The Annals of Statistics
2017, Vol. 45, No. 4, 1431–1461
DOI: 10.1214/16-AOS1495

TESTS FOR SEPARABILITY IN NONPARAMETRIC COVARIANCE OPERATORS OF RANDOM SURFACES

BY JOHN A. D. ASTON[1], DAVIDE PIGOLI AND SHAHIN TAVAKOLI

University of Cambridge

The assumption of separability of the covariance operator for a random image or hypersurface can be of substantial use in applications, especially in situations where the accurate estimation of the full covariance structure is unfeasible, either for computational reasons, or due to a small sample size. However, inferential tools to verify this assumption are somewhat lacking in high-dimensional or functional data analysis settings, where this assumption is most relevant. We propose here to test separability by focusing on K-dimensional projections of the difference between the covariance operator and a nonparametric separable approximation. The subspace we project onto is one generated by the eigenfunctions of the covariance operator estimated under the separability hypothesis, negating the need to ever estimate the full nonseparable covariance. We show that the rescaled difference of the sample covariance operator with its separable approximation is asymptotically Gaussian. As a by-product of this result, we derive asymptotically pivotal tests under Gaussian assumptions, and propose bootstrap methods for approximating the distribution of the test statistics. We probe the finite sample performance through simulations studies, and present an application to log-spectrogram images from a phonetic linguistics dataset.

1. Introduction. Many applications involve hypersurface data, data that is both functional [as in functional data analysis, see, e.g., Ferraty and Vieu (2006), Horváth and Kokoszka (2012), Ramsay and Silverman (2005), Wang, Chiou and Mueller (2016)] and multidimensional. Examples abound and include images from medical devices such as MRI [Lindquist (2008)] or PET [Worsley et al. (1996)], spectrograms derived from audio signals [Rabiner and Schafer (1978), and as in the application we consider in Section 4] or geolocalized data [see, e.g., Secchi, Vantini and Vitelli (2015)]. In these kinds of problem, the number of available observations (hypersurfaces) is often small relative to the high-dimensional nature of the individual observation, and not usually large enough to estimate a full multivariate covariance function.

It is usually, therefore, necessary to make some simplifying assumptions about the data or their covariance structure. If the covariance structure is of interest, such

Received September 2015; revised June 2016.

[1] Supported by EPSRC Grant EP/K021672/2.

MSC2010 subject classifications. Primary 62G10, 62G20; secondary 62F40.

Key words and phrases. Acoustic phonetic data, bootstrap, dimensional reduction, functional data, partial trace, sparsity.

as for PCA or network modeling, for instance, it is usually assumed to have some kind of lower dimensional structure. Traditionally, this translates into a *sparsity* assumption: one assumes that most entries of the covariance matrix or function are zero. Though being relevant for a number of applications [Tibshirani (2014)], this traditional definition of sparsity may not be appropriate in some cases, such as in imaging, as this can give rise to artefacts in the analysis (e.g., holes in an image). In such problems, where the data is multidimensional, a natural assumption that can be made is that the covariance is *separable*. This assumption greatly simplifies both the estimation and the computational cost in dealing with multivariate covariance functions, while still allowing for a positive definite covariance to be specified. In the context of space-time data $X(s,t)$, for instance, where $s \in [-S, S]^d$, $S > 0$, denotes the location in space, and $t \in [0, T]$, $T > 0$, is the time index, the assumption of separability translates into

$$(1.1) \qquad c(s,t,s',t') = c_1(s,s')c_2(t,t'), \qquad s,s' \in [-S,S]^d; t,t' \in [0,T],$$

where c, c_1 and c_2, are respectively the full covariance function, the space covariance function and the time covariance function. In words, this means that the full covariance function factorises as a product of the spatial covariance function with the time covariance function.

The separability assumption [see, e.g., Genton (2007), Gneiting, Genton and Guttorp (2007)] simplifies the covariance structure of the process and makes it far easier to estimate; in some sense, the separability assumption results in a estimator of the covariance which has less variance, at the expense of a possible bias. As an illustrative example, consider that we observe a discretized version of the process through measurements on a two-dimensional grid (without loss of generality, as the same arguments apply for any dimension greater than 2) being a $q \times p$ matrix (of course, the functional data analysis approach taken here does *not* assume that the replications of the process are observed on same grid, nor that they are observed on a grid). Since we are not assuming a parametric form for the covariance, the degrees of freedom in the full covariance are $qp(qp+1)/2$, while the separability assumption reduces them to $q(q+1)/2 + p(p+1)/2$. This reflects a dramatic reduction in the dimension of the problem even for moderate value of q, p and overcomes both computational and estimation problems due to the relatively small sample sizes available in applications. For example, for $q = p = 10$, we have $qp(qp+1)/2 = 5050$ degrees of freedom, however, if the separability holds, then we have only $q(q+1) + p(p+1) = 110$ degrees of freedom. Of course, this is only one example, and our approach is not restricted to data on a grid, but this illustrates the computational savings that such assumptions can possess.

Three related computational classes of problem can be identified. In the first case, the full covariance structure can be computed and stored. In the second one, it is still possible, although burdensome, to compute the full covariance matrix but it cannot be stored, while the last class includes problems where even computation of the full covariance is infeasible. The values of q, p that set the boundaries

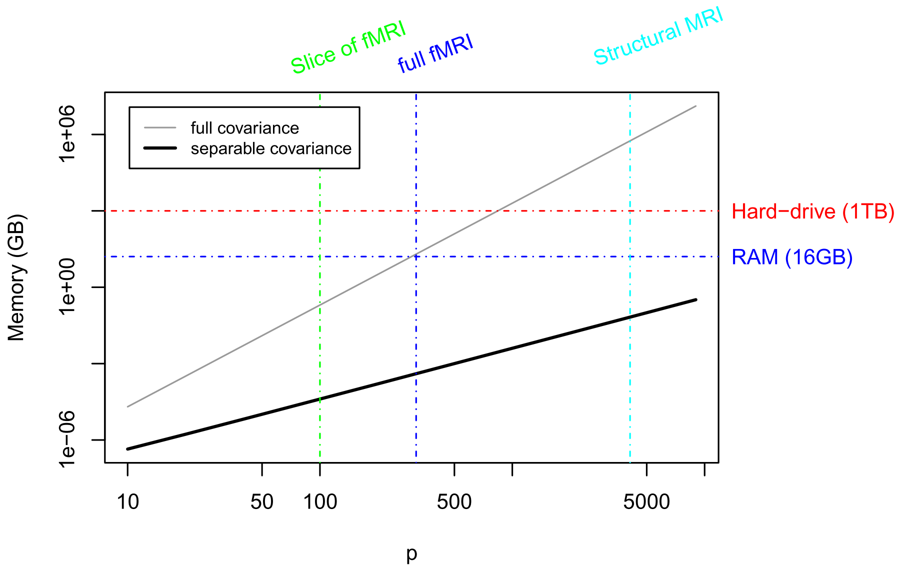


FIG. 1. *Memory required to store the full covariance and the separable covariance of $p \times p$ matrix data, as a function of p. Several types of data related to Neuroimaging (structural and functional Magnetic Resonance Imaging) are used as exemplars of data sizes, as they naturally have multidimensional structure.*

for these classes depend of course on the available hardware and they are rapidly changing. At the present time, however, for widely available systems, storage is feasible up to $q, p \approx 100$ while computation becomes unfeasible when q, p get close to 1000 (see Figure 1). On the contrary, a separable covariance structure can be usually both computed and stored without effort even for these sizes of problem. We would like to stress however that the constraints coming from the need for statistical accuracy are usually tighter. The estimation of the full covariance structure even for $q, p = 100$ presents about 5×10^7 unknown parameters, when typical sample sizes are in the order of hundreds at most. If we are able to assume separability, we can rely on far more accurate estimates.

While the separability assumption can be very useful, and is indeed often implicitly made in many higher dimensional applications when using isotropic smoothing [Lindquist (2008), Worsley et al. (1996)], very little has been done to develop tools to assess its validity on a case by case basis. In the classical multivariate setup, some tests for the separability assumption are available. These have been mainly developed in the field of spatial statistics [see Fuentes (2006), Lu and Zimmerman (2005) and references therein], where the discussion of separable covariance functions is well established, or for applications involving repeated measures [Mitchell, Genton and Gumpertz (2005)]. These methods, however, rely on the estimation of the full multidimensional covariance structure, which can be troublesome. It is sometimes possible to circumvent this problem by considering a parametric model for the full covariance structure [Liu (2014), Simpson (2010), Simpson et al. (2014)]. On the contrary, when the covariance is being nonparametrically specified, as will be the case in this paper, estimation of the full covariance is at best computationally complex with large estimation errors, and in many cases simply computationally infeasible. Indeed, we highlight that, while the focus of this paper is on checking the viability of a separable structure for the covariance, this is done without any parametric assumption on the form of $c_1(s, s')$ and $c_2(t, t')$, thus allowing for the maximum flexibility. This is opposed to assuming a parametric separable form with only few unknown parameters, which is usually too restrictive in many applications, something that has led to separability being rightly criticised and viewed with suspicion in the spatio-temporal statistics literature [Gneiting (2002), Gneiting, Genton and Guttorp (2007)]. Moreover, the methods we develop here are aimed to applications typical of functional data, where replicates from the underlying random process are available. This is different from the spatio-temporal setting, where usually only one realization of the process is observed. See also Constantinou, Kokoszka and Reimherr (2015) for another approach to test for separability in functional data.

It is important to notice that a separable covariance structure (or equivalently, a separable correlation structure) is not necessarily connected with the original data being separable. Furthermore, sums or differences of separable hypersurfaces are not necessarily separable. On the other hand, the error structure may be separable even if the mean is not. Given that in many applications of functional data analysis, the estimation of the covariance is the first step in the analysis, we concentrate on covariance separability. Indeed, covariance separability is an extremely useful assumption as it implies separability of the eigenfunctions, allowing computationally efficient estimation of the eigenfunctions (and principal components). Even if separability is misspecified, separable eigenfunctions can still form a basis representation for the data, they simply no longer carry optimal efficiency guarantees in this case [Aston and Kirch (2012)], but can often have near-optimality under the appropriate assumptions [Chen, Delicado and Müller (2016)].

In this paper, we propose a test to verify if the data at hand are in agreement with a separability assumption. Our test does not require the estimation of the full covariance structure, but only the estimation of the separable structure (1.1), thus avoiding both the computational issues and the diminished accuracy involved in the former. To do this, we rely on a strategy from Functional Data Analysis [Ferraty and Vieu (2006), Horváth and Kokoszka (2012), Ramsay, Graves and Hooker (2009), Ramsay and Silverman (2002, 2005)], which consists in projecting the observations onto a carefully chosen low-dimensional subspace. The key fact for the success of our approach is that, under the null hypothesis, it is possible to determine this subspace using only the marginal covariance functions. While the optimal choice for the dimension of this subspace is a nontrivial problem, some insight can be obtained through our extensive simulation studies (Section 4.1). Ultimately, the proposed test checks the separability in the chosen subspace, which will often be the focus of following analyses.

The paper proceeds as follows. In Section 2, we examine the ideas behind separability, propose a separable approximation of a covariance operator, and study the asymptotics of the difference between the sample covariance operator and its separable approximation. This difference will be the building block of the testing procedures introduced in Section 3, and whose distribution we propose to approximate by bootstrap techniques. In Section 4, we investigate by means of simulation studies the finite sample behaviour of our testing procedures and apply our methods to acoustic phonetic data. A conclusion, given in Section 5, summarizes the main contributions of this paper. Proofs are collected in Appendices A, B and C, while implementation details, theoretical background and additional figures can be found in the supplementary material [Aston, Pigoli and Tavakoli (2016)]. All the tests introduced in the paper are available as an R package `covsep` [Tavakoli (2016)], available on CRAN.

For notational simplicity, the proposed method will be described for two dimensional functional data (e.g., random surfaces), hence a four-dimensional covariance structure (i.e., the covariance of a random surface), but the generalization to higher dimensional cases is straightforward. The methodology is developed in general for data that take values in a Hilbert space, but the case of square integrable surfaces—being relevant for the case of acoustic phonetic data—is used throughout the paper as a demonstration. We recall that the proposed approach is not restricted to data observed on a regular grid, although for simplicity of exposition we consider here the case where data are observed densely and a pre-processing smoothing step allows us to consider the smooth surfaces as our observations, as happens, for example, the case of the acoustic phonetic data described in Section 4. If data are observed sparsely, the proposed approach can still be applied but there may be the need to use more appropriate estimators for the marginal covariance functions [see, e.g., Yao, Müller and Wang (2005)] and these need to satisfy the properties described in Section 2.

2. Separable covariances: Definitions, estimators and asymptotic results.

While the general idea of the factorization of a multidimensional covariance structure as the product of lower dimensional covariances is easy to describe, the development of a testing procedure asks for a rigorous mathematical definition and the introduction of some technical results. In this section, we propose a definition of separability for covariance operators, show how it is possible to estimate a separable version of a covariance operator and evaluate the difference between the empirical covariance operator and its separable version. Moreover, we derive some asymptotic results for these estimators. To do this, we first set the problem in the framework of random elements in Hilbert spaces and their covariance operators. The benefit in doing this is twofold. First, our results become applicable in more general settings (e.g., multidimensional functional data, data on multidimensional grids, fixed size rectangular random matrices) and do not depend on a specific choice of smoothness of the data (which is implicitly assumed when modeling the data as, for example, square integrable surfaces). They only rely on the Hilbert space structure of the space in which the data lie. Second, it highlights the importance of the *partial trace* operator in the estimation of the separable covariance structure, and how the properties of the partial trace (Appendix C) play a crucial role in the asymptotic behavior of the proposed test statistics. However, to ease explanation, we use the case of the Hilbert space of square integrable surfaces (which shall be used in our linguistic application, see Section 4) as an illustration of our testing procedure.

2.1. *Notation.* Let us first introduce some definitions and notation about operators in a Hilbert space [see, e.g., Gohberg, Goldberg and Kaashoek (1990), Kadison and Ringrose (1997), Ringrose (1971)]. Let H be a real separable Hilbert space (i.e., a Hilbert space with a countable orthonormal basis), whose inner product and norm are denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. The space of bounded (linear) operators on H is denoted by $\mathcal{S}_\infty(H)$, and its norm is $|||T|||_\infty = \sup_{x \neq 0} \|Tx\|/\|x\|$. The space of Hilbert–Schmidt operators on H is denoted by $\mathcal{S}_2(H)$, and is a Hilbert space with the inner-product $\langle S, T \rangle_{\mathcal{S}_2} = \sum_{i \geq 1} \langle Se_i, Te_i \rangle$ and induced norm $||| \cdot |||_2$, where $(e_i)_{i \geq 1} \subset H$ is an orthonormal basis of H. The space of trace-class operator on H is denoted by $\mathcal{S}_1(H)$, and consists of all compact operators T with finite trace-norm, that is, $|||T|||_1 = \sum_{n \geq 1} s_n(T) < \infty$, where $s_n(T) \geq 0$ denotes the nth singular value of T. For any trace-class operator $T \in \mathcal{S}_1(H)$, we define its trace by $\mathrm{Tr}(T) = \sum_{i \geq 1} \langle Te_i, e_i \rangle$, where $(e_i)_{i \geq 1} \subset H$ is an orthonormal basis, and the sum is independent of the choice of the orthonormal basis.

If H_1, H_2 are real separable Hilbert spaces, we denote by $H = H_1 \otimes H_2$ their tensor product Hilbert space, which is obtained by the completion of all finite sums $\sum_{i,j=1}^N u_i \otimes v_j$, $u_i \in H_1, v_j \in H_2$, under the inner-product $\langle u \otimes v, z \otimes w \rangle = \langle u, z \rangle \langle v, w \rangle$, $u, z \in H_1$, $z, w \in H_2$ [see, e.g., Kadison and Ringrose (1997)]. If $C_1 \in \mathcal{S}_\infty(H_1)$, $C_2 \in \mathcal{S}_\infty(H_2)$, we denote by $C_1 \tilde{\otimes} C_2$ the unique linear operator on $H_1 \otimes H_2$ satisfying

$$(C_1 \tilde{\otimes} C_2)(u \otimes v) = C_1 u \otimes C_2 v, \qquad \text{for all } u \in H_1, v \in H_2. \tag{2.1}$$

It is a bounded operator on H, with $|||C_1 \tilde{\otimes} C_2|||_\infty = |||C_1|||_\infty |||C_2|||_\infty$. Furthermore, if $C_1 \in \mathcal{S}_1(H_1)$ and $C_2 \in \mathcal{S}_1(H_2)$, then $C_1 \tilde{\otimes} C_2 \in \mathcal{S}_1(H_1 \otimes H_2)$ and $|||C_1 \tilde{\otimes} C_2|||_1 = |||C_1|||_1 |||C_2|||_1$. We denote by $\mathrm{Tr}_1 : \mathcal{S}_1(H_1 \otimes H_2) \to \mathcal{S}_1(H_2)$ the *partial trace with respect to* H_1. It is the unique bounded linear operator satisfying $\mathrm{Tr}_1(A \tilde{\otimes} B) = \mathrm{Tr}(A)B$, for all $A \in \mathcal{S}_1(H_1)$, $B \in \mathcal{S}_1(H_2)$. $\mathrm{Tr}_2 : \mathcal{S}_1(H_1 \otimes H_2) \to \mathcal{S}_1(H_1)$ is defined symmetrically (see Appendix C for more details).

If $X \in H$ is a random element with $\mathbb{E}\|X\| < \infty$, then $\mu = \mathbb{E}X \in H$, the mean of X, is well defined. Furthermore, if $\mathbb{E}\|X\|^2 < \infty$, then $C = \mathbb{E}[(X - \mu) \otimes_2 (X -$

$\mu)$] defines the *covariance operator* of X, where $f \otimes_2 g$ is the operator on H defined by $(f \otimes_2 g)h = \langle h, g\rangle f$, for $f, g, h \in H$. The covariance operator C is a trace-class hermitian operator on H, and encodes all the second-order fluctuations of X around its mean.

Using this nomenclature, we are going to deal with random variables belonging to a tensor product Hilbert space. This framework encompasses the situation where X is a random surface, for example, a space–time indexed data, that is, $X = X(s,t), s \in [-S,S]^d, t \in [0,T], \; S, T > 0$, by setting $H = L^2([-S,S]^d \times [0,T], \mathbb{R})$, for instance (notice however that additional smoothness assumptions on X would lead to assume that X belongs to some other Hilbert space). In this case, the covariance operator of the random element $X \in L^2([-S,S]^d \times [0,T], \mathbb{R})$ satisfies

$$Cf(s,t) = \int_{[-S,S]^d} \int_0^T c(s,t,s',t') f(s',t')\, ds'\, dt', \qquad s \in [-S,S]^d, t \in [0,T],$$

$f \in L^2([-S,S]^d \times [0,T], \mathbb{R})$, where $c(s,t,s',t') = \operatorname{cov}[X(s,t), X(s',t')]$ is the *covariance function* of X. The space of square integrable surfaces,

$$L^2([-S,S]^d \times [0,T], \mathbb{R}),$$

is a tensor product Hilbert space because it can be identified with

$$L^2([-S,S]^d, \mathbb{R}) \otimes L^2([0,T], \mathbb{R}).$$

2.2. *Separability.* We recall now that we want to define separability so that the covariance function can be written as $c(s,t,s',t') = c_1(s,s')c_2(t,t')$, for some $c_1 \in L^2([-S,S]^d \times [-S,S]^d, \mathbb{R})$ and $c_2 \in L^2([0,T] \times [0,T], \mathbb{R})$. This can be extended to the covariance operator of a random elements $X \in H = H_1 \otimes H_2$, where H_1, H_2 are arbitrary separable real Hilbert spaces. We call its covariance operator C *separable* if

$$C = C_1 \tilde{\otimes} C_2, \tag{2.2}$$

where C_1, respectively C_2, are trace-class operators on H_1, respectively on H_2, and $C_1 \tilde{\otimes} C_2$ is defined in (2.1). Notice that though the decomposition (2.2) is not unique, since $C_1 \tilde{\otimes} C_2 = (\alpha C_1) \tilde{\otimes} (\alpha^{-1} C_2)$ for any $\alpha \neq 0$, this will not cause any problem at a later stage since we will ultimately be dealing with the product $C_1 \tilde{\otimes} C_2$, which is identifiable.

In practice, neither C nor $C_1 \tilde{\otimes} C_2$ are known. If $X_1, \ldots, X_N \overset{\text{i.i.d.}}{\sim} X$ and (2.2) holds, the sample covariance operator $\widehat{C}_N$ is not necessarily separable in finite samples. However, we can estimate a separable approximation of it by

$$\widehat{C}_{1,N} \tilde{\otimes} \widehat{C}_{2,N}, \tag{2.3}$$

where $\widehat{C}_{1,N} = \operatorname{Tr}_2(\widehat{C}_N)/\sqrt{\operatorname{Tr}(\widehat{C}_N)}$, $\widehat{C}_{2,N} = \operatorname{Tr}_1(\widehat{C}_N)/\sqrt{\operatorname{Tr}(\widehat{C}_N)}$. The intuition behind (2.3) is that

$$\operatorname{Tr}(T)T = \operatorname{Tr}_2(T)\,\widetilde{\otimes}\,\operatorname{Tr}_1(T),$$

for all $T \in \mathcal{S}_1(H_1 \otimes H_2)$ of the form $T = A \,\widetilde{\otimes}\, B$, $A \in \mathcal{S}_1(H_1)$, $B \in \mathcal{S}_1(H_2)$, with $\operatorname{Tr}(T) \neq 0$.

Let us consider again what this means when X is a random element of $L^2([-S,S]^d \times [0,T],\mathbb{R})$—that is, the realization of a space–time process—of which we observe N i.i.d. replications $X_1,\dots,X_N \sim X$. In this case, Proposition C.2 tells us that if the covariance function is continuous, the operators $\widehat{C}_{1,N}$ and $\widehat{C}_{2,N}$ are defined by

$$\widehat{C}_{1,N} f(s) = \int_{[-S,S]^d} \widehat{c}_{1,N}(s,s') f(s)\,ds, \qquad f \in L^2([-S,S]^d,\mathbb{R}),$$
$$\widehat{C}_{2,N} g(t) = \int_0^T \widehat{c}_{2,N}(t,t') g(t)\,dt, \qquad g \in L^2([0,T],\mathbb{R}),$$

where

$$\widehat{c}_{1,N}(s,s') = \frac{\tilde{c}_{1,N}(s,s')}{\sqrt{\int_{[-S,S]^d} \tilde{c}_{1,N}(s,s)\,ds}},$$
$$\widehat{c}_{2,N}(t,t') = \frac{\tilde{c}_{2,N}(t,t')}{\sqrt{\int_0^T \tilde{c}_{2,N}(t,t)\,dt}},$$

and

$$\begin{aligned}
\tilde{c}_{1,N}(s,s') &= \frac{1}{N}\sum_{i=1}^N \int_0^T (X_i(s,t) - \overline{X}(s,t))(X_i(s',t) - \overline{X}(s',t))\,dt \\
&= \int_0^T c_N(s,t,s',t)\,dt, \\
\tilde{c}_{2,N}(t,t') &= \frac{1}{N}\sum_{i=1}^N \int_{[-S,S]^d} (X_i(s,t) - \overline{X}(s,t))(X_i(s,t') - \overline{X}(s,t'))\,ds \\
&= \int_{[-S,S]^d} c_N(s,t,s,t')\,ds, \\
\overline{X}(s,t) &= \frac{1}{N}\sum_{i=1}^N X_i(s,t), \widehat{c}_N(s,t,s',t') \\
&= \frac{1}{N}\sum_{i=1}^N (X_i(s,t) - \overline{X}(s,t))(X_i(s',t') - \overline{X}(s',t')),
\end{aligned}$$

for all $s, s' \in [-S, S]^d, t, t' \in [0, T]$. The assumption of separability here means that the estimated covariance is written as a product of a purely spatial component and a purely temporal component, thus making both modeling and estimation easier in many practical applications.

We stress again that we aim to develop a test statistic that solely relies on the estimation of the separable components C_1 and C_2, and does not require the estimation of the full covariance C. We can expect that under the null hypothesis $H_0 : C = C_1 \tilde{\otimes} C_2$, the difference $D_N = \widehat{C}_N - \widehat{C}_{1,N} \tilde{\otimes} \widehat{C}_{2,N}$ between the sample covariance operator and its separable approximation should take small values. We propose therefore to construct our test statistic by projecting D_N onto the first eigenfunctions of C, since these encode the directions along which X has the most variability. If we denote by $C_1 = \sum_{i \geq 1} \lambda_i u_i \otimes_2 u_i$ and $C_2 = \sum_{j \geq 1} \gamma_j v_j \otimes_2 v_j$ the Mercer decompositions of C_1 and C_2, we have

$$C = C_1 \tilde{\otimes} C_2 = \sum_{i,j \geq 1} \lambda_i \gamma_j (u_i \otimes v_j) \otimes_2 (u_i \otimes v_j),$$

where we have used results from Section 1.1 of the supplementary material [Aston, Pigoli and Tavakoli (2016)]. The eigenfunctions of C are therefore of the form $u_r \otimes v_s$, where $u_r \in H_1$ is the rth eigenfunction of C_1 and $v_s \in H_2$ is the sth eigenfunction of C_2. We define a test statistic based on the projection

$$T_N(r, s) = \sqrt{N} \langle D_N(\hat{u}_r \otimes \hat{v}_s), \hat{u}_r \otimes \hat{v}_s \rangle, \qquad r, s \geq 1 \text{ fixed}, \tag{2.4}$$

where we have replaced the eigenfunctions of C_1 and C_2 by their empirical counterpart, that is, the Mercer decompositions of $\widehat{C}_{1,N}$, respectively $\widehat{C}_{2,N}$, are given by $\widehat{C}_{1,N} = \sum_{i \geq 1} \hat{\lambda}_i \hat{u}_i \otimes \hat{u}_i$, respectively $\widehat{C}_{2,N} = \sum_{j \geq 1} \hat{\gamma}_j \hat{v}_j \otimes \hat{v}_j$. Notice that though the eigenfunctions of $\widehat{C}_{1,N}$ and $\widehat{C}_{2,N}$ are defined up to a multiplicative constant $\alpha = \pm 1$, our test statistic is well defined. The key fact for the practical implementation of the method is that $T_N(r, s)$ can be computed without the need to estimate (and store in memory) the operator D_N, since $T_N(r, s) = \sqrt{N}(\frac{1}{N} \sum_{k=1}^{N} \langle X_k - \overline{X}_N, \widehat{v}_i \otimes \widehat{u}_j \rangle^2 - \hat{\lambda}_r \hat{\gamma}_s)$. In particular, the computation of $T_N(r, s)$ does *not* require an estimation of the full covariance operator C, but only the estimation of the marginal covariance operators C_1 and C_2, and their eigenstructure.

2.3. *Asymptotics.* The theoretical justification for using a projection of D_N to define a test procedure is that, under the null hypothesis $H_0 : C = C_1 \tilde{\otimes} C_2$, we have $|||D_N|||_1 \xrightarrow{p} 0$ as $N \to \infty$, that is, D_N convergences in probability to zero with respect to the trace norm. In fact, we will show in Theorem 2.3 that $\sqrt{N} D_N$ is asymptotically Gaussian under the following regularity conditions.

CONDITION 2.1. *X is a random element of the real separable Hilbert space H satisfying*

$$\sum_{j=1}^{\infty} \left(\mathbb{E}[\langle X, e_j\rangle^4]\right)^{1/4} < \infty, \tag{2.5}$$

for some orthonormal basis $(e_j)_{j\geq 1}$ of H.

The implications of this condition can be better understood in light of the following remark.

REMARK 2.2 [Mas (2006)]. 1. Condition 2.1 implies that $\mathbb{E}\|X\|^4 < \infty$.

2. If $\mathbb{E}\|X\|^4 < \infty$, then $\sqrt{N}(C_N - C)$ converges in distribution to a Gaussian random element of $\mathcal{S}_2(H)$ for $N \to \infty$, with respect to the Hilbert–Schmidt topology. Under Condition 2.1, a stronger form of convergence holds: $\sqrt{N}(C_N - C)$ converges in distribution to a random element of $\mathcal{S}_1(H)$ for $N \to \infty$, with respect to the trace-norm topology.

3. If X is Gaussian and $(\lambda_j)_{j\geq 1}$ is the sequence of eigenvalues of its covariance operator, a sufficient condition for (2.5) is $\sum_{j\geq 1}\sqrt{\lambda_j} < \infty$.

Condition 2.1 requires fourth-order moments rather than the usual second order moments often assumed in functional data, as in this case we are interested in investigating the variation of the second moment, and hence require assumptions on the fourth-order structure. Recall that $\widehat{C}_N = \frac{1}{N}\sum_{j=1}^{N}(X_i - \overline{X}) \otimes_2 (X_i - \overline{X})$, where $\overline{X} = N^{-1}\sum_{k=1}^{N} X_k$. The following result establishes the asymptotic distribution of $D_N = \widehat{C}_N - \frac{\operatorname{Tr}_2(\widehat{C}_N)\tilde{\otimes}\operatorname{Tr}_1(\widehat{C}_N)}{\operatorname{Tr}(\widehat{C}_N)}$.

THEOREM 2.3. *Let H_1, H_2 be separable real Hilbert spaces, $X_1, \ldots, X_N \sim X$ be i.i.d. random elements on $H_1 \otimes H_2$ with covariance operator C, and* $\operatorname{Tr} C \neq 0$.

If X satisfies Condition 2.1 (with $H = H_1 \otimes H_2$), then under the null hypothesis

$$H_0 : C = C_1 \tilde{\otimes} C_2, \qquad C_1 \in \mathcal{S}_1(H_1), C_2 \in \mathcal{S}_1(H_2),$$

we have

$$\sqrt{N}\left(\widehat{C}_N - \frac{\operatorname{Tr}_2(\widehat{C}_N) \tilde{\otimes} \operatorname{Tr}_1(\widehat{C}_N)}{\operatorname{Tr}(\widehat{C}_N)}\right) \xrightarrow{d} Z \qquad \text{as } N \to \infty, \tag{2.6}$$

where Z is a Gaussian random element of $\mathcal{S}_1(H_1 \otimes H_2)$ with mean zero, whose covariance structure is given in Lemma A.1.

Condition 2.1 is used here because we need $\sqrt{N}(\widehat{C}_N - C)$ to converge in distribution in the topology of the space $\mathcal{S}_1(H_1 \otimes H_2)$; it could be replaced by any (weaker) condition ensuring such convergence. The assumption $\operatorname{Tr} C \neq 0$ is equivalent to assuming that X is not almost surely constant.

PROOF OF THEOREM 2.3. First, notice that $C = C_1 \mathbin{\tilde{\otimes}} C_2 = \frac{\operatorname{Tr}_2(C) \tilde{\otimes} \operatorname{Tr}_1(C)}{\operatorname{Tr}(C)}$ under H_0. Therefore, using the linearity of the partial trace, we get

$$
\begin{aligned}
&\sqrt{N}\biggl(\widehat{C}_N - \frac{\operatorname{Tr}_2(\widehat{C}_N) \mathbin{\tilde{\otimes}} \operatorname{Tr}_1(\widehat{C}_N)}{\operatorname{Tr}(\widehat{C}_N)}\biggr) \\
&\quad = \sqrt{N}(\widehat{C}_N - C) \\
&\qquad + \sqrt{N}\biggl(\frac{\operatorname{Tr}_2(C) \mathbin{\tilde{\otimes}} \operatorname{Tr}_1(C)}{\operatorname{Tr}(C)} + \frac{\operatorname{Tr}_2(\widehat{C}_N) \mathbin{\tilde{\otimes}} \operatorname{Tr}_1(\widehat{C}_N)}{\operatorname{Tr}(\widehat{C}_N)}\biggr) \\
&\quad = \sqrt{N}(\widehat{C}_N - C) + \frac{\operatorname{Tr}(\sqrt{N}(\widehat{C}_N - C))C}{\operatorname{Tr}(\widehat{C}_N)} \\
&\qquad - \frac{\operatorname{Tr}_2(\sqrt{N}(\widehat{C}_N - C)) \mathbin{\tilde{\otimes}} \operatorname{Tr}_1(C)}{\operatorname{Tr}(\widehat{C}_N)} \\
&\qquad - \frac{\operatorname{Tr}_2(\widehat{C}_N) \mathbin{\tilde{\otimes}} \operatorname{Tr}_1(\sqrt{N}(\widehat{C}_N - C))}{\operatorname{Tr}(\widehat{C}_N)} \\
&\quad = \Psi\bigl(\sqrt{N}(\widehat{C}_N - C), \widehat{C}_N\bigr),
\end{aligned}
$$

where

$$
\Psi(T, S) = T + \frac{\operatorname{Tr}(T)C}{\operatorname{Tr}(S)} - \frac{\operatorname{Tr}_2(T) \mathbin{\tilde{\otimes}} \operatorname{Tr}_1(C)}{\operatorname{Tr}(S)} - \frac{\operatorname{Tr}_2(S) \mathbin{\tilde{\otimes}} \operatorname{Tr}_1(T)}{\operatorname{Tr}(S)},
$$

$T, S \in \mathcal{S}_1(H_1 \otimes H_2)$. Notice that the function $\Psi : \mathcal{S}_1(H_1 \otimes H_2) \times \mathcal{S}_1(H_1 \otimes H_2) \to \mathcal{S}_1(H_1 \otimes H_2)$ is continuous at $(T, S) \in \mathcal{S}_1(H_1 \otimes H_2) \times \mathcal{S}_1(H_1 \otimes H_2)$ in each coordinate, with respect to the trace norm, provided $\operatorname{Tr}(S) \neq 0$. Since $\sqrt{N}(\widehat{C}_N - C)$ converges in distribution—under Condition 2.1—to a Gaussian random element $Y \in \mathcal{S}_1(H_1 \otimes H_2)$, with respect to the trace norm $|||\cdot|||_1$ [see Mas (2006), Proposition 5], $\Psi(\sqrt{N}(\widehat{C}_N - C), \widehat{C}_N)$ converges in distribution to

$$
\Psi(Y, C) = Y + \frac{\operatorname{Tr}(Y)C}{\operatorname{Tr}(C)} - \frac{\operatorname{Tr}_2(Y) \mathbin{\tilde{\otimes}} \operatorname{Tr}_1(C)}{\operatorname{Tr}(C)} - \frac{\operatorname{Tr}_2(C) \mathbin{\tilde{\otimes}} \operatorname{Tr}_1(Y)}{\operatorname{Tr}(C)} \tag{2.7}
$$

by the continuous mapping theorem in metric spaces [Billingsley (1999)]. $\Psi(Y, C)$ is Gaussian because each of the summands of (2.7) are Gaussian. Indeed, the first and second summands are obviously Gaussian, and the last two summands are Gaussian by Proposition C.3, and Proposition 1.2 in the supplementary material [Aston, Pigoli and Tavakoli (2016)]. □

We can now give the asymptotic distribution of $T_N(r,s)$, defined in (2.4) as the (scaled) projection of D_N in a direction given by the tensor product of the empirical eigenfunctions $\hat{u}_r$ and $\hat{v}_s$. The proof of the following result is given in Appendix B.

COROLLARY 2.4. *Under the conditions of Theorem* 2.3, *if* $\mathcal{I} \subset \{(i,j) : i, j \geq 1\}$ *is a* finite *set of indices such that* $\lambda_r \gamma_s > 0$ *for each* $(r,s) \in \mathcal{I}$, *then*

$$\bigl(T_N(r,s)\bigr)_{(r,s)\in\mathcal{I}} \xrightarrow{d} N(0,\Sigma) \qquad \textit{as } N \to \infty.$$

This means that the vector $(T_N(r,s))_{(r,s)\in\mathcal{I}}$ *is asymptotically multivariate Gaussian*, *with asymptotic variance-covariance matrix* $\Sigma = (\Sigma_{(r,s),(r',s')})_{(r,s),(r',s')\in\mathcal{I}}$ *is given by*

$$\begin{aligned}
\Sigma_{(r,s),(r',s')} = \tilde{\beta}_{rsr's'} &+ \frac{\alpha_{rs}\tilde{\beta}_{r's'\cdot\cdot} + \alpha_{r's}\tilde{\beta}_{r\cdot\cdot s'} + \alpha_{rs'}\tilde{\beta}_{r'\cdot\cdot s} + \alpha_{r's'}\tilde{\beta}_{rs\cdot\cdot}}{\operatorname{Tr}(C)} \\
&+ \frac{\alpha_{rs}\alpha_{r's'}\tilde{\beta}_{\cdot\cdot\cdot\cdot}}{\operatorname{Tr}(C)^2} + \frac{\lambda_r\lambda_{r'}\tilde{\beta}_{\cdot s\cdot s'}}{\operatorname{Tr}(C_1)^2} + \frac{\gamma_s\gamma_{s'}\tilde{\beta}_{r\cdot r'\cdot}}{\operatorname{Tr}(C_2)^2} \\
&- \frac{\lambda_r\tilde{\beta}_{r's'\cdot s} + \lambda_{r'}\tilde{\beta}_{rs\cdot s'}}{\operatorname{Tr}(C_1)} - \frac{\gamma_s\tilde{\beta}_{r's'r\cdot} + \gamma_{s'}\tilde{\beta}_{rsr'\cdot}}{\operatorname{Tr}(C_2)} \\
&- \frac{\alpha_{rs}}{\operatorname{Tr}(C)}\left(\frac{\gamma_{s'}\tilde{\beta}_{r'\cdot\cdot\cdot}}{\operatorname{Tr}(C_2)} + \frac{\lambda_{r'}\tilde{\beta}_{\cdot s'\cdot\cdot}}{\operatorname{Tr}(C_1)}\right) \\
&- \frac{\alpha_{r's'}}{\operatorname{Tr}(C)}\left(\frac{\gamma_s\tilde{\beta}_{r\cdot\cdot\cdot}}{\operatorname{Tr}(C_2)} + \frac{\lambda_r\tilde{\beta}_{\cdot s\cdot\cdot}}{\operatorname{Tr}(C_1)}\right),
\end{aligned}$$

where $\mu = \mathbb{E}[X]$, $\alpha_{rs} = \lambda_r\gamma_s$,

$$\tilde{\beta}_{ijkl} = \mathbb{E}\bigl[\langle X-\mu, u_i \otimes v_j\rangle^2 \langle X-\mu, u_k \otimes v_l\rangle^2\bigr],$$

and "·" denotes summation over the corresponding index, *that is*, $\tilde{\beta}_{r\cdot jk} = \sum_{i\geq 1}\tilde{\beta}_{rijk}$.

We note that the asymptotic variance-covariance of $(T_N(r,s))_{(r,s)\in\mathcal{I}}$ depends on the second- and fourth-order moments of X, which is not surprising since it is based on estimators of the covariance of X. Under the additional assumption that X is Gaussian, the asymptotic variance-covariance of $(T_N(r,s))_{(r,s)\in\mathcal{I}}$ can be entirely expressed in terms of the covariance operator C. The proof of the following result is given in Appendix B.

COROLLARY 2.5. *Assume the conditions of Theorem* 2.3 *hold*, *and that* X *is Gaussian. If* $\mathcal{I} \subset \{(i,j) : i, j \geq 1\}$ *is a* finite *set of indices such that* $\lambda_r\gamma_s > 0$ *for each* $(r,s) \in \mathcal{I}$, *then*

$$\bigl(T_N(r,s)\bigr)_{(r,s)\in\mathcal{I}} \xrightarrow{d} N(0,\Sigma) \qquad \textit{as } N \to \infty,$$

where

$$\Sigma_{(r,s),(r',s')} = \frac{2\lambda_r\lambda_{r'}\gamma_s\gamma_{s'}}{\operatorname{Tr}(C)^2}\big(\delta_{rr'}\operatorname{Tr}(C_1)^2 + |||C_1|||_2^2 - (\lambda_r+\lambda_{r'})\operatorname{Tr}(C_1)\big)$$
$$\times\big(\delta_{ss'}\operatorname{Tr}(C_2)^2 + |||C_2|||_2^2 - (\gamma_s+\gamma_{s'})\operatorname{Tr}(C_2)\big),$$

and $\delta_{ij}=1$ *if* $i=j$, *and zero otherwise. In particular, notice that* Σ *itself is separable.*

It will be seen in the next section that even in the case where we use a bootstrap test, knowledge of the asymptotic distribution can be very useful to establish a pivotal bootstrap test, which will be seen to have very good performance in simulation.

3. Separability tests and bootstrap approximations.

In this section, we use the estimation procedures and the theoretical results presented in Section 2 to develop a test for $H_0: C = C_1 \widetilde{\otimes} C_2$, against the alternative that C cannot be written as a tensor product.

First, it is straightforward to define a testing procedure when X is Gaussian. Indeed, if we let

$$G_N(r,s) = T_N^2(r,s) = N\left(\frac{1}{N}\sum_{k=1}^{N}\langle X_k-\overline{X}, \widehat{u}_r\otimes\widehat{v}_s\rangle^2 - \hat{\lambda}_r\hat{\gamma}_s\right)^2 \tag{3.1}$$

and

$$\hat{\sigma}^2(r,s) = \big(\operatorname{Tr}(\widehat{C}_{1,N})^2\operatorname{Tr}(\widehat{C}_{2,N})^2\big)^{-1}2\hat{\lambda}_r^2\hat{\gamma}_s^2$$
$$\times\big(\operatorname{Tr}(\widehat{C}_{1,N})^2 + |||\widehat{C}_{1,N}|||_2^2 - 2\hat{\lambda}_r\operatorname{Tr}(\widehat{C}_{1,N})\big) \tag{3.2}$$
$$\times\big(\operatorname{Tr}(\widehat{C}_{2,N})^2 + |||\widehat{C}_{2,N}|||_2^2 - 2\hat{\gamma}_s\operatorname{Tr}(\widehat{C}_{2,N})\big),$$

then $\hat{\sigma}^{-2}(r,s)G_N(r,s)$ is asymptotically χ_1^2 distributed, and $\{G_N^2(r,s) > \hat{\sigma}^2(r,s)\chi_1^2(1-\alpha)\}$, where $\chi_1^2(1-\alpha)$ is the $1-\alpha$ quantile of the χ_1^2 distribution, would be a rejection region of level approximately α, for $\alpha\in[0,1]$ and N large.

Apart for the distributional assumption for X to be Gaussian, this approach suffers also the important limitation that it only tests the separability assumption along *one* eigendirection. It is possible to extend this approach to take into account several eigendirections. For simplicity, let us consider the case $\mathcal{I}=\{1,\dots,p\}\times\{1,\dots,q\}$. Denote by $\mathbf{T}_N(\mathcal{I})$ the $p\times q$ matrix with entries $(\mathbf{T}_N(\mathcal{I}))_{ij} = T_N(i,j)$, and let

$$\widetilde{G}_N(\mathcal{I}) = \big|\hat{\Sigma}_{L,\mathcal{I}}^{-1/2}\mathbf{T}_N(\mathcal{I})\hat{\Sigma}_{R,\mathcal{I}}^{-\mathsf{T}/2}\big|^2, \tag{3.3}$$

where $|A|^2$ denotes the sum of squared entries of a matrix A, $A^{-1/2}$ denotes the inverse of (any) square root of the matrix A, $A^{-\mathsf{T}/2} = (A^{-1/2})^{\mathsf{T}}$, and the matrices $\hat{\Sigma}_{L,\mathcal{I}}$, respectively $\hat{\Sigma}_{R,\mathcal{I}}$, which are estimators of the row, respectively column, asymptotic covariances of $\mathbf{T}_N(\mathcal{I})$, are defined in Section 2 of the supplementary material [Aston, Pigoli and Tavakoli (2016)]. Then $\widetilde{G}_N(\mathcal{I})$ is asymptotically χ^2_{pq} distributed. In the simulation studies (Section 4.1), we consider also an approximate version of this Studentized test statistics, $\widetilde{G}^a_N(\mathcal{I}) = \sum_{(r,s)\in\mathcal{I}} T^2_N(r,s)/\hat{\sigma}^2(r,s)$, which are obtained simply by standardizing marginally each entry $T^2_N(r,s)$, thus ignoring the dependence between the test statistics associated with different directions. In order to assess the advantage of Studentization, we also consider the non-Studentized test statistic

$$G_N(\mathcal{I}) = \sum_{(r,s)\in\mathcal{I}} T^2_N(r,s).$$

The computation details for $\widetilde{G}_N$, T_N, $\hat{\sigma}^2(r,s)$, $\hat{\Sigma}_{L,\mathcal{I}}$ and $\hat{\Sigma}_{R,\mathcal{I}}$ are described in Section 2 of the supplementary material [Aston, Pigoli and Tavakoli (2016)].

REMARK 3.1. Notice that the only test whose asymptotic distribution is parameter-free is $\widetilde{G}_N(\mathcal{I})$, under Gaussian assumptions. It would in principle be possible to construct an analogous test without the Gaussian assumptions (using Corollary 2.4). However, due to the large number of parameters that would need to be estimated in this case, we expect the asymptotics to come into force only for very large sample sizes (this is actually the case under Gaussian assumptions, specially if the set of projections $\mathcal{I}$ is large, as can be seen in Figure S5 of the supplementary material [Aston, Pigoli and Tavakoli (2016)]). For these reasons, we shall investigate bootstrap approximations to the test statistics.

The choice of the number of eigenfunctions K (the number of elements in $\mathcal{I}$) onto which one should project is not trivial. The popular choice of including enough eigenfunctions to explain a fixed percentage of the variability in the dataset may seem inappropriate in this context, because under the alternative hypothesis there is no guarantee that the separable eigenfunctions explain that percentage of variation.

For fixed K, notice that the test at least guarantees the separability in the subspace of the respective K eigenfunctions, which is where the following analysis will be often focused. On the other hand, since our test statistic looks at an estimator of the nonseparable component

$$D = C - \frac{\operatorname{Tr}_2(C) \mathbin{\widetilde{\otimes}} \operatorname{Tr}_1(C)}{\operatorname{Tr}(C)},$$

restricted to the subspace spanned by the eigenfunctions $u_r \otimes v_s$, the test takes small values (and thus lacks power) when

$$\langle D(u_r \otimes v_s), u_r \otimes v_s\rangle = \langle D, (u_r \otimes_2 u_r) \mathbin{\widetilde{\otimes}} (v_s \otimes_2 v_s)\rangle_{\mathcal{S}_2} = 0,$$

that is when the nonseparable component D is orthogonal to

$$(u_r \otimes_2 u_r) \mathbin{\widetilde{\otimes}} (v_s \otimes_2 v_s)$$

with respect to the Hilbert–Schmidt inner product. Thus, the proposed test statistic $G_N(\mathcal{I})$ is powerful when D is not orthogonal to the subspace

$$V_{\mathcal{I}} = \operatorname{span}\{(u_i \otimes_2 u_i) \widetilde{\otimes} (v_j \otimes_2 v_j), (i, j) \in \mathcal{I}\},$$

and in general the power of the test for finite sample size depends on the properly rescaled norm of the projection of D onto $V_{\mathcal{I}}$.

In practice, it seems reasonable to use the subset of eigenfunctions that it is possible to estimate accurately given the available sample sizes. The accuracy of the estimates for the eigendirections can be in turn evaluated with bootstrap methods; see, for example, Hall and Hosseini-Nasab (2006) for the case of functional data. A good strategy may also be to consider more than one subset of eigenfunctions and then summarize the response obtained from the different tests using a Bonferroni correction.

As an alternative to these test statistics (based on projections of $D_N = C_N - C_{1,N} \mathbin{\widetilde{\otimes}} C_{2,N}$), we consider also a test based on the squared Hilbert–Schmidt norm of D_N, that is, $\|\!|D_N|\!\|_2^2$, whose null distribution will be approximated by a bootstrap procedure (this test will be referred to as *Hilbert–Schmidt test* hereafter). Though it seems that such tests would require one to store the full sample covariance of the data (which could be infeasible), we describe in Section 2 of the supplementary material [Aston, Pigoli and Tavakoli (2016)] a way of circumventing such problem, although the computation of each entry of the full covariance is still needed. Therefore, this could be used only for applications in which the dimension of the discretized covariance matrix is not too large.

In the following, we propose also a bootstrap approach to approximate the distribution of the test statistics $\widetilde{G}_N(\mathcal{I})$, $\widetilde{G}_N^a(\mathcal{I})$ and $G_N(\mathcal{I})$, with the aim to improve the finite sample properties of the procedure and to relax the distributional assumption on X.

3.1. *Parametric bootstrap.* If we assume we know the distribution of X up to its mean μ and its covariance operator C, that is, $X \sim F(\mu; C)$, we can approximate the distribution of $\widetilde{G}_N(\mathcal{I})$, $\widetilde{G}_N^a(\mathcal{I})$, $G_N(\mathcal{I})$ and $\|\!|D_N|\!\|_2^2$ under the separability hypothesis via a parametric bootstrap procedure. Since $C_{1,N} \mathbin{\widetilde{\otimes}} C_{2,N}$, respectively $\overline{X}$, is an estimate of C, respectively μ, we simulate B bootstrap samples $X_1^b, \ldots, X_N^b \overset{\text{i.i.d.}}{\sim} F(\overline{X}, C_{1,N} \mathbin{\widetilde{\otimes}} C_{2,N})$, for $b = 1, \ldots, B$. For each sample, we compute $H_N^b = H_N(X_1^b, \ldots, X_N^b)$, where $H_N = G_N(\mathcal{I})$, $H_N = \widetilde{G}_N(\mathcal{I})$, $H_N = \widetilde{G}_N^a(\mathcal{I})$, respectively $H_N = \|\!|D_N|\!\|_2^2$, if we wish to use the non-Studentized projection test, the Studentized projection test, the approximated Studentized version or the Hilbert–Schmidt test, respectively. A formal description of the algorithm for obtaining the p-value of the test based on the statistic $H_N = H_N(X_1, \ldots, X_N)$ with the parametric bootstrap can be found in Section 2 of the supplementary material [Aston, Pigoli and Tavakoli (2016)], along with the details for the computation of H_N. We highlight that this procedure does not ask for the estimation of the full covariance structure, but only of its separable approximation, with the exception of the Hilbert–Schmidt test (and even in this case, it is possible to avoid the storage of the full covariance).

3.2. *Empirical bootstrap.* In many applications, it is not possible to assume a distribution for the random element X, and a nonparametric approach is therefore needed. In this setting, we can use the empirical bootstrap to estimate the distribution of the test statistic $G_N(\mathcal{I}), \widetilde{G}_N(\mathcal{I})$ or $|||D_N|||_2^2$ under the null hypothesis $H_0 : C = C_1 \tilde{\otimes} C_2$. Let H_N denote the test statistic whose distribution is of interest. Based on an i.i.d. sample $X_1, \dots, X_N \sim X$, we wish to approximate the distribution of H_N with the distribution of some test statistic $\Delta_N^* = \Delta_N(X_1^*, \dots, X_N^*)$, where $X_1^*, \dots, X_N^*$ is obtained by drawing with replacement from the set $\{X_1, \dots, X_N\}$. Though it is tempting to use $\Delta_N^* = H_N(X_1^*, \dots, X_N^*)$, this is not an appropriate choice. Indeed, let us look at the case $H_N = G_N(i, j)$. Notice that the true covariance of X is

$$C = \frac{\mathrm{Tr}_2(C) \tilde{\otimes} \mathrm{Tr}_1(C)}{\mathrm{Tr}(C)} + D, \tag{3.4}$$

where D is a possibly nonzero operator, and that

$$H_N^* = G_N\big(i, j | X_1^*, \dots, X_N^*\big) = N\big\langle \big(C_N^* - C_{1,N}^* \tilde{\otimes} C_{2,N}^*\big)(\hat{u}_i \otimes \hat{v}_j), \hat{u}_i \otimes \hat{v}_j\big\rangle^2,$$

where $C_N^* = C_N(X_1^*, \dots, X_N^*)$, $C_{1,N}^* = C_{1,N}(X_1^*, \dots, X_N^*)$, and $C_{2,N}^* = C_{2,N}(X_1^*, \dots, X_N^*)$. Since $(C_N^* - C_{1,N}^* \tilde{\otimes} C_{2,N}^*) \approx (C_N - C_{1,N} \tilde{\otimes} C_{2,N}) \approx D$, the statistic H_N^* would approximate the distribution of H_N under the hypothesis (3.4), which is not what we want. We therefore propose the following choices of $\Delta_N^* = \Delta_n(X_1^*, \dots, X_N^*; X_1, \dots, X_N)$, depending on the choice of H_N:

1. $H_N = G_N(\mathcal{I})$, $\Delta_N^* = \sum_{(i,j) \in \mathcal{I}} (T_N^*(i, j) - T_N(i, j))^2$.
2. $H_N = \widetilde{G}_N(\mathcal{I})$, $\Delta_N^* = |(\hat{\Sigma}_{L,\mathcal{I}}^*)^{-1/2} (\mathbf{T}_N^*(\mathcal{I}) - \mathbf{T}_N(\mathcal{I}))(\hat{\Sigma}_{R,\mathcal{I}}^*)^{-\mathsf{T}/2}|^2$, where $\hat{\Sigma}_{L,\mathcal{I}}^* = \hat{\Sigma}_{L,\mathcal{I}}(X_1^*, \dots, X_N^*)$, and $\hat{\Sigma}_{R,\mathcal{I}}^* = \hat{\Sigma}_{R,\mathcal{I}}(X_1^*, \dots, X_N^*)$ are the row, respectively column, covariances estimated from the bootstrap sample.
3. $H_N = \widetilde{G}_N^a(\mathcal{I})$, $\Delta_N^* = \sum_{(i,j) \in \mathcal{I}} (T_N^*(i, j) - T_N(i, j))^2 / \hat{\sigma}_*^2(i, j)$, where $\hat{\sigma}_*^2(i, j) = \hat{\sigma}^2(i, j | X_1^*, \dots, X_N^*)$.
4. $H_N = |||D_N|||_2^2$, $\Delta_N^* = |||D_N^* - D_N|||_2^2$, where $D_N^* = D_N(X_1^*, \dots, X_N^*)$.

The algorithm to approximate the p-value of H_N by the empirical bootstrap is described in detail in the supplementary material [Aston, Pigoli and Tavakoli (2016)]. The basic idea consists of generating B bootstrap samples, computing Δ_N^* for each bootstrap sample and looking at the proportion of bootstrap samples for which Δ_N^* is larger than the test statistic H_N computed from the original sample.

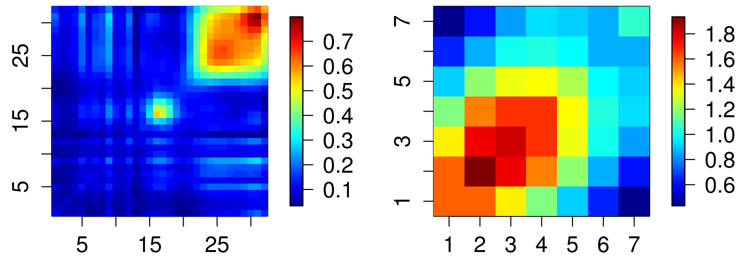


FIG. 2. *Covariance functions c_1 (left) and c_2 (right) used in the simulation study.*

4. Empirical demonstrations of the method.

4.1. *Simulation studies.* We investigated the finite sample behavior of our testing procedures through an intensive reproducible simulation study (its running time is equivalent to approximately 401 days on a single CPU computer). We compared the test based on the asymptotic distribution of (3.1), as well as the tests based on $G_N(\mathcal{I})$, $\widetilde{G}_N(\mathcal{I})$, $\widetilde{G}_N^a(\mathcal{I})$ and $|||D_N|||_2^2$, with the p-values obtained via the parametric bootstrap or the empirical bootstrap.

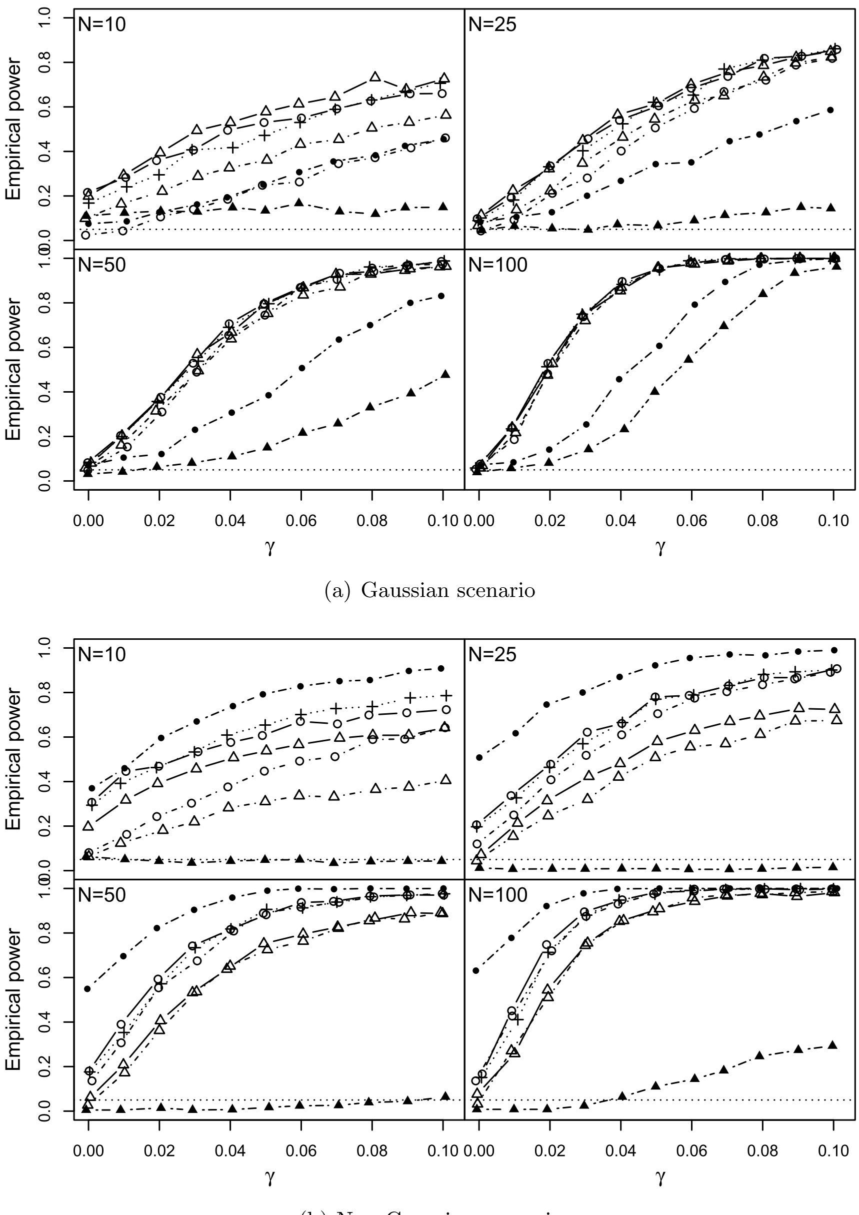
We generated discretized functional data $X_1, \dots, X_N \in \mathbb{R}^{32\times 7}$ under two scenarios. In the first scenario (Gaussian scenario), the data were generated from a multivariate Gaussian distribution $\mathcal{N}(0, \mathbf{C})$. In the second scenario (non-Gaussian scenario), the data were generated from a centered multivariate t distribution with 6 degrees of freedom. In the Gaussian scenario, we set $\mathbf{C} = \mathbf{C}^{(\gamma)}$, where

$$\begin{aligned}\mathbf{C}^{(\gamma)}(i_1, j_1, i_2, j_2) &= (1-\gamma)c_1(i_1, i_2)c_2(j_1, j_2)\\ &\quad + \gamma\frac{1}{(j_1 - j_2)^2 + 1}\exp\left\{-\frac{(i_1 - i_2)^2}{(j_1 - j_2)^2 + 1}\right\},\end{aligned} \tag{4.1}$$

$\gamma \in [0, 1]$; $i_1, i_2 = 1, \dots, 32$; $j_1, j_2 = 1, \dots, 7$. The covariances c_1 and c_2 used in the simulations can be seen in Figure 2. For the non-Gaussian scenario, we chose a multivariate t distribution with the correlation structure implied by $\mathbf{C}^{(\gamma)}$. The parameter $\gamma \in [0, 1]$ controls the departure from the separability of the covariance $\mathbf{C}^{(\gamma)}$: $\gamma = 0$ yields a separable covariance, whereas $\gamma = 1$ yields a complete non-separable covariance structure [Cressie and Huang (1999)]. All the simulations have been performed using the R package `covsep` [Tavakoli (2016)], available on CRAN, which implements the tests presented in the paper.

For each value of $\gamma \in \{0, 0.01, 0.02, \dots, 0.1\}$ and $N \in \{10, 25, 50, 100\}$, we performed 1000 replications for each of the above simulations, and estimated the power of the tests based on the asymptotic distribution of (3.1).

We first also estimated the power of the tests $\widetilde{G}_N(1, 1)$, $G_N(1, 1)$ and $|||D_N|||_2$, with distributions approximated by a Gaussian parametric bootstrap, and the empirical bootstrap, with $B = 1000$. The results are shown in Figure 3. In the Gaussian scenario [Figure 3, panel (a)], the empirical size of all the proposed tests gets



(a) Gaussian scenario

(b) Non-Gaussian scenario

FIG. 3. *Empirical power of the testing procedures in the* Gaussian *scenario* [*panel* (a)] *and* non-Gaussian *scenario* [*panel* (b)], *for* $N = 10, 25, 50, 100$ *and* $\mathcal{I} = \mathcal{I}_1$. *The results shown correspond to the test* (3.1) *based on its asymptotic distribution* ($\cdots + \cdots$), *the Gaussian parametric bootstrap projection tests* $\widetilde{G}_N(\mathcal{I}_1)$(– · –○– · –), *and* $G_N(\mathcal{I}_1)$ (—○—) *the empirical bootstrap projection tests* $\widetilde{G}_N(\mathcal{I}_1)$ (– · –△– · –), *and* $G_N(\mathcal{I}_1)$ (—△—), *the Gaussian parametric Hilbert–Schmidt test* (— · —●— · —) *and the empirical Hilbert–Schmidt test* (— · —▲— · —). *The horizontal dotted line indicates the nominal level* (5%) *of the test. Note that the points have been horizontally jittered for better visibility.*

closer to the nominal level (5%) as N increases (see also Table S1 in the supplementary material [Aston, Pigoli and Tavakoli (2016)]). Nevertheless, the non-Studentized tests $G_N(1, 1)$, for both parametric and empirical bootstrap, seem to have a slower convergence with respect to the Studentized version, and even for $N = 100$ the level of these tests appear still higher than the nominal one (and a CLT-based 95% confidence interval for the true level does not contain the nominal level in both cases). The empirical bootstrap version of the Hilbert–Schmidt test also fails to respect the nominal level at $N = 100$, but its parametric bootstrap counterpart respects the level, even for $N = 25$. For $N = 25, 50, 100$, the most powerful tests (amongst those who respect the nominal level) are the parametric and empirical bootstrap versions of $\widetilde{G}_N(1, 1)$, and they seem to have equal power. The power of the Hilbert–Schmidt test based on the parametric bootstrap seems to be competitive only for $N = 100$ and $\gamma = 0.1$, and is much lower for other values of the parameters. The test based on the asymptotic distribution does not respect the nominal level for small N but it does when N increases. Indeed, the convergence to the nominal level seems remarkably fast and its power is comparable with those of the parametric and empirical bootstrap tests based on $\widetilde{G}_N(1, 1)$. Despite being based on an asymptotic result, its performance is quite good also in finite samples, and it is less computationally demanding than the bootstrap tests.

In the non-Gaussian scenario [Figure 3, panel (b)], only the empirical bootstrap version of $\widetilde{G}_N(1, 1)$ and of the Hilbert–Schmidt test seem to respect the level for $N = 10$ (see also Table S1 in the supplementary material [Aston, Pigoli and Tavakoli (2016)]). Amongst these tests, the most powerful one is clearly the empirical bootstrap test based on $\widetilde{G}_N(1, 1)$. Although the Gaussian parametric bootstrap test has higher empirical power, it does not have the correct level (as expected), and thus cannot be used in a non-Gaussian scenario. Notice also that the test based on the asymptotic distribution of $\widetilde{G}_N(1, 1)$ (under Gaussian assumptions) does not respects the level of the test even for $N = 100$. The same holds for the Gaussian bootstrap version of the Hilbert–Schmidt test. Finally, though the empirical bootstrap version of the Hilbert–Schmidt test respects the level for $N = 10, 25, 50, 100$, it has virtually no power for $N = 10, 25, 50$, and has very low power for $N = 100$ (at most 0.3 for $\gamma = 0.1$).

As mentioned previously, there is no guarantee that a violation in the separability of C is mostly reflected in the first separable eigensubspace. Therefore, we consider also a larger subspace for the test. Figure S4 in the supplementary material [Aston, Pigoli and Tavakoli (2016)] shows the empirical power for the asymptotic test, the parametric and empirical bootstrap tests based on the test statistic $\widetilde{G}_N(\mathcal{I}_2)$, as well as parametric and bootstrap tests based on the test statistics $G_N(\mathcal{I}_2)$, $\widetilde{G}^a_N(\mathcal{I}_2)$ where $\mathcal{I}_2 = \{(i, j) : i, j = 1, 2\}$. In the Gaussian scenario, the asymptotic test is much slower in converging to the correct level compared to its univariate version based on $\widetilde{G}_N(1, 1)$. For larger N, its power is comparable to that of the parametric and empirical bootstrap based on the Studentized test statistics

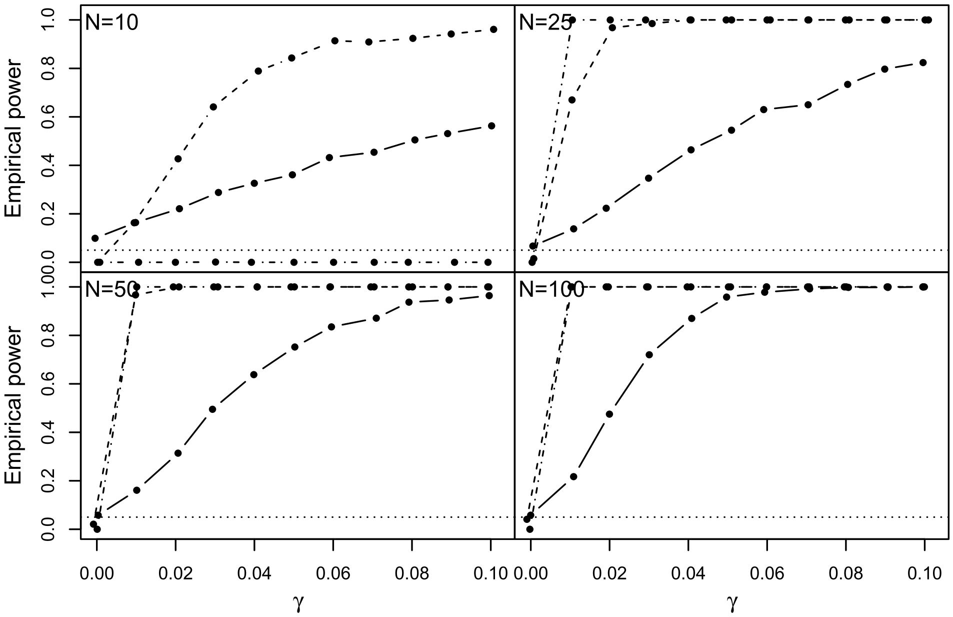


FIG. 4. *Empirical power of the empirical bootstrap version of* $\widetilde{G}_N(\mathcal{I}_l)$, *for* $l = 1$ (—●—), $l = 2$ (– –●– –) *and* $l = 3$ (– · –●– · –), *in the* Gaussian *scenario. The horizontal dotted line indicates the nominal level* (5%) *of the test. Note that the points have been horizontally jittered for better visibility.*

$\widetilde{G}_N(\mathcal{I}_2)$, which in addition respects the nominal level, even for $N = 10$. It is interesting to note that the approximated Studentized bootstrap tests $\widetilde{G}^a_N(\mathcal{I}_2)$ have a performance which is better than the non-Studentized bootstrap tests $G_N(\mathcal{I}_2)$ but far worse than that of the Studentized tests $\widetilde{G}_N(\mathcal{I}_2)$. The Hilbert–Schmidt test is again outperformed by all the other tests, with the exception of the non-Studentized bootstrap test when $N = 10, 25$. The results are similar for the non-Gaussian scenario, apart for the fact that the asymptotic test does not respect the nominal level (as expected, since it asks for X to be Gaussian).

To investigate the difference between projecting on one or several eigensubspaces, we also compare the power of the empirical bootstrap version of the tests $\widetilde{G}_N(\mathcal{I})$ for increasing projection subspaces, that is, for $\mathcal{I} = \mathcal{I}_l, l = 1, 2, 3$, where $\mathcal{I}_1 = \{(1, 1)\}, \mathcal{I}_2 = \{(i, j) : i, j = 1, 2\}$ and $\mathcal{I}_3 = \{(i, j) : i = 1, \dots, 4; j = 1, \dots, 10\}$. The results are shown in Figure 4 for the Gaussian scenario and Figure S1 in the supplementary material [Aston, Pigoli and Tavakoli (2016)] for the non-Gaussian scenario. In the Gaussian scenario, for $N = 10$, the most powerful test is $\widetilde{G}_N(\mathcal{I}_2)$. In this case, projecting onto a larger eigensubspace decreases the power of the test dramatically. However, for $N \geq 25$ the power of the test is the largest for $\widetilde{G}_N(\mathcal{I}_3)$, albeit only significantly larger than that of $\widetilde{G}_N(\mathcal{I}_2)$ when $\gamma = 0.01$. Our interpretation is that when the sample size is too small, including too many eigendirection is bound to add only noise that degrades the performance

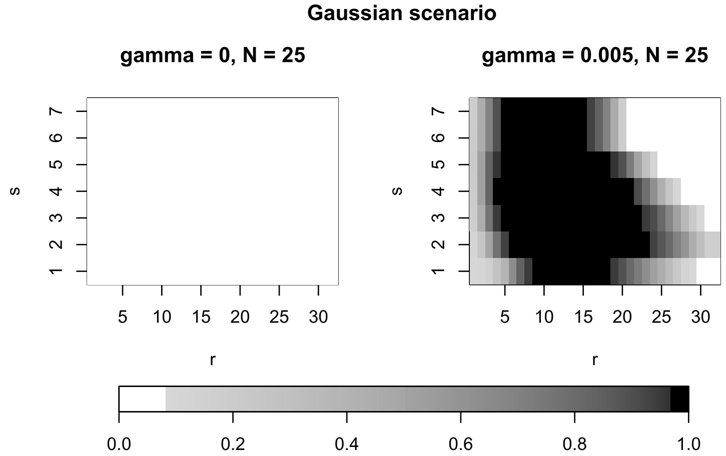


FIG. 5. *Empirical size (left) and power (right) of the separability test as functions of the projection set $\mathcal{I}$. The test used is $\widetilde{G}_N(\mathcal{I})$, with distribution approximated by the empirical bootstrap with* $B = 1000$. *The left plot, respectively the right plot, was simulated from the Gaussian scenario with* $\gamma = 0$, *respectively* $\gamma = 0.005$, *and* $N = 25$. *Each* (r, s) *rectangle represents the level/power of the test based on the projection set* $\mathcal{I} = \{(i, j) : 1 \leq i \leq r, 1 \leq j \leq s\}$.

of the test. However, as long as the separable eigenfunctions are estimated accurately, projecting in a larger eigenspace improves the performance of test. See also Figure S5 in the supplementary material [Aston, Pigoli and Tavakoli (2016)] for the complete simulation results of the projection set $\mathcal{I}_3$.

This prompts us to investigate how the power of the test varies across all projection subsets

$$\mathcal{I}_{r,s} = \{(i, j) : 1 \leq i \leq r, 1 \leq j \leq s\},$$

$r = 1, \dots, 32, s = 1, \dots, 7$. The test used is $\widetilde{G}_N(\mathcal{I})$, with distribution approximated by the empirical bootstrap with $B = 1000$. Figure 5 shows the empirical size and power of the separability test in the Gaussian scenario for sample size $N = 25$, and Figure S2, respectively Figure S3, of the supplementary material [Aston, Pigoli and Tavakoli (2016)] shows the power for different sample sizes in the Gaussian scenario, respectively the non-Gaussian scenario.

4.1.1. *Discussion of simulation studies.* The simulation studies above illustrate how the empirical bootstrap test based on the test statistics $\widetilde{G}_N(\mathcal{I})$ usually outperforms its competitors, albeit it is also much more computationally expensive than the asymptotic test, whose performance are comparable in the Gaussian scenario for large enough number of observations.

The choice of the best set of eigendirections to use in the definition of the test statistics is difficult. It seems that K should be ideally chosen to be increasing with N. This is reasonable, because larger values of N increase the accuracy of the estimation of the eigenfunctions and, therefore, we will be able to detect departures from the separability in more eigendirections, including ones not only associated with the largest eigenvalues. However, the optimal rate at which K should increase with N is still an open problem, and will certainly depend in a complex way on the eigenstructure of the true underlying covariance operator C.

This is confirmed by the results reported in Figure 5 and Figures S2 and S3 of the supplementary material [Aston, Pigoli and Tavakoli (2016)]. These indeed show that taking into account too few eigendirections can result in smaller power, while including too many of them can also decrease the power.

As an alternative to tests based on projections of D_N, the tests based on the squared Hilbert–Schmidt norm of D_N, that is, $|||D_N|||_2^2$, could potentially detect any departure from the separability hypothesis—as opposed to the tests $\widetilde{G}_N(\mathcal{I})$. But as the simulation study illustrates, they might be far less powerful in practice, particularly in situations where the departure from separability is reflected in only in a few eigendirections. Moreover, this approach still requires the computation of the full covariance operator (although not its storage) and is therefore not feasible for all applications.

4.2. *Application to acoustic phonetic data.* An interesting case where the proposed methods can be useful are phonetic spectrograms. These data arise in the analysis of speech records, since relevant features of recorded sounds can be better explored in a two-dimensional time-frequency domain.

In particular, we consider here the dataset of 23 speakers from five different Romance languages that has been first described in Pigoli et al. (2014). The speakers were recorded while pronouncing the words corresponding to the numbers from one to ten in their language and the recordings are converted to a sampling rate of 16,000 samples per second. Since not all these words are available for all the speakers, we have a total of 219 speech records. We focus on the spectrum that speakers produce in each speech recording $x_{ik}^L(t)$, where L is the language, $i = 1, \dots, 10$ the pronounced word and $k = 1, \dots, n_{Li}$ the speaker, n_{Li} being the number of speakers available for language L and word i. We then use a short-time Fourier transform to obtain a two-dimensional log-spectrogram: we use a Gaussian window function $w(\cdot)$ with a window size of 10 milliseconds and we compute the short-time Fourier transform as

$$X_{ik}^L(\omega, t) = \int_{-\infty}^{+\infty} x_{ik}^L(\tau) w(\tau - t) e^{-j\omega\tau}\, d\tau.$$

The spectrogram is defined as the magnitude of the Fourier transform and the log-spectrogram (in decibel) is therefore

$$\mathfrak{S}_{ik}^L(\omega, t) = 10 \log_{10}\bigl(\bigl|X_{ik}^L(\omega, t)\bigr|^2\bigr).$$

The raw log-spectrograms $\mathfrak{S}_{ik}^L$ are then smoothed [with the robust spline smoothing method proposed in Garcia (2010)] and aligned in time using an adaptation to

2-D of the procedure in Tang and Müller (2008), the resulting log-spectrogram is denoted S_{ik}^L. The alignment is needed because a phase distortion can be present in acoustic signals, due to difference in speech velocity between speakers. Since the different words of each language have different mean log-spectrograms, the focus of the linguistic analysis—which is the study cross-linguistics changes—is on the residual log-spectrograms

$$R_{ik}^L(\omega, t) = S_{ik}^L(\omega, t) - \frac{1}{n_{Li}} \sum_{k=1}^{n_{Li}} S_{ik}^L(\omega, t).$$

Assuming that all the words within the language have the same covariance structure, we disregard hereafter the information about the pronounced words that generated the residual log-spectrogram, and use the surface data $R_j^L(\omega, t)$, $j = 1, \ldots, N_L$, that is, the set of observations for the language L including all speakers and words, for the separability test. These observations are measured on an equispaced grid with 81 points in the frequency direction and 100 points in the time direction. This translate on a full covariance structure with about 33×10^6 degrees of freedom. Thus, although the discretized covariance matrix is in principle computable, its storage is a problem. More importantly, the accuracy of its estimate is poor, since we have at most 50 observations within each language. For these reasons, we would like to investigate if a separable approximation of each covariance is appropriate.

We thus apply the Studentized version of the empirical bootstrap test for separability to the residual log-spectrograms for each language individually. Here, we take into consideration different choices for set of eigendirections to be used in the definition of the test statistic $\widetilde{G}_N(\mathcal{I})$, namely $\mathcal{I} = \mathcal{I}_1 = \{(1, 1)\}$, $\mathcal{I} = \mathcal{I}_2 = \{(r, s) : 1 \le r \le 2, 1 \le s \le 3\}$, $\mathcal{I} = \mathcal{I}_3 = \{(r, s) : 1 \le r \le 8, 1 \le s \le 10\}$. For all cases, we use $B = 1000$ bootstrap replicates.

The resulting p-values for each language and for each set of indices can be found in Table 1. Taking into account the multiple testings with a Bonferroni correction, we can conclude that the separability assumption does not appear to hold. We can also see that the departure from separability is caught mainly on the first

TABLE 1

P-values for the test for the separability of the covariance operators of the residual log-spectrograms of the five Romance languages, using the Studentized version of the empirical bootstrap

$\mathcal{I}$	**French**	**Italian**	**Portuguese**	**American Spanish**	**Iberian Spanish**
$\mathcal{I}_1$	0.65	<0.001	<0.001	<0.001	<0.001
$\mathcal{I}_2$	0.078	0.197	0.022	0.36	0.013
$\mathcal{I}_3$	0.001	0.002	0.001	0.001	<0.001

component for the two Spanish varieties. In conclusion, a separable covariance structure is not a good fit for these languages, and thus, when practitioners use this approximation for computational or modeling reasons, they should bear in mind that relevant aspects of the covariance structure may be missed in the analysis.

5. Discussion and conclusions. We presented tests to verify the separability assumption for the covariance operators of random surfaces (or hypersurfaces) through hypothesis testing. These tests are based on the difference between the sample covariance operator and its separable approximation—which we have shown to be asymptotically Gaussian—projected onto subspaces spanned by the eigenfunctions of the covariance of the data. While the optimal choice for this subspace is still an open problem and may depend on the eigenstructure of the full covariance operator, it is possible to give some advice on how to choose $\mathcal{I}$ in practice:

- In many cases, a dimension reduction based on the separable eigenfunctions is needed also for the follow up analysis. It is then recommended to use the same subspace for the test procedure as well, so that it is clear whether the projection of the covariance structure onto the subspace that will be used for the analysis is separable or not, as shown in Section 3.
- As mentioned in Section 3, it is usually better to focus on the subset of eigenfunctions that it is possible to estimate accurately with the available data. These can be again identified with bootstrap methods such as the one described in Hall and Hosseini-Nasab (2006) or considering the dimension of the sample size. As highlighted by the results of the simulation studies in Figure 5 and in Figures S2 and S3 of the supplementary material [Aston, Pigoli and Tavakoli (2016)], the empirical power of the test starts to decline when eigendirections that cannot be reasonably estimated with the available sample size are included.
- When in doubt, it is also possible to apply the test to more than one subset of eigenfunctions and then summarize the response using a Bonferroni correction. We follow this approach in the data application described in Section 4.2.

Though an asymptotic distribution is available in some cases, we also propose to approximate the distribution of our test statistics using either a parametric bootstrap (in case the distribution of the data is known) or an empirical bootstrap. A simulation study suggests that the Studentized version of the empirical bootstrap test gives the highest power in non-Gaussian settings, and has power comparable to its parametric bootstrap counterpart and to the asymptotic test in the Gaussian setting. We therefore use the Studentized empirical bootstrap for the application to linguistic data, since it is not easy to assess the distribution of the data generating process. The bootstrap test leads to the conclusion that the covariance structure is indeed not separable.

Our present approach implicitly assumed that the functional observations (e.g., the hypersurfaces) were densely observed. Though this approach is not restricted to data observed on a grid, it leaves aside the important class of functional data that are sparsely observed [e.g., Yao, Müller and Wang (2005)]. However, the extension of our methodology to the case of sparsely observed functional data is also possible, as long as the estimator used for the full covariance is consistent and satisfies a central limit theorem. While we have only detailed the methods for 2-dimensional surfaces, the extension to higher-order multidimensional functions (such as 3-dimensional volumetric images from applications such as magnetic resonance imaging) is straightforward.

APPENDIX A: THE ASYMPTOTIC COVARIANCE STRUCTURE

LEMMA A.1. *The covariance operator of the random operator* Z, *defined in Theorem* 2.3, *is characterized by the following equality*, *in which* $\Gamma = \mathbb{E}[(X \otimes X - C)\widetilde{\otimes}(X \otimes X - C)]$:

$$
\begin{aligned}
&\mathbb{E}\big[\mathrm{Tr}\big[(A_1 \widetilde{\otimes} A_2)Z\big]\mathrm{Tr}\big[(B_1 \widetilde{\otimes} B_2)Z\big]\big] \\
&\quad = \mathrm{Tr}\big[(A\widetilde{\otimes}B)\Gamma\big] + \frac{\mathrm{Tr}[BC]}{\mathrm{Tr}(C)}\mathrm{Tr}\big[(A\widetilde{\otimes}\mathrm{Id}_H)\Gamma\big] \\
&\qquad - \frac{\mathrm{Tr}[B_2C_2]}{\mathrm{Tr}[C_2]}\mathrm{Tr}\big[\big(A\widetilde{\otimes}(B_1 \widetilde{\otimes} \mathrm{Id}_{H_2})\big)\Gamma\big] \\
&\qquad - \frac{\mathrm{Tr}[B_1C_1]}{\mathrm{Tr}[C_1]}\mathrm{Tr}\big[\big(A\widetilde{\otimes}(\mathrm{Id}_{H_1} \widetilde{\otimes} B_2)\big)\Gamma\big] \\
&\qquad + \frac{\mathrm{Tr}[AC]}{\mathrm{Tr}[C]}\bigg\{\mathrm{Tr}\big[(\mathrm{Id}_H\widetilde{\otimes}B)\Gamma\big] + \frac{\mathrm{Tr}[BC]}{\mathrm{Tr}[C]}\mathrm{Tr}[\Gamma] \\
&\qquad - \frac{\mathrm{Tr}[B_2C_2]}{\mathrm{Tr}[C_2]}\mathrm{Tr}\big[\big(\mathrm{Id}_H\widetilde{\otimes}(B_1 \widetilde{\otimes} \mathrm{Id}_{H_2})\big)\Gamma\big] \\
&\qquad - \frac{\mathrm{Tr}[B_1C_1]}{\mathrm{Tr}[C_1]}\mathrm{Tr}\big[\big(\mathrm{Id}_H\widetilde{\otimes}(\mathrm{Id}_{H_1} \widetilde{\otimes} B_2)\big)\Gamma\big]\bigg\} \\
&\qquad - \frac{\mathrm{Tr}[A_2C_2]}{\mathrm{Tr}[C_2]}\bigg\{\mathrm{Tr}\big[\big((A_1 \widetilde{\otimes} \mathrm{Id}_{H_2})\widetilde{\otimes}B\big)\Gamma\big] \\
&\qquad + \frac{\mathrm{Tr}[BC]}{\mathrm{Tr}[C]}\mathrm{Tr}\big[\big((A_1\widetilde{\otimes}\mathrm{Id}_{H_2})\widetilde{\otimes}\mathrm{Id}_H\big)\Gamma\big] \\
&\qquad - \frac{\mathrm{Tr}[B_2C_2]}{\mathrm{Tr}[C_2]}\mathrm{Tr}\big[\big((A_1\widetilde{\otimes}\mathrm{Id}_{H_2})\widetilde{\otimes}(B_1 \widetilde{\otimes} \mathrm{Id}_{H_2})\big)\Gamma\big] \\
&\qquad - \frac{\mathrm{Tr}[B_1C_1]}{\mathrm{Tr}[C_1]}\mathrm{Tr}\big[\big((A_1\widetilde{\otimes}\mathrm{Id}_{H_2})\widetilde{\otimes}(\mathrm{Id}_{H_1} \widetilde{\otimes} B_2)\big)\Gamma\big]\bigg\} \\
&\qquad - \frac{\mathrm{Tr}[A_1C_1]}{\mathrm{Tr}[C_1]}\bigg\{\mathrm{Tr}\big[\big((\mathrm{Id}_{H_1} \widetilde{\otimes} A_2)\widetilde{\otimes}B\big)\Gamma\big]
\end{aligned}
\tag{A.1}
$$

$$
\begin{aligned}
&+\frac{\mathrm{Tr}[BC]}{\mathrm{Tr}[C]}\mathrm{Tr}\big[\big((\mathrm{Id}_{H_1} \mathbin{\widetilde{\otimes}} A_2)\widetilde{\otimes}\mathrm{Id}_H\big)\Gamma\big] \\
&-\frac{\mathrm{Tr}[B_2C_2]}{\mathrm{Tr}[C_2]}\mathrm{Tr}\big[\big((\mathrm{Id}_{H_1} \mathbin{\widetilde{\otimes}} A_2)\widetilde{\otimes}(B_1 \mathbin{\widetilde{\otimes}} \mathrm{Id}_{H_2})\big)\Gamma\big] \\
&-\frac{\mathrm{Tr}[B_1C_1]}{\mathrm{Tr}[C_1]}\mathrm{Tr}\big[\big((\mathrm{Id}_{H_1} \mathbin{\widetilde{\otimes}} A_2)\widetilde{\otimes}(\mathrm{Id}_{H_1} \mathbin{\widetilde{\otimes}} B_2)\big)\Gamma\big]\Big\},
\end{aligned}
$$

where $A_1, B_1 \in \mathcal{S}_\infty(H_1)$, $A_2, B_2 \in \mathcal{S}_\infty(H_2)$, *and* $A = A_1 \mathbin{\widetilde{\otimes}} A_2$, $B = B_1 \mathbin{\widetilde{\otimes}} B_2$, $H = H_1 \otimes H_2$, *and* Id_H *denotes the identity operator on the Hilbert space* H.

PROOF. By the linearity of the expectation and the trace, and by the properties of the partial trace, the computation of (A.1) boils down to the computation of expressions of the form

$$
\mathbb{E}\big[\mathrm{Tr}\big[(A_1' \mathbin{\widetilde{\otimes}} A_2')Y\big]\mathrm{Tr}\big[(B_1' \mathbin{\widetilde{\otimes}} B_2')Y\big]\big],
$$

for general $A_1', B_1' \in \mathcal{S}_\infty(H_1)$, $A_2', B_2' \in \mathcal{S}_\infty(H_2)$. Since $\mathbb{E}|||Y|||_1^2 < \infty$, we have

$$
\begin{aligned}
&\mathbb{E}\big(\mathrm{Tr}\big[(A_1' \mathbin{\widetilde{\otimes}} A_2')Y\big]\mathrm{Tr}\big[(B_1' \mathbin{\widetilde{\otimes}} B_2')Y\big]\big) \\
&\quad = \mathrm{Tr}\big[\big((A_1'\widetilde{\otimes}A_2')\widetilde{\otimes}(B_1' \mathbin{\widetilde{\otimes}} B_2')\big)\mathbb{E}(Y\widetilde{\otimes}Y)\big] \\
&\quad = \mathrm{Tr}\big[\big((A_1'\widetilde{\otimes}A_2')\widetilde{\otimes}(B_1' \mathbin{\widetilde{\otimes}} B_2')\big)\Gamma\big],
\end{aligned}
$$

where $\Gamma = \mathbb{E}[(X \otimes X - C)\widetilde{\otimes}(X \otimes X - C)]$. The computation of (A.1) follows directly. □

APPENDIX B: PROOFS

PROOF OF COROLLARY 2.4. To alleviate the notation, we shall assume without loss of generality that $\mu = \mathbb{E}X = 0$. Using the properties of the tensor product (see Section 1.1 of the supplementary material [Aston, Pigoli and Tavakoli (2016)]), we get that $T_N(r,s) = \mathrm{Tr}[(\hat{A}_r \mathbin{\widetilde{\otimes}} \hat{B}_s)\sqrt{N}D_N]$, where $\hat{A}_r = (\hat{u}_r \otimes_2 \hat{u}_r)$, $\hat{B}_s = (\hat{v}_s \otimes_2 \hat{v}_s)$. Now notice that though $A_r = u_r \otimes_2 u_r$ and $B_s = v_s \otimes_2 v_s$ are not estimable separately (since C_1 and C_2 are not identifiable), their $\widetilde{\otimes}$-product is identifiable, and is consistently estimated by $\hat{A}_r \mathbin{\widetilde{\otimes}} \hat{B}_s$ (in Trace norm). Slutsky's lemma, Theorem 2.3 and the continuous mapping theorem imply therefore that $(T_N(r,s))_{(r,s)\in\mathcal{I}}$ has the same asymptotic distribution of $(\widetilde{T}_N(r,s))_{(r,s)\in\mathcal{I}}$, where $\widetilde{T}_N(r,s) = \mathrm{Tr}[(A_r \mathbin{\widetilde{\otimes}} B_s)\sqrt{N}D_N]$. This implies that

$$
\big(T_N(r,s)\big)_{(r,s)\in\mathcal{I}} \xrightarrow{d} Z' = \big(\mathrm{Tr}\big[(A_r \mathbin{\widetilde{\otimes}} B_s)Z\big]\big)_{(r,s)\in\mathcal{I}} \qquad \text{as } N \to \infty,
$$

where Z is a mean zero Gaussian random element of $\mathcal{S}_1(H_1 \otimes H_2)$ whose covariance structure is given by Lemma A.1. Z' is therefore also Gaussian random element, with mean zero and covariances

$$
\Sigma_{(r,s),(r',s')} = \mathrm{cov}\big(Z'_{(r,s)}, Z'_{(r',s')}\big) = \mathbb{E}\big[\mathrm{Tr}\big[(A_r\widetilde{\otimes}B_s)Z\big]\mathrm{Tr}\big[(A_{r'} \mathbin{\widetilde{\otimes}} B_{s'})Z\big]\big].
$$

Using Lemma A.1, we see that the computation of $\Sigma_{(r,s),(r',s')}$ depends on the terms $\mathrm{Tr}[(A_r \widetilde{\otimes} B_s)C] = \lambda_r \gamma_s$, $\mathrm{Tr}[A_r C_1] = \lambda_r$, $\mathrm{Tr}[B_s C_2] = \gamma_s$, as well as on the value of

$$\mathrm{Tr}\big[\big((A'_1 \widetilde{\otimes} B'_1)\widetilde{\otimes}(A'_2 \widetilde{\otimes} B'_2)\big)\Gamma\big]$$

for general $A'_1, A'_2 \in \mathcal{S}_\infty(H_1)$, $B'_1, B'_2 \in \mathcal{S}_\infty(H_2)$. Using the Karhunen–Loève expansion $X = \sum_{i,i' \geq 1} \xi_{ii'} u_i \otimes v_{i'}$, where $\xi_{ii'} = \langle X, u_i \otimes v_{i'} \rangle$, we get

$$\begin{aligned}\Gamma &= \mathbb{E}\big((X \otimes_2 X - C)\widetilde{\otimes}(X \otimes_2 X - C)\big) \\ &= \sum_{i,i',j,j',k,k',l,l' \geq 1} \beta_{ii'jj'kk'll'}(u_{ij} \widetilde{\otimes} v_{i'j'})\widetilde{\otimes}(u_{kl} \widetilde{\otimes} v_{k'l'}) \\ &\quad - \sum_{i,i',j,j'} \alpha_{ii'}\alpha_{jj'}(u_{ii} \widetilde{\otimes} v_{i'i'})\widetilde{\otimes}(u_{jj} \widetilde{\otimes} v_{j'j'}),\end{aligned}$$

where we have written $u_{ij} = u_i \otimes_2 u_j \in \mathcal{S}_1(H_1)$, $v_{ij} = v_i \otimes_2 v_j \in \mathcal{S}_1(H_2)$, $\beta_{ii'jj'kk'll'} = \mathbb{E}[\xi_{ii'}\xi_{jj'}\xi_{kk'}\xi_{ll'}]$, $\alpha_{ij} = \lambda_i \gamma_j$ and used the identity $u_{ij} \widetilde{\otimes} v_{i'j'} = (u_i \otimes v_{i'}) \otimes_2 (u_j \otimes v_{j'})$. Therefore,

$$\begin{aligned}&\mathrm{Tr}\big[\big((A'_1 \widetilde{\otimes} A'_2)\widetilde{\otimes}(B'_1 \widetilde{\otimes} B'_2)\big)\Gamma\big] \\ &\quad = \sum_{i,i',j,j',k,k',l,l' \geq 1} \beta_{ii'jj'kk'll'} \mathrm{Tr}\big[A'_1 u_{ij}\big] \mathrm{Tr}\big[A'_2 v_{i'j'}\big] \mathrm{Tr}\big[B'_1 u_{kl}\big] \mathrm{Tr}\big[B'_2 v_{k'l'}\big] \\ &\qquad - \sum_{i,i',j,j'} \alpha_{ii'}\alpha_{jj'} \mathrm{Tr}\big[A'_1 u_{ii}\big] \mathrm{Tr}\big[B'_1 u_{jj}\big] \mathrm{Tr}\big[A'_2 v_{i'i'}\big] \mathrm{Tr}\big[B'_2 v_{j'j'}\big],\end{aligned}$$

and the computation of the variance $\Sigma_{(r,s),(r',s')}$ follows from a straightforward (though tedious) calculation. □

PROOF OF COROLLARY 2.5. We only need to compute and substitute the values of the fourth-order moments terms $\tilde{\beta}_{ijkl}$ in the expression given by Corollary 2.4. Since $\tilde{\beta}_{ijkl} = \mathbb{E}[\xi_{ij}^2 \xi_{kl}^2] = 3\alpha_{kl}^2$ if $(i, j) = (k, l)$, and $\tilde{\beta}_{ijkl} = \alpha_{ij}\alpha_{kl}$ if $(i, j) \neq (k, l)$, straightforward calculations give

$$\begin{aligned}\tilde{\beta}_{rs\cdot\cdot} &= 2\alpha_{rs}^2 + \alpha_{rs} \mathrm{Tr}(C) = \tilde{\beta}_{r\cdot\cdot s}, \\ \tilde{\beta}_{\cdot\cdot\cdot\cdot} &= \mathrm{Tr}(C)^2 + 2|||C|||_2^2, \\ \tilde{\beta}_{\cdot s\cdot s'} &= \gamma_s\gamma_{s'}\big(\mathrm{Tr}(C_1)^2 + 2\delta_{ss'}|||C_1|||_2^2\big), \\ \tilde{\beta}_{r\cdot r'\cdot} &= \lambda_r\lambda_{r'}\big(\mathrm{Tr}(C_2)^2 + 2\delta_{rr'}|||C_2|||_2^2\big), \\ \tilde{\beta}_{rs\cdot s'} &= 2\delta_{ss'}\alpha_{rs}^2 + \alpha_{rs}\gamma_{s'} \mathrm{Tr}(C_1), \\ \tilde{\beta}_{rsr'\cdot} &= 2\delta_{rr'}\alpha_{rs}^2 + \alpha_{rs}\lambda_{r'} \mathrm{Tr}(C_2), \\ \tilde{\beta}_{\cdot\cdot\cdot s} &= 2\gamma_s^2|||C_1|||_2^2 + \gamma_s \mathrm{Tr}(C_1)^2 \mathrm{Tr}(C_2), \\ \tilde{\beta}_{r\cdot\cdot\cdot} &= 2\lambda_r^2|||C_2|||_2^2 + \lambda_r \mathrm{Tr}(C_1) \mathrm{Tr}(C_2)^2,\end{aligned}$$

where $\delta_{ij} = 1$ if $i = j$, and zero otherwise. The proof is completed by direct calculations. □

APPENDIX C: PARTIAL TRACES

Letting $\mathcal{S}_1(H_1 \otimes H_2)$ denote the space of trace-class operators on $H_1 \otimes H_2$, we define the partial trace with respect to H_1 as the unique linear operator $\mathrm{Tr}_1 : \mathcal{S}_1(H_1 \otimes H_2) \to \mathcal{S}_1(H_2)$ satisfying $\mathrm{Tr}_1(A \tilde{\otimes} B) = \mathrm{Tr}(A)B$ for all $A \in \mathcal{S}_1(H_1)$, $B \in \mathcal{S}_1(H_2)$.

PROPOSITION C.1. *The operator* Tr$_1$ *is well defined, linear, continuous and satisfies*

$$\| \mathrm{Tr}_1(A) \|_1 \leq \|A\|_1, \qquad A \in \mathcal{S}_1(H_1 \otimes H_2). \tag{C.1}$$

Furthermore, we have the following characterization of the partial trace. If $T \in \mathcal{S}_1(H_1 \otimes H_2)$,

$$\mathrm{Tr}\big(S \,\mathrm{Tr}_1(T)\big) = \mathrm{Tr}\big((\mathrm{Id}_1 \tilde{\otimes} S)T\big) \qquad \textit{for all } S \in \mathcal{S}_\infty(H_2), \tag{C.2}$$

where Id$_1$ *is the identity operator on* H_1.

This result is proved in Section 4 in the supplementary material [Aston, Pigoli and Tavakoli (2016)]. We can also define $\mathrm{Tr}_2 : \mathcal{S}_1(H_1 \otimes H_2) \to \mathcal{S}_1(H_1)$ analogously. The following result gives an explicit formula for the partial traces of integral operators with continuous kernels.

PROPOSITION C.2. *Let $D_s \subset \mathbb{R}^p, D_t \subset \mathbb{R}^q$ be compact subsets, $H_1 = L^2(D_s, \mathbb{R}), H_2 = L^2(D_t, \mathbb{R})$, and $H = L^2(D_s \times D_t, \mathbb{R}) = H_1 \otimes H_2$. If $C \in \mathcal{S}_1(L^2(D_s \times D_t, \mathbb{R}))$ is a positive definite operator with symmetric continuous kernel $c = c(s, t, s', t')$, that is, $c(s, t, s', t') = c(s', t', s, t)$ for all $s, s' \in D_s, t, t' \in D_t$, and*

$$Cf(s,t) = \iint_{D_s \times D_t} c(s, t, s', t') f(s', t') \, ds' \, dt', \qquad f \in L^2(D_s \times D_t, \mathbb{R}),$$

then Tr$_1(C)$ *is the integral operator on $L^2(D_t, \mathbb{R})$ with kernel $k(t, t') = \int_{D_s} c(s, t, s, t') \, ds$. The analogous result also holds for* Tr$_2(C)$.

This result is proved in Section 4 in the supplementary material [Aston, Pigoli and Tavakoli (2016)]. The next result states that the partial trace of a Gaussian random trace-class operator is also Gaussian.

PROPOSITION C.3. *Let $Z \in \mathcal{S}_1(H_1 \otimes H_2)$ be a Gaussian random element. Then* Tr$_1(Z) \in \mathcal{S}_1(H_2)$ *is a Gaussian random element.*

PROOF. The proof is completed by noticing that $A \in \mathcal{S}_\infty(H_2)$, we have $\mathrm{Tr}(A \,\mathrm{Tr}_1(Z)) = \mathrm{Tr}((\mathrm{Id} \tilde{\otimes} A)Z)$, where the right-hand side is obviously Gaussian. □

Acknowledgements. We wish to thank the Editor, Associate Editor and the referees for their comments that have led to an improved version of the paper. We also wish to thank Victor Panaretos for interesting discussions.

SUPPLEMENTARY MATERIAL

Supplementary material: "Tests for separability in nonparametric covariance operators of random surfaces" (DOI: 10.1214/16-AOS1495SUPPA; .pdf). Background technical results, implementation details, additional simulation studies and additional proofs.

Research data supporting "Tests for separability in nonparametric covariance operators of random surfaces" (DOI: 10.1214/16-AOS1495SUPPB; .zip). R package implementing the methodology of the paper. Data and script for reproducing the numerical simulations and figures of the paper. Data generated for the paper and used in the phonetic application, and script to reproduce it.

REFERENCES

ASTON, J. A. D. and KIRCH, C. (2012). Evaluating stationarity via change-point alternatives with applications to fMRI data. *Ann. Appl. Stat.* **6** 1906–1948. MR3058688

ASTON, J. A. D., PIGOLI, D. and TAVAKOLI, S. (2017). Supplement to "Tests for separability in nonparametric covariance operators of random surfaces." DOI:10.1214/16-AOS1495SUPPA, DOI:10.1214/16-AOS1495SUPPB.

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

CHEN, K., DELICADO, P. and MÜLLER, H.-G. (2016). Modeling function-valued stochastic processes, with applications to fertility dynamics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* To appear. doi:10.1111/rssb.12160.

CONSTANTINOU, P., KOKOSZKA, P. and REIMHERR, M. (2015). Testing separability of space–time functional processes. Available at arXiv:1509.07017.

CRESSIE, N. and HUANG, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.* **94** 1330–1340. MR1731494

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis*: *Theory and Practice*. Springer, New York. MR2229687

FUENTES, M. (2006). Testing for separability of spatial-temporal covariance functions. *J. Statist. Plann. Inference* **136** 447–466. MR2211349

GARCIA, D. (2010). Robust smoothing of gridded data in one and higher dimensions with missing values. *Comput. Statist. Data Anal.* **54** 1167–1178. MR2580947

GENTON, M. G. (2007). Separable approximations of space–time covariance matrices. *Environmetrics* **18** 681–695. MR2408938

GNEITING, T. (2002). Nonseparable, stationary covariance functions for space–time data. *J. Amer. Statist. Assoc.* **97** 590–600. MR1941475

GNEITING, T., GENTON, M. G. and GUTTORP, P. (2007). Geostatistical space–time models, stationarity, separability, and full symmetry. *Monogr. Statist. Appl. Probab.* **107** 151.

GOHBERG, I., GOLDBERG, S. and KAASHOEK, M. A. (1990). *Classes of Linear Operators*. *Vol. I*. *Operator Theory*: *Advances and Applications* **49**. Birkhäuser, Basel. MR1130394

HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 109–126. MR2212577

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer, New York. MR2920735

KADISON, R. V. and RINGROSE, J. R. (1997). *Fundamentals of the Theory of Operator Algebras*: *Elementary Theory. Vol. I. Graduate Studies in Mathematics* **15**. Amer. Math. Soc., Providence, RI. MR1468229

LINDQUIST, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. MR2530545

LIU, C., RAY, S. and HOOKER, G. (2014). Functional principal components analysis of spatially correlated data. Available at arXiv:1411.4681.

LU, N. and ZIMMERMAN, D. L. (2005). The likelihood ratio test for a separable covariance matrix. *Statist. Probab. Lett.* **73** 449–457. MR2187860

MAS, A. (2006). A sufficient condition for the CLT in the space of nuclear operators—Application to covariance of random functions. *Statist. Probab. Lett.* **76** 1503–1509. MR2245571

MITCHELL, M. W., GENTON, M. G. and GUMPERTZ, M. L. (2005). Testing for separability of space–time covariances. *Environmetrics* **16** 819–831. MR2216653

PIGOLI, D., ASTON, J. A. D., DRYDEN, I. L. and SECCHI, P. (2014). Distances and inference for covariance operators. *Biometrika* **101** 409–422. MR3215356

RABINER, L. R. and SCHAFER, R. W. (1978). *Digital Processing of Speech Signals* **100**. Prentice-hall, Englewood Cliffs.

RAMSAY, J. O., GRAVES, S. and HOOKER, G. (2009). *Functional Data Analysis with R and MATLAB*. Springer, New York.

RAMSAY, J. O. and SILVERMAN, B. W. (2002). *Applied Functional Data Analysis*: *Methods and Case Studies*. Springer, New York. MR1910407

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

RINGROSE, J. R. (1971). *Compact Non-self-Adjoint Operators*. Van Nostrand Reinhold, London. MR3075382

SECCHI, P., VANTINI, S. and VITELLI, V. (2015). Analysis of spatio-temporal mobile phone data: A case study in the metropolitan area of Milan. *Stat. Methods Appl.* **24** 279–300. MR3376864

SIMPSON, S. L. (2010). An adjusted likelihood ratio test for separability in unbalanced multivariate repeated measures data. *Stat. Methodol.* **7** 511–519. MR2719948

SIMPSON, S. L., EDWARDS, L. J., STYNER, M. A. and MULLER, K. E. (2014). Separability tests for high-dimensional, low-sample size multivariate repeated measures data. *J. Appl. Stat.* **41** 2450–2461. MR3256397

TANG, R. and MÜLLER, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika* **95** 875–889. MR2461217

TAVAKOLI, S. (2016). covsep: Tests for determining if the covariance structure of 2-dimensional data is separable. R package version 1.0.0. Available at https://CRAN.R-project.org/package=covsep.

TIBSHIRANI, R. J. (2014). Past, present, and future of statistical science. In *Praise of Sparsity and Convexity* (X. Lin, C. Genest, D. L. Banks, G. Molenberghs, D. W. Scott and J.-L. Wang, eds.) 497–506. Chapman & Hall, London.

WANG, J.-L., CHIOU, J.-M. and MUELLER, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application* **3** 257–295.

WORSLEY, K. J., MARRETT, S., NEELIN, P., VANDAL, A. C., FRISTON, K. J. and EVANS, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping* **4** 58–73.

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

STATISTICAL LABORATORY
DEPARTMENT OF PURE MATHS
AND MATHEMATICAL STATISTICS
UNIVERSITY OF CAMBRIDGE
CAMBRIDGE
UNITED KINGDOM
E-MAIL: j.aston@statslab.cam.ac.uk
dp497@cam.ac.uk
s.tavakoli@statslab.cam.ac.uk