SFB 823

A test for separability in covariance operators of random surfaces

Pramita Bagchi, Holger Dette

Nr. 19/2017

Discussion Paper

SFB 823

A test for separability in covariance operators of random surfaces

Pramita Bagchi, Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany

October 21, 2017

Abstract

The assumption of separability is a simplifying and very popular assumption in the analysis of spatio-temporal or hypersurface data structures. It is often made in situations where the covariance structure cannot be easily estimated, for example because of a small sample size or because of computational storage problems. In this paper we propose a new and very simple test to validate this assumption. Our approach is based on a measure of separability which is zero in the case of separability and positive otherwise. The measure can be estimated without calculating the full non-separable covariance operator. We prove asymptotic normality of the corresponding statistic with a limiting variance, which can easily be estimated from the available data. As a consequence quantiles of the standard normal distribution can be used to obtain critical values and the new test of separability is very easy to implement. In particular, our approach does neither require projections on subspaces generated by the eigenfunctions of the covariance operator, nor resampling procedures to obtain critical values nor distributional assumptions as recently used by Aston et al. (2017) and Constantinou et al. (2017) to construct tests for separability. We investigate the finite sample performance by means of a simulation study and also provide a comparison with the currently available methodology. Finally, the new procedure is illustrated analyzing wind speed and temperature data.

Keywords: functional data, minimum distance, separability, space-time processes, surface data structures
AMS Subject classification: 62G10, 62G20

1 Introduction

Data, which is functional **and** multidimensional is usually called surface data and arises in areas such as medical imaging [see Skup (2010); Worsley et al. (1996)], spectrograms derived from audio signals or geolocalized data [see Bar-Hen et al. (2008); Rabiner and Schafer (1978)]. In many of these ultra high-dimensional problems a completely non-parametric estimation of the covariance operator is not possible as the number of available observations is small compared to the dimension. A common approach to obtain reasonable estimates in this context are structural assumptions on the covariance of the underlying process, and in recent years the assumption of separability has become very popular, for example in the analysis of geostatistical space-time models [see Genton (2007); Gneiting et al. (2007), among others]. Roughly speaking, this assumption allows to write the covariance

$$c(s,t,s',t') = \mathbb{E}[X(s,t)X(s',t')]$$

of a (real valued) space-time process $\{X(s,t)\}_{(s,t)\in S\times T}$ as a product of the space and time covariance function, that is

$$c(s,t,s',t') = c_1(s,s')c_2(t,t'). \tag{1.1}$$

It was pointed out by many authors that the assumption of separability yields a substantial simplification of the estimation problem and thus reduces computational costs in the estimation of the covariance in high dimensional problems [see for example Huizenga et al. (2002); Rougier (2017)]. Despite of its importance, there exist only a few tools to validate the assumption of separability for surface data.

Many authors developed tests for spatio-temporal data. For example, Fuentes (2006) proposed a test based on the spectral representation, and Lu and Zimmerman (2005); Mitchell et al. (2005, 2006) investigated likelihood ratio tests under the assumption of a normal distribution. Recently, Constantinou et al. (2017) derived the joint distribution of the three statistics appearing in the likelihood ratio test and used this result to derive the asymptotic distribution of the (log) likelihood ratio. These authors also proposed alternative tests which are based on distances between an estimator of the covariance under the assumption of separability and an estimator which does not use this assumption.

Aston et al. (2017) considered the problem of testing for separability in the context of hypersurface data. These authors pointed out that many available methods require the estimation of the full multidimensional covariance structure, which can become infeasible for high dimensional data. In order to address this issue they developed a bootstrap test for applications, where replicates from the underlying random process are available. To avoid estimation and storage of the full covariance finite-dimensional projections of the difference between the covariance operator and a nonparametric separable approximation are proposed. In particular they suggest to project onto subspaces generated by the eigenfunctions of the covariance operator estimated under the assumption of separability. However, as pointed in the same references the choice of the number of eigenfunctions onto which one should project is not trivial and the test might be sensitive with respect to this choice.

In this paper we present an alternative and very simple test for the hypothesis of separability in hypersurface data. We consider a similar setup as in Aston et al. (2017) and proceed in two steps. First we derive an *explicit* expression for the minimal distance between the covariance operator and its approximation by a separable covariance operator, where the minimum is taken with respect to the second factor of the tensor product. It turns out that this minimum vanishes if and only if the covariance operator is separable. Secondly, we directly estimate the minimal distance (and not the covariance operator itself) from the available data. As a consequence the calculation of the test statistic does neither use an estimate of the full non-separable covariance operator nor requires the specification of subspaces used for a projection. In the main result of this paper we derive the asymptotic distribution of the test statistic, which is normal (after appropriate standardization) under the null hypothesis **and** alternative. The limiting variance under the null hypothesis can easily be estimated and as consequence we obtain a very simple test for the hypothesis of separability, which only requires the quantiles of the normal distribution. Moreover, in contrast to the work of Aston et al. (2017), the test proposed here does not require resampling and - from a theoretical perspective - the limiting theorems are valid under more general and easier to verify moment assumptions.

In Section 2 we review some basic terminology and minimize the distance between the covariance operator and separable covariance operators with respect to the second factor of the tensor product. This minimum distance could also be interpreted as a measure of deviation from separability (it is zero in the case of separability and positive otherwise). In Section 3 we propose an estimator of the minimum distance and prove asymptotic normality of a standardised version of this statistic. We also provide a simple estimate of the limiting variance under the null hypothesis and prove its consistency. Section 4 is devoted to an investigation of the finite sample properties of the new test and a comparison with two alternative tests for this problem, which have recently been proposed by Aston et al. (2017) and Constantinou et al. (2017). In particular we demonstrate that - despite of its simplicity - the new procedure has very competitive properties compared to the currently available methodology. Finally, some technical details are deferred to the Appendix A.

2 Hilbert spaces and a measure of separability

We begin introducing some basic facts about Hilbert spaces, Hilbert-Schmidt operators and tensor products. For more details we refer to the monographs of Weidmann (1980), Dunford and Schwartz (1988) or Gohberg et al. (1990). Let H be a real separable Hilbert space with inner-product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. The space of bounded linear operators on H is denoted by $S_\infty(H)$ with operator norm

$$|||T|||_\infty := \sup_{\|f\| \leq 1} \|Tf\|.$$

A bounded linear operator T is said to be compact if it can be written as

$$T = \sum_{j \geq 1} s_j(T) \langle e_j, \cdot \rangle f_j,$$

where $\{e_j : j \geq 1\}$ and $\{f_j : j \geq 1\}$ are orthonormal sets of H, $\{s_j(T) : j \geq 1\}$ are the singular values of T and the series converges in the operator norm. We say that a compact operator T belongs to the Schatten class of order $p \geq 1$ and write $T \in S_p(H)$ if

$$|||T|||_p = \Big(\sum_{j \geq 1} s_j(A) \Big)^{1/p} < \infty.$$

The Schatten class of order $p \geq 1$ is a Banach space with norm $|||.|||_p$ and with the property that $S_p(H) \subset S_q(H)$ for $p < q$. In particular we are interested in Schatten classes of order $p = 1$ and 2. A compact operator T is called Hilbert-Schmidt operator if $T \in S_2(H)$ and trace class if $T \in S_1(H)$. The space of Hilbert-Schmidt operators $S_2(H)$ is also a Hilbert space with the Hilbert-Schmidt inner product given by

$$\langle A, B \rangle_{HS} = \sum_{j \geq 1} \langle A e_j, B e_j \rangle$$

for each $A, B \in S_2(H)$, where $\{e_j : j \geq 1\}$ is an orthonormal basis (the inner product does not depend on the choice of the basis).

For two real separable Hilbert spaces H_1 and H_2, the tensor product of H_1 and H_2, denoted as $H := H_1 \otimes H_2$, is the Hilbert space 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 $\langle u \otimes v, w \otimes z \rangle = \langle u, w \rangle \langle v, z \rangle$, for $u, w \in H_1$ and $v, z \in H_2$. For $C_1 \in S_\infty(H_1)$ and $C_2 \in S_\infty(H_2)$, the tensor product $C_1 \tilde{\otimes} C_2$ is defined as the unique linear operator on $H_1 \otimes H_2$ satisfying

$$(C_1 \tilde{\otimes} C_2)(u \otimes v) = C_1 u \otimes C_2 v, \ \ u \in H_1, v \in H_2.$$

In fact $C_1\tilde{\otimes}C_2 \in S_\infty(H)$ with $|||C_1\tilde{\otimes}C_2|||_\infty = |||C_1|||_\infty|||C_2|||_\infty$. Moreover, if $C_1 \in S_p(H_1)$ and $C_2 \in S_p(H_2)$, for $p \geq 1$, then $C_1\tilde{\otimes}C_2 \in S_p(H)$ with $|||C_1\tilde{\otimes}C_2|||_p = |||C_1|||_p|||C_2|||_p$. For more details we refer to Chapter 8 of Weidmann (1980). In the sequel we denote the Hilbert-Schmidt inner product on $S_2(H)$ with $H = H_1 \otimes H_2$ as $\langle\cdot,\cdot\rangle_{HS}$ and that of $S_2(H_1)$ and $S_2(H_2)$ as $\langle\cdot,\cdot\rangle_{S_2(H_1)}$ and $\langle\cdot,\cdot\rangle_{S_2(H_2)}$ respectively.

2.1 Measuring separability

We consider random elements X in the Hilbert space H with $\mathbb{E}\|X\|^4 < \infty$. (See Chapter 7 from Hsing and Eubank (2015)) Then the covariance operator of X is defined as $C := \mathbb{E}[(X - \mathbb{E}X) \otimes_o (X - \mathbb{E}X)]$, where for $f, g \in H$ the operator $f \otimes_o g : H \to H$ is defined by

$$(f \otimes_o g)h = \langle h, g\rangle f \text{ for all } h \in H.$$

Under the assumption $\mathbb{E}\|X\|^4 < \infty$ we have $C \in S_2(H)$. We also assume $|||C|||_2 \neq 0$, which essentially means the random variable X is not degenerate. To test separability we consider the hypothesis

$$H_0 : C = C_1 \otimes C_2 \text{ for some } C_1 \in S_2(H_1) \text{ and } C_2 \in S_2(H_2). \tag{2.1}$$

Our approach is based on a approximation of the operator C by a separable operator $C_1\tilde{\otimes}C_2$ with respect to the norm $|||\cdot|||_2$. Ideally, we are looking for

$$D := \inf\left\{ |||C - C_1\tilde{\otimes}C_2|||_2^2 \;\middle|\; C_1 \in S_2(H_1), C_2 \in S_2(H_2)\right\}, \tag{2.2}$$

such that the hypothesis of separability in (2.1) can be rewritten in terms of the distance D, that is

$$H_0 : D = 0 \;\text{ versus }\; H_1 : D > 0\,. \tag{2.3}$$

However, it turns out that it is difficult to express D explicitly in terms of the covariance operator C. For this reason we proceed in a slightly different way in two steps. First we fix C_1 and determine

$$D(C_1) := \inf\left\{ |||C - C_1\tilde{\otimes}C_2|||_2^2 \;\middle|\; C_2 \in S_2(H_2)\right\}, \tag{2.4}$$

that is we are minimizing $|||C - C_1\tilde{\otimes}C_2|||_2^2$ with respect to second factor C_2 of the tensor product. In particular we will show that the infimum is in fact a minimum and derive an explicit expression for $D(C_1)$ and its minimizer. Next we shows that the resulting minimum, $D(C_1)$ vanishes if and only if the hypothesis of separability holds.

For this purpose we have to introduce additional notation and have to prove several auxiliary results. The main statement is given in Theorem 2.1 (whose formulation also requires the new notation). First, consider the bounded linear operator $T_1 : S_2(H) \times S_2(H_1) \mapsto S_2(H_2)$ defined by

$$T_1(A \tilde{\otimes} B, C_1) = \langle A, C_1 \rangle_{S_2(H_1)} B \tag{2.5}$$

for all $C_1 \in S_2(H_1)$. Similarly, let $T_2 : S_2(H) \times S_2(H_2) \to S_2(H_1)$ be the bounded linear operator defined by

$$T_2(A \tilde{\otimes} B, C_2) = \langle B, C_2 \rangle_{S_2(H_2)} A \tag{2.6}$$

for all $C_2 \in S_2(H_2)$.

Proposition 2.1. *The operators T_1 and T_2 are well-defined, bi-linear and continuous with*

$$\langle B, T_1(C, C_1) \rangle_{S_2(H_2)} = \langle C, C_1 \tilde{\otimes} B \rangle_{HS}, \tag{2.7}$$

$$\langle A, T_2(C, C_2) \rangle_{S_2(H_1)} = \langle C, A \tilde{\otimes} C_2 \rangle_{HS}. \tag{2.8}$$

for all $A, C_1 \in S_2(H_1)$, $B, C_2 \in S_2(H_2)$ and $C \in S_2(H)$.

Proof. By Lemma A.1 in Appendix A the space

$$\mathscr{D}_0 := \Big\{ \sum_{i=1}^{n} A_i \tilde{\otimes} B_i : A_i \in S_2(H_1), B_i \in S_2(H_2), n \in \mathbb{N} \Big\} \tag{2.9}$$

is a dense subset of $S_2(H_1 \otimes H_2)$ (note that a similar result for the space $S_1(H_1 \otimes H_2)$ has been established in Lemma 1.6 of the supplementary material in Aston et al. (2017)). For all $B \in S_2(H_2)$, $E \in \mathscr{D}_0$ and $C_1 \in S_2(H_1)$, we have

$$\begin{aligned} \langle B, T_1(E, C_1) \rangle_{S_2(H_2)} &= \Big\langle B, \sum_{i=1}^{n} \langle A_i, C_1 \rangle_{S_2(H_1)} B_i \Big\rangle_{S_2(H_2)} = \sum_{i=1}^{n} \langle A_i, C_1 \rangle_{S_2(H_1)} \langle B, B_i \rangle_{S_2(H_2)} \\ &= \Big\langle \sum_{i=1}^{n} A_i \tilde{\otimes} B_i, C_1 \tilde{\otimes} B \Big\rangle_{HS} = \langle E, C_1 \tilde{\otimes} B \rangle_{HS}. \end{aligned} \tag{2.10}$$

Using the fact that

$$||| T_1(E, C_1) |||_2 \le ||| T_1(E, C_1) |||_1 = \sup \big\{ \langle B, T_1(E, C_1) \rangle_{S_2(H)} : B \in S_2(H_2), |||B|||_\infty \le 1 \big\}, \tag{2.11}$$

(2.10) and the Cauchy Schwarz inequality it follows that

$$||| T_1(E, C_1) |||_2 \le |||C_1|||_2 |||E|||_2. \tag{2.12}$$

Therefore, for each $C_1 \in S_2(H_1)$, we can extend $T_1(., C_1)$ continuously on $S_2(H)$.

Furthermore as (2.7) holds for all $C \in \mathscr{D}_0$ and the maps $C \mapsto \langle B, T_1(C, C_1) \rangle_{S_2(H_1)}$ and $C \mapsto \langle C, C_1 \tilde{\otimes} B \rangle_{HS}$ are continuous for all $B \in S_2(H_2)$ and $C_1 \in S_2(H_1)$, we can conclude that (2.7) holds for all $C \in S_2(H)$. □

Corollary 2.1. *For all* $C \in S_2(H)$, $C_1 \in S_2(H_1)$ *and* $C_2 \in S_2(H_2)$ *we have*

$$|||T_1(C, C_1)|||_2 \le |||C|||_2 |||C_1|||_2 \ \text{ and } \ |||T_2(C, C_2)|||_2 \le |||C|||_2 |||C_2|||_2.$$

Proposition 2.2. *For any* $C \in S_2(H)$ *and* $C_1 \in S_2(H_1)$, *we have*

$$\langle C, C_1 \tilde{\otimes} T_1(C, C_1)\rangle_{HS} = |||T_1(C, C_1)|||_2^2.$$

Proof. Recall the definition of the set $\mathscr{D}_0$ in (2.9) and let $C = \sum_{n=1}^N A_n \tilde{\otimes} B_n \in \mathscr{D}_0$, where $A_n \in S_2(H_1)$, $B_n \in S_2(H_2)$. We write

$$\begin{aligned}
\langle C, C_1 \tilde{\otimes} T_1(C, C_1)\rangle_{HS} &= \left\langle C, C_1 \tilde{\otimes} \sum_{n=1}^N \langle A_n, C_1\rangle_{S_2(H_1)} B_n \right\rangle_{HS} \\
&= \sum_{n=1}^N \langle A_n, C_1\rangle_{S_2(H_1)} \langle C, C_1 \tilde{\otimes} B_n\rangle_{HS} \\
&= \sum_{n=1}^N \sum_{m=1}^N \langle A_n, C_1\rangle_{S_2(H_1)} \langle A_m, C_1\rangle_{S_2(H_1)} \langle B_m, B_n\rangle_{S_2(H_2)}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\langle T_1(C, C_1), T_1(C, C_1)\rangle_{S_2(H_2)} &= \left\langle \sum_{n=1}^N \langle A_n, C_1\rangle_{S_2(H_1)} B_n, \sum_{m=1}^N \langle A_m, C_1\rangle_{S_2(H_1)} B_m \right\rangle_{S_2(H_2)} \\
&= \sum_{n=1}^N \sum_{m=1}^N \langle A_n, C_1\rangle_{S_2(H_1)} \langle A_m, C_1\rangle_{S_2(H_1)} \langle B_m, B_n\rangle_{S_2(H_2)}.
\end{aligned}$$

Therefore, for all $C_1 \in S_2(H_1)$, the functions $f, g : S_2(H) \mapsto \mathbb{R}$ defined by

$$f(C) := \langle C, C_1 \tilde{\otimes} T_1(C, C_1)\rangle_{HS} \ \text{ and } \ g(C) := |||T_1(C, C_1)|||_2^2$$

are continuous and coincide on the dense subset $\mathscr{D}_0$ of $S_2(H)$. So $f(C) = g(C)$ for all $C \in S_2(H)$ and hence the result follows. □

Theorem 2.1. *For each* $C \in S_2(H)$ *and* $C_1 \in S_2(H_1)$ *the distance*

$$D(C_1, C_2) = |||C - C_1 \tilde{\otimes} C_2|||_2 \tag{2.13}$$

is minimized at

$$\widetilde{C}_2 = \frac{T_1(C, C_1)}{|||C_1|||_2^2}. \tag{2.14}$$

Moreover, for $C_1 \in S_2(H_1)$ *the minimum distance in* (2.13) *is given by*

$$D(C_1) = |||C|||_2^2 - \frac{|||T_1(C, C_1)|||_2^2}{|||C_1|||_2^2}. \tag{2.15}$$

In particular $D(C_1)$ *is zero if and only if* $C = C_1 \tilde{\otimes} C_2$ *for some* $C_2 \in S_2(H_2)$.

Proof. We write

$$\begin{aligned}|||C - C_1 \tilde{\otimes} C_2|||_2^2 =& |||C - C_1 \tilde{\otimes} \widetilde{C}_2|||_2^2 + |||C_1 \tilde{\otimes} \widetilde{C}_2 - C_1 \tilde{\otimes} C_2|||_2^2 \\ &+ 2\langle C - C_1 \tilde{\otimes} \widetilde{C}_2, C_1 \tilde{\otimes} \widetilde{C}_2 - C_1 \tilde{\otimes} C_2\rangle_{HS}.\end{aligned}$$

For the last term we obtain from (2.14)

$$\begin{aligned}\langle C - C_1 \tilde{\otimes} \widetilde{C}_2, C_1 \tilde{\otimes} \widetilde{C}_2 - C_1 \tilde{\otimes} C_2\rangle_{HS} =& \langle C, C_1 \tilde{\otimes} \widetilde{C}_2\rangle_{HS} - |||C_1 \tilde{\otimes} \widetilde{C}_2|||_2^2 \\ &- \langle C, C_1 \tilde{\otimes} C_2\rangle_{HS} + \langle C_1 \tilde{\otimes} \widetilde{C}_2, C_1 \tilde{\otimes} C_2\rangle_{HS} \\ =& \frac{1}{|||C_1|||_2^2}\langle C, C_1 \tilde{\otimes} T_1(C, C_1)\rangle_{HS} - \frac{|||T_1(C, C_1)|||_2^2}{|||C_1|||_2^2} \\ &- \langle C, C_1 \tilde{\otimes} C_2\rangle_{HS} + \langle C_2, T_1(C, C_1)\rangle_{HS},\end{aligned}$$

which is zero by (2.7) and Proposition 2.2. Therefore for all $C_2 \in S_2(H_2)$ we have

$$|||C - C_1 \tilde{\otimes} C_2|||_2^2 \geq |||C - C_1 \tilde{\otimes} \widetilde{C}_2|||_2^2$$

which proves the first assertion of Theorem 2.1.

For a proof of the representation (2.15) we substitute the operator $\widetilde{C}_2$ defined in (2.14) for C_2 in the expression of $D(C_1, C_2)$ and obtain

$$\begin{aligned}D(C_1) =& D(C_1, \widetilde{C}_2) = |||C - C_1 \tilde{\otimes} \widetilde{C}_2|||_2^2 = \langle C - C_1 \tilde{\otimes} \widetilde{C}_2, C - C_1 \tilde{\otimes} \widetilde{C}_2\rangle_{HS} \\ =& |||C|||_2^2 + |||C_1 \tilde{\otimes} \widetilde{C}_2|||_2^2 - 2\langle C, C_1 \tilde{\otimes} \widetilde{C}_2\rangle_{HS} \\ =& |||C|||_2^2 + \frac{|||T_1(C, C_1)|||_2^2}{|||C_1|||_2^2} - \frac{2}{|||C_1|||_2^2}\langle C, C_1 \tilde{\otimes} T_1(C, C_1)\rangle_{HS}.\end{aligned}$$

Then the second assertion follows from Proposition 2.2.

Now assume that $C = C_1 \tilde{\otimes} C_2$ for some $C_2 \in S(H_2)$, then (2.5) implies

$$|||T_1(C, C_1)|||_2^2 = (\langle C_1, C_1\rangle_{S_2(H_1)})^2 |||C_2|||_2^2 = |||C_1|||_2^4 |||C_2|||_2^2$$

and therefore $D(C_1) = 0$. Conversely, if $D(C_1) = 0$, we have $C = C_1 \tilde{\otimes} \widetilde{C}_2$, with $|||\widetilde{C}_2|||_2 \leq |||C|||_2$ by Corollary 2.1. □

2.2 Hilbert-Schmidt integral operators

An important case for applications consists is the set $H := L^2(S \times T, \mathbb{R})$ of all square integrable functions defined on $S \times T$, where $S \subset \mathbb{R}^p$, $T \subset \mathbb{R}^q$ are bounded measurable sets. If the covariance operator C of the random variable X is a Hilbert-Schmidt operator it follows from Theorem 6.11 in Weidmann (1980) that there exists a kernel $c \in L^2\big((S \times T) \times (S \times T), \mathbb{R}\big)$ such that C can be characterized as an integral operator, i.e.

$$Cf(s, t) = \int_S \int_T c(s, t, s', t') f(s', t') ds' dt', \quad f \in L^2(S \times T, \mathbb{R}),$$

almost everywhere on $S\times T$. Moreover the kernel is given by the covariance kernel of X, that is $c(s,t,s',t') = \operatorname{Cov}\left[X(s,t), X(s',t')\right]$, and the space H can be identified with the tensor product of $H_1 = L^2(S,\mathbb{R})$ and $H_2 = L^2(T,\mathbb{R})$.

Similarly, if C_1 and C_2 are Hilbert-Schmidt operators on $L^2(S,\mathbb{R})$ and $L^2(T,\mathbb{R})$ respectively there exists symmetric kernels $c_1 \in L^2(H_1\times H_1,\mathbb{R})$ and $c_2 \in L^2(H_2\times H_2,\mathbb{R})$ such that

$$\begin{aligned} C_1 f(s) &= \int_S c_1(s,s') f(s') ds',\ f\in H_1 \\ C_2 g(t) &= \int_T c_2(t,t') g(t') dt',\ g\in H_2 \end{aligned}$$

almost everywhere on S and T, respectively. The following result shows that in this case the operators T_1 and T_2 defined by (2.5) and (2.6), respectively, are also Hilbert-Schmidt (or equivalently integral) operators.

Proposition 2.3. *If C and C_1 are integral operators with kernels $c \in L^2\big((S\times T)\times(S\times T),\mathbb{R}\big)$ and $c_1 \in L^2(S\times S,\mathbb{R})$, then $T_1(C,C_1)$ is a an integral operator with kernel given by*

$$k(t,t') = \int_S\int_S c(s,t,s't') c_1(s,s') ds ds'. \tag{2.16}$$

An analog result is true for the operator T_2.

Proof. By Lemma A.2 in Appendix A for a given $\epsilon>0$ there exists an integral operator C' with kernel c' such that

$$|||C-C'|||_2 < \frac{\epsilon}{2|||C_1|||_2} \text{ and } \|c-c'\|_\infty < \epsilon/2,$$

where $C' = \sum_{n=1}^N A_n\tilde{\otimes}B_n$, and A_n and B_n are finite rank operators with continuous kernels a_n and b_n. Note that $\sum_{n=1}^N a_n b_n$ is the kernel of the operator C'. Let K be the integral operator with the kernel defined in (2.16) and K' be the integral operator with kernel

$$k'(t,t') := \int_S\int_S c'(s,t,s't') c_1(s,s') ds ds',$$

then (note that K' is a Hilbert-Schmidt operator)

$$|||T_1(C,C_1)-K|||_2 \le |||T_1(C,C_1)-T_1(C',C_1)|||_2 + |||T_1(C',C_1)-K'|||_2 + |||K'-K|||_2. \tag{2.17}$$

By (2.12) the first term is bounded by $|||C_1|||_2 |||C-C_1|||_2 < \epsilon/2$. To handle the second term note that for any $f\in H_2$,

$$\begin{aligned} T_1(C',C_1)f(t) =& T_1\left(\sum_{n=1}^N A_n\tilde{\otimes}B_n, C_1\right) f(t) = \sum_{n=1}^N \langle A_n, C_1\rangle_{S_2(H_1)} B_n f(t) \\ &= \sum_{n=1}^N \int_T\int_S\int_S a_n(s,s') c_1(s,s') b_n(t,t') f(t') ds ds' dt' \end{aligned}$$

$$= \int_T \int_S \int_S c'(s,t,s',t') c_1(s,s') ds ds' f(t') dt'$$
$$= \int_T k'(t,t') f(t') dt' = K' f(t).$$

Therefore, the second term in (2.17) is zero. If $|S|$ and $|T|$ denote the Lebesgue measure of the set S and T, respectively, the third term can be written as

$$\left[\int_T \int_T \left(\int_S \int_S (c(s,t,s',t') - c'(s,t,s't')) c_1(s,s') ds ds' \right)^2 dt dt' \right]^{1/2}$$
$$\leq \|c - c'\|_\infty \|c_1\|_2 |S|^2 |T|,$$

which is bounded by $\|c_1\|_2 |S|^2 |T| \epsilon/2$. Since the choice of $\epsilon > 0$ is arbitrary, this proves the assertion of Proposition 2.3. □

Using the explicit formula for T_1 described in Proposition 2.3 the minimum distance can be expressed in terms of the corresponding kernels of the operators, that is

$$\begin{aligned} D(C_1) &= ⦀C⦀_2^2 - \frac{⦀T_1(C, C_1)⦀_2^2}{⦀C_1⦀_2^2} \\ &= \int_T \int_T \int_S \int_S c^2(s,t,s',t') ds ds' dt dt' \\ &\quad - \frac{\int_T \int_T \left[\int_S \int_S c(s,t,s't') c_1(s,s') ds ds' \right]^2 dt dt'}{\int_S \int_S c_1^2(s,s') ds ds'}. \end{aligned} \tag{2.18}$$

3 Estimation and asymptotic properties

We estimate the minimum distance given in (2.15) by plugging in estimators for C and C_1 based on a sample $X_1, X_2, \ldots, X_N$. The covariance operator C is estimated by

$$\widehat{C}_N := \frac{1}{N} \sum_{i=1}^{N} \left[(X_i - \overline{X}) \otimes_o (X_i - \overline{X}) \right]. \tag{3.1}$$

For the estimation of the operator C_1 we first note that it is sufficient to estimate C_1 up to a multiplicative constant, due to the self-normalizing form of the second term of the minimum distance $D(C_1)$. Let Δ denote the identity operator in $S_2(H_2)$, note that under the null hypothesis of separability $H_0 : C = C_1 \otimes C_2$ we have $T_2(C, \Delta) = \langle C_2, \Delta \rangle_{S_2(H_2)} C_1$, and for this choice the minimum distance in (2.15) is given by

$$D_0 := D(T_2(C, \Delta)) = ⦀C⦀_2^2 - \frac{⦀T_1(C, T_2(C, \Delta))⦀^2}{⦀T_2(C, \Delta)⦀_2^2}. \tag{3.2}$$

Therefore, we propose to estimate (a multiple of) the operator C_1 by

$$\widehat{C}_{1N} = T_2(\widehat{C}_N, \Delta), \tag{3.3}$$

and obtain the test statistic

$$\hat{D}_N = |||\widehat{C}_N|||_2^2 - \frac{|||T_1(\widehat{C}_N, T_2(\widehat{C}_N, \Delta))|||^2}{|||T_2(\widehat{C}_N, \Delta)|||_2^2}. \tag{3.4}$$

3.1 Weak convergence

The main result of this section provides the asymptotic properties of the statistic $\hat{D}_N$.

Theorem 3.1. *If $\mathbb{E}\|X\|_2^4 < \infty$, then $\sqrt{N}\big(\hat{D}_N - D_0\big)$ converges weakly to a zero mean normal distribution, where the centering term D_0 is defined in* (3.2).

Proof. Observing (3.2) and (3.4) we write

$$\sqrt{N}\big(\widehat{D}_N - D_0\big) = \sqrt{N}\left(|||\widehat{C}_N|||_2^2 - \frac{|||T_1(\widehat{C}_N, T_2(\widehat{C}_N, \Delta))|||^2}{|||T_2(\widehat{C}_N, \Delta)|||_2^2} - |||C|||_2^2 + \frac{|||T_1(C, T_2(C, \Delta))|||^2}{|||T_2(C, \Delta)|||_2^2}\right) \tag{3.5}$$

and note that $\widehat{D}_N$ and D_0 are functions of the random variables

$$G_N = \big(|||\widehat{C}_N|||_2^2,\ |||T_1(\widehat{C}_N, T_2(\widehat{C}_N, \Delta))|||^2,\ |||T_2(\widehat{C}_N, \Delta)|||_2^2\big)^T,$$
$$G = \big(|||C|||_2^2,\ |||T_1(C, T_2(C, \Delta))|||^2,\ |||T_2(C, \Delta)|||_2^2\big)^T,$$

respectively. Therefore, we first investigate the weak convergence of the vector $\sqrt{N}(G_N - G)$. For this purpose we note that for A and B in $S_2(H)$, the identity

$$|||A|||_2^2 - |||B|||_2^2 = |||A - B|||_2^2 + 2\langle A - B, B\rangle_{HS}$$

holds and introduce the decomposition

$$\sqrt{N}(G_N - G) = I_N + II_N\ ,$$

where the random variables I_N and II_N are defined by

$$\begin{aligned} I_N =& \sqrt{N}\big(|||\widehat{C}_N - C|||_2^2,\ |||T_1(\widehat{C}_N, T_2(\widehat{C}_N, \Delta)) - T_1(C, T_2(C, \Delta))|||_2^2,\ |||T_2(\widehat{C}_N, \Delta) - T_2(C, \Delta)|||_2^2\big)^T, \\ II_N =& 2\sqrt{N}\big(\langle \widehat{C}_N - C, C\rangle_{HS},\ \langle T_1(\widehat{C}_N, T_2(\widehat{C}_N, \Delta)) - T_1(C, T_2(C, \Delta)), T_1(C, T_2(C, \Delta))\rangle_{HS}, \\ & \langle T_2(\widehat{C}_N, \Delta) - T_2(C, \Delta), T_2(C, \Delta)\rangle_{HS}\big)^T. \end{aligned}$$

Using the linearity of T_1 and T_2 we further write

$$\begin{aligned} T_1(\widehat{C}_N, T_2(\widehat{C}_N, \Delta)) - T_1(C, T_2(C, \Delta)) =& T_1(\widehat{C}_N, T_2(\widehat{C}_N, \Delta)) - T_1(C, T_2(\widehat{C}_N, \Delta)) \\ & + T_1(C, T_2(\widehat{C}_N, \Delta)) - T_1(C, T_2(C, \Delta)) \\ =& T_1(\widehat{C}_N - C, T_2(\widehat{C}_N, \Delta)) + T_1(C, T_2(\widehat{C}_N - C, \Delta)). \end{aligned}$$

and obtain the representation

$$
\begin{aligned}
I_N =& \frac{1}{\sqrt{N}} \begin{pmatrix} \left\| \sqrt{N}(\widehat{C}_N - C) \right\|_2^2 \\ \left\| T_1\left(\sqrt{N}(\widehat{C}_N - C), T_2(\widehat{C}_N, \Delta)\right) + T_1\left(C, T_2(\sqrt{N}(\widehat{C}_N - C), \Delta)\right) \right\|_2^2 \\ \left\| T_2\left(\sqrt{N}(\widehat{C}_N - C), \Delta\right) \right\|_2^2 \end{pmatrix} \\
=:& \frac{1}{\sqrt{N}} F_1\left(\sqrt{N}(\widehat{C}_N - C), \widehat{C}_N\right), \\
II_N =& 2 \begin{pmatrix} \left\langle \sqrt{N}(\widehat{C}_N - C), C \right\rangle_{HS} \\ \left\langle T_1\left(\sqrt{N}(\widehat{C}_N - C), T_2(\widehat{C}_N, \Delta)\right) + T_1\left(C, T_2(\sqrt{N}(\widehat{C}_N - C), \Delta)\right), T_1(C, T_2(C, \Delta)) \right\rangle_{HS} \\ \left\langle T_2\left(\sqrt{N}(\widehat{C}_N - C), \Delta\right), T_2(C, \Delta) \right\rangle_{HS} \end{pmatrix} \\
=:& F_2\left(\sqrt{N}(\widehat{C}_N - C), \widehat{C}_N\right),
\end{aligned}
$$

where the last equations define the functions F_1 and F_2 in an obvious manner. Note that

$$
F := (F_1, F_2) : S_2(H) \times S_2(H) \mapsto \mathbb{R}^6
$$

is composition of continuous functions and hence continuous. By Proposition 5 in Dauxois et al. (1982) the random variable $\sqrt{N}(\widehat{C}_N - C)$ converges in distribution to a centered Gaussian random element $\mathcal{G}$ with variance

$$
\Gamma := \lim_{N\to\infty} \operatorname{Var}(\sqrt{N}\widehat{C}_N) \tag{3.6}
$$

in $S_2(H)$ with respect to Hilbert-Schmidt topology. Therefore, using continuous mapping arguments we have

$$
F\left(\sqrt{N}(\widehat{C}_N - C), \widehat{C}_N\right) \xrightarrow{d} F(\mathcal{G}, C) \ ,
$$

and consequently

$$
\sqrt{N}(G_N - G) = \frac{1}{\sqrt{N}} F_1\left(\sqrt{N}(\widehat{C}_N - C), \widehat{C}_N\right) + F_2\left(\sqrt{N}(\widehat{C}_N - C), \widehat{C}_N\right) \xrightarrow{d} F_2(\mathcal{G}, C).
$$

We write

$$
F_2(\mathcal{G}, C) = 2 \begin{pmatrix} \langle \mathcal{G}, C \rangle_{HS} \\ \langle T_1\left(\mathcal{G}, T_2(C, \Delta)\right) + T_1\left(C, T_2(\mathcal{G}, \Delta)\right), T_1(C, T_2(C, \Delta)) \rangle_{HS} \\ \langle T_2\left(\mathcal{G}, \Delta\right), T_2(C, \Delta) \rangle_{HS} \end{pmatrix},
$$

which can be further simplified as

$$
F_2(\mathcal{G}, C) = 2 \begin{pmatrix} \langle \mathcal{G}, C \rangle_{HS} \\ \langle \mathcal{G}, T_1(C, T_2(C, \Delta)) \tilde{\otimes} T_2(C, \Delta) \rangle_{HS} + \langle C, T_1(C, T_2(C, \Delta)) \tilde{\otimes} T_2(\mathcal{G}, \Delta) \rangle_{HS} \\ \langle \mathcal{G}, T_2(C, \Delta) \tilde{\otimes} \Delta \rangle_{HS} \end{pmatrix}. \tag{3.7}
$$

By Proposition 2.1 $T_2(\mathscr{G},\Delta)$ is a Gaussian Process in $S_2(H_2)$. This fact along with Lemma A.3 in Appendix A imply that $F_2(\mathscr{G},C)$ is a normal distributed random vector with mean zero and covariance matrix, say Σ. By (3.5),

$$\sqrt{N}\big(\widehat{D}_N - D_0\big) = \sqrt{N}(f(G_N) - f(G)) \ ,$$

where the function $f:\mathbb{R}^3 \mapsto \mathbb{R}$ is defined by $f(x,y,z) = x - y/z$. Therefore, using the delta method and the fact that

$$\mathbb{P}\big(\|\| T_2(\widehat{C}_N,\Delta)\|\|_2^2 > 0\big) \to \mathbb{P}\big(\|\| T_2(C,\Delta)\|\|_2^2 > 0\big) = 1$$

as $\|\|C\|\|_2 \neq 0$, we finally obtain

$$\sqrt{N}\big(\widehat{D}_N - D_0\big) \xrightarrow{d} N\big(0,(\nabla f(G))^T\Sigma(\nabla f(G))\big) \tag{3.8}$$

as $n\to\infty$ where $\nabla f(x,y,z) = (1,-1/z,-y/z^2)^T$ denotes the gradient of the function f. □

Remark 3.1. A sufficient condition for Theorem 3.1 to hold is $\mathbb{E}\|X\|_2^4 < \infty$. This is weaker than Condition 2.1 in Aston et al. (2017), as indicated in Remark 2.2 (1) of their paper. These authors used this assumption to prove weak convergence under the trace-norm topology, which is required to establish Theorem 2.3 in Aston et al. (2017). In contrast the proof of Theorem 3.1 here only requires weak convergence under the Hilbert-Schmidt topology, which defines a weaker topology.

The following result specializes Theorem 3.1 to the null hypothesis of separability, where a simple expression for the asymptotic variance is available.

Corollary 3.1. *If* $\mathbb{E}\|X\|_2^4 < \infty$ *and the null hypothesis of separability holds, we have*

$$\sqrt{N}\widehat{D}_N \xrightarrow{d} N(0,v^2)$$

as $N\to\infty$*, where the asymptotic variance is given by*

$$v^2 = \frac{16\|\|C\|\|_2^4}{\|\| T_2(C,\Delta)\|\|_2^4}\Big\langle \Gamma T_2(C,\Delta)\tilde{\otimes}\Delta, T_2(C,\Delta)\tilde{\otimes}\Delta\Big\rangle_{HS} \tag{3.9}$$

and the operator Γ *is defined in* (3.6).

Proof. Under the null hypothesis of separability, we have

$$\begin{aligned} T_2(C,\Delta) &= \langle C_2,\Delta\rangle_{S_2(H_2)} C_1, \\ T_1(C,T_2(C,\Delta)) &= \langle C_2,\Delta\rangle_{S_2(H_2)} T_1(C,C_1) = \langle C_2,\Delta\rangle_{S_2(H_2)} \|\|C_1\|\|_2^2 C_2 \end{aligned} \tag{3.10}$$

and as a consequence we obtain

$$D_0 = |||C|||_2^2 - \frac{|||T_1(C, T_2(C,\Delta))|||^2}{|||T_2(C,\Delta)|||_2^2} = |||C|||_2^2 - |||C_1|||_2^2 |||C_2|||_2^2 = 0.$$

Therefore, the convergence to a zero mean normal distribution follows from Theorem 3.1. To calculate the asymptotic variance, note that under the null hypothesis of separability the second coordinate of $F_2(\mathcal{G}, C)$ in (3.7) can be simplified using Proposition 2.1 and (3.10) as follows,

$$\begin{aligned}
\langle T_1(\mathcal{G}, T_2(C,\Delta)), T_1(C, T_2(C,\Delta))\rangle_{HS} =& \langle \mathcal{G}, T_2(C,\Delta) \tilde{\otimes} T_1(C, T_2(C,\Delta))\rangle_{HS} \\
=& \langle C_2, \Delta\rangle^2_{S_2(H_2)} |||C_1|||_2^2 \langle \mathcal{G}, C_1 \tilde{\otimes} C_2\rangle_{HS} \\
=& |||T_2(C,\Delta)|||_2^2 \langle \mathcal{G}, C\rangle_{HS} \\
\langle T_1(C, T_2(\mathcal{G},\Delta)), T_1(C, T_2(C,\Delta))\rangle_{HS} =& \Big\langle \langle C_1, T_2(\mathcal{G},\Delta)\rangle_{HS} C_2, \langle C_2, \Delta\rangle_{S_2(H_2)} |||C_1|||_2^2 C_2 \Big\rangle \\
=& |||C_1|||_2^2 |||C_2|||_2^2 \langle C_2, \Delta\rangle_{S_2(H_2)} \langle \mathcal{G}, C_1 \tilde{\otimes} \Delta\rangle_{HS} \\
=& |||C|||_2^2 \langle \mathcal{G}, T_2(C,\Delta) \tilde{\otimes} \Delta\rangle_{HS}.
\end{aligned}$$

The elements in the covariance matrix $\Sigma = (\Sigma_{ij})$ of $F_2(\mathcal{G}, C)$ are therefore given by

$$\begin{aligned}
\Sigma_{11} =& 4\Big\langle \Gamma C, C\Big\rangle_{HS} \\
\Sigma_{22} =& 4|||T_2(C,\Delta)|||_2^4 \Big\langle \Gamma C, C\Big\rangle_{HS} + 4|||C|||_2^4 \Big\langle \Gamma T_2(C,\Delta)\tilde{\otimes}\Delta, T_2(C,\Delta)\tilde{\otimes}\Delta\Big\rangle_{HS} \\
&+ 8|||T_2(C,\Delta)|||_2^2 |||C|||_2^2 \Big\langle \Gamma C, T_2(C,\Delta)\tilde{\otimes}\Delta\Big\rangle_{HS} \\
\Sigma_{33} =& 4\Big\langle \Gamma T_2(C,\Delta)\tilde{\otimes}\Delta, T_2(C,\Delta)\tilde{\otimes}\Delta\Big\rangle_{HS} \\
\Sigma_{12} =& 4|||T_2(C,\Delta)|||_2^2 \Big\langle \Gamma C, C\Big\rangle_{HS} + 4|||C|||_2^2 \Big\langle \Gamma C, T_2(C,\Delta)\tilde{\otimes}\Delta\Big\rangle_{HS} \\
\Sigma_{13} =& 4\Big\langle \Gamma C, T_2(C,\Delta)\tilde{\otimes}\Delta\Big\rangle_{HS} \\
\Sigma_{23} =& 4|||T_2(C,\Delta)|||_2^2 \Big\langle \Gamma C, T_2(C,\Delta)\tilde{\otimes}\Delta\Big\rangle_{HS} + 4|||C|||_2^2 \Big\langle \Gamma T_2(C,\Delta)\tilde{\otimes}\Delta, T_2(C,\Delta)\tilde{\otimes}\Delta\Big\rangle_{HS}.
\end{aligned}$$

Using this the asymptotic variance $\nu^2 = (\nabla f(G))^T \Sigma (\nabla f(G))$ in the proof of Theorem 3.1 can be simplified by a straightforward but tedious calculation to

$$\nu^2 = \frac{16|||C|||_2^4}{|||T_2(C,\Delta)|||_2^4} \Big\langle \Gamma T_2(C,\Delta)\tilde{\otimes}\Delta, T_2(C,\Delta)\tilde{\otimes}\Delta\Big\rangle_{HS},$$

which proves our assertion. □

Let $u_{1-\alpha}$ denote the $(1-\alpha)$ quantile of the standard normal distribution and let $\hat{\nu}_N^2$ be a consistent estimate of the limiting variance (3.9). Then it follows from Theorem 3.1

and Corollary 3.1 that the test which rejects the null hypothesis of separability, whenever

$$\widehat{D}_N > u_{1-\alpha} \frac{\hat{v}_N}{\sqrt{N}} \tag{3.11}$$

is an asymptotic level α test. The following result shows that this test is also consistent.

Corollary 3.2. *Assume that* $\mathbb{E}\|X\|_2^4 < \infty$. *If the null hypothesis of separability is not true, we have* $\sqrt{N}\widehat{D}_N \xrightarrow{p} \infty$ *as* $N \to \infty$.

Proof. By the third part of Theorem 2.1 we can write

$$D_0 = |||C|||_2^2 - \frac{|||T_1(C, T_2(C,\Delta))|||_2^2}{|||T_2(C,\Delta)|||_2^2}.$$

As in the proof of Theorem 2.1, the last quantity can be shown to be equal to $|||C - T_2(C,\Delta)\tilde{\otimes} T_1(C, T_2(C,\Delta))|||_2^2$. Note that $T_2(C,\Delta) \in S_2(H_1)$ and $T_1(C, T_2(C,\Delta)) \in S_2(H_2)$, therefore D_0 is positive if H_0 is not true. Hence the result follows from Theorem 3.1. □

3.2 Hilbert-Schmidt integral operators

In the remaining part of this section we concentrate on the case, where X is a random element in $H = L^2(S \times T)$, using the explicit formula for the operator T_1 described in Proposition 2.3 the minimum distance can be expressed in terms of the corresponding kernels, that is

$$\begin{aligned} D_0 = D(T_2(C,\Delta)) &= \int_T \int_T \int_S \int_S c^2(s,t,s',t') ds ds' dt dt' \\ &- \frac{\int_T \int_T \left[\int_S \int_S c(s,t,s't')\tilde{c}_1(s,s') ds ds'\right]^2 dt dt'}{\int_S \int_S \tilde{c}_1^2(s,s') ds ds'}. \end{aligned} \tag{3.12}$$

where $\tilde{c}_1$ denote the kernel corresponding to the operator $T_2(C,\Delta)$, that is

$$\tilde{c}_1(s,s') = \int_T c(s,t,s',t) dt .$$

In this case the estimator $\widehat{C}_N$ defined in (3.1) is induced by the kernel

$$\hat{c}_N(s,t,s',t') = \frac{1}{N} \sum_{i=1}^N (X_i(s,t) - \overline{X}(s,t))(X_i(s',t') - \overline{X}(s',t')) ,$$

and the estimator $\widehat{C}_{1N} = T_2(\widehat{C}_N, \Delta)$ defined in (3.3) is induced by the kernel

$$\hat{c}_{1N}(s,s') = \int_T \hat{c}_N(s,t,s',t) dt.$$

The estimator $\hat{D}_N$ of D_0 is calculated by plugging in $\hat{c}_N$ and $\hat{c}_{1N}$ to the expression in (3.12). By Corollary 3.1, under the null hypothesis of separability, the statistic $\sqrt{N}\hat{D}_N$ is asymptotically centered normal distributed with variance given by (3.9). In order to define an estimator of this variance we introduce the following technical lemma which is proved in Appendix A.

Lemma 3.1. *Assume* $\mathbb{E}||X||_2^4 < \infty$ *and that* $H = L^2(S \times T, \mathbb{R})$. *Then the operator* $\Gamma \in S_2(H)$ *defined by* (3.6) *is induced by the kernel*

$$\begin{aligned}\gamma(s_1, t_1, s_2, t_2, s_3, t_3, s_4, t_4) =& k(s_1, t_1, s_2, t_2, s_3, t_3, s_4, t_4) + c(s_1, t_1, s_3, t_3)c(s_2, t_2, s_4, t_4) \\ &+ c(s_1, t_1, s_4, t_4)c(s_2, t_2, s_3, t_3) - 2c(s_1, t_1, s_2, t_2)c(s_3, t_3, s_4, t_4),\end{aligned} \tag{3.13}$$

where the kernel k is given by

$$\begin{aligned}k(s_1, t_1, s_2, t_2, s_3, t_3, s_4, t_4) =& \mathbb{E}\Big((X(s_1, t_1) - \mu(s_1, t_1))(X(s_2, t_2) - \mu(s_2, t_2)) \\ &(X(s_3, t_3) - \mu(s_3, t_3))(X(s_4, t_4) - \mu(s_4, t_4))\Big).\end{aligned} \tag{3.14}$$

By Proposition 5 of Dauxois et al. (1982) if $\mathbb{E}\|X\|_2^8 < \infty$ the 4-th order kernel k defined in (3.14) can be consistently estimated by

$$\begin{aligned}\hat{k}_N(s_1, t_1, s_2, t_2, s_3, t_3, s_4, t_4) =& \frac{1}{N}\sum_{i=1}^{N}\Big((X_i(s_1, t_1) - \overline{X}(s_1, t_1))(X_i(s_2, t_2) - \overline{X}(s_2, t_2)) \\ &(X_i(s_3, t_3) - \overline{X}(s_3, t_3))(X_i(s_4, t_4) - \overline{X}(s_4, t_4))\Big),\end{aligned}$$

i.e., $\hat{k}_N$ converges in probability to k as an element of $L^2\left[(S \times T)^4, \mathbb{R}\right]$. A consistent estimator of the limiting covariance kernel γ is obtained by replacing k and c by $\hat{k}_N$ and $\hat{c}_N$, respectively in (3.13), i.e.,

$$\begin{aligned}\hat{\gamma}_N(s_1, t_1, s_2, t_2, s_3, t_3, s_4, t_4) =& \hat{k}_N(s_1, t_1, s_2, t_2, s_3, t_3, s_4, t_4) + \hat{c}_N(s_1, t_1, s_3, t_3)\hat{c}_N(s_2, t_2, s_4, t_4) \\ &+ \hat{c}_N(s_1, t_1, s_4, t_4)\hat{c}_N(s_2, t_2, s_3, t_3) - 2\hat{c}_N(s_1, t_1, s_2, t_2)\hat{c}_N(s_3, t_3, s_4, t_4).\end{aligned} \tag{3.15}$$

Indeed a consistent estimator of the variance in (3.9) is given by

$$\hat{v}^2 = \frac{16 |||\widehat{C}_N|||_2^4}{|||T_2(\widehat{C}_N, \Delta)|||_2^4}\Big\langle\widehat{\Gamma}_N T_2(\widehat{C}_N, \Delta)\tilde{\otimes}\Delta, T_2(\widehat{C}_N, \Delta)\tilde{\otimes}\Delta\Big\rangle_{HS}. \tag{3.16}$$

where $\widehat{\Gamma}_N$ is the operator induced by $\hat{\gamma}_N$ by right integration.

Remark 3.2. Although the propsoed estimator is based on the norm of the complete covariance kernel c, numerically we do not need to store the complete covariance kernel. For example, we obtain for the first term of the statistic $\hat{D}_N$ the representation

$$\begin{aligned}
|||\widehat{C}_N|||_2^2 &= \frac{1}{N^2}\int_T\int_S\int_T\int_S\left(\sum_{i=1}^N (X_i(s,t)-\overline{X}(s,t))(X_i(s',t')-\overline{X}(s',t'))\right)^2 dsdtds'dt' \\
&= \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\left[\int_T\int_S (X_i(s,t)-\overline{X}(s,t))(X_j(s,t)-\overline{X}(s,t))dsdt\right]^2
\end{aligned}$$

All other terms of the estimator in (3.4) and the variance estimator (3.16) can be represented similarly using simple matrix operations on the data matrix without storing the full or marginal covariance kernels.

4 Finite Sample Properties

In this section we study the finite sample properties of the test for the hypothesis of separability defined by (3.11) by means of a small simulation study. We also compare the new test with the tests proposed by Aston et al. (2017) and Constantinou et al. (2017) and illustrate potential applications with a data example. For this purpose we have implemented the new method proposed in this paper as well as the bootstrap test described in Aston et al. (2017) and the weighted χ^2 test based on the test statistic $\widehat{T}_F$ as described in Theorem 3 of Constantinou et al. (2017). The test due to Aston et al. (2017) requires specification of eigen-subspace $\mathscr{I}$, which was taken to be $\{(1,1)\}$ as in their simulation studies and p-values were obtained by empirical bootstrap. We used the R package "covsep" to implement their method. For the tests proposed by Constantinou et al. (2017) we choose the procedure based on the statistic $\widehat{T}_F$ as in a simulation study it turned out to be the most powerful procedure among the four methods proposed in this paper. The test requires the specifications of the number of spatial and temporal principle components, which were both taken to be 2.

4.1 Simulation Studies

The data were generated from a zero mean Gaussian process with two different covariance kernels. The first is the spatio-temporal covariance kernel

$$c(s,t,s',t') = \frac{\sigma^2}{(a|t-t'|^{2\alpha}+1)^{\tau}}\exp\left(-\frac{c\|s-s'\|^{2\gamma}}{(a|t-t'|^{2\alpha}+1)^{\beta\gamma}}\right), \tag{4.1}$$

introduced by Gneiting (2002). In this covariance function, a and c are nonnegative scaling parameters of time and space, respectively; α and γ are smoothness parameters which take values in the interval $(0,1]$; β is the separability parameter which varies in

Table 1: *Empirical rejection probabilities of different tests for the hypothesis of separability (level* 5%*). M1: the test* (3.11) *proposed in this paper; M2: the test of Aston et al. (2017); M3: the test of Constantinou et al. (2017). The data are generated from model* (4.1)*, where the case* $\beta = 0$ *corresponds to the null hypothesis of separability.*

	$N = 100$			$N = 200$			$N = 500$		
β	M1	M2	M3	M1	M2	M3	M1	M2	M3
0	3.1 %	0.4%	4.3 %	4.8%	4.3%	5.3%	5.1%	4.9 %	3.8%
0.3	34.7 %	45.8%	11.3 %	47.6 %	60.1%	22.1%	59.2%	64.4%	35.6 %
0.5	63.8 %	56.9 %	26.1 %	79.9%	85.7%	41.2%	93.2 %	90.1%	53.6%
0.7	77.1 %	72.8 %	47.9%	90.1 %	89.4 %	58.5 %	98.5 %	99.3%	67.9%
1	93.9%	95.8%	61.7%	99.9%	98.1 %	83.2%	100%	100%	90.1%

the interval $[0,1]$; $\sigma^2 > 0$ is the point-wise variance; and $\tau \geq \beta d/2$, where d is the spatial dimension. If $\beta = 0$, the covariance is separable and the space-time interaction becomes stronger with increasing values of β. We fix $\gamma = 1, \alpha = 1/2, \sigma^2 = 1, a = 1, c = 1$ and $\tau = 1$ in the following discussion and choose different values for the parameter β specifying the level of separability.

As a second example we consider covariance structure

$$c(s,t,s',t') = \frac{\sigma^2 c_0^{d/2}}{(a_0^2(t-t')^2+1)^{1/2}\Big(a_0^2(t-t')^2+c_0\Big)^{d/2}} \times \exp\Big[-b_0\|s-s'\|\Big(\frac{a_0^2(t-t')^2+1}{a_0^2(t-t')^2+c_0}\Big)^{1/2}\Big], \tag{4.2}$$

which was introduced by Cressie and Huang (1999). Here a_0 and b_0 are non negative scaling parameters of time and space respectively; $c_0 > 0$ is the separability parameter, $\sigma^2 > 0$ is the point-wise variance and d is the spatial dimension. The covariance is separable if $c_0 = 1$. For the simulation study we take $a_0 = 2, b_0 = 1, \sigma^2 = 1$, $d = 2$ and consider different values of the parameter c_0.

We generate data at 100 equally spaced time points in $[0,1]$ and 11 space points on the grid $[0,1]\times[0,1]$. The integrals are approximated by average of the function value at grid points. The nominal significance level is taken to be 5% and empirical rejection region are computed based on 1000 Monte-Carlo replications. The results are displayed in Table 1 and 2, where we show the results for the test (3.11) proposed in this paper (M1), the test based on projections of subspaces suggested in Aston et al. (2017) (M2) and the test and the weighted χ^2 test of Constantinou et al. (2017) (M3).

All procedures yield rather similar results under the null hypothesis, and in general

Table 2: *Empirical rejection probabilities of different tests for the hypothesis of separability (level* 5%*). M1: the test* (3.11) *proposed in this paper; M2: the test of Aston et al. (2017); M3: the test of Constantinou et al. (2017). The are data generated from model* (4.2)*, where the case* $c_0 = 1$ *corresponds to the null hypothesis of separability.*

	$N = 100$			$N = 200$			$N = 500$		
c_0	M1	M2	M3	M1	M2	M3	M1	M2	M3
1	4.3 %	5.5%	8.1 %	4.9%	4.2 %	6.3%	3.8%	3.1%	7.2%
3	40.2 %	39.1%	9.3 %	51.3 %	59.7 %	13.4 %	67.3%	66.8%	22.8%
5	62.5 %	61.7 %	38.2 %	82.4 %	82.7%	52.9%	95.4%	94.9 %	73.5%
7	79.1 %	75.4%	59.4%	88.2%	91.4 %	83.3%	99.1 %	99.9%	91.8%
10	94.8%	93.9%	75.2%	99.9 %	100 %	91%	100%	100%	96.3%

the nominal level is well approximated. Only for model (1) and sample size $N = 100$ the bootstrap test of Aston et al. (2017) is very conservative, while nominal level of the test of Constantinou et al. (2017) is slightly too large in model (2). Under the alternative this test yields less power than the test of Aston et al. (2017) and the test (3.11) proposed in this paper. On the other hand we do not observe substantial difference between these two tests. In model (1) the bootstrap test of Aston et al. (2017) shows a slightly better performance in 6 of 11 cases and exactly the opposite behavior is observed in model (2). Here the new test (3.11) slightly outperforms the bootstrap test in 6 of 11 cases.

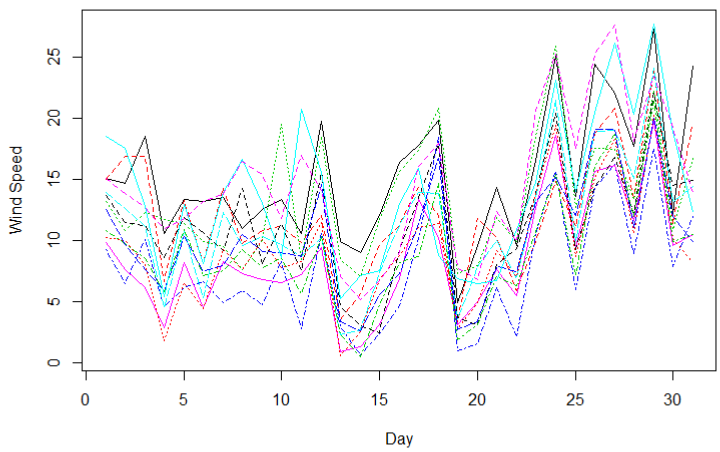


Figure 1: *Average wind speed curves for January 1961*

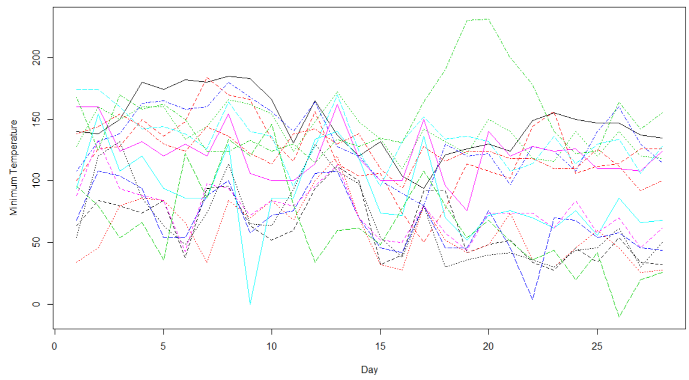
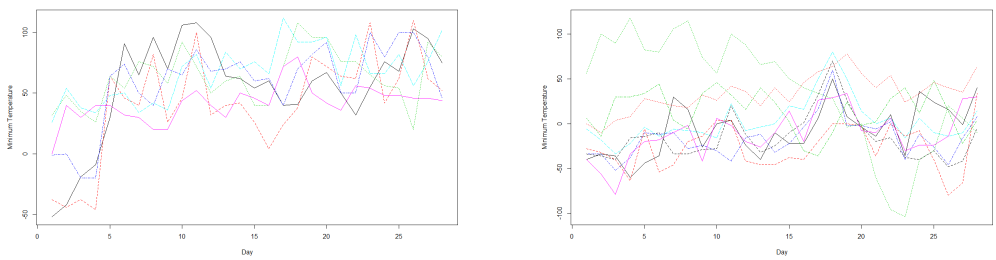


Figure 2: *Minimum Temperature curves for January 1971*



(a) *Coastal Region*

(b) *Inland Region*

Figure 3: *Minimum Temperature curves (Coast and Inland) for January 1971*

4.2 Application to Real Data

4.2.1 Irish Wind Data

We finally apply the new method to the Irish wind data of Haslett and Raftery (1989). The data is available at `http://lib.stat.cmu.edu/datasets/wind.data` and consists of daily average wind-speed at 12 meteorological stations in Ireland during the period 1961–1978. The locations of the stations can be found in Table 3 of Gneiting (2002). We treat the data as functional observations $X_n(s,t)$, for each month n, with $N = 216$. A plot of the data for January 1961 is presented in Figure 1.

The data has a seasonal component. We deseasonalize the data by subtracting the monthly mean from each curve as in Constantinou et al. (2017). Our test is applied to the deseasonalized data which gives a p-value of 0.011. This suggests departure from separability, which agrees with the findings of both Gneiting (2002) and Constantinou et al. (2017).

4.2.2 Spanish Temperature Data

We also analyze the Spanish Temperature Data used in Cuesta-Albertos and Febrero-Bande (2010). The data consists of daily minimum temperature of 15 meteorological sites of northern part of Spain during the period 1971–200. Each of these stations are located at different provinces covering 6 coastal and 9 inland areas of Spain. As before we treat them as functional data $X_n(s,t)$, for each month n with $N = 360$. The temperature of January 1971 is presented in Figure 2. As indicated by Cuesta-Albertos and Febrero-Bande (2010), the data from inland and coastal region behave quite differently, the coastal area being more stable. This is also evident from Figure 3, showing temperature curve for these two regions separately.

The data is deseasonalized before analysis by subtracting the monthly means from each curve. Our test is then applied to both the whole data and to the data restricted to the two regions. All the tests yield p-value very close to 0 (< 0.001). This indicates a strong non-separability in the data. Note that Cuesta-Albertos and Febrero-Bande (2010) also found strong space-time interaction, even after considering the effect of the regions. Therefore the violation of separability assumption is expected in this scenario.

Acknowledgments. This work has been supported in part by the Collaborative Research Center "Statistical modelling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1, C1) of the German Research Foundation (DFG). The authors would like to thank Martina Stein, who typed parts of this paper with considerable technical expertise.

References

Aston, J. A. D., Pigoli, D., and Tavakoli, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics*, 45(4):1431–1461.

Bar-Hen, A., Bel, L., and Cheddadi, R. (2008). Spatio-temporal functional regression on paleoecological data. *Functional and Operatorial Statistics*, pages 54–56.

Constantinou, P., Kokoszka, P., and Reimherr, M. (2017). Testing separability of space-time functional processes. *Biometrika*, 104(2):425–437.

Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448):1330–1339.

Cuesta-Albertos, J. and Febrero-Bande, M. (2010). A simple multiway anova for functional data. *Test*, 19(3):537–557.

Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of multivariate analysis*, 12(1):136–154.

Dunford, N. and Schwartz, J. T. (1988). *Linear Operators. Vols. I, II, III.* Wiley Classics Library. John Wiley & Sons, Inc., New York.

Fuentes, M. (2006). Testing for separability of spatial-temporal covariance functions. *Journal of Statistical Planning and Inference*, 136(2):447–466.

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

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, 97(458):590–600.

Gneiting, T., Genton, M. G., and Guttorp, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry. In Finkenstadt, B., Held, L., and Isham, V., editors, *Statistical Methods for Spatio-Temporal Systems*, pages 151–175. Chapman and Hall/CRC, Boca Raton.

Gohberg, I., Goldberg, S., and Kaashoek, M. A. (1990). *Classes of Linear Operators. Vol. I. Operator Theory: Advances and Applications 49.* Birkhäuser, Basel.

Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing ireland's wind power resource. *Applied Statistics*, pages 1–50.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to linear Operators.* Wiley, New York.

Huizenga, H. M., de Munck, J. C., Waldorf, L. J., and Grasman, R. P. P. P. (2002). Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model. *IEEE Transactions on Biomedical Engineering*, 49(6):533–539.

Lifshits, M. (2012). Lectures on gaussian processes. In *Lectures on Gaussian Processes*, pages 1–117. Springer.

Lu, N. and Zimmerman, D. L. (2005). The likelihood ratio test for a separable covariance matrix. *Statistics & Probability Letters*, 73:449–457.

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2005). Testing for separability of space-time covariances. *Environmetrics*, 16:819–831.

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2006). A likelihood ratio test for separability of covariances. *Journal of Multivariate Analysis*, 97(5):1025–1043.

Rabiner, L. R. and Schafer, R. W. (1978). *Digital Processing of Speech Signals 100.* Prentice-Hall, Englewood Cliffs.

Rougier, J. (2017). A representation theorem for stochastic processes with separable covariance functions, and its implications for emulation. *arXiv preprint arXiv:1702.05599.*

Skup, M. (2010). Longitudinal fmri analysis: A review of methods. *Statistics and its interface*, 3(2):235.

Weidmann, J. (1980). *Linear operators in Hilbert spaces.* Graduate texts in mathematics. Springer-Verlag.

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.

A Some technical results

A.1 Properties of Tensor Product Hilbert Spaces

In this section we present some auxiliary results which have been used in the proofs of the main results. The following two results are generalizations of Lemma 1.6 and Lemma 1.7 in the online supplement of (Aston et al., 2017).

Lemma A.1. *The set*

$$\mathscr{D}_0 := \Big\{ \sum_{i=1}^{n} A_i \tilde{\otimes} B_i : A_i \in S_2(H_1), B_i \in S_2(H_2),\ \textit{are finite rank operators}, n \in \mathbb{N} \Big\} \tag{4.3}$$

is dense in $S_2(H_1 \otimes H_2)$.

Proof. Let $T \in S_2(H_1 \otimes H_2)$, then $T = \sum_{j\geq 1} \lambda_j u_j \otimes_o v_j$ where $(\lambda_j)_{j\geq 1}$ are the singular values of T, with $\sum_j \lambda_j^2 < \infty$; and $\{u_j\}_{j\geq 1}$ and $\{v_j\}_{j\geq 1}$ are orthonormal subsets of $H_1 \otimes H_2$. The vectors u_j and v_j's can be further decomposed as $u_j = \sum_{l\geq 1} s_{j,l} e_l^{(1)} \otimes e_l^{(2)}$ and $v_j = \sum_{l\geq 1} t_{j,l} f_l^{(1)} \otimes f_l^{(2)}$, where the series converge in the norm of $H_1 \otimes H_2$. Let

$$u_j^M = \sum_{l=1}^{M} s_{j,l} e_l^{(1)} \otimes e_l^{(2)} \quad \text{and} \quad v_j^M = \sum_{l=1}^{M} t_{j,l} f_l^{(1)} \otimes f_l^{(2)}.$$

Given $1 > \epsilon > 0$, choose N large enough such that $|||T - \sum_{j=1}^{N} \lambda_j u_j \otimes_o v_j|||_2 \leq \epsilon/2$. With this we write

$$\begin{aligned}
\left|\left|\left| T - \sum_{j=1}^{N} \lambda_j u_j^M \otimes_o v_j^M \right|\right|\right|_2 &\leq \left|\left|\left| T - \sum_{j=1}^{N} \lambda_j u_j \otimes_o v_j \right|\right|\right|_2 \\
&\quad + \left|\left|\left| \sum_{j=1}^{N} \lambda_j \left[(u_j - u_j^M) \otimes_o v_j + u_j^M \otimes_o (v_j - v_j^M) \right] \right|\right|\right|_2 \\
&\leq \epsilon/2 + \sum_{j=1}^{N} \lambda_j \left|\left|\left| (u_j - u_j^M) \otimes_o v_j + u_j^M \otimes_o (v_j - v_j^M) \right|\right|\right|_2 \\
&\leq \epsilon/2 + \left[\sum_{j=1}^{N} \lambda_j^2 \right]^{1/2} \left[\sum_{j=1}^{N} \left|\left|\left| (u_j - u_j^M) \otimes_o v_j + u_j^M \otimes_o (v_j - v_j^M) \right|\right|\right|_2 \right]^{1/2} \\
&\leq \epsilon/2 + |||T|||_2 \left[\sum_{j=1}^{N} \left[\left\| (u_j - u_j^M) \right\| \left\| v_j \right\| + \left\| u_j^M \right\| \left\| (v_j - v_j^M) \right\| \right] \right]^{1/2}.
\end{aligned}$$

Choose $M \geq 1$ such that

$$\| u_j - u_j^M \| \leq \min\left\{ \frac{\epsilon^2}{12N|||T|||_2^2}, 1 \right\} \quad \text{and} \quad \| v_j - v_j^M \| \leq \frac{\epsilon^2}{12N|||T|||_2^2}.$$

As $\|u_j\| = 1$ and $\|u_j^M\| \leq \|u_j\| + \|u_j - u_j^M\|$, with this choice we have

$$\left|\!\left|\!\left| T - \sum_{j=1}^{N} \lambda_j u_j^M \otimes_o v_j^M \right|\!\right|\!\right|_2 \leq \epsilon$$

The proof is now complete by noting that $\sum_{j=1}^{N} \lambda_j u_j^M \otimes_o v_j^M$ is a finite rank operator. □

Lemma A.2. *Let $C \in S_2(L^2(S \times T, \mathbb{R}))$, where T and S are compact subsets of $\mathbb{R}^p$ and $\mathbb{R}^q$ respectively, be an integral operator with symmetric continuous kernel c. For any $\epsilon > 0$, there exists an $C' = \sum_{n=1}^{N} A_n \tilde{\otimes} B_n$, where $A_n : L^2(S, \mathbb{R}) \to L^2(S, \mathbb{R})$, $B_n : L^2(T, \mathbb{R}) \to L^2(T, \mathbb{R})$ are finite rank operators with continuous kernels a_n and b_n respectively, such that*

(a) $|||C - C'|||_2 \leq \epsilon$,

(b) $\sup_{s,s' \in S, t,t' \in T} |c(s,t,s',t') - c'(s,t,s',t')| \leq \epsilon$, *where c' is the kernel of the operator C'.*

Proof. By Mercer's Theorem, there exists continuous orthonormal functions $\{u_n\}_{n \geq 1} \subset L^2(S \times T, \mathbb{R})$ and λ_n is the summable sequence of positive eigenvalues, such that

$$c(s,t,s',t') = \sum_{n \geq 1} \lambda_n u_n(s,t) u_n(s',t')$$

where the convergence is absolute and uniform.

Let U_n be the integral operator with kernel u_n and $C_N := \sum_{n=1}^{N} \lambda_n U_n \otimes_o U_n$. Denote the kernel of C_N as c_N. As u_N can be approximated by sums of tensor products of continuous functions, let $u_N^M := \sum_{l=1}^{M} f_{n,l}^{(1)} \otimes f_{n,l}^{(2)}$, where $f_{n,l}^{(1)} \in L^2(S, \mathbb{R}), f_{n,l}^{(2)} \in L^2(T, \mathbb{R})$ are continuous, and choose M such that

$$\|U_n - U_n^M\| \leq \min\left\{ \frac{\epsilon^2}{12N\kappa |||C|||_2^2}, \kappa \right\},$$

where $\kappa = \max_{n=1,\ldots,N} \|U_n\|_\infty$. Writing $C^{N,M} := \sum_{n=1}^{N} \lambda_n U_n^M \otimes_2 U_n^M$, and denoting the kernel by $c^{N,M}$, we have

$$\|c^N - c^{N,M}\|_\infty \leq \sum_{n=1}^{N} \lambda_n \left[\|U_n - U_n^M\|_\infty \|U_n\|_\infty + \|U_n^M\|_\infty \|U_n - U_n^M\|_\infty \right]$$

An application of the Cauchy-Schwartz inequality along with the choice of M gives an upper bound to the last quantity to be $\epsilon/2$. Similar calculations show that

$$|||C_N - C^{N,M}|||_2 \leq \epsilon/2.$$

Finally, as $C^{N,M}$ is indeed a finite sum of tensor products of finite rank integral operators, we have the desired result. □

We conclude this section with a simple result about Gaussian processes on a Hilbert space. For this purpose recall that a random element $\mathscr{G}$ on a real separable Hilbert space H is said to be Gaussian with mean $\mu \in H$ and covariance operator $\Gamma : H \mapsto H$ if for all $x \in H$, the random variable $\langle \mathscr{G}, x\rangle$ has a normal distribution with mean $\langle \mu, x\rangle$ and variance $\langle \Gamma x, x\rangle$. (See Section 1.3 of Lifshits (2012) for more details)

Lemma A.3. *Let H_1 and H_2 be two real separable Hilbert spaces and $\mathscr{G}$ be a Gaussian process on $S_2(H_2)$. Then for all $A \in S_2(H_1)$, the process $A\tilde{\otimes}\mathscr{G}$ is a Gaussian process in $S_2(H_1 \otimes H_2)$.*

Proof. We will show for any $T \in S_2(H_1 \otimes H_2)$, the random variable $\langle T, A\tilde{\otimes}\mathscr{G}\rangle$ has a normal distribution. By Lemma A.1 and continuity of inner-product it is enough to show the result for $T \in \mathscr{D}_0$. Therefore let $T = \sum_{n=1}^N A_n \tilde{\otimes} B_n$ then

$$\langle T, A\tilde{\otimes}\mathscr{G}\rangle = \sum_{n=1}^N \langle A_n, A\rangle_{S_2(H_1)} \langle B_n, \mathscr{G}\rangle_{S_2(H_2)}$$

which is sum of normal random variables and hence normal. □

A.2 Proof of Lemma 3.1:

Proof. Without loss of generality we assume $\mathbb{E}(X) = \mu = 0$. To obtain the kernel γ, we need to calculate the limit of

$$N\mathrm{Cov}\,(\hat{c}_N(s_1, t_1, s_2, t_2), \hat{c}_N(s_3, t_3, s_4, t_4))\,.$$

To this end we write

$$\mathbb{E}(\hat{c}_N(s_1, t_1, s_2, t_2)\hat{c}_N(s_3, t_3, s_4, t_4)) = \frac{1}{N^2}\mathbb{E}((HX(s_1, t_1))^T (HX(s_2, t_2))(HX(s_3, t_3))^T (HX(s_4, t_4)))$$

with $X(s, t) = (X_1(s, t), X_2(s, t), \cdots, X_N(s, t))^T$ and $H = I - \frac{1}{N}J$ where I is the $N \times N$ identity matrix and J is the $N \times N$ matrix with $(J)_{ij} = 1$ for all i, j. Using the fact that $H^T H = H$, we have

$$\begin{aligned}
\mathbb{E}(\hat{c}_N(s_1, t_1, s_2, t_2)\hat{c}_N(s_3, t_3, s_4, t_4)) =& \frac{1}{N^2}\mathbb{E}(X(s_1, t_1)^T HX(s_2, t_2)X(s_3, t_3)^T HX(s_4, t_4)) \\
=& \frac{1}{N^2}\mathbb{E}\Big(\sum_{i,j} h_{ij} X_i(s_1, t_1)X_j(s_2, t_2)\sum_{k,l} h_{kl} X_k(s_3, t_3)X_k(s_4, t_4)\Big) \\
=& \frac{1}{N^2}\sum_{i,j} h_{ii}h_{jj}\mathbb{E}\big(X_i(s_1, t_1)X_i(s_2, t_2)X_j(s_3, t_3)X_j(s_4, t_4)\big) \\
&+ \frac{1}{N^2}\sum_{i,j} h_{ij}^2\mathbb{E}\big(X_i(s_1, t_1)X_i(s_3, t_3)X_j(s_2, t_2)X_j(s_4, t_4)\big) \\
&+ \frac{1}{N^2}\sum_{i,j} h_{ij}^2\mathbb{E}\big(X_i(s_1, t_1)X_i(s_4, t_4)X_j(s_2, t_2)X_j(s_3, t_3)\big)
\end{aligned}$$

$$+\frac{1}{N^2}\sum_i h_{ii}^2 \mathbb{E}\big(X_i(s_1,t_1)X_i(s_2,t_2)X_i(s_3,t_3)X_i(s_4,t_4)\big)\,.$$

The last equality is due to the symmetry of H, the independence of the X_i and the fact that $\mathbb{E}(X_i(s,t))=0$. Further simplifying the last quantity we have

$$\begin{aligned}
\mathbb{E}(\hat{c}_N(s_1,t_1,s_2,t_2)\hat{c}_N(s_3,t_3,s_4,t_4)) =& \frac{(N-1)^2}{N^2}c(s_1,t_1,s_2,t_2)c(s_3,t_3,s_4,t_4)\\
&+\frac{(N-1)}{N^2}c(s_1,t_1,s_3,t_3)c(s_2,t_2,s_4,t_4)\\
&+\frac{(N-1)}{N^2}c(s_1,t_1,s_4,t_4)c(s_2,t_2,s_3,t_3)\\
&+\frac{(N-1)^2}{N^3}k(s_1,t_1,s_2,t_2,s_3,t_3,s_4,t_4)+O\left(\frac{1}{N^2}\right)\,.
\end{aligned}$$

Therefore we finally obtain

$$\begin{aligned}
N\mathrm{Cov}\,(\hat{c}_N(s_1,t_1,s_2,t_2),\hat{c}_N(s_3,t_3,s_4,t_4)) =& \frac{1-2N}{N}c(s_1,t_1,s_2,t_2)c(s_3,t_3,s_4,t_4)\\
&+\frac{(N-1)}{N}c(s_1,t_1,s_3,t_3)c(s_2,t_2,s_4,t_4)\\
&+\frac{(N-1)}{N}c(s_1,t_1,s_4,t_4)c(s_2,t_2,s_3,t_3)\\
&+\frac{(N-1)^2}{N^2}k(s_1,t_1,s_2,t_2,s_3,t_3,s_4,t_4)+O\left(\frac{1}{N}\right)\,,
\end{aligned}$$

and taking the limit of the last expression as $N\to\infty$, gives the desired result. □