Asymptotic normality of the principal components of functional time series

Piotr Kokoszka[a], Matthew Reimherr[b,*]

[a] *Colorado State University, Fort Collins, USA*
[b] *University of Chicago, Chicago, USA*

Abstract

We establish the asymptotic normality of the sample principal components of functional stochastic processes under nonrestrictive assumptions which admit nonlinear functional time series models. We show that the aforementioned asymptotic depends only on the asymptotic normality of the sample covariance operator, and that the latter condition holds for weakly dependent functional time series which admit expansions as Bernoulli shifts. The weak dependence is quantified by the condition of L^4-m-approximability which includes all functional time series models in practical use. We also demonstrate convergence of the cross covariance operators of the sample functional principal components to their counterparts in the normal limit.
ⓒ 2013 Elsevier B.V. All rights reserved.

Keywords: Asymptotic normality; Functional principal components; Weak dependence

1. Introduction

The objective of this paper is to establish the asymptotic normality of the sample functional principal components (FPCs) under weak assumptions, and for a wide class of functional data. FPCs play a fundamental role in functional data analysis, much more important than the role of the principal components in multivariate analysis. This is due to the fact that functional data are infinitely dimensional, and the FPCs provide a basis for the most important dimension reduction technique though the Karhunen–Loéve expansion. The asymptotic properties of the estimators of the FPCs are therefore central to statistical applications. We cannot give anything approaching a full account of the applications of the FPCs, but examples include human growth patterns, credit

* Correspondence to: Department of Statistics, University of Chicago, Chicago, IL 60637, USA. Tel.: +1 8012317735.
   E-mail address: mreimherr@uchicago.edu (M. Reimherr).

card transaction volumes, asset prices, geomagnetic storm activity, and climate applications such as temperature and rainfall. For more detailed discussions of these applications and more, we recommend the texts [29,19,11]. We first explain the contribution of this paper, and provide further comments later.

Suppose $\{X_n, \ n = 1, 2, \ldots\}$ is a stochastic process taking values in the space $L^2(\mathcal{T})$ of square integrable functions on a compact interval $\mathcal{T}$. We will later assume that this series is stationary, but to explain the problem, it is enough to assume that the X_n have the same distribution. Denote by X a generic random function with the same distribution as each X_n. If $E\,\|X\|^2 < \infty$, then the mean $\mu = EX$ and the covariance operator

$$C(x) = E\left[\langle (X - \mu), x\rangle (X - \mu)\right], \quad x \in L^2(\mathcal{T}),$$

exist. The eigenfunctions of C are called the functional principal components and denoted by $v_k, k \geq 1$, i.e. we have $C(v_k) = \lambda_k v_k$. While important modified estimators for special data structures exist, see e.g. [32,14], the FPCs v_k are usually estimated by the empirical FPCs $\hat{v}_k$ defined as the eigenfunctions of the empirical covariance operator

$$\widehat{C}(x) = N^{-1} \sum_{n=1}^{N} \langle (X_n - \bar{X}_N), x\rangle (X_n - \bar{X}_N), \quad x \in L^2(\mathcal{T}),$$

where $\bar{X}_N$ is the usual sample average. The objective of this paper is to establish general conditions, valid for known functional time series models, under which the empirical FPCs $\hat{v}_k$ are asymptotically normal.

Since the v_k and the $\hat{v}_k$ are defined as eigenfunctions, they are determined only up to multiplicative constants. Following the usual practice, we assume that the v_k and the $\hat{v}_k$ form orthonormal systems in $L^2(\mathcal{T})$. This however still leaves a possibility open that, for large N, $\hat{v}_k$ will point in a roughly opposite direction than v_k. This difficulty is usually overcome by introducing the unobservable random signs $\hat{s}_k = \mathrm{sign}(\langle \hat{v}_k, v_k\rangle)$, and establishing the convergence of $\hat{s}_k \hat{v}_k - v_k$. In the following, we assume that $\hat{s}_k = 1$, as adding these signs does not impact our arguments.

Mas [23] considered a functional linear process $X_n = \sum_k \Psi_k(\varepsilon_{n-k})$, in which Ψ_k are bounded linear operators and $\{\varepsilon_n\}$ is a sequence of i.i.d. mean zero random functions in L^2 with $E\,\|\varepsilon_0\|^4 < \infty$. His main objective was to show that the sample autocovariance operators

$$\widehat{C}_h = N^{-1} \sum_{n=1}^{N} \langle X_{n+h}, \cdot\rangle X_n, \quad h = 0, 1, \ldots, H,$$

are jointly asymptotically normal, where H is some fixed lag. As a corollary, he showed that the asymptotic normality of $\widehat{C} = \widehat{C}_0$ implies that $\sqrt{N}(\hat{\lambda}_k - \lambda_k)$ is asymptotically normal. In the course of his proof he also established the asymptotic normality of $\sqrt{N}(\hat{v}_k - v_k)$. He used complex arguments related to those developed by Dauxois et al. [6] and involving Cauchy contours and resolvents. We will present a much simpler argument which is also valid under much weaker assumptions on the stochastic process $\{X_n, \ n = 1, 2, \ldots\}$, and which leads to a direct description of the asymptotic covariance structure of the $\hat{v}_k$. In particular, our assumptions hold for a general class of nonlinear processes known as Bernoulli shifts; see Definition 1.

Using perturbation theory, [24] established the asymptotic normality of the projection operators, as well as the law of large numbers and a large deviation principle, in a setting very similar to ours, that is they derive these results assuming only the corresponding results for the sample covariance operators. The asymptotic normality of the eigenfunctions does not seem

to follow directly from these results. We use a more direct approach that does not rely on the perturbation theory.

We also note that the asymptotic normality of the $\hat{v}_k$ can be deduced from the results of Hall and Hosseini-Nasab [15] who consider higher order asymptotic expansions of the $\hat{v}_k$. These expansions are however established assuming that the X_n are i.i.d. and satisfy a number of technical moment and Lipschitz continuity conditions. Since our goal is only to establish the asymptotic normality, we will not require any of these technical assumptions. Our work is motivated by the fact that many interesting functional data sets form time series; the functions X_n are observed on consecutive days or years, and are generally dependent. Examples of such functional time series are discussed in [18,19].

The remainder of this paper is organized as follows. Section 2 states the assumptions and the main results. In Section 3, we show how the results of Section 2 are applied to yield the asymptotic normality of the $\hat{v}_k$ for important classes of functional time series. The proofs are collected, respectively, in Sections 4 and 5.

2. Main results

We first state a general assumption on the observed stochastic process. The condition that the first p eigenvalues are distinct and positive is needed to uniquely identify the v_k.

Assumption 1. Assume that $\{X_n, \ n = 1, 2, \ldots\}$ is a strictly stationary sequence of random functions in $L^2(\mathcal{T})$ satisfying $E\|X_1\|^4 < \infty$. Furthermore, assume that the first p eigenvalues of the covariance operator C are distinct and ordered as $\lambda_1 > \lambda_2 > \cdots > \lambda_p > \lambda_{p+1} \geq \cdots$.

The main point of our argument is that to establish the asymptotic normality of the empirical FPCs $\hat{v}_k$, it is enough to show that

$$Z_N = N^{1/2}(\widehat{C} - C),$$

is asymptotically normal in the space of Hilbert–Schmidt operators. One may then obtain the asymptotic normality for the $\hat{v}_k$ by verifying that Assumption 2 holds for a specific class of stationary functional time series.

Assumption 2. Assume that there is a mean zero Gaussian Hilbert–Schmidt operator, such that

$$Z_N \xrightarrow{\mathcal{L}} Z.$$

The covariance operator of Z is denoted by Γ.

The operators $C, \widehat{C}, Z$, and Z_N are (random) Hilbert–Schmidt integral operators acting on $L^2(\mathcal{T})$. Recall that if Ψ is such an operator and $\psi(t, s)$ is its kernel, then

$$\Psi(y)(t) = \int_{\mathcal{T}} \psi(t, s) y(s) \, ds, \quad y \in L^2(\mathcal{T}).$$

If Φ is another Hilbert–Schmidt integral operator with kernel $\phi(t, s)$, then their inner product is

$$\langle \Psi, \Phi \rangle_{\mathcal{S}} = \int_{\mathcal{T}} \int_{\mathcal{T}} \psi(t, s) \phi(t, s) \, dt \, ds.$$

The subscript $\mathcal{S}$ on the inner products will typically be dropped as it will be clear what space is involved. The covariance operator Γ acts on the space $\mathcal{S}$ of Hilbert–Schmidt operators acting on $L^2(\mathcal{T})$. It is itself a Hilbert–Schmidt operator.

For any x and y in $L^2(\mathcal{T})$, we define $x \otimes y$ to be the integral operator with kernel $x(t)y(s)$. For $j = 1, \ldots, p$, define

$$T_{j,N} = \sum_{k \neq j} (\lambda_j - \lambda_k)^{-1} \langle Z_N, v_j \otimes v_k \rangle v_k,$$

and

$$T_j = \sum_{k \neq j} (\lambda_j - \lambda_k)^{-1} \langle Z, v_j \otimes v_k \rangle v_k.$$

Under our assumptions the first p eigenvalues are distinct and all eigenvalues are ordered in the decreasing order, so $T_{j,N}$ and T_j are almost surely well defined for $1 \leq j \leq p$; cf. the proof of Theorem 1.

The following results formalize the claims made in Section 1.

Proposition 1. *Under Assumptions 1 and 2, for $j = 1, \ldots, p$,*

$$\|N^{1/2}(\hat{v}_j - v_j) - T_{j,N}\| \xrightarrow{P} 0,$$

where $\|\cdot\|$ is the L^2 norm.

Theorem 1. *Under Assumptions 1 and 2, $T_{j,N} \xrightarrow{\mathcal{L}} T_j$,*

$$N^{1/2}(\hat{v}_j - v_j) \xrightarrow{\mathcal{L}} T_j, \tag{1}$$

and each $\hat{v}_j$ is asymptotically Gaussian with mean v_j and covariance operator

$$\frac{1}{N} \sum_{k_1 \neq j, k_2 \neq j} \frac{\langle \Gamma, (v_j \otimes v_{k_1}) \otimes (v_j \otimes v_{k_2}) \rangle}{(\lambda_j - \lambda_{k_1})(\lambda_j - \lambda_{k_2})} v_{k_1} \otimes v_{k_2}.$$

Furthermore, the $\hat{v}_j$, $1 \leq j \leq p$, are jointly asymptotically Gaussian (in the product space $L^2(\mathcal{T}) \times \cdots \times L^2(\mathcal{T})$) with cross covariance operators (for $j_1 = 1, \ldots, p$ and $j_2 = 1, \ldots, p$)

$$\frac{1}{N} \sum_{k_1 \neq j_1, k_2 \neq j_2} \frac{\langle \Gamma, (v_{j_1} \otimes v_{k_1}) \otimes (v_{j_2} \otimes v_{k_2}) \rangle}{(\lambda_{j_1} - \lambda_{k_1})(\lambda_{j_2} - \lambda_{k_2})} v_{k_1} \otimes v_{k_2}.$$

Corollary 1. *If Assumptions 1 and 2 hold and the observations are i.i.d. Gaussian random functions then the covariance operator in Corollary 1 simplifies to*

$$\frac{1}{N} \sum_{k \neq j} \frac{\lambda_j \lambda_k}{(\lambda_j - \lambda_k)^2} v_k \otimes v_k,$$

and the cross covariance operators simplify to

$$-\frac{1}{N} \frac{\lambda_{j_1} \lambda_{j_2}}{(\lambda_{j_1} - \lambda_{j_2})^2} v_{j_2} \otimes v_{j_1}.$$

The formulas in Corollary 1 agree with those known in the multivariate setting, see [22, Section 3 p. 48], and in the i.i.d. functional setting, see [6, p. 149]. We remark that the normality in Corollary 1 is not essential. What is needed is that the projections are independent of each other (not simply uncorrelated), which is at least guaranteed in the normal setting. Observe finally that

the sign of $\hat{v}_j$ plays no role in its asymptotic covariance function, but the signs of $\hat{v}_{j_1}$ and $\hat{v}_{j_2}$ play a role in the asymptotic cross covariance functions.

Theorem 1 and Corollary 1 specify the covariances of the limits T_j. In statistical applications, it is useful to know that the covariances of the differences $N^{1/2}(\hat{v}_j - v_j)$ converge to the covariances of these limits. In time series analysis of linear models, see e.g. [1], convergence of the sample covariance matrices is established by imposing summability conditions on the impulse response coefficients. More general results can be obtained by imposing cumulant conditions. It is well-known that the convergence of moments is much more subtle for nonlinear time series; see [7,27], among others. In the remainder of this section, we show that the covariances of the $N^{1/2}(\hat{v}_j - v_j)$ converge under a very simple condition related to Assumption 2. In Section 3, we provide a stronger result for a special class of nonlinear functional time series.

Assumption 2 implies that $Z_N \otimes Z_N \xrightarrow{\mathcal{L}} Z \otimes Z$. Note that $Z_N \otimes Z_N$ is the tensor product of operators. We will not distinguish between tensor products of functions and operators as it will always be clear from the setting which case we are dealing with. Note that the tensor product of two Hilbert–Schmidt operators is now an operator acting on the space of Hilbert–Schmidt operators defined analogously to the tensor product of two functions. Our objective then is to show that $E(Z_N \otimes Z_N) \to E(Z \otimes Z)$. This is achieved by employing Lemma 3 of Section 4 which is a general result useful for determining the convergence of moments in a functional sequence. In light of Lemma 3, perhaps the most direct way of guaranteeing the convergence of moments is stated in the following theorem.

Theorem 2. *Suppose Assumptions 1 and 2 hold and,* $\sup_N E\|Z_N\|^{2+\epsilon} < \infty$*, for some $\varepsilon > 0$.* *Then*

$$E(Z_N \otimes Z_N) \to E(Z \otimes Z).$$

If, in addition, $E\|Z_N \otimes Z_N\| \to E\|Z \otimes Z\|$, then

$$N E((\hat{v}_j - v_j) \otimes (\hat{v}_k - v_k)) \to E(T_j \otimes T_k).$$

3. Applications

The main point of Section 2 is that in order to establish the asymptotic normality of the empirical FPCs $\hat{v}_k$, it is enough to verify that Assumption 2 holds. It thus becomes a separate question to find out for what classes of functional stochastic processes Assumption 2 holds. We could not find ready answers to this question. The results in Chapter 4 of [4] are concerned with the asymptotic behavior of the Hilbert–Schmidt norm of $\widehat{C} - C$. However his representation (4.19), i.e. $\widehat{C} - C = N^{-1}\{Y_1 + Y_2 + \cdots, Y_N\}$, where $\{Y_n\}$ is an AR(1) process in $\mathcal{S}$, together with his Theorem 3.10, which establishes the asymptotic normality of the average of an AR(1) process defined in any Hilbert space, allows us to conclude that Assumption 2 holds for the functional AR(1) processes (with i.i.d. innovations).

In this section, we show that Assumption 2 holds for weakly dependent functional time series defined as Bernoulli shifts of i.i.d. elements in an abstract measurable space. The weak dependence is quantified by the condition known as L^4-m-approximability. Before stating the definition, we emphasize that all stationary functional processes used in practice can be expressed as Bernoulli shifts and are L^4-m-approximable, assuming their parameters are in suitably defined ranges. Hörmann and Kokoszka [17] and Hörmann et al. [16] discuss examples which include bilinear and ARCH-type functional processes.

For $p \geq 1$, we denote by $L_H^p = L_H^p(\Omega, \mathcal{A}, P)$ the space of $H = L^2(\mathcal{T})$ valued random functions X such that

$$v_p(X) = \left(E\|X\|^p\right)^{1/p} = \left(E\left\{\int X^2(t)dt\right\}^{p/2}\right)^{1/p} < \infty.$$

We use H to denote the function space $L^2(\mathcal{T})$ to lighten the notation.

Definition 1. A sequence $\{X_n\} \in L_H^p$ is called L^p-*m-approximable* if each X_n admits the representation

$$X_n = f(\varepsilon_n, \varepsilon_{n-1}, \ldots), \tag{2}$$

where the ε_i are i.i.d. elements taking values in a measurable space S, and f is a measurable function $f : S^\infty \to H$. Moreover we assume that if $\{\varepsilon_i'\}$ is an independent copy of $\{\varepsilon_i\}$ defined on the same probability space, then letting

$$X_n^{(m)} = f(\varepsilon_n, \varepsilon_{n-1}, \ldots, \varepsilon_{n-m+1}, \varepsilon_{n-m}', \varepsilon_{n-m-1}', \ldots) \tag{3}$$

we have

$$\sum_{m=1}^\infty v_p\left(X_n - X_n^{(m)}\right) < \infty. \tag{4}$$

Definition 1 implies that $\{X_n\}$ is strictly stationary. It is clear from the representation of X_n and $X_n^{(m)}$ that $E\|X_m - X_m^{(m)}\|^p = E\|X_1 - X_1^{(m)}\|^p$, so that condition (4) could be formulated solely in terms of X_1 and the approximations $X_1^{(m)}$. Obviously the sequence $\{X_n^{(m)}, n \in \mathbb{Z}\}$ as defined in (3) is *not* m-dependent. To this end we need to define for each n an independent copy $\{\varepsilon_k^{(n)}\}$ of $\{\varepsilon_k\}$ (this can always be achieved by enlarging the probability space) which is then used instead of $\{\varepsilon_k'\}$ to construct $X_n^{(m)}$, i.e. we set

$$X_n^{(m)} = f(\varepsilon_n, \varepsilon_{n-1}, \ldots, \varepsilon_{n-m+1}, \varepsilon_{n-m}^{(n)}, \varepsilon_{n-m-1}^{(n)}, \ldots). \tag{5}$$

Since this modification leaves condition (4) unchanged, we will assume from now on that the $X_n^{(m)}$ are defined by (5). Then, for each $m \geq 1$, the sequences $\{X_n^{(m)}, n \in \mathbb{Z}\}$ are strictly stationary and m-dependent, and each $X_n^{(m)}$ is equal in distribution to X_n.

Definitions of dependence based on representation (2) have been used in "non-functional" contexts by many researchers, a number of references are given in Chapter 16 of [19]. In particular, the notion of physical dependence of Wu [30,31] is closely related (condition (4) is replaced by a different condition). There is no obvious connection between L^p-m-approximability and mixing conditions or the weak dependence of Doukhan and Louhichi [9].

Theorem 3. *Assumption 2 holds for every L^4-m-approximable sequence in $L^2(\mathcal{T})$. Furthermore, the covariance operator Γ is given by*

$$\Gamma = E[(X_1 \otimes X_1 - C) \otimes (X_1 \otimes X_1 - C)] + 2\sum_{i=2}^\infty [(X_1 \otimes X_1 - C) \otimes (X_i \otimes X_i - C)]$$

and satisfies

$$\Gamma = \lim_{N \to \infty} E[Z_N \otimes Z_N]. \tag{6}$$

Corollary 2. *Suppose* $\{X_n\}$ *is an* L^4-m-*approximable sequence in* $L^2(\mathcal{T})$ *such that the first* p *eigenvalues of its covariance operator are distinct and positive. Then there exist normal elements* T_j, $1 \le j \le p$, *of* $L^2(\mathcal{T})$ *such that*

$$\left\{ N^{1/2}(\hat{v}_j - v_j), \ 1 \le j \le p \right\} \overset{\mathcal{L}}{\to} \left\{ T_j, \ 1 \le j \le p \right\}. \tag{7}$$

Consider a functional linear process

$$X_n = \sum_{j=0}^{\infty} \Psi_j(\varepsilon_{n-j}), \tag{8}$$

where the Ψ_j are bounded linear operators acting in $L^2(\mathcal{T})$, and the ε_j are i.i.d. random elements of $L^2(\mathcal{T})$ with $E \|\varepsilon_0\|^2 < \infty$. A sufficient condition for the existence of such a process is $\sum_{j=0}^{\infty} \|\Psi_j\|_{\mathcal{L}}^2 < \infty$, where the subscript $\mathcal{L}$ indicates the operator norm (recall that $\|\Psi\|_{\mathcal{L}} \le \|\Psi\|_{\mathcal{S}}$). A direct verification shows that the linear process (8) is L^p-m-approximable for any $p \ge 2$ if $\sum_{j=1}^{\infty} j \|\Psi_j\|_{\mathcal{L}} < \infty$. This condition clearly holds for the functional autoregressive process with the autoregressive operator Ψ because in that case $\Psi_j = \Psi^j$, and $\|\Psi^j\|_{\mathcal{L}} \le ab^j$, for some $a > 0$ and $b < 1$; see Chapter 3 of [4] or Chapter 13 of [19]. The above discussion leads to the following corollary.

Corollary 3. *Suppose* $\{X_n\}$ *is a functional* AR(1) *process defined by the equations* $X_n = \Psi(X_{n-1}) + \varepsilon_n$ *with i.i.d.* ε_j *satisfying* $E \|\varepsilon_0\|^4 < \infty$. *If the first* p *eigenvalues of the covariance operator of* X_0 *are different and positive, then relation* (7) *holds.*

Similar corollaries can be stated for other functional models for which L^4-m-approximability has been verified. The point of Corollary 2 is however that asymptotic normality of the $\hat{v}_k$ holds for very general weakly dependent functional time series, whether linear or nonlinear.

We note however that the condition $\sum_{j=1}^{\infty} j \|\Psi_j\|_{\mathcal{L}} < \infty$ is more restrictive than the condition $\sum_{j=1}^{\infty} \|\Psi_j\|_{\mathcal{L}} < \infty$ considered by Mas [23]. Thus we cannot conclude his results from Corollary 2. The proof of Theorem 3 relies on the fact that the projections $\langle X_n, h \rangle$ of an L^4-m-approximable functional sequence form an L^4-m-approximable scalar sequence. While very strong results exist for scalar linear sequences, see [28], we cannot use them directly because if $\{X_n\}$ is linear, $\{\langle X_n, h \rangle\}$ generally is not a linear process with scalar innovations.

The central limit theorem in Assumption 2 can be established under a number of other dependence structures and there is a large literature on this topic. As stating these types of results would involve a significant introduction and a good deal of extra notation, we restrict ourselves to providing relevant references. To verify Assumption 2 under mixing conditions, the work of Dehling [8], Merlevéde et al. [26], or Merlevéde [25] could be used. Conditions related to the martingale property could also be used; see [21,5]. We focused on L^4-m-approximability because it has already been used in the functional context, in addition to Hörmann and Kokoszka [17]; see also [12,20]. Time series models are typically defined by structural equations which involve i.i.d. innovations ε_i, and these equations admit solutions as Bernoulli shifts. Using moment conditions to quantify dependence is also more convenient than using mixing conditions, because moment conditions are easier to verify for specific models; L^4-m-approximability can be viewed as broadly analogous to cumulant conditions.

Using L^4-m-approximability, the extra ε moment condition in Theorem 2 can be dropped. This has been established in the course of the proof of Theorem 3, but we state this result for completeness.

Theorem 4. *Suppose Assumptions* 1 *and* 2 *hold. If* $\{X_n\}$ *is an* L^4-m *approximable sequence, then* $E(Z_N \otimes Z_N) \to E(Z \otimes Z)$.

4. Proofs of the results of Section 2

We begin with presenting some identities which will be used throughout the proofs without further mention. Recall the definition of an integral Hilbert–Schmidt operator given in Section 2. If Ψ is a symmetric integral Hilbert–Schmidt operator, then

$$\langle \Psi(x), y \rangle = \int \psi(t, s) x(s) y(t) \, ds \, dt = \langle \Psi, y \otimes x \rangle_{\mathcal{S}} = \langle \Psi, x \otimes y \rangle_{\mathcal{S}}.$$

The subscript $\mathcal{S}$ on the inner products will often be dropped. For functions x_1, x_2, x_3, and x_4 in $L^2(\mathcal{T})$ we will make heavy use of relations such as

$$\langle \Gamma(x_1 \otimes x_2), x_3 \otimes x_4 \rangle = \iiiint \Gamma(t_3, t_4, t_1, t_2) x_1(t_1) x_2(t_2) x_3(t_3) x_4(t_4) \, dt_1 \, dt_2 \, dt_3 \, dt_4$$
$$= \langle \Gamma, (x_3 \otimes x_4) \otimes (x_1 \otimes x_2) \rangle.$$

Note that $(x_3 \otimes x_4) \otimes (x_1 \otimes x_2)$ indicates the tensor product of two Hilbert–Schmidt operators $x_3 \otimes x_4$ and $x_1 \otimes x_2$, and is therefore an operator acting on the space of Hilbert–Schmidt operators. Again, we do not notationally distinguish between tensor products of functions and operators as it will always be contextually clear which we are working with. If Γ is a covariance operator, then it is symmetric and we can rearrange terms to conclude

$$\langle \Gamma(x_1 \otimes x_2), x_3 \otimes x_4 \rangle = \langle \Gamma, (x_1 \otimes x_2) \otimes (x_3 \otimes x_4) \rangle.$$

We now list a few lemmas which are used in the proof of Proposition 1. Lemma 1 is proven by algebraic manipulations, but it leads to a simple general argument that allows to establish the claim. Lemma 2 is well-known, and is given for ease of reference. Throughout this section, we suppose that Assumptions 1 and 2 hold.

Lemma 1. *For any* $j = 1, 2, \ldots,$

$$\langle \hat{v}_j - v_j, v_j \rangle = -\frac{1}{2} \| \hat{v}_j - v_j \|^2.$$

For any $j = 1, 2, \ldots, k = 1, 2, \ldots$ *and* $k \neq j$,

$$\langle \hat{v}_j - v_j, v_k \rangle = \frac{N^{-1/2}}{\hat{\lambda}_j - \lambda_k} \langle Z_N, \hat{v}_j \otimes v_k \rangle,$$

as long as $\hat{\lambda}_j \neq \lambda_k$.

Proof. Expressing the norm using inner products we have that

$$2\langle \hat{v}_j - v_j, v_j \rangle + \| \hat{v}_j - v_j \|^2 = \langle \hat{v}_j - v_j, 2v_j \rangle + \langle \hat{v}_j - v_j, \hat{v}_j - v_j \rangle$$
$$= \langle \hat{v}_j - v_j, v_j + \hat{v}_j \rangle$$
$$= \langle \hat{v}_j, v_j \rangle - \langle v_j, v_j \rangle + \langle \hat{v}_j, \hat{v}_j \rangle - \langle v_j, \hat{v}_j \rangle.$$

The above is zero since the inner product is symmetric and $\hat{v}_j$ and v_j have norm 1.

By adding and subtracting terms we can rewrite

$$\hat{\lambda}_j(\hat{v}_j - v_j) = C(\hat{v}_j - v_j) + \widehat{C}(\hat{v}_j) - C(\hat{v}_j) - (\hat{\lambda}_j - \lambda_j)v_j.$$

Taking the inner product of both sides with respect to v_k we have

$$\hat{\lambda}_j\langle\hat{v}_j - v_j, v_k\rangle = \langle C, (\hat{v}_j - v_j) \otimes v_k\rangle + \langle\widehat{C}, \hat{v}_j \otimes v_k\rangle - \langle C, \hat{v}_j \otimes v_k\rangle$$
$$= \lambda_k\langle\hat{v}_j - v_j, v_k\rangle + \langle\widehat{C} - C, \hat{v}_j \otimes v_k\rangle.$$

Which implies that

$$\langle\hat{v}_j - v_j, v_k\rangle = (\hat{\lambda}_j - \lambda_k)^{-1}\langle\widehat{C} - C, \hat{v}_j \otimes v_k\rangle = \frac{N^{-1/2}}{\hat{\lambda}_j - \lambda_k}\langle Z_N, \hat{v}_j \otimes v_k\rangle. \quad \square$$

Lemma 2. *For any $1 \leq j \leq p$, under Assumptions 1 and 2 we have*

$$\|C - \widehat{C}\| = O_P(N^{1/2}), \qquad |\hat{\lambda}_j - \lambda_j| = O_P(N^{1/2}), \quad and \quad \|\hat{v}_j - v_j\| = O_P(N^{1/2}).$$

Proof. The first property follows by Assumption 2. The next two claims follow since

$$|\hat{\lambda}_j - \lambda_j| \leq \|C - \widehat{C}\| \quad \text{and} \quad \|\hat{v}_j - v_j\| \leq \frac{2\sqrt{2}}{\alpha_j}\|C - \widehat{C}\|, \tag{9}$$

where $\alpha_j = \min\{\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}\}$ (with $\alpha_1 = \lambda_1 - \lambda_2$). Inequalities (9) are proven in Chapter 2 of [19]. (They follow from more general results proved in [10,13], and used by many authors including [4,23].) $\square$

Proof of Proposition 1. Since $v_1, v_2, \ldots$ is an orthonormal basis of $L^2(\mathcal{T})$, we have by Parseval's identity

$$\|N^{1/2}(\hat{v}_j - v_j) - T_{j,N}\|^2 = \sum_k\langle N^{1/2}(\hat{v}_j - v_j) - T_{j,N}, v_k\rangle^2.$$

So we consider the terms

$$\langle N^{1/2}(\hat{v}_j - v_j) - T_{j,N}, v_k\rangle,$$

for every $k = 1, 2, \ldots$. The arguments that follow do not depend on the index j, so, without loss of generality, we take $j = 1$.

When $k = 1$ we have that

$$\langle T_{1,N}, v_1\rangle = \sum_{l \geq 2}(\lambda_1 - \lambda_l)^{-1}\langle Z_N, v_1 \otimes v_l\rangle\langle v_l, v_1\rangle = 0.$$

Consequently,

$$\langle N^{1/2}(\hat{v}_1 - v_1) - T_{1,N}, v_1\rangle = N^{1/2}\langle\hat{v}_1 - v_1, v_1\rangle.$$

Applying Lemma 1, and then Lemma 2, we have

$$N^{1/2}\langle\hat{v}_1 - v_1, v_1\rangle = -N^{1/2}\|\hat{v}_1 - v_1\|^2 \xrightarrow{P} 0.$$

When $k \geq 2$

$$\langle T_{1,N}, v_k\rangle = \sum_{l \geq 2}(\lambda_1 - \lambda_l)^{-1}\langle Z_N, v_1 \otimes v_l\rangle\langle v_l, v_k\rangle = (\lambda_1 - \lambda_k)^{-1}\langle Z_N, v_1 \otimes v_k\rangle.$$

By Lemma 1,

$$N^{1/2}\langle \hat{v}_1 - v_1, v_k \rangle - \langle T_{1,N}, v_k \rangle = \frac{1}{\hat{\lambda}_1 - \lambda_k}\langle Z_N, \hat{v}_1 \otimes v_k \rangle - \frac{1}{\lambda_1 - \lambda_k}\langle Z_N, v_1 \otimes v_k \rangle.$$

Putting both terms under the same denominator, we obtain

$$\frac{1}{\hat{\lambda}_1 - \lambda_k}\langle Z_N, \hat{v}_1 \otimes v_k \rangle - \frac{1}{\lambda_1 - \lambda_k}\langle Z_N, v_1 \otimes v_k \rangle$$

$$= \frac{(\lambda_1 - \lambda_k)\langle Z_N, \hat{v}_1 \otimes v_k \rangle - (\hat{\lambda}_1 - \lambda_k)\langle Z_N, v_1 \otimes v_k \rangle}{(\hat{\lambda}_1 - \lambda_k)(\lambda_1 - \lambda_k)}.$$

Reordering the numerator we have

$$(\lambda_1 - \lambda_k)\langle Z_N, \hat{v}_1 \otimes v_k \rangle - (\hat{\lambda}_1 - \lambda_k)\langle Z_N, v_1 \otimes v_k \rangle$$

$$= (\lambda_1 - \lambda_k)\langle Z_N(\hat{v}_1), v_k \rangle - (\hat{\lambda}_1 - \lambda_k)\langle Z_N(v_1), v_k \rangle$$

$$= \langle (\lambda_1 - \lambda_k)Z_N(\hat{v}_1) - (\hat{\lambda}_1 - \lambda_k)Z_N(v_1), v_k \rangle$$

$$= \langle (\lambda_1 - \lambda_k)Z_N(\hat{v}_1 - v_1) + (\lambda_1 - \lambda_k)Z_N(v_1) - (\hat{\lambda}_1 - \lambda_k)Z_N(v_1), v_k \rangle$$

$$= \langle (\lambda_1 - \lambda_k)Z_N(\hat{v}_1 - v_1) + (\lambda_1 - \hat{\lambda}_1)Z_N(v_1), v_k \rangle.$$

Now, notice that

$$\sum_{k \geq 2}\langle N^{1/2}(\hat{v}_j - v_j) - T_{j,N}, v_k \rangle^2$$

$$= \sum_{k \geq 2}\frac{\langle (\lambda_1 - \lambda_k)Z_N(\hat{v}_1 - v_1) + (\lambda_1 - \hat{\lambda}_1)Z_N(v_1), v_k \rangle^2}{(\hat{\lambda}_1 - \lambda_k)^2(\lambda_1 - \lambda_k)^2}.$$

It is convenient to split the above sum as follows:

$$\sum_{k \geq 2}\frac{\langle Z_N(\hat{v}_1 - v_1), v_k \rangle^2}{(\hat{\lambda}_1 - \lambda_k)^2} \tag{10}$$

$$+ \sum_{k \geq 2}\frac{2(\lambda_1 - \hat{\lambda}_1)\langle Z_N(\hat{v}_1 - v_1), v_k \rangle\langle Z_N(v_1), v_k \rangle}{(\hat{\lambda}_1 - \lambda_k)^2(\lambda_1 - \lambda_k)} \tag{11}$$

$$+ \sum_{k \geq 2}\frac{(\lambda_1 - \hat{\lambda}_1)^2\langle Z_N(v_1), v_k \rangle^2}{(\hat{\lambda}_1 - \lambda_k)^2(\lambda_1 - \lambda_k)^2}. \tag{12}$$

Considering (10), for N large, $(\hat{\lambda}_1 - \lambda_k)^2$ is the smallest when $k = 2$. Note that when $j \neq 1$ the maximum will be attained at either λ_{j-1} or λ_{j+1}. So, by Parseval's identity, for N large,

$$\sum_{k \geq 2}\frac{\langle Z_N(\hat{v}_1 - v_1), v_k \rangle^2}{(\hat{\lambda}_1 - \lambda_k)^2} \leq \frac{\|Z_N(v_1 - \hat{v}_1)\|^2}{(\hat{\lambda}_1 - \lambda_2)^2} \leq \frac{\|Z_N\|^2\|v_1 - \hat{v}_1\|^2}{(\hat{\lambda}_1 - \lambda_2)^2} \overset{P}{\to} 0,$$

by Lemma 2.

Turning to (11) we have by Parseval's identity and the Cauchy–Schwarz inequality, for N large,

$$\left| \sum_{k \geq 2}\frac{2(\lambda_1 - \hat{\lambda}_1)\langle Z_N(\hat{v}_1 - v_1), v_k \rangle\langle Z_N(v_1), v_k \rangle}{(\hat{\lambda}_1 - \lambda_k)^2(\lambda_1 - \lambda_k)} \right|$$

$$\leq \left| \frac{2(\lambda_1 - \hat{\lambda}_1)}{(\hat{\lambda}_1 - \lambda_2)^2 (\lambda_1 - \lambda_2)} \right| \|Z_N(\hat{v}_1 - v_1)\| \|Z_N(v_1)\|$$

$$\leq \left| \frac{2(\lambda_1 - \hat{\lambda}_1)}{(\hat{\lambda}_1 - \lambda_2)^2 (\lambda_1 - \lambda_2)} \right| \|\hat{v}_1 - v_1\| \|Z_N\|^2 \xrightarrow{P} 0,$$

by Lemma 2. Finally, considering (12) we have by Parseval's identity, for N large,

$$\sum_{k \geq 2} \frac{(\lambda_1 - \hat{\lambda}_1)^2 \langle Z_N(v_1), v_k \rangle^2}{(\hat{\lambda}_1 - \lambda_k)^2 (\lambda_1 - \lambda_k)^2} \leq \frac{(\lambda_1 - \hat{\lambda}_1)^2}{(\hat{\lambda}_1 - \lambda_2)^2 (\lambda_1 - \lambda_2)^2} \|Z_N(v_1)\|^2$$

$$\leq \frac{(\lambda_1 - \hat{\lambda}_1)^2}{(\hat{\lambda}_1 - \lambda_2)^2 (\lambda_1 - \lambda_2)^2} \|Z_N\|^2 \xrightarrow{P} 0.$$

Therefore we can conclude that

$$\|N^{1/2}(\hat{v}_j - v_j) - T_{j,N}\|^2 = \langle N^{1/2}(\hat{v}_j - v_j) - T_{j,N}, v_1 \rangle^2$$
$$+ \sum_{k \geq 2} \langle N^{1/2}(\hat{v}_j - v_j) - T_{j,N}, v_k \rangle^2 \xrightarrow{P} 0. \quad \square$$

Proof of Theorem 1. For $j = 1, 2, \ldots, p$, consider a function g_j which maps a Hilbert–Schmidt operator Ψ (acting on $L^2(\mathcal{T})$) to an element of $L^2(\mathcal{T})$, defined as

$$g_j(\Psi) = \sum_{k \neq j} (\lambda_j - \lambda_k)^{-1} \langle \Psi, v_j \otimes v_k \rangle v_k.$$

To see that $g_j(\Psi)$ is an element of $L^2(\mathcal{T})$, observe that

$$\|g_j(\Psi)\|^2 = \sum_{k \neq j} (\lambda_j - \lambda_k)^{-2} \langle \Psi, v_j \otimes v_k \rangle^2 \leq \frac{1}{\alpha_j^2} \sum_{k \neq j} \langle \Psi, v_j \otimes v_k \rangle^2 \leq \frac{\|\Psi\|}{\alpha_j^2},$$

where $\alpha_1 = \lambda_1 - \lambda_2$ and $\alpha_j = \min\{\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}\}$, for $j \geq 2$. The above bound also shows that g_j is a bounded operator. Since g_j is linear, it is also continuous. By the continuous mapping theorem, see e.g. [3], we can conclude from Assumption 2 that

$$T_{j,N} = g_j(Z_N) \xrightarrow{\mathcal{L}} g_j(Z) = T_j.$$

We now compute the covariance operator of T_j:

$$E[T_j \otimes T_j] = E\left[\left(\sum_{k_1 \neq j} (\lambda_j - \lambda_{k_1})^{-1} \langle Z, v_j \otimes v_{k_1} \rangle v_{k_1} \right) \right.$$
$$\left. \otimes \left(\sum_{k_2 \neq j} (\lambda_j - \lambda_{k_2})^{-1} \langle Z, v_j \otimes v_{k_2} \rangle v_{k_2} \right) \right]$$

$$= \sum_{k_1 \neq j, k_2 \neq j} \frac{\langle E[Z \otimes Z], (v_j \otimes v_{k_1}) \otimes (v_j \otimes v_{k_2}) \rangle}{(\lambda_j - \lambda_{k_1})(\lambda_j - \lambda_{k_2})} v_{k_1} \otimes v_{k_2}$$

$$= \sum_{k_1 \neq j, k_2 \neq j} \frac{\langle \Gamma, (v_j \otimes v_{k_1}) \otimes (v_j \otimes v_{k_2}) \rangle}{(\lambda_j - \lambda_{k_1})(\lambda_j - \lambda_{k_2})} v_{k_1} \otimes v_{k_2}.$$

This verifies the corollary. $\quad \square$

Proof of Corollary 2. If the functions are i.i.d., then

$$\Gamma = E[(X_1 \otimes X_1) \otimes (X_1 \otimes X_1)] - C \otimes C.$$

This is established by direct verification; see also Section 2.6 p. 33 of [19]. Therefore, for $j \neq k_1$ and $j \neq k_2$

$$\langle \Gamma, (v_j \otimes v_{k_1}) \otimes (v_j \otimes v_{k_1}) \rangle = E[\langle X_1, v_j \rangle \langle X_1, v_{k_1} \rangle \langle X_1, v_j \rangle \langle X_1, v_{k_2} \rangle] \\ - \langle C, v_j \otimes v_{k_1} \rangle \langle C, v_j \otimes v_{k_2} \rangle.$$

Since $j \neq k_1$ and $j \neq k_2$, we have that

$$\langle C, v_j \otimes v_{k_1} \rangle \langle C, v_j \otimes v_{k_2} \rangle = 0.$$

Since the functions are assumed to be Gaussian, we have that

$$E[\langle X_1, v_j \rangle \langle X_1, v_{k_1} \rangle \langle X_1, v_j \rangle \langle X_1, v_{k_2} \rangle] = \lambda_j \lambda_{k_1} 1_{k_1 = k_2},$$

and implies that

$$\langle \Gamma, (v_j \otimes v_{k_1}) \otimes (v_j \otimes v_{k_1}) \rangle = \lambda_j \lambda_{k_1} 1_{k_1 = k_2}.$$

For the cross covariance operators, since $j_1 \neq j_2$, $j_1 \neq k_1$ and $j_2 \neq k_2$, we have that

$$E[\langle X_1, v_{j_1} \rangle \langle X_1, v_{k_1} \rangle \langle X_1, v_{j_2} \rangle \langle X_1, v_{k_2} \rangle] = \lambda_{j_1} \lambda_{j_2} 1_{j_1 = k_2} 1_{j_2 = k_1},$$

which, upon plugging into the formulas given in Corollary 1, completes the proof.   □

The following lemma is useful for determining the convergence of moments in a functional sequence.

Lemma 3. *Let $\{X_i\}$, X, be random elements of a Hilbert space $\mathbb{H}$ and let $\{Y_i\}$, Y be real valued random variables.*

1. *Suppose that $X_i \overset{\mathcal{L}}{\to} X$ and $E\|X_i\| \to E\|X\|$, then $EX_i \to EX$.*
2. *(Sandwich Theorem) Suppose $\|X_i\| \leq |Y_i|$, $X_i \overset{\mathcal{L}}{\to} X$, $Y_i \overset{\mathcal{L}}{\to} Y$, and $E|Y_i| \to E|Y|$, then $EX_i \to EX$.*
3. *Suppose $X_i \overset{\mathcal{L}}{\to} X$ and $\sup E\|X_i\|^{1+\varepsilon} < \infty$, for some $\varepsilon > 0$, then $EX_i \to EX$.*

Proof. The key to this lemma is part (1) which relates the convergence of moments of the X_i to those of the $\|X_i\|$, and thus allows us to apply univariate results. So, starting with (1), assume that $X_i \overset{\mathcal{L}}{\to} X$ and $E\|X_i\| \to E\|X\|$. Since $E\|X_i\| \to E\|X\|$, we have that the $\|X_i\|$ are uniformly integrable, Billingsley [3, Theorem 3.6]. So, for $\alpha > 0$, we split the expected values as follows

$$\|EX_i - EX\| = \|E(X_i 1_{\|X_i\| \leq \alpha}) - E(X 1_{\|X\| \leq \alpha}) + E(X_i 1_{\|X_i\| > \alpha}) - E(X 1_{\|X\| > \alpha})\| \\ \leq \|E(X_i 1_{\|X_i\| \leq \alpha}) - E(X 1_{\|X\| \leq \alpha})\| + \|E(X_i 1_{\|X_i\| > \alpha}) - E(X 1_{\|X\| > \alpha})\| \\ \leq \|E(X_i 1_{\|X_i\| \leq \alpha}) - E(X 1_{\|X\| \leq \alpha})\| + E(\|X_i\| 1_{\|X_i\| > \alpha}) + E(\|X\| 1_{\|X\| > \alpha}).$$

Now, fix an $\varepsilon > 0$. Using the uniform integrability of the $\|X_i\|$, we can choose α such that

$$E(\|X_i\| 1_{\|X_i\| > \alpha}) + E(\|X\| 1_{\|X\| > \alpha}) < \frac{\varepsilon}{2},$$

uniformly in i. Using the dominated convergence theorem and Dudley–Skorohod embedding theorem for random functions, Bosq [4, pp. 29 and 46], we can find I, such that for all $i > I$,

$$\|E(X_i 1_{\|X_i\| \leq \alpha}) - E(X 1_{\|X\| \leq \alpha})\| < \frac{\varepsilon}{2}.$$

Since $\|EX_i - EX\|$ does not depend on α, we conclude that for all $i > I$,

$$\|EX_i - EX\| < \varepsilon,$$

which completes the proof of (1).

If the assumptions of part (2) hold, then the $\|X_i\|$ are uniformly integrable, so $E\|X_i\| \to E\|X_i\|$. Part (2) thus follows from part (1).

The moment condition in part (3) also implies the uniform integrability of the $\|X_i\|$, see e.g. [3, pp. 31–32], so part (3) also follows from part (1).   □

Proof of Theorem 2. Under Assumption 2, we have that $Z_N \otimes Z_N \overset{\mathcal{L}}{\to} Z \otimes Z$, so, by Lemma 3-3, we will have $E[Z_N \otimes Z_N] \to E[Z \otimes Z]$, if we can verify the uniform moment condition. Observe that

$$\|Z_N \otimes Z_N\| = \|Z_N\|^2,$$

which implies

$$E\|Z_N \otimes Z_N\|^{1+(1/2)\varepsilon} = E\|Z_N\|^{2+\varepsilon} < \infty,$$

which completes the proof of the first claim.

To prove the second claim, we apply Lemma 3-2. By (1), define

$$X_N := N(\hat{v}_j - v_j) \otimes (\hat{v}_k - v_k) \overset{\mathcal{L}}{\to} T_j \otimes T_k =: X.$$

By Assumption 2, we have $Y_N \overset{\mathcal{L}}{\to} Y$ and $EY_N \to EY$, with $Y_N = \|Z_N \otimes Z_N\|$ and $Y = \|Z \otimes Z\|$. Using (9), we obtain

$$\|N(\hat{v}_j - v_j) \otimes (\hat{v}_k - v_k)\| \le \frac{8}{\alpha_j \alpha_k} N\|\hat{C} - C\|^2 = \frac{8}{\alpha_j \alpha_k}\|Z_N \otimes Z_N\|.$$

The claim thus follows from Lemma 3-2.   □

5. Proofs of the results of Section 3

We begin this section with the following lemma which is a direct consequence of a much more general functional CLT for random vectors established in [2].

Lemma 4. *Suppose $\{\xi_n\}$ is an L^2-m-approximable sequence of mean zero scalar random variables. Then, the sequence*

$$\gamma = \sum_{h=-\infty}^{\infty} E[\xi_0 \xi_h]$$

converges absolutely, and

$$N^{-1/2} \sum_{n=1}^{N} \xi_n \overset{\mathcal{L}}{\to} W,$$

where W is a mean zero normal random variable with variance γ.

Next we give a result from [5] that provides a very useful form of the central limit theorem for stationary processes. Combining their Lemma 3.1 and Remark 3.3, we obtain the following lemma.

Lemma 5. *Let* $\{Y_{iN} : 1 \leq i \leq N, 1 \leq N \leq \infty\}$ *be an array of random elements of a Hilbert space* $\mathbb{H}$, *strictly stationary for each N, and with* $EY_{iN} = 0$ *and* $E\|Y_{iN}\|^2 < \infty$. *If*

1. *for every* $h \in \mathbb{H}$ *we have* $\langle h, \sum_i Y_{iN} \rangle \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma(h))$, *where* $\sigma(h)$ *is a positive number depending on h, and*
2. $\lim_{N \to \infty} E\| \sum_i Y_{iN} \|^2$ *exists and is finite,*

then, there exists a covariance operator Γ *such that*

$$\sum_{i=1}^{N} Y_{iN} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma).$$

The first condition of Lemma 5 simply checks that the projections are asymptotically normal, while the second is a useful tightness condition for stationary Hilbert space valued processes.

Proof of Theorem 3. We start by showing that if the sequence $\{X_i\}$ is L^4-m-approximable then $\{X_i \otimes X_i\}$ is L^2-m-approximable. Let $X_i^{(m)}$ be the m-dependent approximation to X_i defined by (5). Direct verification shows that

$$\|X_i \otimes X_i - X_i^{(m)} \otimes X_i^{(m)}\| \leq \sqrt{2} \left(\|X_i\| \|X_i - X_i^{(m)}\| + \|X_i^{(m)}\| \|X_i - X_i^{(m)}\| \right).$$

Therefore,

$$
\begin{aligned}
E\|X_i \otimes X_i - X_i^{(m)} \otimes X_i^{(m)}\|^2 &\leq 2E \left(\|X_i\| \|X_i - X_i^{(m)}\| + \|X_i^{(m)}\| \|X_i - X_i^{(m)}\| \right)^2 \\
&= 2 \left(E\|X_i\|^2 \|X_i - X_i^{(m)}\|^2 + E\|X_i^{(m)}\|^2 \|X_i - X_i^{(m)}\|^2 \right. \\
&\qquad \left. + 2E\|X_i\| \|X_i^{(m)}\| \|X_i - X_i^{(m)}\|^2 \right).
\end{aligned}
$$

Applying the Cauchy–Schwarz inequality multiple times, and using the fact that $X_i \overset{\mathcal{L}}{=} X_i^{(m)}$, yields

$$
\begin{aligned}
E\|X_i \otimes X_i - X_i^{(m)} \otimes X_i^{(m)}\|^2 &\leq 2 \left(2(E\|X_i\|^4 E\|X_i - X_i^{(m)}\|^4)^{1/2} \right. \\
&\qquad \left. + 2(E\|X_i\|^4 E\|X_i - X_i^{(m)}\|^4)^{1/2} \right) \\
&= 8(E\|X_i\|^4 E\|X_i - X_i^{(m)}\|^4)^{1/2}.
\end{aligned}
$$

Therefore, we have that

$$\sum_{m=1}^{\infty} \nu_2(X_i \otimes X_i - X_i^{(m)} \otimes X_i^{(m)}) \leq \sum_{m=1}^{\infty} \sqrt{8} \, \nu_4(X_i)\nu_4(X_i - X_i^{(m)}) < \infty,$$

since the X_i are L^4-m-approximable.

Next we will show that the L^2-m-approximability of the operators $X_i \otimes X_i$ implies that $\sqrt{N}(\widehat{C} - C)$ is asymptotically normal. This is a delicate point because L^2-m-approximability implies only the second moment of $X_i \otimes X_i$ and these operators are dependent. We will use the results of Aue et al. [2] and Chen et al. [5]. Since the $X_i \otimes X_i$ are L^2-m-approximable random operators, this implies that the projections $\langle X_i \otimes X_i, h \rangle$ form an L^2-m-approximable sequence of

scalars for any $h \in L^2(\mathcal{T} \times \mathcal{T})$. Consequently, by Lemma 4, $\langle \sqrt{N}(\widehat{C} - C), h \rangle$ is asymptotically normal for any $h \in L^2(\mathcal{T} \times \mathcal{T})$. If we let $Y_i = X_i \otimes X_i - C$, then we can write

$$\langle \sqrt{N}(\widehat{C} - C), h \rangle = \sqrt{N} \sum_{i=1}^{N} \langle Y_i, h \rangle.$$

Since the $\langle Y_i, h \rangle$ are L^2-m approximable, we have by Lemma 4 that the asymptotic covariance of $\sqrt{N} \langle \widehat{C} - C, h \rangle$ is

$$E\left(\langle Y_1, h \rangle \langle Y_1, h \rangle \right) + 2 \sum_{i=2}^{\infty} E\left(\langle Y_1, h \rangle \langle Y_i, h \rangle \right)$$

$$= \left\langle E[Y_1 \otimes Y_1] + 2 \sum_{i=2}^{\infty} E[Y_1 \otimes Y_i], h \otimes h \right\rangle.$$

Thus we will have that

$$\Gamma = E[Y_1 \otimes Y_1] + 2 \sum_{i=2}^{\infty} E[Y_1 \otimes Y_i] = \lim_{N \to \infty} E[Z_N \otimes Z_N],$$

if the sum is convergent. Using the fact that $Y_1^{(i-1)}$ is independent of Y_i we have

$$\Gamma = E[Y_1 \otimes Y_1] + 2 \sum_{i=2}^{\infty} E[(Y_1 - Y_1^{(i-1)}) \otimes Y_i].$$

Applying the Cauchy–Schwarz inequality we have that

$$\| E[(Y_1 - Y_1^{(i-1)}) \otimes Y_i] \| \le v_2(Y_1 - Y_1^{i-1}) v_2(Y_i),$$

which implies that

$$\| E[Y_1 \otimes Y_1] \| + 2 \sum_{i=2}^{\infty} \| E[(Y_1 - Y_1^{(i-1)}) \otimes Y_i] \|$$

$$\le \| E[Y_1 \otimes Y_1] \| + 2 v_2(Y_i) \sum_{i=2}^{\infty} v_2(Y_1 - Y_1^{i-1}) < \infty,$$

since the Y_i are L^2-m approximable. Therefore the sum is convergent and we have that

$$\Gamma = E[Y_1 \otimes Y_1] + 2 \sum_{i=2}^{\infty} E[Y_1 \otimes Y_i].$$

By Lemma 5, this will imply that $\sqrt{N}(\widehat{C} - C)$ is asymptotically normal with covariance operator Γ if we can prove that

$$\lim_{N \to \infty} E\|\sqrt{N}(\widehat{C} - C)\|^2,$$

exists and is finite. By definition we have

$$E\|\sqrt{N}(\widehat{C} - C)\|^2 = N^{-1} E \int \left(\sum_{i=1}^{N} (X_i(t)X_i(s) - E(X_i(t)X_i(s))) \right)^2 dt \, ds.$$

Letting $Y_i(t, s) = X_i(t)X_i(s) - E(X_i(t)X_i(s))$, we have

$$E\|\sqrt{N}(\widehat{C} - C)\|^2 = N^{-1} \int E \sum_{i,k=1}^{N} Y_i(t, s)Y_k(t, s) \, dt \, ds$$

$$= \int \left(\Gamma_0(t, s) + 2 \sum_{i=1}^{N-1} \frac{N-i}{N} \Gamma_i(t, s) \right) dt \, ds,$$

where $\Gamma_i(t, s) = E(Y_1(t, s)Y_{i+1}(t, s)) = \mathrm{Cov}(X_1(t)X_1(s), X_{1+i}(t)X_{1+i}(s))$. So, if the infinite series

$$\int \left(\Gamma_0(t, s) + 2 \sum_{i=1}^{\infty} \Gamma_i(t, s) \right) dt \, ds$$

converges and is finite, then $\lim_{N \to \infty} E\|\sqrt{N}(\widehat{C} - C)\|^2$ will exist and be finite. Since $Y_1^{(m)}$ is independent of Y_{m+1} we have that

$$E[Y_1(t, s)Y_{i+1}(t, s)] = E[(Y_1(t, s) - Y_1^{(i)}(t, s))Y_{i+1}(t, s)].$$

This in turn implies, by the Cauchy–Schwarz inequality,

$$\int |\Gamma_i(t, s)| \, dt \, ds \le v_2 \left(Y_1 - Y_1^{(i)} \right) v_2(Y_1).$$

Since the Y_i are L^2-m-approximable, we have for all N

$$\int \left(|\Gamma_0(t, s)| + 2 \sum_{i=1}^{\infty} |\Gamma_i(t, s)| \right) dt \, ds$$

$$\le \int |\Gamma_0(t, s)| \, dt \, ds + 2 \sum_{i=1}^{\infty} v_2 \left(Y_1 - Y_1^{(i)} \right) v_2(Y_1) < \infty,$$

which completes the proof. $\square$

Acknowledgment

References

[1] T.W. Anderson, The Statistical Analysis of Time Series, Wiley and Sons, 1971.
[2] A. Aue, S. Hörmann, L. Horváth, M. Reimherr, Break detection in the covariance structure of multivariate time series models, The Annals of Statistics 37 (2009) 4046–4087.
[3] P. Billingsley, Convergence of Probability Measures, second ed., Wiley, New York, 1999.
[4] D. Bosq, Linear Processes in Function Spaces, Springer, 2000.
[5] X. Chen, H. White, Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications, Econometric Theory 14 (1998) 260–284.
[6] J. Dauxois, A. Pousse, Y. Romain, Asymptotic theory for principal component analysis of a vector random function, Journal of Multivariate Analysis 12 (1982) 136–154.
[7] R.A. Davis, T. Mikosch, The sample autocorrelations of heavy-tailed processes with applications to ARCH, The Annals of Statistics 26 (1998) 2049–2080.
[8] H. Dehling, Limit theorems for sums of weakly dependent Banach space valued random variables, Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete 63 (1983) 393–432.

[9] P. Doukhan, S. Louhichi, A new weak dependence and applications to moment inequalities, Stochastic Processes and their Applications 84 (1999) 313–343.

[10] N. Dunford, J.T. Schwartz, Linear Operators, Parts I and II, Wiley, 1988.

[11] F. Ferraty, Y. Romain (Eds.), The Oxford Handbook of Functional Data Analysis, Oxford University Press, 2011.

[12] R. Gabrys, L. Horváth, P. Kokoszka, Tests for error correlation in the functional linear model, Journal of the American Statistical Association 105 (2010) 1113–1125.

[13] I. Gohberg, S. Golberg, M.A. Kaashoek, Classes of Linear Operators, in: Operator Theory: Advances and Applications, vol. 49, Birkhäuser, 1990.

[14] O. Gromenko, P. Kokoszka, L. Zhu, J. Sojka, Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends, The Annals of Applied Statistics 6 (2012) 669–696.

[15] P. Hall, M. Hosseini-Nasab, On properties of functional principal components, Journal of the Royal Statistical Society: Series B 68 (2006) 109–126.

[16] S. Hörmann, L. Horváth, R. Reeder, A functional version of the ARCH model. Econometric Theory (2013) (forthcoming).

[17] S. Hörmann, P. Kokoszka, Weakly dependent functional data, The Annals of Statistics 38 (2010) 1845–1884.

[18] S. Hörmann, P. Kokoszka, Functional time series, in: C.R. Rao, T. Subba Rao (Eds.), Time Series, in: Handbook of Statistics, vol. 30, Elsevier, 2012.

[19] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, Springer, 2012.

[20] L. Horváth, P. Kokoszka, R. Reeder, Estimation of the mean of functional time series and a two sample problem, Journal of the Royal Statistical Society: Series B 75 (2013) 103–122.

[21] A. Jakubowski, On Limit Theorems for Sums of Dependent Hilbert Space Valued Random Variables, Springer-Verlag, 1980.

[22] I.T. Jolliffe, Principal Component Analysis, Springer, 2002.

[23] A. Mas, Weak convergence for the covariance operators of a Hilbertian linear process, Stochastic Processes and their Applications 99 (2002) 117–135.

[24] A. Mas, L. Menneteau, Perturbation approach applied to the asymptotic study of random operators, in: J. Hoffmann-Jørgensen, M.B. Marcus, J.A. Wellner (Eds.), High Dimensional Probability. III, Birkhäuser, Basel, 2003, pp. 127–134.

[25] Florence Merlevéde, On the central limit theorem and its weak invariance principle for strongly mixing sequences with values in a Hilbert space via martingale approximation, Journal of Theoretical Probability 16 (2003) 625–653.

[26] F. Merlevéde, M. Peligrad, S. Utev, Sharp conditions for the CLT of linear processes in a Hilbert space, Journal of Theoretical Probability 10 (1997) 681–693.

[27] T. Mikosch, C. Stărică, Limit theory for the sample autocorrelation and extremes of a GARCH(1, 1) process, The Annals of Statistics 28 (2000) 1427–1451.

[28] M. Peligrad, S. Utev, Central limit theorem for linear processes, The Annals of Probability 25 (1997) 443–456.

[29] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, Springer, 2005.

[30] W. Wu, Nonlinear system theory: another look at dependence, Proceedings of the National Academy of Sciences of the United States of America 102 (2005) National Academy of Sciences.

[31] W. Wu, Strong invariance principles for dependent random variables, The Annals of Probability 35 (2007) 2294–2320.

[32] F. Yao, H.-G. Müller, J.-L. Wang, Functional data analysis for sparse longitudinal data, Journal of the American Statistical Association 100 (2005) 577–590.