Asymptotics of Functional Spectral Component Analysis

Bo Hu*
Department of Economics, Indiana University

Abstract

This paper develops a unified asymptotic distribution theory for functional spectral component analysis—the analysis of spectral characteristics of certain operators, referred to as base operators, constructed from functional data. Using a functional delta method, I derive the limiting distributions of spectral quantities under both i.i.d. and weakly dependent functional data. The framework accommodates non-self-adjoint base operators as well as those with non-simple spectra. Applications include functional principal component analysis, functional partial least squares, and regularized inverses in the functional setting. I provide explicit conditions for the weak convergence of sample base operators and for the consistent estimation of long-run variances. This general approach offers a flexible foundation for analyzing a broad range of spectral methods in functional data analysis.

Keywords: Functional spectral component analysis, asymptotic distribution theory, functional data, central limit theorem, weak dependence

*Corresponding author. hu21@iu.edu

1 Introduction

A fundamental tool in functional data analysis (FDA) is functional principal component analysis (FPCA), the infinite-dimensional analogue of principal component analysis. FPCA plays a central role not only in understanding the properties of the sample variance operator but also in reducing dimensionality in functional settings. For example, it is widely used in functional regression, both under independence and dependence. See Bosq (2000), Ramsey & Silverman (2005), Ferraty & Vieu (2006), Horväth & Kokoszka (2012), and Kokoszka & Reimherr (2017) for examples.

The theory of FPCA is built on the analysis of the eigenvalues and eigenfunctions of the sample variance operator, which are often interpreted as the spectral information of the operator.[1] Many other estimators and quantities of interest in functional data analysis can also be represented in terms of the spectral information of operators constructed from the data. I refer to such operators as *base operators* in this paper, including, but not limited to, variance, covariance, autocovariance, higher-moment operators, and operators arising from other primitive procedures (e.g., first-step estimators). Unlike in FPCA, in general these operators need not be self-adjoint. As examples of analysis based on spectral information, functional partial least squares (FPLS) involves finding directions that correspond to the leading singular vectors of certain covariance operators, and regularization techniques in FDA and machine learning—such as Tikhonov and spectral cut-off regularization—operate by stabilizing the inverse spectrum of operators constructed from the data.

To unify such analyses, I use the term *functional spectral component analysis (FSCA)* to refer to the study of statistical quantities defined via the spectral characteristics of base operators in the functional setting.

[1] Here, "spectral" refers to the operator-theoretic sense, not to spectral (frequency-domain) analysis in time series.

This paper develops a general asymptotic distribution theory for FSCA quantities using a version of the functional delta method. The proposed framework accommodates both i.i.d. and various forms of weakly dependent functional data. As illustrations, I derive asymptotic results for: (i) the eigenvalues, eigenfunctions, and spectral projections of the sample variance operator; (ii) the singular values and singular vectors of general compact operators, including sample covariance and autocovariance operators; (iii) the weight functions in functional partial least squares; and (iv) regularized inverses of the sample variance operator, where the data takes values in real separable Hilbert spaces. The majority of empirical implementations adopt such spaces, with many taking $L^2[0,1]$ as the underlying domain. The framework is general and can be extended to other statistical procedures in functional data analysis, such as functional canonical correlation analysis.

A seminal study of the asymptotic distribution of spectral quantities in functional data analysis is Dauxois et al. (1982), who analyzed FPCA under i.i.d. functional data using tools from functional calculus and the Rubin–Billingsley theorem. Their approach relied on the almost sure convergence of the sample covariance operator. Mas (2002) extended the asymptotic theory of FPCA to the setting of functional linear processes. In this paper, I observe that many spectral quantities, being continuous mappings of certain base operators, inherit the asymptotic distributional properties of those operators via a functional version of the delta method. This approach requires only weak convergence of the sample base operators, rather than almost sure convergence. Provided that an appropriate Fréchet derivative exists, explicit expressions for the limiting distributions of spectral quantities can be obtained. In particular, I build on the Fréchet derivative results for eigen-elements established by Gilliam et al. (2009) to derive joint asymptotic distributions for the eigenvalues, eigenfunctions, and projections of FPCA. Kokoszka & Reimherr (2013) also developed asymptotic theory for FPCA based on weak convergence of the sample variance operator in the Gaussian case, noting that Gaussianity is not essential so long as certain projection-based independence conditions hold. By contrast, the delta-method-based framework in this paper accommodates arbitrary distributions in the limiting law of the sample variance operator without requiring additional structural assumptions. Other contributions to the asymptotic theory of FPCA include Boente & Fraiman (2000), Hall & Hosseini-Nasab (2006), Panaretos & Tavakoli (2013), and Hörmann et al. (2015), among others.

Beyond FPCA, the framework in this paper applies to broader scenarios. In particular, I derive asymptotic distribution theory for functional partial least squares (FPLS). While FPLS has been considered by Preda & Saporta (2005), and its asymptotic theory was developed in Delaigle & Hall (2012), my setting generalizes theirs to allow both predictors and responses to be functional. This includes as special cases finite-dimensional regressors and/or responses. Also, whereas Delaigle & Hall (2012) analyzed an alternative formulation of FPLS, I establish asymptotic theory directly for the iterative algorithm. As another example, I apply the theory to regularized inverses of variance operators, which are widely used in high-dimensional and machine learning contexts such as ridge regression, PCA regression, and kernel methods in reproducing kernel Hilbert spaces (Vito et al. 2005, Bauer et al. 2007, Dicker et al. 2017).

While much of the FPCA literature assumes simple (non-repeated) eigenvalues—Dauxois et al. (1982) being a notable exception—the framework in this paper accommodates the case of non-simple eigenvalues and singular values. Furthermore, whereas the existing literature has primarily focused on the i.i.d. case, I allow for weak dependence in the data. This generalization is natural, as the framework here only requires weak convergence of the sample base operators, which typically follows from central limit theorems under weak dependence. I provide explicit conditions under which such operators converge weakly to

Gaussian limits, and offer consistent estimators for the long-run variance operator appearing in their limiting distribution.

Finally, I comment on the case of strongly dependent functional data. The theory for FSCA becomes even simpler in such settings: once a functional central limit theorem (i.e., an invariance principle) is established for the partial sum process of the sample base operators, the asymptotic distribution of FSCA components follows directly by the continuous mapping theorem, without requiring the functional delta method.

2 Preliminaries

In this section, I introduce some preliminaries for the development of the theory.[2] Let the probability space be $(\Omega, \mathcal{F}, \mathbb{P})$. Throughout the paper, let H, H_1, and H_2 denote real separable Hilbert spaces, each equipped with an inner product $\langle \cdot, \cdot \rangle$ and the corresponding norm $\| \cdot \|$. For simplicity, I use the same notation for inner products and norms across different spaces; this should not cause confusion when the context is clear.

Let ξ be an H-valued random element, i.e., a measurable mapping from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(H, \mathcal{B}(H))$, where $\mathcal{B}(\cdot)$ denotes the Borel σ-algebra. We say that ξ is of *strong order* p if $\mathbb{E}\|\xi\|^p < \infty$. The space of (equivalence classes of) all H-valued random elements of strong order p is denoted by $L^p(H)$. Equipped with the norm $\|\xi\|_p = (\mathbb{E}\|\xi\|^p)^{1/p}$, the space $L^p(H)$ is a Banach space for $p \geq 1$; in particular, $L^2(H)$ is a Hilbert space.

The expectation of ξ is defined as the unique vector $\mathbb{E}\xi \in H$ such that $\langle \mathbb{E}\xi, x \rangle = \mathbb{E}\langle \xi, x \rangle$ for all $x \in H$. If $\xi \in L^p(H)$ for some $p \geq 1$, then $\mathbb{E}\xi$ exists and is unique. The expectation operator $\mathbb{E}$, viewed as a mapping from $L^p(H)$ to H, is linear and continuous.

[2] For further details, see, e.g., Rudin (1991), Conway (2000), Gohberg et al. (1990), and Vakhania et al. (1987).

The tensor product $H_1 \otimes H_2$ is defined as the vector space spanned by elementary tensors $x_1 \otimes x_2$, with $x_1 \in H_1$, $x_2 \in H_2$, where each elementary tensor is interpreted as a bilinear form on $H_1 \times H_2$ given by $(x_1 \otimes x_2)(v_1, v_2) = \langle x_1, v_1 \rangle \langle x_2, v_2 \rangle$. It is known that such a tensor product exists and is unique up to isomorphism. If we define an inner product on $H_1 \otimes H_2$ by $\langle x_1 \otimes x_2, y_1 \otimes y_2 \rangle = \langle x_1, y_1 \rangle \langle x_2, y_2 \rangle$ and complete the space under the norm induced by this inner product, then $H_1 \otimes H_2$ becomes a separable Hilbert space.

Let ξ and η be mean-zero random elements taking values in H_1 and H_2, respectively. The covariance operator of ξ and η is defined as $C_{\xi\eta} = \mathbb{E}(\xi \otimes \eta)$. Since $H_1 \otimes H_2$ is a separable Hilbert space, the expectation $\mathbb{E}(\xi \otimes \eta)$ is well defined and uniquely determined whenever ξ and η are of strong order 2. The variance operator of ξ is naturally defined as $C_{\xi\xi} = \mathbb{E}(\xi \otimes \xi)$. For random elements that are not mean zero, the variance and covariance operators are defined analogously after centering the variables by subtracting their expectations.

We say that a random element ξ taking values in H is Gaussian if, for every $x \in H$, the real-valued random variable $\langle \xi, x \rangle$ is Gaussian. It is well known that the distribution of an H-valued Gaussian random element is completely determined by its mean and covariance operator. In what follows, we use the notation $\mathbb{N}(m, K)$ to denote the distribution of an H-valued Gaussian random element with mean m and variance operator K.

Let $\mathcal{L}(H_1, H_2)$ denote the space of all bounded linear operators from H_1 to H_2, equipped with the operator norm $\|\cdot\|$. We write $\mathcal{L}(H)$ as shorthand for $\mathcal{L}(H, H)$. Since H_1 is separable, it admits a countable orthonormal basis $\{e_i\}_{i=1}^{\infty}$. For operators $A, B \in \mathcal{L}(H_1, H_2)$, we define the inner product $\langle A, B \rangle = \sum_{i=1}^{\infty} \langle Ae_i, Be_i \rangle$, and denote the corresponding norm by $\|A\|_{HS}$. It is well known that this inner product, and hence the norm, is independent of the choice of orthonormal basis. An operator $A \in \mathcal{L}(H_1, H_2)$ is called a Hilbert–Schmidt operator if $\|A\|_{HS} < \infty$. The collection of all such operators forms a separable Hilbert space $\mathcal{L}_{\mathrm{HS}}(H_1, H_2) \subset \mathcal{L}(H_1, H_2)$ with the above inner product. It holds that $\|A\| \leq \|A\|_{HS}$ for all

$A \in \mathcal{L}_{HS}(H_1, H_2)$.

There exists an isometric isomorphism between $H_1 \otimes H_2$ and $\mathcal{L}_{HS}(H_1, H_2)$, given by identifying the elementary tensor $x_1 \otimes x_2$ with the rank-one operator $v \mapsto \langle x_1, v \rangle x_2$ for $v \in H_1$. Under this identification, the covariance operator $\mathbb{E}(\xi \otimes \eta)$ can be viewed as a bounded linear operator satisfying $\mathbb{E}(\xi \otimes \eta)v = \mathbb{E}\langle \xi, v \rangle \eta, \quad \text{for all } v \in H_1$. In what follows, we will more frequently adopt this operator-theoretic point of view, rather than the bilinear mapping interpretation of tensors.

An operator $A \in \mathcal{L}(H_1, H_2)$ is said to be compact if the closure of the image $A(B_{H_1}(0, 1))$ is compact in H_2, where $B_{H_1}(0, 1)$ denotes the open unit ball centered at 0 in H_1. The space of all compact linear operators from H_1 to H_2 is denoted by $\mathcal{L}_C(H_1, H_2)$. It is known that $\mathcal{L}_{\mathrm{HS}}(H_1, H_2) \subset \mathcal{L}_C(H_1, H_2) \subset \mathcal{L}(H_1, H_2)$.

The spectrum $\sigma(A)$ of an operator $A \in \mathcal{L}(H)$ is the set of all scalars $\lambda \in \mathbb{C}$ such that $A - \lambda I$ is not invertible, where I denotes the identity operator on H. A scalar λ for which $A - \lambda I$ is not injective, i.e., $\mathcal{N}(A - \lambda I) \neq \{0\}$, is called an eigenvalue of A. The set of all eigenvalues is called the point spectrum of A and is denoted by $\sigma_p(A)$. The corresponding kernel space $\mathcal{N}(A - \lambda I)$ is called the eigenspace associated with λ, and any nonzero element of this space is referred to as an eigenvector corresponding to λ. An eigenvector with unit norm is called a normalized eigenvector.

It is well known that if A is compact, then its spectrum $\sigma(A)$ is nonempty, compact, and at most countable, with at most one possible accumulation point, which must be 0. Moreover, for compact operators, every nonzero element of the spectrum is an eigenvalue. An operator $A \in \mathcal{L}(H)$ is said to be of finite rank if its range $\mathcal{R}(A)$ is finite-dimensional. An operator is compact if and only if it is the norm-limit of a sequence of finite-rank operators. In particular, every finite-rank operator is compact.

For any operator $A \in \mathcal{L}(H_1, H_2)$, there exists a unique operator $A^* \in \mathcal{L}(H_2, H_1)$ such that $\langle Ax, y \rangle = \langle x, A^* y \rangle$ for all $x \in H_1,\ y \in H_2$. The operator A^* is called the adjoint of A. An operator $A \in \mathcal{L}(H)$ is said to be self-adjoint if $A^* = A$. If $A \in \mathcal{L}(H)$ is a compact, self-adjoint operator, then it admits a spectral decomposition of the form

$$A = \sum_{i=1}^{\infty} \lambda_i (e_i \otimes e_i), \tag{1}$$

where $\{\lambda_i\}$ is a sequence of (possibly repeated) eigenvalues of A, and $\{e_i\}$ is an orthonormal basis for $\mathcal{N}(A)^\perp$ consisting of corresponding eigenvectors.

Functional calculus can be used to express operators in terms of their spectral information. To apply functional calculus, we need to work with vector spaces over the complex field. Since we consider random elements taking values in a real separable Hilbert space H, we introduce its complexification $\mathbb{H}$—a complex Hilbert space into which H is isometrically embedded. Every orthonormal basis of H remains an orthonormal basis of $\mathbb{H}$. Each operator $A \in \mathcal{L}(H)$ admits a unique complexification $\mathbb{A} \in \mathcal{L}(\mathbb{H})$ that coincides with A on H and satisfies $\|A\| = \|\mathbb{A}\|$. Moreover, the adjoint of the complexified operator coincides with the complexified operator of the adjoint. If A is self-adjoint, then A and $\mathbb{A}$ share the same spectral decomposition in their respective spaces. In particular, if A has the spectral decomposition $A = \sum_{i=1}^{\infty} \lambda_i (e_i \otimes e_i)$ where $\{e_i\}$ is an orthonormal basis for $\mathcal{N}(A)^\perp \subset H$, then the complexified operator $\mathbb{A}$ has the spectral decomposition $\mathbb{A} = \sum_{i=1}^{\infty} \lambda_i (e_i \otimes e_i)$ where $\{e_i\}$ is now viewed as an orthonormal system in $\mathbb{H}$. As we shall see, the development of the theory in this paper relies essentially on spectral representations. Therefore, the passage from A to its complexification $\mathbb{A}$ is a purely technical device for applying functional calculus and does not alter the nature of the problem. For notational simplicity, we will continue to write A in place of $\mathbb{A}$.[3]

[3] An alternative approach is to begin with a complex Hilbert space H and restrict the random elements to take values in a real subspace of H. For example, one may consider H to be the space of complex-valued

Let Γ be a path in $\mathbb{C}$, that is, Γ is a piecewise continuously differentiable mapping of a compact interval $[a, b]$ of the real line into the complex plane. Let f be a continuous function from the range of Γ into a complex Hilbert space $\mathbb{H}$. We define the integral $\int_\Gamma f(x)\,\mathrm{d}x$ of f over the path Γ to be the limit of $\sum_{j=1}^n f(\Gamma(s_j))(\Gamma(s_j) - \Gamma(s_{j-1}))$ as the partition $a = s_0 \leq s_1 \leq \cdots \leq s_n = b$ becomes finer and finer, where the convergence is in the norm of $\mathbb{H}$. This is a direct generalization of the Riemann–Stieltjes integral for functions with values in a Banach space, and it can be shown that the vector-valued integral satisfies the usual properties of the Riemann–Stieltjes integral for complex-valued functions. For example, the limit in the definition is independent of the sequence of partitions used to evaluate it, and therefore the integral is well defined. Also, it follows that $\int_\Gamma f(x)\,\mathrm{d}x = \int_a^b f(\Gamma(s))\Gamma'(s)\,\mathrm{d}s$.

Let D be a region in the complex plane, i.e., a nonempty connected open subset of $\mathbb{C}$. Define the set $\mathcal{L}_D(\mathbb{H}) = \{A \in \mathcal{L}(\mathbb{H}) : \sigma(A) \subset D\}$, where $\sigma(A)$ denotes the spectrum of A. Then, for any holomorphic function f defined on D and any $A \in \mathcal{L}_D(\mathbb{H})$, we define

$$f(A) = \frac{1}{2\pi i} \oint_{\Gamma_A} f(z)(zI - A)^{-1}\,\mathrm{d}z, \tag{2}$$

where I is the identity operator on $\mathcal{L}(\mathbb{H})$, and Γ_A is any finite collection of closed paths such that

$$\frac{1}{2\pi i} \int_{\Gamma_A} (z - \lambda)^{-1}\,\mathrm{d}z = \begin{cases} 1 & \text{if } \lambda \in \sigma(A), \\ 0 & \text{if } \lambda \in D^c. \end{cases}$$

In this case, we say that Γ_A is a contour that surrounds $\sigma(A)$. It is known that such a contour Γ_A exists, and that the definition of $f(A)$ is independent of the choice of Γ_A.

---

square-integrable functions and restrict attention to the subspace of real-valued functions. In this setting, the distribution of an H-valued random element is concentrated on a proper real subspace. When using functionals in the dual space (e.g., when defining Gaussianity), one must restrict attention to real-linear functionals. The two approaches are equivalent. Following the standard probabilistic literature, we assume H is real from the outset and complexify it only when needed.

3 Functional Spectral Component Analysis

As introduced in the Introduction, FSCA concerns statistical quantities defined through the spectral characteristics of base operators. Such quantities can often be expressed in the form $f(A)$, as in (2), for appropriately chosen base operator A and mapping f, or as functions of $f(A)$. Many analyses in the functional data setting fall within this framework. In this section, I present several examples, with FPCA being the most familiar.

Let $\{X_t\}_{t=-\infty}^{\infty}$ be a sequence of H-valued random elements, where $X_t \in L^2(H)$ for each t. We say that $\{X_t\}$ is weakly stationary if $\mathbb{E}X_t$ is independent of t, and $C_{X_1,X_{1-k}} = C_{X_h,X_{h-k}}$ for all $h, k \in \mathbb{Z}$. In this case, we write $\gamma(k) = C_{X_1,X_{1-k}}$. For simplicity, I first consider a mean-zero weakly stationary process $\{X_t\}$ and a sample $(X_1, X_2, \ldots, X_T)$ drawn from it. The mean-zero assumption is made here only to simplify the illustration of FSCA examples; we shall drop this assumption in later sections.

3.1 Functional Principal Component Analysis

FPCA is the infinite-dimensional counterpart of principal component analysis, which seeks a set of orthogonal basis functions whose span captures the majority of the variation in the data in an optimal way (Kleffe 1973). It is also commonly used as a dimension-reduction tool, particularly in high-dimensional settings. FPCA is based on the spectral representation of the variance operator

$$V = \mathbb{E}(X_t \otimes X_t),$$

which is compact, self-adjoint, and positive semi-definite. The spectral representation is given by

$$V = \sum_{i=1}^{\infty} \lambda_i (v_i \otimes v_i), \tag{3}$$

where (λ_i, v_i) are the eigenvalue–eigenvector pairs of V. Since V is self-adjoint, the eigenvalues λ_i are real and the eigenvectors v_i are mutually orthogonal. Recall that the eigenvalues constitute the point spectrum of V.

The estimation of FPCA is based on the sample variance operator, defined as

$$\widehat{V} = \frac{1}{T}\sum_{t=1}^{T}(X_t \otimes X_t).$$

In this paper, estimated quantities are denoted with hats and are functions of the sample size T, which is omitted from the notation for convenience. All asymptotic results are derived under the limit $T \to \infty$. Note that $\widehat{V}$ has finite rank and is therefore compact. Moreover, $\widehat{V}$ is self-adjoint and thus admits the spectral representation

$$\widehat{V} = \sum_{i=1}^{T} \widehat{\lambda}_i \, (\widehat{v}_i \otimes \widehat{v}_i).$$

With an appropriate ordering and identification scheme, the eigenpairs $(\widehat{\lambda}_i, \widehat{v}_i)$ can be used to estimate (λ_i, v_i). The detailed identification assumptions will be introduced in Section 4.2.

The central quantities of interest are the eigenprojections $P_i = v_i \otimes v_i$ and their empirical counterparts $\widehat{P}_i = \widehat{v}_i \otimes \widehat{v}_i$. Other quantities in FPCA can be expressed as functions of the eigenprojections. As we shall see in Section 4.2, P_i and $\widehat{P}_i$ can be written in the form of (2) with $A = V$ and $A = \widehat{V}$, respectively, and with an appropriately chosen mapping f.

3.2 Singular Value Decomposition

Sometimes we wish to analyze base operators that are similar to the variance operator but are not self-adjoint. For example, the covariance operator of two functional random variables, or the autocovariance $\gamma(k)$ of a weakly stationary functional time series $\{X_t\}$, is not self-adjoint in general. We may be interested, for instance, in the magnitude of such covariance operators. A non-self-adjoint operator $A \in \mathcal{L}(H_1, H_2)$ admits the singular value decomposition

$$A = \sum_{i=1}^{\infty} \mu_i \, (u_i \otimes w_i),$$

where $\{u_i\}$ and $\{w_i\}$ are orthonormal systems in H_1 and H_2, respectively, and $\{\mu_i\}$ are real, nonnegative singular values, arranged in descending order. Under this ordering, $\|A\| = \mu_1$. If A denotes the autocovariance operators of a functional time series, the above relationship implies that the magnitude of the series' second-order dependence is captured by the singular values of the autocovariance operators.

It is known that

$$\mu_i = \lambda_i^{1/2}(A^*A) = \lambda_i^{1/2}(AA^*),$$

that u_i is an eigenvector of A^*A associated with the eigenvalue λ_i, and that w_i is an eigenvector of AA^* associated with the same eigenvalue λ_i. Using these relationships, quantities of interest in the singular value decomposition can be expressed as functions of the spectral quantities of A^*A or AA^*.

3.3 Functional Partial Least Squares

FPLS can be formulated within the framework of functional spectral component analysis. Let $\{(Y_t, X_t)\}$ be a sequence of weakly stationary $H_1 \times H_2$-valued random elements, and assume for now that X_t and Y_t are mean zero. The FPLS problem seeks $\mathsf{u}^{(k)} \in H_1$ and $\mathsf{w}^{(k)} \in H_2$ for $k = 1, \ldots, K$ that solve

$$\max_{\left\|\mathsf{w}^{(k)}\right\| \leq 1, \left\|\mathsf{u}^{(k)}\right\| \leq 1} \left\langle \mathsf{w}^{(k)}, \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{Y}_t^{(k)} \otimes \widehat{X}_t^{(k)} \right) \mathsf{u}^{(k)} \right\rangle \tag{4}$$

iteratively, where $\widehat{X}_t^{(1)} = X_t$ and $\widehat{Y}_t^{(1)} = Y_t$, and for $k \geq 2$, the updates are

$$\begin{aligned}
\widehat{Z}_t^{(k)} &= \left\langle \widehat{X}_t^{(k-1)}, \widehat{\mathsf{u}}^{(k-1)} \right\rangle, \\
\widehat{X}_t^{(k)} &= \widehat{X}_t^{(k-1)} - \frac{\sum_{t=1}^{T} \widehat{Z}_t^{(k)} \widehat{X}_t^{(k-1)}}{\sum_{t=1}^{T} \left(\widehat{Z}_t^{(k)}\right)^2} \widehat{Z}_t^{(k)}, \\
\widehat{Y}_t^{(k)} &= \widehat{Y}_t^{(k-1)} - \frac{\sum_{t=1}^{T} \widehat{Z}_t^{(k)} \widehat{Y}_t^{(k-1)}}{\sum_{t=1}^{T} \left(\widehat{Z}_t^{(k)}\right)^2} \widehat{Z}_t^{(k)}.
\end{aligned}$$

The sample covariance operator in step k admits the singular value decomposition

$$\widehat{C}^{(k)} := \frac{1}{T}\sum_{t=1}^{T} Y_t^{(k)} \otimes X_t^{(k)} = \sum_{i=1}^{T} \widehat{\mu}_i^{(k)} \left(\widehat{u}_i^{(k)} \otimes \widehat{w}_i^{(k)}\right),$$

where the singular values $\widehat{\mu}_i^{(k)}$ are non-negative and arranged in descending order. The solution to (4) is

$$\widehat{\mathsf{w}}^{(k)} = \widehat{w}_1^{(k)}, \quad \widehat{\mathsf{u}}^{(k)} = \widehat{u}_1^{(k)},$$

so that the FPLS weight function $\widehat{\mathsf{w}}^{(k)}$ is the eigenvector of $\widehat{C}^{(k)}\widehat{C}^{(k)*}$ corresponding to its largest eigenvalue, and $\widehat{\mathsf{u}}^{(k)}$ is the eigenvector of $\widehat{C}^{(k)*}\widehat{C}^{(k)}$ corresponding to its largest eigenvalue.

The population counterpart of the FPLS problem is as follows: the kth population FPLS direction $\mathsf{w}^{(k)}$ solves

$$\max_{\left\|\mathsf{w}^{(k)}\right\|\leq 1, \left\|\mathsf{u}^{(k)}\right\|\leq 1} \left\langle \mathsf{w}^{(k)}, \left(\mathbb{E}\, Y_t^{(k)} \otimes X_t^{(k)}\right) \mathsf{u}^{(k)} \right\rangle,$$

where $X_t^{(1)} = X_t$ and $Y_t^{(1)} = Y_t$, and for $k \geq 2$, the updates are

$$\begin{aligned}
Z_t^{(k)} &= \left\langle X_t^{(k-1)}, \mathsf{u}^{(k-1)} \right\rangle, \\
X_t^{(k)} &= X_t^{(k-1)} - \frac{\mathbb{E}\left[Z_t^{(k)} X_t^{(k-1)}\right]}{\mathbb{E}\left[(Z_t^{(k)})^2\right]} Z_t^{(k)}, \\
Y_t^{(k)} &= Y_t^{(k-1)} - \frac{\mathbb{E}\left[Z_t^{(k)} Y_t^{(k-1)}\right]}{\mathbb{E}\left[(Z_t^{(k)})^2\right]} Z_t^{(k)}.
\end{aligned}$$

The population covariance operator in step k admits the singular value decomposition

$$C^{(k)} := \mathbb{E}\, Y_t^{(k)} \otimes X_t^{(k)} = \sum_{i=1}^{\infty} \mu_i^{(k)} \left(u_i^{(k)} \otimes w_i^{(k)}\right),$$

where the singular values $\mu_i^{(k)}$ are non-negative and arranged in descending order. The solution to the population FPLS problem is

$$\mathsf{w}^{(k)} = w_1^{(k)}, \quad \mathsf{u}^{(k)} = u_1^{(k)},$$

so that $\mathsf{w}^{(k)}$ is the eigenvector of $C^{(k)}C^{(k)*}$ corresponding to its largest eigenvalue, and $\mathsf{u}^{(k)}$ is the eigenvector of $C^{(k)*}C^{(k)}$ corresponding to its largest eigenvalue.

3.4 Inverse Problem for Functional Data

Statistical methods in many settings involve solving inverse problems. For example, ordinary linear regression requires inverting the covariance matrix of the covariates. In the functional setting, however, where the data are infinite-dimensional, the inverse problem is typically ill-posed and a direct inverse is not usable. A common remedy is to employ regularized inverses based on the spectral representation of the corresponding operators.

For instance, given the spectral representation of the variance operator V in (3), one may consider the regularized inverse

$$V^+ = \sum_{i=1}^{\infty} q(\lambda_i)\,(v_i \otimes v_i), \tag{5}$$

where $q\colon \mathbb{R}_+ \to \mathbb{R}$ is a bounded regularization function on $(0, \|V\|]$ that replaces the unbounded function $x \mapsto 1/x$. This regularized inverse fits naturally into our framework, since it can be expressed in the form (2) by setting $A = V$ and $f = q$. In particular, if $q(x) = \frac{1}{x}1_{(\alpha,\infty)}(x)$ for some $\alpha > 0$, the inverse is said to be cut-off regularized, whereas if $q(x) = \frac{1}{x+\alpha}$ for some $\alpha > 0$, the inverse is referred to as Tikhonov regularized. Such regularized inverses of V arise, for example, in inference for functional linear regression and in functional canonical correlation analysis.

4 Asymptotic Theory

In this section, I develop a general asymptotic theory for FSCA using a functional delta method. The proposed framework applies directly to all examples presented in the preceding section, as well as to other related settings.

4.1 Delta Method for FSCA

In this section, our objective is to derive the asymptotic distribution of $f(\widehat{A})$, where $\widehat{A}$ is a consistent estimator of a general base operator A that converges in distribution, and $f(A)$ is defined as in (2). A natural tool for this purpose is the functional delta method. Its application requires an appropriate derivative of $f(A)$. While Hadamard differentiability is typically sufficient in the context of the functional delta method (van der Vaart 1998), the setting of functional calculus in (2) admits a direct derivation of the stronger Fréchet derivative.

Let $\mathbb{D}$ and $\mathbb{E}$ be normed vector spaces, and let $g : \mathbb{D}_g \to \mathbb{E}$ be a mapping with domain $\mathbb{D}_g \subset \mathbb{D}$. The map g is said to be Fréchet differentiable at A if there exists a bounded linear operator $g'(A) : \mathbb{D} \to \mathbb{E}$ such that

$$\|g(A + \Pi) - g(A) - g'(A)\Pi\| \to 0 \quad \text{as} \quad \|\Pi\| \to 0.$$

To obtain the derivative of $f(A)$ in (2), note that for $A, B \in \mathcal{L}(H)$ and $z \notin \sigma(A) \cup \sigma(B)$,

$$(zI - B)^{-1} - (zI - A)^{-1} = (zI - B)^{-1}(B - A)(zI - A)^{-1}.$$

This identity implies that if $g(A) = (zI - A)^{-1}$, then

$$g'(A)\Pi = (zI - A)^{-1}\Pi(zI - A)^{-1}, \quad \Pi \in \mathcal{L}(H).$$

Since $\sigma(A)$ is compact and Γ_A is closed, the distance between $\sigma(A)$ and Γ_A is strictly positive, and that $\|(zI - A)^{-1}\|$ is uniformly bounded for $z \in \Gamma_A$. By linearity of the integral and the dominated convergence theorem (Rudin 1991, Theorem 3.29), for a general holomorphic function f,

$$f'(A)\Pi = \frac{1}{2\pi i} \oint_{\Gamma_A} f(z)(zI - A)^{-1}\Pi(zI - A)^{-1}\,\mathrm{d}z, \quad \Pi \in \mathcal{L}(H). \tag{6}$$

A holomorphic function is a function that is complex differentiable at every point in a neighborhood of each point in its domain.

We therefore obtain the following functional delta method for FSCA.

Theorem 1. *Let A, D, f, Γ_A be defined as in (2). Suppose there exists an increasing sequence r_T of numbers and an estimator $\widehat{A}$ of A such that $r_T(\widehat{A} - A) \to_d \Xi$ as $T \to \infty$. Let f_T be a sequence of (possibly random) holomorphic functions on D such that $\sup_{z \in \Gamma_A} |f_T(z) - f(z)| = o_p(r_T^{-1})$, then*

$$r_T(f_T(\widehat{A}) - f(A)) \to_d \frac{1}{2\pi i} \oint_{\Gamma_A} f(z)(zI - A)^{-1} \Xi (zI - A)^{-1} \mathrm{d}z.$$

We make the following remarks.

1. In spectral component analysis procedures, as illustrated in the previous section, it is generally much easier to establish the weak convergence of some base operator $\widehat{A}$ than to derive the weak convergence of the associated spectral quantities directly. Once suitable A and f are identified, Theorem 1 provides a unified framework for handling the weak convergence of spectral component analysis quantities.
2. The limit distribution is expressed in the form of a contour integral. In some cases, the Cauchy integral formula (Rudin 1987, Theorem 10.15) can be applied to simplify this expression further, as will be demonstrated in the next section.
3. The theorem holds for general bounded linear operators in $\mathcal{L}(\mathbb{H})$. In many applications—such as in FPCA—A is compact and self-adjoint. In that case, one can use the spectral decomposition $A = \sum_{i=1}^{\infty} \lambda_i P_i$ and the resolvent expansion $(zI - A)^{-1} = \sum_{i=1}^{\infty} \frac{1}{z - \lambda_i} P_i$ to rewrite the contour integral on the right-hand side, which often leads to further simplifications.
4. The convergence condition $|f_T(z) - f(z)| = o_p(r_T^{-1})$ need only hold uniformly on Γ_A, rather than on D. Since Γ_A is compact in $\mathbb{C}$ and the value of the integral in (2) is independent of the particular choice of Γ_A, we may select a suitably regular contour to facilitate verification of the uniform convergence condition. In most applications, Γ_A can be chosen as a union of circles enclosing certain eigenvalues.

5. In many practical situations, we have $f_T = f$ for all T, in which case the uniform convergence condition is automatically satisfied.

4.2 Asymptotics of FPCA

Let the FPCA setting be as in Section 3.1. Recall that our focus is on the variance operator V and its estimator $\widehat{V}$, where

$$V = \mathbb{E}(X_t \otimes X_t) = \sum_{i=1}^{\infty} \lambda_i (v_i \otimes v_i), \quad \widehat{V} = \frac{1}{T} \sum_{t=1}^{T} (X_t \otimes X_t) = \sum_{i=1}^{T} \widehat{\lambda}_i (\widehat{v}_i \otimes \widehat{v}_i).$$

For identification, we impose the following assumption.

Assumption 1. *$\mathcal{N}(V) = 0$, and V has simple eigenvalues, so that the eigenvalues can be ordered as $\lambda_1 > \lambda_2 > \cdots > 0$.*

This assumption is standard in the functional data analysis literature. The condition $\mathcal{N}(V) = 0$ ensures that 0 is not an eigenvalue of V. This is innocuous, since if 0 were an eigenvalue, we could replace the Hilbert space H by its quotient with respect to the null space of V. As V is compact, self-adjoint, and positive definite, its eigenvalues are positive, isolated, and converge to 0. The "simple eigenvalues" condition guarantees that the ordering $\lambda_1 > \lambda_2 > \ldots$ is well-defined and that each eigenspace is simple (i.e., one-dimensional).

Let $\widehat{V}$ be a positive semi-definite, self-adjoint estimator of V. In the general setting, we assume that a central limit theorem holds for $\widehat{V}$. Sufficient conditions for this assumption will be provided in Section 5.

Assumption 2. *$\sqrt{T}(\widehat{V} - V) \to_d U =_d \mathbb{N}(0, K)$ for some $K \in (H \otimes H) \otimes (H \otimes H)$ as $T \to \infty$.*

Assumption 2 implies that $\widehat{V} \to_p V$ in the Hilbert–Schmidt norm. Since $\widehat{V}$ is of finite rank, its spectrum consists of finitely many nonnegative eigenvalues. Ordering these as $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq$

$\cdots \geq \widehat{\lambda}_T$ and setting $\widehat{\lambda}_i = 0$ for all $i > T$, we have $|\widehat{\lambda}_i - \lambda_i| \leq \|\widehat{V} - V\| \leq \|\widehat{V} - V\|_{HS}$, so that $\widehat{\lambda}_i \to_p \lambda_i$ uniformly in i. Consequently, for any $\epsilon > 0$, there exists T_0 sufficiently large such that $\mathbb{P}\Big(\widehat{\lambda}_1 > \widehat{\lambda}_2 > \cdots > \widehat{\lambda}_T > 0\Big) > 1 - \epsilon$ for all $T > T_0$. Moreover, if $\widehat{V} \to_{a.s.} V$, then $\mathbb{P}\Big(\widehat{\lambda}_1 > \widehat{\lambda}_2 > \cdots\Big) \to 1$ as $T \to \infty$. The sign of each eigenvector $\widehat{v}_i$ is set so that $\langle \widehat{v}_i, v_i \rangle \geq 0$.

To apply Theorem 1, we fix i and construct D and D_i as follows. Take D_i to be the open disk centered at λ_i with radius smaller than $\delta_i = \min\{\lambda_{i-1} - \lambda_i, \lambda_i - \lambda_{i+1}\}$. Let D_i^0 be the open disk centered at 0 with radius greater than λ_{i+1} but smaller than $\lambda_i - \delta_i$, and let D_i^1 be the open disk centered at λ_1 with radius greater than $\lambda_1 - \lambda_{i-1}$ but smaller than $\lambda_1 - \lambda_i - \delta_i$ (set $D_1^1 = \varnothing$). Define $D = D_i \cup D_i^0 \cup D_i^1$. By construction, $\sigma(V) \subset D$, $\sigma(V) \cap D_i = \lambda_i$, and $\overline{D_i} \cap \overline{D \setminus D_i} = \varnothing$. Let f be the indicator function 1_{D_i}. Note that f is holomorphic and bounded. Then, by the resolvent expansion and the Cauchy integral formula,

$$
\begin{aligned}
f(V) &= \frac{1}{2\pi i} \oint_{\Gamma_V} 1_{D_i}(z)(zI - A)^{-1} \mathrm{d}z \\
&= \frac{1}{2\pi i} \oint_{\Gamma_V} 1_{D_i}(z) \sum_{j=1}^{\infty} \frac{1}{z - \lambda_j} P_j \mathrm{d}z \\
&= P_i.
\end{aligned}
$$

Recall that $P_i = v_i \otimes v_i$ is the eigenprojection onto the eigenspace corresponding to λ_i, and $\widehat{P}_i = \widehat{v}_i \otimes \widehat{v}_i$ is its estimator.

Theorem 2. *Under Assumptions 1 and 2,*

$$\sqrt{T}(\widehat{\lambda}_i - \lambda_i) \to_d \langle U v_i, v_i \rangle,$$

$$\sqrt{T}(\widehat{v}_i - v_i) \to_d Q_i U v_i,$$

and

$$\sqrt{T}\left(\widehat{P}_i - P_i\right) \to_d P_i U Q_i + Q_i U P_i$$

where $P_i = v_i \otimes v_i$ *and* $Q_i = \sum_{l \neq i} \frac{1}{\lambda_i - \lambda_l} v_l \otimes v_l$. *The above convergences hold jointly.*

Although the assumption that the eigenvalues are simple is commonly made in the functional data analysis literature, our framework can readily accommodate the case of repeated eigenvalues. Recall that since V is self-adjoint, all its eigenvalues are real. We now replace Assumption 1 with the following.

Assumption 3. *$\mathcal{N}(V) = \{0\}$, and V has eigenvalues (with repetitions) $\lambda_1 \geq \lambda_2 \geq \dots$. The algebraic multiplicity of λ_i is denoted by m_i.*

The presence of eigenvalues with algebraic multiplicities greater than one introduces two complications. First, the eigenspace corresponding to an eigenvalue λ_i may be more than one dimensional. If we write $V = \sum_{i=1}^{\infty} \lambda_i v_i \otimes v_i$ and let $J_i = \{j \in \mathbb{N} : \lambda_j = \lambda_i\}$, then the eigenspace corresponding to λ_i is the subspace spanned by $\{v_j : j \in J_i\}$, and the projection onto this eigenspace is given by $P_i = \sum_{j \in J_i} v_j \otimes v_j$. In this case, the eigenvectors are not uniquely identified, even if restrictions on their length and sign are imposed. What is uniquely identified is the eigenspace, or equivalently, the eigenprojection P_i. We therefore focus on eigenvalues and eigenprojections in the asymptotic analysis rather than on individual eigenvectors.

Second, the algebraic multiplicities must be estimated in order to link the estimated eigenvalues with their associated projections. To account for these complications, we define the estimator $\widehat{P}_i$ as follows. Let c_T be a sequence of numbers such that $c_T \to 0$ and $\sqrt{T} c_T \to \infty$. Define $\widehat{J}_i = \{j \in \mathbb{N} : |\widehat{\lambda}_i - \widehat{\lambda}_j| < c_T\}$ and

$$\widehat{P}_i = \sum_{j \in \widehat{J}_i} \widehat{v}_j \otimes \widehat{v}_j.$$

To state the result, we use the following notation: For any positive semi-definite, self-adjoint, compact operator A, let $\lambda_j^{\circ}(A)$ denote its j-th eigenvalue and $P_j^{\circ}(A)$ denote the projection onto the corresponding eigenspace. Note that the eigenspace may be multidimensional if the eigenvalue is repeated.

Theorem 3. *Under Assumptions 2 and 3,*

$$\sqrt{T}\left(\widehat{P}_i - P_i\right) \to_d P_i U Q_i + Q_i U P_i$$

where $P_i = \sum_{l \in J_i} v_l \otimes v_l$ *and* $Q_i = \sum_{l \notin J_i} \frac{1}{\lambda_i - \lambda_l} v_l \otimes v_l$, *and for any* $j \in J_i$,

$$\sqrt{T}(\widehat{\lambda}_j - \lambda_i) \to_d \lambda^{\circ}_{j(i)}(P_i U P_i)$$

where $j(i) = j - k_i + 1, k_i = \min J_i$. *The above convergences hold jointly.*

4.3 Asymptotics of Singular Value Decomposition

The results in the previous section can also be applied to obtain asymptotic distributions of quantities in singular value decompositions of base operators that are compact but not necessarily self-adjoint. We follow the setting in Section 3.2 and suppose that $\widehat{A}$ is an estimator for a compact operator $A \in \mathcal{L}(H_1, H_2)$, and that both operators admit singular value decompositions

$$A = \sum_{i=1}^{\infty} \mu_i (u_i \otimes w_i) \quad \text{and} \quad \widehat{A} = \sum_{i=1}^{\infty} \widehat{\mu}_i (\widehat{u}_i \otimes \widehat{w}_i),$$

where μ_i and $\widehat{\mu}_i$ are arranged in nonnegative, descending order. We identify the sign of $\widehat{u}_i$ by requiring that $\langle \widehat{u}_i, u_i \rangle \geq 0$.

We make the following assumption on the asymptotic distribution of the estimator.

Assumption 4. $\sqrt{T}(\widehat{A} - A) \to_d \mathcal{U} =_d \mathbb{N}(0, \mathcal{K})$ *for some* $\mathcal{K} \in (H_1 \otimes H_2) \otimes (H_1 \otimes H_2)$ *as* $T \to \infty$.

To establish the asymptotic properties of the singular value decomposition, we first state the following lemma.

Lemma 1. *Under Assumption 4,*

$$\sqrt{T}(\widehat{A}^* \widehat{A} - A^* A) \to_d A^* \mathcal{U} + \mathcal{U}^* A,$$

and

$$\sqrt{T}(\widehat{A}\widehat{A}^* - AA^*) \to_d A\mathcal{U}^* + \mathcal{U}A^*.$$

We introduce the following identification assumptions for the singular value decomposition setting, corresponding to Assumption 1 in the FPCA setting.

Assumption 5. *$\mathcal{N}(A) = \{0\}, \mathcal{N}(A^*) = \{0\}$, and A has simple singular values so that we may order the singular values as $\mu_1 > \mu_2 > \cdots > 0$.*

Theorem 4. *Under Assumptions 4 and 5,*

$$\sqrt{T}(\widehat{\mu}_i - \mu_i) \to_d \frac{1}{2\mu_i}\langle (A^*\mathcal{U} + \mathcal{U}^*A)u_i, u_i\rangle =_d \frac{1}{2\mu_i}\langle (A\mathcal{U}^* + \mathcal{U}A^*)w_i, w_i\rangle,$$

$$\sqrt{T}(\widehat{u}_i - u_i) \to_d Q_i^u(A^*\mathcal{U} + \mathcal{U}^*A)u_i,$$

and

$$\sqrt{T}(\widehat{w}_i - w_i) \to_d Q_i^w(A\mathcal{U}^* + \mathcal{U}A^*)w_i$$

where $Q_i^u = \sum_{l\neq i} \frac{u_l \otimes u_l}{\mu_i^2 - \mu_l^2}$, and $Q_i^w = \sum_{l\neq i} \frac{w_l \otimes w_l}{\mu_i^2 - \mu_l^2}$. The above convergences hold jointly.

As in the previous section, we can also handle the case where A has repeated singular values.

Assumption 6. *$\mathcal{N}(A) = \{0\}, \mathcal{N}(A^*) = \{0\}$, and A has singular values (with repetitions) $\mu_1 \geq \mu_2 \geq \dots$. The algebraic multiplicity of μ_i is denoted by m_i.*

Let $\mathcal{J}_i = \{j \in \mathbb{N} : \mu_j = \mu_i\}$. As a consequence of Theorem 3 and Lemma 1, we obtain the following result.

Theorem 5. *Under Assumptions 4 and 6, for any $j \in \mathcal{J}_i$,*

$$\sqrt{T}(\widehat{\mu}_j - \mu_i) \to_d \frac{1}{2\mu_i}\lambda^\circ_{j(i)}\left(P_i^u(A^*\mathcal{U} + \mathcal{U}^*A)P_i^u\right) =_d \frac{1}{2\mu_i}\lambda^\circ_{j(i)}(P_i^w(A\mathcal{U}^* + \mathcal{U}A^*)P_i^w)$$

where $P_i^u = P_i^\circ(A^*A), P_i^w = P_i^\circ(AA^*)$, $j(i) = j - k_i + 1, k_i = \min \mathcal{J}_i$.

4.4 Asymptotics of FPLS

We follow the setting in Section 3.3. Recall that the problem of FPLS is essentially the problem of singular value decomposition of the operators

$$C^{(k)} = \mathbb{E}Y_t^{(k)} \otimes X_t^{(k)} = \sum_{i=1}^{\infty} \mu_i^{(k)} \left(u_i^{(k)} \otimes w_i^{(k)}\right)$$

and

$$\widehat{C}^{(k)} = \frac{1}{T}\sum_{t=1}^{T} Y_t^{(k)} \otimes X_t^{(k)} = \sum_{i=1}^{T} \widehat{\mu}_i^{(k)} \left(\widehat{u}_i^{(k)} \otimes \widehat{w}_i^{(k)}\right),$$

so we can directly apply the results from the previous section.

In this section, we focus on the case where the largest singular value of $C^{(k)}$ is simple. When the largest singular value is repeated, the k-th FPLS direction is not uniquely determined, and its asymptotic behavior depends on the specific direction chosen by the algorithm during implementation. Nonetheless, once a particular selection scheme is specified, its asymptotic behavior can be analyzed using the theory developed in the preceding sections.

Assumption 7. *For each k, $\mathcal{N}(C^{(k)}) = \{0\}$, $\mathcal{N}(C^{(k)*}) = \{0\}$, $C^{(k)}$ has no repeated singular values so that we can order as $\mu_1 > \mu_2 > \dots$, and $\sqrt{T}(\widehat{C}^{(k)} - C^{(k)}) \to_d \mathcal{U}^{(k)} =_d \mathbb{N}(0, \mathcal{K}^{(k)})$ for some $\mathcal{K}^{(k)} \in (H_1 \otimes H_2) \otimes (H_1 \otimes H_2)$ for each k as $T \to \infty$.*

Theorem 6. *Under Assumption 7,*

$$\sqrt{T}\left(\widehat{\mathsf{u}}^{(k)} - \mathsf{u}^{(k)}\right) \to_d Q^{uk}\left(C^{(k)*}\mathcal{U}^{(k)} + \mathcal{U}^{(k)*}C^{(k)}\right) u_1^{(k)},$$

and

$$\sqrt{T}\left(\widehat{\mathsf{w}}^{(k)} - \mathsf{w}^{(k)}\right) \to_d Q^{wk}\left(C^{(k)}\mathcal{U}^{(k)*} + \mathcal{U}^{(k)}C^{(k)*}\right) w_1^{k}$$

where $Q^{uk} = \sum_{l\geq 2} \frac{u_l^{(k)} \otimes u_l^{(k)}}{\mu_1^2 - \mu_l^2}$, and $Q^{wk} = \sum_{l\geq 2} \frac{w_l^{(k)} \otimes w_l^{(k)}}{\mu_1^2 - \mu_l^2}$. The above convergences hold jointly.

Assumption 7 is more involved than Assumption 4 because $C^{(k)}$ and $\widehat{C}^{(k)}$ are defined iteratively. However, as we shall show in Section 5, the weak convergence of $\widehat{C}^{(k)}$ can be established under mild conditions. Likewise, the limit distributions in Theorem 6 are defined iteratively, but in applications they can be estimated using their sample analogue. This makes inference based on these iteratively defined limit distributions entirely feasible.

4.5 Inverse Problem for Functional Data

We follow the setting in Section 3.4 and consider the inverse $V^+ = \sum_{i=1}^{\infty} q(\lambda_i)(v_i \otimes v_i)$ of the variance operator V as in (5), where $q(\cdot)$ is a regularized version of the (unbounded) reciprocal function $\frac{1}{\cdot}$. Let $\widehat{V}$ be a self-adjoint estimator of V. We estimate V^+ by

$$\widehat{V}^+ = \sum_{i=1}^{\infty} \widehat{q}(\widehat{\lambda}_i)(\widehat{v}_i \otimes \widehat{v}_i),$$

where $(\widehat{\lambda}_i, \widehat{v}_i)$ are the eigenpairs of $\widehat{V}$, and $\widehat{q}$ is a function (possibly data-dependent and therefore random) that approximates q.

Assumption 8. *The function q is such that $\oint_\Gamma q(z)(zI - V)^{-1}\mathrm{d}z$ is well defined where Γ is a contour that surrounds $\sigma(V)$ and $\sup_{z\in\Gamma} |\widehat{q}(z) - q(z)| = o_p(T^{-1/2})$.*

The first part of the assumption guarantees that the operator V^+ is well defined. For example, in the case where q takes the form $q(z) = \frac{1}{z}1_{(\alpha,\infty)}(z)$ or $q(z) = \frac{1}{z+\alpha}$, the assumption is satisfied. The second part of the assumption controls the convergence rate of a possibly estimated $\widehat{q}$. In many applications, we have $\widehat{q} = q$, in which case this assumption holds trivially. Noting that $V^+ = q(V)$ and $\widehat{V}^+ = \widehat{q}(\widehat{V})$, we may apply Theorem 1 under Assumptions 1, 2, and 8 to obtain

$$\sqrt{T}\left(\widehat{V}^+ - V^+\right) \to_d \frac{1}{2\pi i}\oint_\Gamma q(z)(zI - V)^{-1}U(zI - V)^{-1}\,\mathrm{d}z,$$

where U is an $\mathbb{N}(0, K)$ random element defined as in Assumption 2.

A more explicit expression of the limit distribution can be derived once a particular form of q is specified.

Theorem 7. *Under Assumptions 1, 2 and 8, if $q(z) = \frac{1}{z+\alpha}$ for some $\alpha \neq 0$, then*

$$\sqrt{T}\left(\widehat{V}^+ - V^+\right) \to_d \mathcal{S}U\mathcal{S}$$

where $\mathcal{S} = \sum_{i=1}^{\infty} \frac{1}{\lambda_i + \alpha} P_i$*. If* $q(z) = \frac{1}{z} 1_{(\alpha,\infty)}(z)$ *for some* $\alpha > 0$*, then*

$$\sqrt{T}\left(\widehat{V}^{+} - V^{+}\right) \to_d \sum_{\{i:\lambda_i > \alpha\}} (\mathcal{P}_i U Q_i + Q_i U \mathcal{P}_i - \mathcal{P}_i U \mathcal{P}_i)$$

where $\mathcal{P}_i = \frac{1}{\lambda_i} P_i$*, and* $Q_i = \sum_{l \neq i} \frac{1}{\lambda_i - \lambda_l} P_i$*.*

We make the following remarks.

1. In the case $q(z) = \frac{1}{z+\alpha}$, if we allow $\alpha = \alpha_T$ to vary with the sample size T, then $\|\widehat{V}^{+} - V^{+}\| = O_p(\alpha_T^{-2} T^{-1/2})$.
2. In the case $q(z) = \frac{1}{z} 1_{(\alpha,\infty)}$, if we allow $\alpha = \alpha_T$ to vary with sample size T, and define $K_T = \max\{i \in \mathbb{N} : \lambda_i > 1/\alpha_T\}$, then $\|\widehat{V}^{+} - V^{+}\| = O_p(T^{-1/2} \sum_{i=1}^{K_T} \lambda_i^{-1} \delta_i^{-1})$ where $\delta_i = \min\{\lambda_i - \lambda_{i+1}, \lambda_{i-1} - \lambda_i\}$.

4.6 Representation

It would be useful to represent the limit quantities that involve operators in the above theorems in more convenient one-dimensional forms. To do so, we utilize the Karhunen-Loeve expansion, which says that for any H-valued random element X_t, it can be represented as

$$X_t = \mathbb{E}X_t + \sum_{i=1}^{\infty} Z_{ti} v_i$$

where Z_{ti} is an array of real valued random variables such that $\mathbb{E}Z_{ti}^2 = \lambda_i$ and $\mathbb{E}Z_{ti}Z_{tj} = 0$ for $i \neq j$, and (λ_i, v_i) is the eigen-pair of the variance operator $V = \mathbb{E}(X_t - \mathbb{E}X_t) \otimes (X_t - \mathbb{E}X_t)$ of X_t. Note that $\mathbb{E}Z_{ti}^4 - (\mathbb{E}Z_{ti}^2)^2$ is the variance of Z_{ti}^2 and that $\langle X_t - \mathbb{E}X_t, v_i \rangle = Z_{ti}$.

I shall give results corresponding to Theorem 2 as examples. The representation results corresponding to the other theorems can be derived as long as one find the correct X_t and V to work with. For a general process $\{X_t\}$, its long run variance is defined by $\Omega(X_t) = \sum_{h=-\infty}^{\infty} (\mathbb{E}X_h X_0 - \mathbb{E}X_t \mathbb{E}X_0)$.

Theorem 8. *Suppose that X_t has the Karhunen-Loeve expansion as above. Then under the assumptions of Theorem 2, we have that*

1. $\langle Uv_i, v_i\rangle =_d \mathbb{N}(0, \Omega(Z_{ti}^2))$;
2. *For any $v \in H$,*
$$\langle Q_i U v_i, v\rangle =_d \mathbb{N}\left(0, \Omega\left(\sum_{j\neq i} \frac{Z_{ti}Z_{tj}\langle v_j, v\rangle}{\lambda_i - \lambda_j}\right)\right).$$
In particular, $\langle Q_i U v_i, v_j\rangle =_d \mathbb{N}\left(0, \frac{Z_{ti}Z_{tj}}{\lambda_i-\lambda_j}\right)$ if $j \neq i$, and $\langle Q_i U v_i, v_j\rangle$ is degenerate at 0 if $j = i$;
3. *For any $u, v \in H$,*
$$\Big\langle (P_i U Q_i + Q_i U P_i)v, u\Big\rangle =_d \mathbb{N}\left(0, \Omega\left(\sum_{j\neq i} \frac{Z_{ti}Z_{tj}\Big(\langle v_i, v\rangle\langle v_j, u\rangle + \langle v_j, v\rangle\langle v_i, u\rangle\Big)}{\lambda_i - \lambda_j}\right)\right).$$
In particular, $\Big\langle (P_i U Q_i + Q_i U P_i)v_j, v_k\Big\rangle =_d \mathbb{N}\left(0, \Omega\left(\frac{Z_{ti}Z_{tk}}{\lambda_i-\lambda_k}\right)\right)$ if $j = i, k \neq i$, $\Big\langle (P_i U Q_i + Q_i U P_i)v_j, v_k\Big\rangle =_d \mathbb{N}\left(0, \Omega\left(\frac{Z_{ti}Z_{tj}}{\lambda_i-\lambda_j}\right)\right)$ if $j \neq i, k = i$, and $\Big\langle (P_i U Q_i + Q_i U P_i)v_j, v_k\Big\rangle$ is degenerate at 0 for other combinations of j and k. Note that all the distributions in this theorem are normal distributions on $\mathbb{R}$.

5 Central Limit Theorems

In this section I give a central limit theorem for the sample base operators under weak dependence conditions.

There are different schemes to introduce weak dependence in the data that does not break the usual limit theorems. We employ two schemes in the following. One is through mixing conditions, which is frequently used in the literature of dependent data. The other is L^p-m-approximability developed by Kokoszka & Reimherr (2013).

Let $\{X_t\}$ be a sequence of H-valued random elements. Let $\mathcal{F}_m^n = \sigma(X_t, m \leq t \leq n)$, where m, n could be possibly $\pm\infty$. We say that the sequence $\{X_t\}$ is α-mixing if $\alpha(k) = \sup_n \sup_{A\in\mathcal{F}_{-\infty}^n, B\in\mathcal{F}_{n+k}^\infty} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \to 0$ as $k\to\infty$. We say that $\alpha(k)$ is of size $-r$ if $\alpha(k) = O(k^{-r})$. We say that the sequence $\{X_t\}$ is ϕ-mixing if $\phi(k) = \sup_n \sup_{A\in\mathcal{F}_{-\infty}^n, B\in\mathcal{F}_{n+k}^\infty, \mathbb{P}(A)>0} |\mathbb{P}(B|A) - \mathbb{P}(B)| \to 0$ as $k\to\infty$. The size of $\phi(k)$ is defined similarly.

Let X_t be a sequence of H-valued random elements in $L^P(H)$ such that it admits the representation $X_t = f(\varepsilon_t, \varepsilon_{t-1}, \cdots)$ for an independent and identically distributed S-valued random elements, where S is a measurable space and f is a measurable mapping. Suppose for each t there is an independent copy $\{\varepsilon_i^{(t)}\}$ of $\{\varepsilon_i\}$ such that $X_t^{(m)}$ defined by $X_t^{(m)} = f(\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-m+1}, \varepsilon_{t-m}^{(t)}, \varepsilon_{t-m-1}^{(t)}, \cdots)$ satisfies $\sum_{m=1}^\infty \left(\mathbb{E}\|X_t - X_t^{(m)}\|^p\right)^{1/p} < \infty$. In this case we call $\{X_t\}$ an L^p-m-approximable process. Kokoszka & Reimherr (2013) show that functional linear processes and functional ARCH processes with iid innovations are L^p-m-approximable under suitable regularity conditions.

Central limit theorems for α- or ϕ-mixing strictly stationary H-valued random elements are established in Dehling (1983). The author also gives conditions under which central limit theorem holds for weakly stationary α- or ϕ-mixing sequences. The central limit theorem for L^4-m-approximable sequence is given in Kokoszka & Reimherr (2013). Note that an L^4-m-approximable sequence is by definition strictly stationary. Other possible conditions includes the mixingale condition as in Chen & White (1998).

In the next I establish weak convergences of variance and covariance operators of such processes. Let $\{Y_t\}$ and $\{X_t\}$ be processes of H_1 and H_2-valued random elements.

Assumption 9. *Suppose that one of the following conditions hold.*

1. $\{(Y_t, X_t)\}$ *is an α-mixing strictly stationary sequence such that* $\mathbb{E}\|Y_t\|^{4+2\delta} < \infty$, $\mathbb{E}\|X_t\|^{4+2\delta} < \infty$ *for some* $\delta > 0$, *and the α-mixing coefficients α_k satisfies* $\sum_{k=1}^\infty \alpha_k^{\frac{\delta}{2+\delta}} < \infty$.

2. $\{(Y_t, X_t)\}$ *is a* ϕ*-mixing strictly stationary sequence such that* $\mathbb{E}\,\|Y_t\|^4 < \infty$, $\mathbb{E}\,\|X_t\|^4 < \infty$, *and the* ϕ*-mixing coefficients* $\phi(k)$ *satisfies* $\sum_{k=1}^{\infty} \phi_k^{\frac{1}{2}} < \infty$.

3. $\{(Y_t, X_t)\}$ *is an* L^4*-m-approximable sequence.*

Write $\overline{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$ and $\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$. The estimator for the covariance operator $\mathbb{E}[(Y_t - \mathbb{E}Y_t) \otimes (X_t - \mathbb{E}X_t)]$ is given by

$$\widehat{A} = \frac{1}{T}\sum_{t=1}^{T}\left[\left(Y_t - \overline{Y}\right) \otimes \left(X_t - \overline{X}\right)\right].$$

Theorem 9. *Under Assumptions 9,*

$$\sqrt{T}(\widehat{A} - A) \to_d \mathbb{N}\left(0, \sum_{h=-\infty}^{\infty} \kappa(h)\right)$$

where $\kappa : \mathbb{Z} \to (H_1 \otimes H_2) \otimes (H_1 \otimes H_2)$ *is defined by*

$$\begin{aligned}\kappa(h) =& \mathbb{E}\Bigg([(Y_h - \mathbb{E}Y_h) \otimes (X_h - \mathbb{E}X_h)] \otimes [(Y_0 - \mathbb{E}Y_0) \otimes (X_0 - \mathbb{E}X_0)]\Bigg) \\ &- \Bigg(\mathbb{E}[(Y_h - \mathbb{E}Y_h) \otimes (X_h - \mathbb{E}X_h)]\Bigg) \otimes \Bigg(\mathbb{E}[(Y_0 - \mathbb{E}Y_0) \otimes (X_0 - \mathbb{E}X_0)]\Bigg).\end{aligned}$$

Apparently, this theorem incorporates the weak convergence of sample variance opeartor $\widehat{V} = \frac{1}{T}\sum_{t=1}^{T}[(X_t - \overline{X}) \otimes (X_t - \overline{X})]$ by taking $Y_t = X_t$. Recall that we identify $H_1 \otimes H_2$ with the space $\mathcal{L}_{HS}(H_1, H_2)$ of all Hilbert-Schmidt operators on H. Since $\mathcal{L}_{HS}(H_1, H_2)$ is again a separable Hilbert space, quantities such as $[(Y_h - \mathbb{E}Y_h) \otimes (X_h - \mathbb{E}X_h)] \otimes [(Y_0 - \mathbb{E}Y_0) \otimes (X_0 - \mathbb{E}X_0)]$ and therefore $\kappa(h)$ are well defined and can be viewed as operators in the space $\mathcal{L}_{HS}(\mathcal{L}_{HS}(H_1, H_2))$ of Hilbert-Schmidt operators on $\mathcal{L}_{HS}(H_1, H_2)$. Apparently, $\kappa(h)$ can be viewed as the autocovariance function of the $H_1 \otimes H_2$-valued process $\{Y_t \otimes X_t\}$, and $\sum_{h=-\infty}^{\infty} \kappa(h)$ can be viewed as the its long-run variance.

Using the above result, we can iteratively show that in the FPLS setting, $\widehat{A}^{(k)}$ converges weakly as in Assumption 7.

Theorem 10. *Let $\widehat{C}^{(k)}$ and $C^{(k)}$ be defined as in Section 3.3. Under Assumption 9,*

$$\sqrt{T}(\widehat{C}^{(k)} - C^{(k)}) \to_d \mathbb{N}\left(0, \sum_{h=-\infty}^{\infty} \kappa^{(k)}(h)\right)$$

where $\kappa^{(k)} : \mathbb{Z} \to (H_1 \otimes H_2) \otimes (H_1 \otimes H_2)$ is defined by

$$\begin{aligned}\kappa(h) = \mathbb{E}\Bigg([(Y_h^{(k)} - \mathbb{E}Y_h^{(k)}) \otimes (X_h^{(k)} - \mathbb{E}X_h^{(k)})] \otimes [(Y_0^{(k)} - \mathbb{E}Y_0^{(k)}) \otimes (X_0^{(k)} - \mathbb{E}X_0^{(k)})]\Bigg) \\ - \Bigg(\mathbb{E}[(Y_h^{(k)} - \mathbb{E}Y_h^{(k)}) \otimes (X_h^{(k)} - \mathbb{E}X_h^{(k)})]\Bigg) \otimes \Bigg(\mathbb{E}[(Y_0^{(k)} - \mathbb{E}Y_0^{(k)}) \otimes (X_0^{(k)} - \mathbb{E}X_0^{(k)})]\Bigg).\end{aligned}$$

To conduct statistical inferences using the above results, we need to estimate the long run variance operator $\Omega_{Y\otimes X} = \sum_{h=-\infty}^{\infty} \kappa(h)$. For example, the autocovariance operator in the limit distribution in Theorem 9 can be estimated by

$$\begin{aligned}\widehat{\kappa}(h) =& \frac{1}{T}\sum_{t=h+1}^{T}\Bigg[\Bigg(\left(Y_t - \overline{Y}\right) \otimes \left(X_t - \overline{X}\right) - \frac{1}{T}\sum_{s=1}^{T}\left[\left(Y_s - \overline{Y}\right) \otimes \left(X_s - \overline{X}\right)\right]\Bigg) \\ &\otimes \Bigg(\left(Y_{t-h} - \overline{Y}\right) \otimes \left(X_{t-h} - \overline{X}\right) - \frac{1}{T}\sum_{s=1}^{T}\left[\left(Y_s - \overline{Y}\right) \otimes \left(X_s - \overline{X}\right)\right]\Bigg)\Bigg]\end{aligned}$$

for $0 \le h \le T-1$, and $\widehat{\kappa}(h) = \widehat{\kappa}(-h)^*$ for $-(T-1) \le h < 0$. We then estimate the long run variance by

$$\widehat{\Omega}_{Y\otimes X} = \sum_{|h|\le(T-1)} w(b_T h)\widehat{\kappa}(h)$$

where w is a suitable window function and b_T is the bandwidth parameter. To establish consistency of the long run variance operator, we review the definition of cumulants.

Let $\{X_t\}$ be an H-valued processes. We define its p-th order cumulants ($p \le n$) by $\Theta(t_1, \ldots, t_p) = \sum_\pi (|\pi| - 1)!(-1)^{|\pi|-1} \bigotimes_{\Phi\in\pi} \mathbb{E} \bigotimes_{i\in\Phi} X_{t_i}$ where π runs through all partitions of $\{1, \ldots, p\}$, Φ runs through all partition blocks in π, and $|\pi|$ is the number of partition blocks in π. When $\{X_t\}$ is strictly stationary, we have $\Theta(t_1, t_2, \ldots, t_p) = \Theta(0, t_2-t_1, \ldots, t_p-t_1)$. We write this common value as $Q(t_2-t_1, \ldots, t_p-t_1)$. Note that $Q(k_1, k_2, \ldots, k_q) : \mathbb{Z}^q \to \bigotimes_{i=1}^{q+1} H$ is a $q+1$-th order cumulant.

Assumption 10. *Suppose that the followings hold.*

1. $Y_t \in L^8(H_1), X_t \in L^8(H)$ *and the fourth order cumulant* $Q(r,s,t)$ *of the process* $\{(Y_t - \mathbb{E}Y_t) \otimes (X_t - \mathbb{E}X_t)\}$ *is absolutely summable, i.e.,* $\sum_{r,s,t=-\infty}^{\infty} \|Q(r,s,t)\| < \infty$.
2. $w : \mathbb{R} \to \mathbb{R}_+$ *is an even, bounded, square integrable function such that* $w(0) = 1$ *and that for every* b *and* T *we have* $b \sum_{|h|<T} w(bh) \leq C(bT)^{1/2-\epsilon}$ *for some* $\epsilon > 0$
3. $\sum_{h=-\infty}^{\infty} |h|^q \|\kappa(h)\| < \infty$ *for some* $q > 0$.
4. *There exists positive integer* $r \geq q$ *such that* $\lim_{z\to 0} \frac{1-w(z)}{|z|^r} < \infty$ *and is nonzero.*
5. $b_T \to 0, b_T T \to \infty$, *and* $0 < \lim_{T\to\infty} b_T^{1+2q} T < \infty$.

Theorem 11. *Under Assumptions 9 and 10, we have*

$$\widehat{\Omega}_{Y\otimes X} \to_p \sum_{h=-\infty}^{\infty} \kappa(h).$$

We note here that the convergence in probability in the above theorem should be viewed as in the Hilbert-Schmidt norm in the space $\mathcal{L}_{HS}(\mathcal{L}_{HS}(H_1, H_2))$. The long run variance in the limit distribution of Theorem 10 can be estimated similarly.

6 Discussion

In this paper, I have developed a general asymptotic theory of spectral component analysis for weakly stationary functional data. In cases where the functional data are strongly dependent so that the usual central limit theorem does not hold for estimators of base operators such as variance and covariance operators, the delta method developed in the previous sections cannot be applied. However, for such nonstationary processes, if some form of functional central limit theorem holds for the partial sum process, then the continuous mapping theorem can be used directly to obtain the asymptotic distributions of spectral components. This approach has been explored in Chang et al. (2016) and Chang et al. (2025).

As an example, let $\{\varepsilon_t\}$ be an i.i.d. sequence of H-valued random elements with mean zero and variance operator Σ, and define $X_t = \sum_{s=1}^{t} \varepsilon_s$. The process $\{X_t\}$, being an H-valued random walk, is clearly neither stationary nor weakly dependent. However, a functional central limit theorem holds for X_t in this case. If we define

$$W_T(r) = \frac{1}{\sqrt{T}} X_{\lfloor Tr \rfloor} := \frac{1}{\sqrt{T}} \sum_{s=1}^{\lfloor Tr \rfloor} \varepsilon_s, \quad r \in [0,1],$$

where $\lfloor Tr \rfloor$ denotes the integer part of Tr, then $W_T(r) \to_d W$ as $T \to \infty$, where W is a Brownian motion on H with variance structure Σ. This weak convergence is understood as the convergence of H-valued stochastic processes to an H-valued limiting process.

If we formulate the sample variance operator as $\widehat{V} = \frac{1}{T} \sum_{t=1}^{T} X_t \otimes X_t$, then instead of $\sqrt{T}(\widehat{V} - V)$ converging weakly to a Gaussian random element, we have

$$\frac{1}{T} \widehat{V} \xrightarrow{d} \int_0^1 (W \otimes W)(r) \, \mathrm{d}r := \mathsf{W},$$

where $W \otimes W$ is an $H \otimes H$-valued stochastic process and the integral is in the Riemann–Stieltjes sense. Since the operations of taking eigenvalues and eigenvectors are continuous, the continuous mapping theorem can be applied directly to yield

$$\left(\frac{1}{T} \widehat{\lambda}_i, \widehat{P}_i \right) \to_d \left(\lambda_i^{\circ}(\mathsf{W}), P_i^{\circ}(\mathsf{W}) \right).$$

7 Disclosure statement

The authors report there are no competing interests to declare.

SUPPLEMENTARY MATERIAL

Proofs of all lemmas and theorems are provided in the supplementary materials.

References

Bauer, F., Pereverzev, S. & Rosasco, L. (2007), 'On regularization algorithms in learning theory', *Journal of Complexity* **23**(1), 52–72.

Boente, G. & Fraiman, R. (2000), 'Kernel-based functional principal components', *Statistics & Probability Letters* **48**(4), 335–345.

Bosq, D. (2000), *Linear Processes in Function Spaces: Theory and Applications*, Lecture Notes in Statistics 149, Springer, New York.

Chang, Y., Hu, B. & Park, J. Y. (2025), 'Econometric analysis of persistent functional dynamics', Manuscript.

Chang, Y., Kim, C. S. & Park, J. Y. (2016), 'Nonstationarity in time series of state densities', *Journal of Econometrics* **192**, 152–167.

Chen, X. & White, H. (1998), 'Central limit and functional central limit theorems for hilbert-valued dependent heterogeneous arrays with applications', *Econometric Theory* **14**(2), 260–284.

Conway, J. B. (2000), *A Course in Functional Analysis*, 2nd edn, Springer.

Dauxois, J., Pousse, A. & 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**, 136–154.

Dehling, H. (1983), 'Limit theorems for sums of weakly dependent banach space valued random variables', *Zeitschrift ftir Wahrscheinlichkeitstheorie und verwandte Gebiete* **63**, 393–432.

Delaigle, A. & Hall, P. (2012), 'Methodology and theory for partial least squares applied to functional data', *Annals of Statistics* **40**(1), 322–352.

Dicker, L. H., Foster, D. P. & Hsu, D. (2017), 'Kernel ridge vs. principal component regression: Minimax bounds and the qualification of regularization operators', *Electronic Journal of Statistics* **11**(1), 1022–1047.

Ferraty, F. & Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer, New York.

Gilliam, D., Hohage, T., Ji, X. & Ruymgaart, F. (2009), 'The frechet derivative of an analytic function of a bounded operator with some applications', *International Journal of Mathematics and Mathematical Sciences* **2009**.

Gohberg, I., Goldberg, S. & Kaashoek, M. A. (1990), *Classes of Linear Operators*, Vol. I, Birkhäuser, Basel.

Hall, P. & Hosseini-Nasab, M. (2006), 'On properties of functional principal components analysis', *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **68**(1), 109–126.

Hörmann, S., Kidziński, Ł. & Hallin, M. (2015), 'Dynamic functional principal components', *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **77**(2), 319–348.

Horväth, L. & Kokoszka, P. (2012), *Inference for Functional Data with Applications*, Springer Series in Statistics, Springer, New York.

Kleffe, J. (1973), 'Principal components of random variables with values in a seperable Hilbert space', *Mathematische Operationsforschung und Statistik* **4**(5), 391–406.

Kokoszka, P. & Reimherr, M. (2013), 'Asymptotic normality of the principal components of functional time series', *Stochastic Processes and Their Applications* **123**, 1546–1562.

Kokoszka, P. & Reimherr, M. (2017), *Introduction to Functional Data Analysis*, Texts in Statistical Science, CRC Press.

Mas, A. (2002), 'Weak convergence for the covariance operators of a Hilbertian linear process', *Stochastic Processes and Their Applications* **99**(1), 117–135.

Panaretos, V. M. & Tavakoli, S. (2013), 'Cramér–karhunen–loève representation and harmonic principal component analysis of functional time series', *Stochastic Processes and Their Applications* **123**(7), 2779–2807.

Preda, C. & Saporta, G. (2005), 'Pls regression on a stochastic process', *Computational Statistics & Data Analysis* **48**(1), 149–158.

Ramsey, J. O. & Silverman, B. W. (2005), *Functional Data Analysis*, Springer Series in Statistics, 2nd edn, Springer-Verlag, New York.

Rudin, W. (1987), *Real and Complex Analysis*, 3rd edn, McGraw-Hill Companies, Inc.

Rudin, W. (1991), *Functional Analysis*, 2nd edn, McGraw-Hill.

Vakhania, N. N., Tarieladze, V. I. & Chobanyan, S. A. (1987), *Probability Distributions on Banach Spaces*, D. Reidel Publishing Company.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press.

Vito, E. D., Rosasco, L., Caponnetto, A., Giovannini, U. D. & Odone, F. (2005), 'Learning from examples as an inverse problem', *Journal of Machine Learning Research* **6**(30), 883–904.