Nonparametric Estimation of Smoothed Principal Components Analysis of Sampled Noisy Functions

Hervé Cardot

Unité de Biométrie et Intelligence Artificielle,

INRA Toulouse, BP 27

F-31326 Castanet-Tolosan Cedex, FRANCE

e-mail: cardot@toulouse.inra.fr

(To appear in the Journal of Nonparametric Statistics)

Abstract

This study deals with the simultaneous nonparametric estimations of n curves or observations of a random process corrupted by noise in which sample paths belong to a finite dimension functional subspace. The estimation, by means of B-splines, leads to a new kind of functional principal components analysis. Asymptotic rates of convergence are given for the mean and the eigenelements of the empirical covariance operator. Heuristic arguments show that a well chosen smoothing parameter may improve the estimation of the subspace which contains the sample path of the process. Finally, simulations suggest that the estimation method studied here is advantageous when there are a small number of design points.

Key words: functional principal component analysis, nonparametric regression, rates of convergence, mean square error, asymptotic expansion, hybrid splines, B-splines.

1 Introduction

This paper aims at giving the asymptotic rates of convergence in mean integrated square error (MISE) of the nonparametric estimation of the second order characteristics of several independent and identically distributed sampled curves lying in a common finite dimensional subspace of smooth functions. Futhermore, we suppose that curves are corrupted by a white noise at the (non random) design points. In applications, this may be the case when the data are sampled growth curves

(Boularan *et al.* 1993), meteorological data (Ramsay & Dalzell, 1991) or econometric data (Kneip, 1995). The reader is also referred to the book of Ramsay & Silverman (1997) which presents many statistical models for such functional data.

The method of estimation examinated here has been previously studied, in a pratical way, by Besse *et al.* (1997). Their estimates are constructed by means of hybrid splines (see Kelly & Rice (1990) for a definition and Diack & Thomas-Agnan (1997) for the use of these functions to test convexity) and lead to a new functional Principal Components Analysis (PCA). Besides, this PCA is adapted to the case where the data are unbalanced. They show by simulations that this new method can improve the estimation of the signal compared to usual nonparametric methods such as smoothing splines.

Asymptotic properties of the PCA of Hilbert valued random variables have already been investigated by Dauxois *et al.* (1982) and various authors (see e.g. Bosq 1991, Pezzulli & Silverman 1993, Silverman 1996). When the observations are discretized, Biritxinaga (1987) and Besse (1991) have demonstrated the convergence of the spline estimates of functional PCA. However, they did not assume that the curves where corrupted by sampling noise at the design points and they did not discuss the links between the asymptotic behaviour of the MISE and the different parameters such as the number of curves and design points and the smoothing parameter value. In this article, we demonstrate the convergence of our estimates and bound the MISE of the mean and the eigenelements of the smooth covariance operator as a function of all these parameters. Then we demonstrate how smoothing can improve the accuracy of the estimates.

The organization of the paper is as follows. In section 2, we establish notations, describe the model and explain how the estimates are constructed. In section 3, we give rates of convergence for the mean and the eigenfunctions of the covariance operator. In section 4, heuristic arguments based on perturbation theory (as in Pezzulli & Silverman, 1993 and Silverman, 1996) are used to show how smoothing may improve our estimates. A simulation study is presented in section 5 where it is noted that it can be advantageous to smooth when there is a small number of design points. Finally, section 6 summarizes the proofs of our propositions. More details may be found in Cardot (1997).

2 Model and estimation

2.1 Notations

Consider a random function Z defined on the probability space (Ω, A, P) and taking values in the separable Hilbert space $L^2[0,1]$ equipped with the inner product:

$$\forall f, g \in L^2[0,1] \quad \langle f, g \rangle = \int_0^1 f(t)g(t)\, dt.$$

Let's denote by $||.||_{L^2}$ the norm induced by this inner product and define $f \otimes g$ to be the rank one operator which satisfies:

$$\forall f, g, h \in L^2[0,1] \quad f \otimes g(h) = \langle f, h \rangle \, g.$$

In order to approximate the sample paths, we use normalized B-splines (De Boor 1978, Schumaker 1981). Denote by $\mathcal{S}_{k\nu}$ the space of splines functions of degree ν with k equispaced knots. It is well known that $\mathcal{S}_{k\nu}$ has a basis consisting of $r = k + \nu$ normalized B-splines, $B_{kj}(t),\ j = 1, \ldots, k+\nu$. Let $\mathbf{B}_k(t)$ be the vector[1] of $\{B_{kj}(t),\ j = 1, \ldots, k+\nu\}$ and consider the Gram matrices

$$[\mathbf{C}_k]_{jl} \; = \; \int_0^1 B_{kj}(t) B_{kl}(t) dt \ , \ j,l = 1, \ldots, r \ , \tag{1}$$

$$\left[\widetilde{\mathbf{C}}_k\right]_{jl} \; = \; \frac{1}{p} \sum_{i=1}^{p} B_{kj}(t_i) B_{kl}(t_i) \ , \ j,l = 1, \ldots, r, \tag{2}$$

where $t_1, \ldots, t_p$ are the design points.

Define for all integer m smaller than ν, the matrix $\mathbf{G}_k$:

$$[\mathbf{G}_k]_{jl} \; = \; \int B_{kj}^{(m)}(t) B_{kl}^{(m)}(t) dt \ , \ j,l = 1, \ldots, r \ . \tag{3}$$

It is the matrix associated to the semi-norm defined by the differential operator D^m which measure the smoothness of any function belonging to $\mathcal{S}_{k\nu}$.

In this article, the norm $||.||$ will denote either

- the euclidian norm, $\quad \forall \mathbf{b} \in \mathbb{R}^p, \ \ ||\mathbf{b}||^2 = \sum_{j=1}^p b_j^2,$
- the usual matrix norm, that is for any real matrix $\mathbf{A}$, we have $||\mathbf{A}|| = \sup_{||\mathbf{x}||=1} ||\mathbf{A} \ \mathbf{x}||$ where $||\mathbf{x}||$ is the euclidian norm of vector $\mathbf{x}$.
- the usual norm for bounded linear operators defined on $L^2[0,1]$.

2.2 Model

Let $S = \{S(t), t \in [0,1]\}$, be a "smooth" second order random function composed of a deterministic component $\mu(t)$ and a centered random residual part $Z(t)$:

$$S(t) \; = \; \mu(t) + Z(t), \quad t \in [0,1]. \tag{4}$$

The function μ is the mean of S and the random part Z is supposed to belong to a finite dimensional space, say E_q, of smooth functions.

Suppose we observe n independent realizations of this signal corrupted by a white noise ϵ at the sampling points $0 \leq t_1 < \ldots < t_p \leq 1$. Then, we have the data $\mathbf{Y}_i \in \mathbb{R}^p$:

$$\mathbf{Y}_i \; = \; \boldsymbol{\mu} + \mathbf{Z}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n \tag{5}$$

[1] In the succeding, we will use bold face letters to denote vector or matrices

where :

$$\begin{cases} Y_i(t_j) = \mu(t_j) + Z_i(t_j) + \epsilon_{ij} & i = 1, \dots, n\ ,\ j = 1, \dots, p; \\ Z_i \in E_q,\ \mathbb{E}(Z_i) = 0,\ \mathbb{E}\, \|Z_i\|^2_{L^2} < +\infty & E_q \subset L^2[0,1],\ \dim E_q = q; \\ Z_i \text{ independent of } \epsilon_{ij} & \mathbb{E}(\epsilon_{ij}) = 0,\ \ E(\epsilon_{ij}\epsilon_{i'j'}) = \sigma^2 \delta_{ii'}\delta_{jj'}. \end{cases} \tag{6}$$

For sake of simplicity, the distribution of the design points is supposed to be uniform on $[0, 1]$:

$(\mathbf{A}_1)$ $\quad t_j = \dfrac{j-1}{p-1}, \qquad j = 1, 2, \dots, p.$

It is also possible to get similar convergence results as those obtained in section 3 by assuming a weaker condition on the design points (Agarwall & Studden, 1980), at the expense, however, of more complicated expressions. Nevertheless, the estimators defined in next section are perfectly adapted to this situation.

The process Z is a second order process and therefore it can be written, in quadratic mean, for a basis of orthonormal functions $\{\phi_1, \dots, \phi_q\}$ of the subspace E_q :

$$Z(t) = \sum_{j=1}^{q} \langle Z, \phi_j \rangle\, \phi_j(t). \tag{7}$$

The smoothness assumption on the signal S is then ensured by assuming that μ and each function ϕ_j satisfy the condition:

$(\mathbf{A}_2)$ *μ and $\phi_1, \dots, \phi_q$ belong to $C^m[0,1]$, that is to say they have m continuous derivatives.*

This article aims at demonstrating the convergence of the estimates of μ and $\phi_1, \dots, \phi_q$ deduced from the observation of the vectors $\{\mathbf{Y}_i, i = 1, \dots, n\}$.

One way to compute the functions ϕ_j is to perform the spectral analysis of the covariance operator of the random function Z:

$$\Gamma = \mathbb{E}\,(Z \otimes Z)\,. \tag{8}$$

In that case, these functions are the eigenfunctions of Γ associated to the eigenvalues $\lambda_j = \mathbb{E}\left(\langle Z, \phi_j\rangle^2\right)$ and it is called the Principal Components Analysis (PCA) or the Karhunen-Loeve expansion of Z (Dauxois *et al.* 1982).

2.3 Construction of the estimates

Fix $\nu = m + 1$, and denote by $\widehat{y}_i$ the least squares estimate of the ith sample path in the B-splines basis. It can be written as follows:

$$\begin{aligned} \widehat{y}_i(t) &= \hat{\mathbf{b}}_i' \widetilde{\mathbf{C}}_k^{-1} \mathbf{B}_k(t) \\ &= \hat{\mathbf{s}}_i' \mathbf{B}_k(t), \end{aligned} \tag{9}$$

where $\hat{\mathbf{b}}_i = \frac{1}{p}\sum_{j=1}^p Y_{ij}\mathbf{B}_k(t_j)$ and $\hat{\mathbf{s}}_i = \widetilde{\mathbf{C}}_k^{-1}\hat{\mathbf{b}}_i$. The number of knots k depends on p and n but these indices will omitted for sake of simplicity.

The data specificity suggests two kinds of constraints in the estimation procedure. On the one hand, a dimensionality constraint must assume that curves only span a finite dimensional subspace. On the other hand, each real curve has to be sufficiently smooth to satisfy the condition $\mathbf{A}_2$.

To estimate the sample paths and in order to take these constraints into account, Besse *et al.* (1997) considered the following optimization problem, which may be interpreted as a Tikhonov regularization of the least squares estimates $\widehat{y}_i$:

$$\min_{\widetilde{y}_i \in H_q^r} \left\{ \frac{1}{n}\sum_{i=1}^n \left(\|\widehat{y}_i - \widetilde{y}_i\|_{L^2}^2 + \rho \left\| \widetilde{y}_i^{(m)} \right\|_{L^2}^2 \right) \right\}; \tag{10}$$

where ρ is a smoothing parameter and H_q^r is a q-dimensional subspace of $\mathcal{S}_{k\nu}$.

Since $\hat{\mathbf{s}}_i'\mathbf{C}_k\hat{\mathbf{s}}_j = \int \widehat{y}_i(t)\widehat{y}_j(t)dt$ and $\hat{\mathbf{s}}_i'\mathbf{G}_k\hat{\mathbf{s}}_j = \int \widehat{y}_i^{(m)}(t)\widehat{y}_j^{(m)}(t)dt$, the optimization problem (10) is equivalent to:

$$\min_{u_i, A_q} \left\{ \frac{1}{n}\sum_{i=1}^n \left(\|\hat{\mathbf{s}}_i - \mathbf{u}_i\|_{\mathbf{C}_k}^2 + \rho \|\mathbf{u}_i\|_{\mathbf{G}_k}^2 \right) \;;\mathbf{u} \in A_q \;,\; \dim\, A_q = q \right\} \tag{11}$$

where A_q is a q-dimensional subspace of $\mathbb{R}^r$.

Denote by $\hat{\mathbf{s}}_{k,n} = \frac{1}{n}\sum_{i=1}^n \hat{\mathbf{s}}_i$ the coordinates of the empirical mean in the B-splines basis and then

$$\mathbf{S}_n = \frac{1}{n}\sum_{i=1}^n (\hat{\mathbf{s}}_i - \hat{\mathbf{s}}_{k,n})(\hat{\mathbf{s}}_i - \hat{\mathbf{s}}_{k,n})'$$

is the empirical covariance matrix. Finally, define

$$\mathbf{H}_{k,\rho} \;=\; (\mathbf{C}_k + \rho\mathbf{G}_k)^{-1} \tag{12}$$

which acts as a "Hat matrix".

The solution [2] of problem (10) is given by:

$$\begin{aligned} \widehat{\mathbf{u}}_{i,\rho} &= \widehat{\mathbf{P}}_{q,\rho}\mathbf{H}_{k,\rho}\mathbf{C}_k(\hat{\mathbf{s}}_i - \hat{\mathbf{s}}_{k,n}) + \mathbf{H}_{k,\rho}\mathbf{C}_k\hat{\mathbf{s}}_{k,n}\;, \quad i = 1\ldots,n, \\ \widehat{y}_{i,\rho}(t) &= \widehat{\mathbf{u}}_{i,\rho}'\mathbf{B}_k(t), \quad t \in [0,1] \end{aligned}$$

where $\widehat{\mathbf{P}}_{q,\rho} = \sum_{j=1}^q \widehat{\mathbf{v}}_{j,\rho}\widehat{\mathbf{v}}_{j,\rho}'\mathbf{H}_{k,\rho}^{-1}$ is the $\mathbf{H}_{k,\rho}^{-1}$-orthogonal projection onto the subspace generated by the first q eigenvectors $(\widehat{\mathbf{v}}_{1,\rho},\ldots,\widehat{\mathbf{v}}_{q,\rho})$, normalized with respect to the metric $\mathbf{H}_{k,\rho}^{-1}$, of the matrix:

$$\mathbf{H}_{k,\rho}\mathbf{C}_k\mathbf{S}_n\mathbf{C}_k. \tag{13}$$

Each estimated sample path can also be written as follows:

$$\widehat{y}_{i,\rho}(t) = \widehat{\mu}_\rho(t) + \widehat{z}_{i,\rho}(t), \quad t \in [0,1], \tag{14}$$

[2] `S+` programs for carrying out the estimation are available by anonymous `FTP` from `ftp.cict.fr` in the directory `pub/lstatprob/ferraty`.

where $\widehat{\mu}_\rho = (\mathbf{H}_{k,\rho}\mathbf{C}_k\hat{\mathbf{s}}_{k,n})'\,\mathbf{B}_k$ is the smoothed estimate of the mean function and

$$\widehat{z}_{i,\rho} = \left(\widehat{\mathbf{P}}_{q,\rho}\mathbf{H}_{k,\rho}\mathbf{C}_k(\hat{\mathbf{s}}_i - \hat{\mathbf{s}}_{k,n})\right)'\mathbf{B}_k$$

is the smooth estimate of the rank constrained individual random effect. The estimates of the ϕ_j's defined in (7) are written as follows in the B-splines basis:

$$\widehat{\phi}^{\rho}_{j,n}(t) = \widehat{\mathbf{v}}'_{j,\rho}\mathbf{B}_k(t), \quad t \in [0,1], \quad j = 1,\ldots,q.$$

Let's notice that these estimates of the eigenfunctions are not orthonormal with respect to the usual inner product in $L^2[0,1]$ (excepted if $\rho = 0$) since their coordinates are normalized with respect to $\mathbf{H}^{-1}_{k,\rho}$.

Equivalently, the solution is given by:

$$\hat{\mathbf{u}}_{i,\rho} = \mathbf{H}^{1/2}_{k,\rho}\widehat{\mathbf{Q}}_{q,\rho}\mathbf{H}^{1/2}_{k,\rho}\mathbf{C}_k(\hat{\mathbf{s}}_i - \hat{\mathbf{s}}_{k,n}) + \mathbf{H}_{k,\rho}\mathbf{C}_k\hat{\mathbf{s}}_{k,n}\ ,\ i = 1\ldots,n.$$

where $\widehat{\mathbf{Q}}_{q,\rho} = \sum_{j=1}^q \mathbf{v}_{j,\rho}\mathbf{v}'_{j,\rho}$ is the orthogonal projection onto the subspace generated by the first q eigenvectors of the symmetric matrix:

$$\mathbf{H}^{1/2}_{k,\rho}\mathbf{C}_k\mathbf{S}_n\mathbf{C}_k\mathbf{H}^{1/2}_{k,\rho}. \tag{15}$$

Remark : with a different goal, the estimation of smooth eigenfunctions of the covariance operator, the method proposed by Silverman (1996) leads to the same estimates of the eigenelements. The main difference is that Silverman (1996) assumed the sample paths to be continuoulsy observed and not contaminated with noise. Silverman's estimates, defined as the solution of the following problem:

$$\widehat{\mathbf{v}}'_{j,\rho}\mathbf{C}_k\mathbf{S}_n\mathbf{C}_k\widehat{\mathbf{v}}_{j,\rho} = \widehat{\lambda}_{j,\rho} \quad \text{and} \quad \widehat{\mathbf{v}}'_{j,\rho}\mathbf{H}^{-1}_{k,\rho}\widehat{\mathbf{v}}_{i,\rho} = \delta_{ij}$$

are also the eigenvectors $\widehat{\mathbf{v}}_{j,\rho}$, normalized with respect to the metric $\mathbf{H}^{-1}_{k,\rho}$, of the matrix defined in (13).

3 Convergence in mean integrated square error

This section is divided into three parts. Firstly, we give rates of convergence for the mean function. Then we bound the MISE of the eigenvectors of the covariance operator *without smoothing*, that is to say by assuming that the smoothing parameter ρ is fixed and equals zero. Finally, the use of perturbation theory (Kato, 1976), for small values of ρ allows us to derive the MISE for the smooth eigenvectors.

Asymptotic properties rely upon approximation properties of B-splines functions and on the existence of fourth order moments of random variables. We also assume that the growth of p and k can be controlled when n tends to infinity.

3.1 Smoothed estimates of the mean function

The expression for the smooth estimate $\widehat{\mu}_\rho$, as defined in (14), is given by:

$$\widehat{\mu}_\rho(t) = \widehat{\mathbf{s}}'_\rho \mathbf{B}_k(t) \quad \text{where } \widehat{\mathbf{s}}_\rho = \mathbf{H}_{k,\rho}\mathbf{C}_k \widehat{\mathbf{s}}_{k,n} = \mathbf{H}_{k,\rho}\mathbf{C}_k \widetilde{\mathbf{C}}_k^{-1} \widehat{\mathbf{b}}_{k,n},$$

and let's denote its expectation by $\widetilde{\mu}_\rho = \mathbb{E}\,\widehat{\mu}_\rho$. The mean square error $\mathbb{E}\,||\mu - \widehat{\mu}_\rho||^2_{L^2}$ is written, as usual, as square bias plus variance:

$$\mathbb{E}\,||\mu - \widehat{\mu}_\rho||^2_{L^2} = ||\mu - \widetilde{\mu}_\rho||^2_{L^2} + \mathbb{E}\,||\widetilde{\mu}_\rho - \widehat{\mu}_\rho||^2_{L^2}\,.$$

Theorem 3.1 *Under assumptions $\mathbf{A}_1$, and $\mathbf{A}_2$, if ρ satisfies $\rho = o(k^{-2m})$ and $k = o(p)$ then we can bound bias and variance as follows:*

$$||\mu - \widetilde{\mu}_\rho||^2_{L^2} = \mathcal{O}(\rho^2\, k^{4m}) + \mathcal{O}(k^{-2m}),$$

$$\mathbb{E}\,||\widetilde{\mu}_\rho - \widehat{\mu}_\rho||^2_{L^2} = \mathcal{O}(n^{-1}).$$

Consequently, we have:

$$\mathbb{E}\,||\widehat{\mu}_\rho - \mu||^2_{L^2} = \mathcal{O}(n^{-1}) + \mathcal{O}(k^{-2m}) + \mathcal{O}(\rho^2 k^{4m}).$$

This theorem implies that if k, p and ρ are chosen as follows:

$$k = \mathcal{O}\left(n^{\frac{1}{2m}}\right), \quad \rho = \mathcal{O}(n^{-3/2}), \quad \frac{k}{p} = o(1),$$

then the MISE is bounded above by:

$$\mathbb{E}\,||\widehat{\mu}_\rho - \mu||^2_{L^2} = \mathcal{O}(n^{-1}).$$

Furthermore, if we suppose that $Z = 0$ almost surely, this model is just the usual nonparametric model. It is then easy to see that the variance is of order $\mathcal{O}(k/(np))$ (see Cardot & Diack, 1998) and we obtain the optimal rate of convergence for nonparametric estimates. One can also get intermediate rates of convergence by assuming, for example, that Z satisfies mixing conditions (Burman, 1991).

Remark: the parameters k and ρ both act as smoothing tools and consequently have to be chosen simultaneously. We think that the cross validation method proposed by Rice & Silverman (1991) and Hart & Werhly (1993) obtained by leaving one entire curve out, could be usefully adapted to this problem. However, an asymptotic study should be performed to justify such a pratice.

3.2 Results on the covariance operator without smoothing

Henceforth, we suppose for simplicity that the mean has been subtracted off, and thus the sample paths are centered.

Consider the empirical covariance operator, say $\widehat{\Gamma}_{np}$, of the estimates $\widehat{y}_i(t)$ defined in equation (9). Split each random vector, $\mathbf{Y}_i$, into a random signal $\mathbf{Z}_i$ plus noise $\boldsymbol{\epsilon}_i$: $\mathbf{Y}_i = \mathbf{Z}_i + \boldsymbol{\epsilon}_i \quad i = 1, \ldots, n$. Denote by $\widetilde{z}_i$, the B-splines estimate of the signal deduced from $\mathbf{Z}_i$ and by $\widetilde{\epsilon}_i$ the B-splines estimate of the noise obtained from $\boldsymbol{\epsilon}_i$. Then, each nonparametric estimate of the sample paths can be written:

$$\widehat{y}_i = \widetilde{z}_i + \widetilde{\epsilon}_i, \quad i = 1, \ldots, n,$$

and the empirical covariance operator $\widehat{\Gamma}_{np}$ may be expressed as follows:

$$\begin{aligned}
\widehat{\Gamma}_{np} &= \frac{1}{n}\sum_{i=1}^{n} \widehat{y}_i \otimes \widehat{y}_i \\
&= \frac{1}{n}\sum_{i=1}^{n} (\widetilde{z}_i + \widetilde{\epsilon}_i) \otimes (\widetilde{z}_i + \widetilde{\epsilon}_i) \\
&= \widetilde{\Gamma}_{np} + \widetilde{\Gamma}_{1,np} + \widetilde{\Gamma}_{2,np} + \widetilde{\Gamma}_{\epsilon,np},
\end{aligned}$$

where

$$\widetilde{\Gamma}_{np} = \frac{1}{n}\sum_{i=1}^{n} \widetilde{z}_i \otimes \widetilde{z}_i\ , \qquad \widetilde{\Gamma}_{1,np} = \frac{1}{n}\sum_{i=1}^{n} \widetilde{z}_i \otimes \widetilde{\epsilon}_i,$$
$$\widetilde{\Gamma}_{2,np} = \frac{1}{n}\sum_{i=1}^{n} \widetilde{\epsilon}_i \otimes \widetilde{z}_i\ , \qquad \widetilde{\Gamma}_{\epsilon,np} = \frac{1}{n}\sum_{i=1}^{n} \widetilde{\epsilon}_i \otimes \widetilde{\epsilon}_i\ .$$

The study of the asymptotic convergence is organized as follows:

- the sampling effect is studied in lemmas 3.2, 3.3, 3.4 and theorem 3.5,
- the discretization and B-splines approximation effect is measured in lemma 3.6.

The first step can be viewed as a study of the asymptotic variance whereas the second one is a study of the asymptotic behaviour of the bias.

Some of the following results are based on the Hilbert-Schmidt properties of covariance operators of second order Hilbert valued random variables. A Hilbert-Schmidt operator T defined on $L^2[0,1]$ is a bounded linear operator satisfying:

$$\sum_{i \in \mathbb{N}^*} \langle Te_i, Te_i \rangle < +\infty \tag{16}$$

for any orthonormal basis $(e_i)_{i\in\mathbb{N}}$ of $L^2[0,1]$. This relation defines a norm, say $\|T\|_{\mathcal{H}}$, which satisfies $\|T\|_{\mathcal{H}} \geq \|T\|$.

The covariance operator Γ of a second order random variable Z which takes values in $L^2[0,1]$ is a Hilbert-Schmidt operator. Futhermore, there exists a function $\Gamma(s,t)$ such that:

$$\Gamma(f)\,(t) \;=\; \int_O^1 \Gamma(s,t)f(s)\,ds, \quad \forall f \in L^2[0,1]. \tag{17}$$

Actually, $\Gamma(s,t)$ is the covariance of the underlying continuous time process:

$$\Gamma(s,t) = \mathbb{E}\left(Z(s), Z(t)\right).$$

The Hilbert-Schmidt norm of Γ can therefore be expressed equivalently as follows:

$$\|\Gamma\|_{\mathcal{H}}^2 \;=\; \int_0^1 \int_0^1 \left(\Gamma(s,t)\right)^2 \, ds\, dt. \tag{18}$$

The following lemma allows us to measure the norm between a covariance operator of a second order random variable and its empirical estimate. Its demonstration is similar to the real case and consequently omitted.

Lemma 3.2 *If $(X_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d random variables which take values in $L^2[0,1]$ and satisfy $\mathbb{E}\|X\|_{L^2}^4 < \infty$ then the following holds:*

$$\mathbb{E}\|\Gamma_n - \Gamma\|_{\mathcal{H}}^2 \;=\; \frac{1}{n}\mathbb{E}\|X\|_{L^2}^4 - \frac{1}{n}\|\Gamma\|_{\mathcal{H}}^2 \tag{19}$$

where $\Gamma_n = \frac{1}{n}\sum_{i=1}^n X_i \otimes X_i$ and $\Gamma = \mathbb{E}(X \otimes X)$.
Furthermore, if the sequence of i.i.d second order random variables $(Y_i)_{i\in\mathbb{N}}$ is independent of $(X_i)_{i\in\mathbb{N}}$, then the cross covariance operator and its empirical estimate satisfy:

$$\mathbb{E}\|\Delta_n - \Delta\|_{\mathcal{H}}^2 \;=\; \frac{\mathbb{E}\|Y\|_{L^2}^2 \;\, \mathbb{E}\|X\|_{L^2}^2}{n} - \frac{1}{n}\|\Delta\|_{\mathcal{H}}^2 \tag{20}$$

where $\Delta_n = \frac{1}{n}\sum_{i=1}^n Y_i \otimes X_i$ and $\Delta = \mathbb{E}(Y \otimes X)$.

Lemma 3.3 *If $\sup_{t\in[0,1]} \mathbb{E}[Z(t)^4] < \infty$, we have:*

$$\mathbb{E}\left\|\widetilde{\Gamma}_{np} - \widetilde{\Gamma}_p\right\|_{\mathcal{H}}^2 = \mathcal{O}\left(\frac{1}{n}\right), \tag{21}$$

where $\mathcal{O}()$ does not depend on p and k.

Remark : Using relation (7) and the regularity of the functions ϕ_j, it is easy to check that condition $\sup_t \mathbb{E}[Z(t)^4] < +\infty$ of lemma 3.3 is equivalent to the following assumption: $\mathbb{E}\|Z\|_{L^2}^4 < +\infty$.

Let's define $\widetilde{P}_{k,p} = \sum_{l,j=1}^r \widetilde{a}_{lj} e_{ij}$ where $\widetilde{a}_{lj} = \left[\widetilde{\mathbf{C}}_k^{-1}\right]_{l,j}$, $l,j = 1,\ldots,r$. It is the discrete approximation of the projection onto the space $\mathcal{S}_{k\nu}$ in the basis generated by the rank one operators

$$e_{ij} = B_{kj} \otimes B_{ki}, \; i,j = 1,\ldots,r. \tag{22}$$

Lemma 3.4 *We have*

$$\mathbb{E}(\widetilde{\Gamma}_{\epsilon,np}) = \frac{\sigma^2}{p}\widetilde{P}_{k,p}.$$

Moreover, if $\mathbb{E}(\epsilon^4) < \infty$ *then the following holds:*

$$\mathbb{E}\left\|\widetilde{\Gamma}_{\epsilon,np} - \frac{\sigma^2}{p}\widetilde{P}_{k,p}\right\|_{\mathcal{H}}^2 = \mathcal{O}\left(\frac{k^2}{n\,p^2}\right) .$$

Theorem 3.5 *If* $\sup_t \mathbb{E}[Z(t)^4] < \infty$ *,* $\mathbb{E}(\epsilon^4) < \infty$ *and* $k \leq p$*, then we have*

$$\mathbb{E}\left\|\widehat{\Gamma}_{np} - (\widetilde{\Gamma}_p + \frac{\sigma^2}{p}\widetilde{P}_{k,p})\right\|_{\mathcal{H}}^2 = \mathcal{O}\left(\frac{1}{n}\right) .$$

The following lemma measures the loss induced by the discretisation of the curves and shows that the covariance operator of the noise tends to zero.

Lemma 3.6 *Under assumption* $\mathbf{A}_1$ *and if* $k = o(p)$ *when* k *and* p *tend to infinity, we have*

$$\begin{aligned} \left\|\Gamma - \widetilde{\Gamma}_p\right\|_{\mathcal{H}} &= \mathcal{O}\left(k^{-m}\right), \\ \left\|\frac{\sigma^2}{p}\widetilde{P}_{k,p}\right\| &= \mathcal{O}\left(\frac{1}{p}\right). \end{aligned}$$

Hence

$$\left\|\Gamma - \left(\widetilde{\Gamma}_p + \frac{\sigma^2}{p}\widetilde{P}_{k,p}\right)\right\| = \mathcal{O}\left(k^{-m}\right) + \mathcal{O}\left(\frac{1}{p}\right).$$

Let's denote by $\{\widehat{\phi}_{1n}, \ldots, \widehat{\phi}_{qn}\}$ the eigenvectors of $\widehat{\Gamma}_{np}$ associated to the decreasing sequence of eigenvalues. Since these functions are determined up to sign change, we will suppose that $< \widehat{\phi}_{jn}, \phi_j > \geq 0$. We can now state the main theorem of this section:

Theorem 3.7 *Under assumptions* $\mathbf{A}_1$*,* $\mathbf{A}_2$*, if* $k = o(p)$ *when* n *and* k *tend to infinity, then we have*

$$\left(\mathbb{E}\left\|\widehat{\Gamma}_{np} - \Gamma\right\|^2\right)^{1/2} = \mathcal{O}(k^{-m}) + \mathcal{O}(n^{-1/2}) + \mathcal{O}(p^{-1}). \tag{23}$$

Moreover, if the eigenvalues of Γ *are distinct,* $\lambda_1 > \ldots > \lambda_q > 0$*, then the following holds:*

$$\left(\mathbb{E}\left\|\widehat{\phi}_{jn} - \phi_j\right\|_{L^2}^2\right)^{1/2} = \mathcal{O}(k^{-m}) + \mathcal{O}(n^{-1/2}) + \mathcal{O}(p^{-1}), \quad j = 1, \ldots, q. \tag{24}$$

Remarks:

1. The rate of convergence depends explicitly on the number of design points p. If we choose $p \approx \sqrt{n}$, $k = \mathcal{O}(n^{1/2m})$, the condition $k = o(p)$ is satisfied and

$$\left(\mathbb{E} \left\| \widehat{\phi}_{jn} - \phi_j \right\|_{L^2}^2 \right)^{1/2} = \mathcal{O}(n^{-1/2}).$$

2. We have supposed, for simplicity, that the eigenvalues of Γ were distinct. This condition could be relaxed by considering the subpaces generated by the eigenvectors of Γ. Then one should work with projections instead of eigenvectors such as in Dauxois *et al.* (1982).

3. The assumption $q < \infty$ is crucial in lemma 3.6 to bound $\left\| \Gamma - \widetilde{\Gamma}_p \right\|$. If it wasn't satisfied, it would still be possible to obtain some results of convergence. We could show for instance the convergence of $\widetilde{\Gamma}$ towards Γ by using compactness properties of covariance operators (Besse 1991) but in counterpart, one would loose rates of convergence. On the other hand, we could add a condition on the decrease of the eigenvalues such as $\lambda_j = ar^j, a > 0, r \in]0,1[$, and control the growth of q with n, p and k.

3.3 Smooth estimates of the eigenelements

In order to prove the consistency of the smooth estimates, we will study the asymptotic behaviour of the eigenvectors of the operator $\widehat{\Gamma}_{n,p,\rho}$, whose matrix representation in the B-splines basis is defined by (13). The demonstration of this result is based on an asymptotic expansion, for small ρ, of $\widehat{\Gamma}_{n,p,\rho}$.

Let $\{\widehat{\phi}^{\rho}_{j,n},\ j = 1, \ldots q\}$ be the first q eigenvectors of $\widehat{\Gamma}_{n,p,\rho}$. Since they are determined only up to a sign change, we suppose that $(< \widehat{\phi}^{\rho}_{j,n}, \phi_j >) \geq 0$.

We can now state the main theorem of this section:

Theorem 3.8 *Under assumptions of theorem 3.7 and if $\rho = o(k^{-2m})$, the following holds:*

$$\left(\mathbb{E} \left\| \widehat{\phi}^{\rho}_{j,n} - \phi_j \right\|_{L^2}^2 \right)^{1/2} = \mathcal{O}(n^{-1/2}) + \mathcal{O}(k^{-m}) + \mathcal{O}(\rho\, k^{2m}) + \mathcal{O}(p^{-1}), \quad j = 1, \ldots, q.$$

The mean square error of smooth estimates of the eigenfunctions tends to zero at the same rate if ρ is chosen as follows:

$$\rho = \mathcal{O}(n^{-3/2}).$$

4 Heuristic investigation of the effect of smoothing

In this section, we carry out some heuristic calculations based on perturbation theory (Kato, 1976), for large n and small ρ to determine whether, and when, smoothing has an advantageous effect on the estimation of any particular eigenfunction. In order to simplify calculations, parameters k and p were supposed to be fixed. Similar calculus have already been performed by various authors such as Pezzulli & Silverman (1993) and Silverman (1996). We investigate here, within the subspace of splines functions $\mathcal{S}_{k,\nu}$, the effect of smoothing on the mean square error of estimation in order to establish a link between the optimal smoothing parameter and the number of observed curves. This asymptotic expansion is justified by the convergence of our estimates demonstrated in theorem 3.8. Nevertheless, we have to make the additional assumption on the distribution of the random vectors $\mathbf{Z}_i$ and $\boldsymbol{\epsilon}_i$:

($\mathbf{A}_3$) for all i, $\mathbf{Z}_i$ and $\boldsymbol{\epsilon}_i$ are independent gaussian vectors.

Since the eigenelements of the matrix $\mathbf{M}_{n,\rho}$ defined in (13) are equal to those obtained by the Silverman (1996) procedure (remark in section 2.3), we only recall the method used to get an asymptotic expansion of the eigenelements. Then, we are able to approximate the quadratic error of estimation, within $\mathcal{S}_{k,\nu}$, of a new and independent curve.

4.1 Asymptotic expansions

Using the matrix representation of the covariance operator

$$\begin{aligned}\mathbf{M}_{n,\rho} &= \mathbf{H}_{k,\rho}\mathbf{C}\mathbf{S}_n\mathbf{C} \\ &= \left(\mathbf{I}+\rho\mathbf{C}^{-1}\mathbf{G}\right)^{-1}\mathbf{S}_n\mathbf{C}\end{aligned}$$

and recall its eigenvectors $\widehat{\mathbf{v}}_{1,\rho},\ldots,\widehat{\mathbf{v}}_{q,\rho}$ are orthonormal with respect to metric $\mathbf{H}_{k,\rho}^{-1}$. Define $\mathbf{M}_n = \mathbf{S}_n\mathbf{C}$ and $\mathbf{M} = \mathbb{E}(\mathbf{M}_n)$. The eigenelements $\left(\widehat{\lambda}_{\ell,\rho},\widehat{\mathbf{v}}_{\ell,\rho}\right)$ of $\mathbf{M}_{n,\rho}$ satisfy:

$$\mathbf{M}_n\widehat{\mathbf{v}}_{\ell,\rho} = \widehat{\lambda}_{\ell,\rho}\left(\mathbf{I}+\rho\mathbf{C}^{-1}\mathbf{G}\right)\widehat{\mathbf{v}}_{\ell,\rho}.$$

We can deduce from the central limit theorem that the covariance structure of $\sqrt{n}(\mathbf{M}_n-\mathbf{M})$ does not vary with n (Silverman, 1996) and p and k since they are supposed to be fixed. Then, one can express $\mathbf{M}_n = \mathbf{M}+n^{-1/2}\mathbf{R}$, where $\mathbf{R}$ is a random matrix with mean zero.

Assuming n is large enough, ρ is small enough and using bounds obtained in theorem 3.8, one can write the asymptotic expansion of the eigenelements of $\mathbf{M}_{n,\rho}$ as:

$$\widehat{\lambda}_\rho = \lambda + n^{-1/2}\mu_1 + \rho\,\mu_2 + \rho n^{-1/2}\mu_{12} + n^{-1}\mu_{11} + \rho^2\mu_{22} + \cdots \qquad (25)$$

$$\widehat{\mathbf{v}}_\rho = \mathbf{v} + n^{-1/2}\mathbf{u}_1 + \rho\,\mathbf{u}_2 + \rho n^{-1/2}\mathbf{u}_{12} + n^{-1}\mathbf{u}_{11} + \rho^2\mathbf{u}_{22} + \cdots \qquad (26)$$

with

$$\left\{ \begin{array}{lll} \mathbf{M}_n \widehat{\mathbf{v}}_\rho = \widehat{\lambda}_\rho \left(\mathbf{I} + \rho \mathbf{C}^{-1}\mathbf{G}\right) \widehat{\mathbf{v}}_\rho & \text{and} & \widehat{\mathbf{v}}_\rho' \mathbf{H}_{k,\rho}^{-1} \widehat{\mathbf{v}}_\rho = 1 \\ \mathbf{M}\mathbf{v} = \lambda \mathbf{v} & \text{and} & \mathbf{v}'\mathbf{C}\mathbf{v} = 1 \end{array} \right. \tag{27}$$

Subscript ℓ is omitted for sake of simplicity. By identifying the orders in $n^{-1/2}, \rho, n^{-1/2}\rho, \ldots$ and defining $\mathbf{A} = \mathbf{C}^{-1}\mathbf{G}$, one obtain the two following sets of equations:

$$\left\{ \begin{array}{lll} \mathbf{u}_1'\mathbf{C}\mathbf{v} & = & 0 \\ 2\mathbf{u}_2'\mathbf{C}\mathbf{v} + \mathbf{v}'\mathbf{G}\mathbf{v} & = & 0 \\ \mathbf{u}_1'\mathbf{G}\mathbf{v} + \mathbf{u}_{12}'\mathbf{C}\mathbf{v} + \mathbf{u}_1'\mathbf{C}\mathbf{u}_2 & = & 0 \\ 2\mathbf{u}_{11}'\mathbf{C}\mathbf{v} + \mathbf{u}_1'\mathbf{C}\mathbf{u}_1 & = & 0 \\ 2\mathbf{u}_{22}'\mathbf{C}\mathbf{v} + 2\mathbf{u}_2'\mathbf{G}\mathbf{v} + \mathbf{u}_2'\mathbf{C}\mathbf{u}_2 & = & 0 \end{array} \right. \tag{28}$$

and

$$\left\{ \begin{array}{lll} \mathbf{R}\mathbf{v} + \mathbf{M}\mathbf{u}_1 & = & \mu_1 \mathbf{v} + \lambda \mathbf{u}_1 \\ \mathbf{M}\mathbf{u}_2 & = & \mu_2 \mathbf{v} + \lambda \mathbf{u}_2 + \lambda \mathbf{A}\mathbf{v} \\ \mathbf{R}\mathbf{u}_2 + \mathbf{M}\mathbf{u}_{12} & = & \mu_2 \mathbf{u}_1 + \mu_1 \mathbf{u}_2 + \mu_{12}\mathbf{v} + \lambda \mathbf{u}_{12} + \mu_{12}\mathbf{v} + \lambda\ \mathbf{A}\mathbf{u}_1 + \mu_1 \mathbf{A}\mathbf{v} \\ \mathbf{M}\mathbf{u}_{11} + \mathbf{R}_1\mathbf{u}_1 & = & \mu_{11}\mathbf{v} + \lambda \mathbf{u}_{11} + \mu_1 \mathbf{u}_1 \\ \mathbf{M}\mathbf{u}_{22} & = & \mu_{22}\mathbf{v} + \lambda \mathbf{u}_{22} + \mu_2 \mathbf{u}_2 + \mu_2 \mathbf{A}\mathbf{v} + \lambda \mathbf{A}\mathbf{u}_2 \end{array} \right. \tag{29}$$

Define $\mathbf{P}_j = \mathbf{v}_j\mathbf{v}_j'\mathbf{C}$, the $\mathbf{C}$-orthogonal projection onto the subspace generated by the vector $\mathbf{v}_j$ and

$$\mathbf{T}_j = \sum_{\nu \neq j} \frac{\mathbf{P}_\nu}{\lambda_j - \lambda_\nu}.$$

These operators commute since they have the same eigenvectors and are symmetric with respect to metric $\mathbf{C}$.

We then get from above:

$$\left\{ \begin{array}{lll} \mu_1 & = & \mathrm{tr}(\mathbf{P}_\ell \mathbf{R}) = \langle \mathbf{v}, \mathbf{R}\mathbf{v} \rangle_{\mathbf{C}} \\ \mu_2 & = & -\lambda\ \mathrm{tr}(\mathbf{P}_\ell \mathbf{A}) = -\lambda\ d_\ell^2 \\ \mathbf{u}_1 & = & \mathbf{T}_\ell \mathbf{R}\mathbf{v}_\ell \\ \mathbf{u}_2 & = & -\lambda\ \mathbf{T}_\ell \mathbf{A}\mathbf{v}_\ell - \frac{1}{2} d_\ell^2 \mathbf{v}_\ell \\ \mathbf{u}_{11} & = & \mathbf{T}_\ell \mathbf{R}\mathbf{u}_1 - \mu_1 \mathbf{T}_\ell \mathbf{u}_1 - \frac{1}{2}\mathbf{u}_1'\mathbf{C}\mathbf{u}_1 \mathbf{v}_\ell \\ \mathbf{u}_{12} & = & \mathbf{T}_\ell \mathbf{R}\mathbf{u}_2 - \lambda\ \mathbf{T}_\ell \mathbf{A}\mathbf{u}_1 - \mu_1 \mathbf{T}_\ell \mathbf{u}_2 - \mu_2 \mathbf{T}_\ell \mathbf{A}\mathbf{v} - \mu_2 \mathbf{T}_\ell \mathbf{u}_1 - (\mathbf{u}_1'\mathbf{C}\mathbf{u}_2 + \mathbf{u}_1'\mathbf{G}\mathbf{v}_\ell)\,\mathbf{v}_\ell \\ \mathbf{u}_{22} & = & -\lambda \mathbf{T}_\ell \mathbf{A}\mathbf{u}_2 - \mu_2 \mathbf{T}_\ell \mathbf{A}\mathbf{v}_\ell - \mu_2 \mathbf{T}_\ell \mathbf{u}_2 - \left(\frac{1}{2}\mathbf{u}_2'\mathbf{C}\mathbf{u}_2 + \mathbf{u}_2'\mathbf{G}\mathbf{v}_\ell\right) \mathbf{v}_\ell \end{array} \right. \tag{30}$$

where $d_l^2 = \mathbf{v}_\ell'\mathbf{G}\mathbf{v}_\ell$ is the semi-norm of the vector $\mathbf{v}_\ell$ which measures the smoothness of the corresponding function belonging to $\mathcal{S}_{k,\nu}$.

4.2 Is smoothing advantageous ?

In this section we will expand the MISE, in terms of n and ρ, for the estimates of:

- eigenvectors,
- a new and independent sampled curve.

4.2.1 Approximation of the eigenfunctions

Consider the following risk function:

$$\text{MISE}(\rho) = \mathbb{E}\left(\|\widehat{\mathbf{v}}_{\ell,\rho} - \mathbf{v}_\ell\|^2_{\mathbf{C}}\right)$$

as a criterion for measuring the accuracy of the estimated eigenvectors. It is sufficient to compare MISE(ρ) with MISE(0) to determine when smoothing has an advantageous effect. Neglecting terms of order $n^{-3/2}, \rho^2/n, \rho^3 n^{-1/2}, \ldots$, one obtains the following approximation:

$$\begin{aligned}\mathbb{E}\left(\|\widehat{\mathbf{v}}_{\ell,\rho} - \mathbf{v}_\ell\|^2_{\mathbf{C}}\right) \;\approx\; & \mathbb{E}\left(n^{-1}\|\mathbf{u}_1\|^2_{\mathbf{C}} + 2\rho n^{-1/2}\mathbf{u}_1'\mathbf{C}\mathbf{u}_2 + 2\rho n^{-1}\left\{\mathbf{u}_{11}'\mathbf{C}\mathbf{u}_2 + \mathbf{u}_{12}'\mathbf{C}\mathbf{u}_1\right\}\right. \\ & \left. + \rho^2\|\mathbf{u}_2\|^2_{\mathbf{C}}\right).\end{aligned} \tag{31}$$

Using $\mathbb{E}\,\mathbf{R} = 0$, $\mathbf{T}_\ell\mathbf{v}_\ell = 0$ and the assumption of normality $\mathbf{A}_3$ we obtain

$$\mathbb{E}\,(\mathbf{RAR}) = \mathbf{MAM} + \text{tr}(\mathbf{AM})\mathbf{M}$$

for any symmetric matrix $\mathbf{A}$ (see Pezzulli & Silverman 1993). Hence, we can obtain for instance:

$$\begin{aligned}\mathbb{E}\,\|\mathbf{u}_1\|^2_{\mathbf{C}} &= \mathbb{E}\,\langle \mathbf{T}_\ell\mathbf{R}\mathbf{v}, \mathbf{T}_\ell\mathbf{R}\mathbf{v}\rangle_{\mathbf{C}} \\ &= \mathbb{E}\,\left\langle \mathbf{v}, \mathbf{R}\mathbf{T}_\ell^2\mathbf{R}\mathbf{v}\right\rangle_{\mathbf{C}} \\ &= \left\langle \mathbf{v}, \mathbf{M}\mathbf{T}_\ell^2\mathbf{M}\mathbf{v}\right\rangle_{\mathbf{C}} + \text{tr}\left(\mathbf{T}_\ell^2\mathbf{M}\right)\langle \mathbf{v}, \mathbf{M}\mathbf{v}\rangle_{\mathbf{C}} \\ &= \lambda_\ell \sum_{\nu\neq\ell}\frac{\lambda_\nu}{(\lambda_\ell-\lambda_\nu)^2}.\end{aligned}$$

Pursuing these types of manipulations, we finally obtain:

$$\begin{cases}\mathbb{E}\,\langle \mathbf{u}_1, \mathbf{u}_2\rangle_{\mathbf{C}} = 0 \\ \mathbb{E}\,\langle \mathbf{u}_2, \mathbf{u}_{11}\rangle_{\mathbf{C}} = \frac{1}{4}\lambda_\ell d_\ell^2 \sum_{v\neq\ell}\frac{\lambda_\nu}{(\lambda_\ell-\lambda_\nu)^2} \\ \mathbb{E}\,\langle \mathbf{u}_1, \mathbf{u}_{12}\rangle_{\mathbf{C}} = -\frac{1}{2}\lambda_\ell d_\ell^2 \sum_{v\neq\ell}\frac{\lambda_\nu}{(\lambda_\ell-\lambda_\nu)^2} + \lambda_\ell^2\sum_{v\neq\ell}\frac{\lambda_\nu(d_\ell^2-d_\nu^2)}{(\lambda_\ell-\lambda_\nu)^3} \\ \mathbb{E}\,\|\mathbf{u}_2\|^2_{\mathbf{C}} = \frac{1}{4}d_\ell^4 + \lambda_\ell^2\sum_{\nu\neq\ell}\frac{d_{\nu\ell}^2}{(\lambda_\ell-\lambda_\nu)^2}\end{cases} \tag{32}$$

where $d_{\ell j} = \mathbf{v}_\ell'\mathbf{G}\mathbf{v}_j$. The asymptotic approximation of the MISE

$$\text{MISE}(\rho) - \text{MISE}(0) = 2\rho n^{-1}\mathbb{E}\left\{\langle \mathbf{u}_{11}, \mathbf{u}_2\rangle_{\mathbf{C}} + \langle \mathbf{u}_{12}, \mathbf{u}_1\rangle_{\mathbf{C}}\right\} + \rho^2\mathbb{E}\,\|\mathbf{u}_2\|^2_{\mathbf{C}}$$

can be viewed as a quadratic polynomial in ρ. Hence, it is helpful to use a smoothing parameter, for small ρ when

$$\mathbb{E}\left\{\langle \mathbf{u}_{11}, \mathbf{u}_2\rangle_{\mathbf{C}} + \langle \mathbf{u}_{12}, \mathbf{u}_1\rangle_{\mathbf{C}}\right\} \;\leq\; 0. \tag{33}$$

At the optimum value

$$\rho^* = \frac{-1}{n}\frac{\mathbb{E}\left\{\langle \mathbf{u}_{11}, \mathbf{u}_2\rangle_{\mathbf{C}} + \langle \mathbf{u}_{12}, \mathbf{u}_1\rangle_{\mathbf{C}}\right\}}{\mathbb{E}\,\|\mathbf{u}_2\|^2_{\mathbf{C}}}$$

then

$$\text{MISE}(\rho^*) - \text{MISE}(0) = \frac{-1}{n^2} \frac{\left(\mathbb{E}\left\{\langle \mathbf{u}_{11}, \mathbf{u}_2 \rangle_{\mathbf{C}} + \langle \mathbf{u}_{12}, \mathbf{u}_1 \rangle_{\mathbf{C}}\right\}\right)^2}{\mathbb{E}\left\|\mathbf{u}_2\right\|^2_{\mathbf{C}}}.$$

A sufficient condition for smoothing to be advantageous, is that the ordered eigenvectors are more and more "rough" ($\ell < \nu \Rightarrow d_\ell \leq d_\nu$). For real data sets, this condition is often satisfied. It is to be noticed that the condition obtained here is less restrictive than the one of Pezzulli & Silverman (1993) (see Silverman 1996 for further explanations).

4.2.2 Approximation of a new curve

One can also perform the same type of expansion of the MISE for a new and independent centered sample path. That is to say, suppose we have constructed our estimates from a sample of n curves and we then observe a new trajectory, independent of the sample. By writing its coordinates, $\widehat{\mathbf{s}}$, in the B-splines basis and then splitting them into two parts, signal plus noise, $\widehat{\mathbf{s}} = \mathbf{s} + \boldsymbol{\epsilon}$, and separating its covariance operator into signal and noise components

$$\begin{aligned} \mathbb{E}\left(\widehat{\mathbf{s}}\widehat{\mathbf{s}}'\mathbf{C}\right) &= \boldsymbol{\Gamma} + \frac{\sigma^2}{p}\mathbf{C}^{-1}\widetilde{\mathbf{C}} \\ &= \sum_\nu \lambda_\nu^{(z)}\mathbf{P}_\nu + \frac{\sigma^2}{p}\mathbf{I} + o\left(\frac{1}{p}\right). \end{aligned} \tag{34}$$

We take into account that $\left\|\mathbf{C}^{-1}\widetilde{\mathbf{C}} - \mathbf{I}\right\| = \left\|\mathbf{C}^{-1}(\widetilde{\mathbf{C}} - \mathbf{C})\right\| = \mathcal{O}(k/p)$ is negligible (Lemma 6.2). Hence, the eigenvalues $\left(\lambda_\nu^{(z)}\right)$ of the covariance operator of the signal Z satisfy:

$$\lambda_\nu = \lambda_\nu^{(z)} + \frac{\sigma^2}{p} + o\left(\frac{1}{p}\right), \qquad \nu = 1, \ldots, q. \tag{35}$$

Writing

$$\begin{aligned} \widehat{\mathbf{s}}_{\rho,n} &= \widehat{\mathbf{v}}_{\ell,\rho}\widehat{\mathbf{v}}'_{\ell,\rho}\mathbf{H}^{-1}_{k,\rho}\mathbf{H}_{k,\rho}\mathbf{C}\widehat{\mathbf{s}} \\ &= \langle \widehat{\mathbf{v}}_{\ell,\rho}, \widehat{\mathbf{s}} \rangle_{\mathbf{C}}\, \widehat{\mathbf{v}}_{\ell,\rho} \end{aligned}$$

gives the approximation of the coordinates of this new curve in the space spanned by the ℓth eigenvector. $\widehat{\mathbf{v}}_{\ell,\rho}\widehat{\mathbf{v}}'_{\ell,\rho}\mathbf{H}^{-1}_{k,\rho}$ is the $\mathbf{H}^{-1}_{k,\rho}$-orthogonal projection onto the subspace generated by $\widehat{\mathbf{v}}_{\ell,\rho}$. The following loss function

$$\mathcal{L}(\rho) = \mathbb{E}\left\|\widehat{\mathbf{s}}_{\rho,n} - \mathbf{P}_\ell\mathbf{s}\right\|^2_{\mathbf{C}} \tag{36}$$

allows us to evaluate the approximation error for reconstructing the curve in the subspace generated by $\widehat{\mathbf{v}}_{\ell,\rho}$.

To examine the effect of smoothing, one must compare $\mathcal{L}(\rho)$ with $\mathcal{L}(0)$. For this, define $\widehat{\mathbf{v}}$ the estimate of $\mathbf{v}_\ell$ without smoothing and $\widehat{\mathbf{s}}_n$ the corresponding estimate of the coordinates:

$$\begin{aligned} \widehat{\mathbf{s}}_n &= \widehat{\mathbf{v}}\widehat{\mathbf{v}}'\mathbf{C}\widehat{\mathbf{s}} \\ &= \mathbf{P}_\ell\widehat{\mathbf{s}} + n^{-1/2}\left(\mathbf{T}_\ell\mathbf{R}\mathbf{P}_\ell + \mathbf{P}_\ell\mathbf{R}\mathbf{T}_\ell\right)\widehat{\mathbf{s}} + \cdots \end{aligned}$$

Writing $\widehat{\mathbf{s}}_{\rho,n} = \widehat{\mathbf{v}}_\ell \widehat{\mathbf{v}}_\ell' \mathbf{C}\widehat{\mathbf{s}}$ and expanding it as follows

$$\widehat{\mathbf{s}}_{\rho,n} = \widehat{\mathbf{s}}_n + \left\{\rho\left(\mathbf{u}_2\mathbf{v}' + \mathbf{v}\mathbf{u}_2'\right) + \rho . n^{-1/2}\left(\mathbf{u}_{12}\mathbf{v}' + \mathbf{v}\mathbf{u}_{12}'\right)\right\}\mathbf{C}\widehat{\mathbf{s}} \tag{37}$$

gives the following approximation

$$\begin{aligned}\mathcal{L}(\rho) \approx\ & \mathcal{L}(0) + \mathbb{E}\Big\{\rho^2 \left\|\left(\mathbf{u}_2\mathbf{v}' + \mathbf{v}\mathbf{u}_2'\right)\mathbf{C}\widehat{\mathbf{s}}\right\|_{\mathbf{C}}^2 \\ & + 2\rho\ \left\langle \widehat{\mathbf{s}}_n - \mathbf{P}_\ell \mathbf{s}\ ,\ \left(\mathbf{u}_2\mathbf{v}' + \mathbf{v}\mathbf{u}_2'\right)\mathbf{C}\widehat{\mathbf{s}} + n^{-1/2}\left(\mathbf{u}_{12}\mathbf{v}' + \mathbf{v}\mathbf{u}_{12}' + \mathbf{u}_1\mathbf{u}_2' + \mathbf{u}_2\mathbf{u}_1'\right)\mathbf{C}\widehat{\mathbf{s}}\right\rangle_{\mathbf{C}}\Big\}\end{aligned} \tag{38}$$

The real term of interest in this equation is the last term, say $2W$, since it is advantageous to smooth if, and only if, it is negative. In this case, $\mathcal{L}(\rho)$ will be a decreasing function near zero. Using the fact that

$$\widehat{\mathbf{s}}_n - \mathbf{P}_\ell \mathbf{s} = \mathbf{P}_\ell \boldsymbol{\epsilon} + n^{-1/2}\left(\mathbf{T}_\ell \mathbf{R}\mathbf{P}_\ell + \mathbf{P}_\ell \mathbf{R}\mathbf{T}_\ell\right)(\mathbf{s} + \boldsymbol{\epsilon}) + \cdots$$

and $\widehat{\mathbf{s}}$, $\boldsymbol{\epsilon}$ and $\widehat{\mathbf{v}}$ are independent, we get:

$$\begin{aligned} W =\ & \rho\mathbb{E}\left\langle \mathbf{P}\boldsymbol{\epsilon}, \left(\mathbf{u}_2\mathbf{v}' + \mathbf{v}\mathbf{u}_2'\right)\mathbf{C}\boldsymbol{\epsilon}\right\rangle_{\mathbf{C}} \\ & + \rho n^{-1/2}\mathbb{E}\left\langle \mathbf{P}\boldsymbol{\epsilon}, \left(\mathbf{u}_{12}\mathbf{v}' + \mathbf{v}\mathbf{u}_{12}' + \mathbf{u}_2\mathbf{u}_1' + \mathbf{u}_1\mathbf{u}_2'\right)\mathbf{C}\boldsymbol{\epsilon}\right\rangle_{\mathbf{C}} \\ & + \rho n^{-1/2}\mathbb{E}\left\langle (\mathbf{T}\mathbf{R}\mathbf{P} + \ \mathbf{P}\mathbf{T}\mathbf{R})\widehat{\mathbf{s}}, \left(\mathbf{u}_2\mathbf{v}' + \mathbf{v}\mathbf{u}_2'\right)\mathbf{C}\widehat{\mathbf{s}}\right\rangle_{\mathbf{C}} . \end{aligned} \tag{39}$$

Since $\mathbb{E}\ \mathbf{T}\mathbf{R}\mathbf{P} = 0$ and $\mathbb{E}\mathbf{u}_1 = \mathbb{E}\mathbf{u}_{12} = 0$, the terms in $\rho n^{-1/2}$ are zero and

$$W \approx -\rho\frac{\sigma^2}{p}d_\ell^2. \tag{40}$$

This expression is *always negative* and consequently, it is always better to smooth to estimate a new curve. Nevertheless, we must keep in mind that expression (40) is based on an approximation of the loss function $\mathcal{L}(\rho)$. One can notice as well that W decreases when the number of design points p increases.

One can ask the question what is the ideal amount of smoothing to estimate optimally this new curve. For this, we need to calculate $\mathbb{E}\left\|(\mathbf{u}_2\mathbf{v}' + \mathbf{v}\mathbf{u}_2')\mathbf{C}\widehat{\mathbf{s}}\right\|_{\mathbf{C}}^2$. Consider for example

$$\begin{aligned}\mathbb{E}\left\langle \mathbf{u}_2\mathbf{v}'\mathbf{C}\widehat{\mathbf{s}}, \mathbf{u}_2\mathbf{v}'\mathbf{C}\widehat{\mathbf{s}}\right\rangle_{\mathbf{C}} &= \mathbb{E}\left\langle (\lambda\mathbf{T}_\ell\mathbf{A} + \tfrac{d_\ell^2}{2}\mathbf{I})\mathbf{P}_\ell\,\widehat{\mathbf{s}}, (\lambda\mathbf{T}_\ell\mathbf{A} + \tfrac{d_\ell^2}{2}\mathbf{I})\mathbf{P}_\ell\,\widehat{\mathbf{s}}\right\rangle_{\mathbf{C}} \\ &= \mathbb{E}\left\langle \widehat{\mathbf{s}}, \mathbf{P}_\ell(\lambda\mathbf{A}\mathbf{T}_\ell + \tfrac{d_\ell^2}{2}\mathbf{I})(\lambda\mathbf{T}_\ell\mathbf{A} + \tfrac{d_\ell^2}{2}\mathbf{I})\mathbf{P}_\ell\widehat{\mathbf{s}}\right\rangle_{\mathbf{C}} \\ &= \operatorname{tr}\left\{\mathbf{P}_\ell\left(\lambda^2\mathbf{A}\mathbf{T}_\ell^2\mathbf{A} + \lambda\tfrac{d_\ell^2}{2}(\mathbf{T}_\ell\mathbf{A} + \mathbf{A}\mathbf{T}_\ell) + \tfrac{d_\ell^4}{4}\mathbf{I}\right)\mathbf{P}_\ell\,\mathbb{E}\left(\widehat{\mathbf{s}}\widehat{\mathbf{s}}'\mathbf{C}\right)\right\} \\ &= \lambda\left(\lambda^2\operatorname{tr}\left\{\mathbf{P}_\ell\mathbf{A}\mathbf{T}_\ell^2\mathbf{A}\mathbf{P}_\ell\right\} + \tfrac{d_\ell^4}{4}\operatorname{tr}\mathbf{P}_\ell\right) \\ &= \lambda\left(\frac{d_\ell^4}{4} + \lambda^2\sum_{\nu\neq\ell}\frac{d_{\ell\nu}^2}{(\lambda_\ell - \lambda_\nu)^2}\right)\end{aligned}$$

which finally yields:

$$\begin{aligned}\mathbb{E}\left\|(\mathbf{u}_2\mathbf{v}' + \mathbf{v}\mathbf{u}_2')\mathbf{C}\widehat{\mathbf{s}}\right\|_{\mathbf{C}} =\ & \frac{\sigma^2}{p}\left(d_\ell^4 + 3(\lambda_\ell^{(z)})^2\sum_{\nu\neq\ell}\frac{d_{\ell\nu}^2}{(\lambda_\ell - \lambda_\nu)^2}\right. \\ & \left. + \lambda_\ell^{(z)}d_\ell^4 + (\lambda_\ell^{(z)})^2\sum_{\nu\neq\ell}\frac{(\lambda_\ell^{(z)} + \lambda_\nu^{(z)})d_{\ell\nu}^2}{(\lambda_\ell - \lambda_\nu)^2}\right) + o\left(\frac{1}{p}\right).\end{aligned}$$

Hence, the ideal amount of smoothing may be expressed as follows:

$$\rho^* \approx \frac{1}{p}\,\frac{\sigma^2 d_\ell^2}{\frac{\sigma^2}{p}a + b} \qquad (a > 0,\ b > 0). \tag{41}$$

The more the data are rough, the larger the optimal smoothing parameter has to be and the more the noise is dispersed, the more the data have to be smoothed.

5 Simulations

In previous sections, the links between the optimal smoothing parameter ρ^* and the number of curves n were studied. In this section we suggest to put the stress, by means of simulations, on the relation between the number of knots, the number of design points and the optimal smoothing parameter in the mean square error sense.

Artificial data sets were generated as follows:

$$Z_i(t) = C_{1i}\ \sin(1.5\pi t) + C_{2i}\ \sin(2.5\pi t) + C_{3i}\ \sin(3.5\pi t) \quad i = 1,\ldots,n;\ t \in [0,1], \tag{42}$$

where $\{C_{1i}, C_{2i}, C_{3i}, i = 1,\ldots,n\}$ are $3n$ pseudo random numbers indepedently and normaly distributed with zero mean and variance $\sigma_1^2 = 2$, $\sigma_2^2 = 1.3$, $\sigma_3^2 = 0.9$. Clearly, the subspace E_q which contains the random process Z is of dimension three and is generated by the set of functions $\{\ \sin(1.5\pi t), \sin(2.5\pi t), \sin(3.5\pi t)\ \}$.

Afterwards, these curves were sampled uniformly and corrupted by adding a white noise ϵ, normally distributed with zero mean and unit variance, at the design points. We finally observe the n random vectors belonging to $\mathbb{R}^p$:

$$\mathbf{Y}_i = \mathbf{Z}_i + \boldsymbol{\epsilon}_i, \quad i = 1,\ldots,n;$$

where $Y_{ij} = Z_i(\frac{j}{p-1}) + \epsilon_{ij}$, $j = 0,\ldots,p-1$.

We performed nine simulations by considering different values for n and p:

- n equal to $n_1 = 50$, $n_2 = 100$ and $n_3 = 200$.
- p equal to $p_1 = 30$, $p_2 = 60$ and $p_3 = 120$.

Estimates were constructed from the observations $\mathbf{Y}_i$ using the method described in section 2. We have decided to choose $m = 2$, that is to say the roughness penalty is the $L^2[0,1]$ norm of the second derivative, and $\nu = 3$ which means that curves where approximated by cubic B-splines. Let's denote by $\widehat{\mathbf{Y}}_i(\rho, q, k)$ the rank constrained estimation of the ith curve with smoothing parameter ρ in a q-dimensional subspace of $\mathcal{S}_{k\nu}$. For k and p fixed, the optimal smoothing parameters q^* and ρ^* where chosen by minimizing the true quadratic risk function:

$$R_k(\rho, q) \ = \ \frac{1}{n}\sum_{i=1}^{n}\frac{1}{p}\left\|\widehat{\mathbf{Y}}_i(\rho, q, k) - \mathbf{Z}_i\right\|^2. \tag{43}$$

k=8	q^*	ρ^*	$R(q^*, \rho^*)$	$R(q^{**}, 0)$
p=30	3	5.0e-06	0.057	0.107
p=60	3	2.4e-06	0.024	0.033
p=120	3	1.2e-06	0.011	0.013

Table 1: *Quadratic risk,* $k = 8$ *and* $n = 200$.

k=16	q^*	ρ^*	$R(q^*, \rho^*)$	$R(q^{**}, 0)$
p=30	3	9.8e-06	0.054	0.144
p=60	3	5.1e-06	0.024	0.039
p=120	3	1.2e-06	0.011	0.014

Table 2: *Quadratic risk,* $k = 16$ *and* $n = 50$.

In order to be able to compare the efficiency of these smooth estimates with the B-splines estimates obtained without using a smoothing paramater, we have chosen another optimal dimension, q^{**}, by minimizing $R_k(0, q)$ with respect to q. Then, we have compared $R_k(\rho*, q^*)$ with $R_k(0, q^{**})$.

The analysis of the results of these simulations leads us to make the following remarks which can give ideas for future work. It seems that:

- the number of knots k has no real effect, on these simulations, on the accuracy of the smooth estimates. A too large k is compensated the roughness penalty controlled by the smoothing parameter value (see Table 2 and Table 3).

- for small numbers of design points, the use of a smoothing parameter may improve significantly the efficiency of these estimates (see Table 1, Table 2 and Table 3). On the other hand, if p is large, smoothing does not seem to be justifed since it does not improve significantly the estimation. These remarks agree with results of the previous section, in particular equation (40).

- smoothing gives more flexibility to the estimates and allows to compensate a bad dimension choice (see figure 1). Actually, the true dimension is 3 but an estimate in a subspace of dimension 4 can nearly attain the optimal risk provided that ρ is well chosen.

k=24	q^*	ρ^*	$R(q^*, \rho^*)$	$R(q^{**}, 0)$
p=30	3	9.8e-06	0.059	0.158
p=60	3	2.4e-06	0.025	0.039
p=120	3	1.2e-06	0.011	0.014

Table 3: *Quadratic risk,* $k = 24$ *and* $n = 100$.

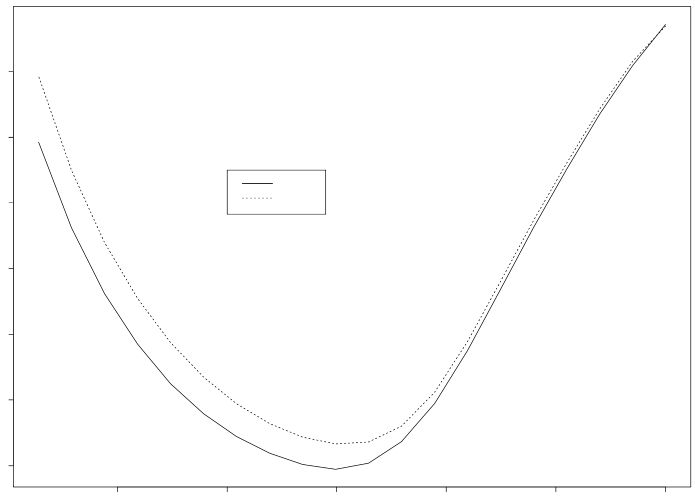

Figure 1: *Risk function R_k with respect to ρ and dimension q ($k = 24$, $p = 30$ and $n = 100$).*

- the optimal dimension q^* doesn't seem to be directly linked with the optimal smoothing parameter. Actually, we notice that q^* is always equal to q^{**}. We think that it may be possible to choose the dimension and the smoothing parameter separately by using two different criteria. For instance, one could choose q^* in a first step, and then ρ^*.

6 Proofs

6.1 Technical lemmas on B-splines

The following lemmas on B-splines will be stated since they will be needed later on to prove results of convergence.

Lemma 6.1
B-splines are nonnegative functions which satisfy the following equation:

$$\sum_{j=1}^{r} B_{kj}(t) = 1, \quad \forall t \in [0,1].$$

Futhermore, the support of each B-spline B_{kj} is included in $[\delta_j, \delta_{j+\nu}]$ where $\delta_1 = \cdots = \delta_\nu = 0,\ \delta_{j+\nu} = j/(k+1),\ j = 1,\ldots,k,\ \delta_{k+1+\nu} = \cdots = \delta_{k+2\nu} = 1.$

Lemma 6.2 *Under assumption $\mathbf{A}_1$ and if $k < p$, we have:*

- $\|\mathbf{C}_k\| = \mathcal{O}(k^{-1})$ *and* $\left\|\mathbf{C}_k^{-1}\right\| = \mathcal{O}(k)$
- $\left\|\widetilde{\mathbf{C}}_k\right\| = \mathcal{O}(k^{-1})$ *and* $\left\|\widetilde{\mathbf{C}}_k^{-1}\right\| = \mathcal{O}(k)$

- $\left\|\mathbf{C}_k - \widetilde{\mathbf{C}}_k\right\| = \mathcal{O}(p^{-1})$,
- $\|\mathbf{G}_k\| = \mathcal{O}(k^{2m-1})$,

where $\|.\|$ *is the usual matrix norm.*

Proof of lemma 6.2
The first three points were demonstrated by Agarwall & Studden (1980) and Burman (1991). They lie on compact support and regularity properties of splines functions.

On the other hand, each B_{ki} has m continuous derivatives for $m < \nu$ and satisfies (Schumaker 1981, theorem 4.22) :

$$\sup_{t\in[0,1]} |B_{ki}^{(m)}(t)| \leq ck^m,$$

where constant c does not depend on k and i. Furthermore, each B-spline has a support length of order k^{-1} (lemma 6.1). Thus, the elements of matrix $\mathbf{G}_k$ satisfy:

$$|[\mathbf{G}_k]_{lj}| \leq \int_0^1 \left|B_{kj}^{(m)}(t)\right| \left|B_{kl}^{(m)}(t)\right| \, dt = \begin{cases} 0 & \text{if } |l-j| > \nu \\ \mathcal{O}(k^{-1}k^{2m}) & \text{else.} \end{cases}$$

Therefore, applying theorem 1.19 of Chatelin (1983) to the symmetric matrix $\mathbf{G}_k$ we obtain:

$$\|\mathbf{G}_k\| \leq \sup_{j=1,\dots,r} \sum_{l=1}^{r} |[\mathbf{G}_k]_{lj}| = \mathcal{O}(k^{2m-1}).$$

□

We also need to bound the distance between the two matrices $\mathbf{H}_{k,\rho}^{1/2}$ and $\mathbf{C}_k^{-1/2}$ for "small" smoothing parameters ρ.

Lemma 6.3 *For small* ρ *and for all* k *we have:*

$$\left\|\mathbf{H}_{k,\rho}^{1/2} - \mathbf{C}_k^{-1/2}\right\| = \mathcal{O}(\rho\, k^{\frac{4m+1}{2}}).$$

Proof *of lemma* 6.3
The proof of this lemma is based on perturbation theory of linear operators (Kato 1976, Chatelin 1983). Since the matrix $\mathbf{C}_k$ is positive, its resolvent $\mathbf{r}(\lambda) = (\mathbf{C}_k + \lambda\mathbf{I})^{-1}$ exists for $\lambda > 0$ and we have (Kato, 1976 pp 282):

$$\mathbf{C}_k^{-1/2} = \frac{1}{\pi}\int_0^{+\infty} \frac{\mathbf{r}(\lambda)}{\sqrt{\lambda}}\, d\lambda. \tag{44}$$

Let's consider $\widetilde{\mathbf{r}}(\lambda) = \left(\mathbf{H}_{k,\rho}^{-1} + \lambda\mathbf{I}\right)^{-1}$. The second resolvent equation gives us:

$$\begin{aligned} \widetilde{\mathbf{r}}(\lambda) - \mathbf{r}(\lambda) &= -\mathbf{r}(\lambda)(\mathbf{H}_{k,\rho}^{-1} - \mathbf{C}_k)\widetilde{\mathbf{r}}(\lambda) \\ &= -\rho\, \mathbf{r}(\lambda)\mathbf{G}_k\widetilde{\mathbf{r}}(\lambda), \end{aligned} \tag{45}$$

and thus, using (44) and (45), we obtain:

$$\mathbf{H}_{k,\rho}^{1/2} - \mathbf{C}_k^{-1/2} = -\frac{\rho}{\pi}\int_0^{+\infty} \mathbf{r}(\lambda)\mathbf{G}_k\widetilde{\mathbf{r}}(\lambda)\frac{d\lambda}{\sqrt{\lambda}}. \tag{46}$$

Let λ_0 be the smallest eigenvalue of the matrices $\mathbf{C}_k$ and $\mathbf{H}_{k,\rho}^{-1}$. We have $\|\mathbf{r}(\lambda)\| \leq (\lambda_0+\lambda)^{-1}$ and from lemma 6.2, we get $\lambda_0 = \mathcal{O}(k^{-1})$. Hence, using (46), we can bound:

$$\begin{aligned}\left\|\mathbf{H}_{k,\rho}^{1/2} - \mathbf{C}_k^{-1/2}\right\| &= \frac{\rho}{\pi}\left\|\int_0^{+\infty} \mathbf{r}(\lambda)\mathbf{G}_k\widetilde{\mathbf{r}}(\lambda)\frac{d\lambda}{\sqrt{\lambda}}\right\| \\ &\leq \frac{\rho}{\pi}\left(\int_0^{+\infty} \frac{(\lambda_0+\lambda)^{-2}}{\sqrt{\lambda}}\ d\lambda\right)\|\mathbf{G}_k\| \\ &= \mathcal{O}(\rho\ k^{3/2}\ k^{2m-1}), \end{aligned} \tag{47}$$

since $\|\mathbf{G}_k\| = \mathcal{O}(k^{2m-1})$ and $\displaystyle\int_0^{+\infty} \frac{(\lambda_0+\lambda)^{-2}}{\sqrt{\lambda}}\ d\lambda \leq C\lambda_0^{-3/2} = \mathcal{O}(k^{3/2})$.

□

6.2 Asymptotic behaviour of the MISE for the mean function

Proof *of theorem 3.1*

Beginning with the study of the asymptotic behavior of the bias, denote by $\widehat{\mu}_{k,n}$ the least squares estimate of the mean function in the B-splines basis which can be written:

$$\widehat{\mu}_{k,n}(t) = \widehat{\mathbf{b}}_{k,n}'\widetilde{\mathbf{C}}_k^{-1}\mathbf{B}_k(t), \tag{48}$$

where $\widehat{\mathbf{b}}_{k,n} = \frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{b}}_i$. Considering the expectation, $\widetilde{\mu}_k(t)$, as follows:

$$\widetilde{\mu}_k(t) = \widetilde{\mathbf{b}}_k'\widetilde{\mathbf{C}}_k^{-1}\mathbf{B}_k(t), \tag{49}$$

where $\widetilde{\mathbf{b}}_k = \frac{1}{p}\sum_{j=1}^p \mu(t_j)\mathbf{B}_k(t_j)$, we have for the bias:

$$\|\mu - \widetilde{\mu}_\rho\|_{L^2}^2 \ \leq\ 2\left(\|\widetilde{\mu}_\rho - \widetilde{\mu}_k\|_{L^2}^2 + \|\mu - \widetilde{\mu}_k\|_{L^2}^2\right).$$

Under assumption $\mathbf{A}_1$ and condition $k = o(p)$, one obtains from Agarwall & Studden (1980) (theorem 3.1B):

$$\|\widetilde{\mu}_k - \mu\|_{L^2}^2 \ \approx\ k^{-2m}\int_0^1 \left(D^m\mu(t)\right)^2\ dt. \tag{50}$$

Furthermore, if $\rho = o(k^{-2m})$ it is possible to expand the "hat matrix" $\mathbf{H}_{k,\rho}$:

$$\begin{aligned}\mathbf{H}_{k,\rho}\mathbf{C}_k &= \left(\mathbf{I} + \rho\mathbf{C}_k^{-1}\mathbf{G}_k\right)^{-1} \\ &= \mathbf{I} - \rho\mathbf{C}_k^{-1}\mathbf{G}_k + o(\rho k^{2m}),\end{aligned} \tag{51}$$

because we have $\left\|\mathbf{C}_k^{-1}\mathbf{G}_k\right\| = \mathcal{O}(k^{2m})$ from lemma 6.2. Using this expansion we get

$$\begin{aligned}
\|\widetilde{\mu}_\rho - \widetilde{\mu}_k\|_{L^2}^2 &= \int_0^1 [(\mathbb{E}\widehat{\mathbf{s}}_\rho - \mathbb{E}\widehat{\mathbf{s}}_{k,n})' \mathbf{B}_k(t)]^2 \, dt \\
&= (\mathbb{E}\widehat{\mathbf{s}}_\rho - \mathbb{E}\widehat{\mathbf{s}}_{k,n})' \mathbf{C}_k (\mathbb{E}\widehat{\mathbf{s}}_\rho - \mathbb{E}\widehat{\mathbf{s}}_{k,n}) \\
&= \mathbb{E}\widehat{\mathbf{s}}_{k,n}' (\mathbf{H}_{k,\rho}\mathbf{C}_k - \mathbf{I})' \mathbf{C}_k (\mathbf{H}_{k,\rho}\mathbf{C}_k - \mathbf{I}) \mathbb{E}\widehat{\mathbf{s}}_{k,n} \\
&= \rho^2 \mathbb{E}\widehat{\mathbf{s}}_{k,n}' \mathbf{G}_k \mathbf{C}_k^{-1} \mathbf{G}_k \mathbb{E}\widehat{\mathbf{s}}_{k,n} + o(\rho^2 k^{4m}). \qquad (52)
\end{aligned}$$

It is easy to check that $\|\mathbb{E}\widehat{\mathbf{s}}_{k,n}\|^2 = \mathcal{O}(k)$ and using lemma 6.2 again we have $\left\|\mathbf{G}_k\mathbf{C}_k^{-1}\mathbf{G}_k\right\| = \mathcal{O}(k^{4m-1})$. Therefore the bias is bounded above as follows:

$$\|\mu - \widetilde{\mu}_\rho\|_{L^2}^2 = \mathcal{O}(k^{-2m}) + \mathcal{O}(\rho^2 k^{4m}).$$

Writing the variance term as follows:

$$\begin{aligned}
\mathbb{E}\,\|\widetilde{\mu}_\rho - \widehat{\mu}_\rho\|_{L^2}^2 &= \mathbb{E} \int_0^1 (\widetilde{\mathbf{s}}_\rho - \widehat{\mathbf{s}}_\rho)' \mathbf{B}_k(t) \, dt \\
&= \mathbb{E}\,(\widetilde{\mathbf{s}}_\rho - \widehat{\mathbf{s}}_\rho)' \mathbf{C}_k (\widetilde{\mathbf{s}}_\rho - \widehat{\mathbf{s}}_\rho) \\
&= \mathbb{E} \left(\widetilde{\mathbf{b}}_k - \widehat{\mathbf{b}}_{k,n}\right)' \mathbf{H}_{k,\rho}\mathbf{C}_k\mathbf{H}_{k,\rho} \left(\widetilde{\mathbf{b}}_k - \widehat{\mathbf{b}}_{k,n}\right) \\
&\leq \|\mathbf{H}_{k,\rho}\mathbf{C}_k\mathbf{H}_{k,\rho}\| \, \mathbb{E} \left\|\widetilde{\mathbf{b}}_k - \widehat{\mathbf{b}}_{k,n}\right\|^2. \qquad (53)
\end{aligned}$$

Matrix $\mathbf{C}_k$ is positive and definite, matrix $\mathbf{G}_k$ is nonnegative, and the smallest eigenvalue of $\mathbf{C}_k$ is of order k^{-1} (lemma 6.2) whereas $\mathbf{G}_k$ has m null eigenvalues since the derivative of order m of polynomials whose degree is less than m is null. Therefore the smallest eigenvalue of $\mathbf{H}_{k,\rho}^{-1} = \mathbf{C}_k + \rho\mathbf{G}_k$ is of order k^{-1} and $\|\mathbf{H}_{k,\rho}\| = \mathcal{O}(k)$. Thus

$$\begin{aligned}
\|\mathbf{H}_{k,\rho}\mathbf{C}_k\mathbf{H}_{k,\rho}\| &\leq \|\mathbf{H}_{k,\rho}\|^2 \|\mathbf{C}_k\| \\
&= \mathcal{O}(k^2)\mathcal{O}(k^{-1}) \\
&= \mathcal{O}(k), \qquad (54)
\end{aligned}$$

and it remains to bound $\mathbb{E}\left\|\widetilde{\mathbf{b}}_k - \widehat{\mathbf{b}}_{k,n}\right\|^2$. Let's consider $\bar{\mathbf{Z}}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{Z}_i$ and $\bar{\boldsymbol{\epsilon}}_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\epsilon}_i$ and write:

$$\widehat{\mathbf{b}}_{k,n} - \widetilde{\mathbf{b}}_k = \frac{1}{p}\sum_{j=1}^p [\bar{\mathbf{Z}}_n + \bar{\boldsymbol{\epsilon}}_n]_j \; \mathbf{B}_k(t_j).$$

By assumptions, we have $\mathbb{E}(\bar{\mathbf{Z}}_n) = E(\bar{\boldsymbol{\epsilon}}_n) = 0$, $\mathbb{E}\bar{\mathbf{Z}}_n\bar{\mathbf{Z}}_n' = \frac{1}{n}\boldsymbol{\Gamma}_p$, $\mathbb{E}\bar{\boldsymbol{\epsilon}}_n\bar{\boldsymbol{\epsilon}}_n' = \frac{\sigma^2}{n}\mathbf{I}_p$, $\mathbb{E}\bar{\mathbf{Z}}_n\bar{\boldsymbol{\epsilon}}_n' = 0$,

$$\text{and} \qquad \mathbb{E}[(\bar{\mathbf{Z}}_n + \bar{\boldsymbol{\epsilon}}_n)(\bar{\mathbf{Z}}_n + \bar{\boldsymbol{\epsilon}}_n)'] = \frac{1}{n}(\boldsymbol{\Gamma}_p + \sigma^2\mathbf{I}).$$

From equation (7) and assumption $\mathbf{A}_2$ on the regularity of Z one can check that $\sup_{t\in[0,1]}\{\mathbb{E}(Z(t)^2\} < +\infty$. Let's define $\Delta = \frac{1}{n}(\boldsymbol{\Gamma}_p + \sigma^2\mathbf{I}_p)$ and $D = \max(\Delta_{ij})$. Then we have $D = \mathcal{O}(1/n)$. Furthermore,

$$\begin{aligned}\mathbb{E}\left\|\widehat{\mathbf{b}}_{k,n} - \widetilde{\mathbf{b}}_k\right\|^2 &= \frac{1}{p^2}\sum_{1\leq l,j\leq p}\Delta_{l,j}\sum_{i=1}^{k+\nu}B_{ki}(t_l)B_{ki}(t_j)\\ &\leq \frac{D}{p^2}\sum_{1\leq l,j\leq p}\sum_{i=1}^{k+\nu}B_{ki}(t_l)B_{ki}(t_j).\end{aligned}$$

Let's write $\frac{1}{p^2}\sum_{1\leq l,j\leq p}\sum_{i=1}^{k+\nu}B_{ki}(t_l)B_{ki}(t_j) = \sum_{i=1}^{k+\nu}\left(\frac{1}{p}\sum_{j=1}^{p}B_{ki}(t_j)\right)\left(\frac{1}{p}\sum_{l=1}^{p}B_{ki}(t_l)\right)$. From lemma 6.1, we know that each B-spline has a support length of order k^{-1} and therefore the cardinal of the set $\{l \mid B_{ki}(t_l) \neq 0\}$ is of order pk^{-1}, and $\frac{1}{p}\sum_{j=1}^{p}B_{ki}(t_j) = \mathcal{O}(k^{-1})$. Hence

$$\frac{1}{p^2}\sum_{1\leq l,j\leq p}\sum_{i=1}^{k+\nu}B_{ki}(t_l)B_{ki}(t_j) = \mathcal{O}(k^{-1}), \tag{55}$$

and we obtain:

$$\mathbb{E}\left\|\widehat{\mathbf{b}}_{k,n} - \widetilde{\mathbf{b}}_k\right\|^2 = \mathcal{O}(k^{-1}n^{-1}). \tag{56}$$

Finally, we get the desired result by using (56), (54) and (53). □

6.3 Results on the covariance operator and its eigenvectors

Proof *of lemma 3.3*

Denote by $\widetilde{z}$ the B-splines estimation of the signal z deduced from $\mathbf{Z}$. Using lemma 3.2, we get the following bound:

$$\mathbb{E}\left\|\widetilde{\Gamma}_{np} - \widetilde{\Gamma}_p\right\|_{\mathcal{H}}^2 \leq \frac{\mathbb{E}\left\|\widetilde{z}\right\|_{L^2}^4}{n}.$$

We only need to study $\mathbb{E}\left\|\widetilde{z}\right\|_{L^2}^4$; let's define $\widetilde{\mathbf{b}} = 1/p\sum_{j=1}^{p}Z(t_j)\mathbf{B}_k(t_j)$ and write $\widetilde{z}$ into the B-splines basis to get:

$$\begin{aligned}\mathbb{E}\left\|\widetilde{z}\right\|_{L^2}^4 &= \mathbb{E}\left\{\int\left(\widetilde{\mathbf{b}}'\widetilde{\mathbf{C}}_k^{-1}\mathbf{B}_k(t)\right)^2 dt\right\}^2\\ &\leq \left\|\widetilde{\mathbf{C}}_k^{-1}\right\|^4\left\|\mathbf{C}_k\right\|^2\mathbb{E}\left\{\widetilde{\mathbf{b}}'\widetilde{\mathbf{b}}\right\}^2\\ &\leq \mathcal{O}(k^2)\mathbb{E}\left\{\widetilde{\mathbf{b}}'\widetilde{\mathbf{b}}\right\}^2,\end{aligned}$$

using lemma 6.2. Let's consider $D = \sup_{t\in[0,1]}\mathbb{E}[Z(t)^4] < \infty$. It yields

$$\begin{aligned}\mathbb{E}\left\{\widetilde{\mathbf{b}}'\widetilde{\mathbf{b}}\right\}^2 &= \frac{1}{p^4}\sum_{a,b,c,d=1,\ldots,p}\mathbb{E}(Z(t_a)Z(t_b)Z(t_c)Z(t_d))\mathbf{B}_k(t_a)'\mathbf{B}_k(t_b)\mathbf{B}_k(t_c)'\mathbf{B}_k(t_d)\\ &\leq D\frac{1}{p^4}\sum_{a,b,c,d=1,\ldots,p}\mathbf{B}_k(t_a)'\mathbf{B}_k(t_b)\mathbf{B}_k(t_c)'\mathbf{B}_k(t_d)\\ &\leq D\frac{1}{p^4}\sum_{a,b=1,\ldots,p}\mathbf{B}_k(t_a)'\mathbf{B}_k(t_b)\sum_{c,d=1,\ldots,p}\mathbf{B}_k(t_c)'\mathbf{B}_k(t_d).\end{aligned}$$

Then, using (55), we obtain:

$$\frac{1}{p^2}\sum_{a,b=1,\ldots,p}\mathbf{B}_k(t_a)'\mathbf{B}_k(t_b)=\frac{1}{p^2}\sum_{i=1}^{k+\nu}\sum_{a=1}^{p}\sum_{b=1}^{p}B_{ki}(t_a)B_{ki}(t_b)=\mathcal{O}(k^{-1}),$$

and $\mathbb{E}\left\{\widetilde{\mathbf{b}}'\widetilde{\mathbf{b}}\right\}^2=\mathcal{O}(k^{-2})$. Finally, we deduce from above that $\mathbb{E}\left\|\widetilde{z}\right\|_{L^2}^4=\mathcal{O}(1)$, and the desired result

$$\mathbb{E}\left\|\widetilde{\Gamma}_{np}-\widetilde{\Gamma}_p\right\|_{\mathcal{H}}^2=\mathcal{O}(n^{-1}).$$

□

Proof *of lemma 3.4*
Let's define, for $i=1,\ldots,n$, $\widetilde{\mathbf{b}}_{i,\epsilon}=1/p\sum_{j=1}^p[\boldsymbol{\epsilon}_i]_j\mathbf{B}_k(t_j)$ and $\widetilde{\mathbf{s}}_{i,\epsilon}=\widetilde{\mathbf{C}}_k^{-1}\widetilde{\mathbf{b}}_{i,\epsilon}$. Let's write

$$\begin{aligned}\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{\mathbf{b}}_{i,\epsilon}\widetilde{\mathbf{b}}_{i,\epsilon}'\right) &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{p^2}\sum_{j,\ell=1}^{p}\mathbb{E}([\boldsymbol{\epsilon}_i]_j[\boldsymbol{\epsilon}_i]_\ell)\mathbf{B}_k(t_j)\mathbf{B}_k(t_\ell)' \\ &= \frac{1}{p^2}\sum_{j}^{p}\sigma^2\mathbf{B}_k(t_j)\mathbf{B}_k(t_j)' \\ &= \frac{\sigma^2}{p}\widetilde{\mathbf{C}}_k.\end{aligned}$$

Hence, we have $\mathbb{E}(\widetilde{\mathbf{s}}_{i,\epsilon}\widetilde{\mathbf{s}}_{i,\epsilon}')=\sigma^2/p\ \widetilde{\mathbf{C}}_k^{-1}$ and the first part of the proof is obtained readily by expressing $\widetilde{\Gamma}_{\epsilon,np}$ in the basis $(e_{ij})_{i,j=1,\ldots,r}$.

Let's consider $d_4=\mathbb{E}(\epsilon^4)$ and define $\mathbf{b}_\epsilon=1/p\sum_{j=1}^p[\boldsymbol{\epsilon}]_j\mathbf{B}_k(t_j)$. By writing the B-splines approximation of the white noise, say $\widetilde{\epsilon}$, deduced from the vector $\boldsymbol{\epsilon}$, we get $\mathbb{E}\left\|\widetilde{\epsilon}\right\|^4=\mathcal{O}(k^2)\mathbb{E}\{\mathbf{b}_\epsilon'\mathbf{b}_\epsilon\}^2$ and

$$\begin{aligned}\mathbb{E}\left\{\mathbf{b}_\epsilon'\mathbf{b}_\epsilon\right\}^2 &= \frac{1}{p^4}\sum_{a,b,c,d=1\ldots p}\mathbb{E}\{\epsilon_a\epsilon_b\epsilon_c\epsilon_d\}\mathbf{B}_k(t_a)'\mathbf{B}_k(t_b)\mathbf{B}_k(t_c)'\mathbf{B}_k(t_d) \\ &= \frac{d_4}{p^4}\sum_{a=1}^{p}(\mathbf{B}_k(t_a)'\mathbf{B}_k(t_a))^2+\frac{(\sigma^2)^2}{p^4}\sum_{a=1}^{p}\sum_{b\neq a}(\mathbf{B}_k(t_a)'\mathbf{B}_k(t_b))^2 \\ &\quad+\frac{(\sigma^2)^2}{p^4}\sum_{a=1}^{p}\sum_{b\neq a}\mathbf{B}_k(t_a)'\mathbf{B}_k(t_b)\mathbf{B}_k(t_b)'\mathbf{B}_k(t_a) \\ &\quad+\frac{(\sigma^2)^2}{p^4}\sum_{a=1}^{p}\sum_{b\neq a}\mathbf{B}_k(t_a)'\mathbf{B}_k(t_a)\mathbf{B}_k(t_b)'\mathbf{B}_k(t_b).\end{aligned}$$

The first term is negligible since $\mathbf{B}_k(t_a)'\mathbf{B}_k(t_a)\leq 1$. Moreover, using the following equality

$$\frac{1}{p}\sum_{a=1}^{p}\mathbf{B}_k(t_a)'\mathbf{B}_k(t_a)=\mathrm{tr}(\widetilde{\mathbf{C}}_k)$$

and $\left\|\widetilde{\mathbf{C}}_k\right\|=\mathcal{O}(k^{-1})$ we finally get $\mathbb{E}\{\mathbf{b}_\epsilon'\mathbf{b}_\epsilon\}^2=\mathcal{O}(p^{-2})$ and the result. □

Proof *of theorem 3.5*

Since for all $i = 1, \ldots, n$, we have $\mathbb{E}(\widetilde{\epsilon}_i) = 0$ and $\widetilde{\epsilon}_i$ is independent of $\widetilde{z}_i$, it is easy to check that $\mathbb{E}\,\widetilde{\Gamma}_{1,np} = 0$ and $\mathbb{E}\,\widetilde{\Gamma}_{2,np} = 0$. Hence, we have:

$$\mathbb{E}\,\widehat{\Gamma}_{np} = \widetilde{\Gamma}_p + \frac{\sigma^2}{p}\widetilde{P}_{k,p}\,.$$

Furthermore, since $\widetilde{\Gamma}_{1,np}$ is the transposed of $\widetilde{\Gamma}_{2,np}$, these operators have the same properties. Using (20) and the independence between $\mathbf{Z}$ and $\boldsymbol{\epsilon}$, one can bound $\mathbb{E}\left\|\widetilde{\Gamma}_{1,np}\right\|_{\mathcal{H}}^2$ as follows:

$$\begin{aligned}
\mathbb{E}\left\|\widetilde{\Gamma}_{1,np}\right\|_{\mathcal{H}}^2 &\leq \frac{\mathbb{E}\,\|\widetilde{z}\|_{L^2}^2\;\mathbb{E}\,\|\widetilde{\epsilon}\|_{L^2}^2}{n} \\
&= \mathcal{O}\left(\frac{k}{np}\right) \qquad (57)
\end{aligned}$$

because $\mathbb{E}\,\|\widetilde{\epsilon}\|_{L^2}^2 = \mathcal{O}\left(\frac{k}{p}\right)$.
Let's write now:

$$\begin{aligned}
\widehat{\Gamma}_{np} - \left(\widetilde{\Gamma}_p + \frac{\sigma^2}{p}\widetilde{P}_{k,p}\right) &= \left(\widetilde{\Gamma}_{np} - \widetilde{\Gamma}_p\right) + \left(\widetilde{\Gamma}_{1,np} + \widetilde{\Gamma}_{2,np}\right) \qquad (58) \\
&\quad + \left(\widetilde{\Gamma}_{\epsilon,np} - \frac{\sigma^2}{p}\widetilde{P}_{k,p}\right),
\end{aligned}$$

and use lemmas 3.3, 3.4 and equation (57), in order to get

$$\left(\mathbb{E}\left\|\widehat{\Gamma}_{np} - \left(\widetilde{\Gamma}_p + \frac{\sigma^2}{p}\widetilde{P}_{k,p}\right)\right\|_{\mathcal{H}}^2\right)^{1/2} = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\left(1 + \mathcal{O}\left(\frac{\sqrt{k}}{\sqrt{p}}\right) + \left(\frac{k}{p}\right)\right).$$

The result is obtained readily since the terms of order k/p are negligible. □

Proof *of lemma 3.6*
Let's express $\widetilde{\Gamma}_p$ in the basis (e_{lj}) defined in (22):

$$\widetilde{\Gamma}_p = \sum_{l,j=1}^{r} \widetilde{a}_{lj} e_{l,j} \quad \text{where} \quad [\widetilde{a}_{lj}]_{l,j} = \widetilde{\mathbf{A}} = \widetilde{\mathbf{C}}_k^{-1}\left[\frac{1}{p^2}\sum_{j,l=1}^{p} \mathbb{E}(Z(t_j)Z(t_l))\mathbf{B}_k(t_j)\mathbf{B}_k'(t_l)\right]\widetilde{\mathbf{C}}_k^{-1}.$$

Furthermore, we get from expansion (7) of the random function Z, that its covariance may be expressed as follows:

$$\mathbb{E}\left(Z(t_j)Z(t_l)\right) = \sum_{i=1}^{p} \lambda_i \phi_i(t_j)\phi_i(t_l).$$

Let's denote by $\widetilde{\phi}_{p,i}$ the B-splines approximation of the function ϕ_i deduced from its observation at the design points $(t_j)_j$. One can see easily that $\widetilde{\Gamma}_p$ may be written equivalently as follows:

$$\widetilde{\Gamma}_p = \sum_{i=1}^{q} \lambda_i \widetilde{\phi}_{p,i} \otimes \widetilde{\phi}_{p,i}. \qquad (59)$$

Let's notice that the functions $\widetilde{\phi}_{p,i}$, which are approximations of the eigenvectors of Γ, are not *a priori*, the eigenvectors of the operator $\widetilde{\Gamma}_p$. By assumption, functions ϕ_i satisfy the regularity condition $\mathbf{A}_2$, and so we get (theorem 3.1B, Agarwall & Studden 1980):

$$\left\|\phi_i - \widetilde{\phi}_{p,i}\right\|_{L^2}^2 \approx k^{-2m} \left\|D^m \phi_i\right\|_{L^2}^2 . \tag{60}$$

Thus,

$$\begin{aligned}
\left\|\Gamma - \widetilde{\Gamma}_p\right\|_{\mathcal{H}} &= \left\|\sum_{i=1}^{q} \lambda_i \left(\phi_i \otimes \phi_i - \widetilde{\phi}_{p,i} \otimes \widetilde{\phi}_{p,i}\right)\right\|_{\mathcal{H}} \\
&= \left\|\sum_{i=1}^{q} \lambda_i \left[\left(\phi_i - \widetilde{\phi}_{p,i}\right) \otimes \phi_i + \widetilde{\phi}_{p,i} \otimes \left(\phi_i - \widetilde{\phi}_{p,i}\right)\right]\right\|_{\mathcal{H}} \\
&\leq \sum_{i=1}^{q} \lambda_i \left\|\phi_i - \widetilde{\phi}_{p,i}\right\|_{L^2} \left(\|\phi_i\|_{L^2} + \left\|\widetilde{\phi}_{p,i}\right\|_{L^2}\right) .
\end{aligned} \tag{61}$$

The first part of the proof is obtained by applying (60) to the q functions $\widetilde{\phi}_{p,i}$.

By definition of the norm we have:

$$\left\|\widetilde{P}_{k,p}\right\| = \sup\{| < (\widetilde{P}_{k,p})f, f > |, \|f\|_{L^2} = 1\}.$$

Let's define the vector $\mathbf{f}_k$ whose elements are $\{< f, B_{kj} >,\ j = 1, \ldots, r\}$. The term $< (\widetilde{P}_{k,p})f, f >$ may be expressed in a matrix form as follows:

$$< (\widetilde{P}_{k,p})f, f >= \mathbf{f}_k' \widetilde{\mathbf{C}}_k^{-1} \mathbf{f}_k.$$

Since $\|f\|_{L^2} = 1$, we have $\mathbf{f}_k'\mathbf{f}_k = \mathcal{O}(k^{-1})$ and using lemma 6.2, we get: $\left\|\widetilde{\mathbf{C}}_k^{-1}\right\| = \mathcal{O}(k)$. Hence, $\left\|\widetilde{P}_{k,p}\right\| = \mathcal{O}(1)$ and

$$\left\|\frac{\sigma^2}{p} \widetilde{P}_{k,p}\right\| = \mathcal{O}\left(\frac{1}{p}\right).$$

□

Proof *of theorem 3.7*
Bound (23) is obtained readily by applying theorem 3.5 and lemma 3.6.

We get the convergence of $\widehat{\phi}_{jn}$ towards ϕ_j by using lemma 3.1 of Bosq (1991). Indeed, it states that

$$\begin{aligned}
\left\|\phi_1 - \widehat{\phi}_{1n}\right\|_{L^2} &\leq \frac{2\sqrt{2}}{(\lambda_1 - \lambda_2)} \left\|\Gamma - \widehat{\Gamma}_{np}\right\| && \text{if } j = 1 \\
\left\|\phi_j' - \widehat{\phi}_{jn}\right\|_{L^2} &\leq \frac{2\sqrt{2}}{\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})} \left\|\Gamma - \widehat{\Gamma}_{np}\right\| && \text{if } j > 1
\end{aligned}$$

and allows us to obtain readily the second result.

□

Before embarking on the proof of theorem 3.8, we have to establish some notations. Let's denote by $\widehat{\mathbf{s}}_j^\rho$ the coordinates in the B-splines basis of the vectors $\widehat{\phi}_{j,n}^\rho$. These are obtained by applying $\mathbf{H}_{k,\rho}^{1/2}$ to the eigenvectors, say $\{\mathbf{v}_{j,n}^\rho,\ j=1,\ldots,q\}$, of matrix $\mathbf{H}_{k,\rho}^{1/2}\mathbf{C}_k\mathbf{S}_n\mathbf{C}_k\mathbf{H}_{k,\rho}^{1/2}$ defined in (15). Let's denote by $\{\mathbf{v}_j^\rho,\ j=1,\ldots q\}$ (respectively $\{\mathbf{v}_j,\ j=1,\ldots,q\}$) the first q eigenvectors of $\mathbf{H}_{k,\rho}^{1/2}\mathbf{C}_k\mathbb{E}\left(\mathbf{S}_n\right)\mathbf{C}_k\mathbf{H}_{k,\rho}^{1/2}$ (respectively of the matrix $\mathbf{C}_k^{1/2}\mathbb{E}\left(\mathbf{S}_n\right)\mathbf{C}_k^{1/2}$). These are obtained by applying $\mathbf{C}_k^{-1/2}$ to the coordinates of the eigenvectors of $\widehat{\Gamma}_{n,p}$. Finally, let's define the matrix $\mathbf{S}=\mathbb{E}\left(\mathbf{S}_n\right)$.

Proof *of theorem* 3.8

This proof is split into two parts. At first we prove the convergence of $\mathbf{v}_{j,n}^\rho$ towards $\mathbf{v}_j$. Then we show that the eigenfunctions $\widehat{\phi}_{j,n}^\rho$ converge towards ϕ_j.

At first, let's notice that

$$\mathbb{E}\left\|\mathbf{H}_{k,\rho}^{1/2}\mathbf{C}_k\mathbf{S}_n\mathbf{C}_k\mathbf{H}_{k,\rho}^{1/2}-\mathbf{H}_{k,\rho}^{1/2}\mathbf{C}_k\mathbf{S}\mathbf{C}_k\mathbf{H}_{k,\rho}^{1/2}\right\|^2 = \mathcal{O}\left(\frac{1}{n}\right) \tag{62}$$

by applying to the matrix $\mathbf{H}_{k,\rho}^{1/2}\mathbf{C}_k\mathbf{S}_n\mathbf{C}_k\mathbf{H}_{k,\rho}^{1/2}$ the same arguments as in lemma 3.2. Let's compare now the deterministic matrices $\mathbf{H}_{k,\rho}^{1/2}\mathbf{C}_k\mathbf{S}\mathbf{C}_k\mathbf{H}_{k,\rho}^{1/2}$ and $\mathbf{C}_k^{1/2}\mathbf{S}\mathbf{C}_k^{1/2}$. For this, let's write $\mathbf{C}_k^{1/2}\mathbf{S}\mathbf{C}_k^{1/2}=\mathbf{C}_k^{-1/2}\mathbf{C}_k\mathbf{S}\mathbf{C}_k\mathbf{C}_k^{-1/2}$ and define $I=\left\|\mathbf{H}_{k,\rho}^{1/2}\mathbf{C}_k\mathbf{S}\mathbf{C}_k\mathbf{H}_{k,\rho}^{1/2}-\mathbf{C}_k^{1/2}\mathbf{S}\mathbf{C}_k^{1/2}\right\|$. It yields

$$\begin{aligned} I &= \left\|\left(\mathbf{H}_{k,\rho}^{1/2}-\mathbf{C}_k^{-1/2}\right)\mathbf{C}_k\mathbf{S}\mathbf{C}_k\mathbf{H}_{k,\rho}^{1/2}+\mathbf{C}_k^{1/2}\mathbf{S}\mathbf{C}_k\left(\mathbf{H}_{k,\rho}^{1/2}-\mathbf{C}_k^{-1/2}\right)\right\| \\ &\leq \left\|\mathbf{H}_{k,\rho}^{1/2}-\mathbf{C}_k^{-1/2}\right\|\left(\left\|\mathbf{C}_k\mathbf{S}\mathbf{C}_k\mathbf{H}_{k,\rho}^{1/2}\right\|+\left\|\mathbf{C}_k^{1/2}\mathbf{S}\mathbf{C}_k\right\|\right). \end{aligned} \tag{63}$$

We have from lemma 6.3, $\left\|\mathbf{H}_{k,\rho}^{1/2}-\mathbf{C}_k^{-1/2}\right\|=\mathcal{O}(\rho\ k^{(4m+1)/2})$. Furthermore, it is easy to check that $\left\|\mathbf{H}_{k,\rho}^{1/2}\right\|=\mathcal{O}(k^{-1/2})$ and $\left\|\mathbf{C}_k^{1/2}\right\|=\mathcal{O}(k^{-1/2})$. If we write

$$\begin{aligned} \left\|\mathbf{C}_k^{1/2}\mathbf{S}\mathbf{C}_k\right\| &\leq \left\|\mathbf{C}_k^{1/2}\right\|\|\mathbf{S}\|\|\mathbf{C}_k\|, \\ \left\|\mathbf{C}_k\mathbf{S}\mathbf{C}_k\mathbf{H}_{k,\rho}^{1/2}\right\| &\leq \|\mathbf{C}_k\|^2\|\mathbf{S}\|\left\|\mathbf{H}_{k,\rho}^{1/2}\right\|, \end{aligned}$$

it remains to bound $\|\mathbf{S}\|$ to get a bound for I. In order to do this, let's write $\|\mathbf{S}\|=\|\mathbb{E}1/n\sum_{i=1}^n\mathbf{s}_i\mathbf{s}_i'\|\leq 1/n\sum_{i=1}^n\|\mathbb{E}\mathbf{s}_i\mathbf{s}_i'\|=\mathcal{O}(k)$ since $\|\mathbb{E}\mathbf{s}_i\mathbf{s}_i'\|=\mathcal{O}(k)$ by similar arguments as those used in the proof of theorem 3.1. Thus, it yields

$$\left\|\mathbf{H}_{k,\rho}^{1/2}\mathbf{C}_k\mathbf{S}\mathbf{C}_k\mathbf{H}_{k,\rho}^{1/2}-\mathbf{C}_k^{1/2}\mathbf{S}\mathbf{C}_k^{1/2}\right\| = \mathcal{O}(\rho k^{2m}). \tag{64}$$

Furthermore, the eigenvalues of $\mathbf{C}_k^{1/2}\mathbf{S}\mathbf{C}_k^{1/2}$ are also the eigenvalues of $\widetilde{\Gamma}_p$, and thus are distinct, for small k/p and large k (lemma 3.6 and theorem 3.7). Hence, the eigenvectors of $\mathbf{C}_k^{1/2}\mathbf{S}_n\mathbf{C}_k^{1/2}$ satisfy

$$\left(\mathbb{E}\left\|\mathbf{v}_{j,n}^\rho-\mathbf{v}_j\right\|^2\right)^{1/2} = \mathcal{O}(n^{-1/2})+\mathcal{O}(\rho\ k^{2m}), \tag{65}$$

by applying lemma 3.1 of Bosq (1991) and relations (62) and (64).

Let's compare now the eigenfunctions $\{\widehat{\phi}^{\rho}_{j,n},\ j=1,\dots,q\}$ of $\widehat{\Gamma}_{n,p,\rho}$ with those of $\widetilde{\Gamma}_p$, say $\{\widetilde{\phi}_{j,p},\ j=1,\dots,q\}$. We have $\widehat{\phi}^{\rho}_{j,n}(t)=(\widehat{\mathbf{s}}^{\rho}_j)'\mathbf{B}_k(t),\ t\in[0,1]$, where $\widehat{\mathbf{s}}^{\rho}_j=\mathbf{H}^{1/2}_{k,\rho}\mathbf{v}^{\rho}_{j,n}$ and $\widetilde{\phi}_{j,p}(t)=(\mathbf{C}_k^{-1/2}\mathbf{v}_j)'\mathbf{B}_k(t),\ t\in[0,1]$. Thus, we can bound

$$
\begin{aligned}
\mathbb{E}\left\|\widehat{\phi}^{\rho}_{j,n}-\widetilde{\phi}_{j,p}\right\|^2_{L^2} &= \mathbb{E}\left\|\mathbf{H}^{1/2}_{k,\rho}\mathbf{v}^{\rho}_{j,n}-\mathbf{C}_k^{-1/2}\mathbf{v}_j\right\|_{\mathbf{C}_k} \\
&\leq \mathbb{E}\left\{(\mathbf{v}^{\rho}_{j,n})'\left(\mathbf{H}^{1/2}_{k,\rho}-\mathbf{C}_k^{-1/2}\right)\mathbf{C}_k\left(\mathbf{H}^{1/2}_{k,\rho}-\mathbf{C}_k^{-1/2}\right)\mathbf{v}^{\rho}_{j,n}\right\} \\
&\quad +\mathbb{E}\left\{\left(\mathbf{v}^{\rho}_{j,n}-\mathbf{v}_j\right)'\mathbf{C}_k^{-1/2}\mathbf{C}_k\mathbf{C}_k^{-1/2}\left(\mathbf{v}^{\rho}_{j,n}-\mathbf{v}_j\right)\right\} \\
&\quad +2\mathbb{E}\left\{\left(\mathbf{v}^{\rho}_{j,n}-\mathbf{v}_j\right)'\mathbf{C}_k^{-1/2}\mathbf{C}_k\left(\mathbf{H}^{1/2}_{k,\rho}-\mathbf{C}_k^{-1/2}\right)\mathbf{v}^{\rho}_{j,n}\right\}
\end{aligned}
$$

by writing

$$
\mathbf{H}^{1/2}_{k,\rho}\mathbf{v}^{\rho}_{j,n}-\mathbf{C}_k^{-1/2}\mathbf{v}_j=\left(\mathbf{H}^{1/2}_{k,\rho}-\mathbf{C}_k^{-1/2}\right)\mathbf{v}^{\rho}_{j,n}+\mathbf{C}_k^{-1/2}\left(\mathbf{v}^{\rho}_{j,n}-\mathbf{v}_j\right).
$$

Using relation (65) and lemmas 6.3 and 6.2, we finally get

$$
\left(\mathbb{E}\left\|\widehat{\phi}^{\rho}_{j,n}-\widetilde{\phi}_{j,p}\right\|^2_{L^2}\right)^{1/2} = \mathcal{O}(n^{-1/2})+\mathcal{O}(\rho\, k^{2m}). \tag{66}
$$

The end of the proof is obtained by applying again lemma 3.1 of Bosq (1991) to the operators $\widetilde{\Gamma}_p$ and Γ:

$$
\left\|\widetilde{\phi}_{j,p}-\phi_j\right\|_{L^2}=\mathcal{O}(p^{-1})+\mathcal{O}(k^{-m}),
$$

□

Acknowledgments : The author is indebted to Professor P.C. Besse, to Drs. C. Diack and D.B. Stephenson and an anonymous referee for critical comments which considerably improved this article in various ways.

References

Agarwal, G. and Studden, W (1980). Asymptotic integrated mean square error using least squares and bias minimizing splines. *The Annals of Statistics*, 8:1307–1325.

Besse, P. (1991). Approximation spline de l'analyse en composantes principales d'une variable aléatoire hilbertienne. *Annales de la Faculté des sciences de Toulouse*, 12:329–346.

Besse, P., Cardot, H., and Ferraty, F. (1997). Simultaneous nonparametric regressions of unbalanced longitudinal data. *Computational Statistics and Data Analysis*, 24:255–270.

Biritxinaga, E. (1987). *Estimation spline de la moyenne d'une fonction aléatoire.* Thèse de troisième cycle, Université de Pau, France.

Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. In Roussas, G., editor, *Nonparametric Functional Estimation and Related Topics*, pages 509–529. Nato, Asi series.

Boularan, J., Ferré, L., and Vieu, P. (1993). Growth curves: a two stage nonparametric approach. *Journal of Statistical Planning and Inference*, 38:327–350.

Burman, P. (1991). Regression function estimation from dependant observations. *Journal of Multivariate Analysis*, 36:263–279.

Cardot, H. (1997). *Contribution à l'estimation et à la prévision statistique de données fonctionnelles.* Thèse de troisième cycle, Université Paul Sabatier, Toulouse, France.

Cardot, H. and Diack, C. (1998). *Convergence en moyenne quadratique de l'estimateur de la régression par splines hybrides. C. R. Acad. Sci. Paris, Série I*, t. 326, 615-618.

Chatelin, F. (1983). *Spectral Approximation of Linear Operators.* Academic Press.

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

Diack, C. and Thomas-Agnan, C. (1997). A nonparametric test of non-convexity of the regression. To appear in *Journal of Nonparametric Statistics.*

De Boor, C. (1978). *A Practical Guide to Splines.* Springer Verlag.

Hart, J. and Wehrly, T. (1993). Consistency of cross-validation when data are curves. *Stochastic processes and their applications*, 45:351–361.

Kato, T. (1976). *Perturbation Theory for Linear Operator.* Springer-Verlag.

Kelly, C. and Rice, J. (1990). Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics*, 46:1071–1085.

Kneip, A. (1995). Nonparametric estimation of common regressors for similar curve data. *The Annals of Statistics*, 22:1386–1427.

Pezzulli, S. and Silverman, B. (1993). On smoothed principal components analysis. *Computational Statistics*, 8:1–16.

Ramsay, J. and Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, B*, 53:539–572. with discussion.

Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis.* Springer-Verlag.

Rice, J. and Silverman, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, B*, 53:233–243.

Schumaker, L. (1981). *Spline Functions: Basic Theory.* Wiley-Interscience.

Silverman, B. (1996). Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics*, 24:1–24.