FUNCTIONAL SLICED INVERSE REGRESSION ANALYSIS

L. FERRÉ AND A. F. YAO

UNIVERSITÉ TOULOUSE 2 AND UNIVERSITÉ TOULOUSE 3, FRANCE

ABSTRACT. Most of the usual multivariate methods have been extended to the context of functional data analysis. Our contribution concerns the study of Sliced Inverse Regression when the response variable is real but the regressor is a function. In a first part, we show how the relevant properties of the SIR model remain essentially the same in the functional context under suitable conditions. Unfortunately, and due to troubles with the inverse of the covariance operator, the estimation procedure used in the multivariate case cannot be directly transposed to the functional one. Then, we propose a way to overcome this difficulty and we show the consistency of the estimates of the parameters of the model.

1. INTRODUCTION

In numerous area (for instance in economics, chemistry, language recognition, medicine,...) measuring instruments allow the registration of curves. Then, functional data are now usual. This fact has required to focus on the extension of the classical statistical methods

Address for correspondence: Groupe de Recherche en Informatique et Mathématiques du Mirail, Université Toulouse 2, 5, allées A. Machado, F-31058 Toulouse.

for multivariate data to the domain of functional data. From a theoretical point of view, the works of Dauxois and Pousse (1976) lay the foundation for asymptotic studies of functional factorial method such as Principal Component Analysis, Canonical Correlation Analysis and Discriminant Analysis. Since then, many other methods have been studied from both theoretical and practical point of view and numerous corresponding estimates proposed (see e.g. Ramsay and Silverman (1997) for an extensive review). Particularly, the functional linear model where the response is scalar has been tackled for many ways. On the one hand, without really including the functional aspect of the data, (Massy, 1965) uses the regression on principal component, (Wold, 1975) a partial least squares method (P.L.S) and (Hoel et Kenneard, 1970) the ridge regression method. On the other hand, Hastie and Mallows (1993) and Marx and Eilers (1996) use spline (respectively, cubic spline and P-spline) and a penalized least square criterion.

More recently, Cardot and al. (1999) have proposed a method based on Functional Principal Component Analysis and Preda and Saporta (2000) investigate P.L.S for functional data. However those approaches are restricted to the linear model which might not be relevant in some case. For that reason, Ferraty and Vieu (2000) have proposed a functional non parametric model which, of course, exclude any model shape. Our purpose is somewhat a compromise between both by considering a semi-parametric functional model.

We will assume that the curves are squared integrable functions from an interval $[a, b]$ into $\mathbb{R}$. The space $L^2([a, b])$ of function of that type can be regarded with the usual inner product $(.,.)$, defined by $(f, g) = \int_a^b f(t)g(t)dt$. The associated norm will be denote by $||.||$. The model that we are dealing with, assume the existence of a function f, a real random variable ϵ and K linearly independent functions β_k, $k = 1, ..., K$, in $L^2([a, b])$ such that

$$Y = f(\int_a^b \beta_1(t)X(t)dt, ..., \int_a^b \beta_K(t)X(t)dt, \varepsilon). \tag{1.1}$$

The random variable ε is centered and independent of X. By using the inner product above, the model becomes:

$$Y = f((\beta_1, X), ..., (\beta_K, X), \varepsilon). \tag{1.2}$$

This model is clearly the functional version of the one explored by Li (1991) in the Sliced Inverse Regression (SIR) method. In classical SIR recall that the explanatory variable X belongs to a finite dimensional space, say $\mathbb{R}^p$. This method has motivated a large amount of works by considering various estimates of $\mathrm{var}(E(X|Y))$ (Hsing and Carrol, 1992 or Zhu and Fang,1996), by proposing several improvements (for instance SIR II Li, 1991 or Principal Hessian Direction, Li, 1992) or by proposing methods for determining the value of K (Schott(1994), Ferré (1998)).

Here the curse of dimensionality may not be a reason for our choice (Ferraty and Vieu estimates is insensitive to that trouble). However, we think that keeping some parametric aspect may be relevant for practical purpose. For instance, it might be helpful to know at which points the explanatory curves have the most (or the less) effect on the response.

As it is well-know, SIR is not directly a method for estimating the vectors $\beta_k' s$ but rather a way for estimating the K dimensional subspace they span, the Effective Direction Reduction space (EDR space). This space is also call sufficient reduction dimension space (See Cook, 1998), since the model we are dealing with states the existence of a subspace such that the projection of the curve onto that subspace contains as much information on the response variable as the initial curve.

Then a first question arise: can SIR be transposed to functional data? To answer this question, it is convenient to extend our purpose beyond functional data by considering Hilbert space random variables. Indeed, the space $L^2([a,b])$ with the inner product above defined is an Hilbert space and the properties of Hilbert random variables can be applied. Then we will deal with a couple of random variables (Y, X) where Y is the scalar response and X a random variable with values in a general Hilbert space, say H (X will be a continuous random process when

$H = L^2([a, b])$). While simplifying the proofs, it has the advantage of making our purpose more general even if the reader should keep in mind that our vectors are essentially curves.

This will be developed in Section 2. We will show that under a condition similar to the one required in classical SIR, the procedure to derive estimates of the EDR space still remains. Particularly, we show that this condition stands also for Hilbert elliptic random variables. Then, the principle of inverse regression still remains valid from the theoretical point of view. However, a main difference occurs when dealing with estimation, because the estimates are not available in practice since they depend on a non compact operator. Indeed, this method involves the inverse of covariance operator, and, it is well know that, under usual assumptions, this inverse is not bounded. This phenomenon is encountered in many situations in which data are curves. The main effect is that it produces an estimate of the covariance operator which is ill conditioned, and the aim of the above mentioned methods is to overcome this problem (Ridge regression (Hoel and Kennard, 1970), PCA regression (Massy, 1965), P.L.S in regression (H. Wold, 1975), smoothing in canonical analysis (Leurgans and al, 1993)).

In Section 3, we suggest to use a strategy similar to the one presented in Bosq (1991) for ARH(1) models and applied to the functional linear model by Cardot and al. (1999). Basically, it consists in projecting first the infinite dimensional data set, onto a series of finite dimensional subspaces derived from the spectral decomposition of the operator var(X).

Then, SIR is performed. As previously mentioned, several estimates of var$(E(X|Y))$ could have been proposed. In this paper, we retain the one obtained by slicing Y, i.e, the one corresponding to Li's Sliced Inverse Regression (1991). In that case, our main result is the consistency of the estimates of the functions β_k's.

Finally, Section 4 is devoted to applications. We consider some simulations in which the random variable X is a Brownian motion. These simulations confirm the capability of the method to yield correct estimations. Furthermore, since in this context, the method relies on the choice of three parameters, H the number of slices, k_n, the dimension of the a priori projecting space, and K the dimension of the EDR space, the simulations give an insight on the way those parameters can be selected in practice.

2. Functional Sliced Inverse Regression

As previously mentioned, the model we are dealing with, is the following one:

suppose there exist K independent vectors in H, say $\beta_1, ..., \beta_K$, a scalar random variable ε, independent of X and f a function from $\mathbb{R}^{K+1}$ into $\mathbb{R}$ such that:

$$Y = f((\beta_1, X), ..., (\beta_K, X), \varepsilon). \tag{2.1}$$

Without loss of generality and for the sake of simplicity, we will assume, all along the paper, that the variable X is centered. Concerning X, we recall that its expectation is defined (see, e.g. Fortet, 1995) as the only element in H, $E(X)$, satisfying, for all vector b in H, $(b, E(X)) = E((b, X))$. When X is a continuo random process, i.e $H = L^2([a, b])$, this definition coincide with the usual one (i.e for any s in $[a, b]$, $E(X)(s) = E(X(s)))$. We will denote by Γ the covariance operator of X. For the sake of simplicity, this operator will be assumed regular. If we recall that the tensor product between two vectors in H is the operator which associates to any z in H the vector $(x \otimes y)(z) = (x, z)y$, Γ is defined by $\Gamma = E(X \otimes X)$. Again, if $H = L^2([a, b])$, recall that we can define the covariance function such that for all s and t in $[a, b]$, $\Gamma(s, t) = E(X(s)X(t))$. In that case the definition is equivalent to the usual one when Γ is defined as the operator which associated to any function f in $L^2([a, b])$, the function Γf, such that for any s in $[a, b]$: $\Gamma f(s) = \int_a^b \Gamma(s, t)f(t)dt$.

We also need the conditional expectation of X given Y. It will be the random variable $E(X \mid Y)$ in H satisfying for any b in H, $(b,\ E(X \mid Y)) = E((b,\ X) \mid Y)$. Finally, we denote by Γ_e the covariance operator of $E(X \mid Y)$.

In the following, we will need another inner product on H. Indeed, since Γ is a positive self-adjoint linear operator, we will denote by $(.,.)_\Gamma$

the inner product on H defined by : for all x and y in H, $(x,y)_\Gamma = (\Gamma x, y)$. The associated tensor product will be denoted by $\otimes_\Gamma$.

We focus here on the estimation of the functional parameters $\beta_1, ..., \beta_K$. The key of the estimation method is derived from the following theorem which is the Hilbert version of Li's Theorem 2.1:

Theorem 2.1. *If the variable X satisfies : for all b in H, there exists a vector C in $\mathbb{R}^k$, such that, if*

$B' = ((\beta_1, X), ..., (\beta_K, X)),\ E((b,X)|B) = C'B,$

then, we have: $E\left(X \mid Y\right)$ belongs to the subspace spanned by

$\Gamma\beta_1, \Gamma\beta_2, \ldots, \Gamma\beta_K.$

The proof which is essentially the same as in the multivariate case is given in the Appendix. We also give in the Appendix, the proof of the corollary :

Corollary 2.2. *The operator Γ_e is degenerated in any direction $\Gamma-$orthogonal to the EDR Space.*

This shows that the operator $\Gamma^{-1}\Gamma_e$ is a finite rank operator whose image is contained into the EDR space. This operator is a compact operator and it is self-adjoint with respect to the inner product induced by Γ. Then, as in the multivariate case, a way to achieve the EDR space, is to perform the spectral decomposition of the operator $\Gamma^{-1}\Gamma_e$ and to retained the Γ-orthonormal eigenvectors $(b_1, ..., b_K)$ associated with the K non null eigenvalues $(\lambda_1, ..., \lambda_K)$.

What does the Inverse Regression do? In the functional context it is possible to present the Inverse Regression as a way to derive a Karhunen-Loève like decomposition of the conditional stochastic process $E(X|Y)$. This result is given in the following Theorem:

Theorem 2.3. *Let X be a random process and Y a real random variable. Under the same conditions as in Theorem 2.1, we have:*

$$E(X|Y) = \sum_{k=1}^{K} (u_k, E(X|Y))_{\Gamma^{-1}}\, u_k = \sum_{k=1}^{K} \phi_k u_k,$$

where $u_k = \Gamma b_k$, $\phi_k = (u_k, E(X|Y))_{\Gamma^{-1}}$, and this is the Karhunen-Loève decomposition of $E(X|Y)$ when H is considered with the inner product defined by Γ^{-1}.

Remark 1 The condition required for Theorem 2.1 is similar to the one used in the finite dimensional case. As it is well-known, it is verified when the explanatory variable has symmetric elliptical distribution. Then, the question is: does this property remains in the Hilbert case? The answer is positive. But before showing this, let us recall the definition of an elliptically distributed Hilbert variable. An Hilbert random variable with mean μ_X and covariance operator Γ is said to be elliptically distributed if its characteristic function satisfies for all t in H,

$E(exp(i(t,X))) = exp(i(t,\mu_X))h((t,\Phi t))$ where h is a derivable, real value function and Φ is an operator such that $\Gamma = h'(0)\Phi$. This definition is only a partial generalization of the multivariate definition since some, but not all, the elliptical multivariate variables have such an expression for their characteristic function. Nevertheless, the symmetric distributions and the normal distribution are included in that definition. Particularly, it implies that any finite dimensional projection of X has an elliptical distribution. This fact allows us to propose a proof which is essentially the same as in the multivariate case. To save place, the proof is omitted here but it is available on request.

Remark 2 As discussed in Li (1991) and Ferré (1998), SIR is an appealing method for explanatory analysis by providing a way to derive the direction in which X is the best explained by Y. This feature remains in the Hilbert context. Indeed, if we look for the first direction, solution of that problem, we are looking for an element β_1 of H which satisfies $\beta_1 = argsup_\beta \frac{Var(E(\beta,X)|Y)}{Var((\beta,X))}$. But since $Var(E(\beta,X)|Y) = (\Gamma_e\beta,\beta)$ and $Var((\beta,X)) = (\Gamma\beta,\beta)$, we have

$$\beta_1 = argsup_\beta \frac{(\Gamma_e\beta,\beta)}{(\Gamma\beta,\beta)} = argsup_\beta \frac{(\Gamma^{-1}\Gamma_e\beta,\beta)_\Gamma}{(\beta,\beta)_\Gamma}$$

which shows that β_1 is the Γ-normed eigenvector associated with the largest eigenvalue of the operator $\Gamma^{-1}\Gamma_e$. The following functions are obtained by a similar procedure under orthogonal constraint.

3. Estimation of the EDR Space

Let (X_i, Y_i) be a sample of size n of i.i.d. variables with same distribution as (X, Y). As in Li(1991), we consider a S-partition of the domain of Y, $(I_s)_{s\in\{1,...,S\}}$. We denote $p_s = P(Y \in I_s)$, $\mu_s = E(X|Y \in I_s)$ and $\tilde{\Gamma}_e = \sum_{s=1}^{S} p_s \mu_s \otimes \mu_s$. For the covariance operators, we consider the usual estimates $\Gamma_n = \frac{1}{n}\sum_i X_i \otimes X_i - \bar{X} \otimes \bar{X}$ and $\Gamma_{en} = \sum_{s=1}^{S} \hat{p}_s \hat{\mu}_s \otimes \hat{\mu}_s - \bar{X} \otimes \bar{X}$, where $\bar{X} = \frac{1}{n}\sum_i X_i$ and $\hat{\mu}_s = \frac{1}{n_s}\sum_i X_i \, 1_{Y_i \in I_s}$.

All those estimates are convergent. It is sufficient to show that to apply the weak law of large number for i.i.d Hilbert variables (see e.g., Ito and Nisio(1968)). Then, from the Corollary 2.2, the estimation procedure seems to be a straightforward replication of the multivariate case. However, while it exits, the inverse of Γ is not bounded. Indeed, as noticed by Bosq (1991), the operator Γ has finite trace so that the sequence of its eigenvalue has limit 0. Then, we suggest here to proceed as Bosq (1991) by considering, instead of Γ, a sequence of finite rank operators, with bounded inverse and converging to Γ.

Then, let (k_n) be a non convergent sequence of number. Let $(\alpha_j)_{j=1}^{\infty}$ be the decreasing sequence of eigenvalues of the positive definite and compact operator Γ, with associated sequence of eigenvectors $(V_j)_{j=1}^{\infty}$. For each k, let Π_k be the projector onto the subspace spanned by the vectors $(V_1, ..., V_k)$, i.e. $\Pi_k = \sum_{j=1}^{k} V_j \otimes V_j$. Now define $\Gamma_{k_n} = \Pi_{k_n}\Gamma\Pi_{k_n}$. Thus, Γ_{k_n} is a sequence of compact operator, converging to Γ as n tends to ∞, and with bounded generalized-inverse since Γ_{k_n} has finite rank.

For practical purpose, we deal with estimate achieved by applying the above procedure to the empirical version of Γ. Let $(\hat{\alpha}_j)_{j=1}^{\infty}$ be the decreasing sequence of eigenvalues of Γ_n, with associated sequence of eigenvectors $(\hat{V}_j)_{j=1}^{\infty}$. For each k, let $\hat{\Pi}_k$ be the projector onto the subspace spanned by the vectors $(\hat{V}_1, ..., \hat{V}_k)$, i. e. $\hat{\Pi}_k = \sum_{j=1}^{k} \hat{V}_j \otimes \hat{V}_j$. Let $\Gamma_{n,k_n} = \hat{\Pi}_{k_n} \Gamma_n \hat{\Pi}_{k_n}$ and define $(\hat{\beta}_{k,k_n})_{k=1,...,K}$ to be the sequence of Γ_{n,k_n}-orthonormal eigenvectors associated with the K largest eigenvalues of $\Gamma_{n,k_n}^{+} \Gamma_{en}$ denoted by $\widehat{\lambda}_{1,k_n}, ..., \widehat{\lambda}_{K,k_n}$. Recall that this operator has rank K with probability one.

In order to derive the asymptotic properties, we need some technical assumptions. Let $(a_j)_{j=1}^{\infty}$ be the sequence defined by $a_1 = \frac{2\sqrt{2}}{\alpha_1 - \alpha_2}$ and $a_j = \frac{2\sqrt{2}}{\min\left(\alpha_{i-1} - \alpha_i, \alpha_i - \alpha_{i+1}\right)}$.

We assume that

(H-1) the eigenvalues of Γ are distinct,

(H-2) the variable X satisfies $E\left(\|X\|^4\right) < +\infty$,

the sequences a_j and α_{k_n} are linked to n by the following relations:

(H-3) $\lim_{n\to\infty} \frac{1}{\sqrt{n\alpha_{k_n}}} \sum_{j=1}^{k_n} a_j = 0$,

(H-4) $\lim_{n\to\infty} \frac{1}{\sqrt{n}\,\alpha_{k_n}^2} = 0$,

(H-6) the eigenvalues of $\Gamma^{-1}\Gamma_e$ are distinct and for, all i, $\lambda_i > \lambda_{i+1}$,

Theorem 3.1. *Under assumptions (H-1), (H-2), (H-3) and (H-4), we have, for all $k \in 1, ..., K$:*

$||\widehat{\beta}_{k,k_n} - b_k|| \longrightarrow_p 0$ *as* $n \longrightarrow 0$.

The proof of this theorem is given in Appendix 5.3.

Remark 3 This result only states the convergence of the estimators $\widehat{\beta}_{k,k_n}$ without specifying the rate of convergence. Let us just mention that our approach can not lead to achieve the $\sqrt{n}-$convergence. Indeed, as our assumptions suggest it, the rate of convergence will be penalized by the one of the sequence (α_{k_n}) .

Remark 4 The extension of the inverse regression method to the functional case could have been tackled from different ways. In Ferré and Yao (2000), we use the fact that $\Gamma^{-1/2}\Gamma_e\Gamma^{-1/2}$ is a finite rank operator whose eigenvectors are the one of $\Gamma^{1/2}\Gamma_e^{+}\Gamma^{1/2}$and we show the consistency of the estimators of the functions β. Another study in progress, concerns the introducing of smoothing parameters. Similarly as Leurgans, Moyeed and Silverman (1993) for the canonical correlation analysis, we suggest to add a roughness penalty term and estimate the β's by maximizing the criterion: $\frac{(\Gamma^{-1}\Gamma_e\beta,\beta)_\Gamma}{(\beta,\beta)_\Gamma+\nu_1||D^2\beta||^2}$; where D^2 is the second order derivative operator and ν_1 is a scalar smoothing parameter.

Remark 5 The previous result insure the efficiency of the method for estimating the functions β's. However, its implementation requires the selection of several parameters. The number of slices, S, the dimension of the a priori projection, k_n, and the dimension of the E.D.R space, K. We will not study this problem in details here. However, similarly as in the multivariate case, the selection of those parameters can be obtained through the study of a criterion measuring the quality of the estimation of the β's. The same arguments as in Ferré (1998), can be used in favor of: $\frac{1}{K}E(tr(\widehat{\Pi}_K^{\Gamma_n}\Pi_K^{\Gamma}))$ where Π_K^{Γ} (resp. $\widehat{\Pi}_K^{\Gamma_n}$) is the projector $\Gamma-$orthogonal (resp. Γ_n-orthogonal) onto the subspace spanned by the vectors $(\beta_1, \dots, \beta_K)$ (resp. $(\widehat{\beta}_1, \dots, \widehat{\beta}_K)$). But no direct use of the criterion is possible in practice, because, since now, no expansion of that criterion has been obtained in that case because perturbation theory do not apply. Note that this last point also prevent the use of a statistical test based on the remaining eigenvalues as in Li(1991). Then, a solution could be the estimation of $\frac{1}{K}E(tr(\widehat{\Pi}_K^{\Gamma_n}\Pi_K^{\Gamma}))$ by, for instance, bootstrapping.

4. Simulations

Let $H = L^2([0,1])$. We evaluate the results of functional SIR by simulated data.

The model used to generate the data is:

$$Y = \exp((\beta_1, X)) + \exp(|(\beta_2, X)|) + \epsilon, \tag{4.1}$$

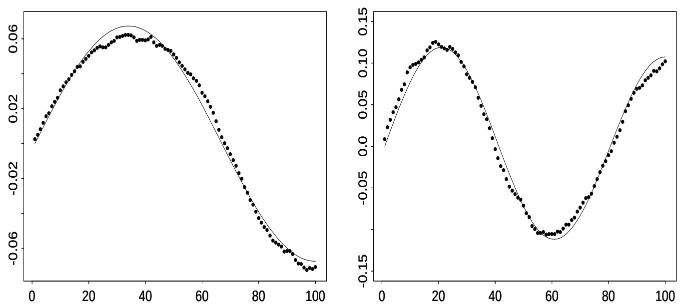
where $(X_t)_{t\in[0,1]}$ is a standard Brownian motion and $\epsilon \sim N(0, 0.1)$. The sample size is $n = 500$, and the curves are observed at $p = 100$ times at the same date. In order to simplify the comparisons, we have $\Gamma-$normed the functions β_i's.

According to Remark 5, the choice of k_n, K and S is derived from the criterion $tr(\widehat{\Pi}_K^{\Gamma_n}\Pi_K^{\Gamma})/K$. For the simulations, this criterion is computable and the selection of the above parameters is obtained by maximizing it. This will guide our choice all along the simulations. As a general observation, we note that while the method is sensitive to the choice of k_n (which is 3 or 4 according to the model here), S can be choosen with more flexibility.

The eigenfunctions of the operator Γ are (see e.g Ash and Gardner, 1975): $V_i = \sqrt{2}\sin((2i-1)\pi t/2)$, $i = 1, 2, \ldots$ For that reason, we first take the functions β_1 and β_2 to be elements of the eigensubspace of Γ : $\beta_1(t) = \sin(3\pi t/2)$ and $\beta_2(t) = \sin(5\pi t/2)$.

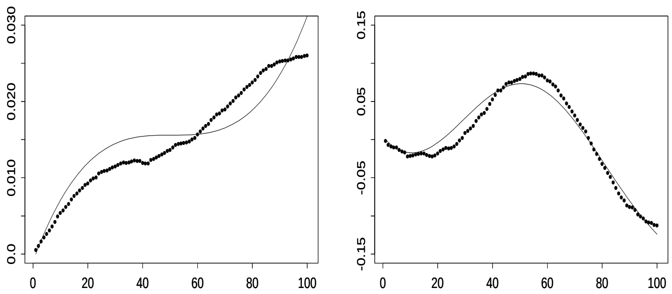
Figure 4.1 gives the plots of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ together with β_1 and β_2. The estimates in that case coincides almost perfectly with the generated functions. Not surprising is that result since this case corresponds to the ideal situation where the E.D.R space is included into the a priori projection subspace. Consistant is the result that returns $k_n = 3$ as solution because, the functions β's are respectively the second and third eigenfunctions of the operator Γ.

The second example takes its motivation for the choice of the β_i's in the fact that, for all integer k, the space spanned by $(V_1, \ldots, V_k)$ is a space of functions null at 0. Then, we constrain $\beta_i(0)$ to be 0 for $i = 1, 2$ and we take $\beta_1(t) = (2t-1)^3 + 1, \quad \beta_2(t) = \cos(\pi(2t-1)) + 1.$ and these functions are plotted in Figure 4.2.



(a) The first direction (b) The second direction

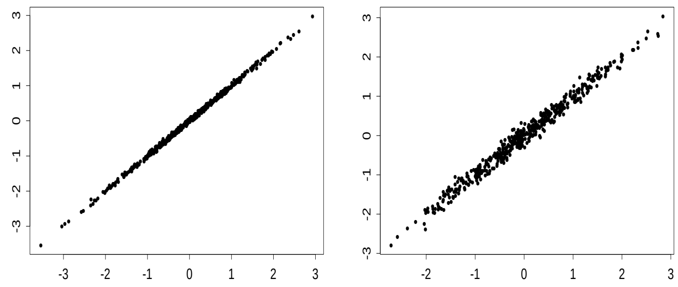
FIGURE 4.1. Superposition of the plots for $\beta_1(t) = \sin(3\pi t/2)$, and $\beta_2(t) = \sin(5\pi t/2)$



(a) The first direction (b) The second direction

FIGURE 4.2. The plot of the $\beta' s$ (in full line) and the $\widehat{\beta}' s$ (in dotted line) when $\beta_i(0) = 0$.

In Figure 4.2, we give the plot of these functions together with their estimates. For the polynomial function which is the first eigenfunction of the operator we are dealing with, the estimation is quite satisfactory even if some difficulties appear at the inflexion point. On the other hand, the second eigenfunction is accurately estimated. Whatever it is, what we are interested in, is to obtain projections of the data the closest as possible to the "true" projections. In order to verify this



(a) The first indexes (b) The second indexes

FIGURE 4.3. The plots of $(\widehat{\beta}_i, X)$ versus (β_i, X), for $i = 1, 2$ when $\beta_i(0) = 0$.

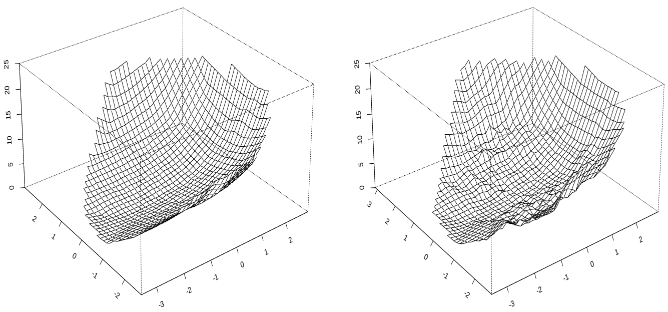
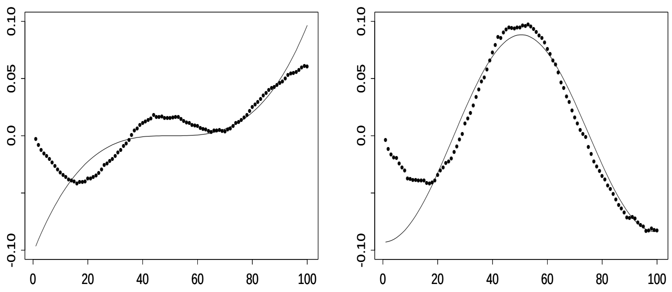


FIGURE 4.4. The plot of Y versus $((\beta_1, X), (\beta_2, X))$ (on the right side) and Y versus $(\widehat{\beta}_1, X), (\widehat{\beta}_2, X)$ (on the left side).

point, we plot, in Figure 4.3, the indexes $(\widehat{\beta}_i, X)$ versus (β_i, X), for $i = 1, 2$. We can see that, while the first eigenfunction is not so well estimated, the projection of X onto it, is particularly relevant. For the second eigenfunction, the estimated index perform with less accuracy.

Finally, we give, in Figure 4.4, the graphical display of the Y versus the whole set of indexes. Not surprising is the result that shows that the estimated response curves is really close to the simulated one.

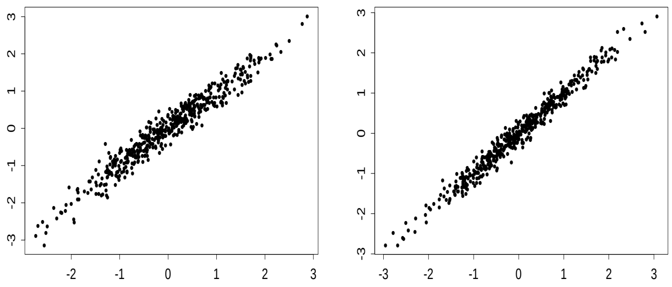


(a) The first direction (b) The second direction

FIGURE 4.5. The plots of the β's (in full line) and the $\widehat{\beta}$'s (in dotted line) when $\beta_i(0)$ is different from zero.

In this next simulations, we remove the constraint $\beta_i(0) = 0$ and we take $\beta_1(t) = (2t-1)^3 \quad \beta_2(t) = \cos(\pi(2t-1))$. Of course, we cannot expect now a relevant estimation of the eigenfunction in a neighborhood of zero and this is confirmed by Figure 4.5. However, as we have already mentioned it, the projection is more interesting than the function itself.

Indeed, Figure 4.6 reveals that the defect concerning the eigenfunction does not affect the estimation of the indexes. This feature is confirmed by scatter plots of Figure 4.7 and response curves of Figure 4.8. Particularly, the scatter plots of Figure 4.7 are graphical displays of the marginal relationship; one can see that the structures underlined by the scatter plots, could be properly fitted by the initial functions say, $\exp x$ and $\exp |x|$.



(a) The first indexes (b) The second indexes

FIGURE 4.6. The plots of the index $(\widehat{\beta}_i, X)$ versus (β_i, X), for $i = 1,\, 2$ when $\beta_i(0) \neq 0$.

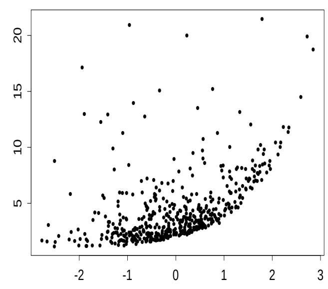
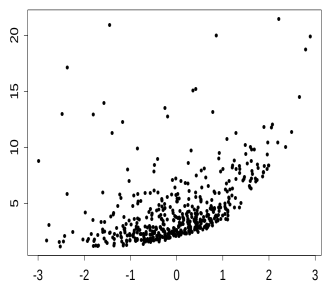
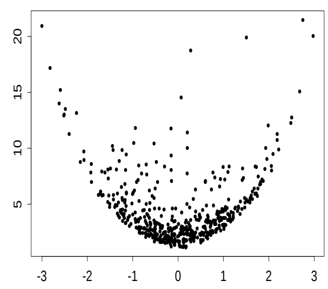
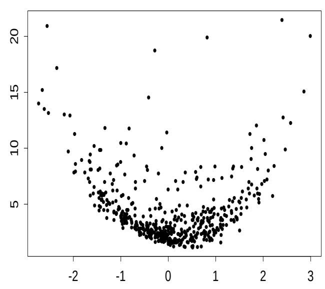
	
The plot of Y versus (β_1, X)	The plot of Y versus $(\widehat{\beta}_1, X)$
	
The plot of Y versus (β_2, X)	The plot of Y versus $(\widehat{\beta}_2, X)$

FIGURE 4.7. The marginal plots of Y versus $(\beta_i,\, X)$ for $i = 1, 2$

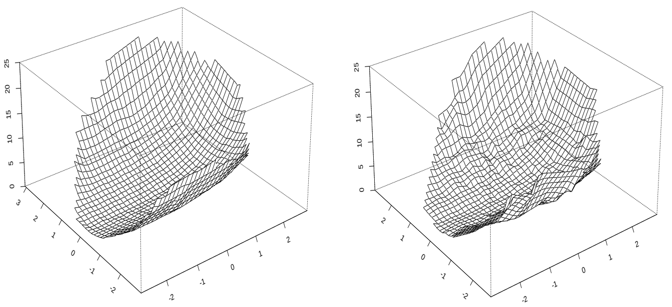


FIGURE 4.8. The plot of Y versus the set of (β_i, X) (on the right side) and Y versus the set of $(\widehat{\beta}_i, X)$ (on the left side) when $\beta_i(0) \neq 0$.

5. Appendix

5.1. **Proof of Theorem 2.1 and Corollary 2.2.**

The proof of Theorem 2.1 is very similar to the one in finite dimension. Let b be in H such that for all k $(b,\, \beta_k)_\Gamma = 0$.We want to show that $(b, E(X \mid Y)) = 0$. We first have, from the definition of the conditional expectation,

$$(b,\, E(X \mid Y)) = E((b,\, X) \mid Y) = E(E((b, X)|(\beta_1, X), \ldots, (\beta_K, X))|Y)$$

From, the condition of Theorem 2.1, we have:

$$E(E((b,\, X)\, |(\beta_1,\, X)\,,\, \ldots\,,\, (\beta_K,\, X))|\, Y){=}E(\textstyle\sum_{k=1}^{K} C_k\, (\beta_k,\, X)\, (b,\, X)).$$

But, for all k, $(\beta_k,\, X)\, (b,\, X) = (b,\, (\beta_k,\, X)\, X) = (b,\, X \otimes X \beta_k)$ and then $E(\sum_{k=1}^{K} C_k\, (\beta_k,\, X)\, (b,\, X)) = \sum_{k=1}^{K} C_k\, (b,\, E(X \otimes X)\beta_k) = 0$ which achieve the proof of Theorem 2.1.

Proof of Corollary 2.2 is straightforward and is also similar to the one in finite dimension: let b be a vector of H, orthogonal to the sub-space spanned by $(\Gamma\beta_1, \ldots \Gamma\beta_1)$ then we have $(b, E(X \mid Y)) = 0$. Then,

$E\left((b,\, E\,(X \mid Y))^2\right) = 0$ implies that $E\left((b,\, E\,(X \mid Y))\; b, E\,(X \mid Y)\right) =$ $E((b, (b, E\,(X \mid Y))\, E\,(X \mid Y)) = E\left((b,\, E(X \mid Y) \otimes E(X \mid Y)\, b)\right) =$ $(b,\, \Gamma_e b) = 0.$

5.2. **Proof of Theorem 2.3.**

Let $u_k = \Gamma \beta_k$. We have shown that $E(X|Y)$ belongs to the subspace spanned by the vectors $u_1, ..., u_K$. These vectors are Γ^{-1}-orthonormal, then, we have $E(X|Y) = \Pi_K^{\Gamma^{-1}}(E(X|Y))$ where $\Pi_K^{\Gamma^{-1}}$ is the projector Γ^{-1}-orthogonal onto the subspace spanned by the vectors $u_1, ..., u_K$. Then, $E(X|Y) = \sum_{k=1}^{K} u_k \otimes_{\Gamma^{-1}} u_k E(X|Y) = \sum_{k=1}^{K} (u_k, E(X|Y))_{\Gamma^{-1}} u_k =$ $\sum_{k=1}^{K} \phi_k u_k$, where $\phi_k = (u_k, E(X|Y))_{\Gamma^{-1}}$. Furthermore, for $k = 1, ...K$, u_k is the Γ-normed eigenfunction associated with the kth largest eigenvalue of the operator $\Gamma_e \Gamma^{-1}$, $E(\phi_k^2)$ and the proof is complete.

5.3. **Proof of Theorem 3.1.**

For the proof, we need the following lemma:

Lemma 5.1. *Let G and E to be two linear self-adjoint and compact operators defined in a Hilbert space. Let $(g_i)_{i \in \mathbf{N}*}$and $(e_i)_{i \in \mathbf{N}^*}$ the respective decreasing sequences of eigenvalues; let $(U_i)_{i \in \mathbf{N}*}$ the sequence of orthonormal eigenvectors associated with $(g_i)_{i \in \mathbf{N}*}$ and $(E_i)_{i \in \mathbf{N}*}$ the sequence of orthonormal eigenvectors associated with $(e_i)_{i \in \mathbf{N}^*}$. Then, for all $i \in \mathbf{N}^*$,*

(*i*) $|g_i - e_i| \leq ||G - E||_\infty$

(ii) if $g_i > g_{i+1}$,

$2\,||G-E||_\infty \geq (\sum_{k=1}^{\infty}(g_k - g_i)^2 (E_i, U_k)^2)^{1/2}$

(iii) $||U_i - E_i|| \leq a_i\,||G-E||_\infty$ *where*

$$a_i = \begin{cases} \frac{2\sqrt{2}}{g_1 - g_2} & \text{if } i = 1 \\ \frac{2\sqrt{2}}{\min(g_{i-1}-g_i\,,\,g_i - g_{i+1})} & \text{otherwise} \end{cases}.$$

The proof of this lemma can be found in Bosq (1991) in the particular case where $G = \Gamma$ and $E = \Gamma_n$.

Now, for all $i = 1, ..., K$, let $\eta_i = \Gamma^{1/2} b_i$, and $\widehat{\eta}_{i,k_n} = (\Gamma_{n,k_n})^{1/2}\,\widehat{\beta}_{i,kn}$. From their definitions, we know that $b_i = (1/\lambda_i)\Gamma^{-1}\Gamma_e b_i$ and $\widehat{\beta}_{i,k_n} = (1/\widehat{\lambda}_{i,k_n})\Gamma^{+}_{n,k_n}\Gamma_{en}\widehat{\beta}_{i,k_n}$.

Then, we have the following inequality:

$$\begin{aligned} ||\widehat{\beta}_{i,k_n} - b_i|| &= \left\| \frac{1}{\widehat{\lambda}_{i,k_n}} \Gamma^{+}_{n,k_n}\Gamma_{en}(\Gamma^{+}_{n,k_n})^{1/2}\widehat{\eta}_{i,k_n} - \frac{1}{\lambda_i}\Gamma^{-1}\Gamma_e\Gamma^{-1/2}\eta_i \right\| \\ &\leq \left\| \frac{1}{\widehat{\lambda}_{i,k_n}} \Gamma^{+}_{n,k_n}\Gamma_{en}(\Gamma^{+}_{n,k_n})^{1/2} - \frac{1}{\lambda_i}\Gamma^{-1}\Gamma_e\Gamma^{-1/2} \right\|_\infty \\ &\quad + \frac{||\Gamma^{-1}\Gamma_e\Gamma^{-1/2}||_\infty}{\lambda_i}||\widehat{\eta}_{i,k_n} - \eta_i||. \\ &\leq \frac{1}{\widehat{\lambda}_{i,k_n}} ||\Gamma^{+}_{n,k_n}\Gamma_{en}(\Gamma^{+}_{n,k_n})^{1/2} - \Gamma^{-1}\Gamma_e\Gamma^{-1/2}||_\infty \\ &\quad + ||\Gamma^{-1}\Gamma_e\Gamma^{-1/2}||_\infty \left(|\frac{1}{\widehat{\lambda}_{i,k_n}} - \frac{1}{\lambda_i}| + \frac{1}{\lambda_i}||\widehat{\eta}_{i,k_n} - \eta_i|| \right). \end{aligned}$$

Noting that $\frac{1}{\widehat{\lambda}_{i,k_n}} \leq \frac{1}{\lambda_i} + |\frac{1}{\widehat{\lambda}_{i,k_n}} - \frac{1}{\lambda_i}|$ and applying Lemma 5.1 with $G = (\Gamma^+_{n,k_n})^{1/2}\Gamma_{en}(\Gamma^+_{n,k_n})^{1/2}$ and $E = \Gamma^{-1/2}\Gamma_e\Gamma^{-1/2}$ we get:

$$||\widehat{\beta}_{i,k_n} - b_i|| \leq C_n(D_n^{'} + \frac{1}{\lambda_i}) + ||\Gamma^{-1}\Gamma_e\Gamma^{-1/2}||_\infty(D_n^{'} + \frac{\tilde{a}_i D_n}{\lambda_i}). \tag{5.1}$$

where $C_n = ||\Gamma^+_{n,k_n}\Gamma_{en}(\Gamma^+_{n,k_n})^{1/2} - \Gamma^{-1}\Gamma_e\Gamma^{-1/2}||_\infty$, $D_n^{'} = |\frac{1}{\widehat{\lambda}_{i,k_n}} - \frac{1}{\lambda_i}|$ $D_n = ||(\Gamma^+_{n,k_n})^{1/2}\Gamma_{en}(\Gamma^+_{n,k_n})^{1/2} - \Gamma^{-1/2}\Gamma_e\Gamma^{-1/2}||_\infty$ and $\tilde{a}_i = \begin{cases} \frac{2\sqrt{2}}{\lambda_1-\lambda_2} & \text{if } i = 1 \\ \frac{2\sqrt{2}}{\min(\lambda_{i-1}-\lambda_i\,,\,\lambda_i-\lambda_{i+1})} & \text{otherwise} \end{cases}$.

Then, from inequality (5.1), it suffices to show that $C_n \longrightarrow_p 0$, $D_n \longrightarrow_p 0$ and $D_n' \longrightarrow_p 0$ to end the proof of the theorem.

Now, we show that:

$$C_n \longrightarrow_p 0. \tag{5.2}$$

For that, we get the following inequality:

$$\begin{aligned} C_n \leq\; & ||\Gamma^+_{n,k_n}\Gamma_e(\Gamma^+_{n,k_n})^{1/2} - \Gamma^+_{k_n}\Gamma_e(\Gamma^+_{k_n})^{1/2}||_\infty \\ & + ||\Gamma^+_{k_n}\Gamma_e(\Gamma^+_{k_n})^{1/2} - \Gamma^{-1}\Gamma_e\Gamma^{-1/2}||_\infty \\ & + \frac{||\Gamma_{en} - \Gamma_e||_\infty}{\widehat{\alpha}_{k_n}^{3/2}}, \end{aligned} \tag{5.3}$$

We first prove that:

$$||\Gamma^+_{k_n}\Gamma_e(\Gamma^+_{k_n})^{1/2} - \Gamma^{-1}\Gamma_e\Gamma^{-1/2}||_\infty \longrightarrow 0. \tag{5.4}$$

Indeed, note that,

$$\Gamma_{k_n}^{+}\Gamma_e(\Gamma_{k_n}^{+})^{1/2} = \Pi_{k_n}\Gamma^{-1}\Gamma_e\Gamma^{-1/2}\Pi_{k_n},$$

and let $\Pi_{k_n}^{\perp}$ be the projector onto the eigensubspace associated with the eigenvalues smaller than α_{k_n}. Then, we have:

$$\begin{aligned} ||\Gamma_{k_n}^{+}\Gamma_e(\Gamma_{k_n}^{+})^{1/2} - \Gamma^{-1}\Gamma_e\Gamma^{-1/2}||_\infty \;\leq\; & ||\Pi_{k_n}^{\perp}\Gamma^{-1}\Gamma_e\Gamma^{-1/2}||_\infty \\ & +||\Gamma^{-1}\Gamma_e\Gamma^{-1/2}\Pi_{k_n}^{\perp}||_\infty. \end{aligned}$$

Now, $||\Pi_{k_n}^{\perp}\Gamma^{-1}\Gamma_e\Gamma^{-1/2}||_\infty \leq \sum_{i=1}^{K}\lambda_i||\Pi_{k_n}^{\perp}b_i||$, and $||\Pi_{k_n}^{\perp}\Gamma^{-1}\Gamma_e\Gamma^{-1/2}||_\infty$ tend to zero as soon as $||\Pi_{k_n}^{\perp}b_i|| \longrightarrow 0$ for $i = 1,..,K$; this is realised since $\Pi_{k_n}^{\perp}$ converges weakly to 0. Similar arguments lead to the convergence of $||\Gamma^{-1}\Gamma_e\Gamma^{-1/2}\Pi_{k_n}^{\perp}||_\infty$ to zero and we get result (5.4) by the way.

Secondly, we show that:

$$||\Gamma_{n,k_n}^{+}\Gamma_e(\Gamma_{n,k_n}^{+})^{1/2} - \Gamma_{k_n}^{+}\Gamma_e(\Gamma_{k_n}^{+})^{1/2}||_\infty \longrightarrow_p 0. \tag{5.5}$$

For that, we use the following inequality:

$$\begin{aligned} &||\Gamma_{n,k_n}^{+}\Gamma_e(\Gamma_{n,k_n}^{+})^{1/2} - \Gamma_{k_n}^{+}\Gamma_e(\Gamma_{k_n}^{+})^{1/2}||_\infty \\ &\qquad\leq ||\left(\Gamma_{n,k_n}^{+} - \Gamma_{k_n}^{+}\right)\Gamma_e(\Gamma_{n,k_n}^{+})^{1/2}||_\infty \\ &\qquad +||\Gamma_{k_n}^{+}\Gamma_e\left((\Gamma_{n,k_n}^{+})^{1/2} - (\Gamma_{k_n}^{+})^{1/2}\right)||_\infty. \end{aligned}$$

Since $||\Gamma_e(\Gamma_{k_n}^+)^{1/2}||_\infty \leq ||\Gamma_e\Gamma^{-1/2}||_\infty$ and $||\Gamma_{k_n}^+\Gamma_e||_\infty \leq ||\Gamma^{-1}\Gamma_e||_\infty$, it leads to:

$$(5.6)\; ||\Gamma_{n,k_n}^+\Gamma_e(\Gamma_{n,k_n}^+)^{1/2} - \Gamma_{k_n}^+\Gamma_e(\Gamma_{k_n}^+)^{1/2}||_\infty \leq E_n||\Gamma^{-1}\Gamma_e||_\infty$$
$$+E_n^{'}(||\Gamma_e||_\infty E_n + ||\Gamma_e\Gamma^{-1/2}||_\infty)$$

where $E_n = ||(\Gamma_{n,k_n}^+)^{1/2} - (\Gamma_{k_n}^+)^{1/2}||_\infty$ and $E_n^{'} = ||\Gamma_{n,k_n}^+ - \Gamma_{k_n}^+||_\infty$, and it suffices to show that $E_n \longrightarrow_p 0$ and $E_n^{'} \longrightarrow_p 0$ to conclude for (5.5).

To establish the former result, we start with the inequality:

$$(5.7) \qquad E_n = ||(\Gamma_{k_n}^+)^{1/2} - (\Gamma_{n,k_n}^+)^{1/2}||_\infty \;\; \leq \;\; A_n + B_n$$

where $A_n = ||\sum_{j=1}^{k_n}\frac{1}{\sqrt{\alpha_j}}(V_j \otimes V_j - \widehat{V}_j \otimes \widehat{V}_j)||_\infty$ and
$B_n = ||\sum_{j=1}^{k_n}(\frac{1}{\sqrt{\alpha_j}} - \frac{1}{\sqrt{\widehat{\alpha}_j}})\widehat{V}_j \otimes \widehat{V}_j||_\infty$.

But, on the one hand, we have:

$$\begin{aligned} A_n &\leq \sum_{j=1}^{k_n}\frac{1}{\sqrt{\alpha_j}}||(V_j \otimes V_j - \widehat{V}_j \otimes \widehat{V}_j)||_\infty \\ &\leq 2\sum_{j=1}^{k_n}\frac{1}{\sqrt{\alpha_j}}||V_j - \widehat{V}_j||, \end{aligned}$$

and then, by Lemma 5.1 ,

$$A_n \leq 2\frac{1}{\sqrt{\alpha_{k_n}}}||\Gamma - \Gamma_n||_\infty \sum_{j=1}^{k_n} a_j\, .$$

Now, for a given $\epsilon > 0$, we have

$$
\begin{aligned}
P(A_n > \epsilon) &= P\left(\frac{||\Gamma - \Gamma_n||_\infty}{\sqrt{\alpha_{k_n}}} \sum_{j=1}^{k_n} a_j > \epsilon\right) \\
&= P\left(||\Gamma - \Gamma_n||_\infty > (\sum_{j=1}^{k_n} a_j)^{-1} \epsilon \sqrt{\alpha_{k_n}}\right)
\end{aligned}
$$

which, by Tchebychev inequality, leads to:

$$P(A_n > \epsilon) \leq E(||X||^4)(\sum_{j=1}^{k_n} a_j)^2 / \epsilon^2 n \alpha_{k_n}$$

which shows that, by assumption (H-3),

$$A_n \longrightarrow_p 0. \tag{5.8}$$

On the other hand, we get

$$
\begin{aligned}
B_n^2 &= \sup_{||x||=1} \left\| \sum_{j=1}^{k_n} (\frac{1}{\sqrt{\alpha_j}} - \frac{1}{\sqrt{\widehat{\alpha}_j}}) \widehat{V}_j \otimes \widehat{V}_j\,(x) \right\|^2 \\
&\leq \sup_{||x||=1} \sum_{j=1}^{k_n} \frac{\left(\sqrt{\alpha_j} - \sqrt{\widehat{\alpha}_j}\right)^2}{\alpha_j\, \widehat{\alpha}_j} \left\| \widehat{V}_j \otimes \widehat{V}_j\,(x) \right\|^2 \\
&\leq \frac{\|\Gamma - \Gamma_n\|_\infty}{\alpha_{k_n}\, \widehat{\alpha}_{k_n}} \sup_{||x||=1} \sum_{j=1}^{k_n} \left\| \widehat{V}_j \otimes \widehat{V}_j\,(x) \right\|^2 \\
&\leq \frac{\|\Gamma - \Gamma_n\|_\infty}{\alpha_{k_n}\, \widehat{\alpha}_{k_n}} \left\| \widehat{\Pi}_{k_n} \right\|_\infty^2 \\
&= \frac{\|\Gamma - \Gamma_n\|_\infty}{\alpha_{k_n}\, \widehat{\alpha}_{k_n}}.
\end{aligned}
$$

Then, for all $\epsilon > 0$ we have,

$$
\begin{aligned}
P(B_n > \epsilon) &= P(\frac{\|\Gamma - \Gamma_n\|_\infty^{1/2}}{\sqrt{\alpha_{k_n}}\sqrt{\widehat{\alpha}_{k_n}}} > \epsilon) \\
&= P(\frac{\|\Gamma - \Gamma_n\|_\infty}{\alpha_{k_n}\,\widehat{\alpha}_{k_n}} > \epsilon^2, \widehat{\alpha}_{k_n} > \frac{\alpha_{k_n}}{2}) \\
&\quad + P(\frac{\|\Gamma - \Gamma_n\|_\infty}{\alpha_{k_n}\,\widehat{\alpha}_{k_n}} > \epsilon^2, \widehat{\alpha}_{k_n} \leq \frac{\alpha_{k_n}}{2}) \\
&\leq P(\|\Gamma - \Gamma_n\|_\infty > \epsilon^2 \alpha_{k_n}^2/2) + P(\widehat{\alpha}_{k_n} \leq \frac{\alpha_{k_n}}{2}).
\end{aligned}
$$

But, by inequality (i) of Lemma 5.1,

$$
P(\widehat{\alpha}_{k_n} \leq \frac{\alpha_{k_n}}{2}) = P(\alpha_{k_n} - \widehat{\alpha}_{k_n} \geq \alpha_{k_n}/2) \leq P(\|\Gamma - \Gamma_n\|_\infty > \frac{\alpha_{k_n}}{2}),
$$

hence,

$$
\begin{aligned}
P(B_n > \epsilon) &\leq P(\|\Gamma - \Gamma_n\|_\infty > \epsilon^2 \frac{\alpha_{k_n}^2}{2}) + P(\|\Gamma - \Gamma_n\|_\infty > \frac{\alpha_{k_n}}{2}) \\
&\leq \frac{4E(||X||^4)}{\epsilon^4}\frac{1}{n\alpha_{k_n}^4} + \frac{4E(||X||^4)}{n\alpha_{k_n}^2}
\end{aligned}
$$

and finally,

$$
P(B_n > \epsilon) \leq \frac{4E(||X||^4)}{n\alpha_{k_n}^4}(1/\epsilon^4 + \alpha_{k_n}^2),
$$

so, $B_n \longrightarrow_p 0$ by assumption (H-4) and

$$
E_n \longrightarrow_p 0. \tag{5.9}
$$

Similar calculus lead to

$$E_n^{'} = ||\Gamma_{k_n}^{+} - \Gamma_{n,k_n}^{+}||_{\infty} \longrightarrow_p 0, \tag{5.10}$$

and then, we get (5.5).

To conclude for the convergence of C_n to zero, it remains to show that:

$$\frac{||\Gamma_{en} - \Gamma_e||_{\infty}}{\widehat{\alpha}_{k_n}^{3/2}} \longrightarrow_p 0. \tag{5.11}$$

Which is achieved from the inequality:

$$\frac{||\Gamma_{en} - \Gamma_e||_{\infty}}{\widehat{\alpha}_{k_n}^{3/2}} \leq ||\Gamma_{en} - \Gamma_e||_{\infty} \left(E_n + \frac{1}{\sqrt{\alpha_{k_n}}} \right)^3,$$

together with result 5.9, equality: $||\Gamma_{en} - \Gamma_e||_{\infty} = O_p(\frac{1}{\sqrt{n}})$, and assumption (H-4).

To end the proof of Theorem 3.1, we have to show that

$$D_n \longrightarrow_p 0. \tag{5.12}$$

Indeed, simples calculus lead to:

$$\begin{aligned} D_n &= ||\Gamma_{n,k_n}^{1/2} \Gamma_{n,k_n}^{+} \Gamma_{en} (\Gamma_{n,k_n}^{+})^{1/2} - \Gamma^{1/2} \Gamma^{-1} \Gamma_e \Gamma^{-1/2}||_{\infty} \\ &\leq ||\Gamma^{-1} \Gamma_e \Gamma^{-1/2}||_{\infty} ||\Gamma_{n,k_n}^{1/2} - \Gamma^{1/2}||_{\infty} + ||\Gamma_{n,k_n}^{1/2}||_{\infty} C_n. \end{aligned} \tag{5.13}$$

Futhermore ,

$$\begin{aligned} \|\Gamma_{n,k_n}^{1/2} - \Gamma^{1/2}\|_\infty &\leq \|\Gamma_{n,k_n}^{1/2} - \Gamma_n^{1/2}\|_\infty + \|\Gamma_n^{1/2} - \Gamma^{1/2}\|_\infty \\ &= \widehat{\alpha}_{k_n}^{1/2} + \|\Gamma_n^{1/2} - \Gamma^{1/2}\|_\infty \end{aligned}$$

and by Lemma 5.1 we get:

$$\|\Gamma_{n,k_n}^{1/2} - \Gamma^{1/2}\|_\infty \leq (\|\Gamma_n - \Gamma\|_\infty + \alpha_{k_n})^{1/2} + \|\Gamma_n^{1/2} - \Gamma^{1/2}\|_\infty. \tag{5.14}$$

But terms on the right side clearly tend in probability to zero so,

$$\|\Gamma_{n,k_n}^{1/2} - \Gamma^{1/2}\|_\infty \longrightarrow_p 0. \tag{5.15}$$

Now, since $\|\Gamma_{n,k_n}^{1/2}\|_\infty = \|\Gamma_n^{1/2}\|_\infty$ converges to the constant $\|\Gamma\|_\infty$, we get (5.12) because of inequality (5.13), results (5.2) and (5.15). This implies from Lemma 5.1 that $\widehat{\lambda}_{i,k_n} \longrightarrow_p \lambda_i$ so, $1/\widehat{\lambda}_{i,k_n} \longrightarrow_p 1/\lambda_i$ and,

$$D_n^{'} \longrightarrow_p 0 \tag{5.16}$$

Which complete the proof of Theorem 3.1 □

References

[1] Ash, R. B., Gardner, M. F. (1975). *Topics in Stochastic Processes.* Academic Press, New York.

[2] Bosq, D. (1991). Modelization, non-parametric estimation and prediction for continuous time processes. In: Roussas, G. (Ed.), Nonparametric Functional estimation and related Topics, NATO, ASI Series, pp. 509-529.

[3] Cardot, H. , Ferraty. F and Sarda, P. (1999) Functional linear model. *Statist.Probab.Let.* **45** 11-22.

[4] Cook, R.D. (1998). *Regression graphics: ideas for studying regressions through graphics.* New York: Wiley.

[5] Dauxois, J. and Pousse, A. (1976). Les analyses factorielles en calcul des probabilités et en statistique: essai d'étude synthétique. *Thèse Toulouse III.*

[6] Dunford, N. and Schwartz, J. T. (1964). *Linear Operators.* Intersience publishers, New-York.

[7] Ferraty, F. and Vieu P. (2000). Functional nonparametric model for scalar response. Preprint.

[8] Ferré, L. (1998). Determining the Dimension in Sliced Inverse Regression and Related Methods. *J. Amer. Statist. Assoc.* **93,** 132-140.

[9] Ferré, L. and Yao, A. F. (2000). Kernel estimate based on Functional the Inverse Regression. Preprint.

[10] Fortet, R. (1995). *Vecteurs, fonctions et distributions aléatoires dans les espaces de Hilbert.* Hermes. Paris.

[11] Hastie, T. and Mallows, C. (1993). A discussion of "A Statistical View of Some Chemometrics Regression Tools" by I. E. Frank and J. H. Friedman. *Technometrics,* **35,** 140-143.

[12] Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression: Based Estimation for Nonorthogonal Problems. *Technometrics,* **8,** 27-51.

[13] Hsing, T. and Caroll, R. J. (1992). An asymptotic theory for Sliced Inverse Regression. *Ann. Statist.*, **20**, 2, 1040-1061.

[14] Ito, K. and Nisio, M. (1968). On the convergence of sums of independent Banach space valued random variables. Osaka J. Maths. 5, 35-48.

[15] Leurgans , S.E., Moyeed, R.A. and Silverman, B.W. (1993). Canonical Correlation Analysis when the data are curves. *J.R.Statis. Soc., B,* **55,** 725-740.

[16] Li, K. C. (1991). Sliced Inverse Regression for dimension reduction. *J. Amer. Statist. Assoc.,* **86,** 316-342.

[17] Li, K. C. (1992). On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein 's Lemma. *J. Amer. Statist. Assoc.,* **87,** 1025-1039.

[18] Marx, B. D. and Eirlers P. H. (1996). Generalized Linear Regression on Sampled Signals with Penalized Likelihood. In: A. Forcina, G. M. Marcheti, R. Hartzinger, G. Galmacci. *Statistical Modeling. Proceedings of the 11th International workshop on Statistical modelling,* Orvietto.

[19] Massy, W. F. (1965). Principal Components Regression in Exploratory Statistical Research. *J. Amer. Statist. Assoc.* **60,** 234-246.

[20] Preda, C. (1999). Analyse factorielle d'un processus: problèmes d'approximation et de régression, *Thèse de doctorat de l'Université de Lille 1.*

[21] Preda, C. and Saporta, G. (2000). Régression PLS sur un processus stochastique. pub. IRMA, Lille 2000, **40,** VII.

[22] Ramsay, J. O. and Silverman, B. W. (1997). Functional Data Analysis. Springer-Verglag.

[23] Schott, J. R. (1994). Determining the Dimension in Sliced Inverse Regression. *J. Amer. Statist. Assoc.* **89,** 141-148.

[24] Wold, H. (1975). Soft Modeling by Latent Variables; the Nonlinear Iterative Partial Least Square Approach. *in perspectives in Probability and Statistics, Papers in Honour of M. S. Bartlett, ed. J. Gani, London: Academic Press.*

[25] Zhu L. X. and Fang K. T. (1996). Asymptotics for kernel estimate of Sliced Inverse Regression. *Ann. Statist.*, **3**, 1053-1068.

L. Ferré
Equipe GRIMM
Université Toulouse le Mirail,
5, Allées Antonio Machado,
31058 Toulouse cedex, France.

A.F. Yao
Laboratoire de Statistique
et Probabilités, URMR C 5583
Université Paul Sabatier,
118, route de Narbonne,
31062 Toulouse cedex, France.