



**HAL**  
open science

## High Dimensional Principal Projections

André Mas, Frits Ruymgaart

► **To cite this version:**

| André Mas, Frits Ruymgaart. High Dimensional Principal Projections. 2013. hal-00772880v1

**HAL Id: hal-00772880**

**<https://hal.science/hal-00772880v1>**

Preprint submitted on 11 Jan 2013 (v1), last revised 17 Jun 2018 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# HIGH-DIMENSIONAL PRINCIPAL PROJECTIONS

André MAS<sup>1</sup>

Institut de Mathématiques et de Modélisation de Montpellier  
Université Montpellier II

Frits RUYMGAART <sup>2</sup>

Department of Mathematics and Statistics  
Texas Tech University  
Lubbock, TX 79409

## Abstract

The Principal Component Analysis (PCA) is a famous technique from multivariate statistics. It is frequently carried out in dimension reduction either for functional data or in a high dimensional framework. To that aim PCA yields the eigenvectors  $(\widehat{\varphi}_i)_i$  of the covariance operator of a sample of interest. Dimension reduction is obtained by projecting on the eigenspaces spanned by the  $\widehat{\varphi}_i$ 's usually endowed with nice properties in term of optimal information. We focus on the empirical eigenprojectors in the functional PCA of a  $n$ -sample and prove several non asymptotic results. More specifically we provide an upper bound for their mean square risk. This rate does not depend on the rate of decrease of the eigenvalues which seems to be a new result. We also derive a lower bound on the risk. The latter matches the upper bound up to a  $\log n$  term. The results are applied to improve a technique of nonparametric functional estimation.

*Index Terms* — Functional principal Component Analysis, Dimension reduction, Nonparametric functional regression, Covariance operator, Perturbation theory.

*AMS 2010 Classification:* primary : 62H25, 62G08 ; secondary : 47A55.

## 1 Introduction

The theoretical covariance operator and its empirical analogue are objects of fundamental importance in the theory of functional data. Principal component analysis is of importance in its own right, and as a dimension reduction technique is often the first step to make finite-dimensional procedures suitable for high dimensional or functional data. In this paper we will elaborate on both aspects.

Let  $\mathbf{H}$  denote a real separable Hilbert space endowed with inner product  $\langle \cdot, \cdot \rangle$  and subsequent norm  $\|\cdot\|$  and let  $X_1, X_2, \dots, X_n$  be a sample of independent and identically distributed random elements with values in  $\mathbf{H}$ . Our first purpose is to establish sharp bounds on the prediction error of functional principal components in terms of the sample size  $n$  on the one hand and the index of the component on the other. These bounds allow us to obtain rates of convergence to zero of the prediction errors when the number of components increases to infinity. Secondly this result is applied to nonparametric functional regression. Biau and Mas (2012) reduce the dimension of the regressor by exploring a fixed, finite number of principal components. This naturally causes some information loss. In this paper we will be able to let the number of components and the corresponding dimension of the panel increase with the sample size.

As has been pointed out by Watson (1983) for matrices and by Dauxois Pousse and Romain (1982) for operators, the elegant techniques of perturbation theory (Dunford and Schwartz (1988) and Kato (1976)) seem to be tailored to the analysis of the spectral characteristics of random operators when they can be considered as small perturbations of an underlying target operator, as in the case with the sample covariance operators. For general theory with a special view towards applications in statistics we refer, for instance to Ruymgaart, Yang (1997) Mas, Menneteau (2003) or Cupidon et al. (2007). The interested reader should also note that kernel PCA was also

---

<sup>1</sup>Corresponding author, [mas@math.univ-montp2.fr](mailto:mas@math.univ-montp2.fr)

<sup>2</sup>[h.ruymgaart@ttu.edu](mailto:h.ruymgaart@ttu.edu)

investigated with similar tools in Koltchinskii (1998) and in Blanchard, Bousquet and Zwald (2007) for instance.

It is known that the covariance operator is Hermitian, positive and trace-class, and hence compact. It will be assumed to be strictly positive with all eigenvalues simple and, as usual, arranged in decreasing order with zero as limit. An essential difficulty which is typical for the infinite dimensional case is that the spacings between the eigenvalues become arbitrarily small. Relations between the spectral characteristics of the sample covariance operator such as the cumulative eigenprojector of order  $k$  and those of the population covariance operator are based on a Neumann series expansion for the sample covariance operator. For this expansion to be valid, however, the perturbation should be sufficiently small and in the present situation 'how small' depends on the spacing between the  $k^{\text{th}}$  and the  $(k+1)^{\text{st}}$  eigenvalue. Because these spacings tend to zero, for a given sample size  $n$  sufficiently small prediction errors can only be obtained for cumulative projections up to a finite order  $k = k(n)$ . We will see however that  $k(n) \rightarrow +\infty$  as  $n \rightarrow +\infty$ .

The organization of this article is as follows : in the next section we provide basic material about operators and perturbations, then in Section 3 the main theorems about random projectors are given and applied to nonparametric regression in a high dimensional or functional framework. The last section collects the proofs.

## 2 Notation, elements of operator theory and functional calculus

The space  $\mathbf{H}$  plays a crucial role since the data will be sampled in such a space. It is clearly suited to high dimensional Euclidean data. But many functional datasets may be also embedded in a Hilbert space : either in a space of non smooth functions such as  $\mathbf{L}^2([0, 1])$  or with tunable degree of regularity like the Sobolev space of functions  $\mathbf{H}^m([0, 1]) = \{f \in \mathbf{L}^2([0, 1]) : f^{(m)} \in \mathbf{L}^2([0, 1])\}$ . Another reason rather of mathematical nature highlights this choice. We focus here on the covariance structure and PCA for functional data. The trouble is that, for a Banach-valued random element  $X \in \mathbf{B}$ , the covariance operator is an operator from  $\mathbf{B}^*$  the dual of  $\mathbf{B}$  onto  $\mathbf{B}$  (see, for instance, Ledoux and Talagrand (1991) or Vakhania, Tarieladze, Chobanyan (1987)). The structure of the dual space  $\mathbf{B}^*$  may not be simple (not even for rather basic examples of  $\mathbf{B}$  such as the space of continuous functions on a compact set) except in the case when  $\mathbf{B} = \mathbf{H}$  is a Hilbert space where  $\mathbf{B}^* = \mathbf{H}$ .

### 2.1 Linear and covariance operators

We introduce the following two operator spaces and associated norms. The Banach space  $\mathcal{L}(\mathbf{H}, \mathbf{H}) = \mathcal{L}$  is the classical space of bounded linear operators mapping  $\mathbf{H}$  onto itself endowed with the norm defined for each  $T$  in  $\mathcal{L}$  by :

$$\|T\|_\infty = \sup_{x \in \mathcal{B}_1} \|Tx\|,$$

where  $\mathcal{B}_1$  is the unit sphere of  $\mathbf{H}$ . The Hilbert space  $\mathcal{L}_2$  is the space of Hilbert-Schmidt operators, ( $\mathcal{L}_2 \subset \mathcal{L}$ ) i.e. the space of those operators  $T$  such that, for any basis  $(e_k)_{k \in \mathbb{N}}$  of  $\mathbf{H}$ ,

$$\|T\|_2 = \sum_{k=1}^{+\infty} \|Te_k\|^2 < +\infty.$$

It is a well-known fact that  $\mathcal{L}_2$  is a separable Hilbert space whenever  $\mathbf{H}$  is. The inner product in  $\mathcal{L}_2$  is  $\langle T, S \rangle_2 = \sum_{k=1}^{+\infty} \langle Te_k, Se_k \rangle$  and does not depend on the choice of the basis. The space  $\mathcal{L}_1$  of finite trace (or nuclear) operators will be mentioned sometimes in the paper but is of secondary importance. We just mention that  $\mathcal{L}_1 \subset \mathcal{L}_2$ ; The two norms mentioned above are not equivalent and

$$\|\cdot\|_\infty \leq \|\cdot\|_2.$$

Consequently the canonical injection from  $\mathcal{L}_2$  onto  $\mathcal{L}$  is continuous. For further information on linear operators we refer to Schmeidler (1965), Weidman (1980), Dunford and Schwartz (1988), Gohberg, Goldberg and Kaashoek (1991) amongst many others.

Let the tensor product between  $u$  and  $v$  in  $\mathbf{H}$  stand for the one-rank operator from  $\mathbf{H}$  to  $\mathbf{H}$  by :

$$(u \otimes v)(t) = \langle u, t \rangle v,$$

for all  $t$  in  $\mathbf{H}$ .

Since we are concerned with functional PCA we will for the most part be dealing with covariance operators, and a precise definition and presentation of some of their main features is in order. Whenever  $\mathbb{E} \|X_1\|^2 < +\infty$ , the theoretical covariance operator  $\Sigma$  and its empirical counterpart,  $\widehat{\Sigma}$ , based on the sample are symmetric, positive, trace-class operators from  $\mathbf{H}$  to  $\mathbf{H}$  defined by :

$$\Sigma = \mathbb{E} ((X_1 - \mathbb{E}X_1) \otimes (X_1 - \mathbb{E}X_1)), \quad (1)$$

$$\widehat{\Sigma} = \frac{1}{n} \sum_{k=1}^n (X_k - \overline{X}_n) \otimes (X_k - \overline{X}_n), \quad (2)$$

where

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

When  $X$  is centered -which will be assumed throughout for mathematical convenience-  $\mathbb{E}X = 0$ ,  $\Sigma = \mathbb{E}(X \otimes X)$  and  $\widehat{\Sigma} = \frac{1}{n} \sum_{k=1}^n X_k \otimes X_k$ . By  $(\lambda_k, \varphi_k)$  we denote the  $k^{th}$  eigenelement (eigenvalue and eigenvector) of  $\Sigma$ . The  $\lambda_k$ 's are positive and we set  $\lambda_1 > \lambda_2 > \dots$ , and  $(\lambda_k)_{k \in \mathbb{N}} \in l_1$ .

In the Hilbert setting, the distribution of a centered random element  $X$  may be characterized in a very simple way. Indeed if  $=_d$  denotes equality in distribution, we have :

$$X =_d \sum_{j=1}^{+\infty} \sqrt{\lambda_j} \eta_j \varphi_j(\cdot) \quad (3)$$

where the  $\eta_j$ 's are centered non-correlated real random variables with unit variance. The above decomposition is often referred to as the Karhunen-Loève development or development of  $X$  with respect to its reproducing kernel Hilbert space (RKHS). Besides the  $\eta_j$ 's are independent when  $X$  is a Gaussian random function. For definition and studies of RKHS we refer to Berlinet, Thomas-Agnan (2004).

We add hats to denote the empirical versions of the eigenelements and we get  $\widehat{\Sigma} \widehat{\varphi}_i = \widehat{\lambda}_i \widehat{\varphi}_i$  for all  $i \in \mathbb{N}$ , with all  $\widehat{\lambda}_i = 0$  except for a finite number not exceeding  $n$  or  $n - 1$ .

## 2.2 Basics of functional calculus for operators

We refer the reader to Kato (1976), Weidman (1980), Dunford and Schwartz (1988), Gohberg, Goldberg and Kaashoek (1991) for complete presentation of the perturbation-based functional calculus for operators. Roughly speaking this theory allows to define  $f(T)$  where  $T$  is a linear operator and  $f$  is function defined on and with values in a suitable part of the complex plane. Note that the mathematical nature of  $f$  and  $T$  make them incompatible so that a special definition of  $f(T)$  is needed. It will be in particular possible to consider the eigenprojections of  $T$  as functions of this operator. For this purpose the resolvent  $(\zeta I - T)^{-1}$  defined for each  $\zeta$  outside the spectrum of  $T$  will be needed.

Let  $\partial\Omega_k$  be the contour of the complex plane associated to the connected domain  $\Omega_k$  separating the  $k$  first eigenvalues of  $\Sigma$  from the others and defined in the following way :  $\Omega_k = \cup_{j=1}^k \mathcal{C}_j$  where  $\mathcal{C}_j$  is a circle centered at  $\lambda_j$  and with radius  $\delta_j = \frac{1}{2} \min(\lambda_j - \lambda_{j+1}, \lambda_{j-1} - \lambda_j)$ . In the proofs we will introduce another contour but at this point we just need to focus on  $\partial\Omega_k$  in order to illustrate the definitions of projectors by perturbation techniques.

Results from perturbation theory yield :

$$\mathbf{P}_k = \frac{1}{2\pi i} \int_{\partial\Omega_k} (\zeta I - \Sigma)^{-1} d\zeta = \frac{1}{2\pi i} \sum_{j=1}^k \int_{\partial\mathcal{C}_j} (\zeta I - \Sigma)^{-1} d\zeta = \sum_{j=1}^k \pi_j, \quad (4)$$

where  $\pi_j$  is the orthogonal projector associated with the single eigenvector  $\varphi_j$ . This may be rewritten  $\pi_j = \varphi_j \otimes \varphi_j$  when  $\lambda_j$  is of order one. The empirical counterpart of (4) involves a

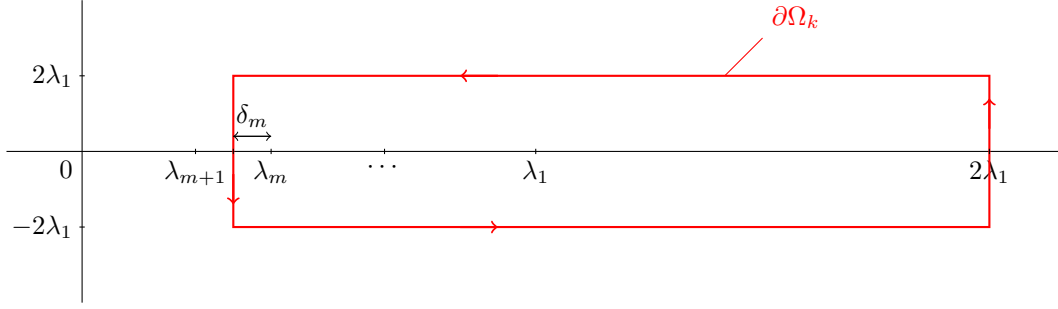


Figure 1: Rectangular contour

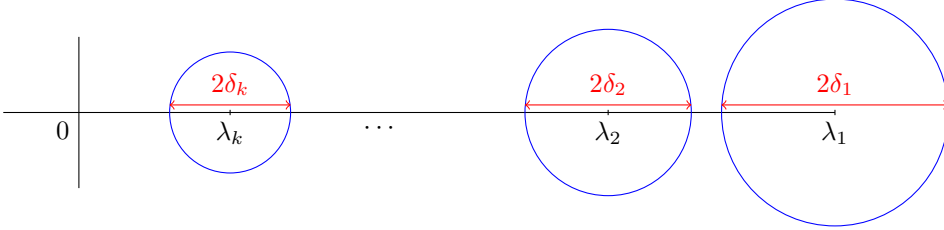


Figure 2: Contour made of disjoint circles

random contour, say,  $\partial\widehat{\Omega}_k$  similar in shape to  $\partial\Omega_k$  which contains the  $k$  first eigenvalues of  $\widehat{\Sigma}$ . In fact only the left vertex of  $\widehat{\Omega}_k$  may be considered as random :

$$\widehat{\mathbf{P}}_k = \frac{1}{2\pi i} \int_{\partial\widehat{\Omega}_k} (\zeta I - \widehat{\Sigma})^{-1} d\zeta. \quad (5)$$

In fact  $\widehat{\mathbf{P}}_k$  as defined in (5) equals the sum of all eigenprojections of  $\widehat{\Sigma}$  with eigenvalues inside  $\partial\widehat{\Omega}_k$  each denoted  $\widehat{\pi}_j$ .

### 3 Main results

We state first the assumptions needed to derive our results. Throughout the sequel the symbols  $c$  and  $C$  will denote universal constants.

#### Assumptions on $X$

Consider  $X$  a centered random function with Karhunen-Loeve development (3). We assume that the  $\eta_i$ 's are independent, that the eigenvalues  $\lambda_i$  are all of order one and that the sequence  $(\lambda_i)_{i \in \mathbb{N}}$  is decreasing. We also denote  $\delta_i = (\lambda_i - \lambda_{i+1})/2$ . We need higher moment assumptions because we need to apply Bernstein's exponential inequality to functionals of  $\widehat{\Sigma} - \Sigma$ . More specifically we assume that for all  $j, \ell \in \mathbb{N}$  there exists a constant  $b$  such that

$$\max_j \mathbb{E} \left( |\eta_j|^{2\ell} \right) \leq \ell! b^{\ell-1}. \quad (6)$$

These assumptions hold for instance for Gaussian  $X$  (with  $b = 4$  for instance) and when the  $\eta_j$ 's are compactly supported.

#### Assumptions on the spectrum of $\Sigma$

Let the function  $\lambda : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  be defined by  $\lambda(j) = \lambda_j$  for any  $j \in \mathbb{N}$  and by continuous interpolation between  $j$  and  $j+1$ . From the assumption above we already know that  $\sum_j \lambda_j < +\infty$ . Indeed the summability of the eigenvalues of  $\Sigma$  is ensured whenever  $\mathbb{E}\|X\|^2 < +\infty$ . Moreover, we assume that one of the two following assumptions hold for the eigenvalues

$$\lambda_j = c j^{-1-\alpha}, \quad (7)$$

$$\lambda_j = c \exp(-\alpha j), \quad (8)$$

where  $c$  and  $\alpha$  are some strictly positive constants. The spacings  $\delta_j$  can be obtained easily. These last conditions are rather mild and match a large class of eigenvalues. The generalization to other classes of eigenvalues sequences like Laurent series  $\lambda_j = cj^{-1-\alpha}(\log j)^{-\beta}$  or even  $\lambda_j = cj^{-1}(\log j)^{-1-\beta}$  with  $\beta > 0$  is straightforward and will be omitted.

### 3.1 Eigenprojectors

We start with a first result. Consider  $\widehat{\varphi}_k$  the  $k$ -th empirical eigenvector associated with the  $k$ -th empirical eigenvalue. With  $\widehat{\varphi}_k$  at hand it is possible to construct the projector  $\widehat{\pi}_k = \widehat{\varphi}_k \otimes \widehat{\varphi}_k$  associated to the  $k^{\text{th}}$  principal space of the functional PCA. We derive a rate of convergence for  $\widehat{\pi}_k$ .

**Proposition 1** *Let (6) and (7) or (8) hold. Then :*

$$\mathbb{E} \|\widehat{\pi}_k - \pi_k\|_{\infty}^2 \leq \exp(-c_1 \log^2 n) + \frac{c_2 k^2 \log^2 n}{n}$$

The proof of Proposition 1 is omitted because the derivation is essentially included in the proof of Theorem 5 below about the approximation of the projection involving the  $k$  first eigenvectors. It is easy to derive a similar result with the eigenfunctions or eigenvectors. Recall that  $\widehat{\pi}_k = \widehat{\varphi}_k \otimes \widehat{\varphi}_k$ .

**Corollary 2** *Under the assumptions of Proposition 1*

$$\mathbb{E} \|\widehat{\varphi}_k - \varphi_k\|^2 \leq 4 \exp(-c_1 \log^2 n) + 4 \frac{c_2 k^2 \log^2 n}{n}.$$

This result could be compared with Theorem 1 p.114 in Hall and Hosseini-Nasab (2006). These authors obtained in a similar framework but under milder moment assumptions  $\|\widehat{\varphi}_k - \varphi_k\| \leq c/(\delta_k \sqrt{n})$  with high probability. The corollary above may be viewed as some kind of improvement because the right hand side does not depend on the eigenvalue decay rate and is always sharper since  $k^2 \delta_k \rightarrow 0$  when  $k \rightarrow +\infty$ .

However an important question arises about the previous Proposition : is this approximation rate sharp ? We try to develop an answer below. To that end we focus on the simple projector encountered above  $\widehat{\pi}_k$ . We get :

**Proposition 3** *When (7) or (8) hold :*

$$\mathbb{E} \|\widehat{\pi}_k - \pi_k\|_{\infty}^2 \geq c_5 \frac{k^2}{n}.$$

This proves that the upper bound obtained in Proposition 1 is sharp and cannot be improved up to a  $\log^2 n$  term, at least in our framework and with our set of assumptions.

**Remark 4** *The proof of Proposition 3 is strongly connected to functional calculus techniques for operators. We write  $\widehat{\pi}_k = f_k(\widehat{\Sigma})$  and  $\pi_k = f_k(\Sigma)$ . Then considering  $\widehat{\Sigma}$  as a (random) perturbation of  $\Sigma$  the lower bound is derived through some development up to the second order of  $\widehat{\pi}_k - \pi_k$ , namely*

$$f_k(\widehat{\Sigma}) - f_k(\Sigma) = g_k^1(\widehat{\Sigma} - \Sigma) + g_k^2(\widehat{\Sigma} - \Sigma, \widehat{\Sigma} - \Sigma).$$

*We show that the second order term  $g_k^2(\cdot, \cdot)$  is negligible with respect to the first  $g_k^1(\widehat{\Sigma} - \Sigma)$  and we compute the exact value of  $\mathbb{E} \left\| g_k^1(\widehat{\Sigma} - \Sigma) \right\|_{\infty}^2$ .*

We are almost ready to give the main results about the cumulative projectors  $\widehat{\mathbf{P}}_k$  and  $\mathbf{P}_k$ . But for the sake of clarity we need to distinguish three situations.

**Case 1 :** (7) with  $\alpha > 1$  or (8) hold,

**Case 2 :** (7) holds with  $0 < \alpha \leq 1$ ,

**Case 3 :** (7) holds.

**Theorem 5** *Let (6) hold. Take  $\mathbf{u} = X_i$  or  $\mathbf{u} = X_{n+1}$  or  $\mathbf{u}$  nonrandom with the additional assumption that  $\sup_i |\langle \mathbf{u}, \varphi_i \rangle| / \lambda_i < +\infty$  then for all  $n$  :*

$$\begin{aligned} \text{In Case 1 : } \mathbb{E} \left\| \left( \widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 &\leq \exp(-c_3 \log^2 n) + \frac{c_4 k^2 \log^2 n}{n} \\ \text{In Case 2 : } \mathbb{E} \left\| \left( \widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 &\leq \exp(-c_3 \log^2 n) + \frac{c'_4 k^2 \log^2 k \log^2 n}{n} \\ \text{In Case 3 : } \mathbb{E} \left\| \widehat{\mathbf{P}}_k - \mathbf{P}_k \right\|_\infty^2 &\leq \exp(-c_3 \log^2 n) + \frac{c'_4 k^2 \log^2 k \log^2 n}{n} \end{aligned}$$

**Remark 6** *We have to underline that the perturbation techniques carried out to get the right hand side in the theorem above are specific. They are an efficient way to link the covariance sequence with the projection sequence. However since  $\widehat{\mathbf{P}}_k$  and  $\mathbf{P}_k$  are orthogonal projectors  $\|\widehat{\mathbf{P}}_k\| = \|\mathbf{P}_k\| = 1$  and  $\|\widehat{\mathbf{P}}_k - \mathbf{P}_k\|_\infty \leq 2$ . The problem is that nothing ensures that the bound derived in Theorem 5 is (significantly) smaller than 2. However this remark may be done about most of the results about eigenprojectors or unit eigenvectors .*

**Remark 7** *It is important to notice that we could not assess Theorem 5 in Case 3 when the eigenvalues decay very rapidly i.e. with exponential rate. Actually this result may be attainable but we could not get it with our method of proof. Roughly speaking, the assumption  $\sup_i |\langle \mathbf{u}, \varphi_i \rangle| / \lambda_i < +\infty$  is crucial to get the bound above with a fast decay of the  $\lambda_p$ 's. This assumption prevents us to obtain the uniform bound. This pathology is closely connected with the inverse problems aspects of the estimation of  $\widehat{\varphi}_k$  hence of  $\widehat{\pi}_k$ .*

### 3.2 Application to high-dimensional kernel estimation

We propose to apply the preceding results in the specific context of a two-step estimation algorithm adapted to the regression when  $X$  lies in a high-dimensional or functional space. Let  $(y, X)$  be a couple of random elements and let  $r(x) = \mathbb{E}(y|X = x)$  be the regression function evaluated at a fixed  $x$ . From the sample  $(y_i, X_i)_{1 \leq i \leq n} \in \mathbb{R} \times H$  where  $H$  stands like above for a Hilbert space of functions or the Euclidean space  $\mathbb{R}^d$  where  $d$  is large (typically with respect to the sample size) we investigate the classical issue of the estimation of  $r(x)$ . General regression estimation has been intensively studied along the past years (Masry (2005), Ferraty and Vieu. (2004, 2006), Ferraty, Mas, Vieu (2007), Ferraty, Laksaci, Tadj, Vieu (2010). Several methods may be carried to solve numerical or implementation issues like the computation of small ball probabilities, a crucial step in asymptotic theory. However the complexity of the general model leads to degenerate minimax rates of convergence, typically of order  $[\log n]^{-\alpha}$  for some positive  $\alpha$  as shown in Mas (2012). This may be seen as the ultimate side effect of the curse of dimensionality. This fact should lead to different strategies when modelling a nonlinear link with a functional data variable as predictor, paving the way to alternatives such as additive models, multiple indices models or mixed approach such as projection pursuit.

However here the method is the following. First compute the (potentially functional) PCA of the sample  $(X_i)_{1 \leq i \leq n}$  and retrieve the projectors  $\widehat{\mathbf{P}}_k$  with a prescribed  $k$ . Here  $k$  will be chosen so that  $k \ll d$ . Then compute an estimate of  $r(x)$  inspired from the classical Nadarya-Watson approach :

$$\widehat{r}_{NW}(x) = \frac{\sum_{i=1}^n Y_i K(\|X_i - x\|/h)}{\sum_{i=1}^n K(\|X_i - x\|/h)}.$$

By replacing the usual norm on the space  $\mathbf{H}$  by the 'projection' semi-norm  $\|\cdot\|_k$  defined by  $\|u\|_k = \|\widehat{\mathbf{P}}_k u\|_{\mathbb{R}^k}$  the estimate becomes :

$$\widehat{r}(x) = \frac{\sum_{i=1}^n Y_i K\left(\left\|\widehat{\mathbf{P}}_k(X_i - x)\right\|/h\right)}{\sum_{i=1}^n K\left(\left\|\widehat{\mathbf{P}}_k(X_i - x)\right\|/h\right)}.$$

This proposal is intuitive and rather simple to implement through standard statistical packages from most softwares. The problems arise from the mathematical properties of this estimate. First

if  $k$  is fixed  $\hat{r}(x)$  will **not** estimate  $r(x)$  in general. The best we may expect is to estimate fairly  $\mathbb{E}[y|\mathbf{P}_k(X) = \mathbf{P}_k(x)]$  or even  $r \circ \mathbf{P}_k(x)$ . As a consequence a bias occurs in general. This bias may be modest, if  $r \circ \mathbf{P}_k(x)$  were a good approximation of  $r(x)$  for instance which in turn will be true if either  $x$  is close to  $\mathbf{P}_k(x)$  or if  $r \circ \mathbf{P}_k$  approximates  $r$  well. This would justify an asymptotic framework, that is when  $k$  tends to infinity. But we intend here again to give non-asymptotic results.

Furthermore the numerator and denominator of  $\hat{r}(x)$  are not sums of independent random variables and the dependence structure is, up to the authors, very confused. However, convergence theorems could be stated at the expense of assumptions on the dependence structure of the sequence of random variables  $\left\| \left\| \hat{\mathbf{P}}_k(X_i - x) \right\| \right\|_{1 \leq i \leq n}$ .

We introduce

$$r^*(x) = \frac{\sum_{i=1}^n Y_i K(\|\mathbf{P}_k(X_i - x)\|/h)}{\sum_{i=1}^n K(\|\mathbf{P}_k(X_i - x)\|/h)}.$$

Contrary to  $\hat{r}(x)$  the random variable  $r^*(x)$  is **not** an estimate since  $\mathbf{P}_k$  is the population projector of the PCA and is consequently unknown. But its numerator and denominator are both sums of independent random variables. It may be viewed as an oracle in the sense that the risk for  $r^*(x)$  is expected to bound below the risk for  $\hat{r}(x)$ . A good way for investigating this issue would be to show that for some target  $r_k(x)$  :

$$\frac{r^*(x) - r_k(x)}{\hat{r}(x) - r_k(x)} \rightarrow 1,$$

for some stochastic mode of convergence such as almost sure convergence or convergence in mean square. But as we underlined just above the target  $r_k(x)$  may not be well defined. It is not either precisely the goal of our work to address this issue. Biau and Mas (2012) proceed differently. They introduce  $\tau_n$  the minimax rate of convergence in the non-parametric regression model over a well-chosen class of functions and show that :

$$\frac{\hat{r}(x) - r^*(x)}{\tau_n} \xrightarrow{n \rightarrow +\infty} 0, \tag{9}$$

as the sample size increases and for a fixed dimension of projection  $D$  (denoted here by  $k$  and that may increase with the sample size in our framework). We cannot follow the same route here because the dimension is not fixed and consequently the minimax rate  $\tau_n$  is unknown. Besides in (9) constants depending on  $D$  are used for bounding  $(\hat{r}(x) - r^*(x))/\tau_n$ . These constants are made explicit here and we must take them into account because they usually tend to infinity when  $k$  does.

For later use let :

$$S^* = \sum_{i=1}^n K(\|\mathbf{P}_k(X_i - x)\|), \quad \hat{S}_n = \sum_{i=1}^n K\left(\left\|\hat{\mathbf{P}}_k(X_i - x)\right\|\right),$$

$$Z^* = \sum_{i=1}^n Y_i K(\|\mathbf{P}_k(X_i - x)\|/h), \quad \hat{Z}_n = \sum_{i=1}^n Y_i K\left(\left\|\hat{\mathbf{P}}_k(X_i - x)\right\|/h\right).$$

We introduce now the set of assumptions we need to derive our main Proposition.

**Assumptions on the small ball probabilities [A1]** :  $F_k(s) = \mathbb{P}(\|\mathbf{P}_k(X_i - x)\| < s)$  with  $F_k(s) > 0$  in a right neighborhood of 0 and  $\sup_{h \in [0,1]} \frac{F_k(hu)}{F_k(h)} \leq c_F u^p$  with  $p \leq k$ .

**Assumptions on the kernel K [A2]** : The kernel  $K$  is bounded above and below on its support  $[0, 1]$  with  $0 < \inf_{u \in [0,1]} K(u) = \underline{K} \leq |K|_\infty$ ,  $u^p K'(u) \in L^1([0, 1])$  where  $p$  as in [A1].

**Assumptions on the regression function [A3]** :  $r$  is bounded.

At this point some comments are needed about the assumptions : [A2] and [A3] are classical. The reader familiar with nonparametric methods will remark that we do not need regularity assumptions about  $r$ . This is due to the fact that below we are only interested in bounding the difference  $\hat{r}(x) - r^*(x)$  and not  $\hat{r}(x) - r(x)$ . As a consequence the approximation of  $r$  -in its functional analysis sense- is not our aim here and ensuing smoothness conditions are no more required. Assumption [A1] though not quite standard is rather mild. Indeed we consider there the cumulative density function (in a neighborhood of zero) of the positive real random variable  $\|\mathbf{P}_k(X_i - x)\|$  which in turn is the norm of the  $\mathbb{R}^k$  valued random vector  $\mathbf{P}_k(X_i - x)$ . It is natural



to think that the local behaviour of  $F_k(s)$  is polynomial with degree less than  $s$  in a neighborhood of 0. Smoothness of the density of  $\mathbf{P}_k X_i$  evaluated at  $\mathbf{P}_k x$  would ensure this fact. And if  $F_k(s) \sim s^p L(s)$  with  $L$  a slowly-varying function at 0 then [A1] holds.

**Proposition 8** *The following non asymptotic bound holds under the assumptions (6), (8) or (7) with  $\alpha > 1$  and [A1-3] above :*

$$\mathbb{E} [\widehat{r}(x) - r^*(x)]^2 \leq c_6 \frac{k^4}{nh^2} \log^2 n \log \left( \frac{h\sqrt{n}}{k^2} \right). \quad (10)$$

A first difference appears with Biau and Mas (2012). The bound of Proposition 8 displays explicitly the dimension  $k$  whereas this dimension was fixed and did not appear in the work of these authors. A comparison with their Corollary 4.1 shows that, up to a  $\log n$  term, the numerator  $k^2 \log^2 k$  is new. It accounts for the price to pay to let the dimension increase with the sample size.

Like in Biau and Mas (2012) a crucial issue consists in comparing the bound above with a sort of benchmark rate. For instance the comparison with the minimax rate was pertinent then but it is no more clear if this comparison is still possible. We develop this point in the Remark below. We remind that in a non asymptotic framework the minimax rate for the risk in nonparametric regression (see for instance the monograph by Tsybakov (2004) and references therein) is up to constants  $\tau_n^*(\alpha, k) = n^{-2\alpha/(2\alpha+k)}$  where  $\alpha$  is a smoothness parameter unimportant here. When a Nadarya-Watson estimate is selected the optimal bandwidth is an  $h^* = O(n^{-1/(2\alpha+k)})$ , and we can write  $\tau_n^*(p, k) = [n \cdot (h^*)^k]^{-1}$ .

**Remark 9** *Plugging directly a  $k_n$  in the rate  $\tau_n^*(\alpha, \cdot)$  may not necessarily make sense. If  $k_n$  is too large  $\tau_n^*(\alpha, k_n)$  may not tend to zero (in an asymptotic sense) or even become small. Typically here for  $\tau_n^*(\alpha, k_n)$  to tend to zero we should take  $k_n := k_n^* = o(\log n)$ . Plugging this  $k_n^*$  into (10) we see that  $\mathbb{E} [\widehat{r}(x) - r^*(x)]^2$  is up to a power of  $\log n$  negligible with respect to  $\tau_n^*(p, k_n)$ . But the route described here is specific. It presupposes that the bandwidth is picked -and optimal here- before the dimension is selected. Besides the minimax approach is not clear in this context as shown in the beginning of this remark.*

We propose to base the next result on another argument. After the remark above we intend not to privilege any of the bandwidth or the projection dimension. The ideal would be to check that uniformly with respect to  $h$  and  $k$  (10) does not finally change much the pointwise risk  $\mathbb{E} [\widehat{r}(x) - r(x)]^2$ . Note that :

$$\begin{aligned} r(x) &= \mathbb{E}(Y|X=x) = r(x) - r \circ \mathbf{P}_k(x) + r \circ \mathbf{P}_k(x) \\ &= \mathcal{A}_k + r \circ \mathbf{P}_k(x) \end{aligned}$$

where  $\mathcal{A}_k$  is clearly an approximation error and  $r \circ \mathbf{P}_k(x)$  is the conditional expectation of  $y$  when the data are projected onto the image of  $\mathbf{P}_k$  and will be approximated empirically by  $\mathbb{E}(y|\mathbf{P}_k X_1, \dots, \mathbf{P}_k X_n)$ . In addition we should consider here that we undergo a blurred context where  $\mathbf{P}_k$  is an unknown operator and estimated. The decomposition above becomes

$$r(x) = \mathcal{A}_k + r \circ \widehat{\mathbf{P}}_k(x) + r \circ \mathbf{P}_k(x) - r \circ \widehat{\mathbf{P}}_k(x),$$

where  $r \circ \mathbf{P}_k(x) - r \circ \widehat{\mathbf{P}}_k(x) = \eta_k$  may be viewed as an additional noise. Besides  $\widehat{r}(x)$  is a natural candidate to estimate  $r \circ \widehat{\mathbf{P}}_k(x)$ . Consequently the initial model could be rewritten in additive form and with an abuse of notation :

$$y = r \circ \widehat{\mathbf{P}}_k(x) + \mathcal{A}_k + \eta_k + \varepsilon, \quad (11)$$

where  $\varepsilon$  is the initial noise  $y - \mathbb{E}(Y|X=x)$ . The preceding decomposition of  $y$  is split in three terms : the unknown parameter  $r \circ \widehat{\mathbf{P}}_k(x)$  which is artificially depending on the data here, a nonrandom approximating  $\mathcal{A}_k$  and the noise part shared between  $\varepsilon$  and  $\eta_k$ . It turns out that Proposition 8 bounds the second moment of  $\eta_k$ . Now the classical Bias-Variance decomposition gives us some tracks to deal accurately with (10) : we should just make sure that the variance

part due to  $\eta_k$  does not exceed or is even negligible with respect to the variance part due to  $\varepsilon$  and whatever the values of  $h$  and  $k$ . In other words a fair way to go is to prove that :

$$\sup_{h,k} nh^k \mathbb{E} [\widehat{r}(x) - r^*(x)]^2 \leq c_7.$$

**Proposition 10** *Under the assumptions of the previous proposition and assuming that  $k > 2$ ,  $0 \leq h \leq h_{\max} < 1$  and that  $h^\tau \log^3 n$  is uniformly bounded with respect to  $n$  for any  $\tau > 0$  there exists a constant  $c_7(h_{\max})$  such that :*

$$\sup_{h,k} nh^k \mathbb{E} [\widehat{r}(x) - r^*(x)]^2 \leq c_7.$$

*In an asymptotic perspective  $nh^k \mathbb{E} [\widehat{r}(x) - r^*(x)]^2 \rightarrow_{n \rightarrow +\infty} 0$  whenever  $k_n / \log \log n$  is bounded above and for any bandwidth  $h$  depending either on  $k_n$  or  $n$  such that  $h \rightarrow 0$ .*

The proof of Proposition 10 is omitted. For the first part it suffices to see that  $h^{k-2} k^4 \log^3 n \leq h^{k-2-\tau} k^4 h^\tau \log^3 n \leq h_{\max}^{k-2-\tau} k^4 \leq c_7$  where here and elsewhere  $x \leq y$  means that  $x \leq cy$  for some constant  $c$ . The second part follows from straightforward calculations.

**Remark 11** *This avoids the considerations of Remark 9 and allows independent choices of  $h$  and  $k$  by comparing  $\mathbb{E} [\widehat{r}(x) - r^*(x)]^2$  only with the variance part of the estimate in the ideal model  $y = r \circ \mathbf{P}_k(x) + \varepsilon$  and not with the bias and the purely deterministic approximation error.*

## 4 Mathematical derivations

We introduce the following events for later use. Let  $0 < c_0 < 1$

$$\begin{aligned} \mathcal{A}_k &= \left\{ \widehat{\lambda}_k \in \Omega_k \right\} \cap \left\{ \widehat{\lambda}_{k+1} \notin \Omega_k \right\}, \\ \mathcal{B}_k(c_0) &= \left\{ \sup_{\zeta \in \partial \Omega_k} \left\| (\zeta I - \Sigma)^{-1/2} (\widehat{\Sigma} - \Sigma) (\zeta I - \Sigma)^{-1/2} \right\|_\infty < c_0 \right\}. \end{aligned}$$

The event  $\mathcal{B}_k(c_0)$  will be introduced within the proof of our main result. It will be shown just below in Lemma 14 that for a well-chosen and small enough  $c_0 < 1/2$ ,  $\mathcal{B}_k(c_0) \subset \mathcal{A}_k$ . When  $\mathcal{A}_k$  holds (hence when  $\mathcal{B}_k(c_0)$  does) the  $k$  first empirical eigenvalues (and only the  $k$  first) are inside  $\Omega_k$ . Then subject to these events we can derive the linearization :

$$\widehat{\mathbf{P}}_k - \mathbf{P}_k = \frac{1}{2\pi i} \int_{\partial \Omega_k} \left[ (\zeta I - \widehat{\Sigma})^{-1} - (\zeta I - \Sigma)^{-1} \right] d\zeta. \quad (12)$$

The formula above comes down to removing the random contour  $\partial \widehat{\Omega}_k$  in the definition of  $\widehat{\mathbf{P}}_k$  at (5). But we have to condition with respect to  $\mathcal{A}_k$  first. On  $\overline{\mathcal{A}}_k$  the strategy will be to see that  $\mathbb{E} \left\| \widehat{\mathbf{P}}_k - \mathbf{P}_k \right\|_\infty 1_{\overline{\mathcal{A}}_k} \leq 2\mathbb{P}(\overline{\mathcal{A}}_k)$  and to bound accurately the latter probability. Consequently we will restrict ourselves to bounding  $\mathbb{P}(\overline{\mathcal{B}}_k(c_0))$ .

### 4.1 Proof of Theorem 5

The proof takes two steps. First we apply the linearization formula (12) above but we have to control  $\mathbb{P}(\overline{\mathcal{A}}_k)$  non asymptotically which is achieved at Lemma 13 below. Second we will use perturbation theory to bound accurately the linearized difference between  $\widehat{\mathbf{P}}_k$  and  $\mathbf{P}_k$ .

We state or recall now two technical Lemmas.

**Lemma 12** *When (7) or (8) hold there exists a constant  $c_\alpha$  such that for all  $k \geq 1$*

$$\sum_{i \neq k} \frac{\lambda_i}{|\lambda_i - \lambda_k|} + \frac{\lambda_k}{\delta_k} \leq c_\alpha k. \quad (13)$$

The two next Lemmas aim at proving that the events  $\mathcal{A}_k$  and  $\mathcal{B}_k(c_0)$  hold with a high probability. We derive in fact two exponential inequalities of the same kind showing finally that  $\mathbb{P}(\overline{\mathcal{B}}_k(t)) \simeq \mathbb{P}(\overline{\mathcal{A}}_k) \simeq \exp(-k^2/n)$ .

**Lemma 13** Denote  $\mathbf{a}_k = \left[ \sum_{i \neq k} \frac{\lambda_i}{|\lambda_i - \lambda_k|} + \frac{\lambda_k}{\delta_k} \right]$ . We have :

$$\mathbb{P}(\overline{\mathcal{B}}_k(t)) \leq 2 \exp \left( - \frac{nt^2}{2\mathbf{a}_k^2(2b-1) + \frac{256b^3}{(2b-1)\mathbf{a}_k}t} \right).$$

**Proof of Lemma 13 :** In this subsection, we use Bernstein's exponential inequality for Hilbert-valued random variables (see for instance Bosq (2000) p.49) and references therein). First denote  $\mu_k = \lambda_k - \delta_k$  and  $\tilde{\Pi}(\mu_k) = (\mu_k I - \Sigma)^{-1/2} \left( \widehat{\Sigma} - \Sigma \right) (\mu_k I - \Sigma)^{-1/2}$  and rewrite :

$$(\zeta I - \Sigma)^{-1/2} \left( \widehat{\Sigma} - \Sigma \right) (\zeta I - \Sigma)^{-1/2} = G_k(\zeta) \tilde{\Pi}(\mu_k) G_k(\zeta),$$

with  $G_k(\zeta) = (\zeta I - \Sigma)^{-1/2} (\mu_k I - \Sigma)^{1/2}$  we get :

$$\begin{aligned} & \sup_{\zeta \in \partial\Omega_k} \left\| (\zeta I - \Sigma)^{-1/2} \left( \widehat{\Sigma} - \Sigma \right) (\zeta I - \Sigma)^{-1/2} \right\|_{\infty} \\ & \leq \sup_{\zeta \in \partial\Omega_k} \|G_k(\zeta)\|_{\infty} \cdot \left\| \tilde{\Pi}(\mu_k) \right\|_{\infty}, \end{aligned}$$

and note that  $\sup_k \sup_{\zeta \in \partial\Omega_k} \|G_k(\zeta)\|_{\infty} \leq 1$  whatever the contour (rectangle or union of circles) chosen. Then :

$$\begin{aligned} \left\| \tilde{\Pi}(\mu_k) \right\|_{\infty}^2 & \leq \left\| \tilde{\Pi}(\mu_k) \right\|_2^2 \\ & = \sum_{p,q \neq k}^{+\infty} \frac{\langle (\Sigma_n - \Sigma)(\varphi_p), \varphi_q \rangle^2}{|\mu_k - \lambda_p| |\mu_k - \lambda_q|} + \frac{\langle (\Sigma_n - \Sigma)(\varphi_k), \varphi_k \rangle^2}{\delta_k^2} \\ & \leq 2 \sum_{p,q \neq k}^{+\infty} \frac{\langle (\Sigma_n - \Sigma)(\varphi_p), \varphi_q \rangle^2}{|\lambda_k - \lambda_p| |\lambda_k - \lambda_q|} + 2 \frac{\langle (\Sigma_n - \Sigma)(\varphi_k), \varphi_k \rangle^2}{\delta_k^2}, \end{aligned} \quad (14)$$

since for  $p \neq k$   $|\mu_k - \lambda_p| \geq |\lambda_k - \lambda_p| - \delta_k \geq |\lambda_k - \lambda_p|/2$ .

Finally our only task is to give an exponential bound for the probability :

$$\mathbb{P} \left( \sum_{p,q \neq k}^{+\infty} \frac{\langle (\Sigma_n - \Sigma)(\varphi_p), \varphi_q \rangle^2}{|\lambda_k - \lambda_p| |\lambda_k - \lambda_q|} + \frac{\langle (\Sigma_n - \Sigma)(\varphi_k), \varphi_k \rangle^2}{\delta_k^2} > t \right).$$

Consider the linear bounded symmetric operator  $S_k$  defined in the basis  $(\varphi_i)_{1 \leq i}$  by  $S_k(\varphi_i) = \varphi_i / \sqrt{|\lambda_k - \lambda_i|}$  for  $i \neq k$  and  $S_k(\varphi_k) = \varphi_k / \sqrt{\delta_k}$ . It is plain that :

$$\sum_{p,q \neq k}^{+\infty} \frac{\langle (\Sigma_n - \Sigma)(\varphi_p), \varphi_q \rangle^2}{|\lambda_k - \lambda_p| |\lambda_k - \lambda_q|} + \frac{\langle (\Sigma_n - \Sigma)(\varphi_k), \varphi_k \rangle^2}{\delta_k^2} = \|S_k(\Sigma_n - \Sigma)S_k\|_2^2,$$

with  $S_k(\Sigma_n - \Sigma)S_k = (1/n) \sum_{i=1}^n Z_{i,k}$  with  $Z_{i,k} = S_k X_i \otimes S_k X_i - S_k \Sigma S_k$  hence  $\mathbb{E} Z_{i,k} = 0$ . In order to apply Theorem 2.5 in Bosq (2000) mentioned above we have to identify  $l_k$  and  $b_k$  such that for all integer  $m$ ,  $n^{-m+1} \mathbb{E} \|Z_{1,k}\|_2^m \leq m! l_k^2 b_k^{m-2} / 2$ . We claim that the previous inequality holds with  $b_k = \mathbf{a}_k 256b^3 / [(2b-1)n]$  and  $l_k = (2b-1) \mathbf{a}_k^2 / n$ . We prove it now.

Let  $m = 2$  we get  $\mathbb{E} \|Z_{1,k}\|_2^2 = \mathbb{E} \|S_k X_1\|^4 - \|S_k \Sigma S_k\|_2^2$  where  $\|S_k \Sigma S_k\|_2^2 = \sum_{i=1}^{+\infty} \lambda_i^2 / |\lambda_k - \lambda_i|^2$  and :

$$\begin{aligned} \mathbb{E} \|S_k X_1\|^4 & = \mathbb{E} \left( \sum_{i=1}^{+\infty} \lambda_i \eta_i^2 / |\lambda_k - \lambda_i| \right)^2 = \sum_{i=1}^{+\infty} \frac{\lambda_i^2 \mathbb{E} \eta_i^4}{|\lambda_k - \lambda_i|^2} + \sum_{i \neq j}^{+\infty} \frac{\lambda_i \lambda_j}{|\lambda_k - \lambda_i| |\lambda_k - \lambda_j|} \\ & \leq 2b \sum_{i=1}^{+\infty} \frac{\lambda_i^2}{|\lambda_k - \lambda_i|^2} + \left( \sum_{i=1}^{+\infty} \frac{\lambda_i}{|\lambda_k - \lambda_i|} \right)^2 - \sum_{i=1}^{+\infty} \frac{\lambda_i^2}{|\lambda_k - \lambda_i|^2}. \end{aligned}$$

Then

$$\begin{aligned} \mathbb{E} \|S_k X_1\|^4 - \|S_k \Sigma S_k\|_2^2 &= 2(b-1) \sum_{i=1}^{+\infty} \frac{\lambda_i^2}{|\lambda_k - \lambda_i|^2} + \left( \sum_{i=1}^{+\infty} \frac{\lambda_i}{|\lambda_k - \lambda_i|} \right)^2 \\ &\leq (2b-1) \left( \sum_{i=1}^{+\infty} \frac{\lambda_i}{|\lambda_k - \lambda_i|} \right)^2. \end{aligned}$$

From this we see that we can choose  $l_k^2 = (2b-1) \mathbf{a}_k^2/n$ .

By classical properties of the norm in Hilbert spaces and Minkowski's inequality we get :  $\mathbb{E} \|Z_{1,k}\|_2^m \leq 2^{m/2} \mathbb{E} \left( \|S_k X_1\|^4 + \|S_k \Sigma S_k\|_2^2 \right)^{m/2} \leq 2^m \mathbb{E} \left( \|S_k X_1\|^{2m} + \|S_k \Sigma S_k\|_2^m \right)$ . We keep this inequality in mind. Now we bound more generally  $\mathbb{E} \|S_k X_1\|^{2m}$ . We have to compute :

$$\sum_{i_1, \dots, i_m} \frac{\lambda_{i_1} \dots \lambda_{i_m}}{|\lambda_k - \lambda_{i_1}| \dots |\lambda_k - \lambda_{i_m}|} \mathbb{E} [\eta_{i_1}^2 \dots \eta_{i_m}^2]$$

We focus on  $\mathbb{E} [\eta_{i_1}^2 \dots \eta_{i_m}^2]$  and recall that we assumed just above (6) that the  $\eta_j$ 's are independent. In order to bound accurately this expectation we consider the rearrangement of the  $\eta_i$ 's involved and introduce  $D$  as the numbers of distinct indices in the product  $\eta_{i_1}^2 \dots \eta_{i_m}^2$ . Clearly  $1 \leq D \leq m$ . We write  $\eta_{i_1}^2 \dots \eta_{i_m}^2 = \prod_{l=1}^D \eta_{p_l}^{2\alpha_l}$  where  $p_l$  are all distinct indices (hence the  $\eta_{p_l}$  are independent) and the sequence of exponents  $(\alpha_1, \dots, \alpha_D)$  adds up to  $m$ . Hence  $\mathbb{E} [\eta_{i_1}^2 \dots \eta_{i_m}^2] = \prod_{l=1}^D \mathbb{E} \eta_{p_l}^{2\alpha_l} \leq \prod_{l=1}^D (\alpha_l!) \cdot b^{\alpha_l - 1} = b^{m-D} \prod_{l=1}^D \alpha_l!$ . Since  $\prod_{l=1}^D \alpha_l! / m! \leq 1$  we get

$$\mathbb{E} [\eta_{i_1}^2 \dots \eta_{i_m}^2] \leq m! b^m \frac{1}{b^D}.$$

From the moment bound (6) and by Jensen's inequality we get  $1 = \mathbb{E} (|\eta_1|^2)^2 \leq \mathbb{E} (|\eta_1|^4) \leq \max_j \mathbb{E} (|\eta_j|^4) \leq 2b$  hence  $1/b^D \leq 2^D \leq 2^m$ . Collecting all these facts we get first :

$$\mathbb{E} \|S_k X_1\|^{2m} \leq m! (2b)^m \left( \sum_i \frac{\lambda_i}{|\lambda_k - \lambda_i|} \right)^m$$

From  $\|S_k \Sigma S_k\|_2^m = \left( \sum_{i=1}^{+\infty} \lambda_i^2 / |\lambda_k - \lambda_i|^2 \right)^{m/2}$  we can derive the moment inequality involving  $Z_{1,k}$

$$\begin{aligned} \mathbb{E} \|Z_{1,k}\|_2^m &\leq 2^m \left[ m! (2b)^m \left( \sum_{i \neq k} \frac{\lambda_i}{|\lambda_k - \lambda_i|} \right)^m + \left( \sum_{i=1}^{+\infty} \lambda_i^2 / |\lambda_k - \lambda_i|^2 \right)^{m/2} \right] \\ &\leq 2^m m! (2b)^m \left( \sum_{i \neq k} \frac{\lambda_i}{|\lambda_k - \lambda_i|} \right)^m \left[ 1 + \frac{\left( \sum_{i=1}^{+\infty} \lambda_i^2 / |\lambda_k - \lambda_i|^2 \right)^{m/2}}{m! (2b)^m \left( \sum_i \frac{\lambda_i}{|\lambda_k - \lambda_i|} \right)^m} \right] \\ &\leq 2^{m+1} m! (2b)^m \left( \sum_i \frac{\lambda_i}{|\lambda_k - \lambda_i|} \right)^m \\ &= 2^{m+1} m! (2b)^m [\mathbf{a}_k]^m. \end{aligned}$$

It is now a simple computation to identify this last term with  $b_k$  and we get  $b_k = \mathbf{a}_k 256b^3 / [(2b-1)n]$ . The exponential inequality in the Lemma is derived from Theorem 2.21 p.49, equation (2.21) in Bosq (2000).

**Lemma 14** *We have :*

$$\mathbb{P} (\overline{\mathcal{A}}_k) \leq C \exp \left( -c \frac{n}{k^2} \right).$$

whenever  $\Omega_k$  is the contour drawn at figures 1 or 2.

**Proof of Lemma 14 :**

We should make sure that, with a high probability of order about  $1 - \exp(-k^2/n)$   $\mathcal{A}_k = \{\widehat{\lambda}_k \in \Omega_k\} \cap \{\widehat{\lambda}_{k+1} \notin \Omega_k\}$  holds. The event  $\mathcal{A}_k$  is defined here for  $\Omega_k$  rectangle but the method will (should) apply with only slight modifications to the union of  $k$  circles centered on the  $\lambda_k$  and with radii  $\delta_k$ .

Since  $\mathbb{P}(\overline{\mathcal{A}}_k) \leq \mathbb{P}(\widehat{\lambda}_k \leq \lambda_k - \delta_k) + \mathbb{P}(\widehat{\lambda}_{k+1} \geq \lambda_k - \delta_k)$  we focus and giving a bound for  $\mathbb{P}(\widehat{\lambda}_k \leq \lambda_k - \delta_k)$  and we let the reader check that the method applies to proving a bound for the other probability involving  $\widehat{\lambda}_{k+1}$ .

So we focus on  $\mathbb{P}(\widehat{\lambda}_k \leq \lambda_k - \delta_k)$  and are going to apply the minimax theorem for compact operators eigenvalues namely here :

$$\widehat{\lambda}_k = \max_{\dim E_k = k} \min_{x \in E_k, \|x\|=1} \langle \Sigma_n x, x \rangle$$

where  $E_k$  is any  $k$ -dimensional vector space in  $\mathbf{H}$ . Define  $E_k^* = \text{span}\{\varphi_i : 1 \leq i \leq k\}$  then

$$\begin{aligned} \widehat{\lambda}_k &\geq \min_{x \in E_k^*, \|x\| \leq 1} \langle (\Sigma_n - \Sigma)x, x \rangle + \min_{x \in E_k^*} \langle \Sigma x, x \rangle \\ &= \min_{x \in E_k^*, \|x\| \leq 1} \langle (\Sigma_n - \Sigma)x, x \rangle + \lambda_k \end{aligned}$$

hence  $\mathbb{P}(\widehat{\lambda}_k \leq \lambda_k - \delta_k) \leq \mathbb{P}(\min_{x \in E_k^*} \langle (\Sigma_n - \Sigma)x, x \rangle \leq -\delta_k)$ . Notice that even in  $\Sigma_n$  and  $\Sigma$  are positive operators  $\Sigma_n - \Sigma$  is not, though symmetric and  $\min_{x \in E_k^*} \langle (\Sigma_n - \Sigma)x, x \rangle$  may take negative values.

Let  $M_n$  be the  $k$  dimensional matrix defined by  $[M_n]_{i,j} = \langle (\Sigma_n - \Sigma)\varphi_i, \varphi_j \rangle$  and  $D_\lambda$  be the diagonal matrix of size  $k$  with  $[D_\lambda]_{ii} = \sqrt{|\lambda_{k+1} - \lambda_i|}$ . It is a well known fact from matrix analysis that

$$\min_{x \in E_k^*, \|x\| \leq 1} \langle (\Sigma_n - \Sigma)x, x \rangle = l_k(M_n)$$

where  $l_k(M_n)$  stands for the smallest eigenvalue of  $M_n$ . Now we get successively :

$$\mathbb{P}(\widehat{\lambda}_k \leq \lambda_k - \delta_k) \leq \mathbb{P}(l_k(M_n) \leq -\delta_k) \leq \mathbb{P}(|l_k(M_n)| \geq \delta_k)$$

We follow the proof of Proposition 1.3 page 98 in Gohberg, Goldbert and Kaashoek (1991) which is close to our Lemma. Let  $M$  be a  $k - 1$ -dimensional space in  $\mathbb{R}^k$  and pick a  $u \perp M$  with unit norm :

$$\begin{aligned} \langle M_n u, u \rangle &= \langle D_\lambda^{-1} M_n D_\lambda^{-1} D_\lambda u, D_\lambda u \rangle \\ &= \frac{\langle D_\lambda^{-1} M_n D_\lambda^{-1} D_\lambda u, D_\lambda u \rangle}{\|D_\lambda u\|^2} \|D_\lambda u\|^2 \end{aligned}$$

Hence

$$\max_{u \in M^\perp} |\langle M_n u, u \rangle| \leq \|D_\lambda^{-1} M_n D_\lambda^{-1}\|_\infty \max_{u \in M^\perp} \|D_\lambda u\|^2$$

Taking the min for all  $k - 1$ -dimensional subspaces  $M$  we get :

$$|l_k(M_n)| \leq \min_{\dim M = k-1} \max_{u \in M^\perp} |\langle M_n u, u \rangle| \leq \delta_k \|D_\lambda^{-1} M_n D_\lambda^{-1}\|_\infty$$

because  $\min_{\dim M = k-1} \max_{u \in M^\perp} \|D_\lambda u\|^2 = |\lambda_{k+1} - \lambda_k|$ . So that now we have

$$\begin{aligned} \mathbb{P}(\widehat{\lambda}_k \leq \lambda_k - \delta_k) &\leq \mathbb{P}(\|D_\lambda^{-1} M_n D_\lambda^{-1}\|_\infty \geq \delta_k / |\lambda_{k+1} - \lambda_k|) \\ &\leq \mathbb{P}(\|D_\lambda^{-1} M_n D_\lambda^{-1}\|_\infty \geq c) \end{aligned}$$

for some fixed  $c$ .

We can focus on

$$\|D_\lambda^{-1} M_n D_\lambda^{-1}\|_\infty^2 \leq \sum_{j,l} \frac{\langle (\Sigma_n - \Sigma) \varphi_j, \varphi_l \rangle^2}{|\lambda_{k+1} - \lambda_j| |\lambda_{k+1} - \lambda_l|}$$

It is proved in the previous Lemma (see equation (14) ) that

$$\mathbb{P} (\|D_\lambda^{-1} M_n D_\lambda^{-1}\|_\infty \geq c) \leq \exp(-cn/k^2).$$

This finishes the proof.

**Proof of Proposition 5:**

The outline of the proof is the following. We split the situation in two sets. We first choose an accurate  $\mathbf{l}_{k,n} = \frac{\mathbf{a}_k}{\sqrt{n}} \log n < 1/2$  and consider  $\mathcal{B}_k(\mathbf{l}_{k,n})$ . When  $\mathcal{B}_k(\mathbf{l}_{k,n})$  does not hold we can bound roughly  $\|\widehat{\mathbf{P}}_k - \mathbf{P}_k\|_\infty$  by  $2\mathbb{P}(\overline{\mathcal{B}}_k(\mathbf{l}_{k,n}))$  and use Lemma 13 above to get the exponential inequality

$$\mathbb{P}(\overline{\mathcal{B}}_k(\mathbf{l}_{k,n})) \leq 2 \exp\left(-\frac{\log^2 n}{2} \frac{1}{(2b-1) + \frac{256b^3 \log n}{\sqrt{n}}}\right) \leq 2 \exp(-c^\diamond \log^2 n)$$

where  $c^\diamond$  is a positive constant.

Now when  $\mathcal{B}_k(\mathbf{l}_{k,n})$  holds we take advantage of the linearization formula (12). Denote  $R(\zeta) = (\zeta I - \Sigma)^{-1}$  and  $\widehat{R}(\zeta) = (\zeta I - \widehat{\Sigma})^{-1}$ . It is easy to see that  $\widehat{R}(\zeta) = R^{1/2}(\zeta) T(\zeta) R^{1/2}(\zeta)$  with

$$T(\zeta) = (\zeta I - \Sigma)^{1/2} \widehat{R}(\zeta) (\zeta I - \Sigma)^{1/2},$$

hence  $T(\zeta) - I = T(\zeta) [R^{1/2}(\zeta) \Pi R^{1/2}(\zeta)]$  and  $T(\zeta) = [I - R^{1/2}(\zeta) \Pi R^{1/2}(\zeta)]^{-1}$  whenever  $\sup_{\zeta \in \partial\Omega_k} \|\widetilde{\Pi}(\zeta)\|_\infty \leq \mathbf{l}_{k,n} < 1$  where  $\widetilde{\Pi}(\zeta) = R^{1/2}(\zeta) \Pi R^{1/2}(\zeta)$ . Then :

$$\widehat{R}(\zeta) - R(\zeta) = R^{1/2}(\zeta) [I - \widetilde{\Pi}(\zeta)]^{-1} \widetilde{\Pi}(\zeta) R^{1/2}(\zeta). \quad (15)$$

We underline here that at this point the contour considered in the lines below is given in Figure 2. As announced earlier :

$$\|(\widehat{\mathbf{P}}_k - \mathbf{P}_k) \mathbf{u}\| \leq 2 \|\mathbf{u}\| \mathbb{1}_{\overline{\mathcal{B}}_k(\mathbf{l}_{k,n})} + \left\| \frac{1}{2\pi\iota} \oint_{\partial\Omega_k} R^{1/2}(\zeta) [I - \widetilde{\Pi}(\zeta)]^{-1} \widetilde{\Pi}(\zeta) R^{1/2}(\zeta) \mathbf{u} d\zeta \right\| \mathbb{1}_{\mathcal{B}_k(\mathbf{l}_{k,n})}. \quad (16)$$

We focus on the second term in the left hand side of the above equation :

$$\begin{aligned} & \left\| \frac{1}{2\pi\iota} \oint_{\partial\Omega_k} R^{1/2}(\zeta) [I - \widetilde{\Pi}(\zeta)]^{-1} \widetilde{\Pi}(\zeta) R^{1/2}(\zeta) \mathbf{u} d\zeta \right\| \mathbb{1}_{\mathcal{B}_k(\mathbf{l}_{k,n})} \\ & \leq \frac{\mathbf{l}_{k,n}}{2\pi(1 - \mathbf{l}_{k,n})} \oint_{\partial\Omega_k} \|R^{1/2}(\zeta)\|_\infty \|R^{1/2}(\zeta) \mathbf{u}\| d\zeta. \end{aligned}$$

At this point we must consider different situations depending on the nature of  $\mathbf{u}$ . First notice that  $\|R^{1/2}(\zeta)\|_\infty = \sup_p |\zeta - \lambda_p|^{-1/2} \leq c' |\zeta - \lambda_k|^{-1/2}$  due to the design of the contour  $\partial\Omega_k$ . Taking expectation and applying Cauchy-Schwartz inequality to  $\mathbb{E}[\|\mathbf{u}\|^2 \mathbb{1}_{\overline{\mathcal{B}}_k(\mathbf{l}_{k,n})}]$  we get :

$$\begin{aligned} \mathbb{E} \left\| (\widehat{\mathbf{P}}_k - \mathbf{P}_k) \mathbf{u} \right\|^2 & \leq 4\mathbb{E}[\|\mathbf{u}\|^4]^{1/2} \mathbb{P}^{1/2}(\overline{\mathcal{B}}_k(\mathbf{l}_{k,n})) + 2 \left( \frac{\mathbf{l}_{k,n}}{2\pi(1 - \mathbf{l}_{k,n})} \right)^2 \mathbb{E} \left[ \oint_{\partial\Omega_k} |\zeta - \lambda_k|^{-1/2} \|R^{1/2}(\zeta) \mathbf{u}\| d\zeta \right]^2 \\ & \leq 4 \exp(-c \log^2 n) + \frac{2}{\pi^2} \mathbf{l}_{k,n}^2 \mathbb{E} \left[ \oint_{\partial\Omega_k} |\zeta - \lambda_k|^{-1/2} \|R^{1/2}(\zeta) \mathbf{u}\| d\zeta \right]^2, \quad (17) \end{aligned}$$

for some  $c$  and  $c'$ . Now we split the contour  $\partial\Omega_k$  and compute the integral in the brackets for each

circle  $\partial C_j$ . We have :

$$\begin{aligned} \oint_{\partial C_j} |\zeta - \lambda_j|^{-1/2} \left\| R^{1/2}(\zeta) \mathbf{u} \right\| d\zeta &= \oint_{\partial C_j} \sqrt{\sum_{p=1}^{+\infty} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{|\zeta - \lambda_j| |\zeta - \lambda_p|}} d\zeta \\ &\leq \int_0^{2\pi} \sqrt{\sum_{p=1}^{+\infty} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{\delta_j |\lambda_j - \lambda_p + \delta_j \varphi^{i\theta}|}} \delta_j d\theta \\ &= \sqrt{\delta_j} \int_0^{2\pi} \sqrt{\sum_{p=1}^{+\infty} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{\left| |\lambda_j - \lambda_p| - \delta_j \right|}} d\theta \leq C \sqrt{\sum_{p=1}^{+\infty} \frac{\delta_j \lambda_p}{|\lambda_j - \lambda_p|} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{\lambda_p}}, \end{aligned}$$

where we set abusively  $|\lambda_j - \lambda_p| = \delta_j$  when  $p = j$ . Plugging this bound into (17) we obtain :

$$\mathbb{E} \left\| \left( \widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 \leq 4 \exp(-c \log^2 n) + C \cdot \mathbf{I}_{k,n}^2 \mathbb{E} \left[ \sum_{j=1}^k \sqrt{\sum_{p=1}^{+\infty} \frac{\delta_j \lambda_p}{|\lambda_j - \lambda_p|} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{\lambda_p}} \right]^2. \quad (18)$$

A last effort is needed : we develop the square in the second right hand side term and use a simple Cauchy-Schwartz inequality to obtain successively :

$$\begin{aligned} \mathbb{E} \left[ \sum_{j=1}^k \sqrt{\sum_{p=1}^{+\infty} \frac{\delta_j \lambda_p}{|\lambda_j - \lambda_p|} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{\lambda_p}} \right]^2 &= \left[ \sum_{j=1}^k \sum_{p=1}^{+\infty} \frac{\delta_j \lambda_p}{|\lambda_j - \lambda_p|} \frac{\mathbb{E} \langle \mathbf{u}, \varphi_p \rangle^2}{\lambda_p} \right] \\ &\quad + \sum_{j \neq m=1}^k \delta_j \delta_m \mathbb{E} \sqrt{\sum_{p=1}^{+\infty} \frac{\lambda_p}{|\lambda_j - \lambda_p|} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{\lambda_p} \sum_{p=1}^{+\infty} \frac{\lambda_p}{|\lambda_m - \lambda_p|} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{\lambda_p}}, \end{aligned}$$

where

$$\begin{aligned} &\mathbb{E} \sqrt{\sum_{p=1}^{+\infty} \frac{\lambda_p}{|\lambda_j - \lambda_p|} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{\lambda_p} \sum_{p=1}^{+\infty} \frac{\lambda_p}{|\lambda_m - \lambda_p|} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{\lambda_p}} \\ &\leq \sqrt{\sum_{p=1}^{+\infty} \frac{\lambda_p}{|\lambda_j - \lambda_p|} \frac{\mathbb{E} \langle \mathbf{u}, \varphi_p \rangle^2}{\lambda_p} \sum_{p=1}^{+\infty} \frac{\lambda_p}{|\lambda_m - \lambda_p|} \frac{\mathbb{E} \langle \mathbf{u}, \varphi_p \rangle^2}{\lambda_p}} \\ &\leq C \sqrt{\sum_{p=1}^{+\infty} \frac{\lambda_p}{|\lambda_j - \lambda_p|} \sum_{p=1}^{+\infty} \frac{\lambda_p}{|\lambda_m - \lambda_p|}}, \end{aligned}$$

and it is reminded that  $\sup_p \mathbb{E} \langle \mathbf{u}, \varphi_p \rangle^2 / \lambda_p < +\infty$  if  $\mathbf{u}$  has the same distribution as  $X$ , and  $\sup_p \langle \mathbf{u}, \varphi_p \rangle^2 / \lambda_p < +\infty$  if  $\mathbf{u}$  is a non random point. Finally :

$$\mathbb{E} \left[ \sum_{j=1}^k \sqrt{\sum_{p=1}^{+\infty} \frac{\delta_j \lambda_p}{|\lambda_j - \lambda_p|} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{\lambda_p}} \right]^2 \leq \left( \sum_{j=1}^k \sqrt{j \delta_j} \right)^2$$

which can be plugged in (18) :

$$\mathbb{E} \left\| \left( \widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 \leq 4 \exp(-c \log^2 n) + C \cdot \mathbf{I}_{k,n}^2 \left( \sum_{j=1}^k \sqrt{j \delta_j} \right)^2.$$

Obviously this bound is of interest when  $\sum_{j=1}^k \sqrt{j \delta_j}$  is reasonably low (typically bounded uniformly with respect to  $k$  or  $O(\log k)$ ) which in turns occurs when  $\lambda_j = c \exp(-\alpha j)$  or when  $\lambda_j = c j^{-1-\alpha}$  with  $\alpha \geq 1$ . However when  $\lambda_j = c j^{-1-\alpha}$  with  $0 < \alpha < 1$  or when  $\lambda_j = c j^{-1} \log^{1+\alpha} j$   $\sum_{j=1}^k \sqrt{j \delta_j} \geq k^{(1-\alpha)/2}$  and gets large when  $k$  does the bound above loses some interest.

It is possible to circumvent this problem just by changing the contour used to define and compute the projectors. It turns out that this new contours yields a control of the norm  $\mathbb{E} \left\| \widehat{\mathbf{P}}_k - \mathbf{P}_k \right\|_{\infty}^2$ , which is a deeper result than the pointwise control above in the special situation :  $\lambda_j = cj^{-1-\alpha}$  with  $0 < \alpha$ .

Indeed consider the new contour  $\partial\Omega_k$  obtained as the boundary of the rectangle of the complex plane with left vertice at  $x = \lambda_k - \delta_k$ , right vertice at  $\lambda_1 + 1$  and horizontal vertices at  $y = \pm 1$ . This contour is drawn at Figure 1. The initial formula (12) remains unchanged changed to  $\widehat{\mathbf{P}}_k - \mathbf{P}_k = \frac{1}{2\pi i} \int_{\partial\Omega_k} \left[ (\zeta I - \widehat{\Sigma})^{-1} - (\zeta I - \Sigma)^{-1} \right] d\zeta$ . It is also obvious that all the preliminary computations carried out to control the probability of the events  $\mathcal{A}_k = \{\widehat{\lambda}_k \in \Omega_k\} \cap \{\widehat{\lambda}_{k+1} \notin \Omega_k\}$  hence  $\mathcal{B}_k(\mathbf{l}_{k,n})$  remain true when changing the contour. Consequently we can turn to the bound (17) where this time :

$$\oint_{\partial\Omega_k} |\zeta - \lambda_k|^{-1/2} \left\| R^{1/2}(\zeta) \mathbf{u} \right\| d\zeta \leq A + B \int_0^1 \sqrt{\sum_{p=1}^{+\infty} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{(y + \delta_k)(y + |\lambda_k - \lambda_p|)}} dy.$$

The number 1 in the integral of the right hand side is unimportant. We have :

$$\begin{aligned} \int_0^1 \sqrt{\sum_{p=1}^{+\infty} \frac{\langle u, \varphi_p \rangle^2}{(y + \delta_k)(y + |\lambda_k - \lambda_p|)}} dy &= \int_0^{1/\delta_k} \sqrt{\sum_{p=1}^{+\infty} \frac{\delta_k \langle \mathbf{u}, \varphi_p \rangle^2}{(s+1)(\delta_k s + |\lambda_k - \lambda_p|)}} ds \\ &\leq \int_0^1 \sqrt{\sum_{p=1}^{+\infty} \frac{\delta_k \langle \mathbf{u}, \varphi_p \rangle^2}{(\delta_k s + |\lambda_k - \lambda_p|)}} ds + \int_1^{1/\delta_k} \sqrt{\sum_{p=1}^{+\infty} \frac{\delta_k \langle \mathbf{u}, \varphi_p \rangle^2}{\delta_k s^2}} ds \\ &\leq \int_0^1 \sqrt{\sum_{p=1}^{+\infty} \frac{\delta_k \langle \mathbf{u}, \varphi_p \rangle^2}{|\lambda_k - \lambda_p|}} ds + \ln(1/\delta_k) \sqrt{\sum_{p=1}^{+\infty} \langle \mathbf{u}, \varphi_p \rangle^2} \\ &\leq 2 \ln(1/\delta_k) \sqrt{\sum_{p=1}^{+\infty} \langle \mathbf{u}, \varphi_p \rangle^2}, \end{aligned}$$

and (17) becomes :

$$\mathbb{E} \left\| \left( \widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 \leq 4 \exp(-c \log^2 n) + c \mathbf{l}_{k,n}^2 \ln^2(1/\delta_k) \mathbb{E} \|\mathbf{u}\|^2.$$

This time the bound obtained is of interest when the assumption in (7) holds. Then  $\ln(1/\delta_k) \leq c \ln k$ .

As a consequence of the line above and of (16) we see that for all nonrandom  $\mathbf{u}$  with norm 1,  $\left\| \left( \widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 \leq 2 \mathbf{l}_{\overline{\mathcal{B}}_k(\mathbf{l}_{k,n})} + c \mathbf{l}_{k,n}^2 \ln^2(1/\delta_k)$  hence that

$$\left\| \left( \widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \right\|_{\infty}^2 \leq 2 \mathbf{l}_{\overline{\mathcal{B}}_k(\mathbf{l}_{k,n})} + c \mathbf{l}_{k,n}^2 \ln^2(1/\delta_k).$$

This finishes the proof of Proposition 5.

### Proof of Corollary 2 :

First notice that  $\|(\widehat{\pi}_k - \pi_k)(\varphi_k)\|^2 \leq \|\widehat{\pi}_k - \pi_k\|_{\infty}^2$  where  $(\widehat{\pi}_k - \pi_k)(\varphi_k) = \langle \widehat{\varphi}_k, \varphi_k \rangle \widehat{\varphi}_k - \varphi_k$  but we also have  $\langle (\widehat{\pi}_k - \pi_k) \varphi_k, \varphi_k \rangle = \langle \widehat{\varphi}_k, \varphi_k \rangle^2 - 1$ . Now

$$\begin{aligned} \|\widehat{\varphi}_k - \varphi_k\| &\leq \|\widehat{\varphi}_k - \langle \widehat{\varphi}_k, \varphi_k \rangle \widehat{\varphi}_k\| + \|\langle \widehat{\varphi}_k, \varphi_k \rangle \widehat{\varphi}_k - \varphi_k\| \\ &\leq 1 - \langle \widehat{\varphi}_k, \varphi_k \rangle + \|\widehat{\pi}_k - \pi_k\|_{\infty} \end{aligned}$$

At last  $\|\widehat{\varphi}_k - \varphi_k\|^2 \leq 2(1 - \langle \widehat{\varphi}_k, \varphi_k \rangle)^2 + 2\|\widehat{\pi}_k - \pi_k\|_{\infty}^2$  and we obtain the desired bound.

### Proof of Proposition 3.



From all that was done above we restrict ourselves to proving the lower bound for  $(\widehat{\pi}_k - \pi_k)$  when  $\sup_{\zeta \in \mathcal{C}_k} \left\| (\zeta I - \Sigma)^{-1/2} (\widehat{\Sigma} - \Sigma) (\zeta I - \Sigma)^{-1/2} \right\|_\infty < a_k/\sqrt{n}$ . So that we get :

$$\begin{aligned} \widehat{\pi}_k - \pi_k &= S_k (\Sigma_n - \Sigma) \pi_k + \pi_k (\Sigma_n - \Sigma) S_k + T_{k,n} \\ T_{k,n} &= \frac{1}{2\pi i} \oint_{\mathcal{C}_k} R^{1/2}(\zeta) \left[ I - \widetilde{\Pi}(\zeta) \right]^{-1} \widetilde{\Pi}^2(\zeta) R^{1/2}(\zeta) d\zeta, \end{aligned} \quad (19)$$

with  $S_k = \sum_{j \neq k} \varphi_j \otimes \varphi_k / (\lambda_j - \lambda_k)$ . Below we just developed  $\widehat{\pi}_k - \pi_k$  up to the order two instead of the first order used to derive the bound in Theorem 5. The second order term is  $T_{k,n}$  with

$$\|T_{k,n}\|_\infty \leq \delta_k \frac{\left\| \widetilde{\Pi}(\zeta) \right\|_\infty^2}{1 - \left\| \widetilde{\Pi}(\zeta) \right\|_\infty} \sup_{\zeta \in \mathcal{C}_k} \|R(\zeta)\|_\infty = \frac{\left\| \widetilde{\Pi}(\zeta) \right\|_\infty^2}{1 - \left\| \widetilde{\Pi}(\zeta) \right\|_\infty} \leq \frac{a_k}{\sqrt{n}} \frac{\frac{a_k}{\sqrt{n}}}{1 - \frac{a_k}{\sqrt{n}}} = d_{k,n},$$

since we assume that  $\sup_{\zeta \in \mathcal{C}_k} \left\| \widetilde{\Pi}(\zeta) \right\|_\infty < a_k/\sqrt{n} < 1$ . We turn to the first order term namely  $S_k (\Sigma_n - \Sigma) \pi_k + \pi_k (\Sigma_n - \Sigma) S_k$ . It is simple to see that its norm is exactly

$$2 \|\pi_k (\Sigma_n - \Sigma) S_k\|_\infty = 2 \|S_k (\Sigma_n - \Sigma) \varphi_k\| = 2 \|S_k \Sigma_n \varphi_k\| = 2 \sqrt{\frac{1}{n} \left\| \sum_{i=1}^n \langle X_i, \varphi_k \rangle S_k X_i \right\|^2}.$$

Then from (19) we get

$$\begin{aligned} 2 \|S_k \Sigma_n \varphi_k\| &\leq \|\widehat{\pi}_k - \pi_k\|_\infty + d_{k,n}, \\ 2 \|S_k \Sigma_n \varphi_k\|^2 &\leq \|\widehat{\pi}_k - \pi_k\|_\infty^2 + d_{k,n}^2. \end{aligned}$$

Taking expectation we finally get

$$\mathbb{E} \|\widehat{\pi}_k - \pi_k\|_\infty^2 \geq 2 \mathbb{E} \|S_k \Sigma_n \varphi_k\|^2 - d_{k,n}^2.$$

Now from calculations similar to those carried out within the proof of Lemma 13 we get  $\mathbb{E} S_k \Sigma_n \varphi_k = 0$ ,  $\mathbb{E} \|S_k \Sigma_n \varphi_k\|^2 = \frac{1}{n} \lambda_k \sum_{j \neq k} \lambda_j / (\lambda_j - \lambda_k)^2$  hence

$$\mathbb{E} \|\widehat{\pi}_k - \pi_k\|_\infty^2 \geq \frac{2}{n} \lambda_k \sum_{j \neq k} \lambda_j / (\lambda_j - \lambda_k)^2 - d_{k,n}^2.$$

Besides when (7) holds  $a_k^2 \asymp \lambda_k \sum_{j \neq k} \lambda_j / (\lambda_j - \lambda_k)^2 \asymp k^2$ . This together with the fact that  $d_{k,n}^2$  is negligible with respect to  $a_k^2/n$  yields the desired result :

$$\mathbb{E} \|\widehat{\pi}_k - \pi_k\|_\infty^2 \geq c \frac{a_k^2}{n},$$

for some absolute constant  $c$ .

## 4.2 Derivation of results of Section 3.2

The next two lemmas aim at providing the reader with basic inequalities which will be extensively used in the sequel.

**Lemma 15** *We have for all  $h > 0$*

$$\begin{aligned} \underline{K} F_k(h) &\leq \mathbb{E} K(\| \mathbf{P}_k(X_i - x) \| / h) \leq \left[ c_F \int u^k |K'(u)| du \right] F_k(h) \leq \bar{c} F_k(h), \\ \underline{K} n F_k(h) &\leq \mathbb{E} S^* \leq \bar{c} n F_k(h). \end{aligned}$$

The proof of the Lemma is simple. The lower bound in the equation above is obtained from  $\mathbb{E}K(\|\mathbf{P}_k(X_i - x)\|/h) \geq \mathbb{E}K(\|\mathbf{P}_k(X_i - x)\|/h) 1_{\{\|\mathbf{P}_k(X_i - x)\| < h\}}$ . The upper bound is derived from Fubini's theorem

$$\begin{aligned} \mathbb{E}K(\|\mathbf{P}_k(X_i - x)\|/h) &= - \int K'(u) F_k(hu) du \\ &\leq F_k(h) \int |K'(u)| \frac{F_k(hu)}{F_k(h)} du \leq \left[ c_F \int u^p |K'(u)| du \right] F_k(h) \end{aligned}$$

We refer to Ferraty, Mas, Vieu (2007) and Biau, Mas (2012) where this method was already used in an asymptotic framework.

**Lemma 16** For all  $t > 0$ ,

$$\mathbb{P}\left(\left|\frac{S^* - \mathbb{E}S^*}{\mathbb{E}S^*}\right| > t\right) \leq 2 \exp - [nF_k(h) t^2 \underline{K}^2 / 2 |K|_\infty (\bar{c} + t\underline{K})].$$

The proof is a simple consequence of Bernstein's theorem, of the boundedness of  $K$  and of Lemma 15 hence omitted.

**Preliminary facts for the proof of Proposition 8 :**

We deal with  $\hat{r}(x) - r^*(x) = \frac{\hat{Z}}{S} - \frac{Z^*}{S^*}$ . It is simple to get the four following lines given here without proof for further purpose :

$$\begin{aligned} \hat{r}(x) - r^*(x) &= \hat{r}(x) \frac{S^* - \hat{S}}{S^*} - \frac{Z^* - \hat{Z}}{S^*}, \\ \max\{|\hat{r}(x)|, |r^*(x)|\} &\leq |r|_\infty + \sum_{i=1}^n |\varepsilon_i|, \\ \mathbb{E} \left[ \sum_{i=1}^n \varepsilon_i K(\|\mathbf{P}_k(X_i - x)\|/h) |X_1, \dots, X_n \right]^2 &= \sigma_\varepsilon^2 \sum_{i=1}^n K^2(\|\mathbf{P}_k(X_i - x)\|/h) \leq \sigma_\varepsilon^2 |K|_\infty S^*. \end{aligned}$$

We have

$$\hat{r}(x) - r^*(x) = [\hat{r}(x) - r^*(x)] \frac{S^* - \hat{S}}{S^*} + r^*(x) \frac{S^* - \hat{S}}{S^*} - \frac{Z^* - \hat{Z}}{S^*}. \quad (20)$$

It is straightforward to see that  $\mathcal{G} = \left\{ \left| \frac{S^* - \mathbb{E}S^*}{S^*} \right| < 1/4 \right\} = \left\{ 3/4 < \frac{\mathbb{E}S^*}{S^*} < 5/4 \right\} = \left\{ -1/5 < \frac{S^* - \mathbb{E}S^*}{\mathbb{E}S^*} < 1/3 \right\}$  and that  $\bar{\mathcal{G}} \subset \left\{ \left| \frac{S^* - \mathbb{E}S^*}{\mathbb{E}S^*} \right| > 1/5 \right\}$ . We split  $\hat{r}(x) - r^*(x)$  in two parts. We start with :

$$\begin{aligned} \mathbb{E} |\hat{r}(x) - r^*(x)|^2 1_{\bar{\mathcal{G}}} &\leq 2|r|_\infty^2 \mathbb{P}(\bar{\mathcal{G}}) + 2\mathbb{E} \left[ 1_{\bar{\mathcal{G}}} \left( \sum_{i=1}^n |\varepsilon_i| \right)^2 \right] \\ &= 2\mathbb{P}(\bar{\mathcal{G}}) \left[ |r|_\infty^2 + n\mathbb{E}\varepsilon^2 + n(n-1)(\mathbb{E}|\varepsilon|)^2 \right] \\ &\leq 2\mathbb{P} \left( \left| \frac{S^* - \mathbb{E}S^*}{\mathbb{E}S^*} \right| > 1/5 \right) \left[ |r|_\infty^2 + n\mathbb{E}\varepsilon^2 + n(n-1)(\mathbb{E}|\varepsilon|)^2 \right], \end{aligned} \quad (21)$$

where the last line is controlled by the exponential inequality in Lemma 16. When  $nF_k(h)/\log n$  is larger than a prescribed and universal constant we get  $n^2\mathbb{P} \left( \left| \frac{S^* - \mathbb{E}S^*}{\mathbb{E}S^*} \right| > 1/5 \right) \leq 1/n$  and the line above becomes negligible.

On the other hand it is simple to get :

$$\begin{aligned}
|\widehat{r}(x) - r^*(x)| 1_{\mathcal{G}} &\leq \frac{4}{3} \left| r^*(x) \frac{S^* - \widehat{S}}{S^*} - \frac{Z^* - \widehat{Z}}{S^*} \right| 1_{\mathcal{G}} \\
&\leq \frac{4}{3} |r^*(x)| \left| \frac{S^* - \widehat{S}}{S^*} \right| 1_{\mathcal{G}} + \frac{4}{3} \left| \frac{Z^* - \widehat{Z}}{S^*} \right| 1_{\mathcal{G}} \\
&\leq \frac{5}{3} |r^*(x)| \left| \frac{S^* - \widehat{S}}{\mathbb{E}S^*} \right| 1_{\mathcal{G}} + \frac{5}{3} \left| \frac{Z^* - \widehat{Z}}{\mathbb{E}S^*} \right| 1_{\mathcal{G}} \\
&\leq \frac{5}{3} |r^*(x)| \left| \frac{S^* - \widehat{S}}{\mathbb{E}S^*} \right| 1_{\mathcal{G}} + \frac{5}{3} \left| \frac{Z^* - \widehat{Z}}{\mathbb{E}S^*} \right| 1_{\mathcal{G}}. \tag{22}
\end{aligned}$$

In view of (22) and (21) the conclusion will follow from an appropriate control of both terms :

$\mathbb{E} \left| r^*(x) (S^* - \widehat{S}) \right|^2 1_{\mathcal{G}}$ ,  $\mathbb{E} \left| Z^* - \widehat{Z} \right|^2 1_{\mathcal{G}}$ . These are addressed in the forthcoming Lemmas. For further purpose we notice that

$$\mathbb{E} |\widehat{r}(x) - r^*(x)|^2 1_{\mathcal{G}} \leq \frac{50}{9K^2 n^2 F_k^2(h)} \left[ \mathbb{E} \left| r^*(x) (S^* - \widehat{S}) \right|^2 1_{\mathcal{G}} + \mathbb{E} \left| Z^* - \widehat{Z} \right|^2 1_{\mathcal{G}} \right]. \tag{23}$$

**Lemma 17**

$$\mathbb{E} \left[ \left| r^*(x) (S^* - \widehat{S}) \right|^2 \mathbb{1}_{\left\{ \left| \frac{S^* - \widehat{S}}{\mathbb{E}S^*} \right| \leq 1/3 \right\}} \middle| X_1, \dots, X_n \right] \leq |S^* - \widehat{S}|^2 \left[ 2|r|_{\infty}^2 + \frac{5}{4} \frac{\sigma_{\varepsilon}^2 |K|_{\infty}}{\mathbb{E}S^*} \right] \quad a.s.$$

**Proof :**

Keep in mind that  $\mathcal{G}$  holds :

$$\begin{aligned}
&\mathbb{E} \left[ \left| r^*(x) (S^* - \widehat{S}) \right|^2 \middle| X_1, \dots, X_n \right] \\
&\leq 2|r|_{\infty}^2 |S^* - \widehat{S}|^2 + 2 \left( \frac{S^* - \widehat{S}}{S^*} \right)^2 \mathbb{E} \left[ \left( \sum_{i=1}^n \varepsilon_i K(\|\mathbf{P}_k(X_i - x)\|/h) \right)^2 \middle| X_1, \dots, X_n \right] \\
&= 2 |S^* - \widehat{S}|^2 \left[ |r|_{\infty}^2 + \frac{\sigma_{\varepsilon}^2 |K|_{\infty}}{S^*} \right].
\end{aligned}$$

Taking into account that  $\mathcal{G}$  holds we get the announced result.

**Lemma 18** *We have*

$$\begin{aligned}
\mathbb{E} |S^* - \widehat{S}|^2 &\leq \mathbf{A} \cdot n \frac{k^4}{h^2} F_k^2(h) \log^2 n \log \left( \frac{h\sqrt{n}}{k^2} \right), \\
\mathbb{E} |Z^* - \widehat{Z}|^2 &\leq |r|_{\infty} \mathbf{B} \cdot n \frac{k^4}{h^2} F_k^2(h) \log^2 n \log \left( \frac{h\sqrt{n}}{k^2} \right),
\end{aligned}$$

where  $\mathbf{A}$  and  $\mathbf{B}$  are constants depending only on  $K$ .

**Proof :**

We deal first with  $S^* - \widehat{S}$  and limit ourselves to carrying out computations when  $K$  is the naive kernel defined by  $K(u) = \mathbb{1}_{0 \leq u \leq 1}$ . The interested reader will easily check that the derivation may be adapted to more general kernels under assumption **A2**. The method by Biau and Mas (2012) is not enough to get the bound announced and we have to follow another route with sharper bound :

$$\begin{aligned}
\mathbb{E} |S^* - \widehat{S}|^2 &= \mathbb{E} \left( \sum_{i=1}^n \mathbb{1}_{\|\mathbf{P}_k(X_i - x)\| \leq h} - \mathbb{1}_{\|\widehat{\mathbf{P}}_k(X_i - x)\| \leq h} \right)^2 \\
&= \mathbb{E} \left( \sum_{i=1}^n \mathbb{1}_{\|\mathbf{P}_k(X_i - x)\| \leq h, \|\widehat{\mathbf{P}}_k(X_i - x)\| > h} - \mathbb{1}_{\|\mathbf{P}_k(X_i - x)\| > h, \|\widehat{\mathbf{P}}_k(X_i - x)\| \leq h} \right)^2.
\end{aligned}$$

Then

$$\mathbb{E} \left| S^* - \widehat{S} \right|^2 \leq 2\mathbb{E} \left( \sum_{i=1}^n \mathbb{1}_{\|\mathbf{P}_k(X_i-x)\| \leq h, \|\widehat{\mathbf{P}}_k(X_i-x)\| > h} \right)^2 + 2\mathbb{E} \left( \sum_{i=1}^n \mathbb{1}_{\|\mathbf{P}_k(X_i-x)\| > h, \|\widehat{\mathbf{P}}_k(X_i-x)\| \leq h} \right)^2.$$

Let  $0 < \eta < h$  to be fixed later. We treat only the first series since our method applies to the other :

$$\sum_{i=1}^n \mathbb{1}_{\|\mathbf{P}_k(X_i-x)\| \leq h, \|\widehat{\mathbf{P}}_k(X_i-x)\| > h} \leq \sum_{i=1}^n \mathbb{1}_{h-\eta \leq \|\mathbf{P}_k(X_i-x)\| \leq h} + \sum_{i=1}^n \mathbb{1}_{\|(\mathbf{P}_k - \widehat{\mathbf{P}}_k)(X_i-x)\| > \eta, \|\mathbf{P}_k(X_i-x)\| \leq h}.$$

The distribution of the new first series above is Binomial hence  $\mathbb{E} \left( \sum_{i=1}^n \mathbb{1}_{h-\eta \leq \|\mathbf{P}_k(X_i-x)\| \leq h} \right)^2 = np_k + n^2 p_k^2$  with  $p_k = F_k(h) - F_k(h-\eta) \leq c\eta k F_k(h)/h$  and

$$\mathbb{E} \left( \sum_{i=1}^n \mathbb{1}_{h-\eta \leq \|\mathbf{P}_k(X_i-x)\| \leq h} \right)^2 \leq n^2 \eta^2 k^2 \frac{F_k^2(h)}{h^2}. \quad (24)$$

We turn to the other term. From

$$\begin{aligned} & \left\{ \left\| (\mathbf{P}_k - \widehat{\mathbf{P}}_k)(X_i-x) \right\| > \eta, \|\mathbf{P}_k(X_i-x)\| \leq h \right\} \\ & \subset \{ \|(X_i-x)\| > M \log n \} \cup \left\{ \left\| \mathbf{P}_k - \widehat{\mathbf{P}}_k \right\| > \frac{\eta}{M \log n}, \|\mathbf{P}_k(X_i-x)\| \leq h \right\}, \end{aligned}$$

and the fact that  $\mathbb{E} \exp(\lambda \|X\|) < +\infty$  for some  $\lambda$  we get that  $\mathbb{E} \left( \sum_{i=1}^n \mathbb{1}_{\|(X_i-x)\| > M \log n} \right)^2$  is bounded by a  $n^{-p}$  for some large  $p$  (depending on a good choice of  $M$ ). Then

$$\begin{aligned} \sum_{i=1}^n \mathbb{1}_{\|(\mathbf{P}_k - \widehat{\mathbf{P}}_k)(X_i-x)\| > \eta, \|\mathbf{P}_k(X_i-x)\| \leq h} & \leq \mathbb{1}_{\|(\mathbf{P}_k - \widehat{\mathbf{P}}_k)\| > \eta/M \log n} \sum_{i=1}^n \mathbb{1}_{\|\mathbf{P}_k(X_i-x)\| \leq h} \\ & = \mathbb{1}_{\|(\mathbf{P}_k - \widehat{\mathbf{P}}_k)\| > \eta/M \log n} \sum_{i=1}^n [\mathbb{1}_{\|\mathbf{P}_k(X_i-x)\| \leq h} - \mathbb{P}(\|\mathbf{P}_k(X_i-x)\| \leq h)] \\ & \quad + n \mathbb{1}_{\|(\mathbf{P}_k - \widehat{\mathbf{P}}_k)\| > \eta/M \log n} \mathbb{P}(\|\mathbf{P}_k(X_1-x)\| \leq h). \end{aligned}$$

Denote  $U_i = \mathbb{1}_{\|\mathbf{P}_k(X_i-x)\| \leq h} - \mathbb{P}(\|\mathbf{P}_k(X_i-x)\| \leq h)$ ,

$$\begin{aligned} & \left( \sum_{i=1}^n \mathbb{1}_{\|(\mathbf{P}_k - \widehat{\mathbf{P}}_k)(X_i-x)\| > \eta, \|\mathbf{P}_k(X_i-x)\| \leq h} \right)^2 \\ & \leq 2n^2 \mathbb{P}^2(\|\mathbf{P}_k(X_1-x)\| \leq h) \mathbb{P} \left( \left\| (\mathbf{P}_k - \widehat{\mathbf{P}}_k) \right\| > \eta/M \log n \right) \\ & \quad + 2\mathbb{E} \mathbb{1}_{\|(\mathbf{P}_k - \widehat{\mathbf{P}}_k)\| > \eta/M \log n} \left\{ \sum_{i=1}^n [\mathbb{1}_{\|\mathbf{P}_k(X_i-x)\| \leq h} - \mathbb{P}(\|\mathbf{P}_k(X_i-x)\| \leq h)] \right\}^2 \\ & \leq 2n^2 F^2(h) \exp(-n\eta^2/Mk^2 \log^2 n) + 2\mathbb{P}^{1/2} \left( \left\| \mathbf{P}_k - \widehat{\mathbf{P}}_k \right\| > \eta/M \log n \right) \mathbb{E}^{1/4} \left( \sum_{i=1}^n U_i \right)^4 \\ & \leq cn^2 F^2(h) \exp(-n\eta^2/Mk^2 \log^2 n), \end{aligned}$$

because the second term is negligible with respect to the first. The optimal  $\eta$  is chosen now so that :

$$\frac{n^2 \eta^2 k^2 F_k^2(h)}{h^2} \asymp n^2 F^2(h) \exp(-n\eta^2/Mk^2 \log^2 n),$$

which yields

$$\eta^* \asymp \frac{k}{\sqrt{n}} \log n \log^{1/2} \frac{h\sqrt{n}}{k^2}.$$

Plugging this optimal  $\eta^*$  in (24) gives the first result of the Lemma.

Recall that we left initially  $Z^* - \widehat{Z}$  to treat  $S^* - \widehat{S}$ . We have

$$\begin{aligned} \mathbb{E} \left| Z^* - \widehat{Z} \right|^2 &= \mathbb{E} \left\{ \sum_{i=1}^n r(X_i) \left[ K(\|\mathbf{P}_k(X_i - x)\|/h) - K(\|\widehat{\mathbf{P}}_k(X_i - x)\|/h) \right] \right\}^2 \\ &\quad + \sigma_\varepsilon^2 \sum_{i=1}^n \mathbb{E} \left[ K(\|\mathbf{P}_k(X_i - x)\|/h) - K(\|\widehat{\mathbf{P}}_k(X_i - x)\|/h) \right]^2. \end{aligned}$$

Focusing here again on the specific important case of a naive kernel we check that

$$\begin{aligned} &\sum_{i=1}^n \mathbb{E} \left[ K(\|\mathbf{P}_k(X_i - x)\|/h) - K(\|\widehat{\mathbf{P}}_k(X_i - x)\|/h) \right]^2 \\ &= n \mathbb{E} \left[ \mathbb{1}_{\|\mathbf{P}_k(X_1 - x)\| \leq h, \|\widehat{\mathbf{P}}_k(X_1 - x)\| > h} - \mathbb{1}_{\|\mathbf{P}_k(X_1 - x)\| > h, \|\widehat{\mathbf{P}}_k(X_1 - x)\| \leq h} \right]^2 \end{aligned}$$

may be treated with exactly the same tools as above so that we can upper bound it up to constants with  $n\eta k F_k(h)/h + nF(h) \exp(-n\eta^2/Mk^2 \log^2 n)$ . These terms are unimportant since they are negligible with respect to the contribution of the other part of the decomposition of  $Z^* - \widehat{Z}$  namely :

$$\begin{aligned} &\mathbb{E} \left\{ \sum_{i=1}^n r(X_i) \left[ K(\|\mathbf{P}_k(X_i - x)\|/h) - K(\|\widehat{\mathbf{P}}_k(X_i - x)\|/h) \right] \right\}^2 \\ &= \mathbb{E} \left\{ \sum_{i=1}^n r(X_i) \left[ \mathbb{1}_{\|\mathbf{P}_k(X_i - x)\| \leq h, \|\widehat{\mathbf{P}}_k(X_i - x)\| > h} - \mathbb{1}_{\|\mathbf{P}_k(X_i - x)\| > h, \|\widehat{\mathbf{P}}_k(X_i - x)\| \leq h} \right] \right\}^2 \\ &\leq 2 \mathbb{E} \left\{ \sum_{i=1}^n r(X_i) \mathbb{1}_{\|\mathbf{P}_k(X_i - x)\| \leq h, \|\widehat{\mathbf{P}}_k(X_i - x)\| > h} \right\}^2 + 2 \left\{ \sum_{i=1}^n r(X_i) \mathbb{1}_{\|\mathbf{P}_k(X_i - x)\| > h, \|\widehat{\mathbf{P}}_k(X_i - x)\| \leq h} \right\}^2 \\ &\leq 2|r|_\infty \left[ \sum_{i=1}^n \mathbb{1}_{\|\mathbf{P}_k(X_i - x)\| \leq h, \|\widehat{\mathbf{P}}_k(X_i - x)\| > h} \right]^2 + 2|r|_\infty \left[ \sum_{i=1}^n \mathbb{1}_{\|\mathbf{P}_k(X_i - x)\| > h, \|\widehat{\mathbf{P}}_k(X_i - x)\| \leq h} \right]^2, \end{aligned}$$

and we are faced with exactly the same computations as above for  $S^* - \widehat{S}$ .

### Proof of Proposition 8:

Combining (20) with Lemma 18 we get :

$$\begin{aligned} \mathbb{E} |\widehat{r}(x) - r^*(x)|^2 1_{\mathcal{G}} &\leq \frac{50}{9\underline{K}^2 n^2 F_k^2(h)} \left[ \mathbb{E} \left| r^*(x) (S^* - \widehat{S}) \right|^2 1_{\mathcal{G}} + \mathbb{E} \left| Z^* - \widehat{Z} \right|^2 1_{\mathcal{G}} \right] \\ &\leq \mathbf{C} \cdot \frac{k^4}{nh^2} \log^2 n \log \left( \frac{h\sqrt{n}}{k^2} \right). \end{aligned}$$

This finishes the proof of the proposition.

## References

- [1] Berlinet A., Thomas-Agnan C. (2004) : *Reproducing kernel Hilbert spaces in Probability and Statistics*, Kluwer Academic Publishers.
- [2] Biau G., Mas A. (2012) : PCA-Kernel estimation, *Statistics & Risk Modeling*, **29**, 19–46.
- [3] Blanchard G., Bousquet O., Zwald L. : Statistical Properties of Kernel Principal Component Analysis, *Mach. Learn.*, **66**, 259-294, 2007.
- [4] Cupidon J., Eubank R., Gilliam T., Ruymgaart F. (2007) : The Delta-method for analytic functions of random operators with application to functional data. *Bernoulli*, **13**, 1179-1194.

- [5] Dauxois J., Pousse A., 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.
- [6] Dunford N., Schwartz, J.T. (1988) : *Linear Operators, Vol. I & II*. Wiley Classics Library.
- [7] Ferraty F., Laksaci A., Tadj A., Vieu P., (2010) : Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, **140**, n° 2, 335–352.
- [8] Ferraty F., Mas A., Vieu P. (2007). Advances on nonparametric regression from functional data, *Aust. New Zealand J. Stat.*, **49** n°3, 1-20.
- [9] Ferraty F., Vieu P. (2004) Nonparametric models for functional data, with applications in regression, time series prediction and curve discrimination. *J. Nonparametr. Stat.*, **16**, 111-125.
- [10] Ferraty F., Vieu P. (2006) *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, Berlin.
- [11] Gohberg I., Goldberg S., Kaashoek, M.A. (1991) : *Classes of Linear Operators Vol I & II*, Operator Theory : Advances and Applications, Birkhäuser.
- [12] Hall P., Hosseini-Nasab M. (2006) :On properties of functional principal components analysis, *Journal of the Royal Statistical Society, Series B*, **68**, n°1, 109-126.
- [13] Kato T. (1976) : *Perturbation theory for linear operators*. Grundlehren der mathematischen Wissenschaften. 132. Berlin -Heidelberg - New York: Springer-Verlag. 2nd Ed.
- [14] Koltchinskii V. (1998) : Asymptotics of spectral projections of some random matrices approximating integral operators. In Ernst Eberlein, Marjorie Hahn, and Michel Talagrand, editors, High Dimensional Probability, volume 43 of Progress in Probability, pages 191–227. Birkhäuser.
- [15] Ledoux M., Talagrand M. (1991), *Probability in Banach Spaces*, Springer-Verlag.
- [16] Mas A. (2012) : Lower bound in regression for functional data by small ball probability representation in Hilbert space, *Electron. J. Statist*, **6**, 1745-1778.
- [17] Mas A., Menneteau L. (2003) : Perturbation approach applied to the asymptotic study of random operators, *Progress in Probability*, **55**, 127-134, Birkhäuser.
- [18] Masry, E. (2005) : Nonparametric regression estimation for dependent functional data: Asymptotic normality. *Stochastic Process. Appl.* **115**, 155–177.
- [19] Ruymgaart F., Yang, S. (1997) : Some applications of Watson’s perturbation approach to random matrices. *J. Multivar. An.*, **60**, 48-60.
- [20] Schmeidler W. (1965) : *Linear Operators in Hilbert Spaces*, Academic Press.
- [21] Tsybakov, A.B. (2004) *Introduction a l’estimation non-paramétrique*. Springer.
- [22] Vakhania, N.N., Tarieladze, V.I., Chobanyan, S.M.Vakhania, N.N., Tarieladze, V.I., Chobanyan, S.M. (1987) : *Probability Distributions on Banach Spaces*. Mathematics and its Applications. (D. Reidel Publishing).
- [23] Weidman J. (1980) : *Linear Operators in Hilbert Spaces*, Graduate Texts in Mathematics , Springer.