



**HAL**  
open science

## High Dimensional Principal Projections

André Mas, Frits Ruymgaart

► **To cite this version:**

André Mas, Frits Ruymgaart. High Dimensional Principal Projections. *Complex Analysis and Operator Theory*, 2015, 9 (1), pp.35 - 63. 10.1007/s11785-014-0371-5 . hal-00772880v2

**HAL Id: hal-00772880**

**<https://hal.science/hal-00772880v2>**

Submitted on 17 Jun 2018

**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  $(\hat{\varphi}_i)_i$  of the covariance operator of a sample of interest. Dimension reduction is obtained by projecting on the eigenspaces spanned by the  $\hat{\varphi}_i$ 's usually endowed with nice properties in terms 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 in a nonparametric functional estimation model.

*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 associated norm  $\|\cdot\|$  and let  $X_1, X_2, \dots, X_n$  be a sample of independent and identically distributed random elements with values in  $\mathbf{H}$ . 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,  $\Sigma_n$ , based on the sample are symmetric, positive, trace-class operators from  $\mathbf{H}$  to  $\mathbf{H}$  defined by :

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

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

where the tensor product between  $u$  and  $v$  in  $\mathbf{H}$  stands for the one-rank operator from  $\mathbf{H}$  to  $\mathbf{H}$  defined by  $(u \otimes v)(t) = \langle u, t \rangle v$ , for all  $t$  in  $\mathbf{H}$  and  $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ .

When  $X$  is centered -which will be assumed throughout for mathematical convenience-  $\mathbb{E}X = 0$ ,  $\Sigma = \mathbb{E}(X \otimes X)$  and  $\Sigma_n = \frac{1}{n} \sum_{k=1}^n X_k \otimes X_k$ . By  $(\lambda_k, \varphi_k)$  we denote the  $k^{\text{th}}$  eigenelement (eigenvalue and eigenvector) of  $\Sigma$ . The  $\lambda_k$ 's are positive and simple by assumption and hence

---

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

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

satisfy  $\lambda_1 > \lambda_2 > \dots > 0$ , and  $(\lambda_k)_{k \in \mathbb{N}} \in l_1$ . The eigenprojectors are of major importance in the sequel. Since the eigenvalues are of order one we may define  $\pi_i = \varphi_i \otimes \varphi_i$  the rank one projection operator onto the vector space spanned by  $\varphi_i$  and  $\mathbf{P}_k$  the projector onto  $\text{span}\{\varphi_1, \dots, \varphi_k\}$ . We add hats to denote the empirical versions of the eigenelements and we get  $\Sigma_n \widehat{\varphi}_i = \widehat{\lambda}_i \widehat{\varphi}_i$  for all  $i \in \mathbb{N}$ , with all  $\widehat{\lambda}_i = 0$  except for a finite number not exceeding  $n$ . We also set  $\widehat{\pi}_i = \widehat{\varphi}_i \otimes \widehat{\varphi}_i$  and  $\widehat{\mathbf{P}}_k$ .

Let  $\delta(\cdot, \cdot)$  be some distance, pseudo distance or dissimilarity measure for projectors. The initial purpose of this work is to provide sharp upper and lower non asymptotic bounds for  $\mathbb{E}\delta(\widehat{\pi}_i, \pi_i)$  and for  $\mathbb{E}\delta(\widehat{\mathbf{P}}_k, \mathbf{P}_k)$ . We intend to display the dependence in  $i$  and  $k$  in the previous expectations. We also expect these bounds to be adaptive with respect to the eigenvalues decay rate.

Secondly these results will be 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 of the sample covariance operators. For general theory with a special view towards applications in statistics we refer, for instance to Mas, Menneteau (2003), Cupidon et al. (2007), or Gilliam et al. (2009). The interested reader should also note that kernel PCA was also investigated with similar tools in Koltchinskii (1998) and in Blanchard, Bousquet and Zwald (2007) for instance. We make an intensive use of perturbation techniques in the proofs.

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 we can let  $k_n \rightarrow +\infty$  as  $n \rightarrow +\infty$ .

At this point let us briefly elaborate on one of the main theoretical results of the paper : the rate of convergence of the risk of the  $k^{\text{th}}$  sample eigenprojector (i.e the eigenprojector corresponding to the  $k^{\text{th}}$  largest eigenvalue). For each  $n$  and  $k$  an upper bound for this risk is given in Proposition 7. This upper bound is, moreover, shown to be essentially optimal apart from a logarithmic factor in Proposition 2. Apparently the upper bound for the risk depends directly on  $k$  but turns out to be independent of the particular decay of the eigenvalues of the population covariance operator as long as some mild conditions are fulfilled. This seems to be new and an improvement over similar results by Hall and Hosseini-Nasab (2005) and Zwald and Blanchard (2005). These authors show dependence on the inverse of the  $k^{\text{th}}$  spacing of the eigenvalues. Since these spacings can become very small, this dependence may have a serious negative effect on the convergence rate of the risk. The reason for the discrepancy between our result and that in Hall and Hosseini-Nasab (2005) probably lies in differences of approach. Our method of proof is based on an exact expansion of the empirical eigenprojector in terms of the theoretical covariance operator and the perturbation, exploiting the associated Neumann series. As has already been observed above the norm of the perturbation operator  $\Sigma_n - \Sigma$  should be sufficiently small for the expansion to be valid. Since this perturbation is random we need to single out the part of the sample space where this requirement is fulfilled. Although, indeed, the requirement for the  $k^{\text{th}}$  eigenprojection depends on the  $k^{\text{th}}$  spacing, very precise concentration inequalities allow us to get rid of this dependence in our final result.

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 Notations, elements of operator theory and functional calculus for operators

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 (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  $\mathbf{B}^* = \mathbf{H}$ .

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 orthonormal basis  $(e_k)_{k \in \mathbb{N}}$  of  $\mathbf{H}$ ,

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

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 either. 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.

Finally we briefly introduce some basic facts about functional calculus for operators since all this material makes the core of the mathematical techniques involved in the proofs. The notion of functional calculus for operators is twofold. We mention the book by A.V. Skorohod (1984) dealing with random spectral measures and leading to a specific theory. Here by functional calculus we will strictly allude to perturbation theory and the functions  $f(T)$  we consider have to be analytic in a neighborhood of the spectrum of  $T$ .

We refer the reader to Kato (1976), Weidman (1980), Dunford and Schwartz (1988), Gohberg, Goldberg and Kaashoek (1991) for a 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  $\partial\Lambda_k$  be the rectangular contour of the complex plane defined at Figure 1. And 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  $\delta_j = \frac{1}{2} \min(\lambda_j - \lambda_{j+1}, \lambda_{j-1} - \lambda_j)$ .

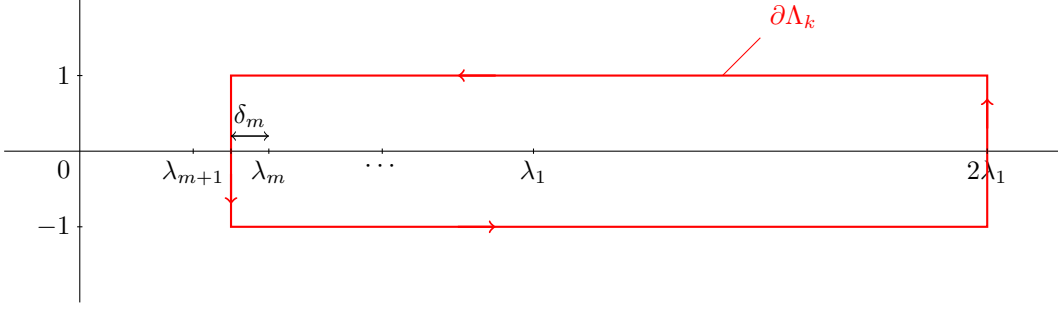


Figure 1: Rectangular contour

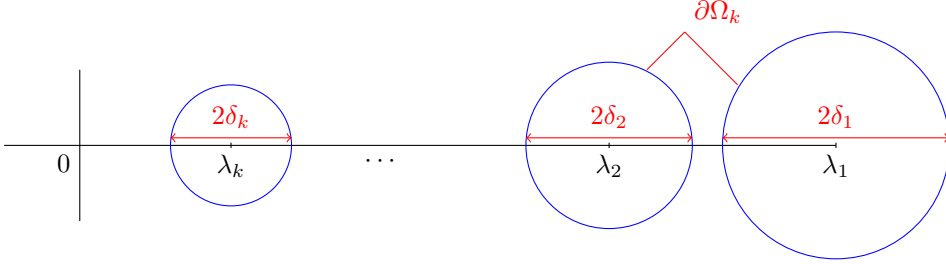


Figure 2: Contour made of disjoint circles

Results from perturbation theory yield :

$$\begin{aligned} \mathbf{P}_k &= \frac{1}{2\pi\iota} \int_{\partial\Omega_k} (\zeta I - \Sigma)^{-1} d\zeta = \frac{1}{2\pi\iota} \sum_{j=1}^k \int_{\partial\mathcal{C}_j} (\zeta I - \Sigma)^{-1} d\zeta = \sum_{j=1}^k \pi_j, \\ \mathbf{P}_k &= \frac{1}{2\pi\iota} \int_{\partial\Lambda_k} (\zeta I - \Sigma)^{-1} d\zeta, \end{aligned} \quad (3)$$

where  $\iota^2 = -1$  and  $\pi_j$  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 (3) involves a random contour, say,  $\partial\hat{\Omega}_k$  (resp.  $\partial\hat{\Lambda}_k$ ) similar in shape to  $\partial\Omega_k$  (resp.  $\partial\Lambda_k$ ) which contains the  $k$  first eigenvalues of  $\Sigma_n$ . In fact only the left edge of  $\partial\hat{\Lambda}_k$  may be considered as random :

$$\hat{\mathbf{P}}_k = \frac{1}{2\pi\iota} \int_{\partial\hat{\Omega}_k} (\zeta I - \Sigma_n)^{-1} d\zeta, \quad (4)$$

and  $\hat{\mathbf{P}}_k$  as defined in (4) equals the sum of all eigenprojections of  $\Sigma_n$  with eigenvalues inside  $\partial\hat{\Omega}_k$  each denoted  $\hat{\pi}_j$ .

### 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 $X$

In this 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 :

$$X =_d \sum_{j=1}^{+\infty} \sqrt{\lambda_j} \eta_j \varphi_j(\cdot) \quad (5)$$

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).

Consider  $X$  a centered random function with Karhunen-Loève development (5). 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 need higher moment assumptions because we need to apply Bernstein's exponential inequality to functionals of  $\Sigma_n - \Sigma$ . 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.

**Remark 1** *Assuming the independence of the  $\eta_i$ 's is a rather strong assumption. It stems here from the use of well-known exponential inequalities. The computations may be possible however by replacing independence by cross moments conditions or by invoking exponential bounds for mixing sequences. These issues go beyond the scope of the article.*

### 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 deduced 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  $\lambda_j = c j^{-1-\alpha} (\log j)^{-\beta}$  or even  $\lambda_j = c j^{-1} (\log j)^{-1-\beta}$  with  $\beta > 0$  is straightforward and will be omitted.

## 3.1 Eigenprojectors

As announced in the Introduction we aim at estimating the risk for principal projectors and recall that in our framework  $\hat{\pi}_k$  is associated to the  $k^{\text{th}}$  principal space of the functional PCA and defined by  $\hat{\pi}_k = \hat{\varphi}_k \otimes \hat{\varphi}_k$  where  $\hat{\varphi}_k$  is the  $k$ -th empirical eigenvector associated with the  $k$ -th empirical eigenvalue.

Introducing risk or loss functions for projectors should be carried out with caution. Indeed proximity between projection operators is intimately connected with proximity between subspaces which in turn may be slightly more intricate to measure than for vectors or even functions. To briefly illustrate these facts consider a fixed basis in  $\mathbf{H}$  say  $(\varphi_j)_{j \in \mathbb{N}}$  and consider  $\mathbf{P}_k$  that projects on  $\text{span}\{\varphi_j : 1 \leq j \leq k\}$ . Obviously for all  $x$  in  $\mathbf{H}$  we have  $\mathbf{P}_k x \rightarrow x$  when  $k$  increases to infinity but  $\|\mathbf{P}_k - \mathbf{I}\|_\infty = 1$  for all  $k$ . The same arguments holds when comparing  $\mathbf{P}_k$  and  $\mathbf{P}_{k+1}$  for large  $k$ . Consequently the choice of a proximity measure between  $\hat{\pi}_k$  and  $\pi_k$  for instance may have a major influence on the final rates. We provide below a lower bound and several upper bounds for the differences  $\hat{\pi}_k - \pi_k$  and  $\hat{\mathbf{P}}_k - \mathbf{P}_k$  and investigate the usual sup norm and evaluation semi-norms at fixed or random points. In the following a series plays a crucial role. Let :

$$\mathbf{a}_k = \left[ \sum_{i \neq k} \frac{\lambda_i}{|\lambda_i - \lambda_k|} + \frac{\lambda_k}{\delta_k} \right].$$

Our main results may be generally stated by means of  $\mathbf{a}_k$  and with no specific assumption on the eigenvalues. We will prove in Lemma 15 that under (7) or (8),  $\mathbf{a}_k \leq ck$  for some positive constant  $c$ . This will provide a simpler statement of the theorems. We start with a lower bound for  $\hat{\pi}_k$ .

**Proposition 2** *The following lower bound holds :*

$$\mathbb{E} \|\widehat{\pi}_k - \pi_k\|_\infty^2 \geq \frac{1}{2n} \sum_{j \neq k} \lambda_k \lambda_j / (\lambda_j - \lambda_k)^2 - 4 \frac{\mathbf{a}_k^4}{n^2}.$$

When (7) or (8) holds it becomes :

$$\mathbb{E} \|\widehat{\pi}_k - \pi_k\|_\infty^2 \geq c_3 \frac{k^2}{n}.$$

We notice that the lower bound seems to be non-sensitive to the eigenvalue decay rate.

**Remark 3** *The proof of Proposition 2 is strongly connected to functional calculus techniques for operators. We write  $\widehat{\pi}_k = f_k(\Sigma_n)$  and  $\pi_k = f_k(\Sigma)$ . Then considering  $\Sigma_n$  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*

$$f_k(\Sigma_n) - f_k(\Sigma) = g_k^1(\Sigma_n - \Sigma) + g_k^2(\Sigma_n - \Sigma, \Sigma_n - \Sigma).$$

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

**Remark 4** *Dauvois, Pousse and Romain (1982) in their pioneering work derived the asymptotic distribution of projectors such as  $\widehat{\pi}_k$ . The covariance operator of the Gaussian limiting distribution appears at page 145 and is detailed on page 146 of their article. It is intimately connected with the first order term of our lower bound given above.*

Now we are ready to turn to the issue of the upper bound. We start with 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  $\sup_i |\langle \mathbf{u}, \varphi_i \rangle| / \lambda_i < 1$  then for all  $n \geq 2$  and  $k \geq 2$ :*

$$\mathbb{E} \left\| \left( \widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 \leq \kappa c_0 \exp \left( -c_1 \frac{n}{\mathbf{a}_k^2} \right) + c_3 \cdot \frac{\mathbf{a}_k^2}{n} \log^2 n \cdot \left( \sum_{j=1}^k \sqrt{j \delta_j} \right)^2$$

where  $\kappa = (2b - 1) \sum_i \lambda_i^2 + (\text{tr} \Gamma)^2$ . Then we can derive :

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

**Remark 6** *It is important to notice that we could not assess Theorem 5 in sup norm in Cases 1 and 2, 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  $\sup_i |\langle \mathbf{u}, \varphi_i \rangle| / \lambda_i < 1$  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  $\widehat{\pi}_k$ . However the forthcoming Proposition 7 overcomes this problem when  $\widehat{\mathbf{P}}_k$  is replaced by  $\widehat{\pi}_k$ .*

The two next results may be viewed as consequences of Theorem 5. Their proofs are consequently omitted.

**Proposition 7** *Let (6) hold. Then :*

$$\mathbb{E} \|\widehat{\pi}_k - \pi_k\|_\infty^2 \leq c'_0 \exp\left(-c'_1 \frac{n}{\mathbf{a}_k^2}\right) + c'_3 \cdot \frac{\mathbf{a}_k^2}{n} \log^2 n.$$

When (7) or (8) hold this bound becomes :

$$\mathbb{E} \|\widehat{\pi}_k - \pi_k\|_\infty^2 \leq \exp(-c'_1 \log^2 n) + \frac{c'_2 k^2 \log^2 n}{n}$$

This latter bound is rather sharp when compared with the lower bound derived in Proposition 2 at least in our framework and with our set of assumptions. It differs from a  $\log^2 n$  term.

We turn to the eigenvectors in the next Corollary and recall that  $\widehat{\pi}_k = \widehat{\varphi}_k \otimes \widehat{\varphi}_k$  where  $\widehat{\varphi}_k$  is determined up to a sign. In order to fix this point  $\widehat{\varphi}_k$  must be chosen so that the sign of  $\langle \widehat{\varphi}_k, \varphi_k \rangle$  is positive (see Hall and Hosseini-Nasab (2006)).

**Corollary 8** *Select  $\widehat{\varphi}_k$  so that the sign of  $\langle \widehat{\varphi}_k, \varphi_k \rangle$  is positive the under the assumptions of Proposition 7 :*

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

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  $k \rightarrow +\infty$ . In the framework of Kernel PCA, Bousquet, Blanchard and Zwald (2007) obtain similar results involving  $\delta_k^{-1}$  in their Theorem 3.7. It would be an interesting issue to know if our approach extends to Kernel PCA but it is out of the scope of the current paper. Their bounds however were somewhat surprising in the following sense : if the eigenvalues decay quickly to zero so do the  $\delta_k$ 's and this means that the data are close to a finite dimensional sample. Then we would expect a rate always closer to  $\sqrt{n}$  for the convergence rate of these eigenprojectors. Yet these bound based on  $\delta_k^{-1}$  get always worse. In the case of situation (8), resp. (7) these bounds are  $O(e^k/\sqrt{n})$  resp.  $O(k^{2+\alpha}/\sqrt{n})$ . This problem does not remain here, at least under the assumptions we consider for the  $\lambda_k$ 's.

**Remark 9** *The method of proof adapts to projections on any set of eigenvectors chosen amongst the  $k$  first. This will not significantly improve the rate of convergence however. This fact, underlined by a referee, paves the way to introducing some kind of high-pass filters (with an obvious Fourier analogy) by projecting on  $(\varphi_m, \dots, \varphi_k)$  with  $m = \lfloor \alpha k \rfloor$ ,  $\alpha \in (1/2, 1)$  for instance. This may be of interest when studying spectrometric curves for which the high-frequency components are highly informative.*

### 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 in 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  $\alpha$  -which may be seen as the ultimate side effect of the curse of dimensionality- as shown in Mas (2012). This paves the way to different strategies for nonlinear models involving functional data. Such alternatives are additive models, multiple indices models or mixed approach such as projection pursuit. Several authors



studied these variants of the general regression : Amato, Antoniadis, De Feis (2006) and Ferraty (2011) for the single index model, Chen, Hall and Müller (2011) for the multiple index model and Ferraty, Goïa, Salinelli, Vieu (2012) for the Functional Projection Pursuit Regression amongst others.

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  $\hat{\mathbf{P}}_k$  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 :

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

This is not really fair to call  $\hat{r}_{NW}(x)$  above an estimate of  $r(x)$  since the norm  $\|\cdot\|$  cannot be practically calculated and should be numerically approximated for instance when  $\mathbf{H} = L^2([0, 1])$ . By replacing the theoretical norm on the space  $\mathbf{H}$  by the 'projection' semi-norm  $\|\cdot\|_k$  defined by  $\|u\|_k = \left\| \hat{\mathbf{P}}_k u \right\|_{\mathbb{R}^k}$  the estimate becomes :

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

This proposal is intuitive and rather simple to implement through standard statistical packages from most softwares. The problems arise from the dependence between  $X_i$  and  $\hat{\mathbf{P}}_k$  : the numerator and denominator of  $\hat{r}(x)$  are not sums of independent random variables and the dependence structure is, in our opinion rather confusing.

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 remain below the risk for  $\hat{r}(x)$ . In order to investigate  $\hat{r}(x)$  Biau and Mas (2012) 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 have to adapt their method here because the dimension  $k$  is not fixed. 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.

**Remark 10** *A more classical route would be to investigate the risk for the estimate  $\hat{r}(x)$  and by the way to prove the convergence of  $\hat{r}(x)$ . Considering (9) circumvents this difficult issue which is beyond the scope of the paper.*

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  $p$  appears 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  $p$  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 11** *The following non asymptotic bound holds under the assumptions (6), (8) or (7) with  $\alpha > 1$  and [A1-3] above :*

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

A first difference appears with Biau and Mas (2012). The bound of Proposition 11 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  $k^4 \log k$  is new. It accounts for the price to pay to let the dimension increase with the sample size.

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  $\tau_n^*(p, k) = [n \cdot (h^*)^k]^{-1}$ .

The next Proposition develops the comparison between  $\mathbb{E} [\hat{r}(x) - r^*(x)]^2$  and  $\tau_n^*(p, k)$ .

**Proposition 12** *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 :*

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

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

The proof of Proposition 12 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 13** *The statement of Proposition 12 allows independent choices of  $h$  (tending to 0) and  $k$  (going to infinity) by comparing  $\mathbb{E} [\hat{r}(x) - r^*(x)]^2$  only with the variance part of the estimate in the new projected model  $Y = r \circ \mathbf{P}_k(x) + \varepsilon$ .*

**Remark 14** *As pointed out by a referee our results remain true when  $Y$  is functional-valued, say  $Y \in \mathcal{H}$  where  $\mathcal{H}$  is some Hilbert space. Obviously the mean squares such as  $\mathbb{E} [\hat{r}(x) - r^*(x)]^2$  should then be replaced by  $\mathbb{E} \|\hat{r}(x) - r^*(x)\|_{\mathcal{H}}^2$  where  $\|\cdot\|_{\mathcal{H}}$  is the norm on  $\mathcal{H}$ .*

## 4 Mathematical derivations

We introduce the following events for later use. Let  $0 < c_0 < 1/2$

$$\mathcal{A}_k = \left\{ \widehat{\lambda}_k \in \Omega_k \right\} \cap \left\{ \widehat{\lambda}_{k+1} \notin \Omega_k \right\},$$

$$\mathcal{B}_k(c_0) = \left\{ \sup_{\zeta \in \partial\Omega_k} \left\| (\zeta I - \Sigma)^{-1/2} (\Sigma_n - \Sigma) (\zeta I - \Sigma)^{-1/2} \right\|_\infty < c_0 \right\}.$$

The event  $\mathcal{B}_k(c_0)$  will be introduced within the proof of our main result. Lemma 18 will detail the connection between  $\mathcal{B}_k(c_0)$  and  $\mathcal{A}_k$ . When  $\mathcal{A}_k$  holds the  $k$  first empirical eigenvalues (and only the  $k$  first) are inside  $\Omega_k$ . Then subject to these events we can derive the linearization :

$$\widehat{\mathbf{P}}_k - \mathbf{P}_k = \frac{1}{2\pi i} \int_{\partial\Omega_k} \left[ (\zeta I - \Sigma_n)^{-1} - (\zeta I - \Sigma)^{-1} \right] d\zeta. \quad (11)$$

The formula above comes down to removing the random contour  $\partial\widehat{\Omega}_k$  in the definition of  $\widehat{\mathbf{P}}_k$  at (4). 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. Thanks to lemma 18 we will restrict ourselves to bounding  $\mathbb{P}(\overline{\mathcal{B}}_k(c_0))$ .

We also assume in all the sequel that  $\mathbf{a}_k \log n / \sqrt{n} < 1/2$ .

### 4.1 Proof of Theorem 5

The proof takes two steps. First we apply the linearization formula (11) above but we have to control  $\mathbb{P}(\overline{\mathcal{A}}_k)$  non asymptotically which is achieved at Lemma 16 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 15** *When (7) or (8) hold there exists a constant  $c_\alpha$  such that for all  $k \geq 1$*

$$\sum_{i \neq k} \frac{\lambda_i}{|\lambda_i - \lambda_k|} + \frac{\lambda_k}{\delta_k} \leq c_\alpha k. \quad (12)$$

The proof stems from classical Riemann approximations. 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 16** *Denote  $\mathbf{a}_k = \left[ \sum_{i \neq k} \frac{\lambda_i}{|\lambda_i - \lambda_k|} + \frac{\lambda_k}{\delta_k} \right]$  We have :*

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

**Remark 17** *Earlier work focus on a probabilistic control of  $\Sigma_n - \Sigma$  to obtain bounds on the projectors. Here the idea consists in studying a normalized -through perturbation and resolvent-variant of the latter to derive a more precise result.*

**Proof of Lemma 16 :** 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  $\mu_k = \lambda_k - \delta_k$   $\Pi = \Sigma_n - \Sigma$  and  $\widetilde{\Pi}(\mu_k) = (\mu_k I - \Sigma)^{-1/2} (\Sigma_n - \Sigma) (\mu_k I - \Sigma)^{-1/2}$  and rewrite :

$$(\zeta I - \Sigma)^{-1/2} (\Sigma_n - \Sigma) (\zeta I - \Sigma)^{-1/2} = G_k(\zeta) \widetilde{\Pi}(\mu_k) G_k(\zeta),$$

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

$$\begin{aligned} & \sup_{\zeta \in \partial\Omega_k} \left\| (\zeta I - \Sigma)^{-1/2} (\Sigma_n - \Sigma) (\zeta I - \Sigma)^{-1/2} \right\|_\infty \\ & \leq \sup_{\zeta \in \partial\Omega_k} \|G_k(\zeta)\|_\infty^2 \cdot \left\| \widetilde{\Pi}(\mu_k) \right\|_\infty \leq 9 \left\| \widetilde{\Pi}(\mu_k) \right\|_\infty \end{aligned}$$

since  $\sup_k \sup_{\zeta \in \partial\Omega_k} \|G_k(\zeta)\|_\infty \leq 3$  whatever the contour chosen. Then :

$$\begin{aligned}
\left\| \tilde{\Pi}(\mu_k) \right\|_\infty^2 &\leq \left\| \tilde{\Pi}(\mu_k) \right\|_2^2 \\
&= \sum_{p,q \neq k}^{+\infty} \frac{\langle (\Sigma_n - \Sigma)(\varphi_p), \varphi_q \rangle^2}{|\mu_k - \lambda_p| |\mu_k - \lambda_q|} + \frac{\langle (\Sigma_n - \Sigma)(\varphi_k), \varphi_k \rangle^2}{\delta_k^2} \\
&\quad + \sum_{p \neq k}^{+\infty} \frac{\langle (\Sigma_n - \Sigma)(\varphi_p), \varphi_k \rangle^2}{|\mu_k - \lambda_p| \delta_k} \\
&\leq 4 \sum_{p,q \neq k}^{+\infty} \frac{\langle (\Sigma_n - \Sigma)(\varphi_p), \varphi_q \rangle^2}{|\lambda_k - \lambda_p| |\lambda_k - \lambda_q|} + \frac{\langle (\Sigma_n - \Sigma)(\varphi_k), \varphi_k \rangle^2}{\delta_k^2} \\
&\quad + 2 \sum_{p \neq k}^{+\infty} \frac{\langle (\Sigma_n - \Sigma)(\varphi_p), \varphi_k \rangle^2}{|\lambda_k - \lambda_p| \delta_k},
\end{aligned} \tag{13}$$

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

Finally, ignoring the constants 4 and 2 in (13) and 9 that appeared above, our only task is to give an exponential bound for the probability :

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

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 :

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

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  $b_k = \mathbf{a}_k 256 b^3 / [(2b-1)n]$  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  $\|s_k \Sigma s_k\|_2^2 = \sum_{i=1}^{+\infty} \lambda_i^2 / |\lambda_k - \lambda_i|^2$ . Set abusively  $|\lambda_i - \lambda_k| = \delta_k$  when  $i = k$  along the next five lines then :

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

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 :

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

where we set abusively  $|\lambda_j - \lambda_p| = \delta_j$  when  $p = j$  for the sake of simplicity. 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 distincts 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 :

$$\mathbb{E} \|s_k X_1\|^{2m} \leq m! (2b)^m \mathbf{a}_k^m.$$

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

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

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 18** *For some  $c_0 < 1/2$  we have :*

$$\mathbb{P}(\overline{\mathcal{A}}_k) \leq 3\mathbb{P}(\overline{\mathcal{B}}_k(c_0)).$$

**Proof of Lemma 18 :**

The proof of the Lemma consists essentially in proving  $\overline{\mathcal{A}}_k \subset \overline{\mathcal{B}}_k^*(c_0)$  for some  $c_0$  where  $\overline{\mathcal{B}}_k^*(c_0)$  is a set containing at least  $\overline{\mathcal{B}}_k(c_0)$ . The proof is similar to Lemmas E1 and E2 pages 7 and 8 of the supplementary material of Hilgert, Mas, Verzelen (2013). The contour used by these authors differs slightly from  $\Omega_k$  (another circle was added to make sure that any  $\hat{\lambda}_j$   $j \leq k$  falls inside the contour). We recall that the choice of the contour does not change the projector whenever  $k$  eigenvalues are located inside (see for instance Lemma XVIII.2.31 p.2255 of the part III in Dunford, Schwartz (1988)).

**Proof of Theorem 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  $\left\| \widehat{\mathbf{P}}_k - \mathbf{P}_k \right\|_\infty$  by  $2\mathbb{P}(\overline{\mathcal{B}}_k(\mathbf{l}_{k,n}))$  and use Lemma 16 above to get the exponential inequality

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

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

Now when  $\mathcal{B}_k(\mathbf{l}_{k,n})$  holds we take advantage of the linearization formula (11). Recall that  $\Pi = \Sigma_n - \Sigma$  and  $\tilde{\Pi}(\mu_k) = (\mu_k I - \Sigma)^{-1/2} (\Sigma_n - \Sigma) (\mu_k I - \Sigma)^{-1/2}$ . Denote  $R(\zeta) = (\zeta I - \Sigma)^{-1}$  and  $\widehat{R}(\zeta) = (\zeta I - \Sigma_n)^{-1}$ . 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  $\sup_{\zeta \in \partial\Omega_k} \left\| \tilde{\Pi}(\zeta) \right\|_\infty \leq \mathbf{l}_{k,n} < 1$ . Then :

$$\widehat{R}(\zeta) - R(\zeta) = R^{1/2}(\zeta) \left[ I - \tilde{\Pi}(\zeta) \right]^{-1} \tilde{\Pi}(\zeta) R^{1/2}(\zeta). \quad (14)$$

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

$$\left\| \left( \widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\| \leq 2 \|\mathbf{u}\| \mathbb{1}_{\overline{\mathcal{B}}_k(\mathbf{l}_{k,n})} + \left\| \frac{1}{2\pi i} \oint_{\partial\Omega_k} R^{1/2}(\zeta) \left[ I - \tilde{\Pi}(\zeta) \right]^{-1} \tilde{\Pi}(\zeta) R^{1/2}(\zeta) \mathbf{u} d\zeta \right\| \mathbb{1}_{\mathcal{B}_k(\mathbf{l}_{k,n})}. \quad (15)$$

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

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

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

$$\begin{aligned} \mathbb{E} \left\| \left( \widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 & \leq 4 \mathbb{E} \left[ \|\mathbf{u}\|^4 \right]^{1/2} \mathbb{P}^{1/2}(\overline{\mathcal{B}}_k(\mathbf{l}_{k,n})) \\ & \quad + 2 \left( \frac{\mathbf{l}_{k,n}}{2\pi(1-\mathbf{l}_{k,n})} \right)^2 \mathbb{E} \left[ \oint_{\partial\Omega_k} |\zeta - \lambda_k|^{-1/2} \left\| R^{1/2}(\zeta) \mathbf{u} \right\| d\zeta \right]^2 \\ & \leq c \exp(-c' \log^2 n) + \frac{2}{\pi^2} \mathbf{l}_{k,n}^2 \mathbb{E} \left[ \oint_{\partial\Omega_k} |\zeta - \lambda_k|^{-1/2} \left\| R^{1/2}(\zeta) \mathbf{u} \right\| d\zeta \right]^2, \quad (16) \end{aligned}$$

for some  $c$  and  $c'$ . Now we split the contour  $\partial\Omega_k$  and compute the integral in the brackets for each circle  $\partial C_j$ . Consider two cases. First if  $j = k$  :

$$\begin{aligned} \mathbb{E} \oint_{\partial C_k} |\zeta - \lambda_k|^{-1/2} \left\| R^{1/2}(\zeta) \mathbf{u} \right\| d\zeta & = |\delta_k|^{-1/2} \oint_{\partial C_k} \mathbb{E} \sqrt{\sum_{p=1}^{+\infty} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{|\zeta - \lambda_p|}} d\zeta \\ & \leq |\delta_k|^{1/2} \sup_{\zeta \in \partial C_k} \sqrt{\sum_{p=1}^{+\infty} \frac{\lambda_p}{|\zeta - \lambda_p|}} \\ & \leq c |\delta_k|^{1/2} \sqrt{\sum_{p=1, p \neq k}^{+\infty} \frac{\lambda_p}{|\lambda_k - \lambda_p|} + \frac{\lambda_k}{\delta_k}} \leq c |k \delta_k|^{1/2} \end{aligned}$$

since  $\sup_i |\langle \mathbf{u}, \varphi_i \rangle|^2 / \lambda_i < 1$ . Now if  $j \neq k$  we have quite similarly :

$$\begin{aligned}
\mathbb{E} \oint_{\partial C_j} |\zeta - \lambda_k|^{-1/2} \left\| R^{1/2}(\zeta) \mathbf{u} \right\| d\zeta &\leq c \oint_{\partial C_j} |\lambda_j - \lambda_k|^{-1/2} \mathbb{E} \left\| R^{1/2}(\zeta) \mathbf{u} \right\| d\zeta \\
&\leq |\lambda_j - \lambda_k|^{-1/2} \oint_{\partial C_j} \mathbb{E} \sqrt{\sum_{p=1}^{+\infty} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{|\zeta - \lambda_p|}} d\zeta \\
&\leq |\lambda_j - \lambda_k|^{-1/2} \delta_j \mathbb{E} \sqrt{\sum_{p=1, p \neq j}^{+\infty} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{|\lambda_j - \lambda_p|} + \frac{\langle \mathbf{u}, \varphi_j \rangle^2}{\delta_j}} \\
&\leq |\lambda_j - \lambda_k|^{-1/2} \delta_j \sqrt{\sum_{p=1, p \neq j}^{+\infty} \frac{\lambda_p}{|\lambda_j - \lambda_p|} + \frac{\lambda_j}{\delta_j}} \\
&\leq |\lambda_j - \lambda_k|^{-1/2} \delta_j \sqrt{j} \leq \sqrt{j \delta_j}.
\end{aligned}$$

Plugging these bounds into (16) we obtain :

$$\mathbb{E} \left\| \left( \widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 \leq c \exp(-c' \log^2 n) + C \cdot \mathbf{l}_{k,n}^2 \left[ \sum_{j=1}^k \sqrt{j \delta_j} \right]^2. \quad (17)$$

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 :  $\lambda_j = c j^{-1-\alpha}$  with  $0 < \alpha$ .

Indeed consider the new contour  $\partial \Lambda_k$  obtained as the boundary of the rectangle of the complex plane with left edge at  $x = \lambda_k - \delta_k$ , right edge at  $\lambda_1 + 1$  and horizontal edge at  $y = \pm 1$ . This contour is drawn at Figure 1. The initial formula (11) remains unchanged  $\widehat{\mathbf{P}}_k - \mathbf{P}_k = \frac{1}{2\pi i} \int_{\partial \Lambda_k} \left[ (\zeta I - \Sigma_n)^{-1} - (\zeta I - \Sigma)^{-1} \right] d\zeta$ . It is also obvious that all the preliminary computations carried out to control the probability of the events  $\mathcal{A}_k = \left\{ \widehat{\lambda}_k \in \Omega_k \right\} \cap \left\{ \widehat{\lambda}_{k+1} \notin \Omega_k \right\}$  hence  $\mathcal{B}_k(\mathbf{l}_{k,n})$  remain true when changing the contour. Consequently we can turn to the bound (16). Denote  $\partial \Lambda_k = \partial \Lambda_k^{h\pm} \cup \partial \Lambda_k^r \cup \partial \Lambda_k^l$  where  $\partial \Lambda_k^r$  denotes the right edge,  $\partial \Lambda_k^l$  the left edge and  $\partial \Lambda_k^{h\pm}$  the union of the two disjoint horizontal edges. It is plain that

$$\oint_{\partial \Lambda_k^{h\pm} \cup \partial \Lambda_k^r} |\zeta - \lambda_k|^{-1/2} \left\| R^{1/2}(\zeta) \mathbf{u} \right\| d\zeta \leq C$$

where  $C$  is some constant since  $|\zeta - \lambda_k|^{-1/2} \leq |1 - \lambda_k|^{-1/2}$  for  $\zeta \in \partial \Lambda_k^{h\pm} \cup \partial \Lambda_k^r$ . The remaining term is :

$$\begin{aligned}
\oint_{\partial \Lambda_k^l} |\zeta - \lambda_k|^{-1/2} \left\| R^{1/2}(\zeta) \mathbf{u} \right\| d\zeta &= \int_{-1}^1 \sqrt{\sum_{p=1}^{+\infty} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{|iy - \delta_k| (iy + |\lambda_k - \lambda_p| - \delta_k)}} dy \\
&\leq c \int_0^1 \sqrt{\sum_{p=1}^{+\infty} \frac{\langle \mathbf{u}, \varphi_p \rangle^2}{(y + \delta_k)(y + |\lambda_k - \lambda_p|)}} dy
\end{aligned}$$

where  $c$  is a universal constant and we used bounds like  $|y - \delta_k| \geq (y + \delta_k)/2$ . We have :

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

and (16) becomes :

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

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 (15) we see that for all nonrandom  $\mathbf{u}$  with norm 1,  $\left\| \left( \widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 \leq 2 \mathbb{1}_{\overline{\mathcal{B}}_k(\mathbf{1}_{k,n})} + c l_{k,n}^2 \ln^2(1/\delta_k)$  hence that

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

This finishes the proof of Theorem 5.

### Proof of Corollary 8 :

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

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

since  $\langle \widehat{\varphi}_k, \varphi_k \rangle \leq 1$ . 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 2.

From all that was done above we restrict ourselves to proving the lower bound for  $(\widehat{\pi}_k - \pi_k)$  when  $\sup_{\zeta \in \mathcal{C}_k} \left\| (\zeta I - \Sigma)^{-1/2} (\Sigma_n - \Sigma) (\zeta I - \Sigma)^{-1/2} \right\|_{\infty} < \mathbf{a}_k / \sqrt{n}$  that is when we can readily linearize  $\widehat{\pi}_k - \pi_k$  up to order 2 :

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

with  $T_{k,n} = \frac{1}{2\pi\iota} \oint_{\mathcal{C}_k} R^{1/2}(\zeta) \left[ I - \widetilde{\Pi}(\zeta) \right]^{-1} \widetilde{\Pi}^2(\zeta) R^{1/2}(\zeta) d\zeta$ . So that we get :

$$\widehat{\pi}_k - \pi_k = S_k (\Sigma_n - \Sigma) \pi_k + \pi_k (\Sigma_n - \Sigma) S_k + T_{k,n} \quad (18)$$



with  $S_k = \sum_{j \neq k} \varphi_j \otimes \varphi_j / (\lambda_j - \lambda_k)$ . The computation of  $S_k$  appears in several papers dealing with projectors or resolvent-based computations (e.g. p.347 in Cardot, Mas, Sarda (2007)). The second order term is  $T_{k,n}$  with :

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

since we assume that  $\sup_{\zeta \in \mathcal{C}_k} \left\| \tilde{\Pi}(\zeta) \right\|_\infty < \mathbf{a}_k / \sqrt{n} < 1/2$ . We turn to the first order term namely  $S_k (\Sigma_n - \Sigma) \pi_k + \pi_k (\Sigma_n - \Sigma) S_k$ . It is plain that :

$$\begin{aligned} \|S_k (\Sigma_n - \Sigma) \pi_k + \pi_k (\Sigma_n - \Sigma) S_k\|_\infty &\geq \|S_k (\Sigma_n - \Sigma) \pi_k \varphi_k + \pi_k (\Sigma_n - \Sigma) S_k \varphi_k\| \\ &= \|S_k (\Sigma_n - \Sigma) \pi_k \varphi_k\| = \|S_k \Sigma_n \varphi_k\| \end{aligned}$$

since  $\pi_k = \varphi_k \otimes \varphi_k$  and

$$\|S_k \Sigma_n \varphi_k\|^2 = \frac{1}{n^2} \left\| \sum_{i=1}^n \langle X_i, \varphi_k \rangle S_k X_i \right\|^2.$$

Then from (18) we get :

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

Taking expectation we finally get

$$\mathbb{E} \|\hat{\pi}_k - \pi_k\|_\infty^2 \geq (1/2) \mathbb{E} \|S_k \Sigma_n \varphi_k\|^2 - d_{k,n}^2.$$

Now from calculations similar to those carried out within the proof of Lemma 16 we get :

$$\begin{aligned} \mathbb{E} \|S_k \Sigma_n \varphi_k\|^2 &= \frac{1}{n^2} \mathbb{E} \left\| \sum_{i=1}^n S_k [\langle X_i, \varphi_k \rangle X_i] \right\|^2 \\ &= \frac{1}{n^2} \mathbb{E} \sum_{i,j=1}^n \langle X_i, \varphi_k \rangle \langle X_j, \varphi_k \rangle \langle S_k(X_i), S_k(X_j) \rangle = \\ &= \frac{1}{n} \mathbb{E} \langle X_1, \varphi_k \rangle^2 \|S_k(X_1)\|^2 + \frac{n-1}{2n} \mathbb{E} [\langle X_1, \varphi_k \rangle \langle X_2, \varphi_k \rangle \langle S_k(X_1), S_k(X_2) \rangle]. \end{aligned}$$

Conditioning by  $X_2$  and taking expectation with respect to  $X_1$  it is simple to see that the second term in the right hand side above is null. The first one is :

$$\frac{1}{n} \mathbb{E} \langle X_1, \varphi_k \rangle^2 \|S_k(X_1)\|^2 = \frac{\lambda_k}{n} \mathbb{E} \|S_k(X_1)\|^2 = \frac{1}{n} \lambda_k \sum_{j \neq k} \lambda_j / (\lambda_j - \lambda_k)^2$$

because  $S_k(X_1) = \varphi_k \sum_{j \neq k} \langle \varphi_j, X_1 \rangle / (\lambda_j - \lambda_k)$ . Hence :

$$\mathbb{E} \|\hat{\pi}_k - \pi_k\|_\infty^2 \geq \frac{1}{2n} \lambda_k \sum_{j \neq k} \lambda_j / (\lambda_j - \lambda_k)^2 - d_{k,n}^2.$$

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

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

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 19** *We have for all  $h > 0$*

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

The lower bound in the equation above is obtained from :

$$\mathbb{E}K(\|\mathbf{P}_k(X_i - x)\|/h) \geq \mathbb{E}K(\|\mathbf{P}_k(X_i - x)\|/h) \mathbb{1}_{\{\|\mathbf{P}_k(X_i - x)\| < h\}}.$$

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 20** *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 consequence of Bernstein's theorem, of the boundedness of  $K$  and of Lemma 19 hence omitted.

**Preliminary facts for the proof of Proposition 11 :**

We deal with  $\hat{r}(x) - r^*(x) = \frac{\hat{Z}}{\hat{S}} - \frac{Z^*}{S^*}$ . 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 :

$$\hat{r}(x) - r^*(x) = [\hat{r}(x) - r^*(x)] \frac{S^* - \hat{S}}{S^*} + r^*(x) \frac{S^* - \hat{S}}{S^*} - \frac{Z^* - \hat{Z}}{S^*}. \quad (19)$$

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 :

$$\begin{aligned} \mathbb{E} |\hat{r}(x) - r^*(x)|^2 \mathbb{1}_{\bar{\mathcal{G}}} &\leq 2|r|_\infty^2 \mathbb{P}(\bar{\mathcal{G}}) + 2\mathbb{E} \left[ \mathbb{1}_{\bar{\mathcal{G}}} \left( \sum_{i=1}^n |\varepsilon_i| \right)^2 \right] \\ &= 2\mathbb{P}(\bar{\mathcal{G}}) \left[ |r|_\infty^2 + n\mathbb{E}\varepsilon^2 + n(n-1)(\mathbb{E}|\varepsilon|)^2 \right] \\ &\leq 2\mathbb{P} \left( \left| \frac{S^* - \mathbb{E}S^*}{\mathbb{E}S^*} \right| > 1/5 \right) \left[ |r|_\infty^2 + n\mathbb{E}\varepsilon^2 + n(n-1)(\mathbb{E}|\varepsilon|)^2 \right], \end{aligned} \quad (20)$$

where the last line is controlled by the exponential inequality in Lemma 20. 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 :

$$\begin{aligned}
|\widehat{r}(x) - r^*(x)| \mathbb{1}_{\mathcal{G}} &\leq \frac{4}{3} \left| r^*(x) \frac{S^* - \widehat{S}}{S^*} - \frac{Z^* - \widehat{Z}}{S^*} \right| \mathbb{1}_{\mathcal{G}} \\
&\leq \frac{4}{3} |r^*(x)| \left| \frac{S^* - \widehat{S}}{S^*} \right| \mathbb{1}_{\mathcal{G}} + \frac{4}{3} \left| \frac{Z^* - \widehat{Z}}{S^*} \right| \mathbb{1}_{\mathcal{G}} \\
&\leq \frac{5}{3} |r^*(x)| \left| \frac{S^* - \widehat{S}}{\mathbb{E}S^*} \right| \mathbb{1}_{\mathcal{G}} + \frac{5}{3} \left| \frac{Z^* - \widehat{Z}}{\mathbb{E}S^*} \right| \mathbb{1}_{\mathcal{G}} \\
&\leq \frac{5}{3} |r^*(x)| \left| \frac{S^* - \widehat{S}}{\mathbb{E}S^*} \right| \mathbb{1}_{\mathcal{G}} + \frac{5}{3} \left| \frac{Z^* - \widehat{Z}}{\mathbb{E}S^*} \right| \mathbb{1}_{\mathcal{G}}. \tag{21}
\end{aligned}$$

In view of (21) and (20) the conclusion will follow from an appropriate control of both terms :

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

$$\mathbb{E} |\widehat{r}(x) - r^*(x)|^2 \mathbb{1}_{\mathcal{G}} \leq \frac{50}{9\underline{K}^2 n^2 F_k^2(h)} \left[ \mathbb{E} \left| r^*(x) (S^* - \widehat{S}) \right|^2 \mathbb{1}_{\mathcal{G}} + \mathbb{E} \left| Z^* - \widehat{Z} \right|^2 \mathbb{1}_{\mathcal{G}} \right]. \tag{22}$$

**Lemma 21**

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

**Proof :**

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

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

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

**Lemma 22** *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 sufficient to get the bound announced and we have to follow another route with sharper bound :

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

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

$$\mathbb{E} \left( \sum_{i=1}^n \mathbb{1}_{h-\eta \leq \|\mathbf{P}_k(X_i-x)\| \leq h} \right)^2 \leq n^2 \eta^2 k^2 \frac{F_k^2(h)}{h^2}. \quad (23)$$

We turn to the other term. From :

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

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

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

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

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

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 (23) gives the first result of the Lemma.

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

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

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

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

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 :

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

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

**Proof of Proposition 11:**

Combining (19) with Lemma 22 we get :

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

This finishes the proof of the proposition.

**Acknowledgements :** We are very grateful to two referees whose remarks and careful reading considerably enhanced the paper.

## References

- [1] Amato, U., Antoniadis, A., and De Feis, I. (2006). Dimension reduction in functional regression with applications. *Comp.Statist Data Anal.*, **50**, 2422-2446.
- [2] Berline, A., Thomas-Agnan, C. (2004) : *Reproducing kernel Hilbert spaces in Probability and Statistics*, Kluwer Academic Publishers.
- [3] Biau, G., Mas, A. (2012) : PCA-Kernel estimation, *Statistics & Risk Modeling*, **29**, 19–46.
- [4] Blanchard, G., Bousquet, O., Zwald, L. (2007) : Statistical Properties of Kernel Principal Component Analysis, *Mach. Learn.*, **66**, 259-294.
- [5] Bosq, D. (2000) : *Linear processes in function spaces*. Lectures notes in statistics, Springer.
- [6] Cardot, H., Mas, A., Sarda, P. (2007) : CLT in functional linear models, *Probab. Theory Relat. Fields*, **138**, 325–361.
- [7] Chen, D., Hall, P., Müller, H.G. (2011) : Single and multiple index functional regression models with nonparametric link, *Ann. Statist.*, **39**, 1720-1747.
- [8] 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.
- [9] Dauxois, J., Pousse, A., Romain, Y. (1982) : Asymptotic theory for the principal component analysis of a random vector function : some applications to statistical inference, *J. Multivariate Anal.*, **12**, 136-154.
- [10] Dunford, N., Schwartz, J.T. (1988) : *Linear Operators, Vol. I, II & III.*, Wiley Classics Library.
- [11] Ferraty, F. (2011). *Recent Advances on Functional Data analysis and Related Fields* (Ed.).Contributions to Statistics, Physica-Verlag, Heidelberg.
- [12] Ferraty, F., Goïa, A., Salinelli, E., Vieu, P. (2013) : Functional projection pursuit regression, *Test*, **22**, 293-320.
- [13] Ferraty, F., Laksaci A., Tadj, A., Vieu, P. (2010) : Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, **140**, 335–352.
- [14] Ferraty, F., Laksaci A., Tadj, A., Vieu, P. (2011) : Kernel regression with functional response, *Electron. J. Stat*, **5**, 159-171.
- [15] Ferraty, F., Mas, A., Vieu, P. (2007). Advances on nonparametric regression from functional data, *Aust. New Zealand J. Stat.*, **49**, 1-20.
- [16] 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.
- [17] Ferraty, F., Vieu, P. (2006) *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, Berlin.
- [18] Gilliam, J.S., Hohage, T., Ji, X., Ruymgaart, F. (2009) : The Fréchet derivative of an analytic function of a bounded operator with some applications. *Internat J. Math. Sci.*, Art. ID 239025.
- [19] Gohberg, I., Goldberg, S., Kaashoek, M.A. (1991) : *Classes of Linear Operators Vol I & II*, Operator Theory : Advances and Applications, Birkhäuser.
- [20] Hall, P., Hosseini-Nasab, M. (2006) : On properties of functional principal components analysis, *J. Royal Statist Society, Series B*, **68**, 109-126.
- [21] Hilgert, N., Mas, A., Verzelen, N. (2013) : Minimax adaptive tests for the functional linear model, *Ann. Statist.*, **41**, 838-869.

- [22] Kato, T. (1976) : *Perturbation Theory for Linear Operators*. Grundlehren der mathematischen Wissenschaften, 132, Berlin-Heidelberg-New York: Springer-Verlag. 2nd Ed.
- [23] 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.
- [24] Ledoux, M., Talagrand, M. (1991) : *Probability in Banach Spaces*, Springer-Verlag.
- [25] Mas, A. (2012) : Lower bound in regression for functional data by small ball probability representation in Hilbert space, *Electron. J. Statist*, **6**, 1745-1778.
- [26] Mas, A., Menneteau, L. (2003) : Perturbation approach applied to the asymptotic study of random operators, *Progress in Probability*, **55**, 127-134, Birkhäuser.
- [27] Masry, E. (2005) : Nonparametric regression estimation for dependent functional data: Asymptotic normality. *Stochastic Process. Appl.*, **115**, 155–177.
- [28] Schmiedler, W. (1965) : *Linear Operators in Hilbert Spaces*. Academic Press.
- [29] Skorohod A.V. (1984) : *Random Linear Operators*. D. Reidel Publishing.
- [30] Tsybakov, A.B. (2004) *Introduction a l'estimation non-paramétrique*, Springer.
- [31] Vakhania, N.N., Tarieladze, V.I., Chobanyan, S.A. (1987) : *Probability Distributions on Banach Spaces*, Mathematics and its Applications, D. Reidel Publishing.
- [32] Watson, G. S. (1983). *Statistics on Spheres*, Wiley, New York.
- [33] Weidman. J. (1980) : *Linear Operators in Hilbert Spaces*, Graduate Texts in Mathematics, Springer.
- [34] Zwald L. , Blanchard G. (2005) : On the convergence of eigenspaces in kernel principal component analysis . In *Advances in Neural Information Processing Systems* **18**, NIPS.