Probab. Theory Relat. Fields (2007) 138:325–361
DOI 10.1007/s00440-006-0025-2

# CLT in functional linear regression models

**Hervé Cardot · André Mas · Pascal Sarda**

Received: 12 June 2005 / Revised: 7 July 2006 / Published online: 23 August 2006


**Abstract** We propose in this work to derive a CLT in the functional linear regression model. The main difficulty is due to the fact that estimation of the functional parameter leads to a kind of ill-posed inverse problem. We consider estimators that belong to a large class of regularizing methods and we first show that, contrary to the multivariate case, it is not possible to state a CLT in the topology of the considered functional space. However, we show that we can get a CLT for the weak topology under mild hypotheses and in particular without assuming any strong assumptions on the decay of the eigenvalues of the covariance operator. Rates of convergence depend on the smoothness of the functional coefficient and on the point in which the prediction is made.

**Keywords** Central limit theorem · Hilbertian random variables · Functional data analysis · Covariance operator · Inverse problem · Regularization · Perturbation theory

H. Cardot
CESAER, 26, bd Docteur Petitjean, BP 87999, 21079 Dijon Cedex, France
e-mail: herve.cardot@enesad.inra.fr

A. Mas (✉)
Institut de Modélisation Mathématique de Montpellier, Université Montpellier II,
Place Eugène Bataillon, 34095 Montpellier Cedex 5, France
e-mail: mas@math.univ-montp2.fr

P. Sarda
Laboratoire de Statistique et Probabilités, Université Paul Sabatier,
118, route de Narbonne, 31062 Toulouse Cedex 4, France
e-mail: sarda@cict.fr

P. Sarda
GRIMM, EA 2254, Université Toulouse-le-Mirail,
5, Allées Antonio-Machado, 31058 Toulouse Cedex 1, France

**Mathematics Subject Classification (2000)** 60F05 · 60H05 · 45Q05 · 62J07

## 1 Introduction

For several years, there has been a considerable interest in *Functional Data Analysis*. Indeed, a consequence of advances in technology is the collection of many data sets on dense grids (e.g. in remote sensing, spectrometry or medicine) adding in some sense more and more information. The question is then: can we do something specific with this new information? It is the merit of the books by Ramsay and Silverman [31,32] to have prepared the ground for answers to this question. They, and other authors after them, have shown the practical benefits of using ad hoc statistical methods for these data sets: the point of view is clearly to take into account the functional nature of the data. This means that one considers the data as objects belonging to functional spaces with infinite dimension and one has to use adapted probabilistic and functional analysis tools to derive properties of estimators in such a context.

This emulates a need for developing theoretical/practical aspects on the ground of functional data analysis. It is the aim of this paper to contribute to this kind of development. The framework of our study is itself an important part of functional data problems. We are interested in the properties of an estimator of the linear regression in a functional framework, that is to say the linear regression of a real random variable on a functional variable. The two main motivations of this work are to study rigorously the asymptotic distribution of the estimator and from a statistical point of view to deduce asymptotic confidence intervals for prediction based on functional linear regression.

Estimation of regression with a functional predictor is not new and has many potential applications such as Chemometrics as it can be noticed in the paper by Frank and Friedman [13]. Whereas chemometricians have mainly used adaptations of statistical multivariate methods, *functional* procedures have gained in popularity more recently as said above. For instance Hastie and Mallows [18] have raised, in the discussion of the paper by Frank and Friedman [13], the question of functional alternative methods. Thus, for this case of estimating a regression, two main approaches have been considered, (1) estimating the functional linear regression which is a "continuous" version of multivariate linear regression and was first considered in Ramsay and Dalzell [30] and (2) proposing a complete nonparametric point of view developed by Ferraty and Vieu [12]. We consider the former approach hereafter: see Sect. 2, for the definition of the functional linear regression. Contrary to the multivariate linear regression where the vector of parameters is identifiable provided that the covariance matrix is non-singular, identifiability of the corresponding functional parameter is not ensured unless a sufficient and necessary condition is satisfied (see Sect. 2).

Different estimators for this functional parameter have been considered: see for instance [5,6,8,15]. Upper bounds for the $L^2$ rate of convergence found until now lead to the conjecture that the transition from the finite dimension to the infinite dimension does not lead to the usual parametric rate of convergence (see [6]). As a matter of fact, the main difficulty is that estimating the functional parameter of the functional linear regression can be viewed as an ill-conditioned inverse problem since it relies on the inversion of the covariance operator which is compact [3,5,19]. It is even a class of complicated ill-conditioned inverse problem since the covariance operator is unknown. However the problem of approximating inverses of covariance operators or of selfadjoint compact operators is not new. It is adressed in Nashed and Wahba [29], Arsenin and Tikhonov [2], Groetsch [16] among many others. The main point is always to regularize a matrix $M$ (respectively an operator $S$) which is invertible but "not by much" (respectively unbounded). This property implies that for any vector $x$, $Mx$ (respectively $Sx$) may have large variations even when $x$ does not vary much. Numerous procedures were proposed. Such procedures appear especially in image analysis or deconvolution or in specific M-estimation problems for instance.

In Sect. 3, we consider a class of regularization methods for inverting the covariance operator that leads to a quite general class of estimators with the aim of investigating CLT for prediction and as a by-product producing confidence sets for prediction. The Central Limit Theorem for i.i.d. Hilbert valued random variables play a central role in deriving the main results of this paper. The monograph by Araujo and Giné [1] or Chapter 5 in the book by Ledoux and Talagrand [22] deal with this crucial theorem of probability theory and provide deep studies about the CLT in infinite dimensional spaces. Müller and Stadtmüller [28] derived a CLT in the setting of Generalized Linear Models (including linear regression). For the non independent case we also mention three recent works by Dedecker and Merlevède [10] and Merlevède [27] and Mas [25].

Section 4 is devoted to the asymptotic behavior of these estimators relaxing as much as possible the set of assumptions (moment assumptions, assumptions on the spectrum of $\Gamma$) and considering a large class of regularizing methods for inverting the empirical covariance operator. We first derive an important result which shows that it is not possible to state a CLT for the functional coefficient with respect to the norm topology of the functional space. Nevertheless, we show that it is possible to get a CLT if we consider the behavior of the predictor with respect to the weak topology, that is to say for point-wise prediction. We show that the results depend on the nature of the predictor and fixed or random design lead to different CLT. Whereas when the predictor is random it is not possible to reach a parametric rate of convergence, this rate can be obtained depending on the value and the smoothness properties of the fixed predictor: we obtain a parametric rate for pointwise convergence at $x$ whenever $x$ belongs to the reproducing kernel Hilbert space associated to the covariance operator. As said above, the estimation procedure is quite general and includes in particular estimation based on spectral truncation (regression on functional principal components) studied by Müller and Stadtmüller [28] in the paper mentioned above. These authors derive asymptotic normality according to a kind of $L^2$ distance between the estimated and true functional coefficient of the model whereas we are here interested in proving a CLT for prediction both in the fixed and random design cases.

The proofs depend heavily on perturbation theory for linear operators to get, as accurate as possible, approximations of the eigenelements of the empirical covariance operators. Similar methods based on functional calculus have been used for deriving asymptotic properties of the functional principal components analysis by Dauxois et al. [9], Kneip [20], Kneip and Utikal [21] or Mas and Menneteau [26]. Section 5 proposes a brief discussion about possible extensions and statistical applications of these results. Finally Sect. 6 is devoted to the proofs.

## 2 Functional linear regression

We consider a sample $(X_i, Y_i), i = 1, \ldots, n$ of independent and identically distributed random variables drawn from a pair $(X, Y)$. The variables $X$ and $Y$ are defined on the same probability space and $Y$ (the response) is valued in $\mathbb{R}$. The variable $X$ (the predictor) is a random variable taking values in a general real separable Hilbert space $H$ with an inner product denoted in the following by $\langle ., . \rangle$ and an associated norm denoted by $\|.\|$. As a matter of fact $H$ may be the Sobolev space $W^{m,2}(\mathcal{C})$ of functions defined on some compact interval $\mathcal{C}$ of $\mathbb{R}$ having $m$ square integrable derivatives, $m$ being a positive integer. In that case the inner product $\langle ., . \rangle$ is the usual inner product on this space i.e.

$$\langle f, g \rangle = \sum_{p=0}^{m} \int_{\mathcal{C}} f^{(p)}(x) g^{(p)}(x) \mathrm{d}x, \quad f, g \in H.$$

Note that this special case is particularly interesting for modelling situations where we have functional data as shown by the numerous applications given in Ramsay and Silverman [31,32]. Although we develop below theory for general Hilbertian random variables, we keep in mind this special situation and then use the word *functional* variable to qualify $X$.

In the following we assume that $\mathbb{E}Y^2 < +\infty$ and that $X$ is a $H$-valued random variable such that

$$\mathbb{E}(\|X\|^4) < +\infty. \tag{H.1}$$

Then $X$ is of second order and one can define the expectation of $X$, namely $\mathbb{E}(X)$, that we suppose in order to simplify notations to be the null element of $H$, $(\mathbb{E}(X) = 0)$. Moreover the covariance operator of $X$ is defined as the linear operator $\Gamma$ defined on $H$ such that

$$\Gamma h = \mathbb{E}(X \otimes X(h)), \quad h \in H,$$

where $X \otimes X$ is the tensor product operator defined, for every $h$ belonging to $H$, as $X \otimes X(h) = \langle h, X \rangle X$. It is known that $\Gamma$ is a self-adjoint, positive and nuclear operator hence it is Hilbert–Schmidt and hence compact [9]. We denote by $(\lambda_j)_j$ the sorted sequence of non null distinct eigenvalues of $\Gamma$, $\lambda_1 > \lambda_2 > \cdots > 0$, and $(e_j)_j$ a sequence of orthonormal associated eigenvectors. We assume that the multiplicity of each $\lambda_j$ is one (remind that since $\Gamma$ is compact the multiplicity of each $\lambda_j \neq 0$ is finite). We could consider the more general case of multiple eigenvalues without affecting our forthcoming results but the price would be more complicated proofs and a poor gain with respect to the main objectives of the paper. Let us also define the cross-covariance operator of $X$ and $Y$ as the functional $\Delta$ defined on $H$ by

$$\Delta h = \mathbb{E}(X \otimes Y(h)), \quad h \in H,$$

where $X \otimes Y$ is the tensor product functional defined, for every $h$ belonging to $H$, as $X \otimes Y(h) = \langle h, X\rangle Y$. Now, we aim at considering the *functional linear regression* of the variable $Y$ on $X$. This means that we are seeking the solution $\rho \in H$ of the following minimization problem

$$\inf_{\beta \in H} \mathbb{E}\left(|Y - \langle \beta, X\rangle|^2\right). \tag{1}$$

When a solution $\rho$ exists and is uniquely determined, we can write

$$Y = \langle \rho, X\rangle + \varepsilon, \tag{2}$$

where $\varepsilon$ is a centered real random variable with variance $\sigma_\varepsilon^2$ such that $E(\varepsilon X) = 0$. It is quite easy to show that it is equivalent that $\rho$ satisfies Eq. (2) and that it satisfies the following moment equation (see for instance [4])

$$\Delta = \Gamma\rho.$$

However, when the dimension of $H$ is infinite, existence and uniqueness of $\rho$ is not ensured since a bounded inverse of $\Gamma$ does not exist: we need an additional condition to get existence and uniqueness of $\rho$, namely

**Condition** $\mathcal{U}$. The variables $X$ and $Y$ satisfy

$$\sum_j \frac{\langle \mathbb{E}(XY), e_j\rangle^2}{\lambda_j^2} < +\infty.$$

Under condition $\mathcal{U}$, Cardot et al. [4] show that a unique solution to Eq. (2) exists in $((Ker(\Gamma))^\perp$ and that this solution is of the form

$$\rho = \sum_j \frac{\langle \mathbb{E}(XY), e_j\rangle}{\lambda_j} e_j.$$

Then, identifiability is true only in $(Ker(\Gamma))^{\perp}$ or in other words the set of solution of (2) is of the form $\rho + Ker(\Gamma)$. Again, to simplify further developments we assume from now on that the following condition is satisfied

$$Ker(\Gamma) = \{0\}. \tag{H.2}$$

Finally, we assume from now on that the first and second moment of $\varepsilon$ given $X$ are respectively equal to $\mathbb{E}(\varepsilon|X) = 0$ and $\mathbb{E}(\varepsilon^2|X) = \sigma_\varepsilon^2$.

## 3 Inverse problem and regularization procedure

Once we get identifiability through condition $\mathcal{U}$, we turn to the problem of estimating the "functional" parameter $\rho$ from the sample $(X_i, Y_i)$, $i = 1, \ldots, n$. The first step is to define the empirical versions of $\Gamma$ and $\Delta$ which are

$$\Gamma_n = \frac{1}{n}\sum_{i=1}^{n} X_i \otimes X_i, \quad \Delta_n = \frac{1}{n}\sum_{i=1}^{n} X_i \otimes Y_i.$$

We have

$$\Delta_n = \Gamma_n \rho + U_n,$$

where $U_n = n^{-1}\sum_{i=1}^{n} X_i \otimes \varepsilon_i$ and taking the expectation we get

$$\mathbb{E}(\Delta_n) = \Delta = \Gamma\rho.$$

As shown in the previous section, inversion of $\Gamma$ can be viewed as a kind of ill-conditioned inverse problem (unlike in usual ill-conditioned inverse problems the operator $\Gamma$ is unknown). Also, the inverse of $\Gamma_n$ does not exist because $\Gamma_n$ is almost surely a finite rank operator. As usually for ill-conditioned inverse problem we need regularization and our aim is now to propose a general and unified method to get a sequence of continuous estimators for $\Gamma^{-1}$ based on $\Gamma_n$.

The method is theoretically based on the functional calculus for operators (see [11] or [14], for instance).

For further purpose we first define the sequence $\delta_j, j = 1, \ldots,$ of the smallest differences between distinct eigenvalues of $\Gamma$ as

$$\begin{cases} \delta_1 = \lambda_1 - \lambda_2, \\ \delta_j = \min(\lambda_j - \lambda_{j+1}, \lambda_{j-1} - \lambda_j). \end{cases}$$

Now take for $(c_n)_{n\in\mathbb{N}}$ a sequence of strictly positive numbers tending to zero such that $c_n < \lambda_1$ and set

$$k_n = \sup\left\{p : \lambda_p + \delta_p/2 \geq c_n\right\}. \tag{3}$$

Then define a class of sequences of positive functions $(f_n)_{n\in\mathbb{N}}$ with support $[c_n, +\infty)$ such that

$$f_n \text{ is decreasing on } [c_n, \lambda_1 + \delta_1], \tag{F.1}$$

$$\lim_{n\to+\infty} \sup_{x\geq c_n} |xf_n(x) - 1| = 0, \tag{F.2}$$

$$f_n'(x) \text{ exists for } x \in [c_n, +\infty). \tag{F.3}$$

Moreover, we will make in some cases the additional assumption below which will be helpful to reduce the bias of our estimator

$$\sup_{s\geq c_n} |sf_n(s) - 1| = o(1/\sqrt{n}). \tag{H.3}$$

Now we describe practically the regularization procedure. The eigenvalues of $\Gamma_n$ are denoted by $\hat{\lambda}_j$ and the associated eigenfunctions by $\hat{e}_j$. The bounded linear operator $\Gamma_n^\dagger$ is defined the following way:

- Choose a threshold $c_n$,
- Choose a sequence of functions $(f_n)_n$ satisfying (F.1)–(F.3),
- Compute the (functional) PCA of $\Gamma_n$ (*i.e.* calculate the eigenvalues $\hat{\lambda}_j$ and the eigenvectors $\hat{e}_j$),
- Compute the finite rank operator $\Gamma_n^\dagger$ with the same eigenvectors as $\Gamma_n$ and associated eigenvalues $f_n(\hat{\lambda}_j)$ (*i.e.* $\Gamma_n^\dagger = \sum_{j=1}^n f_n(\hat{\lambda}_j)\hat{e}_j \otimes \hat{e}_j$).

Obviously $c_n$ must be larger than the smallest significatively non-null eigenvalue of $\Gamma_n$. Once the threshold $c_n$ and the function $f_n$ (both depending on the sample size $n$) have been chosen, we see that the computation of the estimator of $\rho$ is quite easy through the relation

$$\hat{\rho} = \Gamma_n^\dagger \Delta_n. \tag{4}$$

Now let us give some examples of functions $f_n$ and the derived estimators of $\rho$.

*Example 1* If $f_n(x) = 1/x$ when $x \geq c_n$ and 0 elsewhere, condition (H.3) holds and $\Gamma_n^\dagger$ is obtained by simple spectral truncation with threshold $c_n$. The operator $\Gamma_n^\dagger\Gamma_n$ is nothing but the projection on a finite dimensional space. Note however that the random dimension of this space, say $d_n$, is not necessarily equal to $k_n$ (see (3)): for instance we may be in the situation where $\hat{\lambda}_{k_n+1} > c_n$ and then $d_n \geq k_n + 1$. Unlike $d_n$, $k_n$ is non random and was introduced because, as will be seen in the proofs, $P(d_n \neq k_n)$ tends to zero fast enough to consider essentially the situation when $d_n = k_n$. In other words the derived estimator for $\rho$ is asymptotically equivalent to the one considered in [5].

*Example 2* Let $\alpha_n$ be some scalar parameter. If $f_n(x) = 1/(x + \alpha_n)$ when $x \geq c_n$ and 0 elsewhere, we get a ridge-type estimator. Condition (H.3) is satisfied whenever $\alpha_n\sqrt{n}/c_n \longrightarrow 0$.

*Example 3* Let $\alpha_n$ be some scalar parameter. If $f_n(x) = x/(x^2 + \alpha_n)$ on its support, $\Gamma_n^\dagger$ is nothing but the Tikhonov regularization of $\Gamma_n$. Once more (H.3) holds if $\alpha_n \sqrt{n}/c_n^2 \longrightarrow 0$.

We may define as well, following Mas [24], a class of approximate for $\Gamma_n$ introducing $f_{n,p}(x) = x^p/(x + \alpha_n)^{p+1}$ or $f_{n,p}(x) = x^p/(x^{p+1} + \alpha_n)$, where again $\alpha_n$ is some scalar parameter and $p$ some integer.

This procedure is quite general to define regularized version or pseudo inverses for $\Gamma_n$. Up to the authors knowledge, all standard techniques for regularizing ill-conditioned matrices or unbounded operators stem from the above functional calculus.

We conclude this section by some remarks on computational aspects. Even if our results are derived when observing the whole curves, in practical situations the curves are only known in a finite number of discretization points. Well known numerical methods for approximating inner products (integrals) can be used to compute the estimators when the discretization points are numerous and dense enough. Under such a framework, we believe that our theoretical developments below remain valid under some additional conditions on the regularity of the trajectories and on the design (deterministic or random) of the discretization points. This issue, which certainly deserves further attention, is beyond the scope of the paper and one can find a preliminary discussion on this topic in Cardot et al. [4]. We must say that our work should be seen in this context. It may occur however that only few observations are available for each curve. A recent paper by Yao et al. [33] deals with this specific problem for sparse longitudinal data by means of a non-parametric regression approach.

Another problem in practice is the choice of the threshold $c_n$. For the estimator based on spectral truncation in example 1 above, this choice is in some sense the same as the choice of a dimension space $k_n$ where the data are projected (see relation (3)). A cross-validation criterion may be used to select this dimension space as done in Cardot et al. [4]. Müller and Stadtmüller [28] use an AIC criterion for choosing the dimension $k_n$ in a generalized functional linear model (including functional linear regression). Then, from (3), replacing the eigenvalues $\lambda_j$ by their empirical counterparts $\widehat{\lambda}_j$ provides a practical way to select $c_n$.

## 4 Asymptotic results

In this section, we mainly state weak convergence results for the statistical predictor of $Y_{n+1}$ for a new value $X_{n+1}$ obtained by means of estimator defined in (4), namely $\widehat{Y}_{n+1} = \langle \hat{\rho}, X_{n+1} \rangle$. Hence, we should study the stochastic convergence of

$$\langle \hat{\rho}, X_{n+1} \rangle - \langle \rho, X_{n+1} \rangle. \tag{5}$$

We also look at prediction for a given value of $x \in H$ and study the stochastic convergence of

$$\langle \hat{\rho}, x \rangle - \langle \rho, x \rangle. \tag{6}$$

It is important to note that all the results are obtained without assuming any prior knowledge for the rate of decay of the eigenvalues $\lambda_j$ of $\Gamma$ to zero. We will see that unfortunately a bias term appears which cannot be removed without very specific assumptions on the sequence on the spectrum of $\Gamma$ and on the smoothness properties of $\rho$.

We begin to investigate the weak convergence for the norm topology on $H$ for our estimate. The next and important result underlines the limits of the functional approach. It tells us that it is not possible to get a general result that would allow to build confidence sets in the functional setting. This highlights the fact that when considering functional data one must take care and multivariate classical results are not necessarily true anymore.

**Theorem 1** *It is impossible for $\widehat{\rho} - \rho$ to converge in distribution to a non-degenerate r.e. in the norm topology of $H$.*

The proof of Theorem 1 is postponed to Sect. 6.4: it is shown actually that for any normalizing sequence $\alpha_n \uparrow +\infty, \alpha_n (\widehat{\rho} - \rho)$ does not converge in distribution for the norm topology but to a degenerate random element.

Nevertheless this negative result does not mean that it is not possible to get some confidence sets. We have to consider a weak topology (with respect to the inner product), that is to say point-wise confidence bands, and study separately the cases of deterministic and random points. We first give results for the prediction approach.

We define $\Gamma^\dagger$ as $f_n(\Gamma)$. It is important to note that $\Gamma^\dagger$ depends on the sample size $n$ through the sequence $k_n$. From this we take in the following

$$t_{n,x} = \sqrt{\sum_{j=1}^{k_n} \lambda_j \left[f_n(\lambda_j)\right]^2 \langle x, e_j\rangle^2} = \sqrt{\left\|\Gamma^{1/2}\Gamma^\dagger x\right\|^2}, \quad x \in H,$$

$$s_n = \sqrt{\sum_{j=1}^{k_n} \left[\lambda_j f_n(\lambda_j)\right]^2} = \sqrt{tr(\Gamma^\dagger\Gamma)},$$

and denote by $\hat{t}_{n,x}$ and by $\hat{s}_n$ their empirical counterparts based on the estimated eigenvalues $\hat{\lambda}_j$'s. Note that the sequence $t_{n,x}$ may either converge or diverge depending on whether $\sum_{j=1}^{+\infty} \lambda_j^{-1}\langle x, e_j\rangle^2 = \left\|\Gamma^{-1/2}x\right\|^2$ is finite or not (i.e. whether $x$ is in the range of $\Gamma^{-1/2}$ or not). At the opposite, the term $s_n$ always tends to infinity.

### 4.1 Weak convergence for the predictor

We state a weak convergence theorem for the predictor given in (5). We denote by $\Pi_{k_n}$ the projector onto the eigenspace associated to the $k_n$ first

eigenvalues, and by $\widehat{\Pi}_{k_n}$ its empirical counterpart, i.e. the projector on the eigenspace associated to $\widehat{\lambda}_1, \widehat{\lambda}_2, ..., \widehat{\lambda}_{k_n}$.
Assumptions (H.1)–(H.3) are truly basic. They just ensure that the statistical problem is correctly posed. In order to get deep asymptotic results we introduce extra assumptions denoted by (A.1)–(A.3).

$$\sum_{l=1}^{+\infty} |\langle \rho, e_l \rangle| < +\infty. \tag{A.1}$$

$$\begin{gathered}\text{There exists a convex positive function } \lambda \text{, such that,} \\ \text{at least for } j \text{ large, } \lambda_j = \lambda\,(j)\,. \end{gathered} \tag{A.2}$$

We recall the Karhunen–Loève expansion of $X$, that is

$$X = \sum_{l=1}^{+\infty} \sqrt{\lambda_l} \xi_l e_l,$$

where the $\xi_l$'s are centered r.r.v such that $E\xi_l \xi_{l'} = 1$ if $l = l'$ and 0 otherwise. We assume the following assumption for variables $\xi_l$

$$\sup_l \mathbb{E}\xi_l^4 \ \leq \ M \ < \ +\infty. \tag{A.3}$$

*Remark 1* Assumption (A.2) is clearly unrestrictive since it holds for standard rates of decrease for the eigenvalues, polynomial or exponential. It implies that

$$\delta_k = \min\left(\lambda_k - \lambda_{k+1}, \lambda_{k-1} - \lambda_k\right) = \lambda_k - \lambda_{k+1}.$$

*Remark 2* Simple calculations show that assumption (A.3) implies assumption (H.1), namely that $\mathbb{E}\,\|X\|^4 < +\infty$ and does not require any condition on the stochastic dependence within the $\xi_l$'s. Besides (A.3) holds for a very large class of real-valued random variables (remind that the $\xi_l$'s are subject to $\mathbb{E}\xi_l = 0$ and $\mathbb{E}\xi_l^2 = 1$).

**Theorem 2** *When assumptions* (H.2)–(H.3) *and* (A.1)–(A.3) *hold and if*

$$\frac{k_n^{5/2}\,(\log k_n)^2}{\sqrt{n}} \to 0, \tag{7}$$

*then*

$$\frac{\sqrt{n}}{s_n}\left(\langle \widehat{\rho}, X_{n+1}\rangle - \langle \Pi_{k_n}\rho, X_{n+1}\rangle\right) \ \stackrel{w}{\to} \ N\left(0, \sigma_\varepsilon^2\right).$$

*If moreover*

$$n^{1/6} = O(k_n), \tag{8}$$

*and either* $\sup_p \left(\left|\langle \rho, e_p\rangle\right| p^{5/2}\right) < +\infty$ *or* $\sup_p \left(p^4 \lambda_p\right) < +\infty$ *then*

$$\frac{\sqrt{n}}{s_n}\left(\langle \widehat{\rho}, X_{n+1}\rangle - \langle \rho, X_{n+1}\rangle\right) \xrightarrow{w} N\left(0, \sigma_\varepsilon^2\right).$$

*Remark 3* The term $s_n$ always tends to infinity and hence we cannot obtain a "parametric" rate of decay in probability. Using the fact that under assumption $(A.2)$ $s_n$ is of the same order as $\sqrt{k_n}$, conditions (8) implies that the more favorable rate of convergence in our result is $n^{-5/12}$ which is quite good compared to what has been obtained previously. On another side, we think that there is still a large field of investigation about rates of convergence in functional linear regression models. An emulating perspective is to seek for optimal rates of convergence under appropriate assumptions (regularity of $X$, rate of decay of the eigenvalues $\lambda_j$, ...).

In the results above $s_n$ depends on the unknown eigenvalues. It is worth trying to get an "adaptive" version: replacing the $\lambda_i$'s with the $\widehat{\lambda}_i$'s in the first result of Theorem 2 leads to a new result with both a random bias and a random normalization term.

**Corollary 1** *Under Assumptions* (H.2)–(H.3) *and* (A.1)–(A.3) *and if* $k_n$ *satisfies* (7)*, we have*

$$\frac{\sqrt{n}}{\widehat{s}_n \sigma_\varepsilon}\left(\langle \widehat{\rho}, X_{n+1}\rangle - \langle \Pi_{k_n}\rho, X_{n+1}\rangle\right) \xrightarrow{w} N(0,1),$$

*where*

$$\widehat{s}_n = \sqrt{\sum_{j=1}^{k_n}\left[\widehat{\lambda}_j f_n\left(\widehat{\lambda}_j\right)\right]^2}.$$

*Remark 4* In all the previous results, the variance of the white noise $\sigma_\varepsilon^2$ is unknown. Replacing $\sigma_\varepsilon$ with a convergent estimate of $\sigma_\varepsilon$ does not change the Theorems.

## 4.2 Weak convergence for the estimate of $\rho$

We are now giving weak convergence results for the prediction at a given value $x$ in $H$.

**Theorem 3** *Fix any $x$ in $H$. When the assumptions of Theorem* 2 *hold and if*

$$\sup_p \frac{\left|\langle x, e_p\rangle\right|^2}{\lambda_p} < +\infty \quad \text{and} \quad \frac{k_n^3 (\log k_n)^2}{t_{n,x}\sqrt{n}} \to 0,$$

*then*

$$\frac{\sqrt{n}}{t_{n,x}\sigma_\varepsilon}\left(\langle\widehat{\rho},x\rangle-\left\langle\widehat{\Pi}_{k_n}\rho,x\right\rangle\right)\xrightarrow{w}N(0,1).$$

*Remark 5* The bias term here is random. It can be seen from the proof of the crucial Proposition 2 that the situation cannot be improved without very specific (maybe artificial) assumptions either on $\rho$ or on the $\lambda_i$'s.

The normalizing sequence $\sqrt{n}/t_{n,x}$ depends on the unknown $\lambda_j$'s. It is worth trying to get again an adaptive version of the above theorem (i.e. replace $t_{n,x}$ with $\widehat{t}_{n,x}=\sqrt{\sum_{j=1}^{k_n}\left(\widehat{\lambda}_j\left[f_n\left(\widehat{\lambda}_j\right)\right]^2\left\langle x,\widehat{e}_j\right\rangle^2\right)}$.

**Corollary 2** *Theorem 3 still holds if $t_{n,x}$ is replaced with its empirical counterpart $\widehat{t}_{n,x}$.*

The following Remark is crucial since it brings out once more what seems to be a typical feature of the functional setting.

*Remark 6* As seen before the sequence $t_{n,x}$ may either converge or diverge. Indeed, if $\left\|\Gamma^{-1/2}x\right\|$ is finite the normalization sequence grows surprisingly at a parametric rate (i.e $\sqrt{n}$). This could be understood as an extra-smoothing of the estimate $\widehat{\rho}$ through the integrals involving the scalar product. But in terms of prediction this fact could be misleading. This "extra-smoothing" is indeed an excessive and artificial smoothing since $P\left(\left\|\Gamma^{-1/2}X_{n+1}\right\|<+\infty\right)=0$. This also means the realizations of $X$ do not belong with probability one to the reproducing kernel Hilbert space associated to its covariance function [17]. In other words the results of this section are given for the sake of completeness and to explore the analytical properties of our estimates. For these reasons and if prediction is under concern, only $\left\langle\widehat{\rho},X_{n+1}\right\rangle$ should be considered and studied. In a multivariate setting all these considerations make no sense, since the situation is simpler (in fact, usually $P\left(\left\|\Gamma^{-1/2}X_{n+1}\right\|<+\infty\right)=1$ because $\Gamma^{-1/2}$ is bounded when $\Gamma$ is a full rank covariance matrix).

Within the proofs it is readily seen that assumption $(A.1)$ plays a crucial role in getting a non random bias term. The next Proposition illustrates this situation.

**Proposition 1** *Assume that $\lambda_k=k^{-1-\alpha}$ and $\langle x,e_k\rangle^2=k^{-1-\beta}$ with $\beta>1+\alpha$. Then, if $\sum_{j=1}^{+\infty}j^{1-\beta}\left\langle\rho,e_j\right\rangle^2=+\infty$, the sequence $\frac{\sqrt{n}}{t_{n,x}\sigma_\varepsilon}\left\langle\left(\widehat{\Pi}_{k_n}-\Pi_{k_n}\right)\rho,x\right\rangle$ may not be bounded in probability even if the random variables $X_i$, $i=1,\ldots,n$, are i.i.d. centered, Gaussian.*

*Remark 7* The condition $\beta>1+\alpha$ just ensures that $\left\|\Gamma^{-1/2}x\right\|<+\infty$. Besides if $\left\langle\rho,e_j\right\rangle^2=j^{-1-\gamma}$, $\sum_{j=1}^{n}j^{1-\beta}\left\langle\rho,e_j\right\rangle^2$ diverges whenever $\beta+\gamma<1$ which implies that $\sum_{j=1}^{+\infty}\left|\left\langle\rho,e_j\right\rangle\right|=+\infty$.

In fact the assumption on the location of $\rho$ mentioned in the Proposition should be understood as smoothness conditions.

## 5 Concluding remarks

One important application of previous results is the construction of confidence sets for prediction. In real life problems, the regression function $\rho$ is unknown but Theorem 2 allows us to build confidence sets. Let $q_\alpha$ be the quantile of order $1-\alpha/2$ of a Gaussian random variable with mean 0 and unit variance, we get under previous assumptions the following confidence set for prediction,

$$\lim_{n\to\infty} P\left(\frac{\sqrt{n}}{\hat{\sigma}\,\widehat{s}_n}\left|\langle\widehat{\rho}, X_{n+1}\rangle - \langle\rho, X_{n+1}\rangle\right| \geq q_\alpha\right) = 1-\alpha\,. \tag{9}$$

A simulation study [7] has shown that such confidence sets are accurate even for moderate sample sizes, i.e. for $n$ around 100.

From a mathematical point of view, one of the main novelty of this work relies on the facts that no prior information on the eigenvalues is assumed and the dimension sequence $k_n$ does not depend on the rate of decrease of these eigenvalues. As a consequence $k_n$ increase rather slowly, but not so much for a non parametric model. Nevetheless, let us notice that this situation may be significantly improved if some information on the eigenvalues is available.

From Theorem 2 it is possible to derive a general bound for the $L^2$ prediction error. Simple calculations (see the proof of Theorem 2) lead to:

$$\langle\widehat{\rho}-\rho, X_{n+1}\rangle^2 = O_{\mathbb{P}}\left(\frac{s_n}{n}\right) + O_{\mathbb{P}}\left(\sum_{j=k_n+1}^{\infty}\lambda_j\langle\rho, e_j\rangle^2\right). \tag{10}$$

Thus, it is not possible to go further without imposing more precise hypotheses on the smoothness of function $\rho$ with respect to the basis of eigenfunctions $e_j$ and the rate of decay of the eigenvalues $\lambda_j$ as remarked sooner in the article. Nevertheless, it was seen that the second term on the right in (10) can converge rapidly to zero in some situations. Besides assumption (A.1) provides us with some kind of uniformity with respect to $\rho$ when the latter belongs to a subset of $H$. Naturally, with these remarks we have in mind the study of the minimax rate of $L^2$ risk for the class of our predictor.

## 6 Proofs

Along the proofs we suppose that (H.1)–(H.3) hold. The letter $C$ will always stand for any (nonrandom and universal) positive constant. For any bounded operator $T$ defined and with values in $H$ we classically set

$$\|T\|_\infty = \sup_{x\in B_1}\|Tx\|\,,$$

where $B_1$ is the unit ball of $H$. We will quite often make use of the following facts.

- For any $u$ in $H$,

$$\mathbb{E}\langle X,u\rangle^2=\langle \Gamma u,u\rangle=\left\|\Gamma^{1/2}u\right\|^2.$$

- For a sufficiently large $i$, $\lambda_i\le\frac{C}{i\log i}$.
- The Hilbert–Schmidt norm is more precise than the classical norm for operators. Hence if $T$ is Hilbert–Schmidt

$$\|T\|_\infty\le\|T\|_{HS}=\sqrt{\sum_p\|Tu_p\|^2},$$

where $(u_p)_{p\in\mathbb{N}}$ is any complete orthonormal sequence in $H$.

From definitions of $\hat{\rho}$ and $U_n$ we have

$$\widehat{\rho}=\Gamma_n^\dagger\Gamma_n\rho+\left(\Gamma_n^\dagger-\Gamma^\dagger\right)U_n+\Gamma^\dagger U_n,$$

from which the forthcoming decomposition is trivial

$$\widehat{\rho}-\Pi_{k_n}\rho=T_n+S_n+R_n+Y_n, \tag{11}$$

where

$$T_n=\left(\Gamma_n^\dagger\Gamma_n-\widehat{\Pi}_{k_n}\right)\rho,\quad S_n=\left(\Gamma_n^\dagger-\Gamma^\dagger\right)U_n,$$
$$R_n=\Gamma^\dagger U_n,\quad Y_n=\left(\widehat{\Pi}_{k_n}-\Pi_{k_n}\right)\rho.$$

We also define

$$L_n=\Pi_{k_n}\rho-\rho.$$

The proofs are tiled into four subsections. After a brief introduction on operator-valued analytic functions, we begin with providing useful convexity inequalities for the eigenvalues and subsequent moment bounds. The second part shows that all the bias terms but $L_n$, say $T_n,S_n$ and $Y_n$ tend to zero in probability when correctly normalized. Weak convergence of $R_n$ is proved in the short third subsection. The last part provides the main results of the paper by collecting the Lemmas and Propositions previously proved.

## 6.1 Preliminary results

All along the proofs we will need auxiliary results from perturbation theory for bounded operators. It is of much help to have basic notions about spectral representation of bounded operators and perturbation theory. We refer to Dunford-Schwartz [11, Chapter VII. 3] or to Gohberg et al. [14] for an introduction to functional calculus for operators related with Riesz integrals.

Let us denote by $\mathcal{B}_i$ the oriented circle of the complex plane with center $\lambda_i$ and radius $\delta_i/2$ and define

$$\mathcal{C}_n = \bigcup_{i=1}^{k_n} \mathcal{B}_i \, .$$

The open domain whose boundary is $\mathcal{C}_n$ is not connected but however we can apply the functional calculus for bounded operators (see Dunford–Schwartz Section VII.3 Definitions 8 and 9). We also need to change slightly the definition of the sequence of functions $(f_n)_n$ by extending it to the complex plane, more precisely to $\mathcal{C}_n$. We admit that it is possible to extend $f_n$ to an analytic function $\widetilde{f_n}$ defined on the interior of $\mathcal{C}_n$ (in the plane) such that $\sup_{z\in\mathcal{C}_n} \left|\widetilde{f_n}(z)\right| \leq C \sup_{x\in[c_n,\lambda_1+\delta_1]} |f_n(x)|$. For instance if $f_n(x) = (1/x)\mathbb{1}_{[c_n,+\infty)}(x)$, take $\widetilde{f_n}(z) = (1/z)\mathbb{1}_{\mathcal{C}_n}(z)$. Results from perturbation theory yield

$$\Pi_{k_n} = \frac{1}{2\pi\iota} \int_{\mathcal{C}_n} (z-\Gamma)^{-1} \, \mathrm{d}z, \tag{12}$$

$$\Gamma^{\dagger} = \frac{1}{2\pi\iota} \int_{\mathcal{C}_n} (z-\Gamma)^{-1} f_n(z) \, \mathrm{d}z, \tag{13}$$

where $\iota^2 = -1$.

We also need to introduce the square root of symmetric operators: if $T$ is a positive self-adjoint operator (random or not), we denote by $(zI-T)^{1/2}$ the symmetric operator whose eigenvectors are the same as $T$ and whose eigenvalues are the complex square root of $z-\lambda_k$, $k\in\mathbb{N}$, denoted $(z-\lambda_k)^{1/2}$.

**Lemma 1** *Consider two large enough positive integers $j$ and $k$ such that $k > j$. Then*

$$j\lambda_j \;\geq\; k\lambda_k \quad \textit{and} \quad \lambda_j - \lambda_k \geq \left(1-\frac{j}{k}\right)\lambda_j. \tag{14}$$

*Besides*

$$\sum_{j\geq k} \lambda_j \leq (k+1)\,\lambda_k. \tag{15}$$

*Proof* We set for notational convenience $\lambda_j = \varphi(1/j)$ where $\varphi$ is, by assumption (A.2), a convex function defined on the interval $[0,1]$ such that $\varphi(0) = 0$ and $\varphi(1) = \lambda_1$.

The two inequalities in (14) follows directly from the well known inequalities for convex functions

$$\frac{\varphi(x_1)-\varphi(x_0)}{x_1-x_0} \leq \frac{\varphi(x_2)-\varphi(x_0)}{x_2-x_0} \leq \frac{\varphi(x_2)-\varphi(x_1)}{x_2-x_1}, \quad 0\leq x_0 < x_1 < x_2 \leq 1,$$

and by taking $x_0 = 0$, $x_1 = 1/k$ and $x_2 = 1/j$.

Set $\mu_k = \sum_{l \geq k} \lambda_l$. It is easy to see that the sequence $(\mu_k)_k$ satisfies assumption (A.2). Indeed for all $k$

$$\mu_k - \mu_{k+1} \leq \mu_{k-1} - \mu_k,$$

which is a sufficient condition to construct a convex function $\mu(k) = \mu_k$. We can then apply the second part of (14) with $\mu_{k+1}$ instead of $\lambda_k$ and $\mu_k$ instead of $\lambda_j$, which yields

$$\mu_k - \mu_{k+1} = \lambda_k \geq \frac{1}{k+1}\mu_k,$$

and (15) is proved. □

**Lemma 2** *The following is true for j large enough*

$$\sum_{l \neq j} \frac{\lambda_l}{|\lambda_l - \lambda_j|} \leq Cj \log j.$$

*Proof* We are first going to decompose the sum into three terms

$$\sum_{l \neq j} \frac{\lambda_l}{|\lambda_l - \lambda_j|} = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3,$$

where

$$\mathcal{T}_1 = \sum_{l=1}^{j-1} \frac{\lambda_l}{\lambda_l - \lambda_j}, \quad \mathcal{T}_2 = \sum_{l=j+1}^{2j} \frac{\lambda_l}{\lambda_j - \lambda_l}, \quad \mathcal{T}_3 = \sum_{l=2j+1}^{+\infty} \frac{\lambda_l}{\lambda_j - \lambda_l}.$$

Applying Lemma 1 we get

$$\mathcal{T}_1 = \sum_{l=1}^{j-1} \frac{\lambda_l}{\lambda_l - \lambda_j} \leq j \sum_{l=1}^{j-1} \frac{1}{j-l} \leq C_1 j \log j,$$

where $C_1$ is some positive constant. Also, applying once more (14) then (15), we get

$$\mathcal{T}_2 = \sum_{l=j+1}^{2j} \frac{\lambda_l}{\lambda_j - \lambda_l} \leq \sum_{l=j+1}^{2j} \frac{\lambda_l}{\lambda_j} \frac{l}{l-j}$$
$$\leq 2j \sum_{l=j+1}^{2j} \frac{1}{l-j} \leq C_2 j \log j,$$

and

$$\mathcal{T}_3 \leq \sum_{l=2j+1}^{+\infty} \frac{\lambda_l}{\lambda_j - \lambda_l} \leq \frac{\sum_{l=2j+1}^{+\infty} \lambda_l}{\lambda_j - \lambda_{2j}} \leq 2\frac{\sum_{l=2j+1}^{+\infty} \lambda_l}{\lambda_j} \leq C_3 j.$$

Hence the result follows and Lemma 2 is proved. □

**Lemma 3** *We have for $j$ large enough*

$$\mathbb{E} \sup_{z \in \mathcal{B}_j} \left\| (zI - \Gamma)^{-1/2} (\Gamma_n - \Gamma) (zI - \Gamma)^{-1/2} \right\|_\infty^2 \leq \frac{C}{n} (j \log j)^2 , \tag{16}$$

*and*

$$\mathbb{E} \sup_{z \in \mathcal{B}_j} \left\| (zI - \Gamma)^{-1/2} X_1 \right\|^2 \leq Cj \log j.$$

*Proof* Take $z \in \mathcal{B}_j$. By bounding the sup norm by the Hilbert–Schmidt one (see above), we get

$$\begin{aligned}
&\left\| (zI - \Gamma)^{-1/2} (\Gamma_n - \Gamma) (zI - \Gamma)^{-1/2} \right\|_\infty^2 \\
&\leq \sum_{l=1}^{+\infty} \sum_{k=1}^{+\infty} \left\langle (zI - \Gamma)^{-1/2} (\Gamma_n - \Gamma) (zI - \Gamma)^{-1/2} (e_l) , e_k \right\rangle^2 \\
&\leq \sum_{l,k=1}^{+\infty} \frac{\langle (\Gamma_n - \Gamma) (e_l) , e_k \rangle^2}{|z - \lambda_l| \, |z - \lambda_k|} \\
&\leq 4 \sum_{\substack{l,k=1,\\ l,k \neq j}}^{+\infty} \frac{\langle (\Gamma_n - \Gamma) (e_l) , e_k \rangle^2}{\left|\lambda_j - \lambda_l\right| \left|\lambda_j - \lambda_k\right|} + 2 \sum_{\substack{k=1,\\ k \neq j}}^{+\infty} \frac{\left\langle (\Gamma_n - \Gamma) \left(e_j\right) , e_k \right\rangle^2}{\delta_j \, |z - \lambda_k|} \\
&\quad + \frac{\left\langle (\Gamma_n - \Gamma) \left(e_j\right) , e_j \right\rangle^2}{\delta_j^2},
\end{aligned}$$

since it can be checked that whenever $z = \lambda_j + \frac{\delta_j}{2} e^{\iota\theta} \in \mathcal{B}_j$ and $i \neq j$

$$|z - \lambda_i| = \left| \lambda_j - \lambda_i + \frac{\delta_j}{2} e^{\iota\theta} \right| \geq \left|\lambda_j - \lambda_i\right| - \frac{\delta_j}{2} \geq \left|\lambda_j - \lambda_i\right| /2.$$

Besides

$$\mathbb{E} \left\langle (\Gamma_n - \Gamma) (e_l) , e_k \right\rangle^2 \; = \; \frac{1}{n} \left[ \mathbb{E} \left( \langle X_1, e_k \rangle^2 \langle X_1, e_l \rangle^2 \right) - \langle \Gamma (e_l) , e_k \rangle^2 \right] \leq \frac{M}{n} \lambda_l \lambda_k,$$

when assumption (A.3) holds. Finally

$$
\begin{aligned}
&\mathbb{E}\sup_{z\in\mathcal{B}_j}\left\|(zI-\Gamma)^{-1/2}(\Gamma_n-\Gamma)(zI-\Gamma)^{-1/2}\right\|_\infty^2\\
&\quad\le\frac{M}{n}\left[\sum_{\substack{l,k=1,\\ l,k\neq j}}^{+\infty}\frac{\lambda_l\lambda_k}{\left|\lambda_j-\lambda_l\right|\left|\lambda_j-\lambda_k\right|}+\frac{\lambda_j}{\delta_j}\sum_{k=1,k\neq j}^{+\infty}\frac{\lambda_k}{\left|\lambda_j-\lambda_l\right|}+\left(\frac{\lambda_j}{\delta_j}\right)^2\right]\\
&\quad=\frac{M}{n}\left[\left(\sum_{k=1,k\neq k}^{+\infty}\frac{\lambda_k}{\left|\lambda_j-\lambda_k\right|}\right)^2+\frac{\lambda_j}{\delta_j}\sum_{k=1,k\neq j}^{+\infty}\frac{\lambda_k}{\left|\lambda_j-\lambda_l\right|}+\left(\frac{\lambda_j}{\delta_j}\right)^2\right].
\end{aligned}
$$

It suffices now to apply Lemmas 1 and 2 to get the desired result. The same method leads to proving the second part of the display. Lemma 3 is proved. □

**Lemma 4** *Denoting*

$$
\mathcal{E}_j(z)=\left\{\left\|(zI-\Gamma)^{-1/2}(\Gamma_n-\Gamma)(zI-\Gamma)^{-1/2}\right\|_\infty<1/2, z\in\mathcal{B}_j\right\},
$$

*the following holds*

$$
\left\|(zI-\Gamma)^{1/2}(zI-\Gamma_n)^{-1}(zI-\Gamma)^{1/2}\right\|_\infty \mathbb{1}_{\mathcal{E}_j(z)}\le C,\quad a.s.
$$

*where C is some positive constant. Besides*

$$
\mathbb{P}\left(\mathcal{E}_j^c(z)\right)\le\frac{j\log j}{\sqrt{n}}. \tag{17}
$$

*Proof* We have successively

$$
(zI-\Gamma_n)^{-1}=(zI-\Gamma)^{-1}+(zI-\Gamma)^{-1}(\Gamma-\Gamma_n)(zI-\Gamma_n)^{-1},
$$

hence

$$
\begin{aligned}
&(zI-\Gamma)^{1/2}(zI-\Gamma_n)^{-1}(zI-\Gamma)^{1/2}\\
&\quad=I+(zI-\Gamma)^{-1/2}(\Gamma-\Gamma_n)(zI-\Gamma_n)^{-1}(zI-\Gamma)^{1/2},
\end{aligned}\tag{18}
$$

and

$$
\begin{aligned}
&\left[I+(zI-\Gamma)^{-1/2}(\Gamma_n-\Gamma)(zI-\Gamma)^{-1/2}\right]\\
&\quad\times(zI-\Gamma)^{1/2}(zI-\Gamma_n)^{-1}(zI-\Gamma)^{1/2}=I.
\end{aligned}\tag{19}
$$

It is a well known fact that if the linear operator $T$ satisfies $\|T\|_\infty < 1$ then $I+T$ is an invertible and its inverse is bounded and given by formula

$$(I+T)^{-1} = I - T + T^2 - \cdots.$$

From (18) and (19) we deduce that

$$\begin{aligned}&\left\|(zI-\Gamma)^{1/2}(zI-\Gamma_n)^{-1}(zI-\Gamma)^{1/2}\right\|_\infty \mathbb{1}_{\mathcal{E}_j(z)}\\&\quad = \left\|\left[I+(zI-\Gamma)^{-1/2}(\Gamma_n-\Gamma)(zI-\Gamma)^{-1/2}\right]^{-1}\right\|_\infty \mathbb{1}_{\mathcal{E}_j(z)} \le 2, \quad a.s.\end{aligned}$$

Now, the bound in (17) stems easily from Markov inequality and (16) in Lemma 3. This finishes the proof of the Lemma. □

The empirical counterparts of (12) and (13)-mentioned above- involve a random contour, say $\widehat{\mathcal{B}}_i$, centered at $\widehat{\lambda}_i$. It should be noted that these contours cannot be replaced by the $\mathcal{B}_i$'s since the latter may contain more than $k_n$ eigenvalues of $\Gamma_n$. The aim of the following Lemma is to find sufficient conditions under which $\widehat{\mathcal{B}}_i$ may be replaced with $\mathcal{B}_i$. In other words, we have to check that for a sufficiently large $n$ the $p$th eigenvalue of $\Gamma_n$ is close enough to the $p$th eigenvalue of $\Gamma$. Before stating this first lemma, we introduce the following event

$$\mathcal{A}_n = \left\{\forall j \in \{1,...,k_n\} \,|\, \frac{\left|\widehat{\lambda}_j - \lambda_j\right|}{\delta_j} < 1/2\right\}.$$

**Lemma 5** *If* $\frac{k_n^2 \log k_n}{\sqrt{n}} \to 0$, *then*

$$\frac{1}{2\pi\iota}\int_{\mathcal{C}_n} (z-\Gamma_n)^{-1}\, dz = \widehat{\Pi}_{k_n}\mathbb{1}_{\mathcal{A}_n} + r_n,$$

*where* $r_n$ *is a random operator satisfying* $\sqrt{n} r_n \xrightarrow{\mathbb{P}} 0$ *in the operator norm.*

*Proof* When the event $\mathcal{A}_n$ holds, the $k_n$ first empirical eigenvalues $\widehat{\lambda}_j$ lie in $\mathcal{B}_j$ and then

$$\widehat{\Pi}_{k_n} = \frac{1}{2\pi\iota}\int_{\widehat{\mathcal{C}}_n} (z-\Gamma_n)^{-1}\,\mathrm{d}z = \frac{1}{2\pi\iota}\int_{\mathcal{C}_n} (z-\Gamma_n)^{-1}\,\mathrm{d}z.$$

From this it is clear that

$$\frac{1}{2\pi\iota}\int_{\mathcal{C}_n} (z-\Gamma_n)^{-1}\,\mathrm{d}z = \widehat{\Pi}_{k_n}\mathbb{1}_{\mathcal{A}_n} + \mathbb{1}_{\mathcal{A}_n^c}\frac{1}{2\pi\iota}\int_{\mathcal{C}_n} (z-\Gamma_n)^{-1}\,\mathrm{d}z.$$

Denoting $r_n = \mathbb{1}_{\mathcal{A}_n^c} \frac{1}{2\pi\iota} \int_{\mathcal{C}_n} (z - \Gamma_n)^{-1} \,\mathrm{d}z$, we see that, since $\left\| \frac{1}{2\pi\iota} \int_{\mathcal{C}_n} (z - \Gamma_n)^{-1} \,\mathrm{d}z \right\|_\infty = 1$, we have for $\varepsilon > 0$

$$\mathbb{P}\left(\sqrt{n}\, \|r_n\|_\infty > \varepsilon\right) \leq \mathbb{P}\left(\mathbb{1}_{\mathcal{A}_n^c} > \varepsilon\right) = \mathbb{P}\left(\mathcal{A}_n^c\right).$$

It remains to find a bound for $\mathbb{P}\left(\mathcal{A}_n^c\right)$. We have

$$\begin{aligned}\mathbb{P}\left(\mathcal{A}_n^c\right) &\leq \sum_{j=1}^{k_n} \mathbb{P}\left(\left|\widehat{\lambda}_j - \lambda_j\right| > \delta_j/2\right) \\ &\leq 2\sum_{j=1}^{k_n} \frac{\mathbb{E}\left|\widehat{\lambda}_j - \lambda_j\right|}{\delta_j} = \frac{2}{\sqrt{n}} \sum_{j=1}^{k_n} \frac{\lambda_j}{\delta_j} \frac{\sqrt{n}\mathbb{E}\left|\widehat{\lambda}_j - \lambda_j\right|}{\lambda_j}.\end{aligned} \tag{20}$$

In order to get a uniform bound with respect to $j$ of the latter expectation we follow the same arguments as Bosq [3], proof of Theorem 4.10 p. 122–123. In Bosq, the setting is quite more general but however his Theorem 4.10 ensures that in our framework the asymptotic behaviour of $\sqrt{n}\left(\left|\widehat{\lambda}_j - \lambda_j\right|/\lambda_j\right)$ is the same as $\sqrt{n}\left(\left|\langle(\Gamma_n - \Gamma)\, e_j, e_j\rangle\right|/\lambda_j\right)$. From assumption (A.3), we get

$$\sqrt{n}\frac{\mathbb{E}\left|\langle(\Gamma_n - \Gamma)\, e_j, e_j\rangle\right|}{\lambda_j} \leq \frac{\sqrt{\mathbb{E}\left|\langle X_1 e_j\rangle^4 - \lambda_j^2\right|}}{\lambda_j} \leq C, \tag{21}$$

where $C$ does not depend on $j$. From (20) and (21) we deduce, applying Lemma 1 once more, that

$$\mathbb{P}\left(\mathcal{A}_n^c\right) \leq \frac{C}{\sqrt{n}} \sum_{j=1}^{k_n} \frac{\lambda_j}{\delta_j} \leq \frac{C}{\sqrt{n}} \sum_{j=1}^{k_n} j \log j \leq \frac{C}{\sqrt{n}} k_n^2 \log k_n,$$

from which the result follows. □

It may be easily proved that the same result as in the preceding Lemma holds with $\Gamma_n^\dagger$ instead of $\widehat{\Pi}_{k_n}$. From now on we will implicitly work on the space $\mathcal{A}_n$ and then write

$$\widehat{\Pi}_{k_n} = \left(\frac{1}{2\pi\iota} \int_{\mathcal{C}_n} (z - \Gamma_n)^{-1} \,\mathrm{d}z\right)$$

and

$$\Gamma_n^\dagger = \left( \frac{1}{2\pi\iota} \int_{\mathcal{C}_n} (z - \Gamma_n)^{-1} \widetilde{f}_n (z)\, \mathrm{d}z \right).$$

We will also abusively denote $\Pi_{k_n} \mathbb{1}_{\mathcal{A}_n}$ by $\Pi_{k_n}$ and $\Gamma^\dagger \mathbb{1}_{\mathcal{A}_n}$ by $\Gamma^\dagger$.

*Remark 8* In fact thanks to Lemma 5, we can deal with all our random elements as if almost surely all the random eigenvalues were in their associated circles $\mathcal{B}_j$. The reader should keep this fact in mind all along the forthcoming proofs. The condition on $k_n$ needed on the Lemma is clearly weaker that the ones which appear for the main results to hold.

## 6.2 Bias terms

This subsection is devoted to the study of the bias terms $S_n$, $T_n$ and $Y_n$. A bound is also given for $L_n$ for further purpose. We first begin with the term $T_n$ for which we have the following lemma.

**Lemma 6** *If* $(H.3)$ *holds*

$$\|T_n\|_\infty = \left\| \left( \Gamma_n^\dagger \Gamma_n - \widehat{\Pi}_{k_n} \right) \rho \right\|_\infty = o_P \left( \frac{1}{\sqrt{n}} \right).$$

*Proof* Obviously $\Gamma_n^\dagger \Gamma_n - \widehat{\Pi}_{k_n}$ is a self-adjoint random operator whose eigenvalues are the $\left(\widehat{\lambda}_j f_n \left(\widehat{\lambda}_j\right) - 1\right)_{1 \le j \le k_n}$ and 0 otherwise. So we have

$$\left\| \Gamma_n^\dagger \Gamma_n - \widehat{\Pi}_{k_n} \rho \right\|_\infty \le C \sup_{s \ge c_n} \left( |s f_n (s) - 1| \right).$$

If assumption (H.3) holds, the last term above is an $o\left(1/\sqrt{n}\right)$, which proves the second equality. □

**Lemma 7** *The two following bounds are valid*

$$\sqrt{\frac{n}{k_n}} \mathbb{E} \left| \langle L_n, X_{n+1} \rangle \right| \le \sqrt{\frac{n}{k_n}} \left| \langle \rho, e_{k_n} \rangle \right| \sqrt{\sum_{l \ge k_n + 1} \lambda_l},$$

$$\sqrt{\frac{n}{k_n}} \mathbb{E} \left| \langle L_n, X_{n+1} \rangle \right| \le \frac{\lambda_{k_n}}{k_n} \sqrt{\frac{n}{\log k_n}} \sqrt{\sum_{l \ge k_n + 1} \langle \rho, e_l \rangle}.$$

*Proof* We have

$$\begin{aligned}
\mathbb{E}\left|\left\langle\left(I-\Pi_{k_n}\right)\rho, X_{n+1}\right\rangle\right| &\leq \sqrt{\mathbb{E}\sum_{l=k_n+1}\langle\rho,e_l\rangle^2\left\langle X_{n+1},e_l\right\rangle^2}\\
&=\sqrt{\sum_{l\geq k_n+1}\lambda_l\langle\rho,e_l\rangle^2}\\
&\leq\begin{cases}\left|\left\langle\rho,e_{k_n}\right\rangle\right|\sqrt{\sum_{l\geq k_n+1}\lambda_l}\\ \frac{\lambda_{k_n}}{\sqrt{k_n\log k_n}}\sqrt{\sum_{l\geq k_n+1}\langle\rho,e_l\rangle},\end{cases}
\end{aligned}$$

since $\lambda_l$ and $|\langle\rho,e_l\rangle|$ are absolutely summing sequences. □

**Proposition 2** *If* $\frac{1}{\sqrt{n}}k_n^{5/2}(\log k_n)^2\to 0$ *as* $n$ *goes to infinity, then*

$$\sqrt{\frac{n}{k_n}}\left\langle\left(\widehat{\Pi}_{k_n}-\Pi_{k_n}\right)\rho,X_{n+1}\right\rangle\xrightarrow{\mathbb{P}}0.$$

*Proof* The proof of the Proposition is the keystone of the paper. We begin with

$$\begin{aligned}
\left(\widehat{\Pi}_{k_n}-\Pi_{k_n}\right) &= \frac{1}{2\pi\iota}\sum_{j=1}^{k_n}\int_{\mathcal{B}_j}\left[(zI-\Gamma_n)^{-1}-(zI-\Gamma)^{-1}\right]\mathrm{d}z\\
&=\frac{1}{2\pi\iota}\sum_{j=1}^{k_n}\int_{\mathcal{B}_j}\left[(zI-\Gamma_n)^{-1}(\Gamma_n-\Gamma)(zI-\Gamma)^{-1}\right]\mathrm{d}z\\
&=\mathcal{S}_n+\mathcal{R}_n,
\end{aligned}$$

where

$$\mathcal{S}_n=\frac{1}{2\iota\pi}\sum_{j=1}^{k_n}\int_{\mathcal{B}_j}\left[(zI-\Gamma)^{-1}(\Gamma_n-\Gamma)(zI-\Gamma)^{-1}\right]\mathrm{d}z,$$

and

$$\mathcal{R}_n=\frac{1}{2\iota\pi}\sum_{j=1}^{k_n}\int_{\mathcal{B}_j}\left[(zI-\Gamma)^{-1}(\Gamma_n-\Gamma)(zI-\Gamma)^{-1}(\Gamma_n-\Gamma)(zI-\Gamma_n)^{-1}\right]\mathrm{d}z. \tag{22}$$

Result (28) below will provide us with a sufficient condition for $\mathcal{R}_n$ to be negligible. At first, we turn to $\mathcal{S}_n$. We have

$$\begin{aligned}
\mathbb{E}\left\langle \mathcal{S}_n\rho, X_{n+1}\right\rangle^2 &= \mathbb{E}\left(\sum_{l,l'=1}^{+\infty} \langle \mathcal{S}_n\rho, e_l\rangle \left\langle X_{n+1}, e_l\right\rangle \langle \mathcal{S}_n\rho, e_{l'}\rangle \left\langle X_{n+1}, e_{l'}\right\rangle\right) \\
&= \mathbb{E}\left(\sum_{l=1}^{+\infty} \langle \mathcal{S}_n\rho, e_l\rangle^2 \left\langle X_{n+1}, e_l\right\rangle^2\right) = \left(\sum_{l=1}^{+\infty} \lambda_l \mathbb{E}\, \langle \mathcal{S}_n\rho, e_l\rangle^2\right),
\end{aligned}$$

since $\mathbb{E}\left(\left\langle X_{n+1}, e_l\right\rangle\left\langle X_{n+1}, e_{l'}\right\rangle\right) = 0$ if $l \neq l'$ and $X_{n+1}$ is independent from $\mathcal{S}_n$. Now

$$\mathbb{E}\, \langle \mathcal{S}_n\rho, e_l\rangle^2 = \mathbb{E}\, \langle \rho, \mathcal{S}_n e_l\rangle^2 = \mathbb{E}\left(\sum_{l'=1}^{+\infty} \langle \rho, e_{l'}\rangle \langle \mathcal{S}_n e_l, e_{l'}\rangle\right)^2.$$

The operator $\mathcal{S}_n$ was explicitly computed by Dauxois et al. [9]. More precisely

$$\frac{1}{2\pi\iota}\int_{\mathcal{B}_j} \left[(zI-\Gamma)^{-1}(\Gamma_n-\Gamma)(zI-\Gamma)^{-1}\right]\mathrm{d}z = v_j(\Gamma_n-\Gamma)\pi_j + \pi_j(\Gamma_n-\Gamma)v_j,$$

with $v_j = \sum_{j'\neq j} \frac{1}{\lambda_{j'}-\lambda_j}\pi_{j'}$ where $\pi_j$ is the projector on the eigenspace associated to the $j^{th}$ eigenfunction of $\Gamma$. Hence

$$\begin{aligned}
\langle \mathcal{S}_n e_l, e_{l'}\rangle &= \sum_{j=1}^{k_n}\left[\left\langle(\Gamma_n-\Gamma)\pi_j e_l, v_j e_{l'}\right\rangle + \left\langle(\Gamma_n-\Gamma)v_j e_l, \pi_j e_{l'}\right\rangle\right] \\
&= \begin{cases} 0 \text{ if } (l' \leq k_n \text{ and } l \leq k_n) \text{ or if } (l' > k_n \text{ and } l > k_n), \\ \frac{\left\langle(\Gamma_n-\Gamma)e_l, e_{l'}\right\rangle}{\lambda_{l'}-\lambda_l} \text{ if } l' > k_n \text{ and } l \leq k_n, \\ \frac{\left\langle(\Gamma_n-\Gamma)e_l, e_{l'}\right\rangle}{\lambda_l-\lambda_{l'}} \text{ if } l' \leq k_n \text{ and } l > k_n. \end{cases}
\end{aligned} \tag{23}$$

Finally, if we take for instance $l \leq k_n$

$$\begin{aligned}
\mathbb{E}\, \langle \mathcal{S}_n\rho, e_l\rangle^2 &= \mathbb{E}\left(\sum_{l' \geq k_n+1}^{+\infty} \langle \rho, e_{l'}\rangle \frac{\langle(\Gamma_n-\Gamma)e_l, e_{l'}\rangle}{\lambda_{l'}-\lambda_l}\right)^2 \\
&= \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n}\sum_{l' \geq k_n+1}^{+\infty} \langle \rho, e_{l'}\rangle \frac{\left\langle(X_j\otimes X_j-\Gamma)e_l, e_{l'}\right\rangle}{\lambda_{l'}-\lambda_l}\right)^2 \\
&= \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n} Z_{j,l,n}^*\right)^2,
\end{aligned}$$

where

$$Z^*_{j,l,n} = \sum_{l' \geq k_n+1}^{+\infty} \langle \rho, e_{l'} \rangle \frac{\left\langle \left(X_j \otimes X_j - \Gamma\right) e_l, e_{l'} \right\rangle}{\lambda_{l'} - \lambda_l},$$

and the $(Z^*_{j,l,n})_{j\geq 1}$ are centered and uncorrelated random variables. Hence

$$\mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n} Z^*_{j,l,n}\right)^2 = \frac{1}{n}\mathbb{E}\left(\sum_{l' \geq k_n+1}^{+\infty} \langle \rho, e_{l'} \rangle \frac{\langle (X_1, e_l)\rangle \langle (X_1, e_{l'})\rangle}{\lambda_{l'} - \lambda_l}\right)^2.$$

Since $l \leq k_n < l'$, by using the Karhunen–Loève expansion of $X_1$, we get

$$\sum_{l' \geq k_n+1}^{+\infty} \langle \rho, e_{l'} \rangle \frac{\langle (X_1, e_l)\rangle \langle (X_1, e_{l'})\rangle}{\lambda_{l'} - \lambda_l} = \sum_{l' \geq k_n+1}^{+\infty} \langle \rho, e_{l'} \rangle \frac{\sqrt{\lambda_l \lambda_{l'}} \xi_l \xi_{l'}}{\lambda_{l'} - \lambda_l}.$$

and then

$$\begin{aligned}
&\mathbb{E}\left(\sum_{l' \geq k_n+1}^{+\infty} \langle \rho, e_{l'} \rangle \frac{\langle (X_1, e_l)\rangle \langle (X_1, e_{l'})\rangle}{\lambda_{l'} - \lambda_l}\right)^2 \\
&\quad = \sum_{l', m \geq k_n+1}^{+\infty} \langle \rho, e_{l'} \rangle \langle \rho, e_m \rangle \frac{\sqrt{\lambda_l^2 \lambda_l \lambda_m} \mathbb{E}\left(\xi_l^2 \xi_{l'} \xi_m\right)}{\left(\lambda_{l'} - \lambda_l\right)\left(\lambda_m - \lambda_l\right)}.
\end{aligned}$$

By applying twice Cauchy–Schwarz inequality to the $\xi_k$'s and under assumption (A.3), we get

$$\begin{aligned}
\mathbb{E}\left(\xi_l^2 \xi_{l'} \xi_m\right) &\leq \sqrt{\mathbb{E}\left(\xi_l^4\right)} \sqrt{\mathbb{E}\left(\xi_{l'}^2 \xi_m^2\right)} \\
&\leq \sqrt{M}\sqrt{M}.
\end{aligned}$$

Summing up what we made above we get

$$\mathbb{E}\left(\sum_{l' \geq k_n+1}^{+\infty} \langle \rho, e_{l'} \rangle \frac{\langle (X_1, e_l)\rangle \langle (X_1, e_{l'})\rangle}{\lambda_{l'} - \lambda_l}\right)^2 \leq M\left(\sum_{l' \geq k_n+1}^{+\infty} \langle \rho, e_{l'} \rangle \frac{\sqrt{\lambda_l \lambda_{l'}}}{\lambda_{l'} - \lambda_l}\right)^2.$$

Remember that we had fixed $l \leq k_n$. Now, if we take $l > k_n$ similar calculations lead to

$$\mathbb{E}\left(\sum_{l'=1}^{k_n} \langle \rho, e_{l'} \rangle \frac{\langle (X_1, e_l)\rangle \langle (X_1, e_{l'})\rangle}{\lambda_{l'} - \lambda_l}\right)^2 \leq M\left(\sum_{l'=1}^{k_n} \langle \rho, e_{l'} \rangle \frac{\sqrt{\lambda_l \lambda_{l'}}}{\lambda_{l'} - \lambda_l}\right)^2.$$

At last

$$\frac{n}{k_n}\mathbb{E}\left\langle \mathcal{S}_n\rho, X_{n+1}\right\rangle^2 \leq \frac{M}{k_n}\sum_{l=1}^{k_n}\lambda_l\left(\sum_{l'\geq k_n+1}^{+\infty}\langle\rho, e_{l'}\rangle\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_{l'}-\lambda_l}\right)^2 \tag{24}$$

$$+\frac{M}{k_n}\sum_{l>k_n}\lambda_l\left(\sum_{l'\geq 1}^{k_n}\langle\rho, e_{l'}\rangle\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_{l'}-\lambda_l}\right)^2. \tag{25}$$

We apply Lemma 1 first to bound (24)

$$\begin{aligned}\frac{M}{k_n}\sum_{l=1}^{k_n}\lambda_l\left(\sum_{l'\geq k_n+1}^{+\infty}\langle\rho, e_{l'}\rangle\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right)^2 &\leq \frac{M}{k_n}\sum_{l=1}^{k_n}\lambda_l\left(\sum_{l'\geq k_n+1}^{+\infty}\langle\rho, e_{l'}\rangle\sqrt{\frac{\lambda_{l'}}{\lambda_l}}\frac{1}{1-\frac{l}{l'}}\right)^2\\ &\leq \frac{M}{k_n}\sum_{l=1}^{k_n}\left(\sum_{l'\geq k_n+1}^{+\infty}\langle\rho, e_{l'}\rangle\sqrt{\lambda_{l'}}\frac{1}{1-\frac{l}{l'}}\right)^2.\end{aligned}$$

Now we set $h_n=\left[\sqrt{\frac{k_n}{\log k_n}}\right]$ where $[u]\,, u\in\mathbb{R}$, denotes the largest integer smaller than $u$. Note that the last inequality in the display above may be split as follows

$$\begin{aligned}\frac{M}{k_n}\sum_{l=1}^{k_n}\lambda_l\left(\sum_{l'\geq k_n+1}^{+\infty}\langle\rho, e_{l'}\rangle\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right)^2 &\leq \frac{2M}{k_n}\sum_{l=1}^{k_n}\left(\sum_{l'\geq k_n+1}^{k_n+h_n}|\langle\rho, e_{l'}\rangle|\sqrt{\lambda_{l'}}\frac{1}{1-\frac{l}{l'}}\right)^2\\ &\quad+\frac{2M}{k_n}\sum_{l=1}^{k_n}\left(\sum_{l'\geq k_n+h_n}^{+\infty}|\langle\rho, e_{l'}\rangle|\sqrt{\lambda_{l'}}\frac{1}{1-\frac{l}{l'}}\right)^2.\end{aligned} \tag{26}$$

Dealing with the second term we get for $l'\geq k_n+h_n$

$$1-\frac{l}{l'}\geq 1-\frac{k_n}{k_n+h_n}=\frac{h_n}{k_n+h_n},$$

and hence

$$\begin{aligned}\sum_{l'\geq k_n+h_n}^{+\infty}|\langle\rho, e_{l'}\rangle|\sqrt{\lambda_{l'}}\frac{1}{1-\frac{l}{l'}} &\leq \sum_{l'\geq k_n+h_n}^{+\infty}|\langle\rho, e_{l'}\rangle|\sqrt{\lambda_{l'}}\left(1+\frac{k_n}{h_n}\right)\\ &\leq \sum_{l'\geq k_n+h_n}^{+\infty}|\langle\rho, e_{l'}\rangle|\sqrt{\lambda_{l'}}\left(1+\sqrt{k_n\log k_n}\right).\end{aligned}$$

Now obviously

$$\sup_{l'\geq k_n+h_n} \sqrt{\lambda_{l'}} \left(1+\sqrt{k_n \log k_n}\right) \leq K,$$

since $\sqrt{\lambda_{l'} l' \log l'} \to 0$ from which we deduce that

$$\frac{2M}{k_n}\sum_{l=1}^{k_n}\left(\sum_{l'\geq k_n+h_n}^{+\infty} \langle \rho, e_{l'}\rangle \sqrt{\lambda_{l'}} \frac{1}{1-\frac{l}{l'}}\right)^2 \leq \frac{2MK}{k_n}\sum_{l=1}^{k_n}\left(\sum_{l'\geq k_n+h_n}^{+\infty} |\langle \rho, e_{l'}\rangle|\right)^2.$$

When assumption (A.3) holds, Cesaro's mean Theorem ensures that the term on the left above tends to zero. We turn to the first term in equation (26)

$$\begin{aligned}\left(\sum_{l'\geq k_n+1}^{k_n+h_n} |\langle \rho, e_{l'}\rangle| \sqrt{\lambda_{l'}} \frac{1}{1-\frac{l}{l'}}\right)^2 &\leq h_n^2 \max_{\substack{k_n+1\leq l'\leq k_n+h_n,\\ 1\leq l\leq k_n}} \left\{|\langle \rho, e_{l'}\rangle| \sqrt{\lambda_{l'}} \frac{1}{1-\frac{l}{l'}}\right\}^2 \\ &\leq \frac{k_n}{\log k_n}\lambda_{k_n} k_n^2 \max_{k_n+1\leq l'\leq k_n+h_n,} \left(|\langle \rho, e_{l'}\rangle|^2\right).\end{aligned}$$

Now $\lambda_{k_n} k_n$ as well as $k_n \max_{k_n+1\leq l'\leq k_n+h_n,} (|\langle \rho, e_{l'}\rangle|)$ tend to zero when assumption (A.3) holds. We get once more

$$\frac{2M}{k_n}\sum_{l=1}^{k_n}\left(\sum_{l'\geq k_n+1}^{k_n+h_n} |\langle \rho, e_{l'}\rangle| \sqrt{\lambda_{l'}} \frac{1}{1-\frac{l}{l'}}\right)^2 \to 0.$$

A similar truncating technique would prove that the term in (25) also tends to zero as $n$ goes to infinity which leads to

$$\frac{n}{k_n}\mathbb{E}\left\langle \mathcal{S}_n \rho, X_{n+1}\right\rangle^2 \to 0. \tag{27}$$

In order to finish the proof of the Proposition we must deal with the term introduced in (22). We have the following result

$$\sqrt{\frac{n}{k_n}}\left|\left\langle \mathcal{R}_n \rho, X_{n+1}\right\rangle\right| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}} k_n^{5/2} (\log k_n)^2\right), \tag{28}$$

when $\frac{k_n^2 \log k_n}{\sqrt{n}} \to 0$. Indeed, consider

$$T_{j,n} = \int_{\mathcal{B}_j} \left[(zI-\Gamma)^{-1}(\Gamma_n-\Gamma)(zI-\Gamma)^{-1}(\Gamma_n-\Gamma)(zI-\Gamma_n)^{-1}\right] \mathrm{d}z.$$

Then setting

$$G_n(z) = (zI - \Gamma)^{-1/2} (\Gamma_n - \Gamma) (zI - \Gamma)^{-1/2},$$

we have

$$\begin{aligned}
&\left|\left\langle T_{j,n}\rho, X_{n+1}\right\rangle\right| \\
&\quad = \left| \int_{\mathcal{B}_j} \left\langle (zI - \Gamma)^{-1/2} (\Gamma_n - \Gamma) (zI - \Gamma)^{-1} (\Gamma_n - \Gamma) (zI - \Gamma_n)^{-1} \rho, \right.\right. \\
&\qquad \left.\left. (zI - \Gamma)^{-1/2} X_{n+1} \right\rangle \mathrm{d}z \right| \\
&\quad \leq \int_{\mathcal{B}_j} \left| \left\langle G_n^2(z) (zI - \Gamma)^{1/2} (zI - \Gamma_n)^{-1} (zI - \Gamma)^{1/2} (zI - \Gamma)^{-1/2} \rho, \right.\right. \\
&\qquad \left.\left. (zI - \Gamma)^{-1/2} X_{n+1} \right\rangle \right| \mathrm{d}z \\
&\quad \leq \int_{\mathcal{B}_j} \|G_n(z)\|_\infty^2 \left\| (zI - \Gamma)^{1/2} (zI - \Gamma_n)^{-1} (zI - \Gamma)^{1/2} \right\|_\infty \\
&\qquad \times \left\| (zI - \Gamma)^{-1/2} X_{n+1} \right\| \left\| (zI - \Gamma)^{-1/2} \rho \right\| \mathrm{d}z. \qquad (29)
\end{aligned}$$

Following Lemma 4, the random variable $\left\| (zI - \Gamma)^{1/2} (zI - \Gamma_n)^{-1} (zI - \Gamma)^{1/2} \right\|_\infty$ is decomposed in two terms

$$\left\| (zI - \Gamma)^{1/2} (zI - \Gamma_n)^{-1} (zI - \Gamma)^{1/2} \right\|_\infty \left( \mathbb{1}_{\mathcal{E}_j(z)} + \mathbb{1}_{\mathcal{E}_j^c(z)} \right).$$

On the one hand when $\mathcal{E}_j(z)$ holds it was proved in Lemma 4 that

$$\|(zI - \Gamma)\|_\infty \left\| (zI - \Gamma_n)^{-1} \right\|_\infty \leq C. \qquad (30)$$

On the other hand when $\mathcal{E}_j^c(z)$ holds we may write for all $\eta > 0$ thanks to bound (17)

$$\mathbb{P}\left( \left|\left\langle T_{j,n}\rho, X_{n+1}\right\rangle\right| \mathbb{1}_{\mathcal{E}_j^c(z)} > \eta \right) \leq \mathbb{P}\left( \mathcal{E}_j^c(z) \right) \leq \frac{M}{\sqrt{n}} (j \log j),$$

which entails that

$$\mathbb{P}\left(\sum_{j=1}^{k_n}\left|\left\langle T_{j,n}\rho, X_{n+1}\right\rangle\right| \mathbb{1}_{\mathcal{E}_j^c(z)} > \eta\right)$$
$$\leq M\sum_{j=1}^{k_n}\frac{1}{\sqrt{n}}(j\log j) \leq \frac{k_n^2\log k_n}{\sqrt{n}} \to 0.$$

Consequently we can deal with all $T_{j,n}$ as if the event $\mathcal{E}_j(z)$-hence the bound (30)-holds almost surely. We take expectation and note that $G_n(z)$ and $X_{n+1}$ are independent

$$\mathbb{E}\left|\left\langle T_{j,n}\rho, X_{n+1}\right\rangle\right| \leq C\int_{\mathcal{B}_j}\mathbb{E}\left\|G_n(z)\right\|_\infty^2\mathbb{E}\left\|(zI-\Gamma)^{-1/2}X_{n+1}\right\|\left\|(zI-\Gamma)^{-1/2}\rho\right\|\mathrm{d}z.$$

By Lemma 3 we have

$$\mathbb{E}\left|\left\langle T_{j,n}\rho, X_{n+1}\right\rangle\right| \leq \frac{C}{n}\mathrm{diam}\left(\mathcal{B}_j\right)\cdot(j\log j)^{5/2}\sup_{z\in\mathcal{B}_j}\left\|(zI-\Gamma)^{-1/2}\rho\right\|$$
$$\leq C\sqrt{\delta_j}\cdot(j\log j)^{5/2}\|\rho\| \leq C(j\log j)^2,$$

since $\delta_j \leq C(j\log j)^{-1}$ at least for a sufficiently large $j$. Finally summing over all the $j$'s from 1 to $k_n$ leads to

$$\sqrt{\frac{n}{k_n}}\mathbb{E}\left|\left\langle \mathcal{R}_n\rho, X_{n+1}\right\rangle\right| \leq C\frac{1}{\sqrt{nk_n}}\sum_{j=1}^{k_n}(j\log j)^2 \leq \frac{C}{\sqrt{n}}k_n^{5/2}(\log k_n)^2$$

which proves (28) and achieves the proof of the proposition. □

The methods used to prove the next Proposition are close to those developed above.

**Proposition 3** *If* $\frac{1}{\sqrt{n}}k_n^{5/2}(\log k_n)^2 \to 0$, *then*

$$\sqrt{\frac{n}{k_n}}\left|\left\langle\left(\Gamma_n^\dagger - \Gamma^\dagger\right)U_n, X_{n+1}\right\rangle\right| \xrightarrow{\mathbb{P}} 0.$$

*Besides if $x$ is a fixed vector in $H$ such that* $\sup_p \frac{|\langle x, e_p\rangle|^2}{\lambda_p} < +\infty$ *and* $\frac{k_n^3(\log k_n)^2}{t_{n,x}\sqrt{n}} \to 0$,

$$\frac{\sqrt{n}}{t_{n,x}}\left|\left\langle\left(\Gamma_n^\dagger - \Gamma^\dagger\right)U_n, x\right\rangle\right| \xrightarrow{\mathbb{P}} 0.$$

*Proof* Once again we develop the expression above by means of complex integrals for operator-valued analytic functions. Hence

$$\Gamma_n^{\dagger}-\Gamma^{\dagger}=\frac{1}{2\pi\iota}\int_{\mathcal{C}_n}\widetilde{f}_n(z)\left[(zI-\Gamma_n)^{-1}(\Gamma-\Gamma_n)(zI-\Gamma)^{-1}\right]\mathrm{d}z$$
$$=\sum_{j=1}^{k_n}\frac{1}{2\pi\iota}\int_{\mathcal{B}_j}\widetilde{f}_n(z)\left[(zI-\Gamma_n)^{-1}(\Gamma-\Gamma_n)(zI-\Gamma)^{-1}\right]\mathrm{d}z,$$

and

$$\left|\left\langle\left(\Gamma_n^{\dagger}-\Gamma^{\dagger}\right)U_n,X_{n+1}\right\rangle\right|\leq C\sum_{j=1}^{k_n}H_{j,n},$$

where

$$H_{j,n}=\int_{\mathcal{B}_j}\left|\widetilde{f}_n(z)\left\langle(zI-\Gamma)^{1/2}(zI-\Gamma_n)^{-1}(zI-\Gamma)^{1/2}G_n\right.\right.$$
$$\left.\left.(z)(zI-\Gamma)^{-1/2}U_n,(zI-\Gamma)^{-1/2}X_{n+1}\right\rangle\right|\mathrm{d}z.$$

We copy verbatim the arguments used to bound (29) : first of all we reintroduce the operator $G_n(z)$ below and

$$(zI-\Gamma)^{1/2}(zI-\Gamma_n)^{-1}(zI-\Gamma)^{1/2},$$

remains almost surely bounded by a constant which does not depend on $n$ or $j$ plus a negligible term as was proved just below (30). Hence

$$H_{j,n}\leq C\int_{\mathcal{B}_j}\left|\widetilde{f}_n(z)\right|\|G_n(z)\|\left\|(zI-\Gamma)^{-1/2}U_n\right\|\left\|(zI-\Gamma)^{-1/2}X_{n+1}\right\|\mathrm{d}z.$$

We take expectation

$$\mathbb{E}H_{j,n}\leq C\int_{\mathcal{B}_j}\left|\widetilde{f}_n(z)\right|\mathbb{E}\left(\|G_n(z)\|\left\|(zI-\Gamma)^{-1/2}U_n\right\|\right)\mathbb{E}\left\|(zI-\Gamma)^{-1/2}X_{n+1}\right\|\mathrm{d}z$$
$$\leq C\mathrm{diam}\left(\mathcal{B}_j\right)\sup_{z\in\mathcal{B}_j}\left(\left|\widetilde{f}_n(z)\right|\mathbb{E}\left\|(zI-\Gamma)^{-1/2}X_{n+1}\right\|\sqrt{\mathbb{E}\|G_n(z)\|^2}\right.$$
$$\left.\times\sqrt{\mathbb{E}\left\|(zI-\Gamma)^{-1/2}U_n\right\|^2}\right),$$

where Cauchy–Schwarz inequality was applied. Now invoking Lemma 3 yields

$$\mathbb{E}H_{j,n} \leq C\frac{\mathrm{diam}\,(\mathcal{B}_j)}{\sqrt{n}}\,(j\log j)^{3/2}\sup_{z\in\mathcal{B}_j}\left(\left|\widetilde{f}_n\,(z)\right|\sqrt{\mathbb{E}\left\|(zI-\Gamma)^{-1/2}\,U_n\right\|^2}\right).$$

Obviously

$$\begin{aligned}\mathbb{E}\left\|(zI-\Gamma)^{-1/2}\,U_n\right\|^2 &= \frac{\sigma_\varepsilon^2}{n}\mathbb{E}\left\|(zI-\Gamma)^{-1/2}\,X_1\right\|^2\\ &= \frac{\sigma_\varepsilon^2}{n}\sum_{l=1}^{+\infty}\frac{\lambda_l}{|z-\lambda_l|},\end{aligned}$$

hence

$$\sup_{z\in\mathcal{B}_j}\left(\sqrt{\mathbb{E}\left\|(zI-\Gamma)^{-1/2}\,U_n\right\|^2}\right)\leq\frac{1}{\sqrt{n}}\,(j\log j)^{1/2}\,.$$

At last

$$\mathbb{E}H_{j,n}\leq C\frac{\delta_j}{\lambda_j n}\,(j\log j)^2\leq\frac{C}{n}\,(j\log j)^2\,,$$

and

$$\mathbb{E}\left|\left\langle\left(\Gamma_n^{\dagger}-\Gamma^{\dagger}\right)U_n,X_{n+1}\right\rangle\right|\leq\frac{C}{n}k_n^3\,(\log k_n)^2\,,$$

which proves the first part of the Proposition. Replacing $X_{n+1}$ with a fixed $x$ in $H$, means replacing $\mathbb{E}\left\|(zI-\Gamma)^{-1/2}\,X_{n+1}\right\|$ with

$$\left\|(zI-\Gamma)^{-1/2}\,x\right\|\leq\sqrt{\sum_{p=1}^{+\infty}\frac{\langle x,e_p\rangle^2}{\left|z-\lambda_p\right|}}\leq\sqrt{\sup_p\frac{\left|\langle x,e_p\rangle\right|}{\lambda_p}}\sqrt{\sum_{p=1}^{+\infty}\frac{\lambda_p}{\left|z-\lambda_p\right|}},$$

and the derivation of the second part of the Proposition stems from the first part. □

### 6.3 Weakly convergent terms

This subsection is quite short but was separated from the others for the sake of clarity and in order to give a logical structure to the proofs.

**Lemma 8** *We have*

$$\sqrt{\frac{n}{t_{n,x}}}\langle R_n, x\rangle \xrightarrow{w} N\left(0,\sigma_\varepsilon^2\right), x \in H,$$

*and*

$$\sqrt{\frac{n}{s_n}}\left\langle R_n, X_{n+1}\right\rangle \xrightarrow{w} N\left(0,\sigma_\varepsilon^2\right).$$

*Proof* We have

$$\langle R_n, x\rangle = \left\langle \Gamma^\dagger U_n, x\right\rangle = \frac{1}{n}\sum_{i=1}^{n}\left\langle \Gamma^\dagger X_i, x\right\rangle \varepsilon_i,$$

which is an array – $\Gamma^\dagger$ implicitly depends on $n$ – of independent real r.v. The Central Limit Theorem holds for this sequence and leads to the first announced result. We turn to the second display

$$\begin{aligned}\left\langle R_n, X_{n+1}\right\rangle &= \left\langle \Gamma^\dagger U_n, X_{n+1}\right\rangle \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\langle \Gamma^\dagger X_i, X_{n+1}\right\rangle \varepsilon_i = \sum_{i=1}^{n} Z_{i,n}.\end{aligned}$$

Denoting by $\mathcal{F}_i$ the $\sigma$-algebra generated by $(X_1, \varepsilon_1, ..., X_i, \varepsilon_i)$, we see that $Z_{i,n}$ is a martingale difference sequence w.r.t. $\mathcal{F}_i$. Also note that

$$\mathbb{E}\left(Z_{i,n}^2|\mathcal{F}_i\right) = \frac{\varepsilon_i^2}{n^2}\left\|\Gamma^{1/2}\Gamma^\dagger X_i\right\|^2,$$

and that

$$\begin{aligned}\mathbb{E}\left[\varepsilon_i^2\left\|\Gamma^{1/2}\Gamma^\dagger X_i\right\|^2\right] &= \mathbb{E}\left[\left\|\Gamma^{1/2}\Gamma^\dagger X_i\right\|^2 \mathbb{E}\left(\varepsilon_i^2|X_i\right)\right] \\ &= \sigma_\varepsilon^2 \mathbb{E}\left\|\Gamma^{1/2}\Gamma^\dagger X_i\right\|^2 \\ &= \sigma_\varepsilon^2 \sum_{j=1}^{k_n}\left[\lambda_j f_n\left(\lambda_j\right)\right]^2 = \sigma_\varepsilon^2 s_n^2.\end{aligned}$$

Applying the Central Limit Theorem for real valued martingale difference arrays (see e.g. [23]) we get the second result. □

### 6.4 Proofs of the main results

The careful reader has noted that within the preceding steps of the proofs $s_n^2$ was replaced with $k_n$ in the normalizing sequence. Very simple computations prove that under (A.2) and if $k_n/\sqrt{n}$ tends to zero this permutation is possible in the sense that $s_n^2$ and $k_n$ are asymptotically equivalent. enough to prove that $k_n \geq Cs_n^2$ for some constant $C$). Using this argument, we give now the proof of main results of the paper, some of them being straightforward consequences of Lemmas and Propositions established above.

*Proof of Theorem 1* From (11) and all that was made above it suffices to prove that the Theorem holds with $U_n$ replacing $\widehat{\rho} - \widehat{\Pi}_{k_n}\rho$. Now suppose that for a given normalizing sequence $\alpha_n > 0$, $\alpha_n U_n$ converges weakly in the norm topology of $H$. For all $x$ in $H$, $\alpha_n \langle U_n, x\rangle$ converges weakly too and

$$\alpha_n \langle U_n, x\rangle = \frac{\alpha_n}{n} \sum_{i=1}^{n} \left\langle X_i, \Gamma^{\dagger} x\right\rangle \varepsilon_i,$$

is an array of real independent random variable. Suppose that $x$ belongs to the domain of $\Gamma^{-1}$, namely that

$$\sum_{j=1}^{+\infty} \frac{\left\langle x, e_j\right\rangle^2}{\lambda_j^2} < +\infty,$$

then

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\langle X_j, \Gamma^{\dagger} x\right\rangle \varepsilon_j \xrightarrow{w} N\left(0, \beta_x \sigma_\varepsilon^2\right),$$

where $\beta_x$ depends on $x$ and on the eigenvalues of $\Gamma$. Consequently $\alpha_n = \sqrt{n}$. Now if $\sum_j \left\langle x, e_j\right\rangle^2 / \lambda_j^2$ is divergent, $\mathbb{E}(\langle X_i, \Gamma^{\dagger} x\rangle^2 \varepsilon_i^2) \uparrow +\infty$ and $\alpha_n \langle U_n, x\rangle$ cannot converge in distribution. This finishes the proof of the Theorem. □

*Proof of Theorem 2* The proof of the first part of Theorem 2 stems from the decomposition (11), Lemma 6, Proposition 2, the first part of Proposition 3 and from the second result of Lemma 8.

To prove the second part of Theorem 2, it suffices to apply Lemma 7 with condition (8) and the suitable assumption on the eigenelements of $\Gamma$. □

*Proof of Corollary 1* It suffices to prove that $\frac{\left|\widehat{s}_n^2 - s_n^2\right|}{s_n^2} \xrightarrow{\mathbb{P}} 0$, or equivalently that

$$\frac{\sum_{j=1}^{k_n} \left|\lambda_j f_n\left(\lambda_j\right) - \widehat{\lambda}_j f_n\left(\widehat{\lambda}_j\right)\right| \left(\lambda_j f_n\left(\lambda_j\right) + \widehat{\lambda}_j f_n\left(\widehat{\lambda}_j\right)\right)}{\sum_{j=1}^{k_n} \left[\lambda_j f_n\left(\lambda_j\right)\right]^2} \xrightarrow{\mathbb{P}} 0.$$

Clearly since $\sup_{j\in\mathbb{N}}\left|\widehat{\lambda}_j-\lambda_j\right| = O_P\left(1/\sqrt{n}\right)$ and $xf_n(x)$ is bounded for $x > c_n$ it is enough to get

$$\frac{\sum_{j=1}^{k_n}\left|\lambda_j f_n\left(\lambda_j\right)-\widehat{\lambda}_j f_n\left(\widehat{\lambda}_j\right)\right|}{\sum_{j=1}^{k_n}\left[\lambda_j f_n\left(\lambda_j\right)\right]^2} \xrightarrow{\mathbb{P}} 0. \tag{31}$$

But by assumption (H.3)

$$\begin{aligned}\sum_{j=1}^{k_n}\left|\lambda_j f_n\left(\lambda_j\right)-\widehat{\lambda}_j f_n\left(\widehat{\lambda}_j\right)\right| &\le \sum_{j=1}^{k_n}\left|\lambda_j f_n\left(\lambda_j\right)-1\right| + \sum_{j=1}^{k_n}\left|1-\widehat{\lambda}_j f_n\left(\widehat{\lambda}_j\right)\right| \\ &= o_P\left(k_n/\sqrt{n}\right),\end{aligned}$$

and $k_n/\sqrt{n} \to 0$. □

*Proof of Theorem 3* The proof of Theorem 3 stems from (11), Proposition 3 and from Lemma 8. □

*Proof of Corollary 2* We have to prove that

$$\frac{\widehat{t}_{n,x}^2 - t_{n,x}^2}{t_{n,x}^2} = \frac{\sum_{j=1}^{k_n}\widehat{\lambda}_j\left[f_n\left(\widehat{\lambda}_j\right)\right]^2\left\langle x,\widehat{e}_j\right\rangle^2 - \lambda_j\left[f_n\left(\lambda_j\right)\right]^2\left\langle x,e_j\right\rangle^2}{\sum_{j=1}^{k_n}\lambda_j\left[f_n\left(\lambda_j\right)\right]^2\left\langle x,e_j\right\rangle^2} \xrightarrow{\mathbb{P}} 0.$$

We split the expression into two terms

$$\begin{aligned} w_{n1} &= \frac{\sum_{j=1}^{k_n}\left(\widehat{\lambda}_j\left[f_n\left(\widehat{\lambda}_j\right)^2\right]-\lambda_j\left[f_n\left(\lambda_j\right)\right]^2\right)\left\langle x,\widehat{e}_j\right\rangle^2}{\sum_{j=1}^{k_n}\lambda_j\left[f_n\left(\lambda_j\right)\right]^2\left\langle x,e_j\right\rangle^2}, \\ w_{n2} &= \frac{\sum_{j=1}^{k_n}\lambda_j\left[f_n\left(\lambda_j\right)\right]^2\left(\left\langle x,\widehat{e}_j\right\rangle^2-\left\langle x,e_j\right\rangle^2\right)}{\sum_{j=1}^{k_n}\lambda_j\left[f_n\left(\lambda_j\right)\right]^2\left\langle x,e_j\right\rangle^2}.\end{aligned}$$

Copying what was done for the proof of Corollary 1, we can easily prove that $w_{n1} \xrightarrow{\mathbb{P}} 0$. In order to alleviate formulas and displays, we are going to prove that $w_{n2} \xrightarrow{\mathbb{P}} 0$ in the special case when $f_n\left(\lambda_j\right) = 1/\lambda_j$. The general situation stems easily from this special case. Thus, we have now

$$w_{n2} = \frac{\sum_{j=1}^{k_n}\left(\left\langle x,\widehat{e}_j\right\rangle^2-\left\langle x,e_j\right\rangle^2\right)/\lambda_j}{\sum_{p=1}^{k_n}\left\langle x,e_j\right\rangle^2/\lambda_j}.$$

We denote by $\widehat{\pi}_j$ the projector on the eigenspace associated to the $j^{th}$ eigenfunction of $\Gamma_n$. Then, with this notation, we can write $\left\langle x,\widehat{e}_j\right\rangle^2-\left\langle x,e_j\right\rangle^2 = \left\|\widehat{\pi}_j x\right\|_\infty^2 - \left\|\pi_p x\right\|_\infty^2 = \left\langle\left(\widehat{\pi}_j-\pi_j\right)x;x\right\rangle$ and we have

$$\left|\langle x,\widehat{e}_j\rangle^2 - \langle x,e_j\rangle^2\right| \leq \left\|\widehat{\pi}_j - \pi_j\right\|_\infty \|x\|^2,$$
$$\widehat{\pi}_j - \pi_j = \frac{1}{2\pi\iota}\int_{\mathcal{B}_j}\left[(zI-\Gamma_n)^{-1} - (zI-\Gamma)^{-1}\right]\mathrm{d}z$$
$$= \frac{1}{2\pi\iota}\int_{\mathcal{B}_j}\left[(zI-\Gamma_n)^{-1}(\Gamma_n-\Gamma)(zI-\Gamma)^{-1}\right]\mathrm{d}z,$$

and

$$\mathbb{E}\left\|\widehat{\pi}_j - \pi_j\right\|_\infty \leq C\frac{j\log j}{\sqrt{n}}.$$

Finally

$$|w_{n2}| \leq C\frac{1}{\sqrt{n}}\sum_{j=1}^{k_n} j\log j \leq C\frac{k_n^2\log k_n}{\sqrt{n}} \to 0,$$

which finishes the proof of the Corollary. □

*Proof of Proposition 1* Take $x = \sum x_i e_i$ and $\rho = \sum \rho_i e_i$ in $H$. Obviously, it suffices to prove that the Proposition holds when $\widehat{\Pi}_{k_n} - \Pi_{k_n}$ is replaced with

$$\varphi_{k_n}\left((\Gamma_n-\Gamma)\right) = \sum_{j=1}^{k_n}\left[\mathcal{S}_j(\Gamma_n-\Gamma)\Pi_j + \Pi_j(\Gamma_n-\Gamma)\mathcal{S}_j\right].$$

Following Dauxois et al. [9] p 143–144, we can check that when $X_1$ is Gaussian, $\sqrt{n}(\Gamma_n-\Gamma)$ converges weakly to the Gaussian random operator $G$ defined by

$$G = \sum_{j\leq j'}\sqrt{\lambda_j\lambda_{j'}}\xi_{j,j'}\left(e_j\otimes e_{j'} + e_{j'}\otimes e_j\right) + \sqrt{2}\sum_j \lambda_j\left(e_j\otimes e_{j'}\right)\xi_{j,j},$$

where $\xi_{j,j'}$'s are i.i.d. Gaussian centered r.r.v. with variance equal to 1. Thus, we replace once more $\sqrt{n}(\Gamma_n-\Gamma)$ with $G$ (the situation is indeed the same as if the operator $X_1\otimes X_1$ was assumed to be Gaussian). We are going to prove that $\frac{\langle\varphi_{k_n}(G)\rho,x\rangle}{t_{n,x}}$ is not bounded in probability, whatever the sequence $k_n\to+\infty$, by choosing a special $\rho$. We focus on the *jth* term of the above sum.

The exact computation of $<\left(\Pi_j G\mathcal{S}_j + \mathcal{S}_j G\Pi_j\right)(x), \rho>$ may be deduced from Dauxois et al. [9] p. 146. Assuming that all the $\lambda_j$'s have all the same order of multiplicity equals to 1, we easily get

$$\left\langle \left(\Pi_j G\mathcal{S}_j + \mathcal{S}_j G\Pi_j\right)(x), \rho\right\rangle = \sum_{l\neq j} \frac{\sqrt{\lambda_l\lambda_j}}{\lambda_j - \lambda_l}\left(x_j\rho_l + x_l\rho_j\right)\xi_{jl}.$$

The previous sum is a real centered Gaussian random variable with variance

$$\sum_{l\neq j} \frac{\lambda_l\lambda_j}{\left(\lambda_j - \lambda_l\right)^2}\left(x_j\rho_l + x_l\rho_j\right)^2.$$

Summing over $j$ provides the variance of $\left\langle \varphi_{k_n}(G)\rho, x\right\rangle$

$$\begin{aligned}\sum_{j=1}^{k_n}\sum_{l\neq j} \frac{\lambda_l\lambda_j}{\left(\lambda_j - \lambda_l\right)^2}\left(x_j\rho_l + x_l\rho_j\right)^2 &\geq \sum_{j=1}^{k_n}\lambda_j\rho_j^2\sum_{l\neq j}\frac{\lambda_l x_l^2}{\left(\lambda_j - \lambda_l\right)^2}\\ &\geq \sum_{j=1}^{k_n}\lambda_j\rho_j^2\sum_{l=1}^{j-1}\frac{\lambda_l x_l^2}{\left(\lambda_j - \lambda_l\right)^2}.\end{aligned}$$

For the sake of simplicity we assume that $x_k > 0$ and $\rho_k > 0$. Now if $x_l^2 = l^{-1-\beta}$ and $\lambda_l = l^{-1-\alpha}$ the computation of the second sum stems from

$$\sum_{l=1}^{j-1}\frac{\lambda_l x_l^2}{\left(\lambda_j - \lambda_l\right)^2} \sim \int_1^{j-1}\frac{s^{\alpha-\beta}}{\left(1-\left(\frac{s}{j}\right)^{1+\alpha}\right)^2}\mathrm{d}s \sim Cj^{2+\alpha-\beta}.$$

Finally

$$\sum_{j=1}^{k_n}\lambda_j\rho_j^2\sum_{l=1}^{j-1}\frac{\lambda_l x_l^2}{\left(\lambda_j - \lambda_l\right)^2} \geq C\sum_{j=1}^{k_n} j^{1-\beta}\rho_j^2 \to +\infty.$$

We see that the variance of $\frac{\langle \varphi_{k_n}(G)\rho, x\rangle}{t_{n,x}}$ explodes and that this random variable cannot converge in distribution. □

**Acknowledgements** The authors would like to thank the organizers and participants of the working group STAPH on functional statistics in Toulouse for fruitful discussions.

## References

1. Araujo, A., Giné, E.: The central limit theorem for real and Banach valued random variables. Wiley Series in Probability and Mathematical Statistics, Wiley New York (1980)
2. Arsenin, V., Tikhonov, A.: Solutions of ill-posed problems. Winston and Sons, Washington DC (1977)

3. Bosq, D.: Linear processes in function spaces. Lecture notes in statistics, 149, Springer, Berlin Heidelberg New York (2000)
4. Cardot, H., Ferraty, F., Mas, A., Sarda, A.: Testing hypotheses in the functional linear model. Scand. J. Stat. **30**, 241–255 (2003)
5. Cardot, H., Ferraty, F., Sarda, P.: Functional linear model. Stat. Probab. Lett. **45**, 11–22 (1999)
6. Cardot, H., Ferraty, F., Sarda, P.: Spline estimators for the functional linear model. Stat. Sin. **13**, 571–591 (2003)
7. Cardot, H., Mas, A., Sarda, P.: CLT in functional linear regression models : application to confidence sets for prediction. Technical Report, University Montpellier II, Dept. of Mathematics, 42 p. Available at http://www.math.univ-montp2.fr/∼mas/rapp-inra-um2.pdf (2004)
8. Cuevas, A., Febrero, M., Fraiman, R.: Linear functional regression: The case of fixed design and functional response. Can. J. Stat. **30**, 285–300 (2002)
9. Dauxois, J., Pousse, A., Romain, Y.: Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. J. Mult. Anal. **12**, 136–154 (1982)
10. Dedecker, J., Merlevède, F.: Necessary and sufficient conditions for the conditional central limit theorem. Ann. Probab. **30**, 1044–1081 (2002)
11. Dunford, N., Schwartz, J.T.: Linear Operators, Vol. I & II. Wiley Classics Library (1988)
12. Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis, Theory and Applications. Springer Series in Statistics, Springer, Berlin Heidelberg New York (2006)
13. Frank, I.E., Friedman, J.H.: A statistical view of some chemometrics regression tools (with discussion). Technometrics **35**, 109–148 (1993)
14. Gohberg, I., Goldberg, S., Kaashoek, M.A.: Classes of linear operators. Vol I & II. Operator Theory : advances and applications. Birkhaüser Verlag (1991)
15. Goutis, C.: Second-derivative functional regression with applications to near infra-red spectroscopy. J. R. Stat. Soc. Ser. B **60**, 103–114 (1998)
16. Groetsch, C.: Inverse Problems in the Mathematical Sciences. Vieweg, Wiesbaden (1993)
17. Hajek, J.: On linear statistical problems in stochastic processes. Chekoslovak Math. J. **12**, 404–443 (1962)
18. Hastie, T., Mallows, C.: A discussion of "A statistical view of some chemometrics regression tools" by I.E. Frank and J.H. Friedman. Technometrics **35**, 140–143 (1993)
19. He, G., Müller, H.G., Wang, J.L.: Functional canonical analysis for square integrable stochastic processes. J. Multi. Anal. **85**, 54–77 (2003)
20. Kneip, A.: Nonparametric estimation of common regressors for similar curve data. Ann Stat. **22**, 1386–1427 (1995)
21. Kneip, A., Utikal, K.J.: Inference for density families using functional principal component analysis. J. Am. Stat. Ass. **96**, 519–542 (2001)
22. Ledoux, M., Talagrand, M.: Probability in Banach spaces. Isoperimetry and processes. Springer, Berlin Heidelberg New York (1991)
23. Mc Leish, D.L.: Dependent central limit theorem and invariance principles. Ann. Probab. **2**, 620–628 (1974)
24. Mas, A.: Estimation d'opérateurs de corrélation de processus fonctionnels : lois limites, tests, déviations modérées (Estimation of correlation operators for functional processes: weak convergence, tests, moderate deviations). Ph. D. Thesis, University Paris 6 (1999)
25. Mas, A.: Weak convergence in the functional autoregressive model. J. Multi. Anal. (2006) (to appear)
26. Mas, A., Menneteau, L.: Perturbation approach applied to the asymptotic study of random operators. Progr. Probab. **55**, 127–123 (2003)
27. Merlevède, F.: On the central limit theorem and its weak invariance via martingale approximation. J. Theoret. Probab. **3**, 625–653 (2003)
28. Müller, H.-G., Stadtmüller, U.: Genralized functional linear models. Ann. Stat. **33**, 774–805 (2005)
29. Nashed, M.Z., Wahba, G.: Generalized inverses in reproducing kernel spaces: An appraoch to regularization of linear operator equations. SIAM J. Math. Anal. **6**, 974–987 (1974)
30. Ramsay, J.O., Dalzell, B.W.: Some tools for functional data analysis (with discussion), J. R. Stat. Soc. B **53**, 539–572 (1991)
31. Ramsay, J.O., Silverman, B.W.: Applied Functional Data Analysis: Methods and Case Studies. Springer, Berlin Heidelberg New York (2002)

32. Ramsay, J.O., Silverman, B.W.: Functional Data Analysis, 2nd edn. Springer, Berlin Heidelberg New York (2005)
33. Yao, F., Müller, H.-G., Wang, J.-L.: Functional data analysis for sparse longitudinal data. J. Am. Stat. Assoc. **100**, 577–590 (2005)