



HAL
open science

Weak convergence in the functional autoregressive model

André Mas

► **To cite this version:**

André Mas. Weak convergence in the functional autoregressive model. *Journal of Multivariate Analysis*, 2007, 98 (6), pp.1231-1261. 10.1016/j.jmva.2006.05.010 . hal-00008621

HAL Id: hal-00008621

<https://hal.science/hal-00008621>

Submitted on 12 Sep 2005

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.

Weak convergence in the functional autoregressive model.

André Mas*

Abstract

The functional autoregressive model is a Markov model tailored for data of functional nature. It revealed fruitful when attempting to model samples of dependent random curves and has been widely studied along the past few years. This article aims at completing the theoretical study of the model by addressing the crucial issue of weak convergence for estimates from the model. The main difficulties stem from an underlying inverse problem as well as from dependence between the data. Traditional facts about weak convergence in non parametric models appear : the normalizing sequence is not an $o(\sqrt{n})$, a bias term appears. Several original features of the functional framework are pointed out.

Keywords : Functional data, autoregressive model, Hilbert space, weak convergence, random operator, perturbation theory, linear inverse problem, martingale difference arrays.

1 Introduction

1.1 The model and its history

The Functional Autoregressive Model of order 1 (FAR1) generalizes to random elements with values in an infinite dimensional space the classical AR(1) model belonging to the celebrated class of ARMA process, widely used in time series analysis. This model was introduced by Bosq [9], then studied by several authors. Several chapters in Bosq [10] are dedicated to a thorough study of this strictly stationary process $(X_n)_{n \in \mathbb{Z}}$ defined by

$$X_n - m = \rho(X_{n-1} - m) + \varepsilon_n, \quad n \in \mathbb{Z}, \quad (1)$$

where the X_k 's and the ε_k 's are random elements with values in an infinite dimensional vector space \mathcal{E} , ρ is an unknown linear operator from \mathcal{E} to \mathcal{E} and $m \in \mathcal{E}$ is the expectation of the process. In all the following we will assume that for all n ε_n is independent of X_{n-1} . The process $(X_n)_{n \in \mathbb{Z}}$ is Markov whenever the ε_n 's are such that $\mathbb{E}(\varepsilon_n | X_{n-1}) = 0$ where \mathbb{E} denotes expectation.

The model was extended in Mourid [24] considering autoregressive processes of higher orders. Besse and Cardot [6] proved that the model is adapted to splines techniques. Then Pumo [25] studied autoregressive processes with values in the Banach space of continuous functions on $[0, 1]$. The PhD Thesis by Mas [20] was partly devoted to the topic. Besse, Cardot and Stephenson [7]

*Institut de Modélisation Mathématiques de Montpellier, UMR 5149, Université Montpellier 2, CC051, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France. (Phone : +33 467143956, Fax : +33 467143558) mas@math.univ-montp2.fr

developped a method based on kernels. Recently Mas and Menneteau [22] announced large and moderate deviations theorems for the process or its covariance sequence whereas Antoniadis and Sapatinas [3] implemented wavelet methods which considerably improved the prevision mean square error. Even more recently Menneteau [23] proved laws of the iterated logarithm for statistics arising from functional PCA of the process.

The model revealed fruitful in several areas of applied statistics : electrical engineering (Cavallini et alii, [12]), climatology (Besse et alii, [7], Antoniadis and Sapatinas [3]), medicine (Marion and Pumo, [18]).

The main interest of (1) relies in its predictive power. Estimating the correlation operator only aims at providing an estimate, say ρ_n yielding a predictor for the unknown X_{n+1} , $\rho_n(X_n)$ based on the sample (X_1, \dots, X_n) .

However if convergence of $\rho_n(X_n)$ to $\rho(X_n)$ for instance was often studied either in probability or almost surely, the issue of weak convergence has not been truly tackled yet. An attempt was proposed in Mas [19] but the conditions under which the result holds are extremely restricting. The problem of weak convergence is especially intricate due to the functional framework and to an underlying inverse problem (see next section). A weak convergence result implies obtaining the sharpest rate for convergence in probability. Authors studying rates of convergence for the predictor usually just give bounds... Besides a weak convergence result would be of much help in getting confidence sets for $\rho(X_n) = E(X_{n+1}|X_n, X_{n-1}, \dots)$. Maybe a bootstrap procedure could be proposed to achieve the same goal but on a one hand I did not find any real and reliable bootstrap procedure adapted to this pure functional framework in the literature. On the other hand even if a bootstrap approach may be satisfactory on a practical viewpoint, it will just provide an approximate distribution. Here the *exact* asymptotic distribution is given. Besides the scope of the paper is rather theoretical. The surprising Theorem 3.1 for instance is -to me at last- really food for thought for people dealing with functional data. However a promising approach would be to compare the results of this paper and those obtained by a bootstrap procedure, if any is available.

One of the other interests of the model is its simplicity. However in the general framework mentioned above, a first problem arises : in the case of a general space \mathcal{E} , not much is known about the mathematical description and properties of the linear space, say $\mathcal{L}(\mathcal{E})$, of bounded linear operators from \mathcal{E} to \mathcal{E} . Estimating ρ requires to build a sequence ρ_n of random linear operators in $\mathcal{L}(\mathcal{E})$, and we may face serious troubles if the space $\mathcal{L}(\mathcal{E})$ is too complex.

Usually authors focus on special cases and take for instance $\mathcal{E} = \mathcal{C}^m[0, T]$, a Banach space of functions defined on $[0, T]$ and with several continuous derivatives (in Mourid, [24]) or $\mathcal{E} = W^{m,p}[0, T]$, a space of Sobolev functions on a real interval (see for instance Adams [2]) for definitions and properties of Sobolev spaces). There are practical reasons for these choices. Indeed, the curves X_n are observed at discretized times and must be first reconstructed by implementing splines or wavelets for instance. These techniques provide explicit functions belonging to the spaces mentioned above.

Here appears the second problem : studying weak convergence for random elements, such as our predictor $\rho_n(X_n)$, in general infinite dimensional spaces is especially difficult, sometimes tricky. The most general tool is the Portmanteau Theorem (see Billingsley, [8]) but it is rather a general definition than a criterion to check the convergence of measures. Even if we consider the

Central Limit Theorem which is a very important but special case of convergence in distribution for measures, there are only a few spaces for which sufficient conditions are available (even fewer for a necessary condition). We refer to Ledoux and Talagrand [16] for a review on the CLT in Banach spaces. However, if \mathcal{E} is a **separable Hilbert space**, the situation becomes more favourable. Take Z_i a sequence of random elements in \mathcal{E} . It is a well-known fact that the CLT holds for i.i.d. Z_i if and only if the strong second moment is finite (i.e. $\mathbb{E} \|Z_i\|^2 < +\infty$). Besides many authors studied the CLT under different sorts of dependence assumptions (m-dependence, mixing, martingale differences, etc). We refer to Araujo-Gine [4] for a monograph on the CLT. The Hilbertian setting is quite comfortable for several other well-known mathematical reasons :

- All Hilbert spaces are isometrically isomorphic to the sequence space l^2 , hence have the same underlying geometric structure. They appear as the most natural generalization of the Euclidean space to the infinite dimensional setting.
- The bases are denumerable, the parallelogram identity is valid, the projection on convex sets is uniquely defined.
- The operator ρ belongs to the Banach space of linear operators on a Hilbert space. This space is widely used in several areas of mathematics. Spectral decompositions are available for compact operators.

In all the sequel we will set once and for all $\mathcal{E} = \mathcal{H}$ and \mathcal{H} will usually be a space $W^{m,2}$ where the smoothness index m belongs to \mathbb{N} ($W^{0,2} = L^2$).

The next remark is related to ρ and also aims at restricting the field of our research in order to gain some accuracy in the forthcoming results. In fact the space $\mathcal{L}(\mathcal{H})$ is much too large : this Banach space is not separable. This could turn out to be a serious problem as far as measurability is concerned (remind that we need to define a sequence of estimates ρ_n for ρ taking values in $\mathcal{L}(\mathcal{H})$). For other reasons mentioned in the next section, we will suppose that ρ is a **compact operator**. The space $\mathcal{K}(\mathcal{H})$ of compact operators is separable, its properties are closed to those of (finite size) matrices. Many features of linear operators on finite dimensional spaces are generalized to $\mathcal{K}(\mathcal{H})$ in a kind way.

The space \mathcal{H} is endowed with norm $\|\cdot\|$, derived from the scalar product $\langle \cdot, \cdot \rangle$. In the case where $\mathcal{H} = W^{m,2}$ we have

$$\langle f, g \rangle = \sum_{j=0}^m \int f^{(j)}(s) g^{(j)}(s) ds.$$

Spaces of continuous operators on \mathcal{H} are endowed with the classical sup-norm defined for all bounded operator T by

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

where \mathcal{H}_1 is the unit ball of \mathcal{H} .

The space of Hilbert-Schmidt operators denoted $\mathcal{K}_2(\mathcal{H})$ is endowed with norm $\|T\|_2 = \sum_p \|T(e_p)\|^2$ where e_p is any c.o.n.s. in \mathcal{H} . The spaces $\mathcal{K}_2(\mathcal{H})$ is a subspace of $\mathcal{K}(\mathcal{H})$. Note that up to the author's knowledge, the literature on model (1) or its close alternatives in an Hilbertian framework assumes that $\rho \in \mathcal{K}_2(\mathcal{H})$. Consequently we consider in this article a larger class for the unknown parameter.

The tensor product notation is of much use. It enables to define finite rank operators. For $u, v \in \mathcal{H}$,

$$(u \otimes v)(h) = \langle u, h \rangle v.$$

We may have to deal with another space of operators : the space of trace class operators $\mathcal{K}_1(\mathcal{H}) \subset \mathcal{K}(\mathcal{H})$ (the $\|\cdot\|_1$ norm on this space will not be fully defined here but I just mention that $\|u \otimes v\|_1 = \|u\| \|v\|$). Finally we will sometimes use the following norm bound :

$$\|\cdot\|_\infty \leq \|\cdot\|_2 \leq \|\cdot\|_1.$$

2 Identification and covariance regularization

In this Hilbert space setting, Bosq [10] proved that whenever it exists j_0 such that $\|\rho^{j_0}\|_\infty < 1$ and when $\mathbb{E}\|\varepsilon_1\|^2$ is finite, X_n is a strictly stationary sequence. For the sake of simplicity and in order to alleviate calculations within the proofs we will assume that $\|\rho\|_\infty < 1$. In the sequel we will assume that $\mathbb{E}(X_n) = m = 0$ i.e. we will not address the problem of estimating the mean since this issue was extensively treated in the literature. But we have to face two other serious issues.

2.1 Identifiability

As the data are of functional nature, the inference on ρ cannot be based on likelihood. Lebesgue's measure does not exist on non locally compact spaces and up to the author's knowledge the classical notion of density has not been extended to functional random elements. A classical **moment method** provides the following normal equation :

$$\Delta = \rho\Gamma \tag{2}$$

where

$$\begin{aligned} \Gamma &= \mathbb{E}(X_1 \otimes X_1), \\ \Delta &= \mathbb{E}(X_1 \otimes X_2) \end{aligned}$$

are the covariance operator (resp. the cross covariance operator of order one) of the process $(X_n)_{n \in \mathbb{Z}}$.

It is a well known fact that whenever $\mathbb{E}(\|X_1\|^2)$ is finite Γ is a selfadjoint positive, trace class operator (hence compact). In other words, Γ admits the following Schmidt (i.e. spectral) decomposition :

$$\Gamma = \sum_{l \in \mathbb{N}} \lambda_l \pi_l, \quad \sum_{l \in \mathbb{N}} \lambda_l < +\infty \tag{3}$$

where $(\lambda_l)_{l \geq 1}$ is the sequence of the positive eigenvalues of Γ and $(\pi_l)_{l \geq 1}$ is the associated sequence of projectors. In the sequel the eigenvectors of Γ are denoted $(e_l)_{l \geq 1}$ hence $\pi_l = e_l \otimes e_l$ and if x is any vector of \mathcal{H} we set $x_p = \langle x, e_p \rangle$. For further purpose $\Gamma_\varepsilon = \mathbb{E}(\varepsilon_1 \otimes \varepsilon_1)$ will stand for the covariance operator of ε_1 .

The first step consists in checking that equation (2) correctly defines the unknown parameter ρ .

Proposition 2.1 *When the inference on ρ is based on the moment equation (2), identifiability holds if and only if $\ker \Gamma = \{0\}$.*

The proof of the Proposition is simple. Let us give a sketch of it now. Assume that $\ker \Gamma \neq \{0\}$ and pick $v \in \ker \Gamma$. Setting $\rho_{v,u} = \rho + v \otimes u$ where u is any vector in \mathcal{H} it is basic to see that $\Delta = \rho_{v,u}\Gamma$ again. In other words the moment equation may not be able to distinguish between ρ and $\rho_{v,u}$.

Remark 2.1 *The condition $\ker \Gamma = \{0\}$ implies that all the eigenvalues are strictly positive. In the sequel we will assume that $\lambda_1 \geq \lambda_2 \geq \dots > 0$.*

2.2 Regularizing the inverse covariance operator

Even if the identifiability of ρ is ensured by assumption \mathbf{A}_0 , we must remain cautious when building an estimator. Several serious problems appear.

First it is crucial to note that we **cannot** deduce from (2) that $\Delta\Gamma^{-1} = \rho$. Indeed Γ^{-1} does not necessarily exist. A necessary and sufficient condition for Γ^{-1} to be defined as a linear mapping is : $\ker \Gamma = \{0\}$. Then Γ^{-1} is an unbounded symmetric operator on \mathcal{H} . The consequences are the following :

- Γ^{-1} is just defined on the dense vector space

$$\mathcal{D}(\Gamma^{-1}) = \text{Im}\Gamma = \left\{ x \in \mathcal{H}, \sum_{i=1}^n \frac{x_p^2}{\lambda_p^2} < +\infty \right\}$$

and $\mathcal{D}(\Gamma^{-1}) \subsetneq \mathcal{H}$.

- Γ^{-1} is a measurable linear mapping but is not continuous, in other words it is continuous at no point for which is it defined or "the domain of Γ^{-1} is also the set of its discontinuities".
- $\Gamma\Gamma^{-1}$ is not the identity operator on \mathcal{H} but on $\mathcal{D}(\Gamma^{-1})$ which entails that (2) implies $\Delta\Gamma^{-1} = \rho|_{\text{Im}\Gamma} \neq \rho$

The previous facts are very well-known in operator theory and give rise here to an ill-posed problem (or an inverse problem). Since Γ^{-1} is extremely irregular, we should propose a way to regularize it i.e. find out Γ^\dagger say, a linear operator "close" to Γ^{-1} and having additional continuity properties. There are several ways to deal with this problem. We refer to Arsenin and Tikhonov [5] and Groetsch [15], amongst many others, for famous books about this topic.

Here the approach is quite intuitive and classical : when (3) holds,

$$\Gamma^{-1}(x) = \sum_{l \in \mathbb{N}} \frac{1}{\lambda_l} \pi_l(x)$$

for all x in $\mathcal{D}(\Gamma^{-1})$. We just set

$$\Gamma^\dagger(x) = \sum_{l \leq k_n} \frac{1}{\lambda_l} \pi_l(x)$$

where $(k_n)_{n \in \mathbb{N}}$ is an increasing sequence tending to infinity. It may be proved that whenever $x \in \mathcal{D}(\Gamma^{-1})$ and $n \uparrow +\infty$,

$$\Gamma^\dagger(x) \rightarrow \Gamma^{-1}(x).$$

Besides Γ^\dagger is a continuous operator with $\|\Gamma^\dagger\|_\infty = \lambda_{k_n}^{-1}$ and implicitly depends on n .

If (2) is the starting point in our estimation procedure, replacing the unknown operators by their empirical counterparts gives :

$$\Delta_n = \rho_n^{imp} \Gamma_n$$

where

$$\Gamma_n = \frac{1}{n} \sum_{k=1}^n X_k \otimes X_k,$$

$$\Delta_n = \frac{1}{n-1} \sum_{k=1}^{n-1} X_k \otimes X_{k+1}$$

and ρ_n^{imp} just **implicitly** defines our estimate for ρ .

The preceding remarks give some clues to reach the end of the estimation step. Setting

$$\Gamma_n^\dagger = \sum_{l \leq k_n} \frac{1}{\hat{\lambda}_l} \hat{\pi}_l \tag{4}$$

where $\hat{\lambda}_l$ and $\hat{\pi}_l$ are the empirical counterparts of λ_l and π_l we get :

Definition 2.1 *The estimate of ρ is ρ_n given by $\rho_n = \Delta_n \Gamma_n^\dagger$.*

For further purpose we denote $\hat{\Pi}_{k_n} = \sum_{j=1}^{k_n} \hat{\pi}_j$ the projector on the space spanned by the k_n first eigenvectors of Γ_n .

Remark 2.2 *The $\hat{\lambda}_l$'s and the $\hat{\pi}_l$'s are obtained as by-products of the functional PCA of the sample (X_1, \dots, X_n) .*

2.3 A smoothness condition on the autocorrelation operator

In order to get the main results given in the next section we need to develop one of the crucial assumptions needed further. This subsection is devoted to explaining it. This condition must be understood as a smoothness condition on the unknown operator ρ . But what do we mean by "smoothness" for a linear operator? The notion of smoothness is intuitively related to functions or mapping and should be made more clear in our setting. In order to be more illustrative let us consider for ρ a diagonal operator on \mathcal{H} . Say in any complete orthonormal system :

$$\rho = \text{diag} \left[(\mu_i)_{i \geq 1} \right]$$

with $\mu_i \geq \mu_{i+1}$. Obviously if $\mu_i = 1$ $\rho = I$ and if the sequence μ_i is bounded $\left((\mu_i)_{i \geq 1} \in l^\infty \right)$, ρ is a bounded operator. If $(\mu_i)_{i \geq 1} \in c_0$, ρ is a compact operator. If $(\mu_i)_{i \geq 1} \in l^2$, ρ is a Hilbert-Schmidt operator, etc. The degree of smoothness of ρ will be strictly determined by the rate of decrease to zero of $(|\mu_i|)_{i \geq 1}$ or, generally speaking of its eigenvalues or characteristic numbers. When the μ_i 's decrease quickly ρ is "close" to any finite dimensional approximation based on the n first μ_i 's (when n gets large). Conversely imagine that the $|\mu_i|$'s tend to infinity, then ρ is unbounded hence not continuous hence not smooth.

The next assumption

$$\mathbf{A}_1 : \left\| \Gamma^{-1/2} \rho \right\|_\infty < +\infty \quad (5)$$

tells us that ρ should be at least as "smooth" as $\Gamma^{1/2}$. Indeed let us try to be more illustrative and assume that ρ is symmetric and has the same basis of eigenvectors as Γ . Assumption (5) implies that the sequence $(\mu_i / \sqrt{\lambda_i})_{i \in \mathbb{N}}$ is bounded. We set $\tilde{\rho} = \Gamma^{-1/2} \rho$.

Remark 2.3 *As a consequence of the above we remark for further purpose that if $\tilde{\rho}$ is bounded, so is $\tilde{\rho}^*$. But for the reasons mentioned in the previous subsection $\tilde{\rho}^* \neq \rho^* \Gamma^{-1/2}$. In fact $\rho^* \Gamma^{-1/2}$ is a bounded operator defined on $\mathcal{D}(\Gamma^{-1/2})$. Like any bounded operator on a dense domain it may be uniquely extended to a bounded operator defined on the whole \mathcal{H} . This operator precisely coincides with $\tilde{\rho}^*$. I just point out the following : from (5) we deduce that*

$$\sup_p \left\| \rho^* \Gamma^{-1/2} (e_p) \right\|^2 = \sup_p \frac{\left\| \rho^* (e_p) \right\|^2}{\lambda_p} \leq M = \|\tilde{\rho}^*\|_\infty. \quad (6)$$

3 Main results

The main results of this work are collected in two theorems below. We first recapitulate three seminal assumptions under the same label :

$$\mathbf{A}_0 : \begin{cases} \ker \Gamma = \{0\} \\ \mathbb{E} \|\varepsilon\|^2 < +\infty \\ \|\rho\|_\infty < 1. \end{cases}$$

The subscript 0 was given on purpose since this set of assumptions is minimal in order to begin any statistical inference on the model.

Then I remind the reader the so-called Karhunen-Loève (KL) extension of the random element X : the distribution of X (i.e. of X_n for all n since the sequence is strictly stationary) is :

$$X =_d \sum_{k=1}^{+\infty} \xi_k \sqrt{\lambda_k} e_k \quad (7)$$

where $=_d$ denotes equality of distributions and the ξ_k 's are non correlated real valued random variables with null expectation and unit variance (the ξ_k 's are i.i.d. gaussian if X_1 is). We will make use of (7) within the proofs.

The following moment assumption is mild :

$$\mathbf{A}_2 : \sup_k \mathbb{E} \xi_k^4 < M \quad (8)$$

It is fulfilled by large families of r.v. ξ_k 's (subject to $\mathbb{E}\xi_k = 0$ and $\mathbb{E}\xi_k^2 = 1$) with thin enough queues : gaussian, uniform, two sided exponential, etc, but will fail for certain classes of two sided Pareto random variables for instance. Remember that we study weak convergence for $\rho_n(X_{n+1})$, that ρ_n depends on Γ_n and consequently that assumptions on functionals of the fourth moment of X_1 (like **A₂**) are unavoidable.

The next assumption is related to the eigenvalues of Γ .

Let $\lambda_j = \lambda(j)$ where λ is a positive function defined on and with values in \mathbb{R}^+ . Clearly function λ is decreasing if the eigenvalues are ordered decreasingly and $\lim_{t \rightarrow +\infty} \lambda(t) = 0$. We assume that :

A₃ : The function λ is convex

Remark 3.1 *Actually we just need **A₃** to hold for large values of j . This assumption is finally not constraining at all since it is suited to many classical cases : when the rate of decay to zero is arithmetic (say $\lambda_j = \text{Const}/j^{1+\alpha}$, $\alpha > 0$) or exponential ($\lambda_j = \text{Const} \cdot \exp(-\alpha j)$, $\alpha > 0$) and in several other less standard situations such as Laurent series ($\lambda_j = \text{Const}/(j^\alpha \log^{1+\beta} j)$, $\alpha, \beta > 0$).*

Remark 3.2 *Assumption **A₃** implies that $\lambda_j - \lambda_{j+1} \leq \lambda_{j-1} - \lambda_j$.*

The next and first theorem assesses that :

Theorem 3.1 *It is impossible for $\hat{\rho}_n - \rho$ to converge in distribution for the norm topology on \mathcal{K} .*

Remark 3.3 *What is actually proved is : for any normalizing sequence $\alpha_n \uparrow +\infty$, $\alpha_n(\hat{\rho}_n - \rho)$ either diverges or converges in distribution to the Dirac distribution on the null element in \mathcal{K} . Also note that weak convergence cannot take place for the Hilbert-Schmidt topology either since the embedding from \mathcal{K}_2 to \mathcal{K} is continuous.*

For technical reasons, we will focus on a slightly modified version of the prediction problem. We will assume that ρ_n is built from (X_1, \dots, X_n) and that X_{n+2} is to be predicted from ρ_n and X_{n+1} . In other word the sample is tiled, the last observed curve (X_{n+1} here) is taken into account to predict X_{n+2} but not to construct ρ_n .

Here is the main result of the paper. Remind that $\hat{\Pi}_{k_n}$ was introduced just before Definition 2.1.

Theorem 3.2 *When assumptions **A₀** – **A₃** hold and if $k_n = o\left(\frac{n^{1/4}}{\log n}\right)$,*

$$\sqrt{\frac{n}{k_n}} \left(\hat{\rho}_n(X_{n+1}) - \rho \hat{\Pi}_{k_n}(X_{n+1}) \right) \xrightarrow{w} \mathcal{G}$$

where \mathcal{G} is a \mathcal{H} -valued gaussian centered random variable with covariance operator Γ_ε .

Remark 3.4 *Theorem 3.1 remains unchanged if ρ is changed into $\rho \hat{\Pi}_{k_n}$ which appears more "natural" in view of Theorem 3.2.*

This central result should be commented. First of all the normalizing sequence is typically nonparametric : $\sqrt{n/k_n}$. Second a bias term appears. Recently, Cardot, Mas and Sarda [11] obtained a similar result in a much simpler regression model, based on i.i.d. observations unlike here. A non random bias was obtained -namely the random projector $\widehat{\Pi}_{k_n}$ was replaced by a non random one- but this could not be carried out here. Also note that since ε is the innovation of process X , the best target we can hope to reach is $\rho(X_{n+1})$ i.e. the conditional expectation of y_{n+1} , which is random in any case. However it is simple to prove that $\left\| \rho \widehat{\Pi}_{k_n}(X_{n+1}) - \rho(X_{n+1}) \right\|$ tends to zero in probability when n tends to infinity. Finally even if the random term $\rho \widehat{\Pi}_{k_n}(X_{n+1})$ is not quite satisfactory on a theoretical viewpoint, it may be easily interpreted by practitioners since $\widehat{\Pi}_{k_n}(X_{n+1})$ is the projection of the new input onto the k_n first axes of the functional PCA of the sample. These axes have optimality properties w.r.t. the decomposition of variance for the process X .

4 Concluding remarks

As seen from the literature on the subject, two modes of stochastic convergence had already been investigated for estimates of ρ in model (1) : convergence in probability and almost sure convergence. Weak convergence was the missing one essentially because it is more intricate.

In fact from

$$\|\rho_n(X_{n+1}) - \rho(X_{n+1})\| \leq \|\rho_n - \rho\|_\infty \|X_{n+1}\|$$

it is plain that convergence (almost sure or in probability) for $\|\rho_n - \rho\|_\infty$ implies convergence for the predictor. Theorem 3.1 proves that the situation is much more different as far as convergence in distribution is addressed.

It should be also stressed that assumptions $\mathbf{A}_0 - \mathbf{A}_3$ are truly mild. For instance all theoretical articles dealing with the problem of asymptotics for the predictor assume that ρ is symmetric and that the rate of decay of the sequence of eigenvalues is known.

The main advance relies undoubtedly on the fact that the dimension sequence k_n does not depend anymore on the eigenvalues (previously such conditions as $n^\alpha \lambda_{k_n} \rightarrow +\infty$ for some α where necessary). The existence of a universal k_n enables to revisit all previous results on the topic and sheds a new light on this model. Indeed in view of Theorem 3.2, it is tempting to postulate that a L^2 minimax rate of convergence could be k_n/n when ρ belongs to the set defined by assumption \mathbf{A}_1 (this set is nothing but an ellipsoid of \mathcal{K}). But these considerations are beyond the scope of this article.

5 Mathematical derivations

Assumptions $\mathbf{A}_0 - \mathbf{A}_3$ are supposed to hold throughout the proofs. The generic notation M will be used to denote universal constants. The next equation is straightforward from (1), links Γ , Γ_ε and ρ , and will soon be needed :

$$\Gamma = \rho \Gamma \rho^* + \Gamma_\varepsilon. \tag{9}$$

We start with letting

$$S_n = \sum_{k=1}^n X_{k-1} \otimes \varepsilon_k.$$

Easy calculations give

$$\begin{aligned} \rho_n &= \Delta_n \Gamma_n^\dagger = \rho \Gamma_n \Gamma_n^\dagger + S_n \Gamma_n^\dagger, \\ \rho_n &= \rho \widehat{\Pi}_{k_n} + S_n \Gamma_n^\dagger. \end{aligned} \tag{10}$$

It is plain by (4) that $\Gamma_n \Gamma_n^\dagger = \widehat{\Pi}_{k_n}$. Hence :

$$\rho_n - \rho \widehat{\Pi}_{k_n} = S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) + S_n \Gamma^\dagger \tag{11}$$

which is the starting point.

This section is decomposed into three subsections. In the first one preliminary results and tools connected with the theory of perturbation for operators on Hilbert spaces are provided. In the second part I prove that $S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right)$ is a vanishing term if the dimension sequence k_n is well chosen. The third part is devoted to studying weak convergence and proving Theorem 3.2. The proof of Theorem 3.1 is postponed to the end of the paper.

5.1 Preliminary results

5.1.1 Some inequalities

We first deal with a crucial Lemma.

Lemma 5.1 *We have :*

$$\sup_{m,p} n \frac{\mathbb{E} \langle (\Gamma_n - \Gamma)(e_p), e_m \rangle^2}{\lambda_p \lambda_m} \leq M \tag{12}$$

$$\sup_{m,p} n \frac{\mathbb{E} \langle S_n(e_p), e_m \rangle^2}{\lambda_p \lambda_m} \leq M \tag{13}$$

Proof. We begin with proving (12).

$$\begin{aligned} \langle (\Gamma_n - \Gamma)(e_p), e_m \rangle^2 &= \frac{1}{n^2} \left(\sum_{k=1}^n \langle X_k, e_p \rangle \langle X_k, e_m \rangle \right)^2 \\ \mathbb{E} \langle (\Gamma_n - \Gamma)(e_p), e_m \rangle^2 &= \frac{1}{n} \mathbb{E} \left(\langle X_1, e_p \rangle^2 \langle X_1, e_m \rangle^2 \right) \\ &\quad + \frac{2}{n^2} \mathbb{E} \sum_{1 \leq i < k \leq n} (\langle X_i, e_p \rangle \langle X_i, e_m \rangle \langle X_k, e_p \rangle \langle X_k, e_m \rangle) \end{aligned}$$

It is easily seen by KL decomposition (7) and assumption \mathbf{A}_2 that the first term may be bounded by

$$\frac{1}{n} \lambda_p \lambda_m \mathbb{E} (\xi_p^2 \xi_m^2) \leq M \frac{\lambda_p \lambda_m}{n} \tag{14}$$

whenever $p = m$ or $p \neq m$.

Now assume that $p \neq m$. We study the second :

$$\begin{aligned} X_k &= \varepsilon_k + \dots + \rho^{k-i-1}(\varepsilon_{i+1}) + \rho^{k-i}(X_i) \\ X_k &= E_{k,i} + \rho^{k-i}(X_i) \end{aligned}$$

where

$$E_{k,i} = \varepsilon_k + \dots + \rho^{k-i-1}(\varepsilon_{i+1})$$

hence

$$\begin{aligned} & \mathbb{E} \sum_{i < k} (\langle X_i, e_p \rangle \langle X_i, e_m \rangle \langle X_k, e_p \rangle \langle X_k, e_m \rangle) \\ &= \mathbb{E} \sum_{i < k} (\langle X_i, e_p \rangle \langle X_i, e_m \rangle \langle \rho^{k-i}(X_i), e_p \rangle \langle \rho^{k-i}(X_i), e_m \rangle) \\ &+ \mathbb{E} \sum_{i < k} (\langle X_i, e_p \rangle \langle X_i, e_m \rangle \langle E_{k,i}, e_p \rangle \langle E_{k,i}, e_m \rangle) \tag{i} \\ &= \mathbb{E} \sum_{i < k} (\langle X_i, e_p \rangle \langle X_i, e_m \rangle \langle \rho^{k-i}(X_i), e_p \rangle \langle \rho^{k-i}(X_i), e_m \rangle) \tag{ii} \\ &= \mathbb{E} \sum_{i < k} (\langle X_1, e_p \rangle \langle X_1, e_m \rangle \langle \rho^{k-i}(X_1), e_p \rangle \langle \rho^{k-i}(X_1), e_m \rangle) \tag{iii} \\ &= \mathbb{E} \left[\langle X_1, e_p \rangle \langle X_1, e_m \rangle \sum_{i < k} \langle \rho^{k-i}(X_1), e_p \rangle \langle \rho^{k-i}(X_1), e_m \rangle \right] \\ &= \mathbb{E} \left[\langle X_1, e_p \rangle \langle X_1, e_m \rangle \sum_{k=1}^{n-1} (n-k) \langle \rho^k(X_1), e_p \rangle \langle \rho^k(X_1), e_m \rangle \right] \tag{iv} \end{aligned}$$

where (ii) stems from(i) because if $p \neq m$

$$\begin{aligned} & \mathbb{E} (\langle X_i, e_p \rangle \langle X_i, e_m \rangle \langle E_{k,i}, e_p \rangle \langle E_{k,i}, e_m \rangle) \\ &= \mathbb{E} (\langle X_i, e_p \rangle \langle X_i, e_m \rangle) \mathbb{E} (\langle E_{k,i}, e_p \rangle \langle E_{k,i}, e_m \rangle) \\ &= 0. \end{aligned}$$

and (iii) stems from (ii) by stationarity. Now by (iv),

$$\begin{aligned} & \frac{1}{n} \left| \mathbb{E} \sum_{1 \leq i < k \leq n} (\langle X_i, e_p \rangle \langle X_i, e_m \rangle \langle X_k, e_p \rangle \langle X_k, e_m \rangle) \right| \\ & \leq \mathbb{E} \left[|\langle X_1, e_p \rangle \langle X_1, e_m \rangle| \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \left| \langle \rho^k(X_1), e_p \rangle \langle \rho^k(X_1), e_m \rangle \right| \right] \tag{15} \end{aligned}$$

Let us fix $k \geq 1$ and develop

$$\begin{aligned} \langle \rho^k(X_1), e_p \rangle \langle \rho^k(X_1), e_m \rangle &= \sqrt{\lambda_p \lambda_m} \langle \Gamma^{-1/2} \rho^k(X_1), e_p \rangle \langle \Gamma^{-1/2} \rho^k(X_1), e_m \rangle \\ &= \sqrt{\lambda_p \lambda_m} \left\langle \rho^{k-1}(X_1), \frac{\rho^*(e_p)}{\sqrt{\lambda_p}} \right\rangle \left\langle \rho^{k-1}(X_1), \frac{\rho^*(e_m)}{\sqrt{\lambda_m}} \right\rangle \end{aligned}$$

and denoting $u_p = \rho^*(e_p) / \sqrt{\lambda_p}$,

$$\begin{aligned} \left| \langle \rho^k(X_1), e_p \rangle \langle \rho^k(X_1), e_m \rangle \right| &\leq \sqrt{\lambda_p \lambda_m} \left\| \rho^{k-1} \right\|^2 \|X_1\|^2 \|u_p\| \|u_m\| \\ &\leq M \sqrt{\lambda_p \lambda_m} \left\| \rho^{k-1} \right\|^2 \|X_1\|^2 \end{aligned} \quad (16)$$

since by (6) $\|u_p\|$ and $\|u_m\|$ may be bounded uniformly wrt p and m by $\|\text{ext}(\tilde{\rho}^*)\|_\infty$ (see Remark 2.3.below) Then

$$\begin{aligned} &\frac{1}{n} \left| \mathbb{E} \sum_{1 \leq i < k \leq n} (\langle X_i, e_p \rangle \langle X_i, e_m \rangle \langle X_k, e_p \rangle \langle X_k, e_m \rangle) \right| \\ &\leq M \sqrt{\lambda_p \lambda_m} \mathbb{E} \left(\|X_1\|^2 |\langle X_1, e_p \rangle \langle X_1, e_m \rangle| \right) \sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \right) \left\| \rho^{k-1} \right\|^2 \end{aligned}$$

And

$$\mathbb{E} \left(\|X_1\|^2 |\langle X_1, e_p \rangle \langle X_1, e_m \rangle| \right) = \sqrt{\lambda_p \lambda_m} \left(\sum_{l=1}^{+\infty} \lambda_l \mathbb{E} (\xi_l^2 \xi_p \xi_m) \right) \quad (17)$$

by (7) again. Applying twice Cauchy-Schwarz inequality we bound the infinite sum by a constant which does not depend on p and m . Collecting (14), (15), (16) and (17) we get

$$n \sup_{p \neq m} \frac{\mathbb{E} \langle (\Gamma_n - \Gamma)(e_p), e_m \rangle^2}{\lambda_p \lambda_m} \leq M$$

In order to complete the proof (remember that we assumed that $p \neq m$ just below (14)) we can check that our computations remain valid if we take $p = m$.

The proof of (13) is similar but simpler. We have

$$\begin{aligned} \mathbb{E} \langle S_n(e_p), e_m \rangle^2 &= \frac{1}{n} \mathbb{E} \left(\langle X_1, e_p \rangle^2 \langle \varepsilon_2, e_m \rangle^2 \right) \\ &= \frac{1}{n} \mathbb{E} \left(\langle X_1, e_p \rangle^2 \right) \mathbb{E} \left(\langle \varepsilon_2, e_m \rangle^2 \right) = \frac{\lambda_p \langle \Gamma_\varepsilon e_m, e_m \rangle}{n} \\ &= \frac{\lambda_p (\lambda_m - \langle \rho \Gamma \rho^* e_m, e_m \rangle)}{n} \leq \frac{\lambda_p \lambda_m}{n} \end{aligned}$$

since $\Gamma = \rho \Gamma \rho^* + \Gamma_\varepsilon$. ■

The proof of the three following Lemmas may be found in Cardot, Mas, Sarda [11].

Lemma 5.2 *Consider two positive integers j and k large enough and such that $k > j$. Then*

$$j \lambda_j \geq k \lambda_k \quad \text{and} \quad \lambda_j - \lambda_k \geq \left(1 - \frac{j}{k} \right) \lambda_j. \quad (18)$$

Besides

$$\sum_{j \geq k} \lambda_j \leq (k+1) \lambda_k. \quad (19)$$

Lemma 5.3 *The following is true for j large enough*

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

5.1.2 A few basic facts about perturbation theory

Perturbation theory for bounded operators is a powerful tool all along our study and is of much help when dealing with random (or not) covariance operators. It features several theoretical interests : for instance eigenprojectors or pseudo inverses of Γ may be expressed as functions of Γ only (without introducing the eigenvectors). However this theory is not widely used in statistics although the only mathematical prerequisite is the theory of holomorphic functions and of integrals on contours in the complex plane. We refer to Dunford-Schwartz [13] (Chapter VII.3) or to Gohberg, Goldberg and Kaashoek [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 λ_i and radius $\delta_i/3$ 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 [13], Section VII.3 Definitions 8 and 9). Results from perturbation theory yield :

$$\Pi_{k_n} = \frac{1}{2\pi\iota} \int_{\mathcal{C}_n} (zI - \Gamma)^{-1} dz.$$

where $\iota^2 = -1$, Π_{k_n} is defined similarly to $\widehat{\Pi}_{k_n}$ (see Theorem 3.2) and stands for the projector on the space spanned by the k_n first eigenvectors of Γ . The integral is defined on the complex plane. Note that the random counterparts (*i.e.* where Π_{k_n} and Γ are respectively replaced by $\widehat{\Pi}_{k_n}$ and Γ_n) of the previous equation is just :

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

and the contour $\widehat{\mathcal{C}}_n$ is random and depends on the $\widehat{\lambda}_i$'s. The following equalities are also valid

$$\begin{aligned} \Gamma^\dagger &= \int_{\mathcal{C}_n} z^{-1} (zI - \Gamma)^{-1} dz = \sum_{j=1}^{k_n} \int_{\mathcal{B}_j} z^{-1} (zI - \Gamma)^{-1} dz. \\ \Gamma_n^\dagger &= \int_{\widehat{\mathcal{C}}_n} z^{-1} (zI - \Gamma_n)^{-1} dz = \sum_{j=1}^{k_n} \int_{\widehat{\mathcal{B}}_j} z^{-1} (zI - \Gamma_n)^{-1} dz. \end{aligned}$$

and

$$\begin{aligned} &S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) \\ &= \int_{\widehat{\mathcal{C}}_n} z^{-1} S_n (z - \Gamma_n)^{-1} (X_{n+1}) dz - \int_{\mathcal{C}_n} z^{-1} S_n (z - \Gamma)^{-1} (X_{n+1}) dz \end{aligned} \quad (20)$$

As announced at the beginning of the proof section we will prove in the next subsection that (20) -correctly normalized by $\sqrt{n/k_n}$ - tends to zero in probability, hence is negligible. We need two Lemmas to start. In these Lemmas the square root of a symmetric operator T , say $T^{1/2}$ appears. The bounded operator $T^{1/2}$ has the same eigenvectors as T . Its eigenvalues are the complex square roots of those of T .

Lemma 5.4 *We have for j large enough*

$$\mathbb{E} \sup_{z \in \mathcal{B}_j} \left\| (zI - \Gamma)^{-1/2} (\Gamma_n - \Gamma) (zI - \Gamma)^{-1/2} \right\|_{\infty}^2 \leq \frac{M}{n} (j \log j)^2, \quad (21)$$

$$\mathbb{E} \sup_{z \in \mathcal{B}_j} \left\| z^{-1/2} S_n (zI - \Gamma)^{-1/2} \right\|_{\infty}^2 \leq \frac{M}{n} j \log j \quad (22)$$

$$\mathbb{E} \sup_{z \in \mathcal{B}_j} \left\| (zI - \Gamma)^{-1/2} \varepsilon_1 \right\|_{\infty}^2 \leq M j \log j \quad (23)$$

In fact this last Lemma was proved in Cardot, Mas, Sarda [11] in an i.i.d framework. However a quick inspection of the proof shows that, by Lemma 5.1 the same result holds in this dependent setting for (21) and (22). In order to convince the suspicious reader I give now the derivation of (23) which uses basically the same technique as for (21) and (22) but is shorter. We have :

$$\begin{aligned} \left\| (zI - \Gamma)^{-1/2} \varepsilon_1 \right\|_{\infty}^2 &= \sum_{p=1, p \neq j}^{+\infty} \frac{\langle \varepsilon_1, e_p \rangle^2}{|z - \lambda_p|} \\ \sup_{z \in \mathcal{B}_j} \left\| (zI - \Gamma)^{-1/2} \varepsilon_1 \right\|_{\infty}^2 &= \sum_{p=1, p \neq j}^{+\infty} \frac{\langle \varepsilon_1, e_p \rangle^2}{|\lambda_j - \lambda_p|} \end{aligned}$$

since obviously for all $p \neq j$, $|z - \lambda_p| \geq |\lambda_j - \lambda_p|$ when $z \in \mathcal{B}_j$. Then

$$\mathbb{E} \sup_{z \in \mathcal{B}_j} \left\| (zI - \Gamma)^{-1/2} \varepsilon_1 \right\|_{\infty}^2 = \sum_{p=1, p \neq j}^{+\infty} \frac{\mathbb{E} \langle \varepsilon_1, e_p \rangle^2}{|\lambda_j - \lambda_p|}$$

Now from $\Gamma = \Gamma_{\varepsilon} + \rho \Gamma \rho^*$ we see that $\mathbb{E} \langle \varepsilon_1, e_p \rangle^2 = \langle \Gamma_{\varepsilon} e_p, e_p \rangle \leq \langle \Gamma e_p, e_p \rangle = \lambda_p$ hence

$$\mathbb{E} \sup_{z \in \mathcal{B}_j} \left\| (zI - \Gamma)^{-1/2} \varepsilon_1 \right\|_{\infty}^2 \leq \sum_{p=1, p \neq j}^{+\infty} \frac{\lambda_p}{|\lambda_j - \lambda_p|} \leq M j \log j$$

by Lemma 5.3.

This last Lemma will be used when dealing with residual terms $S_n \left(\Gamma_n^{\dagger} - \Gamma^{\dagger} \right)$ appearing in (11).

Lemma 5.5 *Denoting*

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

The following holds

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

where M is some positive constant. Besides

$$\mathbb{P}(\mathcal{E}_j^c) \leq M \frac{j \log j}{\sqrt{n}}. \quad (24)$$

Proof. We have successively

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

hence

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

and

$$\left[I + (zI - \Gamma)^{-1/2} (\Gamma_n - \Gamma) (zI - \Gamma)^{-1/2} \right] (zI - \Gamma)^{1/2} (zI - \Gamma_n)^{-1} (zI - \Gamma)^{1/2} = I. \quad (25)$$

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

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

and

$$\left\| (I + T)^{-1} \right\|_\infty \leq \frac{1}{1 - \|T\|_\infty}$$

From (25) we deduce that

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

Now, the bound in (24) stems easily from Markov inequality and (21) in Lemma 5.4. This finishes the proof of the Lemma. ■

5.2 Residual term

This first lemma only aims at proving that the random contour $\widehat{\mathcal{C}}_n$ can be replaced by the non random one \mathcal{C}_n in (20) in order to merge both integrals.

Lemma 5.6 *When $\frac{1}{\sqrt{n}} k_n^2 \log k_n \rightarrow 0$,*

$$S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) = \int_{\mathcal{C}_n} z^{-1} S_n \left[(z - \Gamma_n)^{-1} - (z - \Gamma)^{-1} \right] (X_{n+1}) dz + L_n$$

where $\sqrt{n} \|L_n\|$ vanishes in probability.

Proof. We introduce the following event :

$$\mathcal{A}_n = \left\{ \forall j \in \{1, \dots, k_n\} \left| \frac{\widehat{\lambda}_j - \lambda_j}{\delta_j} \right| < 1/8 \right\},$$

and $\mathbb{1}_{\mathcal{A}_n}$ is the indicator function of the set \mathcal{A}_n .

Introducing the set \mathcal{A}_n enables to consider the situation when all the ordered eigenvalues of Γ_n are close enough to those of Γ . In fact when \mathcal{A}_n holds all the k_n first empirical eigenvalues $\widehat{\lambda}_j$ lie in the circle of center λ_j and radius $\delta_j/8$, say $\widetilde{\mathcal{B}}_j$ (included in \mathcal{B}_j). Consequently none of the $\widehat{\lambda}_j$ is located in the annulus between \mathcal{B}_j and $\widetilde{\mathcal{B}}_j$ and when \mathcal{A}_n holds $\widehat{\mathcal{C}}_n$ may be replaced by \mathcal{C}_n . It is clear from previous remarks that

$$\begin{aligned} S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) &= S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) \left(\mathbb{1}_{\mathcal{A}_n} + \mathbb{1}_{\mathcal{A}_n^c} \right) \\ &= \left(\int_{\mathcal{C}_n} z^{-1} S_n \left[(zI - \Gamma_n)^{-1} - (z - \Gamma)^{-1} \right] (X_{n+1}) dz \right) \\ &\quad - \left(\int_{\mathcal{C}_n} z^{-1} S_n \left[(zI - \Gamma_n)^{-1} - (z - \Gamma)^{-1} \right] (X_{n+1}) dz \right) \mathbb{1}_{\mathcal{A}_n^c} \\ &\quad + S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) \mathbb{1}_{\mathcal{A}_n^c} \end{aligned}$$

We set

$$\begin{aligned} L_n &= S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) \mathbb{1}_{\mathcal{A}_n^c} - \left(\int_{\mathcal{C}_n} z^{-1} S_n \left[(zI - \Gamma_n)^{-1} - (zI - \Gamma)^{-1} \right] (X_{n+1}) dz \right) \mathbb{1}_{\mathcal{A}_n^c} \\ &= \left[S_n \Gamma_n^\dagger (X_{n+1}) - \left(\int_{\mathcal{C}_n} z^{-1} S_n (zI - \Gamma_n)^{-1} (X_{n+1}) dz \right) \right] \mathbb{1}_{\mathcal{A}_n^c} \end{aligned}$$

and we see that

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

It suffices to get $\mathbb{P} \left(\mathcal{A}_n^c \right) \rightarrow 0$. But

$$\mathbb{P} \left(\mathcal{A}_n^c \right) \leq \sum_{i=1}^{k_n} \mathbb{P} \left(\left| \widehat{\lambda}_i - \lambda_i \right| > \delta_i/8 \right).$$

Now we refer to Theorem 4.10 of Bosq [10]. Following the proof of this Theorem along p.122 and 123 it is proved that the asymptotic behaviour of $\left| \widehat{\lambda}_i - \lambda_i \right|$ is the same as $|\langle (\Gamma_n - \Gamma) e_i, e_i \rangle|$. Then

$$\mathbb{P} \left(\left| \widehat{\lambda}_i - \lambda_i \right| > \delta_i/8 \right) \leq 8 \frac{\lambda_i}{\delta_i} \mathbb{E} \left(\frac{\left| \widehat{\lambda}_i - \lambda_i \right|}{\lambda_i} \right) \sim 8 \frac{\lambda_i}{\delta_i} \mathbb{E} \frac{|\langle (\Gamma_n - \Gamma) e_i, e_i \rangle|}{\lambda_i}$$

By assumption **A**₂ we get

$$\mathbb{E} \frac{|\langle (\Gamma_n - \Gamma) e_i, e_i \rangle|}{\lambda_i} \leq \sqrt{\mathbb{E} \frac{|\langle (\Gamma_n - \Gamma) e_i, e_i \rangle|^2}{\lambda_i^2}} \leq \frac{M}{\sqrt{n}}$$

by (12). At last

$$\mathbb{P}(\mathcal{A}_n^c) \leq 8 \frac{M}{\sqrt{n}} \sum_{i=1}^{k_n} \frac{\lambda_i}{\delta_i} \leq \frac{M'}{\sqrt{n}} \sum_{i=1}^{k_n} i \log i \leq \frac{M''}{\sqrt{n}} k_n^2 \log k_n.$$

This concludes the proof of the lemma. ■

For the sake of clarity, from now on we will abusively note

$$S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) = \int_{\mathcal{C}_n} z^{-1} S_n \left[(z - \Gamma_n)^{-1} - (z - \Gamma)^{-1} \right] (X_{n+1}) dz$$

but Lemma 5.6 above shows that this does not change anything to the validity of our forthcoming results.

The next Proposition is the central result of this subsection.

Proposition 5.1 *If $\frac{1}{\sqrt{n}} k_n^2 (\log k_n)^2 \rightarrow 0$ (which is true if $k_n = o\left(\frac{n^{1/4}}{\log n}\right)$) we have :*

$$\sqrt{\frac{n}{k_n}} S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) \xrightarrow{\mathbb{P}} 0$$

in \mathcal{H} .

Proof of Proposition 5.1 :

We develop :

$$\begin{aligned} S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) &= \int_{\mathcal{C}_n} z^{-1} S_n \left[(zI - \Gamma_n)^{-1} - (zI - \Gamma)^{-1} \right] (X_{n+1}) dz \\ &= \int_{\mathcal{C}_n} z^{-1} S_n (zI - \Gamma)^{-1} (\Gamma - \Gamma_n) (zI - \Gamma_n)^{-1} (X_{n+1}) dz \\ &= \int_{\mathcal{C}_n} z^{-1} S_n (zI - \Gamma)^{-1} (\Gamma - \Gamma_n) (zI - \Gamma)^{-1/2} \\ &\quad \times (zI - \Gamma)^{1/2} (zI - \Gamma_n)^{-1} (zI - \Gamma)^{1/2} (zI - \Gamma)^{-1/2} (X_{n+1}) dz \end{aligned}$$

and

$$\begin{aligned} &\left\| S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) \right\| \\ &\leq \int_{\mathcal{C}_n} \left| z^{-1/2} \right| \left\| z^{-1/2} S_n (zI - \Gamma)^{-1/2} \right\|_\infty \left\| (zI - \Gamma)^{-1/2} (\Gamma - \Gamma_n) (zI - \Gamma)^{-1/2} \right\|_\infty \\ &\quad \times \left\| (zI - \Gamma)^{1/2} (zI - \Gamma_n)^{-1} (zI - \Gamma)^{1/2} \right\|_\infty \left\| (zI - \Gamma)^{-1/2} (X_{n+1}) \right\| dz. \end{aligned}$$

By Lemma (5.5),

$$\begin{aligned}
& \left\| S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) \right\| \\
&= \left\| S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) \right\| \mathbb{1}_{\{\cap_j \mathcal{E}_j\}} + \left\| S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) \right\| \mathbb{1}_{\{\cup_j \mathcal{E}_j^c\}} \\
&\leq 2 \int_{\mathcal{C}_n} \left| z^{-1/2} \right| \left\| z^{-1} S_n (zI - \Gamma)^{-1/2} \right\|_\infty \left\| (zI - \Gamma)^{-1/2} (\Gamma - \Gamma_n) (zI - \Gamma)^{-1/2} \right\|_\infty \\
&\quad \times \left\| (zI - \Gamma)^{-1/2} (X_{n+1}) \right\| dz + \left\| S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) \right\| \mathbb{1}_{\cup_j \mathcal{E}_j^c}
\end{aligned} \tag{26}$$

Obviously $\sqrt{n} \left\| S_n \left(\Gamma_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) \right\| \mathbb{1}_{\cup_j \mathcal{E}_j^c}$ decays to zero in probability whenever $\sum_{j=1}^{k_n} \mathbb{P} \left(\mathcal{E}_j^c \right) \rightarrow 0$ i.e. when $\frac{1}{\sqrt{n}} \sum_{j=1}^{k_n} (j \log j) \sim \frac{k_n^2 \log k_n}{\sqrt{n}}$ does.

Let us turn to (26), tile it into two terms by decomposing X_{n+1} :

$$\begin{aligned}
W_1 &= \sum_{j=1}^{k_n} \int_{\mathcal{B}_j} \left| z^{-1/2} \right| \left\| z^{-1} S_n (zI - \Gamma)^{-1/2} \right\|_\infty \\
&\quad \times \left\| (zI - \Gamma)^{-1/2} (\Gamma - \Gamma_n) (zI - \Gamma)^{-1/2} \right\|_\infty \left\| (zI - \Gamma)^{-1/2} (\varepsilon_{n+1}) \right\| dz \\
W_2 &= \sum_{j=1}^{k_n} \int_{\mathcal{B}_j} \left| z^{-1/2} \right| \left\| z^{-1} S_n (zI - \Gamma)^{-1/2} \right\|_\infty \\
&\quad \times \left\| (zI - \Gamma)^{-1/2} (\Gamma - \Gamma_n) (zI - \Gamma)^{-1/2} \right\|_\infty \left\| (zI - \Gamma)^{-1/2} \rho(X_n) \right\| dz
\end{aligned}$$

and first prove that $\sqrt{n/k_n} W_1$ tends in probability to zero. Let us simplify this first term.

$$\begin{aligned}
W_1 &\leq \sum_{j=1}^{k_n} \frac{\delta_j}{\sqrt{|\lambda_j - \delta_j|}} \sup_{z \in \mathcal{B}_j} \left\{ \left\| z^{-1/2} S_n (zI - \Gamma)^{-1/2} \right\|_\infty \left\| (zI - \Gamma)^{-1/2} (\varepsilon_{n+1}) \right\| \right\} \\
&\quad \times \sup_{z \in \mathcal{B}_j} \left\{ \left\| (zI - \Gamma)^{-1/2} (\Gamma - \Gamma_n) (zI - \Gamma)^{-1/2} \right\|_\infty \right\}
\end{aligned}$$

hence

$$\begin{aligned}
\mathbb{E} W_1 &\leq \sum_{j=1}^{k_n} \frac{\delta_j}{\sqrt{|\lambda_j - \delta_j|}} \sqrt{\mathbb{E} \sup_{z \in \mathcal{B}_j} \left\{ \left\| (zI - \Gamma)^{-1/2} (\Gamma - \Gamma_n) (zI - \Gamma)^{-1/2} \right\|_\infty \right\}^2} \\
&\quad \times \sqrt{\mathbb{E} \sup_{z \in \mathcal{B}_j} \left\{ \left\| z^{-1/2} S_n (zI - \Gamma)^{-1/2} \right\|_\infty \right\}^2} \mathbb{E} \sup_{z \in \mathcal{B}_j} \left\{ \left\| (zI - \Gamma)^{-1/2} (\varepsilon_{n+1}) \right\| \right\}^2
\end{aligned} \tag{i}$$

$$\leq \frac{M}{n} \sum_{j=1}^{k_n} \frac{\delta_j}{\sqrt{|\lambda_j - \delta_j|}} (j \log j)^2 \tag{ii}$$

$$\leq \frac{M}{n} \sum_{j=1}^{k_n} \sqrt{\delta_j} j^2 (\log j)^2 \leq \frac{M}{n} k_n^{5/2} (\log k_n)^2$$

From (i) to (ii) I invoke Lemma 5.4, $\frac{\delta_j}{\sqrt{|\lambda_j - \delta_j|}}$ was bounded by $\sqrt{\delta_j}$, at last it is plain that $\sqrt{j\delta_j}$ is bounded. As a consequence of the above if one chooses k_n such that

$$\sqrt{\frac{n}{k_n}} \frac{1}{n} k_n^{5/2} (\log k_n)^2 = \frac{1}{\sqrt{n}} k_n^2 (\log k_n)^2 \rightarrow 0$$

we see that $\sqrt{\frac{n}{k_n}} W_1$ tends in probability to zero. We turn to the second term W_2 and like above

$$\begin{aligned} W_2 &\leq \sum_{j=1}^{k_n} \frac{\delta_j}{\sqrt{|\lambda_j - \delta_j|}} \sup_{z \in \mathcal{B}_j} \left\{ \left\| z^{-1/2} S_n (zI - \Gamma)^{-1/2} \right\|_{\infty} \left\| (zI - \Gamma)^{-1/2} \Gamma^{1/2} \tilde{\rho}(X_n) \right\| \right\} \\ &\quad \times \sup_{z \in \mathcal{B}_j} \left\{ \left\| (zI - \Gamma)^{-1/2} (\Gamma - \Gamma_n) (zI - \Gamma)^{-1/2} \right\|_{\infty} \right\} \end{aligned}$$

The situation is slightly more complicated than above since S_n is not independent from X_n . We introduce a truncation. Assume that τ_n is an increasing sequence tending to infinity.

$$\begin{aligned} W_2 &= W_2 \mathbf{I}_{\{\|X_n\| < \tau_n\}} + W_2 \mathbf{I}_{\{\|X_n\| \geq \tau_n\}} \\ &= W_2^- + W_2^+. \end{aligned}$$

Obviously $\sqrt{\frac{n}{k_n}} W_2^+$ tends in probability to zero since for all $\varepsilon > 0$

$$\mathbb{P} \left(\sqrt{\frac{n}{k_n}} W_2^+ > \varepsilon \right) \leq \mathbb{P} (\|X_n\| \geq \tau_n) \leq \frac{\mathbb{E} \|X_1\|}{\tau_n}.$$

We turn to

$$\begin{aligned} W_2^- &\leq \|\tilde{\rho}\| \tau_n \sum_{j=1}^{k_n} \frac{\delta_j}{\sqrt{|\lambda_j - \delta_j|}} \sup_{z \in \mathcal{B}_j} \left\{ \left\| z^{-1/2} S_n (zI - \Gamma)^{-1/2} \right\|_{\infty} \left\| (zI - \Gamma)^{-1/2} \Gamma^{1/2} \right\|_{\infty} \right\} \\ &\quad \times \sup_{z \in \mathcal{B}_j} \left\{ \left\| (zI - \Gamma)^{-1/2} (\Gamma - \Gamma_n) (zI - \Gamma)^{-1/2} \right\|_{\infty} \right\}, \\ \mathbb{E} W_2^- &\leq \|\tilde{\rho}\| \tau_n \sum_{j=1}^{k_n} \frac{\delta_j}{\sqrt{|\lambda_j - \delta_j|}} \sup_{z \in \mathcal{B}_j} \left\{ \left\| (zI - \Gamma)^{-1/2} \Gamma^{1/2} \right\|_{\infty} \right\} \sqrt{\mathbb{E} \sup_{z \in \mathcal{B}_j} \left\{ \left\| z^{-1/2} S_n (zI - \Gamma)^{-1/2} \right\|_{\infty}^2 \right\}} \\ &\quad \times \sqrt{\mathbb{E} \sup_{z \in \mathcal{B}_j} \left\{ \left\| (zI - \Gamma)^{-1/2} (\Gamma - \Gamma_n) (zI - \Gamma)^{-1/2} \right\|_{\infty}^2 \right\}} \\ &\leq M \frac{\tau_n}{n} \sum_{j=1}^{k_n} \frac{\delta_j}{\sqrt{|\lambda_j - \delta_j|}} j^2 (\log j)^{3/2} \leq M \frac{\tau_n k_n^{5/2} (\log k_n)^{3/2}}{n} \end{aligned}$$

hence $\sqrt{\frac{n}{k_n}} W_2^-$ tends in probability to zero whenever $\frac{\tau_n k_n^2 (\log k_n)^{3/2}}{\sqrt{n}} \rightarrow 0$. Now we choose $\tau_n = \sqrt{\log k_n}$ with k_n as above for W_1 . This finishes the proof of Proposition 5.1.

5.3 Weakly convergent term

As seen from (11) and from previous subsection $S_n \Gamma^\dagger (X_{n+1})$ will fully determine the asymptotics of the predictor :

$$\begin{aligned} S_n \Gamma^\dagger (X_{n+1}) &= \sum_{k=1}^n \left\langle X_{k-1}, \Gamma^\dagger (X_{n+1}) \right\rangle \varepsilon_k \\ &= \sum_{k=1}^n Z_{k,n}. \end{aligned}$$

We decompose $Z_{k,n}$ in three terms

$$\begin{aligned} Z_{k,n} &= Z_{k,n}^+ + Z_{k,n}^0 + Z_{k,n}^- \\ Z_{k,n}^+ &= \left\langle \Gamma^\dagger X_{k-1}, \varepsilon_{n+1} + \rho(\varepsilon_n) + \dots + \rho^{n-k}(\varepsilon_{k+1}) \right\rangle \varepsilon_k \\ Z_{k,n}^0 &= \left\langle \Gamma^\dagger X_{k-1}, \rho^{n+1-k} \varepsilon_k \right\rangle \varepsilon_k \\ Z_{k,n}^- &= \left\langle \Gamma^\dagger X_{k-1}, \rho^{n+2-k}(X_{k-1}) \right\rangle \varepsilon_k \end{aligned}$$

stemming from

$$X_{n+1} = \varepsilon_{n+1} + \rho(\varepsilon_n) + \dots + \rho^{n+1-k}(\varepsilon_k) + \rho^{n+2-k}(X_{k-1}).$$

We will show in Lemma 5.10 below that the series involving $Z_{k,n}^0$ and $Z_{k,n}^-$ are negligible ; weak convergence is strictly determined by $\sum_{k=1}^n Z_{k,n}^+$. The asymptotic distribution is given at Proposition 5.2 below. We begin with an important Lemma.

Lemma 5.7 *The random sequences $Z_{k,n}^+$ and $Z_{k,n}^-$ are Hilbert-valued martingale difference arrays w.r.t. the sequence $(\mathcal{F}_i)_{i \leq k}$ where \mathcal{F}_i is the σ -algebra generated by $(\varepsilon_l)_{l \leq i}$*

Proof :

Denoting

$$\begin{aligned} X_{k,n}^\sharp &= \varepsilon_{n+1} + \rho(\varepsilon_n) + \dots + \rho^{n-k}(\varepsilon_{k+1}), \\ \mathbb{E} \left(Z_{k,n}^+ | \mathcal{F}_{k-1} \right) &= \mathbb{E} \left(\left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\sharp \right\rangle \varepsilon_k | \mathcal{F}_{k-1} \right). \end{aligned}$$

Since ε_k is independent from $X_{k,n}^\sharp$ and both sequences of random elements are centered we deduce that

$$\mathbb{E} \left(Z_{k,n}^+ | \mathcal{F}_{k-1} \right) = 0.$$

Then

$$\begin{aligned} \mathbb{E} \left(Z_{k,n}^- | \mathcal{F}_{k-1} \right) &= \mathbb{E} \left(\left\langle \Gamma^\dagger X_{k-1}, \rho^{n+2-k}(X_{k-1}) \right\rangle \varepsilon_k | \mathcal{F}_{k-1} \right) \\ &= \left\langle \Gamma^\dagger X_{k-1}, \rho^{n+2-k}(X_{k-1}) \right\rangle \mathbb{E}(\varepsilon_k | \mathcal{F}_{k-1}) \\ &= 0 \end{aligned}$$

Proposition 5.2

$$S_n^+ = \frac{1}{\sqrt{nk_n}} \sum_{k=1}^n Z_{k,n}^+ \xrightarrow{w} \mathcal{G}(0, \Gamma_\varepsilon).$$

Proof of the Proposition :

Since $\sum_{k=1}^n Z_{k,n}^+$ is a \mathcal{H} -valued martingale difference array we first could hope to apply existing criteria for weak convergence of such sequences. Most of these criteria (see Walk [29] or Rackauskas [26]) rely on convergence in probability for the conditional covariance operator. They do not seem to be adapted in this context (I could not go through with it...). I propose the reader to come back to the "sources" of the Central Limit Theorem on infinite dimensional vector spaces. We will simply prove that S_n^+ is a uniformly tight sequence and that finite distributions, when computed on a sufficiently large set of functionals converge to gaussian limits, hence characterizing the limiting covariance operator Γ_ε . In order to understand this approach I refer to the paper by A. de Acosta [1], especially to Theorem 2.3 p.279.

For further purpose we begin with a first Lemma in which covariance and cross-covariance operators for the array $Z_{k,n}^+$ are computed.

Lemma 5.8 *If $k < i$, $\mathbb{E} \left(Z_{k,n}^+ \otimes Z_{i,n}^+ \right) = 0$ and*

$$\mathbb{E} \left(Z_{k,n}^+ \otimes Z_{k,n}^+ \right) = \Gamma_\varepsilon \left(k_n - \text{tr} \left(\Gamma^\dagger \rho^{n-k+1} \Gamma (\rho^*)^{n-k+1} \right) \right).$$

Proof.

$$Z_{k,n}^+ \otimes Z_{i,n}^+ = \left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle \left\langle \Gamma^\dagger X_{i-1}, X_{i,n}^\# \right\rangle (\varepsilon_k \otimes \varepsilon_i)$$

and since $X_{i-1} = \rho^{i-k} (X_{k-1}) + \varepsilon_{i-1} + \dots + \rho^{i-1-k} (\varepsilon_k)$. We tile $Z_{k,n}^+ \otimes Z_{i,n}^+$ into two terms. We see that

$$\mathbb{E} \left[\left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle \left\langle \Gamma^\dagger \rho^{i-k} (X_{k-1}), X_{i,n}^\# \right\rangle (\varepsilon_k \otimes \varepsilon_i) \right] = 0$$

since ε_k is independent from all the other terms. The second term is :

$$\left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle \left\langle \Gamma^\dagger \left(\varepsilon_{i-1} + \dots + \rho^{i-1-k} (\varepsilon_k) \right), X_{i,n}^\# \right\rangle (\varepsilon_k \otimes \varepsilon_i).$$

Its expectation is null since X_{k-1} is centered and independent from all the other terms. We focus on the second part of the Lemma.

We have

$$\begin{aligned} \mathbb{E} \left(Z_{k,n}^+ \otimes Z_{k,n}^+ \right) &= \left(\mathbb{E} \left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle^2 \right) \mathbb{E} (\varepsilon_k \otimes \varepsilon_k) \\ &= \left(\mathbb{E} \left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle^2 \right) \Gamma_\varepsilon \end{aligned}$$

and

$$\begin{aligned}
\mathbb{E} \left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle^2 &= \mathbb{E} \left(\mathbb{E} \left\langle X_{k-1}, \Gamma^\dagger X_{k,n}^\# \right\rangle^2 \mid X_{k,n}^\# \right) \\
&= \mathbb{E} \left\| \Gamma^{1/2} \Gamma^\dagger X_{k,n}^\# \right\|^2 \\
&= \mathbb{E} \left\| \Gamma^{\dagger 1/2} X_{k,n}^\# \right\|^2 \\
&= \text{tr} \left(\Gamma^\dagger \Gamma_{k,n}^\# \right)
\end{aligned}$$

where

$$\begin{aligned}
\Gamma_{k,n}^\# &= \mathbb{E} \left(X_{k,n}^\# \otimes X_{k,n}^\# \right) \\
&= \Gamma_\varepsilon + \rho \Gamma_\varepsilon \rho^* + \dots + \rho^{n-k} \Gamma_\varepsilon (\rho^*)^{n-k} \\
&= \Gamma - \rho^{n-k+1} \Gamma (\rho^*)^{n-k+1}.
\end{aligned}$$

Then

$$\begin{aligned}
\text{tr} \left(\Gamma^\dagger \Gamma_{k,n}^\# \right) &= \text{tr} \left(\Gamma^\dagger \Gamma \right) - \text{tr} \left(\Gamma^\dagger \rho^{n-k+1} \Gamma (\rho^*)^{n-k+1} \right) \\
&= k_n - \text{tr} \left(\Gamma^\dagger \rho^{n-k+1} \Gamma (\rho^*)^{n-k+1} \right).
\end{aligned}$$

The proof of Lemma 5.8 is complete. \blacksquare

Now we prove that all the finite-dimensional distributions converge to a gaussian limit. It suffices to get, for all x in \mathcal{H} ,

$$\frac{1}{\sqrt{nk_n}} \sum_{k=1}^n \left\langle Z_{k,n}^+, x \right\rangle \xrightarrow{w} \mathcal{N} \left(0, \sigma_{\varepsilon,x}^2 \right) \quad (27)$$

where $\sigma_{\varepsilon,x}^2 = \mathbb{E} \langle \varepsilon_k, x \rangle^2$.

Since $\langle Z_{k,n}^+, x \rangle$ is a real valued MDA it suffices to apply the criteria given in Mac Leish [17].

In view of Lemma (5.8) it is enough to prove that $\sum_{k=1}^n \text{tr} \left(\Gamma^\dagger \Gamma_{k,n}^\# \right) \sim nk_n$ that is

$$\frac{\sum_{k=1}^n \text{tr} \left(\Gamma^\dagger \Gamma_{k,n}^\# \right) - nk_n}{nk_n} = \frac{\sum_{k=1}^n \text{tr} \left(\Gamma^\dagger \rho^{n-k+1} \Gamma (\rho^*)^{n-k+1} \right)}{nk_n} \rightarrow 0$$

The usual properties of the trace provide

$$\begin{aligned}
\left| \text{tr} \left(\Gamma^\dagger \rho^{n-k+1} \Gamma (\rho^*)^{n-k+1} \right) \right| &= \left| \text{tr} \left((\rho^*)^{n-k+1} \Gamma^\dagger \rho^{n-k+1} \Gamma \right) \right| \\
&= \left| \text{tr} \left((\rho^*)^{n-k+1} \Gamma^\dagger \rho^{n-k+1} \Gamma \right) \right| \\
&\leq \left\| (\rho^*)^{n-k+1} \Gamma^\dagger \rho^{n-k+1} \right\|_\infty |\text{tr} \Gamma| \\
&= \left\| (\rho^*)^{n-k} \tilde{\rho}^* \Gamma^{1/2} \Gamma^\dagger \Gamma^{1/2} \tilde{\rho} \rho^{n-k} \right\|_\infty |\text{tr} \Gamma| \\
&\leq \left\| \rho^{n-k} \right\|^2 \|\tilde{\rho}^*\|_\infty \|\tilde{\rho}\|_\infty |\text{tr} \Gamma|
\end{aligned}$$

and we see that whenever $nk_n \rightarrow +\infty$

$$\sum_{k=1}^n \left(\mathbb{E} \left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle^2 \right) \sim nk_n \quad (28)$$

which ensures (27).

Now we turn to the second part of the proof, namely : "the sequence $(S_n^+)_{n \in \mathbb{N}}$ is tight". Once more we go through a Lemma.

Lemma 5.9 *By \mathcal{P}_m we denote the projector associated to the m first eigenvectors of the covariance operator Γ_ε of ε_1 . Then,*

$$\limsup_{m \rightarrow +\infty} \sup_n \mathbb{P} \left(\|(I - \mathcal{P}_m) S_n^+\| > \varepsilon \right) = 0. \quad (29)$$

Remark 5.1 *What we prove is "with prescribed probability the sequence S_n^+ is concentrated in the ε -neighborhood of a finite dimensional space -i.e. $\text{Im}(\mathcal{P}_m)$ ". This phenomenon is called flat concentration and ensures the tightness of $(S_n^+)_{n \in \mathbb{N}}$ (see de Acosta (1970), Definition 2.1 p.279).*

Proof of Lemma 5.9 :

$$\mathbb{P} \left(\|(I - \mathcal{P}_m) S_n^+\| > \varepsilon \right) \leq \frac{\mathbb{E} \left(\|(I - \mathcal{P}_m) S_n^+\|^2 \right)}{\varepsilon^2}$$

where

$$\begin{aligned} \mathbb{E} \left(\|(I - \mathcal{P}_m) S_n^+\|^2 \right) &= \frac{1}{nk_n} \mathbb{E} \left(\left\| \sum_{k=1}^n \left\langle \Gamma^\dagger X_{k-1}, \varepsilon_{n+1} + \rho(\varepsilon_n) + \dots + \rho^{n-k}(\varepsilon_{k+1}) \right\rangle (I - \mathcal{P}_m) \varepsilon_k \right\|^2 \right) \quad (i) \\ &= \frac{1}{nk_n} \left(\sum_{k=1}^n \mathbb{E} \left(\left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle^2 \|(I - \mathcal{P}_m) \varepsilon_k\|^2 \right) \right) \quad (ii) \\ &= \frac{1}{nk_n} \mathbb{E} \|(I - \mathcal{P}_m) \varepsilon_k\|^2 \left(\sum_{k=1}^n \mathbb{E} \left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle^2 \right) \\ &= \frac{1}{nk_n} \text{tr} \left((I - \mathcal{P}_m) \Gamma_\varepsilon \right) \left(\sum_{k=1}^n \mathbb{E} \left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle^2 \right). \end{aligned}$$

On line (ii) the expectation of all the cross products is null. I skip through these calculations since they are exactly alike thoses carried within Lemma 5.8 above. The computations made in the first part of the proof (see display (28)) are useful here. They ensure that

$$\sup_n \frac{1}{nk_n} \left(\sum_{k=1}^n \mathbb{E} \left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle^2 \right) < M$$

where M is some universal constant. At last letting m tend to infinity we get

$$\lim_{m \rightarrow +\infty} \text{tr} \left((I - \mathcal{P}_m) \Gamma_\varepsilon \right) = 0$$

which proves Lemma 5.9.

It remains to conclude. Lemma 5.9 ensures that the centered sequence S_n^+ is tight. By (27) we know that the weak limit is gaussian and that its covariance function (hence its covariance operator) is fully characterized : the same as ε_1 . We invoke for instance A. de Acosta (1970) to conclude the proof of Proposition 5.2.

Lemma 5.10

$$\frac{1}{\sqrt{nk_n}} \sum_{k=1}^n Z_{k,n}^- \xrightarrow{\mathbb{P}} 0, \tag{30}$$

$$\frac{1}{\sqrt{nk_n}} \sum_{k=1}^n Z_{k,n}^0 \xrightarrow{\mathbb{P}} 0. \tag{31}$$

Proof :

It is plain that $Z_{k,n}^-$ is an array of non-correlated random elements. We prove that

$$\frac{1}{nk_n} \mathbb{E} \left\| \sum_{k=1}^n Z_{k,n}^- \right\|^2 \rightarrow 0$$

$$\begin{aligned} \mathbb{E} \left\| \sum_{k=1}^n Z_{k,n}^- \right\|^2 &= \mathbb{E} \|\varepsilon_1\|^2 \sum_{k=1}^n \mathbb{E} \left\langle \Gamma^\dagger X_{k-1}, \rho^{n+2-k} (X_{k-1}) \right\rangle^2 \\ &= \mathbb{E} \|\varepsilon_1\|^2 \sum_{k=1}^n \mathbb{E} \left\langle \left(\Gamma^\dagger\right)^{1/2} X_{k-1}, \Gamma^{-1/2} \rho^{n+2-k} (X_{k-1}) \right\rangle^2 \\ &\leq \mathbb{E} \|\varepsilon_1\|^2 \sum_{k=1}^n \mathbb{E} \left[\left\| \left(\Gamma^\dagger\right)^{1/2} X_{k-1} \right\|^2 \|X_{k-1}\|^2 \right] \left\| \Gamma^{-1/2} \rho^{n+2-k} \right\|_\infty \\ &= \|\tilde{\rho}\|_\infty \mathbb{E} \|\varepsilon_1\|^2 \mathbb{E} \left[\left\| \left(\Gamma^\dagger\right)^{1/2} X_1 \right\|^2 \|X_1\|^2 \right] \sum_{k=1}^n \left\| \rho^{n+1-k} \right\|_\infty. \end{aligned}$$

Since KL expansion yields

$$\left\| \left(\Gamma^\dagger\right)^{1/2} X_1 \right\|^2 \|X_1\|^2 =_d \sum_{i=1}^{k_n} \xi_i^2 \sum_{j=1}^{+\infty} \lambda_j \xi_j^2$$

we easily see by assumption **A**₂ that

$$\mathbb{E} \left[\left\| \left(\Gamma^\dagger\right)^{1/2} X_1 \right\|^2 \|X_1\|^2 \right] = O(k_n) \tag{32}$$

hence (30).

We turn to obtaining a bound for the second term. With $Z_{k,n}^0 = \langle \Gamma^\dagger X_{k-1}, \rho^{n+1-k} \varepsilon_k \rangle \varepsilon_k$ we

get :

$$\begin{aligned}
\mathbb{E} \left\| \sum_{k=1}^n Z_{k,n}^0 \right\|^2 &= \sum_{k=1}^n \mathbb{E} \|Z_{k,n}^0\|^2 + 2 \sum_{1 \leq i < j \leq n} \mathbb{E} \langle Z_{i,n}^0, Z_{j,n}^0 \rangle \\
&= \sum_{k=1}^n \mathbb{E} \left\langle \Gamma^\dagger X_{k-1}, \rho^{n+1-k} \varepsilon_k \right\rangle^2 \|\varepsilon_k\|^2 \\
&\quad + 2 \sum_{1 \leq i < j \leq n} \mathbb{E} \left(\left\langle \Gamma^\dagger X_{i-1}, \rho^{n+1-i} \varepsilon_i \right\rangle \langle \varepsilon_i, \varepsilon_j \rangle \left\langle \Gamma^\dagger X_{j-1}, \rho^{n+1-j} \varepsilon_j \right\rangle \right).
\end{aligned}$$

The first term may be bounded by

$$\begin{aligned}
&\sum_{k=1}^n \mathbb{E} \left[\left\| \left(\Gamma^\dagger \right)^{1/2} \rho^{n+1-k} \varepsilon_k \right\|^2 \|\varepsilon_k\|^2 \right] \\
&\leq \|\tilde{\rho}\|_\infty \mathbb{E} \left(\|\varepsilon_1\|^4 \right) \sum_{k=1}^n \left\| \rho^{n-k} \right\|_\infty.
\end{aligned}$$

The second term may be rewritten :

$$\begin{aligned}
&\sum_{1 \leq i < j \leq n} \mathbb{E} \left\langle \Gamma^\dagger X_{i-1}, \rho^{n+1-i} \varepsilon_i \right\rangle \langle \varepsilon_i, \varepsilon_j \rangle \left\langle \Gamma^\dagger X_{j-1}, \rho^{n+1-j} \varepsilon_j \right\rangle \\
&= \sum_{1 \leq i < j \leq n} \mathbb{E} \left\langle \Gamma^\dagger X_{i-1}, \rho^{n+1-i} \varepsilon_i \right\rangle \left\langle \Gamma^\dagger \rho^{n+1-j} \Gamma_\varepsilon(\varepsilon_i), X_{j-1} \right\rangle \\
&= \sum_{1 \leq i < j \leq n} \mathbb{E} \left\langle \Gamma^\dagger X_{i-1}, \rho^{n+1-i} \varepsilon_i \right\rangle \left\langle \Gamma^\dagger \rho^{n+1-j} \Gamma_\varepsilon(\varepsilon_i), \rho^{j-i} X_{i-1} \right\rangle \\
&= \sum_{1 \leq i < j \leq n} \mathbb{E} \left\langle \left(\Gamma^\dagger \right)^{1/2} X_{i-1}, \tilde{\rho} \rho^{n-i} \varepsilon_i \right\rangle \left\langle \tilde{\rho} \rho^{n-j} \Gamma_\varepsilon(\varepsilon_i), \tilde{\rho} \rho^{j-i-1} X_{i-1} \right\rangle.
\end{aligned}$$

Taking absolute values we get the bound

$$\|\tilde{\rho}\|_\infty^3 \mathbb{E} \left(\|\varepsilon_1\|^2 \right) \mathbb{E} \left[\|X_1\| \left\| \left(\Gamma^\dagger \right)^{1/2} X_1 \right\| \right] \|\Gamma_\varepsilon\|_\infty \sum_{1 \leq i < j \leq n} \|\rho^{n-i}\|_\infty \|\rho^{n-j}\|_\infty \|\rho^{j-i-1}\|_\infty.$$

Once again invoking (32) we get (31) in Lemma 5.10.

Proof of Theorem 3.1 :

From all that was done above it is straightforward to deduce that weak convergence for $\rho_n - \rho$ depends only on the term $S_n \Gamma^\dagger$ in (11). We recall it : $S_n \Gamma^\dagger = \sum_{k=1}^n \Gamma^\dagger X_{k-1} \otimes \varepsilon_k$. I guess the reader will agree with the following sentences : "Assume that $(\varepsilon_k)_{k \in \mathbb{Z}}$ and $(X_k)_{k \in \mathbb{Z}}$ are independent sequences of independent random elements in \mathcal{H} . Then if in this framework $S_n \Gamma^\dagger$ does not converge weakly $S_n \Gamma^\dagger$ will not converge weakly in the setting of model (1)". Obviously the situation is much favourable assuming independence "everywhere".

Let us assume that $\frac{\alpha_n}{n} S_n \Gamma^\dagger$ converges weakly to some random variable Z for some increasing sequence α_n . We deduce that, for any $f \in \mathcal{K}^*$, the dual space of \mathcal{K}^* ,

$$f \left(\frac{\alpha_n}{n} S_n \Gamma^\dagger \right) = \frac{\alpha_n}{n} f \left(S_n \Gamma^\dagger \right) = \frac{\alpha_n}{n} \sum_{k=1}^n f \left(\Gamma^\dagger X_{k-1} \otimes \varepsilon_k \right)$$

converges weakly to $f(Z)$. In fact $\mathcal{K}^* = \mathcal{K}_1$ the space of trace class operators (see Dunford-Schwartz [13] for this classical result), the duality bracket is nothing than the usual trace. Consequently we should investigate weak convergence for

$$\frac{\alpha_n}{n} \sum_{k=1}^n \text{tr} \left[T \left(\Gamma^\dagger X_{k-1} \otimes \varepsilon_k \right) \right] = \frac{\alpha_n}{n} \sum_{k=1}^n \left\langle \Gamma^\dagger X_{k-1}, T \varepsilon_k \right\rangle$$

where T is a trace class operator. To prove Theorem 3.1, it is enough to take $T = u \otimes v$, $u, v \in \mathcal{H}$. Indeed

$$f \left(\frac{\alpha_n}{n} S_n \Gamma^\dagger \right) = \frac{\alpha_n}{n} \sum_{k=1}^n \left\langle \Gamma^\dagger X_{k-1}, T \varepsilon_k \right\rangle = \frac{\alpha_n}{n} \sum_{k=1}^n \left\langle X_{k-1}, \Gamma^\dagger v \right\rangle \langle u, \varepsilon_k \rangle$$

Now we consider two cases depending on the location of v :

1. If $v \in D(\Gamma^{-1})$, $\Gamma^\dagger v$ is a bounded sequence that converges to $\Gamma^{-1}v$. It is straightforward to see that $f \left(\frac{1}{\sqrt{n}} S_n \Gamma^\dagger \right)$ converges in distribution to $f(Z)$ (which is gaussian) by the real CLT for i.i.d. r.v. **This means that necessarily** $\alpha_n = \sqrt{n}$.
2. Let us take a general $v \notin D(\Gamma^{-1})$, and compute the variance of the series above with $\alpha_n = \sqrt{n}$

$$\begin{aligned} \mathbb{E} \left[f \left(\frac{1}{\sqrt{n}} S_n \Gamma^\dagger \right) \right]^2 &= \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left\langle X_{k-1}, \Gamma^\dagger v \right\rangle^2 \mathbb{E} \langle u, \varepsilon_k \rangle^2 \\ &= \frac{\sigma_{\varepsilon, u}^2}{n} \sum_{k=1}^n \mathbb{E} \left\langle X_{k-1}, \Gamma^\dagger v \right\rangle^2 \\ &= \sigma_{\varepsilon, u}^2 \left\| \Gamma^{1/2} \Gamma^\dagger v \right\|^2 = \sigma_{\varepsilon, u}^2 \left\| \left(\Gamma^\dagger \right)^{1/2} v \right\|^2 \end{aligned}$$

where $\sigma_{\varepsilon, u}^2 = \mathbb{E} \langle u, \varepsilon_1 \rangle^2$ and

$$\left\| \left(\Gamma^\dagger \right)^{1/2} v \right\|^2 = \sum_{i=1}^{k_n} \frac{\langle v, e_i \rangle^2}{\lambda_i}$$

Choosing $\langle v, e_i \rangle^2 = \lambda_i$ or $\langle v, e_i \rangle^2 = \lambda_i \beta_i$ where $\beta_i \rightarrow \beta > 0$ we see that $\left\| \left(\Gamma^\dagger \right)^{1/2} v \right\|^2 \rightarrow +\infty$ and the real valued random variable $f \left((1/\sqrt{n}) S_n \Gamma^\dagger \right)$ cannot converge weakly since its variance tends to infinity. This shows that the marginals of $(\alpha_n/n) S_n \Gamma^\dagger$ do not all converge to the same limiting measure and not all at the same rate, which prevents weak convergence in the topology of \mathcal{K} . Hence Theorem 3.1.

References

- [1] A. de Acosta, Existence and convergence of probability measures in Banach spaces, Trans. Amer. Math. Soc. 152 (1970) 273-298.
- [2] R.A. Adams, Sobolev spaces, Academic Press, (1975).

- [3] A. Antoniadis, T. Sapatinas, Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes, *J. Multivariate. Anal.*, 87 (2003) 133-158.
- [4] A. Araujo, E. Gine, *The Central Limit Theorem for Real and Banach Valued Random Variables*, Wiley Series in Probability and Mathematical Statistics, 1980.
- [5] V. Arsenin, A. Tikhonov, *Solutions of ill-posed problems*, Winston and Sons, Washington D.C.,1977.
- [6] P. Besse, H. Cardot, Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1, *Canad. J. Statist*, 24 (1996) 467-487.
- [7] P. Besse, H. Cardot, D. Stephenson, Autoregressive forecasting of some climatic variations, *Scand. J. Statist*, 27 (2000), 673-687.
- [8] P. Billingsley, *Convergence of probability measures*. Wiley Series in Probability and Mathematical Statistics, 1968.
- [9] D. Bosq, Modelization , nonparametric estimation and prediction for continuous time processes. in: Roussas (Ed) *Nato Asi Series C*, 335 (1991) 509-529.
- [10] D. Bosq, *Linear processes in function spaces*. Lectures notes in statistics. Springer Verlag, 2000.
- [11] H. Cardot, A. Mas, P. Sarda, CLT in functional linear models, submitted manuscript. Available at http://fr.arxiv.org/PS_cache/math/pdf/0508/0508073.pdf
- [12] A. Cavallini, G.C. Montanari, M. Loggini, O. Lessi, M. Cacciari, Nonparametric prediction of harmonic levels in electrical networks. *Proceed. IEEE ICHPS VI*, Bologna (1994) 165-171.
- [13] N. Dunford, J.T. Schwartz, *Linear Operators Vol I,II,III*. Wiley Classics Library, 1988.
- [14] I. Gohberg, S. Goldberg, M.A. Kaashoek, *Classes of linear operators, vol I,II*, Operator theory : advances and applications, Birkhäuser Verlag, 1991.
- [15] C. Groetsch.: *Inverse Problems in the Mathematical Sciences*, Vieweg, Wiesbaden, 1993.
- [16] M. Ledoux, M. Talagrand, *Probability on Banach spaces-Isoperimetry and Processes*, Springer Verlag, Berlin, 1991.
- [17] D.L Mc Leish, Dependent central limit theorem and invariance principles, *Ann. Probab.* 2 (1974) 620-628.
- [18] J.M. Marion, B. Pumo, Comparaison des modèles ARH(1) et ARHD(1) sur des données physiologiques (in French), *Ann. Isup*, (2004) 29-38.
- [19] A. Mas, Normalité asymptotique de l'estimateur empirique de l'opérateur d'autocorrélation d'un processus ARH(1). *C.R. Acad.Sci.*, t.329, Sér. I (1999) 899-902.
- [20] A. Mas, Estimation d'opérateurs de corrélation de processus linéaires fonctionnels : lois limites, tests, déviations modérées, PhD Thesis (in French and English), Université Paris 6, 2000.

- [21] A. Mas, Weak convergence for the covariance operators of a Hilbertian linear process. *Stoch Process. App.* 99 (2002), 117-135.
- [22] A. Mas, L. Menneteau, Large and moderate deviations for infinite-dimensional autoregressive processes, *J. Multivar. Anal.*, 87 (2003), 241-260.
- [23] L. Menneteau, Some laws of the iterated logarithm in Hilbertian autoregressive models *J. Multivariate. Anal.* 92 (2005) 405-425.
- [24] T. Mourid, Processus autorégressifs d'ordre supérieur. *C.R. Acad.Sci.*, t.317, Sér. I (1993), 1167-1172.
- [25] B. Pumo, Prediction of continuous time processes by $C[0, 1]$ -valued autoregressive process, *Stat. Infer. Stoch. Processes*, 3 (1999) 1-13.
- [26] A. Rackauskas, On the conditional covariance condition in the martingale CLT, *Lith. Math. Journal*, 35, n°1, (1995) 93-104.
- [27] J.A. Rice, B.W. Silverman, Estimating the mean and covariance structure nonparametrically when the data are curves, *J.R.S.S. Ser. B*, 53 (1991), 233-243.
- [28] N.N. Vakhania, V.I. Tarieladze, S.M. Chobanyan, *Probability Distributions on Banach Spaces. Mathematics and its Applications.* D. Reidel Publishing, 1987.
- [29] H. Walk, An invariance principle for the Robbins-monroe process in a Hilbert space, *Z. Wahrsch. verw. Geb.* 39, (1977) 135-150.