

Estimation error for blind Gaussian time series filtering

T. Espinasse, F. Gamboa and J-M. Loubes

February 1, 2010

Abstract

In the frame of time series analysis, we compute the quadratic error in the blind estimation of the projection operator for prediction with infinite past. The estimation is made using only a single finite sample of the time series. It is performed by plugging the empirical covariance in a clever Schur complement decomposition of the projector.

Keywords: Asymptotic Statistics, Covariance Estimation, Time Series.

Subject Class. MSC-2000: 62G05, 62G20.

Introduction

In many concrete situations the statistician observes a finite path X_1, \dots, X_n of a real temporal phenomena. A common modeling is to assume that the observation is a finite path of a second order weak stationary process $\mathbf{X} := (X_t)_{t \in \mathbb{Z}}$ (we refer, for example, to [10], [14] and references therein). This means that the random variable (r.v.) X_t is, for any $t \in \mathbb{Z}$, square integrable and that the mean (supposed to be equal to zero from now) and the covariance structure of the process is invariant by any translation on the time index. That is, for any $t, s \in \mathbb{Z}$, $\mathbb{E}(X_t) = 0$ and $\mathbb{E}(X_t X_s)$ only depends on the distance between t and s . A more popular frame is the Gaussian case where the additional Gaussianity assumption on all finite marginal distributions of the process $(X_t)_{t \in \mathbb{Z}}$ is added. In this case, as the multidimensional Gaussian distribution only depends on moments of order one and two, the process is also strongly stationary. This means that the law of all finite dimensional marginal distributions are invariant if the time is shifted:

$$(X_1, \dots, X_n) \stackrel{\mathcal{L}}{=} (X_{t+1}, \dots, X_{t+n}), \quad (t \in \mathbb{Z}, n \in \mathbb{N}).$$

Gaussian stationary process are very popular because they share plenty of very nice properties concerning their statistical identification and prediction (see, for example, [2] or [18]). For instance, a well known property of Gaussian time series is that linear prediction is optimal. Hence, if one wish to predict for $t \geq 0$, X_t from X_{-N}, \dots, X_{-1} ($N > 0$) the r.v. minimising the prediction error is just a linear combination of X_{-N}, \dots, X_{-1} involving a linear projector operator onto the complete infinite past $\text{Proj}_{H_{\mathbb{Z}^-}^X}$ defined on the Hilbert space H_{∞}^X generated by the process \mathbf{X} . Furthermore, $\text{Proj}_{H_{\mathbb{Z}^-}^X}$ may be computed from the covariance function of the process (see Section 1 for precise statements). In this paper, we work with a Gaussian stationary process and we will address the problem of blind filtering.

This means that observing the finite sample path X_{-N}, \dots, X_{-1} of the process we wish to predict the future values X_t , $t \geq 0$ without knowing the covariance structure (blind means that the covariance of \mathbf{X} is unknown). The prediction problem is classical (see for example [9],[1], [3]). The blind one is less classical and have been very few studied. The particular case of Kriging (see [15], [18]) in which a parametric model on the covariance is assumed is discussed and studied in [18]. An interesting work of Bickel and al [4] consider the case where many samples are available. In our work, we do not assume any parametric model and set and solve the problem in a nonparametric way. Our method relies on the estimation of the covariance function. This function is estimated and an empirical regularisation method jointly with a clever Schur decomposition of the inverse covariance operator (see Theorem 2.1) allow to build an accurate estimate of $\text{Proj}_{H_{\mathbb{Z}^-}^X}$. The asymptotic property of this estimate is study and rate of convergence are stated. Roughly speaking, the rate of convergence stated in the main Theorem 3.3 depends of the regularity of the covariance function through the spectral density of the process (one of the main assumption is that the spectral measure of \mathbf{X} has a density).

The paper is organized as follows. The next Section is devoted to some recalls and technical results on time series analysis that will be used throughout the paper. In Section 2 we set and discuss the estimation procedure. Further, the asymptotic behavior and the rate of convergence of the estimator are tackled in Section 3. Section 4 is devoted to some numerical illustrations of the estimation procedure. All the proofs are postponed to the Appendix in Section 5.

1 Notations and preliminary definitions

Recall that $X = (X_k)_{k \in \mathbb{Z}}$, is a zero-mean Gaussian stationary process. Denote $r(k-j) := \mathbb{E}[X_k X_j]$. We assume that the sequence $(r_k)_{k \in \mathbb{Z}}$ is in $l_2(\mathbb{Z})$. This implies the existence of the $\mathbb{L}_2(\mathbb{T})$ function

$$f^*(t) := \sum_{k=-\infty}^{\infty} r(k) e^{ikt},$$

where $\mathbb{T} = [0, 2\pi)$. This function, called the spectral density of the time series, is a real even non negative function. As X is a Gaussian process, it conveys all the information on the process distribution. Furthermore, the covariance operator is a Toeplitz operator operating on $l_2(\mathbb{Z})$ associated to f^* , denoted $R = T_{\infty}(f^*)$ (for a thorough overview on the subject, we refer to [8]). Let $T_N(f^*)$ be the covariance matrix of the process observed at N following times. $(X_{i+1}, \dots, X_{i+N}), \forall i \in \mathbb{Z}$.

In this section, we recall some useful definitions and results about spectral analysis of time series.

For any finite or infinite subset $B \subset \mathbb{Z}$, we denote by H_B^X the partial past of the time series X for the values indexed by B , that is the information conveyed by the observation of the X_k , $k \in B$ (that is the closure of the vectorial space generated by the X_k , for $k \in B$).

$$H_B^X := \overline{\text{Vect}(X_k)_{k \in B}}$$

In particular, let $H_{\infty}^X := H_{\mathbb{Z}}^X$, be an Hilbert space with corresponding scalar product

$$\langle Y, Z \rangle_{H_{\infty}^X} = \mathbb{E}_{\mathbb{P}_X} [YZ].$$

Define also

$$H_{-\infty}^X := \bigcap_{n \in \mathbb{Z}} H_n^X$$

This is the asymptotic past of X . The process is said to be regular if $H_{-\infty}^X = \{0\}$, which is ensured as soon as $\log(f^*)$ is integrable, as pointed out in [8].

In this paper, we make the following classical assumption, that provide the integrability of $\log(f^*)$

Assumption 1.1.

$$\exists m, m' > 0, \forall t \in \mathbb{T}, m < f^*(t) < m'$$

This is the strongest assumption of the paper. It enables to invert the operators $T_\infty(f^*)$ and $T_N(f^*)$, $\forall N > 0$, as shown by the following proposition. Assumption 1.1 provides a controls for a given spectral density f^* on the spectrum of the corresponding Toeplitz operator $T_\infty(f^*)$ (i.e. the set $\{\lambda \in \mathbb{C}, T_\infty(f^*) - \lambda \text{Id} \text{ is not invertible}\}$ denoted $\text{Sp}(T_N(f^*))$). Note that the Toeplitz operator $T_\infty(f^*)$ is viewed here as an $l_2(\mathbb{Z})$ operator and not a $l_2(\mathbb{N})$ one.

Proposition 1.2. [16] *Under assumption 1.1,*

$$\forall N \in \mathbb{Z}^+ \cup \{+\infty\}, \text{Sp}(T_N(f^*)) \subset [m, m']$$

As a consequence, since $m > 0$, we obtain that $0 \notin \text{Sp}(T_N(f^*))$.

Now, we get some propositions about several operator norms we will use in this work.

First, define the operator norm on $l_2(\mathbb{Z})$:

Definition 1.3 (Canonical norm). Let Q be a linear operator, operating on $l^2(\mathbb{Z})$, we define

$$\|Q\|_{l_2} := \sup_{x \in l_2(\mathbb{Z}), \|x\|_2=1} \|Qx\|_2,$$

where

$$(Qx)_j = \sum_{k \in \mathbb{Z}} Q_{jk} x_k$$

As a direct consequence of Proposition 1.2, we get the following bound

$$\|R\|_{l_2} \leq m',$$

Hence, it is possible to define a new scalar product on $l_2(\mathbb{Z})$ with the symmetric bounded operator $R = T_\infty(f^*)$, and the corresponding operator norm. Let us define

Definition 1.4 (Warped norm).

$$\forall x, y \in l_2(\mathbb{Z}), \langle x, y \rangle_{l_2^R} := \langle x, Ry \rangle_{l_2}$$

Let Q be a linear operator, operating on $l_2(\mathbb{Z})$,

$$\|Q\|_{l_2^R} := \sup_{x \in l_2^R(\mathbb{Z}), \|x\|_{l_2^R}=1} \|Qx\|_{l_2^R}.$$

Thanks to Assumption 1.1 on f^* , the norm $\|\cdot\|_{l_2^R}$ given by R is equivalent to the canonical norm, because the distortion due to R is bounded. The following proposition states that the operators are bounded on l_2^R if and only if they are bounded on l_2 , which provides the equivalence of both norms. Indeed for any operator Q , we have

Proposition 1.5.

$$\frac{m}{m'} \|Q\|_{l_2} \leq \|Q\|_{l_2^R} \leq \frac{m'}{m} \|Q\|_{l_2}$$

Proof.

$$\begin{aligned} \|Q\|_{l_2^R} &= \sup_{y \in l_2^R} \frac{\|Qy\|_{l_2^R}}{\|y\|_{l_2^R}} \\ &\leq \frac{m'}{m} \sup_{y \in l_2^R} \frac{\|Qy\|_{l_2}}{\|y\|_{l_2}} = \frac{m'}{m} \|Q\|_{l_2} \end{aligned}$$

and symmetrically,

$$\|Q\|_{l_2^R} \geq \frac{m}{m'} \|Q\|_{l_2}.$$

□

In particular, R is a bounded operator on $l_2(\mathbb{Z})$, in the sense of the warped operator norm.

Proposition 1.6. *There is a canonical isometry between H_∞^X and $(l_2(\mathbb{Z}), \langle \cdot, \cdot \rangle_R)$, and also $(\overline{\text{Vect}}(e^{ikt})_{k \in \mathbb{Z}}, \langle \cdot, \cdot \rangle_{\mathbb{L}_2(f^*)})$ ($\overline{\text{Vect}}(e^{ikt})$ denotes here the closure of the set of trigonometric polynomials), through the following functions :*

$$\begin{aligned} \Phi_1 &: \mathbb{L}_2(\mathbb{P}) \rightarrow l_2(\mathbb{Z}) \\ Y = \sum_{k \in \mathbb{Z}} \beta_k X_k &\mapsto (\beta_k)_{k \in \mathbb{Z}} \end{aligned}$$

and

$$\begin{aligned} \Phi_2 &: \mathbb{L}_2(\mathbb{P}) \rightarrow \mathbb{L}_2(f^*) \\ Y = \sum_{k \in \mathbb{Z}} \beta_k X_k &\mapsto (t \mapsto \sum_{k \in \mathbb{Z}} \beta_k e^{ikt}) \end{aligned}$$

Proof. We can write, for any $Y \in \mathbb{L}_2(\mathbb{P})$ such that $Y = \sum_{k \in \mathbb{Z}} \beta_k X_k$. As a result,

$$\begin{aligned} \|Y\|_{\mathbb{L}_2(\mathbb{P})}^2 &= \sum_{j, k \in \mathbb{Z}} \beta_j \beta_k \mathbb{E}_{\mathbb{P}} [X_j X_k] \\ &= \sum_{j, k \in \mathbb{Z}} \beta_j \beta_k r(|j - k|) \\ &= \|\Phi_1(Y)\|_{l_2^R}^2 \\ &= \int_{\mathbb{T}} |\Phi_2(Y)(t)|^2 f^*(t) dt \\ &= \|\Phi_2(Y)\|_{\mathbb{L}_2(f^*)}^2, \end{aligned}$$

□

All the operators Q used in this paper are defined from H_∞^X into H_∞^X . Thanks to the isometries, we can also see them as operators respectively on $(\overline{\text{Vect}}(e^{ikt})_{k \in \mathbb{Z}}, \langle \cdot, \cdot \rangle_{\mathbb{L}_2(f^\star)})$ or $(l_2(\mathbb{Z}), \langle \cdot, \cdot \rangle_R)$, through $\Phi_k Q \Phi_k^{-1}$.

Finally, we prove a useful proposition which characterizes the inverse of a Toeplitz operator as a Toeplitz operator associated to the inverse.

Proposition 1.7. *Under the assumption 1.1, the Toeplitz operator on $l_2(\mathbb{Z})$, $T_\infty(f^\star)$ is invertible. Furthermore, we have*

$$(T_\infty(f^\star))^{-1} = T_\infty\left(\frac{1}{f^\star}\right)$$

Proof. First let us prove that

$$T_\infty(a)T_\infty(b) = T_\infty(ab)$$

Indeed, the Fourier development of ab may be written as

$$(ab)(t) = \sum_k \left(\sum_j a_j b_{k-j} \right) e^{ikt},$$

so that the equality follows. Then, notice that $T_\infty(1) = I_\infty$, where 1 denotes the constant function with value 1, and I_∞ the identity on the Hilbert space $l_2(\mathbb{Z})$. \square

In the following, we will denote $P = R^{-1} = T_\infty\left(\frac{1}{f^\star}\right)$.

2 Time series prediction with finite past observations

We wish to predict the future values X_k , for $k \geq 0$ of the process X while observing a finite number of past values X_{-N}, \dots, X_{-1} . Since we consider a Gaussian process, the best linear predictor, defined as the projection of X_k onto its past is also the best predictor. This projector can be written using the following notations. Define

$$\begin{aligned} z_N^k(i) &= r(|i - k|), \quad \forall i = -1 \dots -N, \\ (T_\infty^-(f^\star))_{ij} &= r(|i - j|), \quad \forall -\infty \leq i, j \leq -1 \end{aligned}$$

Hence the projection of X_k onto the complete past $H_{\mathbb{Z}^-}^X$, denoted by $p_\infty := \text{Proj}_{H_{\mathbb{Z}^-}^X}$, can be written as

$$p_\infty(X_k) = (T_\infty^-(f^\star))^{-1} z_\infty^k$$

Here, we are facing a more difficult issue since the whole past is not observed but only a finite number of observations. Hence our aim is twofold

- Compute the projector onto $H_{[-N, -1]}^X$, the linear hull of X_{-N}, \dots, X_{-1} . This solution can be written as $(T_N(f^\star))^{-1} z_N^k$.
- Since its expression depends on the spectral density which is unknown, we will use an estimate of f^\star to find an estimator of the projection operator.

The efficiency of the estimation procedure will be assessed using the quadratic operator norm between the estimated projector with finite past observations and the true projector onto the complete past of the process. More precisely, if p_∞ is the projection operator from $H_{\mathbb{Z}^+}^X$ onto $H_{\mathbb{Z}^-}^X$, and \hat{p} denotes the current estimate of the projector onto the infinite past built using the N observations X_{-N}, \dots, X_{-1} , we will compute

$$\|\hat{p} - p_\infty\| = \sup_{Y \in H_{[-N, N]}^X, \|Y\|_{H_{[-N, N]}^X} = 1} \sqrt{\mathbb{E} [(\hat{p}(Y) - p_\infty(Y))^2]} \quad (1)$$

Usually, the forecast is investigated in term of prediction error. If \hat{X}_k denotes the prediction of X_k , the natural error should be

$$\text{Prediction error} = \sqrt{\mathbb{E}(\hat{X}_k - X_k)^2}$$

Actually, this does not enable to study the consistency nor the asymptotic efficiency of the estimator, since this error term is macroscopic, that is of same order than the innovation. That is the reason why, to avoid this problem, we consider the loss function (1), computing the quadratic error between the prediction knowing the covariance and the one with the estimated covariance.

2.1 Projection onto finite observations with known covariance

We aim at providing an expression of the projector of an observation onto a finite set of the past of a time series. For this, we generalize the expression provided by Bondon ([6], [7]) of a projector with infinite past. For any linear operator Q from H_∞^X to H_∞^X , we denote by Q_{CB} the operator define on H_∞^X by truncating Q at right on H_B^X and at left on H_C^X .

The following theorem provides an alternative expression of any projection operators.

Theorem 2.1. *Let $A \subset \mathbb{Z}$. Recall that R is the covariance operator of the process, and $P = R^{-1}$. Then, writing the operators blockwise with blocks $A, M = A^C$, we have*

$$\begin{aligned} Proj_{H_A^X} &= \begin{bmatrix} Id_A & R_A^{-1} R_{AM} \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} Id_A & -P_{AM} P_M^{-1} \\ 0 & 0 \end{bmatrix} \end{aligned}$$

Furthermore, the prediction error $\mathbb{E} \left[Proj_{H_A^X}(Y) - Y \right]^2 = a^T P_M^{-1} a$ where $Y = a^T X$ for any $a \in \mathbb{R}^{\mathbb{Z}}$. Hence the quadratic error is given by the quadratic form P_M^{-1} .

We provide in the appendix the proof of this theorem.

We point out that this theorem is helpful for the computation of the bias, since it replaces an inversion of an infinite operator by an inversion of a finite matrix.

2.2 Construction of the empirical projection operator

Previous theorem provides a general expression of the projection which depends on the unknown spectral density of the process. Hence, we need to estimate it to build an estimator of the projection operator p_∞ .

Recall that the expression of the empirical unbiased covariance estimator is given by (see for example [2])

$$\forall 0 < p \leq N, \hat{r}_N(p) = \frac{1}{N-p} \sum_{k=-N}^{-p-1} X_k X_{k+p}$$

Notice when p is close to N , the estimation is hampered since we only sum $N-p$ terms. Hence we will not use the complete available data but rather use a cut-off level. So consider a sequence $K(N)$, growing to infinity with N that will be specified later and define the empirical spectral density as

$$\hat{f}_K^N(t) = \sum_{p=-K(N)}^{K(N)} \hat{r}_N(p) e^{ipt} \quad (2)$$

To approximate the projection onto the observed variables $X_{-1}, \dots, X_{-K(N)}$, we will consider an estimator $\hat{p}_{K(N)}^N$, defined for any random variable in $H_{[1, K(N)]}^X$, such that, when $K(N)$ is large, this estimate converges to p_∞ the projector of $H_{\mathbb{Z}^+}^X$ onto $H_{\mathbb{Z}^-}^X$. Moreover, we want to use a single sample to build the estimator and forecast the future at the same time.

To this aim, we divide the index space \mathbb{Z} into $M_K \cup O_K \cup B_K \cup F_K$ where :

- $M_K = \{\dots, -K-2, -K-1\}$ denotes the index of the past data that will not be used for the prediction (missing data)
- $O_K = -K, \dots, -1$ the index of the data used for the prediction (observed data)
- $B_K = 0, \dots, K-1$ the index of the data we currently want to project (blind data)
- $F_K = K, K+1, \dots$ the remaining index (future data)

In this framework, Theorem 2.1 shows that the projection operator of $H_{B_K}^X$ onto $H_{O_K}^X$ has the following expression

$$p_K = p_{O_K} \Big|_{B_K} = (R_{O_K})^{-1} R_{O_K B_K}$$

Hence, the two quantities $R_{O_K B_K}$ and $(R_{O_K})^{-1}$ must be estimated. On the one hand, a natural estimator of the first projector is given by

$$\begin{aligned} \hat{R}_{O_K}^N &= T_\infty(\hat{f}_K^N) \\ \text{i.e. } \left(\hat{R}_{O_K}^N \right)_{ij} &= \mathbb{1}_{|j-i| \leq K} \hat{r}_N(|j-i|) \end{aligned}$$

On the other hand, a natural estimator of $(R_{O_K})^{-1}$ could be first to estimate the projector by $(\hat{R}_{O_K}^N)$ and then to invert it. However, it is not obvious that this matrix is invertible. So, we will consider an empirical regularized version by setting

$$\tilde{R}^N = \hat{R}_{O_K}^N + \hat{\alpha} I_{O_K},$$

for a well chosen $\hat{\alpha}$. Set

$$\hat{\alpha} = -\min \hat{f}_K^N \mathbb{1}_{\min \hat{f}_K^N \leq 0} + \frac{m}{4} \mathbb{1}_{\min \hat{f}_K^N \leq \frac{m}{4}}$$

so that $\left\| \left(\tilde{R}_{O_K}^N \right)^{-1} \right\|_{l_2} \leq \frac{m}{4}$. Remark that \tilde{R}^N is the Toeplitz matrix associated to the function $\tilde{f}^N = \hat{f}_K^N + \hat{\alpha}$, which was taylored to ensure that, \tilde{f}^N is always greater than $\frac{m}{4}$, yielding the desired control on \tilde{R}^{-1} . Other regularization schemes could have been investigated. Nevertheless note that adding a translation factor makes computation easier than if we have used for instance a threshold on \hat{f}_K^N since we only modify the diagonal coefficients of the covariance matrix in this case.

Finally, we will consider the following estimator

Definition 2.2. The estimator $\hat{p}_{K(N)}^N$ of p_∞ at window $K(N)$ is defined as follow :

$$\hat{p}_{K(N)}^N = \left(\tilde{R}_{O_K}^N \right)^{-1} \hat{R}_{O_K, B_K}^N, \quad (3)$$

3 Asymptotic behaviour of the empirical projection operator

We consider Sobolev's type regularity by setting

$$\forall s > 1, W_s = \left\{ g \in \mathbb{L}_2(\mathbb{T}), g(t) = \sum_{k \in \mathbb{Z}} a_k e^{ikt}, \sum_{k \in \mathbb{Z}} k^{2s} a_k^2 < \infty \right\}$$

and define

$$\forall g \in W_s, \|g\|_{W_s} = \inf \left\{ M, \sum_{k \in \mathbb{Z}} k^{2s} a_k^2 \leq M \right\}$$

Assumption 3.1. There exists $s \geq 1$ such that $f^* \in W_s$.

Recall that the rate of convergence is given with respect to the operator norm defined in (1). The following proposition describes the behaviour of the bias and variance of the estimated projector.

Proposition 3.2. *Under assumptions 1.1 and 2.1, for N large enough, the empirical estimator with window $K(N)$ satisfies*

$$\left\| \hat{p}_{K(N)}^N - p_\infty \right\| \leq C_1 \frac{K(N)^2 \sqrt{\log(K(N))}}{\sqrt{N}} + C_2 \frac{1}{K(N)^{\frac{2s-1}{2}}},$$

where C_1 and C_2 are given in annex.

Obviously, the best rate of convergence is obtained by balancing the variance and the bias and finding the best window $K(N)$. Indeed, the variance increases as $K(N)$ is larger while the bias decreases with the window $K(N)$. Proposition 3.2 leads immediately to the following theorem.

Theorem 3.3 (Rate of convergence of the prediction estimator). *Under assumptions 1.1 and 2.1, for N large enough choosing $K^*(N) = \left(\frac{N}{\log(N)} \right)^{\frac{1}{2(2s+3)}}$ gives to*

$$\left\| \hat{p}_{K^*}^N - p_\infty \right\| \leq O \left(\left(\frac{\log(N)}{N} \right)^{\frac{2s-1}{2(2s+3)}} \right) \quad (4)$$

Notice that the more regular is the function (i.e when s increases), the faster is the estimation rate. In particular when $s \rightarrow \infty$, we obtain $(\frac{\log(N)}{N})^{\frac{1}{2}}$, which is, up to the log-term, the optimal speed. As a matter of fact, in this case, estimating the first coefficients of the covariance matrix is enough, hence the bias is very small. Proving a lower bound on the mean error, that could lead to a minimax result is a difficult task since the tools used to design the estimator are far from the usual estimation technics.

However, there are two main advantages to use our methodology. First, our blind prediction scheme provides a fully numerically tractable predictor that is easily computable. Note also that rate may be extended to non Gaussian series for linear prediction extending the exponential inequalities of Lemma 5.5. This will be done in a forthcoming paper following the ideas developped in [5] and [12]. Another extension will be considered in [13] in the context of Gaussian processes on general graphs.

4 Numerical Simulations

In this section, we study the estimation procedure with some simulations.

First of all, consider the spectral density defined by $f^* = (\frac{x}{\pi})^5 \sin(1/x) + 2$ on $[-\pi, \pi]$. Note that this density satisfies Assumptions 1.1 and 3.1, with for instance $m = 1$, $m' = 3$, $s = 1$. Consider the Toeplitz operator associated to f^* and compute $T_{1000}(f^*)$. The simulation is obtained by the stepwise procedure

- Simulate a sample $X = (X_k)$ for $k = -999, \dots, 0$.
- Then, for N growing from 4 to 200, 4 by 4, compute the empirical covariance $R_{K(N)}$ observing X_k for $k = -N, \dots, -1$. Here we have chosen the best $K(N) = N^{-\frac{1}{3}}$ ranging from $N^{-\frac{1}{3}}$ to the theoretical one $N^{-\frac{1}{10}}$ using cross validation.
- Compute the quadratic error between this empirical prediction of X_0 , and the long past prediction using the true covariance and the whole data set $X = (X_k)_{k \in [-999, -1]}$.

We observe 500 iterations of previous algorithm and give a plot in Figure 4 of the estimation error. The x axis concerns the time, the y axis shows the different iterations and the error is given by the third axis. The different realizations are sorted by the decreasing order for the last values. Then, we compute in Figure 4 the empirical mean and variance of the observations (the mean is given by the dash-dot line and the variance by the continuous one). As expected, we observe that the estimation error decreases but we also observe the decrease of the variance of this error. Hence, this implies that the prediction is more and more reliable while n is growing.

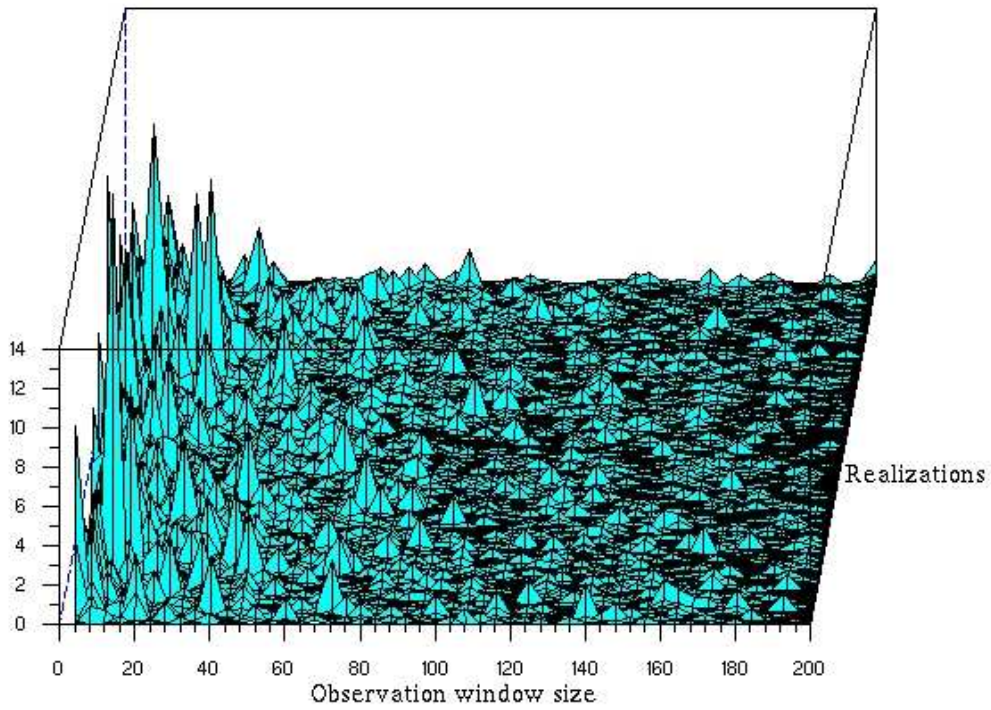


Figure 1: Prediction error evolution for X_0 on 500 paths

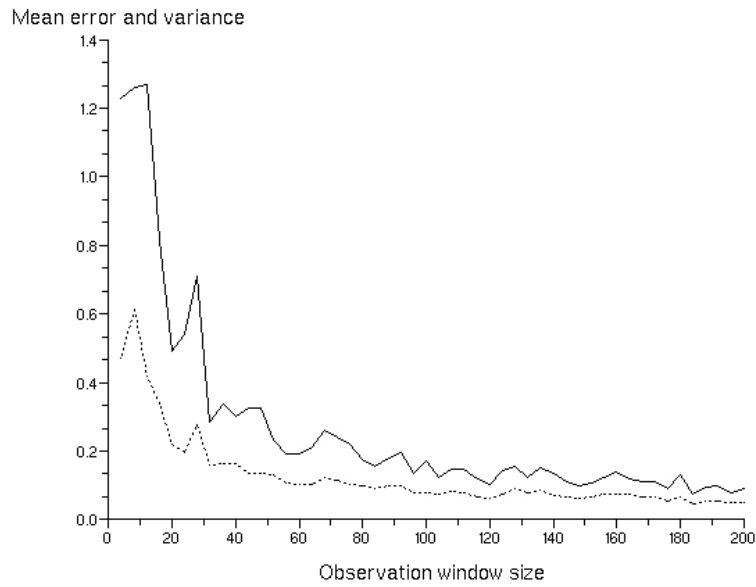


Figure 2: Prediction error for X_0 . Mean and variance on 500 paths

N	12	20	80	140	200
Mean	0.42	0.22	0.097	0.072	0.050
Variance	1.3	0.50	0.17	0.13	0.092

Table 1: Some values of the prediction error for X_0 . Mean and variance on 500 paths

5 Appendix

5.1 Proof of Theorem 2.1

Proof. First of all, $P = R^{-1}$ is a Toeplitz operator from H_∞^X to H_∞^X with eigenvalues in $[\frac{1}{m'}, \frac{1}{m}]$ and P_M may be inverted as a principal minor of P . Let us define the Schur complement of P on H_M^X : $S = P_A - P_{AM}P_M^{-1}P_{MA}$. The next lemma provides an expression of R_A^{-1} , with a inversion blockwise.

Lemma 5.1.

$$\begin{aligned} R_A^{-1} &= S \\ &= P_A - P_{AM}P_M^{-1}P_{MA} \end{aligned}$$

Proof. One can check that, written blockwise,

$$\begin{aligned} &\begin{bmatrix} P_A & P_{AM} \\ P_{MA} & P_M \end{bmatrix} \begin{bmatrix} S^{-1} & -S^{-1}P_{AM}P_M^{-1} \\ -P_M^{-1}P_{MA}S^{-1} & P_M^{-1} + P_M^{-1}P_{MA}S^{-1}P_{AM}P_M^{-1} \end{bmatrix} \\ &= \begin{bmatrix} P_AS^{-1} - P_{AM}P_M^{-1}P_{MA}S^{-1} & -P_AS^{-1}P_{AM}P_M^{-1} + P_{AM}(P_M^{-1} + P_M^{-1}P_{MA}S^{-1}P_{AM}P_M^{-1}) \\ P_{MA}S^{-1} - P_M^{-1}P_{MA}S^{-1} & -P_{MA}P_M^{-1}P_{MA}S^{-1} + P_M(P_M^{-1} + P_M^{-1}P_{MA}S^{-1}P_{AM}P_M^{-1}) \end{bmatrix} \\ &= \begin{bmatrix} SS^{-1} & (P_{AM}P_M^{-1}P_{MA}S^{-1} + I_A - P_AS^{-1})P_{AM}P_M^{-1} \\ P_{MA}S^{-1} - P_M^{-1}P_{MA}S^{-1} & -P_{MA}S^{-1}P_{AM}P_M^{-1} + I_M + P_{MA}S^{-1}P_{AM}P_M^{-1} \end{bmatrix} \\ &= \begin{bmatrix} I_A & 0 \\ 0 & I_M \end{bmatrix} \end{aligned}$$

And since the matrix are symmetric, we can transpose the last equality, so we obtain that

$$\begin{aligned} &\begin{bmatrix} S^{-1} & -S^{-1}P_{AM}P_M^{-1} \\ -P_M^{-1}P_{MA}S^{-1} & P_M^{-1} + P_M^{-1}P_{MA}S^{-1}P_{AM}P_M^{-1} \end{bmatrix} = P^{-1} \\ &= R \end{aligned}$$

So that $R_A = S^{-1}$. □

We can then compute the projection operator :

$$\begin{aligned} p_A &= \begin{bmatrix} R_A^{-1} & 0 \\ 0 & 0 \end{bmatrix} R \\ &= \begin{bmatrix} Id_A & R_A^{-1}R_{AM} \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} Id_A & SR_{AM} \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} Id_A & (P_A - P_{AM}P_M^{-1}P_{MA})R_{AM} \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} Id_A & P_AR_{AM} - P_{AM}P_M^{-1}(Id_M - P_MR_M) \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} Id_A & P_AR_{AM} - P_{AM}P_M^{-1} + P_{AM}R_M \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} Id_A & -P_{AM}P_M^{-1} \\ 0 & 0 \end{bmatrix} \end{aligned}$$

where we have used $PK = Id$ in the last two lines.

Now consider the quadratic error operator ($Q : X \mapsto X^T Q X$ the quadratic error).

$$\begin{aligned} Q &:= p_{H_A}^T R p_{H_A} \\ &= P_M^{-1} \end{aligned}$$

Q can be obtained by a direct computation (writing the product right above), but it is easier to use the expression of the variance of a projector in the Gaussian case given for instance by [18].

$$Q = R_M - R_{MA} R_A^{-1} R A M$$

Again, notice that Q is the Schur complement of R on H_A^X , and thanks to Lemma 5.1 applied to P instead of R , we get

$$Q = P_M^{-1}.$$

□

5.2 Proof of Proposition 3.2

Proof. We will use the operator norm to compute the error. Notice that the operators studied here may be random. In this case we assume that we have an independent copy X' (defined on Ω' the corresponding probability space) of the process X (with associated probability space Ω). Then, $\|\cdot\|_{l_2^R}$ denotes the random norm such that

$$\forall \omega \in \Omega \quad \|A(\omega)\|_{l_2^R} = \sup_{Y \in H_\infty^X \|\cdot\|_2=1} \int_{\omega'} (A(\omega)Y)^2(\omega') d\mathbb{P}(\omega')$$

and not, as we could have expected

$$\|A\|_{l_2^R} = \sup_{Y \in H_\infty^X \|\cdot\|_2=1} \int_{\omega} (A(\omega)Y)^2(\omega) d\mathbb{P}(\omega)$$

This norm measures the performance of the prediction. We could have obtained it if we had at hand one sample for the estimation, and another independent sample dedicated to the prediction.

Then, the proof falls into two steps : we compute independently the bias and the variance :

$$\|\hat{p}_{K(N)}^N - p_\infty\| \leq \underbrace{\|\hat{p}_{K(N)}^N - p_{K(N)}\|}_{\text{Variance}} + \underbrace{\|p_{K(N)} - p_\infty\|}_{\text{Bias}}$$

We prove in the next part two next lemmas that contain the major ideas of the proof. The bias is given by the following lemma

Lemma 5.2. *The following upper bound holds, for N large enough,*

$$\|\hat{p}_{K(N)}^N - p_\infty\|_{l_2^R} \leq C_2 \frac{1}{K(N)^{\frac{2s-1}{2}}},$$

where $C_2 = \left\| \frac{1}{f^*} \right\|_{H_{2s}} m'(1 + \frac{m'}{m})$

Set $A = \hat{p}_{K(N)}^N - p_{K(N)}$, the control for the variance is due to this lemma:

Lemma 5.3.

$$\int_0^\infty \mathbb{P} \left(\|A\|_{l_2^R}^4 > t \right) dt \leq C_0^4 K(N)^4 \left(\frac{\log(K(N))}{N} \right)^2 + o(K(N)^4 \left(\frac{\log(K(N))}{N} \right)^2),$$

where $C_0 = 4m' \left(\frac{6m'}{m^2} + \frac{4}{m} + 2 \right)$

Then we write, for any window parameter $K(N)$ and centered Y in $H_{B_K}^X$ such that $\mathbb{E}[Y^2] = 1$,

$$\begin{aligned} \sqrt{\mathbb{E} \left[(p_{K(N)}Y - p_\infty Y)^2 \right]} &\leq \|p_{K(N)} - p_\infty\|_{l_2^R} \sqrt{\mathbb{E}[Y^2]} \\ &\leq C_2 \frac{1}{K(N)^{\frac{s-1}{2}}} \end{aligned}$$

And, for the variance, we notice first that

$$1 = \mathbb{E}[Y^2] = \beta' R_{M_K} \beta \geq m \beta' \beta = m \sum_{i=0}^{K(N)-1} \beta_i^2,$$

Further,

$$\begin{aligned} \mathbb{E} \left[(A(Y))^2 \right] &= \int_\omega \left(\sum_{i=0 \dots K(N)-1, j=-K(N), \dots, -1} A_{ij}(\omega) \beta_i X_j(\omega) \right)^2 d\mathbb{P}(\omega) \\ &\leq \int_\omega \sum_{j=-K(N)}^{-1} \left(\sum_{i=0}^{K(N)} A_{ij}(\omega) \beta_i \right)^2 \sum_{j=-K(N)}^{-1} X_j^2(\omega) d\mathbb{P}(\omega) \\ &\leq \int_\omega \sum_{i=0 \dots K(N)-1, j=-K(N), \dots, -1} A_{ij}^2(\omega) \sum_{i=0}^{K(N)} \beta_i^2 \sum_{j=-K(N)}^{-1} X_j^2(\omega) d\mathbb{P}(\omega), \end{aligned}$$

by applying twice Cauchy-Swartz's inequality. So that,

$$\begin{aligned} \mathbb{E} \left[(A(Y))^2 \right] &\leq \int_\omega \|A(\omega)\|_2^2 \frac{1}{m} \sum_{j=n_0+1}^{K(N)+n_0} X_j^2 d\mathbb{P}(\omega) \\ &\leq \frac{K(N)}{m} \int_\omega \|A(\omega)\|_{l_2}^2 \sum_{j=n_0+1}^{K(N)+n_0} X_j^2(\omega) d\mathbb{P}(\omega), \end{aligned}$$

Using the equivalence between this two norms for finite matrix with size (n, m) (see for instance [17]), we obtain

$$\|A\|_2 \leq \sqrt{n} \|A\|_{l_2}.$$

Further, using Proposition 1.5, we get

$$\begin{aligned}
&\leq \frac{K(N)}{m} \int_{\omega} \|A(\omega)\|_{l_2}^2 \sum_{j=n_0+1}^{K(N)+n_0} X_j^2(\omega) d\mathbb{P}(\omega) \\
&\leq \frac{K(N)}{m} \sqrt{\int_{\omega} \|A(\omega)\|_{l_2}^4 d\mathbb{P}(\omega)} \sqrt{\int_{\omega} \left(\sum_{j=n_0+1}^{K(N)+n_0} X_j^2(\omega) \right) d\mathbb{P}(\omega)} \\
&\leq \frac{K(N)}{m} \sqrt{\int_{\mathbb{R}^+} \mathbb{P}(\|A\|_{l_2}^4 > t) dt} \sqrt{K(N)^2 \int_{\omega} (X_j^4) d\mathbb{P}(\omega)},
\end{aligned}$$

We use here again Cauchy-Schwartz's inequality and the fact that, for all nonnegative random variable Y ,

$$\mathbb{E}[Y] = \int_{\mathbb{R}^+} \mathbb{P}(Y > t) dt$$

Since X_0 is Gaussian, its four order moment is finite, say r_4 . Then Lemma 5.3 yields that, for N large enough,

$$\mathbb{E} \left[\left(A(Y) \right)^2 \right] \leq \frac{C_0^2 \sqrt{r_4} K(N)^4 \log(K(N))}{mN},$$

and so,

$$\|\hat{p}_{K(N)}^N - p_{K(N)}^N\| \leq \frac{C_1 K(N)^2 \sqrt{\log(K(N))}}{\sqrt{N}},$$

with $C_1 = \frac{C_0 \sqrt[4]{r_4}}{\sqrt{m}}$, which proves the theorem. \square

5.3 Proofs of Concentration and regularity lemmas

First we compute the bias and prove Lemma 5.2 :

Proof. of Lemma 5.2

Using Theorem 2.1, we can write

$$\text{Proj}_{H_{\mathbb{Z}^-}}|_{B_K} = p_{\infty}|_{\mathbb{Z}_+} = -P_{\mathbb{Z}^- \mathbb{Z}_+} (P_{\mathbb{Z}_+})^{-1}$$

But $p.|_{B_K} = p.|_{\mathbb{Z}_+}|_{B_K}$ So,

$$\left\| p_{K(N)}|_{B_K} - p_{\infty}|_{\mathbb{Z}_+} \right\|_{l_2^R} \leq \left\| p_{K(N)}|_{\mathbb{Z}_+} - p_{\infty}|_{\mathbb{Z}_+} \right\|_{l_2^R}$$

We can now compute

$$\begin{aligned}
\left\| p_{K(N)}|_{\mathbb{Z}_+} - p_\infty|_{\mathbb{Z}_+} \right\|_{l_2^R} &\leq \frac{m'}{m} \left\| p_{K(N)}|_{\mathbb{Z}_+} - p_\infty|_{\mathbb{Z}_+} \right\|_{l_2} \\
&\leq \frac{m'}{m} \left\| \begin{bmatrix} (R_{O_K})^{-1} R_{O_K \mathbb{Z}_+} \\ 0 \end{bmatrix} - \begin{bmatrix} -P_{O_K \mathbb{Z}_+} (P_{\mathbb{Z}_+})^{-1} \\ -P_{M_K^- \mathbb{Z}_+} (P_{\mathbb{Z}_+})^{-1} \end{bmatrix} \right\|_{l_2} \\
&\leq \frac{m'}{m} \left\| \begin{bmatrix} (R_{O_K})^{-1} R_{O_K \mathbb{Z}_+} + P_{O_K \mathbb{Z}_+} (P_{\mathbb{Z}_+})^{-1} \\ P_{M_K^- \mathbb{Z}_+} (P_{\mathbb{Z}_+})^{-1} \end{bmatrix} \right\|_{l_2} \\
&\leq \frac{m'}{m} \left\| \begin{bmatrix} (R_{O_K})^{-1} R_{O_K \mathbb{Z}_+} P_{\mathbb{Z}_+} + P_{O_K \mathbb{Z}_+} \\ P_{M_K^- \mathbb{Z}_+} \end{bmatrix} \right\|_{l_2} \left\| (P_{\mathbb{Z}_+})^{-1} \right\|_{l_2} \\
&\leq \frac{m'}{m} \left\| (P_{\mathbb{Z}_+})^{-1} \right\|_{l_2} \left(\left\| (R_{O_K})^{-1} R_{O_K \mathbb{Z}_+} P_{\mathbb{Z}_+} + P_{O_K \mathbb{Z}_+} \right\|_{l_2} + \left\| P_{M_K^- \mathbb{Z}_+} \right\|_{l_2} \right)
\end{aligned}$$

The last step used only the inequality

$$\left\| \begin{bmatrix} A \\ B \end{bmatrix} \right\|_{l_2} \leq \left\| \begin{bmatrix} A \\ 0 \end{bmatrix} \right\|_{l_2} + \left\| \begin{bmatrix} 0 \\ B \end{bmatrix} \right\|_{l_2} = \|A\|_{l_2} + \|B\|_{l_2}$$

But, since $P = R^{-1}$,

$$R_{O_K \mathbb{Z}_+} P_{\mathbb{Z}_+} + R_{O_K} P_{O_K \mathbb{Z}_+} = -R_{O_K M_K^-} P_{M_K^- \mathbb{Z}_+}$$

So, we obtain,

$$\begin{aligned}
\left\| p_{K(N)}|_{\mathbb{Z}_+} - p_\infty|_{\mathbb{Z}_+} \right\|_{l_2^R} &\leq \frac{m'}{m} \left\| (P_{\mathbb{Z}_+})^{-1} \right\|_{l_2} \left(\left\| (R_{O_K})^{-1} \left(-R_{O_K M_K^-} P_{M_K^- \mathbb{Z}_+} \right) \right\|_{l_2} + \left\| P_{M_K^- \mathbb{Z}_+} \right\|_{l_2} \right) \\
&\leq \frac{m'}{m} \left\| (P_{\mathbb{Z}_+})^{-1} \right\|_{l_2} \left(\left\| (R_{O_K})^{-1} \right\|_{l_2} \left\| -R_{O_K M_K^-} P_{M_K^- \mathbb{Z}_+} \right\|_{l_2} + \left\| P_{M_K^- \mathbb{Z}_+} \right\|_{l_2} \right) \\
&\leq \frac{m'}{m} \left\| (P_{\mathbb{Z}_+})^{-1} \right\|_{l_2} \left(\left\| (R_{O_K})^{-1} \right\|_{l_2} \left\| R_{O_K M_K^-} \right\|_{l_2} \left\| P_{M_K^- \mathbb{Z}_+} \right\|_{l_2} + \left\| P_{M_K^- \mathbb{Z}_+} \right\|_{l_2} \right) \\
&\leq \frac{m'}{m} \left\| (P_{\mathbb{Z}_+})^{-1} \right\|_{l_2} \left(\left\| (R_{O_K})^{-1} \right\|_{l_2} \left\| R_{O_K M_K^-} \right\|_{l_2} + 1 \right) \left\| P_{M_K^- \mathbb{Z}_+} \right\|_{l_2}
\end{aligned}$$

But, we have,

$$\left\| (P_{\mathbb{Z}_+})^{-1} \right\|_{l_2} \leq m',$$

as the inverse of a principal minor of P .

$$\left\| (R_{O_K})^{-1} \right\|_{l_2} \leq \frac{1}{m},$$

since it is the inverse of a principal minor of R .

$$\left\| R_{O_K M_K^-} \right\|_{l_2} \leq m'$$

as an extracted operator of R .

Thus, we get

$$\left\| p_{K(N)}|_{\mathbb{Z}_+} - p_\infty|_{\mathbb{Z}_+} \right\|_{l_2^R} \leq C_4 \left\| P_{M_K^- \mathbb{Z}_+} \right\|_{l_2},$$

where $C_4 = \frac{m'^2}{m} (1 + \frac{m'}{m})$. Since $f^* \in W_s$ (Assumption 2.1), and $f^* \geq m > 0$, we have also $\frac{1}{f^*} \in W_s$. If we denote $p(k) = P_{i,i+k}$ the Fourier coefficient of $\frac{1}{f^*}$, we get

$$\begin{aligned}
\| \| P_{M_K^- \mathbb{Z}_+} \| \|_{l_2} &\leq \| P_{M_K^- \mathbb{Z}_+} \|_2 \\
&\leq \sqrt{\sum_{i \leq -K(N); 0 \leq j} p(j-i)^2} \\
&\leq \sqrt{\sum_{i=K(N)}^{\infty} \sum_{j=i}^{\infty} p(j)^2} \\
&\leq \sqrt{\sum_{i=K(N)}^{\infty} \frac{\| \frac{1}{f^*} \|_{W_s}^2}{i^s}} \\
&\leq \sqrt{\| \frac{1}{f^*} \|_{W_s} \frac{1}{K(N)^{s-1}}}
\end{aligned}$$

□

So that the lemma is proved and the bias is given by

$$\| \| p_{K(N)}|_{B_K} - p_{\infty}|_{z_+} \| \|_{l_2^R} \leq C_4 \sqrt{\| \frac{1}{f^*} \|_{W_s} \frac{1}{K(N)^{\frac{2s-1}{2}}}}.$$

Actually, the rate of convergence for the bias is given by the regularity of the spectral density, since it depends on the coefficients far away from the principal diagonal.

Now, we prove Lemma 5.3, which achieves the proof of the theorem.

Proof. of Lemma 5.3

At first,

$$\begin{aligned}
\| A \|_{l_2} &= \| \| (\tilde{R}_{O_K})^{-1} \hat{R}_{O_K B_K} - (R_{O_K})^{-1} R_{O_K B_K} \| \|_{l_2} \\
&\leq \| R_{O_K B_K} \|_{l_2} \| \| (\tilde{R}_{O_K})^{-1} - (R_{O_K})^{-1} \| \|_{l_2} + \| \| (\tilde{R}_{O_K})^{-1} \| \|_{l_2} \| \| \hat{R}_{O_K B_K} - R_{O_K B_K} \| \|_{l_2} \\
&\leq \| R_{O_K B_K} \|_{l_2} \| \| (\tilde{R}_{O_K})^{-1} \| \|_{l_2} \| \| (R_{O_K})^{-1} \| \|_{l_2} \| \| \tilde{R}_{O_K} - R_{O_K} \| \|_{l_2} \\
&\quad + \| \| (\tilde{R}_{O_K})^{-1} \| \|_{l_2} \| \| \hat{R}_{O_K B_K} - R_{O_K B_K} \| \|_{l_2}
\end{aligned}$$

But, we have,

$$\| R_{O_K B_K} \|_{l_2} \leq m',$$

as an extracted operator of R .

$$\| \| (R_{O_K})^{-1} \| \|_{l_2} \leq \frac{1}{m},$$

as the inverse of a principal minor of R .

$$\| \| (\tilde{R}_{O_K})^{-1} \| \|_{l_2} \leq \frac{4}{m},$$

thanks to the regularization.

$$\left\| \tilde{R}_{O_K} - R_{O_K} \right\|_{l_2} \leq K(N) \sup_{p \leq 2K(N)} \{|\hat{r}_N(p) - r(p)|\} + |\hat{\alpha}|,$$

$$\left\| \hat{R}_{O_K B_K} - R_{O_K B_K} \right\|_{l_2} \leq K(N) \left(\sup_{p \leq 2K(N)} \{|\hat{r}_N(p) - r(p)|\} \right)$$

thanks to the regularization.

$$|\hat{\alpha}| = \left| -\min \hat{f}_K^N \mathbb{1}_{\min \hat{f}_K^N \leq 0} + \frac{m}{4} \mathbb{1}_{\min \hat{f}_K^N \leq \frac{m}{4}} \right|$$

So,

$$|\hat{\alpha}| \leq (2K(N) + 1) \sup_{p \leq 2K(N)} \{\hat{r}_N(p) - r(p)\} + \frac{m}{4} \mathbb{1}_{\min \hat{f}_K^N \leq \frac{m}{4}}$$

For the last inequality, we used the following lemma, proved in the next section

Lemma 5.4. *The empirical spectral density is such that, for N large enough*

$$\left\| \hat{f}_{K(N)}^N - f^* \right\|_{\infty} \leq (2K(N) + 1) \sup_{p \leq 2K(N)} \{\hat{r}_N(p) - r(p)\} + \frac{m}{4},$$

which implies

$$\left| \min \hat{f}_K^N \mathbb{1}_{\min \hat{f}_K^N \leq 0} \right| \leq (2K(N) + 1) \sup_{p \leq 2K(N)} \{\hat{r}_N(p) - r(p)\}$$

So, we obtain,

$$\begin{aligned} \|A\|_{l_2} &\leq \frac{4m'}{m^2} \left(K(N) \sup_{p \leq 2K(N)} \{|\hat{r}_N(p) - r(p)|\} + |\hat{\alpha}| \right) + \frac{4}{m} K(N) \left(\sup_{p \leq 2K(N)} \{|\hat{r}_N(p) - r(p)|\} \right) \\ &\leq \left(\frac{6m'}{m^2} + \frac{4}{m} + 2 + \frac{1}{K(N)} \right) K(N) \left(\sup_{p \leq 2K(N)} \{|\hat{r}_N(p) - r(p)|\} \right) + \frac{m'}{m} \mathbb{1}_{\min \hat{f}_K^N \leq \frac{m}{4}} \end{aligned}$$

We will use here some other technical lemmas whose proofs are also given in the last section. The first one gives an uniform concentration result on the estimator $\hat{r}_N(p)$:

Lemma 5.5. *For all window $K(N)$ such that $K(N) \rightarrow \infty$ and $K(N) \log(K(N)) = o(N)$, there exists N_0 such that, for all $N \geq N_0$, and $x \geq 0$,*

$$\forall p \leq 2K(N), |\hat{r}_N(p) - r(p)| > 4m' \left(\sqrt{\frac{(\log(K(N)) + x)}{N}} + \frac{x}{N} \right),$$

with probability at least $1 - e^{-x}$

For ease of notations, we set $C_0 = 4m' \left(\frac{6m'}{m^2} + \frac{4}{m} + 2 \right)$ and $C_3 = \frac{m'}{m}$. For the computation of the mean, the interval $[0, +\infty]$ will be divided into three parts, where only the first contribution is significant, thanks to the exponential concentration. We prove by computation, that the two other parts are negligible.

We just obtained, for all $x \geq 0$

$$\|A\|_{l_2^R} \leq (C_0 + o(1)) K(N) \left(\sqrt{\frac{\log(K(N)) + x}{N}} + \frac{x}{N} \right) + C_3 \mathbb{1}_{\min \hat{f}_K^N \leq \frac{m}{4}},$$

with probability at least $1 - e^{-x}$

$$\text{Set } t_1 = \left(C_0 K(N) \sqrt{\frac{\log(K(N))}{N}} \right)^4$$

For $t \in [0, t_1]$, we use the inequality

$$\mathbb{P} \left(\|A\|_{l_2^R}^4 > t \right) \leq 1$$

We obtain the first contribution to the integral, which is also the non negligible part.

$$\int_0^{t_1} \mathbb{P} \left(\|A\|_{l_2^R}^4 > t \right) dt = \left(C_0 K(N) \sqrt{\frac{\log(K(N))}{N}} \right)^4$$

$$\text{Now, set } t_2 = \left(C_0 K(N) \sqrt{\frac{\log(K(N)) + N}{N}} + C_3 \right)^4$$

For $t \in [t_1, t_2]$, we use

$$\mathbb{P} \left(\|A\|_{l_2^R}^4 > \sup \left(C_0^4 K(N)^4 \left(\frac{\log(K(N)) + x}{N} \right)^2, C_0^4 K(N)^4 \left(\frac{x}{N} \right)^4 \right) \right) \leq e^{-x} + \mathbb{P} \left(\min \hat{f}_K^N \leq \frac{m}{4} \right)$$

But, notice that the last lemma provides

$$\mathbb{P} \left(2K(N) \sup_{p \leq 2K(N)} \{|\hat{r}_N(p) - r(p)|\} > \frac{m}{2} \right) \leq e^{-\frac{Nm^2}{(64K(N)m')^2}}$$

Indeed, set $x_0(N) = \frac{Nm^2}{(64K(N)m')^2}$.

One can compute that with probability at least $1 - e^{-x_0(N)}$,

$$\begin{aligned} \sup_{p \leq 2K(N)} \{|\hat{r}_N(p) - r(p)|\} &\leq 4m' \left(\sqrt{\frac{\log(K(N)) + x_0(N)}{N}} + \frac{x_0(N)}{N} \right) \\ &\leq 4m' \left(\sqrt{\frac{\log(K(N))}{N}} + \frac{m^2}{(64K(N)m')^2} + \frac{m^2}{(64K(N)m')^2} \right) \\ &\leq 4m' \left(\sqrt{\frac{\log(K(N))}{N}} + \sqrt{\frac{m^2}{(64K(N)m')^2}} + \frac{m^2}{(64K(N)m')^2} \right) \\ &\leq 4m' \left(\sqrt{\frac{\log(K(N))}{N}} + \frac{m}{(64K(N)m')} + \frac{m^2}{(64K(N)m')^2} \right) \\ &\leq \frac{m}{8K(N)}, \end{aligned}$$

for N large enough. Hence,

$$\mathbb{P} \left(\min \hat{f}_K^N \leq \frac{m}{4} \right) \leq e^{-\frac{Nm^2}{(64K(N)m')^2}}$$

So, we have

$$\mathbb{P} \left(\|A\|_{l_2}^4 > \max \left(C_0^4 K(N)^4 \left(\frac{\log(K(N)) + x}{N} \right)^2, C_0^4 K(N)^4 \left(\frac{x}{N} \right)^4 \right) \right) \leq e^{-x} + e^{-\frac{Nm^2}{(64K(N)m')^2}}$$

Then, the following lemma will be useful for changing a probability inequality into an \mathbb{L}^2 inequality.

Lemma 5.6. *Let X be a nonnegative random variable such that there exists two one to one maps f_1 and f_2 and a $C > 0$ with*

$$\forall x \geq 0 \mathbb{P}(X > \sup(f_1(x), f_2(x))) \leq e^{-x} + C,$$

then

$$\mathbb{P}(X > t) \leq e^{-f_1^{-1}(t)} + e^{-f_2^{-1}(t)} + C$$

So, thanks to lemma 5.6, we have

$$\mathbb{P}\left(\|A\|_{l_2^R}^4 > t\right) \leq e^{-N\sqrt{\frac{t}{C_0^4 K(N)^4} + \log(K(N))}} + e^{-N^4\sqrt{\frac{Nt}{C_0^4 K(N)^4}}} + e^{-\frac{Nm^2}{(64K(N)m')^2}}$$

Now, we will prove that each term can be neglected. Integrating by part, we obtain

$$\begin{aligned} \int_{t_1}^{t_2} e^{-\sqrt{\frac{t}{C_0^4 K(N)^4} + \log(K(N))}} dt &\leq \int_{t_1}^{\infty} e^{-N\sqrt{\frac{t}{C_0^4 K(N)^4} + \log(K(N))}} dt \\ &\leq \left[\frac{-2\sqrt{t}C_0^2 K(N)^2}{N} e^{-N\sqrt{\frac{t}{C_0^4 K(N)^4} + \log(K(N))}} \right]_{t_1}^{\infty} \\ &\quad + \int_{t_1}^{\infty} \frac{C_0^2 K(N)^2}{N\sqrt{t}} e^{-N\sqrt{\frac{t}{C_0^4 K(N)^4} + \log(K(N))}} dt \\ &\leq \frac{2\log(K(N))C_0^4 K(N)^4}{N^2} + \frac{2C_0^4 K(N)^4}{N^2} \\ &= o\left(\left(C_0 K(N)\sqrt{\frac{\log(K(N))}{N}}\right)^4\right) \end{aligned}$$

Then,

$$\begin{aligned} \int_{t_1}^{t_2} e^{-4\sqrt{\frac{Nt}{C_0^4 K(N)^4}}} dt &\leq t_2 e^{-N^4\sqrt{\frac{t_1}{C_0^4 K(N)^4}}} \\ &\leq t_2 e^{-\sqrt{N\log(K(N))}} \\ &= o\left(\left(C_0 K(N)\sqrt{\frac{\log(K(N))}{N}}\right)^4\right) \end{aligned}$$

So that,

$$\begin{aligned} \int_{t_1}^{t_2} e^{-x_0(N)} dt &\leq t_2 e^{-\frac{Nm^2}{(64K(N)m')^2}} \\ &= o\left(\left(C_0 K(N)\sqrt{\frac{\log(K(N))}{N}}\right)^4\right). \end{aligned}$$

Leading to

$$\int_{t_1}^{t_2} \mathbb{P}\left(\|A\|_{l_2}^4 > t\right) dt = o\left(\left(C_0 K(N)\sqrt{\frac{\log(K(N))}{N}}\right)^4\right),$$

Finally, for $t \in [t_2, +\infty]$, we use

$$\mathbb{P} \left(\|A\|_{l_2}^4 > \max \left(\left(C_0 K(N) \sqrt{\frac{\log(K(N)) + x}{N}} + C_3 \right)^4, \left(C_0 K(N) \frac{x}{N} + C_3 \right)^4 \right) \right) \leq e^{-x}$$

Thanks to lemma 5.6, we get

$$\mathbb{P} \left(\|A\|_{l_2}^4 > t \right) \leq e^{-N \left(\frac{\sqrt[4]{t} - C_3}{C_0 K(N)} \right)^2 + \log(K(N))} + e^{-N \frac{(\sqrt[4]{t} - C_3)}{C_0 K(N)}}$$

So, by an integrating by part, we obtain

$$\begin{aligned} \int_{t_2}^{+\infty} e^{-N \left(\frac{\sqrt[4]{t} - C_3}{C_0 K(N)} \right)^2 + \log(K(N))} dt &\leq \int_{\sqrt[4]{t_2} - C_3}^{+\infty} 4(u + C_3)^3 e^{-N \left(\frac{u}{C_0 K(N)} \right)^2 + \log(K(N))} du \\ &\leq \left[P(u, N, K(N)) e^{-N \left(\frac{u}{C_0 K(N)} \right)^2 + \log(K(N))} \right]_{\sqrt[4]{t_2} - C_3}^{+\infty} \\ &\leq P(u, N, K(N)) e^{-N} \\ &\leq o \left(\left(C_0 K(N) \sqrt{\frac{\log(K(N))}{N}} \right)^4 \right). \end{aligned}$$

Here, $P(u, N, K(N))$ is a polynomial of degree 3 in u and is rational function in N and $K(n)$.

Furthermore,

$$\begin{aligned} \int_{t_2}^{+\infty} e^{-N \frac{(\sqrt[4]{t} - C_3)}{C_0 K(N)}} dt &\leq \int_{\sqrt[4]{t_2} - C_3}^{+\infty} 4(u + C_3)^3 e^{-N \frac{u}{C_0 K(N)}} du \\ &\leq \left[P(u, N, K(N)) e^{-N \frac{u}{C_0 K(N)}} \right]_{\sqrt[4]{t_2} - C_3}^{+\infty} \\ &\leq P(u, N, K(N)) e^{-\sqrt{N(\log(K(N)) + N)}} \\ &\leq o \left(\left(C_0 K(N) \sqrt{\frac{\log(K(N))}{N}} \right)^4 \right), \end{aligned}$$

where $P(u, N, K(N))$ is polynomial of degree 3 in u and is rational function in N and $K(n)$.

We proved here

$$\int_0^{+\infty} \mathbb{P} \left(\|A\|_{l_2}^4 > t \right) \leq C_0^4 K(N)^4 \left(\frac{\log(K(N))}{N} \right)^2 + o(K(N)^4 \left(\frac{\log(K(N))}{N} \right)^2),$$

which achieve the proof. \square

5.4 Technicals lemmas

We prove now the technicals lemmas :

Proof. of Lemma 5.5

Notice that $\hat{r}^N(p) = T_N(u)$ with $u = \frac{N}{N-p} \cos(pt)$. We use the following proposition from [11]. Let X_1, \dots, X_n be a centered Gaussian stationary sequence and u a bounded function such that $T_n(u)$ is a symmetric non negative matrix. Then the following concentration inequality holds for $Z_n(u) = \frac{1}{n} (X' T_n(u) X - \mathbb{E}[X' T_n(u) X])$:

$$\mathbb{P} \left(Z_n(u) \geq 2 \|f\|_\infty (\|u\|_2 \sqrt{x} + \|u\|_\infty x) \right) \leq e^{-nx}$$

By applying this result respectively with u and $-u$ and we obtain

$$\mathbb{P} \left(\left| \hat{r}^N(p) - r(p) \right| > 2m' \frac{N}{N-p} (\sqrt{x} + x) \right) \leq 2e^{-Nx}$$

or, equivalently,

$$\left| \hat{r}^N(p) - r(p) \right| > 2m' \frac{N}{N-p} \left(\sqrt{\frac{x + \log(K(N)) + 2 \log(2)}{N}} + \frac{x + \log(K(N)) + 2 \log(2)}{N} \right),$$

with probability lower than $\frac{e^{-x}}{2K(N)}$. By taking an equivalent, we obtain that there exists N_0 such that, for all $N \geq N_0$, for all $p \leq 2K(N)$

$$\mathbb{P} \left(\left| \hat{r}^N(p) - r(p) \right| > 4m' \sqrt{\frac{x + \log(K(N))}{N}} + \frac{x}{N} \right) \leq \frac{e^{-x}}{2K(N)}$$

□

Proof. of Lemma 5.6

We set $t = \sup(f_1(x), f_2(x))$. If $t = f_1(x)$ then

$$\mathbb{P}(X > t) \leq e^{-f_1^{-1}(t)} + C \leq e^{-f_1^{-1}(t)} + e^{-f_2^{-1}(t)} + C$$

Symmetrically, if $t = f_2(x)$ we have

$$\mathbb{P}(X > t) \leq e^{-f_1^{-1}(t)} + e^{-f_2^{-1}(t)} + C$$

□

Proof. of Lemma 5.4 It is sufficient to ensure that the bias is small enough. Choose N_0 such that

$$2 \|f^*\|_{W_s} K(N)^{-s+1} \leq \frac{m}{4}$$

Then we use

$$\begin{aligned} \left\| \hat{f}_{K(N)}^N - f^* \right\|_{\infty} &\leq \sum_{p=-K(N)}^{K(N)} \left| \hat{r}^N(p) - r(p) \right| + 2 \sum_{p>K(N)} |r(p)| \\ &\leq (2K(N) + 1) \sup_{p \leq 2K(N)} \{ \hat{r}_N(p) - r(p) \} + 2 \|f^*\|_{W_s} K(N)^{-s+1} \\ &\leq (2K(N) + 1) \sup_{p \leq 2K(N)} \{ \hat{r}_N(p) - r(p) \} + \frac{m}{4} \end{aligned}$$

□

References

- [1] Anestis Antoniadis, Efstathios Paparoditis, and Theofanis Sapatinas. A functional wavelet-kernel approach for time series prediction. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(5):837–857, 2006.

- [2] Robert Azencott and Didier Dacunha-Castelle. *Series of irregular observations*. Applied Probability. A Series of the Applied Probability Trust. Springer-Verlag, New York, 1986. Forecasting and model building.
- [3] R. J. Bhansali. Parameter estimation and model selection for multistep prediction of a time series: a review. In *Asymptotics, nonparametrics, and time series*, volume 158 of *Statist. Textbooks Monogr.*, pages 201–225. Dekker, New York, 1999.
- [4] Peter J. Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1):199–227, 2008.
- [5] Jeremie Bigot, Rolando Biscay, Jean-Michel Loubes, and Lilian Muniz Alvarez. Non-parametric estimation of covariance functions by model selection, 2009.
- [6] Pascal Bondon. Prediction with incomplete past of a stationary process. *Stochastic Process. Appl.*, 98(1):67–76, 2002.
- [7] Pascal Bondon. Influence of missing values on the prediction of a stationary time series. *J. Time Ser. Anal.*, 26(4):519–525, 2005.
- [8] Albrecht Böttcher and Bernd Silbermann. *Analysis of Toeplitz operators*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, second edition, 2006. Prepared jointly with Alexei Karlovich.
- [9] Sofiane Brahim-Belhouari and Amine Bermak. Gaussian process for nonstationary time series prediction. *Comput. Statist. Data Anal.*, 47(4):705–712, 2004.
- [10] Peter J. Brockwell and Richard A. Davis. *Introduction to time series and forecasting*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2002. With 1 CD-ROM (Windows).
- [11] Fabienne Comte. Adaptive estimation of the spectrum of a stationary Gaussian sequence. *Bernoulli*, 7(2):267–298, 2001.
- [12] H. Djellout, A. Guillin, and L. Wu. Moderate deviations of empirical periodogram and non-linear functionals of moving average processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 42(4):393–416, 2006.
- [13] T. Espinasse F. Gamboa and J.M. Loubes. Gaussian processes on graphs, some statistics for the prediction.
- [14] Neil Gershenfeld. *The nature of mathematical modeling*. Cambridge University Press, Cambridge, 1999.
- [15] Arthur S. Goldberger. Best linear unbiased prediction in the generalized linear regression model. *J. Amer. Statist. Assoc.*, 57:369–375, 1962.
- [16] Ulf Grenander and Gábor Szegő. *Toeplitz forms and their applications*. Chelsea Publishing Co., New York, second edition, 1984.
- [17] George A. F. Seber. *A matrix handbook for statisticians*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2008.
- [18] Michael L. Stein. *Interpolation of spatial data*. Springer Series in Statistics. Springer-Verlag, New York, 1999. Some theory for Kriging.