

# DRIFT ESTIMATION ON NON COMPACT SUPPORT FOR DIFFUSION MODELS

F. COMTE AND V. GENON-CATALOT

**ABSTRACT.** In this paper, we study non parametric drift estimation for an ergodic and  $\beta$ -mixing diffusion process from discrete observation of the sample path. The drift is estimated on a set  $A$  using an approximate regression equation by a least squares contrast which is minimized over a finite dimensional subspace  $S_m$  of  $\mathbb{L}^2(A, dx)$ . This yields a collection of estimators indexed by the dimension of the projection space. The novelty here is that, contrary to previous works, the set  $A$  is general and may be non compact and the diffusion coefficient may be unbounded. This induces a restricted set for the possible dimensions of the projection spaces. Under mild assumptions, risk bounds of a  $\mathbb{L}^2$ -risk are provided where new variance terms are exhibited. A data-driven selection procedure by penalization is proposed where the dimension of the projection space is chosen within a random set contrary to usual selection procedures. Risk bounds are obtained showing that the resulting estimator is adaptive in the sense that its risk achieves automatically the bias variance compromise. The estimation method is illustrated on simulated data for several diffusion models. November 8, 2018

## 1. INTRODUCTION

Non parametric drift estimation for diffusion processes from discrete observations is a widely investigated subject and to some extent thoroughly known (see *e.g.* Hoffmann, 1999, Spokoiny, 2000, Gobet *et al.*, 2004, Kutoyants, 2004, Dalalyan, 2005, Kessler *et al.*, 2012 among many authors). Nevertheless, gaps in this field remain and the aim of this paper is to fill some of these.

Consider discrete observations with sampling interval  $\Delta$ ,  $(X_{i\Delta})_{1 \leq i \leq n+1}$ , of the one-dimensional diffusion process  $(X_t)_{t \geq 0}$ , solution of

$$(1) \quad dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = \eta,$$

where  $(W_t)$  is a Wiener process and  $\eta$  is a random variable independent of  $(W_t)$ . Assume that model (1) is in stationary regime with marginal distribution  $\pi(x)dx$  and exponentially  $\beta$ -mixing. The asymptotic framework is that, as  $n$  tends to infinity, the sampling interval  $\Delta = \Delta_n$  tends to 0 while the total length time of observations  $n\Delta_n$  tends to infinity (high frequency data). The functions  $b, \sigma$  are unknown and we are concerned here with nonparametric estimation of  $b$ .

Drift estimation for diffusion models especially in high frequency framework is closely related to heteroscedastic regression function estimation. Indeed, using equation (1), one

---

*Key words and phrases.* Discrete time observation. Mean square estimator. Model selection. Nonparametric drift estimation. Stochastic differential equations.

Université Paris Descartes, Laboratoire MAP5, email: fabienne.comte@parisdescartes.fr, valentine.genon-catalot@parisdescartes.fr.

can build an approximate regression equation. Setting

$$(2) \quad Y_{i\Delta} = \frac{X_{(i+1)\Delta} - X_{i\Delta}}{\Delta}, \quad Z_{i\Delta} = \frac{1}{\Delta} \sigma(X_{i\Delta}) (W_{(i+1)\Delta} - W_{i\Delta})$$

yields  $Y_{i\Delta} = b(X_{i\Delta}) + Z_{i\Delta} + \text{remainder}$ , where the remainder term should be small for small  $\Delta$ . This is why regression function estimation methods have been applied for drift estimation by several authors (see above). A first approach is to estimate both  $b\pi$  and  $\pi$  and then do the quotient to get an estimator of  $b$ , for instance using kernels (see *e.g.*, Efroymovich, 1999, Tsybakov, 2009, for regression and Kutoyants, 2004 for diffusion models). There are no support constraints with kernels but the fact that two functions are to be estimated is a drawback as two smoothing parameters are to be chosen and the quotient may lead to unstable estimators. The alternative approach (see *e.g.* Birgé and Massart, 1998, Barron *et al.*, 1999, Baraud, 2002, for regression and Comte *et al.* (2007) for diffusions) is based on least-squares and sieves. An estimation set  $A$  is fixed and a collection of finite-dimensional subspaces of  $\mathbb{L}^2(A, dx)$  is chosen. This leads directly, by minimization of a least-squares contrast on each subspace, to a collection of estimators of  $b\mathbf{1}_A$  indexed by the dimension of the projection space. Then, a data-driven selection procedure is applied to choose adequately an estimator among the collection. For this method, authors generally assume that the set  $A$  is compact and proofs rely on this constraint.

Recently, the accuracy of least squares approximation in regression estimation is addressed by Cohen *et al.* (1993) without any support constraint. Then adaptive regression function estimation on non compact support is developed in Comte and Genon-Catalot (2018). These two papers consider only an homoscedastic model. In here, we adapt these results to drift estimation in diffusion models estimated on non compact support. Moreover, we consider an unbounded diffusion coefficient  $\sigma(\cdot)$ . We show how the results of Comte *et al.* (2007) can be generalized to these two extensions.

Section 2 gives the framework and assumptions on the model. Section 3 concerns the estimation of the drift function and is divided in several subsections. First, the approximate regression model is precised. Then, the projection estimators of  $b\mathbf{1}_A$  are built using a collection of finite-dimensional subspaces of  $\mathbb{L}^2(A, dx)$  where the estimation set  $A$  is a general Borel subset of  $\mathbb{R}$ . Due to the non compactness of the estimation set, the possible dimensions of the projection spaces are restricted to a set involving the inverse of a matrix, a constraint called "regularization" by Cohen *et al.* (2013). Risk bounds are obtained based on the expectation of an empirical norm and of the  $\mathbb{L}^2(A, \pi(x)dx)$ - norm. The variance term is completely new and differs from the one obtained in Comte *et al.* (2007). Examples of non compactly supported bases are provided and a data-driven procedure to select among the collection of estimators is proposed. In the selection procedure, the dimension is selected via a penalization criterion within a random set which is non standard in these methods and induces difficulties in proofs. Section 4 is devoted to a simulation study. Some concluding remarks are given in Section 5. Proofs are given in Section 6.

## 2. ASSUMPTIONS ON THE DIFFUSION MODEL

Consider discrete observations with sampling interval  $\Delta$ ,  $(X_{i\Delta})_{1 \leq i \leq n+1}$ , of the diffusion process  $(X_t)_{t \geq 0}$ , solution of (1) where  $(W_t)$  is a Wiener process and  $\eta$  is a random variable independent of  $(W_t)$ . The drift function  $b(\cdot)$  is unknown and our aim is to propose non-parametric estimators for it, relying on the sample  $(X_{i\Delta})_{1 \leq i \leq n+1}$ . The diffusion coefficient

$\sigma^2(\cdot)$  is also unknown and our estimation procedure will not depend on it. The asymptotic setting is:  $\Delta = \Delta_n$  tends to 0 and  $n\Delta_n$  tends to infinity as  $n$  tends to infinity. Without loss of generality, we assume  $\log n\Delta_n \geq 1$ . To simplify notations, we only write  $\Delta$  without the subscript  $n$ . However, when speaking of constants, we mean quantities that depend neither on  $n$  nor on  $\Delta$ . We consider the following assumptions.

- (A1)  $b, \sigma \in C^1(\mathbb{R})$  and there exists  $L \geq 0$ , such that, for all  $x \in \mathbb{R}$ ,  $|b'(x)| + |\sigma'(x)| \leq L$ .  
 (A2) The scale density

$$s(x) = \exp \left\{ -2 \int_0^x \frac{b(u)}{\sigma^2(u)} du \right\}$$

satisfies  $\int_{-\infty}^{\infty} s(x) dx = +\infty = \int_{-\infty}^{\infty} s(x) dx$  and the speed density  $m(x) = 1/(\sigma^2(x)s(x))$  satisfies  $\int_{-\infty}^{\infty} m(x) dx = M < +\infty$ .

- (A3)  $X_0 = \eta$  has distribution  $\pi(x)dx$  given by  $\pi(x) = M^{-1}m(x)$ .

Under (A1), Equation (1) has a unique strong solution adapted to the filtration  $(\mathcal{F}_t = \sigma(\eta, W_s, s \leq t), t \geq 0)$ . The functions  $b, \sigma$  have linear growth:

$$(3) \quad \exists K, \forall x \in \mathbb{R}, |b(x)| + |\sigma(x)| \leq K(1 + |x|).$$

Under the additional assumption (A2), Model (1) admits a unique invariant probability  $\pi(x)dx$ . And under (A3),  $(X_t)$  is strictly stationary and ergodic.

- (A4)  $(X_t)$  is geometrically  $\beta$ -mixing: there exist constants  $K > 0$  and  $\theta > 0$  such that:

$$(4) \quad \beta_X(t) \leq K e^{-\theta t},$$

where  $\beta_X(t) = \int_{-\infty}^{\infty} \pi(x) dx \|P_t(x, dx') - \pi(x') dx'\|_{TV}$  denotes the  $\beta$ -mixing coefficient of  $(X_t)$ . The norm  $\|\cdot\|_{TV}$  is the total variation norm and  $P_t$  denotes the transition probability.

- (A5)  $\|\pi\|_{\infty} < +\infty$  ( $\|\pi\|_{\infty} = \sup_{x \in A} \pi(x)$  denotes the sup norm on  $A$ ).

In Veretennikov (1988) or Pardoux and Veretennikov (2001), sufficient conditions ensuring (A4) may be found. Assumption (A5) is only used in Section 3.4. The following result is used below (see Proposition A, in Gloter (2000)):

**Proposition 2.1.** *Assume (A1)-(A3) and let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be  $C^1$  and such that there exists a constant  $\gamma \geq 0$  such that, for all  $x \in \mathbb{R}$ ,  $|f'(x)| \leq c(1 + |x|^{\gamma})$ . If  $\mathbb{E}|\eta|^{k(1+\gamma)} < +\infty$ ,*

$$(5) \quad \mathbb{E} \left( \sup_{s \in [t, t+h]} |f(X_s) - f(X_t)|^k \right) \leq c h^{k/2} (1 + \mathbb{E}|\eta|^{k(1+\gamma)}).$$

In particular, Proposition 2.1 applies for  $b$  and  $\sigma$  with  $\gamma = 0$ . In what follows, for a function  $h$ , we denote  $h_A := h \mathbf{1}_A$ ,  $\|\cdot\|$  denotes the  $\mathbb{L}^2$ -norm,  $\|\cdot\|_{\pi}$  the norm in  $\mathbb{L}^2(\pi(x)dx)$  and  $\|\cdot\|_{2,m}$  the Euclidian norm in  $\mathbb{R}^m$ .

### 3. DRIFT ESTIMATION

**3.1. Approximate regression model for the drift.** Consider  $Y_{i\Delta}, Z_{i\Delta}$  defined in (2) and set  $R_{i\Delta} := R_{i\Delta,1} + R_{i\Delta,2}$ ,

$$R_{i\Delta,1} = \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} (\sigma(X_s) - \sigma(X_{i\Delta})) dW_s, \quad R_{i\Delta,2} = \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} (b(X_s) - b(X_{i\Delta})) ds.$$

Then, process (1) satisfies the following relation:

$$(6) \quad Y_{i\Delta} = b(X_{i\Delta}) + Z_{i\Delta} + R_{i\Delta}.$$

Equation (6) is close to an heteroscedastic regression equation where  $Z_{i\Delta}$  plays the role of the noise and  $R_{i\Delta}$  is an additional residual term to take into account. This leads us to apply part of the tools proposed for regression function estimation on non compact support in Comte and Genon-Catalot (2018) to the present diffusion context. Note that in this paper, only homoscedastic regression is studied.

**Remark 3.1.** Decomposition (6) is slightly different from the one used in Comte *et al.* (2007) which was  $Y_{i\Delta} = b(X_{i\Delta}) + \tilde{Z}_{i\Delta} + R_{i\Delta,2}$ , where  $\tilde{Z}_{i\Delta} = \Delta^{-1} \int_{i\Delta}^{(i+1)\Delta} \sigma(X_s) dW_s$ . The "new" noise  $Z_{i\Delta}$  has simpler structure than  $\tilde{Z}_{i\Delta}$ . This is important for the technical tools used in the proofs (Talagrand deviation inequalities with coupling method instead of direct martingale deviations).

**3.2. Definition of the projection estimator of the drift.** In this section, several definitions and notations are common or close to the ones used in Comte and Genon-Catalot (2018) inducing some unavoidable repetitions for the text to be self-contained. Consider model (1) with observations  $(X_{i\Delta})_{1 \leq i \leq n+1}$  decomposed as in (6). Let  $A \subset \mathbb{R}$  and let  $(\varphi_j, j = 0, \dots, m-1)$  be an orthonormal system of  $A$ -supported functions belonging to  $\mathbb{L}^2(A, dx)$ . Define  $S_m = \text{span}(\varphi_0, \dots, \varphi_{m-1})$ , the linear space spanned by  $(\varphi_0, \dots, \varphi_{m-1})$ . The  $\varphi_j$ 's may depend on  $m$  but this is omitted in notations for simplicity. We assume that for all  $j$ ,  $\int \varphi_j^2(x) \pi(x) dx < +\infty$ , thus  $S_m \subset \mathbb{L}^2(A, \pi(x) dx)$ , and define an estimator of the drift function  $b$  on  $A$ , element of  $S_m$ , by:

$$\hat{b}_m = \arg \min_{t \in S_m} \gamma_n(t)$$

where  $\gamma_n(t)$  is a least-squares contrast given by

$$(7) \quad \gamma_n(t) = \frac{1}{n} \sum_{i=1}^n [t^2(X_{i\Delta}) - 2Y_{i\Delta}t(X_{i\Delta})] = n^{-1} \sum_{i=1}^n [Y_{i\Delta} - t(X_{i\Delta})]^2 - n^{-1} \sum_{i=1}^n Y_{i\Delta}^2.$$

For functions  $s, t$ , we set

$$\|t\|_n^2 = \frac{1}{n} \sum_{i=1}^n t^2(X_{i\Delta}), \quad \langle s, t \rangle_n := \frac{1}{n} \sum_{i=1}^n s(X_{i\Delta})t(X_{i\Delta}) \text{ and } \langle \vec{u}, t \rangle_n = \frac{1}{n} \sum_{i=1}^n u_i t(X_{i\Delta})$$

when  $\vec{u}$  is the vector  $(u_1, \dots, u_n)'$ ,  $\vec{u}'$  the transpose of  $\vec{u}$ , and  $t$  a function. Let

$$\hat{\Phi}_m = (\varphi_j(X_{i\Delta}))_{1 \leq i \leq n, 0 \leq j \leq m-1},$$

and

$$(8) \quad \hat{\Psi}_m = (\langle \varphi_j, \varphi_k \rangle_n)_{0 \leq j, k \leq m-1} = \frac{1}{n} \hat{\Phi}_m' \hat{\Phi}_m, \quad \Psi_m = \left( \int \varphi_j(x) \varphi_k(x) \pi(x) dx \right)_{0 \leq j, k \leq m-1} = \mathbb{E}(\hat{\Psi}_m).$$

Set  $\vec{Y} = (Y_\Delta, \dots, Y_{n\Delta})'$  and define  $\vec{a}^{(m)} = (\hat{a}_0^{(m)}, \dots, \hat{a}_{m-1}^{(m)})'$  as the  $m$ -dimensional vector such that  $\hat{b}_m = \sum_{j=0}^{m-1} \hat{a}_j^{(m)} \varphi_j$ . Assuming that  $\hat{\Psi}_m$  is invertible almost surely (a.s.) yields

$$(9) \quad \hat{b}_m = \sum_{j=0}^{m-1} \hat{a}_j^{(m)} \varphi_j, \quad \text{with} \quad \vec{a}^{(m)} = (\hat{\Phi}_m' \hat{\Phi}_m)^{-1} \hat{\Phi}_m' \vec{Y} = \frac{1}{n} \hat{\Psi}_m^{-1} \hat{\Phi}_m' \vec{Y}.$$

In what follows, the matrices  $\widehat{\Psi}_m$  and  $\Psi_m$  play a central role for the comparability of the norms  $\|\cdot\|_\pi$  and  $\|\cdot\|_n$  uniformly over a space  $S_m$ . Key tools are deviation inequalities proved in Cohen *al.* (2013) and Comte and Genon-Catalot (2018) for independent sequences of random variables. We extend these to discretely observed diffusion process.

### 3.3. Risk bounds for the drift estimator.

**Notations.** For  $M$  a matrix, we denote by  $\|M\|_{\text{op}}$  the operator norm defined as the square root of the largest eigenvalue of  $MM'$ . If  $M$  is symmetric, it coincides with  $\sup\{|\lambda_i|\}$  where  $\lambda_i$  are the eigenvalues of  $M$ .

Decomposition (6) allows to handle a non necessarily bounded volatility function. It involves the empirical processes:

$$\nu_n(t) = \frac{1}{n\Delta} \sum_{i=1}^n t(X_{i\Delta}) \sigma(X_{i\Delta}) (W_{(i+1)\Delta} - W_{i\Delta}), \quad R_{n,k}(t) = \frac{1}{n} \sum_{i=1}^n t(X_{i\Delta}) R_{i\Delta,k}, \quad k = 1, 2$$

( $R_{n,2}$  is the same as in Comte *et al.* (2007)). The following assumption is required:

$$(10) \quad L(m) := \sup_{x \in A} \sum_{j=0}^{m-1} \varphi_j^2(x) < +\infty.$$

One easily checks that  $L(m)$  does not depend on the choice of the  $\mathbb{L}^2(dx)$ -orthonormal basis of  $S_m$ . If the spaces  $S_m$  are nested, i.e.  $m \leq m' \Rightarrow S_m \subset S_{m'}$ , the map  $m \mapsto L(m)$  is increasing. Assuming  $\mathbb{E}\eta^2 < +\infty$  and using (3) and (10), we define

$$(11) \quad \Psi_{m,\sigma^2} := \left( \int \varphi_j(x) \varphi_k(x) \sigma^2(x) \pi(x) dx \right)_{0 \leq j,k \leq m-1}.$$

To ensure the stability of the least-squares estimator, we must consider a truncated version of  $\hat{b}_m$  given by

$$(12) \quad \widetilde{b}_m = \hat{b}_m \mathbf{1}_{\{L(m)(\|\widehat{\Psi}_m^{-1}\|_{\text{op}} \vee 1) \leq \mathfrak{c} n \Delta / \log^2(n\Delta)\}}, \quad \mathfrak{c} = \frac{\theta(1 - \log 2)}{C_0}$$

where  $C_0$  is a numerical constant,  $C_0 \geq 72$ . The choice of  $\mathfrak{c}$  is done using Proposition 6.1 (i) in Lemma 6.1. Due to  $\theta$ , the constant  $\mathfrak{c}$  is unknown. To avoid this problem, for  $n$  large enough, we can simply change  $\log^2(n\Delta)$  into  $\log^{2+\epsilon}(n\Delta)$  with  $\epsilon > 0$ , and take a constant  $\mathfrak{c}$  equal to 1.

Note that, when the spaces  $S_m$  are nested and  $\Psi_m$  (resp.  $\widehat{\Psi}_m$ ) is invertible,  $m \mapsto \|\Psi_m^{-1}\|_{\text{op}}$  (resp.  $m \mapsto \|\widehat{\Psi}_m^{-1}\|_{\text{op}}$ ) is non-decreasing (see Proposition 2.2 in Comte and Genon-Catalot (2018)). We prove:

**Proposition 3.1.** *Let  $(X_{i\Delta})_{1 \leq i \leq n}$  be observations drawn from model (6) under assumptions (A1)-(A4) and (10), with  $\Delta \leq 1$ ,  $\Delta = \Delta_n \rightarrow 0$  and  $n\Delta \rightarrow +\infty$  when  $n \rightarrow +\infty$ . Assume that  $\mathbb{E}(\eta^4) < +\infty$ . Consider the estimator  $\tilde{b}_m$  of  $b_A$ . Then for  $m$  such that*

$$(13) \quad L(m)(\|\Psi_m^{-1}\|_{\text{op}} \vee 1) \leq \frac{\mathfrak{c} n \Delta}{2 \log^2(n\Delta)} \quad \text{and} \quad m \leq n\Delta$$

with  $\mathfrak{c}$  given in (12), we have

$$\mathbb{E}[\|\tilde{b}_m - b\|_n^2] \leq 7 \inf_{t \in S_m} \|b_A - t\|_\pi^2 + \frac{64}{n\Delta} \text{Tr}[\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}] + c_1 \Delta + \frac{c_2}{n\Delta},$$

$$\mathbb{E}[\|\tilde{b}_m - b\|_\pi^2] \leq c_3 \left\{ \inf_{t \in S_m} \|b_A - t\|_\pi^2 + \frac{1}{n\Delta} \text{Tr}[\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}] + \Delta + \frac{1}{n\Delta} \right\},$$

where  $c_1, c_2, c_3$  are positive constants.

Note that, for  $m$  satisfying condition (13),  $\Psi_m$  is invertible and its eigenvalues are lower bounded by  $2m \log^2(n\Delta)/cn\Delta$ . Condition  $m \leq n\Delta$  is actually included in the first part of (13): indeed, if  $(\theta_j, j = 0, \dots, m-1)$  is an orthonormal basis of  $S_m$  with respect to  $\mathbb{L}^2(A, \pi(x)dx)$ , and  $K(m) = \sup_{x \in A} \sum_{j=0}^{m-1} \theta_j^2(x)$ , then  $K(m) \geq m$  and one can prove that  $K(m) \leq L(m) \|\Psi_m^{-1}\|_{\text{op}}$  (see Lemma 6.2 in Comte and Genon-Catalot (2018)).

As the spaces  $S_m$  are nested, the bias term,  $\inf_{t \in S_m} \|b_A - t\|_\pi^2$  decreases when  $m$  increases. The terms  $c_1\Delta + c_2/(n\Delta)$  are residual terms tending to zero under our asymptotic framework. The novelty is the variance term  $\text{Tr}[\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}]$ . It is non-decreasing and can be upper bounded in several manners (see Proposition below). Note that if  $\sigma(x) \equiv \sigma$ , then  $\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2} = \sigma^2 \text{Id}_m$  and  $\text{Tr}[\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}] = \sigma^2 m$ .

**Proposition 3.2.** *Let  $m$  be an integer. Assume that  $\Psi_m$  is invertible and  $\mathbb{E}\eta^2 < +\infty$ .*

- (1) *If the spaces  $S_m$  are nested, then  $m \mapsto \text{Tr}[\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}]$  is non-decreasing.*
- (2) *If  $\sigma$  is bounded on  $A$ , then  $\text{Tr}[\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}] \leq \|\sigma_A\|_\infty^2 m$ .*
- (3) *Assume that (10) holds, then  $\text{Tr}[\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}] \leq \mathbb{E}[\sigma_A^2(X_0)] L(m) \|\Psi_m^{-1}\|_{\text{op}}$ .*

From (2), if  $\sigma$  is bounded on  $A$  (compact or not), we recover the result of Comte *et al.* (2007), see the penalty term in Theorem 1 therein (which has order  $\|\sigma\|_\infty m/(n\Delta)$ ).

If in addition,  $A$  is compact, we can assume that  $\pi(x) \geq \pi_0 > 0, \forall x \in A$ . By Proposition 4.1 in Comte and Genon-Catalot (2018), this implies that  $\|\Psi_m^{-1}\|_{\text{op}} \leq 1/\pi_0$ . In such a case, condition (13) contains all indices up to order  $n\Delta/\log^2(n\Delta)$ . Therefore, the result of Proposition 3.1 encompasses the result given in Proposition 1 of Comte *et al.* (2007). Note that, in the latter paper, only the empirical risk is studied.

On the other hand, it is proved in Comte and Genon-Catalot (2018) that, with the non compactly supported Laguerre and Hermite bases,

- a) for all  $m \leq n$ ,  $\widehat{\Psi}_m$  is a.s. invertible;
- b) for all  $m$ ,  $\Psi_m$  is invertible and there exists a constant  $c^*$  such that,  $\|\Psi_m^{-1}\|_{\text{op}}^2 \geq c^* m$ .

Property b) enlightens that  $\|\Psi_m^{-1}\|_{\text{op}}$  has a real weight and increases the variance.

**3.4. Model selection.** We precise condition (10) as follows:

- (B1)** The collection of spaces  $S_m$  is nested (that is  $S_m \subset S_{m'}$  for  $m \leq m'$ ) and such that, for each  $m$ , the basis  $(\varphi_0, \dots, \varphi_{m-1})$  of  $S_m$  satisfies

$$(14) \quad L(m) = \left\| \sum_{j=0}^{m-1} \varphi_j^2 \right\|_\infty \leq c_\varphi^2 m \quad \text{for } c_\varphi^2 > 0 \quad \text{a constant.}$$

This assumption is shared by most classical bases on a compact support (histograms, trigonometric polynomials). For non compact support, we have in mind concrete examples of orthonormal bases. First, for  $A = \mathbb{R}^+$ , the basis of  $\mathbb{L}^2(\mathbb{R}^+, dx)$  composed of Laguerre functions  $(\ell_j, j \geq 0)$ , and  $S_m = \text{span}(\ell_0, \dots, \ell_{m-1})$ ; second, for  $A = \mathbb{R}$ , the basis of  $\mathbb{L}^2(\mathbb{R}, dx)$  composed of Hermite functions  $(h_j, j \geq 0)$ , and  $S_m = \text{span}(h_0, \dots, h_{m-1})$  (see Section 4.2 in Comte and Genon-Catalot (2018) and the simulation study below). Laguerre

and Hermite functions being uniformly bounded functions, condition (14) holds.

We define, for  $\pi$  upper-bounded on  $A$  by  $\|\pi\|_\infty$  (that is under **(A5)**), the collection:

$$(15) \quad \mathcal{M}_{n\Delta} = \left\{ m \in \mathbb{N}, c_\varphi^2 m (\|\Psi_m^{-1}\|_{\text{op}}^2 \vee 1) \leq \frac{\mathfrak{d}}{4} \frac{n\Delta}{\log^2(n\Delta)} \right\}, \quad \mathfrak{d} = \frac{\theta}{8C_0 (\|\pi\|_\infty \vee 1 + \frac{1}{3})},$$

where  $\theta$  is defined in (4) and  $C_0 \geq 72$  is the same as in  $\mathfrak{c}$ . The choice of  $\mathfrak{d}$  comes from Lemma 6.4 and uses Proposition 6.1, (ii). Due to  $\|\pi\|_\infty$  and  $\theta$ , the constant  $\mathfrak{d}$  is unknown. As previously, for  $n$  large enough, we can simply change  $\log^2(n\Delta)$  into  $\log^{2+\epsilon}(n\Delta)$  with  $\epsilon > 0$  and take a constant  $\mathfrak{d}$  equal to 1. Note that the constraint on  $m$  in  $\mathcal{M}_{n\Delta}$  is stronger than the one in (13) as  $m (\|\Psi_m^{-1}\|_{\text{op}} \vee 1) \leq m (\|\Psi_m^{-1}\|_{\text{op}}^2 \vee 1)$ .

Introducing the random collection of models  $\widehat{\mathcal{M}}_{n\Delta}$  given by

$$(16) \quad \widehat{\mathcal{M}}_{n\Delta} = \left\{ m \in \mathbb{N}, c_\varphi^2 m (\|\widehat{\Psi}_m^{-1}\|_{\text{op}}^2 \vee 1) \leq \mathfrak{d} \frac{n\Delta}{\log^2(n\Delta)} \right\},$$

with  $\mathfrak{d}$  defined in (15), we define the data-driven selection of  $m$  by

$$(17) \quad \hat{m} = \arg \min_{m \in \widehat{\mathcal{M}}_{n\Delta}} \left\{ -\|\hat{b}_m\|_n^2 + \kappa c_\varphi^2 s^2 \frac{m \|\widehat{\Psi}_m^{-1}\|_{\text{op}}}{n\Delta} \right\}, \quad s^2 = \mathbb{E}[\sigma^2(X_0)],$$

where  $\kappa$  is a numerical constant. The set  $\widehat{\mathcal{M}}_{n\Delta}$  is the empirical counterpart of  $\mathcal{M}_{n\Delta}$  defined by (15), with constant multiplied by 4. This is different from the usual selection procedures where the set of possible values for choosing the dimension  $m$  is nonrandom. Note that for  $m \in \widehat{\mathcal{M}}_{n\Delta}$ ,  $\hat{b}_m = \tilde{b}_m$ , and these are the only  $m$  which are considered.

**Theorem 3.1.** *Let  $(X_{i\Delta})_{1 \leq i \leq n}$  be observations from model (6). Assume that **(A1)**-(**A5**) and **(B1)** hold and that  $\mathbb{E}\eta^6 < +\infty$ . Assume that  $\Delta = \Delta_n \rightarrow 0$  and  $n\Delta \rightarrow +\infty$  when  $n \rightarrow +\infty$ . Then, there exists a numerical constant  $\kappa_0$  such that for  $\kappa \geq \kappa_0$ , we have*

$$(18) \quad \mathbb{E}(\|\hat{b}_{\hat{m}} - b_A\|_n^2) \leq C \inf_{m \in \mathcal{M}_{n\Delta}} \left[ \inf_{t \in S_m} \|b_A - t\|_\pi^2 + \kappa c_\varphi^2 s^2 \frac{m \|\Psi_m^{-1}\|_{\text{op}}}{n\Delta} \right] + c_1 \Delta + \frac{c_1 \log^2(n\Delta)}{n\Delta},$$

and

$$(19) \quad \mathbb{E}(\|\hat{b}_{\hat{m}} - b_A\|_\pi^2) \leq C_1 \inf_{m \in \mathcal{M}_{n\Delta}} \left[ \inf_{t \in S_m} \|b_A - t\|_\pi^2 + \kappa c_\varphi^2 s^2 \frac{m \|\Psi_m^{-1}\|_{\text{op}}}{n\Delta} \right] + c'_1 \Delta + \frac{c'_2 \log^2(n\Delta)}{n\Delta}$$

where  $C, C_1$  are numerical constants and  $c_1, c_2, c'_1, c'_2$  are constants depending on  $\pi, b, \sigma$ .

Inequalities (18) and (19) show that the estimator  $\hat{b}_{\hat{m}}$  automatically realizes the compromise between the squared bias and the variance bound. The results are a substantial generalization of Theorem 1 in Comte *et al.* (2007). In section 4, we explain how to estimate  $s^2$  and how to fix  $\kappa$ .

#### 4. SIMULATION STUDY

Samples  $(X_{i\Delta})_{1 \leq i \leq n}$  from the following models were generated for  $(n, \Delta) = (100000, 0.02)$  ( $n\Delta = 2000$ ) and for  $(n, \Delta) = (50000, 0.01)$  ( $n\Delta = 500$ ).

**Example 1.** Hyperbolic diffusion. The model  $dX_t = -\theta X_t dt + \gamma \sqrt{1 + X_t^2} dW_t$ ,  $X_0 = \eta$ , is simulated by a Euler scheme with step  $\delta$ . We keep one out of 10 observations i.e.

$\Delta = 10\delta$ . Assumptions **(A1)**-(**A2**) hold for  $\theta > -\gamma^2/2$ . The invariant density is proportional to  $1/(1+x^2)^{1+(\theta/\gamma^2)}$  and  $\int x^4 \pi(x) dx < +\infty$  for  $\theta > 3\gamma^2/2$ . We chose  $\theta = 2$  and  $\gamma = \sqrt{1/2}$ . Setting  $Y_t = \arg sh X_t$ , we see that the process  $(Y_t)$  satisfies the conditions of Pardoux and Veretennikov (2001) ensuring the exponential  $\beta$ -mixing property. Therefore,  $(X_t)$  satisfies **(A4)**. For simplicity, as the invariant distribution is centered, instead of starting the process with the invariant distribution, we set  $X_0 = 0$ .

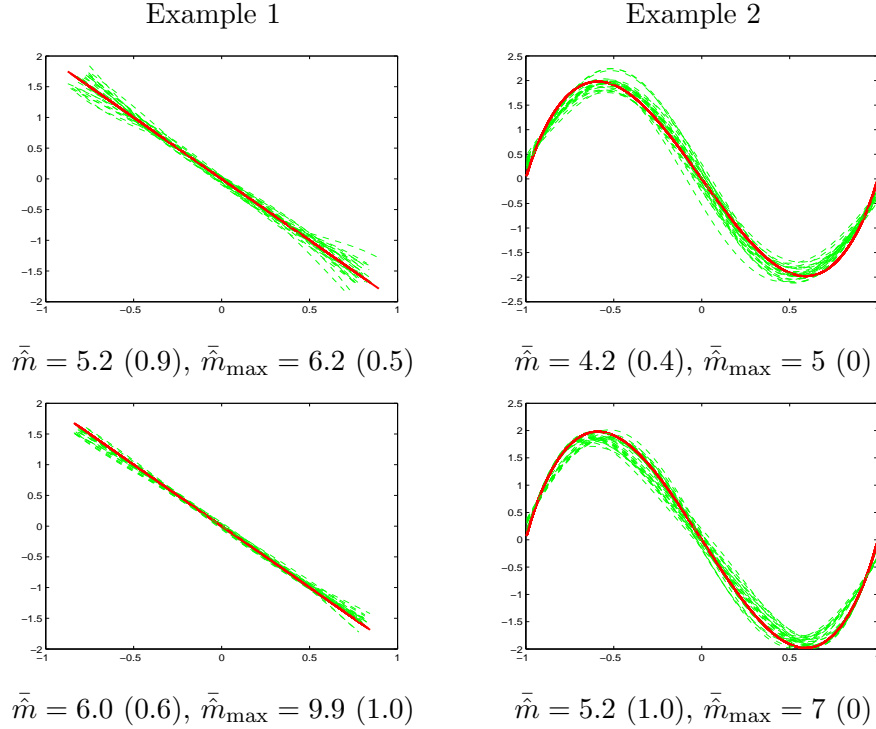


FIGURE 1. 25 estimated curves in Hermite basis (dotted-green), the true in bold (red),  $n\Delta = 500$ , top and  $n\Delta = 2000$ , bottom.

	Example 1		Example 2	
	$\hat{s}^2$	$\tilde{s}^2$	$\hat{s}^2$	$\tilde{s}^2$
$n\Delta = 500$	0.56 (0.005)	0.57 (0.004)	1.38 (0.046)	1.410 (0.048)
$n\Delta = 2000$	0.55 (0.003)	0.57 (0.002)	1.33 (0.021)	1.40 (0.024)

TABLE 1. Estimation of  $s^2 = \mathbb{E}[\sigma^2(X)]$  associated with the paths in Figure 1.

The other examples are obtained from a  $d$ -dimensional Ornstein-Uhlenbeck process  $(U_t)_{t \geq 0}$ , with dynamics given by

$$(20) \quad dU_t = -\frac{k}{2} U_t dt + \frac{\gamma}{2} dW_t, \quad U_0 \sim \mathcal{N}(0, \frac{\gamma^2}{4k} I_d).$$



Exact simulation is generated with step  $\Delta$  by computing

$$U_{(p+1)\Delta} = e^{-\frac{k\Delta}{2}} U_{p\Delta} + \varepsilon_{k\Delta} + \varepsilon_{(p+1)\Delta}, \quad \varepsilon_{k\Delta} \sim_{\text{iid}} \mathcal{N}(0, \frac{\gamma^2(1 - e^{-k\Delta})}{4k} I_d).$$

**Example 2.**  $X_t = \tanh(U_t)$  where  $U_t$  is defined by (20) with  $d = 1$  is solution of (1) with

$$b(x) = (1 - x^2) \left( -\frac{k}{2} \operatorname{atanh}(x) - \frac{\gamma^2}{4} x \right), \quad \sigma(x) = \frac{\gamma}{2} (1 - x^2), \quad \text{with } k = 4 \text{ and } \gamma = 2.$$

**Example 3.**  $X_t = \exp(U_t)$  where  $U_t$  is defined by (20) with  $d = 1$  is solution of (1) with

$$b(x) = x \left( -\frac{k}{2} \log(x) + \frac{\gamma^2}{8} \right), \quad \sigma(x) = \frac{\gamma}{2} x, \quad \text{with } k = 1 \text{ and } \gamma = 1/2.$$

**Example 4.** Cox-Ingersoll-Ross or square-root process. We take  $X_t = \|U_t\|_{2,d}^2$  where  $U_t$  is defined by (20) with  $d = 3$  is solution of (1) with

$$dX_t = \left( \frac{d\gamma^2}{4} - kX_t \right) dt + \gamma \sqrt{X_t} dW_t^*,$$

where  $W_t^*$  is a standard brownian motion. We take  $k = 2$  and  $\gamma = 1$ .

In examples 2,3,4, the models are strictly stationary, ergodic and  $\beta$ -mixing but the functions  $b, \sigma$  do not satisfy **(A1)**. Nevertheless, we implement the estimation method as a specific study would show that our results hold. Examples 3,4 provide nonnegative processes and allow to use Laguerre basis.

Implementation is done with either the Laguerre basis ( $A = \mathbb{R}^+$ ) or the Hermite basis ( $A = \mathbb{R}$ ) which are easy to handle in practice.

- Laguerre basis,  $A = \mathbb{R}^+$ . The Laguerre polynomials ( $L_j$ ) and the Laguerre functions ( $\ell_j$ ) are given by

$$(21) \quad L_j(x) = \sum_{k=0}^j (-1)^k \binom{j}{k} \frac{x^k}{k!}, \quad \ell_j(x) = \sqrt{2} L_j(2x) e^{-x} \mathbf{1}_{x \geq 0}, \quad j \geq 0.$$

The collection  $(\ell_j)_{j \geq 0}$  constitutes a complete orthonormal system on  $\mathbb{L}^2(\mathbb{R}^+)$  satisfying (see Abramowitz and Stegun (1964)):  $\forall j \geq 0, \forall x \in \mathbb{R}^+, |\ell_j(x)| \leq \sqrt{2}$ . The collection of models ( $S_m = \operatorname{span}\{\ell_0, \dots, \ell_{m-1}\}$ ) is nested and obviously (14) holds with  $c_\varphi^2 = 2$ .

- Hermite basis,  $A = \mathbb{R}$ . The Hermite polynomial and the Hermite function of order  $j$  are given, for  $j \geq 0$ , by:

$$(22) \quad H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j} (e^{-x^2}), \quad h_j(x) = c_j H_j(x) e^{-x^2/2}, \quad c_j = (2^j j! \sqrt{\pi})^{-1/2}$$

The sequence  $(h_j, j \geq 0)$  is an orthonormal basis of  $\mathbb{L}^2(\mathbb{R}, dx)$ . Moreover (see Abramowitz and Stegun (1964), Szegő (1959) p.242),  $\|h_j\|_\infty \leq \Phi_0, \Phi_0 \simeq 1,086435/\pi^{1/4} \simeq 0.8160$ , so that (14) holds with  $c_\varphi^2 = \Phi_0^2$ . The collection of models ( $S_m = \operatorname{span}\{h_0, \dots, h_{m-1}\}$ ) is obviously nested. Laguerre polynomials were computed using formula (21) and Hermite polynomials with  $H_0(x) \equiv 1, H_1(x) = x$  and the recursion  $H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$ .

The computation of the set  $\widehat{\mathcal{M}}_n$  raises a problem: the value given by the theory, namely  $c_\varphi^2 m \|\widehat{\Psi}_m^{-1}\|_{\text{op}}^2$  is always very large so that the set of possible  $m$ 's is reduced to  $\{1\}$ . Therefore

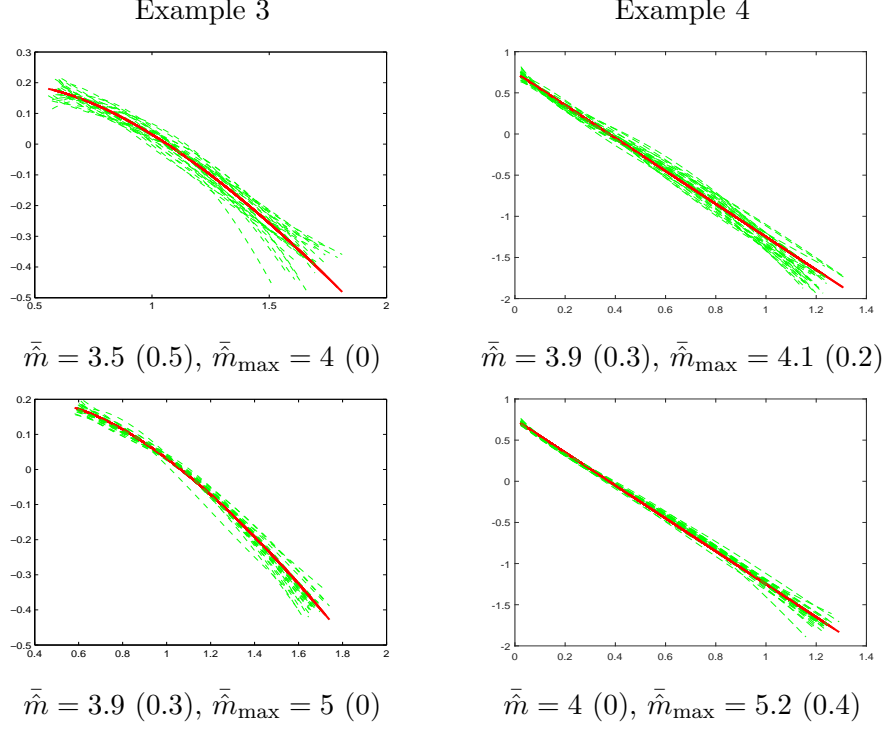


FIGURE 2. 25 estimated curves in Laguerre basis (dotted-green), the true in bold (red),  $n\Delta = 500$ , top and  $n\Delta = 2000$ , bottom.

	Example 3		Example 4	
	$\hat{s}^2$	$\tilde{s}^2$	$\hat{s}^2$	$\tilde{s}^2$
$n\Delta = 500$	0.07 (0.004)	0.07 (0.004)	0.37 (0.01)	0.37 (0.01)
$n\Delta = 2000$	0.07 (0.002)	0.07 (0.002)	0.36 (0.007)	0.37 (0.007)

TABLE 2. Estimation of  $s^2 = \mathbb{E}[\sigma^2(X_0)]$  associated with the paths of Figure 2.

we choose a larger set given by  $\widehat{\mathcal{M}}_n^* = \{m, \|\widehat{\Psi}_m^{-1}\|_{\text{op}}^{1/4} \leq [n\Delta/(\log(n\Delta))^{2.1}]\}$ . Numerically, we obtain  $\widehat{\mathcal{M}}_n^* = \{1, \dots, \hat{m}_{\max}\}$ , with  $\hat{m}_{\max} \leq 10$ , which is more satisfactory.

The term  $s^2 := \mathbb{E}[\sigma^2(X_\Delta)]$  is estimated by residual least squares corresponding to this maximal dimension,

$$\hat{s}^2 = \frac{\Delta}{n} \sum_{i=1}^n [Y_{i\Delta} - \hat{b}_{\hat{m}_{\max}}(X_{i\Delta})]^2.$$

In Tables 1-2, we compare it with  $\tilde{s}^2 = (1/n) \sum_{i=1}^n \sigma^2(X_{i\Delta})$ , a pseudo-estimator using the (unavailable) knowledge of the function  $\sigma^2$ . The comparison is done for the 25 paths generated for Figures 1-2: we can see that the estimation is quite good.

The constant  $\kappa$  is standardly calibrated by preliminary simulations and taken equal to  $\kappa = 2.10^{-4}$ . It is not surprising that  $\kappa$  must be chosen very small: this is due to the

fact that  $m\|\widehat{\Psi}_m^{-1}\|_{\text{op}}$  is large, as noted above. The penalty is taken equal to  $\widehat{\text{pen}}(m) = \kappa c_\varphi^2 \widehat{s}^2 m \|\widehat{\Psi}_m^{-1}\|_{\text{op}} / (n\Delta)$  and  $\hat{m}$  is selected as the minimizer of  $-\|\hat{b}_m\|_n^2 + \widehat{\text{pen}}(m)$ .

We present in Figures 1-2 beams of 25 estimators  $\hat{b}_{\hat{m}}$  corresponding to 25 simulated trajectories of each model using the Hermite basis for examples 1 and 2 and Laguerre basis for examples 3 and 4. We stress that the value of  $\hat{m}$  is rather small: under each graph, we give the mean  $\bar{m}$  computed over the 25 estimators and the mean of the maximal value  $\hat{m}_{\text{max}}$ , both with standard deviation in parenthesis. It is noteworthy that the function is very well reconstructed using a small number of coefficients.

## 5. CONCLUDING REMARKS

In this paper, we revisit the problem of nonparametric drift estimation for an ergodic diffusion from discrete observations of the sample path. As in Comte *et al.*, 2007, an estimation set  $A$  is chosen. The drift function is estimated on  $A$  using an approximate regression equation by a least squares contrast which is minimized over a finite dimensional subspace  $S_m$  of  $\mathbb{L}^2(A, dx)$ . This yields a collection of estimators indexed by the dimension of the projection space. A data-driven procedure is proposed to select the best dimension using a penalization criterion. While in Comte *et al.*, 2007, the set  $A$  must be compact and the diffusion coefficient must be uniformly bounded, the novelty of the present paper is to get rid of these two assumptions. This leads to considerable modifications in the method of estimation and complicates a lot proofs. First, we rely on a slightly different approximate regression equation. This allows to modify the way of dealing with the adaptive procedure. Second, the possible dimensions to define the projection estimators are restricted to a set involving the inversion of a matrix which does not even appear when the set  $A$  is compact. In the risk bounds, the variance term is different from the case where  $A$  is compact and  $\sigma(\cdot)$  bounded. Moreover, to define the adaptive procedure, the adequate dimension is to be selected within a random set. This induces difficulties and a non standard treatment of the classical method of penalization. The estimator obtained is nevertheless adaptive in the sense that its  $\mathbb{L}^2$ -risk achieves the best compromise between the square bias and the new variance term.

An important question may be to look at rates of convergence and optimality in this new setting. This is treated in the case of the simple regression model with independent data on non compact support in Comte and Genon-Catalot (2018) and is worth of interest in the diffusion context.

Estimation of  $\sigma^2$  could be investigated too under the same set of assumptions but leads to rather lengthy developments.

## 6. PROOFS

**6.1. Preliminary properties.** Consider the set where the empirical and  $\mathbb{L}^2(A, \pi(x)dx)$ -norms on  $S_m$  are equivalent:

$$(23) \quad \Omega_m(u) = \left\{ \sup_{t \in S_m, t \neq 0} \left| \frac{\|t\|_n^2}{\|t\|_\pi^2} - 1 \right| \leq u \right\}.$$

We generalize to the diffusion context Proposition 2.3 of Comte and Genon-Catalot (2018) (see also Theorem 1 of Cohen *et al.* (1993) for part (i)).

**Proposition 6.1.** *Let  $(X_{i\Delta})_i$  be a discrete sampling of the process  $(X_t)$  given by (1) and assume **(A1)**-**(A4)** (thus  $(X_{i\Delta})_i$  is strictly stationary and geometrically  $\beta$ -mixing with  $\beta$ -mixing coefficients satisfying  $\beta(i) = \beta_X(i\Delta) \leq Ke^{-\theta i\Delta}$  for some constants  $K > 0, \theta > 0$ ). Consider a basis satisfying (10) and let  $\text{Id}_m$  denotes the  $m \times m$  identity matrix.*

(i) *Assume that  $\Psi_m$  is invertible. For  $\widehat{\Psi}_m$  defined by Equation (8), for all  $u \in [0, 1]$*

$$\begin{aligned} \mathbb{P}(\Omega_m(u)^c) &= \mathbb{P}\left[\|\Psi_m^{-1/2}\widehat{\Psi}_m\Psi_m^{-1/2} - \text{Id}_m\|_{\text{op}} > u\right] \\ &\leq 4m \exp\left(-\frac{n\Delta\theta c(u)}{12\log(n\Delta) L(m)(\|\Psi_m^{-1}\|_{\text{op}} \vee 1)}\right) + \frac{\theta}{6(n\Delta)^5}, \end{aligned}$$

where  $c(u) = u + (1 - u)\log(1 - u)$ .

(ii) *If in addition **(A5)** holds, then, for all  $u > 0$ ,*

$$\mathbb{P}\left[\|\Psi_m - \widehat{\Psi}_m\|_{\text{op}} \geq u\right] \leq 4m \exp\left(-\frac{n\Delta\theta u^2/2}{12L(m)\log(n\Delta)(\|\pi\|_{\infty} \vee 1 + 2u/3)}\right) + \frac{\theta}{6(n\Delta)^5}.$$

**6.2. Proof of Proposition 6.1.** For  $t = \sum_{j=0}^{m-1} x_j \varphi_j$  in  $S_m$ ,  $\|t\|_{\pi}^2 = \vec{x}'\Psi_m\vec{x}$ ,  $\|t\|_n^2 = \vec{x}'\widehat{\Psi}_m\vec{x}$ . Thus,

$$\begin{aligned} \sup_{t \in S_m, \|t\|_{\pi}=1} \left| \frac{1}{n} \sum_{i=1}^n [t^2(X_i) - \mathbb{E}t^2(X_i)] \right| &= \sup_{\vec{x} \in \mathbb{R}^m, \|\Psi_m^{1/2}\vec{x}\|_{2,m}=1} \left| \vec{x}'\widehat{\Psi}_m\vec{x} - \vec{x}'\Psi_m\vec{x} \right| \\ &= \sup_{\vec{u} \in \mathbb{R}^m, \|\vec{u}\|_{2,m}=1} \left| \vec{u}'\Psi_m^{-1/2}(\widehat{\Psi}_m - \Psi_m)\Psi_m^{-1/2}\vec{u} \right| = \|\Psi_m^{-1/2}\widehat{\Psi}_m\Psi_m^{-1/2} - \text{Id}_m\|_{\text{op}}. \end{aligned}$$

Hence,

$$(24) \quad \Omega_m(u)^c = \left\{ \|\Psi_m^{-1/2}\widehat{\Psi}_m\Psi_m^{-1/2} - \text{Id}_m\|_{\text{op}} > u \right\}.$$

Now, we consider the coupling method and the associated variables  $(X_{i\Delta}^*)$  with Berbee's Lemma, see Berbee (1979), with the method described in Viennet (1997, Prop.5.1 and its proof p.484). Assume for simplicity that  $n = 2p_n q_n$  for integers  $p_n, q_n$ . Then there exist random variables  $X_{i\Delta}^*$ ,  $i = 1, \dots, n$  satisfying the following properties:

- For  $\ell = 0, \dots, p_n - 1$ , the random vectors

$$\vec{X}_{\ell,1} = (X_{(2\ell q_n+1)\Delta}, \dots, X_{(2\ell+1)q_n\Delta})' \text{ and } \vec{X}_{\ell,1}^* = (X_{(2\ell q_n+1)\Delta}^*, \dots, X_{(2\ell+1)q_n\Delta}^*)'$$

have the same distribution, and so have the random vectors

$$\vec{X}_{\ell,2} = (X_{[(2\ell+1)q_n+1]\Delta}, \dots, X_{(2\ell+2)q_n\Delta})' \text{ and } \vec{X}_{\ell,2}^* = (X_{[(2\ell+1)q_n+1]\Delta}^*, \dots, X_{(2\ell+2)q_n\Delta}^*)'.$$

- For  $\ell = 0, \dots, p_n - 1$ ,

$$(25) \quad \mathbb{P}\left[\vec{X}_{\ell,1} \neq \vec{X}_{\ell,1}^*\right] \leq \beta(q_n) = \beta_X(q_n\Delta) \text{ and } \mathbb{P}\left[\vec{X}_{\ell,2} \neq \vec{X}_{\ell,2}^*\right] \leq \beta_X(q_n\Delta).$$

- For each  $\delta \in \{1, 2\}$ , the random vectors  $\vec{X}_{0,\delta}^*, \dots, \vec{X}_{p_n-1,\delta}^*$  are independent.

Then let  $\Omega^* = \{X_{i\Delta} = X_{i\Delta}^*, i = 1, \dots, n\}$  and write that

$$(26) \quad \begin{aligned} \mathbb{P}\left[\|\text{Id}_m - \Psi_m^{-1/2}\widehat{\Psi}_m\Psi_m^{-1/2}\|_{\text{op}} > u\right] &\leq \mathbb{P}\left[\{\|\text{Id}_m - \Psi_m^{-1/2}\widehat{\Psi}_m\Psi_m^{-1/2}\|_{\text{op}} > u\} \cap \Omega^*\right] \\ &\quad + \mathbb{P}[(\Omega^*)^c]. \end{aligned}$$

Using the definition of the variables  $X_{i\Delta}^*$ , we get  $\mathbb{P}[(\Omega^*)^c] \leq 2p_n\beta_X(q_n\Delta) \leq 2p_ne^{-\theta q_n\Delta}$ . Choosing  $q_n\Delta = 6\log(n\Delta)/\theta$ , thus  $2p_n = n/q_n = n\Delta\theta/6\log(n\Delta)$ , yields

$$(27) \quad \mathbb{P}[(\Omega^*)^c] \leq \frac{\theta n\Delta}{6\log(n\Delta)(n\Delta)^6} \leq \frac{\theta}{6(n\Delta)^5}.$$

Now, we write  $\mathbf{S}_m = (1/2)(\mathbf{S}_{m,1} + \mathbf{S}_{m,2})$  where  $\mathbf{S}_m$  is given by

$$\mathbf{S}_m = \Psi_m^{-1/2} \widehat{\Psi}_m \Psi_m^{-1/2} - \text{Id}_m = \frac{1}{n} \sum_{i=1}^n \{\mathbf{K}_m(X_{i\Delta}) - \mathbb{E}[\mathbf{K}_m(X_{i\Delta})]\},$$

$$(28) \quad \mathbf{K}_m(X_{i\Delta}) = \Psi_m^{-1/2} \widetilde{\mathbf{K}}_m(X_{i\Delta}) \Psi_m^{-1/2}, \quad \widetilde{\mathbf{K}}_m(X_{i\Delta}) = (\varphi_j(X_{i\Delta}) \varphi_k(X_{i\Delta}))_{0 \leq j, k \leq m-1},$$

with obviously  $\mathbb{E}[\mathbf{K}_m(X_{i\Delta})] = \text{Id}_m$ . Here,  $\mathbf{S}_{m,1}$  is built with the  $\vec{X}_{\ell,1}$ :

$$\mathbf{S}_{m,1} = \frac{1}{p_n} \sum_{\ell=0}^{p_n-1} \frac{1}{q_n} \sum_{r=1}^{q_n} \{\mathbf{K}_m(X_{(2\ell q_n+r)\Delta}) - \mathbb{E}(\mathbf{K}_m(X_{(2\ell q_n+r)\Delta}))\}$$

and  $\mathbf{S}_{m,2}$  is analogously defined with the  $\vec{X}_{\ell,2}$ . We have

$$(29) \quad \begin{aligned} \mathbb{P}\left[\|\text{Id}_m - \Psi_m^{-1/2} \widehat{\Psi}_m \Psi_m^{-1/2}\|_{\text{op}} > u\right] \cap \Omega^* &= \mathbb{P}\left[\|\mathbf{S}_{m,1} + \mathbf{S}_{m,2}\|_{\text{op}} > 2u\right] \cap \Omega^* \\ &\leq \mathbb{P}\left[\|\mathbf{S}_{m,1}^*\|_{\text{op}} > u\right] + \mathbb{P}\left[\|\mathbf{S}_{m,2}^*\|_{\text{op}} > u\right], \end{aligned}$$

where  $\mathbf{S}_{m,\delta}^*$ ,  $\delta = 1, 2$  are built on the  $\vec{X}_{\ell,\delta}^*$ . The two terms are similar so we only treat one. We write

$$\mathbf{S}_{m,1}^* = \sum_{\ell=0}^{p_n-1} \mathbf{G}_\ell^* - \text{Id}_m \quad \text{where} \quad \mathbf{G}_\ell^* = \frac{1}{p_n} \frac{1}{q_n} \sum_{r=1}^{q_n} \mathbf{K}_m(X_{(2\ell q_n+r)\Delta}^*).$$

As  $\|\mathbf{G}_\ell^*\|_{\text{op}} \leq L(m)\|\Psi_m^{-1}\|_{\text{op}}/p_n$ , by Theorem 1 in Cohen *et al.* (2013) and Lemma 6.2 in Comte and Genon-Catalot (2018), we get

$$\mathbb{P}(\|\mathbf{S}_{m,1}^*\|_{\text{op}} > u) \leq 2m \exp\left(-\frac{c(u)p_n}{L(m)\|\Psi_m^{-1}\|_{\text{op}}}\right).$$

Using the definition of  $p_n$ , gives the result (i) of Proposition 6.1.

To prove (ii), we proceed similarly and bound  $\mathbb{P}\left[\|\Psi_m - \widehat{\Psi}_m\|_{\text{op}} \geq u\right] \cap \Omega^*$ . We write  $\widetilde{\mathbf{S}}_m = (1/2)(\widetilde{\mathbf{S}}_{m,1} + \widetilde{\mathbf{S}}_{m,2})$  where  $(\widetilde{\mathbf{K}}_m)$  is defined by (28):

$$\begin{aligned} \widetilde{\mathbf{S}}_m &= \frac{1}{n} \sum_{i=1}^n \{\widetilde{\mathbf{K}}_m(X_{i\Delta}) - \mathbb{E}[\widetilde{\mathbf{K}}_m(X_{i\Delta})]\} = \widehat{\Psi}_m - \Psi_m, \\ \widetilde{\mathbf{S}}_{m,1} &= \frac{1}{p_n} \sum_{\ell=0}^{p_n-1} \frac{1}{q_n} \sum_{r=1}^{q_n} \{\widetilde{\mathbf{K}}_m(X_{(2\ell q_n+r)\Delta}) - \mathbb{E}(\widetilde{\mathbf{K}}_m(X_{(2\ell q_n+r)\Delta}))\} \end{aligned}$$

is built with the  $\vec{X}_{\ell,1}$  and  $\widetilde{\mathbf{S}}_{m,2}$  is analogously defined with the  $\vec{X}_{\ell,2}$ . As above,

$$(30) \quad \mathbb{P}\left[\|\Psi_m - \widehat{\Psi}_m\|_{\text{op}} \geq u\right] \cap \Omega^* \leq \mathbb{P}\left[\|\widetilde{\mathbf{S}}_{m,1}^*\|_{\text{op}} \geq u\right] + \mathbb{P}\left[\|\widetilde{\mathbf{S}}_{m,2}^*\|_{\text{op}} \geq u\right],$$

where  $\tilde{\mathbf{S}}_{m,\delta}^*$ , for  $\delta = 1, 2$  are built on the  $\vec{X}_{\ell,\delta}^*$ . We treat only the first term applying Tropp's result to  $\tilde{\mathbf{S}}_{m,1}^*$  which is a sum of  $p_n$  independent matrices. It is clear that for all  $\ell, r$ ,  $\|\tilde{\mathbf{K}}_m(X_{(2\ell q_n+r)\Delta}^*)\|_{\text{op}} \leq L(m)$  a.s. and thus

$$\frac{1}{p_n q_n} \left\| \sum_{r=1}^{q_n} \tilde{\mathbf{K}}_m(X_{(2\ell q_n+r)\Delta}^*) - \mathbb{E}(\tilde{\mathbf{K}}_m(X_{(2\ell q_n+r)\Delta}^*)) \right\|_{\text{op}} \leq 2 \frac{L(m)}{p_n} = \frac{24}{\theta} \frac{L(m) \log(n\Delta)}{n\Delta} := \mathbf{L}.$$

Now, we must bound the variance of  $\mathbf{S}_{m,1}^*$ . We have

$$\nu(\tilde{\mathbf{S}}_{m,1}^*) = \frac{1}{p_n} \sup_{\|\vec{x}\|_{2,m}=1} \mathbb{E} \left( \frac{1}{q_n^2} \left\| \left[ \sum_{r=1}^{q_n} (\tilde{\mathbf{K}}_m(X_{r\Delta}^*) - \mathbb{E}(\tilde{\mathbf{K}}_m(X_{r\Delta}^*))) \right] \vec{x} \right\|_{2,m}^2 \right)$$

Next,

$$\begin{aligned} \tilde{\mathbb{E}}_1 &= \mathbb{E} \left( \frac{1}{q_n^2} \left\| \left[ \sum_{r=1}^{q_n} (\tilde{\mathbf{K}}_m(X_{r\Delta}^*) - \mathbb{E}(\tilde{\mathbf{K}}_m(X_{r\Delta}^*))) \right] \vec{x} \right\|_{2,m}^2 \right) \\ &= \frac{1}{q_n^2} \sum_{j=0}^{m-1} \text{Var} \left[ \sum_{r=1}^{q_n} \sum_{k=0}^{m-1} \varphi_j(X_{r\Delta}) \varphi_k(X_{r\Delta}) x_k \right] \\ &= \frac{1}{q_n} \sum_{j=0}^{m-1} \sum_{r=1}^{q_n} \text{Var} \left[ \sum_{k=0}^{m-1} \varphi_j(X_{r\Delta}) \varphi_k(X_{r\Delta}) x_k \right] = \sum_{j=0}^{m-1} \text{Var} \left[ \sum_{k=0}^{m-1} \varphi_j(X_{r\Delta}) \varphi_k(X_{r\Delta}) x_k \right] \end{aligned}$$

Therefore, for  $\|x\|_{2,m} = 1$ ,

$$\begin{aligned} \tilde{\mathbb{E}}_1 &\leq \sum_{j=0}^{m-1} \mathbb{E} \left[ \left( \sum_{k=0}^{m-1} \varphi_j(X_{r\Delta}) \varphi_k(X_{r\Delta}) x_k \right)^2 \right] \leq L(m) \mathbb{E} \left[ \left( \sum_{k=0}^{m-1} \varphi_k(X_{r\Delta}) x_k \right)^2 \right] \\ &= L(m) \int \left( \sum_{k=0}^{m-1} \varphi_k(u) x_k \right)^2 \pi(u) du \leq L(m) \|\pi\|_{\infty}. \end{aligned}$$

Thus,  $\nu(\tilde{\mathbf{S}}_{m,1}^*) \leq L(m) \|\pi\|_{\infty}/p_n = (12/\theta)(L(m) \log(n\Delta)/n\Delta)$  using the definition of  $q_n$ . Finally, applying Theorem 6.1 and joining the analogous of (26), (27), (30), the value of  $\mathbf{L}$  and the bound on  $\nu(\tilde{\mathbf{S}}_{m,1}^*)$  gives (ii).  $\square$

**6.3. Proof of Proposition 3.1.** We define the sets

$$\Lambda_m = \left\{ L(m)(\|\hat{\Psi}_m^{-1}\|_{\text{op}} \vee 1) \leq \mathfrak{c} \frac{n\Delta}{\log^2(n\Delta)} \right\} \text{ and } \Omega_m := \Omega_m(1/2) = \left\{ \left| \frac{\|t\|_n^2}{\|t\|_{\pi}^2} - 1 \right| \leq \frac{1}{2}, \forall t \in S_m \right\}.$$

Below, we prove the following lemma

**Lemma 6.1.** *Under the assumptions of Proposition 3.1, for  $m$  satisfying (13), we have*

$$\mathbb{P}(\Lambda_m^c) \leq c/(n\Delta)^5, \quad \mathbb{P}(\Omega_m^c) \leq c/(n\Delta)^5$$

where  $c$  is a positive constant, for any  $\mathfrak{c} \leq \frac{\theta(1-\log 2)}{6 \times 12}$ .

Now, we write

$$\begin{aligned}
 \|\tilde{b}_m - b_A\|_n^2 &= \|\hat{b}_m - b_A\|_n^2 \mathbf{1}_{\Lambda_m} + \|b_A\|_n^2 \mathbf{1}_{\Lambda_m^c} \\
 (31) \quad &= \|\hat{b}_m - b_A\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} + \|\hat{b}_m - b_A\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m^c} + \|b_A\|_n^2 \mathbf{1}_{\Lambda_m^c}.
 \end{aligned}$$

We bound successively the expectation of the three terms.

• Study of  $\mathbb{E}(\|\hat{b}_m - b_A\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m})$ .

The following equality holds, for all functions  $s, t$  in  $S_m$ :

$$(32) \quad \gamma_n(t) - \gamma_n(s) = \|t - b\|_n^2 - \|s - b\|_n^2 - 2\nu_n(t - s) - 2R_{n,1}(t - s) - 2R_{n,2}(t - s).$$

Thus  $\gamma_n(\hat{b}_m) \leq \gamma_n(b_m)$ , for any function  $b_m$  in  $S_m$ , implies

$$(33) \quad \|\hat{b}_m - b\|_n^2 \leq \|b_m - b\|_n^2 + 2\nu_n(\hat{b}_m - b_m) + 2R_{n,1}(\hat{b}_m - b_m) + 2R_{n,2}(\hat{b}_m - b_m).$$

We first study the last term. In the same way as Comte *et al.* (2007) p.531, we have

$$2R_{n,2}(\hat{b}_m - b_m) \leq \frac{1}{8}\|\hat{b}_m - b_m\|_n^2 + \frac{8}{n\Delta^2} \sum_{k=1}^n \left( \int_{k\Delta}^{(k+1)\Delta} (b(X_s) - b(X_{k\Delta})) ds \right)^2$$

Now, using (5) for  $f = b$ , we get

$$\begin{aligned}
 \mathbb{E} \left( \int_{k\Delta}^{(k+1)\Delta} (b(X_s) - b(X_{k\Delta})) ds \right)^2 &\leq \Delta \int_{k\Delta}^{(k+1)\Delta} \mathbb{E}[(b(X_s) - b(X_{k\Delta}))^2] ds \\
 &\leq c\Delta^3(1 + \mathbb{E}\eta^2).
 \end{aligned}$$

Thus, we have:

$$\begin{aligned}
 2\mathbb{E} \left[ R_{n,2}(\hat{b}_m - b_m) \mathbf{1}_{\Lambda_m \cap \Omega_m} \right] &\leq \frac{1}{8}\mathbb{E}[\|\hat{b}_m - b_m\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}] + 8c'\Delta \\
 (34) \quad &\leq \frac{1}{4}\mathbb{E}[\|\hat{b}_m - b\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}] + \frac{1}{4}\mathbb{E}[\|b - b_m\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}] + 8c'\Delta.
 \end{aligned}$$

For the two other terms to study in (33), we write

$$\begin{aligned}
 &2\mathbb{E} \left[ (\nu_n(\hat{b}_m - b_m) + R_{n,1}(\hat{b}_m - b_m)) \mathbf{1}_{\Lambda_m \cap \Omega_m} \right] \\
 &\leq \mathbb{E} \left[ \frac{1}{8}\|\hat{b}_m - b_m\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} + 8 \sup_{t \in S_m, \|t\|_\pi=1} [\nu_n(t) + R_{n,1}(t)]^2 \right]
 \end{aligned}$$

Then, by the definition of  $\Omega_m$ , we have

$$\begin{aligned}
 \mathbb{E} \left[ \frac{1}{8}\|\hat{b}_m - b_m\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} \right] &\leq \mathbb{E} \left[ \frac{1}{4}\|\hat{b}_m - b_m\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} \right] \\
 (35) \quad &\leq \mathbb{E} \left[ \frac{1}{2}\|\hat{b}_m - b\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} \right] + \mathbb{E} \left[ \frac{1}{2}\|b - b_m\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} \right].
 \end{aligned}$$

Moreover

$$8\mathbb{E} \left( \sup_{t \in S_m, \|t\|_\pi=1} [\nu_n(t) + R_{n,1}(t)]^2 \right) \leq 16\mathbb{E} \left( \sup_{t \in S_m, \|t\|_\pi=1} \nu_n^2(t) + \sup_{t \in S_m, \|t\|_\pi=1} R_{n,1}^2(t) \right).$$

Now the following result holds.

**Lemma 6.2.** *Under the assumptions of Proposition 3.1, we have*

$$(36) \quad \mathbb{E} \left( \sup_{t \in S_m, \|t\|_\pi = 1} \nu_n^2(t) \right) = \frac{1}{n\Delta} \text{Tr}[\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}],$$

$$(37) \quad \mathbb{E} \left( \sup_{t \in S_m, \|t\|_\pi = 1} R_{n,1}^2(t) \right) \leq C\Delta.$$

Therefore, gathering (33), (34), (35), (36) and (37) we get, for  $m$  satisfying (13),

$$\frac{1}{4} \mathbb{E}[\|\hat{b}_m - b\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}] \leq \frac{7}{4} \mathbb{E}[\|b_m - b\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}] + \frac{16}{n\Delta} \text{Tr}[\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}] + c_1 \Delta.$$

- Study of  $\mathbb{E}(\|b_A\|_n^2 \mathbf{1}_{\Lambda_m^c})$ .

We use Lemma 6.1. By the Cauchy-Schwarz inequality,  $\mathbb{E}(\|b_A\|_n^2 \mathbf{1}_{\Lambda_m^c}) \leq c \mathbb{E}^{1/2}[b^4(X_0)]/(n\Delta)^2$  with the bound on  $\mathbb{P}(\Lambda_m^c)$  given in Lemma 6.1.

- Study of  $\mathbb{E}(\|\hat{b}_m - b_A\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m^c})$ .

We introduce the operator  $\Pi_m$  of orthogonal projection, for the scalar product of  $\mathbb{R}^n$ , onto the subspace  $\{(t(X_\Delta), \dots, t(X_{n\Delta}), t \in S_m\}$  of  $\mathbb{R}^n$  and denote by  $\Pi_m b_A$  the projection of  $\vec{b}_A = (b_A(X_\Delta), \dots, b_A(X_{n\Delta}))'$ . We can write:

$$(38) \quad \|\hat{b}_m - b_A\|_n^2 = \|\hat{b}_m - \Pi_m b_A\|_n^2 + \|\Pi_m b_A - b_A\|_n^2 \leq \|\hat{b}_m - \Pi_m b_A\|_n^2 + \|b_A\|_n^2.$$

Recall that (see (6))  $Y_{k\Delta} = b(X_{k\Delta}) + E_{k\Delta}$  with  $E_{k\Delta} = Z_{k\Delta} + R_{k\Delta}$ . Elementary computations yield:

$$\Pi_m b_A = \left( \sum_{j=0}^{m-1} \tilde{a}_j^{(m)} \varphi_j(X_{k\Delta}), k = 1, \dots, n \right)' \quad \text{with} \quad \vec{\tilde{a}}^{(m)} = \frac{1}{n} \hat{\Psi}_m^{-1} \hat{\Phi}_m \vec{b}_A,$$

while  $\hat{b}_m = (\sum_{j=0}^{m-1} \hat{a}_j^{(m)} \varphi_j(X_{k\Delta}), k = 1, \dots, n)'$  with  $\vec{\hat{a}}^{(m)} = \frac{1}{n} \hat{\Psi}_m^{-1} \hat{\Phi}_m \vec{Y}$ . Therefore, setting  $\vec{E} = (R_\Delta + Z_\Delta, \dots, R_{n\Delta} + Z_{n\Delta})'$ , we have

$$\begin{aligned} \|\hat{b}_m - \Pi_m b_A\|_n^2 &= \frac{1}{n} \sum_{i=1}^n \left( \sum_{j=0}^{m-1} (\hat{a}_j^{(m)} - \tilde{a}_j^{(m)}) \varphi_j(X_{i\Delta}) \right)^2 = (\vec{\hat{a}}^{(m)} - \vec{\tilde{a}}^{(m)})' \hat{\Psi}_m (\vec{\hat{a}}^{(m)} - \vec{\tilde{a}}^{(m)}) \\ &= \frac{1}{n^2} \vec{E}' \hat{\Phi}_m \hat{\Psi}_m^{-1} (\hat{\Phi}_m)' \vec{E} \leq \|\hat{\Psi}_m^{-1}\|_{\text{op}} \|(\hat{\Phi}_m)' \vec{E}/n\|_{2,m}^2. \end{aligned}$$

On  $\Lambda_m$ ,  $\|\hat{\Psi}_m^{-1}\|_{\text{op}} \leq \mathfrak{c} n \Delta / L(m) \log^2(n\Delta)$ . Consequently,

$$\mathbb{E}[\|\hat{b}_m - \Pi_m b_A\|_n^2 \mathbf{1}_{\Lambda_m \cap \Omega_m^c}] \leq \mathfrak{c} \frac{n\Delta}{L(m) \log^2(n\Delta)} \mathbb{E}^{1/2}(\|(\hat{\Phi}_m)' \vec{E}/n\|_{2,m}^4) \mathbb{P}^{1/2}(\Omega_m^c)$$

We can prove the following Lemma:

**Lemma 6.3.** *Under the assumptions of Proposition 3.1, we have for  $C > 0$  a constant,*

$$\mathbb{E}(\|(\hat{\Phi}_m)' \vec{E}/n\|_{2,m}^4) \leq C m L^2(m) \left( \frac{\Delta^2}{n^3} + \frac{1}{(n\Delta)^2} \right).$$



Using Lemma 6.3, the bound on  $\mathbb{P}(\Omega_m^c)$  given in Lemma 6.1 and the fact that  $m \leq n\Delta$ , we get  $\mathbb{E}(\|\hat{b}_m - b_A\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m^c}) \leq c/(n\Delta)$  for a constant  $c > 0$ . Gathering the three bounds and plugging them in (31) implies the first result of result of Proposition 3.1.

To get the result in  $\mathbb{L}^2(\pi)$ -norm, we write analogously:

$$(39) \quad \|\tilde{b}_m - b_A\|_\pi^2 = \|\hat{b}_m - b_A\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} + \|\hat{b}_m - b_A\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m^c} + \|b_A\|_\pi^2 \mathbf{1}_{\Lambda_m^c}.$$

For any  $t \in S_m$ , we have using  $(x + y)^2 \leq (1 + 1/\theta)x^2 + (1 + \theta)y^2$  with  $\theta = 4$ ,

$$\begin{aligned} \|\hat{b}_m - b_A\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} &\leq \frac{5}{4} \|\hat{b}_m - t\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} + 5 \|t - b_A\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} \\ &\leq \frac{5}{2} \|\hat{b}_m - t\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} + 5 \|t - b_A\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}, \end{aligned}$$

by using the definition of  $\Omega_m$ . We insert  $b_A$  again and get:

$$\|\hat{b}_m - b_A\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} \leq 5 \|\hat{b}_m - b_A\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} + 5 \|b_A - t\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} + 5 \|t - b_A\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}.$$

Therefore taking expectation and applying the first result of Proposition 3.1 yield

$$\mathbb{E} \left( \|\hat{b}_m - b_A\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} \right) \leq 45 \inf_{t \in S_m} (\|t - b_A\|_\pi^2) + 5 \times 64 \frac{\text{Tr}[\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}]}{n\Delta}.$$

Next, we study  $\mathbb{E} \left( \|\hat{b}_m - b_A\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m^c} \right) + \mathbb{E}(\|b_A\|_\pi^2 \mathbf{1}_{\Lambda_m^c})$ . For the first term, we write that  $\|\hat{b}_m - b_A\|_\pi^2 \leq 2(\|\hat{b}_m\|_\pi^2 + \|b_A\|_\pi^2)$  and

$$\|\hat{b}_m\|_\pi^2 = \int \left( \sum_{j=0}^{\hat{m}-1} \hat{a}_j \varphi_j(x) \right)^2 \pi(x) dx = (\vec{\hat{a}}^{(m)})' \Psi_m \vec{\hat{a}}^{(m)} \leq \|\Psi_m\|_{\text{op}} \|\vec{\hat{a}}^{(m)}\|_{2,m}^2.$$

First, under  $\|\sum_{j=0}^m \varphi_j^2\|_\infty \leq L(m)$ , we get

$$\begin{aligned} \|\Psi_m\|_{\text{op}} &= \sup_{\|\vec{x}\|_{2,m}=1} \vec{x}' \Psi_m \vec{x} = \sup_{\|\vec{x}\|_{2,m}=1} \int \left( \sum_{j=0}^{m-1} x_j \varphi_j(u) \right)^2 \pi(u) du \\ &\leq \sup_{\|\vec{x}\|_{2,m}=1} \int \left( \sum_{j=0}^{m-1} x_j^2 \sum_{j=0}^{m-1} \varphi_j^2(u) \right) \pi(u) du \leq L(m). \end{aligned}$$

Next,  $\|\vec{\hat{a}}^{(m)}\|_{2,m}^2 = (1/n^2) \|\hat{\Psi}_m^{-1} \hat{\Phi}_m' \vec{Y}\|_{2,m}^2 \leq (1/n^2) \|\hat{\Psi}_m^{-1} \hat{\Phi}_m'\|_{\text{op}}^2 \|\vec{Y}\|_{2,n}^2$  and

$$\|\hat{\Psi}_m^{-1} \hat{\Phi}_m'\|_{\text{op}}^2 = \lambda_{\max} \left( \hat{\Psi}_m^{-1} \hat{\Phi}_m' \hat{\Phi}_m \hat{\Psi}_m^{-1} \right) = n \lambda_{\max}(\hat{\Psi}_m^{-1}) = n \|\hat{\Psi}_m^{-1}\|_{\text{op}}$$

Therefore, on  $\Lambda_m$  we have  $L(m)(\|\hat{\Psi}_m^{-1}\|_{\text{op}} \vee 1) \leq cn\Delta$ , and thus

$$\|\hat{b}_m\|_\pi^2 \leq \frac{L(m) \|\hat{\Psi}_m^{-1}\|_{\text{op}}}{n} \left( \sum_{i=1}^n Y_{i\Delta}^2 \right) \leq C\Delta \left( \sum_{i=1}^n Y_{i\Delta}^2 \right).$$

Then as  $\mathbb{E}[(\sum_{i=1}^n Y_{i\Delta}^2)^2] \leq n^2 \mathbb{E}(Y_\Delta^4)$ , we get

$$\mathbb{E}(\|\hat{b}_m\|_\pi^2 \mathbf{1}_{\Lambda_m \cap \Omega_m^c}) \leq \sqrt{\mathbb{E}(\|\hat{b}_m\|_\pi^4) \mathbb{P}(\Omega_m^c)} \leq C \mathbb{E}^{1/2}(Y_\Delta^4) n \Delta \mathbb{P}^{1/2}(\Omega_m^c) \leq c'/(n\Delta)$$

as  $\mathbb{E}[Y_\Delta^4] \leq c\Delta^2$ . On the other hand  $\mathbb{E}(\|b_A\|_\pi^2 \mathbf{1}_{\Omega_m^c}) \leq \|b_A\|_\pi^2 \mathbb{P}(\Omega_m^c) \leq c''/(n\Delta)^5$ . Thus  $\mathbb{E}(\|\hat{b}_m - b_A\|_\pi^2 \mathbf{1}_{\Omega_m^c}) \leq c_1/(n\Delta)$ . Joining the bounds for the three terms of (39) ends the proof of the second Inequality of Proposition 3.1.  $\square$

**6.4. Proof of Lemma 6.1.** We use Proposition 6.1, (i) for  $u = 1/2$  and condition (10) to get that

$$\mathbb{P}(\Omega_m^c) = \mathbb{P}(\Omega_m(1/2)^c) \leq 4n\Delta \exp \left[ -\frac{\theta(1 - \log 2) \log n\Delta}{24\mathfrak{c}/2} \right] + \frac{\theta}{6(n\Delta)^5} \leq \frac{C}{(n\Delta)^5}$$

for  $\mathfrak{c} \leq (\theta(1 - \log 2))/(6 \times 12)$ . Besides, in Comte and Genon-Catalot (2018), Lemma 6.3, it is proved that  $\mathbb{P}(\Lambda_m^c) \leq \mathbb{P}(\Omega_m^c) = \mathbb{P}[\|\Psi_m^{-1/2} \hat{\Psi}_m \Psi_m^{-1/2} - \text{Id}_m\|_{\text{op}} > 1/2]$ . The proof of Lemma 6.1 is thus complete.  $\square$

**6.5. Proof of Lemma 6.2.** • Proof of equality (36).

Write that

$$\sup_{t \in S_m, \|t\|_\pi=1} \nu_n^2(t) = \sup_{t \in S_m, \|\Psi_m^{1/2} \vec{a}\|_{2,m}=1, t=\sum_{j=0}^{m-1} a_j \varphi_j} \langle \vec{\sigma \hat{\varepsilon}}, t \rangle_n^2$$

where  $\vec{\sigma \hat{\varepsilon}}$  is the  $n$ -dimensional vector with coordinates  $\sigma(X_{i\Delta})\varepsilon_i$ ,  $i = 1, \dots, n$  and  $\varepsilon_i = (W_{(i+1)\Delta} - W_{i\Delta})/\Delta$ . Let  $t = \sum_{j=0}^{m-1} a_j \varphi_j$  where  $\vec{a} = \Psi_m^{-1/2} \vec{u}$ , that is  $a_j = \sum_{k=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} u_k$  and  $\|\vec{u}\|_{2,m} = 1$ . Then  $t = \sum_{k=0}^{m-1} u_k (\sum_{j=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} \varphi_j)$ . By the Cauchy-Schwarz inequality, we have  $\langle \vec{\sigma \hat{\varepsilon}}, t \rangle_n^2 \leq \sum_{k=0}^{m-1} \langle \vec{\sigma \hat{\varepsilon}}, \sum_{j=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} \varphi_j \rangle_n^2$ , and more precisely,

$$\sup_{t \in S_m, \|\Psi_m^{1/2} \vec{a}\|_{2,m}=1} \langle \vec{\sigma \hat{\varepsilon}}, t \rangle_n^2 = \sum_{k=0}^{m-1} \langle \vec{\sigma \hat{\varepsilon}}, \sum_{j=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} \varphi_j \rangle_n^2.$$

Therefore,

$$\mathbb{E} \left( \sup_{t \in S_m, \|\Psi_m^{1/2} \vec{a}\|_{2,m}=1} \langle \vec{\sigma \hat{\varepsilon}}, t \rangle_n^2 \right) = \sum_{k=0}^{m-1} \mathbb{E} \left( \langle \vec{\sigma \hat{\varepsilon}}, \sum_{j=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} \varphi_j \rangle_n^2 \right).$$

Then using that, for any function  $\psi$  bounded,  $\mathbb{E}[\varepsilon_i \varepsilon_k \psi(X_{i\Delta}) \psi(X_{k\Delta})]$  is equal to 0 if  $i \neq k$  and to  $\mathbb{E}(\psi^2(X_0))/\Delta$  if  $i = k$ , we get

$$\begin{aligned} \mathbb{E} \left( \langle \vec{\sigma \hat{\varepsilon}}, \sum_{j=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} \varphi_j \rangle_n^2 \right) &= \frac{1}{n\Delta} \mathbb{E} \left[ \sigma^2(X_0) \left( \sum_{j=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} \varphi_j(X_0) \right)^2 \right] \\ &= \frac{1}{n\Delta} \sum_{0 \leq j, \ell \leq m-1} [\Psi_m^{-1/2}]_{j,k} [\Psi_m^{-1/2}]_{\ell,k} [\Psi_{m,\sigma^2}]_{j,\ell}. \end{aligned}$$

We thus obtain equality (36) as

$$\begin{aligned} \mathbb{E} \left( \sup_{t \in S_m, \|\Psi_m^{1/2} \vec{a}\|_{2,m}=1} \langle \vec{\sigma \hat{\varepsilon}}, t \rangle_n^2 \right) &= \frac{1}{n\Delta} \sum_{0 \leq j, k, \ell \leq m-1} [\Psi_m^{-1/2}]_{j,k} [\Psi_m^{-1/2}]_{\ell,k} [\Psi_{m,\sigma^2}]_{j,\ell} \\ &= \frac{1}{n\Delta} \text{Tr} \left[ \Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2} \right]. \end{aligned}$$

• Proof of bound (37). Now write

$$\sup_{t \in S_m, \|t\|_\pi=1} R_{n,1}^2(t) = \sup_{t \in S_m, \|\Psi_m^{1/2} \vec{a}\|_{2,m}=1, t=\sum_{j=0}^{m-1} a_j \varphi_j} R_{n,1}^2(t).$$

Let  $t = \sum_{j=0}^{m-1} a_j \varphi_j$  where  $\vec{a} = \Psi_m^{-1/2} \vec{u}$ ,  $\|\vec{u}\|_{2,m} = 1$ . Then (as above):

$$t = \sum_{k=0}^{m-1} u_k \left( \sum_{j=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} \varphi_j \right) \quad \text{and} \quad R_{n,1}^2(t) \leq \sum_{k=0}^{m-1} R_{n,1}^2 \left( \sum_{j=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} \varphi_j \right)$$

Therefore,

$$\mathbb{E} \left( \sup_{t \in S_m, \|\Psi_m^{1/2} \vec{a}\|_{2,m}=1} R_{n,1}^2(t) \right) \leq \sum_{k=0}^{m-1} \mathbb{E} \left[ R_{n,1}^2 \left( \sum_{j=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} \varphi_j \right) \right].$$

Then using that the  $\int_{i\Delta}^{(i+1)\Delta} (\sigma(X_s) - \sigma(X_{i\Delta})) dW_s$  and  $X_{j\Delta}, j \leq i$  are uncorrelated and the terms are centered, and that the process  $(X_t)$  is stationary, we get

$$\mathbb{E} \left[ R_{n,1}^2 \left( \sum_{j=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} \varphi_j \right) \right] = \frac{1}{n} \mathbb{E} \left[ R_{\Delta,1}^2 \left( \sum_{j=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} \varphi_j(X_\Delta) \right)^2 \right].$$

Therefore,

$$\begin{aligned} \mathbb{E} \left( \sup_{t \in S_m, \|t\|_\pi=1} R_{n,1}^2(t) \right) &\leq \frac{1}{n} \sum_{k=0}^{m-1} \sum_{0 \leq j, \ell \leq m-1} [\Psi_m^{-1/2}]_{j,k} [\Psi_m^{-1/2}]_{\ell,k} \mathbb{E} (\varphi_j(X_\Delta) \varphi_\ell(X_\Delta) R_{\Delta,1}^2) \\ &\leq \frac{1}{n} \sum_{0 \leq j, \ell \leq m-1} \left( \sum_{k=0}^{m-1} [\Psi_m^{-1/2}]_{j,k} [\Psi_m^{-1/2}]_{\ell,k} \right) \mathbb{E} (\varphi_j(X_\Delta) \varphi_\ell(X_\Delta) R_{\Delta,1}^2) \\ &= \frac{1}{n} \sum_{0 \leq j, \ell \leq m-1} [\Psi_m^{-1}]_{j,\ell} \mathbb{E} (\varphi_j(X_\Delta) \varphi_\ell(X_\Delta) R_{\Delta,1}^2) \\ &\leq \frac{L(m) \|\Psi_m^{-1}\|_{\text{op}}}{n \Delta^2} \mathbb{E} \left[ \left( \int_{\Delta}^{2\Delta} (\sigma(X_s) - \sigma(X_\Delta)) dW_s \right)^2 \right] \leq C \frac{L(m) \|\Psi_m^{-1}\|_{\text{op}}}{n} \end{aligned}$$

since by (5) for  $f = \sigma$ ,  $\mathbb{E} \left[ \int_{\Delta}^{2\Delta} (\sigma(X_s) - \sigma(X_\Delta))^2 ds \right] \leq c \Delta^2 (1 + \mathbb{E} \eta^2)$ . Now, for  $m$  satisfying (13),  $L(m) \|\Psi_m^{-1}\|_{\text{op}} \leq cn \Delta / 2 \log^2(n \Delta) \leq cn \Delta$ . Thus,

$$\mathbb{E} \left( \sup_{t \in S_m, \|t\|_\pi=1} R_{n,1}^2(t) \right) \leq C' \Delta.$$

This ends the proof of (37) and of Lemma 6.2.  $\square$

**6.6. Proof of Lemma 6.3.** Set

$$N^2 = \|n^{-1} \widehat{\Phi}'_m \vec{E}\|_{2,m}^2 = n^{-2} \sum_{j=0}^{m-1} \left( \sum_{i=1}^n \varphi_j(X_{i\Delta}) E_{i\Delta} \right)^2$$

where  $E_{i\Delta} = \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} (b(X_s) - b(X_{i\Delta})) ds + \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} \sigma(X_s) dW_s$ . We can write  $\sum_{i=1}^n \varphi_j(X_{i\Delta}) E_{i\Delta} = \int_0^{(n+1)\Delta} H_s^{(j)} ds + \int_0^{(n+1)\Delta} K_s^{(j)} dW_s$  with

$$H_s^{(j)} = \frac{1}{\Delta} \sum_{i=1}^n \mathbf{1}_{[i\Delta, (i+1)\Delta]}(s) \varphi_j(X_{i\Delta}) (b(X_s) - b(X_{i\Delta})), \quad K_s^{(j)} = \frac{1}{\Delta} \sum_{i=1}^n \mathbf{1}_{[i\Delta, (i+1)\Delta]}(s) \varphi_j(X_{i\Delta}) \sigma(X_s).$$

Therefore,

$$\begin{aligned} N^4 &= \frac{1}{(n\Delta)^4} \left[ \sum_{j=0}^{m-1} \left( \int_0^{(n+1)\Delta} H_s^{(j)} ds + \int_0^{(n+1)\Delta} K_s^{(j)} dW_s \right)^2 \right]^2 \\ &\leq 8 \frac{m}{(n\Delta)^4} \sum_{j=0}^{m-1} \left[ \left( \int_0^{(n+1)\Delta} H_s^{(j)} ds \right)^4 + \left( \int_0^{(n+1)\Delta} K_s^{(j)} dW_s \right)^4 \right] \\ &\leq 8 \frac{m}{n^4} \sum_{j=0}^{m-1} \left[ \frac{1}{\Delta} \int_0^{(n+1)\Delta} (H_s^{(j)})^4 ds + \frac{1}{\Delta^4} \left( \int_0^{(n+1)\Delta} K_s^{(j)} dW_s \right)^4 \right]. \end{aligned}$$

We bound successively the expectation of the two terms. By (10),  $\sum_{j=0}^{m-1} \varphi_j^4 \leq L^2(m)$ , so:

$$\begin{aligned} \frac{1}{\Delta} \mathbb{E} \sum_{j=0}^{m-1} \int_0^{(n+1)\Delta} (H_s^{(j)})^4 ds &\leq L^2(m) \frac{1}{\Delta} \mathbb{E} \sum_{i=1}^n \int_{i\Delta}^{(i+1)\Delta} (b(X_s) - b(X_{i\Delta}))^4 ds \\ &\leq cL^2(m)n\Delta^2(1 + \mathbb{E}(\eta^4)). \end{aligned}$$

Next, using the Burkholder-Davis-Gundy and the Cauchy-Schwarz inequalities yields

$$\begin{aligned} \frac{1}{\Delta^4} \mathbb{E} \sum_{j=0}^{m-1} \left( \int_0^{(n+1)\Delta} K_s^{(j)} dW_s \right)^4 &\leq \frac{c}{\Delta^4} \mathbb{E} \sum_{j=0}^{m-1} \left( \int_0^{(n+1)\Delta} (K_s^{(j)})^2 ds \right)^2 \\ &\leq \frac{c}{\Delta^4} \sum_{j=0}^{m-1} \mathbb{E} \sum_{j=0}^{m-1} n\Delta \int_{i\Delta}^{(i+1)\Delta} (K_s^{(j)})^4 ds \leq \frac{c}{\Delta^4} L^2(m)n\Delta \mathbb{E} \sum_{i=1}^n \int_{i\Delta}^{(i+1)\Delta} \sigma^4(X_s) ds \\ &\leq \frac{c}{\Delta^4} L^2(m)(n\Delta)^2 \mathbb{E} \sigma^4(\eta) \leq \frac{c'}{\Delta^4} L^2(m)(n\Delta)^2 \mathbb{E}(1 + \eta^4) \end{aligned}$$

Finally,  $\mathbb{E}N^4 \leq CmL^2(m)(\Delta^2 n^{-3} + (n\Delta)^{-2})$ , which is the result of Lemma 6.3.  $\square$

**6.7. Proof of Proposition 3.2.** (1) The result follows from equality (36) and the fact that the spaces are nested.

(2) For the second point, we use:  $\text{Tr}[\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}] \leq m \|\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}\|_{\text{op}}$ . Then,

$$\|\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}\|_{\text{op}} = \sup_{\|x\|_{2,m}=1} x' \Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2} x = \sup_{y, \|\Psi_m^{1/2} y\|_{2,m}=1} y' \Psi_{m,\sigma^2} y.$$

Now, if  $\sigma$  is bounded on  $A$ ,  $y' \Psi_{m,\sigma^2} y$  is equal to

$$\int \left( \sum_{j=0}^{m-1} y_j \varphi_j(x) \right)^2 \sigma^2(x) \pi(x) dx \leq \|\sigma_A^2\|_{\infty} \int \left( \sum_{j=0}^{m-1} y_j \varphi_j(x) \right)^2 \pi(x) dx = \|\sigma_A^2\|_{\infty} \|\Psi_m^{1/2} y\|_{2,m}^2.$$

Thus,  $\text{Tr}[\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}] \leq m\|\sigma_A^2\|_\infty$ .

(3) Let us prove the other variance bound. We recall that  $\|A\|_{\mathbf{F}}^2 = \text{Tr}(AA') = \text{Tr}(A'A)$ . Writing that  $\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2} = (\Psi_m^{-1/2}\Psi_{m,\sigma^2}^{1/2})(\Psi_m^{-1/2}\Psi_{m,\sigma^2}^{1/2})'$ , we have

$$\text{Tr} \left[ \Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2} \right] = \|\Psi_m^{-1/2}\Psi_{m,\sigma^2}^{1/2}\|_F^2$$

Using  $\|AB\|_{\mathbf{F}}^2 \leq \|A\|_{\mathbf{F}}^2\|B\|_{\text{op}}^2$ ,

$$(40) \quad \text{Tr} \left[ \Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2} \right] \leq \|\Psi_m^{-1/2}\|_{\text{op}}^2 \|\Psi_{m,\sigma^2}^{1/2}\|_F^2 = \|\Psi_m^{-1}\|_{\text{op}} \text{Tr}(\Psi_{m,\sigma^2}).$$

Lastly  $\text{Tr}(\Psi_{m,\sigma^2}) = \sum_{j=0}^{m-1} \int \varphi_j^2(x) \sigma^2(x) \pi(x) dx \leq L(m) \mathbb{E}[\sigma_A^2(X_0)]$  gives bound (3).

**6.8. Proof of Theorem 3.1.** We follow the scheme of Theorem 4.1 in Comte and Genon-Catalot (2018). But here, the variables are not independent, the function  $\sigma(\cdot)$  is unbounded and there are two other main differences:

- the penalty

$$(41) \quad \widehat{\text{pen}}(m) = \kappa c_\varphi^2 s^2 \frac{m \|\widehat{\Psi}_m^{-1}\|_{\text{op}}}{n}$$

is random and has to be compared to its deterministic counterpart,  $\text{pen}(m) = \kappa' c_\varphi^2 s^2 m \|\Psi_m^{-1}\|_{\text{op}}/n$ ,

- there are the two additional terms,  $R_{n,1}$  and  $R_{n,2}$ .

We denote by  $\widehat{M}_{n\Delta}$  the maximal element of  $\widehat{\mathcal{M}}_{n\Delta}$  defined by (16), by  $M_{n\Delta}$  the maximal element of  $\mathcal{M}_{n\Delta}$  defined by (15) and by  $M_{n\Delta}^+$  the maximal element of the set defined by

$$(42) \quad \mathcal{M}_{n\Delta}^+ = \left\{ m \in \mathbb{N}, \quad c_\varphi^2 m (\|\Psi_m^{-1}\|_{\text{op}}^2 \vee 1) \leq 4\mathfrak{d} \frac{n\Delta}{\log^2(n\Delta)} \right\}, \text{ with } \mathfrak{d} \text{ is given in (16).}$$

The value  $\widehat{M}_{n\Delta}$  is random but thanks to the constants associated with the sets, with large probability, we prove  $M_{n\Delta} \leq \widehat{M}_{n\Delta} \leq M_{n\Delta}^+$  or equivalently  $\mathcal{M}_{n\Delta} \subset \widehat{\mathcal{M}}_{n\Delta} \subset \mathcal{M}_{n\Delta}^+$ . Set

$$(43) \quad \Xi_{n\Delta} := \left\{ \mathcal{M}_{n\Delta} \subset \widehat{\mathcal{M}}_{n\Delta} \subset \mathcal{M}_{n\Delta}^+ \right\}, \quad \Omega_{n\Delta} = \bigcap_{m \in \mathcal{M}_{n\Delta}^+} \Omega_m.$$

**Lemma 6.4.** *Under the Assumptions of Theorem 3.1,  $\mathbb{P}(\Omega_{n\Delta}^c) \leq c/(n\Delta)^4$  and  $\mathbb{P}(\Xi_{n\Delta}^c) \leq c'/(n\Delta)^4$ , where  $c, c'$  are positive constants.*

We do not give a detailed proof of this Lemma. As  $4\mathfrak{d} \leq \mathfrak{c}/2$ , the first bound of Lemma 6.4 is a simple consequence of Lemma 6.1. The proof of the second bound is not immediate but quite similar to the one of Lemma 6.6 in Comte and Genon-Catalot (2018). It relies on Inequality (ii) of Proposition 6.1 and this is the only place where this inequality is applied.

Now we write the decomposition:

$$(44) \quad \begin{aligned} \hat{b}_{\hat{m}} - b_A &= (\hat{b}_{\hat{m}} - b_A) \mathbf{1}_{\Xi_{n\Delta}} + (\hat{b}_{\hat{m}} - b_A) \mathbf{1}_{\Xi_{n\Delta}^c} \\ &= (\hat{b}_{\hat{m}} - b_A) \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}} + (\hat{b}_{\hat{m}} - b_A) \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}^c} + (\hat{b}_{\hat{m}} - b_A) \mathbf{1}_{\Xi_{n\Delta}^c} \end{aligned}$$

**Lemma 6.5.** *Under the Assumptions of Theorem 3.1,  $\mathbb{E} \left[ \|\hat{b}_{\hat{m}} - b_A\|_n^4 \right] \leq c(n\Delta)^2$ .*

As in addition that  $\mathbb{P}(\Xi_{n\Delta}^c) \leq c/(n\Delta)^4$  and  $\mathbb{P}(\Omega_{n\Delta}^c) \leq c/(n\Delta)^4$ , we get

$$\mathbb{E} \left[ \|\hat{b}_{\hat{m}} - b_A\|_n^2 (\mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}^c} + \mathbf{1}_{\Xi_{n\Delta}^c}) \right] \leq \frac{c}{n\Delta}.$$

Therefore it remains to study  $\mathbb{E}(\|\hat{b}_{\hat{m}} - b_A\|_n^2 \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}})$ . We have

$$\hat{m} = \arg \min_{m \in \widehat{\mathcal{M}}_{n\Delta}} \{ \gamma_n(\hat{b}_m) + \widehat{\text{pen}}(m) \},$$

with  $\widehat{\text{pen}}(m)$  defined by (41). Thus, using the definition of the contrast, we have, for any  $m \in \widehat{\mathcal{M}}_{n\Delta}$ , and any  $b_m \in S_m$ ,

$$(45) \quad \gamma_n(\hat{b}_{\hat{m}}) + \widehat{\text{pen}}(\hat{m}) \leq \gamma_n(b_m) + \widehat{\text{pen}}(m).$$

Now, on the set  $\Xi_{n\Delta} = \left\{ \mathcal{M}_{n\Delta} \subset \widehat{\mathcal{M}}_{n\Delta} \subset \mathcal{M}_{n\Delta}^+ \right\}$ , we have in all cases that  $\hat{m} \leq \widehat{M}_{n\Delta} \leq M_{n\Delta}^+$  and either  $M_{n\Delta} \leq \hat{m} \leq \widehat{M}_{n\Delta} \leq M_{n\Delta}^+$  or  $\hat{m} < M_{n\Delta} \leq \widehat{M}_{n\Delta} \leq M_{n\Delta}^+$ . In the first case,  $\hat{m}$  is upper and lower bounded by deterministic bounds, and in the second,

$$\hat{m} = \arg \min_{m \in \mathcal{M}_{n\Delta}} \{ \gamma_n(\hat{b}_m) + \widehat{\text{pen}}(m) \}.$$

Thus, on  $\Xi_{n\Delta}$ , Inequality (45) holds for any  $m \in \mathcal{M}_{n\Delta}$  and any  $b_m \in S_m$ . With decomposition (32), it yields, for any  $m \in \mathcal{M}_{n\Delta}$  and any  $b_m \in S_m$ , on  $\Xi_{n\Delta} \cap \Omega_{n\Delta}$ ,

$$\begin{aligned} \|\hat{b}_{\hat{m}} - b\|_n^2 &\leq \|b_m - b\|_n^2 + \frac{1}{8} \|\hat{b}_{\hat{m}} - b_m\|_f^2 + 16 \sup_{t \in B_{\hat{m},m}^\pi(0,1)} \nu_n^2(t) + \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\hat{m}) \\ &\quad + 16 \sup_{t \in B_{\hat{m},m}^\pi(0,1)} R_{n,1}^2(t) \mathbf{1}_{\Xi_{n\Delta}} + 2R_{n,2}(\hat{b}_{\hat{m}} - b) \\ &\leq \left(1 + \frac{1}{2}\right) \|b_m - b\|_n^2 + \frac{1}{2} \|\hat{b}_{\hat{m}} - b\|_n^2 + 16 \left( \sup_{t \in B_{\hat{m},m}^\pi(0,1)} \nu_n^2(t) - p(m, \hat{m}) \right)_+ \\ (46) \quad &+ 16 \sup_{t \in B_{\hat{m},m}^\pi(0,1)} R_{n,1}^2(t) \mathbf{1}_{\Xi_{n\Delta}} + 2R_{n,2}(\hat{b}_{\hat{m}} - b_m) + \widehat{\text{pen}}(m) + 16p(m, \hat{m}) - \widehat{\text{pen}}(\hat{m}). \end{aligned}$$

where  $B_{m,m'}^\pi(0,1) = \{t \in S_m + S_{m'}, \|t\|_\pi = 1\}$ .

Then, for  $\hat{m}$  a random index in  $\widehat{\mathcal{M}}_{n\Delta}$ , using (37),

$$\mathbb{E} \left( \sup_{t \in S_m + S_{\hat{m}}, \|t\|_\pi = 1} R_{n,1}^2(t) \mathbf{1}_{\Xi_{n\Delta}} \right) \leq \mathbb{E} \left( \sup_{t \in S_{M_n^+}, \|t\|_\pi = 1} R_{n,1}^2(t) \right) \leq C \Delta.$$

The bound on  $R_{n,2}$  is straightforward (see the proof of Proposition 3.1, non adaptive case) and we get

$$(47) \quad \mathbb{E} \left[ R_{n,2}(\hat{b}_{\hat{m}} - b_m) \mathbf{1}_{\Xi_{n\Delta}} \right] \leq \frac{1}{8} \mathbb{E}[\|\hat{b}_{\hat{m}} - b_m\|_n^2 \mathbf{1}_{\Xi_{n\Delta}}] + 8c'\Delta.$$

The main point is the study of  $\nu_n(t)$ .

**Lemma 6.6.** *Let  $(X_{i\Delta}, i = 1, \dots, n)$  be observations from model (1) under (A1)-(A4), with basis satisfying (B1). Assume that  $\mathbb{E}\eta^6 < +\infty$ . Then there exists  $\kappa_0$  such that  $\nu_n(t)$  satisfies*

$$\mathbb{E} \left[ \left( \sup_{t \in B_{\hat{m},m}^\pi(0,1)} \nu_n^2(t) - p(m, \hat{m}) \right)_+ \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}} \right] \leq C \frac{\log^2(n\Delta)}{n\Delta}$$

where

$$p(m, m') = \kappa_0 s^2 c_\varphi^2 \frac{(m \vee m') \|\Psi_{m \vee m'}^{-1}\|_{\text{op}}}{n\Delta}, \quad s^2 = \mathbb{E}[\sigma_A^2(X_0)].$$

For  $\kappa' \geq 16\kappa_0$ ,  $16p(m, m') \leq \text{pen}(m) + \text{pen}(m')$ . Therefore, plugging the result of Lemma 6.6 and (47) in (46) and taking expectation yield that

$$\begin{aligned} \frac{1}{4} \mathbb{E}(\|\hat{b}_{\hat{m}} - b\|_n^2 \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}}) &\leq \frac{7}{4} \|b_m - b\|_n^2 + \text{pen}(m) + C \frac{\log^2(n\Delta)}{n\Delta} + C' \Delta \\ &\quad + \mathbb{E}(\widehat{\text{pen}}(m) \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}}) + \mathbb{E}[(\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m}))_+ \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}}]. \end{aligned}$$

**Lemma 6.7.** *Under the assumptions of Theorem 3.1, there exist constants  $c_1, c_2 > 0$  such that for  $m \in \mathcal{M}_{n\Delta}$  and  $\hat{m} \in \widehat{\mathcal{M}}_{n\Delta}$ ,*

$$(48) \quad \mathbb{E}(\widehat{\text{pen}}(m) \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}}) \leq c_1 \text{pen}(m) + \frac{c_2}{n\Delta}$$

$$(49) \quad \mathbb{E}[(\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m}))_+ \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}}] \leq \frac{c_2}{n\Delta}.$$

Lemma 6.7 concludes the study of the expectation of the empirical risk on  $\Xi_{n\Delta} \cap \Omega_{n\Delta}$ . The proof of (18) is now complete. For the step from the empirical norm to the  $\mathbb{L}^2(\pi)$ -norm, we proceed as in the proof of Proposition 3.1 and get (19).  $\square$

**6.9. Proof of Lemma 6.5.** We start as in (38) and equations following, for  $m$  replaced by  $\hat{m}$ . We have  $\|\hat{b}_{\hat{m}} - \Pi_{\hat{m}} b_A\|_n^2 = (1/n^2) \vec{E}' \hat{\Phi}_{\hat{m}} \hat{\Psi}_{\hat{m}}^{-1} \hat{\Phi}_{\hat{m}}' \vec{E} \leq \|\hat{\Psi}_{\hat{m}}^{-1}\|_{\text{op}} \|\hat{\Phi}_{\hat{m}}' \vec{E}/n\|_{2, \hat{m}}^2$ . Now we have  $\|\hat{\Psi}_{\hat{m}}^{-1}\|_{\text{op}} \lesssim \sqrt{n\Delta}$  as  $\hat{m} \in \widehat{\mathcal{M}}_{n\Delta}$ , and  $m \mapsto \|\hat{\Phi}_m' \vec{E}/n\|_{2, m}^2$  is increasing, so

$$\|\hat{b}_{\hat{m}} - \Pi_{\hat{m}} b_A\|_n^2 \leq \sqrt{n\Delta} \|\hat{\Phi}_{n\Delta}' \vec{E}/n\|_{2, n\Delta}^2.$$

Thus, using the bound proved in Lemma 6.3, we get  $\mathbb{E}(\|\hat{b}_{\hat{m}} - \Pi_{\hat{m}} b_A\|_n^4) \leq c(n\Delta)^2$ .  $\square$

**6.10. Proof of Lemma 6.6.** To apply the Talagrand Inequality, we make the following decompositions. Let  $u_i = W_{(i+1)\Delta} - W_{i\Delta}$ ,

$$u_i = u_i^{(1)} + u_i^{(2)}, \quad u_i^{(1)} = u_i \mathbf{1}_{|u_i| \leq k_n \sqrt{\Delta}} - \mathbb{E}[u_i \mathbf{1}_{|u_i| \leq k_n \sqrt{\Delta}}],$$

set  $\tau(x) = \sigma_A(x) \mathbf{1}_{\sigma_A^2(x) \leq \ell_n \sqrt{\Delta}}$  and  $\sigma_A(x) = \tau(x) + \theta(x)$ . We have  $\nu_n(t) = \nu_{n,1}(t) + \nu_{n,2}(t) + \nu_{n,3}(t)$ , where

$$\nu_{n,1}(t) = \frac{1}{n\Delta} \sum_{i=1}^n t(X_{i\Delta}) \tau(X_{i\Delta}) u_i^{(1)}, \quad \nu_{n,2}(t) = \frac{1}{n\Delta} \sum_{i=1}^n t(X_{i\Delta}) \theta(X_{i\Delta}) u_i^{(1)}, \text{ and}$$

$$\nu_{n,3}(t) = \frac{1}{n\Delta} \sum_{i=1}^n t(X_{i\Delta}) \sigma(X_{i\Delta}) u_i^{(2)}.$$

Then we write

$$\begin{aligned} \left( \sup_{t \in B_{\hat{m}, m}^\pi(0,1)} \nu_n^2(t) - p(m, \hat{m}) \right)_+ &\leq \left( \sup_{t \in B_{\hat{m}, m}^\pi(0,1)} 3\nu_{n,1}^2(t) - p(m, \hat{m}) \right)_+ \\ (50) \quad &\quad + 3 \sup_{t \in B_{\hat{m}, m}^\pi(0,1)} \nu_{n,2}^2(t) + 3 \sup_{t \in B_{\hat{m}, m}^\pi(0,1)} \nu_{n,3}^2(t), \end{aligned}$$

and we bound the three terms.

• First, we study the second term in (50). Recall that  $M_{n\Delta}^+ \leq 4\mathfrak{d}n\Delta/\log^2(n\Delta)$  is the dimension of the largest space of the collection  $\mathcal{M}_{n\Delta}^+$ . Then we have

$$\begin{aligned} & \mathbb{E}\left[\left(\sup_{t \in B_{m,m}^\pi(0,1)} \nu_{n,2}^2(t)\right)_+ \mathbf{1}_{\Xi_{n\Delta}}\right] \leq \|\Psi_{M_{n\Delta}^+}^{-1}\|_{\text{op}} \sum_{j=0}^{M_{n\Delta}^+-1} \mathbb{E}[\nu_{n,2}^2(\varphi_j)] \\ &= \|\Psi_{M_{n\Delta}^+}^{-1}\|_{\text{op}} \sum_{j=0}^{M_{n\Delta}^+-1} \text{Var}\left(\frac{1}{n\Delta^2} \sum_{i=1}^n u_i^{(1)} \theta(X_{i\Delta}) \varphi_j(X_{i\Delta})\right) \leq \frac{c_\varphi^2 M_{n\Delta}^+ \|\Psi_{M_{n\Delta}^+}^{-1}\|_{\text{op}}}{n\Delta^2} \mathbb{E}[(u_1^{(1)})^2] \mathbb{E}[\theta_A^2(X_0)] \end{aligned}$$

Now we use that  $\mathbb{E}[(u_1^{(1)})^2] \leq \mathbb{E}[u_1^2] = \Delta$  and that  $M_{n\Delta}^+$  is in  $\mathcal{M}_{n\Delta}^+$ , i.e.

$$c_\varphi^2 M_{n\Delta}^+ \|\Psi_{M_{n\Delta}^+}^{-1}\|_{\text{op}} = c_\varphi^2 \sqrt{M_{n\Delta}^+} \sqrt{M_{n\Delta}^+ \|\Psi_{M_{n\Delta}^+}^{-1}\|_{\text{op}}^2} \leq 4\mathfrak{d} \frac{n\Delta}{\log^2(n\Delta)}$$

and we get

$$\begin{aligned} & \mathbb{E}\left[\left(\sup_{t \in B_{m,m}^\pi(0,1)} \nu_{n,2}^2(t)\right)_+\right] \leq c_\varphi^2 \frac{1}{\log^2(n\Delta)} \mathbb{E}[\sigma_A^2(X_0) \mathbf{1}_{\sigma_A^2(X_0) > \ell_n \sqrt{\Delta}}] \leq C \frac{\mathbb{E}[|\sigma_A(X_0)|^{2+q}]}{\log^2(n\Delta) (\ell_n \sqrt{\Delta})^{q/2}} \\ &= C \mathbb{E}[|\sigma_A(X_1)|^{2+q}] \frac{\log^{q-2}(n\Delta)}{(n\Delta)^{q/4}} = C \mathbb{E}[|\sigma_A(X_1)|^6] \frac{\log^2(n\Delta)}{n\Delta}, \end{aligned}$$

by taking  $q = 4$ , and

$$(51) \quad \ell_n \sqrt{\Delta} = c \frac{\sqrt{n\Delta}}{\log^2(n\Delta)}.$$

• Let us now study the third term in (50). We have, relying on similar arguments,

$$\begin{aligned} & \mathbb{E}\left[\left(\sup_{t \in B_{m,m}^f(0,1)} \nu_{n,3}^2(t) \mathbf{1}_{\Xi_n}\right)_+\right] \leq \|\Psi_{M_{n\Delta}^+}^{-1}\|_{\text{op}} \sum_{j=0}^{M_{n\Delta}^+-1} \mathbb{E}[\nu_{n,3}^2(\varphi_j)] \\ &= \|\Psi_{M_{n\Delta}^+}^{-1}\|_{\text{op}} \sum_{j=0}^{M_{n\Delta}^+-1} \text{Var}\left(\frac{1}{n\Delta} \sum_{i=1}^n u_i^{(2)} \sigma_A(X_{i\Delta}) \varphi_j(X_{i\Delta})\right) \leq \frac{c_\varphi^2 M_{n\Delta}^+ \|\Psi_{M_{n\Delta}^+}^{-1}\|_{\text{op}}}{n\Delta^2} \mathbb{E}[\sigma_A^2(X_0)] \mathbb{E}[(u_1^{(2)})^2] \\ &\leq \frac{c_\varphi^2 \mathbb{E}[\sigma_A^2(X_0)]}{\Delta \log^2(n\Delta)} \mathbb{E}[u_1^2 \mathbf{1}_{|u_1| > k_n \sqrt{\Delta}}] \leq \frac{c_\varphi^2 \mathbb{E}[\sigma_A^2(X_0)] \mathbb{E}[u_1^6]}{\Delta \log^2(n\Delta) (k_n \sqrt{\Delta})^4} = C' \mu_6 \frac{\log^2(n\Delta)}{n\Delta}, \end{aligned}$$

where the last line follows from the Markov inequality,  $\mu_6 = 15 = \mathbb{E}[u_1^6/\Delta^3]$  (sixth moment of the standard gaussian) and the choices  $p = 4$  and

$$(52) \quad k_n = c' \frac{(n\Delta)^{1/4}}{\log(n\Delta)}.$$

• To bound the first term, we use the Talagrand inequality applied to the process  $\nu_{n,1}$ . As the variables are not independent we must split again this term into several parts.

We proceed by the coupling strategy used in the proof of Proposition 6.1, applied to  $v_i = (\eta_i, X_{i\Delta})$  which is also a  $\beta$ -mixing sequence with mixing coefficient such that  $\beta_k =$



$\beta_X(k\Delta) \leq ce^{-\theta k\Delta}$ , as in Baraud *et al.* (2001a). We denote by  $\Omega^* = \{v_i = v_i^*, i = 1, \dots, n\}$ . We still have  $\mathbb{P}((\Omega^*)^c) \leq p_n \beta_X(q_n \Delta) \leq c/(n\Delta)^4$  for  $q_n \Delta = 5 \log(n\Delta)/\theta$ .

On  $\Omega^*$ , we replace the  $v_i$  by the  $v_i^*$  and split the term between odd and even blocks. We have to bound, say

$$\mathbb{E} \left( \sup_{t \in B_{\hat{m}, m}^\pi(0,1)} (\nu_{n,1}^{*,1})^2(t) - \bar{p}(m, \hat{m}) \right)_+$$

by using Talagrand inequality applied to mean of  $p_n$  independent random variables

$$\nu_{n,1}^{*,1}(t) = \frac{1}{p_n} \sum_{\ell=0}^{p_n-1} \left( \frac{1}{2q_n \Delta} \sum_{r=1}^{q_n} u_{2\ell q_n+r}^{(1)*} \tau(X_{(2\ell q_n+r)\Delta}^*) t(X_{(2\ell q_n+r)\Delta}^*) \right).$$

Note that the random variables inside the sum in large brackets are not independent but uncorrelated.

Set  $Y_\ell = (u_{\ell,1}^{(1)*}, X_{\ell,1}^*) \in \mathbb{R}^{q_n} \times \mathbb{R}^{q_n}$ , where  $u_{\ell,1}^{(1)*} = (u_{2\ell q_n+r}^{(1)*})_{1 \leq \ell \leq q_n}$  and  $X_{\ell,1}^* = (X_{(2\ell q_n+r)\Delta}^*)_{1 \leq \ell \leq q_n}$ . Then we have  $\nu_{n,1}^{*,1}(t) = \frac{1}{p_n} \sum_{\ell=0}^{p_n-1} [f^{(t)}(Y_\ell) - \mathbb{E} f^{(t)}(Y_\ell)]$ ,  $f^{(t)} : \mathbb{R}^{q_n} \times \mathbb{R}^{q_n} \rightarrow \mathbb{R}$  with

$$f^{(t)}(z, x) = \frac{1}{2q_n \Delta} \sum_{r=1}^{q_n} z_r \mathbf{1}_{|z_r| \leq k_n \sqrt{\Delta}} t(x_r) \tau(x_r), \quad z = (z_1, \dots, z_{q_n}), x = (x_1, \dots, x_{q_n}),$$

and  $\mathcal{F} = \{f^{(t)}, t \in B_{m', m}^\pi(0,1)\}$ . Using analogous tools as above, we get

$$\begin{aligned} & \mathbb{E} \left[ \sup_{t \in B_{m', m}^\pi(0,1)} [\nu_{n,1}^{*,1}(t)]^2 \right] \leq \|\Psi_{m \vee m'}^{-1}\|_{\text{op}} \sum_{j \leq m \vee m'} \frac{1}{n\Delta^2} \text{Var} \left( u_1^{(1)} \tau(X_0) \varphi_j(X_0) \right) \\ & \leq \|\Psi_{m \vee m'}^{-1}\|_{\text{op}} \sum_{j \leq m \vee m'} \frac{1}{n\Delta^2} \mathbb{E} \left[ \left( u_1^{(1)} \tau(X_0) \varphi_j(X_0) \right)^2 \right] \\ & \leq \frac{1}{n} \mathbb{E} [u_1^2] \|\Psi_{m \vee m'}^{-1}\|_{\text{op}} \sum_{j \leq m \vee m'} \mathbb{E} [\tau^2(X_0) \varphi_j^2(X_0)] \leq \mathbb{E} [\sigma_A^2(X_0)] c_\varphi^2 \frac{(m \vee m') \|\Psi_{m \vee m'}^{-1}\|_{\text{op}}}{n\Delta} := H^2. \end{aligned}$$

Next, we have

$$\begin{aligned} & \sup_{t \in B_{m, m'}^\pi(0,1)} \text{Var} \left( \frac{1}{q_n \Delta} \sum_{r=1}^{q_n} \eta_r^* t(X_{r\Delta}^*) \sigma_A(X_{r\Delta}^*) \right) \\ & = \sup_{t \in B_{m, m'}^\pi(0,1)} \text{Var} \left( \frac{1}{q_n \Delta} \sum_{r=1}^{q_n} \eta_r t(X_{r\Delta}^*) \sigma_A(X_{r\Delta}^*) \right) = \frac{\mathbb{E}(\eta_1^2)}{q_n \Delta^2} \sup_{t \in B_{m, m'}^\pi(0,1)} \mathbb{E}(t^2(X_0) \sigma^2(X_0)) \\ & \leq \frac{\mathbb{E}(\varepsilon_1^2)}{q_n \Delta^2} \sup_{t \in B_{m, m'}^\pi(0,1)} \mathbb{E}^{1/2}[\sigma_A^4(X_0)] \mathbb{E}^{1/2}[t^2(X_0)] \|t\|_\infty \\ & \leq \frac{c_\varphi}{q_n \Delta} \mathbb{E}^{1/2}[\sigma_A^4(X_0)] \sqrt{(m \vee m') \|\Psi_{m \vee m'}^{-1}\|_{\text{op}}} := v \end{aligned}$$

$$\begin{aligned}
\text{Lastly} \quad & \sup_{t \in B_{m',m}^\pi(0,1)} \sup_{\vec{z} \in \mathbb{R}^{q_n}, \vec{x} \in \mathbb{R}^{q_n}} \left( \frac{1}{q_n \Delta} \sum_{r=1}^{q_n} |z_r| \mathbf{1}_{|z_r| \leq k_n \sqrt{\Delta}} |\sigma(x_r) \mathbf{1}_{|\sigma^2(x_r)| \leq \ell_n \sqrt{\Delta}} |t(x_r)| \right) \\
& \leq c \frac{k_n \sqrt{\ell_n}}{\Delta^{1/4}} \sup_{t \in B_{m',m}^\pi(0,1)} \sup_x |t(x)| \leq c_\varphi \frac{k_n \sqrt{\ell_n}}{\Delta^{1/4}} \sqrt{(m \vee m') \|\Psi_{m \vee m'}^{-1}\|_{\text{op}}} := M_1.
\end{aligned}$$

Therefore, by applying Talagrand Inequality:

$$\mathbb{E} \left( \sup_{t \in B_{m,m'}^\pi(0,1)} (\nu_{n,1}^{\star,1})^2(t) - 2H^2 \right)_+ \leq C_1 \left( \frac{v}{p_n} \exp(-C_2 \frac{p_n H^2}{v}) + \frac{M_1^2}{p_n^2} \exp(-C_3 \frac{p_n H}{M_1}) \right)$$

we obtain, recalling that  $2p_n q_n = n$  and  $q_n = (5/\theta)(\log(n\Delta)/\Delta)$ , and  $m^* = m \vee m'$ ,

$$\begin{aligned}
& \mathbb{E} \left( \sup_{t \in B_{m,m'}^\pi(0,1)} (\nu_{n,1}^{\star,1})^2(t) - 2H^2 \right)_+ \\
& \leq C'_1 \left( \frac{\sqrt{m^* \|\Psi_{m^*}^{-1}\|_{\text{op}}}}{p_n q_n \Delta} \exp(-C'_2 \sqrt{m^* \|\Psi_{m^*}^{-1}\|_{\text{op}}}) + \frac{k_n^2 \sqrt{\ell_n} m^* \|\Psi_{m^*}^{-1}\|_{\text{op}}}{p_n^2 \sqrt{\Delta}} \exp \left( -C'_3 \frac{p_n}{\sqrt{n\Delta} k_n \sqrt{\ell_n}} \right) \right),
\end{aligned}$$

Now we use that, for the first right-hand-side term,  $\sqrt{x} e^{-C_2 \sqrt{x}} \leq c' e^{-(C_2/2)\sqrt{x}}$  and  $c_\varphi^2 m^* \|\Psi_{m^*}^{-1}\|_{\text{op}} \leq 4\mathfrak{d}n\Delta/\log^2(n\Delta)$ . For the second right-hand-side term, we use the definition (51) and (52) of  $\ell_n$  and  $k_n$ , the value of  $q_n$  and  $c_\varphi^2 m^* \|\Psi_{m^*}^{-1}\|_{\text{op}} \leq 4\mathfrak{d}n\Delta/\log^2(n\Delta)$ . This implies

$$\mathbb{E} \left( \sup_{t \in B_{m,m'}^\pi(0,1)} [\nu_{n,1}^{\star,1}]^2(t) - \frac{1}{6} p(m, m') \right)_+ \leq \frac{C''_1}{n\Delta} \left( e^{-(C'_2/2)\sqrt{m^* \|\Psi_{m^*}^{-1}\|_{\text{op}}}} + \frac{n\sqrt{\Delta}}{\log^4(n\Delta)} e^{-C''_3 \log(n\Delta)} \right).$$

where  $p(m, m') = 12H^2$ . Next note that  $\|\Psi_m^{-1}\|_{\text{op}} \geq 1/\|\Psi_m\|_{\text{op}} \geq 1/\|\pi\|_\infty$ , and choose  $c, c'$  in the definition of  $k_n$  and  $\ell_n$  so that  $C''_3 = 2$ . This yields

$$\mathbb{E} \left( \sup_{t \in B_{m',m}^\pi(0,1)} [\nu_{n,1}^{\star,1}]^2(t) - \frac{1}{3} p(m, m') \right)_+ \leq \frac{C''_1}{n\Delta} \left( \exp(-C_4 \sqrt{m^*}) + \frac{1}{n \log^4(n\Delta)} \right).$$

By summing up all terms over  $m' \in \mathcal{M}_{n\Delta}^+$ , we deduce

$$\begin{aligned}
\mathbb{E} \left( \sup_{t \in B_{\hat{n},m}^\pi(0,1)} [\nu_{n,1}^{\star,1}]^2(t) - \frac{1}{6} p(m, \hat{m}) \right)_+ & \leq \sum_{m'} \mathbb{E} \left( \sup_{t \in B_{m',m}^\pi(0,1)} [\nu_{n,1}^{\star,1}]^2(t) - \frac{1}{6} p(m, m') \right)_+ \\
(53) \quad & \leq \frac{C}{n\Delta}.
\end{aligned}$$

It remains to bound  $\mathbb{E}[(\sup_{t \in B_{\hat{n},m}^\pi(0,1)} (\nu_{n,1})^2(t) - \bar{p}(m, \hat{m})) \mathbf{1}_{(\Omega^*)^c}]_+$ . We use the infinite norm computed to evaluate  $M_1$  and the bound on  $\mathbb{P}[(\Omega^*)^c]$ .  $\square$

**6.11. Proof of Lemma 6.7.** First write that

$$\widehat{\text{pen}}(m) \leq \kappa c_\varphi^2 s^2 \frac{m \|\hat{\Psi}_m^{-1} - \Psi_m^{-1}\|_{\text{op}}}{n\Delta} + \frac{\kappa}{\kappa'} \text{pen}(m)$$

Moreover, for  $m \in \mathcal{M}_{n\Delta}$  and on  $\Xi_{n\Delta}$  (thus  $m \in \widehat{\mathcal{M}}_{n\Delta}$ ),  $c_\varphi^2 m \|\Psi_m^{-1}\|_{\text{op}} \leq (\mathfrak{d}/4)n\Delta / \log^2(n\Delta)$  and  $c_\varphi^2 m \|\widehat{\Psi}_m^{-1}\|_{\text{op}} \leq \mathfrak{d}n\Delta / \log^2(n\Delta)$ . Thus,

$$\begin{aligned} c_\varphi^2 m \|\widehat{\Psi}_m^{-1} - \Psi_m^{-1}\|_{\text{op}} \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}} &= c_\varphi^2 m \|\widehat{\Psi}_m^{-1} - \Psi_m^{-1}\|_{\text{op}} \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}} \mathbf{1}_{\{\|\widehat{\Psi}_m^{-1} - \Psi_m^{-1}\|_{\text{op}} \leq \|\Psi_m^{-1}\|_{\text{op}}\}} \\ &\quad + c_\varphi^2 m \|\widehat{\Psi}_m^{-1} - \Psi_m^{-1}\|_{\text{op}} \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}} \mathbf{1}_{\{\|\widehat{\Psi}_m^{-1} - \Psi_m^{-1}\|_{\text{op}} > \|\Psi_m^{-1}\|_{\text{op}}\}} \\ &\leq c_\varphi^2 m \|\Psi_m^{-1}\|_{\text{op}} \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}} + \frac{5}{4} \frac{\mathfrak{d}n\Delta}{\log^2(n\Delta)} \mathbf{1}_{\{\|\widehat{\Psi}_m^{-1} - \Psi_m^{-1}\|_{\text{op}} > \|\Psi_m^{-1}\|_{\text{op}}\}} \end{aligned}$$

We obtain:

$$\mathbb{E}(\widehat{\text{pen}}(m) \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}}) \leq 2 \frac{\kappa}{\kappa'} \text{pen}(m) + \frac{5}{4} \frac{\mathfrak{d}n\Delta}{\log^2(n\Delta)} \mathbb{P}\left(\|\widehat{\Psi}_m^{-1} - \Psi_m^{-1}\|_{\text{op}} > \|\Psi_m^{-1}\|_{\text{op}}\right).$$

Now by Proposition 2.4 in Comte and Genon-Catalot (2018),

$$\mathbb{P}\left(\|\widehat{\Psi}_m^{-1} - \Psi_m^{-1}\|_{\text{op}} > \|\Psi_m^{-1}\|_{\text{op}}\right) \leq \mathbb{P}\left(\|\Psi_m^{-1/2} \widehat{\Psi}_m \Psi_m^{-1/2} - \text{Id}_m\|_{\text{op}} > \frac{1}{2}\right) \leq c/(n\Delta)^5$$

for  $m \in \mathcal{M}_{n\Delta}$ , by Lemma 6.1. This completes the proof of (48).

Now we turn to the proof of (49). Writing that  $\|\widehat{\Psi}_{\hat{m}}^{-1}\|_{\text{op}} \geq \|\Psi_{\hat{m}}^{-1}\|_{\text{op}} - \|\widehat{\Psi}_{\hat{m}}^{-1} - \Psi_{\hat{m}}^{-1}\|_{\text{op}}$ , we get

$$\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m}) \leq \kappa c_\varphi^2 s^2 \frac{\hat{m} \|\widehat{\Psi}_{\hat{m}}^{-1} - \Psi_{\hat{m}}^{-1}\|_{\text{op}}}{n\Delta} + (\kappa' - \kappa) \frac{s^2 c_\varphi^2 \hat{m} \|\Psi_{\hat{m}}^{-1}\|_{\text{op}}}{n}.$$

Next we decompose similarly to previously, with a change in the cutoff,

$$\begin{aligned} c_\varphi^2 \hat{m} \|\widehat{\Psi}_{\hat{m}}^{-1} - \Psi_{\hat{m}}^{-1}\|_{\text{op}} \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}} &= c_\varphi^2 \hat{m} \|\widehat{\Psi}_{\hat{m}}^{-1} - \Psi_{\hat{m}}^{-1}\|_{\text{op}} \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}} \mathbf{1}_{\{\|\widehat{\Psi}_{\hat{m}}^{-1} - \Psi_{\hat{m}}^{-1}\|_{\text{op}} \leq \frac{1}{2} \|\Psi_{\hat{m}}^{-1}\|_{\text{op}}\}} \\ &\quad + c_\varphi^2 \hat{m} \|\widehat{\Psi}_{\hat{m}}^{-1} - \Psi_{\hat{m}}^{-1}\|_{\text{op}} \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}} \mathbf{1}_{\{\|\widehat{\Psi}_{\hat{m}}^{-1} - \Psi_{\hat{m}}^{-1}\|_{\text{op}} > \frac{1}{2} \|\Psi_{\hat{m}}^{-1}\|_{\text{op}}\}} \\ &\leq \frac{1}{2} c_\varphi^2 \hat{m} \|\Psi_{\hat{m}}^{-1}\|_{\text{op}} \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}} + 5 \frac{\mathfrak{d}n\Delta}{\log^2(n\Delta)} \mathbf{1}_{\{\|\widehat{\Psi}_{\hat{m}}^{-1} - \Psi_{\hat{m}}^{-1}\|_{\text{op}} > \frac{1}{2} \|\Psi_{\hat{m}}^{-1}\|_{\text{op}}\}}. \end{aligned}$$

Now,  $\hat{m} \in \widehat{\mathcal{M}}_n \subset \mathcal{M}_n^+$ , implies that  $c_\varphi^2 \hat{m} \|\Psi_{\hat{m}}^{-1}\|_{\text{op}} \leq 4\mathfrak{d}n \log^2(n\Delta)$ . We get

$$\begin{aligned} (\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m}))_+ \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}} &\leq 4\kappa c_\varphi^2 s^2 \frac{\mathfrak{d}n\Delta}{\log^2(n\Delta)} \mathbf{1}_{\{\|\widehat{\Psi}_{\hat{m}}^{-1} - \Psi_{\hat{m}}^{-1}\|_{\text{op}} > \frac{1}{2} \|\Psi_{\hat{m}}^{-1}\|_{\text{op}}\}} \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}} \\ &\quad + (\kappa' - \frac{\kappa}{2}) \frac{s^2 c_\varphi^2 \hat{m} \|\Psi_{\hat{m}}^{-1}\|_{\text{op}}}{n}. \end{aligned}$$

We choose  $\kappa' - \frac{\kappa}{2} \leq 0$  that is  $\kappa \geq 2\kappa'$ , so that the last term vanishes and then obtain:

$$\begin{aligned} &\mathbb{E}[(\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m}))_+ \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}}] \\ &\leq 4\kappa c_\varphi^2 s^2 \frac{\mathfrak{d}n\Delta}{\log^2(n\Delta)} \mathbb{P}\left(\{\|\widehat{\Psi}_{\hat{m}}^{-1} - \Psi_{\hat{m}}^{-1}\|_{\text{op}} > \frac{1}{2} \|\Psi_{\hat{m}}^{-1}\|_{\text{op}}\} \cap \Xi_{n\Delta} \cap \Omega_{n\Delta}\right). \end{aligned}$$

Now

$$\mathbb{P}\left(\{\|\widehat{\Psi}_{\hat{m}}^{-1} - \Psi_{\hat{m}}^{-1}\|_{\text{op}} > \frac{1}{2} \|\Psi_{\hat{m}}^{-1}\|_{\text{op}}\} \cap \Xi_{n\Delta} \cap \Omega_{n\Delta}\right) \leq \sum_{m \in \mathcal{M}_{n\Delta}^+} \mathbb{P}\left(\|\widehat{\Psi}_m^{-1} - \Psi_m^{-1}\|_{\text{op}} > \frac{1}{2} \|\Psi_m^{-1}\|_{\text{op}}\right)$$

Then we use Proposition 2.4 (ii) in Comte and Genon-Catalot (2018), to get

$$\mathbb{P} \left( \left\{ \|\hat{\Psi}_{\hat{m}}^{-1} - \Psi_{\hat{m}}^{-1}\|_{\text{op}} > \frac{1}{2} \|\Psi_{\hat{m}}^{-1}\|_{\text{op}} \right\} \cap \Xi_{n\Delta} \cap \Omega_{n\Delta} \right) \leq \mathbb{P} \left( \|\Psi_m^{-1/2} \hat{\Psi}_m \Psi_m^{-1/2} - \text{Id}_m\|_{\text{op}} > \frac{1}{4} \right).$$

The choice of  $\mathfrak{d}$  implies that this probability is less than  $K/(n\Delta)^5$ . This leads to

$$\mathbb{E}[(\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m}))_+ \mathbf{1}_{\Xi_{n\Delta} \cap \Omega_{n\Delta}}] \leq c/(n\Delta).$$

This ends the proof of (49) and we can set  $\kappa' = \kappa/2$  and take  $\kappa \geq 2 \times 12 \times 16$ .  $\square$

**6.12. Appendix.** We recall the following result of Tropp (2015) and the Talagrand concentration inequality given in Klein and Rio (2005).

**Theorem 6.1** (Bernstein Matrix inequality). *Consider a finite sequence  $\{\mathbf{S}_k\}$  of independent, random matrices with common dimension  $d_1 \times d_2$ . Assume that  $\mathbb{E}\mathbf{S}_k = \mathbf{0}$  and  $\|\mathbf{S}_k\|_{\text{op}} \leq L$  for all  $k$ . Introduce the random matrix  $\mathbf{Z} = \sum_k \mathbf{S}_k$ . Let  $\nu(\mathbf{Z})$  be the variance statistic of the sum:  $\nu(\mathbf{Z}) = \max\{\lambda_{\max}(\mathbb{E}[\mathbf{Z}'\mathbf{Z}]), \lambda_{\max}(\mathbb{E}[\mathbf{Z}\mathbf{Z}'])\}$ . Then for all  $t \geq 0$*

$$\mathbb{P}[\|\mathbf{Z}\|_{\text{op}} \geq t] \leq (d_1 + d_2) \exp \left( -\frac{t^2/2}{\nu(\mathbf{Z}) + Lt/3} \right).$$

**Theorem 6.2.** *Consider  $n \in \mathbb{N}^*$ ,  $\mathcal{F}$  a class at most countable of measurable functions, and  $(X_i)_{i \in \{1, \dots, n\}}$  a family of real independent random variables. Define, for  $f \in \mathcal{F}$ ,  $\nu_n(f) = (1/n) \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)])$ , and assume that there are three positive constants  $M$ ,  $H$  and  $v$  such that  $\sup_{f \in \mathcal{F}} \|f\|_{\infty} \leq M$ ,  $\mathbb{E}[\sup_{f \in \mathcal{F}} |\nu_n(f)|] \leq H$ , and  $\sup_{f \in \mathcal{F}} (1/n) \sum_{i=1}^n \text{Var}(f(X_i)) \leq v$ .*

*Then for all  $\alpha > 0$ , with  $C(\alpha) = (\sqrt{1 + \alpha} - 1) \wedge 1$ , and  $b = \frac{1}{6}$ ,*

$$\mathbb{E} \left[ \left( \sup_{f \in \mathcal{F}} |\nu_n(f)|^2 - 2(1 + 2\alpha)H^2 \right)_+ \right] \leq \frac{4}{b} \left( \frac{v}{n} e^{-b\alpha \frac{nH^2}{v}} + \frac{49M^2}{bC^2(\alpha)n^2} e^{-\frac{\sqrt{2}bC(\alpha)\sqrt{\alpha}}{7} \frac{nH}{M}} \right).$$

By density arguments, this result can be extended to the case where  $\mathcal{F}$  is a unit ball of a linear normed space.

## REFERENCES

- [1] Baraud, Y. (2002) Model selection for regression on a random design. *ESAIM Probab. Statist.* **6**, 127-146.
- [2] Barron, A., Birgé, L. and Massart, P. (1999) Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113**, 301-413.
- [3] Berbee, H.C.P. (1979). Random walks with stationary increments and renewal theory. *Math. Tracts. Mathematisch Centrum*, Amsterdam, **112**.
- [4] Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329-375.
- [5] Cohen, A., Davenport, M.A. and Leviatan, D. (2013). On the stability and accuracy of least squares approximations. *Found. Comput. math.* **13**, 819-834.
- [6] Comte, F. and Genon-Catalot, V. (2018). Regression function estimation on non compact support as a partly inverse problem. Preprint hal-01690856v3.
- [7] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* **13**, 514-543.
- [8] Dalalyan, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Statist.* **33**(6), 2507-2528.
- [9] Efromovich, S. (1999) *Nonparametric curve estimation. Methods, theory, and applications*. Springer Series in Statistics. Springer-Verlag, New York.

- [10] Gobet, E., Hoffmann, M. and Reiss, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32**, 2223–2253.
- [11] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.* **79**, 135–163.
- [12] Kessler, M., Lindner, A. and Sørensen, M., Editors (2012). *Statistical methods for stochastic differential equations*. CRC Press, Taylor & Francis Group, Boca Raton.
- [13] Klein, T. and Rio, E. (2005) Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33**, no. 3, 1060–1077.
- [14] Kutoyants, Y.A. (2004). Statistical inference for ergodic diffusion processes. Springer Series in Statistics. Springer-Verlag London, Ltd., London.
- [15] Pardoux, E. and Veretennikov, A. Yu. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29**, 3, 1061–1085.
- [16] Spokoiny, V.G. (2000). Adaptive drift estimation for nonparametric diffusion model. *Ann. Statist.* **28**, 815–836.
- [17] Szegő, G. (1975) *Orthogonal polynomials*. Fourth edition. American Mathematical Society, Colloquium Publications, Vol. XXIII. American mathematical Society, Providence, R.I.
- [18] Tropp, J. A. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.*, 8(1-2):1–230.
- [19] Tsybakov, A. B. (2009) Introduction to nonparametric estimation. Springer Series in Statistics. Springer, New York.
- [20] Veretennikov, A. Y. (1988). Bounds for the mixing rate in the theory of stochastic differential equation, *Theory Probab. Appl.*, **32**, 273–281.
- [21] Viennet, G. (1997). Inequalities for absolutely regular processes: application to density estimation. *Probab. Theory Relat. Fields* **107**, 467–492.