

# NONPARAMETRIC DRIFT ESTIMATION FOR I.I.D. PATHS OF STOCHASTIC DIFFERENTIAL EQUATIONS

F. COMTE<sup>(1)</sup>, V. GENON-CATALOT<sup>(1)</sup>

**ABSTRACT.** We consider  $N$  independent stochastic processes  $(X_i(t), t \in [0, T])$ ,  $i = 1, \dots, N$ , defined by a one-dimensional stochastic differential equation which are continuously observed throughout a time interval  $[0, T]$  where  $T$  is fixed. We study nonparametric estimation of the drift function on a given subset  $A$  of  $\mathbb{R}$ . Projection estimators are defined on finite dimensional subsets of  $\mathbb{L}^2(A, dx)$ . We stress that the set  $A$  may be compact or not and the diffusion coefficient may be bounded or not. A data-driven procedure to select the dimension of the projection space is proposed where the dimension is chosen within a random collection of models. Upper bounds of risks are obtained. March 28, 2019

*Keywords and phrases:* Diffusion process, Hermite basis, Laguerre basis, model selection, non-parametric drift estimation, projection estimators.

*AMS Classification.* 62G07–62M05.

## 1. INTRODUCTION

Consider  $N$  independent stochastic processes  $(X_i(t), t \in [0, T])$ ,  $i = 1, \dots, N$  with dynamics ruled by the following one-dimensional stochastic differential equation:

$$(1) \quad dX_i(t) = b(X_i(t))dt + \sigma(X_i(t))dW_i(t), \quad X_i(0) = x_0, \quad i = 1, \dots, N,$$

where  $x_0 \in \mathbb{R}$  is known,  $(W_1, \dots, W_N)$  are independent standard Brownian motions. The drift function  $b : \mathbb{R} \rightarrow \mathbb{R}$  is unknown and our aim is to study nonparametric estimation of  $b$  from the continuous observation of the  $N$  sample paths throughout a fixed time interval  $[0, T]$ . This problem is typically part of functional data analysis which is devoted to analysis of samples of infinite dimensional data (see *e.g.* Ramsay and Silverman, 2007, Wang *et al.*, 2016). In econometrics, authors also refer to panel or longitudinal data analysis where data from a sample of individuals are collected over time (see *e.g.* Hsiao, 2003). In most cases, functional data are modeled with parametric approaches, often using mixed effects non-linear models. In particular, several recent contributions concern *i.i.d.* parametric models of stochastic differential equations with mixed effects (see *e.g.* Ditlevsen and De Gaetano, 2005, Overgaard *et al.*, 2005, Piccini *et al.*, 2010, Piccini and Ditlevsen, 2011, Comte *et al.*, 2013, Delattre and Lavielle, 2013, Delattre *et al.*, 2013, Dion and Genon-Catalot, 2016, Delattre *et al.*, 2018). Note that *i.i.d.* samples of stochastic differential equations have been used recently for multiclass classification of diffusions (see Denis *et al.*, 2018). However, the need of flexibility to deal with the information contained in functional data analysis make it preferable to use a nonparametric approach.

---

<sup>(1)</sup>: Sorbonne Paris Cité, MAP5, UMR 8145 CNRS, Université Paris Descartes, FRANCE,  
email: fabienne.comte@parisdescartes.fr,  
valentine.genon-catalot@parisdescartes.fr.

Drift estimation for one-dimensional diffusion processes has been widely investigated since the 80's. Whether by a parametric or a nonparametric approach, authors have focused on estimation from one trajectory observed on a time interval  $[0, T]$  with continuous or discrete sampling. An asymptotic framework is standardly chosen for the study: either  $T$  is fixed and the diffusion coefficient tends to 0, or  $T$  tends to infinity and ergodicity assumptions on the model are generally required. Moreover, when nonparametric estimation is performed by projection methods, the drift function is generally estimated on a fixed compact subset of  $\mathbb{R}$ . Nevertheless, when practical implementation is done, the compact set is chosen equal to the random data range which contradicts the theoretical results (see, for reference books, *e.g.* Kutoyants, 1984, 2004, Iacus, 2008, Kessler *et al.*, 2012).

In our context, ergodicity is not required for Model (1),  $T$  is fixed and the asymptotic framework is  $N$  tends to infinity. The diffusion coefficient  $\sigma$  is supposed to be known as it is identified from a continuous observation of the sample paths. We fix a subset  $A$  of  $\mathbb{R}$  and consider the estimation of  $b_A := b\mathbf{1}_A$  by a projection method on finite dimensional subspaces of  $\mathbb{L}^2(A, dx)$ . The set  $A$  may be compact or not and the drift function  $b_A$  need not be square-integrable. When  $A = \mathbb{R}^+$  or  $\mathbb{R}$ , we consider subspaces of  $\mathbb{L}^2(A, dx)$  generated respectively by Laguerre functions or Hermite functions. These subspaces have been recently used for nonparametric density or regression function estimation (see *e.g.* Comte and Genon-Catalot (2018a-b, 2019)). We propose nonparametric projection estimators of  $b_A$  and evaluate risk bounds for their  $\mathbb{L}^2$ -risk. This risk is defined either as the expectation of an empirical norm or as the expectation of a  $\mathbb{L}^2(A, f_T(x)dx)$ -norm where the density  $f_T(x)$  is equal to  $T^{-1} \int_0^T dt p_t(x_0, x)$  and  $p_t(x, y)$  is the transition density of the diffusion model. A data-driven procedure is proposed to select the dimension of the projection space. Due to the non compactness of the set  $A$ , specific bounds for the risks are obtained.

In Section 2, the projection estimators are defined and their risks are studied on a fixed projection space, assumptions and rates of convergence are discussed. Section 3 concerns the adaptive procedure. A specific difficulty arising from the non compactness of  $A$  is that the data-driven selection of the projection space dimension must be chosen within a random set. The case where  $\sigma$  is bounded on  $A$  is easier. The penalty term has the usual form and depends on  $\sigma$  only through a single upper bound,  $\|\sigma\mathbf{1}_A\|_\infty$ . For unbounded  $\sigma$ , the study is complicated by the fact that the penalty has an unusual form and is random. A short recap on Laguerre and Hermite bases is given in Section 4 and numerical simulations illustrate the estimations method. Section 5 gives some concluding remarks. Section 6 contains proofs. Some parts of proofs and technical lemmas are borrowed from Comte and Genon-Catalot (2018b, 2019). A Chernoff-type inequality for random matrices (see Tropp (2012)) used in proofs is recalled in Section 7.

## 2. PROJECTION ESTIMATORS OF THE DRIFT ON A FIXED SPACE.

**2.1. Assumptions.** We consider the usual assumptions ensuring that equation (1) admits a unique strong solution adapted to the filtration  $(\mathcal{F}_t = \sigma(W_i(s), s \leq t, i = 1, \dots, N), t \geq 0)$ :

- Either (H1): The functions  $x \mapsto b(x)$  is  $C^1$  and  $x \mapsto \sigma(x)$  is  $C^2$  on  $\mathbb{R}$ , and both have linear growth.
- Or (H2): The function  $x \mapsto b(x)$  is Lipschitz and the function  $x \mapsto \sigma(x)$  is Hölder with exponent  $\alpha \in [1/2, 1]$ . This implies that both  $b$  and  $\sigma$  have linear growth.

Thus

$$(2) \quad \exists K > 0, \forall x \in \mathbb{R}, \quad b^2(x) + \sigma^2(x) \leq K(1 + x^2).$$

Assumption (H1) is standard and Assumption (H2) is fulfilled *e.g.* by  $\sigma(x) = \sqrt{x_+}$  (Cox-Ingersoll-Ross process). Under (H1) or (H2), the Markov process  $(X_i(t))$  admits a transition

density  $p_t(x, y)$  jointly continuous in  $(t, x, y)$  on  $\mathbb{R}^+ \times (\ell, r) \times (\ell, r)$  where  $(\ell, r)$  is the state space of (1) (see *e.g.* Rogers and Williams, 1990, chap. V, Section 7). Moreover, as the initial condition  $x_0$  is deterministic,

$$(3) \quad \forall k \geq 0, \forall t \geq 0 \quad \sup_{0 \leq u \leq t} \mathbb{E}(X_1(u))^{2k} = \sup_{0 \leq u \leq t} \int y^{2k} p_u(x_0, y) dy < +\infty.$$

The following density which is well defined plays an important role in the sequel:

$$(4) \quad f_T(y) = \frac{1}{T} \int_0^T p_u(x_0, y) du.$$

By (3),  $f_T$  has moments of any order. From assumptions (H1) or (H2) and (3), we have, for all  $k$ :

$$(5) \quad \frac{1}{T} \mathbb{E} \left[ \int_0^T \left( b^{2k}(X_1(u)) + \sigma^{2k}(X_1(u)) \right) du \right] = \int (b^{2k}(y) + \sigma^{2k}(y)) f_T(y) dy < +\infty.$$

**2.2. Definition of projection estimators.** The following notations are used below. For  $h$  a function, we denote  $\|h\|$  the  $\mathbb{L}^2$ -norm of  $\mathbb{L}^2(A, dx)$ ,  $\|h\|_{f_T}$  the  $\mathbb{L}^2$ -norm of  $\mathbb{L}^2(A, f_T(x)dx)$  and set  $h_A = h\mathbf{1}_A$  and  $\|h\|_\infty = \sup_{x \in A} |h(x)|$  for the sup-norm on  $A$ . The Euclidean norm in  $\mathbb{R}^m$  is denoted by  $\|\cdot\|_{2,m}$ .

To define nonparametric estimators of the drift function  $b$ , we proceed by a projection method. Consider a set  $A \subset \mathbb{R}$  and a family  $(S_m, m \geq 0)$  of finite-dimensional subspaces of  $\mathbb{L}^2(A, dx)$ , where each  $S_m$  is endowed with an orthonormal basis  $(\varphi_j, j = 0, \dots, m-1)$  of  $A$ -supported functions and we estimate  $b_A := b\mathbf{1}_A$ . The basis of  $S_m$  may depend on  $m$  but for simplicity, we omit this dependence in the notations. We assume that the basis functions  $\varphi_j$  are bounded so that  $S_m \subset \mathbb{L}^2(A, f_T(x)dx)$ .

Then, for  $t : \mathbb{R} \rightarrow \mathbb{R}$  a function, we introduce the contrast:

$$(6) \quad \gamma_N(t) = \frac{1}{NT} \sum_{i=1}^N \left( \int_0^T t^2(X_i(u)) du - 2 \int_0^T t(X_i(u)) dX_i(u) \right)$$

and note that, for any bounded  $t$ , as  $\mathbb{E} \int_0^T t^2(X_1(u)) \sigma^2(X_1(u)) du < +\infty$ ,

$$\begin{aligned} \mathbb{E} \gamma_N(t) &= \frac{1}{T} \mathbb{E} \int_0^T [t(X_1(u)) - b(X_1(u))]^2 du - \frac{1}{T} \mathbb{E} \int_0^T b^2(X_1(u)) du \\ &= \int (t(y) - b(y))^2 f_T(y) dy - \int b^2(y) f_T(y) dy. \end{aligned}$$

This property justifies the definition of a collection of estimators  $\hat{b}_m, m \geq 0$  of  $b_A := b\mathbf{1}_A$  by setting:

$$(7) \quad \hat{b}_m = \arg \min_{t \in S_m} \gamma_N(t)$$

Thus, for each  $m$ ,

$$(8) \quad \hat{b}_m = \sum_{j=0}^{m-1} \hat{\theta}_j \varphi_j$$

where the vector of coefficients  $\hat{\theta}_{(m)} = (\hat{\theta}_0, \dots, \hat{\theta}_{m-1})'$  can be easily computed. Indeed, define the  $m \times 1$ -vector

$$(9) \quad \hat{Z}_m = \left( \frac{1}{NT} \sum_{i=1}^N \int_0^T \varphi_j(X_i(u)) dX_i(u) \right)_{j=0, \dots, m-1}$$

and the  $m \times m$ -matrix

$$(10) \quad \hat{\Psi}_m = \left( \frac{1}{NT} \sum_{i=1}^N \int_0^T \varphi_j(X_i(u)) \varphi_\ell(X_i(u)) du \right)_{j, \ell=0, \dots, m-1}.$$

Then, provided that  $\hat{\Psi}_m$  is a.s. invertible,

$$(11) \quad \hat{\theta}_{(m)} = \hat{\Psi}_m^{-1} \hat{Z}_m.$$

We introduce the empirical norm and the empirical scalar product associated with our observations. For  $t(\cdot), s(\cdot)$  two bounded functions, we set

$$(12) \quad \|t\|_N^2 = \frac{1}{NT} \sum_{i=1}^N \int_0^T t^2(X_i(u)) du, \quad \langle s, t \rangle_N = \frac{1}{NT} \sum_{i=1}^N \int_0^T t(X_i(u)) s(X_i(u)) du,$$

$$(13) \quad \nu_N(t) = \frac{1}{NT} \sum_{i=1}^N \int_0^T t(X_i(u)) \sigma(X_i(u)) dW_i(u).$$

Therefore,  $\mathbb{E}\|t\|_N^2 = \|t\|_{f_T}^2$ ,  $\mathbb{E}\langle s, t \rangle_N = \langle s, t \rangle_{f_T}$  and  $\mathbb{E}\nu_N(t) = 0$ ,  $\mathbb{E}\nu_N^2(t) = \|t\sigma\|_{f_T}^2/NT$ . Using these notations, we obtain:

$$\hat{Z}_m = (\langle \varphi_j, b \rangle_N, j = 0, \dots, m-1)' + E_m \quad \hat{\Psi}_m = (\langle \varphi_j, \varphi_\ell \rangle_N, j, \ell = 0, \dots, m-1)$$

where

$$(14) \quad E_m = (\nu_N(\varphi_j), j = 0, \dots, m-1)'$$

is a centered vector. Using (8)-(11), one easily checks that  $\gamma_N(\hat{b}_m) = -\|\hat{b}_m\|_N^2$ .

**2.3. Risk bound.** For  $M$  a matrix, we denote by  $\text{Tr}(M)$  the trace of  $M$  and 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  $M$ . Moreover, if  $M, N$  are two matrices with compatible product  $MN$ , then,  $\|MN\|_{\text{op}} \leq \|M\|_{\text{op}}\|N\|_{\text{op}}$ . For  $M$  a symmetric nonnegative matrix, we denote  $M^{1/2}$  a symmetric square root of  $M$ .

Let us set the following assumption:

$$(15) \quad L(m) := \sup_{x \in A} \sum_{j=0}^{m-1} \varphi_j^2(x) < +\infty.$$

It is easy to see that the quantity  $L(m)$  depends on the space  $S_m$  but not on the choice of the  $\mathbb{L}^2(A, dx)$ -orthonormal basis of  $S_m$  used to compute it. Indeed,  $L(m) = \sup_{t \in S_m, \|t\|=1} \sup_{x \in A} t^2(x)$ . If the spaces  $S_m$  are nested, i.e.  $m \leq m' \Rightarrow S_m \subset S_{m'}$ , then the map  $m \mapsto L(m)$  is increasing. Throughout the paper, the length-time interval  $T$  is fixed and the asymptotic framework is  $N$  tends to infinity. Without loss of generality, we assume that  $T$  is an integer with  $T \geq 1$ . Though fixed, the value of  $T$  may have an impact on the performances of the estimators. This is why all bounds will be expressed as negative powers of  $NT$ .

To ensure the existence and stability of the estimator, we insert a cutoff and define, for  $m \geq 1$ ,

$$(16) \quad \tilde{b}_m = \hat{b}_m \mathbf{1}_{\{L(m)(\|\hat{\Psi}_m^{-1}\|_{\text{op}} \vee 1) \leq c_T NT / \log(NT)\}}, \quad c_T = \frac{1 - \log(2)}{8T}.$$

Let us define the following  $m \times m$  matrices:

$$(17) \quad \Psi_m = \mathbb{E} \hat{\Psi}_m = (\langle \varphi_j, \varphi_\ell \rangle_{f_T}, j, \ell = 0, \dots, m-1),$$

$$(18) \quad \Psi_{m, \sigma^2} = \mathbb{E}(E_m E_m') = (\langle \sigma \varphi_j, \sigma \varphi_\ell \rangle_{f_T}, j, \ell = 0, \dots, m-1)$$

(see (14)). Under mild assumptions on the basis  $(\varphi_j)$ , the matrix  $\Psi_m$  is invertible as for instance the ones given in the following lemma.

**Lemma 1.** *Assume that  $\lambda(A \cap \text{supp}(f_T)) > 0$  where  $\lambda$  is the Lebesgue measure and  $\text{supp}(f_T)$  the support of  $f_T$ , that the  $(\varphi_j)_{0 \leq j \leq m-1}$  are continuous, and that there exist  $x_0, \dots, x_{m-1} \in A \cap \text{supp}(f_T)$  such that  $\det[(\varphi_j(x_k))_{0 \leq j, k \leq m-1}] \neq 0$ . Then,  $\Psi_m$  is invertible.*

The proof is elementary using that, for  $u = (u_0, \dots, u_{m-1})'$ ,

$$u' \Psi_m u = \int \left( \sum_{j=0}^{m-1} u_j \varphi_j(y) \right)^2 f_T(y) dy.$$

In particular, if  $(\varphi_j)_{0 \leq j \leq m-1}$  is the Laguerre or the Hermite basis (see Section 4),  $\Psi_m$  is invertible.

By convention, when  $M$  is a symmetric non negative and non invertible matrix, we set  $\|M^{-1}\|_{\text{op}} = +\infty$ , a convention which is coherent as when  $M$  is invertible,  $\|M^{-1}\|_{\text{op}} = 1/\inf\{\lambda_j\}$  where  $\{\lambda_j\}$  are the eigenvalues of  $M$ .

**Proposition 1.** *Consider the estimator  $\tilde{b}_m$  of  $b_A$ . Then for  $m$  such that*

$$(19) \quad L(m)(\|\Psi_m^{-1}\|_{\text{op}} \vee 1) \leq \frac{c_T}{2} \frac{NT}{\log(NT)} \quad \text{and} \quad m \leq NT$$

with  $c_T$  given in (16), we have

$$(20) \quad \mathbb{E}[\|\tilde{b}_m - b_A\|_N^2] \leq \inf_{t \in S_m} \|t - b_A\|_{f_T}^2 + \frac{2}{NT} \text{Tr}[\Psi_m^{-1} \Psi_{m, \sigma^2}] + \frac{c_1(T)}{NT},$$

and

$$(21) \quad \mathbb{E}[\|\tilde{b}_m - b_A\|_{f_T}^2] \leq \left(1 + \frac{1 - \log(2)}{2 \log(NT)}\right) \inf_{t \in S_m} \|t - b_A\|_{f_T}^2 + 8 \frac{\text{Tr}[\Psi_m^{-1} \Psi_{m, \sigma^2}]}{NT} + \frac{c_2(T)}{NT},$$

where  $c_1(T)$ ,  $c_2(T)$  depend on  $T$  through  $\int \sigma_A^4(y) f_T(y) dy$  and  $\int b_A^4(y) f_T(y) dy$ .

Actually, we can prove that  $m \leq L(m)\|\Psi_m^{-1}\|_{\text{op}}$  and  $m \lesssim NT$  is automatically satisfied (see Lemma 4 in Comte and Genon-Catalot (2018b)).

In the framework of standard regression with independent data,  $Y_i = b(X_i) + \varepsilon_i$ ,  $i = 1, \dots, n$ , Cohen *et al.* (2013) introduced condition (15) on the space  $S_m$  and (19) on the possible dimensions (see also Comte and Genon-Catalot (2018b, 2019)). The restrictions on the choices of  $m$  imposed by (19) have the effect of stabilizing projection estimators. If  $m$  is too large, then, estimators become very unstable and the precise cutoff for stability is proportional to  $n/\log n$  in the regression model, or  $NT/\log(NT)$  in our case.

Note that  $\|\Psi_m^{-1}\|_{\text{op}} = \sup_{t \in S_m, \|t\|_{f_T}=1} \|t\|^2$  (see Proposition 2 in Comte and Genon-Catalot, 2018b) so that, for nested spaces,  $m \mapsto \|\Psi_m^{-1}\|_{\text{op}}$  is increasing.

From the variance bound in (20), we cannot deduce a precise rate as a function of  $m$ . Nevertheless, this bound verifies:

**Proposition 2.**

- (i) Let  $S_m$  be nested spaces, then  $m \mapsto \text{Tr}[\Psi_m^{-1}\Psi_{m,\sigma^2}]$  is increasing with  $m$ .  
(ii) If  $\sigma$  is bounded on  $A$ ,  $\text{Tr}[\Psi_m^{-1}\Psi_{m,\sigma^2}] \leq m\|\sigma_A\|_\infty^2$ .

Classically, in projection methods, the set  $A$  is chosen to be compact. If  $A$  is compact,  $\sigma_A$  is automatically bounded, Proposition 2 applies, and we obtain a variance bound of order  $m/(NT)$ .

In addition, if  $A$  is compact, it can be assumed that  $f_T$  is lower bounded on  $A$ , say by  $f_0$ . Then we have  $\|\Psi_m^{-1}\|_{\text{op}} \leq 1/f_0$ . Indeed for  $\vec{u} = (u_0, \dots, u_{m-1})'$  a vector of  $\mathbb{R}^m$ ,  $\vec{u}'\Psi_m\vec{u}$  is equal to

$$\int_A \left( \sum_{j=0}^{m-1} u_j \varphi_j(x) \right)^2 f_T(x) dx \geq f_0 \int_A \left( \sum_{j=0}^{m-1} u_j \varphi_j(x) \right)^2 dx = f_0 \|\vec{u}\|_{2,m}^2.$$

Therefore, the stability condition (19) simplifies into  $m \leq cNT/\log(NT)$  where  $c$  depends on  $T$  and  $f_0$ .

If  $A$  is not compact,  $\|\Psi_m^{-1}\|_{\text{op}}$  may be unbounded as a function of  $m$  and may increase the variance rate. For instance, in the case where  $(\varphi_j)$  is the Laguerre basis on  $A = \mathbb{R}^+$  or the Hermite basis on  $A = \mathbb{R}$ , it is proved in the above quoted paper, Proposition 8, that for any underlying density  $f_T$ ,  $\|\Psi_m^{-1}\|_{\text{op}} \geq c\sqrt{m}$  for some constant  $c$  (see Section 4 for the definitions of the Laguerre and Hermite bases).

**2.4. Rates of convergence.** Some conclusions can be drawn from Propositions 1 and 2 concerning the rates of convergence of the projection estimators. In Comte and Genon-Catalot (2018b), to assess the bias rate, the following regularity set is proposed and justified:

$$W_{f_T}^s(A, R) = \left\{ h \in \mathbb{L}^2(A, f_T(x)dx), \forall \ell \geq 1, \|h - h_\ell^{f_T}\|_{f_T}^2 \leq R\ell^{-s} \right\},$$

where  $h_\ell^{f_T}$  is the  $\mathbb{L}^2(A, f_T(x)dx)$ -orthogonal projection of  $h$  on  $S_\ell$ . If  $b_A$  belongs to  $W_{f_T}^s(A, R)$ , then the square bias satisfies  $\|b_m^{f_T} - b_A\|_{f_T}^2 \leq Rm^{-s}$ .

Then, the best compromise between bias and variance terms is obtained defining  $m^*$  by the implicit relation  $(m^*)^{-s} = \text{Tr}[\Psi_{m^*}^{-1}\Psi_{m^*,\sigma^2}]/NT$  and yields a rate of order  $(m^*)^{-s}$ .

If  $\sigma$  is bounded on  $A$  (see Proposition 2), and if  $m^* = (NT)^{1/(s+1)}$  satisfies (19), we find the rate

$$\mathbb{E}[\|\tilde{b}_{m^*} - b_A\|_{f_T}^2] \lesssim (NT)^{-s/(s+1)}.$$

Let us stress that our context is hitherto unstudied and although this new rate looks familiar, the optimal rate for this problem is not known.

The next section is devoted to data-driven choices of the dimension of the projection space and yields an adaptive estimator, *i.e.* achieving automatically the best compromise between square bias and variance terms. This is especially interesting in our case where the exact rate is implicit.

### 3. DATA-DRIVEN PROCEDURE

Let us consider now the following assumption.

- (A1)** 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

$$(22) \quad L(m) = \left\| \sum_{j=0}^{m-1} \varphi_j^2 \right\|_\infty \leq c_\varphi^2 m, \quad \text{for } c_\varphi > 0 \text{ a constant.}$$

- (A2)**  $\|f_T\|_\infty < +\infty$ .

Clearly, Assumption **(A1)** is fulfilled by classical compactly supported bases, such as histograms and trigonometric polynomials, and also by Laguerre and Hermite bases, which are non compactly supported, see Section 4. Note that  $L(m)$  does not depend on the basis, but the bound  $c_\varphi^2 m$  does depend on it. In Section 3.3, we give sufficient conditions ensuring that **(A2)** holds. We consider the following collection of models, for  $\theta$  a positive constant specified below:

$$(23) \quad \widehat{\mathcal{M}}_N(\theta) = \left\{ m \in \{1, 2, \dots, NT\}, c_\varphi^2 m (\|\widehat{\Psi}_m^{-1}\|_{\text{op}}^2 \vee 1) \leq \theta \frac{NT}{\log(NT)} \right\},$$

and its theoretical counterpart

$$(24) \quad \mathcal{M}_N(\theta) = \left\{ m \in \{1, \dots, NT\}, c_\varphi^2 m (\|\Psi_m^{-1}\|_{\text{op}}^2 \vee 1) \leq \frac{\theta}{4} \frac{NT}{\log(NT)} \right\}.$$

Note that, analogously as for  $\|\Psi_m^{-1}\|_{\text{op}}$ ,  $m \mapsto \|\widehat{\Psi}_m^{-1}\|_{\text{op}}$  is increasing.

Under **(A1)**, the condition in the definition of  $\widehat{\mathcal{M}}_N(\theta)$  is to be compared with the stability condition (19) which writes  $c_\varphi^2 m (\|\Psi_m^{-1}\|_{\text{op}} \vee 1) \leq (c_T/2)(NT/\log(NT))$ . The condition imposed in  $\mathcal{M}_N(\theta)$  is thus stronger as, clearly,  $(\|\Psi_m^{-1}\|_{\text{op}} \vee 1) \leq (\|\Psi_m^{-1}\|_{\text{op}}^2 \vee 1)$ . The same remark holds between  $\widehat{\mathcal{M}}_N(\theta)$  and the cutoff used to define  $\tilde{b}_m$  (see (16)).

The aim here is to define a data-driven procedure for selecting the dimension  $m$  of the projection space in such a way that the resulting estimator is adaptive, *i.e.* that its  $\mathbb{L}^2$ -risk realizes automatically the best compromise between the bias and the variance term. For this, we distinguish the case where  $\sigma_A$  is bounded or not as the method is different. In both cases, we need the Bernstein Inequality for continuous local martingales, see Revuz and Yor (1999, p.153), that we state in our context.

**Lemma 2.** *Let  $M_T := NT\nu_N(t)$  (see (13)) and  $\langle M \rangle_T = \int_0^T \sum_{i=1}^N t^2(X_i(u))\sigma^2(X_i(u))du$ . Then,*

$$\mathbb{P}(M_T \geq NT\varepsilon, \langle M \rangle_T \leq NTv^2) \leq \exp\left(-\frac{NT\varepsilon^2}{2v^2}\right).$$

**3.1. Case of bounded  $\sigma_A$ .** If  $\sigma$  is bounded on  $A$ , proofs are simpler. We have that  $\langle M \rangle_T \leq NT\|\sigma_A\|_\infty^2 \|t\|_N^2$  and from Proposition 2, the variance term of the risk bound is upper bounded by  $\|\sigma_A\|_\infty^2 m/NT$ .

Let us define, under **(A2)**,

$$(25) \quad \vartheta_T = \left(3 \wedge \frac{1}{\|f_T\|_\infty}\right) \frac{1}{c_0 T},$$

where  $c_0$  is a numerical constant computed in the proof of Theorem 1. Now we set

$$(26) \quad \hat{m} = \arg \min_{m \in \widehat{\mathcal{M}}_N(\vartheta_T)} \left\{ -\|\hat{b}_m\|_N^2 + \text{pen}_1(m) \right\}, \quad \text{with} \quad \text{pen}_1(m) = \kappa \|\sigma_A\|_\infty^2 \frac{m}{NT},$$

where  $\kappa$  is a numerical constant. Note that  $\|\sigma_A\|_\infty^2 m/NT$  is an upper bound on the variance term obtained in Proposition 1 (see Proposition 2).

**Theorem 1.** *Let  $(X_i(t), t \in [0, T])_{1 \leq i \leq N}$  be observations ruled by model (1). Assume that **(A1)**, **(A2)** hold and that  $\|\sigma_A\|_\infty < \infty$ . Then, there exists a numerical constant  $\kappa_0$  such that for  $\kappa \geq \kappa_0$ , we have*

$$\mathbb{E}[\|\hat{b}_{\hat{m}} - b_A\|_N^2] \leq C \inf_{m \in \widehat{\mathcal{M}}_N(\vartheta_T)} \left( \inf_{t \in S_m} \|b_A - t\|_{f_T}^2 + \text{pen}_1(m) \right) + \frac{C'}{NT},$$

and

$$\mathbb{E}[\|\hat{b}_{\hat{m}} - b_A\|_{f_T}^2] \leq C_1 \inf_{m \in \widehat{\mathcal{M}}_N(\vartheta_T)} \left( \inf_{t \in S_m} \|b_A - t\|_{f_T}^2 + \text{pen}_1(m) \right) + \frac{C'_1}{NT}$$

where  $C, C_1$  are a numerical constants and  $C', C'_1$  are constants depending on  $T$  through  $\|f_T\|_\infty$ ,  $\int b_A^4(y) f_T(y) dy$ ,  $\int \sigma_A^4(y) f_T(y) dy$ .

Theorem 1 says that  $\hat{b}_{\hat{m}}$  automatically realizes the compromise between the squared-bias term and the variance term, on the collection  $\mathcal{M}_N(\mathfrak{d}_T)$ .

The penalty contains  $\|\sigma_A\|_\infty$ . As we assume a continuous observation of each sample path, it is well known that the function  $\sigma$  is identified from such an observation. Therefore,  $\sigma$  can be assumed to be known. Note that the estimation procedure for  $\hat{b}_m$  and  $\hat{m}$  only depends on  $\sigma$  through  $\|\sigma_A\|_\infty$ . In practice, to implement the adaptive procedure, we can use a simple estimator of  $\|\sigma_A\|_\infty$  built from discretisations of the observed trajectories with very small sample step. In the definition of the sets  $\mathcal{M}_N(\mathfrak{d}_T)$  and  $\widehat{\mathcal{M}}_N(\mathfrak{d}_T)$ , there appears  $\|f_T\|_\infty$ , which is unknown. In theory and in practical implementation, we can simply replace  $\mathfrak{d}_T(NT)/\log(NT)$  by  $NT/\log^{1+\epsilon}(NT)$ ,  $\epsilon > 0$ , provided that  $N$  is large enough.

The constant  $\kappa$  is a specific feature of the model selection method. Theorem 1 states that, under the assumptions of the theorem, for any function  $b$ , there exists a numerical (universal) constant  $\kappa_0$  such that the inequalities hold for all  $\kappa \geq \kappa_0$ . The proof provides a numerical value  $\kappa_0$  which is too large. Finding the best value  $\kappa_0$  for a given statistical problem is not easy. For instance, this topic is the subject of Birgé and Massart (2007) paper in the Gaussian white noise model where the authors prove that  $\kappa > 1$  is required in this case. Thus, for practical implementation of the adaptive estimator, it is standard and commonly done that one starts by preliminary simulations to obtain a value of  $\kappa$  closer to the true one. Afterwards, this value is fixed once and for all.

**3.2. Case of unbounded  $\sigma_A$ .** Here, the estimation procedure depends on the complete knowledge of  $\sigma$  and of the constant  $K$  such that  $\sigma^2(x) \leq K(1+x^2)$  (see (2)).

To study the case of unbounded  $\sigma_A$ , it is natural to consider that  $A$  is non compact. In the following, we consider the Laguerre and Hermite bases (see Section 4), and introduce the specific assumptions:

**(A3)** There exists  $c > 0$  such that for all  $m \geq 1$ ,  $\|\Psi_m^{-1}\|_{\text{op}}^2 \geq cm^\beta$  with  $\beta = 4$  for Laguerre basis and  $\beta = 5/3$  for Hermite basis.

**(A4)** The function  $\sigma^2$  is lower bounded on  $A$ :  $\sigma^2(x) \geq \sigma_0^2 > 0$ .

For the Hermite and Laguerre bases,  $\|\Psi_m^{-1}\|_{\text{op}}^2 \geq cm$ , see Comte and Genon-Catalot (2018b), Proposition 8. Consequently, **(A3)** is a stronger constraint: our conjecture, based on numerical simulations, is that it is related to the rate of decay of  $f_T$  near infinity. Under **(A4)**, the state space of the processes  $X_i(t)$  is  $\mathbb{R}$ ; nevertheless, it is possible to estimate  $b$  on  $\mathbb{R}^+$  using the Laguerre basis. Moreover, if  $\sigma$  is not lower bounded, the result below still holds replacing **(A4)** by the technical condition (56).

Let

$$(27) \quad \mathfrak{f}_T = \mathfrak{d}_T \wedge \frac{1 - \log(2)}{14TB\sigma_0^2}$$

with  $\mathfrak{d}_T$  is defined in (26),  $B = 21K$  for Laguerre basis,  $B = 2KC_\infty^2$  for Hermite basis (see the definition of  $C_\infty$  in Section 4),  $K$  is defined in (2) and set

$$(28) \quad \hat{m} = \arg \min_{m \in \mathcal{M}_N(\mathfrak{f}_T)} \left\{ -\|\hat{b}_m\|_N^2 + \widehat{\text{pen}}_2(m) \right\}$$



where

$$(29) \quad \widehat{\text{pen}}_2(m) = \kappa_1 \frac{m(1 + \ell_m)(1 + \|\widehat{\Psi}_m^{-1/2} \widehat{\Psi}_{m,\sigma^2} \widehat{\Psi}_m^{-1/2}\|_{\text{op}})}{NT},$$

with  $\kappa_1$  a numerical constant,  $(\ell_m)$  is a sequence of nonnegative numbers. The matrix  $\widehat{\Psi}_{m,\sigma^2}$  is the empirical counterpart of  $\Psi_{m,\sigma^2}$  (see (18)):

$$\widehat{\Psi}_{m,\sigma^2} = (\langle \sigma \varphi_j, \sigma \varphi_\ell \rangle_N, j, \ell = 0, \dots, m-1).$$

**Theorem 2.** *Let  $(X_i(t), t \in [0, T])_{1 \leq i \leq N}$  be observations ruled by model (1). Assume that **(A1)**-**(A4)** hold. Let the sequence  $(\ell_m)$  be such that  $1 \leq \ell_m \leq NT$ , for  $m \in \mathcal{M}_N(f_T)$  and*

$$(30) \quad \sum_{m \in \mathcal{M}_N(16f_T)} m \|\Psi_m^{-1}\|_{\text{op}} e^{-m\ell_m} \leq \Sigma.$$

Let

$$(31) \quad \text{pen}_2(m) = \kappa_1 \frac{m(1 + \ell_m)(1 + \|\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}\|_{\text{op}})}{NT}.$$

Then, there exists a numerical constant  $\tilde{\kappa}_0$  such that for  $\kappa_1 \geq \tilde{\kappa}_0$ , we have

$$\mathbb{E}[\|\hat{b}_{\widehat{m}} - b_A\|_N^2] \leq C \inf_{m \in \mathcal{M}_N(f_T)} \left( \inf_{t \in S_m} \|b_A - t\|_{f_T}^2 + \text{pen}_2(m) \right) + \frac{C'}{NT}$$

and

$$\mathbb{E}[\|\hat{b}_{\widehat{m}} - b_A\|_{f_T}^2] \leq C_1 \inf_{m \in \mathcal{M}_N(f_T)} \left( \inf_{t \in S_m} \|b_A - t\|_{f_T}^2 + \text{pen}_2(m) \right) + \frac{C'_1}{NT},$$

where  $C, C_1$  are a numerical constants and  $C', C'_1$  are constants depending on  $\Sigma$  and depending on  $T$  through  $\int b_A^4(y) f_T(y) dy$ ,  $\int \sigma_A^{4+56/\beta}(y) f_T(y) dy$ ,  $\|f_T\|_\infty$ .

As previously, the penalty is obtained using an upper bound  $m(1+\ell_m)(1+\|\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\|_{\text{op}})/NT$  of the variance term given in Proposition 1. Theorem 2 thus states that the compromise between the squared-bias term and the variance term, is automatically realized by  $\hat{b}_{\widehat{m}}$ , on the collection  $\mathcal{M}_N(f_T)$ .

Under **(A4)**,  $\text{Tr}(\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}) \geq \sigma_0^2 m$  and  $\|\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\|_{\text{op}} \geq \sigma_0^2$ , thus  $\text{pen}_2(m) \gtrsim \text{pen}_1(m)$ .

In Comte and Genon-Catalot (2018b), examples of densities for which  $\|\Psi_m^{-1}\|_{\text{op}}$  is upper bounded by  $O(m^k)$  are given. In such a case, we can take  $\ell_m = 1$  for all  $m$ , and (30) holds.

**3.3. About assumption (A2) and some extensions.** Recall that, for  $h$  continuous and bounded,  $s \rightarrow \mathbb{E}h(X(s))$  is continuous and therefore,  $T^{-1} \int_0^T \mathbb{E}h(X(s)) ds = \int h(y) f_T(y) dy$  is well defined so that the density  $f_T$  is always well defined.

When the transition density is explicit, we can check **(A2)** directly. For instance, assumption **(A2)** holds for the Brownian motion with drift, for the Ornstein-Uhlenbeck process or for the geometric Brownian motion. More generally, the following result holds.

**Proposition 3.** (i) *If  $\sigma(x) = 1$ ,  $b$  is  $C^1$ ,  $|b| + |b'| \leq M$ , then  $\|f_T\|_\infty < +\infty$ .*

(ii) *If  $\sigma$  is  $C^2$ ,  $\sigma', \sigma''$  bounded,  $\sigma$  lower bounded by  $\sigma_0 > 0$ ,  $b$  is  $C^1$  and  $b, b'$  bounded, then  $\|f_T\|_\infty < +\infty$ .*

Our study concerns a fixed initial condition  $x_0$  for the diffusion model. This is not mandatory. We may also consider the model

$$(32) \quad dX_i(t) = b(X_i(t))dt + \sigma(X_i(t))dW_i(t), \quad X_i(0) = \eta_i, \quad i = 1, \dots, N.$$

where the initial conditions  $\eta_i$  are i.i.d. random variables independent of  $(W_1, \dots, W_N)$ , with common distribution  $\mu$  on  $\mathbb{R}$ , such that  $\mathbb{E}\eta^{2k} < +\infty$ , for  $k$  large enough. In this case,  $X_i(s)$  has distribution  $\int_{\mathbb{R}} \mu(dx)p_s(x, y)dy$ . It is enough to replace  $f_T = f_T^{x_0}$  by

$$f_T^\mu(y) = \frac{1}{T} \int_0^T ds \int_{\mathbb{R}} \mu(dx)p_s(x, y).$$

In particular, if model (32) is positive recurrent with invariant distribution  $\pi(y)dy$  and  $\eta$  has distribution  $\pi$ , then  $\int_{\mathbb{R}} \pi(dx)p_s(x, y) = \pi(y)$  for all  $s$ , implying that  $f_T^\pi = \pi$ . The assumption  $\|f_T\|_\infty < +\infty$  becomes  $\|\pi\|_\infty < \infty$ . So  $\|\cdot\|_{f_T} = \|\cdot\|_\pi$  is fixed, the constants  $c_1(T), c_2(T)$  in Proposition 1 no more depend on  $T$ , and thus the risk bound (especially the variance term) is improved when  $T$  gets large.

#### 4. SIMULATION STUDY

We propose a brief simulation study to illustrate the estimation method. Implementation is done with either the Laguerre basis ( $A = \mathbb{R}^+$ ) or the Hermite basis ( $A = \mathbb{R}$ ). We recall their definition.

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

$$(33) \quad L_j(x) = \sum_{k=0}^j (-1)^k \binom{j}{k} \frac{x^k}{k!}, \quad \ell_j(x) = \sqrt{2}L_j(2x)e^{-x}\mathbf{1}_{x \geq 0}, \quad j \geq 0.$$

The collection  $(\ell_j)_{j \geq 0}$  is a complete orthonormal system on  $\mathbb{L}^2(\mathbb{R}^+)$  satisfying:  $\forall j \geq 0, \forall x \in \mathbb{R}^+, |\ell_j(x)| \leq \sqrt{2}$ , see Abramowitz and Stegun (1964, 22.14.12) The collection of models  $(S_m = \text{span}\{\ell_0, \dots, \ell_{m-1}\})$  is nested and obviously (22) 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:

$$(34) \quad H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j} (e^{-x^2}), \quad h_j(x) = c_j H_j(x) e^{-x^2/2}, \quad c_j = (2^j j! \sqrt{\pi})^{-1/2}$$

The sequence  $(h_j, j \geq 0)$  is an orthonormal basis of  $\mathbb{L}^2(\mathbb{R}, dx)$ . Moreover (see Abramowitz and Stegun (1964, 22.14.17), Szegö (1975) p.242),  $\|h_j\|_\infty \leq \Phi_0, \Phi_0 \simeq 1,086435/\pi^{1/4} \simeq 0.8160$ , so that (22) holds with  $c_\varphi^2 = \Phi_0^2$ . The collection of models  $(S_m = \text{span}\{h_0, \dots, h_{m-1}\})$  is obviously nested. Moreover,  $\|h_j\|_\infty \leq C_\infty (j+1)^{-1/12}, j = 0, 1, \dots$  where the constant  $C_\infty$  given is in Szegö (1975). Thus in this case,  $L(m) \leq C_\infty^2 m^{5/6}$ .

Laguerre polynomials are computed using formula  $(j+1)L_{j+1}(x) = (2j+1-x)L_j(x) - jL_{j-1}(x)$ ,  $L_0(x) = 1$ ,  $L_1(x) = 1-x$  and Hermite polynomials with  $H_0(x) = 1$ ,  $H_1(x) = x$  and the recursion  $H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$ , see Abramowitz and Stegun (1964, 22.7).

We simulate discrete sampling of four models, one by Euler scheme, and the others by exact discretization. All our models admit a stationary distribution. When models are randomly initialized, the initial variable follows the stationary density and **(A2)** is fulfilled..

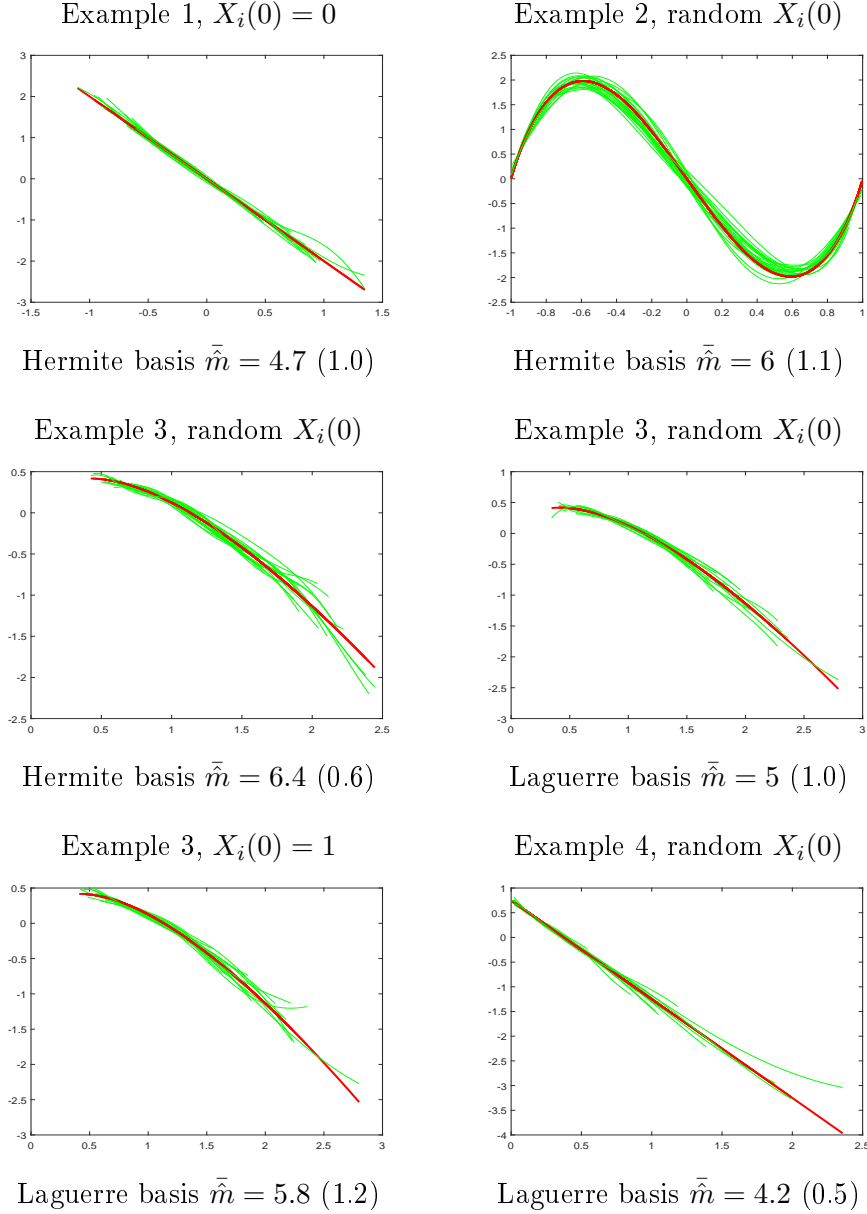


FIGURE 1. 20 estimated curves in Hermite or Laguerre basis (grey-green), the true in bold (black/red),  $N = 100$ ,  $T = 10$ .

**Example 1.** Hyperbolic diffusion. The model  $dX_t = -\theta X_t dt + \gamma \sqrt{1 + X_t^2} dW_t$ ,  $X_0 = 0$ , is simulated by a Euler scheme with step  $\Delta$ . We chose  $\theta = 2$  and  $\gamma = \sqrt{1/2}$ . Model 1 satisfies (H1).

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

$$(35) \quad dU_i(t) = -\frac{r}{2} U_i(t) dt + \frac{\gamma}{2} dW_{i,d}(t), \quad U_i(0) \sim \mathcal{N}_d(0, \frac{\gamma^2}{4r} I_d) \text{ or } U_i(0) = 0.$$

Here  $W_{i,d}$  is a  $d$ -dimensional standard Brownian motion. Exact simulation is generated with step  $\Delta$  by computing

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

**Example 2.**  $X_i(t) = \tanh(U_i(t))$  where  $U_i(t)$  is defined by (35) with  $d = 1$  is solution of (32) with

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

Here,  $X_i(t)$  has state space  $[-1, 1]$ , so that  $b$  and  $\sigma$  are bounded on this domain and (H1) holds.

**Example 3.**  $X_i(t) = \exp(U_i(t))$  where  $U_i(t)$  is defined by (35) with  $d = 1$  is solution of (32) with

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

For example 3, neither (H1) nor (H2) hold for  $b$ .

**Example 4.** Cox-Ingersoll-Ross or square-root process. We take  $X_i(t) = \|U_i(t)\|_{2,d}^2$  where  $U_i(t)$  is defined by (35) with  $d = 3$  is solution of (32) with

$$dX_i(t) = \left( \frac{d\gamma^2}{4} - rX_i(t) \right) dt + \gamma \sqrt{X_i(t)} dW_i^*(t),$$

where  $W_i^*(t)$  is a standard brownian motion. We take  $k = 2$  and  $\gamma = 1$ . Model 4 satisfies (H2).

In all cases, samples  $(X_i(k\Delta))_{1 \leq i \leq N, 1 \leq k \leq n}$ ,  $n\Delta = T$  from the above models are generated, with  $N = 100$  and  $T = 10$  obtained with  $n = 1000$  and  $\Delta = 0.01$ . For examples 1 and 2, the Hermite basis is used; in example 3, both Hermite and Laguerre are experienced, and in example 4, we use Laguerre basis. Indeed, examples 3,4 provide nonnegative processes and are well suited to Laguerre basis use.

The set  $\widehat{\mathcal{M}}_N(f_T)$  is generally too small in practice to contain enough values of  $m$  to be visited. Therefore a larger set given by  $\widehat{\mathcal{M}}_N^* = \{m \leq 10, m \|\widehat{\Psi}_m^{-1}\|_{\text{op}}^{1/4} \leq NT\}$  is chosen.

The penalty is taken equal to  $\widehat{\text{pen}}(m) = \kappa \|\widehat{\Psi}_m^{-1} \widehat{\Psi}_{m,\sigma^2}\|_{\text{op}} / (N n \Delta)$  and  $\hat{m}$  is selected as the minimizer of  $-\|\hat{b}_m\|_N^2 + \widehat{\text{pen}}(m)$ . After preliminary simulations, the constant  $\kappa$  is taken equal to  $\kappa = 1$ , see the comment after Theorem 1.

Figure 1 shows 20 estimated drift functions  $\hat{b}_{\hat{m}}$  (green/grey), and the true (red/black). We stress that the value of  $\hat{m}$  is rather small: under each graph, we give the mean  $\bar{m}$  computed over the 20 estimators, with standard deviation in parenthesis. Thus, we see that the function is very well reconstructed using a small number of coefficients.

## 5. CONCLUDING REMARKS

In this paper, we study nonparametric estimation of the unknown drift of a one-dimensional diffusion process from the observation of  $N$  *i.i.d.* sample paths which are continuously observed throughout a time interval  $[0, T]$ . The drift is estimated on a subset  $A$  of  $\mathbb{R}$  by a projection method where the set  $A$  may be compact or not, in the two cases of bounded or unbounded  $\sigma$ . In each case, an adaptive estimator is proposed.

The estimation procedures use some constants, which can be easily estimated, in particular  $\|f_T\|_\infty$  defined by (4). Assuming that  $f_T \in \mathbb{L}^2(A, dx)$ , the estimation of  $f_T$  can be done standardly by projection method. Let  $a_j = \langle f_T, \varphi_j \rangle$ . Then  $\hat{a}_j = n^{-1} \sum_{i=1}^N T^{-1} \int_0^T \varphi_j(X_i(s)) ds$

is an unbiased estimator of  $a_j$  and we can define the projection estimator of  $f_T$  on  $S_m$  by  $\hat{f}_{T,m} = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j$ . This estimator satisfies

$$\mathbb{E} \|\hat{f}_{T,m} - f_T\|^2 \leq \|f - f_m\|^2 + c_\varphi^2 \frac{m}{N}, \quad f_m = \sum_{j=0}^{m-1} a_j \varphi_j$$

where  $\|\cdot\|$  is the usual  $\mathbb{L}^2$ -norm.

As we assume continuous observation of the sample paths, we can consider that  $\sigma$  is known. Nevertheless, it can be estimated by using discrete sampling with very small sampling interval.

For practical implementation, a discretisation with small step is required to compute the quantities  $\hat{\Psi}_m, \hat{Z}_m, \hat{\Psi}_{m,\sigma^2}$ . The theoretical study of drift estimation for discretely observed paths would be of interest but probably tedious; it is left for further investigation.

## 6. PROOFS

We denote by  $x \lesssim y$  if there exists a constant  $c$  such that  $x \leq cy$ .

**6.1. Proof of Proposition 1.** We start by defining the sets

$$(36) \quad \Lambda_m := \left\{ L(m)(\|\hat{\Psi}_m^{-1}\|_{\text{op}} \vee 1) \leq c_T \frac{NT}{\log(NT)} \right\} \quad \text{and} \quad \Omega_m := \left\{ \left| \frac{\|t\|_N^2}{\|t\|_{f_T}^2} - 1 \right| \leq \frac{1}{2}, \forall t \in S_m \right\}.$$

On  $\Omega_m$ , the empirical norm  $\|\cdot\|_N$  and the  $\mathbb{L}^2(A, f_T(x)dx)$ -norm are equivalent for elements of  $S_m$ :  $(2/3)\|t\|_N^2 \leq \|t\|_{f_T}^2 \leq 2\|t\|_N^2$ . Moreover, if  $\vec{x}' = (x_0, \dots, x_{m-1}) \in \mathbb{R}^m$  and  $t = \sum_{j=0}^{m-1} x_j \varphi_j$ , then

$$\begin{aligned} \|t\|_N^2 &= \vec{x}' \hat{\Psi}_m \vec{x} \quad \text{and} \quad \|t\|_{f_T}^2 = \vec{x}' \Psi_m \vec{x} = \|\Psi_m^{1/2} \vec{x}\|_{2,m}^2, \quad \text{so that} \\ \sup_{t \in S_m, \|t\|_{f_T}=1} \left| \|t\|_N^2 - \|t\|_{f_T}^2 \right| &= \sup_{\vec{x} \in \mathbb{R}^m, \|\Psi_m^{1/2} \vec{x}\|_{2,m}=1} \left| \vec{x}' (\hat{\Psi}_m - \Psi_m) \vec{x} \right| \\ &= \sup_{\vec{u} \in \mathbb{R}^m, \|\vec{u}\|_{2,m}=1} \left| \vec{u}' \Psi_m^{-1/2} (\hat{\Psi}_m - \Psi_m) \Psi_m^{-1/2} \vec{u} \right| \\ &= \|\Psi_m^{-1/2} \hat{\Psi}_m \Psi_m^{-1/2} - \text{Id}_m\|_{\text{op}}. \end{aligned}$$

Therefore,

$$\Omega_m = \left\{ \|\Psi_m^{-1/2} \hat{\Psi}_m \Psi_m^{-1/2} - \text{Id}_m\|_{\text{op}} \leq 1/2 \right\}.$$

The following lemma is analogous to Lemma 5 in Comte and Genon-Catalot (2018b) and determines the value of  $c_T$  given in (16). Its proof is omitted.

**Lemma 3.** *Under the assumptions of Proposition 1, for  $m$  satisfying (19) with  $c_T$  given by (16), we have, for  $c$  is a positive constant,*

$$\mathbb{P}(\Lambda_m^c) \leq c/(NT)^7, \quad \mathbb{P}(\Omega_m^c) \leq c/(NT)^7.$$

Now, we prove (20). For this, we write

$$(37) \quad \begin{aligned} \|\tilde{b}_m - b_A\|_N^2 &= \|\hat{b}_m - b_A\|_N^2 \mathbf{1}_{\Lambda_m} + \|b_A\|_N^2 \mathbf{1}_{\Lambda_m^c} \\ &= \|\hat{b}_m - b_A\|_N^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} + \|\hat{b}_m - b_A\|_N^2 \mathbf{1}_{\Lambda_m \cap \Omega_m^c} + \|b_A\|_N^2 \mathbf{1}_{\Lambda_m^c} := T_1 + T_2 + T_3. \end{aligned}$$

We bound the expectation of the three terms above.

The last term  $T_3$  is the easiest:

$$(38) \quad \mathbb{E} T_3 \leq \mathbb{E}^{1/2} (\|b_A\|_N^4) \mathbb{P}^{1/2}(\Lambda_m^c) \lesssim \frac{1}{(NT)^{7/2}} \lesssim \frac{1}{NT}.$$

as

$$(39) \quad \mathbb{E}(\|b_A\|_N^4) \leq \frac{1}{T^2} \mathbb{E} \left( \int_0^T b_A^2(X_1(u)) du \right)^2 \leq \int b_A^4(y) f_T(y) dy < +\infty.$$

To study  $T_1, T_2$ , let us introduce the operator  $\Pi_m : \mathbb{L}^2(A, f_T(x) dx) \rightarrow S_m$  of orthogonal projection with respect to the empirical scalar product  $\langle \cdot, \cdot \rangle_N$ , *i.e.*  $\Pi_m h$  is the function of  $S_m$  given by

$$\|h - \Pi_m h\|_N^2 = \inf_{t \in S_m} \|h - t\|_N^2.$$

Simple computations show that  $\Pi_m h = \sum_{j=0}^{m-1} \tau_j \varphi_j$  where  $\vec{\tau} = (\tau_0, \dots, \tau_{m-1})' = \widehat{\Psi}_m^{-1} (\langle \varphi_j, h \rangle_N)_{0 \leq j \leq m-1}$ . Thus, we can write:

$$(40) \quad \|\hat{b}_m - b_A\|_N^2 = \|\hat{b}_m - \Pi_m b_A\|_N^2 + \|\Pi_m b_A - b_A\|_N^2 = \|\hat{b}_m - \Pi_m b_A\|_N^2 + \inf_{t \in S_m} \|b_A - t\|_N^2.$$

We have  $\Pi_m b_A = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j$  where  $\hat{a}_{(m)} = (\hat{a}_0, \dots, \hat{a}_{m-1})' = \widehat{\Psi}_m^{-1} (\langle \varphi_j, b_A \rangle_N)_{0 \leq j \leq m-1}$ . Recall that  $\hat{b}_m = \sum_{j=0}^{m-1} \hat{\theta}_j \varphi_j$  with  $\hat{\theta}_{(m)} = \widehat{\Psi}_m^{-1} \widehat{Z}_m$  (see (9)). Hence, we have  $\hat{\theta}_{(m)} - \hat{a}_{(m)} = \widehat{\Psi}_m^{-1} E_m$  (see (14)) and

$$\begin{aligned} \|\hat{b}_m - \Pi_m b_A\|_N^2 &= \frac{1}{NT} \sum_{i=1}^N \int_0^T \left( \sum_{j=0}^{m-1} (\hat{\theta}_j - \hat{a}_j) \varphi_j(X_i(u)) \right)^2 du \\ &= \frac{1}{NT} \sum_{i=1}^N \int_0^T \left[ (\hat{\theta}_{(m)} - \hat{a}_{(m)})' (\varphi_j(X_i(u)))_{0 \leq j \leq m-1} \right]^2 du \\ &= (\hat{\theta}_{(m)} - \hat{a}_{(m)})' \widehat{\Psi}_m (\hat{\theta}_{(m)} - \hat{a}_{(m)}) = E_m' \widehat{\Psi}_m^{-1} E_m. \end{aligned}$$

Now, we look at  $T_1 = \|\hat{b}_m - b_A\|_N^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} = (\|\hat{b}_m - \Pi_m b_A\|_N^2 + \inf_{t \in S_m} \|b_A - t\|_N^2) \mathbf{1}_{\Lambda_m \cap \Omega_m}$  (see (37) and (40)).

On  $\Omega_m$ , all the eigenvalues of  $\Psi_m^{-1/2} \widehat{\Psi}_m \Psi_m^{-1/2}$  belong to  $[1/2, 3/2]$  and so all the eigenvalues of  $\Psi_m^{1/2} \widehat{\Psi}_m^{-1} \Psi_m^{1/2}$  belong to  $[2/3, 2]$ . Thus on  $\Omega_m$ , we have, a.s.

$$E_m' \widehat{\Psi}_m^{-1} E_m = E_m' \Psi_m^{-1/2} \Psi_m^{1/2} \widehat{\Psi}_m^{-1} \Psi_m^{1/2} \Psi_m^{-1/2} E_m \leq 2 E_m' \Psi_m^{-1} E_m.$$

Therefore

$$\begin{aligned} \mathbb{E} \left( \|\hat{b}_m - \Pi_m b_A\|_N^2 \mathbf{1}_{\Omega_m \cap \Lambda_m} \right) &\leq 2 \mathbb{E} \left( \sum_{0 \leq j, k \leq m-1} [E_m]_j [E_m]_k [\Psi_m^{-1}]_{j,k} \right) \\ &= \frac{2}{NT^2} \sum_{0 \leq j, k \leq m-1} [\Psi_m^{-1}]_{j,k} \mathbb{E} \left( \int_0^T \varphi_j(X_1(u)) \sigma(X_1(u)) dW_1(u) \int_0^T \varphi_k(X_1(u)) \sigma(X_1(u)) dW_1(u) \right) \\ (41) \quad &= \frac{2}{NT} \sum_{0 \leq j, k \leq m-1} [\Psi_m^{-1}]_{j,k} [\Psi_{m, \sigma^2}]_{j,k} = \frac{2}{NT} \text{Tr}(\Psi_m^{-1} \Psi_{m, \sigma^2}). \end{aligned}$$

So we obtain:

$$\mathbb{E}(T_1) \leq \inf_{t \in S_m} \|b_A - t\|_{f_T}^2 + \frac{2}{NT} \text{Tr}(\Psi_m^{-1} \Psi_{m, \sigma^2}).$$

Now, we look at  $T_2 = \|\hat{b}_m - b_A\|_N^2 \mathbf{1}_{\Lambda_m \cap \Omega_m^c} \leq (\|\hat{b}_m - \Pi_m b_A\|_N^2 + \|b_A\|_N^2) \mathbf{1}_{\Lambda_m \cap \Omega_m^c}$  and find:

$$(42) \quad T_2 \leq (E_m' \widehat{\Psi}_m^{-1} E_m + \|b_A\|_N^2) \mathbf{1}_{\Lambda_m \cap \Omega_m^c}.$$

This yields, using the definition of  $\Lambda_m$  to bound  $\widehat{\Psi}_m^{-1}$  and the Cauchy-Schwarz inequality,

$$(43) \quad \mathbb{E}T_2 \leq \left( \frac{\mathfrak{c}_T NT}{L(m) \log(NT)} \mathbb{E}^{1/2}((E'_m E_m)^2) + \mathbb{E}^{1/2} \|b_A\|_N^4 \right) \mathbb{P}^{1/2}(\Omega_m^c)$$

where we have already seen that  $\mathbb{E}(\|b_A\|_N^4) \leq \int b_A^4(y) f_T(y) dy$ . The term  $\mathbb{E}[(E'_m E_m)^2]$  is ruled by the following lemma which is proved below:

**Lemma 4.** *With  $E_m$  defined in (14) (see also (13)), we have*

$$\mathbb{E}[(E'_m E_m)^2] \leq c \frac{mL^2(m)}{(NT)^2} \int \sigma_A^4(y) f_T(y) dy$$

where  $c$  is a numerical constant.

Plugging the result of Lemma 4 in (43) allows to conclude for all  $m$  satisfying (19), and  $m \leq NT$ , that  $\mathbb{E}(T_2) \leq c/(NT)^3 \leq c/(NT)$ .

Joining the bounds for the expectations of  $T_1, T_2, T_3$  gives Inequality (20).

Now, we prove (21). We have

$$(44) \quad \mathbb{E}(\|\tilde{b}_m - b_A\|_{f_T}^2) = \mathbb{E}(\|\hat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\Omega_m \cap \Lambda_m}) + \mathbb{E}(\|\hat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\Omega_m^c \cap \Lambda_m}) + \|b_A\|_{f_T}^2 \mathbb{P}(\Lambda_m^c).$$

The last r.h.s. term is bounded by applying Lemma 3.

Next, we study the first term  $\mathbb{E}(\|\hat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\Omega_m \cap \Lambda_m})$ .

Let  $b_m^{f_T}$  denote the orthogonal projection of  $b$  on  $S_m$  w.r.t. the  $\mathbb{L}^2(A, f_T(x) dx)$ -norm and set  $g = b_A - b_m^{f_T}$ , so that the bias term is equal to

$$\|g\|_{f_T} = \inf_{t \in S_m} \|t - b_A\|_{f_T}.$$

We have

$$\hat{b}_m - b_A = \hat{b}_m - \Pi_m b_A + \Pi_m b_A - b_A = \hat{b}_m - \Pi_m b_A + \Pi_m g - g.$$

where  $\Pi_m g = \Pi_m b_A - b_m^{f_T}$ . As  $g$  is orthogonal w.r.t. the  $\mathbb{L}^2(A, f_T(x) dx)$ -scalar product to  $S_m$  and thus to  $\hat{b}_m - \Pi_m b_A + \Pi_m b_A$ , we have

$$\|\hat{b}_m - b_A\|_{f_T}^2 = \|\hat{b}_m - \Pi_m b_A + \Pi_m g\|_{f_T}^2 + \|g\|_{f_T}^2.$$

We can write:

$$\mathbb{E}(\|\hat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}) \leq \|g\|_{f_T}^2 + 2\mathbb{E}(\|\hat{b}_m - \Pi_m b_A\|_{f_T}^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}) + 2\mathbb{E}(\|\Pi_m g\|_{f_T}^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}).$$

The first term is the squared bias. The second term satisfies, by definition of  $\Omega_m$  and (41),

$$2\mathbb{E}(\|\hat{b}_m - \Pi_m b_A\|_{f_T}^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}) \leq 4\mathbb{E}(\|\hat{b}_m - \Pi_m b_A\|_N^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}) \leq \frac{8}{NT} \text{Tr}(\Psi_m^{-1} \Psi_{m, \sigma^2}).$$

For the third term, we have the following result which is proved later on.

**Lemma 5.** *Under the assumptions of Proposition 1,*

$$\mathbb{E}(\|\Pi_m g\|_{f_T}^2 \mathbf{1}_{\Omega_m \cap \Lambda_m}) \leq 2 \frac{\mathfrak{c}_T}{\log(NT)} \|g\|_{f_T}^2 = 2 \frac{\mathfrak{c}_T}{\log(NT)} \inf_{t \in S_m} \|t - b_A\|_{f_T}^2.$$

Therefore, we conclude that

$$(45) \quad \mathbb{E}(\|\hat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\Lambda_m \cap \Omega_m}) \leq \left(1 + 4 \frac{\mathfrak{c}_T}{\log(NT)}\right) \inf_{t \in S_m} \|t - b_A\|_{f_T}^2 + \frac{8}{NT} \text{Tr}(\Psi_m^{-1} \Psi_{m, \sigma^2}).$$

Now, we look at  $\mathbb{E}(\|\hat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\Omega_m^c \cap \Lambda_m})$  (see (44)). We have  $\mathbb{P}(\Omega_m^c) \leq c/(NT)^7$  and  $\|\hat{b}_m - b_A\|_{f_T}^2 \leq 2\|\hat{b}_m\|_{f_T}^2 + 2\|b_A\|_{f_T}^2$ . Therefore, only the term  $\mathbb{E}[\|\hat{b}_m\|_{f_T}^2 \mathbf{1}_{\Omega_m^c \cap \Lambda_m}]$  is to be studied. We have:

$$(46) \quad \|\hat{b}_m\|_{f_T}^2 = \int \left( \sum_{j=0}^{m-1} \hat{\theta}_j \varphi_j(y) \right)^2 f_T(y) dy = \hat{\theta}'_{(m)} \Psi_m \hat{\theta}_{(m)} = \hat{Z}'_m \hat{\Psi}_m^{-1} \Psi_m \hat{\Psi}_m^{-1} \hat{Z}_m \leq \|\hat{\Psi}_m^{-1}\|_{\text{op}}^2 \|\Psi_m\|_{\text{op}} \hat{Z}'_m \hat{Z}_m.$$

We have

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

It follows by definition of  $\Lambda_m$  that

$$(47) \quad \mathbb{E}[\|\tilde{b}_m\|_{f_T}^2 \mathbf{1}_{\Lambda_m \cap \Omega_m^c}] \leq \left( \frac{c_T NT}{\log(NT)} \right)^2 \frac{1}{L(m)} \mathbb{E}^{1/2}[(\hat{Z}'_m \hat{Z}_m)^2] \mathbb{P}^{1/2}(\Omega_m^c).$$

Now, we have:  $(\hat{Z}'_m \hat{Z}_m)^2 \leq 4m \sum_{j=0}^{m-1} \langle \varphi_j, b \rangle_N^4 + 4(E'_m E_m)^2$ . By elementary computations,  $\mathbb{E}(\langle \varphi_j, b \rangle_N^4) \leq \int (\varphi_j(x) b_A(x))^4 f_T(x) dx$ . Therefore, by using Lemma 4,

$$\left( \mathbb{E}(\hat{Z}'_m \hat{Z}_m)^2 \right)^{1/2} \leq 2 \left( \sqrt{m} L(m) \left( \int b_A^4(x) f_T(x) dx \right)^{1/2} + \sqrt{c} \frac{\sqrt{m} L(m)}{NT} \left( \int \sigma_A^4(y) f_T(y) dy \right)^{1/2} \right).$$

Joining the above with (47) yields

$$(48) \quad \mathbb{E}[\|\hat{b}_m\|_{f_T}^2 \mathbf{1}_{\Omega_m^c \cap \Lambda_m}] \leq \frac{c}{NT} \left( \left( \int b_A^4(y) f_T(y) dy \right)^{1/2} + \left( \int \sigma_A^4(y) f_T(y) dy \right)^{1/2} \right).$$

So plugging (48) in (44) together with (45) yields the bound (21).  $\square$

**6.2. Proof of Lemma 4.**  $\mathbb{E}((E'_m E_m)^2) = \frac{1}{N^4 T^4} \mathbb{E}(F(M_0(T), \dots, M_{m-1}(T)))$  where  $F(x_0, \dots, x_{m-1}) = (\sum_{j=0}^{m-1} x_j^2)^2$  and  $M_j(T) = \int_0^T \left( \sum_{i=1}^N \varphi_j(X_i(u)) \sigma(X_i(u)) dW_i(u) \right)$ . By the Cauchy-Schwarz and Burkholder-Davis-Gundy inequalities, we get:

$$\begin{aligned} \mathbb{E}((E'_m E_m)^2) &\leq c \frac{m}{(NT)^4} \sum_{j=0}^{m-1} \mathbb{E} \langle M_j \rangle_T^2 \\ &= c \frac{m}{(NT)^4} \sum_{j=0}^{m-1} \mathbb{E} \left[ \left( \int_0^T \sum_{i=1}^N \varphi_j^2(X_i(u)) \sigma^2(X_i(u)) du \right)^2 \right] \\ &\leq c \frac{m}{(NT)^4} TN \sum_{i=1}^N \sum_{j=0}^{m-1} \mathbb{E} \left( \int_0^T \varphi_j^4(X_i(u)) \sigma_A^4(X_i(u)) du \right) \\ &\leq c \frac{m L^2(m)}{(NT)^2} \int \sigma_A^4(y) f_T(y) dy. \quad \square \end{aligned}$$

**6.3. Proof of Lemma 5.** To compute  $\|\Pi_m g\|_{f_T}$ , let  $(\bar{\varphi}_j)_{0 \leq j \leq m-1}$  be an orthonormal basis of  $S_m$  w.r.t. the  $\mathbb{L}^2(A, f_T(x) dx)$ -scalar product. If  $\bar{\varphi}_j = \sum_{k=0}^{m-1} \alpha_{j,k} \varphi_k$  and  $\mathbf{A}_m = (\alpha_{j,k})_{0 \leq j, k \leq m-1}$ , then

$$\text{Id}_m = \left( \int \bar{\varphi}_j \bar{\varphi}_k f_T \right)_{j,k} = \mathbf{A}_m \Psi_m \mathbf{A}'_m$$



so that  $\mathbf{A}_m$  is a square root of  $\Psi_m^{-1}$ . Let  $\widehat{\mathbf{G}}_m = (\langle \bar{\varphi}_j, \bar{\varphi}_k \rangle_N)_{j,k} = \mathbf{A}_m \widehat{\Psi}_m \mathbf{A}_m'$ . The matrix  $\widehat{\mathbf{G}}_m$  and  $\Psi_m^{-1/2} \widehat{\Psi}_m \Psi_m^{-1/2}$  have the same eigenvalues. Therefore, on  $\Omega_m$ ,  $\|\widehat{\mathbf{G}}_m - \text{Id}_m\|_{\text{op}} \leq 1/2$ , and thus  $\|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}} \leq 2$ .

Now if  $\Pi_m g = \sum_{k=0}^{m-1} \beta_k \bar{\varphi}_k$ , as  $\langle g - \Pi_m g, \bar{\varphi}_j \rangle_N = 0$  for  $j = 0, 1, \dots, m-1$ , we get  $\langle g, \bar{\varphi}_j \rangle_N = \langle \Pi_m g, \bar{\varphi}_j \rangle_N = \sum_{k=0}^{m-1} \beta_k \langle \bar{\varphi}_k, \bar{\varphi}_j \rangle_N$  so that

$$\widehat{\mathbf{G}}_m \vec{\beta}_m = (\langle g, \bar{\varphi}_j \rangle_N)_{0 \leq j \leq m-1} := \vec{d}_m,$$

where  $\vec{\beta}_m = (\beta_0 \dots \beta_{m-1})'$ . Therefore, on  $\Omega_m$ ,

$$(49) \quad \|\Pi_m g\|_{f_T}^2 = \|\vec{\beta}_m\|_{2,m}^2 = \|\widehat{\mathbf{G}}_m^{-1} \vec{d}_m\|_{2,m}^2 \leq \|\widehat{\mathbf{G}}_m^{-1}\|_{\text{op}}^2 \|\vec{d}_m\|_{2,m}^2 \leq 4 \sum_{j=0}^{m-1} \langle g, \bar{\varphi}_j \rangle_N^2.$$

Now, we note that

$$\mathbb{E}(\langle g, \bar{\varphi}_j \rangle_N) = \mathbb{E} \left( \frac{1}{NT} \sum_{i=1}^N \int \bar{\varphi}_j(X_i(u)) g(X_i(u)) du \right) = \langle \bar{\varphi}_j, g \rangle_{f_T} = 0$$

as  $g \perp^{(f_T)} \bar{\varphi}_j$ . Thus

$$\mathbb{E}[\langle g, \bar{\varphi}_j \rangle_N^2] = \text{Var} \left( \frac{1}{NT} \sum_{i=1}^N \int_0^T \bar{\varphi}_j(X_i(u)) g(X_i(u)) du \right) = \frac{1}{NT^2} \text{Var} \left( \int_0^T \bar{\varphi}_j(X_1(u)) g(X_1(u)) du \right)$$

and

$$\begin{aligned} \mathbb{E} \left[ \sum_{j=0}^{m-1} \langle g, \bar{\varphi}_j \rangle_N^2 \mathbf{1}_{\Omega_m \cap \Lambda_m} \right] &\leq \frac{1}{NT^2} \sum_{j=0}^{m-1} \mathbb{E} \left[ \left( \int_0^T \bar{\varphi}_j(X_1(u)) g(X_1(u)) du \right)^2 \right] \\ &= \frac{1}{NT^2} \mathbb{E} [\|\mathbf{A}_m \vec{v}\|_{2,m}^2] \end{aligned}$$

where  $\vec{v} = (\int_0^T \varphi_k(X_1(u)) g(X_1(u)) du)_{0 \leq k \leq m-1}$ . As  $\|\mathbf{A}_m\|_{\text{op}}^2 = \|\Psi_m^{-1}\|_{\text{op}}$ , we get

$$\begin{aligned} \mathbb{E} \left[ \sum_{j=0}^{m-1} \langle g, \bar{\varphi}_j \rangle_N^2 \mathbf{1}_{\Omega_m \cap \Lambda_m} \right] &\leq \frac{\|\Psi_m^{-1}\|_{\text{op}}}{NT^2} \mathbb{E}(\|\vec{v}\|_{2,m}^2) \\ &\leq \frac{\|\Psi_m^{-1}\|_{\text{op}}}{NT^2} \mathbb{E} \left( \sum_{j=0}^{m-1} \left( \int_0^T \varphi_j(X_1(u)) g(X_1(u)) du \right)^2 \right) \\ &\leq \frac{\|\Psi_m^{-1}\|_{\text{op}}}{N} L(m) \|g\|_{f_T}^2. \end{aligned}$$

This, under (19) and reminding (49), implies

$$\mathbb{E}(\|\Pi_m g\|_{f_T}^2 \mathbf{1}_{\Omega_m \cap \Lambda_m}) \leq \frac{2T c_T}{\log(NT)} \|g\|_{f_T}^2.$$

This gives the result of Lemma 5.  $\square$

**6.4. Proof of Proposition 2.** Property (i) follows from Proposition 2.4 in Comte and Genon-Catalot (2019)). For (ii), we can write:

$$\mathrm{Tr} \left[ \Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2} \right] \leq m \|\Psi_m^{-1/2} \Psi_{m,\sigma^2} \Psi_m^{-1/2}\|_{\mathrm{op}}$$

where

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

Now, if  $\sigma$  is bounded on  $A$ ,

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

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

**6.5. Proof of Theorem 1 and Theorem 2.** To deal with the random set  $\widehat{\mathcal{M}}_N(\theta)$  (see (23)), we introduce an additional set

$$(50) \quad \mathcal{M}_N^+(\theta) = \left\{ m \in \mathbb{N}, \quad c_{\varphi}^2 m (\|\Psi_m^{-1}\|_{\mathrm{op}}^2 \vee 1) \leq 4\theta \frac{NT}{\log(NT)} \right\} = \mathcal{M}_N(16\theta).$$

In the following, for simplicity, we shall denote  $\mathcal{M}_N, \widehat{\mathcal{M}}_N, \mathcal{M}_N^+$  for  $\mathcal{M}_N(\mathfrak{d}_T), \widehat{\mathcal{M}}_N(\mathfrak{d}_T), \mathcal{M}_N^+(\mathfrak{d}_T)$  if  $\sigma_A$  is bounded (case of Theorem 1), and for  $\mathcal{M}_N(f_T), \widehat{\mathcal{M}}_N(f_T), \mathcal{M}_N^+(f_T)$  otherwise (case of Theorem 2).

We denote by  $\widehat{M}_N$  (resp.  $M_N^+, M_N$ ) the maximal element of  $\widehat{\mathcal{M}}_N$  (resp.  $\mathcal{M}_N^+, \mathcal{M}_N$ , (see (24)). Let

$$(51) \quad \Xi_N := \left\{ \mathcal{M}_N \subset \widehat{\mathcal{M}}_N \subset \mathcal{M}_N^+ \right\}.$$

Proceeding as in Lemma 7 in Comte and Genon-Catalot (2018b) we can prove that, for the choice of  $\mathfrak{d}_T$  given in (25) with  $c_0$  a large enough numerical value ( $c_0 = 96$  suits), and, for  $c$  a positive constant,

$$(52) \quad \mathbb{P}(\Xi_N^c) = \mathbb{P} \left( \left\{ \mathcal{M}_N \not\subset \widehat{\mathcal{M}}_N \text{ or } \widehat{\mathcal{M}}_N \not\subset \mathcal{M}_N^+ \right\} \right) \leq \frac{c}{(NT)^4}.$$

We write the decomposition:  $\hat{b}_m - b_A = (\hat{b}_m - b_A) \mathbf{1}_{\Xi_N} + (\hat{b}_m - b_A) \mathbf{1}_{\Xi_N^c}$ . As for the study of  $T_2$  defined by (37), starting from (42), we get

$$\|b_A - \hat{b}_m\|_N^2 \mathbf{1}_{\Xi_N^c} \leq (E'_{\hat{m}} \widehat{\Psi}_{\hat{m}}^{-1} E_{\hat{m}} + \|b_A\|_N^2) \mathbf{1}_{\Xi_N^c}.$$

Now, as  $\hat{m} \in \widehat{\mathcal{M}}_N$ ,

$$(E'_{\hat{m}} \widehat{\Psi}_{\hat{m}}^{-1} E_{\hat{m}})^2 \leq \|\widehat{\Psi}_{\hat{m}}^{-1}\|_{\mathrm{op}}^2 (E'_{NT} E_{NT})^2 \leq \frac{\mathfrak{d}_T}{c_{\varphi}^2} \frac{NT}{\log(NT)} (E'_{NT} E_{NT})^2.$$

Lemma 4 yields  $\mathbb{E}[(E'_{NT} E_{NT})^2] \leq c c_{\varphi}^4 (NT) \int \sigma_A^4(y) f_T(y) dy$  and thus  $\mathbb{E}[(E'_{\hat{m}} \widehat{\Psi}_{\hat{m}}^{-1} E_{\hat{m}})^2] \lesssim (NT)^2$ . This together with (52) implies, for  $C$  a constant depending on  $\int \sigma_A^4 f_T, \int b_A^4 f_T$ , and  $\mathfrak{d}_T$ ,

$$\mathbb{E}[\|b_A - \hat{b}_m\|_N^2 \mathbf{1}_{\Xi_N^c}] \leq \frac{C}{NT}.$$

It remains to study  $\mathbb{E}[\|\hat{b}_{\hat{m}} - b_A\|_N^2 \mathbf{1}_{\Xi_N}]$ . To begin with, recall that  $\gamma_N(\hat{b}_m) = -\|\hat{b}_m\|_N^2$ . Consequently, we can write

$$\hat{m} = \arg \min_{m \in \widehat{\mathcal{M}}_N} \{\gamma_N(\hat{b}_m) + \widehat{\text{pen}}(m)\},$$

where  $\widehat{\text{pen}}(m) = \text{pen}_1(m)$  defined by (26) if  $\sigma$  is bounded on  $A$  and  $\widehat{\text{pen}}(m) = \widehat{\text{pen}}_2(m)$  defined by (29) otherwise. Thus, we have, for any  $m \in \widehat{\mathcal{M}}_N$ , and any  $b_m \in S_m$ ,

$$(53) \quad \gamma_N(\hat{b}_{\hat{m}}) + \widehat{\text{pen}}(\hat{m}) \leq \gamma_N(b_m) + \widehat{\text{pen}}(m).$$

On  $\Xi_N = \{\mathcal{M}_N \subset \widehat{\mathcal{M}}_N \subset \mathcal{M}_N^+\}$ ,  $\hat{m} \leq \widehat{M}_N \leq M_N^+$  and either  $M_N \leq \hat{m} \leq \widehat{M}_N \leq M_N^+$  or  $\hat{m} < M_N \leq \widehat{M}_N \leq M_N^+$ . In the first case,  $\hat{m}$  is upper and lower bounded by deterministic bounds and (53) a fortiori holds for any  $m \in \mathcal{M}_N$ ; and in the second cas,

$$\hat{m} = \arg \min_{m \in \mathcal{M}_N} \{\gamma_N(\hat{b}_m) + \widehat{\text{pen}}(m)\}.$$

Thus, on  $\Xi_N$ , (53) holds for any  $m \in \mathcal{M}_N$  and any  $b_m \in S_m$ . The decomposition  $\gamma_n(t) - \gamma_n(s) = \|t - b\|_N^2 - \|s - b\|_N^2 + 2\nu_N(t - s)$ , where  $\nu_N(t)$  is defined by (13), yields, for any  $m \in \mathcal{M}_N$  and any  $b_m \in S_m$ ,

$$\|\hat{b}_{\hat{m}} - b_A\|_N^2 \leq \|b_m - b_A\|_N^2 + 2\nu_N(\hat{b}_{\hat{m}} - b_m) + \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\hat{m}).$$

We introduce the unit ball and the set:

$$B_{m,m'}^{f_T}(0,1) = \{t \in S_m + S_{m'}, \|t\|_{f_T} = 1\}, \quad \Omega_N = \bigcap_{m \in \mathcal{M}_N^+} \Omega_m,$$

where  $\Omega_m$  is defined by (36). We split again:

$$\mathbb{E}[\|\hat{b}_{\hat{m}} - b_A\|_N^2 \mathbf{1}_{\Xi_N}] = \mathbb{E}[\|\hat{b}_{\hat{m}} - b_A\|_N^2 \mathbf{1}_{\Xi_N \cap \Omega_N}] + \mathbb{E}[\|\hat{b}_{\hat{m}} - b_A\|_N^2 \mathbf{1}_{\Xi_N \cap \Omega_N^c}].$$

The term  $\mathbb{E}(\|\hat{b}_{\hat{m}} - b_A\|_N^2 \mathbf{1}_{\Omega_N^c \cap \Xi_N})$  is bounded analogously as  $\mathbb{E}(\|\hat{b}_{\hat{m}} - b_A\|_N^2 \mathbf{1}_{\Xi_N^c})$ , using that by Lemma 3,  $\mathbb{P}(\Xi_N \cap \Omega_N^c) \leq \sum_{m \in \mathcal{M}_N^+} \mathbb{P}(\Omega_m^c) \leq c'/(NT)^6$ .

Then, we study the expectation on  $\Xi_N \cap \Omega_N$ . On  $\Omega_N$ , the following inequality holds:  $\|t\|_{f_T}^2 \leq 2\|t\|_N^2$ ,  $\forall t \in S_{M_N^+}$ . We get, on  $\Xi_N \cap \Omega_N$ ,

$$(54) \quad \begin{aligned} \|\hat{b}_{\hat{m}} - b_A\|_N^2 &\leq \|b_m - b_A\|_N^2 + \frac{1}{8} \|\hat{b}_{\hat{m}} - b_m\|_{f_T}^2 + 8 \sup_{t \in B_{m,m}^{f_T}(0,1)} \nu_N^2(t) + \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\hat{m}) \\ &\leq \left(1 + \frac{1}{2}\right) \|b_m - b_A\|_N^2 + \frac{1}{2} \|\hat{b}_{\hat{m}} - b_A\|_N^2 + 8 \left( \sup_{t \in B_{m,m}^{f_T}(0,1)} \nu_N^2(t) - p(m, \hat{m}) \right)_+ \\ &\quad + \widehat{\text{pen}}(m) + 8p(m, \hat{m}) - \widehat{\text{pen}}(\hat{m}). \end{aligned}$$

Note that, in the case  $\|\sigma_A\|_\infty < +\infty$ ,  $\text{pen}_1(m) = \widehat{\text{pen}}(m)$  is deterministic. Therefore, we can complete the proof of the first inequality of Theorem 1 applying the following Lemma.

**Lemma 6.** *Assume that  $\|\sigma_A\|_\infty < +\infty$ . Then there exists a numerical constant  $\tau$  such that for  $p(m, m') = \tau \|\sigma_A\|_\infty^2 (m + m')/(NT)$ ,*

$$\mathbb{E} \left[ \left( \sup_{t \in B_{m,m}^{f_T}(0,1)} \nu_N^2(t) - p(m, \hat{m}) \right)_+ \mathbf{1}_{\Xi_N \cap \Omega_N} \right] \leq c \|\sigma_A\|_\infty^2 \frac{1}{NT}.$$

Indeed, we choose  $\kappa \geq 8\tau$  in  $\text{pen}_1(m)$  and the first inequality of Theorem 1 follows. For the second inequality, we proceed as in the proof of Theorem 2 in Comte and Genon-Catalot (2018b).  $\square$

**Proof of Lemma 6.** When  $\sigma_A$  is bounded, for  $t$  a  $A$ -supported function,

$$\langle M \rangle_T = \int_0^T \sum_{i=1}^N t^2(X_i(u)) \sigma^2(X_i(u)) du \leq NT \|\sigma_A^2\|_\infty \|t\|_N^2.$$

Thus, by Lemma 2, we obtain:  $\mathbb{P}(\nu_N(t) \geq \varepsilon, \|t\|_N^2 \leq v^2) \leq \exp(-NT\varepsilon^2/(2\|\sigma_A^2\|_\infty v^2))$ . Afterwards, as in Comte *et al.* (2007), we use the  $\mathbb{L}^2$ -chaining technique described in Baraud *et al.* (Section 7, p.44-47, Lemma 7.1, with  $s^2 = \|\sigma_A^2\|_\infty/T$ ).  $\square$

Now we no longer assume  $\sigma_A$  bounded and we consider Laguerre and Hermite bases to complete the proof of Theorem 2. We have the following Lemma.

**Lemma 7.** *Assume (A1)-(A4). Then there exists a numerical value  $\tau_1$  such that  $\nu_N(t)$  satisfies*

$$\mathbb{E} \left[ \left( \sup_{t \in B_{\hat{m}, m}^{f_T}(0,1)} \nu_N^2(t) - p(m, \hat{m}) \right)_+ \mathbf{1}_{\Xi_N \cap \Omega_N} \right] \leq \frac{C}{NT}$$

where  $p(m, m') = \sup(p(m), p(m'))$  with

$$p(m) = \tau_1 \frac{m(1 + \ell_m) \|\Psi_m^{-1/2} \Psi_{m, \sigma^2} \Psi_m^{-1/2}\|_{\text{op}}}{NT}.$$

For  $\kappa_1 \geq 8\tau_1$ ,  $8p(m, m') \leq \text{pen}(m) + \text{pen}(m')$ . Therefore, plugging the result of Lemma 7 in (54) and taking expectation yield that

$$\begin{aligned} \frac{1}{2} \mathbb{E}(\|\hat{b}_{\hat{m}} - b_A\|_N^2 \mathbf{1}_{\Xi_N \cap \Omega_N}) &\leq \frac{3}{2} \|b_m - b_A\|_N^2 + \text{pen}(m) + \frac{C}{NT} \\ &\quad + \mathbb{E}(\widehat{\text{pen}}(m) \mathbf{1}_{\Xi_N \cap \Omega_N}) + \mathbb{E}[(\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m}))_+ \mathbf{1}_{\Xi_N \cap \Omega_N}]. \end{aligned}$$

Now we have the following Lemma, the proof of which is omitted as it is similar to Lemma 6.5 in Comte and Genon-Catalot (2019).

**Lemma 8.** *Under the assumptions of Theorem 2, there exist constants  $c_1, c_2 > 0$  such that for  $m \in \mathcal{M}_N$  and  $\hat{m} \in \widehat{\mathcal{M}}_N$ ,*

$$\begin{aligned} \mathbb{E}(\widehat{\text{pen}}(m) \mathbf{1}_{\Xi_N \cap \Omega_N}) &\leq c_1 \text{pen}(m) + \frac{c_2}{NT} \\ \mathbb{E}[(\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m}))_+ \mathbf{1}_{\Xi_N \cap \Omega_N}] &\leq \frac{c_2}{NT}. \end{aligned}$$

Note that  $c_2$  contains  $\int |\sigma_A|^{4+56/\beta} f_T$ . Lemma 8 concludes the study of the expectation of the empirical risk on  $\Xi_N \cap \Omega_N$ . This gives the first inequality of Theorem 2. The second inequality is obtained following the lines of the proof of Theorem 2 in Comte and Genon-Catalot (2018b).  $\square$

**6.6. Proof of Lemma 7.** Define the set

$$\Omega_{m, \sigma^2} = \left\{ \left| \frac{\|t\sigma\|_N^2}{\|t\sigma\|_{f_T}^2} - 1 \right| \leq \frac{1}{2}, \forall t \in S_m \setminus \{0\} \right\}, \quad \Omega_{N, \sigma^2} = \bigcap_{m \in \mathcal{M}_N^+} \Omega_{m, \sigma^2}.$$

We need the following Lemma, similar to Lemma 3, which determines the constant  $f_T$ .

**Lemma 9.** *Consider Laguerre or Hermite basis. Assume that (A1)-(A4) hold. Then,  $\mathbb{P}(\Omega_{m,\sigma^2}^c) \leq c/(NT)^6$  and  $\mathbb{P}(\Omega_{N,\sigma^2}^c) \leq c/(NT)^5$ .*

Note that

$$\sup_{\|t\|_{f_T}=1} \|t\sigma\|_{f_T}^2 = \sup_{\|\Psi_m^{1/2}\bar{a}\|_{2,m}=1} {}^t\bar{a}\Psi_{m,\sigma^2}\bar{a} = \sup_{\|\bar{u}\|_{2,m}=1} {}^t\bar{u}\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\bar{u} = \|\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\|_{\text{op}}.$$

This implies

$$(55) \quad \sup_{t \in S_m, \|t\|_{f_T}=1} \nu_N^2(t) \leq \|\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\|_{\text{op}} \sup_{t \in S_m, \|t\sigma\|_{f_T}=1} \nu_N^2(t).$$

We have

$$\mathbb{E} \left[ \left( \sup_{t \in S_m, \|t\|_{f_T}=1} \nu_n^2(t) - p(m) \right)_+ \mathbf{1}_{\Xi_N \cap \Omega_N} \right] = \mathbb{E}[\mathbf{T}_1^*(m)] + \mathbb{E}[\mathbf{T}_2^*(m)]$$

with  $\mathbf{A}(m) := \left( \sup_{t \in S_m, \|t\|_{f_T}=1} \nu_n^2(t) - p(m) \right)_+$ ,  $\mathbf{T}_1^*(m) = \mathbf{A}(m)\mathbf{1}_{\Xi_N \cap \Omega_N \cap \Omega_{N,\sigma^2}}$ , and  $\mathbf{T}_2^*(m) = \mathbf{A}(m)\mathbf{1}_{\Xi_N \cap \Omega_N \cap \Omega_{N,\sigma^2}^c}$ . Now, by using (55), we have

$$\mathbb{E}[\mathbf{T}_1^*(m)] \leq \|\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\|_{\text{op}} \mathbb{E} \left[ \left( \sup_{t \in S_m, \|t\sigma\|_{f_T}=1} \nu_N^2(t) - q(m) \right)_+ \mathbf{1}_{\Omega_{N,\sigma^2}} \right],$$

with  $q(m) = \tau_1 m(1 + \ell_m)/(NT)$ .

Following the proof of Proposition 3 in Comte *et al.* (2007) (see also Baraud *et al.*, 2001, Theorem 3.1 and Proposition 6.1, in the regression model case), there exists a numerical constant  $\tau_1$  such that

$$\mathbb{E} \left[ \left( \sup_{t \in S_m, \|t\sigma\|_{f_T}=1} \nu_N^2(t) - q(m) \right)_+ \mathbf{1}_{\Omega_{N,\sigma^2}} \right] \leq c \frac{e^{-m\ell_m}}{NT}$$

As a consequence, for the same numerical constant  $\tau_1$ ,

$$\mathbb{E}[\mathbf{T}_1^*(m)] \leq c \frac{e^{-m\ell_m}}{NT} \|\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\|_{\text{op}}.$$

Moreover  $\|\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\|_{\text{op}} \leq \|\Psi_m^{-1}\|_{\text{op}} \|\Psi_{m,\sigma^2}\|_{\text{op}}$ , and we have

$$\|\Psi_{m,\sigma^2}\|_{\text{op}} = \sup_{\|\bar{a}\|_{2,m}=1} {}^t\bar{a}\Psi_{m,\sigma^2}\bar{a} = \sup_{\|\bar{a}\|_{2,m}=1} \int \left( \sum_{j=0}^{m-1} a_j \varphi_j(y) \right)^2 \sigma^2(y) f_T(y) dy \leq c_\varphi^2 m \int \sigma^2 f_T,$$

Therefore, for  $c_1 = c_\varphi^2 \int \sigma^2 f_T$ ,

$$\mathbb{E}[\mathbf{T}_1^*(m \vee \hat{m})] \leq \sum_{m \in \mathcal{M}_N^+} \mathbb{E}[\mathbf{T}_1^*(m)] \leq c_1 \sum_{m \in \mathcal{M}_N^+} m e^{-m\ell_m} \|\Psi_m^{-1}\|_{\text{op}} \leq c_1 \Sigma$$

under condition (30). Thus, we get

$$\mathbb{E} \left[ \left( \sup_{t \in S_{m \vee \hat{m}}, \|t\|_{f_T}=1} \nu_N^2(t) - p(m, \hat{m}) \right)_+ \mathbf{1}_{\Xi_N \cap \Omega_N \cap \Omega_{N,\sigma^2}} \right] \leq \frac{C}{NT}.$$

Now, we have to study  $\mathbb{E}[\mathbf{T}_2^*(m \vee \hat{m})]$ . First,

$$p(m) \leq \kappa_1 c_\varphi^2 \frac{m^2(1 + \ell_m)}{NT} \|\Psi_m^{-1}\|_{\text{op}} \int \sigma^2 f_T \leq Cm \|\Psi_m^{-1}\|_{\text{op}}^2 \leq C' NT$$

as  $\|\Psi_m^{-1}\|_{\text{op}} \geq m$  under **(A3)** and  $m \in \mathcal{M}_N^+$ . This yields

$$\mathbb{E}[p(m, \hat{m}) \mathbf{1}_{\Xi_N \cap \Omega_N \cap \Omega_{N, \sigma^2}^c}] \leq CNT \mathbb{P}(\Omega_{N, \sigma^2}^c) \leq c/(NT)^4.$$

Second,

$$\mathbb{E}[(\sup_{t \in S_{m \vee \hat{m}}, \|t\|_{f_T}=1} \nu_N^2(t)) \mathbf{1}_{\Xi_N \cap \Omega_N \cap \Omega_{N, \sigma^2}^c}] \leq \mathbb{E}^{1/2}[\sup_{t \in S_{M_N^+}, \|t\|_{f_T}=1} \nu_N^4(t)] \mathbb{P}^{1/2}(\Omega_N \cap \Omega_{N, \sigma^2}^c).$$

Then, we write, setting  $M = M_n^+$  for sake of simplicity,

$$\begin{aligned} & \mathbb{E}(\sup_{t \in S_M, \|t\|_{f_T}=1} \nu_N^4(t)) \leq M \sum_{k=0}^{M-1} \mathbb{E} \nu_N^4(\sum_{j=0}^{M-1} [\Psi_M^{-1/2}]_{jk} \varphi_j) = M \sum_{k=0}^{M-1} \mathbb{E} \nu_N^4([\Psi_M^{-1/2} \varphi]_k) \\ & \leq c \frac{M}{(NT)^4} \sum_{k=0}^{M-1} \mathbb{E} \left( \int_0^T \sum_{i=1}^N ([\Psi_M^{-1/2} \varphi(X_i(s))]_k)^2 \sigma^2(X_i(s)) ds \right)^2 \\ & \leq \frac{cM}{(NT)^2} \int (\sum_{k=0}^{M-1} [\Psi_M^{-1/2} \varphi(y)]_k^2)^2 \sigma_A^4(y) f_T(y) dy \leq \frac{cM}{(NT)^2} \int (\sum_{j=0}^{M-1} \varphi_j^2(y))^2 \|\Psi_M^{-1}\|_{\text{op}}^2 \sigma_A^4(y) f_T(y) dy \\ & \leq cc_\varphi^4 \frac{M^3}{(NT)^2} \|\Psi_M^{-1}\|_{\text{op}}^2 \int \sigma_A^4(y) f_T(y) dy \leq CNT \int \sigma_A^4(y) f_T(y) dy. \end{aligned}$$

Thus,

$$\mathbb{E}[(\sup_{t \in S_{m \vee \hat{m}}, \|t\|_{f_T}=1} \nu_N^2(t)) \mathbf{1}_{\Xi_N \cap \Omega_N \cap \Omega_{N, \sigma^2}^c}] \leq c(NT)^{1/2}/(NT)^{5/2} = c/(NT)^2.$$

We obtain  $\mathbb{E}[\mathbf{T}_2^*(m \vee \hat{m})] \lesssim 1/(NT)^2$ . This ends the proof of Lemma 7.  $\square$

**Proof of Lemma 9.** Analogously as for  $\Omega_m$ , we have

$$\Omega_{m, \sigma^2} = \left\{ \|\Psi_{m, \sigma^2}^{-1/2} \widehat{\Psi}_{m, \sigma^2} \Psi_{m, \sigma^2}^{-1/2} - \text{Id}_m\|_{\text{op}} > \frac{1}{2} \right\}.$$

Therefore, we apply the Chernoff matrix inequality stated in Theorem 1.1 of Tropp (2012). To that aim, we write  $\Psi_{m, \sigma^2}^{-1/2} \widehat{\Psi}_{m, \sigma^2} \Psi_{m, \sigma^2}^{-1/2}$  as a sum of independent matrices

$$\Psi_{m, \sigma^2}^{-1/2} \widehat{\Psi}_{m, \sigma^2} \Psi_{m, \sigma^2}^{-1/2} = \frac{1}{N} \sum_{i=1}^N \mathbf{K}_{m, \sigma^2}(X_i),$$

$$\text{with } \mathbf{K}_{m, \sigma^2}(X_i) = \Psi_{m, \sigma^2}^{-1/2} \left( \frac{1}{T} \int_0^T \varphi_j(X_i(u)) \varphi_k(X_i(u)) \sigma^2(X_i(u)) du \right)_{0 \leq j, k \leq m-1} \Psi_{m, \sigma^2}^{-1/2}.$$

Clearly,  $\mathbb{E}(\mathbf{K}_{m, \sigma^2}(X_i)) = \text{Id}_m$ , so that  $\mu_{\min} = \mu_{\max} = 1$  and

$$\mathbb{P}(\Omega_{m, \sigma^2}^c) \leq 2m \exp(-c_T(1/2) \frac{NT}{R})$$

with  $c_T(\delta) = (\delta + (1 - \delta) \log(1 - \delta))/T$  and  $R$  is an upper bound on the largest eigenvalue of  $\mathbf{K}_{m,\sigma^2}(X_1)$ . Now we have a.s.

$$\begin{aligned} \|\mathbf{K}_{m,\sigma^2}(X_1)\|_{\text{op}} &= \sup_{\|\bar{x}\|_{2,m}=1, y=\Psi_{m,\sigma^2}^{-1/2}x} \frac{1}{T} \int_0^T \left( \sum_{j=0}^{m-1} y_j \varphi_j(X_i(u)) \right)^2 \sigma^2(X_i(u)) du \\ &\leq \|\Psi_{m,\sigma^2}^{-1}\|_{\text{op}} \frac{1}{T} \int_0^T \sum_{j=0}^{m-1} \varphi_j^2(X_i(u)) \sigma^2(X_i(u)) du \end{aligned}$$

Now, we use that  $\sigma^2(x) \leq K(1+x^2)$  with  $K$  known. If  $\varphi_j = \ell_j$ , the Laguerre basis on  $A = \mathbb{R}^+$ , we have  $|\ell_j|^2 \leq 2$  and (see *e.g.* Comte and Genon-Catalot, 2018a, Section 8):  $x\ell_j(x) = -\frac{j+1}{2}\ell_{j+1} + (j + \frac{1}{2})\ell_j(x) - \frac{j}{2}\ell_{j-1}(x)$ . This implies,

$$\begin{aligned} \|\mathbf{K}_{m,\sigma^2}(X_1)\|_{\text{op}} &\leq K(2m + 9m^3 + 9m^2 + m) \|\Psi_{m,\sigma^2}^{-1}\|_{\text{op}} \leq K(3m + 18m^3) \|\Psi_{m,\sigma^2}^{-1}\|_{\text{op}} \\ &\leq 21Km^3 \|\Psi_{m,\sigma^2}^{-1}\|_{\text{op}} := R. \end{aligned}$$

If  $\varphi_j = h_j$ , the Hermite basis on  $A = \mathbb{R}$ , we have  $|h_j| \leq C_\infty(j+1)^{-1/12}$ ,  $j = 0, 1, \dots$  (with the constant  $C_\infty$  given in Szegö (1975)) and (see *e.g.* Comte and Genon-Catalot, 2018a, Section 8):

$$2xh_j(x) = \sqrt{2(j+1)}h_{j+1} + \sqrt{2j}h_{j-1}(x).$$

This yields

$$\|\mathbf{K}_{m,\sigma^2}(X_1)\|_{\text{op}} \leq KC_\infty^2(m^{5/6} + 3m^{11/6}) \|\Psi_{m,\sigma^2}^{-1}\|_{\text{op}} \leq 2KC_\infty^2 m^{11/6} \|\Psi_{m,\sigma^2}^{-1}\|_{\text{op}} := R.$$

Let us note  $R = Bm^b \|\Psi_{m,\sigma^2}^{-1}\|_{\text{op}}$  with  $(B, b) = (21K, 3)$  with Laguerre and  $(B, b) = (2KC_\infty^2, 11/6)$  for Hermite basis. We obtain

$$\mathbb{P}(\Omega_{m,\sigma^2}^c) \leq 2m \exp\left(-c_T(1/2) \frac{NT}{Bm^b \|\Psi_{m,\sigma^2}^{-1}\|_{\text{op}}}\right) \leq \frac{1}{(NT)^6}$$

if  $m \leq NT$  and

$$(56) \quad Bm^b \|\Psi_{m,\sigma^2}^{-1}\|_{\text{op}} \leq c_T(1/2) \frac{NT}{7 \log(NT)}.$$

Now, for  $\sigma^2(x) \geq \sigma_0^2$ , we get  $\|\Psi_{m,\sigma^2}^{-1}\|_{\text{op}} \leq \sigma_0^2 \|\Psi_m^{-1}\|_{\text{op}}$ , so that the above condition is satisfied if

$$B\sigma_0^2 m^b \|\Psi_m^{-1}\|_{\text{op}} \leq c_T(1/2) \frac{NT}{7 \log(NT)}.$$

By definition of  $\mathcal{M}_N$  and under **(A3)**, we have  $m^b \|\Psi_m^{-1}\|_{\text{op}} \leq m \|\Psi_m^{-1}\|_{\text{op}}^2$  so that for

$$\mathfrak{f}_T = \mathfrak{d}_T \wedge \frac{c_T(1/2)}{7B\sigma_0^2} = \mathfrak{d}_T \wedge \frac{1 - \log(2)}{14TB\sigma_0^2}$$

condition (56) is fulfilled and the bound is true.  $\square$

**6.7. Proof of Proposition 3.** We use the following representation (see *e.g.* Rogers (1985)). For (i) Set  $B(y) = \int_0^y b(u)du$ . Then,

$$p_t(x, y) = \frac{1}{\sqrt{2\pi t}} \exp(B(y) - B(x) - \frac{(y-x)^2}{2t}) \mathbb{E}(\exp(-\frac{t}{2} \int_0^1 g((1-u)x + uy + \sqrt{t}B_u^0)du)),$$

$g = b^2 + b'$  et  $(B_u^0, u \in [0, 1])$  is a standard Brownian bridge. As  $|b'| \leq M$  and  $|b| \leq M$ , then,

$$p_t(x, y) \leq \frac{1}{\sqrt{2\pi t}} \exp[M|y-x| + M\frac{t}{2} - \frac{(y-x)^2}{2t}] \leq \frac{1}{\sqrt{2\pi t}} \exp[M\frac{t}{2} + 2M^2t - \frac{(y-x)^2}{4t}].$$

it follows that

$$f_T(y) \leq \frac{2T^{1/2}}{\sqrt{2\pi}} \exp[M\frac{T}{2} + 2M^2T]$$

which implies (i).

For (ii), we consider the model  $dX(t) = b(X(t))dt + \sigma(X(t))dW_t$ , where  $b, \sigma$  are functions from  $\mathbb{R}$  to  $\mathbb{R}$ . Setting  $F(\cdot) = \int_0^\cdot \frac{1}{\sigma(u)}du$ , the process  $Y_t = F(X(t))$  satisfies

$$dY_t = \alpha(Y_t)dt + dW_t,$$

with  $\alpha(y) = \frac{b(F^{-1}(y))}{\sigma(F^{-1}(y))} - \frac{1}{2}\sigma'(F^{-1}(y))$ . The transition density  $p_t(x, x')$  of  $X$  is linked to the transition density  $q_t(y, y')$  of  $Y$  by:  $p_t(x, x') = q_t(F(x), F(x'))1/\sigma(x')$ . As  $\sigma', \sigma''$  are bounded and obtain that  $\|f_T\|_\infty < +\infty$ .  $\square$

## 7. A THEORETICAL TOOL

**Theorem 3.** (*Matrix Chernoff, Tropp (2012)*) Consider a finite sequence  $\{\mathbf{X}_k\}$  of independent, random, self-adjoint matrices with dimension  $d$ . Assume that each random matrix satisfies

$$\mathbf{X}_k \succcurlyeq 0 \quad \text{and} \quad \lambda_{\max}(\mathbf{X}_k) \leq R \quad \text{almost surely.}$$

Define  $\mu_{\min} := \lambda_{\min}(\sum_k \mathbb{E}(\mathbf{X}_k))$  and  $\mu_{\max} := \lambda_{\max}(\sum_k \mathbb{E}(\mathbf{X}_k))$ . Then

$$\mathbb{P} \left\{ \lambda_{\min} \left( \sum_k \mathbf{X}_k \right) \leq (1 - \delta)\mu_{\min} \right\} \leq d \left[ \frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right]^{\mu_{\min}/R} \quad \text{for } \delta \in [0, 1] \quad \text{and}$$

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{X}_k \right) \geq (1 + \delta)\mu_{\max} \right\} \leq d \left[ \frac{e^{\delta}}{(1 + \delta)^{1+\delta}} \right]^{\mu_{\max}/R} \quad \text{for } \delta \geq 0.$$

## REFERENCES

- [1] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition.
- [2] Baraud, Y., Comte, F. and Viennet, G. (2001). Model selection for (auto)-regression with dependent data. *ESAIM P & S*, **5**, 33-49.
- [3] Birgé, L. and Massart, P. (2007). Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields* **138**, 33-73.
- [4] Comte, F. and Genon-Catalot, V. (2018a). Laguerre and Hermite bases for inverse problems. *Journal of the Korean Statistical Society*, **47**, 273-296.
- [5] Comte, F. and Genon-Catalot, V. (2018b). Regression function estimation on non compact support as a partly inverse problem. *Preprint MAP5 2018-01*. To appear in the *Annals of the Institute of Statistical Mathematics*.
- [6] Comte, F. and Genon-Catalot, V. (2019). Regression function estimation on non compact support in an heteroskedastic model. Preprint hal-02009555.
- [7] Comte, F., Genon-Catalot, V. and Samson, A. (2013). Nonparametric estimation for stochastic differential equations with random effects. *Stochastic processes and their Applications*, **123**, 2522-2551.



- [8] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, **13**, 514-543.
- [9] Delattre M., Genon-Catalot V. and Samson A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics*, **40**, 322-343.
- [10] Delattre M., Genon-Catalot V. and Larédo, C. (2018). Parametric inference for discrete observations of diffusion processes with mixed effects. *Stochastic processes and their Applications*, **128**, 1929-1957.
- [11] Delattre, M. and Lavielle, M. (2013). Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Stat. Interface* **6**, 519-532.
- [12] Denis, C., Dion, C. and Martinez, M. (2018). Consistent procedures for multiclass classification of discrete diffusion paths. Preprint hal-01869545.
- [13] Dion, C. and Genon-Catalot, V. (2016). Bidimensional random effect estimation in mixed stochastic differential equations. *Statistical Inference for Stochastic processes*, **19**, 131-158.
- [14] Ditlevsen, S. and De Gaetano, A. (2005). Mixed effects in stochastic differential equation models. *REVSTAT* **3**, 137-153.
- [15] Hsiao, C. (2003). *Analysis of panel data*. Cambridge University Press, Second Edition. Cambridge.
- [16] Iacus, S. M., (2010). *Simulation and inference for stochastic differential equations. With R examples*. Springer. maxima of empirical processes. *Ann. Probab.* **33**, 1060-1077.
- [17] Kessler, M., Lindner, A. and Sørensen, M., Editors (2012). *Statistical methods for stochastic differential equations*. CRC press. Taylor & Francis Group. Boca Raton.
- [18] Kutoyants, Y.A., (1984). *Parameter estimation for stochastic processes*. Berlin: Heldermann.
- [19] Kutoyants, Y.A., (2004). *Statistical inference for ergodic diffusion processes*. Springer, London.
- [20] Overgaard, R., Jonsson, N., Tornøe, C. and Madsen, H. (2005). Non-linear mixed effects models with stochastic differential equations: Implementation of an estimation algorithm. *J. Pharmacokinet. Pharmacodyn.* **32**, 85-107.
- [21] Picchini, U., De Gaetano, A. and Ditlevsen, S. (2010). Stochastic differential mixed-effects models. *Scand. J. Statist.* **37**, 67-90.
- [22] Picchini, U. and Ditlevsen, S. (2011). Practical estimation of high dimensional stochastic differential mixed-effects models. *Comput. Statist. Data Anal.* **55**, 1426-1444.
- [23] Ramsay, J.O. and Silverman, B.W. (2007). *Applied functional data analysis: Methods and case studies*. Springer.
- [24] Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*. Third edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 293. Springer-Verlag, Berlin.
- [25] Rogers, L.C.G. (1985). Smooth transition densities for one-dimensional diffusions. *Bulletin of the London Mathematical Society* **17**, 157-161.
- [26] Rogers, L.C.G. and Williams, D. (1990). *Diffusions, Markov Processes and Martingales, Volume 2: Ito Calculus*. John Wiley & Sons, Chichester.
- [27] Szegő, G. (1975) *Orthogonal polynomials*. Fourth edition. American Mathematical Society, Colloquium Publications, Vol. XXIII. American mathematical Society, Providence, R.I.
- [28] Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389-434.
- [29] Wang, J.-L., Chiou, J.-M. and Mueller, H.-G. (2016). Review of functional data analysis. *Annual Review of Statistics and its application*, **3**, 257-295.