The Annals of Statistics
2020, Vol. 48, No. 6, 3336–3365
https://doi.org/10.1214/19-AOS1933

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

BY FABIENNE COMTE* AND VALENTINE GENON-CATALOT†

MAP5 UMR8145, Université de Paris, [*] *fabienne.comte@parisdescartes.fr;* [†] *valentine.genon-catalot@parisdescartes.fr*

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, the assumptions are discussed and simulation experiments are reported.

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:

$$dX_i(t) = b(X_i(t))\,dt + \sigma(X_i(t))\,dW_i(t), \qquad X_i(0) = x_0, \quad i = 1, \dots, N, \tag{1.1}$$

where $x_0 \in \mathbb{R}$ is known, $(W_1, \dots, W_N)$ are independent standard Brownian motions. The drift function $b : \mathbb{R} \to \mathbb{R}$ 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, Chiou and Mueller (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 nonlinear 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), Picchini, De Gaetano and Ditlevsen (2010), Picchini and Ditlevsen (2011), Comte, Genon-Catalot and Samson (2013), Delattre and Lavielle (2013), Delattre, Genon-Catalot and Samson (2013), Dion and Genon-Catalot (2016), Delattre, Genon-Catalot and Larédo (2018)). Note that *i.i.d.* samples of stochastic differential equations have been used recently for multiclass classification of diffusions (see Denis, Dion and Martinez (2018)). However, the need of flexibility to deal with the information contained in functional data analysis make it preferable to use a nonparametric approach.

Drift estimation for one-dimensional diffusion processes has been widely investigated since the 1980s. 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

Received March 2019; revised September 2019.
MSC2020 subject classifications. 62G07, 62M05.
Key words and phrases. Diffusion process, Hermite basis, Laguerre basis, model selection, nonparametric drift estimation, projection estimators.

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, Lindner and Sørensen (2012)).

In our context, ergodicity is not required for Model (1.1), T is fixed and the asymptotic framework is N tends to infinity. The diffusion coefficient σ is supposed to be known as it is identified from a continuous observation of the sample paths. Here, we emphasize that it is possible by empirical estimators to estimate the mean and the covariance operator of the diffusion. But, it is not easy to deduce "good" estimators of the drift from these estimators. This is precisely the reason why we propose here an elaborate method to estimate the drift function.

To that aim, 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 the Laguerre functions or the Hermite functions. These subspaces have been recently used for nonparametric density or regression function estimation (see, e.g., Comte and Genon-Catalot (2018, 2019, 2020)). 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 noncompacity of the set A, specific bounds for the risks are obtained.

In two previous papers (Comte and Genon-Catalot (2019, 2020)), we have studied the nonparametric estimation of the regression function in the classical regression model where observations (X_i, Y_i) are i.i.d. and satisfy $Y_i = b(X_i) + \sigma(X_i)\varepsilon_i$, with noise variables ε_i i.i.d. such that $\mathbb{E}(\varepsilon_i) = 0$, $\mathbb{E}(\varepsilon_i^2) = 1$ and independent of the X_i's. The common features between these two papers and the present one are that we estimate the function b (regression or drift function) on a possibly noncompact set A, by a projection method on subspaces of $\mathbb{L}^2(A, dx)$, relying on a least-squares contrast, without assuming that $b \in \mathbb{L}^2(A, dx)$ and that both functions b and σ may be unbounded. Thus, although the model, the methodology, proofs and results are quite distinct, we can point useful analogies, which we use to substantially shorten the technical part of the present work. First, the noncompacity of the estimation set imposes a restriction on the choice of dimensions for the projection spaces involving the inverse of a matrix depending on the projection basis (Conditions (2.14)–(2.18)). Such a condition was first introduced in the regression framework by Cohen et al. (2013) to obtain stability and accuracy of the least-squares approximations in case of a general reconstruction set. Second, to be able to study properly the risk of the estimator of b, a trimming of the projection estimator is required involving the inversion of a random matrix (Definition (2.15)). Third, for the adaptive procedure, the data-driven dimension is chosen within a random set, also related to the stability condition. This is why some parts of proofs and technical lemmas are borrowed from our two previous papers when there is no need to reproduce proofs. We were careful to choose similar notation to ease the reading.

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. The case where σ is bounded on A is easier. The penalty term has the usual form and depends on σ only through a single upper bound, $\|\sigma\mathbf{1}_A\|_\infty$. For unbounded σ, the study is complicated by the fact that the penalty has an unusual form and is random. A short recap on the 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. 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 (1.1) 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 σ have linear growth.

Thus

$$\exists K > 0, \forall x \in \mathbb{R}, \quad b^2(x) + \sigma^2(x) \leq K(1 + x^2). \tag{2.1}$$

Assumption (H1) is standard and Assumption (H2) is fulfilled, for example, 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 $\mathbb{R}^+ \times (l, r) \times (l, r)$ where (l, r) is the state space of (1.1) (see, e.g., Rogers and Williams (1990), Chapter V, Section 7). Moreover, as the initial condition x_0 is deterministic,

$$\forall k \geq 0, \forall t \geq 0 \quad \sup_{0 \leq u \leq t} \mathbb{E}(X_1(u))^{2k} = \sup_{0 \leq u \leq t} \int y^{2k} p_u(x_0, y)\, dy < +\infty. \tag{2.2}$$

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

$$f_T(y) = \frac{1}{T} \int_0^T p_u(x_0, y)\, du. \tag{2.3}$$

By (2.2), f_T has moments of any order. From assumptions (H1) or (H2) and (2.2), we have, for all k,

$$\frac{1}{T}\mathbb{E}\left[\int_0^T \big(b^{2k}(X_1(u)) + \sigma^{2k}(X_1(u))\big)\, du\right] = \int \big(b^{2k}(y) + \sigma^{2k}(y)\big) f_T(y)\, dy < +\infty. \tag{2.4}$$

2.2. *Definition of projection estimators*. The following notation is 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 $b_A := b\mathbf{1}_A$. It is possible to choose a basis of S_m that depends on m but for simplicity, we omit this dependence in the notation. We assume that the basis functions φ_j are bounded so that $S_m \subset \mathbb{L}^2(A, f_T(x)\, dx)$.

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

$$\gamma_N(t) = \frac{1}{NT} \sum_{i=1}^N \left(\int_0^T t^2(X_i(u))\, du - 2\int_0^T t(X_i(u))\, dX_i(u)\right) \tag{2.5}$$

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

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

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

$$\hat{b}_m = \arg\min_{t\in S_m} \gamma_N(t). \tag{2.6}$$

Thus, for each m,

$$\hat{b}_m = \sum_{j=0}^{m-1} \hat{\theta}_j \varphi_j, \tag{2.7}$$

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

$$\widehat{Z}_m = \left(\frac{1}{NT}\sum_{i=1}^N \int_0^T \varphi_j\big(X_i(u)\big)\,dX_i(u)\right)_{j=0,\dots,m-1} \tag{2.8}$$

and the $m \times m$-matrix

$$\widehat{\Psi}_m = \left(\frac{1}{NT}\sum_{i=1}^N \int_0^T \varphi_j\big(X_i(u)\big)\varphi_\ell\big(X_i(u)\big)\,du\right)_{j,\ell=0,\dots,m-1}. \tag{2.9}$$

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

$$\hat{\theta}_{(m)} = \widehat{\Psi}_m^{-1}\widehat{Z}_m. \tag{2.10}$$

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

$$\|t\|_N^2 = \frac{1}{NT}\sum_{i=1}^N \int_0^T t^2\big(X_i(u)\big)\,du, \qquad \langle s, t\rangle_N = \frac{1}{NT}\sum_{i=1}^N \int_0^T t\big(X_i(u)\big)s\big(X_i(u)\big)\,du, \tag{2.11}$$

$$\nu_N(t) = \frac{1}{NT}\sum_{i=1}^N \int_0^T t\big(X_i(u)\big)\sigma\big(X_i(u)\big)\,dW_i(u). \tag{2.12}$$

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 this notation, we obtain

$$\widehat{Z}_m = \big(\langle\varphi_j, b\rangle_N\big)_{j=0,\dots,m-1} + E_m, \qquad \widehat{\Psi}_m = \big(\langle\varphi_j, \varphi_\ell\rangle_N\big)_{j,\ell=0,\dots,m-1},$$

where

$$E_m = \big(\nu_N(\varphi_j), j = 0, \dots, m-1\big)' \tag{2.13}$$

is a centered vector. Using (2.7)–(2.10), 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 $\mathrm{Tr}(M)$ the trace of M and by $\|M\|_{\mathrm{op}}$ 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 λ_i are the nonnegative eigenvalues of M. Moreover, if M, N are two matrices with compatible product MN, then $\|MN\|_{\mathrm{op}} \le \|M\|_{\mathrm{op}}\|N\|_{\mathrm{op}}$. For M, a symmetric nonnegative matrix, we denote $M^{1/2}$ a symmetric square root of M.

Let us define

$$L(m) := \sup_{x\in A} \sum_{j=0}^{m-1} \varphi_j^2(x), \tag{2.14}$$

and note that $L(m) < +\infty$, as the φ_j's are bounded. 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, that is, $m \le m' \Rightarrow S_m \subset S_{m'}$, 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 $T \ge 1$. 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 $m \ge 1$,

$$\tilde{b}_m = \hat{b}_m \mathbf{1}_{\{L(m)(\|\widehat{\Psi}_m^{-1}\|_{\mathrm{op}}\vee 1)\le \mathfrak{c}_T NT/\log(NT)\}}, \qquad \mathfrak{c}_T = \frac{1-\log(2)}{8T}. \tag{2.15}$$

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

$$\Psi_m = \mathbb{E}\widehat{\Psi}_m = \big(\langle \varphi_j, \varphi_\ell\rangle_{f_T}, j, \ell = 0, \dots, m-1\big), \tag{2.16}$$

$$\Psi_{m,\sigma^2} = \mathbb{E}\big(E_m E_m'\big) = \big(\langle \sigma\varphi_j, \sigma\varphi_\ell\rangle_{f_T}, j, \ell = 0, \dots, m-1\big) \tag{2.17}$$

(see (2.13)). Under mild assumptions on the basis (φ_j), the matrix Ψ_m is invertible as for instance the ones given in the following lemma.

LEMMA 2.1. *Assume that $\lambda(A \cap \mathrm{supp}(f_T)) > 0$ where λ is the Lebesgue measure and $\mathrm{supp}(f_T)$ the support of f_T, that the $(\varphi_j)_{0\le j\le m-1}$ are continuous, and that there exist $x_0, \dots, x_{m-1} \in A \cap \mathrm{supp}(f_T)$, $\det[(\varphi_j(x_k))_{0\le j,k\le m-1}] \ne 0$. Then Ψ_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 $(\varphi_j)_{0\le j\le m-1}$ is the Laguerre or the Hermite basis (see Section 4), Ψ_m is invertible.

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

PROPOSITION 2.1. *Consider the estimator $\tilde{b}_m$ of b_A. Then for m such that*

$$L(m)\big(\|\Psi_m^{-1}\|_{\mathrm{op}} \vee 1\big) \le \frac{\mathfrak{c}_T}{2}\frac{NT}{\log(NT)} \quad \textit{and} \quad m \le NT \tag{2.18}$$

with $\mathfrak{c}_T$ given in (2.15), *we have*

$$\mathbb{E}\big[\|\tilde{b}_m - b_A\|_N^2\big] \le \inf_{t\in S_m} \|t - b_A\|_{f_T}^2 + \frac{2}{NT}\mathrm{Tr}\big[\Psi_m^{-1}\Psi_{m,\sigma^2}\big] + \frac{c_1(T)}{NT} \tag{2.19}$$

and

$$(2.20)\quad \mathbb{E}[\|\tilde{b}_m - b_A\|_{f_T}^2] \le \left(1 + \frac{1-\log(2)}{2\log(NT)}\right) \inf_{t\in S_m} \|t - b_A\|_{f_T}^2 + 8\frac{\mathrm{Tr}[\Psi_m^{-1}\Psi_{m,\sigma^2}]}{NT} + \frac{c_2(T)}{NT},$$

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 $m \le L(m)\|\Psi_m^{-1}\|_{\mathrm{op}}$ and $m \lesssim NT$ is automatically satisfied (see Lemma 4 in Comte and Genon-Catalot (2019)).

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 (2.14) on the space S_m and (2.18) on the possible dimensions (see also Comte and Genon-Catalot (2019, 2020)). The restrictions on the choices of m imposed by (2.18) 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 $\|\Psi_m^{-1}\|_{\mathrm{op}} = \sup_{t\in S_m, \|t\|_{f_T}=1} \|t\|^2$ (see Proposition 2 in Comte and Genon-Catalot (2019)) so that, for nested spaces, $m \mapsto \|\Psi_m^{-1}\|_{\mathrm{op}}$ is increasing.

From the variance bound in (2.19), we cannot deduce a precise rate as a function of m. Nevertheless, this bound verifies the following.

PROPOSITION 2.2.

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

Classically, in projection methods, the set A is chosen to be compact. If A is compact, σ_A is automatically bounded, Proposition 2.2 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 $\|\Psi_m^{-1}\|_{\mathrm{op}} \le 1/f_0$. 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

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

Therefore, the stability condition (2.18) simplifies into $m \le cNT/\log(NT)$ where c depends on T and f_0.

If A is not compact, $\|\Psi_m^{-1}\|_{\mathrm{op}}$ may be unbounded as a function of m and may increase the variance rate. For instance, in the case where (φ_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, $\|\Psi_m^{-1}\|_{\mathrm{op}} \ge c\sqrt{m}$ 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 2.1 and 2.2 concerning the rates of convergence of the projection estimators. In Comte and Genon-Catalot (2019), to assess the bias rate, the following regularity set is proposed and justified:

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

where $h_\ell^{f_T}$ is the $\mathbb{L}^2(A, f_T(x)\,dx)$-orthogonal projection of h on S_ℓ. If b_A has a given (unknown) regularity s in the previous meaning, that is, if b_A belongs to $W_{f_T}^s(A, R)$, the square bias satisfies $\|b_m^{f_T} - b_A\|_{f_T}^2 \le Rm^{-s}$. Then bound (2.20) becomes, for $NT \ge 2$,

$$\mathbb{E}[\|\tilde{b}_m - b_A\|_{f_T}^2] \le 1.23Rm^{-s} + 8\frac{\mathrm{Tr}[\Psi_m^{-1}\Psi_{m,\sigma^2}]}{NT} + \frac{c_2(T)}{NT}.$$

If σ is bounded on A (see Proposition 2.2), and if $m^\star = (NT)^{1/(s+1)}$ satisfies (2.18), we find the rate

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

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.

In the general case, the best compromise between bias and variance terms is obtained defining $m^\star$ by the implicit relation $(m^\star)^{-s} = \mathrm{Tr}[\Psi_{m^\star}^{-1}\Psi_{m^\star,\sigma^2}]/NT$ and yields a rate of implicit order $(m^\star)^{-s}$. The order of the quantity $= \mathrm{Tr}[\Psi_{m^\star}^{-1}\Psi_{m^\star,\sigma^2}]$ is empirically illustrated in Section 4 (see Figure 1), and seems to be of order m in rather general context. In any case, the choice of $m^\star$ is not possible in practice, as s and R are unknown.

The next section is devoted to data-driven choices of the dimension of the projection space and yields an adaptive estimator, that is, 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. Consider now the following assumptions:

(A1) The collection of spaces S_m is nested (i.e., $S_m \subset S_{m'}$ for $m \le m'$) and such that, for each m, the basis $(\varphi_0, \dots, \varphi_{m-1})$ of S_m satisfies

$$L(m) = \left\| \sum_{j=0}^{m-1} \varphi_j^2 \right\|_\infty \le c_\varphi^2 m, \quad \text{for } c_\varphi > 0 \text{ a constant.} \tag{3.1}$$

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

Clearly, Assumption (A1) is fulfilled by classical compactly supported bases, such as histograms and trigonometric polynomials, and also by the Laguerre and Hermite bases, which are noncompactly 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. For instance, if the φ_j's are uniformly bounded over j, we have $c_\varphi^2 = \sup_j \sup_{x\in A} \varphi_j^2(x)$.

In Section 3.3, we give sufficient conditions ensuring that (A2) holds. We consider the following collection of models, for θ a positive constant specified below:

$$\widehat{\mathcal{M}}_N(\theta) = \left\{ m \in \{1, \dots, NT\}, c_\varphi^2 m(\|\widehat{\Psi}_m^{-1}\|_{\mathrm{op}}^2 \vee 1) \le \theta \frac{NT}{\log(NT)} \right\}, \tag{3.2}$$

and its theoretical counterpart

$$\mathcal{M}_N(\theta) = \left\{ m \in \{1, \dots, NT\}, c_\varphi^2 m(\|\Psi_m^{-1}\|_{\mathrm{op}}^2 \vee 1) \le \frac{\theta}{4} \frac{NT}{\log(NT)} \right\}. \tag{3.3}$$

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

Under (A1), the condition in the definition of $\mathcal{M}_N(\theta)$ is to be compared with the stability condition (2.18): $c_\varphi^2 m(\|\Psi_m^{-1}\|_{\mathrm{op}} \vee 1) \le (\mathfrak{c}_T/2)(NT/\log(NT))$. The condition imposed in $\mathcal{M}_N(\theta)$ is thus stronger as, clearly, $(\|\Psi_m^{-1}\|_{\mathrm{op}} \vee 1) \le (\|\Psi_m^{-1}\|_{\mathrm{op}}^2 \vee 1)$. The same remark holds between $\widehat{\mathcal{M}}_N(\theta)$ and the cutoff used to define $\tilde{b}_m$ (see (2.15)).

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, that is, its $\mathbb{L}^2$-risk realizes automatically the best compromise between the bias and the variance term. For this, we distinguish the case where σ_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 3.1. *Consider* $M_T := NT\nu_N(t)$ *(see* (2.12)) *and compute* $\langle M\rangle_T = \int_0^T \sum_{i=1}^N t^2(X_i(u))\sigma^2(X_i(u))\,du$. *Then*

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

3.1. *Case of bounded* σ_A. If σ 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 2.2, the variance term of the risk bound is upper bounded by $\|\sigma_A\|_\infty^2 m/NT$.

Let us define, under (A2),

$$\mathfrak{d}_T = \left(3 \wedge \frac{1}{\|f_T\|_\infty}\right)\frac{1}{c_0 T}, \tag{3.4}$$

where c_0 is a numerical constant computed in the proof of Theorem 3.1. Now we set

$$\hat{m} = \arg\min_{m\in\widehat{\mathcal{M}}_N(\mathfrak{d}_T)}\{-\|\hat{b}_m\|_N^2 + \mathrm{pen}_1(m)\}, \quad \mathrm{pen}_1(m) = \kappa\|\sigma_A^2\|_\infty\frac{m}{NT}, \tag{3.5}$$

where κ is a numerical constant. Note that $\|\sigma_A^2\|_\infty m/NT$ is an upper bound on the variance term obtained in Proposition 2.1 (see Proposition 2.2).

THEOREM 3.1. *Let* $(X_i(t), t \in [0,T])_{1\leq i\leq N}$ *be observations from model* (1.1). *Assume that* (A1), (A2) *hold and that* $\|\sigma_A^2\|_\infty < \infty$. *Then there exists a numerical constant* κ_0 *such that for* $\kappa \geq \kappa_0$, *we have*

$$\mathbb{E}\big[\|\hat{b}_{\hat{m}} - b_A\|_N^2\big] \leq C \inf_{m\in\mathcal{M}_N(\mathfrak{d}_T)}\Big(\inf_{t\in S_m}\|b_A - t\|_{f_T}^2 + \mathrm{pen}_1(m)\Big) + \frac{C'}{NT}$$

and

$$\mathbb{E}\big[\|\hat{b}_{\hat{m}} - b_A\|_{f_T}^2\big] \leq C_1 \inf_{m\in\mathcal{M}_N(\mathfrak{d}_T)}\Big(\inf_{t\in S_m}\|b_A - t\|_{f_T}^2 + \mathrm{pen}_1(m)\Big) + \frac{C_1'}{NT},$$

where C, C_1 *are 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 3.1 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 σ is identified from such an observation. Therefore, σ can be assumed to be known. Note that the estimation procedure for $\hat{b}_m$ and $\hat{m}$ only depends on σ through $\|\sigma_A\|_\infty$. In practice, to implement the adaptive procedure, we can use a simple estimator of $\|\sigma_A\|_\infty$ built from discretizations 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 replace $\mathfrak{d}_T(NT)/\log(NT)$ by $NT/\log^{1+\epsilon}(NT)$, $\epsilon > 0$, provided that N is large enough.

The constant κ is a specific feature of the model selection method. Theorem 3.1 states that, under the assumptions of the theorem, for any function b, there exists a numerical (universal) constant κ_0 such that the inequalities hold for all $\kappa \geq \kappa_0$. The proof provides a numerical value κ_0 which is too large. Finding the best value κ_0 for a given statistical problem is not easy. For instance, this topic is the subject of Birgé and Massart (2007) 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 κ closer to the true one. Afterwards, this value is fixed once and for all.

3.2. *Case of unbounded σ_A.* Here, the estimation procedure depends on the complete knowledge of σ and of the constant K such that $\sigma^2(x) \leq K(1+x^2)$ (see (2.1)).

To study the case of unbounded σ_A, it is natural to consider that A is noncompact. 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 $m \geq 1$, $\|\Psi_m^{-1}\|_{\mathrm{op}}^2 \geq cm^{\beta}$ with $\beta = 4$ for the Laguerre basis and $\beta = 5/3$ for the Hermite basis.

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

For the Hermite and Laguerre bases, $\|\Psi_m^{-1}\|_{\mathrm{op}}^2 \geq cm$, see Comte and Genon-Catalot (2019), 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), if $A = \mathbb{R}$, then 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 σ is not lower bounded, the result below still holds replacing (A4) by the technical condition (6.21).

Let

$$\mathfrak{f}_T = \mathfrak{d}_T \wedge \frac{1-\log(2)}{14TB\sigma_0^2} \tag{3.6}$$

where $\mathfrak{d}_T$ is defined in (3.4), $B = 21K$ for the Laguerre basis, $B = 2KC_\infty^2$ for the Hermite basis (see the definition of C_∞ in Section 4), K is defined in (2.1) and set

$$\widehat{m} = \arg\min_{m\in\widehat{\mathcal{M}}_N(\mathfrak{f}_T)} \{-\|\hat{b}_m\|_N^2 + \widehat{\mathrm{pen}}_2(m)\} \tag{3.7}$$

where

$$\widehat{\mathrm{pen}}_2(m) = \kappa_1 \frac{m(1+\ell_m)(1+\|\widehat{\Psi}_m^{-1/2}\widehat{\Psi}_{m,\sigma^2}\widehat{\Psi}_m^{-1/2})\|_{\mathrm{op}}}{NT}, \tag{3.8}$$

with κ_1 a numerical constant, (ℓ_m) is a sequence of nonnegative numbers. The matrix $\widehat{\Psi}_{m,\sigma^2}$ is the empirical counterpart of Ψ_{m,σ^2} (see (2.17)), $\widehat{\Psi}_{m,\sigma^2} = (\langle\sigma\varphi_j, \sigma\varphi_\ell\rangle_N)_{0\leq j,\ell\leq m-1}$, that is,

$$\widehat{\Psi}_{m,\sigma^2} = \left(\frac{1}{NT}\sum_{i=1}^N \int_0^T \varphi_j(X_i(u))\varphi_\ell(X_i(u))\sigma^2(X_i(u))\,du\right)_{0\leq j,\ell\leq m-1}.$$

THEOREM 3.2. *Let $(X_i(t), t \in [0,T])_{1\leq i\leq N}$ be observations from model* (1.1). *Assume that* (A1)–(A4) *hold. Let the sequence (ℓ_m) be such that $1 \leq \ell_m \leq NT$, for $m \in \mathcal{M}_N(\mathfrak{f}_T)$ and assume that there exists a constant Σ, $0 < \Sigma < +\infty$ such that*

$$\sum_{m\in\mathcal{M}_N(16\mathfrak{f}_T)} m\|\Psi_m^{-1}\|_{\mathrm{op}} e^{-m\ell_m} \leq \Sigma. \tag{3.9}$$

Let

$$\mathrm{pen}_2(m) = \kappa_1 \frac{m(1+\ell_m)(1+\|\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\|_{\mathrm{op}})}{NT}. \tag{3.10}$$

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

$$\mathbb{E}[\|\hat{b}_{\hat{m}} - b_A\|_N^2] \leq C \inf_{m \in \mathcal{M}_N(\mathfrak{f}_T)} \Bigl(\inf_{t \in S_m} \|b_A - t\|_{f_T}^2 + \mathrm{pen}_2(m)\Bigr) + \frac{C'}{NT}$$

and

$$\mathbb{E}[\|\hat{b}_{\hat{m}} - b_A\|_{f_T}^2] \leq C_1 \inf_{m \in \mathcal{M}_N(\mathfrak{f}_T)} \Bigl(\inf_{t \in S_m} \|b_A - t\|_{f_T}^2 + \mathrm{pen}_2(m)\Bigr) + \frac{C_1'}{NT},$$

where C, C_1 are a numerical constants and C', C_1' are constants depending on Σ 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 on the variance term given in Proposition 2.1, $m(1+\ell_m)(1+\|\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\|_{\mathrm{op}})/NT$. Theorem 3.2 thus states that the compromise between the squared-bias term and the variance term, is automatically realized by $\hat{b}_{\hat{m}}$, on the collection $\mathcal{M}_N(\mathfrak{f}_T)$.

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

In Comte and Genon-Catalot (2019), examples of densities for which $\|\Psi_m^{-1}\|_{\mathrm{op}}$ is upper bounded by $O(m^k)$ are given. In such a case, we can take $\ell_m = 1$ for all m, and (3.9) holds.

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

For the square-root process, the transition density is explicit but with a rather intricate expression in the general case involving Bessel functions (for more details see, e.g., Chaleyat-Maurel and Genon-Catalot (2006)). Nevertheless, in special cases (including our Example 4 in Section 4), we can give a rather tractable formula and check Assumption (A2).

LEMMA 3.2. *Consider the process $(X(t))_{t\geq 0}$, solution of*

$$dX(t) = \bigl(2\theta X(t) + \delta\sigma_0^2\bigr)\,dt + 2\sigma_0\sqrt{X(t)}\,dW(t), \qquad X_0 = x_0,$$

where $(W(t))_{t\geq 0}$ is a standard Brownian motion. Then, for any $\delta = 2r+1$ with r an integer, $r \geq 1$, (A2) *is satisfied.*

It is likely that the result holds for any $\delta \geq 2$ but the proof would require an in-depth study of Bessel functions, which is beyond our scope.

More generally, the following result holds.

PROPOSITION 3.1.

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

(ii) *If σ is C^2, σ', σ'' bounded, σ 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

$$(3.11)\qquad dX_i(t) = b\big(X_i(t)\big)\,dt + \sigma\big(X_i(t)\big)\,dW_i(t), \qquad X_i(0) = \eta_i, \quad i = 1, \dots, N,$$

where the initial conditions η_i are i.i.d. random variables independent of $(W_1, \dots, W_N)$, with common distribution μ 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 (3.11) is positive recurrent with invariant distribution $\pi(y)\,dy$ and η has distribution π, 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 2.1 no more depend on T, and thus the risk bound (especially the variance term) is improved when T gets large.

3.4. *Estimation of f_T.* The estimation procedures use some constants, which can be easily estimated, in particular $\|f_T\|_\infty$ defined by (2.3). 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

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

where $\|\cdot\|$ is the usual $\mathbb{L}^2$-norm. For Laguerre and Hermite basis, under additional assumptions on f_T, the factor $c_\varphi^2 m$ in the variance can be improved; see Comte and Genon-Catalot (2018).

4. Simulation study. We conduct 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.

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

$$(4.1)\qquad L_j(x) = \sum_{k=0}^{j} (-1)^k \binom{j}{k} \frac{x^k}{k!}, \qquad \ell_j(x) = \sqrt{2} L_j(2x) e^{-x} \mathbf{1}_{x\ge 0}, \quad j \ge 0.$$

The collection $(\ell_j)_{j\ge 0}$ is a complete orthonormal system on $\mathbb{L}^2(\mathbb{R}^+)$ satisfying: $\forall j \ge 0$, $\forall x \in \mathbb{R}^+$, $|\ell_j(x)| \le \sqrt{2}$; see Abramowitz and Stegun ((1964), 22.14.12). The collection of models $(S_m = \operatorname{span}\{\ell_0, \dots, \ell_{m-1}\})$ is nested and obviously (3.1) holds with $c_\varphi^2 = 2$.

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

$$(4.2)\quad H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j}\big(e^{-x^2}\big), \qquad h_j(x) = c_j H_j(x) e^{-x^2/2}, \quad c_j = \big(2^j j! \sqrt{\pi}\big)^{-1/2}.$$

The sequence $(h_j, j \ge 0)$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R}, dx)$. Moreover (see Abramowitz and Stegun ((1964), 22.14.17), Szegö (1975) p. 242, Indritz (1961)), $\|h_j\|_\infty \le \Phi_0$, $\Phi_0 \simeq 1/\pi^{1/4} \simeq 0.7511$, so that (3.1) holds with $c_\varphi^2 = \Phi_0^2$. The collection of models ($S_m =$

span$\{h_0, \dots, h_{m-1}\}$) is obviously nested. Moreover, $\|h_j\|_\infty \le C_\infty (j+1)^{-1/12}$, $j = 0, 1, \dots$ where the constant C_∞ given is in Szegö (1975). Thus in this case, $L(m) \le C_\infty^2 (6/5) m^{5/6}$, and this order is smaller than order m, for m large enough.

The 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 the 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.

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 Δ. 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 $(U_i(t))_{t\ge 0}$, with dynamics given by

$$dU_i(t) = -\frac{r}{2}U_i(t)\,dt + \frac{\gamma}{2}\,dW_{i,d}(t), \quad U_i(0) \sim \mathcal{N}_d\left(0, \frac{\gamma^2}{4r}I_d\right) \text{ or } U_i(0) = 0. \tag{4.3}$$

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

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

EXAMPLE 2. $X_i(t) = \tanh(U_i(t))$ where $U_i(t)$ is defined by (4.3) with $d = 1$ is solution of (3.11) with

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

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

EXAMPLE 3. $X_i(t) = \exp(U_i(t))$ where $U_i(t)$ is defined by (4.3) with $d = 1$ is solution of (3.11) with

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

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 (4.3) with $d = 3$ is solution of (3.11) 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 $r = 2$ and $\gamma = 1$. Model 4 satisfies (H2).

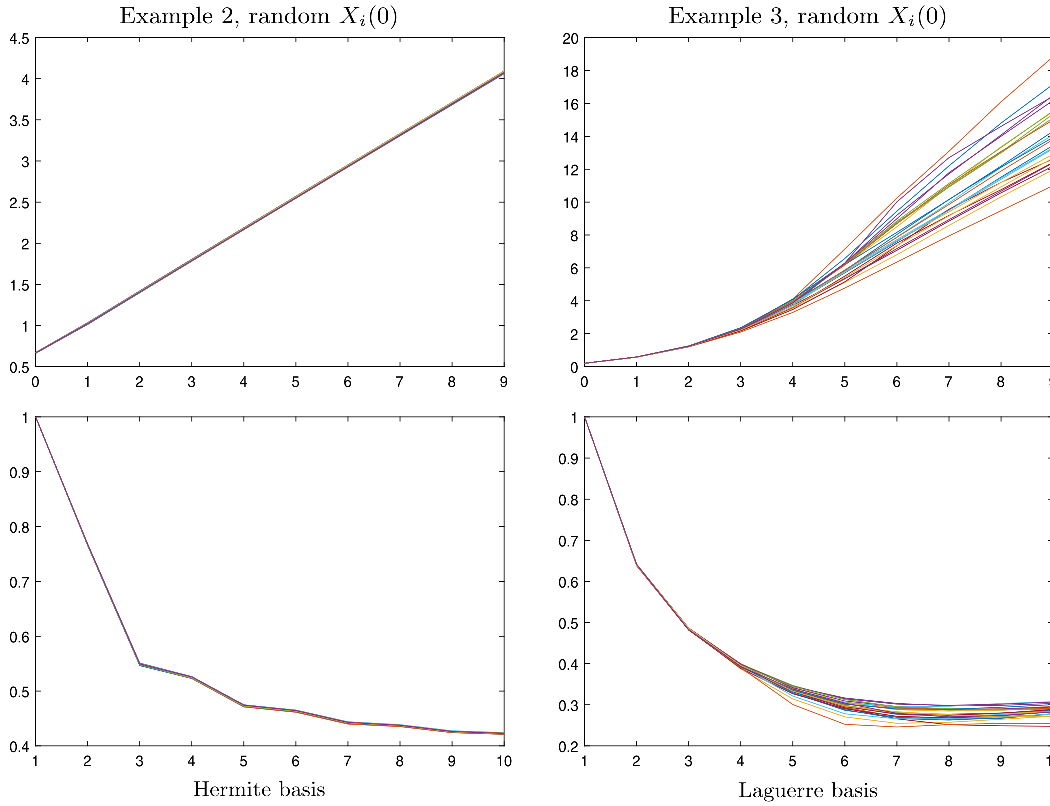

FIG. 1. *Plots of* $m \mapsto \mathrm{Tr}(\widehat{\Psi}_m^{-1}\widehat{\Psi}_{m,\sigma^2})$ *(top) and* $m \mapsto \mathrm{Tr}(\widehat{\Psi}_m^{-1}\widehat{\Psi}_{m,\sigma^2})/[m\|\widehat{\Psi}_m^{-1}\widehat{\Psi}_{m,\sigma^2}\|_{\mathrm{op}}]$ *(bottom) for* 25 *simulated paths*, $m = 1, \dots, 10$ *and* $N = 100$, $T = 100$.

In all cases, samples $(X_i(k\Delta))_{1\le i\le N, 1\le k\le n}$, $n\Delta = T$ from the above models are generated. As almost all our examples (except the first one) correspond to exact discretization, we use only one discretization step Δ, and we compute integrals by replacing them by their standard discrete version. For Examples 1 and 2, the Hermite basis is used; in Example 3, both the Hermite and Laguerre bases are used, and in Example 4, we use the Laguerre basis. Indeed, Examples 3, 4 provide nonnegative processes and are well suited to the Laguerre basis use. The order of $\mathrm{Tr}(\widehat{\Psi}_m^{-1}\widehat{\Psi}_{m,\sigma^2})$ and its comparison with $m\|\hat{\Psi}_m^{-1}\hat{\Psi}_{m,\sigma^2}\|_{\mathrm{op}}$ is illustrated on 25 simulated paths in Figure 1, with $N = 100$, $T = 100$ (corresponding to $n = 1000$ and $\Delta = 0.1$). The lines $m \mapsto \mathrm{Tr}(\widehat{\Psi}_m^{-1}\widehat{\Psi}_{m,\sigma^2})$ (top plots) seem linear (almost exactly for Example 2 which corresponds to a bounded case, and more approximately for Example 3), arguing for a variance term of order m/N. The ratios $m \mapsto \mathrm{Tr}(\widehat{\Psi}_m^{-1}\widehat{\Psi}_{m,\sigma^2})/[m\|\widehat{\Psi}_m^{-1}\widehat{\Psi}_{m,\sigma^2}\|_{\mathrm{op}}]$ (bottom plots) indicate a likely convergence to a value less than 1.

For the model selection part, the set $\widehat{\mathcal{M}}_N(\mathfrak{f}_T)$ is generally too small in practice to contain enough values of m. Therefore, a larger set given by $\widehat{\mathcal{M}}_N^\star = \{m \le 10, m\|\widehat{\Psi}_m^{-1}\|_{\mathrm{op}}^{1/4} \le NT\}$ is chosen. The selected values of m being small, we do not need to look for m larger than 10, but there is no problem to increase this value if it seems useful. The value $\|\widehat{\Psi}_m^{-1}\|_{\mathrm{op}}$, however, increases too fast with m, which may be a numerical artefact: this is why we found mandatory to set a lighter constraint.

The penalty is taken equal to $\widehat{\mathrm{pen}}(m) = \kappa\|\widehat{\Psi}_m^{-1}\widehat{\Psi}_{m,\sigma^2}\|_{\mathrm{op}}/(Nn\Delta)$ and $\hat{m}$ is selected as the minimizer of $-\|\hat{b}_m\|_N^2 + \widehat{\mathrm{pen}}(m)$. After preliminary simulations, the constant κ is taken equal to $\kappa = 2$, see the comment after Theorem 3.1.

Figure 2 shows 25 estimated drift functions $\hat{b}_{\hat{m}}$ (green/grey), and the true (red/black), with $N = 100$ and $T = 100$ (corresponding to $n = 1000$ and $\Delta = 0.1$). We stress that the value

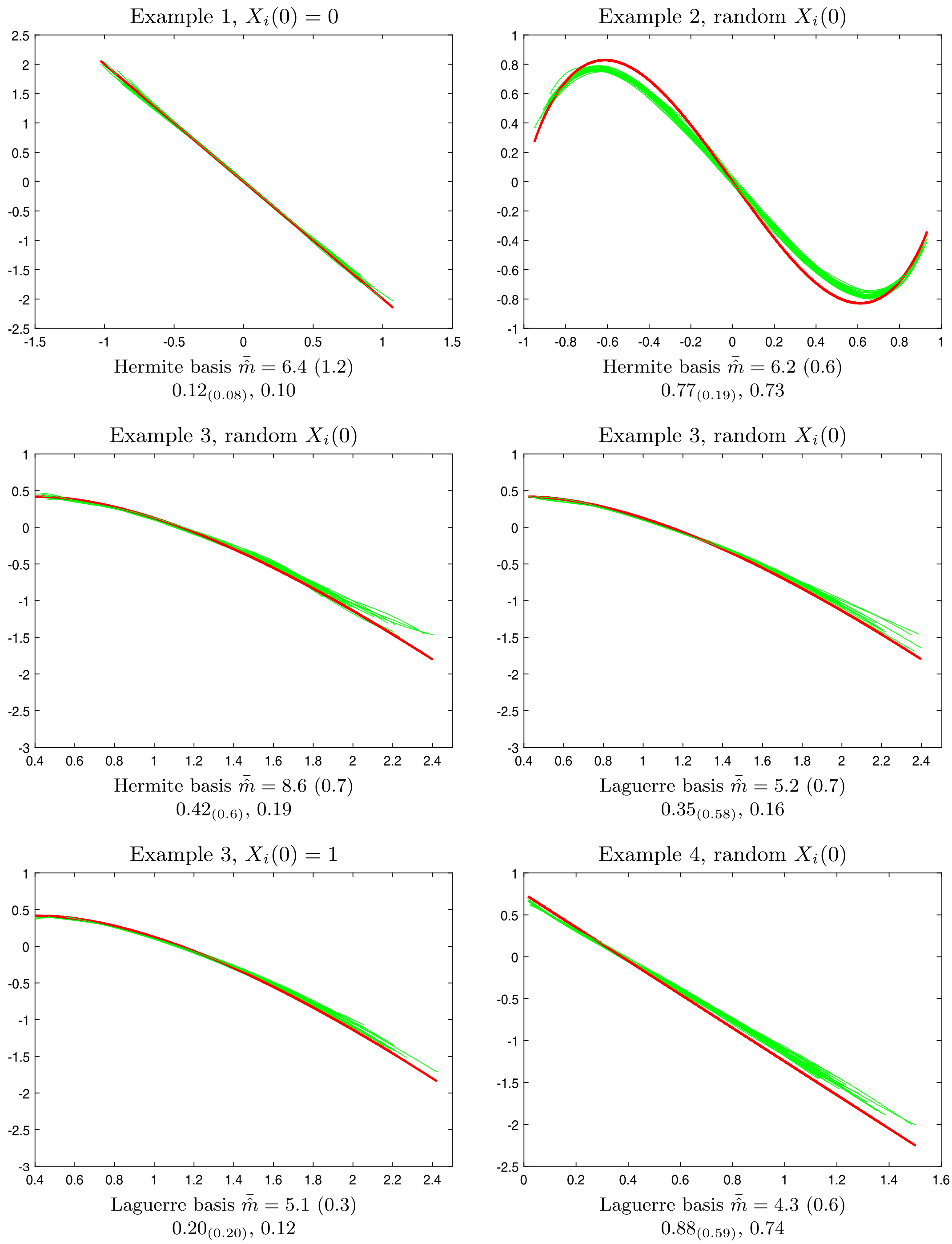

FIG. 2. *25 estimated curves in the Hermite or the Laguerre basis in green, the true in bold red,* $N = 100$, $T = 100$. *Below each plot, the mean of selected dimensions (with std in parentheses), and* $100 * \text{MISE}_{(100*\text{std})}$ *and* $100 *$ *median, over the* 25 *paths.*

of $\hat{m}$ is rather small (between 4 and 9, in all examples): under each graph, we give the mean $\bar{\hat{m}}$ computed over the 25 estimators, with standard deviation in parenthesis, together with the MISE. Thus, we see that the function is very well reconstructed using a small number of coefficients. We present in Table 1 global simulation results correspond to different choices of (N, T) for our four examples. From column 1 to column 2, $N = 100$ is the same but T

TABLE 1
$100 * \text{MISE}_{(100*\text{std})}$, $100 *$ *Median for* 100 *simulations*

(N, n, Δ)	$(100, \underbrace{500, 0.05}_{T=25})$ $NT = 2500$	$(100, \underbrace{1000, 0.1}_{T=100})$ $NT = 10{,}000$	$(1000, \underbrace{100, 0.02}_{T=2})$ $NT = 2000$
Example 1			
Hermite, $X_i(0) = 0$	$1.47_{(8.2)}$, 0.32	$0.13_{(0.12)}$, 0.08	$0.25_{(0.41)}$, 0.11
Example 2			
Hermite, random $X_i(0)$	$0.39_{(0.15)}$, 0.40	$0.77_{(0.19)}$, 0.73	$0.21_{(0.15)}$, 0.18
Example 3			
Hermite, random $X_i(0)$	$0.68_{(1.33)}$, 0.23	$0.37_{(0.87)}$, 0.16	$0.22_{(0.57)}$, 0.06
Laguerre, random $X_i(0)$	$1.07_{(3.67)}$, 0.22	$0.29_{(0.23)}$, 0.20	$0.89_{(6.44)}$, 0.04
Laguerre, $X_i(0) = 1$	$0.32_{(0.59)}$, 0.13	$0.23_{(0.20)}$, 0.17	$0.15_{(0.33)}$, 0.04
Example 4			
Laguerre, random $X_i(0)$	$0.45_{(0.88)}$, 0.21	$0.78_{(0.49)}$, 0.62	$0.17_{(0.35)}$, 0.05

increases from 25 to 100; In the last column, $N = 1000$ is much larger and $T = 2$ is small. In most cases, medians are smaller than means: this probably means that a small number of bad selections degrade the mean risk. The better results, for both mean and median, are obtained in the third column, showing that increasing N is the most decisive action, the value of T being less important.

5. Concluding remarks. In this paper, nonparametric drift estimation for a one-dimensional diffusion process is studied. For this problem, previous papers generally focus on estimation from one trajectory discretely or continuously observed on a time interval $[0, T]$ and the asymptotic framework is either that the diffusion coefficient tends to 0 or that T tends to infinity. The latter case requires, except for special models, ergodicity assumptions (see, e.g., Hoffmann (1999), Dalalyan and Kutoyants (2003), Comte, Genon-Catalot and Rozenholc (2007)).

Here, estimation is performed using the observation of N *i.i.d.* sample paths which are continuously observed throughout a *fixed* time interval $[0, T]$ and the asymptotic framework is N tends to infinity. The assumptions on the model are weak and ergodicity is not required. The drift is estimated on a subset A of $\mathbb{R}$ by a projection method, using a least-squares contrast, on finite dimensional subspaces of $\mathbb{L}^2(A, dx)$. The set A may be compact or not, the drift function need not be square integrable nor bounded and the diffusion coefficient may be bounded or not. We define a trimmed projection estimator and bound its $\mathbb{L}^2$-risk under a restriction of the possible dimensions of projection spaces. Due to the noncompacity of the estimation set, nonstandard bounds are obtained. A data-driven choice of the dimension is proposed leading to an adaptive estimator: its $\mathbb{L}^2$-risk automatically achieves the best compromise between the square bias and the variance terms. The cases of σ bounded or not are studied separately and much more difficulties are encountered in the unbounded case.

In Dalalyan and Reiss (2006, 2007), asymptotically equivalent statistical experiments in the Le Cam sense are obtained for drift estimation based on the observation of one ergodic diffusion model. Finding asymptotically equivalent experiments for drift estimation based on N i.i.d. observations of diffusions is a different problem which is worth of interest and could be a step toward finding the optimal rate for our estimator risk, a rate which is unknown in this setting even in the simplest case of bounded σ.

In the whole paper, we consider that σ is known as it is identified from a continuous observation of the sample paths. Nevertheless, the theoretical study of drift estimation for discretely observed paths with known or unknown σ is of interest. It certainly leads to even more tedious proofs as, due to discretizations, new terms appear and have to be studied. This is left for further investigation.

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

6.1. *Proof of Proposition* 2.1. We start by defining the sets

$$\Lambda_m := \left\{ L(m)(\|\widehat{\Psi}_m^{-1}\|_{\mathrm{op}} \vee 1) \le \mathfrak{c}_T \frac{NT}{\log(NT)} \right\}, \tag{6.1}$$

$$\Omega_m := \left\{ \left| \frac{\|t\|_N^2}{\|t\|_{f_T}^2} - 1 \right| \le \frac{1}{2}, \forall t \in S_m \right\}. \tag{6.2}$$

On Ω_m, the empirical norm $\|\cdot\|_N$ and the $\mathbb{L}^2(A, f_T(x)\,dx)$-norm are equivalent for elements of S_m: $(2/3)\|t\|_N^2 \le \|t\|_{f_T}^2 \le 2\|t\|_N^2$. 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

$$\|t\|_N^2 = \vec{x}'\widehat{\Psi}_m\vec{x} \quad \text{and} \quad \|t\|_{f_T}^2 = \vec{x}'\Psi_m\vec{x} = \|\Psi_m^{1/2}\vec{x}\|_{2,m}^2,$$

so that

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

Therefore,

$$\Omega_m = \left\{ \|\Psi_m^{-1/2}\widehat{\Psi}_m\Psi_m^{-1/2} - \mathrm{Id}_m\|_{\mathrm{op}} \le 1/2 \right\}.$$

The following lemma is analogous to Lemma 5 in Comte and Genon-Catalot (2019) and determines the value of $\mathfrak{c}_T$ given in (2.15). Its proof is omitted.

LEMMA 6.1. *Under the assumptions of Proposition* 2.1, *for m satisfying* (2.18) *with $\mathfrak{c}_T$ given by* (2.15), *we have, for c a positive constant,*

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

Now we prove (2.19). For this, we write

$$\begin{aligned}
\|\tilde{b}_m - b_A\|_N^2 &= \|\hat{b}_m - b_A\|_N^2 \mathbf{1}_{\Lambda_m} + \|b_A\|_N^2 \mathbf{1}_{\Lambda_m^c} \\
&= \|\hat{b}_m - b_A\|_N^2 \mathbf{1}_{\Lambda_m \cap \Omega_m} + \|\hat{b}_m - b_A\|_N^2 \mathbf{1}_{\Lambda_m \cap \Omega_m^c} + \|b_A\|_N^2 \mathbf{1}_{\Lambda_m^c} \\
&:= T_1 + T_2 + T_3.
\end{aligned} \tag{6.3}$$

We bound the expectation of the three terms above.

The last term T_3 is the easiest:

$$\mathbb{E}T_3 \le \mathbb{E}^{1/2}(\|b_A\|_N^4)\mathbb{P}^{1/2}(\Lambda_m^c) \lesssim \frac{1}{(NT)^{7/2}} \lesssim \frac{1}{NT} \tag{6.4}$$

as

$$\mathbb{E}(\|b_A\|_N^4) \le \frac{1}{T^2}\mathbb{E}\left(\int_0^T b_A^2(X_1(u))\,du\right)^2 \le \int b_A^4(y) f_T(y)\,dy < +\infty. \tag{6.5}$$

To study T_1, T_2, let us introduce the operator $\Pi_m : \mathbb{L}^2(A, f_T(x)\,dx) \to S_m$ of orthogonal projection with respect to the empirical scalar product $\langle \cdot, \cdot \rangle_N$, that is, $\Pi_m h$ 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 $\vec{\tau} = (\tau_0, \dots, \tau_{m-1})' = \widehat{\Psi}_m^{-1}(\langle \varphi_j, h \rangle_N)_{0 \le j \le m-1}$. Thus, we can write

$$\|\hat{b}_m - b_A\|_N^2 = \|\hat{b}_m - \Pi_m b_A\|_N^2 + \|\Pi_m b_A - b_A\|_N^2 = \|\hat{b}_m - \Pi_m b_A\|_N^2 + \inf_{t \in S_m} \|b_A - t\|_N^2. \tag{6.6}$$

We have $\Pi_m b_A = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j$ where $\hat{a}_{(m)} = (\hat{a}_0, \dots, \hat{a}_{m-1})' = \widehat{\Psi}_m^{-1}(\langle \varphi_j, b_A \rangle_N)_{0 \le j \le m-1}$. 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 (2.8)). Hence, we have $\hat{\theta}_{(m)} - \hat{a}_{(m)} = \widehat{\Psi}_m^{-1} E_m$ (see (2.13)) and

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

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 (6.3) and (6.6)).

On Ω_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 Ω_m, we have, a.s.,

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

Therefore,

$$\begin{aligned}
\mathbb{E}(\|\hat{b}_m - \Pi_m b_A\|_N^2 \mathbf{1}_{\Omega_m \cap \Lambda_m}) &\le 2\mathbb{E}\Bigl(\sum_{0 \le j,k \le m-1} [E_m]_j [E_m]_k \bigl[\Psi_m^{-1}\bigr]_{j,k} \Bigr) \\
&= \frac{2}{NT} \sum_{0 \le j,k \le m-1} \bigl[\Psi_m^{-1}\bigr]_{j,k} [\Psi_{m,\sigma^2}]_{j,k} \\
&= \frac{2}{NT} \operatorname{Tr}\bigl(\Psi_m^{-1} \Psi_{m,\sigma^2}\bigr),
\end{aligned} \tag{6.7}$$

as $[E_m]_j [E_m]_k$ is equal to

$$\int_0^T \varphi_j\bigl(X_1(u)\bigr) \sigma\bigl(X_1(u)\bigr)\, dW_1(u) \int_0^T \varphi_k\bigl(X_1(u)\bigr) \sigma\bigl(X_1(u)\bigr)\, dW_1(u).$$

So we obtain

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

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

$$T_2 \le \bigl(E_m' \widehat{\Psi}_m^{-1} E_m + \|b_A\|_N^2\bigr) \mathbf{1}_{\Lambda_m \cap \Omega_m^c}. \tag{6.8}$$

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

$$\mathbb{E}T_2 \le \left(\frac{\mathfrak{c}_T NT}{L(m)\log(NT)}\mathbb{E}^{1/2}\big((E_m' E_m)^2\big) + \mathbb{E}^{1/2}\|b_A\|_N^4\right)\mathbb{P}^{1/2}(\Omega_m^c), \tag{6.9}$$

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

LEMMA 6.2. *With E_m defined in (2.13) (see also (2.12)), we have*

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

where c is a numerical constant.

Plugging the result of Lemma 6.2 in (6.9) allows to conclude for all m satisfying (2.18), and $m \le NT$, that $\mathbb{E}(T_2) \le c/(NT)^3 \le c/(NT)$.

Joining the bounds for the expectations of T_1, T_2, T_3 gives inequality (2.19).

Now we prove (2.20). We have

$$\begin{aligned}\mathbb{E}\big(\|\tilde{b}_m - b_A\|_{f_T}^2\big) &= \mathbb{E}\big(\|\hat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\Omega_m\cap\Lambda_m}\big) + \mathbb{E}\big(\|\hat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\Omega_m^c\cap\Lambda_m}\big)\\ &\quad + \|b_A\|_{f_T}^2\mathbb{P}\big(\Lambda_m^c\big).\end{aligned} \tag{6.10}$$

The last right-hand side term is bounded by applying Lemma 6.1.

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

$$\begin{aligned}\mathbb{E}\big(\|\hat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\Lambda_m\cap\Omega_m}\big) &\le 2\mathbb{E}\big(\|\hat{b}_m - \Pi_m b_A\|_{f_T}^2 \mathbf{1}_{\Lambda_m\cap\Omega_m}\big)\\ &\quad + 2\mathbb{E}\big(\|\Pi_m g\|_{f_T}^2 \mathbf{1}_{\Lambda_m\cap\Omega_m}\big) + \|g\|_{f_T}^2.\end{aligned}$$

The first term is the squared bias. The second term satisfies, by definition of Ω_m and (6.7),

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

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

LEMMA 6.3. *Under the assumptions of Proposition 2.1,*

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

Therefore, we conclude that

$$\mathbb{E}(\|\hat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\Lambda_m\cap\Omega_m}) \le \left(1 + 4\frac{\mathfrak{c}_T}{\log(NT)}\right) \inf_{t\in S_m} \|t - b_A\|_{f_T}^2 + \frac{8}{NT}\operatorname{Tr}\bigl(\Psi_m^{-1}\Psi_{m,\sigma^2}\bigr). \tag{6.11}$$

Now we look at $\mathbb{E}(\|\hat{b}_m - b_A\|_{f_T}^2 \mathbf{1}_{\Omega_m^c\cap\Lambda_m})$ (see (6.10)). We have $\mathbb{P}(\Omega_m^c) \le c/(NT)^7$ and $\|\hat{b}_m - b_A\|_{f_T}^2 \le 2\|\hat{b}_m\|_{f_T}^2 + 2\|b_A\|_{f_T}^2$. 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

$$\begin{aligned}\|\hat{b}_m\|_{f_T}^2 &= \int \left(\sum_{j=0}^{m-1} \hat{\theta}_j\varphi_j(y)\right)^2 f_T(y)\,dy = \hat{\theta}'_{(m)}\Psi_m\hat{\theta}_{(m)} = \widehat{Z}'_m\widehat{\Psi}_m^{-1}\Psi_m\widehat{\Psi}_m^{-1}\widehat{Z}_m \\ &\le \|\widehat{\Psi}_m^{-1}\|_{\mathrm{op}}^2\|\Psi_m\|_{\mathrm{op}}\widehat{Z}'_m\widehat{Z}_m\end{aligned}$$

and

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

It follows by definition of Λ_m that

$$\mathbb{E}[\|\tilde{b}_m\|_{f_T}^2 \mathbf{1}_{\Lambda_m\cap\Omega_m^c}] \le \left(\frac{\mathfrak{c}_T NT}{\log(NT)}\right)^2 \frac{1}{L(m)} \mathbb{E}^{1/2}[(\widehat{Z}'_m\widehat{Z}_m)^2]\mathbb{P}^{1/2}(\Omega_m^c). \tag{6.12}$$

Now we have $(\widehat{Z}'_m\widehat{Z}_m)^2 \le 4m\sum_{j=0}^{m-1}\langle\varphi_j, b\rangle_N^4 + 4(E'_mE_m)^2$. By elementary computations, $\mathbb{E}(\langle\varphi_j, b\rangle_N^4) \le \int(\varphi_j(x)b_A(x))^4 f_T(x)\,dx$. Therefore, by using Lemma 6.2,

$$\begin{aligned}&(\mathbb{E}(\widehat{Z}'_m\widehat{Z}_m)^2)^{1/2} \\ &\quad\le 2\left(\sqrt{m}L(m)\left(\int b_A^4(x)f_T(x)\,dx\right)^{1/2} + \sqrt{c}\frac{\sqrt{m}L(m)}{NT}\left(\int \sigma_A^4(y)f_T(y)\,dy\right)^{1/2}\right).\end{aligned}$$

Joining the above with (6.12) yields

$$\mathbb{E}[\|\hat{b}_m\|_{f_T}^2 \mathbf{1}_{\Omega_m^c\cap\Lambda_m}] \le \frac{c}{NT}\left(\left(\int b_A^4(y)f_T(y)\,dy\right)^{1/2} + \left(\int \sigma_A^4(y)f_T(y)\,dy\right)^{1/2}\right). \tag{6.13}$$

So plugging (6.13) in (6.10) together with (6.11) yields the bound (2.20).

6.2. *Proof of Lemma* 6.2. $\mathbb{E}((E'_mE_m)^2) = \frac{1}{N^4T^4}\mathbb{E}(F(M_0(T), \dots, M_{m-1}(T)))$ 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

$$\begin{aligned}\mathbb{E}((E'_mE_m)^2) &\le c\frac{m}{(NT)^4}\sum_{j=0}^{m-1}\mathbb{E}\langle M_j\rangle_T^2 \\ &= c\frac{m}{(NT)^4}\sum_{j=0}^{m-1}\mathbb{E}\left[\left(\int_0^T\sum_{i=1}^N \varphi_j^2(X_i(u))\sigma^2(X_i(u))\,du\right)^2\right]\end{aligned}$$

$$\leq c\frac{m}{(NT)^4}TN\sum_{i=1}^{N}\sum_{j=0}^{m-1}\mathbb{E}\biggl(\int_0^T\varphi_j^4(X_i(u))\sigma_A^4(X_i(u))\,du\biggr)$$

$$\leq c\frac{mL^2(m)}{(NT)^2}\int\sigma_A^4(y)f_T(y)\,dy.$$

6.3. *Proof of Lemma* 6.3. 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

$$\mathrm{Id}_m = \biggl(\int\bar{\varphi}_j\bar{\varphi}_k f_T\biggr)_{j,k} = \mathbf{A}_m\Psi_m\mathbf{A}_m'$$

so that $\mathbf{A}_m$ is a square root of Ψ_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 Ω_m, $\|\widehat{\mathbf{G}}_m - \mathrm{Id}_m\|_{\mathrm{op}} \leq 1/2$, and thus $\|\widehat{\mathbf{G}}_m^{-1}\|_{\mathrm{op}} \leq 2$.

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 $j = 0, 1, \ldots, m-1$, 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

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

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

$$\|\Pi_m g\|_{f_T}^2 = \|\vec{\beta}_m\|_{2,m}^2 = \|\widehat{\mathbf{G}}_m^{-1}\vec{d}_m\|_{2,m}^2 \leq \|\widehat{\mathbf{G}}_m^{-1}\|_{\mathrm{op}}^2\|\vec{d}_m\|_{2,m}^2 \leq 4\sum_{j=0}^{m-1}\langle g, \bar{\varphi}_j\rangle_N^2. \tag{6.14}$$

Now we note that

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

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

$$\mathbb{E}[\langle g, \bar{\varphi}_j\rangle_N^2] = \frac{1}{NT^2}\mathrm{Var}\biggl(\int_0^T\bar{\varphi}_j(X_1(u))g(X_1(u))\,du\biggr)$$

and

$$\mathbb{E}\Biggl[\sum_{j=0}^{m-1}\langle g, \bar{\varphi}_j\rangle_N^2\mathbf{1}_{\Omega_m\cap\Lambda_m}\Biggr] \leq \frac{1}{NT^2}\sum_{j=0}^{m-1}\mathbb{E}\biggl[\biggl(\int_0^T\bar{\varphi}_j(X_1(u))g(X_1(u))\,du\biggr)^2\biggr]$$

$$= \frac{1}{NT^2}\mathbb{E}[\|\mathbf{A}_m\vec{v}\|_{2,m}^2],$$

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

$$\mathbb{E}\Biggl[\sum_{j=0}^{m-1}\langle g, \bar{\varphi}_j\rangle_N^2\mathbf{1}_{\Omega_m\cap\Lambda_m})\Biggr] \leq \frac{\|\Psi_m^{-1}\|_{\mathrm{op}}}{NT^2}\mathbb{E}(\|\vec{v}\|_{2,m}^2)$$

$$\leq \frac{\|\Psi_m^{-1}\|_{\mathrm{op}}}{NT^2}\mathbb{E}\Biggl(\sum_{j=0}^{m-1}\biggl(\int_0^T\varphi_j(X_1(u))g(X_1(u))\,du\biggr)^2\Biggr)$$

$$\leq \frac{\|\Psi_m^{-1}\|_{\mathrm{op}}}{N}L(m)\|g\|_{f_T}^2.$$

This, under (2.18) and reminding (6.14), implies

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

This gives the result of Lemma 6.3. □

6.4. *Proof of Proposition* 2.2. Property (i) follows from Proposition 2.4 in Comte and Genon-Catalot (2020)). For (ii), we can write

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

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 σ is bounded on A,

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

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

6.5. *Proofs of Theorem* 3.1 *and Theorem* 3.2. To deal with the random set $\widehat{\mathcal{M}}_N(\theta)$ (see (3.2)), we introduce an additional set

$$\mathcal{M}_N^+(\theta) = \left\{m\in\mathbb{N}, c_\varphi^2 m\big(\|\Psi_m^{-1}\|_{\mathrm{op}}^2 \vee 1\big) \le 4\theta\frac{NT}{\log(NT)}\right\} = \mathcal{M}_N(16\theta). \tag{6.15}$$

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 σ_A is bounded (case of Theorem 3.1), and for $\mathcal{M}_N(\mathfrak{f}_T)$, $\widehat{\mathcal{M}}_N(\mathfrak{f}_T)$, $\mathcal{M}_N^+(\mathfrak{f}_T)$ otherwise (case of Theorem 3.2).

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 (3.3)). Let

$$\Xi_N := \{\mathcal{M}_N \subset \widehat{\mathcal{M}}_N \subset \mathcal{M}_N^+\}. \tag{6.16}$$

Proceeding as in Lemma 7 in Comte and Genon-Catalot (2019), we can prove that, for the choice of $\mathfrak{d}_T$ given in (3.4) with c_0 a large enough numerical value ($c_0 = 96$ suits), and, for c a positive constant,

$$\mathbb{P}(\Xi_N^c) = \mathbb{P}(\{\mathcal{M}_N \not\subseteq \widehat{\mathcal{M}}_N \text{ or } \widehat{\mathcal{M}}_N \not\subseteq \mathcal{M}_N^+\}) \le \frac{c}{(NT)^4}. \tag{6.17}$$

We write the decomposition: $\hat{b}_{\hat{m}} - b_A = (\hat{b}_{\hat{m}} - b_A)\mathbf{1}_{\Xi_N} + (\hat{b}_{\hat{m}} - b_A)\mathbf{1}_{\Xi_N^c}$. As for the study of T_2 defined by (6.3), starting from (6.8), we get

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

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

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

Lemma 6.2 yields $\mathbb{E}[(E'_{NT}E_{NT})^2] \le cc_\varphi^4(NT)\int \sigma_A^4(y)f_T(y)\,dy$, and thus we have $\mathbb{E}[(E'_{\hat{m}}\widehat{\Psi}_{\hat{m}}^{-1}E_{\hat{m}})^2] \lesssim (NT)^2$. This together with (6.17) implies, for C a constant depending on $\int \sigma_A^4 f_T$, $\int b_A^4 f_T$, and $\mathfrak{d}_T$,

$$\mathbb{E}\big[\|b_A - \hat{b}_{\hat{m}}\|_n^2 \mathbf{1}_{\Xi_N^c}\big] \le \frac{C}{NT}.$$

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

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

where $\widehat{\mathrm{pen}}(m) = \mathrm{pen}_1(m)$ defined by (3.5) if σ is bounded on A and $\widehat{\mathrm{pen}}(m) = \widehat{\mathrm{pen}}_2(m)$ defined by (3.8) otherwise. Thus, we have, for any $m \in \widehat{\mathcal{M}}_N$, and any $b_m \in S_m$,

$$\gamma_N(\hat{b}_{\hat{m}}) + \widehat{\mathrm{pen}}(\hat{m}) \le \gamma_N(b_m) + \widehat{\mathrm{pen}}(m). \tag{6.18}$$

On $\Xi_N = \{\mathcal{M}_N \subset \widehat{\mathcal{M}}_N \subset \mathcal{M}_N^+\}$, $\hat{m} \le \widehat{M}_N \le M_N^+$ and either $M_N \le \hat{m} \le \widehat{M}_N \le M_N^+$ or $\hat{m} < M_N \le \widehat{M}_N \le M_N^+$. In the first case, $\hat{m}$ is upper and lower bounded by deterministic bounds and (6.18) a fortiori holds for any $m \in \mathcal{M}_N$; and in the second case,

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

Thus, on Ξ_N, (6.18) 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 (2.12), yields, for any $m \in \mathcal{M}_N$ and any $b_m \in S_m$,

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

We introduce the unit ball and the set

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

where Ω_m is defined by (6.1). We split again

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

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 6.1, $\mathbb{P}(\Xi_N \cap \Omega_N^c) \le \sum_{m\in\mathcal{M}_N^+} \mathbb{P}(\Omega_m^c) \le c'/(NT)^6$.

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

$$\begin{aligned}
\|\hat{b}_{\hat{m}} - b_A\|_N^2 &\le \|b_m - b_A\|_N^2 + \frac{1}{8}\|\hat{b}_{\hat{m}} - b_m\|_{f_T}^2 \\
&\quad + \Big(8 \sup_{t\in B_{\hat{m},m}^{f_T}(0,1)} \nu_N^2(t) + \widehat{\mathrm{pen}}(m) - \widehat{\mathrm{pen}}(\hat{m})\Big) \\
&\le \Big(1 + \frac{1}{2}\Big)\|b_m - b_A\|_N^2 + \frac{1}{2}\|\hat{b}_{\hat{m}} - b_A\|_N^2 + 8p(m,\hat{m}) \\
&\quad + 8\Big(\sup_{t\in B_{\hat{m},m}^{f_T}(0,1)} \nu_N^2(t) - p(m,\hat{m})\Big)_+ + \widehat{\mathrm{pen}}(m) - \widehat{\mathrm{pen}}(\hat{m}).
\end{aligned} \tag{6.19}$$

Note that, in the case $\|\sigma_A\|_\infty < +\infty$, $\mathrm{pen}_1(m) = \widehat{\mathrm{pen}}(m)$ is deterministic. Therefore, we can complete the proof of the first inequality of Theorem 3.1 applying the following lemma.

LEMMA 6.4. *Assume that* $\|\sigma_A\|_\infty < +\infty$. *Then there exists a numerical constant* τ *such that for* $p(m,m') = \tau\|\sigma_A\|_\infty^2(m+m')/(NT)$,

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

Indeed, we choose $\kappa \ge 8\tau$ in $\mathrm{pen}_1(m)$ and the first inequality of Theorem 3.1 follows. For the second inequality, we proceed as in the proof of Theorem 2 in Comte and Genon-Catalot (2019).

PROOF OF LEMMA 6.4. When σ_A is bounded, for t an A-supported function,

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

Thus, by Lemma 3.1, we obtain

$$\mathbb{P}(\nu_N(t)\ge \varepsilon, \|t\|_N^2 \le v^2) \le \exp(-NT\varepsilon^2/(2\|\sigma_A^2\|_\infty v^2)).$$

Afterwards, as in Comte, Genon-Catalot and Rozenholc (2007), we use the $\mathbb{L}^2$-chaining technique described in Baraud, Comte and Viennet ((2001), Section 7, pp. 44–47, Lemma 7.1, with $s^2 = \|\sigma_A^2\|_\infty/T$). □

Now we no longer assume σ_A bounded and we consider the Laguerre and Hermite bases to complete the proof of Theorem 3.2. We have the following lemma.

LEMMA 6.5. *Assume* (A1)–(A4). *Then there exists a numerical value* τ_1 *such that* $\nu_N(t)$ *satisfies*

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

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

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

For $\kappa_1 \ge 8\tau_1$, $8p(m,m') \le \mathrm{pen}(m) + \mathrm{pen}(m')$. Therefore, plugging the result of Lemma 6.5 in (6.19) and taking expectation yield that

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

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 (2020).

LEMMA 6.6. *Under the assumptions of Theorem* 3.2, *there exist constants* $c_1, c_2 > 0$ *such that for* $m \in \mathcal{M}_N$ *and* $\hat{m} \in \widehat{\mathcal{M}}_N$,

$$\mathbb{E}(\widehat{\mathrm{pen}}(m)\mathbf{1}_{\Xi_N\cap\Omega_N}) \le c_1\mathrm{pen}(m) + \frac{c_2}{NT},$$

$$\mathbb{E}[(\mathrm{pen}(\hat{m}) - \widehat{\mathrm{pen}}(\hat{m}))_+ \mathbf{1}_{\Xi_N\cap\Omega_N}) \le \frac{c_2}{NT}.$$

Note that c_2 contains $\int |\sigma_A|^{4+56/\beta} f_T$. Lemma 6.6 concludes the study of the expectation of the empirical risk on $\Xi_N \cap \Omega_N$. This gives the first inequality of Theorem 3.2. The second inequality is obtained following the lines of the proof of Theorem 2 in Comte and Genon-Catalot (2019).

6.6. *Proof of Lemma* 6.5. Define the set

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

We need the following lemma, similar to Lemma 6.1, which determines the constant $\mathfrak{f}_T$.

LEMMA 6.7. *Consider the Laguerre or the Hermite basis. Assume that* (A1)–(A4) *hold. Then* $\mathbb{P}(\Omega^c_{m,\sigma^2}) \le c/(NT)^6$ *and* $\mathbb{P}(\Omega^c_{N,\sigma^2}) \le c/(NT)^5$.

Note that

$$\begin{aligned}\sup_{\|t\|_{f_T}=1} \|t\sigma\|_{f_T}^2 &= \sup_{\|\Psi_m^{1/2}\vec{a}\|_{2,m}=1} \vec{a}'\Psi_{m,\sigma^2}\vec{a} = \sup_{\|\vec{u}\|_{2,m}=1} \vec{u}'\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\vec{u} \\ &= \|\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\|_{\text{op}}.\end{aligned}$$

This implies

$$\sup_{t\in S_m, \|t\|_{f_T}=1} \nu_N^2(t) \le \|\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\|_{\text{op}} \sup_{t\in S_m, \|t\sigma\|_{f_T}=1} \nu_N^2(t). \tag{6.20}$$

We have

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

with $\mathbf{A}(m) := (\sup_{t\in S_m, \|t\|_{f_T}=1} \nu_n^2(t) - p(m))_+$, $\mathbf{T}_1^\star(m) = \mathbf{A}(m)\mathbf{1}_{\Xi_N\cap\Omega_N\cap\Omega_{N,\sigma^2}}$ and $\mathbf{T}_2^\star(m) = \mathbf{A}(m)\mathbf{1}_{\Xi_N\cap\Omega_N\cap\Omega^c_{N,\sigma^2}}$. Now, by using (6.20), we have

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

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 τ_1 such that

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

As a consequence, for the same numerical constant τ_1,

$$\mathbb{E}[T_1^\star(m)] \le c\frac{e^{-m\ell_m}}{NT} \|\Psi_m^{-1/2}\Psi_{m,\sigma^2}\Psi_m^{-1/2}\|_{\text{op}}.$$

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

$$\begin{aligned}\|\Psi_{m,\sigma^2}\|_{\text{op}} &= \sup_{\|\vec{a}\|_{2,m}=1} \vec{a}'\Psi_{m,\sigma^2}\vec{a} \\ &= \sup_{\|\vec{a}\|_{2,m}=1} \int \left(\sum_{j=0}^{m-1} a_j\varphi_j(y)\right)^2 \sigma^2(y) f_T(y)\, dy \le c_\varphi^2 m \int \sigma^2 f_T.\end{aligned}$$

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

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

under condition (3.9). Thus, we get

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

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

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

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

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

Second,

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

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

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

Thus, the expectation is less than

$$\begin{aligned}
&\frac{cM}{(NT)^2} \int \left(\sum_{j=0}^{M-1} \varphi_j^2(y)\right)^2 \|\Psi_M^{-1}\|_{\mathrm{op}}^2 \sigma_A^4(y) f_T(y)\, dy \\
&\quad \le cc_\varphi^4 \frac{M^3}{(NT)^2} \|\Psi_M^{-1}\|_{\mathrm{op}}^2 \int \sigma_A^4(y) f_T(y)\, dy \le CNT \int \sigma_A^4(y) f_T(y)\, dy,
\end{aligned}$$

and we get

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

We obtain $\mathbb{E}[\mathbf{T}_2^\star(m \vee \hat{m})] \lesssim 1/(NT)^2$. This completes the proof of Lemma 6.5.

PROOF OF LEMMA 6.7. Analogously as for Ω_m, we have

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

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

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

with

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

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

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

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.

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

Now we use that $\sigma^2(x) \le K(1+x^2)$ with K known. If $\varphi_j = \ell_j$, the Laguerre basis on $A = \mathbb{R}^+$, we have $|\ell_j|^2 \le 2$ and (see, e.g., Comte and Genon-Catalot (2018), 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

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

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

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

This yields

$$\begin{aligned} \|\mathbf{K}_{m,\sigma^2}(X_1)\|_{\mathrm{op}} &\le K C_\infty^2 (m^{5/6} + 3m^{11/6}) \|\Psi_{m,\sigma^2}^{-1}\|_{\mathrm{op}} \\ &\le 2K C_\infty^2 m^{11/6} \|\Psi_{m,\sigma^2}^{-1}\|_{\mathrm{op}} := R. \end{aligned}$$

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

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

if $m \le NT$ and

$$Bm^b \|\Psi_{m,\sigma^2}^{-1}\|_{\mathrm{op}} \le c_T(1/2)\frac{NT}{7\log(NT)}. \tag{6.21}$$

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

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

By definition of $\mathcal{M}_N$ and under (A3), we have $m^b\|\Psi_m^{-1}\|_{\mathrm{op}} \le m\|\Psi_m^{-1}\|_{\mathrm{op}}^2$ 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 (6.21) is fulfilled and the bound is true. $\square$

6.7. *Proof of Lemma* 3.2. When δ is an integer, $(X(t))$ has the distribution of $(\sum_{j=1}^{\delta} \xi_j^2(t))$ where $(\xi_j(t))$ are i.i.d. Ornstein–Uhlenbeck processes such that

$$d\xi_j(t) = \theta\xi_j(t)\,dt + \sigma_0\,dW_j(t), \quad \xi_0^j = x_j$$

with $\sum_{j=1}^{\delta} x_j^2 = x_0$. When $\delta = 2r+1$ is an odd integer (with $r \ge 0$, r an integer), the transition density of $(X(t))$ has a tractable expression. Setting

$$a(t) = a = \exp(\theta t), \qquad \beta(t) = \beta = \sigma_0\left(\frac{\exp(2\theta t) - 1}{2\theta}\right)^{1/2}, \tag{6.22}$$

we have

$$p_t^{(2r+1)}(x_0, x) = \mathbf{1}_{x>0}\frac{x^{r-(1/2)}}{\sqrt{2\pi}\beta^{2r+1}}\exp\left[-\left(\frac{x}{2\beta^2} + \frac{a^2x_0}{2\beta^2}\right)\right][\mathcal{T}^r(\cosh)(z)]_{z=a\sqrt{x_0x}/\beta^2}, \tag{6.23}$$

where the operator $\mathcal{T}$ is given by $\mathcal{T}(f) = f'(x)/x$. In particular, when $\delta = 1$ ($r = 0$), assumption (A2) is not satisfied. For $\delta = 3$ ($r = 1$),

$$p_t^{(3)}(x_0, x) = \mathbf{1}_{x>0}\exp\left[-\left(\frac{x}{2\beta^2} + \frac{a^2x_0}{2\beta^2}\right)\right]\frac{1}{\sqrt{2\pi}\beta}\frac{\sinh(a\sqrt{x_0x}/\beta^2)}{a\sqrt{x_0}}. \tag{6.24}$$

We can write, for any $x > 0$, $p_t^{(3)}(x_0, x)$ is smaller than

$$\begin{aligned}
&\frac{1}{2\sqrt{2\pi x_0}a\beta}\exp\left[-\left(\frac{x}{2\beta^2} + \frac{a^2x_0}{2\beta^2}\right)\right][\exp(a\sqrt{x_0x}/\beta^2) + \exp(-a\sqrt{x_0x}/\beta^2)] \\
&\quad \le \frac{1}{2\sqrt{2\pi x_0}a\beta}[\exp\left[-\frac{(\sqrt{x} - a\sqrt{x_0})^2}{2\beta^2} + \exp\left[-\frac{(\sqrt{x} + a\sqrt{x_0})^2}{2\beta^2}\right]\right] \le \frac{1}{\sqrt{2\pi x_0}a\beta},
\end{aligned}$$

where the upper bound only depends on t. Near 0,

$$\frac{1}{a\beta} \sim \frac{1}{\sqrt{t}} \tag{6.25}$$

and it is continuous on $(0, T]$. Therefore, $f_T(x) = (1/T)\int_0^T p_u(x_0, x)\,du$ is bounded and (A2) holds. More generally, it is easy to check that

$$\mathcal{T}^r(\cosh)(z) = \frac{P_r(z)\sinh(z) + Q_r(z)\cosh(z)}{z^{2r-1}},$$

with P_r and Q_r polynomials of degree less than or equal to $r-1$. This can be proved by setting $P_1(z)=1$ and $Q_1(z)=0$ and computing the recursion formula:

$$P_{r+1}(z) = zQ_r(z) + zP_r'(z) - (2r-1)P_r(z),$$
$$Q_{r+1}(z) = zP_r(z) + zQ_r'(z) - (2r-1)Q_r(z).$$

Thus $p_t^{(2r+1)}(x_0,x)$ is equal, for $x>0$, to

$$\frac{x^{r-(1/2)}}{2\sqrt{2\pi}\beta^{2r+1}}\frac{1}{(a\sqrt{x_0x}/\beta^2)^{2r-1}} \times \biggl\{P_r(a\sqrt{x_0x}/\beta^2)\biggl[\exp\biggl[-\frac{(\sqrt{x}-a\sqrt{x_0})^2}{2\beta^2}\biggr] - \exp\biggl[-\frac{(\sqrt{x}+a\sqrt{x_0})^2}{2\beta^2}\biggr]\biggr] + Q_r(a\sqrt{x_0x}/\beta^2)\biggl[\exp\biggl[-\frac{(\sqrt{x}-a\sqrt{x_0})^2}{2\beta^2}\biggr] + \exp\biggl[-\frac{(\sqrt{x}+a\sqrt{x_0})^2}{2\beta^2}\biggr]\biggr]\biggr\}.$$

Therefore, $p_t^{(2r+1)}(x_0,x)$, $x>0$, is bounded by a linear combination of terms of type

$$\frac{x^{r-(1/2)}}{\beta^{2r+1}}\frac{1}{(a\sqrt{x_0x}/\beta^2)^{2r-1-k}}\exp\biggl[-\frac{(\sqrt{x}\pm a\sqrt{x_0})^2}{2\beta^2}\biggr] = \frac{1}{(a\sqrt{x_0})^{2r-1-k}}\beta^{2r-2k-3}(\sqrt{x})^k\exp\biggl[-\frac{(\sqrt{x}\pm a\sqrt{x_0})^2}{2\beta^2}\biggr],$$

where k is an integer, $0\le k\le r-1$. Noticing that for any $u>0$, $u^k e^{-u^2/(2\beta^2)} \le C_k\beta^k$ with $C_k=(2k/e)^{k/2}$, we derive that all the terms in $p_t^{(2r+1)}(x_0,x)$ are uniformly bounded w.r.t. $x\ge 0$ by a combination of terms $C(k,r,x_0)\frac{1}{a^{2r-1-k}}\beta^{2r-k-3}$ where k is an integer, $0\le k\le r-1$. As $2r-k-3\ge r-2\ge -1$, using (6.22)–(6.25), the terms $C(k,r,x_0)\frac{1}{a^{2r-1-k}}\beta^{2r-k-3}$ only depend on t and are integrable near 0; therefore, $f_T(x)=(1/T)\int_0^T p_u(x_0,x)\,du$ is bounded and (A2) holds.

6.8. *Proof of Proposition* 3.1. We use the following representation (see, e.g., Rogers (1985)).

For (i), set $B(y)=\int_0^y b(u)\,du$. Then

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

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

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

it follows that

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

which implies (i).

For (ii), we consider the model $dX(t) = b(X(t))\,dt + \sigma(X(t))\,dW_t$, where b, σ 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 σ', σ'' are bounded and obtain that $\|f_T\|_\infty < +\infty$.

7. A theoretical tool.

THEOREM 7.1 (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*

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

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

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

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

Acknowledgments. We are very grateful to the Associate Editor and the referees for their valuable comments and constructive suggestions.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series* **55**. U.S. Government Printing Office, Washington, D.C. MR0167642

BARAUD, Y., COMTE, F. and VIENNET, G. (2001). Model selection for (auto-)regression with dependent data. *ESAIM Probab. Stat.* **5** 33–49. MR1845321 https://doi.org/10.1051/ps:2001101

BIRGÉ, L. and MASSART, P. (2007). Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields* **138** 33–73. MR2288064 https://doi.org/10.1007/s00440-006-0011-8

CHALEYAT-MAUREL, M. and GENON-CATALOT, V. (2006). Computable infinite-dimensional filters with applications to discretized diffusion processes. *Stochastic Process. Appl.* **116** 1447–1467. MR2260743 https://doi.org/10.1016/j.spa.2006.03.004

COHEN, A., DAVENPORT, M. A. and LEVIATAN, D. (2013). On the stability and accuracy of least squares approximations. *Found. Comput. Math.* **13** 819–834. MR3105946 https://doi.org/10.1007/s10208-013-9142-3

COMTE, F. and GENON-CATALOT, V. (2018). Laguerre and Hermite bases for inverse problems. *J. Korean Statist. Soc.* **47** 273–296. MR3840862 https://doi.org/10.1016/j.jkss.2018.03.001

COMTE, F. and GENON-CATALOT, V. (2019). Regression function estimation on non compact support as a partly inverse problem. *Ann. Inst. Statist. Math.* To appear. Available at https://doi.org/10.1007/s10463-019-00718-2.

COMTE, F. and GENON-CATALOT, V. (2020). Regression function estimation on non compact support in an heteroscesdastic model. *Metrika* **83** 93–128. MR4054254 https://doi.org/10.1007/s00184-019-00727-4

COMTE, F., GENON-CATALOT, V. and ROZENHOLC, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* **13** 514–543. MR2331262 https://doi.org/10.3150/07-BEJ5173

COMTE, F., GENON-CATALOT, V. and SAMSON, A. (2013). Nonparametric estimation for stochastic differential equations with random effects. *Stochastic Process. Appl.* **123** 2522–2551. MR3054535 https://doi.org/10.1016/j.spa.2013.04.009

DALALYAN, A. S. and KUTOYANTS, YU. A. (2003). Asymptotically efficient trend coefficient estimation for ergodic diffusion. *Math. Methods Statist.* **11** 402–427. MR1979742

DALALYAN, A. and REISS, M. (2006). Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab. Theory Related Fields* **134** 248–282. MR2222384 https://doi.org/10.1007/s00440-004-0416-1

DALALYAN, A. and REISS, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: The multidimensional case. *Probab. Theory Related Fields* **137** 25–47. MR2278451 https://doi.org/10.1007/s00440-006-0502-7

DELATTRE, M., GENON-CATALOT, V. and LARÉDO, C. (2018). Parametric inference for discrete observations of diffusion processes with mixed effects. *Stochastic Process. Appl.* **128** 1929–1957. MR3797649 https://doi.org/10.1016/j.spa.2017.08.016

DELATTRE, M., GENON-CATALOT, V. and SAMSON, A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scand. J. Stat.* **40** 322–343. MR3066417 https://doi.org/10.1111/j.1467-9469.2012.00813.x

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. MR3164656 https://doi.org/10.4310/SII.2013.v6.n4.a10

DENIS, C., DION, C. and MARTINEZ, M. (2018). Consistent procedures for multiclass classification of discrete diffusion paths. *Scand. J. Stat.*. To appear. https://doi.org/10.1111/sjos.12415

DION, C. and GENON-CATALOT, V. (2016). Bidimensional random effect estimation in mixed stochastic differential model. *Stat. Inference Stoch. Process.* **19** 131–158. MR3506629 https://doi.org/10.1007/s11203-015-9122-0

DITLEVSEN, S. and DE GAETANO, A. (2005). Mixed effects in stochastic differential equation models. *REVSTAT* **3** 137–153. MR2259358

HOFFMANN, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.* **79** 135–163. MR1670522 https://doi.org/10.1016/S0304-4149(98)00074-X

HSIAO, C. (2003). *Analysis of Panel Data*, 2nd ed. *Econometric Society Monographs* **34**. Cambridge Univ. Press, Cambridge. MR1962511 https://doi.org/10.1017/CBO9780511754203

IACUS, S. M. (2008). *Simulation and Inference for Stochastic Differential Equations*: *With R Examples*. *Springer Series in Statistics*. Springer, New York. MR2410254 https://doi.org/10.1007/978-0-387-75839-8

INDRITZ, J. (1961). An inequality for Hermite polynomials. *Proc. Amer. Math. Soc.* **12** 981–983. MR0132852 https://doi.org/10.2307/2034406

KESSLER, M., LINDNER, A. and SØRENSEN, M., eds. (2012). *Statistical Methods for Stochastic Differential Equations*. *Monographs on Statistics and Applied Probability* **124**. CRC Press, Boca Raton, FL. MR2975799

KUTOYANTS, YU. A. (1984). *Parameter Estimation for Stochastic Processes*. *Research and Exposition in Mathematics* **6**. Heldermann, Berlin. MR0777685

KUTOYANTS, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. *Springer Series in Statistics*. Springer, London. MR2144185 https://doi.org/10.1007/978-1-4471-3866-2

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.

PICCHINI, U., DE GAETANO, A. and DITLEVSEN, S. (2010). Stochastic differential mixed-effects models. *Scand. J. Stat.* **37** 67–90. MR2675940 https://doi.org/10.1111/j.1467-9469.2009.00665.x

PICCHINI, U. and DITLEVSEN, S. (2011). Practical estimation of high dimensional stochastic differential mixed-effects models. *Comput. Statist. Data Anal.* **55** 1426–1444. MR2741425 https://doi.org/10.1016/j.csda.2010.10.003

RAMSAY, J. O. and SILVERMAN, B. W. (2007). *Applied Functional Data Analysis*: *Methods and Case Studies* Springer, New York.

REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Springer, Berlin. MR1725357 https://doi.org/10.1007/978-3-662-06400-9

ROGERS, L. C. G. (1985). Smooth transition densities for one-dimensional diffusions. *Bull. Lond. Math. Soc.* **17** 157–161. MR0806242 https://doi.org/10.1112/blms/17.2.157

ROGERS, L. C. G. and WILLIAMS, D. (1990). *Diffusions*, *Markov Processes*, *and Martingales*. *Vol.* 2: *Ito Calculus*. Wiley, Chichester.

SZEGŐ, G. (1975). *Orthogonal Polynomials*: *Colloquium Publications*, *Vol. XXIII*, 4th ed. Amer. Math. Soc., Providence, R.I. MR0372517

TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12** 389–434. MR2946459 https://doi.org/10.1007/s10208-011-9099-z

WANG, J.-L., CHIOU, J.-M. and MUELLER, H.-G. (2016). Review of functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.