New results for drift estimation in inhomogeneous stochastic differential equations

Fabienne Comte[a,*], Valentine Genon-Catalot[b]

[a] *Université Paris Cité, MAP5, UMR 8145 CNRS, F-75006, FRANCE*
[b] *Université Paris Cité, MAP5, UMR 8145 CNRS, F-75006, FRANCE*

Abstract

We consider N independent and identically distributed (*i.i.d.*) stochastic processes $(X_j(t), t \in [0, T])$, $j \in \{1, \dots, N\}$, defined by a one-dimensional stochastic differential equation (SDE) with time-dependent drift and diffusion coefficient. In this context, the nonparametric estimation of a general drift function $b(t, x)$ from a continuous observation of the N sample paths on $[0, T]$ has never been investigated. Considering a set $\mathbf{I}_\epsilon = [\epsilon, T] \times A$, with $\epsilon \geq 0$ and $A \subset \mathbb{R}$, we build by a projection method an estimator of b on $\mathbf{I}_\epsilon$. As the function is bivariate, this amounts to estimating a matrix of projection coefficients instead of a vector for univariate functions. We make use of Kronecker products, which simplifies the mathematical treatment of the problem. We study the risk of the estimator and distinguish the case where $\epsilon = 0$ and the case $\epsilon > 0$ and $A = [a, b]$ compact. In the latter case, we investigate rates of convergence and prove a lower bound showing that our estimator is minimax. We propose a data-driven choice of the projection space dimension leading to an adaptive estimator. Examples of models and numerical simulation results are proposed. The method is easy to implement and works well, although computationally slower than for the estimation of a univariate function.

Keywords: Adaptive estimator, Diffusion process, Kronecker product, Nonparametric estimation, Projection method, Time-dependent coefficients.
2020 MSC: Primary 62G07, Secondary 62M05

1. Introduction

Statistical inference for the coefficients of stochastic differential equations (SDEs) has a longstanding history and a huge number of contributions deals with the subject. We can refer to the textbooks Kessler *et al.* [31], Höpfner [27], Iacus [29], Kutoyants [33, 34] and the numerous references therein. In most papers, authors assume the observation of one trajectory which may be continuously or discretely observed, on a time interval $[0, T]$. To obtain statistical results, an asymptotic framework is considered which is that, either T is fixed and the diffusion coefficient tends to 0, or T tends to infinity. In the small variance asymptotics, Markov type diffusions, *i.e.* having space and time dependent coefficients, may be considered (see *e.g.* Yoshida, N. [48], Sørensen and Uchida [42], Uchida [46], Gloter and Sørensen [24], Guy *et al.* [25]). In the long time asymptotics, only homogeneous diffusions, *i.e.* with space dependent coefficients, are studied under ergodicity assumptions. For what concerns more precisely nonparametric inference, we refer to Hoffmann [26], Dalalyan [15], Dalalyan and Reiss [16, 17], Comte *et al.* [12], Strauch [43].
In relation with functional data analysis (see *e.g.* Ramsay and Silvermann [39], Wang *et al.* [47], Hsiao [28]), the case of *i.i.d.* paths of stochastic differential equations has received recently considerable attention. Results concerning nonparametric estimation in this setup have been published (see *e.g.* Comte and Genon-Catalot [9], Denis *et al* [19, 20], Marie and Rosier [37], see also Comte and Marie [14] for identically distributed diffusions with correlated Brownian motions). These papers consider homogeneous diffusions, for which the drift and diffusion coefficients do not depend on time but only on space. Recent papers concerned with interacting particle systems assume space-time

*Corresponding author. Email address: `fabienne.comte@u-paris.fr`

dependent coefficients (see *e.g.* Della Maestra and Hoffmann [18], Comte and Genon-Catalot [10]; Belomestny *et al.* [5], Amorino *et al.* [2]). When there is no interaction between particles, these models reduce to *i.i.d.* diffusion processes.
In the foundational paper Della Maestra and Hoffmann [18], the authors study the pointwise risk of an estimator defined as a ratio of two kernel estimators. This requires two bandwidth selection procedures, one for the numerator and one for the denominator. In this paper, we consider a global estimator built by a projection method on sieves and study a global risk.

More precisely, we consider N *i.i.d.* real-valued stochastic processes $(X_i(t), t \geq 0)$, $i \in \{1, \dots, N\}$, with dynamics ruled by:

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

where $x_0 \in \mathbb{R}$ is known and $(W_1, \dots, W_N)$ are independent standard Brownian motions. The functions $b, \sigma : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ are unknown and our aim is to study nonparametric estimation of the drift function $b(t, x)$ from the continuous observation of the N sample paths on a fixed time interval $[0, T]$. We thus generalize the setting of Comte and Genon-Catalot [11] where the drift has the form $b(t, x) = \sum_{i=1}^{K} \alpha_i(t) g_i(x)$ with $g_i(x)$ known functions and $\alpha_i(t)$ unknown functions.
We estimate the function $b(t, x)$ on a set $\mathbf{I}_\epsilon = [\epsilon, T] \times A$ with $0 \leq \epsilon < T$ and $A \subset \mathbb{R}$ without making any assumption on a specific form for b as a bivariate function.

To this end, we define a collection of finite-dimensional subspaces of $\mathbb{L}^2(\mathbf{I}_\epsilon)$, $(S_{m_1} \times \Sigma_{m_2}, m_1, m_2 \geq 0)$, where S_{m_1} is spanned by an orthonormal basis $(\varphi_j, 0 \leq j \leq m_1 - 1)$ of $\mathbb{L}^2([\epsilon, T])$ and Σ_{m_2} is spanned by an orthonormal basis $(\psi_k, 0 \leq k \leq m_2 - 1)$ of $\mathbb{L}^2(A)$. As usual for projection method, we estimate a projection of $b\mathbf{1}_{\mathbf{I}_\epsilon}$ on $(S_{m_1} \times \Sigma_{m_2})$, for $m_1 \geq 1, m_2 \geq 1$ which is a function of the form

$$b_{\mathbf{m}}(t, x) = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} a_{j,k} \varphi_j \odot \psi_k(t, x), \;\; \varphi_j \odot \psi_k(t, x) = \varphi_j(t)\psi_k(x), 0 \leq j \leq m_1 - 1, 0 \leq k \leq m_2 - 1.$$

The specific challenge for estimation of bivariate functions is the fact that we have to estimate a matrix $(a_{j,k})$ of coefficients, instead of a vector for univariate functions. Therefore, the formulae quickly show intractable expressions depending on hypermatrices. The original idea of this paper is to introduce vectorization of matrices which allows to get nice expressions for estimators by means of Kronecker products. In econometrics, this is a usual way of simplifying the mathematical treatment of models, see Kleffe [32], Magnus and Neudecker [36]. In particular, the $m_1 \times m_2$ matrix $A_{\mathbf{m}} = (a_{j,k}), 0 \leq j \leq m_1 - 1, 0 \leq k \leq m_2 - 1$, $\mathbf{m} = (m_1, m_2)$ is transformed into a $m_1 m_2 \times 1$ dimensional vector $\mathrm{vec}(A_{\mathbf{m}})$ by stacking the columns of the matrix $A_{\mathbf{m}}$. Then, the classical regression equation, defining the estimator $\widehat{A}_{\mathbf{m}} = (\widehat{a}_{jk}, 0 \leq j \leq m_1 - 1, 0 \leq k \leq m_2 - 1)$ of the matrix $A_{\mathbf{m}}$, looks like the usual one,

$$\widehat{\Theta}_{m_1,m_2} \; \mathrm{vec}(\widehat{A}_{\mathbf{m}}) = \mathrm{vec}(\widehat{C}_{\mathbf{m}}).$$

where $\widehat{\Theta}_{m_1,m_2}$ is a $m_1 m_2 \times m_1 m_2$ matrix and $\mathrm{vec}(\widehat{C}_{\mathbf{m}})$ is a $m_1 m_2 \times 1$ dimensional vector. Note that both $\widehat{\Theta}_{m_1,m_2}$ and $\mathrm{vec}(\widehat{C}_{\mathbf{m}})$ are computed using the observations.

Let us stress a few points in order to highlight the novelties of our paper. First, we propose a new projection estimator of the inhomogeneous drift, a case for which studies are rare and recent (see Della Maestra and Hoffmann [18] in a more general context, but with kernels). Second, our estimator is globally adaptive with a simple and fast model selection device, for which theoretical upper bounds are obtained. This is partly due to clever algebras tricks. Third, our estimator is direct and not defined by a ratio, contrary to kernel drift estimators for diffusions. We prove its rate optimality, when excluding the neighborhood of 0 in time. The setting is anisotropic, which means that the rates are obtained with distinct regularity indexes in t and x.

In Section 2, we give the assumptions on the model. In Section 3, the projection contrast and the computation of the projection estimator are detailed. In Section 4, we consider the case $\epsilon = 0$. We prove an upper bound on the risk of our estimator for fixed $\mathbf{m}$ (Theorem 1). Then, we propose a data-driven choice of $\mathbf{m}$ leading to an adaptive estimator (Theorem 2). In Section 5, to investigate rate of convergence (Proposition 2) and lower bound (Theorem 3), we need to impose $\epsilon > 0$ and $A = [a, b]$ compact. The lower bound shows that our estimator is minimax optimal in this case. Section 6 is devoted to examples and numerical simulation results, Section 7 gives some concluding remarks and Section 8 contains all proofs. The whole estimation procedure does not depend on $\sigma(t,x)$ which may thus be unknown, except in Theorem 2 where a rough upper bound for $\sigma^2(t,x)$ is involved in the penalty. In Section 9, some properties of Kronecker products are recalled together with a Chernoff matrix inequality. The definitions of the Hermite and trigonometric bases are also recalled.

2. Assumptions

We consider the following assumptions.

- [H1-(i)] The coefficients $b(t,x), \sigma(t,x)$ are continuous real-valued functions on $\mathbb{R}^+ \times \mathbb{R}$,
- [H1-(ii)] For all R, there exists $K_R > 0$ such that, whenever $|x| \leq R, |y| \leq R, 0 \leq s \leq R$,
$$|b(s,x) - b(s,y)| \leq K_R|x-y|, \quad |\sigma(s,x) - \sigma(s,y)| \leq K_R|x-y|,$$
- [H1-(iii)] For all $T > 0$, there exists $C_T > 0$, such that, for $0 \leq s \leq T$, for all x,
$$|b(s,x)| + |\sigma(s,x)| \leq C_T(1+|x|).$$
- [H1-(iv)] For all $T > 0$, there exists $\sigma_0 > 0$, $\sigma_1 > 0$, such that $0 < \sigma_0^2 \leq \sigma^2(t,x) \leq \sigma_1^2 < +\infty$ for all $(t,x) \in [0,T] \times \mathbb{R}$,
- [H1-(v)] The function $b(t,x)$ is of class $C^{1,1}(\mathbb{R}^+ \times \mathbb{R})$, the function $\sigma(t,x)$ is of class $C^{1,2}(\mathbb{R}^+ \times \mathbb{R})$.

Under Assumption [H1(i)-(iv)], equation (1) admits a unique strong solution process $(X_i(t))$ adapted to the filtration $(\mathcal{F}_t = \sigma(W_i(s), s \leq t, i \in \{1,\dots,N\}), t \geq 0)$ (see *e.g.* Rogers and Williams [41], Theorem 12.1). The process $(X_i(t))$ is of Markov type and admits a family of transition densities $p_{s,t}(x,y)$ defined for $0 \leq s < t \leq T, x, y \in \mathbb{R}$, where $p_{s,t}(x,y)$ is equal to the density of $X_i(t)$ given $X_i(s) = x$. These densities satisfy the Kolmogorov backward equation in the backward variables $(s,x) \in [0,T] \times \mathbb{R}$: for fixed (t,y), the function $v(s,x) = p_{s,t}(x,y)$ satisfies

$$-\frac{\partial v}{\partial s} = \frac{1}{2}\sigma^2(s,x)\frac{\partial v}{\partial x^2} + b(s,x)\frac{\partial v}{\partial x}$$

and is of class $C^{1,2}([0,t) \times \mathbb{R})$. Under the additional assumption [H1-(v)], the function $w(t,y) = p_{s,t}(x,y)$ satisfies the Kolmogorov forward equation, for fixed (s,x), in the forward variables (t,y):

$$\frac{\partial w}{\partial t} = \frac{1}{2}\frac{\partial[\sigma^2(t,y)w(t,y)]}{\partial y^2} - \frac{\partial[b(t,y)w(t,y)]}{\partial y}.$$

The function $w(t,y)$ is of class $C^{1,2}((t,T] \times \mathbb{R})$ (see *e.g.* Karatzas and Shreve, p.368-369 [30], Friedman [23], p.141-148). In particular, as $X_i(0) = x_0$,

$$p_{0,t}(x_0,y) := p_t(y)$$

is the density of the random variable $X_i(t)$. The function $(t,y) \to p_t(y)$ is positive and continuous and the following holds:

$$\forall k \geq 0, \forall t \geq 0, \qquad \sup_{0 \leq u \leq t} \mathbb{E}[X_i^{2k}(u)] = \sup_{0 \leq u \leq t} \int y^{2k} p_u(y)dy < +\infty. \tag{2}$$

We also have

$$\forall k \geq 0, \forall s,t \in [0,T], \qquad \mathbb{E}[(X_i(t) - X_i(s))^{2k}] \leq C|t-s|^k, \tag{3}$$

where C is a positive constant depending on k, T, x_0 and the constant C_T of [H1-(iii)] (see *e.g.*, Karatzas and Shreve [30], p.306). Consider the function

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

For h continuous and bounded, $s \to \mathbb{E}[h(X(s))]$ is continuous on $\mathbb{R}^+$ and therefore,

$$T^{-1}\int_0^T \mathbb{E}[h(X(s))]ds = \int h(y)f_T(y)dy$$

is well defined so that the probability measure $f_T(y)dy$ is always well defined, and by (2), has moments of any order. Note that, by Theorem 1.2 of Menozzi *et al.* [38], $p_u(y) \leq Cu^{-1/2}$ so that $f_T(y)$ is finite under our set of assumptions.

3. Definition of projection estimators

Notations. Consider $0 \leq \epsilon < T$ and the set $\mathbf{I}_\epsilon := [\epsilon, T] \times A$ for $A \subset \mathbb{R}$. Let $h \in \mathbb{L}^2(I_\epsilon, dtdx)$, we set $\|h\|^2 = \int_\epsilon^T \int_A h^2(t,x)dt\,dx$. For h_1 a function of t only, $\|h_1\|^2_{[\epsilon,T]} = \int_\epsilon^T h_1^2(t)dt$ and for h_2 a function of x only, $\|h_2\|_A^2 = \int_A h_2^2(x)dx$. For a matrix M, we denote by $M^\top$ the transpose of the matrix M and by $M \otimes N$ the Kronecker product of two matrices. For M a $m \times n$-matrix, we denote by $\text{vec}(M)$ the vector of $\mathbb{R}^{mn}$ composed by stacking the columns of the matrix. In Section 9, some useful properties of Kronecker products are recalled. For $\mathbf{x}$ a vector of $\mathbb{R}^q$, we denote $\|\mathbf{x}\|_{2,q}$ its Euclidean norm. For M a $m \times m$ matrice, $\text{Tr}(M)$ denotes the trace of M, and $\|M\|_{\text{op}}^2 = \sup_{\mathbf{x}\in\mathbb{R}^m} \|M\mathbf{x}\|^2_{2,m}$.

3.1. Projection spaces and estimators

To define nonparametric estimators of the drift function b, we proceed by a projection method. More precisely, we estimate b on I_ϵ, *i.e.*, $b\mathbf{1}_{I_\epsilon}$.

Let $(\varphi_j, 0 \leq j \leq m_1 - 1)$ be an orthonormal system of bounded piecewise continuous functions of $\mathbb{L}^2([\epsilon, T], dt)$ and $(\psi_k, 0 \leq k \leq m_2 - 1)$ an orthonormal system of bounded piecewise continuous functions of $\mathbb{L}^2(A, dx)$. We define $(S_{m_1} \times \Sigma_{m_2}, m_1, m_2 \geq 0)$ a family of finite-dimensional subspaces of $\mathbb{L}^2(\mathbf{I}_\epsilon)$, where S_{m_1} is spanned by $(\varphi_j, 0 \leq j \leq m_1 - 1)$ and Σ_{m_2} is spanned by $(\psi_k, 0 \leq k \leq m_2 - 1)$. The bases of S_{m_1}, Σ_{m_2} may depend on m_1 or m_2 but for simplicity, we omit this dependence in the notations. For $m_1 \geq 1, m_2 \geq 1$, the functions

$$\varphi_j \odot \psi_k(t,x) = \varphi_j(t)\psi_k(x), \quad 0 \leq j \leq m_1 - 1, 0 \leq k \leq m_2 - 1$$

constitute an orthonormal basis of $S_{m_1} \times \Sigma_{m_2}$.

For $h : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ a function, we introduce the contrast inspired by the log-likelihood of the observations:

$$\gamma_N(h) = \frac{1}{NT}\sum_{i=1}^N \left(\int_\epsilon^T h^2(u, X_i(u))du - 2\int_\epsilon^T h(u, X_i(u))dX_i(u) \right). \tag{5}$$

For any bounded h, as $\mathbb{E}\int_\epsilon^T h^2(u, X_1(u))\sigma^2(u, X_1(u))du < +\infty$,

$$\begin{aligned}
\mathbb{E}\gamma_N(h) &= \frac{1}{T}\mathbb{E}\int_\epsilon^T (h(u, X_1(u)) - b(u, X_1(u)))^2\,du - \frac{1}{T}\mathbb{E}\int_\epsilon^T b^2(u, X_1(u))du \\
&= \frac{1}{T}\int_\epsilon^T \int (h(u,y) - b(u,y))^2 p_u(y)dydt - \frac{1}{T}\int_\epsilon^T \int b^2(u,y)p_u(y)dydu,
\end{aligned}$$

which is minimum for $h(u,y) \equiv b(u,y)$. This property justifies the definition of a collection of estimators $\widehat{b}_\mathbf{m}, \mathbf{m} = (m_1, m_2), m_1, m_2 \geq 0$ of $b_{\mathbf{I}_\epsilon} := b\mathbf{1}_{\mathbf{I}_\epsilon}$ by setting:

$$\widehat{b}_\mathbf{m} = \arg\min_{h \in S_{m_1}\times\Sigma_{m_2}} \gamma_N(h). \tag{6}$$

Thus, for each couple $\mathbf{m} = (m_1, m_2)$, we can write

$$\widehat{b}_\mathbf{m}(t,x) = \sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1} \widehat{a}_{j,k}\varphi_j \odot \psi_k(t,x) \tag{7}$$

and must compute the matrix of coefficients:

$$\widehat{A}_{\mathbf{m}} = (\widehat{a}_{j,k})_{0\le j\le m_1-1, 0\le k\le m_2-1}. \tag{8}$$

To this end, define the $m_1 \times m_2$-matrix

$$\widehat{C}_{\mathbf{m}} := \left(\frac{1}{NT}\sum_{i=1}^{N}\int_{\epsilon}^{T}\varphi_j(u)\psi_k(X_i(u))dX_i(u)\right)_{0\le j\le m_1-1, 0\le k\le m_2-1} \tag{9}$$

and the respectively $m_1 \times m_1$ and $m_2 \times m_2$ matrices

$$\Phi_{m_1}(t) \quad := \quad \left(\varphi_j(t)\varphi_{j'}(t)\right)_{1\le j,j'\le m_1-1} \quad \text{and} \quad \widehat{\Psi}_{m_2}(t) := \left(\frac{1}{N}\sum_{i=1}^{N}\psi_k(X_i(t))\psi_{k'}(X_i(t))\right)_{0\le k,k'\le m_2-1}. \tag{10}$$

We also define

$$\Psi_{m_2}(t) := \left(\int \psi_k(x)\psi_{k'}(x)p_t(x)dx\right)_{0\le k,k'\le m_2-1} = \mathbb{E}\widehat{\Psi}_{m_2}(t). \tag{11}$$

The matrices $\Phi_{m_1}(t)$, $\widehat{\Psi}_{m_2}(t)$ and $\Psi_{m_2}(t)$ are symmetric nonnegative. For instance, for $\mathbf{x} = (x_0, \ldots, x_{m_1-1})$, $\mathbf{x}^\top\Phi_{m_1}(t)\mathbf{x} = (\sum_{j=0}^{m_1-1} x_j\varphi_j(t))^2$ and analogously for the other matrices.

Lastly, define the $m_1m_2 \times m_1m_2$ matrices

$$\widehat{\Theta}_{m_1,m_2} := \frac{1}{T}\int_{\epsilon}^{T}\widehat{\Psi}_{m_2}(t)\otimes\Phi_{m_1}(t)dt, \quad \Theta_{m_1,m_2} := \frac{1}{T}\int_{\epsilon}^{T}\Psi_{m_2}(t)\otimes\Phi_{m_1}(t)dt, \tag{12}$$

where clearly $\Theta_{m_1,m_2} = \mathbb{E}\widehat{\Theta}_{m_1,m_2}$. As for all t, the matrices $\Phi_{m_1}(t), \widehat{\Psi}_{m_2}(t), \Psi_{m_2}(t)$ are symmetric, so is $\widehat{\Psi}_{m_2}(t)\otimes\Phi_{m_1}(t)$ (see (39)). As a consequence, $\widehat{\Theta}_{m_1,m_2}$ and Θ_{m_1,m_2} are also symmetric.

The following proposition provides an explicit formula for the matrix of coefficient $\widehat{A}_{\mathbf{m}}$ of our estimator.

Proposition 1. *The matrix $\widehat{A}_{\mathbf{m}}$ satisfies:*

$$\frac{1}{T}\int_{\epsilon}^{T}\Phi_{m_1}(t)\widehat{A}_{\mathbf{m}}\widehat{\Psi}_{m_2}(t)dt = \widehat{C}_{\mathbf{m}}. \tag{13}$$

Moreover (13) *is equivalent to*

$$\widehat{\Theta}_{m_1,m_2}\ \mathrm{vec}(\widehat{A}_{\mathbf{m}}) = \mathrm{vec}(\widehat{C}_{\mathbf{m}}). \tag{14}$$

Equation (14) allows us to compute explicitely the coefficients of the estimator $\widehat{b}_{\mathbf{m}}$ and is obtained by vectorialization of (13).

The vector of coefficients of the function $\widehat{b}_{\mathbf{m}}$ defined by (7) contained in $\mathrm{vec}(\widehat{A}_{\mathbf{m}})$ is uniquely defined if $\widehat{\Theta}_{m_1,m_2}$ is invertible and in this case,

$$\mathrm{vec}(\widehat{A}_{\mathbf{m}}) = \widehat{\Theta}^{-1}_{m_1,m_2}\mathrm{vec}(\widehat{C}_{\mathbf{m}}). \tag{15}$$

Below (see Lemma 1 and condition (18)), we prove that Θ_{m_1,m_2} is invertible and give a condition ensuring that $\widehat{\Theta}_{m_1,m_2}$ is invertible.

For further use, we introduce the empirical norm and scalar product associated with our observations. For $h(.,.), \ell(.,.)$ two bounded functions, we set

$$\|h\|_N^2 = \frac{1}{NT}\sum_{i=1}^{N}\int_{\epsilon}^{T}h^2(u, X_i(u))du, \quad \langle h, \ell\rangle_N = \frac{1}{NT}\sum_{i=1}^{N}\int_{\epsilon}^{T}h(u, X_i(u))\ell(u, X_i(u))du. \tag{16}$$

We also set

$$\nu_N(h) = \frac{1}{NT}\sum_{i=1}^{N}\int_{\epsilon}^{T} h(u, X_i(u))\sigma(u, X_i(u))dW_i(u). \tag{17}$$

Therefore, $\mathbb{E}\|h\|_N^2 = \|h\|_p^2 := \frac{1}{T}\int_{\epsilon}^{T}\int_{\mathbb{R}} h^2(u,y)p_u(y)dy\,du$,

$$\mathbb{E}\langle h, \ell\rangle_N = \langle h, \ell\rangle_p := \frac{1}{T}\int_{\epsilon}^{T}\int_{\mathbb{R}} h(u,y)\ell(u,y)p_u(y)dy\,du$$

and $\mathbb{E}\nu_N(h) = 0$, $\mathbb{E}\nu_N^2(h) = \|h\sigma\|_p^2/NT$.

4. Estimation for $\epsilon = 0$

In this section, we consider $\epsilon = 0$ and set $\mathbf{I}_0 = \mathbf{I}$. We keep everywhere the same notations but with $\epsilon = 0$.

4.1. Identifiability constraint

The following Lemma clarifies the identifiability constraint introduced below and allows us to study the invertibility of Θ_{m_1,m_2} and $\widehat{\Theta}_{m_1,m_2}$.

Lemma 1. *Let $h = \sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1} h_{j,k}\varphi_j \odot \psi_k$ and denote by $H_{\mathbf{m}} = (h_{j,k})_{0\le j\le m_1-1, 0\le k\le m_2-1}$. We have*

$$\begin{aligned}\|h\|_N^2 &= \frac{1}{T}\int_0^T \mathrm{Tr}\left[H_{\mathbf{m}}^\top \Phi_{m_1}(t) H_{\mathbf{m}} \widehat{\Psi}_{m_2}(t)\right]dt = [\mathrm{vec}(H_{\mathbf{m}})]^\top \widehat{\Theta}_{m_1,m_2}\mathrm{vec}(H_{\mathbf{m}}),\\ \|h\|_p^2 &= \mathbb{E}\left(\frac{1}{T}\int_0^T \mathrm{Tr}\left[H_{\mathbf{m}}^\top \Phi_{m_1}(t) H_{\mathbf{m}} \widehat{\Psi}_{m_2}(t)\right]dt\right) = [\mathrm{vec}(H_{\mathbf{m}})]^\top \Theta_{m_1,m_2}\mathrm{vec}(H_{\mathbf{m}}),\end{aligned}$$

where Θ_{m_1,m_2} and $\widehat{\Theta}_{m_1,m_2}$ are defined in (12).

Lemma 1 implies that $\widehat{\Theta}_{m_1,m_2}$ and Θ_{m_1,m_2} are nonnegative. Moreover, as $p_t(y) > 0$, $\forall t \in]0,T]$ and $\forall y \in \mathbb{R}$,

$$\|h\|_p^2 = 0 \Rightarrow h(t,y) = 0 \ \text{ a.e. on }]0,T]\times\mathbb{R}$$

and thus $h_{j,k} = 0$ for $0 \le j \le m_1 - 1$ and $0 \le k \le m_2 - 1$. Consequently, Θ_{m_1,m_2} is positive definite. As $\widehat{\Theta}_{m_1,m_2}$ tends to Θ_{m_1,m_2} as $N \to +\infty$ a.s., $\widehat{\Theta}_{m_1,m_2}$ is invertible for N large enough.

Note that we have

$$\|h\|_N^2 = 0 \Rightarrow \forall i,\ h(u, X_i(u)) = 0 \quad \text{a.s. and a.e. on } [0,T].$$

Therefore, $\widehat{\Theta}_{m_1,m_2}$ is invertible if:

$$\begin{aligned}&\left\{\forall u \in [0,T], \forall i \in \{1,\dots,N\}, \sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1} h_{j,k}\varphi_j \odot \psi_k(u, X_i(u)) = 0 \text{ a.s. and a.e. on } [0,T]\right\} \\ &\Rightarrow \left\{h_{j,k} = 0, \forall j \in \{0,\dots,m_1-1\} \text{ and } \forall k \in \{0,\dots,m_2-1\}\right\}.\end{aligned} \tag{18}$$

Condition (18) is an identifiability constraint linked with the choice of the bases. In Section 6, we use bases for which (18) is fulfilled.

4.2. *Empirical risk of the estimator for fixed* $\mathbf{m}$

We define the risk of the estimator $\widehat{b}_{\mathbf{m}}$ given by (7) as the expectation of the empirical square norm $\|.\|_N^2$ (with $\epsilon = 0$) which is naturally associated with our observations.

Next, we need to define key quantities related to the bases. Let us set:

$$L_\varphi(S_{m_1}) = \sup_{t\in[\epsilon,T]} \sum_{j=0}^{m_1-1} \varphi_j^2(t), \quad L_\psi(\Sigma_{m_2}) = \sup_{x\in A} \sum_{j=0}^{m_2} \psi_j^2(x). \tag{19}$$

These definitions were introduced by Birgé and Massart [6], who also remark that:

$$L_\varphi(S_{m_1}) = \sup_{h_1\in S_{m_1}, \|h_1\|_{[\epsilon,T]}=1} \sup_{t\in[\epsilon,T]} h_1^2(t), \quad L_\psi(\Sigma_{m_2}) = \sup_{h_2\in\Sigma_{m_2}, \|h_2\|_A=1} \sup_{x\in A} h_2^2(x).$$

Therefore, $L_\varphi(S_{m_1})$ and $L_\psi(\Sigma_{m_2})$ only depend on the subspaces and not on the bases chosen to define them. We consider the following assumption:

- [H2] $\exists c_\varphi, c_\psi > 0$ such that $L_\varphi(S_{m_1}) \le c_\varphi^2 m_1$, $L_\psi(\Sigma_{m_2}) \le c_\psi^2 m_2$.

Assumption [H2] holds for several classical bases, for instance, for the trigonometric basis on a compact subset of $\mathbb{R}$ (see Section 5). If we take $A = \mathbb{R}$ and for $(\psi_k)_k$ the Hermite basis, it satisfies the weakened condition $L_\psi(\Sigma_{m_2}) \le c m_2^{1/2}$ (see Comte and Lacour [13], Lemma 1, examples in Section 6 and Appendix).

The following condition restricts the possible choices of m_1, m_2 to ensure the stability of the minimum contrast estimator (see *e.g.* Cohen *et al.* [7, 8]).

$$m_1 m_2 \le NT, \quad L_\psi(m_2)\|\Theta_{m_1,m_2}^{-1}\|_{\mathrm{op}} \le \mathfrak{c}_r \frac{NT}{\log(NT)}, \quad \mathfrak{c}_r = \frac{3\log(3/2)-1}{2+2r}. \tag{20}$$

Under (20), the matrix Θ_{m_1,m_2} is not only invertible but has all its eigenvalues bounded from below.

Condition (18) is not enough for our theory and we need to reinforce (20). For this, we define:

$$\widehat{\Lambda}_{\mathbf{m}} = \left\{ m_1 m_2 \le NT, \quad L_\psi(m_2)\|\widehat{\Theta}_{m_1,m_2}^{-1}\|_{\mathrm{op}} \le 2\mathfrak{c}_r \frac{NT}{\log(NT)} \right\}. \tag{21}$$

On the set $\widehat{\Lambda}_{\mathbf{m}}$, the matrix $\widehat{\Theta}_{m_1,m_2}$ is invertible and its eigenvalues are all bounded from below. Now we define the truncated estimator that we study for fixed (m_1, m_2):

$$\widetilde{b}_{\mathbf{m}} = \widehat{b}_{\mathbf{m}} \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}}. \tag{22}$$

Theorem 1. *Under assumptions [H1], [H2] and condition (20) on* $\mathbf{m}$ *with* $r > 7$ *in (20)-(21), it holds*

$$\mathbb{E}(\|\widetilde{b}_{\mathbf{m}} - b_{\mathbf{I}}\|_N^2) \le \inf_{h\in S_{m_1}\times\Sigma_{m_2}} \|h - b_{\mathbf{I}}\|_p^2 + \frac{2\sigma_1^2 m_1 m_2}{NT} + \frac{C}{NT}, \tag{23}$$

where σ_1^2 *is the upper bound on* σ^2 *defined in* [H1]-(iv)*, and* C *is a positive constant.*

The risk bound (23) is the sum of a squared bias term and a variance term of order $m_1 m_2/NT$, the last term C/NT being negligible.

The variance term is actually

$$\frac{2}{NT}\mathrm{Tr}\left(\Theta_{m_1,m_2}^{-1}\Theta_{m_1,m_2,\sigma^2}\right) \le \frac{2\sigma_1^2 m_1 m_2}{NT},$$

where

$$\Theta_{m_1,m_2,\sigma^2} = \frac{1}{T}\int_0^T \Psi_{m_2,\sigma^2}(t) \otimes \Phi_{m_1}(t)dt, \quad \Psi_{m_2,\sigma^2}(t) = \left(\int \psi_k(x)\psi_{k'}(x)\sigma^2(t,x)p_t(x)dx\right)_{0\le k,k'\le m_2-1}.$$

4.3. Adaptive estimation

In this paragraph, we investigate the possible choice of a data-driven $\mathbf{m}$ which leads to an adaptive estimator realizing automatically the bias-variance compromise.
Consider, for $\mathbf{c}$ a positive constant, the theoretical collection of models:

$$\mathcal{M}_N = \left\{ \mathbf{m} \in \mathbb{N}^2, m_1 m_2 \leq NT,\ L_\psi(m_2)(\|\Theta_{m_1,m_2}\|_{\text{op}}^2 \vee 1) \leq \frac{\mathbf{c}}{2} \frac{NT}{\log^2(NT)} \right\}$$

and its emprirical counterpart:

$$\widehat{\mathcal{M}_N} = \left\{ \mathbf{m} \in \mathbb{N}^2, m_1 m_2 \leq NT,\ L_\psi(m_2)(\|\widehat{\Theta}_{m_1,m_2}\|_{\text{op}}^2 \vee 1) \leq 2\mathbf{c} \frac{NT}{\log^2(NT)} \right\}.$$

Then define

$$\widehat{\mathbf{m}} = \arg \min_{\mathbf{m} \in \widehat{\mathcal{M}_N}} \left(\gamma_N(\widehat{b}_{\mathbf{m}}) + \ \text{pen}(\mathbf{m}) \right), \quad \text{pen}(\mathbf{m}) = \kappa \sigma_1^2 \frac{m_1 m_2}{NT}, \tag{24}$$

where κ is a numerical constant and σ_1^2 is the upper bound on σ^2 (see [H1-iv]). The following result holds:

Theorem 2. *Under assumptions [H1], [H2], there exists a numerical constant κ_0 such that for all $\kappa \geq \kappa_0$,*

$$\mathbb{E}(\|\widehat{b}_{\widehat{\mathbf{m}}} - b_{\mathbf{I}_\epsilon}\|_N^2) \leq C \inf_{\mathbf{m} \in \mathcal{M}_N} \left(\inf_{h \in S_{m_1} \times \Sigma_{m_2}} \|h - b_{\mathbf{I}_\epsilon}\|^2 + \frac{\sigma_1^2 m_1 m_2}{NT} \right) + \frac{C'}{NT},$$

where C and C' are positive constants.

The proof is omitted as it follows closely the analogous result in Comte and Genon-Catalot [9]. It relies on the standard decomposition, for $h, h^\star$ two functions of $S_{m_1} \times \Sigma_{m_2}$,

$$\gamma_N(h) - \gamma_N(h^\star) = \|h - h^\star\|_N^2 + 2\nu_N(h - h^\star)$$

where $\nu_N(h)$ is defined by (17). The other classical point is that $NT\nu_N(h) := M_T$ is a martingale, with bracket

$$\langle M \rangle_T = \int_\epsilon^T \sum_{i=1}^N h^2(u, X_i(u))\sigma^2(u, X_i(u))du,$$

satisfying $\langle M \rangle_T \leq \sigma_1^2 \|h\|_N^2$. Therefore, the following Bernstein inequality for martingales (see *e.g.* Revuz and Yor [40]) holds

$$\mathbb{P}(\nu_N(h) \geq \delta,\ \langle M \rangle_T \leq v^2) \leq \exp\left(-\frac{NT\delta^2}{2\sigma_1^2 v^2} \right).$$

Using this and the chaining method described in Baraud *et al.* [4] gives the result.

5. Optimal rate for compact support estimation with $\epsilon > 0$.

In this section, we study rates of estimation and their optimality, which requires to work in a setting where weighted and standard $\mathbb{L}^2$-norms are equivalent: this can only be done on a compact set, and for $\epsilon > 0$. The first constraint is standard (see Baraud [3]'s first assumption p.130) and the second one specific to our setting.

Therefore here, we consider $\epsilon > 0$ and $A = [a, b],\ a < b$, a compact interval of $\mathbb{R}$. Then there exist two positive constants $c_0(\epsilon, A)$ and $c_1(\epsilon, A)$ such that:

$$\forall t \in [\epsilon, T], \forall x \in A, \quad 0 < c_0(\epsilon, A) \leq p_t(x) \leq c_1(\epsilon, A) < +\infty.$$

5.1. Upper bound

The following Lemma allows to simplify the stability condition (20).

Lemma 2. *Under [H1], $\epsilon > 0$ and A compact, Θ_{m_1,m_2} is invertible and $\|\Theta^{-1}_{m_1,m_2}\|_{\text{op}} \le 1/c_0(\epsilon, A)$. Moreover $\|h\|_p^2 \le c_1(\epsilon, A)\|h\|^2$.*

Now, reminding that c^2_ψ is defined in [H2], the stability condition (20) holds if:

$$m_2 \le \frac{\mathfrak{c}_r c_0(\epsilon, A)}{c^2_\psi} \frac{NT}{\log(NT)}.$$

As $c_0(\epsilon, A)$ is unknown, we can make the choice of (m_1, m_2) within the set

$$\mathcal{M}^\star_N = \left\{(m_1, m_2) \in \mathbb{N}^2,\ m_1 m_2 \le NT, m_2 \le NT/\log^2(NT)\right\}. \tag{25}$$

From Theorem 1 and Lemma 2, under assumption [H1], and for any $\mathbf{m} \in \mathcal{M}^\star_N$, we have

$$\mathbb{E}(\|\widetilde{b}_{\mathbf{m}} - b_{\mathbf{I}_\epsilon}\|^2_N) \le c_1(\epsilon, A) \inf_{h \in S_{m_1} \times \Sigma_{m_2}} \|h - b_{\mathbf{I}_\epsilon}\|^2 + 2\frac{\sigma_1^2 m_1 m_2}{NT} + \frac{C}{NT}, \tag{26}$$

where C is a positive constant.

Let us discuss the rate of convergence of the risk. For $\boldsymbol{\beta} = (\beta_1, \beta_2) \in \mathbb{R}^+ \times \mathbb{R}^+$, and $\mathbf{R} = (R_1, R_2) \in \mathbb{R}^+ \times \mathbb{R}^+$, define the regularity space:

$$W^*(\boldsymbol{\beta}, \mathbf{R}) = \left\{ f \in \mathbb{L}^2([\epsilon, T] \times [a, b]),\ \text{such that} \sum_{j,k \ge 0} c^2_{j,k} j^{2\beta_1} \le R_1^2,\ \sum_{j,k\ge 0} c^2_{j,k} k^{2\beta_2} \le R_2^2 \right\},$$

where $c_{j,k} = \langle f, \varphi_j \odot \psi_k \rangle$.

Proposition 2. *(upper bound) Under assumptions [H1], [H2], if $b_{\mathbf{I}_\epsilon} \in W^*(\boldsymbol{\beta}, \mathbf{R})$, choosing*

$$m_1^\star \propto (NT)^{\beta_2/(\beta_1+\beta_2+2\beta_1\beta_2)}, \quad m_2^\star \propto (NT)^{\beta_1/(\beta_1+\beta_2+2\beta_1\beta_2)}$$

we get

$$\mathbb{E}(\|\widetilde{b}_{\mathbf{m}^\star} - b_{\mathbf{I}_\epsilon}\|^2_N) \lesssim (NT)^{-\frac{2\bar\beta}{2\bar\beta+2}}, \quad \frac{1}{\bar{\boldsymbol{\beta}}} = \frac{1}{2}\left(\frac{1}{\beta_1} + \frac{1}{\beta_2}\right).$$

The resulting rate is the classical nonparametric rate over anisotropic regularity spaces.

5.2. Lower bound

Define for $\boldsymbol{\beta} = (\beta_1, \beta_2) \in \mathbb{N}^2$ and $\mathbf{L} = (L_1, L_2)$,

$$\begin{aligned} W(\boldsymbol{\beta}, \mathbf{L}) = \{ & f \in \mathbb{L}^2([\epsilon, T] \times [a, b]), \\ & f \text{ derivable up to order } \beta_1 \text{ w.r.t. } t, \text{ up to order } \beta_2 \text{ w.r.t. } x, \\ & \iint (\partial^{\beta_1} f(t,x)/\partial t^{\beta_1})^2 dtdx \le L_1^2, \iint (\partial^{\beta_2} f(t,x)/\partial x^{\beta_2})^2 dtdx \le L_2^2 \}. \end{aligned}$$

Then we can prove:

Theorem 3. *Under Assumption [H1], the following lower bound holds:*

$$\liminf_{N\to+\infty} \inf_{T_N} \sup_{b_{\mathbf{I}_\epsilon} \in W(\boldsymbol{\beta}, \mathbf{L})} \mathbb{E}_{b_{\mathbf{I}_\epsilon}}[N^{\frac{2\bar\beta}{2\bar\beta+2}} \|T_N - b_{\mathbf{I}_\epsilon}\|^2] \ge c,$$

where $\inf_{T_N}$ denotes the infimum over all estimators and c is a constant depending on $\mathbf{L}$ and $\boldsymbol{\beta}$.

5.3. *Case of the trigonometric basis*

To illustrate and make more concrete the sets $W(\boldsymbol{\beta}, \mathbf{L})$ and $W^*(\boldsymbol{\beta}, \mathbf{R})$, let us consider trigonometric bases:

$$\begin{cases} \varphi_0(t) = \frac{1}{\sqrt{T-\epsilon}} \mathbf{1}_{[\epsilon,T]}(t), \\ \varphi_{2j-1}(t) = \frac{2}{\sqrt{T-\epsilon}} \cos(2\pi j \frac{t-\epsilon}{T-\epsilon}) \mathbf{1}_{[\epsilon,T]}(t), \\ \varphi_{2j}(t) = \frac{2}{\sqrt{T-\epsilon}} \sin(2\pi j \frac{t-\epsilon}{T-\epsilon}) \mathbf{1}_{[\epsilon,T]}(t), \end{cases} \qquad \begin{cases} \psi_0(x) = \frac{1}{\sqrt{b-a}} \mathbf{1}_{[a,b]}(t), \\ \psi_{2k-1}(x) = \frac{2}{\sqrt{b-a}} \cos(2\pi k \frac{x-a}{b-a}) \mathbf{1}_{[a,b]}(x), \\ \psi_{2k}(x) = \frac{2}{\sqrt{b-a}} \sin(2\pi k \frac{x-a}{b-a}) \mathbf{1}_{[a,b]}(x), \end{cases}$$

for $j \geq 1$ and $k \geq 1$.

Now, to take the boundary conditions into account, define the following set:

$$\begin{aligned} W^{(\mathrm{per})}(\boldsymbol{\beta}, \mathbf{L})) &= \{f \in W(\boldsymbol{\beta}, \mathbf{L}), \forall x \in [a,b], (\partial^{\alpha_1} f(t,x)/(\partial t^{\alpha_1})(\epsilon, x) = (\partial^{\alpha_1} f(t,x)/\partial t^{\alpha_1})(T, x), \\ &\quad \forall t \in [\epsilon, T], (\partial^{\alpha_2} f(t,x)/\partial x^{\alpha_2})(t,a) = (\partial^{\alpha_2} f(t,x)/\partial x^{\alpha_2})(t,b), 0 \leq \alpha_i \leq \beta_i - 1, i \in \{1,2\}\}. \end{aligned}$$

We can prove the result.

Proposition 3. *For integers β_1 and β_2, $\mathbf{L} = (L_1, L_2), \mathbf{R} = (R_1, R_2)$*

$$f \in W^{(\mathrm{per})}(\boldsymbol{\beta}, \mathbf{L}) \quad \Rightarrow \quad f \in W^*(\boldsymbol{\beta}, \mathbf{R})$$

with $R_1^2 = L_1^2 (T-\epsilon)^{2\beta_1} / \pi^{2\beta_1}, R_2^2 = L_2^2 (b-a)^{2\beta_2} / \pi^{2\beta_2}$.

Theorem 3 holds also on $W^{(\mathrm{per})}(\boldsymbol{\beta}, \mathbf{L})$ (see the proof of Theorem 3, where the propositions g_j, h_k belong to this space). Therefore, the lower bound holds for this function space. By Proposition 3, these functions are in $W^*(\boldsymbol{\beta}, \mathbf{R})$ and therefore, the upper bound holds. We conclude that the rates are minimax optimal on this set.

6. Examples and numerical simulation results

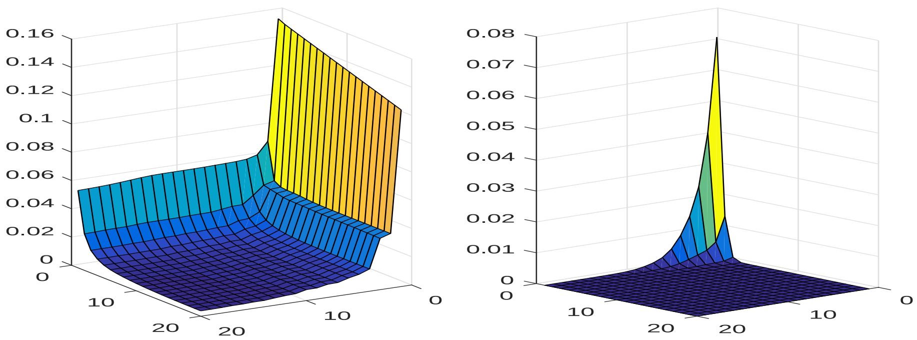

Fig. 1: Smallest eigenvalue of $\widehat{\Theta}_{m_1,m_2}$ defined by (12) in function of m_1 and m_2 going from 1 to 20, on one sample with size $N = 1000$ and function $b_1(t,x) = \cos(\pi x t/2)$. Left: trigonometric basis $\mathbf{T}$, right: Hermite basis $\mathbf{H}$.

We implement the method on some examples. The data are generated by a basic Euler scheme with $T = 2$ and $n = 100$ observations for each path (step $\Delta = 2/100$), with constant σ equal to 0.25 and functions

$$b_1(t,x) = \cos(\pi x\, t/2), \quad b_2(t,x) = \frac{xt}{\sqrt{1+t^2}\sqrt{1+x^2}}, \quad b_3(t,x) = \tanh(xt)$$

where tanh denotes the hyperbolic tangent. They are regular and bounded functions. The results below are given for $N = 1000$.

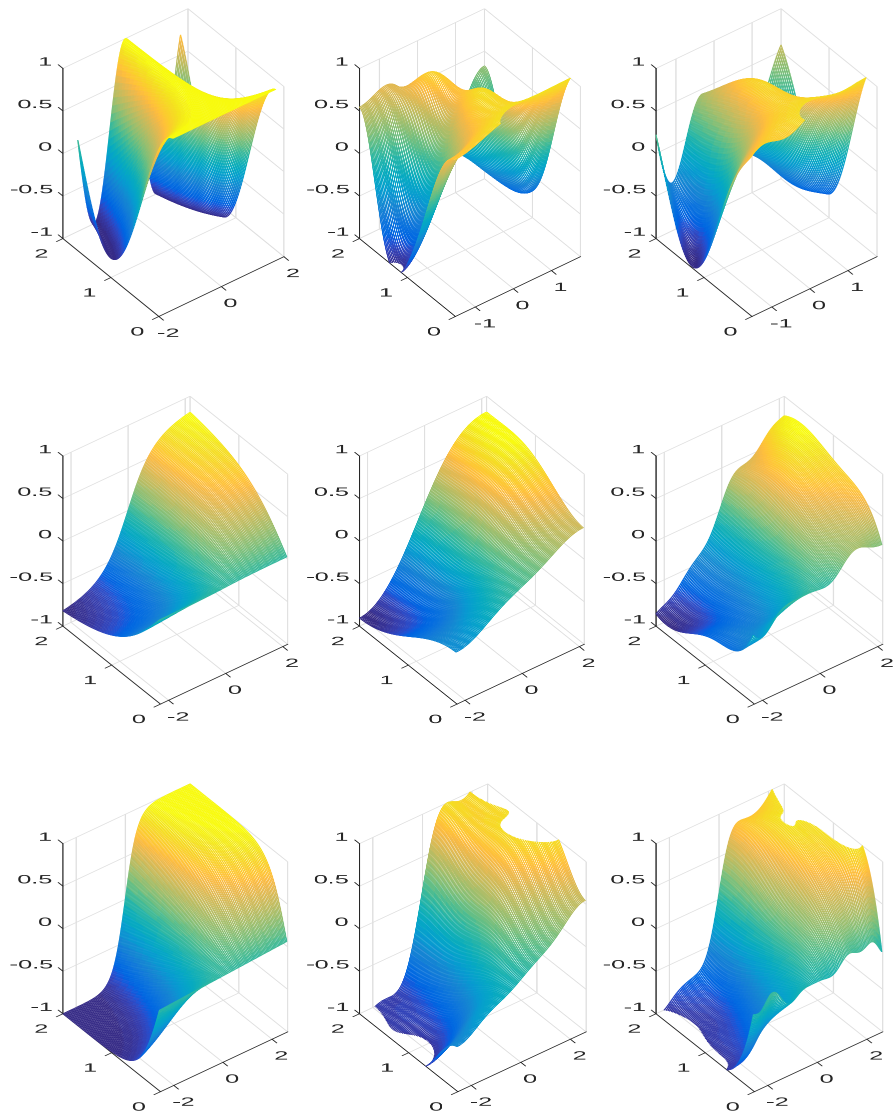

Fig. 2: Adaptive sstimation of $(t, x) \mapsto b_1(t, x) = \cos(\pi x t/2)$ (line 1), $b_2(t, x) = \frac{xt}{\sqrt{1+t^2}\sqrt{1+x^2}}$, (line 2), $b_3(t, x) = \tanh(xt)$ (line 3). Left: True function, Middle: estimated function with trigonometric basis **T**, Right: estimated function with Hermite basis **H**. On sample $N = 1000$ processes, the selected couples of dimensions (m_1, m_2) are $(6, 4)$, $(3, 4)$, $(4, 6)$ for **T** and $(5, 9)$, $(4, 10)$, $(4, 14)$ for **H**, resp. for the estimators of b_1, b_2, b_3.

We compute the estimators using either the half-trigonometric basis $((1, \sqrt{2}\cos(\pi jx), j \geq 1))$, denoted by **T**, or the Hermite basis, denoted by **H**, which is not orthonormal on the compact domain which is considered here (see Section 9). Basis **T** requires the definition of the domain of estimation which is $[1/n = \epsilon, 2] \times [a, b]$ where a is taken as the 5%-quantile of all the data and b as the 95%-quantile. This domain is also used for graphical representations.

The estimator is computed for the selected dimensions obtained from formula (24) with $\kappa_T = 4$ and $\kappa_H = 6$ for the value of κ with basis **T** and **H** respectively. The term σ_1^2 is equal to σ^2 in the constant variance case, and the true value is used (a residual least square estimator may be computed).

We plot in Fig. 1 the surface corresponding to the smallest eigenvalue of $\widehat{\Theta}_{m_1,m_2}$ for each basis and for function b_1, to show that it is decreasing when m_1 or m_2 increases, and this decrease is much faster for Hermite basis. A cutoff is set in the program to compute the estimator only if the inverse of the eigenvalue is less than N^4. The estimator is set to zero otherwise, so that the associated dimensions are not selected.

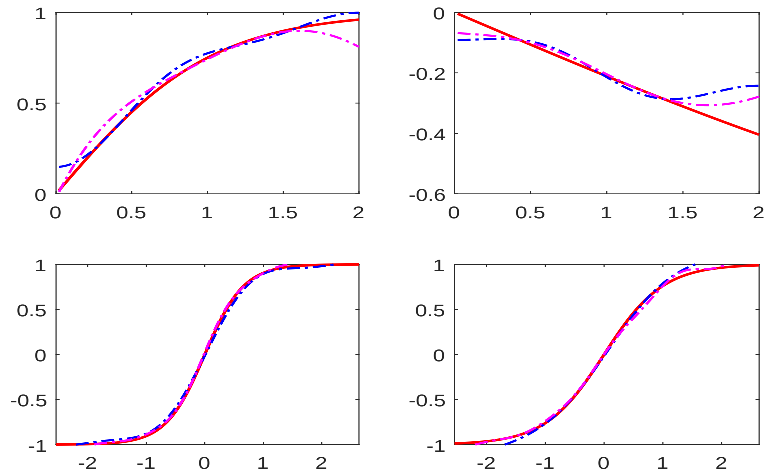

Fig. 3: Estimation of function $b_3 = \tanh(xt)$ for fixed values of t or x, $N = 1000$ paths. True b_3 in full red line, Estimation with trigonometric basis $\mathbf{T}$ in dotted blue, Estimation with Hermite basis $\mathbf{H}$ in dotted magenta. First line: fixed x, $x \simeq 1$ (left) and $x \simeq -0.2$ (right). Second line fixed t, $t = 1.5$ (left) and $t = 1$ (right).

Fig. 2 gives the 3-D representation of the functions $(t, x) \mapsto b_i(t, x)$ for $i \in \{1, 2, 3\}$ and their estimates in the two bases for the dimensions selected by the procedure. Clearly on these examples, the Hermite basis is better for b_1 and the trigonometric for the function b_2. We present in Fig. 3 sections of the last surface corresponding to b_3, for fixed values of x or t (not too near of the boarders). The curves in this case are very good, but some shifts or side effects can occur in other examples.

We can conclude that the method is easy to implement and works well, the main drawback is that it is computationally slower than for univariate estimation.

7. Concluding remarks

In this paper, we consider N independent and identically distributed (*i.i.d.*) stochastic processes $(X_j(t), t \in [0, T])$, $j = 1, \dots, N$, defined by a one-dimensional stochastic differential equation (SDE) with general time-dependent drift and diffusion coefficient. Considering a set $\mathbf{I}_\epsilon = [\epsilon, T] \times A$, with $\epsilon \geq 0$ and $A \subset \mathbb{R}$, we build by a projection method an estimator of b on $\mathbf{I}_\epsilon$. The introduction of Kronecker products simplifies and clarifies the mathematical treatment of this estimation problem. We study the risk of the estimator first in the case where $\epsilon = 0$, second in the case $\epsilon > 0$ and $A = [a, b]$ compact. In the latter case, we investigate rates of convergence and prove a lower bound showing that our estimator is minimax. We propose a data-driven choice of the projection space dimension leading to an adaptive estimator.

Extensions of this work is to consider discrete observations of the sample paths which would lead to an asymptotic condition linking the discretisation step and the size N of the sample. Obviously, the problem of estimating $\sigma(t, x)$ in the context of discrete observations is worth of investigation.

The problem of studying the estimation of a drift $b(t, x_1, \dots, x_d)$ in a d-dimensional diffusion model using *i.i.d.* sample paths is of the utmost interest. Specifically, we would consider $A \subset \mathbb{R}^d$ instead of a subset of $A \subset \mathbb{R}$. This is indeed theoretically possible. For instance, this has been investigated in the case of multivariate regression models by Dussap [21].The formalism is at first sight very similar, but it requires the definition of hypermatrices and associated tools of algebra in this context. In that way, the theory could be extended to multidimensional diffusion models. Nevertheless, implementation would be much more problematic.

8. Proofs

Proof of Proposition 1. For $h = \sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1} h_{j,k}\varphi_j \odot \psi_k$, we note that

$$\frac{\partial h^2(u, X_i(u))}{\partial h_{j_0,k_0}} = 2\varphi_{j_0}(u)\psi_{k_0}(X_i(u))h(u, X_i(u)).$$

So, we have for $j_0 \in \{0, \dots, m_1 - 1\}$, and $k_0 \in \{0, \dots, m_2 - 1\}$

$$\frac{\partial \gamma_N}{\partial h_{j_0,k_0}}(h) = \frac{2}{NT}\sum_{i=1}^{N}\int_{\epsilon}^{T}\left[\varphi_{j_0}(u)\psi_{k_0}(X_i(u))h(u, X_i(u))du - \varphi_{j_0}(u)\psi_{k_0}(X_i(u))dX_i(u)\right].$$

Setting this to 0 for $h = \widehat{b}_{\mathbf{m}}$ defines the coefficients $\widehat{a}_{j,k}$ as satisfying

$$\frac{1}{T}\int_{\epsilon}^{T}\sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1}\underbrace{\varphi_{j_0}(u)\varphi_j(u)}_{[\Phi_{m_1}(u)]_{j_0,j}}\widehat{a}_{j,k}\underbrace{\frac{1}{N}\sum_{i=1}^{N}\psi_{k_0}(X_i(u))\psi_k(X_i(u))\,du}_{[\widehat{\Psi}_{m_2}]_{k,k_0}} = \frac{1}{NT}\sum_{i=1}^{N}\int_{\epsilon}^{T}\varphi_{j_0}(u)\psi_{k_0}(X_i(u))dX_i(u) = [\widehat{C}_{\mathbf{m}}]_{j_0,k_0}$$

for all for $j_0 =\in \{0, \dots, m_1 - 1\}, k_0 \in \{0, \dots, m_2 - 1\}$. This can be summarized by (13).

To clarify further equation (13) defining $\widehat{A}_{\mathbf{m}}$, we introduce the vectorization of matrices $\widehat{A}_{\mathbf{m}}$, $\widehat{C}_{\mathbf{m}}$. We vectorize relation (13) and obtain:

$$\text{vec}\Big(\frac{1}{T}\int_{\epsilon}^{T}\Phi_{m_1}(t)\widehat{A}_{\mathbf{m}}\widehat{\Psi}_{m_2}(t)dt\Big) = \frac{1}{T}\int_{\epsilon}^{T}\text{vec}(\Phi_{m_1}(t)\widehat{A}_{\mathbf{m}}\widehat{\Psi}_{m_2}(t))dt = \text{vec}(\widehat{C}_{\mathbf{m}}).$$

Now using the relation $\text{vec}(ABC) = (C^\top \otimes A)\text{vec}(B)$ (linking the three matrices A, B, C), we get, as $\widehat{A}_{\mathbf{m}}$ does not depend on t,

$$\frac{1}{T}\int_{\epsilon}^{T}\text{vec}(\Phi_{m_1}(t)\widehat{A}_{\mathbf{m}}\widehat{\Psi}_{m_2}(t))dt = \frac{1}{T}\int_{\epsilon}^{T}\widehat{\Psi}_{m_2}(t)\otimes\Phi_{m_1}(t)\text{vec}(\widehat{A}_{\mathbf{m}})dt \;= \widehat{\Theta}_{m_1,m_2}\,\text{vec}(\widehat{A}_{\mathbf{m}}).$$

This explains the second formula of Proposition 1. □

Proof of Lemma 1. The begining of the computation is straightforward.

$$\begin{aligned}
\|h\|_N^2 &= \frac{1}{T}\sum_{0\le j,j'\le m_1-1}\sum_{0\le k,k'\le m_2-1} h_{j,k}h_{j',k'}\int_0^T[\Phi_{m_1}(u)]_{j,j'}[\widehat{\Psi}_{m_2}(u)]_{k,k'}du \\
&= \frac{1}{T}\int_0^T\sum_{k=0}^{m_2-1}\left\{\left(\sum_{j=0}^{m_1-1}[H_{\mathbf{m}}^\top]_{k,j}\left[\sum_{k'=0}^{m_2-1}\left(\sum_{j'=0}^{m_1-1}[\Phi_{m_1}(u)]_{j,j'}[H_{\mathbf{m}}]_{j',k'}\right)[\widehat{\Psi}_{m_2}(u)]_{k',k}\right]\right)\right\}du \\
&= \frac{1}{T}\int_0^T\sum_{k=0}^{m_2-1}\left[H_{\mathbf{m}}^\top\Phi_{m_1}(u)H_{\mathbf{m}}\widehat{\Psi}_{m_2}(u)\right]_{k,k}du = \frac{1}{T}\int_0^T\text{Tr}\left[H_{\mathbf{m}}^\top\Phi_{m_1}(u)H_{\mathbf{m}}\widehat{\Psi}_{m_2}(u)\right]du,
\end{aligned}$$

which is the first equality of the Lemma. We use equality (40) which yields

$$\text{Tr}\left[H_{\mathbf{m}}^\top\Phi_{m_1}(u)H_{\mathbf{m}}\widehat{\Psi}_{m_2}(u)\right] = \text{vec}(H_{\mathbf{m}})^\top\text{vec}(\Phi_{m_1}(u)H_{\mathbf{m}}\widehat{\Psi}_{m_2}(u)).$$

Now with (38), we obtain :

$$\text{vec}(\Phi_{m_1}(u)H_{\mathbf{m}}\widehat{\Psi}_{m_2}(u)) = (\widehat{\Psi}_{m_2}(u)\otimes\Phi_{m_1}(u))\text{vec}(H_{\mathbf{m}}).$$

Integrating wrt u gives

$$\|h\|_N^2 = [\text{vec}(H_{\mathbf{m}})]^\top\,\widehat{\Theta}_{m_1,m_2}\text{vec}(H_{\mathbf{m}}).$$

The last equality is obtained by taking expectation. □

Proof of Theorem 1. Let us define

$$\Omega_{\mathbf{m}} := \left\{ \left| \frac{\|h\|_N^2}{\|h\|_p^2} - 1 \right| \leq \frac{1}{2}, \forall h \in S_{m_1} \times \Sigma_{m_2} \right\}. \tag{27}$$

On $\Omega_{\mathbf{m}}$, the empirical norm $\|.\|_N$ and the p-norm are equivalent for elements of $S_{m_1} \times \Sigma_{m_2}$: $(2/3)\|h\|_N^2 \leq \|h\|_p^2 \leq 2\|h\|_N^2$ and the following result holds.

Lemma 3. *We have*

$$\Omega_{\mathbf{m}} = \left\{ \left\| \Theta_{m_1,m_2}^{-1/2} \widehat{\Theta}_{m_1,m_2} \Theta_{m_1,m_2}^{-1/2} - \mathrm{Id}_{m_1 m_2} \right\|_{\mathrm{op}} \leq \frac{1}{2} \right\}, \tag{28}$$

where Id_K *is the* $K \times K$ *identity matrix.*

Moreover, under $m_1 m_2 \leq NT$ *and under (20), it holds that* $\mathbb{P}(\Omega_{\mathbf{m}}^c) \leq C/(NT)^r$ *and* $\mathbb{P}(\widehat{\Lambda}_{\mathbf{m}}^c) \leq C/(NT)^r$.

Proof of Lemma 3. Let $h = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} h_{j,k} \varphi_j \odot \psi_k$ and denote by $H_{\mathbf{m}} = (h_{j,k})_{0 \leq j \leq m_1 - 1, 0 \leq k \leq m_2 - 1}$. We have

$$\|h\|_N^2 = \mathrm{vec}(H_{\mathbf{m}})^\top \widehat{\Theta}_{m_1,m_2} \mathrm{vec}(H_{\mathbf{m}}) \quad \text{and} \quad \|h\|_p^2 = \mathrm{vec}(H_{\mathbf{m}})^\top \Theta_{m_1,m_2} \mathrm{vec}(H_{\mathbf{m}}) \quad \text{so that}$$

$$\begin{aligned}
\sup_{h \in S_{m_1} \times \Sigma_{m_2}, \|h\|_p = 1} \left| \|h\|_N^2 - \|h\|_p^2 \right| &= \sup_{\substack{\mathrm{vec}(H_{\mathbf{m}}) \in \mathbb{R}^{m_1 m_2}, \\ \|\Theta_{m_1,m_2}^{1/2} \mathrm{vec}(H_{\mathbf{m}})\|_{2,m_1 m_2} = 1}} \left| \mathrm{vec}(H_{\mathbf{m}})^\top (\widehat{\Theta}_{m_1,m_2} - \Theta_{m_1,m_2}) \mathrm{vec}(H_{\mathbf{m}}) \right| \\
&= \sup_{u \in \mathbb{R}^{m_1 m_2}, \|u\|_{2,m_1 m_2} = 1} \left| u^\top \Theta_{m_1,m_2}^{-1/2} (\widehat{\Theta}_{m_1,m_2} - \Theta_{m_1,m_2}) \Theta_{m_1,m_2}^{-1/2} u \right| = \|\Theta_{m_1,m_2}^{-1/2} \widehat{\Theta}_{m_1,m_2} \Theta_{m_1,m_2}^{-1/2} - \mathrm{Id}_{m_1 m_2}\|_{\mathrm{op}}.
\end{aligned}$$

Therefore, (28) holds.

Define $\boldsymbol{\psi}(x) = (\psi_0(x), \ldots, \psi_{m_2-1}(x))^\top$ and $S_\psi(x) = \boldsymbol{\psi}(x)\boldsymbol{\psi}(x)^\top$ so that

$$\widehat{\Psi}_{m_2}(t) = \frac{1}{N} \sum_{i=1}^{N} S_\psi(X_i(t)).$$

We intend to apply Tropp's Inequality [44] (see Theorem 4 in Appendix) to $G - \mathrm{Id}_{m_1 m_2}$ where

$$G = \frac{1}{N} \sum_{i=1}^{N} \mathbb{X}_i, \quad \mathbb{X}_i = \Theta_{m_1,m_2}^{-1/2} \left(\frac{1}{T} \int_0^T S_\psi(X_i(t)) \otimes \Phi_{m_1}(t) dt \right) \Theta_{m_1,m_2}^{-1/2}.$$

Note that $G = \Theta_{m_1,m_2}^{-1/2} \widehat{\Theta}_{m_1,m_2} \Theta_{m_1,m_2}^{-1/2}$ and $\mathbb{E}(G) = \mathrm{Id}_{m_1 m_2}$. Therefore

$$\mathbb{P}(\Omega_{\mathbf{m}}^c) = \mathbb{P}(\|G - \mathrm{Id}_{m_1 m_2}\|_{\mathrm{op}} > 1/2).$$

The matrices $\mathbb{X}_i$, $i = 1, \ldots, N$ are i.i.d. and symmetric nonnegative with $\lambda_{\min}(\mathbb{E}(G)) = \lambda_{\max}(\mathbb{E}(G)) = 1$, thus if $\lambda_{\max}(\mathbb{X}_i) \leq R$, then by Tropp's Inequality

$$\mathbb{P}(\|G - \mathrm{Id}_{m_1 m_2}\|_{\mathrm{op}} > \delta) \leq 2 m_1 m_2 e^{-Nc(\delta)/R}, \quad c(\delta) = (1+\delta)\log(1+\delta) - \delta,$$

as $e^\delta/(1+\delta)^{1+\delta} \geq e^{-\delta}/(1-\delta)^{1-\delta}$, see Cohen *et al.* [7, 8].

It remains to compute the bound R.

$$\lambda_{\max}(\mathbb{X}_i) = \|\mathbb{X}_i\|_{\mathrm{op}} = \sup_{x \in \mathbb{R}^{m_1 m_2}, \|x\|_{2,m_1 m_2} = 1} x^\top \mathbb{X}_i x.$$

$$
\begin{aligned}
x^\top \mathbb{X}_i x &= y^\top \left(\frac{1}{T} \int_0^T S_\psi(X_i(t)) \otimes \Phi_{m_1}(t) dt \right) y \qquad \left(y = \Theta_{m_1,m_2}^{-1/2} x, \ (y = \mathrm{vec}(Y) \right) \\
&= \frac{1}{T} \int_0^T \mathrm{Tr}\left[Y^\top \Phi_{m_1}(t) Y S_\psi(X_i(t)) \right] dt = \frac{1}{T} \int_0^T \sum_{0 \le j,j' \le m_1-1} \sum_{0 \le k,k' \le m_2-1} Y_{j,k} Y_{j',k'} \varphi_j(t) \varphi_{j'}(t) \psi_k(X_i(t)) \psi_{k'}(X_i(t)) dt \\
&= \frac{1}{T} \int_0^T \left(\sum_{k=0}^{m_2-1} \psi_k(X_i(t)) \left(\sum_{j=0}^{m_1-1} Y_{j,k} \varphi_j(t) \right) \right)^2 dt \le \frac{1}{T} \int_0^T \left(\sum_{k=0}^{m_2-1} \psi_k^2(X_i(t)) \right) \sum_{k=0}^{m_2-1} \left[\left(\sum_{j=0}^{m_1-1} Y_{j,k} \varphi_j(t) \right)^2 \right] dt.
\end{aligned}
$$

Therefore

$$
x^\top \mathbb{X}_i x \le L_\psi(m_2) \sum_{k=0}^{m_2-1} \frac{1}{T} \int_0^T \left(\sum_{j=0}^{m_1-1} Y_{j,k} \varphi_j(t) \right)^2 dt = L_\psi(m_2) \|y\|_{2,m_1 m_2}^2 / T.
$$

Now it holds that $\|y\|_{2,m_1m_2}^2 = \|\Theta_{m_1,m_2}^{-1/2} x\|_{2,m_1m_2}^2 \le \|\Theta_{m_1,m_2}^{-1}\|_{\mathrm{op}} \|x\|_{2,m_1m_2}^2$. Consequently

$$
\lambda_{\max}(\mathbb{X}_i) \le L_\psi(m_2) \|\Theta_{m_1,m_2}^{-1}\|_{\mathrm{op}} / T := R.
$$

We obtain

$$
\mathbb{P}(\|G - \mathrm{Id}_{m_1 m_2}\|_{\mathrm{op}} > 1/2) \le 2 m_1 m_2 \exp\left(-\frac{N c(1/2) T}{L_\psi(m_2) \|\Theta_{m_1,m_2}^{-1}\|_{\mathrm{op}}} \right).
$$

For the proof of $\mathbb{P}(\widehat{\Lambda}_{\mathbf{m}}^c) \le C/(NT)^r$, we refer to the proof of Lemma 5 in Comte and Genon-Catalot [9]. □

Using equation (1), we split $\widehat{C}_{\mathbf{m}}$ into the sum of two random matrices:

$$
\widehat{C}_{\mathbf{m}} = C_{\mathbf{m}} + \varepsilon_{\mathbf{m}}, \quad C_{\mathbf{m}} := \left(\langle \varphi_j \odot \psi_k, b \rangle_N \right)_{0 \le j \le m_1-1, 0 \le k \le m_2-1} \tag{29}
$$

and set

$$
\varepsilon_{\mathbf{m}} := \left(\nu_N(\varphi_j \odot \psi_k) \right)_{0 \le j \le m_1-1, 0 \le k \le m_2-1}. \tag{30}
$$

To study the risk of the estimator defined by $\mathbb{E}(\|\widetilde{b}_{\mathbf{m}} - b_{\mathbf{I}}\|_N^2)$,we write

$$
\|\widetilde{b}_{\mathbf{m}} - b_{\mathbf{I}}\|_N^2 = \|\widehat{b}_{\mathbf{m}} - b_{\mathbf{I}}\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}} + \|b_{\mathbf{I}}\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}^c}.
$$

We define the orthogonal projection of $b_{\mathbf{I}}$ on $S_{m_1} \times \Sigma_{m_2}$ wrt the empirical scalar product, denoted by $\Pi_{\mathbf{m}} b$. We find for

$$
\Pi_{\mathbf{m}} b = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} a_{j,k} \varphi_j \odot \psi_k, \quad A_{\mathbf{m}} = (a_{j,k})_{0 \le j \le m_1-1, 0 \le k \le m_2-1}
$$

with

$$
\mathrm{vec}(A_{\mathbf{m}}) = \widehat{\Theta}_{m_1,m_2}^{-1} \mathrm{vec}(C_{\mathbf{m}})
$$

where $C_{\mathbf{m}}$ defined by (29). Then we have by Pythagoras theorem

$$
\|\widehat{b}_{\mathbf{m}} - b_{\mathbf{I}}\|_N^2 = \|\widehat{b}_{\mathbf{m}} - \Pi_{\mathbf{m}} b\|_N^2 + \|\Pi_{\mathbf{m}} b - b_{\mathbf{I}}\|_N^2.
$$

Thus

$$
\|\widetilde{b}_{\mathbf{m}} - b_{\mathbf{I}}\|_N^2 = \left(\|\widehat{b}_{\mathbf{m}} - \Pi_{\mathbf{m}} b\|_N^2 + \|\Pi_{\mathbf{m}} b - b_{\mathbf{I}}\|_N^2 \right) \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}} + \|b_{\mathbf{I}}\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}^c}.
$$

Then we have

$$
\begin{aligned}
\mathbb{E}(\|\widetilde{b}_{\mathbf{m}} - b_{\mathbf{I}}\|_N^2) &= \mathbb{E}\left(\|\Pi_{\mathbf{m}} b - b_{\mathbf{I}}\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}} \right) + \mathbb{E}\left(\|\widehat{b}_{\mathbf{m}} - \Pi_{\mathbf{m}} b\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}} \cap \Omega_{\mathbf{m}}} \right) + \mathbb{E}\left(\|\widehat{b}_{\mathbf{m}} - \Pi_{\mathbf{m}} b\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}} \cap \Omega_{\mathbf{m}}^c} \right) + \mathbb{E}(\|b_{\mathbf{I}}\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}^c}) \\
&:= T_1 + T_2 + T_3 + T_4.
\end{aligned} \tag{31}
$$

For the bias, we have

$$T_1 = \mathbb{E}(\|\Pi_{\mathbf{m}} b - b_{\mathbf{I}}\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}}) = \mathbb{E}\left(\inf_{h\in S_{m_1}\times\Sigma_{m_2}} \|h - b_{\mathbf{I}}\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}}\right) \leq \inf_{h\in S_{m_1}\times\Sigma_{m_2}} \|h - b_{\mathbf{I}}\|_p^2. \tag{32}$$

By Lemma 1,

$$\|\widehat{b}_{\mathbf{m}} - \Pi_{\mathbf{m}} b\|_N^2 = \frac{1}{T}\int_0^T \mathrm{Tr}\left[(\widehat{A}_{\mathbf{m}} - A_{\mathbf{m}})^\top \Phi_{m_1}(t)(\widehat{A}_{\mathbf{m}} - A_{\mathbf{m}})\widehat{\Psi}_{m_2}(t)\right] dt.$$

Now we use (13) and its analogous for $A_{\mathbf{m}}$ and we get

$$\begin{aligned}
\|\widehat{b}_{\mathbf{m}} - \Pi_{\mathbf{m}} b\|_N^2 &= \mathrm{Tr}\left[(\widehat{A}_{\mathbf{m}} - A_{\mathbf{m}})^\top \frac{1}{T}\int_0^T \Phi_{m_1}(t)(\widehat{A}_{\mathbf{m}} - A_{\mathbf{m}})\widehat{\Psi}_{m_2}(t) dt\right] = \mathrm{Tr}\left[(\widehat{A}_{\mathbf{m}} - A_{\mathbf{m}})^\top(\widehat{C}_{\mathbf{m}} - C_{\mathbf{m}})\right] \\
&= \mathrm{vec}(\widehat{A}_{\mathbf{m}} - A_{\mathbf{m}})^\top \mathrm{vec}(\widehat{C}_{\mathbf{m}} - C_{\mathbf{m}}) \text{ with } \mathrm{Tr}(M^\top N) = \mathrm{vec}(M)^\top \mathrm{vec}(N), \\
&= \mathrm{vec}(\widehat{C}_{\mathbf{m}} - C_{\mathbf{m}})^\top \widehat{\Theta}_{m_1,m_2}^{-1} \mathrm{vec}(\widehat{C}_{\mathbf{m}} - C_{\mathbf{m}}) = \mathrm{vec}(\varepsilon_{\mathbf{m}})^\top \widehat{\Theta}_{m_1,m_2}^{-1} \mathrm{vec}(\varepsilon_{\mathbf{m}}).
\end{aligned} \tag{33}$$

On $\Omega_{\mathbf{m}}$, the eigenvalues of $\Theta_{m_1,m_2}^{-1/2}\widehat{\Theta}_{m_1,m_2}\Theta_{m_1,m_2}^{-1/2}$ all belong to $[1/2, 3/2]$. Therefore, the eigenvalues of $\Theta_{m_1,m_2}^{1/2}\widehat{\Theta}_{m_1,m_2}^{-1}\Theta_{m_1,m_2}^{1/2}$ all belong to $[2/3, 2]$. So we write

$$\mathrm{vec}(\varepsilon_{\mathbf{m}})^\top \widehat{\Theta}_{m_1,m_2}^{-1} \mathrm{vec}(\varepsilon_{\mathbf{m}}) = \mathrm{vec}(\varepsilon_{\mathbf{m}})^\top \Theta_{m_1,m_2}^{-1/2}\Theta_{m_1,m_2}^{1/2}\widehat{\Theta}_{m_1,m_2}^{-1}\Theta_{m_1,m_2}^{1/2}\Theta_{m_1,m_2}^{-1/2} \mathrm{vec}(\varepsilon_{\mathbf{m}}).$$

This yields

$$\mathbb{E}(\|\widehat{b}_{\mathbf{m}} - \Pi_{\mathbf{m}} b\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}\cap\Omega_{\mathbf{m}}}) \leq 2\mathbb{E}(\mathrm{vec}(\varepsilon_{\mathbf{m}})^\top \Theta_{m_1,m_2}^{-1} \mathrm{vec}(\varepsilon_{\mathbf{m}})).$$

Now,

$$\begin{aligned}
\mathbb{E}(\mathrm{vec}(\varepsilon_{\mathbf{m}})^\top \Theta_{m_1,m_2}^{-1} \mathrm{vec}(\varepsilon_{\mathbf{m}})) &= \mathbb{E}\left[\mathrm{Tr}\left(\mathrm{vec}(\varepsilon_{\mathbf{m}})^\top \Theta_{m_1,m_2}^{-1} \mathrm{vec}(\varepsilon_{\mathbf{m}})\right)\right] = \mathbb{E}\left[\mathrm{Tr}\left(\Theta_{m_1,m_2}^{-1} \mathrm{vec}(\varepsilon_{\mathbf{m}})\mathrm{vec}(\varepsilon_{\mathbf{m}})^\top\right)\right] \\
&= \mathrm{Tr}\left(\Theta_{m_1,m_2}^{-1}\mathbb{E}\left[\mathrm{vec}(\varepsilon_{\mathbf{m}})\mathrm{vec}(\varepsilon_{\mathbf{m}})^\top\right]\right) = \frac{1}{NT}\mathrm{Tr}\left(\Theta_{m_1,m_2}^{-1}\Theta_{m_1,m_2,\sigma^2}\right).
\end{aligned}$$

Thus, we get

$$T_2 = \mathbb{E}(\|\widehat{b}_{\mathbf{m}} - \Pi_{\mathbf{m}} b\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}\cap\Omega_{\mathbf{m}}}) \leq \frac{2}{NT}\mathrm{Tr}\left(\Theta_{m_1,m_2}^{-1}\Theta_{m_1,m_2,\sigma^2}\right). \tag{34}$$

Now, using (33) and the definition of $\widehat{\Lambda}_{\mathbf{m}}$, we have:

$$\begin{aligned}
T_3 = \mathbb{E}(\|\widehat{b}_{\mathbf{m}} - \Pi_{\mathbf{m}} b\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}\cap\Omega_{\mathbf{m}}^c}) &\leq 2\mathfrak{c}_r \frac{NT}{L_\psi(m_2)\log(NT)}\mathbb{E}\left[\mathrm{vec}(\varepsilon_{\mathbf{m}})^\top \mathrm{vec}(\varepsilon_{\mathbf{m}})\mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}\cap\Omega_{\mathbf{m}}^c})\right] \\
&\leq 2\mathfrak{c}_r \frac{NT}{L_\psi(m_2)\log(NT)}\mathbb{E}^{1/2}\left[\left(\mathrm{vec}(\varepsilon_{\mathbf{m}})^\top \mathrm{vec}(\varepsilon_{\mathbf{m}})\right)^2\right]\mathbb{P}^{1/2}(\Omega_{\mathbf{m}}^c).
\end{aligned}$$

Lemma 4. *Under the assumptions of Theorem 1, we have*

$$\mathbb{E}\left[\left(\mathrm{vec}(\varepsilon_{\mathbf{m}})^\top \mathrm{vec}(\varepsilon_{\mathbf{m}})\right)^2\right] \lesssim \frac{m_1 m_2}{N^2T^2}(L_\varphi(m_1)L_\psi(m_2))^2\left(\frac{1}{T}\int_0^T\int \sigma^4(u,x)p_u(x)dx\,du\right).$$

Applying Lemma 4, we get for $r > 7$, that

$$T_3 = \mathbb{E}(\|\widehat{b}_{\mathbf{m}} - \Pi_{\mathbf{m}} b\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}^c\cap\Omega_{\mathbf{m}}^c}) \lesssim \frac{1}{N}, \tag{35}$$

using $m_1 m_2 \leq NT$, $L_\varphi(m_1) \leq NT$ and $\mathbb{P}^{1/2}(\Omega_{\mathbf{m}}^c) \leq C/N^r$.

Lastly, we notice that

$$\mathbb{E}\left(\|b_{\mathbf{I}}\|_N^4\right) \leq \frac{C_T^4}{T}\mathbb{E}\int_0^T (1 + |X_i(u)|)^4 du$$

and we get, using Lemma 3, that

$$T_4 = \mathbb{E}(\|b_{\mathbf{I}}\|_N^2 \mathbf{1}_{\widehat{\Lambda}_{\mathbf{m}}^c}) \le C_T^2 \left(\frac{1}{T}\mathbb{E}\int_0^T (1+|X_i(u)|)^4 du\right)^{1/2} \mathbb{P}^{1/2}(\widehat{\Lambda}_{\mathbf{m}}^c) \lesssim \frac{1}{N^{r/2}}.$$

This means $T_4 = O(1/N)$ for $r \ge 2$. Therefore, plugging this and (32)-(34)-(35) into (31) gives Inequality (23) of Theorem 1.

Now, we use that σ is uniformly bounded on $[0,T]\times\mathbb{R}$. We exploit the following trick.

$$\begin{aligned}\mathrm{Tr}\left(\Theta_{m_1,m_2}^{-1}\Theta_{m_1,m_2,\sigma^2}\right) &= \mathrm{Tr}\left(\Theta_{m_1,m_2}^{-1/2}\Theta_{m_1,m_2,\sigma^2}\Theta_{m_1,m_2}^{-1/2}\right)\\ &= \mathbb{E}\left[\mathrm{vec}(Z)^\top\Theta_{m_1,m_2}^{-1/2}\Theta_{m_1,m_2,\sigma^2}\Theta_{m_1,m_2}^{-1/2}\mathrm{vec}(Z)\right]\end{aligned}$$

where $Z = (Z_{i,j})$ is a $m_1 \times m_2$-matrix with i.i.d. entries $Z_{i,j}$ such that $\mathbb{E}(Z_{i,j}) = 0$ and $\mathbb{E}(Z_{i,j}^2) = 1$. Let $Y = (Y_{i,j})$ be a $m_1 \times m_2$-matrix with i.i.d. entries $Y_{i,j}$ such that $\mathrm{vec}(Y)^\top = \mathrm{vec}(Z)^\top\Theta_{m_1,m_2}^{-1/2}$, and let us look at $\mathrm{vec}(Y)^\top\Theta_{m_1,m_2,\sigma^2}\mathrm{vec}(Y)$. We have,

$$\begin{aligned}\mathrm{vec}(Y)^\top\Theta_{m_1,m_2,\sigma^2}\mathrm{vec}(Y) &= \frac{1}{T}\sum_{j,k,j',k'} Y_{j,k}Y_{j',k'}\int_0^T du\int \varphi_j(u)\varphi_{j'}(u)\psi_k(x)\psi_{k'}(x)p_u(x)\sigma^2(u,x)dx\\ &= \frac{1}{T}\int_0^T du\int\left(\sum_{j,k} Y_{j,k}\varphi_j(u)\psi_k(x)\right)^2 p_u(x)\sigma^2(u,x)dx \le \sigma_1^2\mathrm{vec}(Y)^\top\Theta_{m_1,m_2}\mathrm{vec}(Y).\end{aligned}$$

This yields

$$\begin{aligned}\mathbb{E}\left[\mathrm{vec}(Z)^\top\Theta_{m_1,m_2}^{-1/2}\Theta_{m_1,m_2,\sigma^2}\Theta_{m_1,m_2}^{-1/2}\mathrm{vec}(Z)\right] &\le \sigma_1^2\mathbb{E}\left[\mathrm{vec}(Z)^\top\Theta_{m_1,m_2}^{-1/2}\Theta_{m_1,m_2,}\Theta_{m_1,m_2}^{-1/2}\mathrm{vec}(Z)\right]\\ &= \sigma_1^2\mathrm{Tr}(\Theta_{m_1,m_2}^{-1/2}\Theta_{m_1,m_2}\Theta_{m_1,m_2}^{-1/2}) = \sigma_1^2 m_1 m_2.\end{aligned}$$

This ends the proof of Theorem 1. □

Proof of Lemma 4 First we have that

$$\mathrm{vec}(\varepsilon_{\mathbf{m}})^\top\mathrm{vec}(\varepsilon_{\mathbf{m}}) = \sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1}\nu_N^2(\varphi_j\odot\psi_k).$$

Thus

$$\mathbb{E}\left[\left(\mathrm{vec}(\varepsilon_{\mathbf{m}})^\top\mathrm{vec}(\varepsilon_{\mathbf{m}})\right)^2\right] \le m_1m_2\sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1}\mathbb{E}\left(\nu_N^4(\varphi_j\odot\psi_k)\right).$$

Now using the Burkholder-Davies-Gundy inequality, we get:

$$\begin{aligned}\mathbb{E}\left(\nu_N^4(\varphi_j\odot\psi_k)\right) &= \frac{1}{(NT)^4}\mathbb{E}\left[\left(\int_0^T\sum_{i=1}^N \varphi_j\odot\psi_k(u,X_i(u))\sigma(u,X_i(u))dW_i(u)\right)^4\right]\\ &\lesssim \frac{1}{(NT)^4}\mathbb{E}\left[\left(\int_0^T\sum_{i=1}^N\left(\varphi_j\odot\psi_k(u,X_i(u))\right)^2\sigma^2(u,X_i(u))du\right)^2\right]\\ &\lesssim \frac{1}{N^2T^2}\frac{1}{T}\int_0^T\mathbb{E}\left[\left(\varphi_j\odot\psi_k(u,X_1(u))\right)^4\sigma^4(u,X_1(u))du\right].\end{aligned}$$

Therefore as $\sum_{j=0}^{m_1-1}\varphi_j^4(u) \le (L_\varphi(m_1))^2$ and $\sum_{k=0}^{m_2-1}\psi_k^4(x) \le (L_\psi(m_2))^2$

$$\mathbb{E}\left[\left(\mathrm{vec}(\varepsilon_{\mathbf{m}})^\top\mathrm{vec}(\varepsilon_{\mathbf{m}})\right)^2\right] \le \frac{m_1m_2(L_\varphi(m_1)L_\psi(m_2))^2}{(NT)^2}\frac{1}{T}\mathbb{E}\left[\int_0^T\sigma^4(u,X_1(u))du\right].$$

This is the announced result. □

Proof of Lemma 2. Let $H_{\mathbf{m}} = (h_{j,k})_{0\le j\le m_1-1, 0\le k\le m_2-1}$, then as by Lemma 1

$$\begin{aligned}[\mathrm{vec}(H_{\mathbf{m}})]^\top \Theta_{m_1,m_2}\mathrm{vec}(H_{\mathbf{m}}) &= \frac{1}{T}\int_\epsilon^T\int_A h^2(t,u)p_t(u)dudt \\ &\ge c_0(\epsilon,A)\frac{1}{T}\int_\epsilon^T\int_A h^2(t,u)dudt = c_0(\epsilon,A)[\mathrm{vec}(H_{\mathbf{m}})]^\top\mathrm{vec}(H_{\mathbf{m}})\end{aligned}$$

where $h(t,u)=\sum_{j,k} h_{j,k}\varphi_j(t)\psi_k(u)$. This proves that any eigenvalue of Θ_{m_1,m_2} is larger than $c_0(\epsilon,A)$. The upper bound for $\|h\|_p^2$ is straightforward. This gives the result of Lemma 2. □

Proof of Proposition 2. In the bound (26), we look at the bias term:

$$\inf_{h\in S_{m_1}\times\Sigma_{m_2}}\|h-b_{\mathbf{I}_\epsilon}\|^2 = \|b_{\mathbf{m}}-b_{\mathbf{I}_\epsilon}\|^2$$

where $b_{\mathbf{m}} = \sum_{j=0}^{m_1-1}\sum_{k=0}^{m_2-1}\langle b,\varphi_j\odot\psi_k\rangle\varphi_j\odot\psi_k$ is the $\mathbb{L}^2$-orthogonal projection of $b_{\mathbf{I}_\epsilon}$ on $S_{m_1}\times\Sigma_{m_2}$. Therefore, if $b_{\mathbf{I}_\epsilon}\in W^*(\boldsymbol{\beta},R)$,

$$\|b_{\mathbf{m}}-b_{\mathbf{I}_\epsilon}\|^2 \le R_1^2 m_1^{-2\beta_1} + R_2^2 m_2^{-2\beta_2}.$$

Making the standard compromise with the variance term of order $m_1m_2/(NT)$ gives the rate. □

Proof of Theorem 3. We follow the scheme of Theorem 2.11 in Tsybakov [45]. Take g and h two regular functions with support $[0,1]$, bounded by K_f and K_g respectively, with g β_1-times derivable and h β_2 times derivable, with square integrable derivatives. Define for $j\in\{0,\dots,M_1-1\}$, and $k\in\{0,\dots,M_2-1\}$, with $A=[a,b]$, $a<b$,

$$g_j(t) = \sqrt{\frac{M_1}{T-\epsilon}}\,g\Big(M_1(\frac{t-\epsilon}{T-\epsilon})-j\Big),\quad h_k(x) = \sqrt{\frac{M_2}{b-a}}\,h\Big(M_2(\frac{x-a}{b-a})-k\Big).$$

Clearly, the g_j have disjoint supports and $g_jg_{j'}=0$ for $j\ne j'$, and for the same reason $h_kh_{k'}=0$ for $k\ne k'$. Denote by $I_j := [\epsilon+j(T-\epsilon)/M_1, \epsilon+(j+1)(T-\epsilon)/M_1]$ and $J_k=[a+k(b-a)/M_2, a+(k+1)(b-a)/M_2]$ the respective supports of g_j, h_k. Let us define proposals: $b_0(t,x)=0$ and for $\theta=(\theta_{j,k})_{j\in\{0,\dots,M_1-1\},k\in\{0,\dots,M_2-1\}}$ with $\theta_{j,k}\in\{0,1\}$ for all $j\in\{0,\dots,M_1-1\}, k\in\{0,\dots,M_2-1\}$,

$$b_\theta(t,x) = \frac{\delta}{\sqrt{NT}}\sum_{j=0}^{M_1-1}\sum_{k=0}^{M_2-1}\theta_{j,k}g_j\odot h_k(t,x).$$

We choose $M_1=(NT)^{\beta_2/(\beta_1+\beta_2+2\beta_1\beta_2)}$, $\quad M_2=(NT)^{\beta_1/(\beta_1+\beta_2+2\beta_1\beta_2)}$.

- As g is β_1-times derivable and h is β_2 times derivable, both with square integrable derivatives, we get that

$$\begin{aligned}&\iint(\partial^{\beta_1}b_\theta(t,x)/\partial t^{\beta_1})^2dtdx + (\partial^{\beta_2}b_\theta(t,x)/\partial x^{\beta_2})^2dtdx \\ =\ & \frac{\delta^2}{NT}\sum_{j=0}^{M_1-1}\sum_{k=0}^{M_2-1}\theta_{j,k}^2\left\{\Big(\frac{M_1}{T-\epsilon}\Big)^{2\beta_1}\int[g^{(\beta_1)}(t)]^2dt\int h_k^2(x)dx + \Big(\frac{M_2}{b-a}\Big)^{2\beta_2}\int g_j^2(t)dt\int[h^{(\beta_2)}(x)]^2dx\right\} \\ \le\ & \frac{\delta^2}{NT}\left[\frac{\int[g^{(\beta_1)}(t)]^2dt\int h^2(x)dx}{(T-\epsilon)^{2\beta_1}}M_1^{2\beta_1+1}M_2 + \frac{\int g^2(t)dt\int[h^{(\beta_2)}(x)]^2dx}{(b-a)^{2\beta_2}}M_1M_2^{2\beta_2+1}\right].\end{aligned}$$

As $M_1^{2\beta_1+1}M_2 = M_1M_2^{2\beta_2+1}=NT$, we obtain that

$$\iint(\partial^{\beta_1}b_\theta(t,x)/\partial t^{\beta_1})^2dtdx + (\partial^{\beta_2}b_\theta(t,x)/\partial x^{\beta_2})^2dtdx \le C^2\delta^2 \le L^2$$

for $\delta \le L/C$ small enough, where

$$C = \frac{\int [g^{(\beta_1)}(t)]^2 dt \int h^2(x)dx}{(T-\epsilon)^{2\beta_1}} + \frac{\int g^2(t)dt \int [h^{(\beta_2)}(x)]^2 dx}{(b-a)^{2\beta_2}}.$$

This implies that b_0 and b_θ belong to $W(\boldsymbol{\beta}, \mathbf{L})$ with $\mathbf{L} = (L, L)$.

• We have that:

$$\|b_\theta - b_{\theta'}\|^2 = \frac{\delta^2}{NT} \sum_{j=0}^{M_1-1} \sum_{k=0}^{M_2-1} (\theta_{j,k} - \theta'_{j,k})^2 \int g_j^2(t) h_k^2(x) dt dx = \frac{\delta^2}{NT} \rho(\theta, \theta') \int g^2(t)dt \int h^2(x)dx$$

where $\rho(\theta, \theta') = \sum_{j,k} \mathbf{1}_{(\theta_{j,k} \neq \theta'_{j,k})}$ is the Hamming distance between θ and θ'.

As a consequence, the Varshamov-Gilbert Lemma (see Lemma 2.9 in Tsybakov [45]) ensures that for $M := M_1 M_2 \ge 8$, there exist $Q \ge 2^{M/8}$ elements say $\{\theta^0, \ldots, \theta^Q\}$ of $\{0,1\}^M$ such that $\rho(\theta^q, \theta^{q'}) \ge M/8$ for all $0 \le q < q' \le Q$, with $\theta^0 = (0, \ldots, 0)$. This leads to:

$$\|b_{\theta^q} - b_{\theta^{q'}}\|^2 \ge \frac{\delta^2}{NT} \frac{M}{8} \|g\|^2 \|h\|^2 = \delta^2 \|g\|^2 \|h\|^2 N^{-\frac{2\beta}{2\beta+2}}. \tag{36}$$

• Lastly, let $\mathbb{P}_\theta$ (resp $\mathbb{P}_0$) denotes the distribution of the process (1) when the drift is equal to $b_\theta(t,x)$ (resp. is equal to 0) and the diffusion coefficient to $\sigma(t,x)$ on the space $C_T = C([0,T])$ of real valued continuous functions on $[0,T]$ endowed with the canonical σ-field $C_T = \sigma(X(t), t \in [0,T])$ where $X(t), t \in [0,T]$ is the canonical process of C_T, *i.e.*, $X_t(x) = x(t)$ for $x \in C_T$. We bound

$$K(\mathbb{P}_\theta^{\otimes N}, \mathbb{P}_0^{\otimes N}) = NK(\mathbb{P}_\theta, \mathbb{P}_0).$$

where $K(P,Q) = \mathbb{E}_P(\log \frac{dP}{dQ})$ is the Kullback-Leibler divergence of P with respect to Q. Under [H1], $\mathbb{P}_\theta$ and $\mathbb{P}_0$ are equivalent and

$$\log \frac{d\mathbb{P}_\theta}{d\mathbb{P}_0} = \int_0^T \frac{b_\theta(t,X(t))}{\sigma^2(t,X(t))} dX(t) - \frac{1}{2} \int_0^T \frac{b_\theta^2(t,X(t))}{\sigma^2(t,X(t))} dt.$$

(see *e.g.* Liptser and Shiryaev [35]). Under $\mathbb{P}_\theta$, $dX(t) = b_\theta(t,X(t))dt + \sigma(t,X(t))dB(t)$ where $(B(t), t \in [0,T])$ is a standard Brownian motion. Therefore,

$$\log \frac{d\mathbb{P}_\theta}{d\mathbb{P}_0} = \frac{1}{2} \int_0^T \frac{b_\theta^2(t,X(t))}{\sigma^2(t,X(t))} dt + \int_0^T \frac{b_\theta(t,X(t))}{\sigma(t,X(t))} dB(t).$$

Thus,

$$K(\mathbb{P}_\theta, \mathbb{P}_0) = \mathbb{E}_{\mathbb{P}_\theta} \frac{1}{2} \int_0^T \frac{b_\theta^2(t,X(t))}{\sigma^2(t,X(t))} dt \le \frac{1}{2\sigma_0^2} \int_\epsilon^T \int_{\mathbb{R}} b_\theta^2(t,x) p_t^\theta(x) dx dt.$$

Then, we have

$$\begin{aligned}
\int_\epsilon^T \int_{\mathbb{R}} b_\theta^2(t,x) p_t^\theta(x) dx dt &= \frac{\delta^2}{NT} \int_\epsilon^T \int_{\mathbb{R}} \sum_{j=0}^{M_1-1} \sum_{k=1}^{M_2-1} \theta_{j,k}^2 g_j^2(t) h_k^2(x) p_t^\theta(x) dx dt \\
&= \frac{\delta^2 M_1 M_2}{NT} \frac{1}{(T-\epsilon)(b-a)} \sum_{j,k} \theta_{j,k}^2 \int_{I_j} \int_{J_k} g^2\left(M_1(\frac{t-\epsilon}{T-\epsilon}) - j\right) h^2\left(M_2(\frac{x-a}{b-a}) - k\right) p_t^\theta(x) dx dt \\
&\le \frac{\delta^2 M_1 M_2}{NT} \frac{K_f^2 K_g^2}{(T-\epsilon)(b-a)} \int_\epsilon^T \int_a^b p_t^\theta(x) dx dt \le \frac{\delta^2 M_1 M_2}{N} \frac{K_f^2 K_g^2}{(T-\epsilon)(b-a)},
\end{aligned}$$

as $\int p_t^\theta(x)dx = 1$ and $(T-\epsilon)/T \le 1$. As a consequence, we get

$$K(\mathbb{P}_\theta, \mathbb{P}_0) \le C\delta^2 \frac{M_1 M_2}{N}.$$

Therefore

$$K(\mathbb{P}_{\theta}^{\otimes N}, \mathbb{P}_{0}^{\otimes N}) \leq C\delta^2 M_1 M_2 \leq \frac{8C\delta^2}{\log(2)} \log(Q).$$

By choosing δ small enough, we obtain, for $\kappa \in (0, \frac{1}{8})$,

$$\frac{1}{Q}\sum_{q=1}^{Q} K(\mathbb{P}_{\theta^q}^{\otimes N}, \mathbb{P}_{0}^{\otimes N}) \leq \kappa \log(Q). \tag{37}$$

We apply Theorem 2.7 in Tsybakov [45] and (36) and (37) imply the lower bound result. □

8.1. Proof of Proposition 3.

Assume that $f \in W^{(\mathrm{per})}(\boldsymbol{\beta}, \mathbf{L})$. Define for $j_1 = 1, \ldots, \beta_1$, the Fourier coefficients of $(\partial^{j_1} f(t,x)/(\partial t^{j_1})(t,x)$ with respect to the trigonometric bases (φ_ℓ, ψ_k):

$$s_{\ell,k}(j_1) = \int_\epsilon^T \int_a^b (\partial^{j_1} f(t,x)/(\partial t^{j_1})(t,x)\varphi_\ell(t)\psi_k(x)dtdx$$

and we set $s_{\ell,k}(0) = c_{jk}$.
We have

$$s_{0,k}(j_1) = \frac{1}{\sqrt{T-\epsilon}}\int_a^b \psi_k(x)dx[(\partial^{j_1-1} f(t,x)/(\partial t^{j_1-1})(\epsilon,x) - (\partial^{j_1-1} f(t,x)/(\partial t^{j_1-1})(T,x)] = 0.$$

Integrating by parts w.r.t. t under the integral, we get for $\ell \geq 1$

$$\begin{aligned} s_{2\ell-1,k}(\beta_1) &= \frac{2}{\sqrt{T-\epsilon}}\int_a^b \psi_k(x)dx \int_\epsilon^T \frac{\partial^{\beta_1-1} f(t,x)}{\partial t^{\beta_1-1}}(t,x)\frac{2\pi\ell}{T-\epsilon}\sin(2\pi\ell\frac{t-\epsilon}{T-\epsilon})dt \\ &= \frac{2\pi\ell}{T-\epsilon}s_{2\ell,k}(\beta_1-1). \end{aligned}$$

Analogously, $s_{2\ell,k}(\beta_1) = -\frac{2\pi\ell}{T-\epsilon}s_{2\ell-1,k}(\beta_1-1)$. This yields:

$$s^2_{2\ell-1,k}(\beta_1) + s^2_{2\ell,k}(\beta_1) = [\frac{2\pi\ell}{T-\epsilon}]^2[s^2_{2\ell-1,k}(\beta_1-1) + s^2_{2\ell,k}(\beta_1-1)].$$

By induction,

$$\sum_{\ell\geq 1}[\frac{2\pi\ell}{T-\epsilon}]^{2\beta_1}[s^2_{2\ell-1,k}(0) + s^2_{2\ell,k}(0)] = [\frac{\pi}{T-\epsilon}]^{2\beta_1}\sum_{\ell\geq 1} a^2_\ell(\beta_1)s_{\ell,k}(0),$$

with $a_\ell(\beta_1) = \ell^{\beta_1}$ if ℓ is even, and $= (\ell+1)^{2\beta_1}$ if ℓ is odd. We deduce

$$\sum_{\ell,k}\int_\epsilon^T \int_a^b \left(\partial^{\beta_1} f(t,x)/(\partial t^{\beta_1})(t,x)\right)^2 \psi_k(x)\varphi_\ell(t)dtdx = [\frac{\pi}{T-\epsilon}]^{2\beta_1}\sum_{\ell,k} a^2_\ell(\beta_1)s^2_{\ell,k}(0).$$

Hence,

$$\int_\epsilon^T \int_a^b \left(\partial^{\beta_1} f(t,x)/(\partial t^{\beta_1})(t,x)\right)^2 dtdx = [\frac{\pi}{T-\epsilon}]^{2\beta_1}\sum_{\ell,k} a^2_\ell(\beta_1)c^2_{\ell,k}.$$

Therefore,

$$\sum_{\ell,k}\ell^{2\beta_1}c^2_{\ell,k} \leq L_1^2(T-\epsilon)^{2\beta_1}/\pi^{2\beta_1} = R_1.$$

Analogously,

$$\sum_{\ell,k}k^{2\beta_2}c^2_{\ell,k} \leq L_2^2(T-\epsilon)^{2\beta_2}/\pi^{2\beta_2} = R_2.$$

This implies that $f \in W^*(\boldsymbol{\beta}, \mathbf{R})$ as announced. □

9. Appendix

9.1. Some properties of Kronecker products (see Magnus and Neudecker [36], chapter 2)

Recall that $M^\top$ denotes the transpose of the matrix M. The Kronecker product of two matrices M, N with respective dimensions $(m \times n)$ and $(p \times q)$ is the $(mp \times nq)$ matrix defined, if $M = (M_{i,j})_{1\le i\le m, 1\le j\le n}$, by

$$M \otimes N = \begin{pmatrix} M_{1,1}N & \dots & M_{m,1}N \\ M_{1,2}N & \dots & M_{m,2}N \\ \vdots & \ddots & \vdots \\ M_{1,n}N & \dots & M_{m,n}N \end{pmatrix}.$$

The Kronecker product has several nice properties when using vectorization of matrices. For a matrix M as above, denote by vec(M) the vector of $\mathbb{R}^{mn}$ given by

$$\mathrm{vec(M)} = (\mathrm{M_{1,1}}, \dots \mathrm{M_{1,n}}, \mathrm{M_{2,1}}, \dots \mathrm{M_{2,n}}, \dots, \mathrm{M_{m,1}}, \dots \mathrm{M_{m,n}})^\top.$$

The following relations hold for matrices M, N, R:

$$\mathrm{vec}(MNR) = (R^\top \otimes M)\mathrm{vec}(N) \tag{38}$$

$$(M \otimes N)^\top = M^\top \otimes N^\top. \tag{39}$$

If $m = p$ and $n = q$, then the product $M^\top N$ is well defined and is a square matrix $n \times n$ matrix and it holds:

$$\mathrm{Tr}(M^\top N) = \mathrm{vec}(M)^\top \mathrm{vec}(N). \tag{40}$$

Lastly, $M \otimes N$ is invertible if and only if M and N are invertible and in this case,

$$(M \otimes N)^{-1} = M^{-1} \otimes N^{-1}. \tag{41}$$

9.2. Tropp's inequality

We recall that for M a self-adjoint $p \times p$ matrix, the notation $M \succcurlyeq 0$ means that for all $\mathbf{x} \in \mathbb{C}^p$, $\mathbf{x}^\top M \bar{\mathbf{x}} \ge 0$ where $\bar{\mathbf{x}}$ is the conjugate of $\mathbf{x}$.

Theorem 4. *(Matrix Chernoff, Tropp [44]) Consider a finite sequence $\{\mathbf{X}_k\}$ of independent, random, self-adjoint matrices with dimension d. Assume that each random matrix satisfies*

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

Define $\mu_{\min} := \lambda_{\min}(\sum_k \mathbb{E}(\mathbf{X}_k))$ and $\mu_{\max} := \lambda_{\max}(\sum_k \mathbb{E}(\mathbf{X}_k))$. (Here $\lambda_{\min}, \lambda_{\max}$ denote the minimum and the maximum eigenvalue of the matrix). Then

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

9.3. The Hermite basis

The Hermite polynomial of order j is given, for $j \ge 0$, by:

$$H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j}(e^{-x^2}).$$

Hermite polynomials are orthogonal with respect to the weight function e^{-x^2} and satisfy:

$$\int_{\mathbb{R}} H_j(x) H_\ell(x) e^{-x^2} dx = 2^j j! \sqrt{\pi} \delta_{j,\ell}$$

(see *e.g.* Abramowitz and Stegun [1]). The Hermite function of order j is given by:

$$h_j(x) = c_j H_j(x) e^{-x^2/2}, \quad c_j = \left(2^j j! \sqrt{\pi}\right)^{-1/2}. \tag{42}$$

The sequence $(h_j, j \ge 0)$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R})$.

9.4. *The half-trigonometric basis*

The trigonometric basis is well-fitted for functions satisfying boundary conditions as described in W^{per}. This is why we rather used the so-called "half-trigonometric" system, namely the cosine basis defined by $\varphi_{0,T}(x) = \sqrt{1/T}\mathbf{1}_{[0,T]}(t)$, $\varphi_{j,T}(t) = \sqrt{2/T}\cos(\pi jt/T)\mathbf{1}_{[0,T]}(t)$, $j = 1,\ldots,m-1$, see Efromovich [22], p.46. It is clearly an orthonormal basis, which is easy to handle and still has good approximation properties, see Efromovich [22] p.32.

Acknowledgments

We thank the Editor, Associate Editor and referee for their work and helpful remarks.

References

[1] M. Abramowitz, and I. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series, 55 For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.
[2] C. Amorino, D. Belomestny, V. Pilipauskaité, M. Podolskij and S.Y. Zhou, Polynomial rates via deconvolution for nonparametric estimation in McKean-Vlasov SDEs. ArXiv preprint arXiv:2401.04667 (2024).
[3] Y. Baraud, Model selection for regression on a random design. ESAIM Probability and Statistics 6 (2002) 127-146.
[4] Y. Baraud, F. Comte and G. Viennet, Model selection for (auto-)regression with dependent data. ESAIM Probability and Statistics 5 (2001) 33-49.
[5] D. Belomestny, V. Pilipauskaitè and M. Podolskij, Semiparametric estimation of McKean-Vlasov SDEs. Ann. Inst. Henri Poincaré Probab. Stat. 59 (1) (2023) 79-96.
[6] L. Birgé and P. Massart, Minimum contrast estimators on sieves: exponential bounds and rates of convergence. Bernoulli 4 (3) (1998) 329-375.
[7] A. Cohen, M.A. Davenport and D. Leviatan, On the stability and accuracy of least squares approximations. Foundations of Computational Mathematics 13 (5) (2013) 819-834.
[8] A. Cohen, M.A. Davenport and D. Leviatan, Correction to: On the stability and accuracy of least squares approximations. Foundations of Computational Mathematics 19 (1) (2019) 239.
[9] F. Comte and V. Genon-Catalot, Nonparametric drift estimation for *i.i.d.* paths of stochastic differential equations. The Annals of Statistics 48 (6) (2020) 3336-3365.
[10] F. Comte and V. Genon-Catalot, Nonparametric adaptive estimation for Interacting particle systems. Scandinavian Journal of Statistics 50 (4) (2023) 1716-1755.
[11] F. Comte and V. Genon-Catalot, Nonparametric estimation for *i.i.d.* stochastic differential equations with space-time dependent coefficients. SIAM/ASA, Journal of Uncertainty Quantification 12 (2) (2024) 377-410.
[12] F. Comte, V. Genon-Catalot and Y. Rozenholc, Penalized nonparametric mean square estimation of the coefficients of diffusion processes. Bernoulli 13 (2) (2007) 514-543.
[13] F. Comte and C. Lacour, Non compact estimation of the conditional density from direct or noisy data. Annales de l'Institut Henri Poincaré, Probability and Statistics 59 (3) (2023) 1463-1507.
[14] F. Comte and N. Marie, Nonparametric drift estimation from diffusions with correlated Brownian motions. Journal of Multivariate Analysis 198 (105222) (2023) 29pp.
[15] A. Dalalyan, Sharp adaptive estimation of the drift function for ergodic diffusions. The Annals of Statistics 33 (6) (2005) 2507-2528.
[16] A. Dalalyan and M. Reiss, . Asymptotic statistical equivalence for scalar ergodic diffusions. Probability Theory and Related Fields 134 (2) (2006) 248-282.
[17] A. Dalalyan and M. Reiss, Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. Probability Theory and Related Fields 137 (1-2) (2007) 25-47.
[18] L. Della Maestra and M. Hoffmann, Nonparametric estimation for interacting particle systems: McKean-Vlasov models. Probability Theory and Related Fields 182 (1-2) (2022) 551-613.
[19] C. Denis, C. Dion and M. Martinez, Consistent procedures for multiclass classification of discrete diffusion paths. Scandinavian Journal of Statistics 47 (2) (2020) 516-554.
[20] C. Denis, C. Dion-Blanc and M. Martinez, A ridge estimator of the drift from discrete repeated observations of the solution of a stochastic differential equation. Bernoulli 27 (4) (2021) 2675-2713.
[21] F. Dussap, Nonparametric multiple regression by projection on non-compactly supported bases. Annals of the Institute of Statistical Mathematics 75 (5) (2023) 731-771.
[22] S. Efromovich, Nonparametric curve estimation. Methods, theory, and applications. Springer Series in Statistics. Springer-Verlag, New York, 1999.
[23] A. Friedman, Stochastic differential equations and Applications, Volume 1. Academic press. New York, 1975.
[24] A. Gloter, A. and M. Sørensen, Estimation for stochastic differential equations with a small diffusion coefficient. Stochastic Processes and Their Applications 119 (3) (2009) 679-699.
[25] R. Guy, C. Larédo and E. Vergu, Parametric inference for discretely observed multidimensional diffusions with small diffusion coefficient. Stochastic Processes and Their Applications 124 (1) (2014) 51-80.
[26] M. Hoffmann, Adaptive estimation in diffusion processes. Stochastic Processes and Their Applications 79 (1) (1999) 135-163.
[27] R. Höpfner, Asymptotic Statistics with a view to stochastic processes. Walter de Gruyter, Berlin/Boston, 2014.

[28] C. Hsiao, Analysis of panel data. Cambridge University Press, Second Edition. Cambridge, 2003.
[29] S. M. Iacus, Simulation and inference for stochastic differential equations. With R examples. Springer, 2008.
[30] I. Karatzas and S. Shreve, Brownian motion and stochastic calculus. Springer-Verlag, New-York, 1997.
[31] M. Kessler, A. Lindner and M. Sørensen, Editors, Statistical methods for stochastic differential equations. CRC press. Taylor & Francis Group. Boca Raton, 2012.
[32] J. Kleffe, On Hsu's Theorem in Multivariate Regression. Journal of Multivariate Analysis 9 (3) (1979) 442-451.
[33] Y.A. Kutoyants, Parameter estimation for stochastic processes. Berlin: Heldermann, 1984.
[34] Y.A Kutoyants, Statistical inference for ergodic diffusion processes. Springer, London, 2004.
[35] R.S. Liptser and A.N. Shiryaev, Statistics of random processes. I. Appl. Math. (N. Y.), 5 Stoch. Model. Appl. Probab. Springer-Verlag, Berlin, xvi+427 pp., 2001.
[36] J.R Magnus and H. Neudecker, Matrix differential calculus with applications in statistics and econometrics. Wiley Ser. Probab. Math. Statist. Appl. Probab. Statist. John Wiley & Sons, Ltd., Chichester, 1988.
[37] N. Marie and A. Rosier, Nadaraya-Watson Estimator for i.i.d. Paths of Diffusion Processes. Scandinavian Journal of Statistics 50 (2) (2023) 589-637.
[38] S. Menozzi, A. Pesce and X. Zhang, Density and Gradient Estimates for Non Degenerate Brownian SDEs with Unbounded Measurable Drift. Journal of Differential Equations 272 (2021) 330-369.
[39] J.O. Ramsay, and B.W. Silverman, Applied functional data analysis: Methods and case studies. Springer, New-york, 2007.
[40] D. Revuz and M. Yor, Continuous martingales and Brownian motion. Third edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 293. Springer-Verlag, Berlin, 1999.
[41] L.C.G. Rogers and D. Williams, Diffusions, Markov processes and Martingales. Volume 2: Itô Calculus. John Wiley & Sons. Chichester, 1990.
[42] M. Sørensen and M. Uchida, Small diffusion asymptotics for discretely sampled stochastic differential equations. Bernoulli 9 (6) (2003) 1051-1069.
[43] C. Strauch, Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. The Annals of Statistics 46 (6B) (2018) 3451-3480.
[44] J. A. Tropp, User-friendly tail bounds for sums of random matrices. Foundation of Computational Mathematics 12 (4) (2012) 389-434.
[45] A. B. Tsybakov, Introduction to nonparametric estimation. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. Springer Series in Statistics. Springer, New York, 2009.
[46] M. Uchida, Estimation for Discretely Observed Small Diffusions Based on Approximate Martingale Estimating Functions. Scandinavian Journal of Statistics 31 (4) (2004) 553-566.
[47] J.-L. Wang, J.-M. Chiou and H.-G. Mueller, Review of functional data analysis. ArKiv preprint 1507.05135, 2015.
[48] N. Yoshida, Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin-Watanabe. Probability Theory and Related Fields 92 (3) (1992) 275-311.