New results for drift estimation in inhomogeneous stochastic differential equations
Fabienne Comte, Valentine Genon-Catalot

To cite this version:
Fabienne Comte, Valentine Genon-Catalot. New results for drift estimation in inhomogeneous stochastic differential equations. 2024. hal-04597162

HAL Id: hal-04597162
https://cnrs.hal.science/hal-04597162
Preprint submitted on 1 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
NEW RESULTS FOR DRIFT ESTIMATION IN INHOMOGENEOUS STOCHASTIC DIFFERENTIAL EQUATIONS.

F. COMTE(1), V. GENON-CATALOT(1)

ABSTRACT. We consider $N$ independent and identically distributed (i.i.d.) stochastic processes $(X_j(t), t \in [0,T])$, $j = 1, \ldots, N$, 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 $I_\epsilon = [\epsilon,T] \times A$, with $\epsilon \geq 0$ and $A \subset \mathbb{R}$, we build by a projection method an estimator of $b$ on $I_\epsilon$. As the function is bivariate, this amounts to estimating a matrix of projection coefficients instead of a vector for univariate functions. Below, the use of Kronecker products 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.

June 1, 2024

Keywords and phrases: Adaptive estimator, Diffusion process, Kronecker product, time-dependent coefficients, nonparametric estimation, projection method.

AMS Classification. 62G07–62M05.

1. Introduction

Statistical inference for the coefficients of stochastic differential equations (SDEs) has a long-standing history and a huge number of contributions deals with the subject. We can refer to the textbooks Kessler et al. (2012), Höpfner (2014), Iacus (2010), Kutoyants, (1984, 2004) 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. (1992), Sørensen and Uchida (2003), Uchida (2004), Gloter and Sørensen (2009), Guy et al. (2014)). 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 (1999), Dalalyan (2005), Dalalyan and Reiss (2006, 2007), Comte et al. (2007), Strauch (2018).

In relation with functional data analysis (see e.g. Ramsay and Silvermann (2007), Wang et al. (2016), Hsiao (2003)), the case of i.i.d. paths of stochastic differential equations has received

(1): Université Paris Cité, MAP5, UMR 8145 CNRS, F-75006, FRANCE, email: fabienne.comte@u-paris.fr, valentine.genon-catalot@mi.parisdescartes.fr.
recently considerable attention. Results concerning nonparametric estimation in this setup have been published (see e.g. Comte and Genon-Catalot (2020), Denis et al (2018, 2021), Marie and Rosier (2023), see also Comte and Marie (2023) 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 dependent coefficients (see e.g. Della Maestra and Hoffmann (2022), Comte and Genon-Catalot (2023)). When there is no interaction between particles, these models reduce to i.i.d. diffusion processes.

In this paper, we consider $N$ i.i.d. real-valued stochastic processes $(X_i(t), t \geq 0)$, $i = 1, \ldots, N$, with dynamics ruled by:

\[
\begin{align*}
\frac{dX_i(t)}{dt} &= b(t, X_i(t)) + \sigma(t, X_i(t)) \frac{dW_i(t)}{dt}, \\
X_i(0) &= x_0, 
\end{align*}
\]

where $x_0 \in \mathbb{R}$ is known and $(W_1, \ldots, W_N)$ are independent standard Brownian motions. The functions $b, \sigma: \mathbb{R}^+ \times \mathbb{R} \rightarrow \mathbb{R}$ 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 (2024) 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 proceed by projection method on sieves and estimate the function $b(t, x)$ on a set $I_\epsilon = [\epsilon, T] \times A$ with $0 < \epsilon < T$ and $A \subset \mathbb{R}$. We thus define a collection of finite-dimensional subspaces of $L^2(I_\epsilon)$, $(S_{m_1} \times S_{m_2}, m_1, m_2 \geq 0)$, where $S_{m_1}$ is spanned by an orthonormal basis $(\varphi_j, 0 \leq j \leq m_1 - 1)$ of $L^2([\epsilon, T])$ and $S_{m_2}$ is spanned by an orthonormal basis $(\psi_k, 0 \leq k \leq m_2 - 1)$ of $L^2(A)$. As usual for projection method, we estimate a projection of $b1_I_\epsilon$ on $(S_{m_1} \times S_{m_2})$, for $m_1 \geq 1, m_2 \geq 1$ which is a function of the form

\[
b_m(t, x) = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} a_{j,k} \varphi_j \psi_k(t, x), \quad \varphi_j \psi_k(t, x) = \varphi_j(t) \psi_k(x), 0 \leq j \leq m_1 - 1, 0 \leq k \leq m_2 - 1
\]

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 (1979), Magnus and Neudecker (1988). In particular, the $m_1 \times m_2$ matrix $A_m = (a_{j,k}), 0 \leq j \leq m_1 - 1, 0 \leq k \leq m_2 - 1$, $m = (m_1, m_2)$ is transformed into a $m_1 m_2 \times 1$ dimensional vector vec$(A_m)$ by stacking the columns of the matrix $A_m$ and the classical regression equation, defining the estimator $\hat{A}_m = (\hat{a}_{jk}, 0 \leq j \leq m_1 - 1, 0 \leq k \leq m_2 - 1)$ of the matrix $A_m$, looks like the usual one,

\[
\hat{\Theta}_{m_1 m_2} \text{vec}(\hat{A}_m) = \text{vec}(\hat{C}_m).
\]

where $\hat{\Theta}_{m_1 m_2}$ is a $m_1 m_2 \times m_1 m_2$ matrix and vec$(\hat{C}_m)$ is a $m_1 m_2 \times 1$ dimensional vector both using the observations.

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. For the study of the risk of the estimator, we distinguish the case where $\epsilon = 0$ (Section 4) and the case $\epsilon > 0$ and $A = [a, b]$ compact (Section 5). In the former case, we propose a data-driven choice of $m$ leading to an adaptive estimator. In the latter case, we investigate rates of convergence and prove a lower bound showing that our estimator is minimax. 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 definition of the Hermite basis is 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 $(\mathcal{F}_i = \sigma(W_i(s), s \leq t, i = 1,\ldots,N), t \geq 0)$ (see e.g. Rogers and Williams, 1990, 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  
  $$-\frac{\partial v}{\partial s} = \frac{1}{2} \sigma^2(s,x) \frac{\partial^2 v}{\partial x^2} + b(s,x) \frac{\partial v}{\partial x}$$

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 $(t,y)$:  
  $$\frac{\partial w}{\partial t} = \frac{1}{2} \sigma^2(t,y) \frac{\partial^2 w}{\partial y^2} - \frac{\partial \rho(t,y)w(t,y)}{\partial y}.$$ 

The function $w(t,y)$ is of class $C^{1,2}([t,T] \times \mathbb{R})$ (see e.g. Karatzas and Shreve, 1986, pp. 368-369, Friedman, 1975, 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 $(t,y) \rightarrow p_t(y)$ is positive and continuous and the following holds:

\begin{equation} \tag{2} \forall k \geq 0, \forall t \geq 0 \quad \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. \end{equation}

We also have

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

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 (1997, p.306)).
Consider the function
\[ f_T(y) = \frac{1}{T} \int_0^T p_u(y) du. \] (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. (2021), \( p_u(y) \leq Cu^{-1/2} \) 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 \( I_\epsilon := [\epsilon, T] \times A \) for \( A \subset \mathbb{R} \). Let \( h \in L^2(I_\epsilon, dt du) \), we set \( \|h\|^2 = \int_{I_\epsilon} h^2(t, x) dt dx \). For a matrix \( M \), we denote by \( M^\perp \) 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 \( x \) a vector of \( \mathbb{R}^q \), we denote \( \|x\|_{2,q} \) its Euclidean norm.

3.1. Projection spaces. To define nonparametric estimators of the drift function \( b \), we proceed by a projection method. Let \( (\varphi_j, 0 \leq j \leq m_1 - 1) \) be an orthonormal system of bounded piecewise continuous functions of \( L^2([\epsilon, T], dt) \) and \( (\psi_k, 0 \leq k \leq m_2 - 1) \) an orthonormal system of bounded piecewise continuous functions of \( L^2(A, dx) \). We define \( (S_{m_1} \times S_{m_2}, m_1, m_2 \geq 0) \) a family of finite-dimensional subspaces of \( L^2(I_t) \), where \( S_{m_1} \) is spanned by \( (\varphi_j, 0 \leq j \leq m_1 - 1) \) and \( S_{m_2} \) is spanned by \( (\psi_k, 0 \leq k \leq m_2 - 1) \). The bases of \( S_{m_1} \), \( S_{m_2} \) 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 \( S_{m_1} \times S_{m_2} \). Let us set:

\[ L_\varphi(S_{m_1}) = \sup_{t \in [\epsilon, T]} \sum_{j=0}^{m_1-1} \varphi_j^2(t), \quad L_\psi(S_{m_2}) = \sup_{x \in A} \sum_{j=0}^{m_2} \psi_j^2(x) \] (5)

These quantities were introduced by Birgé and Massart (1998), and used by Comte and Genon-Catalot (2020) in the framework of regression and drift estimation for diffusions by projection method. By Lemma 1 in Birgé and Massart (1998),
\[ L_\varphi(S_{m_1}) = \sup_{h_1 \in S_{m_1}, \|h_1\|_{[\epsilon, T]} = 1} \sup_{t \in [\epsilon, T]} h_1^2(t), \quad L_\psi(S_{m_2}) = \sup_{h_2 \in S_{m_2}, \|h_2\|_A = 1} \sup_{x \in A} h_2^2(x), \]

where \( \|h_1\|^2_{[\epsilon, T]} = \int_{\epsilon}^T h_1^2(t) dt, \|h_2\|^2_A = \int_A h_2^2(x) dx \). Therefore, \( L_\varphi(S_{m_1}) \) and \( L_\psi(S_{m_2}) \) only depend on the subspaces and not on the bases chosen to define them. In relation with (5), we assume

- \( [H2] \) \( \exists c_\varphi, c_\psi > 0 \) such that \( L_\varphi(S_{m_1}) \leq c_\varphi^2 m_1, L_\psi(S_{m_2}) \leq c_\psi^2 m_2. \)
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 \( L_{\psi}(\Sigma_m) \leq cm_1^{1/2} \) (see Comte and Lacour (2023), Lemma 1, examples in Section 6 and Appendix).

### 3.2. Projection estimator.

For \( h : \mathbb{R}^+ \times \mathbb{R} \rightarrow \mathbb{R} \) a function, we introduce the contrast:

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

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

\[
\mathbb{E}\gamma_N(h) = \frac{1}{T} \mathbb{E} \int_{\epsilon}^{T} (h(u, X_1(u)) - b(u, X_1(u)))^2 du - \frac{1}{T} \mathbb{E} \int_{\epsilon}^{T} b^2(u, X_1(u))du
\]

\[
= \frac{1}{T} \int_{\epsilon}^{T} \int (h(u, y) - b(u, y))^2 p_u(y)dydt - \frac{1}{T} \int_{\epsilon}^{T} b^2(u, y)p_u(y)dydu,
\]

which is minimum for \( h(u, y) \equiv b(u, y) \). This property justifies the definition of a collection of estimators \( \hat{b}_m, m = (m_1, m_2), m_1, m_2 \geq 0 \) of \( b_L := b1_1 \) by setting:

\[
\hat{b}_m = \arg\min_{h \in S_{m_1} \times S_{m_2}} \gamma_N(h)
\]

Thus, for each couple \( m = (m_1, m_2) \), we can write

\[
\hat{b}_m(t, x) = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} \hat{a}_{j,k} \varphi_j \otimes \psi_k(t, x)
\]

where the matrix of coefficients

\[
\hat{A}_m = (\hat{a}_{j,k})_{0 \leq j \leq m_1-1, 0 \leq k \leq m_2-1}
\]

is computed as follows. Define the \( m_1 \times m_2 \)-matrix

\[
\hat{C}_m = \left( \frac{1}{NT} \sum_{i=1}^{N} \int_{\epsilon}^{T} \varphi_j(u)\psi_k(X_i(u))dX_i(u) \right)_{0 \leq j \leq m_1-1, 0 \leq k \leq m_2-1}
\]

and the respectively \( m_1 \times m_1 \) and \( m_2 \times m_2 \) matrices

\[
\Phi_{m_1}(t) = \left( \varphi_j(t) \varphi_j'(t) \right)_{1 \leq j, j' \leq m_1-1}
\]

\[
\hat{\Psi}_{m_2}(t) = \left( \frac{1}{N} \sum_{i=1}^{N} \psi_k(X_i(t))\psi_{k'}(X_i(t)) \right)_{0 \leq k, k' \leq m_2-1}
\]

We also define

\[
\Psi_{m_2}(t) = \left( \int \psi_k(x)\psi_{k'}(x)p_t(x)dx \right)_{0 \leq k, k' \leq m_2-1} = \mathbb{E}\hat{\Psi}_{m_2}(t).
\]

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

For \( h = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} h_{j,k} \varphi_j \otimes \psi_k \), we write that

\[
\frac{\partial \gamma_N}{\partial h_{j_0,k_0}}(\hat{b}_m) = 0, \quad \text{for} \ j_0 = 0, \ldots, m_1 - 1, k_0 = \ldots, m_2 - 1.
\]
This yields the equation defining $\hat{A}_m$:

\begin{equation}
\frac{1}{T} \int_\epsilon^T \Phi_{m_1}(t) \hat{A}_m \Psi_{m_2}(t) dt = \hat{C}_m.
\end{equation}

Define the $m_1m_2 \times m_1m_2$ matrices

\begin{equation}
\hat{\Theta}_{m_1m_2} := \frac{1}{T} \int_\epsilon^T \hat{\Psi}_{m_2}(t) \otimes \Phi_{m_1}(t) dt, \quad \hat{\Theta}_{m_1m_2} := \frac{1}{T} \int_\epsilon^T \Psi_{m_2}(t) \otimes \Phi_{m_1}(t) dt,
\end{equation}

where clearly $\hat{\Theta}_{m_1m_2} = \mathbb{E} \hat{\Theta}_{m_1m_2}$. As for all $t$, the matrices $\Phi_{m_1}(t), \hat{\Psi}_{m_2}(t), \Psi_{m_2}(t)$ are symmetric, so is $\hat{\Psi}_{m_2}(t) \otimes \Phi_{m_1}(t)$ (see (38)). As a consequence, $\hat{\Theta}_{m_1m_2}$ and $\hat{\Theta}_{m_1m_2}$ are also symmetric.

We now introduce the vectorization of matrices $\hat{A}_m, \hat{C}_m$ to clarify equation (13) defining $\hat{A}_m$.

**Proposition 1.** Equation (13) is equivalent to

\[ \hat{\Theta}_{m_1m_2} \operatorname{vec}(\hat{A}_m) = \operatorname{vec}(\hat{C}_m). \]

The vector of coefficients of (8) is uniquely defined if $\hat{\Theta}_{m_1m_2}$ is invertible and in this case,\n
\[ \operatorname{vec}(\hat{A}_m) = (\hat{\Theta}_{m_1m_2})^{-1} \operatorname{vec}(\hat{C}_m). \]

Note that although $\Phi_{m_1}(t)$ has rank 1 for all $t$, this does not prevent $\hat{\Theta}_{m_1m_2}$ or $\hat{\Theta}_{m_1m_2}$ from being invertible.

Let us give some general notation for the sequel. For $h(.,.), \ell(.,.)$ two bounded functions, we set

\begin{equation}
\|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,
\end{equation}

\begin{equation}
\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).
\end{equation}

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

4. Estimation for $\epsilon = 0$

4.1. Empirical risk of the estimator for fixed $m$. In this section, we consider $\epsilon = 0$ and set $I_0 = I$. We keep everywhere the same notations but with $\epsilon = 0$.

We define the risk of the estimator $\hat{b}_m$ given by (8) as the expectation of the empirical square norm $\|.\|_N^2$ (with $\epsilon = 0$) which is naturally associated with our observations.

Using the previous notations and equation (1), we split $\hat{C}_m$ into the sum of two random matrices:

\begin{equation}
\hat{C}_m = C_m + \varepsilon_m, \quad C_m := (\langle \varphi_j \otimes \psi_k, b \rangle_N)_{0 \leq j \leq m_1-1, 0 \leq k \leq m_2-1}
\end{equation}

and

\begin{equation}
\varepsilon_m := (\nu_N(\varphi_j \otimes \psi_k))_{0 \leq j \leq m_1-1, 0 \leq k \leq m_2-1}.
\end{equation}
Lemma 1. Let $h = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} h_{j,k} \varphi_j \otimes \psi_k$ and denote by $H_m = (h_{j,k})_{0 \leq j \leq m_1-1, 0 \leq k \leq m_2-1}$. We have

$$
\|h\|_N^2 = \frac{1}{T} \int_0^T \text{Tr} \left[ H_m^\top \Phi_{m_1}(t) H_m \tilde{\Psi}_{m_2}(t) \right] dt = [\text{vec}(H_m)]^\top \tilde{\Theta}_{m_1,m_2} \text{vec}(H_m),
$$

$$
\|h\|_p^2 = \mathbb{E} \left( \frac{1}{T} \int_0^T \text{Tr} \left[ H_m^\top \Phi_{m_1}(t) H_m \tilde{\Psi}_{m_2}(t) \right] dt \right) = [\text{vec}(H_m)]^\top \Theta_{m_1,m_2} \text{vec}(H_m),
$$

where $\Theta_{m_1,m_2}$ and $\tilde{\Theta}_{m_1,m_2}$ are defined in (14).

Lemma 1 implies that $\tilde{\Theta}_{m_1,m_2}$ and $\Theta_{m_1,m_2}$ 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 $0 \leq j \leq m_1-1$ and $0 \leq k \leq m_2-1$. Consequently, $\Theta_{m_1,m_2}$ is positive definite. As $\tilde{\Theta}_{m_1,m_2}$ tends to $\Theta_{m_1,m_2}$ as $N \to +\infty$ a.s., $\tilde{\Theta}_{m_1,m_2}$ is invertible for $N$ large enough.

Note that we have

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

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

$$
\forall u \in [0,T], \forall i = 1, \ldots, N, \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} h_{j,k} \varphi_j \otimes \psi_k(u, X_i(u)) = 0 \text{ a.s. and a.e. on } [0,T]
$$

$$
\Rightarrow \{ h_{j,k} = 0, j = 0, \ldots, m_1-1, k = 0, \ldots, m_2-1 \}.
$$

Condition (20) is an identifiability constraint linked with the choice of the bases. In Section 6, we use bases for which (20) is fulfilled. Nevertheless, (20) is not enough for our theory and we need to reinforce the identifiability constraint (see (22)).

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., 2013).

$$
(21) \quad m_1 m_2 \leq NT, \quad L_\psi(m_2) \|\Theta_{m_1 m_2}^{-1}\|_\text{op} \leq c_r \frac{NT}{\log(NT)}, \quad c_r = \frac{3 \log(3/2) - 1}{2 + 2r},
$$

The corresponding empirical version allows to define a truncation of the estimator $\tilde{b}_m$.

$$
(22) \quad \tilde{\Lambda}_m = \left\{ m_1 m_2 \leq NT, \quad L_\psi(m_2) \|\Theta_{m_1 m_2}^{-1}\|_\text{op} \leq 2c_r \frac{NT}{\log(NT)} \right\}.
$$

The final estimator that we study for fixed $(m_1, m_2)$ is defined by:

$$
(23) \quad \tilde{b}_m = \bar{b}_m 1_{\tilde{\Lambda}_m}
$$

Theorem 1. Under assumption [H1] and condition (21) on $m$ with $r \geq 2$ in (21)-(22), it holds

$$
E(\|\tilde{b}_m - b_t\|_N^2) \leq \inf_{h \in S_{m_1} \times S_{m_2}} \|h - b_t\|_p^2 + \frac{2 \sigma^2 m_1 m_2}{NT} + \frac{C}{NT},
$$

where $\sigma^2$ is the upper bound on $\sigma^2$ defined in [H1]-[iv], and $C$ is a positive constant.

The risk bound (24) is the sum of 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
\[
\frac{2}{NT} \text{Tr} \left( \Theta_{m_1 m_2}^{-1} \Theta_{m_1 m_2, \sigma^2} \right) \leq \frac{2\sigma_1^2 m_1 m_2}{NT},
\]
where
\[
\Theta_{m_1 m_2, \sigma^2} = \frac{1}{T} \int_0^T \Psi_{m_2, \sigma^2}(t) \Phi_{m_1}(t) dt, \quad \Psi_{m_2, \sigma^2}(t) = \left( \int \psi_k(x) \psi_{k'}(x) \sigma^2(t, x) p_t(x) dx \right)_{0 \leq k, k' \leq m_2 - 1}.
\]

4.2. Adaptive estimation. In this paragraph, we investigate the possible choice of a data-driven \( m \) which leads to an adaptive estimator realizing automatically the bias-variance compromise.

Consider, for a positive constant, the theoretical collection of models:
\[
M_N = \left\{ m \in \mathbb{N}^2, m_1 m_2 \leq NT, L_{\psi}(m_2)(\|\Theta_{m_1 m_2}\|_{\text{op}}^2 \vee 1) \leq \frac{c}{2} \frac{NT}{\log^2(NT)} \right\}
\]
and its empirical counterpart:
\[
\hat{M}_N = \left\{ m \in \mathbb{N}^2, m_1 m_2 \leq NT, L_{\psi}(m_2)(\|\hat{\Theta}_{m_1 m_2}\|_{\text{op}}^2 \vee 1) \leq 2c \frac{NT}{\log^2(NT)} \right\}.
\]

Then define
\[
\hat{m} = \arg \min_{m \in \hat{M}_N} \left( \gamma_N(\hat{b}_m) + \text{pen}(m) \right), \quad \text{pen}(m) = \kappa \sigma_1^2 m_1 m_2 / NT,
\]
where \( \kappa \) is a numerical constant and \( \sigma_1^2 \) is the upper bound on \( \sigma^2 \) (see [H1-iv]). The following result holds:

**Theorem 2.** Under assumption [H1], there exists a numerical constant \( \kappa_0 \) such that for all \( \kappa \geq \kappa_0 \),
\[
\mathbb{E}(\|\hat{b}_m - b_L\|_{\Delta}^2) \leq C \inf_{m \in M_N} \left( \inf_{h \in S_{m_1} \times S_{m_2}} \|h - b_L\|^2 + \frac{\sigma_1^2 m_1 m_2}{NT} \right) + C' \frac{NT}{NT},
\]
where \( C \) and \( C' \) are positive constants.

The proof is omitted as it follows closely the analogous result in Comte and Genon-Catalot (2020). It relies on the standard decomposition, for \( h, h^* \) two functions of \( S_{m_1} \times S_{m_2} \),
\[
\gamma_N(h) - \gamma_N(h^*) = \|h - h^*\|^2 + 2\nu_N(h - h^*)
\]
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 \( \langle M \rangle_T \leq \sigma_1^2 \|h\|^2_{\Delta} \). Therefore, the following Bernstein inequality for martingales (see e.g. Revuz and Yor (1999)) holds
\[
P(\nu_N(h) \geq \delta, \langle M \rangle_T \leq v^2) \leq \exp \left( -\frac{NT\delta^2}{2\sigma_1^2 v^2} \right).
\]

Using this and the chaining method described in Baraud et al. (2001) gives the result.
5. Optimal rate for compact support estimation with $\epsilon > 0$.

5.1. Upper bound. Here we consider $\epsilon > 0$ and $A = [a, b]$, $a < b$, a compact interval of $\mathbb{R}$. Then

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

and the following Lemma holds:

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

Now, the stability condition (21) can be simplified into

$$m_2 \leq \frac{c_0(\epsilon, A)}{c_2^2} \frac{NT}{\log(NT)}.$$ 

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

$$(26) \quad \mathcal{M}_N^* = \{ (m_1, m_2) \in \mathbb{N}^2, m_1m_2 \leq NT, m_2 \leq NT/\log^2(NT) \}.$$ 

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

$$\mathbb{E}(\|\tilde{b}_m - b_L\|_N^2) \leq c_1(\epsilon, A) \inf_{h \in S_{m_1} \times S_{m_2}} \|h - b_L\|^2 + 2\frac{\sigma_1^2m_1m_2}{NT} + \frac{C}{NT},$$

where $C$ is a positive constant.

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

$$W^*(\beta, R) = \{ f \in \mathbb{L}^2([\epsilon, T] \times [a, b]), \text{ such that } \sum_{j,k \geq 0} c_{j,k}^2 j^{2\beta_1} \leq R_1^2, \sum_{j,k \geq 0} c_{j,k}^2 k^{2\beta_2} \leq R_2^2, \text{ where } c_{j,k} = (f, \varphi_j \circ \psi_k), j \geq 0, k \geq 0 \}.$$ 

**Proposition 2.** (upper bound) Under assumption $[H1]$, if $b_L \in W^*(\beta, R)$, choosing

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

we get

$$\mathbb{E}(\|\tilde{b}_{m^*} - b_L\|_N^2) \lesssim (NT)^{-\frac{2\beta}{2\beta_1 + 2\beta_2}}, \quad \frac{1}{\beta} = \frac{1}{2} \left( \frac{1}{\beta_1} + \frac{1}{\beta_2} \right).$$

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

5.2. Lower bound. Define for $\beta = (\beta_1, \beta_2) \in \mathbb{N}^2$ and $L = (L_1, L_2)$,

$$W(\beta, L) = \{ f \in \mathbb{L}^2([\epsilon, T] \times [a, b]), f \text{ derivable up to order } \beta_1 \text{ w.r.t. } t, \text{ up to order } \beta_2 \text{ w.r.t. } x \}$$

with

$$\int \int (\partial^{\beta_1}f(t, x)/\partial t^{\beta_1})^2 dt dx \leq L_1^2, \int \int (\partial^{\beta_2}f(t, x)/\partial x^{\beta_2})^2 dt dx \leq L_2^2.$$ 

Then we can prove:

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

$$\liminf_{N \to +\infty} \inf_{T_N} \sup_{b_L \in W(\beta, L)} \mathbb{E}_{b_L} [N^{\frac{2\beta}{2\beta_1 + 2\beta_2}} \|T_N - b_L\|^2] \geq c$$

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

5.3. **Case of the trigonometric basis.** To illustrate and make more concrete the sets $W(\beta, L)$ and $W^*(\beta, R)$, let us consider trigonometric bases:

\[
\begin{align*}
\varphi_0(t) &= \frac{1}{\sqrt{T - \epsilon}} 1_{[\epsilon, T]}(t) \\
\varphi_{2j-1}(t) &= \frac{1}{\sqrt{T - \epsilon}} \cos(2\pi j \frac{t - \epsilon}{T - \epsilon}) 1_{[\epsilon, T]}(t) \\
\varphi_{2j}(t) &= \frac{1}{\sqrt{T - \epsilon}} \sin(2\pi j \frac{t - \epsilon}{T - \epsilon}) 1_{[\epsilon, T]}(t)
\end{align*}
\]

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

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

\[
W^{(\text{per})}(\beta, L) = \{ f \in W(\beta, L), \forall x \in [a, b], (\partial^{a_1} f(t, x)/(\partial t^{a_1}))(\epsilon, x) = (\partial^{a_1} f(t, x)/(\partial t^{a_1}))(T, x), \forall t \in [\epsilon, T], (\partial^{a_2} f(t, x)/(\partial x^{a_2}))(t, a) = (\partial^{a_2} f(t, x)/(\partial x^{a_2}))(t, b), 0 \leq a_1 \leq \beta_1 - 1, i = 1, 2 \}.
\]

We can prove the result.

**Proposition 3.** For integers $\beta_1$ and $\beta_2$, $L = (L_1, L_2), R = (R_1, R_2)$

\[
f \in W^{(\text{per})}(\beta, L) \Rightarrow f \in W^*(\beta, R)
\]

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 $W^{(\text{per})}(\beta, L)$ (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 $W^*(\beta, R)$ and therefore, the upper bound holds. We conclude that the rates are minimax optimal on this set.

6. **Examples and numerical simulation results**

![Figure 1](image_url)

**Figure 1.** Smallest eigenvalue of $\hat{\Theta}_{m_1, m_2}$ in function of $m_1$ and $m_2$ going from 1 to 20, on one sample $N = 1000$ and function $b_1$. Left: basis $T$, right: basis $H$.

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 $\sigma$ 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 \( N = 1000 \).

![Figure 2. Estimation of \((t, x) \mapsto b_1(t, x)\) (line 1), \(b_2(t, x)\) (line 2), \(b_3(t, x)\) (line 3). Left: True, Middle: estimation with basis \(T\), Right: estimation with basis \(H\). \(N = 1000\), selected couples of dimensions (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 \( ((1, \sqrt{2} \cos(\pi j x), j \geq 1))\), 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 (25) with $\kappa_T = 4$ and $\kappa_H = 6$ for the value of $\kappa$ with basis $T$ and $H$ respectively. The term $\sigma_1^2$ is equal to $\sigma^2$ in the constant variance case, and the true value is used (a residual least square estimator may be computed).

We plot in Figure 1 the surface corresponding to the smallest eigenvalue of $\hat{\Theta}_{m_1m_2}$ 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.

![Figure 1](image1.png)

**Figure 2.** The 3-D representation of the functions $(t,x) \mapsto b_i(t,x)$ for $i = 1, 2, 3$ 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 Figure 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 $(X_j(t), t \in [0,T]), j = 1, \ldots, N$, defined by a one-dimensional stochastic differential equation (SDE) with general time-dependent drift and diffusion coefficient. Considering a set $I_\epsilon = [\epsilon, T] \times A$, with $\epsilon \geq 0$ and $A \subset \mathbb{R}$, we build by a projection method an estimator of $b$ on $I_\epsilon$. 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.

8. Proofs

8.1. Proof of Proposition 1. We vectorize relation (13) and obtain:

$$\text{vec}(\frac{1}{T} \int_0^T \Phi_m(t) \tilde{A}_m \tilde{\Psi}_m(t) dt) = \frac{1}{T} \int_0^T \text{vec}(\Phi_m(t) \tilde{A}_m \tilde{\Psi}_m(t)) dt = \text{vec}(\hat{C}_m).$$

Now using the relation $\text{vec}(ABC) = (C^\perp \otimes A)\text{vec}(B)$ (linking the three matrices $A, B, C$), we get, as $\tilde{A}_m$ does not depend on $t$,

$$\frac{1}{T} \int_0^T \text{vec}(\Phi_m(t) \tilde{A}_m \tilde{\Psi}_m(t)) dt = \frac{1}{T} \int_0^T \tilde{\Psi}_m(t) \otimes \Phi_m(t) \text{vec}(\tilde{A}_m) dt = \hat{\Theta}_{m_1m_2} \text{vec}(\tilde{A}_m).$$

8.2. Proof of Lemma 1. The beginning of the computation is straightforward.

$$\|h\|_N^2 = \frac{1}{T} \sum_{0 \leq j,j' \leq m_1-1} \sum_{0 \leq k,k' \leq m_2-1} h_{j,k} h_{j',k'} \int_0^T [\Phi_m(u)]_{j,j'} [\tilde{\Psi}_m(u)]_{k,k'} du$$

$$= \frac{1}{T} \int_0^T \sum_{k=0}^{m_2-1} \left( \sum_{j=0}^{m_1-1} [H_m^j]_{k,j} \left( \sum_{k'=0}^{m_1-1} [\Phi_m(u)]_{j,j'} [H_m]_{j,j'} [\tilde{\Psi}_m(u)]_{k',k} \right) \right) du$$

$$= \frac{1}{T} \int_0^T \sum_{k=0}^{m_2-1} \left[ H_m^j \Phi_m(u) H_m \tilde{\Psi}_m(u) \right]_{k,k} du$$

$$= \frac{1}{T} \int_0^T \text{Tr} \left[ H_m^j \Phi_m(u) H_m \tilde{\Psi}_m(u) \right] du,$$

which is the first equality of the Lemma. We use equality (39) which yields

$$\text{Tr} \left[ H_m^j \Phi_m(u) H_m \tilde{\Psi}_m(u) \right] = \text{vec}(H_m^j) \text{vec}(\Phi_m(u) H_m \tilde{\Psi}_m(u)).$$

Now with (37), we obtain:

$$\text{vec}(\Phi_m(u) H_m \tilde{\Psi}_m(u)) = (\tilde{\Psi}_m(u) \otimes \Phi_m(u)) \text{vec}(H_m).$$

Integrating wrt $u$ gives

$$\|h\|_N^2 = [\text{vec}(H_m)]^\perp \hat{\Theta}_{m_1m_2} \text{vec}(H_m).$$

The last equalities are obtained by taking expectation. □

8.3. Proof of Theorem 1. Let us define

$$\Omega_m := \left\{ \left( \frac{\|h\|_N^2}{\|h\|_p^2} - 1 \right) \leq \frac{1}{2}, \forall h \in S_{m_1} \times \Sigma_{m_2} \right\}. $$

On $\Omega_m$, the empirical norm $\|\cdot\|_N$ and the $p$-norm are equivalent for elements of $S_{m_1} \times \Sigma_{m_2}$:

$$\frac{2}{3} \|h\|_N^2 \leq \|h\|_p^2 \leq 2\|h\|_N^2$$

and the following result holds.
Lemma 3. We have

\begin{equation}
\Omega_m = \left\{ \left\| \Theta_{m_1m_2}^{-1/2} \hat{\Theta}_{m_1m_2} \Theta_{m_1m_2}^{-1/2} - \text{Id}_{m_1m_2} \right\|_{\text{op}} \leq \frac{1}{2} \right\}.
\end{equation}

Moreover, under $m_1m_2 \leq NT$ and under (21), it holds that $\mathbb{P}(\Omega_m^c) \leq C/(NT)^r$ and $\mathbb{P}(\hat{\Lambda}_m^r) \leq C/(NT)^r$.

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

\[ \|h\|_N^2 = \text{vec}(H_m) \perp \Theta_{m_1m_2} \text{vec}(H_m) \quad \text{and} \quad \|h\|_p^2 = \text{vec}(H_m) \perp \Theta_{m_1m_2} \text{vec}(H_m) \quad \text{so that} \]

\[ \sup_{h \in S_{m_1} \times S_{m_2}, \|h\|_p=1} \left| \|h\|_N^2 - \|h\|_p^2 \right| = \sup_{\text{vec}(H_m) \in \mathbb{R}^{m_1m_2}, \|\Theta_{m_1m_2} \text{vec}(H_m)\|_{2,m_1m_2} = 1} \left| \text{vec}(H_m) \perp \Theta_{m_1m_2} \text{vec}(H_m) \right|
\]

\[ = \sup_{u \in \mathbb{R}^{m_1m_2}, \|u\|_{2,m_1m_2}=1} \left| u \perp \Theta_{m_1m_2} \text{vec}(H_m) \right|
\]

\[ = \|\Theta_{m_1m_2} \perp \Theta_{m_1m_2} \text{vec}(H_m) - \text{Id}_{m_1m_2} \|_{\text{op}}. \]

Therefore, (29) holds.

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

\[ \hat{\psi}_{m_2}(t) = \frac{1}{N} \sum_{i=1}^{N} S_{\psi}(X_i(t)). \]

We intend to apply Tropp’s Inequality (see Theorem 4 in Appendix) to $G - \text{Id}_{m_1m_2}$ where

\[ G = \frac{1}{N} \sum_{i=1}^{N} X_i, \quad X_i = \Theta_{m_1m_2}^{-1/2} \hat{\Theta}_{m_1m_2} \Theta_{m_1m_2}^{-1/2} \left( \frac{1}{T} \int_0^T S_{\psi}(X_i(t)) \otimes \Phi_{m_1}(t) dt \right) \Theta_{m_1m_2}^{-1/2}. \]

Note that $G = \Theta_{m_1m_2}^{-1/2} \hat{\Theta}_{m_1m_2} \Theta_{m_1m_2}^{-1/2}$ and $\mathbb{E}(G) = \text{Id}_{m_1m_2}$. Therefore

\[ \mathbb{P}(\Omega_m^c) = \mathbb{P}(\|G - \text{Id}_{m_1m_2}\|_{\text{op}} > 1/2). \]

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

\[ \mathbb{P}(\|G - \text{Id}_{m_1m_2}\|_{\text{op}} > \delta) \leq 2m_1m_2c^{-Nc(\delta)/R} \text{ where } c(\delta) = (1 + \delta) \log(1 + \delta) - \delta, \]

as $c(\delta)/(1 + \delta)^{1+\delta} \geq e^{-\delta}/(1 - \delta)^{1-\delta}$, see Cohen et al. (2019, 2013).

It remains to compute the bound $R$.

\[ \lambda_{\max}(X_i) = \|X_i\|_{\text{op}} = \sup_{x \in \mathbb{R}^{m_1m_2}, \|x\|_{2,m_1m_2}=1} x^\perp X_i x. \]
$x^\top X_i x = y^\top \left( \frac{1}{T} \int_0^T S_\psi(X_i(t)) \otimes \Phi_m(t) dt \right) y \quad \text{where} \quad y = \Theta_m^{-1/2} x$

$$= \frac{1}{T} \int_0^T \text{Tr} \left[ Y^\top \Phi_m(t) Y S_\psi(X_i(t)) \right] dt \quad \text{where} \quad y = \text{vec}(Y)$$

$$= \frac{1}{T} \int_0^T \sum_{0 \leq j' \leq m_1 - 1} \sum_{0 \leq k' \leq m_2 - 1} Y_{j,k} Y_{j',k'} \varphi_j(t) \varphi_{j'}(t) \psi_k(X_i(t)) \psi_{k'}(X_i(t)) dt$$

$$= \frac{1}{T} \int_0^T \left( \sum_{k=0}^{m_2-1} \psi_k(X_i(t)) \left( \sum_{j=0}^{m_1-1} Y_{j,k} \varphi_j(t) \right) \right)^2 dt$$

$$\leq \frac{1}{T} \int_0^T \left( \sum_{k=0}^{m_2-1} \psi_k^2(X_i(t)) \right) \left( \sum_{k=0}^{m_2-1} \left( \sum_{j=0}^{m_1-1} Y_{j,k} \varphi_j(t) \right)^2 \right) dt.$$

Therefore

$$x^\top X_i x \leq L_\psi(m_2) \sum_{k=0}^{m_2-1} \frac{1}{T} \int_0^T \left( \sum_{j=0}^{m_1-1} Y_{j,k} \varphi_j(t) \right)^2 dt = L_\psi(m_2) \|y\|_{2,m_1m_2}^2 / T.$$

Now it holds that $\|y\|_{2,m_1m_2}^2 = \|\Theta_m^{-1/2} x\|_{2,m_1m_2}^2 \leq \|\Theta_m^{-1}\|_{\text{op}} \|x\|_{2,m_1m_2}^2$. Consequently

$$\lambda_{\text{max}}(X_i) \leq L_\psi(m_2) \|\Theta_m^{-1}\|_{\text{op}} / T := R.$$

We obtain

$$\mathbb{P}(\|G - \text{Id}_{m_1m_2}\|_{\text{op}} > 1/2) \leq 2m_1m_2 \exp \left( - \frac{Nc(1/2)T}{L_\psi(m_2) \|\Theta_m^{-1}\|_{\text{op}}} \right).$$

For the proof of $\mathbb{P}(\hat{\Lambda}_m^\top) \leq C/(NT)^r$, we refer to the proof of Lemma 5 in Comte and Genon-Catalot (2020).

To study the risk of the estimator defined by $\mathbb{E}(\|\tilde{b}_m - b_1\|_N^2)$, we write

$$\|\tilde{b}_m - b_1\|_N^2 = \|\tilde{b}_m - b_1\|_N^2 1_{\Lambda_m^*} + \|b_1\|_N^2 1_{\tilde{\Lambda}_m^*}.$$

We define the orthogonal projection of $b_1$ on $S_{m_1} \times S_{m_2}$ wrt the empirical scalar product, denoted by $\Pi_m b$. We find for

$$\Pi_m b = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} a_{j,k} \varphi_j \otimes \psi_k, \quad A_m = (a_{j,k})_{0 \leq j \leq m_1 - 1, 0 \leq k \leq m_2 - 1}$$

with

$$\text{vec}(A_m) = \hat{\Theta}_m^{-1} \text{vec}(C_m)$$

where $C_m$ defined by (18). Then we have by Pythagoras theorem

$$\|\tilde{b}_m - b_1\|_N^2 = \|\tilde{b}_m - \Pi_m b\|_N^2 + \|\Pi_m b - b_1\|_N^2.$$

Thus

$$\|\tilde{b}_m - b_1\|_N^2 = \left( \|\tilde{b}_m - \Pi_m b\|_N^2 + \|\Pi_m b - b_1\|_N^2 \right) 1_{\Lambda_m^*} + \|b_1\|_N^2 1_{\tilde{\Lambda}_m^*}.$$
Then we have

\[
E(\|b_m - b\|_N^2_{\hat{\Lambda}_m}) = E(\|\Pi_mb - b\|_N^2_{\hat{\Lambda}_m}) + E(\|\hat{b}_m - \Pi_mb\|_N^2_{\hat{\Lambda}_m \cap \Omega_m}) + E(\|b\|_N^2_{\hat{\Lambda}_m}) \leq T_1 + T_2 + T_3 + T_4
\]

For the bias, we have

\[
T_1 = E(\|\Pi_mb - b\|_N^2_{\hat{\Lambda}_m}) = E\left( \inf_{h \in S_{m_1} \times \Omega_{m_2}} \|h - b\|_N^2_{\hat{\Lambda}_m} \right) \leq \inf_{h \in S_{m_1} \times \Omega_{m_2}} \|h - b\|_N^2_{\hat{\Lambda}_m}.
\]

By Lemma 1,

\[
\|\hat{b}_m - \Pi_mb\|_N^2 = \frac{1}{T} \int_0^T \text{Tr}\left( (\hat{A}_m - A_m)^T \Phi_{m_1}(t)(\hat{A}_m - A_m) \hat{\Psi}_{m_2}(t) \right) dt.
\]

Now we use (13) and its analogous for \(A_m\) and we get

\[
\|\hat{b}_m - \Pi_mb\|_N^2 = \text{Tr}\left( (\hat{A}_m - A_m)^T \int_0^T \Phi_{m_1}(t)(\hat{A}_m - A_m) \hat{\Psi}_{m_2}(t) dt \right)
= \text{Tr}\left( (\hat{A}_m - A_m)^T (\hat{C}_m - C_m) \right)
= \text{vec}(\hat{A}_m - A_m)^T \text{vec}(\hat{C}_m - C_m)
= \text{vec}(\hat{C}_m - C_m)^T \text{vec}(\hat{C}_m - C_m)
= \text{vec}(\hat{C}_m - C_m)^T \text{vec}(\hat{C}_m - C_m).
\]

On \(\Omega_m\), the eigenvalues of \(\Theta^{-1/2}_{m_1,m_2} \hat{\Theta}_{m_1,m_2}^{-1/2} \Theta^{-1/2}_{m_1,m_2}\) all belong to \([1/2, 3/2]\). Therefore, the eigenvalues of \(\Theta^{-1/2}_{m_1,m_2} \hat{\Theta}_{m_1,m_2}^{-1/2} \Theta^{-1/2}_{m_1,m_2}\) all belong to \([2/3, 2]\). So we write

\[
\text{vec}(\hat{\varepsilon}_m)^T \Theta_{m_1,m_2}^{-1} \text{vec}(\varepsilon_m) = \text{vec}(\varepsilon_m)^T \Theta_{m_1,m_2}^{-1} \Theta_{m_1,m_2}^{1/2} \hat{\Theta}_{m_1,m_2}^{-1/2} \Theta_{m_1,m_2}^{1/2} \Theta_{m_1,m_2}^{-1} \text{vec}(\varepsilon_m).
\]

This yields

\[
E(\|\hat{b}_m - \Pi_mb\|_N^2_{\hat{\Lambda}_m \cap \Omega_m}) \leq 2E(\text{vec}(\varepsilon_m)^T \Theta_{m_1,m_2}^{1/2} \text{vec}(\varepsilon_m)).
\]

Now,

\[
E(\text{vec}(\varepsilon_m)^T \Theta_{m_1,m_2}^{1/2} \text{vec}(\varepsilon_m)) = E\left[ \text{Tr}\left( \text{vec}(\varepsilon_m)^T \Theta_{m_1,m_2}^{1/2} \text{vec}(\varepsilon_m) \right) \right]
= E\left[ \text{Tr}\left( \Theta_{m_1,m_2}^{-1} \text{vec}(\varepsilon_m)^T \text{vec}(\varepsilon_m) \right) \right]
= \text{Tr}\left( \Theta_{m_1,m_2}^{-1} E\left[ \text{vec}(\varepsilon_m)^T \text{vec}(\varepsilon_m) \right] \right)
= \frac{1}{NT} \text{Tr}\left( \Theta_{m_1,m_2}^{-1} \Theta_{m_1,m_2}^{-1} \right).
\]

Thus, we get

\[
T_2 = E(\|\hat{b}_m - \Pi_mb\|_N^2_{\hat{\Lambda}_m \cap \Omega_m}) \leq \frac{2}{NT} \text{Tr}\left( \Theta_{m_1,m_2}^{-1} \Theta_{m_1,m_2}^{-1} \right).
\]

Now, using (32) and the definition of \(\hat{\Lambda}_m\), we have:

\[
T_3 = E(\|\hat{b}_m - \Pi_mb\|_N^2_{\hat{\Lambda}_m \cap \Omega_m}) \leq \frac{2\epsilon_r}{L_\psi(m_2) \log(NT)} \text{E}^{1/2}\left[ \left( \text{vec}(\varepsilon_m)^T \text{vec}(\varepsilon_m) \right)^2 \right] \frac{NT}{L_\psi(m_2) \log(NT)} \text{E}^{1/2}\left[ \left( \text{vec}(\varepsilon_m)^T \text{vec}(\varepsilon_m) \right)^2 \right] \frac{NT}{L_\psi(m_2) \log(NT)} \text{E}^{1/2}(\varepsilon_m).
\]
Lemma 4. Under the assumptions of Theorem 1, we have

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

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

$$
T_3 = \mathbb{E}(\|b_m - \Pi_m b\|_N^2 1_{\hat{\Lambda}_m^c \cap \Omega_m}) \lesssim \frac{1}{N},
$$

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

Lastly, we notice that

$$
\mathbb{E}(\|b_1\|_N^4) \leq C_4 \frac{1}{T} \mathbb{E} \int_0^T (1 + |X_i(u)|)^4 du
$$

and we get, using Lemma 3, that

$$
T_4 = \mathbb{E}(\|b_1\|_N^2 1_{\hat{\Lambda}_m^c}) \leq C_4 \frac{1}{T} \mathbb{E} \int_0^T (1 + |X_i(u)|)^4 du \mathbb{P}^{1/2}(\hat{\Lambda}_m^c) \lesssim \frac{1}{N^{r/2}}.
$$

This means $T_4 = O(1/N)$ for $r \geq 2$. Therefore, plugging this and (31)-(33)-(34) into (30) gives Inequality (24) of Theorem 1.

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

$$
\text{Tr} \left( \Theta^{-1}_{m_1 m_2} \Theta_{m_1, m_2, \sigma^2} \right) = \text{Tr} \left( \Theta^{-1/2}_{m_1 m_2} \Theta_{m_1, m_2, \sigma^2} \Theta^{-1/2}_{m_1 m_2} \right) = \mathbb{E} \left[ \text{vec}(Z)^T \Theta^{-1/2}_{m_1 m_2} \Theta_{m_1, m_2, \sigma^2} \Theta^{-1/2}_{m_1 m_2} \text{vec}(Z) \right]
$$

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 $\text{vec}(Y)^T = \text{vec}(Z)^T \Theta^{-1/2}_{m_1 m_2}$, and let us look at $\text{vec}(Y)^T \Theta_{m_1, m_2, \sigma^2} \text{vec}(Y)$. We have,

$$
\text{vec}(Y)^T \Theta_{m_1, m_2, \sigma^2} \text{vec}(Y) = \frac{1}{T} \sum_{j,k,j',k'} Y_{j,k} Y_{j',k'} \int_0^T du \int \varphi_j(u) \varphi_{j'}(u) \psi_k(x) \psi_{k'}(x) p_u(x) \sigma^2(u, x) dx
$$

$$
= \frac{1}{T} \int_0^T du \int \left( \sum_{j,k} Y_{j,k} \varphi_j(u) \psi_k(x) \right)^2 p_u(x) \sigma^2(u, x) dx
$$

$$
\leq \sigma_1^2 \text{vec}(Y)^T \Theta_{m_1, m_2} \text{vec}(Y)
$$

This yields

$$
\mathbb{E} \left[ \text{vec}(Z)^T \Theta^{-1/2}_{m_1 m_2} \Theta_{m_1, m_2, \sigma^2} \Theta^{-1/2}_{m_1 m_2} \text{vec}(Z) \right] \leq \sigma_1^2 \mathbb{E} \left[ \text{vec}(Z)^T \Theta^{-1/2}_{m_1 m_2} \Theta_{m_1, m_2, \sigma} \text{vec}(Z) \right] = \sigma_1^2 \text{Tr}(\Theta^{-1/2}_{m_1 m_2} \Theta_{m_1, m_2} \Theta^{-1/2}_{m_1 m_2}) = \sigma_1^2 m_1 m_2.
$$

This ends the proof of Theorem 1. □

8.4. Proof of Lemma 4. First we have that

$$
\text{vec}(\varepsilon_m)^T \text{vec}(\varepsilon_m) = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} \nu_N^2 (\varphi_j \circ \psi_k).
$$
Thus
\[ \mathbb{E} \left[ (\operatorname{vec}(\varepsilon_m)\operatorname{vec}(\varepsilon_m))^T \right] \leq m_1 m_2 \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} \mathbb{E} \left( \nu_N^2(\varphi_j \otimes \psi_k) \right). \]

Now using the Burkholder-Davies-Gundy inequality
\[
\mathbb{E} \left( \nu_N^4(\varphi_j \otimes \psi_k) \right) = \frac{1}{(NT)^4} \mathbb{E} \left[ \left( \int_0^T \sum_{i=1}^N (\varphi_j \otimes \psi_k(u, X_i(u)) \sigma(u, X_i(u))dW_i(u) \right)^4 \right]
\lesssim \frac{1}{(NT)^4} \mathbb{E} \left[ \left( \int_0^T \sum_{i=1}^N (\varphi_j \otimes \psi_k(u, X_i(u)))^2 \sigma^2(u, X_i(u))du \right)^2 \right]
\lesssim \frac{1}{N^2T^2} \frac{1}{T} \int_0^T \mathbb{E} \left[ (\varphi_j \otimes \psi_k(u, X_1(u)))^4 \sigma^4(u, X_1(u))du \right]
\]
Therefore as \( \sum_{j=0}^{m_1-1} \varphi_j^4(u) \leq (L_\varphi(m_1))^2 \) and \( \sum_{k=0}^{m_2-1} \psi_k^4(x) \leq (L_\psi(m_2))^2 \)
\[
\mathbb{E} \left[ (\operatorname{vec}(\varepsilon_m)\operatorname{vec}(\varepsilon_m))^T \right] \leq m_1 m_2 (L_\varphi(m_1)L_\psi(m_2))^2 \frac{1}{T} \int_0^T \sigma^4(u, X_1(u))du.
\]
This is the announced result. □

8.5. **Proof of Lemma 2.** Let \( H_m = (h, j, k)_{0 \leq j \leq m_1-1, 0 \leq k \leq m_2-1}, \) then as by Lemma 1
\[
[\operatorname{vec}(H_m)]^\dagger \Theta_{m_1 m_2} \operatorname{vec}(H_m) = \frac{1}{T} \int_\epsilon^T \int_A h^2(t, u)p_t(u)dudt
\geq c_0(\epsilon, A) \frac{1}{T} \int_\epsilon^T \int_A h^2(t, u)dudt = c_0(\epsilon, A)[\operatorname{vec}(H_m)]^\dagger \operatorname{vec}(H_m)
\]
where \( h(t, u) = \sum_j h_j,k \varphi_j(t) \psi_k(u) \). This proves that any eigenvalue of \( \Theta_{m_1 m_2} \) is larger than \( c_0(\epsilon, A) \). The upper bound for \( ||h||_p^2 \) is straightforward. This gives the result of Lemma 2. □

8.6. **Proof of Proposition 2.** In the bound (27), we look at the bias term:
\[
\inf_{h \in S_{m_1} \times S_{m_2}} ||h - b_L||^2 = ||b_m - b_L||^2
\]
where \( b_m = \sum_{j=0}^{m_1-1} \sum_{k=0}^{m_2-1} \langle b, \varphi_j \otimes \psi_k \rangle \varphi_j \otimes \psi_k \) is the \( L^2 \)-orthogonal projection of \( b_L \) on \( S_{m_1} \times S_{m_2} \). Therefore, if \( b_L \in W^*(\beta, R), \)
\[
||b_m - b_L||^2 \leq R_1^2 m_1^{-2\beta_1} + R_2^2 m_2^{-2\beta_2}.
\]
Making the standard compromise with the variance term of order \( m_1 m_2/(NT) \) gives the rate. □

8.7. **Proof of Theorem 3.** We follow the scheme of Theorem 2.11 in Tsybakov (2009). Take \( g \) and \( h \) two regular functions with support \([0, 1]\), bounded by \( K_g \) and \( K_g \) respectively, with \( g \) \( \beta_1 \)-times derivable and \( h \) \( \beta_2 \) times derivable, with square integrable derivatives. Define for \( j = 0, \ldots, M_1 - 1, \) and \( k = 0, \ldots, M_2 - 1, \) with \( A = [a, b], a < b, \)
\[
g_j(t) = \sqrt{\frac{M_1}{T - \epsilon}} g \left( \frac{M_1(t - \epsilon}{T - \epsilon} - j \right), \quad h_k(x) = \sqrt{\frac{M_2}{b - a}} h \left( \frac{M_2(x - a}{b - a} - k \right).
\]
Clearly, the \( g_j \) have disjoint supports and \( g_j g_{j'} = 0 \) for \( j \neq j' \), and for the same reason \( h_k h_{k'} = 0 \) for \( k \neq k' \). 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 $\theta = (\theta_{j,k})_{j=0,\ldots,M_1-1, k=0,\ldots,M_2-1}$ with $\theta_{j,k} \in \{0, 1\}$ for all $j = 0, \ldots, M_1 - 1, k = 0, \ldots, M_2 - 1$,

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

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

- As $g$ is $\beta_1$-times derivable and $h$ is $\beta_2$ times derivable, both with square integrable derivatives, we get that

$$\int \int (\partial^{\beta_1} b_\theta(t, x)/\partial t^{\beta_1})^2 dt dx + (\partial^{\beta_2} b_\theta(t, x)/\partial x^{\beta_2})^2 dt dx$$

$$= \frac{\delta^2}{NT} \sum_{j=0}^{M_1-1} \sum_{k=0}^{M_2-1} \theta^2_{j,k} \left\{ \left( \frac{M_1}{T - \epsilon} \right)^{2\beta_1} \int [g^{(\beta_1)}(t)]^2 dt \int h_k^2(x) dx \right.$$  
$$\left. + \left( \frac{M_2}{b - a} \right)^{2\beta_2} \int g_j^2(t) dt \int [h^{(\beta_2)}(x)]^2 dx \right\}$$

$$\leq \frac{\delta^2}{NT} \left[ \frac{\int [g^{(\beta_1)}(t)]^2 dt \int h^2(x) dx}{(T - \epsilon)^{2\beta_1}} M_1^{2\beta_1 + 1} M_2 + \frac{\int g^2(t) dt \int [h^{(\beta_1)}(x)]^2 dx}{(b - a)^{2\beta_2}} M_1 M_2^{2\beta_2 + 1} \right].$$

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

$$\int \int (\partial^{\beta_1} b_\theta(t, x)/\partial t^{\beta_1})^2 dt dx + (\partial^{\beta_2} b_\theta(t, x)/\partial x^{\beta_2})^2 dt dx \leq C^2 \delta^2 \leq L^2$$

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

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

This implies that $b_0$ and $b_0'$ belong to $W(\beta, L)$ with $L = (L, L)$.

- We have that:

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

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

As a consequence, the Varshamov-Gilbert Lemma (see Lemma 2.9 in Tsybakov (2009)) ensures that for $M := M_1 M_2 \geq 8$, there exist $Q \geq 2^{M/8}$ elements say $\{\theta^0, \ldots, \theta^Q\}$ of $\{0, 1\}^M$ such that $\rho(\theta^q, \theta^{q'}) \geq M/8$ for all $0 \leq q < q' \leq Q$, with $\theta^0 = (0, \ldots, 0)$. This leads to:

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

- Lastly, let $P_\theta$ (resp $P_0$) 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 $\sigma$-field $C_T = \sigma(\mathcal{X}(t), t \in [0, T])$ where $(\mathcal{X}(t), t \in [0, T])$ is the canonical process of $C_T$, i.e. $\mathcal{X}_t(x) = x(t)$ for $x \in C_T$. We bound

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

$$
\log \frac{d\mathbb{P}_\theta}{d\mathbb{P}_0} = \int_0^T \frac{b_0(t, X(t))}{\sigma^2(t, X(t))} dX(t) - \frac{1}{2} \int_0^T \frac{b_0^2(t, X(t))}{\sigma^2(t, X(t))} dt.
$$

(see e.g. Liptser and Shiryaev (2001)). Under $\mathbb{P}_\theta$, $dX(t) = b_0(t, X(t))dt + \sigma(t, X(t))dB(t)$ where $(B(t), t \in [0, T])$ is a standard Brownian motion. Therefore,

$$
\log \frac{d\mathbb{P}_\theta}{d\mathbb{P}_0} = \frac{1}{2} \int_0^T \frac{b_0^2(t, X(t))}{\sigma^2(t, X(t))} dt + \int_0^T b_0(t, X(t)) dB(t).
$$

Thus,

$$
K(\mathbb{P}_\theta, \mathbb{P}_0) = \mathbb{E}_{\mathbb{P}_0} \int_0^T \frac{b_0^2(t, X(t))}{\sigma^2(t, X(t))} dt \leq \frac{1}{2\sigma_0^2} \int_\mathbb{R} \int \frac{1}{2} \frac{b_0^2(t, x)}{\sigma^2(t, X(t))} dx dt.
$$

Then, we have

$$
\int_\mathbb{R} \int \frac{b_0^2(t, x)}{\sigma^2(t, X(t))} dx dt = \frac{\delta^2}{NT} \int_\mathbb{R} \int \sum_{j=0}^{M_1-1} \sum_{k=1}^{M_2-1} \theta_{j,k}^2 g_j^2(t, \delta^2) h_k^2(x, \delta^2) p_0^2(x) dx dt
$$

$$
= \frac{\delta^2 M_1 M_2}{NT} \frac{K_2^2 K_g^2}{(T - \epsilon)(b - a)} \int_\mathbb{R} \int \int \sum_{j=0}^{M_1-1} \sum_{k=1}^{M_2-1} \theta_{j,k}^2 g_j^2(t, \delta^2) h_k^2(x, \delta^2) p_0^2(x) dx dt
$$

$$
\leq \frac{\delta^2 M_1 M_2}{NT} \frac{K_2^2 K_g^2}{(T - \epsilon)(b - a)} \int_\mathbb{R} \int \int \sum_{j=0}^{M_1-1} \sum_{k=1}^{M_2-1} \theta_{j,k}^2 g_j^2(t, \delta^2) h_k^2(x, \delta^2) p_0^2(x) dx dt
$$

as $\int p_0^2(x) dx = 1$ and $(T - \epsilon)/T \leq 1$. As a consequence, we get

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

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

$$
\frac{1}{Q} \sum_{q=1}^{Q} K(\mathbb{P}_\theta^{\otimes N}, \mathbb{P}_0^{\otimes N}) \leq \kappa \log(Q).
$$

We apply Theorem 2.7 in Tsybakov (2009) and (35) and (36) imply the lower bound result. \hfill $\square$

8.8. **Proof of Proposition 3.** Assume that $f \in W^{(\text{per})}(B, L)$. 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 $(\varphi_\epsilon, \psi_k)$:

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

and we set $s_{j_1, k}(0) = c_{j_1, k}$.

We have

$$
s_{0, k}(t, x) = \frac{1}{\sqrt{T - \epsilon}} \int_a^b \psi_k(x) dx \left( [\partial^{j_1-1} f(t, \epsilon)/(\partial t^{j_1-1})(\epsilon, x) - (\partial^{j_1-1} f(t, \epsilon)/(\partial t^{j_1-1}))(T, x)] = 0. \right.
$$
Integrating by parts w.r.t. $t$ under the integral, we get for $\ell \geq 1$
\[
s_{2\ell-1,k}(\beta_1) = \frac{2}{\sqrt{T-\epsilon}} \int_a^b \psi_k(x)dx \int_\epsilon^T \frac{\partial^{\beta_1-1} f(t,x)}{\partial t^{\beta_1-1}}(t,x) \frac{2\pi \ell}{T-\epsilon} \sin(2\pi \ell t - \epsilon)dt
= \frac{2\pi \ell}{T-\epsilon} s_{2\ell,k}(\beta_1 - 1).
\]
Analogously, $s_{2\ell,k}(\beta_1) = -\frac{2\pi \ell}{T-\epsilon} s_{2\ell-1,k}(\beta_1 - 1)$. This yields:
\[
s_{2\ell-1,k}(\beta_1) + s_{2\ell,k}(\beta_1) = \left[\frac{2\pi \ell}{T-\epsilon}\right]^2 [s_{2\ell-1,k}(\beta_1 - 1) + s_{2\ell,k}(\beta_1 - 1)].
\]
By induction,
\[
\sum_{\ell \geq 1} \left[\frac{2\pi \ell}{T-\epsilon}\right]^{2\beta_1} [s_{2\ell-1,k}(0) + s_{2\ell,k}(0)] = \left[\frac{\pi}{T-\epsilon}\right]^{2\beta_1} \sum_{\ell \geq 1} a^2_\ell(\beta_1) s_{\ell,k}(0),
\]
with $a_\ell(\beta_1) = \lambda_\ell^{\beta_1}$ if $\ell$ is even, and $(\ell + 1)^{2\beta_1}$ if $\ell$ is odd. We deduce
\[
\sum_{\ell,k} \int_\epsilon^T \int_a^b \left(\frac{\partial^{\beta_1} f(t,x)}{\partial t^{\beta_1}}(t,x)\right)^2 \psi_k(x) \varphi_{\ell}(t)dt dx = \left[\frac{\pi}{T-\epsilon}\right]^{2\beta_1} \sum_{\ell,k} a^2_\ell(\beta_1) c^2_{\ell,k}.
\]
Hence,
\[
\int_\epsilon^T \int_a^b \left(\frac{\partial^{\beta_1} f(t,x)}{\partial t^{\beta_1}}(t,x)\right)^2 dt dx = \left[\frac{\pi}{T-\epsilon}\right]^{2\beta_1} \sum_{\ell,k} a^2_\ell(\beta_1) c^2_{\ell,k}.
\]
Therefore,
\[
\sum_{\ell,k} \ell^{2\beta_1} c^2_{\ell,k} \leq L^2_1(T-\epsilon)^{2\beta_1}/\pi^{2\beta_1} = R_1.
\]
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 $f \in W^*(\beta, \mathbb{R})$ as announced. □

References


9. Appendix

9.1. Some properties of Kronecker products (wikipedia). Recall that $M^\perp$ 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 $M = (M_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n}$, by

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

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

$$\text{vec}(M) = (M_{1,1}, \ldots, M_{1,n}, M_{2,1}, \ldots, M_{2,n}, \ldots, M_{m,1}, \ldots, M_{m,n})^\perp.$$

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

$$\text{vec}(MNR) = (R^\perp \otimes M)\text{vec}(N)$$

$$\text{(38) } (M \otimes N)^\perp = M^\perp \otimes N^\perp.$$

As soon as the product $M^\perp N$ is well defined and a square matrix:

$$\text{Tr}(M^\perp N) = \text{vec}(M)^\perp \text{vec}(N).$$

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

$$\text{vec}(M) (M \otimes N)^{-1} = (M^{-1} \otimes N^{-1}).$$

9.2. Tropp’s inequality.

Theorem 4. (Matrix Chernoff, Tropp (2012)) Consider a finite sequence $\{X_k\}$ of independent, random, self-adjoint matrices with dimension $d$. Assume that each random matrix satisfies

$$X_k \succeq 0 \text{ and } \lambda_{\max}(X_k) \leq R \text{ almost surely}.$$

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

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

$$\mathbb{P}\left\{ \lambda_{\max}\left(\sum_k X_k\right) \geq (1+\delta)\mu_{\max}\right\} \leq d \left[ \frac{e^\delta}{(1+\delta)^{1+\delta}} \right]^{\mu_{\max}/R} \text{ for } \delta \geq 0.$$
9.3. **The Hermite basis.** The Hermite polynomial of order \( j \) is given, for \( j \geq 0 \), 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 (1964)). The Hermite function of order \( j \) is given by:

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

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

9.4. **The half-trigonometric basis.** The trigonometric basis is well-fitted for functions satisfying border conditions as described in \( W^{\text{per}} \). This is why we rather used the so-called "half-trigonometric" system, namely the cosine basis defined by

\[
\varphi_{0,T}(x) = \sqrt{\frac{1}{T}} 1_{[0,T]}(t), \quad \varphi_{j,T}(t) = \sqrt{\frac{2}{T}} \cos(\pi j t/T) 1_{[0,T]}(t), \quad j = 1, \ldots, m - 1,
\]

see Efroymovich (1999, p.46). It is clearly an orthonormal basis, which is easy to handle and still has good approximation properties, see Efroymovich (1999, p.32).