NONPARAMETRIC ESTIMATION FOR I.I.D. STOCHASTIC DIFFERENTIAL EQUATIONS WITH SPACE-TIME DEPENDENT COEFFICIENTS

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

Abstract. We consider N *i.i.d.* one-dimensional inhomogeneous diffusion processes $(X_i(t), i = 1, \dots, N)$ with drift $\mu(t,x) = \sum_{j=1}^{K} \alpha_j(t) g_j(x)$ and diffusion coefficient $\sigma(t,x)$, where K, the functions $g_j(x)$ and $\sigma(t,x)$ are known. Our concern is the nonparametric estimation of the K-dimensional unknown function $(\alpha_j(t), j = 1, \dots, k)$ from the continuous observation of the sample paths $(X_i(t))$ throughout a fixed time interval $[0,\tau]$. A collection of projection estimators belonging to a product of finite-dimensional subspaces of $\mathbb{L}^2([0,\tau])$ is built. The $\mathbb{L}^2$-risk is defined by the expectation of either an empirical norm or a deterministic norm fitted to the problem. Rates of convergence for large N are discussed. A data-driven choice of the dimensions of the projection spaces is proposed. The theoretical results are illustrated by numerical experiments on simulated data.

Keywords and phrases: Adaptive estimation. Continuous observation. Inhomogeneous diffusions. Least squares estimator. Nonparametric drift estimation. Projection method. Stochastic differential equations.

June 23, 2023

1. Introduction

In this paper, we consider N independent and identically distributed (*i.i.d.*) processes $(X_i(t))_{1 \le i \le N}$ given by the inhomogeneous stochastic differential equation (SDE)

$$dX_i(t) = \mu(t, X_i(t))dt + \sigma(t, X_i(t))dW_i(t), \text{ with } \mu(t,x) := \sum_{k=1}^{K} \alpha_k(t) g_k(x), \tag{1}$$

with $X_i(0) = \eta_i, i = 1, \dots, N$. The integer K, the deterministic functions $x \mapsto g_k(x)$, $k = 1, \dots, K$ and $(t,x) \mapsto \sigma(t,x)$ are known, $W_i, i = 1 \dots, N$ are N independent Brownian motions, $\eta_i, i = 1, \dots, N$ are *i.i.d.* random variables, independent of $(W_i, i = 1 \dots, N)$. The functions $\alpha_1(t), \dots, \alpha_K(t)$ are deterministic and unknown.
The aim of the paper is the nonparametric estimation of the K-dimensional function $t \in [0, +\infty) \mapsto (\alpha_j(t), j = 1, \dots, k) \in \mathbb{R}^K$ from the continuous observation of the N sample paths thoughout a fixed time interval $[0,\tau]$. The asymptotic framework is $N \to +\infty$.
Inference and especially nonparametric drift estimation for diffusion processes is a well developped topic. Generally authors consider one trajectory, continuously or discretely observed on a time interval $[0,\tau]$. Statistical results are obtained by means of an asymptotic framework: either τ is fixed and the diffusion coefficient tends to 0, or τ tends to infinity. In the former case, space and time dependent coefficients may be considered. In the latter case, ergodicity assumptions

(1): Université de Paris, MAP5, UMR 8145 CNRS,F-75006, FRANCE,
email: fabienne.comte@parisdescartes.fr,
valentine.genon-catalot@mi.parisdescartes.fr.

are required and generally only authorize homogeneous diffusions, see *e.g.* Kutoyants (1984, 2004), Iacus (2008), Kessler *et al.* (2012), Dalalyan and Reiss (2006, 2007), Comte *et al.* (2007), Hoffmann (1999), Strauch (2018), Gloter and Sorensen (2009).

More recently, the interest in inference for *i.i.d.* paths of SDEs has begun to grow. This problem belongs to functional data analysis, *i.e.* analysis of samples of infinite dimensional data (see *e.g.* Ramsay and Silvermann (2007), Wang *et al.* (2016)). Panel or longitudinal data analysis are another name for the study of data collected over time from a sample of individuals (see *e.g.* Hsiao (2003)). Among recent results on nonparametric drift estimation for *i.i.d.* samples of SDEs, one may quote Comte and Genon-Catalot (2020b), Denis *et al* (2020, 2021), Marie and Rosier (2023). See also Comte and Marie (2023) for identically distributed diffusions with correlated Brownian motions. All these papers consider homogeneous diffusions, for which the drift and diffusion coefficients do not depend on time but only on space.

Space-time dependent drifts are considered though, in recent papers dealing with interacting particle systems or their mean field limits. When the coefficients do not depend on the empirical distribution of $(X_i(t), i = 1, \ldots, N)$, *i.e.* when there is no interaction between particles, these models reduce to *i.i.d.* diffusion processes. For instance, Della Maestra and Hoffmann (2022) study a pointwise kernel estimator of a general drift term $\mu(t, x)$. In Comte and Genon-Catalot (2023), an Ornstein-Uhlenbeck interacting particle system with time dependent coefficients is investigated. This study contains, as a particular case, the model $dX_i(t) = \alpha(t)X_i(t)dt + dW_i(t), i = 1, \ldots, N$ and the non parametric estimation of the function $\alpha(t)$ by projection method with data-driven choice of the dimension of the projection space is studied.

In this paper, we extend this case to the general model (1). For $\mathbf{m} = (m_1, \ldots, m_K) \in \mathbb{N}^K$, we consider $S_{\mathbf{m}} = S_{m_1} \times \ldots \times S_{m_K}$ a product of finite-dimensional subspaces of $(\mathbb{L}^2([0, \tau]))^K$ with respective dimensions m_j. We define, for each $\mathbf{m}$, a projection estimator $\widetilde{\mathbf{a}}_{\mathbf{m}}(t) = (\widetilde{\alpha}_j(t), j = 1, \ldots, K)^T$ (^{T} denotes the transpose of the vector) obtained by minimizing a global projection contrast inspired by the log-likelihood of the N processes $(X_i(t), t \in [0, \tau], i = 1, \ldots, N)$. The risk of the estimators is evaluated by the expectation of either the square of an empirical norm or the square of a deterministic norm linked with the projection contrast defined as follows. We introduce for $N \geq 1$ and $t \geq 0$, the $K \times K$ nonnegative symmetric matrices $S_N(t), S(t)$ given by:

$$(2) \quad S_N(t) = \left(\frac{1}{N} \sum_{i=1}^{N} g_j(X_i(t)) g_k(X_i(t)) \right)_{1 \leq j,k \leq K}, \quad S(t) = \left(\mathbb{E}\left[g_j(X_1(t)) g_k(X_1(t)) \right] \right)_{1 \leq j,k \leq K}.$$

For $\mathbf{h} = (h_1, \ldots, h_K)^T \in (\mathbb{L}^2([0, \tau]))^K$, we set

$$(3) \qquad \|\mathbf{h}\|_N^2 = \int_0^\tau \mathbf{h}(t)^T S_N(t) \mathbf{h}(t) dt, \quad \|\mathbf{h}\|_\tau^2 = \int_0^\tau \mathbf{h}(t)^T S(t) \mathbf{h}(t) dt.$$

Under the identifiability assumption that, for all t, the matrices (2) are invertible, $\|.\|_N$ (resp. $\|.\|_\tau$) is a random (resp. deterministic) norm on $(\mathbb{L}^2([0, \tau]))^K$. To bound the estimators risks that is defined as the expectation of these square norms, the key tool is to study the set where the empirical norm $\|\mathbf{h}\|_N$ and the deterministic norm $\|\mathbf{h}\|_\tau$ are equivalent for elements of the space $S_{\mathbf{m}}$. Actually, we are able to compare these norms for all functions of $(\mathbb{L}^2([0, \tau]))^K$. Indeed, we prove that on the set

$$(4) \qquad \mathcal{O}_N = \left\{ \sup_{t \in [0, \tau]} \|S(t)^{-1/2} S_N(t) S(t)^{-1/2} - \mathrm{Id}_K\|_{\mathrm{op}} \leq \frac{1}{2} \right\},$$

for all $\mathbf{h} \in (\mathbb{L}^2([0,\tau]))^K$, $(1/2)\|\mathbf{h}\|_\tau^2 \leq \|\mathbf{h}\|_N^2 \leq (3/2)\|\mathbf{h}\|_\tau^2$ (for a symmetric matrix M, the norm $\|M\|_{\rm op}$ is the supremum of the absolute values of its eigenvalues). By means of the Garsia-Rodemacher-Rumsay (GRR) Lemma as stated in Jourdain and Pagès (2022), we prove that $\mathbb{P}(\mathcal{O}_N^c) \lesssim N^{-p}$ for all $p > 1$ ($\lesssim$ means $\leq$ up to a constant).
After the study of the estimators for fixed $\mathbf{m}$, a data-driven choice of $\mathbf{m}$ is proposed where, for the sake of simplicity, $\sigma(t,x)$ is assumed to be uniformly bounded. The obtained estimator is adaptive, in the sense that it reaches an automatic squared bias-variance compromise.
In Section 2, assumptions and some preliminary results are given. In Section 3, the minimum contrast estimators are defined. Their risks are given in Theorem 2, and the risk bounds show an explicit and clear variance term. We also discuss the rates of convergence: our method estimates all functions simultaneously and the corresponding rate is the estimation rate of one function with regularity associated to the smallest regularity of the K functions. It is interesting to note that the additive drift structure guards against the curse of dimensionality.

Section 3.5 is devoted to the data-driven procedure. Section 4 presents numerical results on simulated data for various examples of models and several orthonormal bases for the projection spaces. Section 5 gives some concluding remarks. In Appendix (Section 7), the GRR Lemma and some useful results on matrices are recalled and examples of orthonormal bases are given.

2. Notation, assumptions and preliminary results.

Notation. For M a matrix, we denote by M^T the transpose of M, by $\mathrm{Tr}(M)$ the trace of M and by $\|M\|_{\rm op}$ the operator norm of M that is the square root of the largest eigenvalue of MM^T. If M is symmetric, $\|M\|_{\rm op} = \sup\{|\lambda_i|\}$ where λ_i are the eigenvalues of M. If, in addition, M is invertible, $\|M^{-1}\|_{\rm op} = \|M\|_{\rm op}^{-1}$.
For $h \in \mathbb{L}_\tau^2 = \mathbb{L}^2([0,\tau])$, we denote by $\|h\| = (\int_0^\tau h^2(t)dt)^{1/2}$ its $\mathbb{L}^2$-norm and $\|\mathbf{x}\|_{2,r}$ denotes the Euclidian norm of the vector $\mathbf{x} = (x_1,\dots,x_r)^T$ of $\mathbb{R}^r$. For $\mathbf{h}(t) = (h_1(t), h_2(t), \dots, h_K(t))^T$ and and $\mathbf{h}^\star(t) = (h_1^\star(t),\dots,h_K^\star(t))^T$ elements of $\mathbb{L}_\tau^2 \times \cdots \times \mathbb{L}_\tau^2$, we set $\|\mathbf{h}\| = (\sum_{k=1}^K \int_0^\tau h_k^2(t)dt)^{1/2}$ and $\langle \mathbf{h}, \mathbf{h}^\star\rangle = \sum_{k=1}^K \int_0^\tau h_k(t)h_k^\star(t)dt$ for respectively the $\mathbb{L}^2$-norm and the scalar product of $\mathbb{L}_\tau^2 \times \cdots \times \mathbb{L}_\tau^2$.

2.1. Assumptions.

We consider *i.i.d.* processes $(X_i(t), t \geq 0, i = 1,\dots,N)$ where $X_i(t)$ is solution of (1) with *i.i.d.* $X_i(0) = \eta_i, i = 1\dots,N$ and *i.i.d.* standard Brownian motions $(W_i(t), t \geq 0)$, $i = 1,\dots,N$, independent of the initial conditions.

We set the following assumptions:
[H1] (i) The functions g_k are Lipschitz with constant L:

$$\forall k = 1,\dots K,\ \exists L > 0, \forall x, y \in \mathbb{R}, |g_k(x) - g_k(y)| \leq L|x-y|.$$

(ii)The function σ is C^1 on $\mathbb{R}^+ \times \mathbb{R}$ and has linear growth w.r.t. x:

$$\forall t > 0,\ \exists C_t > 0,\ \forall s \in [0,t], \forall x \in \mathbb{R},\ |\sigma(s,x)| \leq C_t(1+|x|),$$

where $t \mapsto C_t$ is a continuous function.
(iii) the *i.i.d.* variables η_i have moments of any order.

[H2] The functions $\alpha_k(t) : \mathbb{R}^+ \to \mathbb{R}$, $k = 1,\dots,K$ are continuous on $\mathbb{R}^+$ (and thus belong to $\mathbb{L}_\tau^2$).

[H3] $\forall t \in [0,\tau]$, the matrix $S(t)$ is invertible and $\forall t \in [0,\tau]$, $\forall N \geq 1$, $S_N(t)$ is a.s. invertible (see (2)).

Assumptions **[H1]** and **[H2]** ensure that equation (1) admits a unique strong solution. Under **[H1]-[H2]**, the functions g_j and $x \mapsto \mu(t,x)$ have linear growth:

$$\forall x \in \mathbb{R},\ |g_j(x)| \quad \leq \quad \widetilde{L}(1+|x|), \quad \widetilde{L} = \max\{L, \max_{1\leq j\leq K} |g_j(0)|\}, \tag{5}$$

$$\forall t \in [0,\tau], \forall x \in \mathbb{R},\ |\mu(t,x)| \quad \leq \quad L(\tau)(1+|x|), \quad L(\tau) = K\widetilde{L} \sup_{t\in[0,\tau]} \sup_{1\leq k\leq K} |\alpha_k(t)|. \tag{6}$$

Assumption **[H3]** is an identifiability assumption allowing to estimate the K functions $(\alpha_i(t), i = 1\dots,K)$. For instance, for $K=2$, $\det(S(t)) = \mathbb{E}[g_1^2(X_1(t))]\mathbb{E}[g_2^2(X_1(t))] - \{\mathbb{E}[g_1(X_1(t))g_2(X_1(t))]\}^2$ is nonzero if and only if $g_1(X_1(t))$ is not proportional to $g_2(X_1(t))$ almost surely.
As $S_N(t)$ converges *a.s.* to $S(t)$ as N tends to infinity, if $S(t)$ is invertible, $S_N(t)$ is invertible for N large enough.

The following bounds on the moments of the process are classical and useful in the sequel.

Proposition 1. *Under Assumptions* **[H1]-[H2]**, *for all* $p \geq 0$*:*

$$\mathbb{E}\left[\sup_{t\in[0,\tau]} |X_1(t)|^p\right] < +\infty. \tag{7}$$

For all $r \geq 1$, *there exists a positive constant* $\mathfrak{B}(r,\tau)$ *such that* $\forall s,t \in [0,\tau]$, *with* $|t-s| \leq 1$,

$$\mathbb{E}(|X_1(t) - X_1(s)|^{2r}) \leq \mathfrak{B}(r,\tau)|t-s|^r. \tag{8}$$

Moreover, for $\mathfrak{g} = g_j g_k$, *where* $j,k \in \{1,\dots,K\}$, $\forall r \geq 1$, *there exists a positive constant* $\mathfrak{C}(r,\tau)$ *such that such that* $\forall s,t \in [0,\tau]$, *with* $|t-s| \leq 1$,

$$\mathbb{E}(|\mathfrak{g}(X_1(t)) - \mathfrak{g}(X_1(s))|^{2r}) \leq \mathfrak{C}(r,\tau)|t-s|^r. \tag{9}$$

Details about $\mathfrak{B}(r,\tau)$ and $\mathfrak{C}(r,\tau)$ can be found in the proof of Proposition 1.

2.2. **Different norms in the problem.** Note that, if we set

$$\mathbf{g}(x) = (g_1(x),\dots,g_K(x))^T \text{ and } \quad S_{\mathbf{g}}(x) = \mathbf{g}(x)\,\mathbf{g}(x)^T, \tag{10}$$

we have (recall definition (2))

$$S_N(t) = \frac{1}{N}\sum_{i=1}^N S_{\mathbf{g}}(X_i(t)), \quad S(t) = \mathbb{E}[S_N(t)] = \mathbb{E}[S_{\mathbf{g}}(X_1(t))]. \tag{11}$$

By **[H3]**, the matrices $S_N(t)$ and $S(t)$ are symmetric positive definite. For any $\mathbf{x} = (x_1,\dots,x_K)^T \in \mathbb{R}^K$

$$\mathbf{x}^T S(t)\mathbf{x} = \mathbb{E}\left[\left(\sum_{j=1}^K x_j g_j(X_1(t))\right)^2\right] \geq 0, \quad \mathbf{x}^T S_N(t)\mathbf{x} = \frac{1}{N}\sum_{i=1}^N \left[\left(\sum_{j=1}^K x_j g_j(X_i(t))\right)^2\right] \geq 0.$$

For all $t \geq 0$, $\mathbf{x} \in \mathbb{R}^K \to \mathbf{x}^T S(t)\mathbf{x}$ defines a norm:

$$\|\mathbf{x}\|_{S(t)} := \mathbf{x}^T S(t)\mathbf{x}$$

as $\|\mathbf{x}\|_{S(t)} \geq 0$, $\neq 0$ if and only if $\mathbf{x} \neq 0$. For $\mathbf{x} = (x_1,\dots,x_K)$, $\mathbf{x}^\star = (x_1^\star,\dots,x_K^\star)$, we denote by $\langle \mathbf{x}, \mathbf{x}^\star\rangle_{S(t)} = \mathbf{x}^T\, S(t)\mathbf{x}^\star$ the scalar product associated with the norm $\|\mathbf{x}\|_{S(t)}$.
The empirical version of the norm $\|.\|_{S(t)}$ is given, for $\mathbf{x} \in \mathbb{R}^K$, by

$$\|\mathbf{x}\|^2_{S_N(t)} = \mathbf{x}^T S_N(t)\mathbf{x},$$

with associated scalar product $\langle \mathbf{x}, \mathbf{x}^\star \rangle_{S_N(t)} = \mathbf{x}^T S_N(t)\mathbf{x}^\star$. We have $\mathbb{E}(\|\mathbf{x}\|^2_{S_N(t)}) = \|\mathbf{x}\|^2_{S(t)}$. Lastly, for functions $\mathbf{h} = (h_1, \dots, h_K)$ and $\mathbf{h}^\star = (h_1^\star, \dots, h_K^\star)$ with $h_i, h_i^\star$, $i = 1, \dots, K$ in $\mathbb{L}^2_\tau$, we have (see (3))

$$\|\mathbf{h}\|^2_N = \int_0^\tau \|\mathbf{h}(t)\|^2_{S_N(t)} dt, \quad \langle \mathbf{h}, \mathbf{h}^\star \rangle_N = \int_0^\tau \langle \mathbf{h}(t), \mathbf{h}^\star(t) \rangle_{S_N(t)} dt.$$

Now, $\|\mathbf{h}\|^2_N = 0$ implies that $\|\mathbf{h}(t)\|^2_{S_N(t)} = 0$, *a.e.* on $[0, \tau]$ and thus by **[H3]**, $\mathbf{h}(t) = 0$ in $(\mathbb{L}^2_\tau)^K$. Therefore, $\|.\|_N$ is a norm and $\langle ., . \rangle_N$ a scalar product on $(\mathbb{L}^2_\tau)^K$.
Analogously,

$$\|\mathbf{h}\|^2_\tau := \int_0^\tau \|\mathbf{h}(t)\|^2_{S(t)} dt = \mathbb{E}(\|\mathbf{h}\|^2_N), \quad \langle \mathbf{h}, \mathbf{h}^\star \rangle_\tau = \int_0^\tau \langle \mathbf{h}(t), \mathbf{h}^\star(t) \rangle_{S(t)} dt$$

are respectively a square norm and a scalar product on $(\mathbb{L}^2_\tau)^K$.

As a consequence, three norms are to handle in the problem for a function $\mathbf{h} = (h_1, \dots, h_K)$, the standard $\mathbb{L}^2$-norm on $[0, \tau]$, defined by $\|\mathbf{h}\|^2 = \sum_{i=1}^K \int_0^\tau h_i^2(t) dt$, the $\mathbb{L}^2_\tau$-norm $\|\mathbf{h}\|^2_\tau$ and the empirical norm $\|\mathbf{h}\|^2_N$. The compactness of $[0, \tau]$ and our assumptions allow to compare them.

First, the norm $\|.\|_\tau$ can be compared to the $\mathbb{L}^2$-norm as stated now.

Proposition 2. *Under* **[H1]-[H2]**, $\forall \mathbf{h} \in (\mathbb{L}^2)^K$, $\|\mathbf{h}\|_\tau \leq K G^2 \|\mathbf{h}\|^2$, *where* $\|\mathbf{h}\|^2 = \sum_{j=1}^K \int_0^\tau h_j^2(t) dt$ *and*

$$G^2 := \max_{j=1,\dots,K} \sup_{t \in [0.\tau]} \mathbb{E}[g_j^2(X_1(t))].$$

Proof of Proposition 2. It holds that

$$\|\mathbf{h}\|^2_\tau = \int_0^\tau \mathbf{h}(t)^T S(t) \mathbf{h}(t) dt \leq \sup_{t \in [0, \tau]} \|S(t)\|_{\text{op}} \|\mathbf{h}\|^2.$$

As $\|S(t)\|_{\text{op}} \leq \text{Tr}(S(t)) = \sum_{j=1}^K \mathbb{E}[g_j^2(X_1(t))] \leq K G^2$, we get the result. □

For the link between the empirical and the $\mathbb{L}^2_\tau$ norms, recall the event:

$$\mathcal{O}_N = \left\{ \sup_{t \in [0, \tau]} \|S(t)^{-1/2} S_N(t) S(t)^{-1/2} - \text{Id}_K\|_{\text{op}} \leq \frac{1}{2} \right\}$$

defined by (4). Then, as announced in the Introduction, the following theorem holds

Theorem 1. *Under* **[H1]-[H3]**, $\mathcal{O}_N \subset \{\forall \mathbf{h} \in (\mathbb{L}^2_\tau)^K, (1/2)\|\mathbf{h}\|^2_\tau \leq \|\mathbf{h}\|^2_N \leq (3/2)\|\mathbf{h}\|^2_\tau\}$. *Moreover, for all* $p \geq 1$,

$$\mathbb{P}(\mathcal{O}_N^c) \lesssim N^{-p}. \tag{12}$$

($\lesssim$ means $\leq$ up to a constant).

In other words, on $\mathcal{O}_N$, the empirical norm and its theoretical counterpart are equivalent for functions of $(\mathbb{L}^2_\tau)^K$ and the probability $\mathbb{P}(\mathcal{O}_N^c)$ is as small as we want.

3. Definition and study of estimators of $\alpha_j(t)$, for $j = 1, \ldots, K$.

3.1. **Estimation contrast.** Let $(\varphi_j, j \geq 1)$ be an orthonormal basis of $\mathbb{L}^2_\tau := \mathbb{L}^2([0,\tau])$ composed of continuous functions and S_m be the subspace generated by $(\varphi_j, 1 \leq j \leq m)$.
For $m \geq 1$, let

$$L(S_m) = \sup_{t\in[0,\tau]} \sum_{j=0}^{m-1} \varphi_j^2(t) < +\infty, \tag{13}$$

The quantity $L(S_m)$ was introduced in Comte and Genon-Catalot (2020a, 2020b) in the framework of regression and drift estimation for diffusions by projection method. As

$$L(S_m) = \sup_{h_1\in S_m, \|h_1\|=1} \sup_{t\in[0,\tau]} h_1^2(t)$$

where $\|h_1\|^2 = \int_0^\tau h_1^2(t)dt$, it only depends on the subspace S_m and not on the basis chosen to define it. We assume

[H4] $\exists c > 0$ such that $L(S_m) \leq cm$.

Assumption **[H4]** holds for several classical bases of $\mathbb{L}^2_\tau$. We give examples in Section 7.3. This assumption may be weakened into $L(S_m) \leq cm^\omega$ for any $\omega \geq 1$.

For $\mathbf{h}(t) = (h_1(t), h_2(t), \ldots, h_K(t))^T$ element of $\mathbb{L}^2_\tau \times \cdots \times \mathbb{L}^2_\tau$, we consider the contrast which is inspired by the log-likelihood of the N processes (1) (see notation (10)),

$$U_N(\mathbf{h}) = \frac{1}{N}\int_0^\tau \sum_{i=1}^N \left[\mathbf{h}(t)^T\mathbf{g}(X_i(t))\right]^2 dt - \frac{2}{N}\sum_{i=1}^N \int_0^\tau \left[\mathbf{h}(t)^T\mathbf{g}(X_i(t))\right] dX_i(t). \tag{14}$$

We define the projection estimator of $\mathbf{a}(t) = (\alpha_1(t), \ldots, \alpha_K(t))^T$ on $S_{m_1}\times S_{m_2}\times\cdots\times S_{m_K} := S_{\mathbf{m}}$, for $\mathbf{m} = (m_1, m_2, \ldots, m_K)$, by

$$\widehat{\mathbf{a}}_{\mathbf{m}}(t) = (\widehat{\alpha}_1(t), \ldots, \widehat{\alpha}_K(t))^T = \arg\min_{\mathbf{h}\in S_{\mathbf{m}}} U_N(\mathbf{h}). \tag{15}$$

The choice of $U_N(\mathbf{h})$ for estimating $\mathbf{a}(t)$ is motivated by looking at the expectation:

$$\begin{aligned}
\mathbb{E}(U_N(\mathbf{h})) &= \frac{1}{N}\mathbb{E}\int_0^\tau \sum_{i=1}^N [\sum_{k=1}^K h_k(t)g_k(X_i(t)]^2 dt \\
&- \frac{2}{N}\mathbb{E}\int_0^\tau \sum_{i=1}^N [\sum_{k=1}^N h_k(t)g_k(X_i(t))][\sum_{k=1}^K \alpha_k(t)g_k(X_i(t))]dt \\
&= \mathbb{E}\int_0^\tau [\sum_{k=1}^K h_k(t)g_k(X_1(t)]^2dt - 2\mathbb{E}\int_0^\tau [\sum_{k=1}^K h_k(t)g_k(X_1(t))][\sum_{k=1}^K \alpha_k(t)g_k(X_1(t))]dt \\
&= \|\mathbf{h}\|^2_\tau - 2\langle \mathbf{h}, \mathbf{a}\rangle_\tau = \|\mathbf{h}-\mathbf{a}\|^2_\tau - \|\mathbf{a}\|^2_\tau,
\end{aligned}$$

which is minimal if $h_j = \alpha_j$ for $j = 1, \ldots, K$. Moreover,

$$\begin{aligned}
U_N(\mathbf{h}) &= \int_0^\tau \|\mathbf{h}(t)\|^2_{S_N(t)}\, dt - \frac{2}{N}\sum_{i=1}^N \int_0^\tau \left[\sum_{k=1}^K h_k(t)g_k(X_i(t))\right] dX_i(t) \\
&= \|\mathbf{h}\|^2_N - 2\langle \mathbf{h}, \mathbf{a}\rangle_N - 2\nu_N(\mathbf{h}) \qquad (16)
\end{aligned}$$

where

$$\nu_N(\mathbf{h}) = \frac{1}{N}\sum_{i=1}^{N}\int_0^{\tau}[\sum_{k=1}^{K} h_k(t)g_k(X_i(t))]\,\sigma(t,X_i(t))dW_i(t),$$

is a centered empirical process of interest.

3.2. **Minimum contrast estimator.** Let us now detail the construction and the expression of the estimator (15). Let

$$|\mathbf{m}| := m_1 + \cdots + m_K = \|\mathbf{m}\|_1. \tag{17}$$

Denote by $\widehat{\Psi}_{\mathbf{m}}$ the $|\mathbf{m}|\times|\mathbf{m}|$ symmetric matrix with blocks of size $m_j \times m_k$ denoted by $\widehat{\Psi}_{m_j,m_k}$ given by:

$$\widehat{\Psi}_{\mathbf{m}} = \begin{pmatrix} \widehat{\Psi}_{m_1,m_1} & \dots & \widehat{\Psi}_{m_1,m_K} \\ \vdots & & \vdots \\ \widehat{\Psi}_{m_K,m_1} & \dots & \widehat{\Psi}_{m_K,m_K} \end{pmatrix}, \tag{18}$$

where

$$\widehat{\Psi}_{m_j,m_k} = \left(\int_0^{\tau}\varphi_p(t)\varphi_q(t)\frac{1}{N}\sum_{i=1}^{N} g_j(X_i(t))g_k(X_i(t))dt\right)_{1\le p\le m_j, 1\le q\le m_k}.$$

Set moreover

$$\Psi_{\mathbf{m}} = \mathbb{E}\left(\widehat{\Psi}_{\mathbf{m}}\right).$$

Using definition (15), we can compute

$$\widehat{\mathbf{a}}_{\mathbf{m}}(t) = (\widehat{\alpha}_1(t),\dots,\widehat{\alpha}_K(t))^T \quad \text{where} \quad \widehat{\alpha}_k(t) = \sum_{j=1}^{m_k}\widehat{\alpha}_{k,j}\varphi_j(t) \tag{19}$$

and standardly obtain that the vector $\widehat{A}_{\mathbf{m}} = (\widehat{\alpha}_{1,1},\dots,\widehat{\alpha}_{1,m_1},\widehat{\alpha}_{2,1},\dots,\widehat{\alpha}_{2,m_2},\dots,\widehat{\alpha}_{K,1},\dots,\widehat{\alpha}_{K,m_K})^T$ of $\mathbb{R}^{|\mathbf{m}|}$ is solution of

$$\widehat{\Psi}_{\mathbf{m}}\widehat{A}_{\mathbf{m}} = V_{\mathbf{m}}, \quad \text{with} \quad V_{\mathbf{m}} = \begin{pmatrix} V_{1,m_1} \\ \vdots \\ V_{K,m_K}\end{pmatrix},$$

and V_{j,m_j} are $m_j \times 1$ vectors, $j = 1,\dots,K$, given by

$$V_{j,m_j} = \left(\frac{1}{N}\int_0^{\tau}\varphi_p(t)\sum_{i=1}^{N} g_j(X_i(t))dX_i(t), 1 \le p \le m_j\right)^T.$$

Therefore we need to know if the matrix $\widehat{\Psi}_{\mathbf{m}}$ is invertible, and this is the topic of the following Lemma, which also makes the link between the matrix and the empirical norm.

Lemma 1. *For* $\mathbf{x} = (x_{1,1},\dots,x_{1,m_1},x_{2,1}\dots,x_{2,m_2},\dots,x_{K,1},\dots,x_{K,m_K})^T \in \mathbb{R}^{|\mathbf{m}|}$*, we have*

$$\begin{aligned} \mathbf{x}^T\widehat{\Psi}_{\mathbf{m}}\mathbf{x} &= \int_0^{\tau}\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{K} h_j(t)g_j(X_i(t))\right)^2 dt = \|\mathbf{h}\|_N^2, \\ \mathbf{x}^T\Psi_{\mathbf{m}}\mathbf{x} &= \int_0^{\tau}\mathbb{E}\left(\sum_{j=1}^{K} h_j(t)g_j(X_1(t))\right)^2 dt = \|\mathbf{h}\|_{\tau}^2, \end{aligned}$$

where, for $j = 1, \ldots, K$, $h_j(t) = \sum_{p=1}^{m_j} x_{j,p}\varphi_p(t)$ and $\mathbf{h} = (h_1(t), \ldots, h_K(t))^T$ (see (3)). Under **[H3]***, the matrices $\widehat{\Psi}_{\mathbf{m}}$ and $\Psi_{\mathbf{m}}$ are symmetric positive definite.*

By Lemma 1, under **[H3]**, the matrix $\widehat{\Psi}_{\mathbf{m}}$ is invertible and positive definite. Therefore the estimator (19) can be computed by getting the coefficients $\widehat{A}_{\mathbf{m}}$ as follows:

$$\widehat{A}_{\mathbf{m}} = \widehat{\Psi}_{\mathbf{m}}^{-1} V_{\mathbf{m}}. \tag{20}$$

3.3. **Truncated estimator on a fixed space and risk bounds.** In what follows, we define the risk of any estimator $\bar{\mathbf{a}}_N(t)$ as the expectation of the empirical square norm $\|\bar{\mathbf{a}}_N - \mathbf{a}\|_N^2$ or the deterministic square norm $\|\bar{\mathbf{a}}_N - \mathbf{a}\|_\tau^2$. These definitions of the risk are classically used for problems of regression type by projection method, see e.g. Baraud *et al* (2001), Comte *et al.* (2007), Gendre (2014), Comte *et al.* (2020a), Denis *et al.* (2021).
The following proposition shows that, contrary to other contexts (see Cohen *et. al* (2013, 2019)), we need not introduce a restriction of the choices of the dimension spaces in term of bounding $\|\Psi_{\mathbf{m}}^{-1}\|_{\text{op}}$ by a quantity depending on N. Indeed, it holds that

Proposition 3. *Assume* **[H1]-[H3]**. *Then, for all* $\mathbf{m}$,

$$\|\Psi_{\mathbf{m}}^{-1}\|_{\text{op}} \le \mathfrak{f}_\tau = \sup_{t\in[0,\tau]} \|S(t)^{-1}\|_{\text{op}} \tag{21}$$

Proof of Proposition 3. Let us note that

$$\begin{aligned} \|\Psi_{\mathbf{m}}^{-1}\|_{\text{op}} &= \sup_{\mathbf{x}\in\mathbb{R}^{|m|}, \|\mathbf{x}\|_{2,|\mathbf{m}|}=1} \mathbf{x}^T\Psi_{\mathbf{m}}^{-1}\mathbf{x} = \sup_{\mathbf{y}\in\mathbb{R}^{|m|}, \|\mathbf{y}^T\Psi_{\mathbf{m}}\mathbf{y}\|_{2,|\mathbf{m}|}=1} \mathbf{y}^T\mathbf{y} \\ &= \sup_{\|\mathbf{h}\|_\tau^2=1, \mathbf{h}\in\mathbf{S}_{\mathbf{m}}} \|\mathbf{h}\|^2 = \sup_{\mathbf{h}\in\mathbf{S}_{\mathbf{m}}, \mathbf{h}\neq\mathbf{0}} \frac{\|\mathbf{h}\|^2}{\|\mathbf{h}\|_\tau^2}, \end{aligned} \tag{22}$$

where, for $\mathbf{y} = (y_{j,p}, p = 1, \ldots, m_j, j = 1, \ldots, K)$, $\mathbf{h} = (h_1(t), \ldots, h_K(t))^T$ and for $j = 1, \ldots, K$, $h_j(t) = \sum_{p=1}^{m_j} y_{j,p}\varphi_p(t)$. Recall that

$$\|\mathbf{h}\|_\tau^2 = \int_0^\tau \mathbf{h}(t)^T S(t)\mathbf{h}(t)dt \ge \int_0^\tau \inf_{1\le i\le K} \lambda_i(t)\mathbf{h}(t)^T\mathbf{h}(t)dt$$

where $(\lambda_i(t), i = 1, \ldots K)$ denote the eigenvalues of $S(t)$. Now,

$$\inf_{1\le i\le K} \lambda_i(t) = 1/\|S(t)^{-1}\|_{\text{op}} \ge 1/\mathfrak{f}_\tau.$$

This implies $\|\mathbf{h}\|_\tau^2 \ge \|\mathbf{h}\|^2/\mathfrak{f}_\tau$ which gives the result. □

Remark 1.
- *Combining Propositions 2 and 3, we see that the two norms $\|.\|$ and $\|.\|_\tau$ are equivalent for functions of $S_{\mathbf{m}}$.*
- *Note that, if $\mathbf{m} = (m_1, \ldots, m_K)$ and $\mathbf{m}' = (m_1', \ldots, m_K')$ are such that $m_j \le m_j'$ for $j = 1, \ldots, K$, then by (22), $\|\Psi_{\mathbf{m}}^{-1}\|_{\text{op}} \le \|\Psi_{\mathbf{m}'}^{-1}\|_{\text{op}}$.*
We can do the analogous reasoning for $\|\widehat{\Psi}_{\mathbf{m}}^{-1}\|_{\text{op}}$.

For the estimator (19), a transformation by introducing an adequate truncation is required, in relation with equality (21). For constants $\mathfrak{c}_1, \mathfrak{c}_2 > 0$ that can take any value, let us define

$$\Lambda_N = \{\forall t \in [0,\tau], \ \ \|S_N(t)^{-1}\|_{\text{op}} \le \mathfrak{c}_1 N^{\mathfrak{c}_2}\}. \tag{23}$$

Using (23), we define the trimmed estimator:

$$\widetilde{\mathbf{a}}_{\mathbf{m}} = \widehat{\mathbf{a}}_{\mathbf{m}}\mathbf{1}_{\Lambda_N} \tag{24}$$

The following proposition shows that Λ_N has large probability and guarantees a rough bound on $\|\widehat{\Psi}_{\mathbf{m}}^{-1}\|_{\text{op}}$.

Proposition 4. *Assume that* **[H1]** *to* **[H3]** *are fulfilled. Then, for all $p > 1$, there exists a constant $c_0 > 0$ depending on K, $\mathfrak{f}_\tau$ (see Proposition 3) and p, such that $\mathbb{P}(\Lambda_N^c) \leq c_0 N^{-p}$. Moreover, on Λ_N, it holds that*

$$\forall \mathbf{m}, \quad \|\widehat{\Psi}_{\mathbf{m}}^{-1}\|_{\text{op}} \leq \mathfrak{c}_1 N^{\mathfrak{c}_2}.$$

Denote by $\widehat{\Theta}_{\mathbf{m}}$ the $|\mathbf{m}| \times |\mathbf{m}|$ symmetric matrix built similarly to $\widehat{\Psi}_{\mathbf{m}}$, but given by the blocks $m_j \times m_k$

$$\widehat{\Theta}_{m_j,m_k} = \left(\int_0^\tau \varphi_p(t)\varphi_q(t) \frac{1}{N} \sum_{i=1}^N g_j(X_i(t)) g_k(X_i(t)) \sigma^2(t, X_i(t)) dt \right)_{1 \leq p \leq m_j, 1 \leq q \leq m_k}, \tag{25}$$

The deterministic counterpart is $\Theta_{\mathbf{m}} := \mathbb{E}(\widehat{\Theta}_{\mathbf{m}})$, which is also $|\mathbf{m}| \times |\mathbf{m}|$ and symmetric.

We can prove the following risk bounds with respect to the integrated empirical and deterministic norms.

Theorem 2. *Assume that* **[H1]** *to* **[H4]** *hold and that* $\mathbf{m}$ *satisfies* $|\mathbf{m}| \leq N$. *The estimator* $\widetilde{\mathbf{a}}_{\mathbf{m}}$ *of* $\mathbf{a}(t) = (\alpha_1(t), \dots, \alpha_K(t))^T$ *satisfies, for c is a generic constant,*

$$\mathbb{E}\|\widetilde{\mathbf{a}}_{\mathbf{m}} - \mathbf{a}\|_N^2 \leq \inf_{\mathbf{h}=(h_1,\dots,h_K)^T \in S_{\mathbf{m}}} \|\mathbf{h} - \mathbf{a}\|_\tau^2 + 2\frac{\text{Tr}(\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}})}{N} + \frac{c}{N}, \tag{26}$$

$$\mathbb{E}\|\widetilde{\mathbf{a}}_{\mathbf{m}} - \mathbf{a}\|_\tau^2 \leq 5 \inf_{\mathbf{h}=(h_1,\dots,h_K)^T \in S_{\mathbf{m}}} \|\mathbf{h} - \mathbf{a}\|_\tau^2 + 4\frac{\text{Tr}(\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}})}{N} + \frac{c}{N}. \tag{27}$$

We have $\text{Tr}(\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}}) \leq \mathfrak{C}|\mathbf{m}|$ *where* $\mathfrak{C}$ *is a constant given in the proof and* $|\mathbf{m}|$ *is given in* (17).

An explicit value of the constant $\mathfrak{C}$ above can be given under an additional assumption.

[H5] $\sup_{t\in[0,\tau], x\in\mathbb{R}} \sigma^2(t,x) := \|\sigma\|_\infty^2 < +\infty$.

Corollary 1. *If σ satisfies* **[H5]**, *then* $\text{Tr}(\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}}) \leq \|\sigma\|_\infty^2 |\mathbf{m}|$.

3.4. **Discussion about rates.** To evaluate rates of convergence, we must assess the $\mathbb{L}^2$-norm of the estimators bias within some regularity subspaces of $\mathbb{L}_\tau^2$. Such assessments are standard in nonparametric statistcs, for function α_j belonging to Sobolev spaces associated with the chosen basis (see examples in Comte and Genon-Catalot (2023), section 3.3).

Proposition 5. *Assume* **[H1]** *to* **[H5]** *and that for $j \in \{1, \dots, K\}$, the function α_j belongs to a regularity space such that $\inf_{h\in S_m} \|\alpha_j - h\|^2 \leq R_j m^{-2\alpha_j}$. Choose $m_j = O(N^{1/(2\alpha_j+1)})$ for $j = 1, \dots, K$ and set $\alpha^\star = \min_{j=1,\dots,K} \alpha_j$, then*

$$\mathbb{E}\|\widetilde{\mathbf{a}}_{\mathbf{m}} - \mathbf{a}\|_\tau^2 \lesssim O(N^{-2\alpha^\star/(2\alpha^\star+1)}).$$

If all functions have the same regularity $\alpha_j = \alpha^\star$, for all $j = 1, \dots K$, choosing $m_j = N^{1/(2\alpha^\star+1)} := m^\star$ for $j = 1, \dots, K$, and setting $\mathbf{m}^\star = (m^\star, \dots, m^\star)$, we obtain

$$\mathbb{E}\|\widetilde{\mathbf{a}}_{\mathbf{m}^\star} - \mathbf{a}\|_\tau^2 \lesssim O(N^{-2\alpha^\star/(2\alpha^\star+1)}).$$

Proof of Proposition 5. By Proposition 2, we have

$$\begin{aligned}\inf_{\mathbf{h}=(h_1,\ldots,h_K)^T\in S_{\mathbf{m}}}\|\mathbf{h}-\mathbf{a}\|_\tau^2 &\leq KG^2\inf_{\mathbf{h}\in S_{\mathbf{m}}}\|\mathbf{h}-\mathbf{a}\|^2 = KG^2\inf_{h_j\in S_{m_j},j=1,\ldots,K}\sum_{j=1}^K\|h_j-\alpha_j\|^2\\ &\leq KG^2\sum_{j=1}^K R_j m_j^{-2\alpha_j}.\end{aligned}$$

Under **[H5]**, we find, under the assumptions of Theorem 2 ,

$$\begin{aligned}\mathbb{E}\|\widetilde{\mathbf{a}}_{\mathbf{m}}-\mathbf{a}\|_\tau^2 &\leq 5\inf_{\mathbf{h}\in S_{\mathbf{m}}}\|\mathbf{h}-\mathbf{a}\|_\tau^2+4\|\sigma\|_\infty^2\frac{|\mathbf{m}|}{N}+\frac{c}{N}\\ &\leq 5KG^2\sum_{j=1}^K R_j m_j^{-2\alpha_j}+4\|\sigma\|_\infty^2\frac{m_1+\cdots+m_K}{N}+\frac{c}{N}\\ &\lesssim \sum_{j=1}^K N^{-(2\alpha_j)/(2\alpha_j+1)} = O(N^{-2\alpha^\star/(2\alpha^\star+1)}).\quad\square\end{aligned}$$

Thus, our method has the advantadge of estimating all functions simultaneously and of reaching the rate corresponding to the estimation of one function with regularity $\alpha^\star$. The drawback is that the rate corresponds to the smallest regularity.

3.5. **Model selection.** Now, the choices proposed above are asymptotic and depend on unknown regularity parameters. So, they cannot be implemented. This is why we propose a data driven model selection device. This defines a new estimator, for which we prove a nonasymptotic risk bound.

Consider the collection of models defined by

$$\mathcal{M}_N=\left\{\mathbf{m}\in\{1,\ldots,N\}^K,\quad |\mathbf{m}|\leq N\right\}. \tag{28}$$

Set

$$\widehat{\mathbf{m}}\in\arg\min_{\mathbf{m}\in\mathcal{M}_N}\left[U_N(\widehat{\mathbf{a}}_{\mathbf{m}})+\mathrm{pen}(\mathbf{m})\right],\quad \mathrm{pen}(\mathbf{m})=\kappa\|\sigma\|_\infty^2\frac{|\mathbf{m}|}{N}, \tag{29}$$

and consider the estimator

$$\widetilde{\mathbf{a}}=\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}}\mathbf{1}_{\Lambda_N},$$

where Λ_N is defined by (23).

Theorem 3. *Assume that* **[H1]** *to* **[H5]** *hold. Consider the estimator* $\widetilde{\mathbf{a}}$ *of* $\mathbf{a}(t)=(\alpha_1(t),\ldots,\alpha_K(t))^T$ *with any* $\widehat{\mathbf{m}}$ *defined by (29). Then, there exists a numerical constant* κ_0 *such for all* $\kappa\geq\kappa_0$*, we have*

$$\mathbb{E}\left(\|\widetilde{\mathbf{a}}-\mathbf{a}\|_N^2\right)\leq 4\inf_{\mathbf{m}\in\mathcal{M}_N}\left(\inf_{\mathbf{h}=(h_1,\ldots,h_K)^T\in S_{\mathbf{m}}}\|\mathbf{h}-\mathbf{a}\|_\tau^2+\|\sigma\|_\infty^2\frac{|\mathbf{m}|}{N}\right)+\frac{C}{N}, \tag{30}$$

where C *is a constant depending on* $K, G, \|\sigma\|_\infty$.

As a consequence, Inequality (30) shows that the estimator is performing an automatic finite sample and global square bias/variance compromise. Asymptotically, when $\mathbf{a}$ belongs to a regularity space as described in Section 3.4, the rates given in Proposition 5 follow.

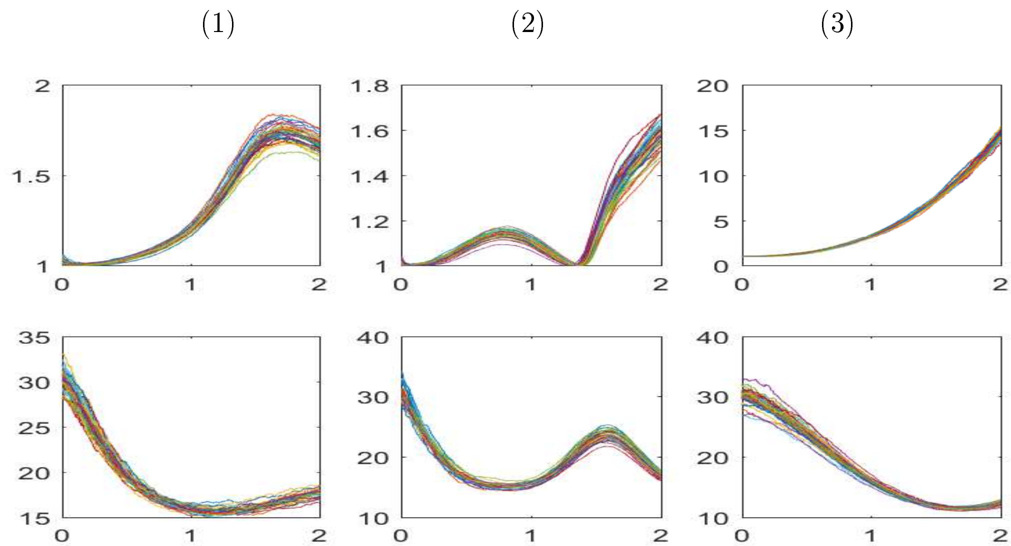

FIGURE 1. Plots of 40 repetitions of $t \mapsto \lambda_{\max}(S_N^{-1}(t))$ on $[0, \tau]$ with $\tau = 2$. Couples (g_1, g_2): first line (a), second line (b). Examples (1)-(2)-(3) in corresponding columns. $N = 2000$

	Triplet (1)				Triplet (2)				Triplet (3)			
	$N = 500$		$N = 2000$		$N = 500$		$N = 2000$		$N = 500$		$N = 2000$	
	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2
MSE A.T.	1.8	.75	.57	.25	1.3	.81	.44	.20	.99	1.6	.30	.40
(std)	(1.0)	(.29)	(.24)	(.08)	(.46)	(.33)	[.15)	(.11)	(.63)	(.62)	(.14)	(.17)
I T	1.8	.78	.56	.26	1.5	.58	.51	.21	.89	1.4	.31	.43
(std)	(1.0)	(.31)	(.25)	(.09)	(.39)	(.23)	(.17)	(.07)	(.66)	(.66)	(.17)	(.17)
MSE A.L.	1.1	.52	.29	.13	.91	.48	.24	.15	.56	.29	.17	.13
(std)	(0.7)	(.27)	(.18)	(.06)	(.39)	(.38)	(.14)	(.07)	(.59)	(.31)	(.11)	(.09)
MSE I.L.	1.3	.56	.34	.14	.95	.40	.29	.11	.62	.53	.18	.16
(std)	(0.7)	(.28)	(.18)	(.07)	(0.38)	(.30)	(.18)	(.08)	(.68)	(.44)	(.12)	(.12)
dim T	6.7	8.6	8.5	11	6.1	4.7	8.1	6.1	3.3	5.0	4.7	6.7
dim L	4.7	7.1	5.2	7.7	6.3	5.0	7.1	5.2	4.0	2.0	4.1	2.1

TABLE 1. First case (g_1, g_2) of Ornstein-Uhlenbeck type (a). Mean squared error (MSE) and standard deviation (std) are both multiplied by 100. A./I. for Anisotropic or Isotropic, T./L. for (half) Trigonometric or Laguerre basis. Dimensions (dim) are averages of selected dimensions in the anisotropic case.

4. NUMERICAL RESULTS ON SIMULATED DATA.

In this simulation section, we consider the case $K = 2$. Two examples of couples (g_1, g_2):

(a) $g_1(x) = 1, g_2(x) = x$
(b) $g_1(x) = x$ and $g_2(x) = x/(1 + x^2)$,

are illustrated, with three examples of triplets $(\alpha_1, \alpha_2, \sigma)$:

(1) $\alpha_1(t) = t(\tau - t)$, $\alpha_2(t) = \sin(4t)$, $\sigma(t,x) = 0.5(1 + \frac{1}{\sqrt{1+x^2}})$,
(2) $\alpha_1(x) = \sin(4t)$, $\alpha_2(t) = \cos(2.5t)$, $\sigma(t,x) = 1/(1+t^2)$,
(3) $\alpha_1(t) = t$, $\alpha_2(t) = -2t/(1+t^2)$, $\sigma(t,x) = 0.5$.

We have generated discrete paths on $[0,\tau]$ with $\tau = 2$ by a basic Euler scheme, with $n = 2000$ observations for a step $\Delta = \tau/n$.

First, we study the behaviour of the largest eigenvalue of $S_N^{-1}(t)$, denoted by $\lambda_{\max}(S_N^{-1}(t))$. The supremum over $[0,\tau]$ of this function is involved in the definition of the estimator, and it corresponds to the empirical version of $\mathfrak{f}_\tau$, which is finite under our assumptions. The results for sample size $N = 2000$ and 40 repetitions are presented in Figure 1. The pictures show that the profiles are very different in the different examples, and their values are also quite different.

Next, we look at the performance of the estimator. For each path, a discrete L^2-distance between the true function and its estimation is computed. The values of MSE are obtained by averaging these results over the $L = 400$ simulated trajectories correponding to each case. To be more precise, for simulation ℓ, we calculate $\left((\widehat{\alpha}_p)^{(\ell)}_{\widehat{m}_p}(k\Delta)\right)_{1\le k\le n}$ for $p = 1,2$ from N independent paths $(X_i^{(\ell)}(k\Delta))_{1\le k\le n}$, for $i = 1,\dots,N$ and compute for $p = 1,2$, the MSE for α_p as

$$\frac{1}{L}\sum_{\ell=1}^{L}\left[\Delta\sum_{k=1}^{n}\left((\widehat{\alpha}_p)^{(\ell)}_{\widehat{m}_p}(k\Delta) - \alpha_p(k\Delta)\right)^2\right].$$

	Triplet (1)				Triplet (2)				Triplet (3)			
	$N=500$		$N=2000$		$N=500$		$N=2000$		$N=500$		$N=2000$	
	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2
MSE A.T.	1.1	19	.37	6.1	1.6	9.5	.51	3.6	.61	9.1	.20	2.1
(std)	(.91)	(13)	(.24)	(3.6)	(1.2)	(9.3)	(.29)	(1.7)	(.42)	(4.2)	(.11)	(.75)
MSE I.T.	1.1	23	.38	8.1	1.9	9.7	.60	3.0	.67	5.0	.22	1.7
(std)	(.94)	(12)	(.24)	(3.5)	(1.2)	(8.3)	(.27)	(2.0)	[.42)	(3.1)	(.11)	(.93)
MSE A.L.	.84	18	.22	5.1	1.6	17	.39	3.1	.44	2.2	.11	.54
(std)	(.88)	(9.3)	(.20)	(2.9)	(1.3)	(11)	(.30)	(1.8)	(.36)	(2.6)	(.09)	[.53)
MSE I.L.	.84	17	.23	4.8	1.7	13	.40	3.3	.48	3.9	.17	1.1
(std)	(.91)	(9.7)	(.21)	(2.6)	(1.3)	(11)	[.30)	(2.3)	(.38)	(3.2)	(.12)	.75)
dim T	9.8	5.8	13	6.7	6.7	3.3	8.6	3.7	6.5	2.1	9.3	3.6
dim L	5.3	5.4	5.4	6.4	6.7	3.1	7.2	4.1	4.3	2.0	5.0	2.0

TABLE 2. Case (b), $(g_1(x), g_2(x)) = (x, x/(1+x^2))$. Mean squared error (MSE) and standard deviation (std) are both multiplied by 100. A./I. for Anisotropic or Isotropic, T./L. for (half) Trigonometric or Laguerre basis. Dimensions (dim) are averages of selected dimensions in the anisotropic case.

We experimented two different samples sizes: $N = 500$ and $N = 2000$ in order to check the improvement brought by increasing N. We also implemented two bases for the estimation: the half trigonometric and the Laguerre basis (see their description in section 7.3 in Appendix). The penalty constant κ in formula (29) is taken equal to 2.5 for half-trigonometric basis and to 3 for Laguerre basis. The true value of $\|\sigma\|_\infty$ is used. Both anisotropic and isotropic model selection

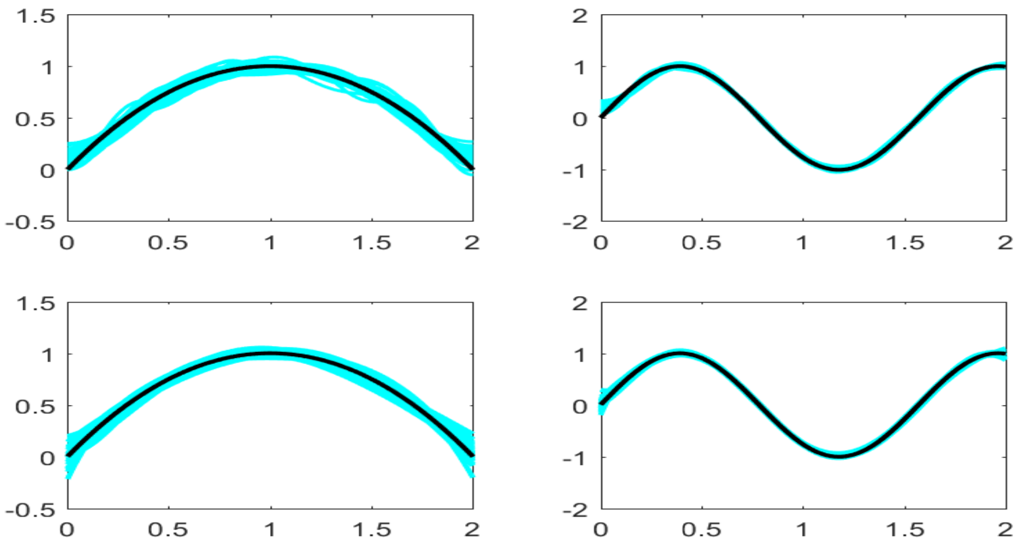

FIGURE 2. Example (1)-(b). True curve in black and 40 estimated functions in cyan, for $N = 2000$. Right: function α_1, 100 MSE 0.52 and 0.31. Left: function α_2, 100 MSE 0.33 and 0.16. Top trigonometric basis, bottom Laguerre bases.

are implemented and, each model is selected among dimensions 1 to $D_{\max}$ with $D_{\max} = 15$ for the half-trigonometric basis and $D_{\max} = 8$ for the Laguerre basis. These maximal dimension are selected to be large enough for all examples (in the sense that much smaller dimensions are always chosen by the algorithm), and in that way, to save computing time (these $D_{\max}$ are not as large as they should).

Let us comment the results given in Tables 1 and 2. Clearly, increasing N always substantially improve the results and decreases the MSE. In the same time, the selected dimensions increase, which is also expected. Most of the time (75%), the anisotropic method gives a better result than the isotropic one, but it is almost always true for α_1 and much more mitigated for α_2. It is likely that, contrary to what the theory says, the global risk is generally improved by the anisotropic model selection, but probably not in a significant order.

Figures 2 and 3 show 40 estimators compared to the true function (in bold black), and illustrate that for different functions (g_1, g_2), the results can be quite different: the function α_2 is well estimated in case (a), but the MSEs are much larger in case (b), and this can be seen on the plots. Figure 4 presents another illustration for smaller sample size, and shows that the estimator can fit a straight line (function α_1), which was not obvious with trigonometric or Laguerre bases.

5. CONCLUDING REMARKS

In this paper, we consider a new setting of N *i.i.d.* one-dimensional inhomogeneous diffusion processes $(X_i(t), i = 1, \ldots, N)$ with drift $\mu(t,x) = \sum_{j=1}^{K} \alpha_j(t) g_j(x)$ and diffusion coefficient $\sigma(t,x)$, where K, the functions $g_j(x)$ and $\sigma(t,x)$ are known. We propose a nonparametric estimation method for the K-dimensional unknown function $(\alpha_j(t), j = 1, \ldots, k)$ from the continuous observation of the N sample paths $(X_i(t))$ throughout a fixed time interval $[0, \tau]$. We proceed by a projection method on finite dimensional subspaces of $\mathbb{L}^2([0,\tau])^K$ and propose a data-driven choice of the dimension of the projection space.

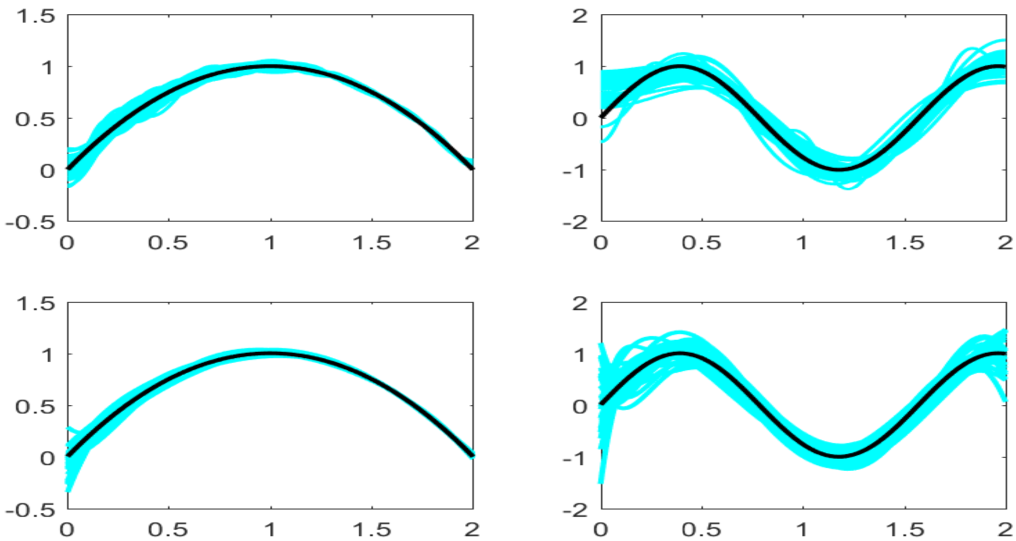

FIGURE 3. Example (1)-(b). True curve in black and 40 estimated functions in cyan, for $N = 2000$. Right: function α_1, 100 MSE 0.28 and 0.16. Left: function α_2, 100 MSE 5.7 and 4.6. Top trigonometric basis, bottom Laguerre bases.

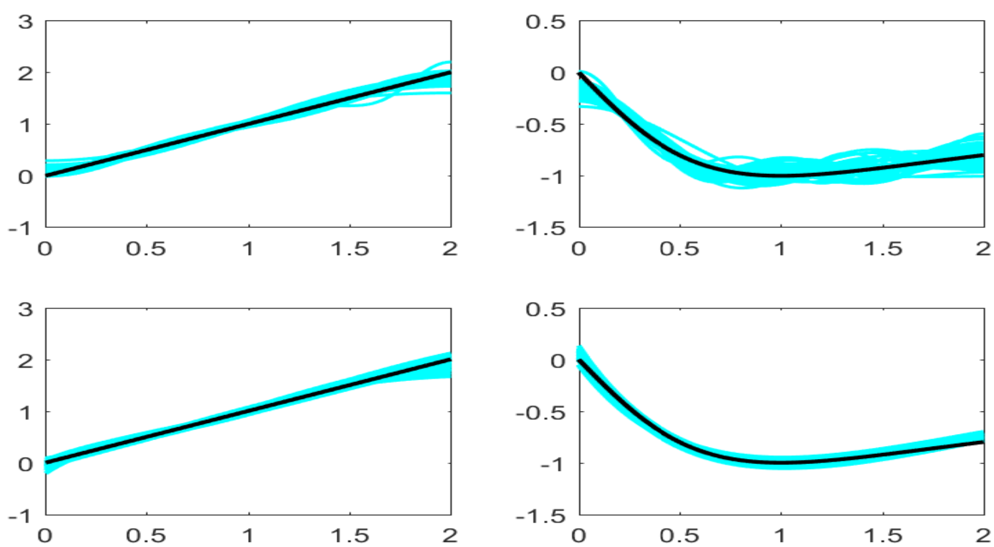

FIGURE 4. Example (3)-(a). True curve in black and 40 estimated functions in cyan, for $N = 500$. Right: function α_1, 100 MSE 0.84 and 0.49. Left: function α_2, 100 MSE 1.0 and 0.22. Top trigonometric basis, bottom Laguerre bases.

We obtain risk bounds for the projection estimator on a fixed space and for the data-driven estimator. Numerical results on simulated data for several models show that the method works in practice.

Discrete time observations could be studied as was done in Comte and Genon-Catalot (2023); it is enough to discretize all formulae and definitions. The handling of residual terms is similar to the Appendix of Comte and Genon-Catalot (2023). This is what is used in simulations.

A possible extension would be to look at the case where τ can be large, but this would be a completely different framework from the point of view of the assumptions on the model.

Another question is: how could we propose methods which may deliver individual estimators for each function α_j catching each regularity? In an additive case for regression, Gendre (2014) studies projections of the observations in order to implement one of the functions of the sum alone. It is not obvious if such a construction may be possible in the present case.

The case of correlated Brownian motions driving the SDEs may also be of interest even if probably difficult, see Comte and Marie (2023). Lastly, the case of general drift $(t,x)\mapsto\mu(t,x)$ by projection method would be worth of investigation and probably related to bivariate rates.

6. Proofs

6.1. **Proof of Proposition 1.** The bound (7) is classical (see e.g. Proposition A in Gloter (2000)) and follows from [**H1**]. The bound (8) is obtained by applying the Burkholder-Davis-Gundy Inequality (see (6) and **[H1]**):

$$\begin{aligned}\mathbb{E}(|X_1(t)-X_1(s)|^{2r}) &\leq 2^{2r-1}\mathbb{E}\left(\left|\int_s^t \mu(u,X_1(u))du\right|^{2r}+\left|\int_s^t \sigma(u,X_1(u))dW_1(u)\right|^{2r}\right)\\ &\leq 2^{2r-1}\left(2^{2r-1}(L(\tau))^{2r}|t-s|^{2r})+2^{2r-1}C_\tau^{2r}|t-s|^r\right)\sup_{0\leq u\leq\tau}(1+\mathbb{E}|X_1(u)|^{2r}).\end{aligned}$$

This yields, for $|t-s|\leq 1$,

$$\mathbb{E}(|X_1(t)-X_1(s)|^{2r})\leq \mathfrak{B}(r,\tau)|t-s|^r$$

with

$$\mathfrak{B}(r,\tau):=2^{4r-2}\left((L(\tau))^{2r}+C_\tau^{2r}\right)\sup_{0\leq u\leq\tau}(1+\mathbb{E}|X_1(u)|^{2r}).$$

For (9), we write that, for $\mathfrak{g}=g_jg_k$,

$$\begin{aligned}|\mathfrak{g}(X_1(t))-\mathfrak{g}(X_1(s))| &\leq |(g_j(X_1(t))-g_j(X_1(s)))g_k(X_1(t))|\\ &\quad +|g_j(X_1(s))(g_k(X_1(t))-g_k(X_1(s)))|\\ &\leq L|X_1(t)-X_1(s)|(|g_k(X_1(t))|+|g_j(X_1(s))|).\end{aligned}$$

$$\mathbb{E}(|\mathfrak{g}(X_1(t))-\mathfrak{g}(X_1(s))|^{2r})\leq 2^{2r-1}L^{2r}\mathbb{E}^{1/2}(|X_1(t)-X_1(s)|^{4r})\max_{k\in\{1,\dots,K\}}\sup_{t\in[0,\tau]}\mathbb{E}^{1/2}(|g_k(X_1(t))|^{4r}).$$

Using (8), we obtain (9) with

$$\mathfrak{C}(r,\tau):=2^{2r-1}L^{2r}\mathfrak{B}^{1/2}(2r,\tau)\max_{k\in\{1,\dots,K\}}\sup_{t\in[0,\tau]}\mathbb{E}^{1/2}(|g_k(X_1(t))|^{4r}).$$

Note that

$$\sup_{t\in[0,\tau]}\mathbb{E}(|g_k(X_1(t))|^{4r})\leq 2^{4r-1}\widetilde{L}^{4r}\left(1+\sup_{t\in[0,\tau]}\mathbb{E}(|X_1(t)|^{4r})\right)<+\infty.\quad\square$$

6.2. **Proof of Theorem 1.** We denote by $S(t)^{1/2}$ a symmetric square root of $S(t)$, invertible under **[H3]**. Let $\mathbf{h} \in (\mathbb{L}^2_\tau)^K$ such that $\|\mathbf{h}\|_\tau^2 = \int_0^\tau \mathbf{h}(t)^T S(t)\mathbf{h}(t)dt = 1$. Then

$$\begin{aligned}
\left|\frac{\|\mathbf{h}\|_N^2}{\|\mathbf{h}\|_\tau^2} - 1\right| &= \left|\|\mathbf{h}\|_N^2 - \|\mathbf{h}\|_\tau^2\right| = \left|\int_0^\tau \mathbf{h}(t)^T(S_N(t) - S(t))\mathbf{h}(t)dt\right| \\
&= \left|\int_0^\tau \mathbf{h}(t)^T S(t)^{1/2}(S(t)^{-1/2}S_N(t)S^{-1/2}(t) - \mathrm{Id}_K))S(t)^{1/2}\mathbf{h}(t)dt\right| \\
&\le \sup_{t\in[0,\tau]} \|S(t)^{-1/2}S_N(t)S(t)^{-1/2} - \mathrm{Id}_K\|_{\mathrm{op}} \int_0^\tau |\mathbf{h}(t)^T S(t)\mathbf{h}(t)|dt \\
&= \sup_{t\in[0,\tau]} \|S(t)^{-1/2}S_N(t)S(t)^{-1/2} - \mathrm{Id}_K\|_{\mathrm{op}},
\end{aligned}$$

using that $\|\mathbf{h}\|_\tau = 1$. As a consequence $\sup_{t\in[0,\tau]} \|S(t)^{-1/2}S_N(t)S(t)^{-1/2} - \mathrm{Id}_2\|_{\mathrm{op}} \le 1/2$ implies for all $\mathbf{h} \in (\mathbb{L}^2_\tau)^K$, $\left|\|\mathbf{h}\|_N^2/\|\mathbf{h}\|_\tau^2 - 1\right| \le 1/2$, which gives the first result.

Next, using **[H3]**, recall that we have set $\mathfrak{f}_\tau := \sup_{t\in[0,\tau]} \|S(t)^{-1}\|_{\mathrm{op}}$. We have

$$\begin{aligned}
\sup_{t\in[0,\tau]} \|S(t)^{-1/2}S_N(t)S(t)^{-1/2} - \mathrm{Id}_K\|_{\mathrm{op}} &\le \sup_{t\in[0,\tau]} \|S(t)^{-1}\|_{\mathrm{op}} \sup_{t\in[0,\tau]} \|S_N(t) - S(t)\|_{\mathrm{op}} \\
&\le \mathfrak{f}_\tau \sup_{t\in[0,\tau]} \sqrt{\mathrm{Tr}((S_N(t) - S(t))^2)}
\end{aligned}$$

Then

$$\begin{aligned}
\mathrm{Tr}((S_N(t) - S(t))^2) &= \sum_{1\le j,k\le K} \left(\frac{1}{N}\sum_{i=1}^N \{g_j(X_i(t))g_k(X_i(t)) - \mathbb{E}[g_j(X_i(t))g_k(X_i(t))]\}\right)^2 \\
&\le K^2 \max_{1\le j,k\le K} \left(\frac{1}{N}\sum_{i=1}^N \{g_j(X_i(t))g_k(X_i(t)) - \mathbb{E}[g_j(X_i(t))g_k(X_i(t))]\}\right)^2
\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathbb{P}(\mathcal{O}_N^c) \le \mathbb{P}\left(\max_{1\le j,k\le K} \sup_{t\in[0,\tau]} \left|\frac{1}{N}\sum_{i=1}^N \{g_j(X_i(t))g_k(X_i(t)) - \mathbb{E}[g_j(X_i(t))g_k(X_i(t))]\}\right| > \frac{1}{2K\mathfrak{f}_\tau}\right), \\
&\le \sum_{1\le j,k\le K} \mathbb{P}\left(\sup_{t\in[0,\tau]} \left|\frac{1}{N}\sum_{i=1}^N \{g_j(X_i(t))g_k(X_i(t)) - \mathbb{E}[g_j(X_i(t))g_k(X_i(t))]\}\right| > \frac{1}{2K\mathfrak{f}_\tau}\right).
\end{aligned}$$

The result of Theorem 1 follows immediately from Lemma 2 below. □

Lemma 2 is obtained by application of the Garsia-Rodemich-Rumsey (1970/71) Lemma (in the formulation stated in Jourdain and Pagès (2022), see Lemma 5 in Section 7).

Lemma 2. *Assume that* **[H1]-[H2]** *holds. Let $\mathfrak{g} = g_j g_k$ for $j, k \in \{1, \ldots, K\}$. Then $\forall p > 1$, there exists a constant $C_{p,\tau}$ such that, for all constant $\mathfrak{a}_\tau > 0$,*

$$\mathbb{P}_0 := \mathbb{P}\left(\sup_{t\in[0,\tau]} \left|\frac{1}{N}\sum_{i=1}^N \mathfrak{g}(X_i(t)) - \mathbb{E}[\mathfrak{g}(X_i(t))]\right| > \mathfrak{a}_\tau\right) \le C_{p,\tau} a_\tau^{-2p} N^{-p}.$$

6.3. **Proof of Lemma 2.** Let us define

$$Y_N(t) := \frac{1}{N}\sum_{i=1}^N [\mathfrak{g}(X_i(t)) - \mathbb{E}(\mathfrak{g}(X_i(t))].$$

First we prove that there exists $a > 1$ and a constant c_τ such that

$$\forall N \geq 1, \forall s,t \in [0,\tau], \quad \mathbb{E}[|Y_N(t) - Y_N(s)|^{2p}] \leq c_\tau |t-s|^a \frac{1}{N^p}.$$

We apply the Rosenthal Inequality and get

$$\begin{aligned}\mathbb{E}[|Y_N(t) - Y_N(s)|^{2p}] &\leq \frac{C(2p)}{N^{2p}} \left(N\mathbb{E}(|\mathfrak{g}(X_1(t)) - \mathfrak{g}(X_1(s))|^{2p}) + (N\mathrm{Var}(\mathfrak{g}(X_1(t)) - \mathfrak{g}(X_1(s))))^p \right) \\ &\leq \frac{C(2p)}{N^{2p}} \left(N\mathbb{E}(|\mathfrak{g}(X_1(t)) - \mathfrak{g}(X_1(s))|^{2p}) + \left(N\mathbb{E}(\mathfrak{g}[(X_1(t)) - \mathfrak{g}(X_1(s)))^2]\right)^p \right),\end{aligned}$$

where $C(2p)$ is the constant of the Rosenthal Inequality. By applying (9) of Proposition 1, we obtain

$$\begin{aligned}\mathbb{E}[|Y_N(t) - Y_N(s)|^{2p}] &\leq \frac{2C(2p)}{N^p} \mathbb{E}(|\mathfrak{g}(X_1(t)) - \mathfrak{g}(X_1(s))|^{2p}) \\ &\leq \frac{2C(2p)}{N^p} \mathfrak{C}(p,\tau)|t-s|^p\end{aligned}$$

Then by Lemma 5 (see Appendix), we get that for $p > 1$, there exists a constant $C_{p,\tau}$ such that

$$\forall N \geq 1, \quad \mathbb{E}\left(\sup_{t\in[0,\tau]} |Y_N(t) - Y_N(0)|^{2p} \right) \leq C_{p,\tau} \frac{1}{N^p}.$$

Next, by the Rosenthal Inequality, we get

$$\mathbb{E}[|Y_N(0)|^{2p}] \leq C(2p) N^{-2p} \{ N\mathbb{E}[|\mathfrak{g}(X_1(0))|^{2p}] + N^p [\mathrm{Var}(\mathfrak{g}(X_1(0)))]^p \}.$$

Therefore for another constant $C_{p,\tau}$,

$$\mathbb{E}\left(\sup_{t\in[0,\tau]} |Y_N(t)|^{2p} \right) \leq C_{p,\tau} \frac{1}{N^p}. \tag{31}$$

Now by the Markov Inequality, $\mathbb{P}_0 \leq \mathfrak{a}_\tau^{-2p} C_{p,\tau} N^{-p}$. □

6.4. **Proof of Lemma 1.** Using the product of matrices by blocks and setting for $j = 1, \ldots, k$, $\mathbf{x}_j^T = (x_{j,1}, \ldots, x_{j,m_j})$, we get:

$$\mathbf{x}^T \widehat{\Psi}_{\mathbf{m}} \mathbf{x} = \sum_{1\leq j,k\leq K} \mathbf{x}_j^T \widehat{\Psi}_{m_j,m_k} \mathbf{x}_k.$$

Using the definition of $\widehat{\Psi}_{m_j,m_k}$ yields, for h_j as defined in Lemma 1:

$$\mathbf{x}_j^T \widehat{\Psi}_{m_j,m_k} \mathbf{x}_k = \int_0^\tau h_j(t) h_k(t) \frac{1}{N} \sum_{i=1}^N g_j(X_i(t)) g_k(X_i(t)) dt.$$

Thus,

$$\mathbf{x}^T \widehat{\Psi}_{\mathbf{m}} \mathbf{x} = \frac{1}{N} \sum_{i=1}^N \int_0^\tau \left(\sum_{j=1}^K h_j(t) g_j(X_i(t)) \right)^2 dt = \int_0^\tau \mathbf{h}(t)^T S_N(t) \mathbf{h}(t) dt.$$

Now, $\mathbf{x}^T \widehat{\Psi}_{\mathbf{m}} \mathbf{x} = 0$ implies that $\mathbf{h}(t)^T S_N(t) \mathbf{h}(t) = 0$ *a.e.* on $[0,\tau]$, by **[H3]**. As, for all j, the functions $(\varphi_j, j = 1, \ldots, m_j)$ are orthonormal on $\mathbb{L}^2_\tau$, this implies that for all j, $\mathbf{x}_j = 0$, therefore, $\mathbf{x} = 0$. This shows that $\widehat{\Psi}_{\mathbf{m}}$ is positive definite. The same holds for $\Psi_{\mathbf{m}}$. □

6.5. **Proof of Proposition 4.** On Λ_N^c, there exists $t_0 \in [0,\tau]$ such that $\|S_N(t_0)^{-1}\|_{\rm op} > \mathfrak{c}_1 N^{\mathfrak{c}_2}$ while $\sup_{t\in[0,\tau}\|S(t)^{-1}\|_{\rm op} \le (\mathfrak{c}_1/2)N^{\mathfrak{c}_2}$ (indeed for $\mathfrak{c}_2>0$, it holds $(\mathfrak{c}_1/2)N^{\mathfrak{c}_2} > \mathfrak{f}_\tau$) and thus $\|S(t_0)^{-1}\|_{\rm op} \le (\mathfrak{c}_1/2)N^{\mathfrak{c}_2}$. It follows that for this t_0, $\|S_N(t_0)^{-1} - S(t_0)^{-1}\|_{\rm op} \ge (\mathfrak{c}_1/2)N^{\mathfrak{c}_2}$. Indeed

$$\mathfrak{c}_1 N^{\mathfrak{c}_2} < \|S_N(t_0)^{-1}\|_{\rm op} \le \|S_N(t_0)^{-1}-S(t_0)^{-1}\|_{\rm op}+\|S(t_0)^{-1}\|_{\rm op} \le \|S_N(t_0)^{-1}-S(t_0)^{-1}\|_{\rm op}+(\mathfrak{c}_1/2)N^{\mathfrak{c}_2}.$$

Therefore

$$\mathbb{P}(\Lambda_N^c) \le \mathbb{P}\left(\exists t_0\in[0,\tau], \|S_N(t_0)^{-1} - S(t_0)^{-1}\|_{\rm op} \ge (\mathfrak{c}_1/2)N^{\mathfrak{c}_2}\right).$$

We use Theorem 4 (of the Appendix) to write that, if $\|S(t)^{-1}(S_N(t)-S(t))\|_{\rm op} < 1$,

$$\|S_N(t)^{-1} - S(t)^{-1}\|_{\rm op} \le \frac{f_\tau^2\|S_N(t)-S(t)\|_{\rm op}}{1-\|S(t)^{-1}(S_N(t)-S(t))\|_{\rm op}}. \tag{32}$$

So we split the event

$$A_t := \{\|S_N(t)^{-1} - S(t)^{-1}\|_{\rm op} \ge c'\} = B_t \cup C_t$$

where

$$B_t = \{\|S_N(t)^{-1} - S(t)^{-1}\|_{\rm op} \ge c',\ \|S(t)^{-1}(S_N(t)-S(t))\|_{\rm op} < 1/2\}$$

and

$$C_t = \{\|S_N(t)^{-1} - S(t)^{-1}\|_{\rm op} \ge c',\ \|S(t)^{-1}(S_N(t)-S(t))\|_{\rm op} \ge 1/2\}.$$

We have

$$\begin{aligned}\mathbb{P}(\exists t_0\in[0,\tau], \|S_N(t_0)^{-1} - S(t_0)^{-1}\|_{\rm op} \ge c') &= \mathbb{P}(\exists t_0\in[0,\tau] \text{ such that } A_{t_0} \text{ holds})\\ &\le \mathbb{P}(\exists t_0\in[0,\tau] \text{ such that } B_{t_0} \text{ holds})\\ &\quad +\mathbb{P}(\exists t_0\in[0,\tau] \text{ such that } C_{t_0} \text{ holds})\end{aligned}$$

Now with (32),

$$\mathbb{P}(\exists t_0\in[0,\tau] \text{ such that } B_{t_0}\text{holds}) \le \mathbb{P}(\exists t_0\in[0,\tau], \|S_N(t_0)-S(t_0)\|_{\rm op} \ge c'/(2f_\tau^2))$$

and by keeping only the second constraint in the other case,

$$\mathbb{P}(\exists t_0\in[0,\tau] \text{ such that } C_{t_0}\text{holds}) \le \mathbb{P}(\exists t_0\in[0,\tau], \|S_N(t_0)-S(t_0)\|_{\rm op} \ge 1/(2f_\tau)).$$

As a consequence,

$$\mathbb{P}(\Lambda_N^c) \le \mathbb{P}(\sup_{t\in[0,\tau]}\|S_N(t)-S(t)\|_{\rm op} \ge \mathfrak{c}_1 N^{\mathfrak{c}_2}/(2f_\tau^2)) + \mathbb{P}(\sup_{t\in[0,\tau]}\|S_N(t)-S(t)\|_{\rm op} \ge 1/(2f_\tau))$$

From the Proof of Theorem 1 and (31), we have for any $p>1$,

$$\mathbb{E}\left(\sup_{t\in[0,\tau]}\|S_N(t)-S(t)\|_{\rm op}^{2p}\right) \le C(p,K,\tau)N^{-p}.$$

Therefore, for any $p>1$, $\mathbb{P}(\Lambda_N^c) \le c_0 N^{-p}$ and the first part of Proposition 4 follows by applying the Markov Inequality.

Next, on Λ_N, we have that $\forall t\in[0,\tau]$,

$$\lambda_{\max}(S_N^{-1}(t)) = \frac{1}{\lambda_{\min}(S_N(t))} \le \mathfrak{c}_1 N^{\mathfrak{c}_2}.$$

($\lambda_{\max}(M)$, resp. $\lambda_{\min}(M)$, denotes the maximal, resp. minimal, eigenvalue of matrix M). In the same way, $\lambda_{\max}(\widehat{\Psi}_m^{-1}) = 1/\lambda_{\min}(\widehat{\Psi}_m)$.

As for any $\mathbf{x} \in \mathbb{R}^{|\mathbf{m}|}$, and for $\mathbf{h}$ associated with $\mathbf{x}$ as in Lemma 1,

$$\mathbf{x}^T \widehat{\Psi}_{\mathbf{m}} \mathbf{x} = \int_0^\tau \mathbf{h}(t)^T S_N(t) \mathbf{h}(t) dt,$$

it follows that on Λ_N, for any eigenvalue λ of $\widehat{\Psi}_{\mathbf{m}}$ and any eigenvector $\mathbf{x}$ with norm equal to 1,

$$\lambda = \mathbf{x}^T \widehat{\Psi}_{\mathbf{m}} \mathbf{x} = \int_0^\tau \mathbf{h}(t)^T S_N(t) \mathbf{h}(t) dt \geq \int_0^\tau \lambda_{\min}(S_N(t)) \|\mathbf{h}(t)\|^2 dt \geq \mathfrak{c}_1^{-1} N^{-\mathfrak{c}_2}.$$

Thus $\lambda_{\min}(\widehat{\Psi}_m) \geq \mathfrak{c}_1^{-1} N^{-\mathfrak{c}_2}$ and it follows that $\|\widehat{\Psi}_{\mathbf{m}}^{-1}\|_{\text{op}} \leq \mathfrak{c}_1 N^{\mathfrak{c}_2}$. $\square$

6.6. **Proof of Theorem 2.** We start with some preliminaries.

6.6.1. *General orthogonal projection w.r.t.* $\langle ., . \rangle_N$. To study the risk of $\widetilde{\mathbf{a}}_{\mathbf{m}}$, we need to have an adequate expression of the orthogonal projection of $\mathbf{a}$ with respect to $\langle ., . \rangle_N$. Let

$$\Phi_{m_1+\cdots+m_{j-1}+k} = (\underbrace{0, \ldots, 0}_{j-1}, \varphi_k, \underbrace{0, \ldots, 0}_{K-j})^T, \quad j = 1, \ldots, K, k = 1, \ldots, m_j$$

The functions $(\Phi_j, j = 1, \ldots, |\mathbf{m}|)$ constitute an orthonormal system of $(\mathbb{L}^2_\tau)^K$ with respect to the scalar product $\langle \mathbf{h}, \mathbf{h}^\star \rangle = \int_0^\tau \sum_{j=1}^K h_j(t) h_j^\star(t)\, dt$ and generate a space $\mathbf{S}_{|\mathbf{m}|}$ (isomorphic to $S_{\mathbf{m}}$) with dimension $|\mathbf{m}| = m_1 + \cdots + m_K$. An element $\mathbf{h} = (h_1, \ldots, h_K)^T$ of $\mathbf{S}_{|\mathbf{m}|}$ can be written as

$$\mathbf{h}(t) = \sum_{i=1}^{|\mathbf{m}|} a_i \Phi_i = (\sum_{i=1}^{m_1} a_i \varphi_i, \sum_{i=1}^{m_2} a_{m_1+i}\ \varphi_i, \ldots, \sum_{i=1}^{m_K} a_{m_1+\ldots,+m_{K-1}+i}\ \varphi_i)^T$$

We have:

$$\widehat{\Psi}_{\mathbf{m}} = \left(\langle \Phi_j, \Phi_\ell \rangle_N\right)_{0 \leq j, \ell \leq |\mathbf{m}|}.$$

Indeed, if $m_1 + \ldots + m_{k-1} + j \leq m_1 + \cdots + m_k$ and $m_1 + \ldots + m_{k'-1} + \ell \leq m_1 \cdots + m_{k'}$,

$$\langle \Phi_j, \Phi_\ell \rangle_N = \Psi_{m_k m_{k'}}.$$

The orthogonal projection $\pi_{\mathbf{m}} \mathbf{a}$ of $\mathbf{a}$ on $\mathbf{S}_{|\mathbf{m}|}$ with respect to the scalar product $\langle ., . \rangle_N$ is characterized by $\pi_{\mathbf{m}} \mathbf{a} - \mathbf{a} \perp \Phi_j, j = 1, \ldots |\mathbf{m}|$. This yields

$$\pi_{\mathbf{m}} \mathbf{a} = \sum_{j=1}^{|\mathbf{m}|} a_j \Phi_j \quad \text{where} \quad \begin{pmatrix} a_1 \\ \vdots \\ a_{|\mathbf{m}|} \end{pmatrix} = \widehat{\Psi}_{\mathbf{m}}^{-1} \begin{pmatrix} \vdots \\ \langle \mathbf{a}, \Phi_j \rangle_N \\ \vdots \end{pmatrix}_{1 \leq j \leq |\mathbf{m}|}. \tag{33}$$

The vector $V_{\mathbf{m}} = (V_{1,m_1}^T, \ldots, V_{K,m_K}^T)^T$ can be written as

$$V_{\mathbf{m}} = \begin{pmatrix} \vdots \\ \langle \mathbf{a}, \Phi_j \rangle_N \\ \vdots \end{pmatrix}_{0 \leq j \leq |\mathbf{m}|} + \mathbb{W}_{\mathbf{m}}, \quad \mathbb{W}_{\mathbf{m}} := \frac{1}{N} \begin{pmatrix} \vdots \\ \int_0^\tau \Phi_j(t)^T d\mathbf{M}_N(t) \\ \vdots \end{pmatrix}_{0 \leq j \leq |\mathbf{m}|}. \tag{34}$$

where

$$\mathbf{M}_N(\tau) = \left(\int_0^\tau \sum_{i=1}^N g_j(X_i(t)) \sigma(t, X_i(t)) dW_i(t) \right)_{1 \leq j \leq K}.$$

Note that, recalling the definition of $\widehat{\Theta}_{\mathbf{m}}$ given in (25), we have

$$\mathbb{E} \mathbb{W}_{\mathbf{m}} \mathbb{W}_{\mathbf{m}}^T = \frac{1}{N} \mathbb{E} \widehat{\Theta}_{\mathbf{m}} := \frac{1}{N} \Theta_{\mathbf{m}}, \tag{35}$$

and $\widehat{\Theta}_{\mathbf{m}} = \widehat{\Psi}_{\mathbf{m}}$ if $\sigma \equiv 1$. The matrices $\widehat{\Theta}_{\mathbf{m}}$ and $\Theta_{\mathbf{m}}$ are symmetric and nonnegative matrices with

$$\mathbf{x}^\tau \Theta_{\mathbf{m}} \mathbf{x} = \int_0^\tau \mathbb{E}\left(\sum_{j=1}^K \underbrace{\sum_{p=1}^{m_j} x_{j,p}\varphi_p(t)}_{h_j(t)} g_j(X_1(t))\sigma(t, X_1(t))\right)^2 dt \geq 0.$$

6.6.2. *A useful Lemma.*

Lemma 3. *Assume* **[H1]** *to* **[H3]**. *Define the set*

$$\Omega_{\mathbf{m}} := \left\{ \left| \frac{\|\mathbf{h}\|_N^2}{\|\mathbf{h}\|_\tau^2} - 1 \right| \leq \frac{1}{2}, \forall \mathbf{h} \in \mathbf{S}_{\mathbf{m}} \right\}. \tag{36}$$

where the empirical norm $\|.\|_N$ and the $\|\cdot\|_\tau$-norm are equivalent for elements of $\mathbf{S}_{\mathbf{m}}$. We have $\mathcal{O}_N \subset \Omega_{\mathbf{m}}$ for all $\mathbf{m}$, and

$$\Omega_{\mathbf{m}} = \left\{ \|\Psi_{\mathbf{m}}^{-1/2}\widehat{\Psi}_{\mathbf{m}}\Psi_{\mathbf{m}}^{-1/2} - \mathrm{Id}_{|\mathbf{m}|}\|_{\mathrm{op}} \leq 1/2 \right\}. \tag{37}$$

Proof of Lemma 3. The inclusion $\mathcal{O}_N \subset \Omega_{\mathbf{m}}$ follows from Theorem 1. On $\Omega_{\mathbf{m}}$, $\forall \mathbf{h} \in \mathbf{S}_{\mathbf{m}}$, $(2/3)\|\mathbf{h}\|_N^2 \leq \|h\|_\tau^2 \leq 2\|\mathbf{h}\|_N^2$. If $\mathbf{x}^T = (x_0, \dots, x_{|\mathbf{m}|}) \in \mathbb{R}^{|\mathbf{m}|}$ and $\mathbf{h} = (\sum_{j=1}^{m_1} x_j\varphi_j, \dots, \sum_{j=1}^{m_K} x_{m_1+\dots+m_{K-1}+j}\varphi_j$ then

$$\|\mathbf{h}\|_N^2 = \mathbf{x}^T\widehat{\Psi}_{\mathbf{m}}\mathbf{x} \quad \text{and} \quad \|\mathbf{h}\|_\tau^2 = \mathbf{x}^T\Psi_{\mathbf{m}}\mathbf{x} = \|\Psi_{\mathbf{m}}^{1/2}\mathbf{x}\|_{2,|\mathbf{m}|}^2, \quad \text{so that} \tag{38}$$

$$\begin{aligned}
\sup_{\mathbf{h}\in\mathbf{S}_{\mathbf{m}}, \|\mathbf{h}\|_\tau=1} \left| \|\mathbf{h}\|_N^2 - \|\mathbf{h}\|_\tau^2 \right| &= \sup_{\vec{x}\in\mathbb{R}^{|\mathbf{m}|}, \|\Psi_{\mathbf{m}}^{1/2}\mathbf{x}\|_{2,|\mathbf{m}|}=1} \left| \mathbf{x}^T(\widehat{\Psi}_{\mathbf{m}} - \Psi_{\mathbf{m}})\mathbf{x} \right| \\
&= \sup_{\mathbf{u}\in\mathbb{R}^{|\mathbf{m}|}, \|\mathbf{u}\|_{2,|\mathbf{m}|}=1} \left| \mathbf{u}^T\Psi_{\mathbf{m}}^{-1/2}(\widehat{\Psi}_{\mathbf{m}} - \Psi_{\mathbf{m}})\Psi_{\mathbf{m}}^{-1/2}\vec{u} \right| \\
&= \|\Psi_{\mathbf{m}}^{-1/2}\widehat{\Psi}_{\mathbf{m}}\Psi_{\mathbf{m}}^{-1/2} - \mathrm{Id}_{|\mathbf{m}|}\|_{\mathrm{op}}.
\end{aligned}$$

Therefore, we get (37). This ends the proof of Lemma 3.□

Now we prove inequalities (26) and (27).

6.6.3. *Proof of inequality* (26). We write, with $\mathbf{a}(t) = (\alpha_j(t), j = 1, \dots, k)$,

$$\begin{aligned}
\|\widetilde{\mathbf{a}}_{\mathbf{m}} - \mathbf{a}\|_N^2 &= \|\widehat{\mathbf{a}}_{\mathbf{m}} - \mathbf{a}\|_N^2 \mathbf{1}_{\Lambda_N} + \|\mathbf{a}\|_N^2 \mathbf{1}_{\Lambda_N^c} \\
&= \|\widehat{\mathbf{a}}_{\mathbf{m}} - \mathbf{a}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N} + \|\widehat{\mathbf{a}}_{\mathbf{m}} - \mathbf{a}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N^c} + \|\mathbf{a}\|_N^2 \mathbf{1}_{\Lambda_N^c} \\
&:= T_1 + T_2 + T_3.
\end{aligned} \tag{39}$$

• Consider the last term $T_3 = \|\mathbf{a}\|_N^2 \mathbf{1}_{\Lambda_N^c}$. We have $\mathbb{E}T_3 \leq \mathbb{E}^{1/2}(\|\mathbf{a}\|_N^4)\mathbb{P}^{1/2}(\Lambda_N^c)$ where

$$\|\mathbf{a}\|_N^2 = \frac{1}{N}\sum_{i=1}^N \int_0^\tau \left(\sum_{j=1}^K \alpha_j(t)g_j(X_i(t))\right)^2 dt.$$

Thus,

$$\begin{aligned}\mathbb{E}[\|\mathbf{a}\|_N^4] &\leq \tau\int_0^\tau \left(\sum_{j=1}^K \alpha_j^2(t)\right)^2 \mathbb{E}\left[\left(\sum_{j=1}^K g_j^2(X_1(t))\right)^2\right] dt\\ &\leq K\tau\int_0^\tau \left(\sum_{k=1}^K \alpha_k^2(t)\right)^2 dt \sum_{j=1}^K \mathbb{E}\left(\sup_{t\in[0,\tau]} g_j^4(X_1(t))\right) := c_K(\tau).\end{aligned}$$

Then Proposition 4 implies $\mathbb{E}T_3 \lesssim \frac{1}{N^{p/2}} \lesssim \frac{1}{N}$ for $p \geq 2$.

• Let us now study of $T_1 = \|\widehat{\mathbf{a}}_{\mathbf{m}} - \mathbf{a}\|_N^2 \mathbf{1}_{\Lambda_N\cap\mathcal{O}_N}$. We can write:

$$\|\widehat{\mathbf{a}}_{\mathbf{m}} - \mathbf{a}\|_N^2 = \|\widehat{\mathbf{a}}_{\mathbf{m}} - \pi_{\mathbf{m}}\mathbf{a}\|_N^2 + \|\pi_{\mathbf{m}}\mathbf{a} - \mathbf{a}\|_N^2 = \|\widehat{\mathbf{a}}_{\mathbf{m}} - \pi_{\mathbf{m}}\mathbf{a}\|_N^2 + \inf_{h\in\mathbf{S}_{\mathbf{m}}} \|\mathbf{a} - h\|_N^2. \tag{40}$$

On one hand, we have $\widehat{\mathbf{a}}_{\mathbf{m}} = \sum_{j=1}^{|\mathbf{m}|}[\widehat{\mathbf{A}}_{\mathbf{m}}]_j\Phi_j$ with $\widehat{\mathbf{A}}_{\mathbf{m}}^T = (\widehat{\alpha}_{1,1},\ldots,\widehat{\alpha}_{1,m_1},\ldots,\ldots,\widehat{\alpha}_{K,m_K}) = \widehat{\Psi}_{\mathbf{m}}^{-1}V_{\mathbf{m}}$. On the other hand, $\pi_{\mathbf{m}}\mathbf{a} = \sum_{j=1}^{\mathbf{M}} a_j\Phi_j$ where (see (33)) $A_{\mathbf{m}} = (a_1,\ldots,a_{|\mathbf{m}|})^T = \widehat{\Psi}_{\mathbf{m}}^{-1}\left(\langle\Phi_j, b\rangle_N\right)_{1\leq j\leq|\mathbf{m}|}$.

Hence, by (34), $\widehat{\mathbf{A}}_{\mathbf{m}} - A_{\mathbf{m}} = \widehat{\Psi}_{\mathbf{m}}^{-1}\mathbb{W}_{\mathbf{m}}$ and using (38),

$$\|\widehat{\mathbf{a}}_{\mathbf{m}} - \pi_{\mathbf{m}}\mathbf{a}\|_N^2 = (\mathbb{W}_{\mathbf{m}})^T\widehat{\Psi}_{\mathbf{m}}^{-1}\widehat{\Psi}_{\mathbf{m}}\widehat{\Psi}_{\mathbf{m}}^{-1}\mathbb{W}_{\mathbf{m}} = (\mathbb{W}_{\mathbf{m}})^T\widehat{\Psi}_{\mathbf{m}}^{-1}\mathbb{W}_{\mathbf{m}}. \tag{41}$$

Recall that by Lemma 3, $\mathcal{O}_N \subset \Omega_{\mathbf{m}}$. On $\Omega_{\mathbf{m}} = \left\{\|\Psi_{\mathbf{m}}^{-1/2}\widehat{\Psi}_{\mathbf{m}}\Psi_{\mathbf{m}}^{-1/2} - \mathrm{Id}_{|\mathbf{m}|}\|_{\mathrm{op}} \leq 1/2\right\}$, all the eigenvalues of $\Psi_{\mathbf{m}}^{-1/2}\widehat{\Psi}_{\mathbf{m}}\Psi_{\mathbf{m}}^{-1/2}$ belong to $[1/2, 3/2]$ and so all the eigenvalues of $\Psi_{\mathbf{m}}^{1/2}\widehat{\Psi}_{\mathbf{m}}^{-1}\Psi_{\mathbf{m}}^{1/2}$ belong to $[2/3, 2]$. Thus, we write

$$\begin{aligned}(\mathbb{W}_{\mathbf{m}})^T\widehat{\Psi}_{\mathbf{m}}^{-1}\mathbb{W}_{\mathbf{m}}\,\mathbf{1}_{\mathcal{O}_N} &= (\mathbb{W}_{\mathbf{m}})^T\Psi_{\mathbf{m}}^{-1/2}\Psi_{\mathbf{m}}^{1/2}\widehat{\Psi}_{\mathbf{m}}^{-1}\Psi_{\mathbf{m}}^{1/2}\Psi_{\mathbf{m}}^{-1/2}\mathbb{W}_{\mathbf{m}}\,\mathbf{1}_{\mathcal{O}_N}\\ &\leq 2(\mathbb{W}_{\mathbf{m}})^T\Psi_{\mathbf{m}}^{-1}\mathbb{W}_{\mathbf{m}}\,\mathbf{1}_{\mathcal{O}_N}.\end{aligned} \tag{42}$$

Therefore , by using equality (35),

$$\begin{aligned}\mathbb{E}\left(\|\widehat{\mathbf{a}}_{\mathbf{m}} - \pi_{\mathbf{m}}\mathbf{a}\|_N^2\mathbf{1}_{\mathcal{O}_N\cap\Lambda_N}\right) &\leq 2\mathbb{E}\left(\sum_{1\leq j,k\leq\mathbf{M}}[\mathbb{W}_{\mathbf{m}}]_j[\mathbb{W}_{\mathbf{m}}]_k[\Psi_{\mathbf{m}}^{-1}]_{j,k}\right)\\ &= \frac{2}{N}\sum_{1\leq j,k\leq\mathbf{M}}[\Psi_{\mathbf{m}}^{-1}]_{j,k}[\Theta_{\mathbf{m}}]_{j,k} = \frac{2}{N}\mathrm{Tr}[\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}}],\end{aligned} \tag{43}$$

So we obtain:

$$\begin{aligned}\mathbb{E}(T_1) &\leq \mathbb{E}(\inf_{h\in\mathbf{S}_{\mathbf{m}}}\|\mathbf{a} - h\|_N^2) + \frac{2}{N}\mathrm{Tr}[\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}}]\\ &\leq \inf_{h\in\mathbf{S}_{\mathbf{m}}}\|\mathbf{a} - h\|_\tau^2 + \frac{2}{N}\mathrm{Tr}[\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}}],\end{aligned}$$

where the second term of the right-hand-side (rhs) above is the variance term appearing in (26).

• Finally, let us study of $T_2 = \|\widehat{\mathbf{a}}_{\mathbf{m}} - \mathbf{a}\|_N^2\mathbf{1}_{\Lambda_N\cap\mathcal{O}_N^c}$. We have $T_2 \leq (\|\widehat{\mathbf{a}}_{\mathbf{m}} - \pi_{\mathbf{m}}\mathbf{a}\|_N^2 + \|\mathbf{a}\|_N^2)\mathbf{1}_{\Lambda_N\cap\mathcal{O}_N^c}$. Using (41) yields

$$T_2 \leq (\mathbb{W}_{\mathbf{m}}^T\widehat{\Psi}_{\mathbf{m}}^{-1}\mathbb{W}_{\mathbf{m}} + \|\mathbf{a}\|_N^2)\mathbf{1}_{\Lambda_N\cap\mathcal{O}_N^c}. \tag{44}$$

By Proposition 4 about Λ_N and the Cauchy-Schwarz inequality, we get,

$$\mathbb{E}T_2 \leq \left(2\mathfrak{c}_1 N^{\mathfrak{c}_2}\mathbb{E}^{1/2}((\mathbb{W}_{\mathbf{m}}^T\mathbb{W}_{\mathbf{m}})^2) + \mathbb{E}^{1/2}\|\mathbf{a}\|_N^4\right)\mathbb{P}^{1/2}(\mathcal{O}_N^c). \tag{45}$$

We have already seen that $\mathbb{E}(\|\mathbf{a}\|_N^4) \leq c_K(\tau)$. For the term $\mathbb{E}[(\mathbb{W}_{\mathbf{m}}^T\mathbb{W}_{\mathbf{m}})^2]$, we prove the following:

Lemma 4. *With $\mathbb{W}_{\mathbf{m}}$ defined in* (34)*, we have, for some constant $c(\tau)$, if the φ_js are bounded:*

$$\mathbb{E}[(\mathbb{W}_{\mathbf{m}}^T\mathbb{W}_{\mathbf{m}})^2] \leq c(\tau)\frac{|\mathbf{m}|\sum_{j=1}^K L(S_{m_j})}{N^2}.$$

Otherwise,

$$\mathbb{E}[(\mathbb{W}_{\mathbf{m}}^T\mathbb{W}_{\mathbf{m}})^2] \leq c(\tau)\frac{|\mathbf{m}|\left(\sum_{j=1}^K L(S_{m_j})\right)^2}{N^2}.$$

Plugging the result of Lemma 4 in (45) allows to conclude, with Inequality (12), that, for all $\mathbf{m}$ satisfying $|\mathbf{m}| \leq N$, $\mathbb{E}(T_2) \lesssim N^{\mathfrak{c}_2+(1/2)-(p/2} \leq N^{-1}$, for $p \geq 2\mathfrak{c}_2 + 3$.
Joining the bounds for the expectations of T_1, T_2, T_3 gives Inequality (26) by choosing $p \geq 2\mathfrak{c}_2+3$.
□

Proof of Lemma 4. Using (34) yields

$$\mathbb{E}[\mathbb{W}_{\mathbf{m}}^T\,\mathbb{W}_{\mathbf{m}}]^2 \;=\; \frac{1}{N^4}\mathbb{E}\left[\sum_{j=1}^{|\mathbf{m}|}\left(\int_0^\tau \Phi_j(t)^T dM_N(t)\right)^2\right]^2 \leq \frac{|\mathbf{m}|}{N^4}\sum_{j=1}^{|\mathbf{m}|}\mathbb{E}\left(\int_0^\tau \Phi_j(t)^T dM_N(t)\right)^4.$$

Now, for $j = 1,\dots,K$, and $k = 1,\dots,m_j$,

$$\int_0^\tau \Phi_{m_1+\cdots+m_{j-1}+k}(t)^T dM_N(t) = \int_0^\tau \varphi_k(t)\sum_{i=1}^N g_j(X_i(t))\sigma(t,X_i(t))dW_i(t).$$

Therefore, using the Burholder-Davies-Gundy inequality yields

$$\begin{aligned}\mathbb{E}[\mathbb{W}_{\mathbf{m}}^T\mathbb{W}_{\mathbf{m}}]^2 &\lesssim \frac{|\mathbf{m}|}{N^4}\left[\sum_{j=1}^K\sum_{k=1}^{m_j}\mathbb{E}\left[\int_0^\tau \varphi_k^2(t)\sum_{i=1}^N g_j^2(X_1(t))\sigma^2(t,X_i(t))dt\right]^2\right]\\ &\leq \frac{\tau|\mathbf{m}|}{N^2}\left(\int_0^\tau\sum_{j=1}^K\sum_{k=1}^{m_j}\varphi_k^4(t)\mathbb{E}(g_j^4(X_1(t))\sigma^4(t,X_1(t))dt\right)\end{aligned}$$

For bounded φ_js, i.e. $|\varphi_j(t)| \leq C_\varphi$, $\forall t\in[0,\tau]$ and under **[H1]** and **(H2]** (see (7)), we obtain

$$\mathbb{E}[\mathbb{W}_{\mathbf{m}}^T\,\mathbb{W}_{\mathbf{m}}]^2 \lesssim \frac{|\mathbf{m}|}{N^2}(\sum_{j=1}^K L(S_{m_j})),$$

as $\sum_{k=1}^{m_j}\varphi_k^4(t) \leq C_\varphi^2\sum_{k=1}^{m_j}\varphi_k^2(t) \leq C_\varphi^2 L(S_{m_j})$.
Without using that the φ_js are bounded, we have $\sum_{j=1}^{m_j}\varphi_j^4(t) \leq (\sum_{j=1}^{m_j}\varphi_j^2(t))^2$ and we obtain

$$\mathbb{E}[\mathbb{W}_{\mathbf{m}}^T\,\mathbb{W}_{\mathbf{m}}]^2 \lesssim \frac{|\mathbf{m}|}{N^2}(\sum_{j=1}^K L^2(S_{m_j})) \lesssim \frac{|\mathbf{m}|}{N^2}(\sum_{j=1}^K L(S_{m_j}))^2,$$

which ends the proof of Lemma 4. □

6.6.4. *Proof of inequality* (27). Similarly to the previous bound, we write

$$\|\widetilde{\mathbf{a}}_{\mathbf{m}}-\mathbf{a}\|_\tau^2 = \|\widehat{\mathbf{a}}_{\mathbf{m}}-\mathbf{a}\|_\tau^2\mathbf{1}_{\Lambda_N\cap\mathcal{O}_N}+\|\widehat{\mathbf{a}}_{\mathbf{m}}-\mathbf{a}\|_\tau^2\mathbf{1}_{\Lambda_N\cap\mathcal{O}_N^c}+\|\mathbf{a}\|_\tau^2\mathbf{1}_{\Lambda_N^c} := T_1'+T_2'+T_3'. \tag{46}$$

It is straightforward that $\mathbb{E}(T_3')=\|\mathbf{a}\|_\tau^2\mathbb{P}(\Lambda_N^c)\lesssim 1/N^p$ for all $p>1$.

Now we turn to T_1'. Let $\mathbf{a}_{\mathbf{m},\tau}$ be the orthogonal projection of $\mathbf{a}$ on $\mathbf{S}_{\mathbf{m}}$ w.r.t. the τ-norm.We have

$$\begin{aligned}\|\widehat{\mathbf{a}}_{\mathbf{m}}-\mathbf{a}\|_\tau^2\mathbf{1}_{\mathcal{O}_N} &= \|\widehat{\mathbf{a}}_{\mathbf{m}}-\mathbf{a}_{\mathbf{m},\tau}\|_\tau^2\mathbf{1}_{\mathcal{O}_N}+\|\mathbf{a}_{\mathbf{m},\tau}-\mathbf{a}\|_\tau^2\mathbf{1}_{\mathcal{O}_N}\\ &\le \|\mathbf{a}_{\mathbf{m},\tau}-\mathbf{a}\|_\tau^2+2\|\widehat{\mathbf{a}}_{\mathbf{m}}-\pi_{\mathbf{m}}\mathbf{a}\|_\tau^2\mathbf{1}_{\mathcal{O}_N}+2\|\mathbf{a}_{\mathbf{m},\tau}-\pi_{\mathbf{m}}\mathbf{a}\|_\tau^2\mathbf{1}_{\mathcal{O}_N}\\ &\le \|\mathbf{a}_{\mathbf{m},\tau}-\mathbf{a}\|_\tau^2+4\|\widehat{\mathbf{a}}_{\mathbf{m}}-\pi_{\mathbf{m}}\mathbf{a}\|_N^2+2\|\mathbf{a}_{\mathbf{m},\tau}-\pi_{\mathbf{m}}\mathbf{a}\|_\tau^2\mathbf{1}_{\mathcal{O}_N}\end{aligned}$$

Thus

$$\mathbb{E}[T_1']\le\|\mathbf{a}_{\mathbf{m},\tau}-\mathbf{a}\|_\tau^2+\frac{4}{N}\mathrm{Tr}[\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}}]+2\mathbb{E}\left(\|\mathbf{a}_{\mathbf{m},\tau}-\pi_{\mathbf{m}}\mathbf{a}\|_\tau^2\mathbf{1}_{\mathcal{O}_N}\right).$$

Now, as $\mathbf{a}_{\mathbf{m},\tau}$ and $\pi_{\mathbf{m}}\mathbf{a}$ belong to $\mathbf{S}_{\mathbf{m}}$,

$$\begin{aligned}\mathbb{E}\left(\|\mathbf{a}_{\mathbf{m},\tau}-\pi_{\mathbf{m}}\mathbf{a}\|_\tau^2\mathbf{1}_{\mathcal{O}_N}\right) &\le 2\mathbb{E}\left(\|\mathbf{a}_{\mathbf{m},\tau}-\pi_{\mathbf{m}}\mathbf{a}\|_N^2\right)=2\mathbb{E}\left(\|\pi_{\mathbf{m}}(\mathbf{a}-\mathbf{a}_{\mathbf{m},\tau})\|_N^2\right)\\ &\le 2\mathbb{E}\left(\|\mathbf{a}-\mathbf{a}_{\mathbf{m},\tau}\|_N^2\right)=2\|\mathbf{a}-\mathbf{a}_{\mathbf{m},\tau}\|_\tau^2\end{aligned}$$

It follows that

$$\mathbb{E}[T_1']\le 5\|\mathbf{a}_{\mathbf{m},\tau}-\mathbf{a}\|_\tau^2+\frac{4}{N}\mathrm{Tr}[\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}}].$$

Let us lastly consider $T_2'=\|\widehat{\mathbf{a}}_{\mathbf{m}}-\mathbf{a}\|_\tau^2\mathbf{1}_{\Lambda_N\cap\mathcal{O}_N^c}$ and write

$$T_2'\le 2(\|\widehat{\mathbf{a}}_{\mathbf{m}}\|_\tau^2+\|\mathbf{a}\|_\tau^2)\mathbf{1}_{\Lambda_N\cap\mathcal{O}_N^c}:=T_{2,1}'+T_{2,2}'.$$

Clearly $\mathbb{E}(T_{2,2}')\le\|\mathbf{a}\|_\tau^2\mathbb{P}(\mathcal{O}_N^c)\le KG^2\|\mathbf{a}\|^2N^{-p}$, by using Proposition 2. Analogously, it holds that $\|\widehat{\mathbf{a}}_{\mathbf{m}}\|_\tau^2\le KG^2\|\widehat{\mathbf{a}}_{\mathbf{m}}\|^2$. Now using formula (20), we get

$$\|\widehat{\mathbf{a}}_{\mathbf{m}}\|^2=\|\widehat{\mathbf{A}}_{\mathbf{m}}\|_{2,|\mathbf{m}|}^2\le\|\widehat{\Psi}_{\mathbf{m}}^{-1}\|_{\mathrm{op}}^2\|V_{\mathbf{m}}\|_{2,|\mathbf{m}|}^2.$$

By Proposition 4, on Λ_N, we have

$$\|\widehat{\Psi}_{\mathbf{m}}^{-1}\|_{\mathrm{op}}^2\le 4\mathfrak{c}_1{}^2N^{2\mathfrak{c}_2}.$$

As a consequence

$$\mathbb{E}[T_{2,1}']\le 4\mathfrak{c}_1{}^2N^{2\mathfrak{c}_2}\mathbb{E}^{1/2}(\|\mathbf{V}_{\mathbf{m}}\|_{2,|\mathbf{m}|}^4)\mathbb{P}^{1/2}(\mathcal{O}_N^c). \tag{47}$$

By formula (34), we write

$$\|\mathbf{V}_{\mathbf{m}}\|_{2,|\mathbf{m}|}^2\le 2\left(\sum_{j=1}^{|\mathbf{m}|}\langle\mathbf{a},\Phi_j\rangle_N^2+\|\mathbb{W}\|_{2,|\mathbf{m}|}^2\right).$$

By Lemma 4, we have a bound on $\mathbb{E}^{1/2}[\|\mathbb{W}\|_{2,|\mathbf{m}|}^4]\le C(\tau)\sqrt{|\mathbf{m}|}\sum_{j=1}^K L(S_{m_j})/N$ and under **[H4]**

$$\mathbb{E}^{1/2}[\|\mathbb{W}\|_{2,|\mathbf{m}|}^4]\lesssim|\mathbf{m}|^{3/2}/N\le|\mathbf{m}|^{1/2}\le N^{1/2} \tag{48}$$

as $|\mathbf{m}|\le N$. We have,

$$\|S_N(t)\|_{\mathrm{op}}\le\frac{1}{N}\sum_{i=1}^N\|S_{\mathbf{g}}(X_i(t))\|_{\mathrm{op}}=\frac{1}{N}\sum_{i=1}^N\mathrm{Tr}(S_{\mathbf{g}}(X_i(t)))=\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^K g_j^2(X_i(t))$$

and thus

$$\mathbb{E}(\|S_N(t)\|_{\rm op}^4) \le K^4 \max_{j\in\{1,\dots,K\}} \sup_{t\in[0,\tau]} \mathbb{E}(g_j^8(X_1(t))) := c_1(\tau,K).$$

Then we get

$$\mathbb{E}\left[\left(\sum_{j=1}^{|\mathbf{m}|}\langle \mathbf{a},\Phi_j\rangle_N^2\right)^2\right] \le |\mathbf{m}|\mathbb{E}\left[\sum_{j=1}^{|\mathbf{m}|}\langle \mathbf{a},\Phi_j\rangle_N^4\right]$$

and

$$\begin{aligned}\mathbb{E}\left(\langle \mathbf{a},\Phi_j\rangle_N^4\right) &\le \mathbb{E}\left(\int_0^\tau \|S_N(t)\|_{\rm op}\|\mathbf{a}(t)\|_{2,K}\,\|\Phi_j(t)\|_{2,K}dt\right)^4\\ &\le \tau^3\int_0^\tau \mathbb{E}(\|S_N(t)\|_{\rm op}^4)\|\mathbf{a}(t)\|_{2,K}^4\,\|\Phi_j(t)\|_{2,K}^4 dt\\ &\le \tau^3 c_1(\tau,K)\left(\sup_{t\in[0,\tau]}\sum_{k=1}^K \alpha_k^2(t)\right)^2\int_0^\tau \varphi_j^4(t)dt.\end{aligned}$$

Thus

$$\mathbb{E}\left[\left(\sum_{j=1}^{|\mathbf{m}|}\langle \mathbf{a},\Phi_j\rangle_N^2\right)^2\right] \le c\tau^3 c_1(\tau,K)\left(\sup_{t\in[0,\tau]}\sum_{k=1}^K \alpha_k^2(t)\right)|\mathbf{m}|^3 \tag{49}$$

Plugging (48) and (49) into (47) yields

$$\mathbb{E}[T'_{2,1}] \lesssim \frac{N^2|\mathbf{m}|^{3/2}}{\log^2(N)}\mathbb{P}^{1/2}(\mathcal{O}_N^c) \lesssim N^{2\mathfrak{c}_2+3/2-p/2}.$$

This term is less than $O(N^{-1})$ for $p \ge 4\mathfrak{c}_2+5$. Having bounded the expectations of T'_1, T'_2, T'_3 yields (27).

Now we bound $\mathrm{Tr}[\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}}]$. As $\Psi_{\mathbf{m}}^{-1}$ and $\Theta_{\mathbf{m}}$ are symmetric and nonnegative, we have (see Lemma 6 in Appendix Section 7):

$$\mathrm{Tr}[\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}}] \le \|\Psi_{\mathbf{m}}^{-1}\|_{op}\mathrm{Tr}[\Theta_{\mathbf{m}}] \le \mathfrak{f}_\tau\sum_{j=1}^K \mathrm{Tr}[\Theta_{m_jm_j}] \quad \text{where}$$

$$\mathrm{Tr}[\Theta_{m_jm_j}] = \sum_{p=1}^{m_j}\int_0^\tau \varphi_p^2(t)\mathbb{E}[g_j^2(X_1(t)\sigma^2(t,X_1(t))]dt \le m_j \sup_{j=1,\dots,K}\sup_{t\in[0,\tau]}\mathbb{E}[g_j^2(X_1(t)\sigma^2(t,X_1(t))].$$

Therefore,

$$\mathrm{Tr}[\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}}] \le \mathfrak{C}|\mathbf{m}|$$

with $\mathfrak{C} = \mathfrak{f}_\tau \sup_{j=1,\dots,K}\sup_{t\in[0,\tau]}\mathbb{E}[g_j^2(X_1(t)\sigma^2(t,X_1(t))]$.

The proof of Theorem 2 is now complete. □

6.7. **Proof of Corollary 1.**

Now, let us prove that, if σ is bounded, then $\mathrm{Tr}[\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}}] \le |\mathbf{m}|\|\sigma\|_\infty^2$.

We use the followig trick. Let $\varepsilon := (\varepsilon_i)_{1\le i\le |\mathbf{m}|}$ be a vector of i.i.d. centered variables with unit variance, independent of $(X_i(t))_{t\ge 0, 1\le i\le N}$. For any $|\mathbf{m}| \times |\mathbf{m}|$ matrix C, it holds that $\mathrm{Tr}(C) = \mathbb{E}(\varepsilon^T C\varepsilon)$. Therefore

$$\mathrm{Tr}[\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}}] = \mathrm{Tr}[\Psi_{\mathbf{m}}^{-1/2}\Theta_{\mathbf{m}}\Psi_{\mathbf{m}}^{-1/2}] = \mathbb{E}\left(\varepsilon^T\Psi_{\mathbf{m}}^{-1/2}\Theta_{\mathbf{m}}\Psi_{\mathbf{m}}^{-1/2}\varepsilon\right).$$

Setting $\mathbf{x} = \Psi_{\mathbf{m}}^{-1/2}\varepsilon$ yields

$$\begin{aligned}&\mathbf{x}^T\Theta_{\mathbf{m}}\mathbf{x}\\ &= \sum_{1\le j,\ell\le K}\ \sum_{\substack{1\le k\le m_j\\ 1\le p\le m_\ell}} \mathbf{x}_{m_1+\dots+m_{j-1}+k}\mathbf{x}_{m_1+\dots+m_{\ell-1}+p}\int_0^\tau \varphi_k(t)\varphi_p(t)\mathbb{E}\left(g_j(X_1(t))g_\ell(X_1(t))\sigma^2(t,X_1(t))\right)dt\\ &= \int_0^\tau \mathbb{E}\left[\left(\sum_{j=1}^K h_j(t)g_j(X_1(t))\right)^2\sigma^2(t,X_1(t))\,\middle|\,\varepsilon\right]dt\end{aligned}$$

where $h_j(t) = \sum_{k=1}^{m_j} x_{m_1+\dots+m_{j-1}+k}\varphi_k(t)$ and $\mathbb{E}(.|\varepsilon)$ is the conditional expectation w.r.t. ε. Thus we get

$$\varepsilon^T\Psi_{\mathbf{m}}^{-1/2}\Theta_{\mathbf{m}}\Psi_{\mathbf{m}}^{-1/2}\varepsilon = \mathbf{x}^T\Theta_{\mathbf{m}}\mathbf{x} \le \|\sigma\|_\infty^2\int_0^T \mathbb{E}\left[\left(\sum_{j=1}^K h_j(t)g_j(X_1(t))\right)^2\,\middle|\,\varepsilon\right]dt.$$

Noticing that

$$\int_0^\tau \mathbb{E}\left[\left(\sum_{j=1}^K h_j(t)g_j(X_1(t))\right)^2\,\middle|\,\varepsilon\right]dt = \mathbf{x}^T\Psi_{\mathbf{m}}\mathbf{x} = \varepsilon^T\Psi_{\mathbf{m}}^{-1/2}\Psi_{\mathbf{m}}\Psi_{\mathbf{m}}^{-1/2}\varepsilon = \|\varepsilon\|_{2,|\mathbf{m}|}^2,$$

we obtain, by taking expectation,

$$\Psi_{\mathbf{m}}^{-1}\Theta_{\mathbf{m}} = \mathrm{Tr}[\Psi_{\mathbf{m}}^{-1/2}\Theta_{\mathbf{m}}\Psi_{\mathbf{m}}^{-1/2}] = \mathbb{E}\left(\varepsilon^T\Psi_{\mathbf{m}}^{-1/2}\Theta_{\mathbf{m}}\Psi_{\mathbf{m}}^{-1/2}\varepsilon\right) \le \|\sigma\|_\infty^2|\mathbf{m}|.$$

Hence, the result. □

6.8. **Proof of Theorem 3.** We write the decomposition

$$\|\widetilde{\mathbf{a}} - \mathbf{a}\|_N^2 = \|\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}} - \mathbf{a}\|_N^2\mathbf{1}_{\Lambda_N} + \|\mathbf{a}\|_N^2\mathbf{1}_{\Lambda_N^c}.$$

The study of the last term is similar to the study of T_2 , see (44)-(45), and yields $\mathbb{E}(\|\mathbf{a}\|_N^2\mathbf{1}_{\Lambda_N^c}) \le C/N$ thanks to Proposition 4, $\mathbb{P}(\Lambda_N^c) \lesssim 1/N^p$ for any $p > 2$.

For the main term $\mathbb{E}(\|\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}} - \mathbf{a}\|_N^2\mathbf{1}_{\Lambda_N})$, we recall that $U_N(\widehat{\mathbf{a}}_{\mathbf{m}}) = -\|\widehat{\mathbf{a}}_{\mathbf{m}}\|_N^2$. By definition of $\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}}$, we have for any $\mathbf{m} \in \mathcal{M}_N$, and any $\mathbf{a}_{\mathbf{m}} \in \mathbf{S}_{\mathbf{m}}$,

$$U_N(\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}}) + \mathrm{pen}(\widehat{\mathbf{m}}) \le U_N(\mathbf{a}_{\mathbf{m}}) + \mathrm{pen}(\mathbf{m}).$$

From (16), we have $U_N(\mathbf{h}) - U_N(\mathbf{h}^\star) = \|\mathbf{h} - \mathbf{a}\|_N^2 - \|\mathbf{h}^\star - \mathbf{a}\|_N^2 - 2\nu_N(\mathbf{h} - \mathbf{h}^\star)$ and therefore for any $\mathbf{m} \in \mathcal{M}_N$, and any $\mathbf{a}_{\mathbf{m}} \in \mathbf{S}_{\mathbf{m}}$, on Λ_N

$$\|\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}} - \mathbf{a}\|_N^2 \le \|\mathbf{a}_{\mathbf{m}} - \mathbf{a}\|_N^2 + \mathrm{pen}(\mathbf{m}) + 2\nu_N(\widehat{\mathbf{a}}_{\mathbf{m}} - \mathbf{a}_{\mathbf{m}}) - \mathrm{pen}(\widehat{\mathbf{m}}).$$

Now we define

$$B_{\mathbf{m},\mathbf{m}'} = \{\mathbf{h} \in \mathbf{S}_{\mathbf{m}} + \mathbf{S}_{\mathbf{m}'},\ \|\mathbf{h}\|_\tau = 1\}.$$

We have

$$\mathbb{E}(\|\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}}-\mathbf{a}\|_N^2\mathbf{1}_{\Lambda_N})=\mathbb{E}(\|\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}}-\mathbf{a}\|_N^2\mathbf{1}_{\Lambda_N\cap\mathcal{O}_N})+\mathbb{E}(\|\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}}-\mathbf{a}\|_N^2\mathbf{1}_{\Lambda_N\cap\mathcal{O}_N^c}).$$

The term $\mathbb{E}(\|\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}}-\mathbf{a}\|_N^2\mathbf{1}_{\Lambda_N\cap\mathcal{O}_N^c})$ is studied analogously as the previous term T_2, with still $\mathbb{P}(\mathcal{O}_N^c)\le c/N^p$ for all p.
Using that, on $\mathcal{O}_N$, $\forall\mathbf{m},\mathbf{m}'\in\{1,\dots,N\}^K,\forall\mathbf{h}\in\mathbf{S}_{\mathbf{m}}+\mathbf{S}_{\mathbf{m}'}$, $\|\mathbf{h}\|_\tau^2\le 2\|\mathbf{h}\|_N^2$, we obtain, on $\mathcal{O}_N\cap\Lambda_N$, the following sequence of inequalities.

$$\begin{aligned}\|\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}}-\mathbf{a}\|_N^2 &\le \|\mathbf{a}_{\mathbf{m}}-\mathbf{a}\|_N^2+\text{pen}(\mathbf{m})+\frac{1}{8}\|\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}}-\mathbf{a}_{\mathbf{m}}\|_\tau^2\\ &\quad +8\sup_{\mathbf{h}\in B_{\mathbf{m},\widehat{\mathbf{m}}}}\nu_N^2(\mathbf{h})-\text{pen}(\widehat{\mathbf{m}})\\ &\le \left(1+\frac{1}{2}\right)\|\mathbf{a}_{\mathbf{m}}-\mathbf{a}\|_N^2+\text{pen}(\mathbf{m})+\frac{1}{2}\|\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}}-\mathbf{a}\|_N^2\\ &\quad +8\left(\sup_{\mathbf{h}\in B_{\mathbf{m},\widehat{\mathbf{m}}}}\nu_N^2(\mathbf{h})-p(\widehat{\mathbf{m}},\mathbf{m})\right)_+ +8p(\widehat{\mathbf{m}},\mathbf{m})-\text{pen}(\widehat{\mathbf{m}}),\end{aligned}$$

where $p(\widehat{\mathbf{m}},\mathbf{m})=\kappa^\star\|\sigma\|_\infty^2(|\widehat{\mathbf{m}}|+|\mathbf{m}|)/N$, where $\kappa^\star$ is a numerical constant (see below). Choosing $\kappa_0\ge 8\kappa^\star$ implies that $8p(\widehat{\mathbf{m}},\mathbf{m})\le\text{pen}(\widehat{\mathbf{m}})+\text{pen}(\mathbf{m})$. Therefore

$$\mathbb{E}(\|\widehat{\mathbf{a}}_{\widehat{\mathbf{m}}}-\mathbf{a}\|_N^2\mathbf{1}_{\Lambda_N\cap\mathcal{O}_N})\le 3\|\mathbf{a}_{\mathbf{m}}-\mathbf{a}\|_\tau^2+4\text{pen}(\mathbf{m})+16\mathbb{E}\left[(\sup_{\mathbf{h}\in B_{\mathbf{m},\widehat{\mathbf{m}}}}\nu_N^2(\mathbf{h})-p(\widehat{\mathbf{m}},\mathbf{m}))_+\mathbf{1}_{\Lambda_N\cap\mathcal{O}_N}\right].$$

To exhibit the numerical value κ^* and achieve the proof of Theorem 3, first, we use Bernstein's inequality for continuous local martingales (see Revuz and Yor, 1999 p. 153): let $M_\tau=N\nu_N(\mathbf{h})$ and

$$\langle M\rangle_\tau=\sum_{i=1}^N\int_0^\tau\left[\sum_{k=1}^K h_k(t)g_k(X_i(t))\right]^2\sigma^2(t,X_i(t))dt.$$

Then,

$$\mathbb{P}\left(M_\tau\ge N\varepsilon,\langle M\rangle_\tau\le Nv^2\right)\le\exp\left(-\frac{N\varepsilon^2}{2v^2}\right).$$

For σ bounded, we have

$$\langle M\rangle_\tau\le N\|\sigma\|_\infty^2\|\mathbf{h}\|_N^2.$$

Therefore

$$\mathbb{P}\left(\nu_N(\mathbf{h})\ge\varepsilon,\|\mathbf{h}\|_N^2\le v^2\right)\le\exp\left(-\frac{N\varepsilon^2}{2\|\sigma\|_\infty^2v^2}\right).$$

This inequality implies that we can apply the $\mathbb{L}^2$-chaining method described in Baraud *et al.*, (2001), Proposition 6.1, p.42 and its proof p. 45-47, which yields that there exists a numerical constant $\kappa^\star$ such that

$$\mathbb{E}\left[\left(\sup_{\mathbf{h}\in B_{\widehat{\mathbf{m}},\mathbf{m}}}\nu_N^2(\mathbf{h})-p(\widehat{\mathbf{m}},\mathbf{m})\right)_+\mathbf{1}_{\widehat{\Lambda}_N\cap\mathcal{O}_N}\right]\le c\frac{\|\sigma\|_\infty^2}{N}$$

with $p(\mathbf{m},\mathbf{m}')=2\kappa\|\sigma\|_\infty^2\frac{|\mathbf{m}|+|\mathbf{m}'|}{N}$, $\kappa^*=2\kappa$ ($\kappa=38$ is the value given in the proof). □

References

[1] Baraud, Y., Comte, F. and Viennet, G. (2001). Model selection for (auto)-regression with dependent data. *ESAIM P & S*, **5**, 33-49.

[2] Cohen, A., Davenport, M.A. and Leviatan, D. (2019). Correction to: On the Stability and Accuracy of Least Squares Approximations. *Foundations of Computational Mathematics*, **19**, p.239.

[3] Cohen, A., Davenport, M.A. and Leviatan, D. (2013). On the Stability and Accuracy of Least Squares Approximations. *Foundations of Computational Mathematics*,**13**, 819-834.

[4] Comte, F. and Genon-Catalot, V. (2018). Laguerre and Hermite bases for inverse problems. *Journal of the Korean Statistical Society*, **47**, 273-296.

[5] Comte, F. and Genon-Catalot, V. (2020a). Regression function estimation on non compact support as a partly inverse problem. *Annals of the Institute of Statistical Mathematics* **72**, 1023-1054.

[6] Comte, F. and Genon-Catalot, V. (2020b). Nonparametric drift estimation for i.i.d. paths of stochastic differential equations. *Annals of Statistics* **48**, 3336-33365.

[7] Comte, F. and Genon-Catalot, V. (2023). Nonparametric adaptive estimation for interacting particle systems. *Scand. J. Stat.*, online version `https://doi.org/10.1111/sjos.12661`.

[8] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, **13**, 514-543.

[9] Comte, F. and Marie, N. (2023) Nonparametric drift estimation from diffusions with correlated Brownian motions. Preprint hal-03736082.

[10] Dalalyan, A. and Reiss, M. (2006). Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab. Theory Relat. Fields* **134**, 248-282.

[11] Dalalyan, A. and Reiss, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probability Theory and Related Fields*, **137**, 25-47.

[12] Della Maestra, L. and Hoffmann, M. (2022). Nonparametric estimation for interacting particle systems: McKean-Vlasov models. *Probability Theory and Related Fields*, **182**, 551-613.

[13] Denis, C., Dion, C. and Martinez, M. (2020). Consistent procedures for multiclass classification of discrete diffusion paths. *Scand. J. Stat.* **47**, 516-554.

[14] Denis, C., Dion-Blanc, C. and Martinez, M. (2021) A ridge estimator of the drift from discrete repeated observations of the solution of a stochastic differential equation. *Bernoulli* **27**, 2675-2713.

[15] Efromovich, S. (1999). *Nonparametric curve estimation. Methods, theory, and applications.* Springer Series in Statistics. Springer-Verlag, New York.

[16] Garsia, A. M., Rodemich, E. and Rumsey, Jr, H. (1970/71) A real variable lemma and the continuity of paths of some Gaussian processes. *Indiana Univ. Math. J.*, 20, 565-578.

[17] Gendre, X. (2014) Model selection and estimation of a component in additive regression. *ESAIM Probab. Stat.* **18**, 77-116.

[18] Gloter, A. (2000) Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Statist.* **4**, 205-227.

[19] Gloter, A. and Sørensen, M. (2009) Estimation for stochastic differential equations with a small diffusion coefficient. *Stochastic Process. Appl.* **119**, 679-699.

[20] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, 79, 135-163.

[21] Hsiao, C. (2003). *Analysis of panel data.* Cambridge University Press, Second Edition. Cambridge.

[22] Jourdain, B. and Pagès, G. (2022). Convex ordering for stochastic Volterra equations and their Euler schemes. Preprint arXiv:2211.10186v1.

[23] Iacus, S. M., (2010). *Simulation and inference for stochastic differential equations. With R examples.* Springer.

[24] Kessler, M., Lindner, A. and Sørensen, M., Editors (2012). *Statistical methods for stochastic differential equations.* CRC press. Taylor & Francis Group. Boca Raton.

[25] Kutoyants, Y.A., (1984). *Parameter estimation for stochastic processes.* Berlin: Heldermann.

[26] Kutoyants, Y.A., (2004). *Statistical inference for ergodic diffusion processes.* Springer, London.

[27] Marie, N. and Rosier, A. (2023) Nadaraya-Watson Estimator for i.i.d. Paths of Diffusion Processes. To appear in *Scand. J. Stat.*, **50**, 589-637.

[28] Ramsay, J.O. and Silverman, B.W. (2007). *Applied functional data analysis: Methods and case studies.* Springer.

[29] Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion.* Third edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 293. Springer-Verlag, Berlin.

[30] Stewart, G. W. and Sun, J.-G. (1990). *Matrix perturbation theory.* Boston etc.: Academic Press, Inc.

[31] Strauch, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *Ann. Statist.* **46**, 3451-3480.
[32] Tsybakov, A. B. (2009). *Introduction to nonparametric estimation.* Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. Springer Series in Statistics. Springer, New York.
[33] Wang, J.-L., Chiou, J.-M. and Mueller, H.-G. (2016). Functional data analysis, *Annual Review of Statistics and Its Application*, **3**, 257-295.

7. Appendix

7.1. **The Garsia-Rodemich-Rumsey (1970/71) Lemma.** We state the version of this lemma given in Jourdain and Pagès (2022).

Lemma 5. *Let $(Y_t^n)_{n\geq 1}$ be a sequence of continuous processes where the processes $Y^n=(Y_t^n)_{t\in[0,T]}$ are defined on a probability space $(\Omega,\mathcal{A},\mathbb{P})$. Let $p\geq 1$. Assume there exists $a>1$, a sequence $(\delta_n)_{n\geq 1}$ of positive real numbers converging to 0 and a real constant $C>0$ such that*

$$\forall n\geq 1, \forall s,t\in[0,T],\quad \mathbb{E}[|Y_t^n-Y_s^n|^p]\leq C|t-s|^a\delta_n^p.$$

Then there exists a real constant $C_{p,T}>0$ such that

$$\forall n\geq 1, \mathbb{E}\left(\sup_{t\in[0,T]}|Y_t^n-Y_0^n|^p\right)\leq C_{p,T}\delta_n^p.$$

7.2. **Useful results from linear algebra.** A proof of the following theorem can be found in Stewart and Sun (1990).

Theorem 4. *Let $\mathbf{A}$, $\mathbf{B}$ be $(m\times m)$ matrices. If $\mathbf{A}$ is invertible and $\|\mathbf{A}^{-1}\mathbf{B}\|_{\rm op}<1$, then $\tilde{\mathbf{A}}:=\mathbf{A}+\mathbf{B}$ is invertible and it holds*

$$\|\tilde{\mathbf{A}}^{-1}-\mathbf{A}^{-1}\|_{\rm op}\leq\frac{\|\mathbf{B}\|_{\rm op}\|\mathbf{A}^{-1}\|_{\rm op}^2}{1-\|\mathbf{A}^{-1}\mathbf{B}\|_{\rm op}} \tag{50}$$

The following Lemma is used in the proofs.

Lemma 6. *Let A,B be two symmetric nonnegative $d\times d$ matrices. Then,*

$$\mathrm{Tr}(AB)\leq\|A\|_{\rm op}\mathrm{Tr}(B) \tag{51}$$

Proof of Lemma 6. Since the matrices are symmetric, there exist two orthogonal matrices P,Q such that $A=P^TDP$, $B=Q\Delta Q^T$ where $D=\mathrm{diag}(\lambda_i(A)),\Delta=\mathrm{diag}(\lambda_i(B))$ are diagonal matrices with diagonal elements equal to the eigenvalues of A (resp. B). As the matrices are nonnegative, $\lambda_i(A)\geq 0,\lambda_i(B)\geq 0$ for all $i=1,\dots d$. Set $P=(p_{ij})$. We have $P^TQ\Delta Q^TP=P^T\mathrm{diag}(\lambda_i(B))P$ and

$$[P^TQ\Delta Q^TP]_{ii}=\sum_j\lambda_j(B)p_{ji}^2\geq 0,$$

which implies

$$\sum_i[P^TQ\Delta Q^TP]_{ii}=\sum_i\sum_j\lambda_j(B)p_{ji}^2=\sum_j\lambda_j(B)\sum_i p_{ji}^2=\sum_j\lambda_j(B)=\mathrm{Tr}(B).$$

Therefore, using nonnegativity of $\lambda_i(A)$ and $\lambda_i(B)$,

$$\mathrm{Tr}(AB)=\sum_i\lambda_i(A)[P^TQ\Delta Q^TP]_{ii}\leq\sup_i\lambda_i(A)\sum_i[PQ\Delta Q^TP^T]_{ii}\leq\|A\|_{\rm op}\mathrm{Tr}(B).\quad\square$$

7.3. **Examples of bases.** In the simulation section, we experimented two bases.

The trigonometric bases and spaces are defined as follows. Let us denote $(S_m^{Trig}, m \geq 0)$ the subspaces of $\mathbb{L}^2([0,\tau])$ such that S_m^{Trig} has odd dimension m and is generated by the orthonormal trigonometric basis. This basis is given by $(\varphi_{j,\tau})$ where $\varphi_{0,\tau}(t) = \sqrt{1/\tau}\mathbf{1}_{[0,\tau]}(t)$,

$$\varphi_{2j-1,\tau}(t) = \sqrt{2/\tau}\cos(2\pi jt/\tau)\mathbf{1}_{[0,\tau]}(t), \quad \varphi_{2j,\tau}(t) = \sqrt{2/\tau}\sin(2\pi jt/\tau)\mathbf{1}_{[0,\tau]}(t)$$

for $j = 1, \dots, (m-1)/2$. It is easy to see that

$$\sum_{j=0}^{m-1} \varphi_{j,\tau}^2(t) = \frac{m}{\tau} \quad \text{and} \quad L(S_m^{Trig}) = \sup_{x\in[0,\tau]} \sum_{j=0}^{m-1} \varphi_{j,\tau}^2(x) = \frac{m}{\tau}.$$

Those properties are adequate, but a function developped in this basis is such that its values at points 0 and τ are equal, and this is not adapted to our examples.

This is why we rather used a trigonometric basis called "half-trigonometric" system, namely the cosine basis defined by $\varphi_{0,T}(x) = \sqrt{1/T}\mathbf{1}_{[0,T]}(t)$, $\varphi_{j,T}(t) = \sqrt{2/T}\cos(\pi jt/T)\mathbf{1}_{[0,T]}(t)$, $j = 1, \dots, m-1$, see Efromovich (1999, p.46). It is clearly an orthonormal basis. For a twice differentiable function, the projection coefficients decrease like $1/j^2$ without border constraints; such constraints are required for higher regularities only, see Efromovich (1999, p.32). In practical implementation, it appears that this basis is more convenient and performant than the complete trigonometric basis.

We also used a basis which does not match to our theoretical conditions, but which revealed to work well while being parsimonious (few coefficients required for good estimation): the Laguerre basis (see Comte and Genon-Catalot (2018)) defined by

$$\ell_j(t) = \sqrt{2}L_j(2t)e^{-t}\mathbf{1}_{t\geq 0}, \quad j \geq 0, \quad L_j(t) = \sum_{k=0}^{j}(-1)^k \binom{j}{k}\frac{t^k}{k!}. \tag{52}$$

We set $S_m^{Lag} = \text{span}\{\ell_j, j = 0, \dots, m-1\}$. We have

$$\forall t \geq 0, \quad \sum_{j=0}^{m-1} \ell_j^2(t) \leq 2m, \quad \text{and} \quad L(S_m^{Lag}) \leq 2m.$$