NONPARAMETRIC ADAPTIVE ESTIMATION FOR INTERACTING PARTICLE SYSTEMS

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

Abstract. We consider a stochastic system $(X_i^N(t), i = 1, \ldots, N)$ of N interacting particles with constant diffusion coefficient and linear drift $b(t, x, \mu) = \alpha(t)x - \beta(t)\int(x-y)\mu(dy)$ depending on two unknown deterministic functions $\alpha(t), \beta(t)$. Our concern here is the nonparametric estimation of these functions from a continuous observation of the process on $[0, T]$ for fixed T and large N. We define two collections of projection estimators $\widetilde{\alpha}_m(t), \widetilde{\gamma}_p^\star(t)$ respectively of $\alpha(t), \gamma(t) = \alpha(t) - \beta(t)$ where for each m (resp. p), $\widetilde{\alpha}_m(t)$ (resp. $\widetilde{\gamma}_p^\star(t)$) belongs to a finite dimensional subspace of $\mathbb{L}^2([0, T])$. We study the $\mathbb{L}^2$-risks of these estimators where the risk is defined either by the expectation of an empirical norm or by the expectation of a deterministic norm. Afterwards, we propose a data-driven choice $\widehat{m}$ (resp. $p^\star$) of the value m (resp. p) and study the risk of the adaptive estimators. The case of $\beta(t) \equiv 0$ is also treated separately. The results are illustrated by numerical experiments on simulated data.

Keywords and phrases: Interacting particle systems, nonparametric inference, projection estimators, adaptive method.

June 16, 2022

1. Introduction

Stochastic systems of N interacting particles have received a lot of attention in the past decades. First arisen in Statistical Physics for the modelling of granular media (Benedetto *et al.*, 1997), these models progressively appear in many other fields of applications such as Mathematical Biology (Molginer and Edelstein-Keshet, 1999, Baladron *et al.*, 2012), Epidemics Dynamics (Britton *et al.*, 2020) or Finance (Giesecke *et al.*, 2020). The probabilistic properties of these models, especially their behaviour as N is large, have been largely studied (see *e.g.* among many references Méléard, 1996, Sznitman, 1991). On the contrary, the statistical inference for interacting particles remained unstudied for many years with the exception of Kasonga (1990) who studied the maximum likelihood estimation of $\theta = (\alpha, \beta)$ from the observation on the interval $[0, T]$ of the N-dimensional system given by: $dX_i^N(t) = \{\alpha X_i^N(t) - \beta[X_i^N(t) - \overline{X}_N^N(t)]\}dt + dW_i(t)$, with $X_i^N(0) = X_0^i, i = 1, \ldots, N$, and $\overline{X}_N^N(t) = N^{-1}\sum_{j=1}^N X_j^N(t)$, $(W_i, i = 1 \ldots, N)$ are N independent Brownian motions, $X_0^i, i = 1, \ldots, N$ are *i.i.d.* random variables independent of $(W_i, i = 1 \ldots, N)$. A multivariate version of Kasonga's model where $X_i^N(t) \in \mathbb{R}^d$ is studied in Chen (2021). The general model can be described as a N-dimensional stochastic differential equation of the form

$$dX_i^N(t) = b(t, X_i^N(t), \mu_N(t))dt + \sigma(t, X_i^N(t))dW_i(t)$$

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

where $\mu_N(t) = N^{-1}\sum_{j=1}^N \delta_{X_j^N(t)}$ is the empirical measure associated with $(X_i^N(t), i = 1,\ldots,N)$. The drift function is often modelled as $b(t,x,\mu)) = V(x) - \int \Phi(x-y)\mu(dy)$ and Φ represents the interaction term between particles. In this context, Lu *et al.* (2019) consider the nonparametric estimation of the interaction function in a deterministic system where $V \equiv 0$ and $\sigma \equiv 0$. Li *et al.* (2021) are interested in characterizing the identifiability of the interaction function. Sharrock *et al.* (2021) study a parametric model for the drift and estimation by maximum likelihood. In Pavliotis and Zanoni (2021), the point of view of martingale estimating equations is developed for parametric inference based on discrete observations of the system. Della Maestra and Hoffmann (2022) study the nonparametric estimation of the function $b(t,x,\mu)$ by a kernel approach. Belomestny *et al.* (2022) consider a semiparametric model for the interaction function and estimate the nonparametric part by a kernel approach.
In this paper, we consider a linear drift $b(t,x,\mu) = \alpha(t)x - \beta(t)\int(x-y)\mu(dy)$, *i.e.* $(X_i^N(t), i = 1,\ldots,N)$ is given by

$$(1)\quad dX_i^N(t) = \left\{\alpha(t)X_i^N(t) - \beta(t)[X_i^N(t) - \overline{X}_N^N(t)]\right\}dt + dW_i(t), \quad X_i^N(0) = X_0^i, i = 1,\ldots,N,$$

where $\alpha(t), \beta(t)$ are deterministic unknown functions and $\overline{X}_N^N(t)$ is the empirical mean of the sample, $(W_i, i = 1\ldots,N)$ are N independent Brownian motions, $X_0^i, i = 1,\ldots,N$ are *i.i.d.* random variables independent of $(W_i, i = 1\ldots,N)$, as in Kasonga (1990). Our concern here is the nonparametric estimation of the functions $(\alpha(t), \beta(t))$ from a continuous observation of the process the process $(X_i^N(t), i = 1,\ldots,N)$ on $[0,T]$ with fixed T and $N \to +\infty$. Note that this model was proposed by Bishwal (2011) as an extension of Kasonga's model. However, in this paper, nothing is done concerning the estimation of the functions $\alpha(t), \beta(t)$. If $\beta(t) \equiv 0$, the processes $X_i^N(t), i = 1,\ldots,N$ are independent. In this context, the nonparametric estimation of the unknown function $\alpha(t)$ by the method of sieves is considered in Nguyen and Pham (1982). If $\beta(t) \not\equiv 0$, then the N processes are no more independent and $(X_i^N(t), i = 1,\ldots,N)$ constitute a system of interacting particles. The interest of model (1) lies is the fact that, contrary to more general models, computations can be done explicitely.
Not surprisingly, two processes play a crucial role, the empirical mean and the empirical variance of $(X_i^N(t), i = 1,\ldots,N)$:

$$(2)\qquad X(t) = \overline{X}_N^N(t) = \frac{1}{N}\sum_{i=1}^N X_i^N(t),$$

$$(3)\qquad V(t) = V_N(t) = \frac{1}{N}\sum_{i=1}^N [X_i^N(t) - \overline{X}_N^N(t)]^2 = \frac{1}{N}\sum_{i=1}^N (X_i^N(t))^2 - (\overline{X}_N^N(t))^2$$

We prove that each one follows an autonomous stochastic differential equation with small diffusion term and that the two equations are driven by independent Brownian motions. The equation of $X(t)$ only depends on $\alpha(t)$ and the equation of $V(t)$ only depends on $\gamma(t) = \alpha(t) - \beta(t)$. This is why we concentrate on estimating $\alpha(t), \gamma(t)$ and this will be done by two separate contrasts. As N tends to infinity, both processes $X(t), V(t)$ converge almost surely uniformly on $[0,T]$ to a deterministic function (respectively $x(t)$ and $v(t)$) (Propositions 1 and 2). We assume that $\mathbb{E}X_i(0) \neq 0$ and $\mathrm{Var}X_i(0) \neq 0$. Under this assumption, we characterize the probability of deviation of $\mathbb{P}(\sup_{t\in[0,T]} |(X^2(t)/x^2(t)) - 1| > \delta)$ and $\mathbb{P}(\sup_{t\in[0,T]} |(V(t)/v(t)) - 1| > \delta)$ (we apply them for $\delta = 1/2$) (Propositions 3 and 4).
We define two collections of minimum contrast estimators $\widetilde{\alpha}_m(t), \widetilde{\gamma}_p^\star(t)$ respectively of $\alpha(t), \gamma(t)$. For each m (resp. p), $\widetilde{\alpha}_m(t)$ (resp. $\widetilde{\gamma}_p^\star(t)$) belongs to a finite dimensional subspace of $\mathbb{L}^2([0,T])$.

We study the $\mathbb{L}^2$-risks of these estimators where the risk is defined either by the expectation of an empirical norm or by the expectation of a deterministic norm (Propositions 7 and 8). For the estimation of $\alpha(t)$, the empirical norm and the deterministic weighted norm are given for a function h of $\mathbb{L}^2([0,T])$ by

$$\|h\|_X^2 = \int_0^T h^2(t)X^2(t)dt, \quad \|h\|_x^2 = \int_0^T h^2(t)x^2(t)dt. \tag{4}$$

For the estimation of $\gamma(t)$, they are given by

$$\|h\|_{\sqrt{V}}^2 = \int_0^T h^2(t)V(t)dt, \quad \|h\|_{\sqrt{v}}^2 = \int_0^T h^2(t)v(t)dt. \tag{5}$$

Thanks to Propositions 3 and 4, these norms are equivalent on $\mathbb{L}^2([0,T])$ outside a set of small probability which is the key tool for bounding the risks of our estimators. Afterwards, we propose data-driven choices $\widehat{m}$, $p^\star$ of the values m, p and study the risk of the adaptive estimators (Theorem 1). In order to have a benchmark for comparison, we also briefly treat the estimation of $\alpha(t)$ when $\beta(t) \equiv 0$ in model (1), that is $\gamma(t) = \alpha(t), \forall t \in [0,T]$.
Section 2 contains our assumptions and the preliminary properties concerning the two processes $X(t), V(t)$. In Section 3, we build and study our projection estimators. Section 4 concerns the adaptive estimators. Section 5 deals with the estimation when $\beta(t) \equiv 0$. In Section 6, we illustrate our theory by numerical experiments on simulated data. Section 7 contains some concluding remarks. Proofs are gathered in Section 8.

2. Assumptions and Preliminary properties.

We set $\gamma(t) = \alpha(t) - \beta(t)$ and consider the following assumptions:

[H1] $X_0^i, i = 1, \dots, N$ are *i.i.d.* random variables such that $\mathbb{E}(X_0^i) = \mu_0$, $\mathbb{E}(X_0^i)^2 = \sigma_0^2 + \mu_0^2$ with $\mu_0 \neq 0$ and $\sigma_0^2 \neq 0$ and $X_i(0)$ has moments of any order.
[H2] The functions $\alpha(t), \gamma(t) : \mathbb{R}^+ \to \mathbb{R}$ are continuous on $\mathbb{R}^+$ (and thus belong to $\mathbb{L}^2([0,T])$), and $\gamma(t) \not\equiv 0$, $\alpha(t) \not\equiv \gamma(t)$.

With the new parameterization, we have

$$dX_i^N(t) = \left\{\alpha(t)\overline{X}_N^N(t) + \gamma(t)[X_i^N(t) - \overline{X}_N^N(t)]\right\} dt + dW_i(t), \quad X_i^N(0) = X_0^i, i = 1, \dots, N. \tag{6}$$

2.1. Study of the empirical mean and empirical variance.

Proposition 1. *The empirical mean satisfies* $d\overline{X}_N^N(t) = \alpha(t)\overline{X}_N^N(t)dt + \frac{1}{\sqrt{N}}dB_{N,1}(t)$, *that is*

$$\overline{X}_N^N(t) = \exp\left(\int_0^t \alpha(s)ds\right) \ \overline{X}_N^N(0) + \frac{1}{\sqrt{N}}g_N(t)$$

where $g_N(t) = \exp\left(\int_0^t \alpha(s)ds\right) \int_0^t \exp\left(-\int_0^s \alpha(u)du\right)dB_{N,1}(s)$ *and* $B_{N,1}$ *is the standard Brownian motion given by* $B_{N,1}(t) = (1/\sqrt{N})\sum_{j=1}^N W_j(t)$.
Let $x(t) = \mu_0 \exp\left(\int_0^t \alpha(s)ds\right)$. *Almost surely, as* N *tends to infinity,* $\sup_{0\le t\le T}|\overline{X}_N^N(t) - x(t)|$ *tends to* 0.

Note that $dg_N(t) = \alpha(t)g_N(t)dt + dB_{N,1}(t), g_N(0) = 0$, (g_N) has a fixed distribution. It follows from Proposition 1 that we have the explicit expression of $(\overline{X}_N^N(t))^2$:

$$(\overline{X}_N^N(t))^2 = \exp\left(2\int_0^t \alpha(s)ds\right) \ (\overline{X}_N^N(0))^2 + 2\frac{1}{\sqrt{N}}g_N(t)\exp\left(\int_0^t \alpha(s)ds\right) \ \overline{X}_N^N(0) + \frac{1}{N}g_N^2(t),$$

where the middle term is centred. By Proposition 1, almost surely, as N tends to infinity, $\sup_{0\le t\le T}|(\overline{X}_N^N(t))^2 - x^2(t)|$ tends to 0. Under [H1], we have

$$\mathbb{E}(\overline{X}_N^N(t))^2 = \exp{(2\int_0^t \alpha(s)ds)} \ (\frac{\sigma_0^2}{N} + \mu_0^2) + \frac{1}{N}\mathbb{E}g_N^2(t) = x^2(t)\left(1 + \frac{\lambda(t)}{N}\right),$$

with $x(t)$ defined in Proposition 1 and

$$\lambda(t) := \frac{1}{\mu_0^2}\left(\sigma_0^2 + \int_0^t \exp{(-2\int_0^s \alpha(u)du)}ds\right). \tag{7}$$

Thus, as N tend to infinity $\mathbb{E}(\overline{X}_N^N(t))^2 \to x^2(t)$.

Proposition 2. *The process $V_N(t)$ defined in (3) satisfies $V_N(t) > 0$ for all $t \ge 0$ and*

$$dV_N(t) \quad = \quad [2\gamma(t)V_N(t) + (1 - \frac{1}{N})]dt + \frac{2}{\sqrt{N}}\sqrt{V_N(t)}dB_{N,2}(t),$$

where $B_{N,2}(t) = \int_0^t \frac{\sum_{i=1}^N (X_i(t) - \overline{X}_N^N(t))dW_i(t)}{\sqrt{V_N(t)}}$ is a Brownian motion. Setting $\Gamma(t) = \int_0^t \gamma(s)ds$, this yields

$$V_N(t) = e^{2\Gamma(t)}\left(V_N(0) + \int_0^t e^{-2\Gamma(s)}(1 - \frac{1}{N})ds + \frac{2}{\sqrt{N}}\int_0^t e^{-2\Gamma(s)}\sqrt{V_N(s)}dB_{N,2}(s)\right).$$

Let $v(t)$ be defined by $dv(t) = [2\gamma(t)v(t) + 1]dt, v(0) = \sigma_0^2$. Then,

$$v(t) = \sigma_0^2 e^{2\Gamma(t)} + e^{2\Gamma(t)}\int_0^t e^{-2\Gamma(s)}ds > 0 \quad \text{for all } t \ge 0$$

and as N tends to infinity, $\sup_{0\le t\le T}|V_N(t) - v(t)| \to_{a.s.} 0$.

As a by-product of Proposition 2, we get the useful property $\mathbb{E}(V_N(t)) = (1 - 1/N)v(t)$.
Note that $\langle B_{N,1}, B_{N,2}\rangle = 0$. We stress that in the equation for $\overline{X}_N^N(t)$, the drift depends on $\alpha(t)$ only and in the equation of $V_N(t)$, the drift depends on $\gamma(t)$ only. This is the interest of the change of parametrisation $(\alpha(t), \beta(t)) \mapsto (\alpha(t), \gamma(t))$.

2.2. **Almost exponential inequalities.** For simplicity, let us set $X(t) = \overline{X}_N^N(t)$ and $V_N(t) = V(t)$ (see (2)-(3)).

For the sequel, we need the following lemma.

Lemma 1. *For all $r \ge 1$, there exists a constant $c_r(T)$ such that, for all $t \le T$,*

$$\mathbb{E}[|X(t)|^r + V^r(t)] \le c_r(T).$$

Proposition 3. *Assume* **[H1]** . *Define the set*

$$\mathcal{O}_{N,1} = \{\sup_{t\in[0,T]}|\frac{X^2(t)}{x^2(t)} - 1| \le 1/2\}. \tag{8}$$

There exist positive constants $C(T), c(T)$ and for all $r \ge 1$, a constant $c(r)$ such that

$$\mathbb{P}(\mathcal{O}_{N,1}^c) \le C(T)\exp{(-c(T)N)} + \frac{c(r)}{N^r}.$$

Proposition 4. *Assume* **[H1]** *. Define the set*

$$\mathcal{O}_{N,2} = \{ \sup_{t\in[0,T]} |\frac{V(t)}{v(t)} - 1| \leq 1/2\}. \tag{9}$$

There exists a constant $c(T)$ and for all $r \geq 1$, a constant $c(r,T)$ such that

$$\mathbb{P}(\mathcal{O}^c_{N,2}) \leq 2\exp\left(-c(T)N\right) + \frac{c(r,T)}{N^r}.$$

3. Estimation of $(\alpha(t), \gamma(t))$.

Recall the notations $X(t) = \overline{X}^N_N(t)$ and $V(t) = V_N(t) = \frac{1}{N}\sum_{i=1}^N [X^N_i(t) - \overline{X}^N_N(t)]^2$ and set for simplicity $X_j = X^N_j$.

3.1. **Estimation contrast.** Consider $(\varphi_j, j \geq 0)$ and $(\psi_j, j \geq 0)$ two orthonormal bases of $\mathbb{L}^2([0,T])$ and let S_m (resp. Σ_p) be the subspace generated by $(\varphi_j, 0 \leq j \leq m-1)$ (resp. by $(\psi_j, 0 \leq j \leq p-1)$). We assume that the functions $(\varphi_j, j \geq 0)$, $(\psi_j, j \geq 0)$ are continuous on $[0,T]$.
Inspired by the log-likelihood of Process (1), for $h(t) = (h_1(t), h_2(t))$ element of $\mathbb{L}^2([0,T]) \times \mathbb{L}^2([0,T])$, we consider the contrast

$$\begin{aligned} U_N(h) &= \frac{1}{N}\int_0^T \sum_{j=1}^N [h_1(t)X(t) + h_2(t)(X_j(t) - X(t)]^2 dt \\ &\qquad - \frac{2}{N}\sum_{j=1}^N \int_0^T \sum_{j=1}^N [h_1(t)X(t) + h_2(t)(X_j(t) - X(t)] dX_j(t). \end{aligned}$$

Developing $U_N(h) = U_N((h_1,h_2))$ and using that $\sum_{j=1}^N (X_j(t) - X(t)) = 0$, we obtain

$$U_N((h_1,h_2)) = U_{N,1}(h_1) + U_{N,2}(h_2)$$

with

$$\begin{aligned} U_{N,1}(h_1) &= \int_0^T h_1^2(t)X^2(t)dt - 2\int_0^T h_1(t)X(t)dX(t), \\ U_{N,2}(h_2) &= \int_0^T h_2^2(t)V(t)dt - \frac{2}{N}\int_0^T h_2(t)\sum_{j=1}^N (X_j(t) - X(t))dX_j(t). \end{aligned}$$

This is why we define the projection estimators of $\alpha(t)$ on S_m and of $\gamma(t))$ on Σ_p by

$$\widehat{\alpha}_m = \arg\min_{h_1\in S_m} U_{N,1}(h_1), \qquad \gamma^\star_p = \arg\min_{h_2\in\Sigma_p} U_{N,2}(h_2). \tag{10}$$

Recall that we defined random and deterministic weighted norms $\|h_1\|_X, \|h_1\|_x$ in (4) and $\|h_2\|_{\sqrt{V}}\|, \|h_2\|_{\sqrt{v}}\|$ in (5), with the associated scalar products. Note that (see (7) for $\lambda(t)$):

$$\mathbb{E}(\|h_1\|^2_X) = \|h_1\|^2_x + \frac{1}{N}\int_0^T h_1^2(t)\lambda(t)x^2(t)dt, \quad \mathbb{E}(\|h_2\|^2_{\sqrt{V}}) = \left(1 - \frac{1}{N}\right)\|h_2\|^2_{\sqrt{v}}.$$

Proposition 5. *We have*

$$\mathbb{E}(U_{N,1}(h_1)) = \mathbb{E}(\|h_1 - \alpha\|^2_X - \|\alpha\|^2_X), \quad \mathbb{E}(U_{N,2}(h_2)) = \mathbb{E}(\|h_2 - \gamma\|^2_{\sqrt{V}} - \|\gamma\|^2_{\sqrt{V}}).$$

Proposition 5 shows that the expectations of the contrasts $U_{N,1}(h_1)$ and $U_{N,2}(h_2)$ are minimum for $h_1=\alpha, h_2=\gamma$ and explains the definition of the estimators.

Proof of Proposition 5. As $\sum_{j=1}^N (X_j(t)-X(t))=0$, we get

$$U_{N,1}(h_1)=\int_0^T h_1^2(t)X^2(t)dt-2\int_0^T h_1(t)\alpha(t)X^2(t)dt-2\int_0^T h_1(t)X(t)\left(\frac{1}{N}\sum_{j=1}^N dW_j(t)\right).$$

and

$$U_{N,2}(h_2)=\int_0^T h_2^2(t)V(t)dt-2\int_0^T h_2(t)\gamma(t)V(t)dt-2\int_0^T h_2(t)\frac{1}{N}\sum_{j=1}^N (X_j(t)-X(t))dW_j(t).$$

We have

$$(11)\quad U_{N,1}(h_1)=\|h_1-\alpha\|_X^2-\|\alpha\|_X^2-2\nu_{N,1}(h_1),\quad U_{N,2}(h_2)=\|h_2-\gamma\|_{\sqrt{V}}^2-\|\gamma\|_{\sqrt{V}}^2-2\nu_{N,2}(h_2),$$

with

$$\nu_{N,1}(h_1)=\int_0^T h_1(t)X(t)\left(\frac{1}{N}\sum_{j=1}^N dW_j(t)\right)=\frac{1}{\sqrt{N}}\int_0^T h_1(t)X(t)dB_{N,1}(t),$$

$$\nu_{N,2}(h_2)=\int_0^T h_2(t)\frac{1}{N}\sum_{j=1}^N (X_j(t)-X(t))dW_j(t)=\frac{1}{\sqrt{N}}\int_0^T h_2(t)\sqrt{V(t)}dB_{N,2}(t).$$

Note that $\mathbb{E}(\nu_{N,i}(h_i))=0$ for $i=1,2$ and thus

$$\mathbb{E}(U_{N,1}(h_1))=\mathbb{E}(\|h_1-\alpha\|_X^2-\|\alpha\|_X^2),\quad \mathbb{E}(U_{N,2}(h_2))=\mathbb{E}(\|h_2-\gamma\|_{\sqrt{V}}^2-\|\gamma\|_{\sqrt{V}}^2).$$

It is also interesting to note that, as $\langle B_{N,1},B_{N,2}\rangle=0$, it holds $\mathbb{E}(\nu_{N,1}(h_1)\nu_{N,2}(h_2))=0$. □

3.2. **Risk of the projection estimators on a fixed space.** For M a matrix, let $\mathrm{Tr}(M)$ denote the trace of the matrix M and let $\|M\|_{\mathrm{op}}$ denote the operator norm of M that is the square root of the largest eigenvalue of $M\,{}^tM$. If M is symmetric, $\|M\|_{\mathrm{op}}=\sup\{|\lambda_i|\}$ where λ_i are the eigenvalues of M and so $\|M^{-1}\|_{\mathrm{op}}=\|M\|_{\mathrm{op}}^{-1}$. For $h_1\in\mathbb{L}^2([0,T])$, we denote by $\|h_1\|=(\int_0^T h_1^2(t)dt)^{1/2}$ its $\mathbb{L}^2$-norm and $\|x\|_{2,r}$ denotes the Euclidean norm of the vector $x={}^t(x_1,\dots,x_r)$ of $\mathbb{R}^r$.
We now detail the construction and the expression of the estimators (10).
Let us define

$$\widehat{\Psi}_{m,1}=\left(\int_0^T \varphi_j(t)\varphi_k(t)X^2(t)dt\right)_{0\le j,k\le m-1}\text{ and }\widehat{\Psi}_{p,2}=\left(\int_0^T \psi_j(t)\psi_k(t)V(t)dt\right)_{0\le j,k\le p-1}$$

and

$$\widehat{Z}_{m,1}=\left(\int_0^T \varphi_j(t)X(t)dX(t)\right)_{0\le j\le m-1}\text{ and }\widehat{Z}_{p,2}=\left(\int_0^T \psi_j(t)\frac{1}{N}\sum_{k=1}^N (X_k(t)-X(t))dX_k(t)\right)_{0\le j\le p-1}.$$

Noting that, for $\mathbf{u}={}^t(u_0,\dots,u_{m-1})\in\mathbb{R}^m$, ${}^t\mathbf{u}\widehat{\Psi}_{m,1}\mathbf{u}=\int_0^T\left(\sum_{j=0}^{m-1}u_j\varphi_j(t)\right)^2X^2(t)dt\ge 0$, we conclude that $\widehat{\Psi}_{m,1}$ is symmetric positive definite, as well as $\widehat{\Psi}_{p,2}$. Indeed the bases functions and $X(t)$ are continuous and non identically zero.

Set $\widehat{\alpha}_m(t) = \sum_{k=0}^{m-1} [\widehat{\alpha}_{(m)})]_k \, \varphi_k(t)$, $\gamma_p^\star(t) = \sum_{\ell=0}^{p-1} [\gamma_{(p)}^\star]_\ell \, \psi_\ell(t)$. By a standard computation, we get that the vectors $\widehat{\alpha}_{(m)} = {}^t([\widehat{\alpha}_{(m)})]_k, k = 0, \dots m-1)$, $\gamma_{(p)}^\star = {}^t([\gamma_{(p)}^\star]_\ell, \ell = 0, \dots, p-1)$ are solution of

$$\widehat{Z}_{m,1} = \widehat{\Psi}_{m,1} \widehat{\alpha}_{(m)} \quad \text{and} \quad \widehat{Z}_{p,2} = \widehat{\Psi}_{p,2} \gamma_{(p)}^\star.$$

Therefore

$$\widehat{\alpha}_{(m)} = \widehat{\Psi}_{m,1}^{-1} \widehat{Z}_{m,1}, \quad \gamma_{(p)}^\star = \widehat{\Psi}_{p,2}^{-1} \widehat{Z}_{p,2}. \tag{12}$$

We define the symmetric positive definite matrices

$$\Psi_{m,1} = \left(\int_0^T \varphi_j(t) \varphi_k(t) x^2(t) dt \right)_{0 \le j,k \le m-1}, \quad \Psi_{p,2} = \left(\int_0^T \psi_j(t) \psi_k(t) v(t) dt \right)_{0 \le j,k \le p-1}.$$

Note that

$$\mathbb{E}\left(\widehat{\Psi}_{m,1}\right) = \Psi_{m,1} + \frac{1}{N} \left(\int_0^T \varphi_j(t) \varphi_k(t) \lambda(t) x^2(t) dt \right)_{0 \le j,k \le m-1} \text{and } \mathbb{E}\left(\widehat{\Psi}_{p,2}\right) = \left(1 - \frac{1}{N}\right) \Psi_{p,2},$$

where $\lambda(t)$ is defined by (7). Lastly, we define

$$L(S_m) = \sup_{t \in [0,T]} \sum_{j=0}^{m-1} \varphi_j^2(t), \quad L(\Sigma_p) = \sup_{t \in [0,T]} \sum_{j=0}^{p-1} \psi_j^2(t).$$

These quantities were introduced in Comte and Genon-Catalot (2020a, 2020b) in the framework of regression and drift estimation for diffusions by projection method. They only depend respectively on the subspace S_m (on the subspace Σ_p) and not on the bases chosen to define them. Indeed, $L(S_m) = \sup_{h_1 \in S_m, \|h_1\|=1} \sup_{t \in [0,T]} h_1^2(t)$, where $\|h_1\|^2 = \int_0^T h_1^2(t) dt$, and analogously for $L(\Sigma_p)$.

Below, we restrict the possible choices of the dimensions m, p by a condition which ensures the stability of least-squares estimators (see *e.g.* Cohen *et al.* (2013), Comte and Genon-Catalot (2020a)). For $\mathfrak{c}$ a numerical constant, that can take any value, we consider dimensions m, p such that

$$m, p \le N \quad \text{and} \quad L(S_m)(\|\Psi_{m,1}\|_{\text{op}}^{-1} \vee 1) \le \frac{\mathfrak{c} N}{2 \log N}, \quad L(\Sigma_p)(\|\Psi_{p,2}\|_{op}^{-1} \vee 1) \le \frac{\mathfrak{c} N}{2 \log N}. \tag{13}$$

In parallel, define the truncated estimators

$$\widetilde{\alpha}_m = \widehat{\alpha}_m \mathbf{1}_{\Lambda_{m,1}}, \qquad \Lambda_{m,1} = \{L(S_m)(\|\widehat{\Psi}_{m,1}\|_{\text{op}}^{-1} \vee 1) \le \mathfrak{c} N / \log N\} \tag{14}$$

and

$$\widetilde{\gamma}_p^\star = \gamma_p^\star \mathbf{1}_{\Lambda_{p,2}}, \qquad \Lambda_{p,2} = \{L(\Sigma_p)(\|\widehat{\Psi}_{p,2}\|_{\text{op}}^{-1} \vee 1) \le \mathfrak{c} N / \log N\}. \tag{15}$$

It is worth noting that the following holds.

Proposition 6. *The mappings* $m \mapsto \|\Psi_{m,1}^{-1}\|_{\text{op}}$, $m \mapsto \|\widehat{\Psi}_{m,1}^{-1}\|_{\text{op}}$, $p \mapsto \|\Psi_{p,2}^{-1}\|_{\text{op}}$, $p \mapsto \|\widehat{\Psi}_{p,2}^{-1}\|_{\text{op}}$ *are increasing.*

Now, we can state the risks bounds of the above estimators for fixed m, p.

Proposition 7. *Assume that* **[H1]** *and* **[H2]** *hold. Consider the estimator* $\widetilde{\alpha}_m(t)$ *of* $\alpha(t)$ *defined by (14), for* m *satisfying* (13).

`(i)` *For the risk based on the empirical* X*-norm, we have*

$$\mathbb{E}\|\widetilde{\alpha}_m - \alpha\|_X^2 \le \left(1 + \frac{\lambda_T^\star}{N}\right) \left(\inf_{h \in S_m} \|h - \alpha\|_x^2 + 2\frac{m}{N} \right) + \frac{c}{N}, \tag{16}$$

where $\lambda_T^\star := \sup_{t\in[0,T]} \lambda(t)$.
(ii) *For the risk based on the deterministic x-norm, we have*

$$\mathbb{E}\|\widetilde{\alpha}_m - \alpha\|_x^2 \le 2\left(1+\frac{\lambda_T^\star}{N}\right)\left(\inf_{h\in S_m}\|h-\alpha\|_x^2 + 2\frac{m}{N}\right) + \frac{c}{N}. \tag{17}$$

Refining the proof of the bias term, we get

$$\mathbb{E}\|\widetilde{\alpha}_m - \alpha\|_x^2 \le \left(1+\frac{\mathfrak{c}}{\log(N)}K_N(T)\right)\inf_{h\in S_m}\|h-\alpha\|_x^2 + 8\left(1+\frac{\lambda_T^\star}{N}\right)\frac{m}{N} + \frac{c}{N} \tag{18}$$

with

$$K_N(T) := 8T\sup_{t\in[0,T]} x^2(t)\left(\frac{C_1(T)}{\mu_0^4} + \frac{(\lambda_T^\star)^2}{N}\right)$$

and $C_1(T)$ *is a constant depending on* T.

Proposition 8. *Assume that* **[H1]** *and* **[H2]** *hold. Consider the estimator* $\widetilde{\gamma}_p^\star(t)$ *of* $\gamma(t)$ *defined by (15), for p satisfying* (13).
(i) *For the risk based on the empirical* $\sqrt{V}$*-norm, we have*

$$\mathbb{E}\|\widetilde{\gamma}_p^\star - \gamma\|_{\sqrt{V}}^2 \le (1-\frac{1}{N})\left(\inf_{h\in\Sigma_p}\|h-\gamma\|_{\sqrt{v}}^2 + \frac{2p}{N}\right) + \frac{c'}{N}. \tag{19}$$

(ii) *For the risk based on the deterministic* $\sqrt{v}$*-norm, we have*

$$\mathbb{E}\|\widetilde{\gamma}_p^\star - \gamma\|_{\sqrt{v}}^2 \le 2\left(1-\frac{1}{N}\right)\left(\inf_{h\in\Sigma_p}\|h-\gamma\|_{\sqrt{v}}^2 + 2\frac{p}{N}\right) + \frac{c}{N}. \tag{20}$$

With a more elaborate proof,

$$\mathbb{E}\|\widetilde{\gamma}_p^\star - \gamma\|_{\sqrt{v}}^2 \le \left(1+\frac{\mathfrak{c}}{\log(N)}R(T)\right)\inf_{h\in\Sigma_p}\|h-\gamma\|_{\sqrt{v}}^2 + 8\left(1-\frac{1}{N}\right)\frac{p}{N} + \frac{c}{N} \tag{21}$$

with

$$R(T) := 2T\sup_{t\in[0,T]} v(t)\frac{C_2(T)}{\sigma_0^4}$$

and $C_2(T)$ *is a constant depending on* T.

In the two previous propositions, for the risks based on the deterministic norms, we obtain two inequalities. The difference between (17) and (18) lies in the evaluation of the bias term where the factor 2 is improved into a factor $1 + o(1)$.
The exact rates of $\|\Psi_{m,1}^{-1}\|_{\mathrm{op}}$ and $\|\Psi_{p,2}^{-1}\|_{\mathrm{op}}$ as functions of m, p are difficult to compute. However, we can prove the following result.

Proposition 9. (1) *If, for all* $t \in [0,T]$, $\alpha(t) \ge 0$, *then* $\|\Psi_{m,1}^{-1}\|_{\mathrm{op}} = \|\Psi_{m,1}\|_{\mathrm{op}}^{-1} \le \mu_0^{-2}$.
(2) *If, for all* $t \in [0,T]$, $\gamma(t) \ge 0$, *then* $\|\Psi_{p,2}^{-1}\|_{\mathrm{op}} = \|\Psi_{p,2}\|_{\mathrm{op}}^{-1} \le \sigma_0^{-2}$.

3.3. **Rates of convergence.** Rates of convergence can be deduced from risk bounds provided that functional regularity conditions on α, γ are set. We give results for α only, but the same type of results holds for γ. Regularity spaces depend on the basis which is used. Below and in the simulation section, examples are presented.

3.3.1. *Rates on trigonometric spaces.* First we consider the collection $(S_m^{Trig}, m \geq 0)$ of subspaces of $\mathbb{L}^2([0,T])$ where S_m^{Trig} has odd dimension m and is generated by the orthonormal trigonometric basis (denoted by [T]) $(\varphi_{j,T})$ with $\varphi_{0,T}(t) = \sqrt{1/T}\mathbf{1}_{[0,T]}(t)$, $\varphi_{2j-1,T}(t) = \sqrt{2/T}\cos(2\pi jt/T)\mathbf{1}_{[0,T]}(t)$ and $\varphi_{2j,T}(t) = \sqrt{2/T}\sin(2\pi jt/T)\mathbf{1}_{[0,T]}(t)$ for $j = 1, \ldots, (m-1)/2$. This basis satisfies $\sum_{j=0}^{m-1}\varphi_{j,T}^2(t) = m/T$. Therefore $L(S_m^{Trig}) = m/T$.
Assume moreover that, for some $c_0^2 > 0$,

$$\forall t \in [0,T],\ x^2(t) \geq c_0^2. \tag{22}$$

This assumption is fulfilled if $\alpha(t) \geq 0$, $\forall t \in [0,T]$, with $c_0 = \mu_0$. Then, it follows from the proof of Proposition 9 that (22) implies $\|\Psi_m^{-1}\|_{\text{op}} \leq 1/c_0^2$. As a consequence, under (22), m satisfies condition (13) as soon as $m \leq \mathfrak{c}c_0^2TN/\log(N)$, which is a weak constraint.

Let r be a positive integer, $L > 0$ and define

$$\begin{aligned} W^{\text{per}}(r,L) \ &:= \ \{g \in C^r([0,T];\mathbb{R}) : g^{(r-1)} \text{ is absolutely continuous,} \\ &\qquad \int_0^T g^{(r)}(x)^2dx \leqslant L \text{ and } g^{(j)}(0) = g^{(j)}(T),\ \forall j = 0, \ldots, r-1\}. \end{aligned}$$

By Proposition 1.14 of Tsybakov (2009), a function $f \in W^{\text{per}}(r,L)$ admits a development

$$f = \sum_{j=0}^{\infty}\theta_{j,T}\varphi_{j,T} \quad \text{such that} \quad \sum_{j\geqslant 0}\theta_{j,T}^2\tau_j^2 \leqslant C(T,L),$$

where $\tau_j = j^r$ for even j, $\tau_j = (j-1)^r$ for odd j, and $C(L) := L^2(T/\pi)^{2r}$. So, with $f = \sum_{j=0}^{m-1}\theta_{j,T}\varphi_{j,T}$,

$$\|f - f_m\|^2 \leqslant K(L,T,r)m^{-2r}.$$

Thus we obtain the following rate

Corollary 1. *Assume that* **[H1]**, **[H2]**, *(22) hold and that* $\alpha \in W^{\text{per}}(r,L)$. *Then choosing* $m^\star = c^\star N^{1/(2r+1)}$ *gives* $\mathbb{E}(\|\widetilde{\alpha}_{m^\star} - \alpha\|^2) \leq C(T,r,L,c_0^2)N^{-2r/(2r+1)}$.

Indeed, $m^\star$ satisfies $m^\star \leq \mathfrak{c}c_0^2TN/\log(N)$ for well chosen $c^\star$.

Note that we also use the cosine basis [C] defined by $\varphi_{0,T}(x) = \sqrt{1/T}\mathbf{1}_{[0,T]}(t)$, $\varphi_{j,T}(t) = \sqrt{2/T}\cos(\pi jt/T)\mathbf{1}_{[0,T]}(t)$, $j = 1, \ldots, m-1$, see Efromovich (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.

3.3.2. *Rates on Sobolev Laguerre spaces.* Consider now a collection of subspaces of $\mathbb{L}^2(\mathbb{R}^+)$, generated by the Laguerre basis [L] defined by

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

We set $S_m^{Lag} = \text{span}\{\ell_j, j = 0, \ldots, m-1\}$. We have $\forall t \geq 0$, $\sum_{j=0}^{m-1}\ell_j^2(t) \leq 2m$ (see Abramowitz and Stegun (1964)) and as $\varphi_j(0) = \sqrt{2}$, $L(S_m^{Lag}) = 2m$.

The basis is orthonormal in $\mathbb{L}^2(\mathbb{R}^+)$, but not in $\mathbb{L}^2([0,T])$. However, the bounds (16)-(19) can be obtained for this basis too as the orthonormality is not used in the proof. Thus we can get rates on Sobolev Laguerre spaces defined by $W^s(D) = \{f \in \mathbb{L}^2(\mathbb{R}^+),\ \sum_{j\geq 1}j^sc_j(f)^2 \leq D < +\infty\}$,

$c_j(f) = \int_0^{+\infty} f(t)\ell_j(t)dt$. A function belonging to $W^s(D)$ has, roughly speaking, regularity properties of order s, see Comte and Genon-Catalot (2018).

Therefore, under **[H1]**, **[H2]**, if $\alpha \in W^s(D)$ and if $m^\diamond = cN^{1/(s+1)}$ satisfies constraint (13), then $\mathbb{E}(\|\tilde{\alpha}_{m^\diamond} - \alpha\|_X^2) \le c(T,D)N^{-s/(s+1)}$, which is an optimal rate for regression type estimation with Laguerre basis, see Comte and Genon-Catalot (2020a, Theorem 1).

4. Adaptive estimators

We will assume in the following that (see examples of bases in section 3.3):

$$L(S_m) \le c_\varphi^2 m, \quad L(\Sigma_p) \le c_\psi^2 p.$$

We set

$$\mathcal{M}_{N,1} = \left\{ m \le N, L(S_m)\|\Psi_{m,1}^{-1}\|_{\rm op} \le \frac{\mathfrak{c}}{2}\frac{N}{\log(N)} \right\}, \ \widehat{\mathcal{M}}_{N,1} = \left\{ m \le N, L(S_m)\|\widehat{\Psi}_{m,1}^{-1}\|_{\rm op} \le \mathfrak{c}\frac{N}{\log(N)} \right\},$$

$$\mathcal{M}_{N,2} = \left\{ p \le N, L(\Sigma_p)\|\Psi_{p,2}^{-1}\|_{\rm op} \le \frac{\mathfrak{c}}{2}\frac{N}{\log(N)} \right\}, \ \widehat{\mathcal{M}}_{N,2} = \left\{ p \le N, L(\Sigma_p)\|\widehat{\Psi}_{p,2}^{-1}\|_{\rm op} \le \mathfrak{c}\frac{N}{\log(N)} \right\}.$$

Now we define

$$\widehat{m} = \arg\min_{m \in \widehat{\mathcal{M}}_{n,1}} \{U_{N,1}(\widehat{\alpha}_m) + \mathrm{pen}_1(m)\}, \quad \mathrm{pen}_1(m) = \kappa_1 \frac{m}{N} \tag{23}$$

$$p^\star = \arg\min_{p \in \widehat{\mathcal{M}}_{p,2}} \{U_{N,2}(\gamma_p^\star) + \mathrm{pen}_2(p)\}, \quad \mathrm{pen}_2(p) = \kappa_2 \frac{p}{N}. \tag{24}$$

Note that $U_{N,1}(\widehat{\alpha}_m) = -\|\widehat{\alpha}_m\|_X^2$ and $U_{N,2}(\gamma_p^\star) = -\|\gamma_p^\star\|_{\sqrt{V}}^2$.

Theorem 1. *Under* **[H1]-[H2]**, *there exists a numerical κ_0 such that for $\kappa_1 \ge \kappa_0$, $\kappa_2 \ge \kappa_0$,*

$$\mathbb{E}\left(\|\widehat{\alpha}_{\widehat{m}} - \alpha\|_X^2\right) \le C_1 \inf_{m \in \mathcal{M}_{N,1}} \left(\left(1 + \frac{\lambda_T^\star}{N}\right) \inf_{h \in S_m} \|h - \alpha\|_x^2 + \kappa_1 \frac{m}{N} \right) + \frac{c}{N},$$

$$\mathbb{E}\left(\|\gamma_{p^\star}^\star - \gamma\|_{\sqrt{V}}^2\right) \le C_2 \inf_{p \in \mathcal{M}_{N,2}} \left((1 - \frac{1}{N}) \inf_{h \in \Sigma_p} \|h - \gamma\|_{\sqrt{v}}^2 + \kappa_2 \frac{p}{N} \right) + \frac{c'}{N},$$

where C_1 and C_2 are numerical constants.

Theorem 1 shows that our estimators are adaptive in the sense that their risk automatically realizes the best compromise between the bias and the variance terms. In practical implementation, we should not use the value κ_0 provided by our proof since it is not the smallest one. The theoretical determination of the best value κ_0 is difficult. Therefore, it is customary to determine this value by preliminary simulations.

5. Estimation of $\alpha(t)$ when $\beta(t) \equiv 0$

To have a benchmark for comparison, we consider here the simpler case where $\beta(t) \equiv 0$ in model (1), *i.e.* the model given by N *i.i.d.* Ornstein-Uhlenbeck processes:

$$dX_i(t) = \alpha(t)X_i(t)dt + dW_i(t), \quad X_i(0) = X_0^i, \quad i = 1, \dots, N. \tag{25}$$

The problem of nonparametric estimation of $\alpha(t)$ from (25) has been first tackled in Nguyen and Pham (1982) who propose a projection estimator using an increasing sequence of subspaces of $\mathbb{L}^2([0,T])$ as we do here for the couple $(\alpha(t), \gamma(t))$. However, this paper neither gives any concrete choice of projection bases and nor studies the $\mathbb{L}^2$-risk of the projection estimator. Moreover, the problem of an adaptive choice of the projection dimension is not raised at all. This is why we complete this study below. Taking into account our previous statements, we do it as briefly as possible to avoid repetitions.

5.1. **Direct estimation of** $\alpha(t)$**.** Consider, for $h \in \mathbb{L}^2([0,T])$, the contrast given by:

$$\Lambda_N(h) = \frac{1}{N}\int_0^T h^2(t)\sum_{j=1}^N X_j^2(t)dt - \frac{2}{N}\int_0^T h(t)\sum_{j=1}^N X_j(t)dX_j(t).$$

Set $Y(t) := Y_N(t) = \frac{1}{N}\sum_{j=1}^N X_j^2(t) = X^2(t) + V(t)$, $y(t) = x^2(t) + v(t)$.

Proposition 10. *The process* $Y(t)$ *satisfies:* $dY(t) = [2\alpha(t)Y(t) + 1]dt + \frac{2}{\sqrt{N}}\sqrt{Y(t)}dB_{N,3}(t)$, *where* $dB_{N,3}(t) = \frac{\sum_{j=1}^N X_i(t)dW_i(t)}{(\sum_{j=1}^N X_i^2(t))^{1/2}}$ *is a Brownian motion. This yields*

$$\begin{aligned} Y(t) &= \exp(2\int_0^t \alpha(s)ds)\Big[Y(0) + \int_0^t \exp(-2\int_0^s \alpha(u)du)ds \\ &\qquad + \frac{2}{\sqrt{N}}\int_0^t \exp(-2\int_0^s \alpha(u)du)\sqrt{Y(s)}dB_{N,3}(s)\Big]. \end{aligned}$$

As $N \to +\infty$, $Y(t)$ *converges uniformly on* $[0,T]$ *to*

$$y(t) = \exp(2\int_0^t \alpha(s)ds)[\sigma_0^2 + \mu_0^2 + \int_0^t \exp(-2\int_0^s \alpha(u)du)ds].$$

We have $\mathbb{E}Y(t) = y(t)$. As previously, we can prove that the probability of the set

$$\mathcal{O}_N = \{\sup_{t\in[0,T]} |\frac{Y(t)}{y(t)} - 1| \le 1/2\}.$$

satisfies the same inequality as $\mathcal{O}_{N,1}$ and $\mathcal{O}_{N,2}$. Consequently, define

$$\widehat{\widehat{\alpha}}_m = \arg\min_{h\in S_m} \Lambda_N(h),$$

$$\widehat{\widehat{\Psi}}_m = \left(\int_0^T \varphi_j(t)\varphi_k(t)Y(t)dt\right)_{0\le j,k\le m-1}, \widehat{\widehat{Z}}_m = \left(\int_0^T \varphi_j(t)\frac{1}{N}\sum_{k=1}^N X_k(t)dX_k(t)\right)_{0\le j\le m-1}$$

and $\widehat{\widehat{\alpha}}_m(t) = \sum_{j=0}^{m-1}[\widehat{\widehat{\alpha}}_{(m)}]_j\varphi_j(t)$, we get that the vector of the coefficients of $\widehat{\widehat{\alpha}}_m(t)$ is equal to $\widehat{\widehat{\alpha}}_{(m)} = (\widehat{\widehat{\Psi}}_m)^{-1}\widehat{\widehat{Z}}_m$. We define analogously the norms $\|h\|_{\sqrt{Y}}$ and $\|h\|_{\sqrt{y}}$ and we can prove:

Proposition 11. *Assume* **[H1]-[H2]**. *Let* m *satisfy* $m \le N$ *and* $L(S_m)(\|\Psi_m\|_{\text{op}}^{-1} \vee 1) \le \frac{\mathfrak{c}N}{2\log N}$ *with* $\Psi_m = \mathbb{E}\widehat{\widehat{\Psi}}_m$ *and* $\mathfrak{c}$ *a numerical constant. Define the truncated estimator*

$$\widetilde{\widetilde{\alpha}}_m = \widehat{\widehat{\alpha}}_m \mathbf{1}_{\Lambda_m}, \qquad \Lambda_m = \{L(S_m)(\|\widehat{\widehat{\Psi}}_m\|_{\text{op}}^{-1} \vee 1) \le \mathfrak{c}N/\log N\}.$$

Then we have

$$\mathbb{E}\|\widetilde{\widetilde{\alpha}}_m - \alpha\|_{\sqrt{Y}}^2 \le \inf_{h\in S_m}\|h-\alpha\|_{\sqrt{y}}^2 + \frac{2m}{N} + \frac{c'}{N}, \quad \mathbb{E}\|\widetilde{\widetilde{\alpha}}_m - \alpha\|_{\sqrt{y}}^2 \le 2\inf_{h\in S_m}\|h-\alpha\|_{\sqrt{y}}^2 + 4\frac{m}{N} + \frac{c}{N}.$$

As previously, we can define an adaptive estimator with a data-driven choice of the dimension m under the assumption that $L(S_m) \le c_\varphi^2 m$. In parallel as above, we set

$$\mathcal{M}_N = \left\{m \le N, L(S_m)\|\Psi_m^{-1}\|_{\text{op}} \le \frac{\mathfrak{c}}{2}\frac{N}{\log(N)}\right\}, \ \widehat{\widehat{\mathcal{M}}}_N = \left\{m \le N, L(S_m)\|\widehat{\widehat{\Psi}}_m^{-1}\|_{\text{op}} \le \mathfrak{c}\frac{N}{\log(N)}\right\},$$

$$\widehat{\widehat{m}} = \arg\min_{m\in\widehat{\widehat{\mathcal{M}}}_n} \left\{\Lambda_N(\widehat{\alpha}_m) + \text{pen}(m)\right\}, \quad \text{pen}(m) = \kappa\frac{m}{N}. \tag{26}$$

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

$$\mathbb{E}\left(\|\widehat{\alpha}_{\widehat{\widehat{m}}} - \alpha\|_{\sqrt{Y}}\right) \leq C \inf_{m\in\mathcal{M}_N}\left(\inf_{h\in S_m}\|h-\alpha\|^2_{\sqrt{y}} + \kappa\frac{m}{N}\right) + \frac{c}{N},$$

where C is a numerical constant.

Proofs are omitted: they follow the same lines of the proofs of the previous parts and are simpler.

5.2. **Estimators of $\alpha(t)$ from Section 3.** When $\alpha(t) = \gamma(t)$ *i.e.* when $\beta(t) = 0$, we also have at disposal two other estimators of $\alpha(t)$: the first estimator of $\alpha(t)$ (denoted below [M1])and the estimator of $\gamma(t)$ (below [M3]), which are distinct and different from the estimator of the previous subsection (below [M2]). These three estimators are implemented in Section 6 and their risks computed in Table 2. It appears that, with the same basis for both, the estimator of α given from $\gamma^\star_{m^\star}$ by [M3] is better than $\widehat{\alpha}_{\widehat{m}}$ given by [M1]. Here is an interpretation of this phenomenon.

Suppose that we are in a parametric model where $\alpha(t) = \sum_{j=0}^{m-1}\varphi_j(t)\alpha_j$, $\gamma(t) = \sum_{j=0}^{m-1}\varphi_j(t)\gamma_j$ with fixed and known m. Then the exact maximum likelihood estimators of $\alpha_{(m)} = {}^t(\alpha_0,\dots,\alpha_{m-1}))$ and $\gamma_{(m)} = {}^t(\gamma_0,\dots,\gamma_{m-1}))$ are respectively (see (12)):

$$\widehat{\alpha}_{(m)} = \widehat{\Psi}^{-1}_{m,1}\widehat{Z}_{m,1}, \quad \gamma^\star_{(m)} = \widehat{\Psi}^{-1}_{m,2}\widehat{Z}_{m,2}.$$

From these exact expressions, a simple computation shows that, as N tends to infinity,

$$\sqrt{N}(\widehat{\alpha}_{(m)} - \alpha_{(m)}) \to_{\mathcal{L}} \mathbb{X}_1 \sim \mathcal{N}_m(0, \Psi^{-1}_{m,1}) \quad \sqrt{N}(\gamma^\star_{(m)} - \gamma_{(m)}) \to_{\mathcal{L}} \mathbb{X}_2 \sim \mathcal{N}(0, \Psi^{-1}_{m,2}).$$

where we recall that

$$\Psi_{m,1} = \mu_0^2\left(\int_0^T \varphi_j(t)\varphi_k(t)\exp{(2\int_0^t\alpha(s)ds)}dt\right)_{0\leq j,k\leq m-1},$$

$$\Psi_{m,2} = \left(\int_0^T \varphi_j(t)\varphi_k(t)[\sigma_0^2\exp{(2\Gamma(t)} + \exp{(2\Gamma(t)}\int_0^t\exp{(-2\Gamma(s))}ds)]dt\right)_{0\leq j,k\leq p-1}.$$

Therefore, if $\mu_0^2 \leq \sigma_0^2$ and $\alpha(t) = \gamma(t)$, $\Psi_{m,1} < \Psi_{m,2}$ (in the sense of inequality between positive symmetric matrices), thus, $\Psi^{-1}_{m,1} > \Psi^{-1}_{m,2}$.

Then $N\|\widehat{\alpha}(\cdot) - \alpha(\cdot)\|^2 = N\ {}^t(\widehat{\alpha}_{(m)} - \alpha_{(m)})(\widehat{\alpha}_{(m)} - \alpha_{(m)}) \to_{\mathcal{L}} Z_1 = {}^t\mathbb{X}_1\mathbb{X}_1$ and $N\|\gamma^\star(\cdot) - \gamma(\cdot)\|^2 \to_{\mathcal{L}} Z_2 = {}^t\mathbb{X}_2\mathbb{X}_2$. Setting $Y_1 = \Psi^{1/2}_{m,1}\mathbb{X}_1$, we have $Y_1 \sim \mathcal{N}_m(0, \text{Id}_m)$ and $Z_1 = {}^tY_1\Psi^{-1}_{m,1}Y_1)$. Thus, for $Y \sim \mathcal{N}_m(0, \text{Id}_m)$, it holds $\mathbb{E}(Z_1) = \mathbb{E}(\ {}^tY\Psi^{-1}_{m,1}Y)$. Analogously $\mathbb{E}(Z_2) = \mathbb{E}(\ {}^tY\Psi^{-1}_{m,2}Y)$ and thus $\mathbb{E}(Z_1) > \mathbb{E}(Z_2)$. This means that, for $\mu_0^2 \leq \sigma_0^2$, the estimator $\gamma^\star_{(m)}$ is asymptotically better than $\widehat{\alpha}_{(m)}$. This is illustrated in the simulation experiments.

6. Numerical experiments

We consider the four couples of functions $(\alpha_\ell, \gamma_\ell)$ for $\ell \in \{1,\dots,4\}$:

(1) $\alpha_1(t) = \frac{1}{2} + \frac{1}{4}t$, $\gamma_1(t) = 1 - t$,
(2) $\alpha_2(t) = -1 + t^2/2$, $\gamma_2(t) = 1 - t^2/2$,
(3) $\alpha_3(t) = \cos(1.2\,\pi t/2)$, $\gamma_3(t) = \sin(1.2\,\pi t/2)$,
(4) $\alpha_4(t) = \exp(-(t-1)^2)$, $\gamma_4(t) = \exp(-(t-3/2)^2/2)$.

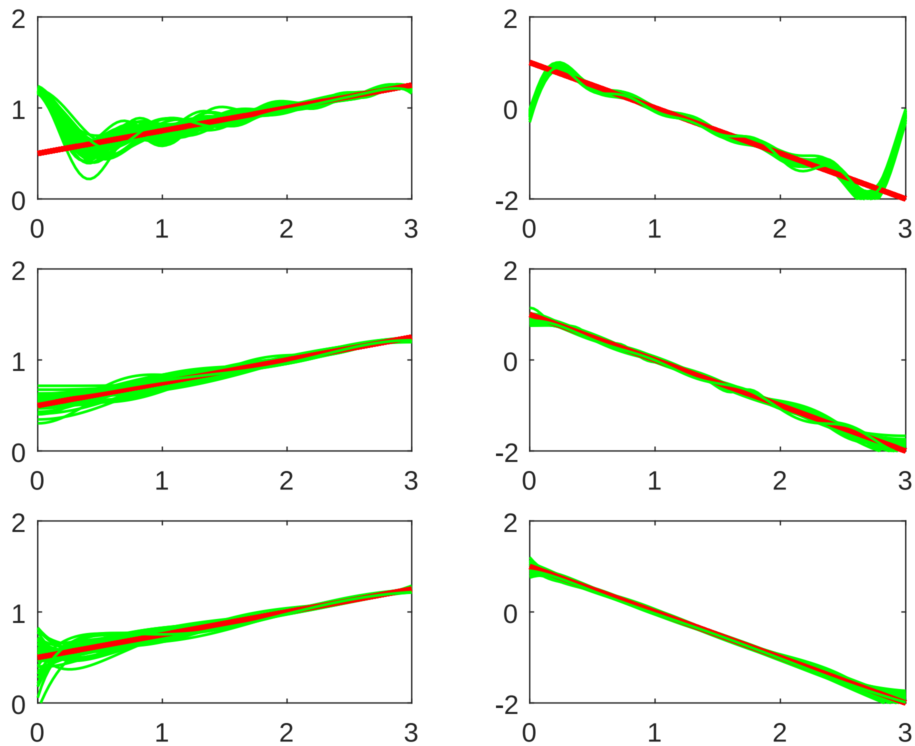

FIGURE 1. True functions in bold red and beam of 40 estimated α (left) and γ (right) with bases [T] (top), [C] (middle) and [L] (bottom) for Example 1 and $N = 1000$. The MISE for α are $0.079, 0.0053, 0.0072$ and the mean of selected dimensions are $4.62, 3.07, 4.90$. The MISE for γ are $0.374, 0.0094, 0.0067$ and mean selected dimensions are $5.0, 5.45, 5.47$.

For $T = 1, 3$, and $N = 250, 1000$, discrete samples are generated with Euler scheme with step $T/1000$, and initial conditions $\mu_0 = 1/2, \sigma_0 = 1$ (note that $\mu_0 \leq \sigma_0$, see Section 5.2). We proceed with 400 repetitions. Three bases are tested: the standard trigonometric basis [T], the cosine-basis [C] and the Laguerre basis [L], see section 3.3.

The cutoff is replaced by a limitation in the collection of models: maximal dimensions are less that 11 for [T], 26 for [C] and 7 for [L]. By doing so, all matrices are numerically invertible and we can check that the maximal dimension is not systematically chosen (otherwise we would enlarge the collection).

The penalty constants are taken as $\kappa_{1,[T]} = 2$, $\kappa_{1,[C]} = 4$ and $\kappa_{1,[L]} = 4$ for the estimation of α, and $\kappa_{2,[T]} = \kappa_{2,[C]} = \kappa_{2,[L]} = 2$ for the estimation of γ (κ_1 is defined in (23), κ_2 is defined in (24) and the additional index determines the basis). For each basis [T], [C], [L], the penalty constants are calibrated from preliminary numerical experiments.

The computed MISE is the mean over the experiments $k = 1, \ldots, 400$ of non-weighted approximated $\mathbb{L}^2$-error, for α:

$$\frac{1}{400}\sum_{k=1}^{400}\frac{T}{100}\sum_{i=1}^{100}\left[\alpha(\frac{iT}{100}) - \widehat{\alpha}^{(k)}_{\widehat{m}}(\frac{iT}{100})\right]^2,$$

where $\widehat{\alpha}_{\widehat{m}}^{(k)}$ is the estimator computed for simulation k, and analogously for γ.

		Estimation of α				Estimation of γ			
		$N=250$		$N=1000$		$N=250$		$N=1000$	
		$T=1$	$T=3$	$T=1$	$T=3$	$T=1$	$T=3$	$T=1$	$T=3$
Ex. 1	[T]	$5.02_{(5.6)}$	$10.6_{(2.3)}$	$1.49_{(1.4)}$	$7.74_{(1.4)}$	$3.63_{(1.0)}$	$39.9_{(3.7)}$	$2.07_{(0.4)}$	$37.7_{(1.7)}$
	[C]	$2.34_{(3.1)}$	$1.62_{(1.9)}$	$0.61_{(0.8)}$	$0.66_{(0.6)}$	$1.01_{(1.3)}$	$2.60_{(1.9)}$	$0.32_{(0.2)}$	$0.86_{(0.5)}$
	[L]	$3.86_{(3.8)}$	$3.64_{(3.5)}$	$0.82_{(0.8)}$	$0.99_{(0.9)}$	$0.97_{(0.9)}$	$2.35_{(1.6)}$	$0.23_{0.2)}$	$0.56_{(0.4)}$
Ex. 2	[T]	$4.35_{(3.6)}$	$791_{(16)}$	$1.98_{(0.8)}$	$792_{(8.1)}$	$7.17_{(3.3)}$	$570_{(46)}$	$3.32_{(0.9)}$	$552_{(14)}$
	[C]	$1.92_{(2.8)}$	$45.3_{(11)}$	$0.63_{(0.5)}$	$39.3_{(4.8)}$	$4.14_{(4.5)}$	$31.2_{(19)}$	$1.17_{(1.1)}$	$10.2_{(5.3)}$
	[L]	$3.21_{2.7)}$	$3.35_{(3.6)}$	$0.84_{(0.9)}$	$0.74_{(0.7)}$	$3.79_{(3.3)}$	$17.1_{(12)}$	$0.82_{(0.7)}$	$4.76_{(3.7)}$
Ex. 3	[T]	$17.3_{(6.9)}$	$18.7_{(12)}$	$7.97_{(1.7)}$	$6.88_{(2.4)}$	$2.44_{(1.0)}$	$6.14_{(1.3)}$	$1.10_{(0.3)}$	$4.79_{(0.5)}$
	[C]	$5.74_{(5.0)}$	$18.5_{(12)}$	$2.07_{(1.2)}$	$5.87_{(2.8)}$	$1.59_{(1.1)}$	$2.57_{(1.2)}$	$0.49_{(0.3)}$	$0.82_{(0.3)}$
	[L]	$7.32_{(6.3)}$	$22.1_{(18)}$	$2.71_{(1.6)}$	$7.68_{(2.0)}$	$1.34_{(0.9)}$	$8.63_{(2.1)}$	$0.35_{(0.2)}$	$7.53_{(0.8)}$
Ex. 4	[T]	$6.76_{(5.5)}$	$3.95_{(2.1)}$	$2.73_{(1.1)}$	$2.09_{(0.8)}$	$2.28_{(0.9)}$	$0.52_{(0.4)}$	$1.09_{(0.3)}$	$0.22_{(0.1)}$
	[C]	$2.72_{[3.7)}$	$3.11_{(3.6)}$	$0.68_{(0.7)}$	$0.81_{(0.5)}$	$0.94_{(1.3)}$	$0.87_{(0.8)}$	$0.26_{(0.3)}$	$0.30_{(0.2)}$
	[L]	$3.36_{(4.6)}$	$5.92_{(4.2)}$	$0.90_{(0.8)}$	$1.02_{(1.0)}$	$1.03_{(0.9)}$	$0.95_{(0.8)}$	$0.25_{(0.2)}$	$023_{(0.2)}$

TABLE 1. 100× MISE for estimation of α and γ (with 100 × standard deviation in parenthesis) in examples 1 to 4 with bases [T], [C] and [L], for $N = 250$ and $N = 1000$, $T = 1$ and $T = 3$ and for $\mu_0 = 1/2$ and $\sigma_0 = 1$.

The global results are given in Table 1. As expected, in all cases, the MISE gets smaller when N increases. Clearly, the trigonometric basis [T] has difficulty for the estimation of non periodic functions (that is, functions which do not take the same value in 0 and T), and gives results which are systematically less good than the two others. This is also illustrated by Figure 1 for example 1, in which beams of 40 estimators (green) for N=1000 and $T = 3$ are compared to the true functions (red): the plots on the first line for basis [T] have clearly important side-effects, while the two other bases seem to correct it. This is the reason why we implemented basis [C]. When T increases, the MISE most of the time increases also, which seems to be a natural scale effect, and the MISE for γ is generally smaller than the MISE for α: it is true that the functions are different, but they are of similar types, so it is likely that γ is easier to estimate than α, see also Figure 1 and compare left plots (estimation of α) and right plots (estimation of γ). This is in accordance with the results of section 5.2, which indicate that the estimator of γ has smaller risk than the estimator of α when $\mu_0^2 \leq \sigma_0^2$. Figures 2 and 3 allow to compare the improvement when going from $N = 250$ (Figure 2) to $N = 1000$ (Figure 3) on the same example 2. Lastly, Figure 4 is a plot for $N = 1000$ and $T = 3$ concerning example 4.

We also experimented the case $\alpha \equiv \gamma$ or $\beta \equiv 0$ described in Section 5. We compare in Table 2 the MISE obtained when estimating α by methods:

- M1 corresponding to the strategy of estimation of α of the general setting,
- M2 corresponding to the specific strategy described in section 5, with constant in (26) chosen as $\kappa_{[C]} = \kappa_{[L]} = 4$,
- M3 corresponding to the strategy of estimation of γ in the general setting.

We took for example $1^\star$, $\alpha_1(t) = \gamma_1(t) = 1/2 + t/4$, for example $2^\star$, $\alpha_2(t) = \gamma_2(t) = 1 - t^2/2$, for example $3^\star$, $\alpha_3(t) = \gamma_3(t) = \cos(1.2\,\pi t)$ and for example $4^\star$, $\alpha_4(t) = \gamma_4(t) = \exp(-(t-1)^2)$. In other words, we kept the same examples of funtions α, and changed γ to take it equal to α, except in example 2. Method M1 is systematically the less good. The two other methods

		$N = 250$			$N = 1000$		
		M1	M2	M3	M1	M2	M3
Ex.1*	[C]	$1.62_{(1.2)}$	$0.45_{(0.4)}$	$0.73_{(0.7)}$	$0.65_{0.6)}$	$0.15_{(0.1)}$	$0.22_{(0.2)}$
	[L]	$3.64_{(3.53)}$	$0.57_{(0.5)}$	$0.76_{(0.7)}$	$0.90_{(0.8)}$	$0.16_{(0.1)}$	$0.19_{(0.1)}$
Ex.2*	[C]	$21.1_{(15)}$	$4.50_{(2.9)}$	$3.74_{(2.5)}$	$9.35_{(6.7)}$	$1.69_{(6.7)}$	$1.24_{(0.7)}$
	[L]	$20.0_{26)}$	$2.10_{(1.4)}$	$2.19_{(1.5)}$	$4.14_{(4.6)}$	$0.54_{0.5)}$	$0.54_{(0.5)}$
Ex. 3*	[C]	20.0(24)	$2.60_{0.9)}$	$2.83_{(1.3)}$	$5.71_{(2.8)}$	$0.91_{(0.3)}$	$0.97_{(0.4)}$
	[L]	$22.5_{(9)}$	$5.87_{(1.0)}$	$60.16_{(1.2)}$	$7.57_{(2.1)}$	$4.83_{(0.4)}$	$4.96_{(0.4)}$
Ex.4*	[C]	$3.24_{(3.6)}$	$0.66_{(0.4)}$	$0.75_{(0.6)}$	$0.87_{(0.5)}$	$0.25_{(0.1)}$	$0.26_{(0.2)}$
	[L]	$5.31_{(3.7)}$	$0.74_{(0.5)}$	0.85(0.7)	$1.09_{(1.0)}$	$0.30_{(0.2)}$	$0.27_{(0.2)}$

TABLE 2. When $\beta(t) \equiv 0$ (and thus $\alpha = \gamma$), $100\times$ MISE for estimation of α (with $100 \times$ standard deviation in parenthesis) in examples 1 to 4 with bases [C] and [L], for $N = 250$ and $N = 1000$, $T = 3$ and for $\mu_0 = 1/2$ and $\sigma_0 = 1$. The three methods are : M1 the method of estimation of α in the complete model, M3 the method of estimation of $\gamma = \alpha$ in the complete model and M2 the specific method of section 5.

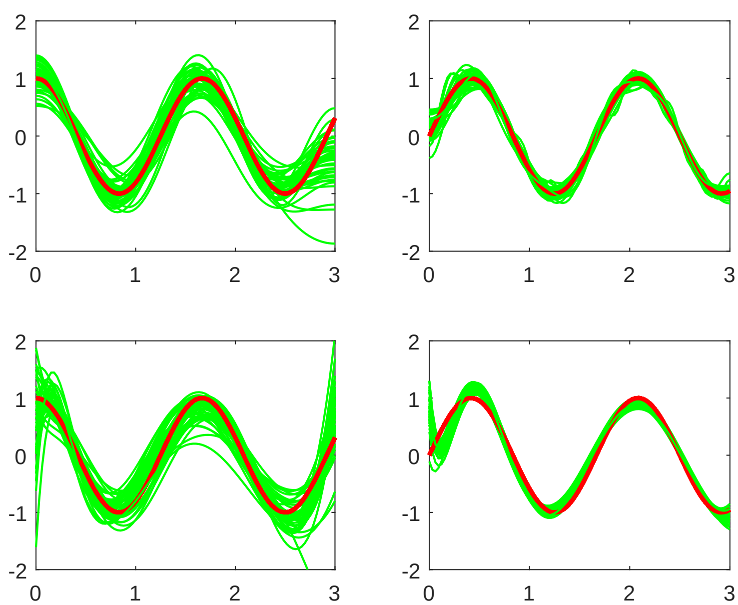

FIGURE 2. True functions in bold red and beam of 40 estimated α (left) and γ (right) with bases [C] (top) and Laguerre (bottom) for Example 3 with $N = 250$. MISE for α: $0.2002, 0.2666$ and mean of selected dimensions: $4.35, 6.65$, MISE for γ: $0.0226, 0.0802$ and mean of selected dimensions $8.2, 7.0$.

give similar results, even if method M2 seems almost all the time better although probably not significantly.

To conclude this section, we can say that the method works globally well in most contexts.

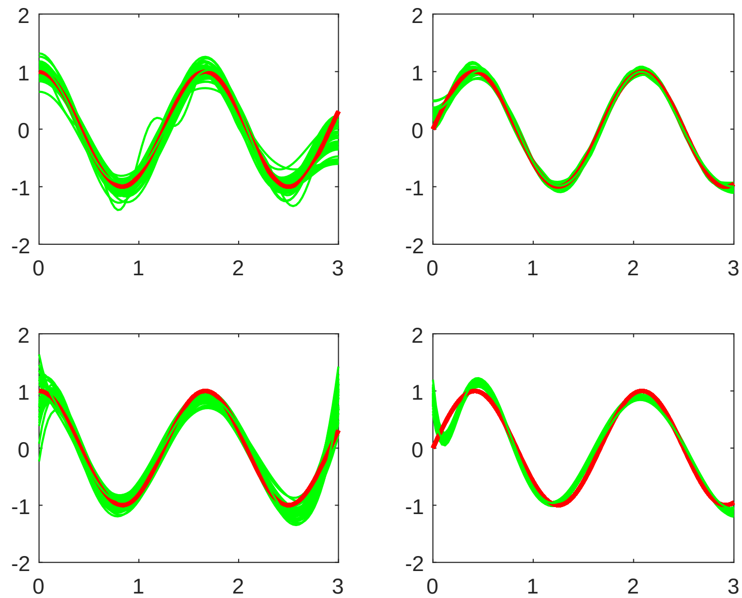

FIGURE 3. True functions in bold red and beam of 40 estimated α (left) and γ (right) with bases [C] (top) and Laguerre (bottom) for Example 3 with $N = 1000$. MISE for α: $0.0622, 0.0827$ and mean of selected dimensions: $5.6, 7.0$. MISE for γ: $0.0239, 0.0817$ and mean of selected dimensions: $9.97, 7.0$.

7. Concluding remarks.

In this paper, we study the nonparametric estimation of the deterministic functions $\alpha(t), \beta(t)$ when the observed process is an interacting system of N particles given by (1). The process is assumed to be continuously observed throughout a time interval $[0, T]$ with fixed T. The number N of particles is large. We build estimators of the functions $\alpha(t), \gamma(t) = \alpha(t) - \beta(t)$ by minimizing projection contrasts deduced from likelihoods, using increasing sequences of finite dimensional subspaces of $\mathbb{L}^2([0, T])$. Bounds for the $\mathbb{L}^2$-risk of the projection estimators are given based either on an empirical norm or a deterministic norm linked with the problem. The bounds of the risks allow to discuss rates of convergence. Then, a data-driven choice of the dimension for the projection space is provided leading to an adaptive result. The case where $\beta(t) \equiv 0$ is briefly treated.
Implementation of the estimators is done based on simulated data for various examples of fonctions $\alpha(t), \gamma(t)$ and two different bases of $\mathbb{L}^2([0, T])$. The numerical results show that the adaptive estimators perform well, the estimation of $\gamma(t)$ being better than the estimation of $\alpha(t)$.

To go further on the topic, the problem of discrete time observation of the processes, with small or fixed sampling interval, may be considered. The generalization of our study to include a diffusion coefficient $sigma(X_i^N(t))$ in (1) with a known $\sigma(\cdot)$ is certainly feasible. More challenging, the study of the estimation of α, β in the general dynamics

$$dX_i^N(t) = \{\alpha(t)X_i^N(t) - \beta(t)\frac{1}{N}\sum_{j=1}^{n}\phi(X_i^N(t) - X_j^N(t))\}dt + dW_i(t), \quad X_i^N(0) = X_0^i, i = 1, \dots, N,$$

with known $\phi(\cdot)$, is under study.

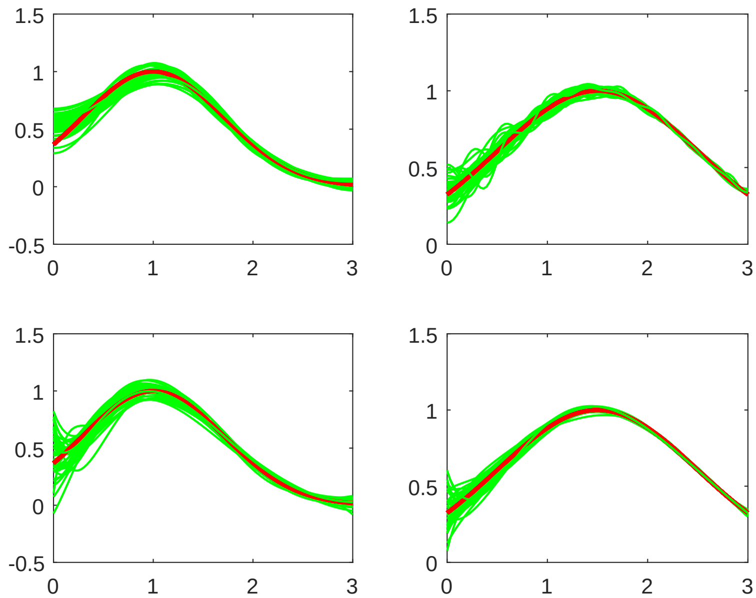

FIGURE 4. True functions in bold red and beam of 40 estimated α (left) and γ (right) with bases [C] (top) and Laguerre (bottom) for Example 4 with $N = 1000$. MISE for α: $0.0084, 0.0091$ and mean of selected dimensions: $3.2, 5.05$. MISE for γ: $0.0031, 0.0022$ and mean of selected dimensions: $11.2, 5.8$.

8. PROOFS

Recall the notations (2)-(3) and $X_j = X_j^N$.

8.1. **Proof of Proposition 1.** We have:

$$d(\sum_{j=1}^N X_j(t)) = [\alpha(t)\sum_{j=1}^N X_j(t) - \beta(t)\sum_{j=1}^N (X_j(t) - X(t))]dt + \sum_{j=1}^N dW_j(t) = \alpha(t)\sum_{j=1}^N X_j(t)dt + \sum_{j=1}^N dW_j(t).$$

Therefore, $dX(t)) = \alpha(t)X(t)dt + \frac{1}{N}\sum_{j=1}^N dW_j(t) = \alpha(t)X(t))dt + \frac{1}{\sqrt{N}}dB_{N,1}(t)$. This equation can be easily solved and yields the solution given in the proposition.
We have:

$$\sup_{t\leq T}|X(t)) - x(t)| \leq \sup_{t\leq T}\exp\left(\int_0^t \alpha(u)du\right)\left(|X(0) - \mu_0| + \sup_{t\leq T}\frac{1}{\sqrt{N}}\left|\int_0^t \exp\left(-\int_0^s \alpha(u)du\right)dB_{N,1}(s)\right|\right)$$

Using the Markov and Burkholder-Davis-Gundy (B-D-G) inequalities yields

$$\mathbb{P}\left(\sup_{t\leq T}\frac{1}{\sqrt{N}}\left|\int_0^t \exp\left(-\int_0^s \alpha(u)du\right)dB_{N,1}(s)\right| > \varepsilon\right) \leq \frac{C_4}{\varepsilon^4 N^2}\left(\int_0^T \exp\left(-2\int_0^s \alpha(u)du\right)ds\right)^2$$

where C_4 is the constant of the B-D-G inequality. We conclude that $\sup_{t\leq T}|\frac{1}{\sqrt{N}}g_N(t)|$ converges a.s. to 0 and using [H1], this yields the result.$\square$

8.2. **Proof of Proposition 2.** We apply Ito's formula to the function $F(x,y)=(x-y)^2$ and use that $\langle dX_i, dX_i\rangle_t = dt$, $\langle dX, dX\rangle_t = \frac{1}{N}dt$, $\langle dX_i, dX\rangle_t = \frac{1}{N}dt$,

$$d[X_i(t)-X(t)]^2 = 2[X_i(t)-X(t)]dX_i(t) - 2[X_i(t)-X(t)]dX(t) + \frac{1}{2}[2dt + \frac{2}{N}dt - \frac{4}{N}dt]$$

Using $\sum_{j=1}^n (X_j(t)-X(t)) = 0$ and (6) yields

$$d\{\sum_{i=1}^N [X_i(t)-X(t)]^2\} = 2\sum_{i=1}^N [X_i(t)-X(t)]dX_i(t) + (N-1)dt$$

$$= 2\sum_{i=1}^N [X_i(t)-X(t)][\alpha(t)X(t)+\gamma(t)[X_i(t)-X(t)]dt + (N-1)dt + 2\sum_{i=1}^N [X_i(t)-X(t)]dW_i(t).$$

Thus, $dV(t) = [2\gamma(t)V(t) + 1 - \frac{1}{N}]dt + dM_N(t)$ where

$$M_N(t) = \frac{2}{N}\int_0^t \sum_{i=1}^N [X_i(s)-X(s)]dW_i(s) = \frac{2}{\sqrt{N}}\int_0^t \sqrt{V(s)}dB_{N,2}(s).$$

By the usual change $V(t) = C(t)\exp(2\int_0^t \gamma(s)ds)$, setting $\Gamma(t) = \int_0^t \gamma(s)ds$, we can obtain the expression:

$$\begin{aligned} V(t) &= \exp(2\Gamma(t))\left(V(0) + (1-\frac{1}{N})\int_0^t \exp(-2\Gamma(s))ds\right) + \exp(2\Gamma(t))\int_0^t \exp(-2\Gamma(s))dM_N(s) \\ &= \exp(2\Gamma(t))\left(V(0) + (1-\frac{1}{N})\int_0^t \exp(-2\Gamma(s))ds\right) + M_N(t) \\ &+ 2\exp(2\Gamma(t))\int_0^t \gamma(s)\exp(-2\Gamma(s)M_N(s)ds. \end{aligned}$$

Note that

$$V(t) = v(t) + A_N(t) + B_N(t) \tag{27}$$

where

$$\begin{aligned} A_N(t) &= \exp(2\Gamma(t))\left(\frac{1}{N}\sum_{i=1}^N [X_i(0)-\mu_0]^2 - \sigma_0^2 - [\mu_0 - X(0)]^2 - \frac{1}{N}\int_0^t \exp(-2\Gamma(s)ds\right), \\ B_N(t) &= M_N(t) + 2\exp(2\Gamma(t))\int_0^t \gamma(s)\exp(-2\Gamma(s)M_N(s)ds. \end{aligned}$$

By [H1], $A_N(t)$ converges to 0 almost surely uniformly on $[0,T]$ as N tends to infinity. To obtain that $V(t)$ converges to $v(t)$ uniformly almost surely on $[0,T]$ to 0, it is enough to prove that

$$\sup_{s\le T}|M_N(s)| \to_{a.s.} 0. \tag{28}$$

For this, we follow Kasonga (1990, p.873).
One of Doob's martingale inequalities states that, for $\alpha>0, \beta>0$, $\mathbb{P}(\sup_{s\le T}(M_N(s) - \frac{\alpha}{2}\langle M_N\rangle_s) > \beta) \le e^{-\alpha\beta}$. Here $\langle M_N\rangle_s = \frac{4}{N}\int_0^s V(u)du$. This yields

$$\mathbb{P}(\sup_{s\le T}|M_N(s)| > \frac{2\alpha}{N}\int_0^T V(u)du + \beta) \le 2e^{-\alpha\beta}$$

Taking $\alpha = N^a, \beta = N^{-b}$ with $0 < b < a < 1$ yields

$$\mathbb{P}(\sup_{s\leq T}|M_N(s)| > \frac{2}{N^{1-a}}\int_0^T V(u)du) + N^{-b}) \leq 2e^{-N^{a-b}}.$$

By (27), we have

$$\int_0^T V(t)dt \leq C(T)\left(\sup_{t\in[0,T]}[v(t)+A_N(t)] + \sup_{t\in[0,T]}|M_N(t)|\right)$$

where $C(T) = 1 + 2\sup_{t\in[0,T]}\exp(2\Gamma(t))\int_0^T|\gamma(s)|\exp(-2\Gamma(s)ds$. Consequently,

$$\begin{aligned}
&\mathbb{P}\left(\left(1-\frac{2C(T)}{N^{1-a}}\right)\sup_{s\leq T}|M_N(s)| > N^{-b} + \frac{2C(T)}{N^{1-a}}(\sup_{t\in[0,T]}\{v(t)+A_N(t)\}\right)\\
&= \mathbb{P}\left(\sup_{s\leq T}|M_N(s)| > N^{-b} + \frac{2C(T)}{N^{1-a}}[\sup_{t\in[0,T]}[v(t)+A_N(t)] + \sup_{t\in[0,T]}|M_N(t)|]\right)\\
&\leq \mathbb{P}\left(\sup_{s\leq T}|M_N(s)| > \frac{2}{N^{1-a}}\int_0^T V(u)du + N^{-b}\right) \leq 2e^{-N^{a-b}}.
\end{aligned}$$

As $\sup_{t\in[0,T]}[v(t)+A_N(t)]$ converges almost surely, by the Borel-Cantelli lemma, we obtain (28). So the proof of Proposition 2 is complete. □

8.3. **Proof of Lemma 1.** We have for $r \geq 1$,

$$\mathbb{E}|X^r(t)| \leq 2^{r-1}\left(\exp(r\int_0^T\alpha(s)ds)\mathbb{E}|X^r(0)| + \frac{1}{N^{r/2}}\mathbb{E}|g_N^r(t)|\right).$$

By the definition of g_N,

$$\mathbb{E}|g_N^r(t)| = C_r\sigma^r(t),\ \text{where } \sigma^2(t) = \exp(2\int_0^t\alpha(s)ds)\int_0^t\exp(-2\int_0^s\alpha(u)du)ds,\ C_r = \mathbb{E}|Z|^r,$$

for Z a standard Gaussian variable. Next, we have, as the $X_i(0)$s are *i.i.d.* and have moments of any order, using the Rosenthal inequality (see Hall and Heyde, 1980, p.23-24),

$$\begin{aligned}
\mathbb{E}|X^r(0)| &\leq 2^{r-1}(|\mu_0|^r + \mathbb{E}|X^r(0)-\mu_0|^r)\\
&\lesssim |\mu_0|^r + \frac{1}{N^{r-1}}\mathbb{E}|X_i(0)-\mu_0|^r + \frac{1}{N^{r/2}}(\mathbb{E}|X_i(0)-\mu_0|^2)^{r/2} \leq c_r.
\end{aligned}$$

Thus, for all $t\in[0,T]$, $\mathbb{E}|X^r(t)| \leq C_r(T)$ for some constant $C_r(T)$.

Note that for all t, $V(t) > 0$. Analogously, by the Rosenthal inequality, we check that, for all $r \geq 1$, $\mathbb{E}V^r(0) \leq C$ where the constant C does not depend on N.
The process $(V(t)$ is solution of a stochastic differential equation with drift $b(t,v) = 2\gamma(t)v + 1 - N^{-1}$ and diffusion coefficient $\sigma(v) = 2\sqrt{v^+}N^{-1/2}$ satisfying $b^2(t,v)+\sigma^2(v) \leq K_T(1+v^2)$ with $K_T = \sup\{2, 8\sup_{t\leq T}\gamma^2(t)\}$. Therefore, for all $r\geq 1$, using the equation for $V(t)$, $\mathbb{E}V^{2r}(0)\leq C$, the Cauchy-Schwarz and the BDG inequalities, we standardly obtain, for another constant $K_r(T)$,

$$\mathbb{E}V^{2r}(t) \leq K_r(T)\left(1+\int_0^t\mathbb{E}V^{2r}(s)ds\right).$$

By the Gronwall lemma, this yields $\mathbb{E}V^{2r}(t) \leq K_r(T)e^{TK_r(T)}$. The conclusion follows, using that, for all $r\geq 1$, $\mathbb{E}V^r(t) \leq \mathbb{E}^{1/2}V^{2r}(t)$. □

8.4. **Proof of Proposition 3.** By Proposition 1, as $\mu_0 \neq 0$, the process $X(t)$ satisfies

$$\frac{X(t)}{x(t)} - 1 = (\frac{X(0)}{\mu_0} - 1) + \frac{1}{\mu_0\sqrt{N}} L_N(t),$$

where $L_N(t) = \int_0^t \exp(-\int_0^s \alpha(u)du)dB_{N,1}(s)$ is a martingale with deterministic bracket $\langle L_N \rangle t = \int_0^t \exp(-2\int_0^s \alpha(u)du)ds := d(t)$. We get,

$$\mathbb{P}(\sup_{t\in[0,T]} |\frac{X(t)}{x(t)} - 1| > \delta) \quad \leq \quad \mathbb{P}(|\frac{X(0)}{\mu_0} - 1| > \delta/2) + \mathbb{P}(\sup_{t\in[0,T]} |L_N(t)| > \delta\mu_0\sqrt{N}/2)$$

The Rosenthal inequality yields, for all $r \geq 1$,

$$\begin{aligned}\mathbb{P}(|\frac{X(0)}{\mu_0} - 1| > \delta/2) \quad &\leq \quad \frac{2^{2r}}{(\delta\mu_0)^{2r}} \mathbb{E}|X(0) - \mu_0|^{2r} \lesssim \frac{1}{N^{2r-1}} \mathbb{E}|X_i(0) - \mu_0|^{2r} + \frac{1}{N^r}(\mathbb{E}|X_i(0) - \mu_0|^2)^r \\ &\lesssim \quad \frac{1}{N^{2r-1}} \mathbb{E}|X_i(0) - \mu_0|^{2r} + \frac{\sigma_0^{2r}}{N^r}\end{aligned}$$

The Bernstein inequality for martingales (see Revuz and Yor (1999), p.153-154) yields

$$\mathbb{P}(\sup_{t\in[0,T]} |L_N(t)| > \delta\mu_0\sqrt{N}/2, \langle L_N \rangle_T \leq d(T)) \leq 2\exp(-\frac{\delta^2\mu_0^2 N}{8d(T)}).$$

Now, we have:

$$\begin{aligned}&\mathbb{P}\left(\sup_{t\in[0,T]} |\frac{X^2(t)}{x^2(t)} - 1| > \delta\right) \leq \mathbb{P}\left(\sup_{t\in[0,T]} |\frac{X(t)}{x(t)} - 1|^2 > \delta/2\right) + \mathbb{P}\left(\sup_{t\in[0,T]} |\frac{X(t)}{x(t)} - 1| > \delta/4\right) \\ \leq \quad &\frac{c(r)}{\delta^{2r} N^r} + 2\left(\exp[-\frac{\delta\mu_0^2 N}{16d(T)}] + \exp[-\frac{\delta^2\mu_0^2 N}{16 \times 8d(T)}]\right).\end{aligned}$$

Thus, for some positive constants $C(T), c(T)$ depending on μ_0, σ_0^2 and T,

$$\mathbb{P}\left(\sup_{t\in[0,T]} |\frac{X^2(t)}{x^2(t)} - 1| > 1/2\right) \leq C(T)\exp(-c(T)N) + \frac{c(r)}{N^r}. \quad \square$$

Remark 1. *If $\mu_0 = 0$, then $x(t) \equiv 0$, which forbides the ratio; if $\sigma_0^2 = 0$, then $v(0) = 0$ and analogous problem arises. If $X_i(0)$ is Gaussian or sub-Gaussian, we have a pure exponential bound.*

8.5. **Proof of Proposition 4.** We have

$$\begin{aligned}\frac{V(t)}{v(t)} - 1 \quad &= \quad \frac{(V(0) - v(0))}{v(0) + \int_0^t e^{(-2\Gamma(s))}ds} - \frac{1}{Nv(0) + \int_0^t e^{-2\Gamma(s)}ds}\int_0^t e^{-2\Gamma(s)}ds \\ &+ \frac{2}{(v(0) + \int_0^t e^{-2\Gamma(s)}ds)\sqrt{N}}\int_0^t e^{-2\Gamma(s)}\sqrt{V(s)}dB_{N,2}(s)\end{aligned}$$

Thus,

$$|\frac{V(t)}{v(t)} - 1| \leq \frac{|V(0) - v(0))|}{v(0)} + \frac{1}{Nv(0)}\int_0^t e^{-2\Gamma(s)}ds + \frac{2}{(v(0))\sqrt{N}}|\int_0^t e^{-2\Gamma(s)}\sqrt{V(s)}dB_{N,2}(s)| \tag{29}$$

We have $V(0) - v(0) = \frac{1}{N}\sum_{i=1}^N (X_i(0) - \mu_0)^2 - \sigma_0^2 - (X(0) - \mu_0)^2$. By the Rosenthal inequality, for all $r \geq 1$,

$$\mathbb{E}|V(0) - \mathbb{E}V(0)|^{2r} \leq \frac{C}{N^r}. \tag{30}$$

Set $K_N(t) = \int_0^t \exp(-2\Gamma(s))\sqrt{V(s)}d\beta'_N(s)$. We have $\langle K_N\rangle_t = \int_0^t \exp(-4\Gamma(s))V(s)ds$. Set $k(T) = \int_0^T \exp(-4\Gamma(s))v(s)ds$. For all $p \geq 1$,

$$\begin{aligned}\mathbb{P}(\sup_{t\leq T}|K_N(t)| \geq c\sqrt{N}) &\leq \mathbb{P}(\sup_{t\leq T}|K_N(t)| \geq c\sqrt{N}), \langle K_N\rangle_T \leq 1+k(T)) + \mathbb{P}(\langle K_N\rangle_T > 1+k(T))\\ &\leq 2\exp(-\frac{c^2N}{2(1+k(T))}) + \mathbb{E}\left(\int_0^T \exp(-4\Gamma(s))(V(s)-v(s))ds\right)^{2r}.\end{aligned}$$

In what follows, the constant $C(T)$ may change for one line to another:

$$\begin{aligned}&\mathbb{E}\left(\int_0^T e^{-4\Gamma(s)}(V(s)-v(s))ds\right)^{2r} \leq C(T)\int_0^T \mathbb{E}[e^{-4\Gamma(s)}(V(s)-v(s))]^{2r}ds\\ &\leq C(T)\left(|V(0)-v(0)| + \frac{1}{N}\int_0^T \exp(-2\Gamma(u))du\right)^{2r} + \frac{1}{N^r}\int_0^T \mathbb{E}\left(\int_0^s \exp(-4\Gamma(u))V(u)du\right)^r ds\\ &\leq C(T)\mathbb{E}\left(|V(0)-v(0)|^{2r} + \frac{1}{N^{2r}} + \frac{1}{N^r}\int_0^T \mathbb{E}[V(u)]^r du\right)\\ &\leq C(T)\left(\mathbb{E}|V(0)-v(0)|^{2r} + \frac{1}{N^{2r}} + \frac{1}{N^r}\right),\end{aligned}$$

applying Lemma 1. Thus, using (29)-(30), for all $\delta, r > 0$, there exist constants $c_\delta(T), C(r,T,\delta)$ such that

$$\mathbb{P}(\sup_{t\leq T}|\frac{V(t)}{v(t)} - 1| \geq \delta) \leq 2\exp(-c_\delta(T)N) + \frac{C(r,T,\delta)}{N^r}. \quad \square$$

8.6. **Proof of Proposition 6.** This proposition is analogous to Proposition 2 of Comte and Genon-Catalot (2020a). Let $t = \sum_{j=0}^{m-1} a_j\varphi_j$, and $\vec{a} = {}^t(a_0,\ldots,a_{m-1})$, then $\|t\|^2 = \|\vec{a}\|_{2,m} = {}^t\vec{a}\,\vec{a}$ and $\|t\|_x^2 = {}^t\vec{a}\Psi_{m,1}\vec{a} = \|\Psi_{m,1}^{1/2}\vec{a}\|_{2,m}^2$, where $\Psi_{m,1}^{1/2}$ is a symmetric square root of $\Psi_{m,1}$. Thus

$$\sup_{t\in S_m,\|t\|_x=1}\|t\|^2 = \sup_{\vec{a}\in\mathbb{R}^m,\|\Psi_{m,1}^{1/2}\vec{a}\|_{2,m}=1} {}^t\vec{a}\,\vec{a}.$$

Set $\vec{b} = \Psi_{m,1}^{1/2}\vec{a}$, that is $\vec{a} = \Psi_{m,1}^{-1/2}\vec{b}$. Then

$$\sup_{t\in S_m,\|t\|_x=1}\|t\|^2 = \sup_{\vec{b}\in\mathbb{R}^m,\|\vec{b}\|_{2,m}=1} {}^t\vec{b}\Psi_{m,1}^{-1}\vec{b} = \|\Psi_{m,1}^{-1}\|_{\mathrm{op}}.$$

As, for $m < m'$, S_m is strictly included in $S_{m'}$, the result follows for the first mapping and analogously for the others. $\square$

8.7. **Proof of Proposition 7.** We start with some preliminaries. On $\mathcal{O}_{N,1}$ defined in equation (8), the empirical norm $\|.\|_X$ and the $\|\cdot\|_x$-norm are equivalent for elements of $\mathbb{L}^2([0,T])$ as on $\mathcal{O}_{N,1}$

$$\forall h \in \mathbb{L}^2([0,T]), \quad (2/3)\|h\|_X^2 \leq \|h\|_x^2 \leq 2\|h\|_X^2.$$

We defined $\Lambda_{m,1}$ in equation (14) and let us set

$$\Omega_{m,1} := \left\{\left|\frac{\|h\|_X^2}{\|h\|_x^2} - 1\right| \leq \frac{1}{2}, \forall h \in S_m\right\}.$$

We note that $\mathcal{O}_{N,1} \subset \Omega_{m,1}$. Now, if $\vec{\mathbf{u}} = {}^t(u_0,\ldots,u_{m-1}) \in \mathbb{R}^m$ and $h = \sum_{j=0}^{m-1} u_j\varphi_j$, then

$$\|h\|_X^2 = {}^t\vec{\mathbf{u}}\widehat{\Psi}_{m,1}\vec{\mathbf{u}} \quad \text{and} \quad \|h\|_x^2 = {}^t\vec{\mathbf{u}}\Psi_{m,1}\vec{\mathbf{u}} = \|\Psi_{m,1}^{1/2}\vec{\mathbf{u}}\|_{2,m}^2, \quad \text{so that} \tag{31}$$

$$\begin{aligned}\sup_{h\in S_m,\|h\|_x=1}\left|\|h\|_X^2-\|h\|_x^2\right| &= \sup_{\vec{\mathbf{u}}\in\mathbb{R}^m,\|\Psi_{m,1}^{1/2}\vec{\mathbf{u}}\|_{2,m}=1}\left|\,{}^t\vec{\mathbf{u}}(\widehat{\Psi}_{m,1}-\Psi_{m,1})\vec{\mathbf{u}}\right|\\ &= \sup_{\vec{\mathbf{z}}\in\mathbb{R}^m,\|\vec{\mathbf{z}}\|_{2,m}=1}\left|\,{}^t\vec{\mathbf{z}}\Psi_{m,1}^{-1/2}(\widehat{\Psi}_{m,1}-\Psi_{m,1})\Psi_{m,1}^{-1/2}\vec{\mathbf{z}}\right|\\ &= \|\Psi_{m,1}^{-1/2}\widehat{\Psi}_{m,1}\Psi_{m,1}^{-1/2}-\mathrm{Id}_m\|_{\mathrm{op}}.\end{aligned}$$

Therefore,

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

Consequently, on $\mathcal{O}_{N,1}$, the eigenvalues of $\Psi_{m,1}^{-1/2}\widehat{\Psi}_{m,1}\Psi_{m,1}^{-1/2}$ belong to $[1/2,3/2]$.
The following lemma, proved in Section 8.8, holds

Lemma 2. *We have, for* $\mathtt{d}$ *a positive constant,* $\mathbb{P}(\mathcal{O}_{N,1}^c)\le \mathtt{d}/N^7$*. Moreover, under the assumptions of Proposition 7, for* m *satisfying (13), we have,* $\mathbb{P}(\Lambda_{m,1}^c)\le \mathtt{d}/N^7$.

Now, we prove inequality (16) of Proposition 7.
To study the risk of $\widetilde{\alpha}_m$, we need to have an adequate expression of the orthogonal projection of α with respect to $\langle .,.\rangle_X$. We have:

$$\widehat{\Psi}_{m,1}=(\langle\varphi_j,\varphi_\ell\rangle_X)_{0\le j,\ell\le m-1}.$$

The orthogonal projection $\pi_{(m)}^X\alpha$ of α on S_m with respect to the scalar product $\langle .,.\rangle_X$ is characterized by $\pi_{(m)}^X\alpha-\alpha\perp\varphi_j, j=0,\dots m-1$. This yields

$$\pi_{(m)}^X\alpha=\sum_{j=0}^{m-1}a_j\varphi_j \quad\text{where}\quad a_{(m)}:=\begin{pmatrix}a_0\\ \vdots\\ a_{m-1}\end{pmatrix}=\widehat{\Psi}_{m,1}^{-1}\begin{pmatrix}\vdots\\ \langle\alpha,\varphi_j\rangle_X\\ \vdots\end{pmatrix}. \tag{32}$$

The vector $\widehat{Z}_{m,1}$ can be written as

$$\widehat{Z}_{m,1}=\begin{pmatrix}\vdots\\ \langle\alpha,\varphi_j\rangle_X\\ \vdots\end{pmatrix}_{0\le j\le m-1}+\mathbb{W}_{m,1},\quad \mathbb{W}_{m,1}:=\begin{pmatrix}\vdots\\ \nu_{N,1}(\varphi_j)\\ \vdots\end{pmatrix}_{0\le j\le m-1}. \tag{33}$$

Note that

$$\mathbb{E}(\mathbb{W}_{m,1}\,{}^t\mathbb{W}_{m,1})=\frac{1}{N}\mathbb{E}\widehat{\Psi}_{m,1}=\frac{1}{N}(\Psi_{m,1}+\frac{1}{N}\mathbf{C}_m) \tag{34}$$

where

$$\mathbf{C}_m=\left(\int_0^T\varphi_j(t)\varphi_k(t)x^2(t)\lambda(t)dt\right)_{0\le j,k\le m-1}.$$

♦ **Proof of inequality (16).**
Now, we prove (16). For this, we write $\|\tilde{\alpha}_m-\alpha\|_X^2=T_1+T_2+T_3$, with

$$T_1:=\|\widehat{\alpha}_m-\alpha\|_X^2\mathbf{1}_{\Lambda_{m,1}\cap\mathcal{O}_{N,1}},\quad T_2:=\|\widehat{\alpha}_m-\alpha\|_X^2\mathbf{1}_{\Lambda_{m,1}\cap\mathcal{O}_{N,1}^c},\quad T_3:=\|\alpha\|_X^2\mathbf{1}_{\Lambda_{m,1}^c}. \tag{35}$$

We bound the expectation of the three terms above.

• The last term $T_3=\|\alpha\|_X^2\mathbf{1}_{\Lambda_{m,1}^c}$ satisfies:

$$\mathbb{E}T_3\le\mathbb{E}^{1/2}(\|\alpha\|_X^4)\mathbb{P}^{1/2}(\Lambda_{m,1}^c). \tag{36}$$

We have

$$\mathbb{E}(\|\alpha\|_X^4) \leq T\mathbb{E}\int_0^T \left[\alpha^2(t)X^2(t)\right]^2 dt \leq T\int_0^T \alpha^4(t)\mathbb{E}X^4(t)dt \leq c(T).$$

Thus, using Lemma 1,

$$\mathbb{E}T_3 \lesssim \frac{1}{N^{7/2}} \lesssim \frac{1}{N}. \tag{37}$$

• Study of $T_1 = \|\widehat{\alpha}_m - \alpha\|_X^2 \mathbf{1}_{\Lambda_{m,1}\cap\mathcal{O}_{N,1}}$. We can write:

$$\|\widehat{\alpha}_m - \alpha\|_X^2 = \|\widehat{\alpha}_m - \pi_{(m)}^X\alpha\|_X^2 + \|\pi_{(m)}^X\alpha - \alpha\|_X^2 = \|\widehat{\alpha}_m - \pi_{(m)}^X\alpha\|_X^2 + \inf_{h\in S_m}\|\alpha - h\|_X^2. \tag{38}$$

On one hand, we have $\widehat{\alpha}_m(t) = \sum_{j=0}^{m-1}[\widehat{\alpha}_{(m)}]_j\varphi_j(t)$ with $(\widehat{\alpha}_{(m)}) = \widehat{\Psi}_{m,1}^{-1}\widehat{Z}_{m,1}$. On the other hand, $\pi_{(m)}^X\alpha = \sum_{j=0}^{m-1}a_j\varphi_j$ where (see (32)) $a_{(m)} = \widehat{\Psi}_{m,1}^{-1}\left(\langle\varphi_j,\alpha\rangle_X\right)_{0\leq j\leq m-1}$.

Hence, by (33), $\widehat{\alpha}_{(m)} - a_{(m)} = \widehat{\Psi}_{m,1}^{-1}\mathbb{W}_{m,1}$ and using (31),

$$\|\widehat{\alpha}_m - \pi_{(m)}^X\alpha\|_X^2 = {}^t\mathbb{W}_{m,1}\widehat{\Psi}_{m,1}^{-1}\widehat{\Psi}_{m,1}\widehat{\Psi}_{m,1}^{-1}\mathbb{W}_{m,1} = {}^t\mathbb{W}_{m,1}\widehat{\Psi}_{m,1}^{-1}\mathbb{W}_{m,1}. \tag{39}$$

Now, $T_1 = (\|\widehat{\alpha}_m - \pi_{(m)}^X\alpha\|_X^2 + \inf_{h\in S_m}\|\alpha - h\|_X^2)\mathbf{1}_{\Lambda_{m,1}\cap\mathcal{O}_{N,1}}$ (see (38)).
On $\mathcal{O}_{N,1}$, all the eigenvalues of $\Psi_{m,1}^{-1/2}\widehat{\Psi}_{m,1}\Psi_{m,1}^{-1/2}$ belong to $[1/2, 3/2]$ and so all the eigenvalues of $\Psi_{m,1}^{1/2}\widehat{\Psi}_{m,1}^{-1}\Psi_{m,1}^{1/2}$ belong to $[2/3, 2]$. Thus, we write

$$\begin{aligned}{}^t\mathbb{W}_{m,1}\widehat{\Psi}_{m,1}^{-1}\mathbb{W}_{m,1}\ \mathbf{1}_{\mathcal{O}_{N,1}} &= {}^t\mathbb{W}_{m,1}\Psi_{m,1}^{-1/2}\Psi_{m,1}^{1/2}\widehat{\Psi}_{m,1}^{-1}\Psi_{m,1}^{1/2}\Psi_{m,1}^{-1/2}\mathbb{W}_{m,1}\ \mathbf{1}_{\mathcal{O}_{N,1}}\\ &\leq 2\ {}^t\mathbb{W}_{m,1}\Psi_{m,1}^{-1}\mathbb{W}_{m,1}\ \mathbf{1}_{\mathcal{O}_{N,1}}.\end{aligned} \tag{40}$$

Therefore

$$\begin{aligned}\mathbb{E}\left(\|\widehat{\alpha}_m - \pi_{(m)}^X\alpha\|_X^2\mathbf{1}_{\mathcal{O}_{N,1}\cap\Lambda_{m,1}}\right) &\leq 2\mathbb{E}\left(\sum_{0\leq j,k\leq m-1}[\mathbb{W}_{m,1}]_j[\mathbb{W}_{m,1}]_k[\Psi_{m,1}^{-1}]_{j,k}\right)\\ &= \frac{2}{N}\sum_{0\leq j,k\leq m-1}[\Psi_{m,1}^{-1}]_{j,k}([\Psi_{m,1}]_{j,k} + \frac{1}{N}[\mathbf{C}_m]_{j,k}) = \frac{2}{N}\mathrm{Tr}[\Psi_{m,1}^{-1}(\Psi_{m,1} + \frac{1}{N}\mathbf{C}_m)]\\ &= \frac{2m}{N} + \frac{2}{N^2}\mathrm{Tr}[\Psi_{m,1}^{-1}\mathbf{C}_m].\end{aligned} \tag{41}$$

by using equality (34).

Now, we bound $\mathrm{Tr}[\Psi_{m,1}^{-1}\mathbf{C}_m)]$. We have

$$\mathrm{Tr}[\Psi_{m,1}^{-1}\mathbf{C}_m)] = \int_0^T\sum_{0\leq j,k\leq m-1}\varphi_j(t)[\Psi_{m,1}^{-1}]_{j,k}\varphi_k(t)x^2(t)\lambda(t)dt = \int_0^T {}^t\varphi_{(m)}(t)\Psi_{m,1}^{-1}\varphi_{(m)}(t)x^2(t)\lambda(t)dt$$

where ${}^t\varphi_{(m)}(t) = (\varphi_0(t),\ldots,\varphi_{m-1}(t))$. As $\Psi_{m,1}^{-1}$ is symmetric positive definite, for all t, ${}^t\varphi_{(m)}(t)\Psi_{m,1}^{-1}\varphi_{(m)}(t) \geq 0$ and thus

$$0 \leq {}^t\varphi_{(m)}(t)\Psi_{m,1}^{-1}\varphi_{(m)}(t)x^2(t)\lambda(t) \leq \lambda_T^\star\ {}^t\varphi_{(m)}(t)\Psi_{m,1}^{-1}\varphi_{(m)}(t)x^2(t).$$

Consequently,

$$\mathrm{Tr}[\Psi_{m,1}^{-1}\mathbf{C}_m)] \leq \lambda_T^\star\mathrm{Tr}[\Psi_{m,1}^{-1}\Psi_{m,1}] = m\lambda_T^\star.$$

So we obtain:

$$\mathbb{E}(T_1) \leq \left(1 + \frac{\lambda_T^\star}{N}\right)\left(\inf_{h\in S_m}\|\alpha - h\|_x^2 + 2\frac{m}{N}\right), \tag{42}$$

using that

$$\mathbb{E}\left(\inf_{h\in S_m}\|\alpha-h\|_X^2\right)\le\left(1+\frac{\lambda_T^\star}{N}\right)\left(\inf_{h\in S_m}\|\alpha-h\|_x^2\right).$$

• Study of $T_2=\|\widehat{\alpha}_m-\alpha\|_X^2\mathbf{1}_{\Lambda_{m,1}\cap\mathcal{O}_{N,1}^c}$. We have $T_2\le(\|\widehat{\alpha}_m-\pi_{(m)}^X\alpha\|_X^2+\|\alpha\|_X^2)\mathbf{1}_{\Lambda_{m,1}\cap\mathcal{O}_{N,1}^c}$. Using (39) yields

$$T_2\le({}^t\mathbb{W}_{m,1}\widehat{\Psi}_{m,1}^{-1}\mathbb{W}_{m,1}+\|\alpha\|_X^2)\mathbf{1}_{\Lambda_{m,1}\cap\mathcal{O}_{N,1}^c}. \tag{43}$$

By the definition of $\Lambda_{m,1}$ and the Cauchy-Schwarz inequality, we get

$$\mathbb{E}T_2\le\left(\frac{\mathfrak{c}N}{L(S_m)\log(N)}\mathbb{E}^{1/2}(({}^t\mathbb{W}_{m,1}\mathbb{W}_{m,1})^2)+\mathbb{E}^{1/2}\|\alpha\|_X^4\right)\mathbb{P}^{1/2}(\mathcal{O}_{N,1}^c). \tag{44}$$

We have already seen that $\mathbb{E}(\|\alpha\|_X^4)\le c(T)$. For the term $\mathbb{E}[({}^t\mathbb{W}_{m,1}\mathbb{W}_{m,1})^2]$, we prove the following lemma:

Lemma 3. *Let the Assumptions of Proposition 7 hold. With $\mathbb{W}_{m,1}$ defined in (33), we have, for some constant $c(T)$, if the φ_js are bounded: $\mathbb{E}[({}^t\mathbb{W}_{m,1}\mathbb{W}_{m,1})^2]\le c(T)(mL(S_m))/N^2$. Otherwise, $\mathbb{E}[({}^t\mathbb{W}_{m,1}\mathbb{W}_{m,1})^2]\le c(T)(mL^2(S_m))/N^2$.*

Plugging the result of Lemma 3 in (44) allows to conclude for all m satisfying (13) that $\mathbb{E}(T_2)\le c/N$. Joining the bounds for the expectations of T_1,T_2,T_3 gives Inequality (16). □

♦ **Proof of Inequality (17).** We have now the following decomposition: $\|\tilde{\alpha}_m-\alpha\|_x^2=T_1'+T_2'+T_3'$ with

$$T_1':=\|\widehat{\alpha}_m-\alpha\|_x^2\mathbf{1}_{\Lambda_{m,1}\cap\mathcal{O}_{N,1}},\quad T_2':=\|\widehat{\alpha}_m-\alpha\|_x^2\mathbf{1}_{\Lambda_{m,1}\cap\mathcal{O}_{N,1}^c},\quad T_3':=\|\alpha\|_x^2\mathbf{1}_{\Lambda_{m,1}^c}. \tag{45}$$

We have $\mathbb{E}(T_3')\le\mathtt{a}\|\alpha\|_x^2/N^7$.
Next, $T_2'\le2(\|\widehat{\alpha}_m\|_x^2+\|\alpha\|_x^2)\mathbf{1}_{\Lambda_{m,1}\cap\mathcal{O}_{N,1}^c}$. We have

$$\|\widehat{\alpha}_m\|_x^2={}^t\widehat{\alpha}_{(m)}\Psi_{m,1}\widehat{\alpha}_{(m)}\le\|\Psi_{m,1}\|_{\mathrm{op}}\|\widehat{\alpha}_{(m)}\|_{2,m}^2\le\sup_{t\in[0,T]}x^2(t)\ \|\widehat{\alpha}_{(m)}\|_{2,m}^2.$$

Moreover, by formula (12),

$$\|\widehat{\alpha}_{(m)}\|_{2,m}^2={}^t\widehat{Z}_{m,1}\widehat{\Psi}_{m,1}^{-1}\widehat{\Psi}_{m,1}^{-1}\widehat{Z}_{m,1}\le\|\widehat{\Psi}_{m,1}^{-1}\|_{\mathrm{op}}^2\|\widehat{Z}_{m,1}\|_{2,m}^2.$$

Now using (14), on $\Lambda_{m,1}$,

$$\|\widehat{\alpha}_{(m)}\|_{2,m}^2\le\left(\frac{\mathfrak{c}N}{\log(N)L(S_m)}\right)^2\|\widehat{Z}_{m,1}\|_{2,m}^2.$$

By (33),

$$\|\widehat{Z}_{m,1}\|_{2,m}^2\le2\sum_{j=0}^{m-1}\langle\alpha,\varphi_j\rangle_X^2+2\,{}^t\mathbb{W}_{m,1}\mathbb{W}_{m,1}\le2\int_0^T\alpha^2(t)X^4(t)dt+2\,{}^t\mathbb{W}_{m,1}\mathbb{W}_{m,1}.$$

So $\mathbb{E}(\|\widehat{Z}_{m,1}\|_{2,m}^4)\le8T\int_0^T\alpha^4(t)\mathbb{E}[X^8(t)]dt+8\mathbb{E}[({}^t\mathbb{W}_{m,1}\mathbb{W}_{m,1})^2]$. By Lemma 1 and Lemma 3, we get that $\mathbb{E}(\|\widehat{Z}_{m,1}\|_{2,m}^4)\lesssim1$ as $m\vee L(S_m)\le N$. As a consequence, $\mathbb{E}[\|\widehat{\alpha}_{(m)}\|_{2,m}^4]\lesssim N^4$ and

$$\mathbb{E}[T_2']\lesssim N^2\mathbb{P}^{1/2}(\mathcal{O}_{N,1}^c)\lesssim N^{-3/2}.$$

For the term T_1', we simply have $\mathbb{E}(T_1')\le2\mathbb{E}(T_1)$ by using the defintion of $\mathcal{O}_{N,1}$. Joining the bounds on $\mathbb{E}(T_j')$, for $j=1,2,3$ gives the result. □

♦ **Proof of Inequality (18).** We propose a more precise study of T_1'. We have

$$\begin{aligned}\|\widehat{\alpha}_m-\alpha\|_x^2 &= \|\widehat{\alpha}_m-\alpha_m^x\|_x^2+\|\alpha_m^x-\alpha\|_x^2\\ &\le \|\alpha_m^x-\alpha\|_x^2+2(\|\widehat{\alpha}_m-\pi_{(m)}^X\alpha\|_x^2+\|\pi_{(m)}^X\alpha-\alpha_m^x\|_x^2)\end{aligned}$$

We get on $\Omega_{m,1}$,

$$\|\widehat{\alpha}_m-\alpha\|_x^2\le\|\alpha_m^x-\alpha\|_x^2+4\|\widehat{\alpha}_m-\pi_{(m)}^X\alpha\|_X^2+2\|\pi_{(m)}^X\alpha-\alpha_m^x\|_x^2.$$

Let $g=\alpha-\alpha_m^x$, then $\pi_{(m)}^Xg=\pi_{(m)}^X\alpha-\alpha_m^x$, and

(46) $$\mathbb{E}(T_1')\le\|\alpha_m^x-\alpha\|_x^2+4\mathbb{E}(\|\widehat{\alpha}_m-\pi_{(m)}^X\alpha\|_X^2\mathbf{1}_{\Lambda_{m,1}\cap\mathcal{O}_{N,1}})+4\mathbb{E}(\|\pi_{(m)}^Xg\|_x^2\mathbf{1}_{\Lambda_{m,1}\cap\Omega_{m,1}})$$

where $\|\alpha_m^x-\alpha\|_x^2=\inf_{h\in S_m}\|\alpha-h\|_x^2$ is the bias term and by (42),

(47) $$\mathbb{E}(\|\widehat{\alpha}_m-\pi_{(m)}^X\alpha\|_X^2\mathbf{1}_{\Lambda_{m,1}\cap\Omega_{m,1}})\le 2\left(1+\frac{\lambda_T^\star}{N}\right)\frac{m}{N}.$$

We have

Lemma 4. *Under the Assumptions of Proposition 7,*

$$\mathbb{E}(\|\pi_{(m)}^Xg\|_x^2\mathbf{1}_{\Lambda_{m,1}\cap\Omega_{m,1}})\le\frac{2\mathfrak{c}T}{\log(N)}\sup_{t\in[0,T]}x^2(t)\left(\frac{C_1(T)}{\mu_0^4}+\frac{(\lambda_T^\star)^2}{N}\right)\|\alpha-\alpha_m^x\|_x^2.$$

Applying Lemma 4 to (46) and using (47) yields

$$\mathbb{E}(T_1')\le\left(1+\frac{8\mathfrak{c}T}{\log(N)}\sup_{t\in[0,T]}x^2(t)\left(\frac{C_1(T)}{\mu_0^4}+\frac{(\lambda_T^\star)^2}{N}\right)\right)\|\alpha_m^x-\alpha\|_x^2+8\left(1+\frac{\lambda_T^\star}{N}\right)\frac{m}{N},$$

which gives the result in (18) by definition of $K_N(T)$. □

Proof of Lemma 4. Let $(\bar\varphi_j)_{0\le j\le m-1}$ an orthonormal basis of S_m w.r.t. $\langle.,.\rangle_x$. We can write $\bar\varphi_j=\sum_{k=0}^{m-1}a_{j,k}\varphi_k$, and we set $A_m=(a_{j,k})_{0\le j,k\le m-1}$. Let $\widehat{G}_{m,1}=(\langle\bar\varphi_j,\bar\varphi_k\rangle_X)_{0\le j,k\le m-1}$, obviously $\widehat{G}_{m,1}={}^tA_m\widehat{\Psi}_{m,1}A_{m,1}$. Then as $\mathrm{Id}_n=(\langle\bar\varphi_j,\bar\varphi_k\rangle)_{0\le j,k\le m-1}={}^tA_m\Psi_{m,1}A_{m,1}$, we know that A_m is a square root of $\Psi_{m,1}^{-1}$. As a consequence,

$$\Omega_{m,1}=\{\|\Psi_{m,1}^{-1/2}\widehat{\Psi}_{m,1}\Psi_{m,1}^{-1/2}\|_{\mathrm{op}}\le\frac12\}=\{\|\widehat{G}_{m,1}-\mathrm{Id}_m\|_{\mathrm{op}}\le1/2\}.$$

Next, we write $\pi_{(m)}^Xg=\sum_{k=0}^{m-1}\beta_k\bar\varphi_k$, with $\langle g-\pi_{(m)}^Xg,\bar\varphi_j\rangle_X=0$ for $j=0,\dots,m-1$. Then on $\Omega_{m,1}$, we have $\|\widehat{G}_{m,1}^{-1}\|_{\mathrm{op}}\le2$. Then $\widehat{G}_{m,1}\beta_{(m)}=(\langle\bar\varphi_j,g\rangle_X)_{0\le j\le m-1}$, where $\beta_{(m)}={}^t(\beta_0,\dots,\beta_{m-1})$. Thus on $\Omega_{m,1}$, we have

$$\|\pi_{(m)}^Xg\|_x^2=\sum_{k=0}^{m-1}\beta_k^2=\|\widehat{G}_{m,1}^{-1}(\langle\bar\varphi_j,g\rangle_X)_{0\le j\le m-1}\|_{2,m}^2\le4\sum_{j=0}^{m-1}\langle\bar\varphi_j,g\rangle_X^2.$$

Note that, as $\int_0^T\bar\varphi_j(t)g(t)x^2(t)dt=\langle g,\bar\varphi_j\rangle_x=0$,

$$\mathbb{E}(\langle\bar\varphi_j,g\rangle_X)=\int_0^T\bar\varphi_j(t)g(t)\mathbb{E}(X^2(t))dt=\frac1N\int_0^T\bar\varphi_j(t)g(t)x^2(t)\lambda(t)dt.$$

We have

$$\sum_{j=0}^{m-1}\bar\varphi_j^2(t)={}^t\varphi_{(m)}(t)\,{}^tA_mA_m\varphi_{(m)}(t)={}^t\varphi_{(m)}(t)\Psi_{m,1}^{-1}\varphi_{(m)}(t)\le L(S_m)\|\Psi_{m,1}^{-1}\|_{\mathrm{op}},$$

where $\varphi_{(m)}(t) = {}^t(\varphi_0(t), \dots, \varphi_{m-1}(t))$. Therefore, recalling that $g = \alpha - \alpha_m^x$, we get

$$\sum_{j=0}^{m-1} [\mathbb{E}(\langle\bar{\varphi}_j, g\rangle_X)]^2 \leq \frac{T}{N^2} L(m) \|\Psi_{m,1}^{-1}\|_{\rm op} \int_0^T x^4(t)\lambda^2(t)(\alpha(t) - \alpha_m^x(t))^2 dt$$

Thus

$$\sum_{j=0}^{m-1} [\mathbb{E}(\langle\bar{\varphi}_j, g\rangle_X)]^2 \leq \frac{\mathfrak{c}T}{2N\log(N)} \sup_{t\in[0,T]} x^2(t)\lambda^2(t)\|\alpha - \alpha_m^x\|_x^2.$$

Now

$$\mathbb{E}(\|\pi_{(m)}^X g\|_x^2 \mathbf{1}_{\Lambda_{m,1}\cap\Omega_{m,1}}) \leq 4\sum_{j=0}^{m-1} {\rm Var}(\langle\bar{\varphi}_j, g\rangle_X) + \frac{2\mathfrak{c}T}{N\log(N)} \sup_{t\in[0,T]} x^2(t)\lambda^2(t)\|\alpha - \alpha_m^x\|_x^2.$$

We have

$$\sum_{j=0}^{m-1} {\rm Var}(\langle\bar{\varphi}_j, g\rangle_X) = \sum_{j=0}^{m-1} \mathbb{E}\left[\left(\int_0^T g(t)\bar{\varphi}_j(t)(X^2(t) - \mathbb{E}(X^2(t)))dt\right)^2\right]$$

$$\leq T\sum_{j=0}^{m-1}\int_0^T g^2(t)\bar{\varphi}_j^2(t){\rm Var}(X^2(t))dt \leq TL(S_m)\|\Psi_{m,1}^{-1}\|_{\rm op}\int_0^T g^2(t){\rm Var}(X^2(t))dt$$

$$\leq \frac{\mathfrak{c}T}{2\log(N)} \frac{\sup_{t\in[0,T]} x^2(t)C_1(T)}{\mu_0^4}\|\alpha - \alpha_m^x\|_x^2$$

as, after some elementary computations ${\rm Var}(X^2(t)) \leq (x^4(t)/\mu_0^4)(C_1(T)/N)$, with

$$C_1(T) = C + 3\left(\int_0^T e^{-\int_0^s \alpha(u)du} ds\right)^2 + 4(\mu_0^2 + \sigma_0^2)\int_0^T e^{-\int_0^s \alpha(u)du} ds,$$

$C = 2\mathbb{E}(Y_1^4) + 4|\mu_0|\mathbb{E}(|Y_1|^3) + 4\mu_0^2\sigma_0^2$ and $Y_1 = X_1(0) - \mu_0$. Therefore we get

$$\mathbb{E}(\|\pi_{(m)}^X g\|_x^2 \mathbf{1}_{\Lambda_{m,1}\cap\Omega_{m,1}}) \leq \frac{2\mathfrak{c}T}{\log(N)} \sup_{t\in[0,T]} x^2(t)\left(\frac{C_1(T)}{\mu_0^4} + \frac{(\lambda_T^\star)^2}{N}\right)\|\alpha - \alpha_m^x\|_x^2.$$

This ends the proof of Lemma 4. □

8.8. **Proof of Lemmas.**

Proof of Lemma 2. On the one hand, $\mathbb{P}(\mathcal{O}_{N,1}^c) \lesssim N^{-7}$ by Proposition 3 with $p = 7$, and on the other hand $\mathcal{O}_{N,1} \subset \Omega_{m,1}$. Therefore, $\mathbb{P}(\Omega_{m,1}^c) \lesssim 1/N^7$.

By the same proof as the one of Proposition 4 (ii) in Comte and Genon-Catalot (2020a), we have that:

$$\{\|\widehat{\Psi}_{m,1}^{-1} - \Psi_{m,1}^{-1}\|_{\rm op} > \alpha\|\Psi_{m,1}^{-1}\|_{\rm op}\} \subset \{\|\Psi_{m,1}^{-1/2}\widehat{\Psi}_{m,1}\Psi_{m,1}^{-1/2} - Id_m\|_{\rm op} > \frac{\inf\{\alpha, 1\}}{2}\}.$$

Then, we mimick Lemma 5 of the same paper to get that, for m satisfying (13),

$$\mathbb{P}(\Lambda_{m,1}^c) \leq \mathbb{P}(\{\|\widehat{\Psi}_{m,1}^{-1} - \Psi_{m,1}^{-1}\|_{\rm op} > \|\Psi_m^{-1}\|_{\rm op}\})$$

$$\leq \mathbb{P}(\{\|\Psi_{m,1}^{-1/2}\widehat{\Psi}_{m,1}\Psi_{m,1}^{-1/2} - Id_m\|_{\rm op} > 1/2\}) = \mathbb{P}(\Omega_{m,1}^c). \quad (48) \quad □$$

Proof of Lemma 3. We have

$$\mathbb{E}[{}^t(\mathbb{W}_{m,1})\mathbb{W}_{m,1}]^2 = \frac{1}{N^2}\mathbb{E}[\sum_{j=0}^{m-1}(\int_0^T \varphi_j(t)X(t)dB_{N,1}(t))^2]^2 \leq \frac{m}{N^2}\sum_{j=0}^{m-1}\mathbb{E}(\int_0^T \varphi_j(t)X(t)dB_{N,1}(t))^4$$

Therefore, using the Burholder-Davies-Gundy inequality yields

$$\mathbb{E}[\,{}^t(\mathbb{W}_{m,1})\mathbb{W}_{m,1}]^2 \lesssim \frac{m}{N^2}\sum_{j=0}^{m-1}\mathbb{E}\left[\int_0^T \varphi_j^2(t)X^2(t)dt\right]^2 \le \frac{Tm}{N^2}\int_0^T\sum_{j=0}^{m-1}\varphi_j^4(t)\mathbb{E}(X^4(t))dt$$

We use the fact that the φ_js are bounded and Lemma 1 to obtain

$$\mathbb{E}[\,{}^t(\mathbb{W}_{m,1})'\mathbb{W}_{m,1}]^2 \lesssim \frac{m}{N^2}L(S_m)$$

Otherwise, we obtain $\mathbb{E}[\,{}^t(\mathbb{W}_{m,1})\mathbb{W}_{m,1}]^2 \lesssim (m/N^2)L^2(S_m)$. □

8.9. **Proof of Proposition 8.** We defined $\Lambda_{p,2}$ in equation (15) and let us set

$$\Omega_{p,2} := \left\{\left|\frac{\|h\|^2_{\sqrt{V}}}{\|h\|^2_{\sqrt{v}}} - 1\right| \le \frac{1}{2}, \forall h\in\Sigma_p\right\}. \tag{49}$$

On $\mathcal{O}_{N,2}$ defined by (9), the empirical norm $\|.\|_{\sqrt{V}}$ and the $\|\cdot\|_{\sqrt{v}}$-norm are equivalent for elements of $\mathbb{L}^2([0,T])$. Moreover

$$\Omega_{p,2} = \left\{\|\Psi_{p,2}^{-1/2}\widehat{\Psi}_{p,2}\Psi_{p,2}^{-1/2} - \mathrm{Id}_p\|_{\mathrm{op}} \le 1/2\right\}. \tag{50}$$

The following lemma holds

Lemma 5. *We have, for* $\mathtt{b}$ *a positive constant,* $\mathbb{P}(\mathcal{O}^c_{N,2}) \le \mathtt{b}/N^7$. *Under the assumptions of Proposition 8, for p satisfying (13), we have,* $\mathbb{P}(\Lambda^c_{p,2}) \le \mathtt{b}/N^7$.

To prove Inequality (19), we proceed as in the proof of Inequality (16), using that $\mathbb{E}(V(t)) = (1-1/N)v(t)$, which makes things easier. The two inequalities also follow and we use that $\mathrm{Var}(V(t)) \le (v^2(t)/\sigma_0^4)(C_2(T)/N)$ with $C_2(T) = \mathbb{E}([X_1(0)-\mu_0]^4) + 4\int_0^T e^{-4\Gamma(s)}v(s)ds$. □

8.10. **Proof of Proposition 9.** Let $\mathbf{u} = {}^t(u_0, u_1, \dots, u_{m-1})$ a vector of $\mathbb{R}^m$ such that $\|\mathbf{u}\|^2_{2,m} = \sum_{j=0}^{m-1}u_j^2 = 1$ and set $h_1(t) = \sum_{j=0}^{m-1}u_j\varphi_j(t)$. We have $\int_0^T h_1^2(t)dt = 1$ and when $\alpha(t)\ge 0$,

$${}^t\mathbf{u}\Psi_{m,1}\mathbf{u} = \int_0^T h_1^2(t)x^2(t)dt \ge \mu_0^2.$$

Analogously, when $\gamma(t)\ge 0$, $v(t)\ge\sigma_0^2$. Then for $\mathbf{u} = {}^t(u_0, u_1, \dots, u_{p-1})$ a vector of $\mathbb{R}^p$ such that $\|\mathbf{u}\|^2_{2,p} = 1$, ${}^t\mathbf{u}\Psi_{p,2}\mathbf{u} \ge \sigma_0^2$. □

8.11. **Proof of Theorem 1.** The proof is given for the estimation of α only, the γ case being very similar.

Lemma 6. *Under the assumptions of Theorem 1, for all* $m, m' \le N$,

$$\mathbb{E}\left[\left(\sup_{h_1\in B_{m\vee m'}}\nu^2_{N,1}(h_1) - (p_1(m)+p_1(m'))\right)_+ \mathbf{1}_{\mathcal{O}_{N,1}}\right] \le 1.6\kappa\frac{e^{-m'}}{N},$$

where $\mathcal{O}_{N,1}$ *is defined by (8),* $B_{m\vee m'} = \{h_1\in S_m + S_{m'}, \|h_1\|_x\le 1\}$ *and* $p_1(m) = \kappa m/N$, *where* κ *is a numerical constant.*

Proof of Lemma 6. By the exponential inequality for martingales (see the Bernstein Inequality for martingales in Revuz and Yor (1999)), we have

$$\mathbb{P}\left(\int_0^T h_1(t)X(t)dB_{N,1}(t)\ge x, \quad \int_0^T h_1^2(t)X^2(t)dt\le y\right) \le \exp\left(-\frac{x^2}{2y}\right).$$

In other words, we have

$$\mathbb{P}\left(\nu_{N,1}(h_1) \geq \varepsilon, \quad \|h_1\|_X^2 \leq \eta^2\right) \leq e^{-\frac{N\varepsilon^2}{2\eta^2}}.$$

Therefore

$$\mathbb{P}\left(\left(\nu_{N,1}(h_1) \geq \sqrt{3}\|h\|_x\sqrt{\varepsilon}\right)\bigcap \mathcal{O}_{N,1}\right) \leq \mathbb{P}\left(\nu_{N,1}(h_1) \geq \sqrt{3}\|h\|_x\sqrt{\varepsilon}, \|h_1\|_X^2 \leq \frac{3}{2}\|h_1\|_x^2\right) \leq e^{-N\varepsilon}.$$

The result follows by aplying the chaining method as in Baraud *et al.* (2001), sections 6-7, Proposition 6.1 with $s^2 = 1$. □

By definition of $\widehat{m}$, we have $\forall m \in \widehat{\mathcal{M}}_{N,1}$, and any $\alpha_m \in S_m$,

$$U_{N,1}(\widehat{\alpha}_{\widehat{m}}) + \text{pen}_1(\widehat{m}) \leq U_{N,1}(\alpha_m) + \text{pen}(m).$$

Moreover, by (11), it holds

$$U_{N,1}(\widehat{\alpha}_{\widehat{m}}) - U_{N,1}(\alpha_m) = \|\widehat{\alpha}_{\widehat{m}} - \alpha\|_X^2 - \|\alpha_m - \alpha\|_X^2 - 2\nu_{N,1}(\widehat{\alpha}_{\widehat{m}} - \alpha_m).$$

Consequently, $\forall m \in \widehat{\mathcal{M}}_{N,1}$, and any $\alpha_m \in S_m$,

$$\begin{aligned} \|\widehat{\alpha}_{\widehat{m}} - \alpha\|_X^2 &\leq \|\alpha_m - \alpha\|_X^2 + \text{pen}_1(m) + 2\nu_{N,1}(\widehat{\alpha}_{\widehat{m}} - \alpha_m) - \text{pen}_1(\widehat{m}) \\ &\leq \|\alpha_m - \alpha\|_X^2 + \text{pen}_1(m) \\ &\quad + \frac{1}{8}\|\widehat{\alpha}_{\widehat{m}} - \alpha_m\|_x^2 + 8 \sup_{h_1 \in S_m + S_{\widehat{m}}, \|h_1\|_x = 1} \nu_{N,1}^2(h_1) - \text{pen}_1(\widehat{m}) \end{aligned} \tag{51}$$

Define

$$\Xi_{N,1} = \{\mathcal{M}_{N,1} \subset \widehat{\mathcal{M}}_{N,1}\}.$$

On $\Xi_{N,1}$, Inequality (51) holds $\forall m \in \mathcal{M}_{N,1}$, and on $\Xi_{N,1} \cap \mathcal{O}_{N,1}$, $\forall m \in \mathcal{M}_{N,1}$,

$$\|\widehat{\alpha}_{\widehat{m}} - \alpha\|_X^2 \leq \|\alpha_m - \alpha\|_X^2 + \text{pen}_1(m) + \frac{1}{4}\|\widehat{\alpha}_{\widehat{m}} - \alpha_m\|_X^2 + 8 \sup_{h_1 \in S_m + S_{\widehat{m}}, \|h_1\|_x = 1} \nu_{N,1}^2(h_1) - \text{pen}_1(\widehat{m})$$

Thus, on $\Xi_{N,1} \cap \mathcal{O}_{N,1}$, $\forall m \in \mathcal{M}_{N,1}$,

$$\begin{aligned} \frac{1}{2}\|\widehat{\alpha}_{\widehat{m}} - \alpha\|_X^2 &\leq \frac{3}{2}\|\alpha_m - \alpha\|_X^2 + \text{pen}_1(m) + 8\left(\sup_{h_1 \in S_m + S_{\widehat{m}}, \|h_1\|_x = 1} \nu_{N,1}^2(h_1) - (p_1(m) + p_1(\widehat{m}))\right) \\ &\quad + 8p_1(m) + 8p_1(\widehat{m}) - \text{pen}_1(\widehat{m}). \end{aligned}$$

By Lemma 6, we get

$$\sum_{m' \leq N} \mathbb{E}\left[\left(\sup_{h_1 \in B_{m \vee m'}} \nu_{N,1}^2(h_1) - (p_1(m) + p_1(m')\right)_+ \mathbf{1}_{\mathcal{O}_{N,1}}\right] \leq 1.6\kappa \frac{\sum_{m'} e^{-m'}}{N} = \frac{C}{N}.$$

For $\kappa_1 \geq 8\kappa$, $8p_1(m) + 8p_1(\widehat{m}) \leq \text{pen}_1(m) + \text{pen}_1(\widehat{m})$. Thus, finally, $\forall m \in \mathcal{M}_{N,1}$, and $\alpha_m \in S_m$, we get for $\kappa_1 \geq 8\kappa$,

$$\mathbb{E}\left(\|\widehat{\alpha}_{\widehat{m}} - \alpha\|_X^2 \mathbf{1}_{\Xi_{N,1} \cap \mathcal{O}_{N,1}}\right) \leq 3\left(1 + \frac{\lambda_T^\star}{N}\right)\|\alpha_m - \alpha\|_x^2 + 4\text{pen}_1(m) + \frac{C}{N}.$$

Now we study $\|\widehat{\alpha}_{\widehat{m}} - \alpha\|_X^2 \mathbf{1}_{(\Xi_{N,1} \cap \mathcal{O}_{N,1})^c}$. We have $T_2'' := \|\widehat{\alpha}_{\widehat{m}} - \alpha\|_X^2 \leq \|\widehat{\alpha}_{\widehat{m}} - \pi_{(\widehat{m})}^X \alpha\|_X^2 + \|\alpha\|_X^2$. Using (39) yields, as $\widehat{m} \in \widehat{\mathcal{M}}_{N,1}$,

$$T_2'' \leq \frac{\mathfrak{c}N}{L(S_{\widehat{m}})\log(N)}\, {}^t\mathbb{W}_{\widehat{m},1}\mathbb{W}_{\widehat{m},1} + \|\alpha\|_X^2.$$

Now, $m \mapsto {}^t\mathbb{W}_{m,1}\mathbb{W}_{m,1}$ is increasing with m, and $L(S_{\widehat{m}}) \geq L(S_1) \geq 1/T$. Therefore

$$T_2'' \leq \frac{\mathfrak{c}TN}{\log(N)} \, {}^t\mathbb{W}_{N,1}\mathbb{W}_{N,1} + \|\alpha\|_X^2.$$

By the Cauchy-Schwarz inequality, we get

$$\mathbb{E}(T_2''\mathbf{1}_{(\Xi_{N,1}\cap\mathcal{O}_{N,1})^c}) \leq \left(\frac{\mathfrak{c}TN}{\log(N)}\mathbb{E}^{1/2}(({}^t\mathbb{W}_{N,1}\mathbb{W}_{N,1})^2) + \mathbb{E}^{1/2}\|\alpha\|_X^4\right)\mathbb{P}^{1/2}(\Xi_{N,1}\cap\mathcal{O}_{N,1})^c).$$

We have seen above that $\mathbb{E}(\|\alpha\|_X^4) \leq c(T)$. For the term $\mathbb{E}[({}^t\mathbb{W}_{N,1}\mathbb{W}_{N,1})^2]$, we apply Lemma 3 to obtain

$$\mathbb{E}[({}^t\mathbb{W}_{N,1}\mathbb{W}_{N,1})^2] \leq c(T)\frac{L^2(S_N)}{N} \lesssim N^3.$$

Thus

$$\mathbb{E}(T_2''\mathbf{1}_{(\Xi_{N,1}\cap\mathcal{O}_{N,1})^c}) \lesssim N^{5/2}\mathbb{P}^{1/2}(\Xi_{N,1}\cap\mathcal{O}_{N,1})^c).$$

By Proposition 3, taking $p = 8$, we have $\mathbb{P}(\mathcal{O}_{N,1}^c) \lesssim N^{-8}$. Now we use the Lemma:

Lemma 7. *It holds that* $\mathbb{P}(\Xi_{N,1}^c) \lesssim N^{-7}$.

This implies that $\mathbb{E}(T_2''\mathbf{1}_{(\Xi_{N,1}\cap\mathcal{O}_{N,1})^c}) \lesssim N^{-1}$. We obtain the first inequality of Theorem 1 for $\kappa_1 \geq 8\kappa := \kappa_0$ and $C_1 = 4$.
We proceed analogously for the second inequality of Theorem 1. □

Proof of Lemma 7. On $\Xi_{N,1}^c$, there exists $k \in \mathcal{M}_{N,1}$ such that $k \notin \widehat{\mathcal{M}}_{N,1}$.

For this index k, we have $L(S_k)\|\Psi_{k,1}^{-1}\|_{\rm op} \leq \mathfrak{c}N/2\log(N)$ and $L(S_k)\|\widehat{\Psi}_{k,1}^{-1}\|_{\rm op} > \mathfrak{c}N/\log(N)$. As

$$\begin{aligned}\mathfrak{c}(N/\log(N)) &< L(S_k)\|\widehat{\Psi}_{k,1}^{-1}\|_{\rm op} \leq L(S_k)\|\Psi_{k,1}^{-1} - \widehat{\Psi}_{k,1}^{-1}\|_{\rm op} + L(S_k)\|\Psi_k^{-1}\|_{\rm op} \\ &\leq L(S_k)\|\Psi_{k,1}^{-1} - \widehat{\Psi}_{k,1}^{-1}\|_{\rm op} + (\mathfrak{c}/2)(N/\log(N)),\end{aligned}$$

we get for this index k that $L(S_k)\|\widehat{\Psi}_{k,1}^{-1} - \Psi_{k,1}^{-1}\|_{\rm op} \geq \mathfrak{c}N/(2\log(N))$.
Let $\Delta_m = \{L(S_m)\|\widehat{\Psi}_{m,1}^{-1} - \Psi_{m,1}^{-1}\|_{\rm op} > (\mathfrak{c}/2)N/\log(N)\}$, we have, using the definition of $\mathcal{M}_{N,1}$,

$$\mathbb{P}(\mathcal{M}_{N,1} \not\subseteq \widehat{\mathcal{M}}_{N,1}) \leq \sum_{m\in\mathcal{M}_{N,1}} \mathbb{P}(\Delta_m) \leq \sum_{m\in\mathcal{M}_{N,1}} \mathbb{P}(\|\widehat{\Psi}_{m,1}^{-1} - \Psi_{m,1}^{-1}\|_{\rm op} > \|\Psi_{m,1}^{-1}\|_{\rm op}).$$

By formula (48),

$$\mathbb{P}(\|\widehat{\Psi}_{m,1}^{-1} - \Psi_{m,1}^{-1}\|_{\rm op} > \|\Psi_{m,1}^{-1}\|_{\rm op}) \leq \mathbb{P}(\Omega_{m,1}^c) \leq \mathbb{P}(\mathcal{O}_{N,1}^c) \lesssim N^{-8}.$$

This implies $\mathbb{P}(\mathcal{M}_{N,1} \not\subseteq \widehat{\mathcal{M}}_{N,1}) \lesssim N^{-7}$. □

References

[1] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York, ninth dover printing, tenth gpo printing edition.

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

[3] Baladron, J., Fasoli, D., Faugeras, O. and Touboul, J. (2012). Mean field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. *The Journal of Mathematical Neuroscience*, **2(1):10**, 1-50.

[4] Belomestny, D., Pilipauskaitė, V. and Podolskij, M. (2021). Semiparametric estimation of McKean-Vlasov SDEs. *Preprint* arXiv:2107.00539.

[5] Benedetto, D., Caglioti, E. and Pulverenti, M. (1997). A kinetic equation for granular media. *Mathematical Modelling and Numerical Analysis*, **31**, 615-641.

[6] Britton,T. and Pardoux, E., Editors (2020). *Stochastic Epidemic Models with Inference.* Lecture Notes in Mathematics, **2255**, Mathematical Biosciences Subseries, Springer.
[7] Bishwal, J. P. N. (2011). Estimation in interacting diffusions: continuous and discrete sampling. *Appl. Math. (Irvine)* **2**, 1154-1158.
[8] Chen, X. (2021). Maximum likelihood estimation of potential energy in interacting particle systems from single-trajectory data. *Electron. Commun. Probab.* **26**, 1-13.
[9] Cohen, A., Davenport, M.A. and Leviatan, D. (2013). On the stability and acuracy of least squares approximations. *Foundations of Computational Mathematics* **13**, 819-834.
[10] Comte, F. and Genon-Catalot, V. (2018). Laguerre and Hermite bases for inverse problems. *Journal of the Korean Statistical Society*, **47**, 273-296.
[11] 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.
[12] 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.
[13] Della Maestra, L. and Hoffmann, M. (2022). Nonparametric estimation for interacting particle systems: McKean-Vlasov models. *Probab. Theory Related Fields* **182**, 551-613.
[14] Efromovich, S. (1999). *Nonparametric curve estimation. Methods, theory, and applications.* Springer Series in Statistics. Springer-Verlag, New York.
[15] Giesecke, K., Schwenkler, G. and Sirignano, J.A. (2020). Inference for large financial systems. *Math. Finance* **30**, 3-46.
[16] Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application.* Probability and Mathematical Statistics. Academic Press, New York-London.
[17] Kasonga, R. A. (1990). Maximum likelihood theory for large interacting systems. *SIAM J. Appl. Math* **50**, 865-875.
[18] Li, Z., Lu, F., Maggioni, M., Tang, S. and Zhang, C. (2021). On the identifiability of interaction functions in systems of interacting particles. *Stochastic Process. Appl.* **132**, 1351163.
[19] Lu, F., Zhong, M., Tang, S. and Maggioni, M. (2019). Nonparametric inference of interaction laws in systems of agents from trajectory data. *Proc. Natl. Acad. Sci.* USA **116**, 14424-14433.
[20] Méléard, S. (1996). Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic Models for Nonlinear Partial Differential Equations*, Lecture Notes in Mathematics, **1627**, 42-95, Springer.
[21] Molginer, A. and Edelstein-Keshet, L. (1999). A non-local model for a swarm. *Journal of Mathematical Biology*, **38**, 534-570.
[22] Nguyen, H. T. and Pham T. D. (1982) Identification of Nonstationary Diffusion Model by the Method of Sieves. *SIAM Journal of Control and Optimization*, **20**, 603-611.
[23] Pavliotis, G.A. and Zanoni, A. (2021). Eigenfunction martingale estimators for interacting particle systems and their mean field limit. *Preprint arXiv*: 2112.04870v1.
[24] Sharrock, L., Kantas, N., Parpas, P. and Pavliotis, G.A. (2021). Parameter Estimation for the McKean-Vlasov Stochastic Differential Equation. *Preprint* arXiv:2106.13751.
[25] Sznitman, A.S. (1991). Topics in propagation of chaos. *Ecole d'été de probabilités de Saint-Flour XIX-1989.* Lecture Notes in Mathematics, **1464**, 165-251, Springer.
[26] 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.
[27] 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.