NONPARAMETRIC ESTIMATION OF THE TRANSITION DENSITY FUNCTION FOR DIFFUSION PROCESSES

FABIENNE COMTE[†] AND NICOLAS MARIE[◇]

[†]Université Paris Cité, CNRS, MAP5, F-75006 Paris, France.
[◇]Université Paris Nanterre, CNRS, Modal'X, 92001 Nanterre, France.

Abstract. We assume that we observe N independent copies of a diffusion process on a time interval $[0, 2T]$. For a given time t, we estimate the transition density $p_t(x, y)$, namely the conditional density of X_{t+s} given $X_s = x$, under conditions on the diffusion coefficients ensuring that this quantity exists. We use a least squares projection method on a product of finite dimensional spaces, prove risk bounds for the estimator and propose an anisotropic model selection method, relying on several reference norms. A simulation study illustrates the theoretical part for Ornstein-Uhlenbeck or square-root (Cox-Ingersoll-Ross) processes.

1. Introduction

Consider the stochastic differential equation

$$X_t = x_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s \ ; \ t \in [0, 2T], \tag{1}$$

where $x_0 \in \mathbb{R}$, $W = (W_t)_{t\in[0,2T]}$ is a Brownian motion, $b, \sigma \in C^1(\mathbb{R})$, and b' and σ' are bounded. Under these conditions on b and σ, Equation (1) has a unique (strong) solution $X = (X_t)_{t\in[0,2T]}$. Under additional conditions, the transition density $p_t(x, .)$ is well defined and can be understood as the conditional density of X_{s+t} given $X_s = x$. The question of estimating such a function, for discrete samples of one path of X, say $X_{k\Delta}$, $k = 1, \dots, n$ considered as a Markov chain, has been studied by several authors. The reader may refer to Lacour [20] and [21], Sart [31], or recently in higher dimension in Löffler and Picard [24]. Nonparametric strategies, based on contrast minimization, ratio of estimators, or singular value decomposition are considered.

Since few years, statistical inference from copies of diffusion processes, especially estimators of the drift function b, has been deeply investigated. This approach of statistical inference in stochastic differential equations (SDE) is part of functional data analysis, which is devoted to samples of infinite dimensional data (see Ramsay and Silverman [30] and Wang *et al.* [32]). From independent copies of X, projection least squares estimators have been studied in Comte and Genon-Catalot [11] for continuous time observations, in Denis *et al.* [16] for discrete time (with small step) observations with a classification purpose in the parametric setting, and in Denis *et al.* [15] in the nonparametric context, for instance. Marie and Rosier [27] propose a kernel based Nadaraya-Watson estimator of the drift function b, with bandwidth selection relying on the Penalized Comparison to Overfitting criterion recently introduced in Lacour *et al.* [22]. Still for independent copies of X, Halconruy and Marie [18] investigates the properties of the projection least squares estimator of b when W is replaced by a Lévy process, and Comte and Genon-Catalot [12] and Marie [25] deal with estimators of the drift function in non-autonomous SDE. More recently, copies-based estimation with dependency has been investigated, see e.g. Della Maestra and Hoffmann [14] or Belomestny *et al.* [6], dealing with nonparametric estimators in interacting particle systems and McKean-Vlasov models.

In the present work, we investigate the question of estimating the transition density $p_t(x, y)$ for any fixed time $t \in (0, T]$, as a function of two variables, from the observation on $[0, 2T]$ of N independent copies of the solution X of Equation (1). Precisely, consider $X^i := \mathcal{I}(x_0, W^i)$ for every $i \in \{1, \dots, N\}$, where $\mathcal{I}(\cdot)$ is the Itô map for Equation (1) and $W^1, \dots, W^N$ are $N \in \mathbb{N}^*$ independent copies of W. What we propose is a projection least squares estimator $\widehat{p}_{\mathbf{m},t}$ of the transition density function p_t of X at time t computed from $X^1, \dots, X^N$, where $\mathbf{m} = (m_1, m_2) \in \{1, \dots, N_T\}^2$ and $N_T := [NT] + 1$. Consider $\mathcal{S}_{\mathbf{m}} := \mathcal{S}_{\varphi,m_1} \otimes \mathcal{S}_{\psi,m_2}$, where $\mathcal{S}_{\varphi,m_1}$ (resp. $\mathcal{S}_{\psi,m_2}$) is a m_1-dimensional (resp. m_2-dimensional) function space defined later. The estimator $\widehat{p}_{\mathbf{m},t}$ is defined as minimizing over functions $\tau \in \mathcal{S}_{\mathbf{m}}$ the objective function

$$\tau \longmapsto \gamma_N(\tau) := \frac{1}{NT} \sum_{i=1}^{N} \left(\int_0^T \int_{-\infty}^{\infty} \tau(X_s^i, y)^2 dy ds - 2 \int_0^T \tau(X_s^i, X_{s+t}^i) ds \right).$$

First, a nonadaptive risk bound is established on $\widehat{p}_{\mathbf{m},t}$, and its rate of convergence is provided when p_t belongs to an anisotropic 2-dimensional Sobolev-Hermite space. Of course, other regularity spaces may be considered, but since the Hermite basis is $\mathbb{R}$-supported, this example is very instructive and allows to apply in the diffusion processes framework some results already established in Comte and Lacour [13] for the conditional density estimation in a different but analogous framework. Then, risk bounds are established on the adaptive estimator $\widehat{p}_{\widehat{\mathbf{m}},t}$, where the couple of dimensions $\widehat{\mathbf{m}} = (\widehat{m}_1, \widehat{m}_2)$ is selected in a random subset of $\{1, \dots, N_T\}^2$ thanks to a Birgé-Massart type criterion, which may be simplified in two detailed special cases.

To start with, we propose in a short preliminary Section 2 two application settings where our estimator of the transition may be used. The nonparametric estimator of the transition density function is precisely defined in Section 3 and nonadaptive risk bounds are established in Section 4. Adaptive estimation methods are studied in Section 5 and the whole procedure is illustrated through simulations in Section 6. Finally, Section 7 provides concluding remarks and proofs are gathered in Appendix A.

2. Two motivating examples

In this section, we briefly present two possible applications of our estimation method of p_t.

- Assume that $\sigma(\cdot)^2 > 0$, and consider the parabolic partial differential equation defined by

$$\frac{\partial u}{\partial t}(t,x) + \frac{1}{2}\sigma(x)^2 \frac{\partial^2 u}{\partial x^2}(t,x) + b(x)\frac{\partial u}{\partial x}(t,x) = 0, \quad u(T,x) = \varphi(x), \tag{2}$$

  where $\varphi : \mathbb{R} \to \mathbb{R}$ is a known twice continuously differentiable function. Moreover, let $X^{t,x}$ be the solution of Equation (1) starting from $x \in \mathbb{R}$ at time $t \in [0,T)$. By Lamberton and Lapeyre [23], Theorem 5.1.7, the solution of Equation (2) is given by

$$F(t,x) := \mathbb{E}(\varphi(X_T^{t,x})) = \int_{-\infty}^{\infty} \varphi(y) p_{T-t}(x,y) dy.$$

  Thus, the solution of Equation (2) can be estimated by

$$\widehat{F}_{\mathbf{m}}(t,x) := \int_{-\infty}^{\infty} \varphi(y) \widehat{p}_{\mathbf{m},T-t}(x,y) dy. \tag{3}$$

  In the same spirit as Milstein *et al.* [29], (3) provides a Monte-Carlo method which may replace the usual finite difference algorithm to solve Equation (2) numerically.
- The options pricing in finance is another possible application of our estimation method of p_t. Let $X = (X_t)_{t \in \mathbb{R}}$ be the prices process of a risky asset, which risk-neutral dynamics are modeled by

$$\frac{dX_t}{X_t} = (r - \delta)dt + \sigma(X_t) dW_t, \tag{4}$$

  where $r > 0$ is the risk-free rate, $\delta > 0$ the dividend rate, $W = (W_t)_{t \in \mathbb{R}}$ is a two-sided Brownian motion, and X_0 is a $\sigma((W_t)_{t \in \mathbb{R}_-})$-measurable square integrable random variable. Consider also the option with payoff $\varphi(X_T^{0,x})$, which price $P(x,T)$ satisfies

$$P(x,T) = e^{-rT} \mathbb{E}(\varphi(X_T^{0,x})) = e^{-rT} \int_0^{\infty} \varphi(y) p_T(x,y) dy.$$

Thus, when $r < \delta$, an estimator of $P(x, T)$ is given by

$$\widehat{P}_{\mathbf{m}}(x, T) := e^{-rT} \int_0^{\infty} \varphi(y)\widehat{p}_{\mathbf{m},T}(x, y)dy,$$

where the copies $X^1, \dots, X^N$ of $(X_t^{0,x})_{t\in[0,2T]}$ are here constructed from $(X_t)_{t\in\mathbb{R}_-}$, the past of the (recurrent Markov) prices process, by following the same line as in Marie [26], Remark 2.3.

3. A projection least squares estimator of the transition density function

3.1. **Assumptions: reminder and additional comments.** First, recall that throughout the paper, we consider copies $X^1, \dots, X^N$ of the solution X of Equation (1) under the condition:

$$x_0 \in \mathbb{R},\ W = (W_t)_{t\in[0,2T]} \text{ is a Brownian motion, } b, \sigma \in C^1(\mathbb{R}) \text{ and } b', \sigma' \text{ are bounded.} \tag{5}$$

As already mentioned, under condition (5), Equation (1) has a unique solution. We also assume that σ satisfies the following non-degeneracy condition:

$$\exists \alpha, A > 0 : \forall x \in \mathbb{R},\ \alpha \leqslant |\sigma(x)| \leqslant \alpha + A. \tag{6}$$

Under the condition (6), the transition density function p_t is well-defined and, for every $x, y \in \mathbb{R}$,

$$p_t(x, y) \leqslant \mathfrak{c}_T t^{-\frac{1}{2}} \exp\left(-\mathfrak{m}_T \frac{(y-x)^2}{t}\right), \tag{7}$$

where $\mathfrak{c}_T$ and $\mathfrak{m}_T$ are positive constants depending on T but not on t, x and y (see Menozzi *et al.* [28], Theorem 1.2). In particular, $t \mapsto p_t(x_0, x)$ belongs to $\mathbb{L}^1([0, T])$, which legitimates to consider the density function f defined by

$$f(x) := \frac{1}{T} \int_0^T p_s(x_0, x)ds \ ;\ \forall x \in \mathbb{R}.$$

Still by Inequality (7):

- f is bounded. Indeed, for every $x \in \mathbb{R}$,

$$f(x) \leqslant \frac{1}{T} \int_0^T \mathfrak{c}_T s^{-\frac{1}{2}} ds = 2\mathfrak{c}_T T^{-\frac{1}{2}}.$$

- Since b' is bounded (and then b has linear growth),

$$|b|^{\kappa} \in \mathbb{L}^2(\mathbb{R}, f(x)dx) \ ;\ \forall \kappa \in \mathbb{R}_+.$$

  Indeed, for every $\upsilon \in \mathbb{R}_+$ and $t \in (0, T]$,

$$\begin{aligned} \mathbb{E}(|X_t|^{\upsilon}) &= \int_{-\infty}^{\infty} |x|^{\upsilon} p_t(x_0, x)dx \\ &\leqslant t^{-\frac{1}{2}} \underbrace{\mathfrak{c}_T \int_{-\infty}^{\infty} |x|^{\upsilon} \exp\left(-\mathfrak{m}_T \frac{(x-x_0)^2}{T}\right) dx}_{=:\mathfrak{c}_{T,\upsilon}<\infty} \end{aligned}$$

  and then, for every $\kappa \in \mathbb{R}_+$,

$$\begin{aligned} \int_{-\infty}^{\infty} |b(x)|^{2\kappa} f(x)dx &= \frac{1}{T} \int_0^T \mathbb{E}(|b(X_s)|^{2\kappa})ds \\ &\leqslant \mathfrak{c}_1 \left(1 + \frac{1}{T} \int_0^T \mathbb{E}(|X_s|^{2\kappa})ds\right) \leqslant \mathfrak{c}_1(1 + 2\mathfrak{c}_{T,2\kappa} T^{-\frac{1}{2}}) < \infty, \end{aligned}$$

  where $\mathfrak{c}_1$ is a positive constant depending only on b and κ.

3.2. **The projection least squares estimator and some related definitions.** Now, let us define rigorously the objective function $\gamma_N(\cdot)$. To that aim, consider $\mathcal{S}_{\mathbf{m}} = \mathcal{S}_{\varphi,m_1}\otimes\mathcal{S}_{\psi,m_2}$, where $\mathcal{S}_{\varphi,m_1} := \text{span}\{\varphi_1,\dots,\varphi_{m_1}\}$ (resp. $\mathcal{S}_{\psi,m_2} := \text{span}\{\psi_1,\dots,\psi_{m_2}\}$), $\varphi_1,\dots,\varphi_{N_T}$ (resp. $\psi_1,\dots,\psi_{N_T}$) are continuous functions from I (resp. J) into $\mathbb{R}$ such that $(\varphi_1,\dots,\varphi_{N_T})$ (resp. $(\psi_1,\dots,\psi_{N_T})$) is an orthonormal family in $\mathbb{L}^2(I,dx)$ (resp. $\mathbb{L}^2(J,dx)$), and $I,J\subset\mathbb{R}$ are non-empty intervals. Then, for any $\tau\in\mathcal{S}_{\mathbf{m}}$, consider

$$\gamma_N(\tau) := \frac{1}{NT}\sum_{i=1}^{N}\left(\int_0^T\int_{-\infty}^{\infty}\tau(X_s^i,y)^2dyds - 2\int_0^T\tau(X_s^i,X_{s+t}^i)ds\right).$$

To understand the relevance of the criterion, we compute its expectation:

$$\mathbb{E}(\gamma_N(\tau)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\tau(x,y)-p_t(x,y))^2f(x)dxdy - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p_t(x,y)^2f(x)dxdy.$$

This shows that, the closer τ is to p_t, the smaller $\mathbb{E}(\gamma_N(\tau))$. This is the reason why our paper deals with the estimator of p_t minimizing $\gamma_N(\cdot)$.

Let us show that γ_N has a unique minimizer in $\mathcal{S}_{\mathbf{m}}$. For $\tau = \sum_{j,\ell}\Theta_{j,\ell}(\varphi_j\otimes\psi_\ell)$ with $\Theta\in\mathcal{M}_{m_1,m_2}(\mathbb{R})$,

$$\nabla_\tau\gamma_N(\tau) = 2(\widehat{\Psi}_{m_1}\Theta - \widehat{Z}_{\mathbf{m},t}),$$

where

$$\widehat{\Psi}_{m_1} := \left(\frac{1}{NT}\sum_{i=1}^{N}\int_0^T\varphi_j(X_s^i)\varphi_{j'}(X_s^i)ds\right)_{j,j'\in\{1,\dots,m_1\}}$$

and

$$\widehat{Z}_{\mathbf{m},t} := \left(\frac{1}{NT}\sum_{i=1}^{N}\int_0^T\varphi_j(X_s^i)\psi_\ell(X_{s+t}^i)ds\right)_{(j,\ell)\in\{1,\dots,m_1\}\times\{1,\dots,m_2\}}.$$

The symmetric matrix $\widehat{\Psi}_{m_1}$ is positive semidefinite because for any $y\in\mathbb{R}^{m_1}$,

$$y^*\widehat{\Psi}_{m_1}y = \frac{1}{NT}\sum_{i=1}^{N}\int_0^T\left(\sum_{j=1}^{m_1}y_j\varphi_j(X_s^i)\right)^2ds \geqslant 0.$$

If in addition $\widehat{\Psi}_{m_1}$ is invertible, it is positive definite, and then

$$\widehat{p}_{\mathbf{m},t} = \sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2}[\widehat{\Theta}_{\mathbf{m},t}]_{j,\ell}(\varphi_j\otimes\psi_\ell) \quad\text{with}\quad \widehat{\Theta}_{\mathbf{m},t} = \widehat{\Psi}_{m_1}^{-1}\widehat{Z}_{\mathbf{m},t} \tag{8}$$

is the only minimizer of γ_N in $\mathcal{S}_{\mathbf{m}}$ called the projection least squares estimator of p_t.

Related definitions/notations:

(1) The empirical inner product $\langle .,.\rangle_N$ is defined by

$$\langle\varphi,\psi\rangle_N := \frac{1}{NT}\sum_{i=1}^{N}\int_0^T\int_{-\infty}^{\infty}\varphi(X_s^i,y)\psi(X_s^i,y)dyds.$$

The empirical norm associated to $\langle .,.\rangle_N$ is denoted by $\|.\|_N$.

(2) The theoretical inner product $\langle .,.\rangle_f$ associated to $\langle .,.\rangle_N$ is defined by

$$\langle\varphi,\psi\rangle_f := \mathbb{E}(\langle\varphi,\psi\rangle_N) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)\psi(x,y)f(x)dxdy.$$

(3) The theoretical counterpart of $\widehat{\Psi}_{m_1}$ is denoted by $\Psi_{m_1} := \mathbb{E}(\widehat{\Psi}_{m_1})$.

4. Risk bounds on the projection least squares estimator

4.1. **Risk bound with respect to the empirical norm.** This subsection deals with a nonadaptive risk bound, with respect to the empirical norm $\|.\|_N$, on our estimator $\widehat{p}_{\mathbf{m},t}$ of p_t. First, consider

$$\mathfrak{L}_\varphi(m) := 1 \vee \left(\sup_{x\in I}\sum_{j=1}^{m}\varphi_j(x)^2\right) \quad \text{and} \quad \mathfrak{L}_\psi(m) := 1 \vee \left(\sup_{x\in I}\sum_{j=1}^{m}\psi_j(x)^2\right)$$

for every $m \in \{1,\dots,N_T\}$, and assume that $\mathbf{m} = (m_1, m_2)$ satisfies the following conditions.

Assumption 4.1. *There exist two constants $\mathfrak{c}_{4.1} > 0$ and $q \in \mathbb{N}^*$, not depending on $\mathbf{m}$ and N, such that*

$$\mathfrak{L}_\psi(m_2) \leqslant \mathfrak{c}_{4.1}N^q.$$

Moreover,

$$\mathfrak{L}_\varphi(m_1)(\|\Psi_{m_1}^{-1}\|_{\mathrm{op}} \vee 1) \leqslant \frac{\mathfrak{c}_\Lambda}{2}\cdot\frac{NT}{\log(NT)} \tag{9}$$

with

$$\mathfrak{c}_\Lambda = \frac{1-\log(2)}{(1+p)T} \quad \textit{and} \quad p = 2(q+4)+1.$$

The first part of Assumption 4.1 is a weak limit on the maximal dimension m_2 that can be considered. The second part (9) of Assumption 4.1 is a generalization of the so-called *stability condition* introduced in the nonparametric regression framework in Cohen *et al.* [8], and already extended to the independent copies of continuous diffusion processes framework in Comte and Genon-Catalot [11].

Theorem 4.2. *Consider $p_{I\times J,t} := (p_t)_{|I\times J}$. Under Assumption 4.1, there exists a constant $\mathfrak{c}_{4.2} > 0$, not depending on $\mathbf{m}$ and N, such that for every $t \in [0,T]$,*

$$\mathbb{E}(\|\widehat{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2) \leqslant \min_{\tau\in\mathcal{S}_\mathbf{m}}\|\tau - p_{I\times J,t}\|_f^2 + \frac{2m_1\mathfrak{L}_\psi(m_2)}{N} + \frac{\mathfrak{c}_{4.2}}{N}. \tag{10}$$

We emphasize that the risk bound in Theorem 4.2 is sharp since the constant in front of the bias term

$$\min_{\tau\in\mathcal{S}_\mathbf{m}}\|\tau - p_{I\times J,t}\|_f^2 \quad \text{in Inequality (10) is 1,}$$

and the constant 2 in the variance term $2m_1\mathfrak{L}_\psi(m_2)/N$ may be $1+\varepsilon$ for $\varepsilon > 0$, up to additional technicalities.

Remark. Let us discuss the order of the variance term in Inequality (10) for some usual bases $(\psi_1,\dots,\psi_{m_2})$. First, for splines, wavelets or trigonometric bases, $\mathfrak{L}_\psi(m_2) \lesssim m_2$, and then the variance term in the risk bound on $\widehat{p}_{\mathbf{m},t}$ is of order m_1m_2/N as for the usual projection estimator of a 2-dimensional density function. Now, for Legendre's basis, $\mathfrak{L}_\psi(m_2) \lesssim m_2^2$, leading to a variance term of order $m_1m_2^2/N$ in Inequality (10). Finally, let us focus on the Hermite basis $(h_j)_{j\in\mathbb{N}}$, defined on $I = \mathbb{R}$ by

$$h_j(x) := \mathfrak{c}_jH_j(x)e^{-\frac{x^2}{2}} \quad \text{with} \quad \mathfrak{c}_j = (2^jj!\sqrt{\pi})^{-\frac{1}{2}} \quad \text{and} \quad H_j(x) = (-1)^je^{x^2}\frac{d^j}{dx^j}e^{-x^2} \tag{11}$$

for every $x \in \mathbb{R}$ and $j \in \mathbb{N}$. The sequence $(h_j)_{j\in\mathbb{N}}$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R}, dx)$. By Lemma 1 in Comte and Lacour [13], we know that there exists a constant $\mathfrak{c}_h > 0$ such that

$$\mathfrak{L}_h(m) \leqslant \mathfrak{c}_h\sqrt{m} \ ; \ \forall m \in \mathbb{N}.$$

Thus, for Hermite's basis, it is worth noting that the variance term in Inequality (10) is of order $m_1\sqrt{m_2}/N$. Therefore, for both Legendre's and Hermite's bases, the variance term in Inequality (10) is not standard.

4.2. **Risk bound with respect to the f-weighted norm, on a truncated version of the projection least squares estimator.** This subsection deals with a nonadaptive risk bound, with respect to the f-weighted norm $\|.\|_f$, on the following truncated version of our estimator $\widehat{p}_{\mathbf{m},t}$:

$$\widetilde{p}_{\mathbf{m},t} := \widehat{p}_{\mathbf{m},t}\mathbf{1}_{\Lambda_{m_1}},$$

where

$$\Lambda_{m_1} := \left\{\mathfrak{L}_{\varphi}(m_1)(\|\widehat{\Psi}_{m_1}^{-1}\|_{\rm op}\vee 1)\leqslant \mathfrak{c}_{\Lambda}\frac{NT}{\log(NT)}\right\}.$$

On the event Λ_{m_1}, $\widehat{\Psi}_{m_1}$ is invertible because

$$\inf\{{\rm sp}(\widehat{\Psi}_{m_1})\}\geqslant \frac{\mathfrak{L}_{\varphi}(m_1)}{\mathfrak{c}_{\Lambda}}\cdot\frac{\log(NT)}{NT},$$

and then $\widetilde{p}_{\mathbf{m},t}$ is well-defined.

Theorem 4.3. *Under Assumption 4.1,*

(1) There exists a constant $\mathfrak{c}_{4.3,1}>0$, not depending on $\mathbf{m}$ and N, such that for every $t\in(0,T]$,

$$\mathbb{E}(\|\widetilde{p}_{\mathbf{m},t}-p_{I\times J,t}\|_f^2)\leqslant 9\min_{\tau\in\mathcal{S}_{\mathbf{m}}}\|\tau-p_{I\times J,t}\|_f^2+\frac{8m_1\mathfrak{L}_{\psi}(m_2)}{N}+\frac{\mathfrak{c}_{4.3,1}(1+R_f(t))}{N}$$

with $R_f(t)=R(t)+\|p_{I\times J,t}\|_f^2$ and

$$R(t)=\frac{1}{T}\mathbb{E}\left[\left(\int_0^T\int_{-\infty}^{\infty}p_t(X_s,y)^2dyds\right)^2\right]^{\frac{1}{2}}.$$

(2) There exists a constant $\mathfrak{c}_{4.3,2}>0$, not depending on $\mathbf{m}$ and N, such that

$$\frac{1}{T}\int_0^T\mathbb{E}(\|\widetilde{p}_{\mathbf{m},t}-p_{I\times J,t}\|_f^2)dt\leqslant 9\min_{\tau\in\mathcal{S}_{\mathbf{m}}}\left\{\frac{1}{T}\int_0^T\|\tau-p_{I\times J,t}\|_f^2dt\right\}+\frac{8m_1\mathfrak{L}_{\psi}(m_2)}{N}+\frac{\mathfrak{c}_{4.3,2}}{N}. \tag{12}$$

Remark. Under the condition (6), for every $x,y\in\mathbb{R}$,

$$p_t(x,y)\geqslant \underline{\mathfrak{c}}_Tt^{-\frac{1}{2}}\exp\left(-\underline{\mathfrak{m}}_T\frac{(y-x)^2}{t}\right), \tag{13}$$

where $\underline{\mathfrak{c}}_T$ and $\underline{\mathfrak{m}}_T$ are positive constants depending on T but not on t, x and y (see Menozzi *et al.* [28], Theorem 1.2). By Jensen's inequality, Inequality (13) and the change of variable formula,

$$\begin{aligned}
R(t) &\geqslant \frac{1}{T}\int_0^T\int_{-\infty}^{\infty}\mathbb{E}(p_t(X_s,y)^2)dyds=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p_t(x,y)^2f(x)dxdy\\
&\geqslant \frac{\underline{\mathfrak{c}}_T^2}{t}\int_{-\infty}^{\infty}f(x)\int_{-\infty}^{\infty}\exp\left(-2\underline{\mathfrak{m}}_T\frac{(y-x)^2}{t}\right)dydx\\
&=\frac{\underline{\mathfrak{c}}_T^2}{t}\int_{-\infty}^{\infty}f(x)dx\int_{-\infty}^{\infty}\exp\left[-2\underline{\mathfrak{m}}_T\left(\frac{y}{t^{1/2}}\right)^2\right]dy\\
&=\frac{\underline{\mathfrak{c}}_T^2}{t^{1/2}}\int_{-\infty}^{\infty}e^{-2\underline{\mathfrak{m}}_Ty^2}dy\xrightarrow[t\to 0^+]{}\infty.
\end{aligned}$$

Therefore,

$$\sup_{t\in(0,T]}R(t)=\infty,$$

and this is the reason why Inequality (12) is also provided. Otherwise, only times $t\in[t_0,T]$, for some fixed $t_0\in(0,T)$, must be considered.

4.3. **Rates in the anisotropic Sobolev-Hermite spaces.** In order to control the bias term in Theorems 4.2 and 4.3, we assume that p_t belongs to a Sobolev-Hermite space. In dimension $d = 1$, these function spaces have been introduced by Bongioanni and Torrea [4]. The connection with Hermite's coefficients was established later (see Belomestny *et al.* [5]) and are summarized in Comte and Genon-Catalot [9]. The definition of these spaces can be extended on $A := \mathbb{R}^d$ for any $d \in \mathbb{N}^*$ (see Comte and Lacour [13], Section 2.3).

Notations:

- For every $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{N}^d$ and $\mathbf{s} = (s_1, \dots, s_d) \in (0, \infty)^d$, $\mathbf{k}^{\mathbf{s}} := k_1^{s_1} \times \cdots \times k_d^{s_d}$.
- For every $g \in \mathbb{L}^2(A)$ and $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{N}^d$, $a_{\mathbf{k}}(g) := \langle g, h_{k_1} \otimes \cdots \otimes h_{k_d} \rangle$.

Throughout this subsection, both $\mathcal{S}_{\varphi,m_1}$ and $\mathcal{S}_{\psi,m_2}$ are generated by the Hermite basis, and then $I = J = \mathbb{R}$. First, let us recall the definition of the Sobolev-Hermite ellipsoid on $A := \mathbb{R}^d$, of order $\mathbf{s} = (s_1, \dots, s_d) \in (0, \infty)^d$ and of radius $L > 0$.

Definition 4.4. *(Sobolev-Hermite ellispoids) The Sobolev-Hermite ellipsoid $W_{\mathbf{s}}^{(d)}(A, L)$ of order $\mathbf{s}$ and of radius L is defined by*

$$W_{\mathbf{s}}^{(d)}(A, L) := \left\{ g \in \mathbb{L}^2(A) : \sum_{\mathbf{k} \in \mathbb{N}^d} a_{\mathbf{k}}(g)^2 \mathbf{k}^{\mathbf{s}} \leqslant L \right\}.$$

Now, by assuming that p_t belongs to $W_{\mathbf{s}}^{(2)}(A, L)$, the bias term in Theorems 4.2 and 4.3 decreases to 0 with polynomial rate. Indeed, noting $p_{\mathbf{m},t}$ the orthogonal projection of p_t on $\mathcal{S}_{\mathbf{m}}$,

$$\begin{aligned} \|p_t - p_{\mathbf{m},t}\|^2 &= \sum_{\mathbf{k} \in \mathbb{N}^2 : \exists q \in \{1,2\}, k_q \geqslant m_q} a_{\mathbf{k}}(p_t)^2 \\ &\leqslant \sum_{q=1}^{2} \sum_{\mathbf{k} \in \mathbb{N}^2 : k_q \geqslant m_q} a_{\mathbf{k}}(p_t)^2 k_q^{s_q} k_q^{-s_q} \leqslant L(m_1^{-s_1} + m_2^{-s_2}) \end{aligned}$$

and then, since $f(\cdot) \leqslant 2\mathfrak{c}_T T^{-1/2}$,

$$\min_{\tau \in \mathcal{S}_{\mathbf{m}}} \|\tau - p_{I \times J,t}\|_f^2 \leqslant \|f\|_\infty \|p_t - p_{\mathbf{m},t}\|^2 \leqslant \frac{2\mathfrak{c}_T L}{\sqrt{T}} (m_1^{-s_1} + m_2^{-s_2}).$$

Thanks to this control of the bias term, and since the variance term in Theorem 4.2 is of order $m_1\sqrt{m_2}/T$ when $\mathcal{S}_{\psi,m_2}$ is generated by the Hermite basis (see Subsection 4.1), one can establish the following proposition.

Proposition 4.5. *Assume that p_t belongs to $W_{\mathbf{s}}^{(2)}(A, L)$, and consider $\mathbf{m}^* = (m_1^*, m_2^*)$ with*

$$m_1^* \propto N^{\frac{s_2}{s_1 s_2 + s_1/2 + s_2}} \quad \text{and} \quad m_2^* \propto N^{\frac{s_1}{s_1 s_2 + s_1/2 + s_2}}.$$

Then,

$$\mathbb{E}(\|\widehat{p}_{\mathbf{m}^*,t} - p_t\|_N^2) = O\left(N^{-\frac{1}{1 + \frac{1}{s_1} + \frac{1}{2s_2}}}\right),$$

provided that m_1^* satisfies the stability condition (9) in Assumption 4.1.

The proof of Proposition 4.5 follows the same line as in Comte and Lacour [13] and is omitted here. Finally, an optimality result for rates in conditional density estimation (see Theorem 1 in Comte and Lacour [13], Section 3.4) suggests that the rate in our Proposition 4.5 is most likely to be optimal.

5. Model selection

5.1. **General case.** In order to introduce an appropriate model selection criterion, throughout this section, the φ_j's and the ψ_ℓ's fulfill the following additional assumption.

Assumption 5.1. *The φ_j's and the ψ_ℓ's fulfill the three following conditions:*

(1) For every $m_1, M_1 \in \{1, \dots, N_T\}$, if $M_1 > m_1$, then $\mathcal{S}_{\varphi,m_1} \subset \mathcal{S}_{\varphi,M_1}$.

(2) For every $m_2, M_2 \in \{1, \dots, N_T\}$, if $M_2 > m_2$, then $\mathcal{S}_{\psi,m_2} \subset \mathcal{S}_{\psi,M_2}$.
(3) There exists a constant $\mathfrak{c}_\varphi \geqslant 1$, not depending on N, such that

$$\mathfrak{L}_\varphi(m_1) \leqslant \mathfrak{c}_\varphi^2 m_1 \ ; \ \forall m_1 \in \{1, \dots, N_T\}.$$

Note that the first two conditions above mean that the two univariate collections of models are nested, which of course does not imply that the product spaces are. For instance, the compactly supported trigonometric basis, and both the non-compactly supported Laguerre's and Hermite's bases, fulfill Assumption 5.1. Let us consider

$$\widehat{\mathbf{m}} = \arg \min_{\mathbf{m} \in \widehat{\mathcal{M}}_N} \{-\|\widehat{p}_{\mathbf{m},t}\|_N^2 + 2\kappa \text{pen}(\mathbf{m})\}, \tag{14}$$

where $\kappa \geqslant \kappa_0$, $\kappa_0 > 0$ is defined later,

$$\text{pen}(\mathbf{m}) := (1 + \log(N)) \frac{m_1 \mathfrak{L}_\psi(m_2)}{N} \ ; \ \forall \mathbf{m} = (m_1, m_2) \in \{1, \dots, N_T\}^2, \tag{15}$$

and $\widehat{\mathcal{M}}_N := \mathcal{U}_N \cap (\widehat{\mathcal{V}}_N \times \mathcal{N})$ with $\mathcal{N} = \{1, \dots, N \wedge N_T\}$,

$$\mathcal{U}_N = \{(m_1, m_2) \in \mathcal{N}^2 : m_1 \mathfrak{L}_\psi(m_2) \leqslant N\},$$
$$\widehat{\mathcal{V}}_N = \left\{ m_1 \in \mathcal{N} : \mathfrak{c}_\varphi^2 m_1 (\|\widehat{\Psi}_{m_1}^{-1}\|_{\text{op}}^2 \vee 1) \leqslant \mathfrak{d} \frac{NT}{\log(NT)} \right\}$$

and

$$\mathfrak{d} = \min\left(\frac{1}{8\mathfrak{c}_\varphi^2 T (\|f\|_\infty + (3\mathfrak{c}_\varphi)^{-1} \sqrt{\mathfrak{c}_\Lambda / 8})(1 + p)}, \frac{\mathfrak{c}_\Lambda}{8} \right).$$

Note that for every $\mathbf{m} = (m_1, m_2) \in \mathcal{U}_N$, m_2 fulfills the first part of Assumption 4.1 with $q = 1$, and then $p = 11$ (recall that $p = 2(q + 4) + 1$) in this section. Consider also the theoretical counterpart $\mathcal{M}_N := \mathcal{U}_N \cap (\mathcal{V}_N \times \mathcal{N})$ of $\widehat{\mathcal{M}}_N$, where

$$\mathcal{V}_N := \left\{ m_1 \in \mathcal{N} : \mathfrak{c}_\varphi^2 m_1 (\|\Psi_{m_1}^{-1}\|_{\text{op}}^2 \vee 1) \leqslant \frac{\mathfrak{d}}{4} \cdot \frac{NT}{\log(NT)} \right\}.$$

The following theorem provides a risk bound on the adaptive estimator $\widehat{p}_{\widehat{\mathbf{m}},t}$.

Theorem 5.2. *Under Assumption 5.1, for $\kappa_0 = 44a$ and $a \geqslant (2 \cdot 84\sqrt{\mathfrak{d}T})^2/2$,*
(1) There exists a constant $\mathfrak{c}_{5.2,1} > 0$, not depending on N, such that for every $t \in (0, T]$,

$$\mathbb{E}(\|\widehat{p}_{\widehat{\mathbf{m}},t} - p_{I\times J,t}\|_N^2) \leqslant 6 \min_{\mathbf{m} \in \mathcal{M}_N} \{\mathbb{E}(\|\widehat{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2) + \kappa \text{pen}(\mathbf{m})\} + \frac{\mathfrak{c}_{5.2,1}(1 + R(t))}{N}.$$

(2) There exists a constant $\mathfrak{c}_{5.2,2} > 0$, not depending on N, such that

$$\frac{1}{T}\int_0^T \mathbb{E}(\|\widehat{p}_{\widehat{\mathbf{m}},t} - p_{I\times J,t}\|_N^2) dt \leqslant 6 \min_{\mathbf{m} \in \mathcal{M}_N} \left\{ \frac{1}{T}\int_0^T \mathbb{E}(\|\widehat{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2) dt + \kappa \text{pen}(\mathbf{m}) \right\} + \frac{\mathfrak{c}_{5.2,2}}{N}.$$

The first result in Theorem 5.2 means that the final estimator $\widehat{p}_{\widehat{\mathbf{m}},t}$ makes automatically (up to a multiplicative constant which may be taken equal to $6 + 2\kappa$) the bias-variance tradeoff by keeping in mind Inequality (10) which provides a risk bound on $\mathbb{E}(\|\widehat{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2)$ for every $\mathbf{m} \in \mathcal{M}_N$.

5.2. **Two special cases: $t > t_0 > 0$ and compactly supported bases.** This subsection deals with two interesting special cases. First, let us consider $t \in [t_0, T]$ with a fixed $t_0 > 0$. This condition on t leads to

$$\sup_{(x,y) \in I\times J} p_t(x, y) \leqslant p_0 := \mathfrak{c}_T t_0^{-\frac{1}{2}} \quad \text{by Inequality (7)},$$

and then the *log term* in the penalty defined by (15) is not required anymore. Precisely, the following model selection criterion, simpler than (14), may be considered:

$$\widetilde{\mathbf{m}} = \arg \min_{\mathbf{m} \in \widehat{\mathcal{M}}_N} \{-\|\widehat{p}_{\mathbf{m},t}\|_N^2 + 2\kappa_b \text{pen}_b(\mathbf{m})\}, \tag{16}$$

where $\kappa_b \geqslant \kappa_{b,0}$, $\kappa_{b,0} > 0$ is defined later, and

$$\text{pen}_b(\mathbf{m}) := \frac{m_1 \mathfrak{L}_\psi(m_2)}{N} \ ; \forall \mathbf{m} = (m_1, m_2) \in \{1, \dots, N_T\}^2.$$

In the sequel, the map $(m_1, m_2) \mapsto m_1 \mathfrak{L}_\psi(m_2)$ fulfills the following additional but usual condition.

Assumption 5.3. *For every $\xi > 0$, there exists $S(\xi) > 0$, not depending on N, such that*

$$\sum_{1 \leqslant m_1, m_2 \leqslant N} \exp(-\xi m_1 \mathfrak{L}_\psi(m_2)) \leqslant S(\xi) < \infty.$$

Note that if $\mathfrak{L}_\psi(m_2) = 1 \vee (\mathfrak{c}_\psi m_2)$, or even if $\mathfrak{L}_\psi(m_2) = 1 \vee (\mathfrak{c}_\psi \sqrt{m_2})$, then $(m_1, m_2) \mapsto m_1 \mathfrak{L}_\psi(m_2)$ fulfills Assumption 5.3 because

$$\sum_{m_1=1}^{N} \exp(-\xi \mathfrak{L}_\psi(m_2))^{m_1} \leqslant \frac{e^{-\xi \mathfrak{L}_\psi(m_2)}}{1 - e^{-\xi}} \ ; \forall m_2 \in \mathbb{N}^*.$$

So, for instance, $(m_1, m_2) \mapsto m_1 \mathfrak{L}_\psi(m_2)$ fulfills Assumption 5.3 when $\mathcal{S}_{\psi, m_2}$ is generated by the trigonometric basis or by Hermite's one.

Theorem 5.4. *Under Assumptions 5.1 and 5.3, for $\kappa_{b,0} = 16.5$, there exists a constant $\mathfrak{c}_{5.4} > 0$, not depending on N, such that for every $t \in [t_0, T]$ with $t_0 > 0$,*

$$\mathbb{E}(\|\widehat{p}_{\widetilde{\mathbf{m}},t} - p_{I \times J,t}\|_N^2) \leqslant 6 \min_{\mathbf{m} \in \mathcal{M}_N} \{\mathbb{E}(\|\widehat{p}_{\mathbf{m},t} - p_{I \times J,t}\|_N^2) + \kappa_b \text{pen}_b(\mathbf{m})\} + \frac{\mathfrak{c}_{5.4}(1 + t_0^{-1/2})}{N}.$$

Now, let us briefly present the second interesting special case by assuming that I is a compact interval. By Inequality (13), there exists $\underline{\mathfrak{m}} > 0$ such that $f(\cdot) \geqslant \underline{\mathfrak{m}}$, and then

$$\begin{aligned}
\|\Psi_{m_1}^{-1}\|_{\text{op}} = \frac{1}{\lambda_{\min}(\Psi_{m_1})} &= \left(\inf_{\theta : \|\theta\|_{2,m_1} = 1} \sum_{j,j'=1}^{m_1} \theta_j \theta_{j'} [\Psi_{m_1}]_{j,j'} \right)^{-1} \\
&= \left[\inf_{\theta : \|\theta\|_{2,m_1} = 1} \int_I \left(\sum_{j=1}^{m_1} \theta_j \varphi_j(x) \right)^2 f(x) dx \right]^{-1} \leqslant \frac{1}{\underline{\mathfrak{m}}} \quad \text{for every } m_1 \in \mathcal{N}.
\end{aligned}$$

So, for $t \in (0, T]$ (resp. $t \in [t_0, T]$ with a fixed $t_0 > 0$), the model selection criterion (14) (resp. (16)) may be simplified another way:

$$\widehat{\mathbf{m}}^* = \arg \min_{\mathbf{m} \in \mathcal{M}_N^*} \{-\|\widehat{p}_{\mathbf{m},t}\|_N^2 + 2\kappa \text{pen}(\mathbf{m})\}$$

$$(\text{resp. } \widetilde{\mathbf{m}}^* = \arg \min_{\mathbf{m} \in \mathcal{M}_N^*} \{-\|\widehat{p}_{\mathbf{m},t}\|_N^2 + 2\kappa_b \text{pen}_b(\mathbf{m})\}),$$

where $\mathcal{M}_N^* := \mathcal{U}_N \cap (\mathcal{V}_N^* \times \mathcal{N})$ and

$$\mathcal{V}_N^* := \left\{ m_1 \in \mathcal{N} : m_1 \leqslant \frac{\mathfrak{c}_\Lambda}{2\mathfrak{c}_\varphi^2(\underline{\mathfrak{m}}^{-1} \vee 1)} \cdot \frac{NT}{\log(NT)} \right\}.$$

A result similar to Theorem 5.2 (resp. Theorem 5.4) may be established on the adaptive estimator $\widehat{p}_{\widehat{\mathbf{m}}^*,t}$ (resp. $\widehat{p}_{\widetilde{\mathbf{m}}^*,t}$) by taking $\widehat{\mathcal{M}}_N = \mathcal{M}_N = \mathfrak{M}_N := \mathcal{M}_N^*$ in the proof (of Theorem 5.2 (resp. Theorem 5.4)).

6. Numerical experiments

We propose a brief simulation study to illustrate our estimation method. The implementation is done using the Hermite basis defined by (11) ($I = J = \mathbb{R}$). The Hermite polynomials are computed thanks to the recursion formula $H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$ with $H_0(x) = 1$ and $H_1(x) = x$ (see Abramowitz and Stegun [1], (22.7)).

We fix $t = 1$, so we can take the penalty involved in Theorem 5.4: $\text{pen}_b(\mathbf{m}) = m_1\sqrt{m_2}/N$. Moreover, we choose $\kappa_b = 2$; a value obtained from preliminary calibration experiments. A cutoff test excludes the dimensions m_1 such that the largest eigenvalue of $\widehat{\Psi}_{m_1}^{-1}$ is too large (see Comte and Genon-Catalot [11]).

We simulate discrete samples in three models, obtained from an exact discretization of d-dimensional Ornstein-Uhlenbeck processes $U_1, \dots, U_N$:

$$dU_i(t) = -\frac{r}{2}U_i(t)dt + \frac{\gamma}{2}dW_{i,d}(t), \quad U_i(0) \sim \mathcal{N}_d\left(0, \frac{\gamma^2}{4r}I_d\right), \tag{17}$$

where $W_{i,d}$ is a d-dimensional standard Brownian motion. An exact simulation is generated with step $\Delta > 0$ by computing

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

In all cases, we take $k \in \{0, \dots, n\}$ with $n = 1000$, $\Delta = 0.01$, and as already mentioned, we fix $t = 1$ in the function $(x, y) \mapsto p_t(x, y)$ to estimate.

Example 1. $X_i(t) = U_i(t)$, where $U_i(t)$ is defined by (17) with $d = 1$. The X_i's are independent copies of the solution of Equation (1) with

$$b(x) = -\frac{rx}{2}, \quad \sigma(x) = \frac{\gamma}{2}, \quad r = 2 \quad \text{and} \quad \gamma = 2.$$

Here, the transition density function is given by

$$p_t^{(1)}(x, y) = \sqrt{\frac{2r}{\pi\gamma^2(1-e^{-rt})}} \exp\left(-\frac{2r}{\gamma^2(1-e^{-rt})}(y - xe^{-\frac{rt}{2}})^2\right).$$

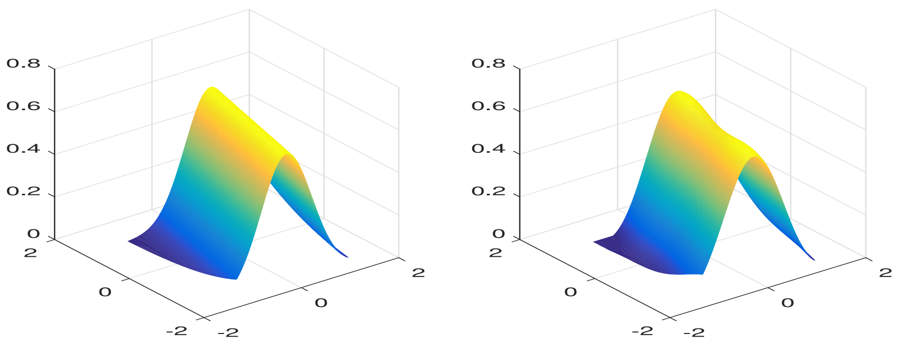

FIGURE 1. Example 1. Transition density (left) and the estimation (right). Selected dimensions (4,5), 100*MISE = 0.22. $N = 200$, $T = 10$, $\Delta = 0.01$, $t = 1$.

Example 2. $X_i(t) = \tanh(U_i(t))$, where $U_i(t)$ is defined by (17) with $d = 1$. The X_i's are independent copies of the solution of Equation (1) with

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

Here, the transition density function is given by

$$p_t^{(2)}(x, y) = \frac{p_t^{(1)}(\text{atanh}(x), \text{atanh}(y))}{1-y^2}.$$

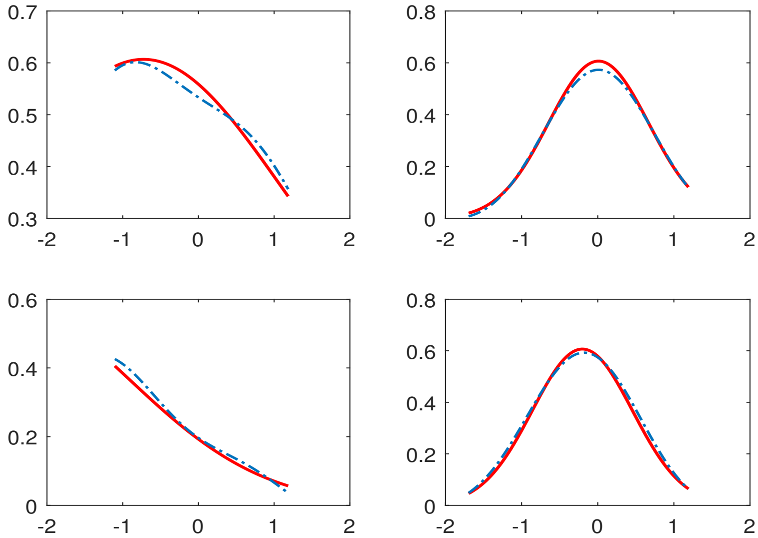

FIGURE 2. Example 1. Full red line, the true and the estimation in dotted blue. Left: $x \mapsto p_t(x, y)$ for a fixed value of y ($y = -0.27$ top and $y = -1$ bottom). Right: $y \mapsto p_t(x, y)$ for a fixed value of x ($x = 0.03$ top and $x = -0.55$ bottom). $N = 200$, $T = 10$, $\Delta = 0.01$, $t = 1$.

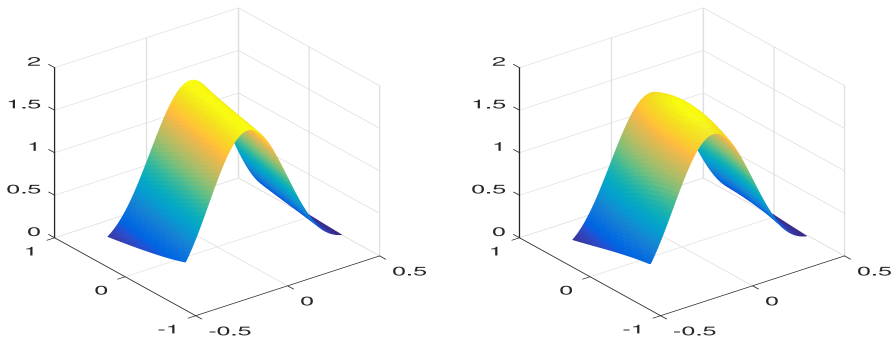

FIGURE 3. Example 2. Transition density (left) and the estimation (right). Selected dimensions (2,41), 100*MISE = 0.16. $N = 200$, $T = 10$, $\Delta = 0.01$, $t = 1$.

Example 3. (Cox-Ingersoll-Ross or square-root process) $X_i(t) = \|U_i(t)\|^2_{2,d}$, where $U_i(t)$ is defined by (17) with $d = 6$. The X_i's are independent copies of the solution of Equation (1) with

$$b(x) = \frac{d\gamma^2}{4} - rx, \quad \sigma(x) = \gamma\sqrt{x}, \quad r = 1 \quad \text{and} \quad \gamma = 1.$$

Here, the transition density function is given by

$$\begin{aligned} p_t^{(3)}(x, y) &= c_t \exp(-c_t(xe^{-rt} + y)) \\ &\quad \times \left(\frac{y}{xe^{-rt}}\right)^{\frac{d}{4} - \frac{1}{2}} \mathcal{I}\left(\frac{d}{2} - 1, 2c_t\sqrt{xye^{-rt}}\right), \quad \text{where} \quad c_t := \frac{2r}{\gamma^2(1 - e^{-rt})} \end{aligned}$$

and $\mathcal{I}(p, x)$ is the modified Bessel function of the first kind of order p at point x (see the formula (20) in Aït-Sahalia [2]).

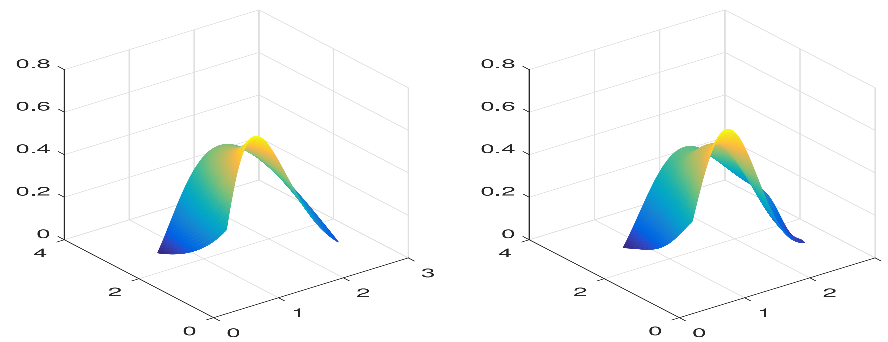

FIGURE 4. Example 3. Transition density (left) and the estimation (right). Selected dimensions (6,9), 100*MISE = 0.29. $N = 200$, $T = 10$, $\Delta = 0.01$, $t = 1$.

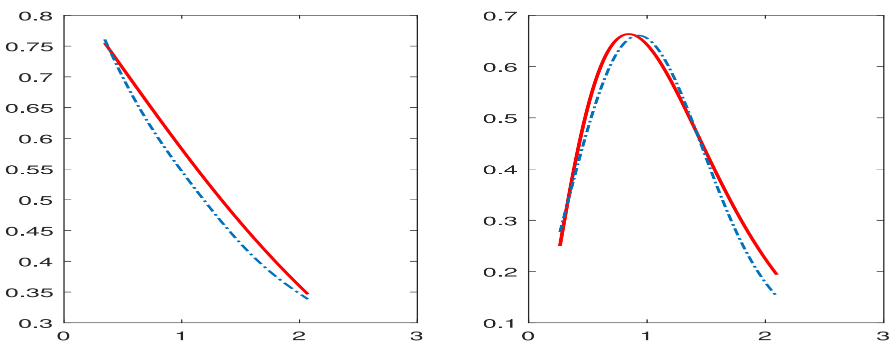

FIGURE 5. Example 3. Full red line, the true and the estimation in dotted blue. Left: $x \mapsto p_t(x, y)$ for a fixed value of y ($y = 0.71$). Right: $y \mapsto p_t(x, y)$ for a fixed value of x ($x = 0.76$). $N = 200$, $T = 10$, $\Delta = 0.01$, $t = 1$.

Figures 1, 3 and 4 show the true surface and the estimated one for Examples 1, 2 and 3 respectively. We see that the *shape* of the estimated transition density is very similar to the true one, and the selected dimensions can be of any orders, especially in the y-direction (see Figure 3, where the selected couple of dimensions is (2,41)). Figures 2 and 5 represent sections of the curve for fixed values of y or x for the same simulated paths as in Figure 1 and 4 respectively. One may notice that the estimation in the y-direction is better than in the x-direction.

In Table 1, we compute normalized squared errors for Models 1 and 3, which are defined by

$$\text{MISE} = \frac{\frac{1}{K}\sum_{k=1}^{K}\frac{DXY^{(k)}}{N_I N_J}\sum_{i=1}^{N_I}\sum_{j=1}^{N_J}(p_t(x_i^{(k)}, y_j^{(k)}) - \widehat{p}_t^{(k)}(x_i^{(k)}, y_j^{(k)}))^2}{\frac{DXY^{(K)}}{N_I N_J}\sum_{i=1}^{N_I}\sum_{j=1}^{N_J} p_t^2(x_i^{(K)}, y_j^{(K)})}, \tag{18}$$

where

$$DXY^{(k)} := (bX^{(k)} - aX^{(k)})(bY^{(k)} - aY^{(k)}),$$

$bX^{(k)}$ and $aX^{(k)}$ are the 98% and 2% quantiles of the $X_t^{(k)}$'s, and $bY^{(k)}$ and $aY^{(k)}$ are the 99% and 1% quantiles of the $X_{t+1}^{(k)}$'s. The super-index k denotes the repetition number, the points $(x_i^{(k)}, y_j^{(k)})$ are equispaced in the range of the observations of the path k, and we take $N_I = N_J = 100$ and $K = 200$. We also give the associated standard deviation, together with the mean of the selected dimensions in each

Model		$N = 100$	$N = 400$	$N = 1000$
1	MISE	1.62 (2.64)	0.48 (0.99)	0.18 (0.40)
	Medians	0.76	0.21	0.10
	Dim (10,12)	(3.05,4.91)	(4.23,5.92)	(5.00,7.00)
3	MISE	6.80 (10.5)	2.14 (5.62)	1.27 (5.62)
	Medians	1.65	0.34	0.19
	dim (12,15)	(5.01, 6.68)	(7.07, 9.07)	(8.94, 12.2)

TABLE 1. Line MISE: 100*MISE (with 100*standard deviation) computed over 200 repetitions, by (18). Line Medians: Median values of 100*MISE. Line Dim with maximal proposals for (m_1, m_2): means of the selected couples $(\widehat{m}_1, \widehat{m}_2)$.

direction. To spare time of computation, we've adjusted maximal proposals for m_1 and m_2 to the choices corresponding to each example in such a way that the largest proposal is never selected (i.e. is always too large).
The results in Table 1 show that, as could be expected, the error is getting smaller when N increases, and in the same time, the selected dimensions are increasing. This is expected from Proposition 4.5. We also notice that medians are much smaller in all cases than means: this indicates that the performance of the estimation is most of the time much better than what the mean indicates. Probably few bad results deteriorate the mean. The variability can be checked in the results to be unrelated to the selected dimensions, which are quite stable.

We've also implemented the (half)-trigonometric basis, or a product basis with $\varphi = t$ and $\psi = h$. They work well, with some questions on the nature of the theoretical bias term and underlying regularity spaces in the mixed (t and h) case.

7. Concluding remarks

In this paper, for a fixed $t \in (0, T]$, we have proposed a least squares contrast estimator $\widehat{p}_{\mathbf{m},t}(x, y)$ of the transition density $p_t(x, y)$ of the solution X to Equation (1). The estimator is defined through the estimators of the coefficients of its development on a basis of a finite dimensional space $\mathcal{S}_{\mathbf{m}} = \mathcal{S}_{\varphi,m_1} \otimes \mathcal{S}_{\psi,m_2}$, $\mathbf{m} = (m_1, m_2)$. In our observation setting, N independent copies $X^1, \dots, X^N$ of X are available. We provide an upper bound on a risk defined as the expectation of the empirical or integrated distance between $\widehat{p}_{\mathbf{m},t}$ and p_t, exhibiting a squared-bias/variance decomposition up to negligible terms. An adaptive procedure is then tailored to automatically select the couple $\mathbf{m} = (m_1, m_2)$, and it is proved to give good results, both in theory and in practice.
Of course, the simulation part immediately faces the question of handling high frequency but discrete samples and the topic may be developed. However, similarly to what happens in functional data analysis (FDA), small step samples often require a specific study from the beginning and can not be directly deduced from continuous-time constructions. In general, the continuous time results are extended to the high frequency one, up to some constraints linking the sample step and the number of observations. This would be worth being investigated. Note that FDA often considers both the influence of noise and discretization, which would mean here observations, for $i = 1, \dots, N$, of $X^i(t_j) + \eta_{i,j}$ with $t_j = j\Delta/n$, $j = 1, \dots, n$ and $\eta_{i,j}$ independent and identically distributed noises with common variance (see Chagny *et al.* [7]). It is not clear if both questions may be simultaneously solved in our setting.
Another question may be to study functions related to the conditional density, the main example being the conditional cumulative distribution function. Lastly, a recent study by Amorino *et al.* [3], defines a concept of local differential privacy in the context of i.i.d. diffusion processes, and it may be worth studying the estimation of f or p_t under such confidentiality condition.

Appendix A. Proofs

A.1. **Proof of Theorem 4.2.** The proof of Theorem 4.2 relies on two technical lemmas stated first.

Lemma A.1. *Consider the event*

$$\Omega_{m_1} := \left\{ \sup_{\tau\in\mathcal{S}_{\varphi,m_1}} \left| \frac{\|\tau\|_{N,1}^2}{\|\tau\|_{f,1}^2} - 1 \right| \leqslant \frac{1}{2} \right\}$$

where, for every $h\in\mathbb{L}^2(\mathbb{R},f(x)dx)$,

$$\|h\|_{N,1}^2 := \frac{1}{NT}\sum_{i=1}^{N}\int_0^T h(X_s^i)^2 ds \quad \textit{and} \quad \|h\|_{f,1}^2 := \int_{-\infty}^{\infty} h(x)^2 f(x)dx.$$

Under Assumption 4.1, there exists a constant $\mathfrak{c}_{A.1}>0$, not depending on m_1 and N, such that

$$\mathbb{P}(\Omega_{m_1}^c) \leqslant \frac{\mathfrak{c}_{A.1}}{N^p} \quad \textit{and} \quad \mathbb{P}(\Lambda_{m_1}^c) \leqslant \frac{\mathfrak{c}_{A.1}}{N^p} \quad \text{with} \quad p = 2(q+4)+1.$$

See Comte and Genon-Catalot [11], Lemma 6.1 for a proof. Now, let us introduce two additional empirical maps:

- The empirical process ν_N, defined by

$$\nu_N(\tau) := \frac{1}{NT}\sum_{i=1}^{N}\int_0^T \left(\tau(X_s^i, X_{s+t}^i) - \int_{-\infty}^{\infty}\tau(X_s^i,y)p_t(X_s^i,y)dy\right)ds$$

  for every $\tau\in\mathcal{S}_{\mathbf{m}}$. Note that

(19) $$[\widehat{Z}_{\mathbf{m},t}]_{j,\ell} = \langle p_t, \varphi_j\otimes\psi_\ell\rangle_N + \nu_N(\varphi_j\otimes\psi_\ell)$$

  for every $j\in\{1,\dots,m_1\}$ and $\ell\in\{1,\dots,m_2\}$.
- The empirical orthogonal projection $\widehat{\Pi}_{\mathbf{m}}$, defined by

(20) $$\widehat{\Pi}_{\mathbf{m}}(\cdot) \in \arg\min_{\tau\in\mathcal{S}_{\mathbf{m}}} \|\tau - \cdot\|_N^2.$$

  For any function h from $\mathbb{R}^2$ into $\mathbb{R}$,

$$\widehat{\Pi}_{\mathbf{m}}(h) = \sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2} [\widehat{\Psi}_{m_1}^{-1}\widehat{P}_{\mathbf{m}}(h)]_{j,\ell}(\varphi_j\otimes\psi_\ell)$$
$$\text{with} \quad \widehat{P}_{\mathbf{m}}(h) = (\langle h,\varphi_j\otimes\psi_\ell\rangle_N)_{(j,\ell)\in\{1,\dots,m_1\}\times\{1,\dots,m_2\}}.$$

Lemma A.2. *For every $t\in[0,T]$,*

$$\|\widehat{p}_{\mathbf{m},t} - \widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2 = \sup_{\tau\in\mathcal{S}_{\mathbf{m}}:\|\tau\|_N=1} \nu_N(\tau)^2.$$

The proof of Lemma A.2 is postponed to Subsubsection A.1.2.

A.1.1. *Steps of the proof.* First of all, by the definition of $\widehat{\Pi}_{\mathbf{m}}$ (see (20)),

$$\|\widehat{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2 = \min_{\tau\in\mathcal{S}_{\mathbf{m}}} \|\tau - p_{I\times J,t}\|_N^2 + \|\widehat{p}_{\mathbf{m},t} - \widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2,$$

and by Lemma A.2,

$$\begin{aligned}\mathbb{E}(\|\widehat{p}_{\mathbf{m},t} - \widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2) &= \mathbb{E}\left(\mathbf{1}_{\Omega_{m_1}} \sup_{\tau\in\mathcal{S}_{\mathbf{m}}:\|\tau\|_N=1}\nu_N(\tau)^2\right) + \mathbb{E}\left(\mathbf{1}_{\Omega_{m_1}^c} \sup_{\tau\in\mathcal{S}_{\mathbf{m}}:\|\tau\|_N=1}\nu_N(\tau)^2\right)\\ &=: A + B.\end{aligned}$$

Let us find suitable bounds on A and B.

Step 1 (bound on A). First, since

$$\|.\|_{f,1}^2\mathbf{1}_{\Omega_{m_1}} \leqslant 2\|.\|_{N,1}^2\mathbf{1}_{\Omega_{m_1}} \quad \text{on} \quad \mathcal{S}_{\varphi,m_1}$$

by the definition of Ω_{m_1}, for every $\tau \in \mathcal{S}_{\mathbf{m}}$,

$$\begin{aligned}\|\tau\|_f^2 \mathbf{1}_{\Omega_{m_1}} &= \int_{-\infty}^{\infty} \|\tau(\cdot, y)\|_{f,1}^2 \mathbf{1}_{\Omega_{m_1}} dy \\ &\qquad\qquad \leqslant 2 \int_{-\infty}^{\infty} \|\tau(\cdot, y)\|_{N,1}^2 \mathbf{1}_{\Omega_{m_1}} dy = 2\|\tau\|_N^2 \mathbf{1}_{\Omega_{m_1}}.\end{aligned}$$

Then,

$$\{\tau \in \mathcal{S}_{\mathbf{m}} : \|\tau\|_N = 1\} \subset \{\tau \in \mathcal{S}_{\mathbf{m}} : \|\tau\|_f^2 \leqslant 2\} \quad \text{on} \quad \Omega_{m_1},$$

leading to

$$A \leqslant \mathbb{E}\left(\sup_{\tau \in \mathcal{S}_{\mathbf{m}} : \|\tau\|_f^2 \leqslant 2} \nu_N(\tau)^2\right).$$

Since $(\varphi_1, \dots, \varphi_{m_1})$ is an orthonormal family of $\mathbb{L}^2(\mathbb{R}, dx)$, $\varphi_1, \dots, \varphi_{m_1}$ are linearly independent, and one may consider the basis $(\varphi_1^f, \dots, \varphi_{m_1}^f)$ of $\mathcal{S}_{\varphi, m_1}$, orthonormal in $\mathbb{L}^2(\mathbb{R}, f(x)dx)$, obtained from $(\varphi_1, \dots, \varphi_{m_1})$ via the Gram-Schmidt process. Thus, $(\varphi_j^f \otimes \psi_\ell)_{j,\ell}$ is an orthonormal basis of $\mathcal{S}_{\mathbf{m}}$ equipped with $\langle ., . \rangle_f$, and noting $\|.\|_{2,\mathbf{m}}$ the Fröbenius norm on $\mathcal{M}_{m_1,m_2}(\mathbb{R})$,

$$A \leqslant \mathbb{E}\left[\sup_{\Theta : \|\Theta\|_{2,\mathbf{m}}^2 \leqslant 2} \left(\sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \Theta_{j,\ell} \nu_N(\varphi_j^f \otimes \psi_\ell)\right)^2\right] \leqslant 2\mathbb{E}\left(\sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \nu_N(\varphi_j^f \otimes \psi_\ell)^2\right).$$

Now, note that for every $j \in \{1, \dots, m_1\}$ and $\ell \in \{1, \dots, m_2\}$,

$$\begin{aligned}\mathbb{E}(\nu_N(\varphi_j^f \otimes \psi_\ell)) &= \frac{1}{NT} \mathbb{E}\left(\sum_{i=1}^{N} \int_0^T \varphi_j^f(X_s^i)(\psi_\ell(X_{s+t}^i) - \mathbb{E}(\psi_\ell(X_{s+t}^i)|X_s^i))ds\right) \\ &= \frac{1}{T} \int_0^T \mathbb{E}(\varphi_j^f(X_s)\psi_\ell(X_{s+t}) - \mathbb{E}(\varphi_j^f(X_s)\psi_\ell(X_{s+t})|X_s))ds = 0.\end{aligned}$$

Therefore, since $\|\varphi_j^f\|_f = 1$ for every $j \in \{1, \dots, m_1\}$,

$$\begin{aligned}A &\leqslant 2\sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \mathrm{var}(\nu_N(\varphi_j^f \otimes \psi_\ell)) \\ &\leqslant \frac{2}{N} \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \mathbb{E}\left[\left(\frac{1}{T} \int_0^T \varphi_j^f(X_s)(\psi_\ell(X_{s+t}) - \mathbb{E}(\psi_\ell(X_{s+t})|X_s))ds\right)^2\right] \\ &\leqslant \frac{2}{NT} \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \int_0^T (\mathbb{E}(\varphi_j^f(X_s)^2 \psi_\ell(X_{s+t})^2) + \mathbb{E}(\varphi_j^f(X_s)^2 \mathbb{E}(\psi_\ell(X_{s+t})|X_s)^2) \\ &\qquad\qquad -2\mathbb{E}(\varphi_j^f(X_s)\psi_\ell(X_{s+t})\mathbb{E}(\varphi_j^f(X_s)\psi_\ell(X_{s+t})|X_s)))ds \\ &\leqslant \frac{2}{NT} \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \int_0^T \mathbb{E}(\varphi_j^f(X_s)^2 \psi_\ell(X_{s+t})^2)ds \leqslant \frac{2m_1 \mathfrak{L}_\psi(m_2)}{N}.\end{aligned}$$

Step 2 (bound on B). Since $\varphi_1, \dots, \varphi_{m_1}$ are linearly independent as mentioned in Step 1, one may consider the basis $(\varphi_1^N, \dots, \varphi_{m_1}^N)$ of $\mathcal{S}_{\varphi, m_1}$, orthonormal for the empirical inner product $\langle ., . \rangle_{N,1}$, obtained from $(\varphi_1, \dots, \varphi_{m_1})$ via the Gram-Schmidt process. Then, $(\varphi_j^N \otimes \psi_\ell)_{j,\ell}$ is an orthonormal basis of $\mathcal{S}_{\mathbf{m}}$

equipped with $\langle .,.\rangle_N$, and by Cauchy-Schwarz's and Jensen's inequalities,

$$\begin{aligned}\sup_{\tau\in\mathcal{S}_{\mathbf{m}}:\|\tau\|_N=1}\nu_N(\tau)^2 &= \sup_{\Theta:\|\Theta\|_{2,\mathbf{m}}^2=1}\left(\sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2}\Theta_{j,\ell}\nu_N(\varphi_j^N\otimes\psi_\ell)\right)^2\\ &\leqslant \sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2}\left(\frac{1}{NT}\sum_{i=1}^{N}\int_0^T\varphi_j^N(X_s^i)(\psi_\ell(X_{s+t}^i)-\mathbb{E}(\psi_\ell(X_{s+t}^i)|X_s^i))ds\right)^2\\ &\leqslant 4\mathfrak{L}_\psi(m_2)\sum_{j=1}^{m_1}\underbrace{\left(\frac{1}{NT}\sum_{i=1}^{N}\int_0^T\varphi_j^N(X_s^i)^2ds\right)}_{=1} = 4m_1\mathfrak{L}_\psi(m_2).\end{aligned}$$

Therefore, by Assumption 4.1 and Lemma A.1, there exists a constant $\mathfrak{c}_1>0$, not depending on $\mathbf{m}$, N and t, such that

$$B\leqslant 4m_1\mathfrak{L}_\psi(m_2)\mathbb{P}(\Omega_{m_1}^c)\leqslant\mathfrak{c}_1N^{-1}.$$

Step 3 (conclusion). By the two previous steps,

$$\mathbb{E}(\|\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2)\leqslant\frac{2m_1\mathfrak{L}_\psi(m_2)}{N}+\frac{\mathfrak{c}_1}{N}.$$

Thus,

$$\begin{aligned}\mathbb{E}(\|\widehat{p}_{\mathbf{m},t}-p_{I\times J,t}\|_N^2) &\leqslant \min_{\tau\in\mathcal{S}_{\mathbf{m}}}\mathbb{E}(\|\tau-p_{I\times J,t}\|_N^2)+\mathbb{E}(\|\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2)\\ &\leqslant \min_{\tau\in\mathcal{S}_{\mathbf{m}}}\|\tau-p_{I\times J,t}\|_f^2+\frac{2m_1\mathfrak{L}_\psi(m_2)}{N}+\frac{\mathfrak{c}_1}{N}.\end{aligned}$$

□

A.1.2. *Proof of Lemma A.2.* First, by Cauchy-Schwarz's inequality,

$$\sup_{\tau\in\mathcal{S}_{\mathbf{m}}:\|\tau\|_N=1}\langle\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t),\tau\rangle_N^2\leqslant\|\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2,$$

and since

$$\tau^*:=\frac{\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t)}{\|\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t)\|_N}\in\mathcal{S}_{\mathbf{m}}$$

satisfies both $\|\tau^*\|_N=1$ and

$$\langle\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t),\tau^*\rangle_N^2=\|\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2,$$

then

$$\|\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2=\sup_{\tau\in\mathcal{S}_{\mathbf{m}}:\|\tau\|_N=1}\langle\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t),\tau\rangle_N^2. \tag{21}$$

Now, let us show that

$$\langle\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t),\tau\rangle_N=\nu_N(\tau)\ ;\ \forall\tau\in\mathcal{S}_{\mathbf{m}}. \tag{22}$$

By the decompositions of $\widehat{p}_{\mathbf{m},t}$ and $\widehat{\Pi}_{\mathbf{m}}(p_t)$ in the basis $(\varphi_j\otimes\psi_\ell)_{j,\ell}$ of $\mathcal{S}_{\mathbf{m}}$, and by (19),

$$\begin{aligned}\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t) &= \sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2}[\widehat{\Psi}_{m_1}^{-1}(\widehat{Z}_{\mathbf{m},t}-\widehat{P}_{\mathbf{m}}(p_t))]_{j,\ell}(\varphi_j\otimes\psi_\ell)\\ &= \sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2}[\widehat{\Psi}_{m_1}^{-1}\widehat{\Delta}_{\mathbf{m}}]_{j,\ell}(\varphi_j\otimes\psi_\ell),\end{aligned}$$

where

$$\widehat{\Delta}_{\mathbf{m}} := \left(\frac{1}{NT}\sum_{i=1}^{N}\int_0^T \varphi_j(X_s^i)\psi_\ell(X_{s+t}^i)ds - \frac{1}{NT}\sum_{i=1}^{N}\int_0^T\int_{-\infty}^{\infty}\varphi_j(X_s^i)\psi_\ell(y)p_t(X_s^i,y)dyds\right)_{j,\ell} = (\nu_N(\varphi_j\otimes\psi_\ell))_{j,\ell}.$$

Then, for any $\tau = \sum_{j,\ell}\Theta_{j,\ell}(\varphi_j\otimes\psi_\ell)$ with $\Theta\in\mathcal{M}_{m_1,m_2}(\mathbb{R})$,

$$\begin{aligned}
\langle\widehat{p}_{\mathbf{m},t} - \widehat{\Pi}_{\mathbf{m}}(p_t),\tau\rangle_N &= \sum_{j,j'=1}^{m_1}\sum_{\ell,\ell'=1}^{m_2}[\widehat{\Psi}_{m_1}^{-1}\widehat{\Delta}_{\mathbf{m}}]_{j,\ell}\Theta_{j',\ell'}\langle\varphi_j\otimes\psi_\ell,\varphi_{j'}\otimes\psi_{\ell'}\rangle_N\\
&= \sum_{j,j'=1}^{m_1}\sum_{\ell,\ell'=1}^{m_2}[\widehat{\Psi}_{m_1}^{-1}\widehat{\Delta}_{\mathbf{m}}]_{j,\ell}\Theta_{j',\ell'}\\
&\qquad\times\underbrace{\frac{1}{NT}\sum_{i=1}^{N}\int_0^T\varphi_j(X_s^i)\varphi_{j'}(X_s^i)ds\int_{-\infty}^{\infty}\psi_\ell(y)\psi_{\ell'}(y)dy}_{=[\widehat{\Psi}_{m_1}]_{j,j'}\delta_{\ell,\ell'}}\\
&= \sum_{j'=1}^{m_1}\sum_{\ell=1}^{m_2}\Theta_{j',\ell}\underbrace{\sum_{j=1}^{m_1}[\widehat{\Psi}_{m_1}]_{j,j'}[\widehat{\Psi}_{m_1}^{-1}\widehat{\Delta}_{\mathbf{m}}]_{j,\ell}}_{=[\widehat{\Delta}_{\mathbf{m}}]_{j',\ell}}\\
&= \sum_{j'=1}^{m_1}\sum_{\ell=1}^{m_2}\Theta_{j',\ell}\nu_N(\varphi_{j'}\otimes\psi_\ell) = \nu_N(\tau).
\end{aligned}$$

Therefore, by Equalities (21) and (22) together,

$$\|\widehat{p}_{\mathbf{m},t} - \widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2 = \sup_{\tau\in\mathcal{S}_{\mathbf{m}}:\|\tau\|_N=1}\nu_N(\tau)^2.$$

□

A.2. **Proof of Theorem 4.3.** First of all,

$$\begin{aligned}
\mathbb{E}(\|\widetilde{p}_{\mathbf{m},t} - p_{I\times J,t}\|_f^2) &= \mathbb{E}(\|\widetilde{p}_{\mathbf{m},t} - p_{I\times J,t}\|_f^2\mathbf{1}_{\Omega_{m_1}}) + \mathbb{E}(\|\widetilde{p}_{\mathbf{m},t} - p_{I\times J,t}\|_f^2\mathbf{1}_{\Omega_{m_1}^c})\\
&=: A + B.
\end{aligned}$$

Let us find suitable bounds on A and B.

Step 1 (bound on A). As already mentioned, since

$$\|.\|_{f,1}^2\mathbf{1}_{\Omega_{m_1}} \leqslant 2\|.\|_{N,1}^2\mathbf{1}_{\Omega_{m_1}} \quad \text{on} \quad \mathcal{S}_{\varphi,m_1}$$

by the definition of Ω_{m_1}, for every $\tau\in\mathcal{S}_{\mathbf{m}}$,

$$\begin{aligned}
\|\tau\|_f^2\mathbf{1}_{\Omega_{m_1}} &= \int_{-\infty}^{\infty}\|\tau(\cdot,y)\|_{f,1}^2\mathbf{1}_{\Omega_{m_1}}dy\\
&\leqslant 2\int_{-\infty}^{\infty}\|\tau(\cdot,y)\|_{N,1}^2\mathbf{1}_{\Omega_{m_1}}dy = 2\|\tau\|_N^2\mathbf{1}_{\Omega_{m_1}}.
\end{aligned}$$

Then, noting $p_{I\times J,t}^f$ as the orthogonal projection of $p_{I\times J,t}$ onto $\mathcal{S}_{\mathbf{m}}$ for the inner product $\langle .,.\rangle_f$,

$$\begin{aligned}
\|\widetilde{p}_{\mathbf{m},t} - p_{I\times J,t}\|_f^2\mathbf{1}_{\Omega_{m_1}} &= (\|\widetilde{p}_{\mathbf{m},t} - p_{I\times J,t}^f\|_f^2 + \|p_{I\times J,t}^f - p_{I\times J,t}\|_f^2)\mathbf{1}_{\Omega_{m_1}}\\
&\leqslant \|p_{I\times J,t}^f - p_{I\times J,t}\|_f^2 + 2\|\widetilde{p}_{\mathbf{m},t} - p_{I\times J,t}^f\|_N^2\mathbf{1}_{\Omega_{m_1}}\\
&\leqslant \min_{\tau\in\mathcal{S}_{\mathbf{m}}}\|\tau - p_{I\times J,t}\|_f^2 + 4\|\widetilde{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2 + 4\|p_{I\times J,t} - p_{I\times J,t}^f\|_N^2.
\end{aligned}$$

Moreover,

$$\|\widetilde{p}_{\mathbf{m},t}-p_{I\times J,t}\|_N^2\leqslant\|\widehat{p}_{\mathbf{m},t}-p_{I\times J,t}\|_N^2+\|p_{I\times J,t}\|_N^2\mathbf{1}_{\Lambda_{m_1}^c}$$

and, by Lemma A.1,

$$\begin{aligned}\mathbb{E}(\|p_{I\times J,t}\|_N^2\mathbf{1}_{\Lambda_{m_1}^c})&\leqslant\mathbb{E}\left[\left(\frac{1}{NT}\sum_{i=1}^N\int_0^T\int_{-\infty}^{\infty}p_t(X_s^i,y)^2dyds\right)^2\right]^{\frac{1}{2}}\mathbb{P}(\Lambda_{m_1}^c)^{\frac{1}{2}}\\&\leqslant\mathfrak{c}_{A.1}^{\frac{1}{2}}\frac{R(t)}{N^{p/2}}\quad\text{with}\quad R(t)=\frac{1}{T}\mathbb{E}\left[\left(\int_0^T\int_{-\infty}^{\infty}p_t(X_s,y)^2dyds\right)^2\right]^{\frac{1}{2}}.\end{aligned}$$

Therefore, by Theorem 4.2, there exists a constant $\mathfrak{c}_1>0$, not depending on $\mathbf{m}$, N and t, such that

$$\begin{aligned}A&\leqslant 5\min_{\tau\in\mathcal{S}_{\mathbf{m}}}\|\tau-p_{I\times J,t}\|_f^2+4\mathbb{E}(\|\widetilde{p}_{\mathbf{m},t}-p_{I\times J,t}\|_N^2)\\&\leqslant 9\min_{\tau\in\mathcal{S}_{\mathbf{m}}}\|\tau-p_{I\times J,t}\|_f^2+\frac{8m_1\mathfrak{L}_{\psi}(m_2)}{N}+\frac{\mathfrak{c}_1(1+R(t))}{N}.\end{aligned}$$

Step 2 (bound on B). Since $\widehat{\Theta}_{\mathbf{m},t}\widehat{\Theta}_{\mathbf{m},t}^*$ is a positive semidefinite (symmetric) matrix,

$$\begin{aligned}\|\widehat{p}_{\mathbf{m},t}\|_f^2&=\int_{-\infty}^{\infty}f(x)\int_{-\infty}^{\infty}\left(\sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2}[\widehat{\Theta}_{\mathbf{m},t}]_{j,\ell}\varphi_j(x)\psi_\ell(y)\right)^2dydx\\&=\int_{-\infty}^{\infty}f(x)\sum_{j,j'=1}^{m_1}\sum_{\ell,\ell'=1}^{m_2}[\widehat{\Theta}_{\mathbf{m},t}]_{j,\ell}[\widehat{\Theta}_{\mathbf{m},t}]_{j',\ell'}\\&\qquad\qquad\times\varphi_j(x)\varphi_{j'}(x)\underbrace{\int_{-\infty}^{\infty}\psi_\ell(y)\psi_{\ell'}(y)dy}_{=\delta_{\ell,\ell'}}dx\\&=\sum_{j,j'=1}^{m_1}\sum_{\ell=1}^{m_2}[\widehat{\Theta}_{\mathbf{m},t}]_{j,\ell}[\widehat{\Theta}_{\mathbf{m},t}]_{j',\ell}\underbrace{\int_{-\infty}^{\infty}\varphi_j(x)\varphi_{j'}(x)f(x)dx}_{=[\Psi_{m_1}]_{j,j'}}\\&=\sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2}[\widehat{\Theta}_{\mathbf{m},t}^*]_{\ell,j}\underbrace{\sum_{j'=1}^{m_1}[\Psi_{m_1}]_{j,j'}[\widehat{\Theta}_{\mathbf{m},t}]_{j',\ell}}_{=[\Psi_{m_1}\widehat{\Theta}_{\mathbf{m},t}]_{j,\ell}}\\&=\sum_{j=1}^{m_1}[\Psi_{m_1}\widehat{\Theta}_{\mathbf{m},t}\widehat{\Theta}_{\mathbf{m},t}^*]_{j,j}=\mathrm{trace}(\Psi_{m_1}\widehat{\Theta}_{\mathbf{m},t}\widehat{\Theta}_{\mathbf{m},t}^*)\leqslant\|\Psi_{m_1}\|_{\mathrm{op}}\mathrm{trace}(\widehat{\Theta}_{\mathbf{m},t}\widehat{\Theta}_{\mathbf{m},t}^*).\end{aligned}$$

Let us find suitable controls on $\|\Psi_{m_1}\|_{\mathrm{op}}^2$ and $\mathbb{E}(\mathrm{trace}(\widehat{\Theta}_{\mathbf{m},t}\widehat{\Theta}_{\mathbf{m},t}^*)^2)$. On the one hand, by Cauchy-Schwarz's and Jensen's inequalities,

$$\begin{aligned}\|\Psi_{m_1}\|_{\mathrm{op}}^2&=\sup_{y:\|y\|_{2,m_1}=1}\sum_{j=1}^{m_1}\left(\sum_{j'=1}^{m_1}y_{j'}\int_{-\infty}^{\infty}\varphi_j(x)\varphi_{j'}(x)f(x)dx\right)^2\\&\leqslant\sum_{j,j'=1}^{m_1}\left(\int_{-\infty}^{\infty}\varphi_j(x)\varphi_{j'}(x)f(x)dx\right)^2\leqslant\mathfrak{L}_{\varphi}(m_1)^2\underbrace{\int_{-\infty}^{\infty}f(x)dx}_{=1}.\end{aligned}$$

On the other hand, since $\widehat{Z}_{\mathbf{m},t}\widehat{Z}_{\mathbf{m},t}^*$ is a positive semidefinite (symmetric) matrix,

$$\mathrm{trace}(\widehat{\Theta}_{\mathbf{m},t}\widehat{\Theta}_{\mathbf{m},t}^*)=\mathrm{trace}((\widehat{\Psi}_{m_1}^{-1})^2\widehat{Z}_{\mathbf{m},t}\widehat{Z}_{\mathbf{m},t}^*)\leqslant\|\widehat{\Psi}_{m_1}^{-1}\|_{\mathrm{op}}^2\mathrm{trace}(\widehat{Z}_{\mathbf{m},t}\widehat{Z}_{\mathbf{m},t}^*).$$

Moreover,

$$\mathbb{E}(\text{trace}(\widehat{Z}_{\mathbf{m},t}\widehat{Z}_{\mathbf{m},t}^*)^2) = \mathbb{E}\left[\left(\sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2}[\widehat{Z}_{\mathbf{m},t}]_{j,\ell}^2\right)^2\right]$$
$$\leqslant \frac{m_1m_2}{N^4T^4}\sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2}\mathbb{E}\left[\left(\sum_{i=1}^{N}\int_0^T\varphi_j(X_s^i)\psi_\ell(X_{s+t}^i)ds\right)^4\right] \leqslant m_1m_2\mathfrak{L}_\varphi(m_1)^2\mathfrak{L}_\psi(m_2)^2$$

and, on the event Λ_{m_1},

$$\|\widehat{\Psi}_{m_1}^{-1}\|_{\text{op}}^4 \leqslant \mathfrak{c}_\Lambda^4\frac{N^4T^4}{\log(NT)^4\mathfrak{L}_\varphi(m_1)^4}.$$

Then, there exists a constant $\mathfrak{c}_2 > 0$, not depending on $\mathbf{m}$, N and t, such that

$$\mathbb{E}(\|\widetilde{p}_{\mathbf{m},t}\|_f^4) \leqslant \mathfrak{c}_2N^4m_1m_2\mathfrak{L}_\psi(m_2)^2.$$

So, by Lemma A.1, there exists a constant $\mathfrak{c}_3 > 0$, not depending on $\mathbf{m}$, N and t, such that

$$B \leqslant \mathbb{E}(\|\widetilde{p}_{\mathbf{m},t} - p_{I\times J,t}\|_f^4)^{\frac{1}{2}}\mathbb{P}(\Omega_{m_1}^c)^{\frac{1}{2}}$$
$$\leqslant \mathfrak{c}_3(1 + \|p_{I\times J,t}\|_f^2)N^{3+q-\frac{p}{2}} \leqslant \frac{\mathfrak{c}_3(1 + \|p_{I\times J,t}\|_f^2)}{N}.$$

Step 3 (conclusion). By the two previous steps, there exists a constant $\mathfrak{c}_4 > 0$, not depending on $\mathbf{m}$, N and t, such that

$$\mathbb{E}(\|\widetilde{p}_{\mathbf{m},t} - p_{I\times J,t}\|_f^2) \leqslant 9\min_{\tau\in\mathcal{S}_\mathbf{m}}\|\tau - p_{I\times J,t}\|_f^2 + \frac{8m_1\mathfrak{L}_\psi(m_2)}{N} + \frac{\mathfrak{c}_4(1 + R_f(t))}{N}$$

with $R_f(t) = R(t) + \|p_{I\times J,t}\|_f^2$. Moreover, by Inequality (7),

$$R(t) = \frac{1}{T}\mathbb{E}\left[\left(\int_0^T\int_{-\infty}^{\infty}p_t(X_s, y)^2dyds\right)^2\right]^{\frac{1}{2}}$$
$$\leqslant \frac{\mathfrak{c}_T}{Tt^{1/2}}\mathbb{E}\left[\left(\int_0^T\int_{-\infty}^{\infty}p_t(X_s, y)dyds\right)^2\right]^{\frac{1}{2}} = \mathfrak{c}_Tt^{-\frac{1}{2}}$$

and

$$\|p_{I\times J,t}\|_f^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p_t(x, y)^2f(x)dxdy$$
$$\leqslant \mathfrak{c}_Tt^{-\frac{1}{2}}\int_{-\infty}^{\infty}f(x)\int_{-\infty}^{\infty}p_t(x, y)dydx = \mathfrak{c}_Tt^{-\frac{1}{2}},$$

leading to

$$\frac{1}{T}\int_0^TR_f(t)dt \leqslant 4\mathfrak{c}_TT^{-\frac{1}{2}}.$$

In conclusion,

$$\frac{1}{T}\int_0^T\mathbb{E}(\|\widetilde{p}_{\mathbf{m},t} - p_{I\times J,t}\|_f^2)dt \leqslant 9\min_{\tau\in\mathcal{S}_\mathbf{m}}\left\{\frac{1}{T}\int_0^T\|\tau - p_{I\times J,t}\|_f^2dt\right\} + \frac{8m_1\mathfrak{L}_\psi(m_2)}{N} + \frac{\mathfrak{c}_4(1 + 4\mathfrak{c}_TT^{-1/2})}{N}.$$

□

A.3. **Proof of Theorem 5.2.** The proof of Theorem 5.2 relies on the two following technical lemmas.

Lemma A.3. *Under Assumption 5.1, for any* $\mathbf{m} = (m_1, m_2)$ *and* $\mathbf{M} = (M_1, M_2)$ *belonging to* $\mathcal{N}^2$*, if* $\mathcal{S}_{\mathbf{m}} \subset \mathcal{S}_{\mathbf{M}}$*, then*

$$\widehat{\Pi}_{\mathbf{m}}(\widehat{p}_{\mathbf{M},t}) = \widehat{p}_{\mathbf{m},t}.$$

Lemma A.4. *Consider the event*

$$\Xi_N := \{\mathcal{M}_N \subset \widehat{\mathcal{M}}_N \subset \mathfrak{M}_N\},$$

where $\mathfrak{M}_N := \mathcal{U}_N \cap (\mathfrak{V}_N \times \mathcal{N})$ *and*

$$\mathfrak{V}_N := \left\{m_1 \in \mathcal{N} : \mathfrak{c}_\varphi^2 m_1(\|\Psi_{m_1}^{-1}\|_{\mathrm{op}}^2 \vee 1) \leqslant 4\mathfrak{d}\frac{NT}{\log(NT)}\right\}.$$

Under Assumption 5.1, there exists a constant $\mathfrak{c}_{A.4} > 0$*, not depending on* N*, such that*

$$\mathbb{P}(\Xi_N^c) \leqslant \frac{\mathfrak{c}_{A.4}}{N^{p-1}}.$$

The proof of Lemma A.3 is postponed to Subsubsection A.3.2, and Lemma A.4 is a straightforward consequence of Comte and Genon-Catalot [11], Inequality (6.17), because

$$\{\mathcal{V}_N \subset \widehat{\mathcal{V}}_N \subset \mathfrak{V}_N\} \subset \Xi_N.$$

A.3.1. *Steps of the proof.* The proof of Theorem 5.2 is dissected in four steps.

Step 1. Let us prove that for any $\mathbf{m} = (m_1, m_2) \in \widehat{\mathcal{M}}_N$,

$$\|\widehat{p}_{\widehat{\mathbf{m}},t} - p_{I\times J,t}\|_N^2 \leqslant 6\|\widehat{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2 + 4\kappa\mathrm{pen}(\mathbf{m}) + 11\left(\|\widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t)\|_N^2 - \frac{2}{11}\kappa\mathrm{pen}(\widehat{\mathbf{m}})\right)_+. \tag{23}$$

Consider $\mathbf{M} = (\max(\mathcal{N}), \max(\mathcal{N}))$ and $m \in \{\mathbf{m}, \widehat{\mathbf{m}}\} \subset \widehat{\mathcal{M}}_N \subset \mathcal{N}^2$. First, let us show that

$$\begin{aligned}\|\widehat{p}_{\widehat{\mathbf{m}},t} - p_{I\times J,t}\|_N^2 \leqslant{} & \|\widehat{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2 - 2(\|\widehat{p}_{\mathbf{m},t} - \widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2 - \kappa\mathrm{pen}(\mathbf{m})) \\ & + 2(\|\widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t)\|_N^2 - \kappa\mathrm{pen}(\widehat{\mathbf{m}})) + R_{\mathbf{M}}\end{aligned} \tag{24}$$

with

$$R_{\mathbf{M}} = 2\langle \widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t) - \widehat{\Pi}_{\mathbf{m}}(p_t), \widehat{p}_{\mathbf{M},t} - \widehat{\Pi}_{\mathbf{M}}(p_t)\rangle_N.$$

Since $\mathcal{S}_m \subset \mathcal{S}_{\mathbf{M}}$ by Assumption 5.1.(1,2),

$$\widehat{\Pi}_m(\widehat{p}_{\mathbf{M},t}) = \widehat{p}_{m,t} \quad \text{by Lemma A.3.}$$

Then,

$$\begin{aligned}\|\widehat{p}_{\mathbf{M},t} - \widehat{p}_{m,t}\|_N^2 &= \|\widehat{p}_{\mathbf{M},t}\|_N^2 + \|\widehat{p}_{m,t}\|_N^2 - 2\langle \widehat{p}_{\mathbf{M},t}, \widehat{p}_{m,t}\rangle_N \\ &= \|\widehat{p}_{\mathbf{M},t}\|_N^2 + \|\widehat{\Pi}_m(\widehat{p}_{\mathbf{M},t})\|_N^2 \\ &\quad -2\underbrace{\langle \widehat{p}_{\mathbf{M},t} - \widehat{\Pi}_m(\widehat{p}_{\mathbf{M},t}), \widehat{\Pi}_m(\widehat{p}_{\mathbf{M},t})\rangle_N}_{=0} - 2\langle \widehat{\Pi}_m(\widehat{p}_{\mathbf{M},t}), \widehat{\Pi}_m(\widehat{p}_{\mathbf{M},t})\rangle_N \\ &= \|\widehat{p}_{\mathbf{M},t}\|_N^2 - \|\widehat{p}_{m,t}\|_N^2.\end{aligned}$$

So, by the definition of $\widehat{\mathbf{m}}$,

$$\begin{aligned}\|\widehat{p}_{\mathbf{M},t} - \widehat{p}_{\widehat{\mathbf{m}},t}\|_N^2 + 2\kappa\mathrm{pen}(\widehat{\mathbf{m}}) &= \|\widehat{p}_{\mathbf{M},t}\|_N^2 - \|\widehat{p}_{\widehat{\mathbf{m}},t}\|_N^2 + 2\kappa\mathrm{pen}(\widehat{\mathbf{m}}) \\ &\leqslant \|\widehat{p}_{\mathbf{M},t}\|_N^2 - \|\widehat{p}_{\mathbf{m},t}\|_N^2 + 2\kappa\mathrm{pen}(\mathbf{m}) = \|\widehat{p}_{\mathbf{M},t} - \widehat{p}_{\mathbf{m},t}\|_N^2 + 2\kappa\mathrm{pen}(\mathbf{m}),\end{aligned}$$

and thus

$$\begin{aligned}&\|\widehat{p}_{\widehat{\mathbf{m}},t} - p_{I\times J,t}\|_N^2 - \|\widehat{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2 \\ &\qquad = \|\widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{p}_{\mathbf{M},t}\|_N^2 - \|\widehat{p}_{\mathbf{m},t} - \widehat{p}_{\mathbf{M},t}\|_N^2 + 2\langle \widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{p}_{\mathbf{m},t}, \widehat{p}_{\mathbf{M},t} - p_{I\times J,t}\rangle_N \\ &\qquad \leqslant 2\kappa\mathrm{pen}(\mathbf{m}) - 2\kappa\mathrm{pen}(\widehat{\mathbf{m}}) + 2\langle \widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{p}_{\mathbf{m},t}, \widehat{p}_{\mathbf{M},t} - p_{I\times J,t}\rangle_N.\end{aligned}$$

Moreover, since $\widehat{\Pi}_m \circ \widehat{\Pi}_{\mathbf{M}} = \widehat{\Pi}_m$ and $\widehat{\Pi}_m(\widehat{p}_{\mathbf{M},t}) = \widehat{p}_{m,t}$,

$$\langle \widehat{p}_{m,t} - \widehat{\Pi}_m(p_t), \widehat{p}_{\mathbf{M},t} - \widehat{\Pi}_{\mathbf{M}}(p_t)\rangle_N = \underbrace{\langle \widehat{\Pi}_m(\widehat{p}_{\mathbf{M},t} - \widehat{\Pi}_{\mathbf{M}}(p_t)), \widehat{p}_{\mathbf{M},t} - \widehat{\Pi}_{\mathbf{M}}(p_t) - \widehat{\Pi}_m(\widehat{p}_{\mathbf{M},t} - \widehat{\Pi}_{\mathbf{M}}(p_t))\rangle_N}_{=0} + \|\widehat{p}_{m,t} - \widehat{\Pi}_m(p_t)\|_N^2,$$

leading to

$$\begin{aligned}
\langle \widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{p}_{\mathbf{m},t}, \widehat{p}_{\mathbf{M},t} - p_{I\times J,t}\rangle_N &= \langle \widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{p}_{\mathbf{m},t}, \widehat{p}_{\mathbf{M},t} - \widehat{\Pi}_{\mathbf{M}}(p_t)\rangle_N + \underbrace{\langle \widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{p}_{\mathbf{m},t}, \widehat{\Pi}_{\mathbf{M}}(p_t) - p_{I\times J,t}\rangle_N}_{=0} \\
&= \langle \widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t), \widehat{p}_{\mathbf{M},t} - \widehat{\Pi}_{\mathbf{M}}(p_t)\rangle_N \\
&\quad + \langle \widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t) - \widehat{\Pi}_{\mathbf{m}}(p_t), \widehat{p}_{\mathbf{M},t} - \widehat{\Pi}_{\mathbf{M}}(p_t)\rangle_N \\
&\quad + \langle \widehat{\Pi}_{\mathbf{m}}(p_t) - \widehat{p}_{\mathbf{m},t}, \widehat{p}_{\mathbf{M},t} - \widehat{\Pi}_{\mathbf{M}}(p_t)\rangle_N \\
&= \|\widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t)\|_N^2 - \|\widehat{p}_{\mathbf{m},t} - \widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2 \\
&\quad + \langle \widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t) - \widehat{\Pi}_{\mathbf{m}}(p_t), \widehat{p}_{\mathbf{M},t} - \widehat{\Pi}_{\mathbf{M}}(p_t)\rangle_N.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|\widehat{p}_{\widehat{\mathbf{m}},t} - p_{I\times J,t}\|_N^2 &\leqslant \|\widehat{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2 + 2\kappa\text{pen}(\mathbf{m}) - 2\kappa\text{pen}(\widehat{\mathbf{m}}) + 2\langle \widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{p}_{\mathbf{m},t}, \widehat{p}_{\mathbf{M},t} - p_{I\times J,t}\rangle_N \\
&\leqslant \|\widehat{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2 - 2(\|\widehat{p}_{\mathbf{m},t} - \widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2 - \kappa\text{pen}(\mathbf{m})) \\
&\quad + 2(\|\widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t)\|_N^2 - \kappa\text{pen}(\widehat{\mathbf{m}})) + R_{\mathbf{M}}.
\end{aligned}$$

Now, let us find a suitable bound on $R_{\mathbf{M}}$. Since $\varphi_1, \dots, \varphi_{\max(\mathcal{N})}$ are linearly independent, one may consider the basis $(\varphi_1^N, \dots, \varphi_{\max(\mathcal{N})}^N)$ of $\mathcal{S}_{\varphi,\max(\mathcal{N})}$, orthonormal for the empirical inner product $\langle .,. \rangle_{N,1}$, obtained from $(\varphi_1, \dots, \varphi_{\max(\mathcal{N})})$ via the Gram-Schmidt process. Then, $(\varphi_j^N \otimes \psi_\ell)_{j,\ell}$ is an orthonormal basis of $\mathcal{S}_{\mathbf{M}}$ equipped with $\langle .,. \rangle_N$, so that

$$\widehat{\Pi}_{\mathbf{M}}(p_t) = \sum_{j=1}^{\max(\mathcal{N})} \sum_{\ell=1}^{\max(\mathcal{N})} \langle p_t, \varphi_j^N \otimes \psi_\ell\rangle_N (\varphi_j^N \otimes \psi_\ell)$$

$$\text{and} \quad \widehat{p}_{\mathbf{M},t} = \widehat{\Pi}_{\mathbf{M}}(\widehat{p}_{\mathbf{M},t}) = \sum_{j=1}^{\max(\mathcal{N})} \sum_{\ell=1}^{\max(\mathcal{N})} \langle \widehat{p}_{\mathbf{M},t}, \varphi_j^N \otimes \psi_\ell\rangle_N (\varphi_j^N \otimes \psi_\ell).$$

Thus,

$$\begin{aligned}
|R_{\mathbf{M}}| &= 2\left|\left\langle \widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t) - \widehat{\Pi}_{\mathbf{m}}(p_t), \sum_{j=1}^{m_1\vee\widehat{m}_1} \sum_{\ell=1}^{m_2\vee\widehat{m}_2} \langle \widehat{p}_{\mathbf{M},t} - p_t, \varphi_j^N \otimes \psi_\ell\rangle_N (\varphi_j^N \otimes \psi_\ell)\right\rangle_N\right| \\
&\leqslant \frac{1}{4}\|\widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t) - \widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2 + 4\left\|\sum_{j=1}^{m_1\vee\widehat{m}_1} \sum_{\ell=1}^{m_2\vee\widehat{m}_2} \langle \widehat{p}_{\mathbf{M},t} - p_t, \varphi_j^N \otimes \psi_\ell\rangle_N (\varphi_j^N \otimes \psi_\ell)\right\|_N^2 \\
&\leqslant \frac{1}{2}\|\widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t) - p_{I\times J,t}\|_N^2 + \frac{1}{2}\|\widehat{\Pi}_{\mathbf{m}}(p_t) - p_{I\times J,t}\|_N^2 + 4\sum_{j=1}^{\widehat{m}_1}\sum_{\ell=1}^{\widehat{m}_2} \langle \widehat{p}_{\mathbf{M},t} - p_t, \varphi_j^N \otimes \psi_\ell\rangle_N^2 \\
&\quad + 4\sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2} \langle \widehat{p}_{\mathbf{M},t} - p_t, \varphi_j^N \otimes \psi_\ell\rangle_N^2 \\
&= \frac{1}{2}\|\widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t) - p_{I\times J,t}\|_N^2 + \frac{1}{2}\|\widehat{\Pi}_{\mathbf{m}}(p_t) - p_{I\times J,t}\|_N^2 \\
&\quad + 4\|\widehat{\Pi}_{\widehat{\mathbf{m}}}(\widehat{p}_{\mathbf{M},t} - p_t)\|_N^2 + 4\|\widehat{\Pi}_{\mathbf{m}}(\widehat{p}_{\mathbf{M},t} - p_t)\|_N^2 \\
&= \frac{1}{2}\|\widehat{p}_{\widehat{\mathbf{m}},t} - p_{I\times J,t}\|_N^2 + \frac{1}{2}\|\widehat{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2 + \frac{7}{2}\|\widehat{p}_{\widehat{\mathbf{m}},t} - \widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t)\|_N^2 + \frac{7}{2}\|\widehat{p}_{\mathbf{m},t} - \widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2. \qquad (25)
\end{aligned}$$

Therefore, by Inequalities (24) and (25),

$$\begin{aligned}\|\widehat{p}_{\widehat{\mathbf{m}},t}-p_{I\times J,t}\|_N^2 &\leqslant 3\|\widehat{p}_{\mathbf{m},t}-p_{I\times J,t}\|_N^2+3\|\widehat{p}_{\mathbf{m},t}-\widehat{\Pi}_{\mathbf{m}}(p_t)\|_N^2\\ &\qquad +4\kappa\mathrm{pen}(\mathbf{m})+11\left(\|\widehat{p}_{\widehat{\mathbf{m}},t}-\widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t)\|_N^2-\frac{2}{11}\kappa\mathrm{pen}(\widehat{\mathbf{m}})\right)\\ &\leqslant 6\|\widehat{p}_{\mathbf{m},t}-p_{I\times J,t}\|_N^2+4\kappa\mathrm{pen}(\mathbf{m})+11\left(\|\widehat{p}_{\widehat{\mathbf{m}},t}-\widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t)\|_N^2-\frac{2}{11}\kappa\mathrm{pen}(\widehat{\mathbf{m}})\right)_+.\end{aligned}$$

Step 2. First of all, consider

$$\Omega_N:=\bigcap_{m_1\in\mathfrak{V}_N}\Omega_{m_1}.$$

For every $m_1\in\mathfrak{V}_N$,

$$\mathfrak{L}_\varphi(m_1)(\|\Psi_{m_1}\|_{\mathrm{op}}\vee 1)\leqslant\mathfrak{c}_\varphi^2 m_1(\|\Psi_{m_1}\|_{\mathrm{op}}^2\vee 1)\leqslant 4\mathfrak{d}\frac{NT}{\log(NT)}\leqslant\frac{\mathfrak{c}_\Lambda}{2}\cdot\frac{NT}{\log(NT)},$$

leading to

$$\mathbb{P}(\Omega_N^c)\leqslant\sum_{m_1\in\mathfrak{V}_N}\mathbb{P}(\Omega_{m_1}^c)\leqslant\frac{\mathfrak{c}_{A.1}}{N^{p-1}}\quad\text{by Lemma A.1 (requiring Assumption 4.1).}\tag{26}$$

Now, since $\Xi_N=\{\mathcal{M}_N\subset\widehat{\mathcal{M}}_N\subset\mathfrak{M}_N\}$, on the event $\Xi_N\cap\Omega_N$, Inequality (23) remains true for every $\mathbf{m}\in\mathcal{M}_N$. Then,

$$\begin{aligned}\|\widehat{p}_{\widehat{\mathbf{m}},t}-p_{I\times J,t}\|_N^2 &= \|\widehat{p}_{\widehat{\mathbf{m}},t}-p_{I\times J,t}\|_N^2\mathbf{1}_{\Xi_N\cap\Omega_N}+\|\widehat{p}_{\widehat{\mathbf{m}},t}-p_{I\times J,t}\|_N^2\mathbf{1}_{\Xi_N^c\cup\Omega_N^c}\\ &\leqslant \min_{\mathbf{m}\in\mathcal{M}_N}\{6\|\widehat{p}_{\mathbf{m},t}-p_{I\times J,t}\|_N^2+4\kappa\mathrm{pen}(\mathbf{m})\}\\ &\qquad +11\left(\|\widehat{p}_{\widehat{\mathbf{m}},t}-\widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t)\|_N^2-\frac{2}{11}\kappa\mathrm{pen}(\widehat{\mathbf{m}})\right)_+\mathbf{1}_{\Xi_N\cap\Omega_N}\\ &\qquad +\left(\|\widehat{p}_{\widehat{\mathbf{m}},t}-\widehat{\Pi}_{\widehat{\mathbf{m}}}(p_t)\|_N^2+\min_{\tau\in\mathcal{S}_{\widehat{\mathbf{m}}}}\|\tau-p_{I\times J,t}\|_N^2\right)(\mathbf{1}_{\Xi_N^c}+\mathbf{1}_{\Omega_N^c})\\ &\leqslant \min_{\mathbf{m}\in\mathcal{M}_N}\{6\|\widehat{p}_{\mathbf{m},t}-p_{I\times J,t}\|_N^2+4\kappa\mathrm{pen}(\mathbf{m})\}+A+B \end{aligned}\tag{27}$$

where, by Lemma A.2,

$$A:=\left(\sup_{\tau\in\mathcal{S}_{\widehat{\mathbf{m}}}:\|\tau\|_N=1}\nu_N(\tau)^2+\|p_{I\times J,t}\|_N^2\right)(\mathbf{1}_{\Xi_N^c}+\mathbf{1}_{\Omega_N^c})$$

$$\text{and}\quad B:=11\left(\sup_{\tau\in\mathcal{S}_{\widehat{\mathbf{m}}}:\|\tau\|_N=1}\nu_N(\tau)^2-\frac{2}{11}\kappa\mathrm{pen}(\widehat{\mathbf{m}})\right)_+\mathbf{1}_{\Xi_N\cap\Omega_N}.$$

Let us find suitable bounds on $\mathbb{E}(A)$ and $\mathbb{E}(B)$. On the one hand, since

$$\sup_{\tau\in\mathcal{S}_{\mathbf{m}}:\|\tau\|_N=1}\nu_N(\tau)^2\leqslant 4m_1\mathfrak{L}_\psi(m_2)\ ;\ \forall\mathbf{m}=(m_1,m_2)\in\mathcal{N}^2$$

as established in the proof of Theorem 4.2 (see Step 2), since $\mathbb{E}(\|p_{I\times J,t}\|_N^4)^{1/2}\leqslant R(t)$ as established in the proof of Theorem 4.3 (see Step 1), and by Inequality (26) and Lemma A.4,

$$\mathbb{E}(A)\leqslant\mathfrak{c}_1\left(R(t)+\sum_{m_1,m_2=1}^{\max(\mathcal{N})}m_1\mathfrak{L}_\psi(m_2)\right)(\mathbb{P}(\Omega_N^c)^{\frac{1}{2}}+\mathbb{P}(\Xi_N^c)^{\frac{1}{2}})\leqslant\frac{\mathfrak{c}_2(1+R(t))}{N},$$

where $\mathfrak{c}_1$ and $\mathfrak{c}_2$ are positive constants not depending on N and t. On the other hand, since $\widehat{\mathbf{m}}\in\mathfrak{M}_N$ on the event $\Xi_N\cap\Omega_N$,

$$\mathbb{E}(B)\leqslant 11\sum_{\mathbf{m}\in\mathfrak{M}_N}\mathbb{E}\left[\left(\sup_{\tau\in\mathcal{S}_{\mathbf{m}}:\|\tau\|_N=1}\nu_N(\tau)^2-\frac{2}{11}\kappa\mathrm{pen}(\mathbf{m})\right)_+\mathbf{1}_{\Xi_N\cap\Omega_N}\right].$$

Moreover, for every $\mathbf{m} = (m_1, m_2) \in \mathfrak{M}_N$,

$$\{\tau \in \mathcal{S}_{\mathbf{m}} : \|\tau\|_N = 1\} \subset \{\tau \in \mathcal{S}_{\mathbf{m}} : \|\tau\|_f^2 \leqslant 2\} \quad \text{on} \quad \Omega_{m_1} \supset \Omega_N.$$

Then,

$$\mathbb{E}(B) \leqslant 11 \sum_{\mathbf{m}\in\mathfrak{M}_N} \underbrace{\mathbb{E}\left[\left(\sup_{\tau\in\mathcal{S}_{\mathbf{m}}:\|\tau\|_f\leqslant 2} \nu_N(\tau)^2 - \frac{2}{11}\kappa\text{pen}(\mathbf{m})\right)_+ \mathbf{1}_{\Xi_N\cap\Omega_N}\right]}_{=:b_{\mathbf{m}}}.$$

Step 3. Thanks to the Klein and Rio version of Talagrand's inequality (see Klein and Rio [19]), this step deals with a suitable control of $b_{\mathbf{m}}$, $\mathbf{m} = (m_1, m_2) \in \mathfrak{M}_N$. First, for every $\tau = \sum_{j,\ell} T_{j,\ell}(\varphi_j \otimes \psi_\ell)$ belonging to $\mathcal{F} := \{\tau \in \mathcal{S}_{\mathbf{m}} : \|\tau\|_f^2 \leqslant 2\}$,

$$\|\tau\|^2 = \text{trace}(TT^*\Psi_{m_1}\Psi_{m_1}^{-1}) \leqslant \|\Psi_{m_1}^{-1}\|_{\text{op}} \underbrace{\text{trace}(T^*\Psi_{m_1}T)}_{=\|\tau\|_f^2} \leqslant 2\|\Psi_{m_1}^{-1}\|_{\text{op}},$$

and then

$$\begin{aligned}
\|\tau\|_\infty &= \sup_{(x,y)\in I\times J}\left\{\sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2} T_{j,\ell}\varphi_j(x)\psi_\ell(y)\right\} \\
&\leqslant \mathfrak{L}_\psi(m_2)^{\frac{1}{2}} \sup_{x\in I}\left\{\sum_{j=1}^{m_1}|\varphi_j(x)|\sum_{\ell=1}^{m_2}|T_{j,\ell}|\right\} \leqslant \sqrt{2}\mathfrak{L}_\varphi(m_1)^{\frac{1}{2}}\mathfrak{L}_\psi(m_2)^{\frac{1}{2}}\|\Psi_{m_1}^{-1}\|_{\text{op}}^{\frac{1}{2}}.
\end{aligned}$$

Since $m_1 \in \mathfrak{V}_N$, and since $\mathfrak{L}_\psi(m_2) \leqslant m_1\mathfrak{L}_\psi(m_2) \leqslant N$ by the definition of $\mathcal{U}_N$,

$$\sup_{\tau\in\mathcal{F}}\|\tau\|_\infty \leqslant M \quad \text{with} \quad M = \frac{\mathfrak{c}_3 N}{\log(N)^{1/2}} \quad \text{and} \quad \mathfrak{c}_3 = 2\sqrt{2\mathfrak{d}T}.$$

Now, since

$$\mathbb{E}\left(\sup_{\tau\in\mathcal{F}}\nu_N(\tau)^2\right) \leqslant \frac{2m_1\mathfrak{L}_\psi(m_2)}{N}$$

as established in the proof of Theorem 4.2 (see Step 1),

$$\mathbb{E}\left(\sup_{\tau\in\mathcal{F}}|\nu_N(\tau)|\right) \leqslant H \quad \text{with} \quad H^2 = \frac{2m_1\mathfrak{L}_\psi(m_2)}{N},$$

and then

$$N\sup_{\tau\in\mathcal{F}}\mathbb{E}(\nu_N(\tau)^2) \leqslant \upsilon \quad \text{with} \quad \upsilon = NH^2.$$

By the aforementioned Talagrand's inequality, for $\alpha := a\log(N)$,

$$\mathbb{E}\left[\left(\sup_{\tau\in\mathcal{F}}\nu_N(\tau)^2 - 2(1+2\alpha)H^2\right)_+\right] \leqslant A(N,\alpha,H,\upsilon) + B(N,\alpha,M,H),$$

where

$$\begin{aligned}
A(N,\alpha,H,\upsilon) &:= \frac{24\upsilon}{N}\exp\left(-\frac{\alpha NH^2}{6\upsilon}\right) \\
&= \frac{48m_1\mathfrak{L}_\psi(m_2)}{N}e^{-\frac{a}{6}\log(N)} \leqslant 48N^{-\frac{a}{6}} \quad \text{because } \mathbf{m}\in\mathcal{U}_N,
\end{aligned}$$

and

$$\begin{aligned}
B(N,\alpha,M,H) &:= \frac{\mathfrak{c}_4 M^2}{N^2}\exp\left(-\frac{\sqrt{2}}{42}\cdot\frac{NH\sqrt{\alpha}}{M}\right) \quad \text{with} \quad \mathfrak{c}_4 = 7056 \\
&= \frac{\mathfrak{c}_3^2\mathfrak{c}_4}{\log(N)}\exp\left(-\frac{\sqrt{2}}{42}\cdot\frac{\sqrt{2a}\log(N)}{\mathfrak{c}_3}\right) \leqslant \mathfrak{c}_3^2\mathfrak{c}_4 N^{-\frac{\sqrt{2a}}{84\sqrt{\mathfrak{d}T}}}.
\end{aligned}$$

Therefore, since $a \geqslant (2 \cdot 84\sqrt{\mathfrak{d}T})^2/2$, and since

$$\kappa \geqslant \kappa_0 = 44a \geqslant \frac{22(1+2a\log(N))}{1+\log(N)} = 11(1+2a\log(N))\frac{H^2}{\text{pen}(\mathbf{m})},$$

there exists a constant $\mathfrak{c}_5 > 0$, not depending on N, m_1, m_2 and t, such that

$$\begin{aligned} b_{\mathbf{m}} &\leqslant \mathbb{E}\left[\left(\sup_{\tau\in\mathcal{S}_{\mathbf{m}}:\|\tau\|_f\leqslant 2}\nu_N(\tau)^2 - \frac{2}{11}\kappa\text{pen}(\mathbf{m})\right)_+\right] \\ &\leqslant \mathbb{E}\left[\left(\sup_{\tau\in\mathcal{F}}\nu_N(\tau)^2 - 2(1+2\alpha)H^2\right)_+\right] \leqslant \frac{\mathfrak{c}_5}{N^2}. \end{aligned}$$

Step 4 (conclusion). By Steps 2 and 3,

$$\mathbb{E}(A) \leqslant \frac{\mathfrak{c}_2(1+R(t))}{N} \quad \text{and} \quad \mathbb{E}(B) \leqslant 11\sum_{\mathbf{m}\in\mathfrak{M}_N} b_{\mathbf{m}} \leqslant \frac{11\mathfrak{c}_5}{N}.$$

In conclusion, by Inequality (27), there exists a constant $\mathfrak{c}_6 > 0$, not depending on N and t, such that

$$\mathbb{E}(\|\widehat{p}_{\widehat{\mathbf{m}},t} - p_{I\times J,t}\|_N^2) \leqslant 6\min_{\mathbf{m}\in\mathcal{M}_N}\{\mathbb{E}(\|\widehat{p}_{\mathbf{m},t} - p_{I\times J,t}\|_N^2) + \kappa\text{pen}(\mathbf{m})\} + \frac{\mathfrak{c}_6(1+R(t))}{N}.$$

A.3.2. *Proof of Lemma A.3.* By the definition of $\widehat{\Pi}_{\mathbf{m}}$ (see (20)),

$$\widehat{\Pi}_{\mathbf{m}}(\widehat{p}_{\mathbf{M},t}) = \sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2}[\widehat{\Psi}_{m_1}^{-1}\widehat{P}_{\mathbf{m}}(\widehat{p}_{\mathbf{M},t})]_{j,\ell}(\varphi_j\otimes\psi_\ell)$$

with, for every $j \in \{1,\dots,m_1\}$ and $\ell \in \{1,\dots,m_2\}$,

$$\begin{aligned} \widehat{P}_{\mathbf{m}}(\widehat{p}_{\mathbf{M},t}) &= \frac{1}{NT}\sum_{i=1}^N\int_0^T\int_{-\infty}^{\infty}\widehat{p}_{\mathbf{M},t}(X_s^i,y)\varphi_j(X_s^i)\psi_\ell(y)dyds \\ &= \sum_{j'=1}^{M_1}\sum_{\ell'=1}^{M_2}[\widehat{\Psi}_{M_1}^{-1}\widehat{Z}_{\mathbf{M},t}]_{j',\ell'} \\ &\qquad\qquad \times\frac{1}{NT}\sum_{i=1}^N\int_0^T\varphi_j(X_s^i)\varphi_{j'}(X_s^i)\underbrace{\int_{-\infty}^{\infty}\psi_\ell(y)\psi_{\ell'}(y)dy}_{=\delta_{\ell,\ell'}}\,ds \\ &= \sum_{j'=1}^{M_1}[\widehat{\Psi}_{M_1}^{-1}\widehat{Z}_{\mathbf{M},t}]_{j',\ell}\underbrace{\frac{1}{NT}\sum_{i=1}^N\int_0^T\varphi_j(X_s^i)\varphi_{j'}(X_s^i)ds}_{=[\widehat{\Psi}_{M_1}]_{j,j'}} \\ &= [\widehat{Z}_{\mathbf{M},t}]_{j,\ell} = [\widehat{Z}_{\mathbf{m},t}]_{j,\ell} \quad \text{because} \quad \mathcal{S}_{\mathbf{m}}\subset\mathcal{S}_{\mathbf{M}}. \end{aligned}$$

Therefore,

$$\widehat{\Pi}_{\mathbf{m}}(\widehat{p}_{\mathbf{M},t}) = \sum_{j=1}^{m_1}\sum_{\ell=1}^{m_2}[\widehat{\Psi}_{m_1}^{-1}\widehat{Z}_{\mathbf{m},t}]_{j,\ell}(\varphi_j\otimes\psi_\ell) = \widehat{p}_{\mathbf{m},t}.$$

A.4. **Proof of Theorem 5.4.** The proof is mainly the same as that of Theorem 5.2, except for the control M of $\sup_{\tau\in\mathcal{F}}\|\tau\|_\infty$ and the control υ of $N\sup_{\tau\in\mathcal{F}}\mathbb{E}(\nu_N(\tau)^2)$ involved in Talagrand's inequality (see Step 3 in Subsubsection A.3.1).

Step 3 (bis). Thanks to the Klein and Rio version of Talagrand's inequality (see Klein and Rio [19]), this step deals with a suitable control of $b_{\mathbf{m}}$, $\mathbf{m} = (m_1, m_2) \in \mathfrak{M}_N$. First, for every $\tau = \sum_{j,\ell} T_{j,\ell}(\varphi_j \otimes \psi_\ell)$ belonging to $\mathcal{F} = \{\tau \in \mathcal{S}_{\mathbf{m}} : \|\tau\|_f^2 \leqslant 2\}$, we still get

$$\|\tau\|_\infty \leqslant \sqrt{2}\mathfrak{L}_\varphi(m_1)^{\frac{1}{2}}\mathfrak{L}_\psi(m_2)^{\frac{1}{2}}\|\Psi_{m_1}^{-1}\|_{\mathrm{op}}^{\frac{1}{2}}.$$

By Assumption 5.1.(3), and since $\mathbf{m} = (m_1, m_2)$ belongs to $\mathfrak{M}_N$,

$$\mathfrak{L}_\varphi(m_1)^{\frac{1}{2}}\|\Psi_{m_1}^{-1}\|_{\mathrm{op}}^{\frac{1}{2}} \leqslant \mathfrak{c}_\varphi^{\frac{1}{2}} m_1^{\frac{1}{4}}(\mathfrak{c}_\varphi^2 m_1\|\Psi_{m_1}^{-1}\|_{\mathrm{op}}^2)^{\frac{1}{4}} \leqslant \mathfrak{c}_\varphi^{\frac{1}{2}} m_1^{\frac{1}{4}}\left(4\mathfrak{d}\frac{NT}{\log(NT)}\right)^{\frac{1}{4}}.$$

Thus, for $NT \geqslant e$,

$$\sup_{\tau\in\mathcal{F}}\|\tau\|_\infty \leqslant M \quad \text{with} \quad M = \mathfrak{c}_{3,b} m_1^{\frac{1}{4}}\mathfrak{L}_\psi(m_2)^{\frac{1}{2}}N^{\frac{1}{4}} \quad \text{and} \quad \mathfrak{c}_{3,b} = 2\sqrt{\mathfrak{c}_\varphi}(\mathfrak{d}T)^{\frac{1}{4}}.$$

Now, since

$$\mathbb{E}\left(\sup_{\tau\in\mathcal{F}}\nu_N(\tau)^2\right) \leqslant \frac{2m_1\mathfrak{L}_\psi(m_2)}{N}$$

as established in the proof of Theorem 4.2 (see Step 1),

$$\mathbb{E}\left(\sup_{\tau\in\mathcal{F}}|\nu_N(\tau)|\right) \leqslant H \quad \text{with} \quad H^2 = \frac{2m_1\mathfrak{L}_\psi(m_2)}{N}.$$

Lastly,

$$\sup_{\tau\in\mathcal{F}}\left\{\mathrm{var}\left(\frac{1}{T}\int_0^T(\tau(X_s, X_{s+t}) - \mathbb{E}(\tau(X_s, X_{s+t})|X_s))ds\right)\right\}$$
$$\leqslant \sup_{\tau\in\mathcal{F}}\mathbb{E}\left[\left(\frac{1}{T}\int_0^T\tau(X_s, X_{s+t})ds\right)^2\right] \leqslant p_0\sup_{\tau\in\mathcal{F}}\|\tau\|_f^2 \leqslant \upsilon \quad \text{with} \quad \upsilon = 2p_0.$$

By the aforementioned Talagrand's inequality, for $\alpha = 1/4$,

$$\mathbb{E}\left[\left(\sup_{\tau\in\mathcal{F}}\nu_N(\tau)^2 - 3H^2\right)_+\right] \leqslant A(N, H, \upsilon) + B(N, M, H),$$

where

$$A(N, H, \upsilon) := \frac{24\upsilon}{N}\exp\left(-\frac{NH^2}{24\upsilon}\right) = \frac{48p_0}{N}\exp\left(-\frac{m_1\mathfrak{L}_\psi(m_2)}{24p_0}\right),$$

and

$$B(N, M, H) := \frac{\mathfrak{c}_{4,b}M^2}{N^2}\exp\left(-\frac{\sqrt{2}}{84}\cdot\frac{NH}{M}\right) \quad \text{with} \quad \mathfrak{c}_{4,b} = 7056$$
$$\leqslant \frac{\mathfrak{c}_{4,b}\mathfrak{c}_{3,b}^2}{\sqrt{N}}\exp\left(-\frac{m_1^{1/4}N^{1/4}}{42\mathfrak{c}_{3,b}}\right) \quad \text{because} \quad \mathbf{m} \in \mathcal{U}_N.$$

Since $m_1 \geqslant 1$,

$$B(N, M, H) \leqslant \frac{\mathfrak{c}_{5,b}}{\sqrt{N}}\exp(-\mathfrak{c}_{6,b}N^{\frac{1}{4}}) \quad \text{with} \quad \mathfrak{c}_{5,b} = \mathfrak{c}_{4,b}\mathfrak{c}_{3,b}^2 \quad \text{and} \quad \mathfrak{c}_{6,b} = \frac{1}{42\mathfrak{c}_{3,b}}.$$

So, for $\kappa_b \geqslant \kappa_{b,0} = 33/2$ and $t \in [t_0, T]$,

$$b_{\mathbf{m}} \leqslant \mathbb{E}\left[\left(\sup_{\tau\in\mathcal{S}_{\mathbf{m}}:\|\tau\|_f\leqslant 2}\nu_N(\tau)^2 - \frac{2}{11}\kappa\mathrm{pen}(\mathbf{m})\right)_+\right] \leqslant \mathbb{E}\left[\left(\sup_{\tau\in\mathcal{F}}\nu_N(\tau)^2 - 3H^2\right)_+\right]$$
$$\leqslant \frac{48p_0}{N}\exp\left(-\frac{m_1\mathfrak{L}_\psi(m_2)}{24p_0}\right) + \frac{\mathfrak{c}_{5,b}}{\sqrt{N}}\exp(-\mathfrak{c}_{6,b}N^{\frac{1}{4}}).$$

Therefore, by Assumption 5.3, there exists a constant $\mathfrak{c}_{7,b} > 0$, not depending on N, such that

$$\mathbb{E}(B) \leqslant 11 \sum_{\mathbf{m}\in\mathfrak{M}_N} b_{\mathbf{m}} \leqslant \frac{\mathfrak{c}_{7,b}}{N}.$$

References

[1] Abramowitz, M. and Stegun, I.A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover Publications, USA.

[2] Aït-Sahalia, Y. (1999). Transition Densities for Interest Rate and Other Nonlinear Diffusions. *The Journal of Finance* LIV, 1361-1395.

[3] Amorino, C., Gloter, A. and Halconruy, H. (2024). Evolving Privacy: Drift Parameter Estimation for Discretely Observed i.i.d. Diffusion Processes under LDP. Preprint arXiv:2401.17829v1.

[4] Bongioanni, B. and Torrea, J.L. (2006). Sobolev Spaces Associated to the Harmonic Oscillator. *Proc. Indian Acad. Sci. (math. Sci.)* 116, 337-360.

[5] Belomestny, D., Comte, F. and Genon-Catalot, V. (2019). Sobolev-Hermite versus Sobolev Nonparametric Density Estimation on $\mathbb{R}$. *Annals of the Institute of Statistical Mathematics* 71, 1, 29-62.

[6] Belomestny, D., Pilipauskaitė, V. and Podolskij, M. (2023). Semiparametric Estimation of McKean-Vlasov SDEs. *Ann. Inst. H. Poincaré Probab. Statist.* 59, 1, 79-96.

[7] Chagny, G., Meynaoui, A. and Roche, A. (2022). Adaptive Nonparametric Estimation in the Functional Linear Model with Functional Output. Preprint hal-03579232.

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

[9] Comte, F. and Genon-Catalot, V. (2018). Laguerre and Hermite Bases for Inverse Problems. *Journal of the Korean Statistical Society* 47, 273-296.

[10] Comte, F. and Genon-Catalot, V. (2020a). Regression Function Estimation as a Partly Inverse Problem. *Annals of the Institute of Statistical Mathematics* 72, 4, 1023-1054.

[11] Comte, F. and Genon-Catalot, V. (2020b). Nonparametric Drift Estimation for i.i.d. Paths of Stochastic Differential Equations. *The Annals of Statistics* 48, 6, 3336-3365.

[12] Comte, F. and Genon-Catalot, V. (2023). Nonparametric Estimation for i.i.d. Stochastic Differential Equations with Space-Time Dependent Coefficients. *SIAM/ASA Journal of Uncertainty Quantification* (accepted).

[13] Comte, F. and Lacour, C. (2023). Non Compact Estimation of the Conditional Density from Direct or Noisy Data. *Annales de l'I.H.P., Probability and Statistics* 59, 1463-1507.

[14] Della Maestra, L. and Hoffmann, M. (2022). Nonparametric Estimation for Interacting Particle Systems: McKean-Vlasov Models. *Probab. Theory Related Fields* 182, 551-613.

[15] Denis, C., Dion-Blanc, C. and Martinez, M. (2020). Consistent Procedures for Multiclass Classification of Discrete Diffusion Paths. *Scand. J. Stat.* 47, 2, 516-554.

[16] Denis, C., Dion-Blanc, C. and Martinez, M. (2021). A Ridge Estimator of the Drift from Discrete Repeated Observations of the Solution of a Stochastic Differential Equation. *Bernoulli* 27, 2675-2713.

[17] DeVore, R.A. and Lorentz, G.G. (1993). *Constructive Approximation.* Springer Berlin, Heidelberg.

[18] Halconruy, H and Marie, N. (2023). On a Projection Least Squares Estimator for Jump Diffusion Processes. *Annals of the Institute of Statistical Mathematics* 76, 2, 209-234.

[19] Klein, T. and Rio, E. (2005). Concentration Around the Mean for Maxima of Empirical Processes. *The Annals of Probability* 33, 3, 1060-1077.

[20] Lacour, C. (2007). Adaptive Estimation of the Transition Density of a Markov Chain. *Ann. Inst. H. Poincaré Probab. Statist.* 43, 571-597.

[21] Lacour, C. (2008) Nonparametric Estimation of the Stationary Density and the Transition Density of a Markov Chain. *Stochastic Process. Appl.* 118, 232-260.

[22] Lacour, C., Massart, P. and Rivoirard, V. (2017). Estimator Selection: a New Method with Applications to Kernel Density Estimation. *Sankhya* 79, 298-335.

[23] Lamberton, D. and Lapeyre, B. (2008). *Introduction to Stochastic Calculus Applied to Finance. Second Edition.* Chapman & Hall/CRC, Financial Mathematics Series.

[24] Löffler, M. and Picard, A. (2021). Spectral Thresholding for the Estimation of Markov Chain Transition Operators. *Electron. J. Stat.* 15, 6281-6310.

[25] Marie, N. (2023). Nonparametric Estimation for I.I.D. Paths of a Martingale Driven Model with Application to Non-Autonomous Financial Models. *Finance and Stochastics* 27, 1, 97-126.

[26] Marie, N. (2024). On a Computable Skorokhod's Integral Based Estimator of the Drift Parameter in Fractional SDE. *Scandinavian Journal of Statistics* (accepted).

[27] Marie, N. and Rosier, A. (2023). Nadaraya-Watson Estimator for I.I.D. Paths of Diffusion Processes. *Scandinavian Journal of Statistics* 50, 2, 589-637.

[28] Menozzi, S., Pesce, A. and Zhang, X. (2021). Density and Gradient Estimates for Non Degenerate Brownian SDEs with Unbounded Measurable Drift. *Journal of Differential Equations* 272, 330-369.

[29] Milstein, G.N., Schoenmakers, J.G.M. and Spokoiny, V. (2004). Transition Density Estimation for Stochastic Differential Equations via Forward-Reverse Representations. *Bernoulli* 10, 2, 281-312.

[30] Ramsay, J.O. and Silverman, B.W. (2007). *Applied Functional Data Analysis: Methods and Case Studies.* Springer, New-York.

[31] Sart, M. (2014). Estimation of the transition density of a Markov chain. *Ann. Inst. Henri Poincaré Probab. Stat.* 50, 1028-1068.

[32] Wang, J-L., Chiou, J-M. and Mueller, H-G. (2016). Review of Functional Data Analysis. *Annu. Rev. Stat. Appl.* 3, 257-295.

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

Email address: `nmarie@parisnanterre.fr`