Nonparametric Estimation of the Transition Density Function for Diffusion Processes

Fabienne Comte, Nicolas Marie

To cite this version:

Fabienne Comte, Nicolas Marie. Nonparametric Estimation of the Transition Density Function for Diffusion Processes. 2024. hal-04527294

HAL Id: hal-04527294
https://hal.science/hal-04527294

Preprint submitted on 29 Mar 2024

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

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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

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

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 $\sigma'$ are bounded. Under these conditions on $b$ and $\sigma$, 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 $X_{t+s}$ given $X_s = x$. The question of estimating such a function, for discrete samples of one path of $X$, say $X_k\Delta$, $k = 1, \ldots, n$ 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 $i \in \{1, \ldots, N\}$, where $\mathcal{I}(\cdot)$ is the Itô map for Equation (1) and $W^1, \ldots, W^N$ are $N \in \mathbb{N}^*$ independent copies of $W$. What we propose is a projection least squares estimator $\hat{p}_{m,t}$ of the transition density function $p_t$ of $X$ at time $t$ computed from $X^1, \ldots, X^N$, where $m = (m_1, m_2) \in \{1, \ldots, N\}^2$ and $N_T := [NT] + 1$. Consider $S_m := S_{x,m_1} \otimes S_{\psi,m_2}$, where $S_{x,m_1}$ (resp. $S_{\psi,m_2}$) is a $m_1$-dimensional (resp. $m_2$-dimensional) function space defined later. The estimator $\hat{p}_{m,t}$ is defined as minimizing over functions $\tau \in S_m$ the objective function

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

First, a nonadaptive risk bound is established on $\hat{p}_{m,t}$, 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 $\hat{p}_{m,t}$, where the couple of dimensions $m = (\hat{m}_1, \hat{m}_2)$ is selected in a random subset of $\{1, \ldots, N_T\}^2$ 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

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

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

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

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

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

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

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

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 $\varphi(X^{0,x}_T)$, which price $P(x, T)$ satisfies

$$P(x, T) = e^{-rT} \mathbb{E}(\varphi(X^{0,x}_T))) = e^{-rT} \int_{-\infty}^{\infty} \varphi(y)p_T(x, y) dy.$$
Thus, when \( r < \delta \), an estimator of \( P(x, T) \) is given by
\[
\hat{P}_m(x, T) := e^{-rT} \int_0^\infty \varphi(y) \hat{P}_m(x, y) dy,
\]
where the copies \( X^1, \ldots, X^N \) of \( (X_t^{0, x})_{t \in [0, T]} \) 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 \( X^1, \ldots, X^N \) of the solution \( X \) of Equation (1) under the condition:
\[
(5) \quad x_0 \in \mathbb{R}, \; W = (W_t)_{t \in [0, T]} \text{ is a Brownian motion, } b, \sigma \in C^1(\mathbb{R}) \text{ and } b', \sigma' \text{ are bounded.}
\]
As already mentioned, under condition (5), Equation (1) has a unique solution. We also assume that \( \sigma \) satisfies the following non-degeneracy condition:
\[
(6) \quad \exists \alpha, A > 0 : \forall x \in \mathbb{R}, \; \alpha \leq |\sigma(x)| \leq \alpha + A.
\]
Under the condition (6), the transition density function \( p_t \) is well-defined and, for every \( x, y \in \mathbb{R} \),
\[
(7) \quad p_t(x, y) \leq \epsilon_T t^{-\frac{d}{2}} \exp \left( \frac{-m_T (y - x)^2}{t} \right),
\]
where \( \epsilon_T \) and \( m_T \) are positive constants depending on \( T \) but not on \( t \) and \( y \) (see Menozzi et al. [28], Theorem 1.2). In particular, \( t \mapsto p_t(x_0, x) \) belongs to \( L^1([0, T]) \), which legitimizes 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 \( x \in \mathbb{R} \),
  \[
f(x) \leq \frac{1}{T} \int_0^T \epsilon_T s^{-\frac{d}{2}} ds = 2\epsilon_T T^{-\frac{d}{2}}.
\]
- Since \( b' \) is bounded (and then \( b \) has linear growth),
  \[
  |b(x)|^\kappa \in L^2(\mathbb{R}, f(x) dx) \; ; \forall \kappa \in \mathbb{R}_+.
  \]
Indeed, for every \( v \in \mathbb{R}_+ \) and \( t \in (0, T] \),
\[
E(|X|^\nu) = \int_{-\infty}^\infty |x|^\nu p_t(x_0, x) dx \\
\leq t^{-\frac{d}{2}} \epsilon_T \int_{-\infty}^\infty |x|^\nu \exp \left( \frac{-m_T (x - x_0)^2}{T} \right) dx \\
= \epsilon_T \kappa T^{-\frac{d}{2}}
\]
and then, for every \( \kappa \in \mathbb{R}_+ \),
\[
\int_{-\infty}^\infty |b(x)|^{2\kappa} f(x) dx = \frac{1}{T} \int_0^T E(|b(X_s)|^{2\kappa}) ds \\
\leq \epsilon_1 \left( 1 + \frac{1}{T} \int_0^T E(|X_s|^{2\kappa}) ds \right) \leq \epsilon_1 (1 + 2\epsilon_T, 2\kappa T^{-\frac{d}{2}}) < \infty,
\]
where \( \epsilon_1 \) is a positive constant depending only on \( b \) and \( \kappa \).
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 \( S_m = S_{\phi,m_1} \otimes S_{\psi,m_2} \), where \( S_{\phi,m_1} := \text{span}\{\phi_1, \ldots, \phi_m\} \) (resp. \( S_{\psi,m_2} := \text{span}\{\psi_1, \ldots, \psi_m\} \)), \( \phi_1, \ldots, \phi_N \) (resp. \( \psi_1, \ldots, \psi_N \)) are continuous functions from \( I \) (resp. \( J \)) into \( \mathbb{R} \) such that \( \{\phi_1, \ldots, \phi_N, \psi_1, \ldots, \psi_N\} \) is an orthonormal family in \( L^2(I, dx) \) (resp. \( L^2(J, dx) \)), and \( I, J \subseteq \mathbb{R} \) are non-empty intervals. Then, for any \( \tau \in S_m \), consider

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

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

\[
E(\gamma_N(\tau)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\tau(x, y) - p_t(x, y))^2 f(x) dx dy - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_t(x, y)^2 f(x) dx dy.
\]

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

Let us show that \( \gamma_N \) has a unique minimizer in \( S_m \). For \( \tau = \sum_{j,t} \Theta_{j,t}(\phi_j \otimes \psi_t) \) with \( \Theta \in \mathcal{M}_{m_1,m_2}(\mathbb{R}) \),

\[
\nabla_{\tau} \gamma_N(\tau) = 2(\hat{\Psi}_{m_1} \Theta - \hat{Z}_{m,t}),
\]

where

\[
\hat{\Psi}_{m_1} := \left( \frac{1}{NT} \sum_{i=1}^N \int_0^T \phi_j(X_i^s)\phi_j(X_i^s) ds \right)_{j,j' \in \{1, \ldots, m_1\}}
\]

and

\[
\hat{Z}_{m,t} := \left( \frac{1}{NT} \sum_{i=1}^N \int_0^T \phi_j(X_i^s)\psi_{t+s} ds \right)_{(j,t) \in \{1, \ldots, m_1\} \times \{1, \ldots, m_2\}}.
\]

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

\[
y^t \hat{\Psi}_{m_1} y = \frac{1}{NT} \sum_{i=1}^N \int_0^T \left( \sum_{j=1}^{m_1} y_j \phi_j(X_i^s) \right)^2 ds \geq 0.
\]

If in addition \( \hat{\Psi}_{m_1} \) is invertible, it is positive definite, and then

(8) \[
\hat{p}_{m,t} = \sum_{j=1}^{m_1} \sum_{t=1}^{m_2} \left[ \hat{\Theta}_{m,t} \right]_{j,t}(\phi_j \otimes \psi_t) \quad \text{with} \quad \hat{\Theta}_{m,t} = \hat{\Psi}_{m_1}^{-1} \hat{Z}_{m,t}
\]

is the only minimizer of \( \gamma_N \) in \( S_m \) called the projection least squares estimator of \( p_t \).

Related definitions/notations:

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

\[
\langle \phi, \psi \rangle_N := \frac{1}{NT} \sum_{i=1}^N \int_0^T \int_{-\infty}^{\infty} \phi(X_i^s, y)\psi(X_i^s, y) dy ds.
\]

The empirical norm associated to \( \langle \cdot, \cdot \rangle_N \) is denoted by \( \|\cdot\|_N \).

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

\[
\langle \phi, \psi \rangle_f := E(\langle \phi, \psi \rangle_N) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x, y)\psi(x, y) f(x) dx dy.
\]

3. The theoretical counterpart of \( \hat{\Psi}_{m_1} \) is denoted by \( \Psi_{m_1} := E(\hat{\Psi}_{m_1}) \).
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 $\| \cdot \|_N$, on our estimator $\hat{p}_{m,t}$ of $p_t$. First, consider

$$\mathcal{L}_\varphi(m) := 1 \vee \left( \sup_{x \in I} \sum_{j=1}^{m} \phi_j(x)^2 \right)$$

and

$$\mathcal{L}_\psi(m) := 1 \vee \left( \sup_{x \in I} \sum_{j=1}^{m} \psi_j(x)^2 \right)$$

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

Assumption 4.1. There exist two constants $\epsilon_{4,1} > 0$ and $q \in \mathbb{N}^*$, not depending on $m$ and $N$, such that

$$\mathcal{L}_\psi(m_2) \leq \epsilon_{4,1} N^q.$$

Moreover,

$$\mathcal{L}_\varphi(m_1) (\| \Psi^{-1}_{m_1} \|_{op} \vee 1) \leq \frac{c_\Lambda}{2} \cdot \frac{N T}{\log(N T)}$$

with

$$c_\Lambda = \frac{1 - \log(2)}{(1 + p) T} \quad \text{and} \quad p = 2(q + 4) + 1.$$

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 $p_{1 \times J,t} := (p_t)_{1 \times J}$. Under Assumption 4.1, there exists a constant $\epsilon_{4,2} > 0$, not depending on $m$ and $N$, such that for every $t \in [0, T]$,

$$E(\| \hat{p}_{m,t} - p_{1 \times J,t} \|_N^2) \leq \min_{\tau \in S_m} \| \tau - p_{1 \times J,t} \|_f^2 + \frac{2m_1 \mathcal{L}_\psi(m_2)}{N} + \frac{\epsilon_{4,2}}{N}.$$

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

$$\min_{\tau \in S_m} \| \tau - p_{1 \times J,t} \|_f^2$$

in Inequality (10) is 1,

and the constant 2 in the variance term $2m_1 \mathcal{L}_\psi(m_2)/N$ 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 $(\psi_1, \ldots, \psi_{m_2})$. First, for splines, wavelets or trigonometric bases, $\mathcal{L}_\psi(m_2) \lesssim m_2$, and then the variance term in the risk bound on $\hat{p}_{m,t}$ is of order $m_1 m_2 / N$ as for the usual projection estimator of a 2-dimensional density function. Now, for Legendre’s basis, $\mathcal{L}_\psi(m_2) \lesssim m_2^2$, leading to a variance term of order $m_1 m_2^2 / N$ in Inequality (10). Finally, let us focus on the Hermite basis $(h_j)_{j \in \mathbb{N}}$, defined on $I = \mathbb{R}$ by

$$h_j(x) := \epsilon_j H_j(x) e^{-x^2} \quad \text{with} \quad \epsilon_j = (2^j j! \sqrt{\pi})^{-\frac{1}{2}} \quad \text{and} \quad H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j} e^{-x^2}$$

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

$$\mathcal{L}_h(m) \leq c_h \sqrt{m} ; \quad \forall m \in \mathbb{N}.$$

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 \( \| \cdot \|_f \), on the following truncated version of our estimator \( \tilde{p}_{m,t} \):

\[
\tilde{p}_{m,t} := \tilde{p}_{m,t} 1_{\Lambda_{m_1}},
\]

where

\[
\Lambda_{m_1} := \left\{ \mathcal{L}_\varphi(m_1)(\| \hat{\psi}_{m_1}^{-1} \|_{\text{op}} \vee 1) \leq c_\Lambda \frac{NT}{\log(NT)} \right\}.
\]

On the event \( \Lambda_{m_1} \), \( \hat{\psi}_{m_1} \) is invertible because

\[
\inf\{\text{sp}(\hat{\psi}_{m_1})\} \geq \frac{\mathcal{L}_\varphi(m_1)}{c_\Lambda}, \frac{\log(NT)}{NT},
\]

and then \( \tilde{p}_{m,t} \) is well-defined.

**Theorem 4.3.** Under Assumption 4.1,

(1) There exists a constant \( c_{4.3.1} > 0 \), not depending on \( m \) and \( N \), such that for every \( t \in (0, T] \),

\[
\mathbb{E}(\| \tilde{p}_{m,t} - p_{I \times J,t} \|_f^2) \leq 9 \min_{\tau \in \mathbb{S}_m} \| \tau - p_{I \times J,t} \|_f^2 + \frac{8m_1 \mathcal{L}_\psi(m_2)}{N} + \frac{c_{4.3.1}(1 + R_f(t))}{N}.
\]

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

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

(2) There exists a constant \( c_{4.3.2} > 0 \), not depending on \( m \) and \( N \), such that

\[
\frac{1}{T} \int_0^T \mathbb{E}(\| \tilde{p}_{m,t} - p_{I \times J,t} \|_f^2) dt \leq 9 \min_{\tau \in \mathbb{S}_m} \left\{ \frac{1}{T} \int_0^T \| \tau - p_{I \times J,t} \|_f^2 dt \right\} + \frac{8m_1 \mathcal{L}_\psi(m_2)}{N} + \frac{c_{4.3.2}}{N}.
\]

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

\[
p_t(x, y) \geq c_T t^{-\frac{1}{2}} \exp \left( -\frac{m_T (y - x)^2}{t} \right),
\]

where \( c_T \) and \( m_T \) 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,

\[
R(t) \geq \frac{1}{T} \int_0^T \int_{-\infty}^\infty \mathbb{E}(p_t(X_s, y)^2 dy dx = \int_{-\infty}^\infty \int_{-\infty}^\infty p_t(x, y)^2 f(x) dx dy
\]

\[
\geq \frac{c_T^2}{T} \int_{-\infty}^\infty f(x) \int_{-\infty}^\infty \exp \left( -2m_T \frac{(y - x)^2}{t} \right) dy dx
\]

\[
= \frac{c_T^2}{T} \int_{-\infty}^\infty f(x) dx \int_{-\infty}^\infty \exp \left( -2m_T \left( \frac{y}{T} \right)^2 \right) dy
\]

\[
= \frac{c_T^2}{T^{1/2}} \int_{-\infty}^\infty e^{-2m_T y^2} dy \xrightarrow{t \to 0^+} \infty.
\]

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 \( k = (k_1, \ldots, k_d) \in \mathbb{N}^d \) and \( s = (s_1, \ldots, s_d) \in (0, \infty)^d \), \( k^s := k_1^{s_1} \times \cdots \times k_d^{s_d} \).
- Throughout this subsection, both \( S_{p, m_1} \) and \( S_{p, m_2} \) are generated by the Hermite basis (see Subsection 4.1), one can establish the following proposition.

**Definition 4.4.** (Sobolev-Hermite ellipsoids) The Sobolev-Hermite ellipsoid \( W_s^{(d)}(A, L) \) of order \( s \) and of radius \( L \) is defined by
\[
W_s^{(d)}(A, L) := \left\{ g \in L^2(A) : \sum_{k \in \mathbb{N}^d} a_k(g)^2 k^s \leq L \right\}.
\]

Now, by assuming that \( p_t \) belongs to \( W_s^{(2)}(A, L) \), the bias term in Theorems 4.2 and 4.3 decreases to 0 with polynomial rate. Indeed, noting \( p_{m,l} \) the orthogonal projection of \( p_t \) on \( S_{m,l} \),
\[
\| p_t - p_{m,l} \|^2 = \sum_{k \in \mathbb{N}^d : \exists q \in \{1, 2\}, k_q \geq m_q} a_k(p_t)^2 \\
\leq \sum_{q=1}^2 \sum_{\mathbf{k} \in \mathbb{N}^d : k_q \geq m_q} a_k(p_t)^2 k_q^{s_q} k_q^{-s_q} \leq L(m_1^{-s_1} + m_2^{-s_2})
\]
and then, since \( f(\cdot) \leq 2cT\tau^{-1/2} \),
\[
\min_{\tau \in \mathbb{S}_m} \| \tau - p_{\mathbb{S}_m} \|_f^2 \leq \| f \|_\infty \| p_t - p_{m,l} \|_f^2 \leq \frac{2cL}{\sqrt{T}} (m_1^{-s_1} + m_2^{-s_2}).
\]

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 \( S_{p, m_2} \) 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 \( W_s^{(2)}(A, L) \), and consider \( m^* = (m_1^*, m_2^*) \) with
\[
m_1^* \propto N^{\frac{1}{1+\frac{s_1}{s_2}+\frac{s_2}{s_1}}} \quad \text{and} \quad m_2^* \propto N^{\frac{1}{1+\frac{s_1}{s_2}+\frac{s_2}{s_1}}}.
\]

Then,
\[
\mathbb{E}(\| \tilde{p}_{m^*,t} - p_t \|_X^2) = O \left( N^{-\frac{1}{1+\frac{s_1}{s_2}+\frac{s_2}{s_1}}} \right),
\]
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 \( \varphi_j \)'s and the \( \psi_j \)'s fulfill the following additional assumption.

**Assumption 5.1.** The \( \varphi_j \)'s and the \( \psi_j \)'s fulfill the three following conditions:

1. For every \( m_1, M_1 \in \{1, \ldots, N_T\} \), if \( M_1 > m_1 \), then \( S_{\varphi, m_1} \subset S_{\varphi, M_1} \).
For every \( m_2, M_2 \in \{1, \ldots, N_T\} \), if \( M_2 > m_2 \), then \( S_{\phi, m_2} \subset S_{\phi, M_2} \).

(3) There exists a constant \( \varepsilon_0 > 1 \), not depending on \( N \), such that
\[
\mathcal{L}_\varepsilon(m_1) \leq \varepsilon_0^2 m_1 \; \forall m_1 \in \{1, \ldots, N_T\}.
\]

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
\[
\tilde{m} = \arg \min_{m \in \mathcal{M}_N} \left\{ -\|\tilde{p}_{m, t}\|_N^2 + 2\kappa \text{pen}(m) \right\},
\]
where \( \kappa \geq \kappa_0, \kappa_0 > 0 \) is defined later,
\[
\text{pen}(m) := (1 + \log(N)) \frac{m_1 \mathcal{L}_\varepsilon(m_2)}{N} \; \forall m = (m_1, m_2) \in \{1, \ldots, N_T\}^2,
\]
and \( \tilde{\mathcal{M}}_N := \mathcal{U}_N \cap (\tilde{\mathcal{V}}_N \times N) \) with \( \mathcal{V}_N = \{1, \ldots, N \wedge N_T\} \)
\[
\mathcal{U}_N = \{(m_1, m_2) \in N^2 : m_1 \mathcal{L}_\varepsilon(m_2) \leq N\},
\]
\[
\tilde{\mathcal{V}}_N = \left\{ m_1 \in \mathcal{N} : \varepsilon_0^2 m_1 (\|\tilde{\Psi}_m^{-1}\|_{\text{op}} \vee 1) \leq \frac{1}{2} \cdot \frac{NT}{\log(NT)} \right\}
\]
and
\[
\theta = \min \left( \frac{1}{8c_2^2 T^{\|f\|_\infty} + (3\varepsilon_0)^{-1} \sqrt{\varepsilon_0/(1 + p)}}, \frac{\varepsilon_0}{8} \right).
\]

Note that for every \( m = (m_1, m_2) \in \mathcal{U}_N, m_2 \) 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 \( \tilde{\mathcal{M}}_N \), where
\[
\mathcal{V}_N := \left\{ m_1 \in \mathcal{N} : \varepsilon_0^2 m_1 (\|\Psi_m^{-1}\|_{\text{op}} \vee 1) \leq \theta \cdot \frac{NT}{\log(NT)} \right\}.
\]
The following theorem provides a risk bound on the adaptive estimator \( \tilde{p}_{m, t} \).

**Theorem 5.2.** Under Assumption 5.1, for \( \kappa_0 = 44a \) and \( a \geq (2 \cdot 84\sqrt{T})^2/2 \),

(1) There exists a constant \( c_{5,2,1} > 0 \), not depending on \( N \), such that for every \( t \in [0, T] \),
\[
\mathbb{E}(\|\tilde{p}_{m, t} - p_{t \times J, t}\|_N^2) \leq \min_{m \in \mathcal{M}_N} \left\{ \mathbb{E}(\|\tilde{p}_{m, t} - p_{t \times J, t}\|_N^2) + \kappa \text{pen}(m) \right\} + \frac{c_{5,2,1}(1 + R(t))}{N}.
\]

(2) There exists a constant \( c_{5,2,2} > 0 \), not depending on \( N \), such that
\[
\frac{1}{T} \int_0^T \mathbb{E}(\|\tilde{p}_{m, t} - p_{t \times J, t}\|_N^2)dt \leq \min_{m \in \mathcal{M}_N} \left\{ \frac{1}{T} \int_0^T \mathbb{E}(\|\tilde{p}_{m, t} - p_{t \times J, t}\|_N^2)dt + \kappa \text{pen}(m) \right\} + \frac{c_{5,2,2}}{N}.
\]

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

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
\[
\sup_{(x, y) \in t \times J} p_t(x, y) \leq p_0 := c_T t_0^{-\frac{1}{2}} \quad \text{by Inequality (7)},
\]
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:
\[
\tilde{m} = \arg \min_{m \in \mathcal{M}_N} \left\{ -\|\tilde{p}_{m, t}\|_N^2 + 2\kappa \text{pen}(m) \right\},
\]
for every \( \xi > 0 \), there exists \( S(\xi) > 0 \), not depending on \( N \), such that
\[
\sum_{1 \leq \xi m_1, m_2 \leq N} \exp(-\xi m_1 E (\xi m_2)) \leq S(\xi) < \infty.
\]

Note that if \( E (\xi m_2) = 1 \) \( (\xi m_2) \), or even if \( E (\xi m_2) = 1 \) \( (\xi m_2) \), then \( (\xi m_2) \) fulfills Assumption 5.3 because
\[
\sum_{m_1 = 1}^{N} \exp(-\xi E (\xi m_2)) m_1 \leq \frac{e^{-\xi E (\xi m_2)}}{1 - e^{-\xi}}; \forall m_2 \in \mathbb{N}^*.
\]

So, for instance, \( (\xi m_2) \) fulfills Assumption 5.3 when \( E (\xi m_2) \) is generated by the trigonometric basis or by Hermite’s one.

**Theorem 5.4.** Under Assumptions 5.1 and 5.3, for \( \kappa_b = 16.5 \), there exists a constant \( \epsilon_{5,4} > 0 \), not depending on \( N \), such that for every \( t \in [t_0, T] \) with \( t_0 > 0 \),
\[
\mathbb{E}(\| \hat{p}_{m,t} - p_{I \times J,t} \|_N^2) \leq 6 \min_{m \in \mathcal{M}_N} \{ \mathbb{E}(\| \hat{p}_{m,t} - p_{I \times J,t} \|_N^2) + \kappa_b \text{pen}_b(m) \} + \frac{\epsilon_{5,4}(1 + t_0^{-1/2})}{N}.
\]

Now, let us briefly present the second interesting special case by assuming that \( I \) is a compact interval. By Inequality (13), there exists \( m > 0 \) such that \( f(t) \geq m \), and then
\[
\| \psi_{m_1} \|_\infty = \frac{1}{\lambda_{\min}(\psi_{m_1})} = \left( \inf_{\| f \|_2 = 1} \sum_{j,j' = 1}^{m} \theta_j \theta_j' [\psi_{m_1}]_{j,j'} \right)^{-1} \leq \frac{1}{m} \quad \text{for every } m_1 \in \mathcal{N}.
\]

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:
\[
\hat{m}^* = \arg \min_{m \in \mathcal{M}_N} \{ -\| \hat{p}_{m,t} \|_N^2 + 2 \kappa_b \text{pen}_b(m) \}
\]

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

where \( \mathcal{M}_N := \mathcal{U}_N \cap (\mathcal{V}_N \times \mathcal{N}) \) and
\[
\mathcal{V}_N := \left\{ m_1 \in \mathcal{N} : m_1 \leq \frac{\epsilon_{5,4}}{2 \epsilon_{5,4} (m_1^{-1} \vee 1)}, \frac{NT}{\log(NT)} \right\}.
\]

A result similar to Theorem 5.2 (resp. Theorem 5.4) may be established on the adaptive estimator \( \hat{p}_{m^*,t} \) (resp. \( \hat{p}_{m^*,t} \)) by taking \( \mathcal{M}_N = \mathcal{M}_N = \mathcal{M}_N^* \) 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: $\text{pen}_b(m) = m_1 \sqrt{m_2}/N$. 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 $\Psi_{m_1}^{-1}$ 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 $U_1, \ldots, U_N$:

\begin{equation}
  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),
\end{equation}

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

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

In all cases, we take $k \in \{0, \ldots, n\}$ 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

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

**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

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

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). $N = 200$, $T = 10$, $\Delta = 0.01$, $t = 1$.

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

Example 3. (Cox-Ingersoll-Ross or square-root process) $X_i(t) = \|U_i(t)\|_d^2$, 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 $b(x) = \frac{d\gamma^2}{4} - rx$, $\sigma(x) = \gamma \sqrt{x}$, $r = 1$ and $\gamma = 1$.

Here, the transition density function is given by

$$p_t^{(3)}(x, y) = c_t \exp(-c_t(xe^{-rt} + y))$$

$$\times \left(\frac{y}{xe^{-rt}}\right)^{\frac{d-1}{2}} I_{\frac{d-1}{2}} \left(d, 2c_t \sqrt{xy} e^{-rt}\right),$$

where $c_t := \frac{2r}{\gamma^2 (1-e^{-rt})}$,

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

\[
MISE = \frac{1}{K} \sum_{k=1}^{K} \frac{DXY^{(k)}}{N_I N_J} \sum_{i=1}^{N_I} \sum_{j=1}^{N_J} \left( p_t(x_i^{(k)}, y_j^{(k)}) - \hat{p}_t(x_i^{(k)}, y_j^{(k)}) \right)^2,
\]

where

\[
DXY^{(k)} := (b_X^{(k)} - a_X^{(k)}) (b_Y^{(k)} - a_Y^{(k)}),
\]

\[
b_X^{(k)} \text{ and } a_X^{(k)} \text{ are the 98% and 2% quantiles of the } X_i^{(k)} \text{'s, and } b_Y^{(k)} \text{ and } a_Y^{(k)} \text{ are the 99% and 1% quantiles of the } X_{i+1}^{(k)} \text{'s. The super-index } k \text{ denotes the repetition number, the points } (x_i^{(k)}, y_j^{(k)}) \text{ are equispaced in the range of the observations of the path } k, \text{ and we take } N_I = N_J = 100 \text{ and } K = 200. \text{ We also give the associated standard deviation, together with the mean of the selected dimensions in each}
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 \( \hat{m}_1, \hat{m}_2 \).

7. Concluding remarks

In this paper, for a fixed \( t \in (0, T] \), we have proposed a least squares contrast estimator \( \hat{p}_{m,t}(x,y) \) 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 \( S_m = S_{\phi,m_1} \otimes S_{\psi,m_2} \), \( m = (m_1, m_2) \). In our observation setting, \( N \) independent copies \( X_1, \ldots, X_N \) of \( X \) are available. We provide an upper bound on a risk defined as the expectation of the empirical or integrated distance between \( \hat{p}_{m,t} \) and \( p_t \), exhibiting a squared-bias/variance decomposition up to negligible terms.

An adaptive procedure is then tailored to automatically select the couple \( m = (m_1, m_2) \), 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 \( i = 1, \ldots, N \), of \( X^i(t_j) + \eta_{i,j} \) with \( t_j = j \Delta/n \), \( j = 1, \ldots, n \) 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.
A.1. Proofs of Theorem 4.2. The proof of Theorem 4.2 relies on two technical lemmas stated first.

Lemma A.1. Consider the event

$$\Omega_{m_1} := \left\{ \sup_{\tau \in S_{m_1}} \left| \frac{\|\tau\|_{N,1}}{\|\tau\|_{f,1}} - 1 \right| \leq \frac{1}{2} \right\}$$

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

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

Under Assumption 4.1, there exists a constant $c_{A,1} > 0$, not depending on $m_1$ and $N$, such that

$$\mathbb{P}(\Omega_{m_1}) \leq \frac{c_{A,1}}{Np} \quad \text{and} \quad \mathbb{P}(\Omega_{m_1}^c) \leq \frac{c_{A,1}}{Np} \quad \text{with} \quad p = 2(q + 4) + 1.$$  

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

- The empirical process $\nu_N$, defined by

$$\nu_N(\tau) := \frac{1}{NT} \sum_{i=1}^{N} \int_0^T \left( \tau(X^i_s, X^{i+1}_s) - \int_{-\infty}^{\infty} \tau(x, y) p_t(x, y) dy \right) \, ds$$

for every $\tau \in S_m$. Note that

$$|\tilde{Z}_{m,1}]_{j, \ell} = (p_t \psi_j \otimes \psi_\ell)_N + \nu_N(\varphi \otimes \psi_\ell)$$

for every $j \in \{1, \ldots, m_1\}$ and $\ell \in \{1, \ldots, m_2\}$.

- The empirical orthogonal projection $\hat{\Pi}_{m}$, defined by

$$\hat{\Pi}_{m}(\cdot) \in \arg \min_{\tau \in S_m} \|\tau - \cdot\|_N^2.$$  

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

$$\hat{\Pi}_{m}(h) = \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \tilde{\varphi}_{-1, j, \ell} \tilde{\varphi}_m(h) \cdot (\varphi_j \otimes \psi_\ell)$$

with

$$\tilde{P}_m(h) = (\langle h, \varphi_j \otimes \psi_\ell \rangle_N)_{(j, \ell) \in \{1, \ldots, m_1\} \times \{1, \ldots, m_2\}}.$$

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

$$\|\hat{\Pi}_{m,t} - \hat{\Pi}_m(p_t)\|_N^2 = \sup_{\tau \in S_m: \|\tau\|_N = 1} \nu_N(\tau)^2.$$  

The proof of Lemma A.2 is postponed to Subsubsection A.1.1.

A.1.1. Steps of the proof. First of all, by the definition of $\hat{\Pi}_m$ (see (20)),

$$\|\hat{\Pi}_{m,t} - \hat{\Pi}_m(p_t)\|_N^2 = \min_{\tau \in S_m} \|\tau - \hat{\Pi}_{m,t}\|_N^2 + \|\hat{\Pi}_{m,t} - \hat{\Pi}_m(p_t)\|_N^2,$$

and by Lemma A.2,

$$\mathbb{E}(\|\hat{\Pi}_{m,t} - \hat{\Pi}_m(p_t)\|_N^2) = \mathbb{E} \left( 1_{\Omega_{m_1}} \sup_{\tau \in S_m: \|\tau\|_N = 1} \nu_N(\tau)^2 \right) + \mathbb{E} \left( 1_{\Omega_{m_1}} \sup_{\tau \in S_m: \|\tau\|_N = 1} \nu_N(\tau)^2 \right) =: A + B.$$  

Let us find suitable bounds on $A$ and $B$.

**Step 1 (bound on $A$).** First, since

$$\|\tau\|_{f,1}^2 1_{\Omega_{m_1}} \leq 2 \|\tau\|_{N,1}^2 1_{\Omega_{m_1}} \quad \text{on} \quad S_{m_1},$$


by the definition of $\Omega_{m_1}$, for every $\tau \in S_m$, 
\[
\|\tau\|_{\Omega_{m_1}}^2 = \int_{-\infty}^{\infty} \|\tau(\cdot, y)\|_{f,1}^2 1_{\Omega_{m_1}} dy \leq 2 \int_{-\infty}^{\infty} \|\tau(\cdot, y)\|_{\mathcal{N},1}^2 1_{\Omega_{m_1}} dy = 2\|\tau\|_{\mathcal{N}}^2 1_{\Omega_{m_1}}.
\]

Then,
\[
\{\tau \in S_m : \|\tau\|_N = 1\} \subset \{\tau \in S_m : \|\tau\|_f^2 \leq 2\} \quad \text{on} \quad \Omega_{m_1},
\]
leading to
\[
A \leq \mathbb{E} \left( \sup_{\tau \in S_m, \|\tau\|_f^2 \leq 2} \nu_N(\tau)^2 \right).
\]

Since $(\varphi_1, \ldots, \varphi_{m_1})$ is an orthonormal family of $L^2(\mathbb{R}, dx)$, $\varphi_1, \ldots, \varphi_{m_1}$ are linearly independent, and one may consider the basis $(\varphi'_1, \ldots, \varphi'_{m_1})$ of $S_{\varphi,m_1}$, orthonormal in $L^2(\mathbb{R}, f(x) dx)$, obtained from $(\varphi_1, \ldots, \varphi_{m_1})$ via the Gram-Schmidt process. Thus, $(\varphi'_j \otimes \psi_{j})_{j,\ell}$ is an orthonormal basis of $S_m$ equipped with $\langle .. \rangle_f$, and noting $\|..\|_{2,m}$ the Fröbenius norm on $M_{m_1,m_2}(\mathbb{R})$,
\[
A \leq \mathbb{E} \left[ \sup_{\Theta : \|\Theta\|_{2,m} \leq 2} \left( \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \Theta_{j,\ell} \nu_N (\varphi'_j \otimes \psi_{\ell}) \right)^2 \right] \leq 2 \mathbb{E} \left( \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \nu_N (\varphi'_j \otimes \psi_{\ell})^2 \right).
\]

Now, note that for every $j \in \{1, \ldots, m_1\}$ and $\ell \in \{1, \ldots, m_2\}$,
\[
\mathbb{E}(\nu_N (\varphi'_j \otimes \psi_{\ell})) = \frac{1}{NT} \mathbb{E} \left( \sum_{i=1}^{N} \int_{0}^{T} \varphi'_j (X^i_s)(\psi_{\ell}(X^i_{s+t}) - \mathbb{E}(\psi_{\ell}(X^i_{s+t})|X^i_s)) ds \right)
\]
\[
= \frac{1}{T} \int_{0}^{T} \mathbb{E}(\varphi'_j (X_s)(\psi_{\ell}(X_{s+t}) - \mathbb{E}(\psi_{\ell}(X_{s+t})|X_s)) ds = 0.
\]

Therefore, since $\|\varphi'_j\|_f = 1$ for every $j \in \{1, \ldots, m_1\}$,
\[
A \leq 2 \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \text{var}(\nu_N (\varphi'_j \otimes \psi_{\ell}))
\]
\[
\leq \frac{2}{N} \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \mathbb{E} \left[ \left( \frac{1}{T} \int_{0}^{T} \varphi'_j (X_s)(\psi_{\ell}(X_{s+t}) - \mathbb{E}(\psi_{\ell}(X_{s+t})|X_s)) ds \right)^2 \right]
\]
\[
\leq \frac{2}{NT} \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \int_{0}^{T} \left( \mathbb{E}(\varphi'_j (X_s)^2 \psi_{\ell}(X_{s+t})^2) + \mathbb{E}(\varphi'_j (X_s)^2 \mathbb{E}(\psi_{\ell}(X_{s+t})|X_s)^2) \right)
\]
\[
- 2\mathbb{E}(\varphi'_j (X_s)\psi_{\ell}(X_{s+t})\mathbb{E}(\varphi'_j (X_s)\psi_{\ell}(X_{s+t})|X_s)) ds \leq \frac{2m_1 L_{\psi}(m_2)}{N}.
\]

**Step 2 (bound on B).** Since $\varphi_1, \ldots, \varphi_{m_1}$ are linearly independent as mentioned in Step 1, one may consider the basis $(\varphi^N_1, \ldots, \varphi^N_{m_1})$ of $S_{\varphi,m_1}$, orthonormal for the empirical inner product $\langle .. \rangle_{N,1}$, obtained from $(\varphi_1, \ldots, \varphi_{m_1})$ via the Gram-Schmidt process. Then, $(\varphi^N_j \otimes \psi_{\ell})_{j,\ell}$ is an orthonormal basis of $S_m$.
equipped with $(\cdot, \cdot)_N$, and by Cauchy-Schwarz’s and Jensen’s inequalities,
\[
\sup_{\tau \in S_m, \|\tau\|_N = 1} \nu_N(\tau)^2 = \sup_{\Theta \in \Theta^2, \Theta_{m,1} = 1} \left( \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \Theta_{j,\ell} \nu_N(\varphi_j \otimes \psi_\ell) \right)^2 \\
\leq \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \left( \frac{1}{N^T} \sum_{i=1}^{N} \int_0^T \varphi_j^N(X_i)(\psi_\ell(X_{i+t})) - E[\psi_\ell(X_{i+t})|X_i]) ds \right)^2 \\
\leq 4 \mathcal{L}_\psi(m_2) \sum_{j=1}^{m_1} \left( \frac{1}{N^T} \sum_{i=1}^{N} \int_0^T \varphi_j^N(X_i)^2 ds \right) = 4m_1 \mathcal{L}_\psi(m_2).
\]

Therefore, by Assumption 4.1 and Lemma A.1, there exists a constant $c_1 > 0$, not depending on $m$, $N$ and $t$, such that
\[
B \leq 4m_1 \mathcal{L}_\psi(m_2) \mathbb{P}(\Omega_{m_1}^c) \leq c_1 N^{-1}.
\]

**Step 3 (conclusion).** By the two previous steps,
\[
\mathbb{E}(\|\hat{\mu}_{m,t} - \tilde{\mu}_m(p_t)\|_N^2) \leq \frac{2m_1 \mathcal{L}_\psi(m_2)}{N} + \frac{c_1}{N}.
\]

Thus,
\[
\mathbb{E}(\|\hat{\mu}_{m,t} - p_{I \times J,t}\|_N^2) \leq \min_{\tau \in S_m} \mathbb{E}(\|\tau - p_{I \times J,t}\|_N^2) + \mathbb{E}(\|\hat{\mu}_{m,t} - \tilde{\mu}_m(p_t)\|_N^2) \\
\leq \min_{\tau \in S_m} \|\tau - p_{I \times J,t}\|_N^2 + \frac{2m_1 \mathcal{L}_\psi(m_2)}{N} + \frac{c_1}{N}.
\]

\[\square\]

A.1.2. **Proof of Lemma A.2.** First, by Cauchy-Schwarz’s inequality,
\[
\sup_{\tau \in S_m, \|\tau\|_N = 1} (\hat{\mu}_{m,t} - \tilde{\mu}_m(p_t), \tau)_N^2 \leq \|\hat{\mu}_{m,t} - \tilde{\mu}_m(p_t)\|_N^2,
\]
and since
\[
\tau^* := \frac{\hat{\mu}_{m,t} - \tilde{\mu}_m(p_t)}{\|\hat{\mu}_{m,t} - \tilde{\mu}_m(p_t)\|_N} \in S_m
\]
satisfies both $\|\tau^*\|_N = 1$ and
\[
(\hat{\mu}_{m,t} - \tilde{\mu}_m(p_t), \tau^*)_N^2 = \|\hat{\mu}_{m,t} - \tilde{\mu}_m(p_t)\|_N^2,
\]
then
\[
(21) \quad \|\hat{\mu}_{m,t} - \tilde{\mu}_m(p_t)\|_N^2 = \sup_{\tau \in S_m, \|\tau\|_N = 1} (\hat{\mu}_{m,t} - \tilde{\mu}_m(p_t), \tau)_N^2.
\]

Now, let us show that
\[
(22) \quad (\hat{\mu}_{m,t} - \tilde{\mu}_m(p_t), \tau)_N = \nu_N(\tau) \quad \forall \tau \in S_m.
\]

By the decompositions of $\hat{\mu}_{m,t}$ and $\tilde{\mu}_m(p_t)$ in the basis $(\varphi_j \otimes \psi_\ell)_{j,\ell}$ of $S_m$, and by (19),
\[
\hat{\mu}_{m,t} - \tilde{\mu}_m(p_t) = \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \hat{\Psi}^{-1}_{m_1,\ell} (\hat{\Delta}_{m_1,\ell} - \tilde{\mu}_m(p_t))_{j,\ell}(\varphi_j \otimes \psi_\ell) \\
= \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \hat{\Psi}^{-1}_{m_1,\ell} (\varphi_j \otimes \psi_\ell),
\]
where
where
\[
\widehat{\Delta}_m := \left( \frac{1}{NT} \sum_{i=1}^N \int_0^T \varphi_f(X_s^i)\psi_t(X_{s+t}^i)ds \right)_{j,\ell} - \frac{1}{NT} \sum_{i=1}^N \int_0^T \int_{-\infty}^{\infty} \varphi_f(X_s^i)\psi_t(y)p_t(X_s^i,y)dyds \right)_{j,\ell} = (\nu_N(\varphi_f \otimes \psi_t))_{j,\ell}.
\]

Then, for any \( \tau = \sum_{j,\ell} \Theta_{j,\ell}(\varphi_f \otimes \psi_t) \) with \( \Theta \in M_{m_1,m_2}(\mathbb{R}) \),
\[
(\bar{p}_{m,t} - \bar{\Pi}_m(p_0), \tau)_N = \sum_{j,j'=1}^{m_1} \sum_{\ell,\ell'=1}^{m_2} \left[ \tilde{\Psi}_{m_1}^{-1}\widehat{\Delta}_m \right]_{j,j',\ell,\ell'} \Theta_{j',\ell'} (\varphi_f \otimes \psi_t) \langle \varphi_f \otimes \psi_t, \varphi_f' \otimes \psi_t' \rangle_N
\]
\[
= \sum_{j,j'=1}^{m_1} \sum_{\ell,\ell'=1}^{m_2} \left[ \tilde{\Psi}_{m_1}^{-1}\widehat{\Delta}_m \right]_{j,j',\ell,\ell'} \Theta_{j',\ell'}
\]
\[
= \sum_{j,j'=1}^{m_1} \Theta_{j',\ell'} \nu_N(\varphi_f \otimes \psi_t) = \nu_N(\tau).
\]

Therefore, by Equalities (21) and (22) together,
\[
\|\bar{p}_{m,t} - \bar{\Pi}_m(p_0)\|_N^2 = \sup_{\tau \in S_m, \|\tau\|_N=1} \nu_N(\tau)^2.
\]

A.2. Proof of Theorem 4.3. First of all,
\[
\mathbb{E}(\|\bar{p}_{m,t} - P_{I \times J,t} \|_N^2) = \mathbb{E}(\|\bar{p}_{m,t} - P_{I \times J,t} \|_N^2 1_{\Omega_m}) + \mathbb{E}(\|\bar{p}_{m,t} - P_{I \times J,t} \|_N^2 1_{\Omega_m^c})
\]
\[
= A + B.
\]

Let us find suitable bounds on \( A \) and \( B \).

**Step 1 (bound on \( A \)).** As already mentioned, since
\[
\|\cdot\|_{N,1}^2 \Omega_m \leq 2\|\cdot\|_{N,1}^2 \Omega_{m_1}
\]
on \( S_{\varphi,m_1} \)
by the definition of \( \Omega_{m_1} \), for every \( \tau \in S_m \),
\[
\|\tau\|_{N,1}^2 \Omega_m = \int_{-\infty}^{\infty} \|\tau(\cdot, y)\|_{N,1}^2 \Omega_{m_1} dy \leq 2 \int_{-\infty}^{\infty} \|\tau(\cdot, y)\|_{N,1}^2 \Omega_{m_1} dy = 2\|\tau\|_{N}^2 \Omega_{m_1}.
\]

Then, noting \( P_{I \times J,t} \) as the orthogonal projection of \( P_{I \times J,t} \) onto \( S_m \) for the inner product \( \langle \cdot, \cdot \rangle_f \),
\[
\|\bar{p}_{m,t} - P_{I \times J,t} \|_{N,1}^2 = \langle \|\bar{p}_{m,t} - P_{I \times J,t} \|_f^2 + \|P_{I \times J,t} - P_{I \times J,t} \|_f^2 \rangle \Omega_{m_1}
\]
\[
\leq \|P_{I \times J,t} - P_{I \times J,t} \|_f^2 + 2\|\bar{p}_{m,t} - P_{I \times J,t} \|_N^2 \Omega_{m_1}
\]
\[
\leq \min_{\tau \in S_m} \|\tau - P_{I \times J,t} \|_f^2 + 4\|\bar{p}_{m,t} - P_{I \times J,t} \|_N^2 + 4\|P_{I \times J,t} - P_{I \times J,t} \|_N^2.
\]
Moreover,
\[ \|\tilde{P}_{m,t} - P_{I \times J,t}\|_N^2 \leq \|\tilde{P}_{m,t} - P_{I \times J,t}\|_N^2 + \|P_{I \times J,t}\|_N^2 - 1 \]
and, by Lemma A.1,
\[
E(\|P_{I \times J,t}\|_N^2 - 1) \leq \mathbb{E} \left[ \left( \frac{1}{N^T} \sum_{i=1}^N \int_0^T \int_{-\infty}^\infty p_t(X_i^*, t, y)^2 dy dt \right)^2 \right]^{\frac{1}{2}} p(Ac) \frac{R(t)}{N^{p/2}} \]
with \( R(t) = \frac{1}{T} \mathbb{E} \left[ \left( \int_0^T \int_{-\infty}^\infty p_t(X, t, y)^2 dy dt \right)^2 \right]^{\frac{1}{2}}. \)

Therefore, by Theorem 4.2, there exists a constant \( c_1 > 0 \), not depending on \( m, N \) and \( t \), such that
\[
A \leq 5 \min_{\tau \in \mathbb{S}_m} \|\tau - P_{I \times J,t}\|_f^2 + 4\mathbb{E}(\|\tilde{P}_{m,t} - P_{I \times J,t}\|_N^2) \leq 9 \min_{\tau \in \mathbb{S}_m} \|\tau - P_{I \times J,t}\|_f^2 + \frac{8m_1 \mathbb{L}_\psi(m_2)}{N} + c_1 (1 + R(t)) \frac{5}{N}.
\]

**Step 2 (bound on \( B \)).** Since \( \Theta_{m,t} \) is a positive semidefinite (symmetric) matrix,\n\[
\|\tilde{P}_{m,t}\|_f^2 = \int_{-\infty}^\infty f(x) \int_{-\infty}^\infty \left( \sum_{j=1}^{m_1} \sum_{t=1}^{m_2} [\Theta_{m,t}]_{j,t} \varphi_j(x) \psi_t(y) \right)^2 dy dx
\]
\[
= \int_{-\infty}^\infty f(x) \sum_{j,j'=1}^{m_1} \sum_{t,t'=1}^{m_2} [\Theta_{m,t}]_{j,t} [\Theta_{m,t}^*]_{j',t'} \varphi_j(x) \varphi_{j'}(x) \psi_t(y) \psi_{t'}(y) dy dx
\]
\[
= \sum_{j,j'=1}^{m_1} \sum_{t,t'=1}^{m_2} [\Theta_{m,t}]_{j,t} [\Theta_{m,t}^*]_{j',t'} \int_{-\infty}^\infty \varphi_j(x) \varphi_{j'}(x) f(x) dx
\]
\[
= \sum_{j=1}^{m_1} \sum_{t=1}^{m_2} [\Theta_{m,t}]_{j,t} \varphi_j(x) \int_{-\infty}^\infty \varphi_j(x) f(x) dx
\]

Let us find suitable controls on \( \|
\]
\[
\|\Psi_{m_1}\|_{op}^2 = \sup_{\|\psi\|_2 = 1} \left( \sum_{j,j'=1}^{m_1} \sum_{t,t'=1}^{m_2} \varphi_j(x) \varphi_{j'}(x) f(x) dx \right)^2
\]
\[
\leq \sum_{j,j'=1}^{m_1} \left( \int_{-\infty}^\infty \varphi_j(x) \varphi_{j'}(x) f(x) dx \right)^2 \leq \mathbb{L}_\psi(m_1)^2 \int_{-\infty}^\infty f(x) dx.
\]

On the other hand, since \( \tilde{Z}_{m,t} \) is a positive semidefinite (symmetric) matrix,
\[
\text{trace}(\tilde{\Theta}_{m,t} \tilde{\Theta}_{m,t}^*) = \text{trace}(\tilde{\Psi}_{m_1}^{-1} \tilde{Z}_{m,t} \tilde{Z}_{m,t}^*) \leq \|
\]
\[
\text{trace}(\tilde{\Theta}_{m,t} \tilde{\Theta}_{m,t}^*) = \text{trace}(\tilde{\Psi}_{m_1}^{-1} \tilde{Z}_{m,t} \tilde{Z}_{m,t}^*) \leq \|
\]


Moreover,
\[
E(\text{trace}(\tilde{Z}_{m,t}\tilde{Z}_{m,t}^*))^2) = E\left[\left(\sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} [\tilde{Z}_{m,t}]_{j,\ell}^2 \right)^2\right] \\
\leq \frac{m_1 m_2}{N^2 T^4} \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} E\left[\left(\sum_{i=1}^N \int_0^T \varphi_j(X_i^t) \psi_\ell(X_i^{t+s}) ds\right)^4\right] \leq m_1 m_2 \mathcal{L}_\varphi(m_1)^2 \mathcal{L}_\psi(m_2)^2 
\]
and, on the event \(\Lambda_{m_1},\)
\[
\|\hat{\psi}_{m_1}^{-1}\|^4_{\text{op}} \leq c_4 \frac{N^4 T^4}{\log(N T)^4 \mathcal{L}_\varphi(m_1)^4}.
\]
Then, there exists a constant \(c_2 > 0\), not depending on \(m, N\) and \(t\), such that
\[
E(\|\tilde{\rho}_{m,t}\|_1^4) \leq c_2 N^4 m_1 m_2 \mathcal{L}_\psi(m_2)^2.
\]
So, by Lemma A.1, there exists a constant \(c_3 > 0\), not depending on \(m, N\) and \(t\), such that
\[
B \leq \mathbb{E}(\|\tilde{\rho}_{m,t} - p_{I \times J,t}\|_1^4) \mathbb{P}(\Omega_{m_1}^c)^\frac{1}{2} \\
\leq c_3 (1 + \|p_{I \times J,t}\|_1^2)^{N^4+q-\frac{3}{2}} \leq \frac{c_3 (1 + \|p_{I \times J,t}\|_1^2)}{N}. 
\]

**Step 3 (conclusion).** By the two previous steps, there exists a constant \(c_4 > 0\), not depending on \(m, N\) and \(t\), such that
\[
E(\|\tilde{\rho}_{m,t} - p_{I \times J,t}\|_1^4) \leq 9 \min_{\tau \in S_m} \|\tau - p_{I \times J,t}\|_1^2 + \frac{8 m_1 \mathcal{L}_\psi(m_2)}{N} + \frac{c_4 (1 + R_f(t))}{N} 
\]
with \(R_f(t) = R(t) + \|p_{I \times J,t}\|_1^2\). Moreover, by Inequality (7),
\[
R(t) = \frac{1}{T} \mathbb{E}\left[\left(\int_0^T \int_{-\infty}^{\infty} p_t(X_s, y)^2 dy dx\right)^{\frac{1}{2}}\right] \leq \frac{c_T}{T^{1/2}} \mathbb{E}\left[\left(\int_0^T \int_{-\infty}^{\infty} p_t(X_s, y) dy dx\right)^{\frac{1}{2}}\right] = c_T t^{-\frac{1}{2}} 
\]
and
\[
\|p_{I \times J,t}\|_1^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_t(x, y)^2 f(x) dx dy \\
\leq c_T t^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{\infty} p_t(x, y) dy dx = c_T t^{-\frac{1}{2}}, 
\]
leading to
\[
\frac{1}{T} \int_0^T R_f(t) dt \leq 4 c_T T^{-\frac{1}{2}}.
\]
In conclusion,
\[
\frac{1}{T} \int_0^T \mathbb{E}(\|\tilde{\rho}_{m,t} - p_{I \times J,t}\|_1^4) dt \leq 9 \min_{\tau \in S_m} \left\{ \frac{1}{T} \int_0^T \|\tau - p_{I \times J,t}\|_1^2 dt \right\} + \frac{8 m_1 \mathcal{L}_\psi(m_2)}{N} + \frac{c_4 (1 + 4 c_T T^{-1/2})}{N}.
\]
\[\square\]
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 \(m = (m_1, m_2)\) and \(M = (M_1, M_2)\) belonging to \(\mathcal{N}^2\), if \(S_m \subset S_M\), then

\[
\tilde{H}_m(\tilde{M}_{m,t}) = \tilde{H}_m. 
\]

**Lemma A.4.** Consider the event

\[
\Xi_N := \{M_N \subset \tilde{M}_N \subset \mathcal{N}_N\},
\]

where \(\mathcal{N}_N := \mathcal{U}_N \cap (\mathcal{G}_N \times \mathcal{N})\) and

\[
\mathcal{G}_N := \left\{m_1 \in \mathcal{N} : c^2 m_1(\|\Psi_{m_1}^{-1}\|^2 \vee 1) \leq \frac{40 NT}{\log(NT)} \right\}.
\]

Under Assumption 5.1, there exists a constant \(c_{A.4} > 0\), not depending on \(N\), such that

\[
P(\Xi_N^c) \leq \frac{c_{A.4}}{N^p}. \]

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

\[
\{V_N \subset \tilde{V}_N \subset \mathcal{G}_N\} \subset \Xi_N.
\]

**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 \(m = (m_1, m_2) \in \tilde{M}_N\),

\[
\|\tilde{p}_{m,t} - p_{I \times J,t}\|_N^2 \leq 6\|\tilde{p}_{m,t} - p_{I \times J,t}\|_N^2 + 4\text{pen}(m) + 11 \left(\|\tilde{p}_{m,t} - \tilde{H}_m(p_t)\|_N^2 - \frac{2}{\Pi} \text{pen}(\tilde{m})\right).
\]

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

\[
\|\tilde{p}_{m,t} - p_{I \times J,t}\|_N^2 \leq \|\tilde{p}_{m,t} - p_{I \times J,t}\|_N^2 - 2(\|\tilde{p}_{m,t} - \tilde{H}_m(p_t)\|_N^2 - \text{pen}(m)) + 2(\|\tilde{p}_{m,t} - \tilde{H}_m(p_t)\|_N^2 - \text{pen}(\tilde{m})) + R_M
\]

with

\[
R_M = 2(\tilde{H}_m(p_t) - \tilde{H}_m(p_t), \tilde{p}_{M,t} - \tilde{H}_m(p_t))_N.
\]

Since \(S_m \subset S_M\) by Assumption 5.1.1,2,

\[
\tilde{H}_m(\tilde{p}_{M,t}) = \tilde{p}_{m,t} \quad \text{by Lemma A.3.}
\]

Then,

\[
\|\tilde{p}_{M,t} - \tilde{p}_{m,t}\|_N^2 = \|\tilde{p}_{M,t}\|_N^2 + \|\tilde{p}_{m,t}\|_N^2 - 2(\tilde{p}_{M,t}, \tilde{p}_{m,t})_N
= \|\tilde{p}_{M,t}\|_N^2 + \|\tilde{H}_m(\tilde{p}_{M,t})\|_N^2
- 2(\tilde{p}_{M,t} - \tilde{H}_m(\tilde{p}_{M,t}), \tilde{H}_m(\tilde{p}_{M,t}))_N - 2(\tilde{H}_m(\tilde{p}_{M,t}), \tilde{H}_m(\tilde{p}_{M,t}))_N = 0
\]

So, by the definition of \(\tilde{m}\),

\[
\|\tilde{p}_{M,t} - \tilde{p}_{m,t}\|_N^2 + 2\text{pen}(\tilde{m}) = \|\tilde{p}_{M,t}\|_N^2 - \|\tilde{p}_{m,t}\|_N^2 + 2\text{pen}(\tilde{m})
\leq \|\tilde{p}_{m,t}\|_N^2 - \|\tilde{p}_{m,t}\|_N^2 + 2\text{pen}(m) = \|\tilde{p}_{M,t} - \tilde{p}_{m,t}\|_N^2 + 2\text{pen}(m),
\]

and thus

\[
\|\tilde{p}_{M,t} - p_{I \times J,t}\|_N^2 \leq \|\tilde{p}_{m,t} - p_{I \times J,t}\|_N^2
= \|\tilde{p}_{m,t} - \tilde{p}_{M,t}\|_N^2 + \|\tilde{p}_{m,t} - \tilde{p}_{M,t}\|_N^2 + 2(\tilde{p}_{m,t} - \tilde{p}_{M,t}, \tilde{p}_{M,t} - p_{I \times J,t})_N
\leq 2\text{pen}(m) - 2\text{pen}(\tilde{m}) + 2(\tilde{p}_{m,t} - \tilde{p}_{M,t}, \tilde{p}_{M,t} - p_{I \times J,t})_N.
\]
Moreover, since \( \tilde{\Pi}_m \circ \tilde{\Pi}_M = \tilde{\Pi}_m \) and \( \tilde{\Pi}_m(\tilde{\Pi}_M(t)) = \tilde{\Pi}_m(t) \),

\[
\langle \tilde{p}_{m,t} - \tilde{\Pi}_M(p_t), \tilde{\Pi}_M(p_t), \tilde{\Pi}_M(p_t) - \tilde{\Pi}_m(p_t) \rangle_N + \| \tilde{p}_{m,t} - \tilde{\Pi}_m(p_t) \|_N^2,
\]

leading to

\[
\langle \tilde{p}_{m,t} - \tilde{\Pi}_M(p_t), \tilde{\Pi}_M(p_t), \tilde{\Pi}_M(p_t) - \tilde{\Pi}_m(p_t) \rangle_N = \langle \tilde{p}_{m,t} - \tilde{\Pi}_M(p_t), \tilde{\Pi}_M(p_t), \tilde{\Pi}_M(p_t) - \tilde{\Pi}_m(p_t) \rangle_N + \| \tilde{p}_{m,t} - \tilde{\Pi}_m(p_t) \|_N^2,
\]

and

\[
\| \tilde{p}_{m,t} - \tilde{\Pi}_m(p_t) \|_N^2 + 2 \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 \leq \| \tilde{p}_{m,t} - \tilde{\Pi}_m(p_t) \|_N^2 + 2 \kappa \text{pen}(\tilde{m}) + 2 \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2.
\]

Therefore,

\[
\| \tilde{p}_{m,t} - \tilde{\Pi}_m(p_t) \|_N^2 + 2 \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 \leq \| \tilde{p}_{m,t} - \tilde{\Pi}_m(p_t) \|_N^2 + 2 \kappa \text{pen}(\tilde{m}) + 2 \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2.
\]

Now, let us find a suitable bound on \( R_{\Pi} \). Since \( \varphi_1, \ldots, \varphi_{\text{max}(N)} \) are linearly independent, one may consider the basis \( \varphi_{1,\ldots,\min(N)} \) of \( S_{\varphi,\text{max}(N)} \), orthonormal for the empirical inner product \( \langle \cdot, \cdot \rangle_N \), obtained from \( \langle \varphi_1, \ldots, \varphi_{\text{max}(N)} \rangle \) via the Gram-Schmidt process. Then, \( \varphi_{j,\ell}^N \otimes \psi_{j,\ell} \) is an orthonormal basis of \( S_M \) equipped with \( \langle \cdot, \cdot \rangle_N \), so that

\[
\tilde{\Pi}_M(p_t) = \sum_{j=1}^{\text{max}(N)} \sum_{\ell=1}^{\text{max}(N)} \langle \varphi_{j,\ell}, \varphi_{j,\ell}^N \rangle_N (\varphi_{j,\ell}^N \otimes \psi_{j,\ell})
\]

Thus,

\[
| R_{\Pi} | = 2 \left( \left| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_m(p_t) \right| \sum_{j=1}^{m_1 \wedge m_2} \sum_{\ell=1}^{m_2 \wedge m_2} \langle \varphi_{j,\ell}, \varphi_{j,\ell}^N \otimes \psi_{j,\ell} \rangle_N (\varphi_{j,\ell}^N \otimes \psi_{j,\ell}) \right)^\frac{1}{2}
\]

\[
= \frac{1}{4} \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_m(p_t) \|_N^2 + 2 \left( \sum_{j=1}^{m_1 \wedge m_2} \sum_{\ell=1}^{m_2 \wedge m_2} \langle \varphi_{j,\ell}, \varphi_{j,\ell}^N \otimes \psi_{j,\ell} \rangle_N (\varphi_{j,\ell}^N \otimes \psi_{j,\ell}) \right)^2
\]

\[
= \frac{1}{2} \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 + \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_m(p_t) \|_N^2
\]

\[
= \frac{1}{2} \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 + \frac{1}{2} \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 + 4 \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \langle \varphi_{j,\ell}, \varphi_{j,\ell}^N \otimes \psi_{j,\ell} \rangle_N (\varphi_{j,\ell}^N \otimes \psi_{j,\ell}) \right)^2
\]

\[
= \frac{1}{2} \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 + \frac{1}{2} \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 + 4 \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 + 4 \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2
\]

\[
= \frac{1}{2} \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 + \frac{1}{2} \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 + \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 + \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2
\]

(25)

\[
= \frac{1}{2} \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 + \frac{1}{2} \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 + \frac{7}{2} \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2 + \frac{7}{2} \| \tilde{\Pi}_m(p_t) - \tilde{\Pi}_M(p_t) \|_N^2.
\]
Therefore, by Inequalities (24) and (25),
\[ \|\hat{p}_{m,t} - p_{I \times J \times t}\|_N^2 \leq 3\|\hat{p}_{m,t} - p_{I \times J \times t}\|_N^2 + 3\|\hat{p}_{m,t} - \hat{p}_{m}(p_t)\|_N^2 \]
\[ + \kappa\text{pen}(m) + 11 \left( \|\hat{p}_{m,t} - \hat{p}_{m}(p_t)\|_N^2 - \frac{2}{11} \kappa\text{pen}(\hat{m}) \right) \]
\[ \leq 6\|\hat{p}_{m,t} - p_{I \times J \times t}\|_N^2 + 4\kappa\text{pen}(m) + 11 \left( \|\hat{p}_{m,t} - \hat{p}_{m}(p_t)\|_N^2 - \frac{2}{11} \kappa\text{pen}(\hat{m}) \right) . \]

**Step 2.** First of all, consider
\[ \Omega_N := \bigcap_{m_1 \in \mathfrak{M}_N} \Omega_{m_1} . \]
For every \( m_1 \in \mathfrak{M}_N \),
\[ \mathcal{L}_\nu(m_1)(\|\Psi_{m_1}\|_{\text{op}} \lor 1) \leq c_2^2 m_1(\|\Psi_{m_1}\|_{\text{op}} \lor 1) \leq 40 \frac{NT}{\log(NT)} \leq \frac{c_4}{2} \cdot \frac{NT}{\log(NT)} , \]
leading to
\[ (26) \quad \mathbb{P}(\Omega_N) \leq \sum_{m_1 \in \mathfrak{M}_N} \mathbb{P}(\Omega_{m_1}) \leq \frac{c_4.1}{N^{p-1}} \quad \text{by Lemma A.1 (requiring Assumption 4.1)} . \]

Now, since \( \Xi_N = \{ M_N \subset \hat{M}_N \subset \mathfrak{M}_N \} \), on the event \( \Xi_N \cap \Omega_N \), Inequality (23) remains true for every \( \hat{m} \in \mathfrak{M}_N \). Thus,
\[ \|\hat{p}_{m,t} - p_{I \times J \times t}\|_N^2 = \|\hat{p}_{m,t} - p_{I \times J \times t}\|_N^2 1_{\Xi_N \cap \Omega_N} + \|\hat{p}_{m,t} - p_{I \times J \times t}\|_N^2 1_{\Xi_N \cap \Omega_N} \]
\[ \leq \min_{\hat{m} \in \mathcal{M}_N} \left\{ 6\|\hat{p}_{m,t} - p_{I \times J \times t}\|_N^2 + 4\kappa\text{pen}(m) \right\} \]
\[ + 11 \left( \|\hat{p}_{m,t} - \hat{p}_{m}(p_t)\|_N^2 - \frac{2}{11} \kappa\text{pen}(\hat{m}) \right) + 1_{\Xi_N \cap \Omega_N} \]
\[ + \left( \|\hat{p}_{m,t} - \hat{p}_{m}(p_t)\|_N^2 + \min_{\tau \in S_m} \|\tau - p_{I \times J \times t}\|_N^2 \right) (1_{\Xi_N} + 1_{\Omega_N}) \]
\[ (27) \quad \leq \min_{\hat{m} \in \mathcal{M}_N} \left\{ 6\|\hat{p}_{m,t} - p_{I \times J \times t}\|_N^2 + 4\kappa\text{pen}(m) \right\} + A + B \]
where, by Lemma A.2,
\[ A := \left( \sup_{\tau \in S_m \mid \tau \parallel N = 1} \nu_N(\tau)^2 + \|p_{I \times J \times t}\|_N^2 \right) (1_{\Xi_N} + 1_{\Omega_N}) \]
and \( B := 11 \left( \sup_{\tau \in S_m \mid \tau \parallel N = 1} \nu_N(\tau)^2 - \frac{2}{11} \kappa\text{pen}(\hat{m}) \right) + 1_{\Xi_N \cap \Omega_N} . \]

Let us find suitable bounds on \( \mathbb{E}(A) \) and \( \mathbb{E}(B) \). On the one hand, since
\[ \sup_{\tau \in S_m \mid \tau \parallel N = 1} \nu_N(\tau)^2 \leq 4m_1 \mathcal{L}_\nu(m_2) \quad \forall m = (m_1, m_2) \in \mathcal{N}^2 \]
as established in the proof of Theorem 4.2 (see Step 2), since \( \mathbb{E}((p_{I \times J \times t})^4_N)^{1/2} \leq R(t) \) as established in the proof of Theorem 4.3 (see Step 1), and by Inequality (26) and Lemma A.4,
\[ \mathbb{E}(A) \leq c_1 \left( R(t) + \sum_{m_1, m_2 = 1}^{\max(N)} m_1 \mathcal{L}_\nu(m_2) \right) (\mathbb{P}(\Omega_N)^{1/2} + \mathbb{P}(\Xi_N)^{1/2}) \leq \frac{c_2 (1 + R(t))}{N} , \]
where \( c_1 \) and \( c_2 \) are positive constants not depending on \( N \) and \( t \). On the other hand, since \( \hat{m} \in \mathfrak{M}_N \) on the event \( \Xi_N \cap \Omega_N \),
\[ \mathbb{E}(B) \leq 11 \sum_{\hat{m} \in \mathfrak{M}_N} \mathbb{E} \left[ \left( \sup_{\tau \in S_m \mid \tau \parallel N = 1} \nu_N(\tau)^2 - \frac{2}{11} \kappa\text{pen}(\hat{m}) \right) + 1_{\Xi_N \cap \Omega_N} \right] . \]
Moreover, for every \( m = (m_1, m_2) \in \mathfrak{M}_N \),
\[
\{ \tau \in \mathcal{S}_m : \| \tau \|_N = 1 \} \subset \{ \tau \in \mathcal{S}_m : \| \tau \|_f^2 \leq 2 \} \quad \text{on} \quad \Omega_m \supset \Omega_N.
\]
Then,
\[
E(B) \leq 11 \sum_{m \in \mathfrak{M}_N} E \left[ \left( \sup_{\tau \in \mathcal{S}_m : \| \tau \|_f \leq 2} \nu_N(\tau)^2 \right) \cdot \frac{2}{11} \kappa_{\text{pen}}(m) \right] \cdot 1_{\Sigma \cap \Omega_N}.
\]

**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 \( b_m, m = (m_1, m_2) \in \mathfrak{M}_N \). First, for every \( \tau = \sum_{j, \ell} T_{j, \ell}(\varphi_j \otimes \psi_\ell) \) belonging to \( \mathcal{F} := \{ \tau \in \mathcal{S}_m : \| \tau \|_f^2 \leq 2 \} \),
\[
\| \tau \|^2 = \text{trace}(TT^* \Psi_{m_1} \Psi_{m_1}^{-1}) \leq \| \Psi_{m_1}^{-1} \|_{\text{op}} \text{trace}(T^* \Psi_{m_1} T) \leq 2\| \Psi_{m_1}^{-1} \|_{\text{op}},
\]
and then
\[
\| \tau \|_\infty = \sup_{(x, y) \in \Omega} \left\{ \sum_{j, \ell=1}^{m_1} \sum_{\ell=1}^{m_2} T_{j, \ell}(\varphi_j(x) \psi_\ell(y)) \right\}
\leq \mathcal{L}_\psi(m_2)^{1/2} \sup_{x \in \Omega} \left\{ \sum_{j=1}^{m_1} |\varphi_j(x)| \sum_{\ell=1}^{m_2} |T_{j, \ell}| \right\} \leq \sqrt{2} \mathcal{L}_\psi(m_1)^{1/2} \mathcal{L}_\psi(m_2)^{1/2} \| \Psi_{m_1}^{-1} \|_{\text{op}}.
\]
Since \( m \in \mathfrak{M}_N \), and since \( \mathcal{L}_\psi(m_2) \leq m_1 \mathcal{L}_\psi(m_2) \leq N \) by the definition of \( \mathcal{U}_N \),
\[
\sup_{\tau \in \mathcal{F}} \| \tau \|_\infty \leq M \quad \text{with} \quad M = \frac{c_3 N}{\log(N)^{1/2}} \quad \text{and} \quad c_3 = 2\sqrt{2} T.
\]
Now, since
\[
E \left( \sup_{\tau \in \mathcal{F}} \nu_N(\tau)^2 \right) \leq \frac{2m_1 \mathcal{L}_\psi(m_2)}{N}
\]
as established in the proof of Theorem 4.2 (see Step 1),
\[
E \left( \sup_{\tau \in \mathcal{F}} |\nu_N(\tau)| \right) \leq H \quad \text{with} \quad H^2 = \frac{2m_1 \mathcal{L}_\psi(m_2)}{N},
\]
and then
\[
N \sup_{\tau \in \mathcal{F}} E(\nu_N(\tau)^2) \leq v \quad \text{with} \quad v = NH^2.
\]
By the aforementioned Talagrand’s inequality, for \( \alpha := a \log(N) \),
\[
E \left( \sup_{\tau \in \mathcal{F}} \nu_N(\tau)^2 - 2(1 + 2\alpha)H^2 \right) \leq A(N, \alpha, H, v) + B(N, \alpha, M, H),
\]
where
\[
A(N, \alpha, H, v) := \frac{24v}{N} \exp \left( - \frac{\alpha NH^2}{6v} \right)
\]
\[
= \frac{48m_1 \mathcal{L}_\psi(m_2)}{N} e^{-\frac{v}{2} \log(N)} \leq 48N^{-\frac{v}{2}} \quad \text{because} \quad m \in \mathcal{U}_N,
\]
and
\[
B(N, \alpha, M, H) := \frac{c_4 M^2}{N^2} \exp \left( - \frac{\sqrt{2}}{42} \cdot \frac{NH \sqrt{\alpha}}{M} \right) \quad \text{with} \quad c_4 = 7056
\]
\[
= \frac{c_3^2 c_4}{\log(N)} \exp \left( - \frac{\sqrt{2}}{42} \cdot \frac{\sqrt{2} c_3 \log(N)}{c_3} \right) \leq c_4^3 c_4 N^{-\frac{\sqrt{2} c_3}{c_3}}.
\]
Therefore, since \( a \geq (2 \cdot 84^{\sqrt{2}})^2 \) and since
\[
\kappa \geq \kappa_0 = 44a \geq \frac{22(1 + 2a \log(N))}{1 + \log(N)} = 11(1 + 2a \log(N)) \frac{H^2_{\text{pen}(m)}}{\kappa_0},
\]
there exists a constant \( c_5 > 0 \), not depending on \( N, m_1, m_2 \) and \( t \), such that
\[
b_m \leq E \left[ \left( \sup_{\tau \in \mathcal{S}_m : \|\tau\| \leq \frac{2}{11} \frac{H^2_{\text{pen}(m)}}{\kappa_0}} \nu_X(\tau)^2 - \frac{2}{11} \kappa_{\text{pen}(m)} \right) \right] + \frac{E}{N} \leq \frac{c_5}{N^2}.
\]

**Step 4 (conclusion).** By Steps 2 and 3,
\[
E(A) \leq \frac{c_4 (1 + R(t))}{N} \quad \text{and} \quad E(B) \leq 11 \sum_{m \in \mathcal{M}_N} b_m \leq \frac{11 c_5}{N}.
\]
In conclusion, by Inequality (27), there exists a constant \( c_6 > 0 \), not depending on \( N \) and \( t \), such that
\[
E(\|\hat{\rho}_{m,t} - p_{I \times J,t}\|_2^2) \leq \min_{m \in \mathcal{M}_N} \left\{ E(\|\hat{\rho}_{m,t} - p_{I \times J,t}\|_2^2) + \kappa_{\text{pen}(m)} \right\} + \frac{c_6 (1 + R(t))}{N}.
\]

**A.3.2. Proof of Lemma A.3.** By the definition of \( \hat{\Pi}_m \) (see (20)),
\[
\hat{\Pi}_m(\hat{\rho}_{M,t}) = \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \left[ \hat{\Psi}_m^{-1} \hat{P}_m(\hat{\rho}_{M,t}) \right]_{j,\ell}(\varphi_j \otimes \psi_\ell)
\]
with, for every \( j \in \{1, \ldots, m_1\} \) and \( \ell \in \{1, \ldots, m_2\} \),
\[
\hat{P}_m(\hat{\rho}_{M,t}) = \frac{1}{NT} \sum_{i=1}^N \int_0^T \int_{-\infty}^\infty \hat{\rho}_{M,t}(X_i^j, y) \varphi_j(X_i^j) \psi_\ell(y) dy ds
\]
\[
= \sum_{j'=1}^{M_1} \sum_{\ell'=1}^{M_2} \left[ \hat{\Psi}_M^{-1} \hat{Z}_{M,t} \right]_{j',\ell'}
\times \frac{1}{NT} \sum_{i=1}^N \int_0^T \varphi_j(X_i^j) \varphi_j(X_i^j) \int_{-\infty}^\infty \psi_\ell(y) \psi_\ell(y) dy ds
\]
\[
= \frac{M_1}{NT} \sum_{i=1}^N \int_0^T \varphi_j(X_i^j) \varphi_j(X_i^j) ds
\]
\[
= \left[ \hat{Z}_{M,t} \right]_{j,\ell} \quad \text{because} \quad \mathcal{S}_m \subset \mathcal{S}_M.
\]
Therefore,
\[
\hat{\Pi}_m(\hat{\rho}_{M,t}) = \sum_{j=1}^{m_1} \sum_{\ell=1}^{m_2} \left[ \hat{\Psi}_m^{-1} \hat{Z}_{m,t} \right]_{j,\ell}(\varphi_j \otimes \psi_\ell) = \hat{\rho}_{m,t}.
\]

**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 \( v \) of \( N \sup_{\tau \in \mathcal{F}} E(\nu_X(\tau)^2) \) 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 \( b_m \), \( m = (m_1, m_2) \in M_N \). First, for every \( \tau = \sum_{j,\ell} T_{j,\ell} (\phi_j \otimes \psi_\ell) \) belonging to \( F = \{ \tau \in S_m : \| \tau \|_f \leq 2 \} \), we still get

\[
\| \tau \|_\infty \leq \sqrt{T} \mathcal{L}_\nu(m_1) \frac{1}{2} \mathcal{L}_\psi(m_2) \frac{1}{2} \| \Psi^{-1} \|_{\text{op}}^{\frac{1}{2}}.
\]

By Assumption 5.1.(3), and since \( m = (m_1, m_2) \) belongs to \( M_N \),

\[
\mathcal{L}_\nu(m_1) \frac{1}{2} \| \Psi^{-1} \|_{\text{op}}^{\frac{1}{2}} \leq c_\nu^2 m_1^{\frac{1}{2}} \left( c_\nu^2 m_1 \| \Psi^{-1} \|_{\text{op}}^2 \right)^{\frac{1}{2}} \leq c_\nu^4 m_1^{\frac{1}{2}} \left( \frac{4\rho_n NT}{\log(NT)} \right)^{\frac{1}{2}}.
\]

Thus, for \( NT \geq e \),

\[
\sup_{\tau \in F} \| \tau \|_\infty \leq M \quad \text{with} \quad M = c_{\nu,b} m_1^{\frac{1}{2}} \mathcal{L}_\psi(m_2)^{\frac{1}{2}} N^4 \quad \text{and} \quad c_{\nu,b} = 2\sqrt{e}(\theta T)^{\frac{1}{2}}.
\]

Now, since

\[
E \left( \sup_{\tau \in F} \nu_N(\tau)^2 \right) \leq \frac{2m_1 \mathcal{L}_\psi(m_2)}{N}
\]

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

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

Lastly,

\[
\sup_{\tau \in F} \left\{ \var{ \frac{1}{T} \int_0^T (\tau(X_s, X_{s+t}) - \mathbb{E}(\tau(X_s, X_{s+t})|X_s))ds } \right\}
\]

\[
\leq \sup_{\tau \in F} E \left[ \left( \frac{1}{T} \int_0^T \tau(X_s, X_{s+t})ds \right)^2 \right] \leq \rho_n \sup_{\tau \in F} \| \tau \|_f \leq v \quad \text{with} \quad v = 2\rho_0.
\]

By the aforementioned Talagrand’s inequality, for \( \alpha = 1/4 \),

\[
E \left[ \sup_{\tau \in F} \nu_N(\tau)^2 - 3H^2 \right]_+ \leq A(N, H, v) + B(N, M, H),
\]

where

\[
A(N, H, v) := \frac{24v}{N} \exp \left( -NH^2 \frac{24v}{N} \right) = \frac{48\rho_0}{N} \exp \left( -\frac{m_1 \mathcal{L}_\psi(m_2)}{24\rho_0} \right),
\]

and

\[
B(N, M, H) := \frac{\epsilon_{4,b} M^2}{N^2} \exp \left( -\frac{\sqrt{2}}{84} \cdot \frac{NH}{M} \right) \quad \text{with} \quad \epsilon_{4,b} = 7056
\]

\[
\leq \frac{\epsilon_{4,b} \epsilon_{6,b}^2}{\sqrt{N}} \exp \left( -\frac{m_1^{1/4} N^{1/4}}{42\epsilon_{4,b}} \right) \quad \text{because} \quad m \in U_N.
\]

Since \( m_1 \geq 1 \),

\[
B(N, M, H) \leq \frac{\epsilon_{5,b}}{\sqrt{N}} \exp(-\epsilon_{6,b} N^{\frac{1}{4}}) \quad \text{with} \quad \epsilon_{5,b} = \epsilon_{4,b} \epsilon_{6,b}^2 \quad \text{and} \quad \epsilon_{6,b} = \frac{1}{42\epsilon_{4,b}}.
\]

So, for \( \kappa_b \geq \kappa_{6,0} = 33/2 \) and \( t \in [t_0, T] \),

\[
b_m \leq E \left[ \left( \sup_{\tau \in S_m \| \tau \|_f \leq 2} \nu_N(\tau)^2 \right)^+ \right] \leq E \left[ \left( \sup_{\tau \in F} \nu_N(\tau)^2 - 3H^2 \right)^+ \right] 
\]

\[
\leq \frac{48\rho_0}{N} \exp \left( -\frac{m_1 \mathcal{L}_\psi(m_2)}{24\rho_0} \right) + \frac{\epsilon_{5,b}}{\sqrt{N}} \exp(-\epsilon_{6,b} N^{\frac{1}{4}}).
\]
Therefore, by Assumption 5.3, there exists a constant $c_{T,b} > 0$, not depending on $N$, such that

$$
\mathbb{E}(B) \leq c_{T,b} \cdot \frac{\sum_{m \in \mathcal{M}_N} b_m}{N}.
$$

**References**


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

*Email address*: nmarie@parisnanterre.fr