The Annals of Applied Probability
2002, Vol. 12, No. 2, 423–476

RATE OF CONVERGENCE OF A PARTICLE METHOD TO THE SOLUTION OF THE MCKEAN–VLASOV EQUATION

BY FABIO ANTONELLI[1] AND ARTURO KOHATSU-HIGA[2]

Università di Chieti and Universitat Pompeu Fabra

This paper studies the rate of convergence of an appropriate discretization scheme of the solution of the McKean–Vlasov equation introduced by Bossy and Talay. More specifically, we consider approximations of the distribution and of the density of the solution of the stochastic differential equation associated to the McKean–Vlasov equation. The scheme adopted here is a mixed one: Euler–weakly interacting particle system. If n is the number of weakly interacting particles and h is the uniform step in the time discretization, we prove that the rate of convergence of the distribution functions of the approximating sequence in the $L^1(\Omega \times \mathbb{R})$ norm is of the order of $\frac{1}{\sqrt{n}} + h$, while for the densities is of the order $h + \frac{1}{\sqrt{n}h^{1/4}}$. The rates of convergence with respect to the supremum norm are also calculated. This result is obtained by carefully employing techniques of Malliavin calculus.

1. Introduction. In a series of articles (see [1, 2, 16]), Bossy and Talay studied the numerical approximation of the solutions to the McKean–Vlasov equation and to the Burgers equation. The McKean–Vlasov equation is obtained as the diffusive limit of a particle system, describing the behavior of a high density gas. Its solution is a probability law density and it can be represented as the law of the solution of an associated nonlinear stochastic differential equation (for further details we refer the reader to [4]).

In their paper, Bossy and Talay choose to approximate the McKean–Vlasov limit by replicating the behavior with a system of n weakly interacting particles, each following a SDE discretized in time with step $h \in (0, 1]$. In [1] it is proved that when $n \to \infty$ and $h \to 0$, then the empirical distribution function of these n particles converges, in a weak sense to be defined later, toward the solution of the McKean–Vlasov limit with a rate at least of the order $\frac{1}{\sqrt{n}} + \sqrt{h}$. Through some simulations it can be clearly seen that the rate in n is optimal but that the rate in h is probably better than $\sqrt{h}$.

In this article, we prove that the rate of convergence of the scheme constructed by Bossy and Talay is actually at least of the order $\frac{1}{\sqrt{n}} + h$, as they also suspected on the basis of some numerical simulations they ran.

Received April 2000; revised April 2001.
[1]Supported in part by CNR-NATO Grant 215.29.
[2]Supported in part by a DGICYT grant.
AMS 2000 subject classifications. Primary 60H10, 60K35; secondary 34F05, 90A12.
Key words and phrases. McKean–Vlasov equation, Malliavin calculus.

To make our introduction more precise, we recall that the McKean–Vlasov equation can be described by means of four Lipschitz kernels $a(x,y)$, $b(x,y)$, $f(x,y)$ and $g(x,y)$ from $\mathbb{R}^2$ to $\mathbb{R}$ and of a differential operator, acting on the probability measures, defined by

$$L(\mu)h(x) = \frac{1}{2}\left[b\left(x, \int_{\mathbb{R}} g(x,y)\,d\mu(y)\right)\right]^2 h''(x) + \left[a\left(x, \int_{\mathbb{R}} f(x,y)\,d\mu(y)\right)\right]h'(x).$$

A family of probability measure $\{\mu_t\}_{t\geq 0}$ is said to be the solution of the McKean–Vlasov equation if it solves

$$\frac{d}{dt}\langle \mu_t, h\rangle = \langle \mu_t, L(\mu_t)h\rangle \qquad \forall\, h \in C_K^\infty(\mathbb{R}) \text{ (compact support)},\ \mu_{t=0} = \mu_0, \tag{1.1}$$

where μ_0 is an initial probability measure. Applications and a general discussion about the above equation can be found in Gärtner [4]. For example, this kind of system is the natural limit equation for the theory of propagation of chaos that allows the interpretation of the above system as the equation satisfied by the limit of the empirical measure associated with a random particle system with long-range weak interaction. Applications of this model arise in Newtonian physics, statistical mechanics, chemical kinetics and segregation problems in biological populations, among others.

By associating a martingale problem to the operator L, μ_t can also be characterized through the stochastic differential equation (SDE),

$$\begin{aligned} X_t = \xi &+ \int_0^t a\left(X_s, \int_{\mathbb{R}} f(X_s,y)\,d\mu_s(y)\right) ds \\ &+ \int_0^t b\left(X_s, \int_{\mathbb{R}} g(X_s,y)\,d\mu_s(y)\right) dW_s, \end{aligned} \tag{1.2}$$

where μ_t denotes the law of the solution X_t, while W is a Wiener process on space, so that the natural filtration generated by W is extended with an initial independent sigma-algebra $\mathcal{G}_0$, to make ξ an $\mathcal{F}_0$-measurable random variable with law μ_0. As shown by Gärtner, under appropriate conditions on the coefficients, there exists a unique strong solution of (1.2), X_t, and its law, μ_t satisfies (1.1).

The SDE (1.2) is sometimes called nonlinear, since its coefficients involve at the same time X_s and its law. In [1], it is suggested that the numerical approximation of (1.2) must act on two levels. On one, the usual time discretization (see [8]) is needed, based on simulations of the increments of the driving process W. On the other, it is necessary to use some empirical measure in order to approximate the measures μ_s that appear in the coefficients. To this purpose, the simulation scheme is expanded introducing n independent driving Wiener processes, each generating a particle through an equation that approximates (1.2) (for details see Section 3). These particles, denoted by X^i, $i = 1, \dots, n$, will interact with each other through their empirical measure, viewed as an approximation of μ_s. By some kind of law of large numbers (or propagation of chaos as it is better known), this interaction tends to disappear as $n \to \infty$.

Bossy and Talay prove that the empirical distribution generated by the X^i converges to the law of X and therefore give a method to approximate the solution of the McKean–Vlasov equation (1.1). More exactly, denoting with h the time step and n the number of particles, they prove the following result, which we report here for the reader's convenience, since we will refer to it for the purpose of comparison.

THEOREM 1.1. *Let $a(x, y) = b(x, y) = y$ and assume*:

(H-1) *There exists a strictly positive constant c such that $g(x, y) \geq c > 0$, $\forall\,(x, y) \in \mathbb{R}^2$.*
(H-2) *The functions f and g are uniformly bounded on $\mathbb{R}^2$; f is globally Lipschitz and g has uniformly bounded first partial derivatives.*
(H-3) *The initial law μ_0 satisfies one of the following*:
   (i) *μ_0 is a Dirac measure at x_0.*
   (ii) *μ_0 has a continuous density p_0 so that there exist constants $M, \alpha > 0$, $\eta \geq 0$ such that $p_0(x) \leq \eta \exp(-\alpha \frac{x^2}{2})$ for $|x| > M$ (if $\eta = 0$, μ_0 has compact support).*

Furthermore, if $u(t, \cdot)$ is the distribution function of X_t and $\overline{u}(t, \cdot)$ the empirical distribution function of the sequence X_t^i for $i = 1, \dots, n$, then for any fixed $t \in [0, T]$,

$$E\|u(t, \cdot) - \overline{u}(t, \cdot)\|_{L^1(\mathbb{R})} \leq C\left(\frac{1}{\sqrt{n}} + \sqrt{h}\right). \tag{1.3}$$

If we substitute (H2) *and* (H3) *with the stronger conditions*:

(H-2′) *$f \in C_b^2(\mathbb{R}^2)$ and $g \in C_b^3(\mathbb{R}^2)$.*
(H-3′) *The initial law μ_0 has a strictly positive density $p_0 \in C^2(\mathbb{R})$ and there exist constants $M, \eta, \alpha > 0$ such that $p_0 + |p_0'(x)| + |p_0''(x)| \leq \eta \exp(-\alpha \frac{x^2}{2})$ for $|x| > M$,*

then μ_t has a density, denoted by $p_t(\cdot)$, and

$$E\left\| p_t(\cdot) - \frac{1}{n}\sum_{j=1}^{n} \phi_\varepsilon(X_t^j - x)\right\|_{L^1(\mathbb{R})} \leq C\left(\varepsilon + \frac{1}{\sqrt{\varepsilon}}\left(\frac{1}{\sqrt{n}} + \sqrt{h}\right)\right), \tag{1.4}$$

where $\phi_\varepsilon(z) = \frac{\exp(-z^2/2\varepsilon)}{\sqrt{2\pi\varepsilon}}$.

The goal of our work is to prove that the rate in (1.3) is actually $\frac{1}{\sqrt{n}} + h$ under conditions comparable to (H1), (H2) and (H3). We will first establish the result for the densities showing that the optimal rate in (1.4) is at least of the order $\frac{1}{\sqrt{nh}} + h$, when $\varepsilon = h$, rather than $h + \frac{1}{\sqrt{nh}} + \sqrt{h} + 1$.

Our efforts clearly drew inspiration from the remarks made by Bossy and Talay (see [1] and [2]), who gave numerical evidence that suggested the rate of convergence was faster than what they proved.

Here we are able to achieve this better rate, by using completely different techniques from those in [1]. Indeed, we carefully employ Malliavin calculus techniques together with some ideas brought to light in a recent work by Kohatsu and Ogawa [7]. The drawback of this method is the high degree of smoothness required on the coefficients of the equation.

Malliavin calculus allows us to establish when the marginal densities of the solution of a SDE exist and are regular, so it is indeed very apt to deal with equations whose coefficients involve probability densities. The introduction of these techniques in this setting enabled us also to weaken slightly the hypotheses on the coefficients as well as those on the initial density function. We establish this result in Section 2 and it is adapted from the similar one obtained by Taniguchi (see [17]), in the study of the smoothness of densities for time dependent systems.

The key idea in the proof of the result is that the coefficients have to verify the so-called restricted Hörmander condition. The main difference between Taniguchi's results and ours is that we do not require any boundedness for the coefficients; indeed, bounded differentiability in all the required derivatives is sufficient; under hypotheses (H0) of the next section, this property is satisfied by the coefficients of (1.2) and we can apply our results of existence and smoothness of the densities to the process under study. Another difference with Taniguchi's paper is the introduction of an initial random variable. If we were to use a uniform restricted Hörmander type condition, as in [17], this difference would be minor. However, applications force the study of the case when the initial random variable is supported on the whole real line. Therefore, such a uniform restricted Hörmander condition would be very restrictive. Here we only require some tail conditions on the initial random variable. In order to carry out the proof in this case one needs to study carefully the behavior of all the bounds with respect to the initial random variable.

In Section 3 we study the approximation errors of the particle method used to approximate the solution of (1.2); this analysis relies on a technique very different from the one used by Bossy and Talay. We try to separate as much as possible the effects of the time discretization and the particle method so that one can find the optimal rate. We believe this to be the main reason why one obtains $\sqrt{h}$ instead of h in (1.3) (see, e.g., Sections 6 and 7 in [16]). One of the key points to do so is an approximation method for the solution of the SDE generated by the difference equation, briefly explained at the end of the proof of Theorem 3.5.

The basic idea is as follows: consider formally the quantity

$$
\begin{aligned}
E\left\| E\big(\delta_x(X_t)\big) - \frac{1}{n}\sum_{j=1}^{n}\phi_\varepsilon(X_t^j - x)\right\|_{L^1(\mathbb{R})} \\
\leq \big\| E\big(\delta_x(X_t)\big) - E\big(\delta_x(X_t^1)\big)\big\|_{L^1(\mathbb{R})} \\
+ E\left\| E\big(\delta_x(X_t^1)\big) - \frac{1}{n}\sum_{j=1}^{n}\phi_\varepsilon(X_t^j - x)\right\|_{L^1(\mathbb{R})}.
\end{aligned}
$$

The second term is about the order $\frac{1}{\sqrt{n}}$ (some correlation structure between the X^j has to be studied). The first is a term of the same kind that arises in classical weak approximation procedures, except that in our case discretization both in time and in space (measure discretization) is used. By analyzing separately the two discretizations one gets a better rate of convergence.

To carry out this idea is not as easy as explained above. It presents some extra complications with respect to the classical case of diffusions and it is essential for our method to work, that we run a separate study of the time and space discretizations.

The results for approximations of the distribution function of X_t are obtained with similar techniques to those used for the density functions. For this reason we decided to explain in detail this second case, technically more demanding, and to sketch the proofs for the first.

We hope the methods exposed here will help develop similar results also for the Burgers equation and in general for nonlinear equations.

In the area of numerical approximations to nonlinear equations the particle method is not the only method available to simulate the solution. There are also other methods that are related to the one presented here (see, e.g., [3, 5, 6, 14]).

The paper is subdivided as follows. In Section 2, we give the preliminary results that enable concluding the existence and smoothness of the densities of the solution of (1.2) under restricted Hörmander conditions. Given that the problem in one dimension simplifies to elliptic conditions we decided to consider the multidimensional case.

In the rest of the paper we concentrate on the one-dimensional case, the general multidimensional case being a straightforward generalization. In Section 3 we establish our approximation results for densities, while in Section 4 we summarize those and we derive the distribution function case.

We adopt the convention of writing the same letter (usually C) for a constant even if it changes from line to line. This constant is always independent of h, n and the partition of the time interval. Unless otherwise stated we will also assume without loss of generality that all constants are bigger than 1.

2. Preliminary results. Let $[0,T]$ be a finite time interval and $(\Omega,\mathcal{F},P)$ a complete probability space, where a standard d-dimensional Brownian motion, W, is defined. We consider the equation in $\mathbb{R}^n$,

$$X_t = \xi + \int_0^t a\big(X_s, F(X_s;\mu_s)\big)\,ds + \int_0^t b\big(X_s, G(X_s;\mu_s)\big)\,dW_s, \tag{2.1}$$

where $F(x;\mu_s)$ or $G(x;\mu_s)$ denote the functions given by $\int_{\mathbb{R}^n}\zeta(x,y)\,d\mu_s(y)$ ($\zeta = f, g$, respectively) and μ_s indicates the distribution of X_s.

We are going to study the existence and smoothness of the density of the solution of (2.1). For ease of writing, we call $\bar{a}(t,x) = a(x, F(x;\mu_t))$ and $\bar{b}(t,x) = b(x, G(x;\mu_t))$, so we rewrite equation (2.1) as

$$X_t = \xi + \int_0^t \bar{a}(s, X_s)\,ds + \sum_{k=1}^d \int_0^t \bar{b}_k(s, X_s)\,dW_s^k. \tag{2.2}$$

Next, we introduce a series of hypotheses that we need for our goal.

ASSUMPTIONS.

(H0) ξ is an $\mathcal{F}_0$-measurable random variable in $\mathbb{R}^n$, such that $\xi \in \bigcap_{p\ge 1} L^p$. The functions

$$a\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n, \qquad b\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}^d, \qquad f, g\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

are all smooth with bounded derivatives, we call M the constant dominating them all.

(H1) There exists an integer m_0 and a positive constant c, which without loss of generality we assume smaller than $1/2$, such that

$$\sum_{i=0}^{m_0}\sum_{v\in I_i}\langle v(0,\xi), \eta\rangle^2 \ge c > 0 \qquad \text{a.e. for all } \eta \in S^{n-1},$$

where the sets I_i are given by $I_0 = \{\bar{b}_1, \dots, \bar{b}_d\}, \dots, I_n = \{[\bar{b}_k, v], v \in I_{n-1}, 1 \le k \le d\}$ and $[\cdot,\cdot]$ denotes the Lie bracket. In this context, the coefficients are to be understood as vector fields, that is $\bar{b}_k(t,x) = \sum_{i=1}^n \bar{b}_k^i(t,x)\frac{\partial}{\partial x_i}$.

(H2) The functions b_k are bounded, let us say by the same constant M as in (H0).

(H3) ξ has a density u_0 for which there exist positive constants η, α, β and ρ such that

$$u_0(x) \le \eta\exp(-\alpha|x|^\beta) \qquad \text{for } |x| \ge \rho.$$

From Hypothesis (H0), it is clear that all the derivatives of $v \in \bigcup_{i=1}^n I_i$ are bounded. Without loss of generality we assume that these derivatives are bounded by the same constant M.

Hypothesis (H1) is the so-called restricted Hörmander condition, as it involves only the diffusion coefficients. In the one-dimensional case this reduces to saying that almost surely $|\bar{b}(0,\xi)| \geq c > 0$, which is very similar to the corresponding (H-1) in Theorem 1.1, requiring $\bar{b}(0,x) \geq c > 0$, for all $x \in \mathbb{R}$. In the multidimensional case, instead Hypothesis (H1) becomes actually much weaker than (H-1), as it may involve the brackets of order higher than one, while (H-1) does not.

Hypothesis (H2) is similar to (H-2′) in Theorem 1.1; note that the smoothness in the coefficients is needed here to allow the study of the smoothness of the density. Finally, Hypothesis (H3) is slightly weaker than the corresponding (H-3′).

Another difference is given by the fact that in Theorem 1.1 all three conditions are assumed, while we are going to show, by means of Malliavin calculus techniques, that it is necessary to assume only Hypothesis (H1) and either Hypothesis (H2) or (H3). Therefore the combination of Hypothesis (H1) and (H2) give that the restricted Hörmander condition does not have to be necessarily uniform as is required in (H-1).

Since all the results in the paper rely heavily on Malliavin calculus, we want to introduce here some of its terminology very briefly.

For $d \in \mathbb{N}$, we denote by $C_b^\infty(\mathbb{R}^d)$ the set of C^∞ bounded functions $f\colon \mathbb{R}^d \to \mathbb{R}$, with bounded derivatives of all orders and we assume that a d-dimensional Wiener process is defined on a probability space.

If we denote by S the class of real random variables F that can be represented as $f(W_{t_1},\dots,W_{t_n})$ for some $n \in \mathbb{N}$, $t_1,\dots,t_n \in [0,T]$ and $f \in C_b^\infty(\mathbb{R}^{nd})$, we can complete this space under the norm $\|\cdot\|_{1,p}$ given by

$$\|F\|_{1,p}^p = E(|F|^p) + \left(\sum_{j=1}^d E\left(\int_0^T |D_s^j F|^2\,ds\right)^{p/2}\right),$$

where D^j is defined as

$$D_s^j F = \sum_{i=1}^n \frac{\partial f}{\partial x_{ij}}(W_{t_1},\dots,W_{t_n}) 1_{[0,t_i]}(s) \qquad \text{for } j = 1,\dots,d,$$

obtaining a Banach space, usually indicated by $\mathbb{D}^{1,p}$. Analogously, we can construct the space $\mathbb{D}^{k,p}$, by completing S under the norm

$$\|F\|_{k,p}^p = E(|F|^p) + \sum_{j=1}^k \sum_{k_1+\cdots+k_d=j} E\left(\left(\int_0^T \cdots \int_0^T |D^{d,k_d}_{s_j\cdots s_{j-k_d}} \cdots D^{1,k_1}_{s_{k_1}\cdots s_1} F|^2\,ds_1\cdots ds_j\right)^{p/2}\right),$$

where $D^{i,l}_{s_1\cdots s_l} F = D^i_{s_1}\cdots D^i_{s_l} F$. Finally, we denote $\mathbb{D}^\infty = \bigcap_{p\geq 1}\bigcap_{k\geq 1}\mathbb{D}^{k,p}$.

The adjoint of the closable unbounded operator $D^j\colon \mathbb{D}^{1,2} \subseteq L^2(\Omega) \to L^2([0,T]\times\Omega)$ is usually denoted by δ^j and it is called the Skorohod integral.

The domain of δ^j is the set of all processes $u \in L^2([0,T]\times\Omega)$ such that

$$\left|E\left(\int_0^T D_t^j F u_t\,dt\right)\right| \le C\|F\|_2 \qquad \forall\ F \in S,$$

for some constant C depending possibly on u.

If $u \in \mathrm{Dom}(\delta^j)$, then $\delta^j(u)$ is the square integrable random variable determined by the duality relation

$$E(\delta^j(u)F) = E\left(\int_0^T D_t^j F u_t\,dt\right) \qquad \forall\ F \in \mathbb{D}^{1,2}.$$

In the multidimensional case we consider $\delta = \sum_j \delta^j$.

Finally, for a possibly m-dimensional random variable F we denote its Malliavin covariance matrix by γ_F and it is defined as

$$\gamma_F^{hk} = \sum_{j=1}^d \int_0^T D_s^j F^h D_s^j F^k\,ds, \qquad h,k = 1,\dots,m.$$

The Malliavin covariance matrix plays a key role when one wants to determine the existence and the smoothness of the densities of the solutions of stochastic differential equations. Namely, following [10] (Proposition 2.1.1, page 78), we have that for any random variable $F \in (\mathbb{D}_{loc}^{1,p})^m$ for some $p > 1$, if γ_F is almost surely invertible, then the law of F is absolutely continuous with respect to Lebesgue measure. Moreover if $F \in \mathbb{D}^{1,2}$ (and it is one-dimensional) and $\gamma_F^{-1}DF$ is in $\mathrm{Dom}(\delta)$ then F has a continuous and bounded density given by

$$f(x) = E\big(1_{\{F>x\}}\delta(\gamma_F^{-1}DF)\big).$$

In particular we will use the fact that if $F \in \mathbb{D}^\infty$ and $|\gamma_F^{-1}| \in \bigcap_{p>1} L^p$ then F has an infinitely differentiable density (see [10], Corollary 2.1.2).

The above gives birth to a general formula known as the integration by parts formula. For any two random variables $F, G \in \mathbb{D}^\infty$, so that $|\gamma_F|^{-1} \in \cap_{p>1} L^p$ and $f \in \mathbb{C}_p^\infty(\mathbb{R})$, the following integration by parts formula holds:

$$E\big(f^{(m)}(F)G\big) = E\big(f(F)H_m(F,G)\big) \qquad \text{for } m \ge 1, \tag{2.3}$$

where $H_m(F,G) = H(F, H_{m-1}(F,G))$ and

$$H_1(F,G) = H(F,G) = \sum_{i=1}^d \delta^i(G\gamma_F^{-1}D^iF),$$

where δ^i denotes the adjoint operator of D^i.

Moreover following carefully the calculations in [10], page 41, one obtains that for any $p > 1$ there exist indices h, k, l, q, q', depending on m and p and a constant $C = C(m, p, h, k, l)$ such that

$$\|H_m(F,G)\|_p \le C\|\gamma_F^{-1}\|_{lp}^q \|F\|_{m+1,kp}^{q'} \|G\|_{m,hp}. \tag{2.4}$$

One can determine exactly l, k and p. For this see [12].

Having introduced all the necessary terminology we first quote a result from [7] about existence and integrability of the solution of (2.1).

THEOREM 2.1. *Let us assume that* (H0) *is satisfied; then there is a unique strong solution of* (2.1) *such that, for all* $p > 1$,

$$E\Big(\sup_{s\le T}|X_s|^p\Big) \le \infty.$$

Furthermore $X_s \in \mathbb{D}^\infty$ *for all* $s \in [0, T]$.

We are now able to state and prove the main result of this section about the marginal densities of X.

THEOREM 2.2. *Assume that Hypotheses* (H0) *and* (H1) *are satisfied together with either Hypotheses* (H2) *or* (H3). *Then* $\gamma_{X_t}^{-1} \in \bigcap_{p\ge 1} L^p$ *and* X_t *has a smooth density*.

REMARK. We would like to point out that the restricted Hörmander condition is needed for the densities to be absolutely continuous with respect to Lebesgue measure when considering time dependent coefficients. It is possible to construct easy examples where the coefficients of a SDE verify an unrestricted Hörmander condition at all times and points, but for which the Malliavin covariance matrix associated to the solution is not invertible (see [17]).

PROOF. Our proof is an adaptation of the method used by [17], extended to the case when the initial point is random and the uniformity on the Hörmander condition is relaxed. Since the argument is basically the same under either Hypothesis (H2) or (H3), we prove the result assuming the latter, being the more difficult one, and we point out the differences with the other case step by step.

In (2.2) the coefficients are time dependent and because of Hypothesis (H0), they are smooth in space with bounded derivatives (hence they are also globally Lipschitz) and globally differentiable in time as many times as needed. Indeed, for $t \in [0, T]$ and any k,

$$\frac{\partial \bar{b}_k}{\partial t}(t,x) = \frac{\partial b_k}{\partial t}\big(x, E\big(g(x,X_t)\big)\big) = \frac{\partial b_k}{\partial y}\big(x, E\big(g(x,X_t)\big)\big)\frac{\partial E(g(x,X_t))}{\partial t}.$$

On the other hand, applying Itô's lemma to $g(x,\cdot)$, we get

$$E\big(g(x,X_t)\big) = E\big(g(x,\xi)\big) + \int_0^t E\Bigg[\sum_{i=1}^n g_{x_i}(x,X_r)\bar{a}^i(r,X_r)\Bigg]dr + \int_0^t E\Bigg[\sum_{i,j=1}^n g_{x_i x_j}(x,X_r)\sum_{k=1}^d \bar{b}_k^i(r,X_r)\bar{b}_k^j(r,X_r)\Bigg]dr$$

and hence we obtain

$$\frac{\partial E(g(x, X_t))}{\partial t} = E\left[\sum_{i=1}^{n} g_{x_i}(x, X_t)\bar{a}^i(t, X_t) + \sum_{i,j=1}^{n} g_{x_i x_j}(x, X_t)\sum_{k=1}^{d}\bar{b}_k^i(t, X_t)\bar{b}_k^j(t, X_t)\right],$$

which is bounded since all the derivatives of the coefficient g are bounded by M by hypothesis and we can bound $E(|\bar{a}^i(r, X_r)|)$ and $E(|\bar{b}_k^i(t, X_t)\bar{b}_k^j(t, X_t)|)$ by exploiting the mean value theorem and Hypothesis (H0). For instance, for $E(|\bar{a}^i(r, X_r)|)$ we have

$$\begin{aligned}
|\bar{a}^i(r, X_r)| &= \left|a^i\left(X_r, \int_{\mathbb{R}^n} f(X_r, y)\mu_r(dy)\right)\right| \\
&= \left|a^i\left(X_r, \int_{\mathbb{R}^n} f(X_r, y)\mu_r(dy)\right) - a^i\left(\xi, \int_{\mathbb{R}^n} f(\xi, y)\mu_0(dy)\right)\right| \\
&\quad + \left|a^i\left(\xi, \int_{\mathbb{R}^n} f(\xi, y)\mu_0(dy)\right) - a^i\left(0, \int_{\mathbb{R}^n} f(0, y)\mu_0(dy)\right)\right| \\
&\quad + \left|a^i\left(0, \int_{\mathbb{R}^n} f(0, y)\mu_0(dy)\right)\right| \\
&\le M\biggl(|X_r - \xi| + M|X_r - \xi| \\
&\qquad + \left|\int_{\mathbb{R}^n} f(\xi, y)\mu_r(dy) - \int_{\mathbb{R}^n} f(\xi, y)\mu_0(dy)\right|\biggr) \\
&\quad + M\left(|\xi| + \int_{\mathbb{R}^n} |f(\xi, y) - f(0, y)|\mu_0(dy)\right) \\
&\quad + \left|a^i\left(0, \int_{\mathbb{R}^n} f(0, y)\mu_0(dy)\right)\right| \\
&\le M(|X_r - \xi| + 2M|X_r - \xi|) + M(|\xi| + M|\xi|) \\
&\quad + \left|a^i\left(0, \int_{\mathbb{R}^n} f(0, y)\mu_0(dy)\right)\right|.
\end{aligned}$$

Taking expectations, $E(|\bar{a}^i(r, X_r)|)$ is finite. The same argument applies to the other term.

Of course, the same procedure may be repeated for $\bar{a}$ and the spatial and time derivatives of any order, proving the smoothness in time of the coefficients of the equation. It is exactly this smoothness that allows us, differently from Taniguchi's paper, to include in our discussion coefficients that might not be bounded. Without loss of generality we assume that all the derivatives of $\bar{a}$ and $\bar{b}$ are bounded by the same constant M.

Let ζ_t denote the derivative of the stochastic flow associated with (2.2) and ζ^{-1} the inverse flow, then both $\sup_{s\le t}|\zeta_t(\xi)|$ and $\sup_{s\le t}|\zeta_t^{-1}(\xi)| \in \bigcap_{p\ge 1} L^p$ and we can write the Malliavin derivative as $D_s^k X_t = \zeta_t(\xi)\zeta_s^{-1}(\xi)\bar{b}_k(s, X_s)$ (see [10], page 109), where by D^k we are denoting the Malliavin derivative with respect to the kth component of the Brownian motion.

Since we already noticed that $X_t \in \mathbb{D}^\infty$, to conclude the existence and regularity of the density, it is enough to check that

$$\gamma_{X_t}^{-1} = \left[\sum_{k=1}^d \zeta_t(\xi)\int_0^t \zeta_s^{-1}(\xi)\bar{b}_k(s, X_s)\bar{b}_k(s, X_s)^T\zeta_s^{-1}(\xi)^T\,ds\,\zeta_t(\xi)^T\right]^{-1} \in \bigcap_{p\ge 1} L^p,$$

where by T we denote the transpose. Using the L^p boundedness of $\zeta_t(\xi)$ and Lemma 2.3.1 in [10], this amounts to showing that for all $p\ge 2$ there exists $\varepsilon_0(p)$ such that for all $\varepsilon \le \varepsilon_0(p)$,

$$P\left(\sum_{k=1}^d \int_0^t \zeta_s^{-1}(\xi)\bar{b}_k(s, X_s)\bar{b}_k(s, X_s)^T\zeta_s^{-1}(\xi)^T\,ds < \varepsilon\right) \le \varepsilon^p. \tag{2.5}$$

Following [17] or [10], to prove (2.5) it suffices to show that for all η in the unit ball S^{n-1},

$$P\left(\sum_{k=1}^d \int_0^t \langle \zeta_s^{-1}(\xi)\bar{b}_k(s, X_s), \eta\rangle^2\,ds < \varepsilon\right) \le \varepsilon^p. \tag{2.6}$$

In particular we will prove that for the same m_0 of Hypothesis (H1), $\nu(m_0) = 5\cdot 4^{m_0}$ and for any sufficiently large $N\in\mathbb{N}$, we have

$$P\left(\sum_{k=1}^d \int_0^t \langle \zeta_s^{-1}(\xi)\bar{b}_k(s, X_s), \eta\rangle^2\,ds < \frac{1}{N^{\nu(m_0)}}\right) \le \frac{1}{N^{p\nu(m_0)}}.$$

In order to do so, we fix $\tau\in\mathbb{R}^+$ and we divide this probability into two parts,

$$\begin{aligned}
&P\left(\sum_{k=1}^d \int_0^t \langle \zeta_s^{-1}(\xi)\bar{b}_k(s, X_s), \eta\rangle^2\,ds < \varepsilon\right)\\
&\quad\le P\left(\sum_{k=1}^d \int_0^t \langle \zeta_s^{-1}(\xi)\bar{b}_k(s, X_s), \eta\rangle^2\,ds < \varepsilon, |\xi|\le\tau\right) + P(|\xi|\ge\tau)\\
&\quad= p_1 + p_2.
\end{aligned}$$

For p_2 we use Hypothesis (H3), which gives $P(|\xi|>\tau)\le C\exp(-\alpha\tau^\beta)$ for $\tau\ge\rho$. At the end of the proof, we will specify how to choose τ to have (2.5) satisfied. From now on, we assume without loss of generality that $M>1$ also is an upper bound for $\max_{0\le i\le m_0, v\in I_i}|v(0,0)|$. As for p_1, we first note that, given Hypothesis (H1) and $|\xi|\le\tau$, one has that for all $\eta\in S^{n-1}$, $|y-\xi|\le R$ and $s\le R$,

$$\sum_{i=0}^{m_0}\sum_{v\in I_i}\left|\langle v(s,y),\eta\rangle^2 - \langle v(0,\xi),\eta\rangle^2\right| \le 4M^2R(R+\tau+1)d^{m_0+2}n.$$

If we impose $4M^2R(R+\tau+1)d^{m_0+2}n < \frac{c}{4}$, then by solving the inequality we get

$$R < \frac{1+\tau}{2}\left[\sqrt{1+\frac{c}{4M^2nd^{m_0+2}(\tau+1)^2}}-1\right].$$

Then choosing $R = \frac{c}{32d^{m_0+2}M^2n(\tau+1)}$ we have that the inequality is satisfied. Therefore we have

$$\sum_{i=0}^{m_0}\sum_{v\in I_i}\langle v(s,y),\eta\rangle^2 \geq \frac{3c}{4} > 0, \tag{2.7}$$

when $|y-\xi| \leq R$ and $s \leq R$.

If we are instead assuming (H2), it is not necessary to split the probability in (2.6) into two parts and (2.7) holds automatically with $R = \frac{c}{16M^2d^{m_0+2}}$ which does not depend on τ.

Let us define the stopping time

$$\sigma_N = \inf\left\{s \in \left[0, \frac{1}{N^3}\right): |X_s - \xi| \geq \frac{R}{2} \text{ or } |\zeta_s^{-1}(\xi) - I| \geq \frac{1}{2}\right\} \wedge \frac{R}{2}.$$

There exists an appropriate constant $C_0 > 1$ such that for $N \geq \max(C_0(\tau+1)^{2/3}, t^{-1/3}) \geq \min(\frac{R}{2}, t)^{-1/3}$ we have $\sigma_N \leq \frac{1}{N^3}$ and

$$p_1 \leq P\left(\sum_{k=1}^{d}\int_0^{\sigma_N}\langle\zeta_s^{-1}(\xi)\bar{b}_k(s,X_s),\eta\rangle^2\,ds < \varepsilon, |\xi| \leq \tau\right). \tag{2.8}$$

Let us remark that by definition of σ_N, for any $q > 1$, we have

$$\begin{aligned}
P\left(\sigma_N < \frac{1}{N^3}\right) &\leq P\left(\sup_{0\leq s\leq 1/N^3}|X_s-\xi| \geq \frac{R}{2}\right) + P\left(\sup_{0\leq s\leq 1/N^3}|\zeta_s^{-1}(\xi) - I| \geq \frac{1}{2}\right)\\
&\leq \left(\frac{2}{R}\right)^q E\left(\sup_{0\leq s\leq 1/N^3}|X_s-\xi|^q\right) + 2^q E\left(\sup_{0\leq s\leq 1/N^3}|\zeta_s^{-1}(\xi)-1|^q\right)\\
&\leq \left(\frac{C_1}{R^q}+C_2\right)2^q\left(\frac{1}{N^3}\right)^{q/2},
\end{aligned}$$

for some constants C_1 and C_2. At this point, let us remark that

$$\sum_{k=1}^{d}\langle\zeta_s^{-1}(\xi)\bar{b}_k(s,X_s),\eta\rangle^2 = \sum_{v\in I_0}\langle\zeta_s^{-1}(\xi)v(s,X_s),\eta\rangle^2$$

and let us introduce, as in [17], the following sets:

$$E_j(N) = \left\{\sum_{i=0}^{j}\sum_{v\in I_i}\int_0^{\sigma_N}\langle\zeta_s^{-1}(\xi)v(s,X_s),\eta\rangle^2\,ds < \frac{d^j}{N^{\gamma_j}}\right\} \cap \{|\xi| \leq \tau\},$$

where $\gamma_j = 5 \cdot 4^{m_0 - j}$, for $j = 0, \dots, m_0$ (note that $\gamma_j = 4\gamma_{j+1}$ and $\gamma_0 = \nu(m_0)$).

We now chose in (2.6), $\varepsilon = \frac{1}{N^{\gamma_0}}$; therefore from (2.8) we get

$$\begin{aligned} p_1 &\le P\left(\sum_{v \in I_0} \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi) \bar{b}_k(s, X_s), \eta\rangle^2 \, ds < \frac{1}{N^{\gamma_0}}, |\xi| \le \tau\right) = P\big(E_0(N)\big) \\ &\le P\left(E_0(N) \cap \left\{\sigma_N < \frac{1}{N^3}\right\}\right) + P\left(E_0(N) \cap \left\{\sigma_N = \frac{1}{N^3}\right\}\right) \\ &\le P\left(\sigma_N < \frac{1}{N^3}\right) + \sum_{j=0}^{m_0 - 1} P\left(E_j(N) \cap E_{j+1}(N)^c \cap \left\{\sigma_N = \frac{1}{N^3}\right\}\right) \\ &\quad + P\left(E_{m_0}(N) \cap \left\{\sigma_N = \frac{1}{N^3}\right\}\right). \end{aligned}$$

We now prove that the last probability is actually zero. In fact, the Hörmander condition (2.7), gives that on the set $\{\sigma_N = \frac{1}{N^3}\}$, we necessarily have that on $\{s \le \sigma_N\}$, $\frac{1}{2} \le |\zeta_s^{-1}(\xi)| \le \frac{3}{2}$ and $|X_s - \xi| \le \frac{R}{2}$, hence (2.7) holds a.s. with X_s in place of y and consequently,

$$\sum_{i=0}^{m_0} \sum_{v \in I_i} \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi) v(s, X_s), \eta\rangle^2 \, ds \ge \frac{3}{8} c \sigma_N.$$

This implies that on the set

$$\begin{aligned} &E_{m_0}(N) \cap \left\{\sigma_N = \frac{1}{N^3}\right\} \\ &\qquad = \left\{\sum_{i=0}^{m_0} \sum_{v \in I_i} \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi) v(s, X_s), \eta\rangle^2 \, ds \le \frac{d^{m_0}}{N^5}, \sigma_N = \frac{1}{N^3}, |\xi| < \tau\right\} \end{aligned}$$

we have

$$\frac{3}{4} \frac{c}{N^3} \le \sum_{i=0}^{m_0} \sum_{v \in I_i} \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi) v(s, X_s), \eta\rangle^2 \, ds \le \frac{d^{m_0}}{N^5}$$

which is clearly not verified, as soon as $N > (\frac{4d^{m_0}}{3c})^{1/2}$.

It remains to analyze the probabilities $P(E_j(N) \cap E_{j+1}(N)^c \cap \{\sigma_N = \frac{1}{N^3}\})$, $j = 0, \dots, m_0 - 1$. On this set, we have

$$\sum_{i=0}^{j} \sum_{v \in I_i} \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi) v(s, X_s), \eta\rangle^2 \, ds \le \frac{d^j}{N^{\gamma_j}}$$

and

$$\sum_{i=0}^{j+1} \sum_{v \in I_i} \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi) v(s, X_s), \eta\rangle^2 \, ds > \frac{d^{j+1}}{N^{\gamma_{j+1}}}.$$

Under these conditions we obtain that for $N \geq (2/d)^{1/15}$,

$$\begin{aligned}
\sum_{i=0}^{j+1} \sum_{v \in I_i} \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi) v(s, X_s), \eta \rangle^2 \, ds
&\geq \sum_{v \in I_{j+1}} \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi) v(s, X_s), \eta \rangle^2 \, ds \\
&= \sum_{i=0}^{j+1} \sum_{v \in I_i} \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi) v(s, X_s), \eta \rangle^2 \, ds \\
&\quad - \sum_{i=0}^{j} \sum_{v \in I_i} \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi) v(s, X_s), \eta \rangle^2 \, ds \\
&\geq \frac{d^{j+1}}{N^{\gamma_{j+1}}} - \frac{d^j}{N^{\gamma_j}} \geq \frac{d^{j+1}}{2N^{\gamma_{j+1}}}.
\end{aligned}$$

On the other hand, the cardinality of the set $\{v \in I_i, i = 0, 1, \ldots, j\}$ is less than d^{j+2}, for any $j = 0, \ldots, m_0$; thus at least one of the terms in the above sum must be greater than or equal to $\frac{d^{j+1}}{2d^{j+2}N^{\gamma_{j+1}}} = \frac{1}{2dN^{\gamma_{j+1}}}$. Moreover

$$\frac{d^j}{N^{\gamma_j}} = \frac{d^j}{N^{4\gamma_{j+1}-9+9}} \leq \frac{1}{N^{4\gamma_{j+1}-9}},$$

if $N^9 > d^{m_0} > d^j$, for all j. Taking all these remarks into account, we may conclude that

$$\begin{aligned}
E_j(N) \cap E_{j+1}(N)^c \subseteq \bigcup_{i=0}^{j} \bigcup_{v \in I_i} &\left\{ \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi) v(s, X_s), \eta \rangle^2 \, ds \leq \frac{1}{N^{4\gamma_{j+1}-9}} \right\} \\
&\cap \left\{ \sum_{k=1}^{d} \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi) [\bar{b}_k, v](s, X_s), \eta \rangle^2 \, ds > \frac{1}{2dN^{\gamma_{j+1}}} \right\} \\
&\cap \{|\xi| \leq \tau\}.
\end{aligned}$$

The coefficients of (2.2) are differentiable in time and space as many times as needed and consequently so are the vector fields $v \in I_i$, $i = 0, \ldots, m_0$. Applying Itô's lemma and the integration by parts to $\zeta_s^{-1}(\xi) v(s, X_s)$ we obtain

$$\begin{aligned}
&\zeta_s^{-1}(\xi) v(s, X_s) \\
&\quad = v(0, \xi) + \int_0^s \zeta_r^{-1}(\xi) \left\{ \frac{\partial v}{\partial s} + [\bar{a}, v] + \frac{1}{2} \sum_{h=1}^{d} [\bar{b}_h, [\bar{b}_h, v]] \right\} (r, X_r) \, dr \\
&\qquad + \sum_{h=1}^{d} \int_0^s \zeta_r^{-1}(\xi) [\bar{b}_h, v](r, X_r) \, dW_r^h.
\end{aligned}$$

Clearly, $\sum_{h=1}^{d} \zeta_r^{-1}(\xi)[\bar{b}_h, v](r, X_r)$ is the diffusion coefficient associated to $\zeta_s^{-1}(\xi)v(s, X_s)$.

With considerations similar to those employed before, it is possible to show that when $s \le \sigma_N$ and $|\xi| \le \tau$, due to the bounds on R we have for $L = 16d^{m_0+2}M^2 n$,

$$
\begin{aligned}
|\zeta_s^{-1}(\xi)| &\le \frac{3}{2}, \\
|\bar{b}_k(s, X_s)| &\le M\bigl(s + |X_s - \xi| + |\xi| + |b(0,0)|\bigr) \\
&\le M(2R + \tau + M) \le M\left(\frac{c}{L} + \tau + M\right).
\end{aligned}
$$

Similarly we may conclude that there exists a constant λ, depending on n, M, m_0, c, d, so that for $s \le \sigma_N$ and $|\xi| \le \tau$,

$$
|\zeta_s^{-1}(\xi)v(s, X_s)|, \qquad \sum_{h=1}^{d} \zeta_s^{-1}(\xi)[\bar{b}_h, v](s, X_s) \le \lambda(1+\tau) \tag{2.9}
$$

for all $v \in I_i$, with $i = 0, \dots, m_0$. In the same way one can treat the drift associated to $\zeta_s^{-1}(\xi)v(s, X_s)$. Without loss of generality we assume that it is bounded also by $\lambda(1+\tau)$.

If we assume (H2) instead of (H3), this whole argument goes through, with the only difference that estimate (2.9) is valid independently of τ.

As shown in [17], if we apply Theorem 8.26 of [15] (with $R = (2d)^{-1}$, $Q = 1$, $M_1 = M_2 = \lambda(1+\tau)$), we obtain for $m = \gamma_{j+1} \ge 5$,

$$
\begin{aligned}
&P\left(E_j(N) \cap E_{j+1}(N)^c, \sigma_N = \frac{1}{N^3}\right) \\
&\quad\le \sum_{i=0}^{j} \sum_{v \in I_i} P\Biggl(|\xi| \le \tau, \sigma_N = \frac{1}{N^3}, \\
&\qquad\qquad \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi)v(s, X_s), \eta\rangle^2 \, ds \le \frac{1}{N^{4\gamma_{j+1}-9}}, \\
&\qquad\qquad \sum_{k=1}^{d} \int_0^{\sigma_N} \langle \zeta_s^{-1}(\xi)[\bar{b}_k, v](s, X_s), \eta\rangle^2 \, ds > \frac{1}{2dN^{\gamma_{j+1}}}\Biggr) \\
&\quad\le \sum_{i=0}^{j} \sum_{v \in I_i} \sqrt{2}N^{\gamma_{j+1}-5} \exp\left\{-\frac{N}{2^{11}d^2\lambda^2(1+\tau)^2}\right\}.
\end{aligned}
$$

Summarizing, we can conclude for $N \geq \max(C_0(\tau+1)^{2/3}, t^{-1/3}, (2/d)^{\lambda/15}, d^{m_0/9}, (\frac{4d^{m_0}}{3c})^{1/2})$,

$$
\begin{aligned}
&P\left(\sum_{k=1}^{d}\int_0^t \langle \zeta_s^{-1}(\xi)\bar{b}_k(s,X_s),\eta\rangle^2\,ds < \frac{1}{N^{5\cdot 4^{m_0}}}\right)\\
&\quad\leq P(|\xi|\geq\tau) + P\left(\sigma_N < \frac{1}{N^3}\right)\\
&\qquad + \sum_{j=0}^{m_0-1} P\left(E_j(N)\cap E_{j+1}(N)^c \cap \left\{\sigma_N = \frac{1}{N^3}\right\}\right)\\
&\quad\leq C\exp(-\alpha\tau^\beta) + \left(\frac{C_1}{R^q}+C_2\right)2^q\left(\frac{1}{N^3}\right)^{q/2}\\
&\qquad + \sum_{j=0}^{m_0-1}\sum_{i=0}^{j}\sum_{v\in I_i}\sqrt{2}N^{\gamma_{j+1}-5}\exp\left\{-\frac{dN}{2^9 m_0^2\lambda^2(1+\tau)^2}\right\}\\
&\quad\leq C\exp(-\alpha\tau^\beta) + \left(\frac{C_1}{R^q}+C_2\right)2^q\left(\frac{1}{N^3}\right)^{q/2}\\
&\qquad + \sum_{j=0}^{m_0-1} d^{j+1}\sqrt{2}N^{\gamma_{j+1}-5}\exp\left\{-\frac{dN}{2^9 m_0^2\lambda^2(1+\tau)^2}\right\}
\end{aligned}
$$

and the result follows by taking $\tau = O(|\log((\frac{1}{N^3})^{q/\alpha})|^{1/\beta})$, for any q which is equivalent to taking N big enough when considering $N \geq C_0(\tau+1)^{2/3}$. $\square$

We would like to remark that it is exactly the use of the restricted Hörmander condition that makes the above theorem true for time dependent coefficients, since it enables the use of Theorem 8.26 of [15], which involves only the diffusion coefficients of the process and therefore does not call for an unrestricted condition. In fact, one can find a counterexample to the general statement of existence of densities under the general Hörmander condition (see page 310 in [17]).

From the previous theorem we know that there exists a unique solution to (2.1) with smooth density that we denote by $p_t(x)$, which we are eventually interested in approximating. From now on, we restrict to the one-dimensional case. So $n = d = 1$.

In order to relate the unique solution of (2.1) to the McKean–Vlasov equation we recall that, under appropriate conditions (see [16] or [1]), the distribution function of X_t, denoted by $u(t,x)$, satisfies the equation

$$
\begin{aligned}
&\frac{\partial u}{\partial t}(t,x) = \frac{1}{2}\frac{\partial}{\partial x}\left[b^2\big(x, G\big(x;\partial_x u(t,\cdot)\big)\big)\frac{\partial u}{\partial x}(t,x)\right] - a\big(x, F\big(x;\partial_x u(t,\cdot)\big)\big)\frac{\partial u}{\partial x}(t,x),\\
&u(0,x) = P(\xi\leq x).
\end{aligned}
$$

Under the same assumptions of Theorem 2.2 the density function, denoted by $p_t(x) \equiv p(t,x)$, exists, is regular and satisfies the following nonlinear equation:

$$\frac{\partial p}{\partial t}(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\big[b^2\big(x, G\big(x; p(t,\cdot)\big)\big)p(t,x)\big] - \frac{\partial}{\partial x}\big[a\big(x, F\big(x; p(t,\cdot)\big)\big)p(t,x)\big],$$
$$u(0,x) = p_0(x).$$

Therefore, it becomes of interest to approximate both the distribution and density function of X_t for fixed $t > 0$.

To do this, in the next section we introduce a particle method described in Bossy and Talay [1] and [2] and we evaluate the rate of convergence of this method to the solution.

3. Particle method. In this section we describe the actual particle method that we use to approximate $p_t(x)$. In order to do so, we proceed by the following steps.

1. Approximate the density $p_t(x)$ by Gaussian densities; that is,
$$p_t(x) = \int_{\mathbb{R}} \delta_x(y)p_t(y)\,dy \sim \int_{\mathbb{R}} \phi_h(y-x)p_t(y)\,dy = E\big(\phi_h(X_t - x)\big),$$
where $\phi_h(z) = \frac{\exp(-z^2/2h)}{\sqrt{2\pi h}}$.
2. Consider the difference $p_t(x) - E(\phi_h(X_t - x))$.
3. Given a partition $\pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$, which without loss of generality we assume to be uniform with mesh h; that is, $h = \Delta t = t_{i+1} - t_i$ for any i, we define the Euler scheme for (2.1) as
$$Y_t = Y_{\eta(t)} + a\big(Y_{\eta(t)}, F(Y_{\eta(t)}; v_{\eta(t)})\big)\big(t - \eta(t)\big) + b\big(Y_{\eta(t)}, G(Y_{\eta(t)}; v_{\eta(t)})\big)(W_t - W_{\eta(t)}), \tag{3.1}$$
where $\eta(t) = \sup\{t_i \le t \colon t_i \in \pi\}$ and $F(x; v_{\eta(t)}) = \int_{\mathbb{R}} f(x,y)\,dv_{\eta(t)}(y)$, with v_s denoting the distribution of Y_s.
4. Consider the difference $E(\phi_h(X_t - x)) - E(\phi_h(Y_t - x))$.
5. Using n one-dimensional independent Brownian motions, W^i, $i = 1, \dots, n$ independent of W, generate n independent copies of the Euler scheme, which we denote by Y^i and consider the difference,
$$E\big(\phi_h(Y_t - x)\big) - \frac{1}{n}\sum_{j=1}^{n} \phi_h(Y_t^j - x).$$
6. Consider the Euler–weakly interacting particle system given by
$$X_t^i = X_{\eta(t)}^i + a\big(X_{\eta(t)}^i, F(X_{\eta(t)}^i; \bar{\mu}_{\eta(t)})\big)\big(t - \eta(t)\big) + b\big(X_{\eta(t)}^i, G(X_{\eta(t)}^i; \bar{\mu}_{\eta(t)})\big)(W_t^i - W_{\eta(t)}^i), \tag{3.2}$$

where

$$\bar{\mu}_{\eta(t)}(dx) = \frac{1}{n}\sum_{j=1}^{n} \delta_{X^j_{\eta(t)}}(dx).$$

7. Consider the difference

$$\frac{1}{n}\sum_{j=1}^{n} \phi_h(Y_t^j - x) - \frac{1}{n}\sum_{j=1}^{n} \phi_h(X_t^j - x).$$

A similar procedure is followed to analyze the approximations for distributions functions, where the role of ϕ_h is played by its distribution function $\Phi_h(x) = \int_{-\infty}^{x} \phi_h(y)\, dy$. Our aim is to show the following result.

THEOREM 3.1. *Assume* (H0), (H1) *and either* (H2) *or* (H3). *Then for any fixed* $t \in (0, T]$,

$$\int_{\mathbb{R}} E\left(\left|u(t,x) - \frac{1}{n}\sum_{j=1}^{n} 1_{\{X_t^j \le x\}}\right|\right) dx \le C\left(h + \frac{1}{\sqrt{n}}\right), \tag{3.3}$$

$$\int_{\mathbb{R}} E\left(\left|p_t(x) - \frac{1}{n}\sum_{j=1}^{n} \phi_h(X_t^j - x)\right|\right) dx \le C\left(h + \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}h^{1/4}}\right). \tag{3.4}$$

Furthermore, if we choose $n = O(\frac{1}{h})^k$ *for some* $k > 0$, *then for each* $p > 1$, *there exists a positive constant* C_p *independent of* h (*and* n) *such that*

$$\sup_{x\in\mathbb{R}} E\left(\left|u(t,x) - \frac{1}{n}\sum_{j=1}^{n} \Phi_h(X_t^j - x)\right|\right) \le C_p\left(h + \frac{1}{\sqrt{n}h^{(p-1)/2}}\right), \tag{3.5}$$

$$\sup_{x\in\mathbb{R}} E\left(\left|p_t(x) - \frac{1}{n}\sum_{j=1}^{n} \phi_h(X_t^j - x)\right|\right) \le C_p\left(h + \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}h^{1-1/2p}}\right). \tag{3.6}$$

Before moving toward this goal, we need to mention a result from [7], that provides an important tool for the subsequent proofs.

LEMMA 3.2 [7]. *Let* X_t *and* Y_t *be defined, respectively, by* (2.1) *and by* (3.1) *and let condition* (H0) *be fulfilled.*

Then $X_t, Y_t \in \mathbb{D}^\infty$ *for any* $t \in [0, T]$ *and for any* $n = 0, 1, \dots$ *and any fixed* $q \ge 1$ *we have*

$$\sup_{s_1,\dots,s_n\le T}\left\|\sup_{t\le T}|D_{s_1}\cdots D_{s_n}X_t|\right\|_{2q} + \sup_{s_1,\dots,s_n\le T}\left\|\sup_{t\le T}|D_{s_1}\cdots D_{s_n}Y_t|\right\|_{2q} \le C,$$

$$\sup_{s_1,\dots,s_n\le T}\left\|\sup_{t\le T}|D_{s_1}\cdots D_{s_n}(X_t - Y_t)|\right\|_{2q} \le Ch^q,$$

with C *a positive constant that depends only on* M, q, n *and* T.

By virtue of this lemma, we can prove the following result about the Malliavin variance that, later on, will help us establish the convergence rate of the approximations toward the solution.

In the rest of the article we will assume that (H0), (H1) are satisfied and that either one of (H2) or (H3) is satisfied.

LEMMA 3.3. *Let ν be a constant in* $[0, 1]$ *and let* $\bar{W}$ *denote a Wiener process independent of* W, *then for any fixed* $s, t \in (0, T]$, $a \in \mathbb{R}^+$ *and* $p \in \mathbb{N}$, *we have*

$$\sup_{\nu\in[0,1]} \|\nu(Y_t - X_t) + a\bar{W}_s\|_{1,p} \le K_1\sqrt{h} + K_2 a,$$

$$\sup_{h\in(0,1]} \sup_{\nu\in[0,1]} \|\gamma^{-1}_{X_t+\nu(Y_t-X_t)+\sqrt{h}\bar{W}_s}\|_p < \infty.$$

PROOF. Let us denote by $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ the canonical space where $\bar{W}$ lives and let us define the Sobolev norms for the product space $\Omega \times \bar{\Omega}$ in the natural manner, that is to say (having denoted by $P' = P \times \bar{P}$ and $E' = E \times \bar{E}$). Also recall all the definitions of Sobolev norms given before Theorem 2.1 also apply here with $D^1 = D$ and $D^2 = \bar{D}$. We start proving the first inequality.

$$\begin{aligned}
\|\nu(Y_t - X_t) + a\bar{W}_s\|^p_{1,p} &= E'\big(|\nu(Y_t - X_t) + a\bar{W}_s|^p\big) \\
&\quad + E\left[\left(\int_0^T \nu^2 |D_r(Y_t - X_t)|^2\,dr\right)^{p/2}\right] \\
&\quad + \bar{E}\left[\left(\int_0^T a^2 |\bar{D}_r\bar{W}_s|^2\,dr\right)^{p/2}\right] \\
&\le 2^{p-1}\big[\nu^p E(|Y_t - X_t|^p) + a^p \bar{E}(|\bar{W}_s|^p)\big] \\
&\quad + \nu^p E\left[\left(\int_0^T |D_r(Y_t - X_t)|^2\,dr\right)^{p/2}\right] + (a^2 s)^{p/2} \\
&\le C\big(\nu^p \|Y_t - X_t\|^p_{1,p} + a^p\|\bar{W}_s\|^p_{1,p}\big).
\end{aligned}$$

But Lemma 3.2 gives that

$$E\left(\sup_{t\le T}|Y_t - X_t|^p\right) + \sup_{r\le T} E\left(\sup_{t\le T}|D_r(Y_t - X_t)|^p\right) \le C_{1,p}h^{p/2}$$

and applying this estimate in the previous inequality, we get

$$\|\nu(Y_t - X_t) + a\bar{W}_s\|^p_{1,p} \le C_{1,p}h^{\frac{p}{2}} + a^p\|\bar{W}_s\|^p_{1,p},$$

so our inequality is satisfied.

For the second inequality, we subdivide the proof in steps.

Step 1. By Theorem 2.2, we have already proved

$$\|\gamma^{-1}_{X_t}\|_p < \infty. \tag{3.7}$$

Step 2. We want to show

$$\|\gamma^{-1}_{X_t+\nu(Y_t-X_t)+a\bar{W}_s}\|_p \le \frac{1}{s}\frac{1}{a^2}. \tag{3.8}$$

By the definition of Malliavin covariance matrix, we have

$$\begin{aligned}\gamma_{X_t+\nu(Y_t-X_t)+a\bar{W}_s} &= \int_0^T \big|D_r\big(X_t+\nu(Y_t-X_t)\big)\big|^2\,dr + \int_0^T a^2|\bar{D}_r\bar{W}_s|^2\,dr \\ &= \int_0^t |D_rX_t(1-\nu)+\nu D_rY_t|^2\,dr + a^2 s \ge a^2 s,\end{aligned}$$

which gives (3.8).

Step 3. Let us consider the set $A = \{|\gamma_{X_t+\nu(Y_t-X_t)+a\bar{W}_s} - \gamma_{X_t}| \le \frac{1}{2}|\gamma_{X_t}|\}$. We have

$$\begin{aligned}&E'\big(|\gamma^{-1}_{X_t+\nu(Y_t-X_t)+a\bar{W}_s}|^p\big)\\ &\quad = E'\big(|\gamma^{-1}_{X_t+\nu(Y_t-X_t)+a\bar{W}_s}|^p 1_A\big) + E'\big(|\gamma^{-1}_{X_t+\nu(Y_t-X_t)+a\bar{W}_s}|^p 1_{A^c}\big)\\ &\quad \le 2^p E'\big(|\gamma^{-1}_{X_t}|^p 1_A\big) + E'\big(|\gamma^{-1}_{X_t+\nu(Y_t-X_t)+a\bar{W}_s}|^p 1_{A^c}\big)\\ &\quad \le 2^p E'\big(|\gamma^{-1}_{X_t}|^p 1_A\big) + P(A^c)^{1/2} E'\big(|\gamma^{-1}_{X_t+\nu(Y_t-X_t)+a\bar{W}_s}|^{2p}\big)^{1/2}.\end{aligned}$$

From (3.8) we know that $E'(|\gamma^{-1}_{X_t+\nu(Y_t-X_t)+a\bar{W}_s}|^{2p})^{1/2} \le \frac{1}{a^2 s}$, so taking $a = \sqrt{h}$ and using (3.7), we can conclude the proof by noticing that

$$P'(A^c) \le 2^k E'\big(|\gamma^{-1}_{X_t}|^k |\gamma_{X_t+\nu(Y_t-X_t)+h\bar{W}_t} - \gamma_{X_t}|^k\big) \le Ch^{k/2},$$

for any k. Taking k big enough, one obtains the result. □

In the light of the previous lemma, we can consider the first step of our approximation procedure and obtain the following.

LEMMA 3.4. *With the above notation and the hypotheses of Theorem* 3.1, *we have*

$$\sup_{x\in\mathbb{R}} \big|p_t(x) - E\big(\phi_h(X_t-x)\big)\big| \le Ch, \tag{3.9}$$

$$\int_{\mathbb{R}} \big|p_t(x) - E\big(\phi_h(X_t-x)\big)\big|\,dx \le Ch, \tag{3.10}$$

with C independent of h.

PROOF. To evaluate (3.9), as we did in Lemma 3.3 let us consider a Brownian motion $\bar{W}$, independent of the original one and let E' denote the expectation on the canonical product space, while D and $\bar{D}$ are the Malliavin derivatives with respect to W and $\bar{W}$.

The difference in (3.9) can be written as

$$\begin{aligned}p_t(x) - E\big(\phi_h(X_t - x)\big) &= p_t(x) - E'\big(\delta_x(X_t + h^{1/2}\bar{W}_1)\big)\\ &= E'\big[\delta_x(X_t) - \delta_x(X_t + h^{1/2}\bar{W}_1)\big].\end{aligned}$$

But as X_t and $X_t + h^{1/2}\bar{W}_1$ have smooth densities, it is known that $\phi_a(y - x) \to \delta_x(y)$ as $a \to 0$, so the last equality leads to

$$\begin{aligned}p_t(x) - E\big(\phi_h(X_t - x)\big) &= \lim_{a\to 0} E'\big[\phi_a(X_t - x) - \phi_a(X_t + h^{1/2}\bar{W}_1 - x)\big]\\ &= -\lim_{a\to 0} E'\big[\phi'_a(X_t - x)h^{1/2}\bar{W}_1 + \tfrac{1}{2}\phi''_a(\xi_t - x)h\bar{W}_1^2\big]\\ &= -\lim_{a\to 0}\big[h^{1/2}E'\big(\phi'_a(X_t - x)\big)E'(\bar{W}_1)\\ &\qquad + \tfrac{1}{2}E'\big(\phi''_a(\xi_t - x)h\bar{W}_1^2\big)\big]\\ &= -\lim_{a\to 0}\tfrac{1}{2}hE'\big(\phi''_a(\xi_t - x)\bar{W}_1^2\big),\end{aligned}$$

where ξ_t represents a midpoint between X_t and $X_t + h^{1/2}\bar{W}_1$ and we used the independence between X and $\bar{W}$.

We remark that for any smooth function f we may rewrite the mean value theorem for two random variables M and N as

$$f(M) - f(N) = \int_0^1 f'\big(M + \nu(N - M)\big)\,d\nu(M - N). \tag{3.11}$$

In our case, using Fubini's theorem, we have

$$E'\big(\phi''_a(\xi_t - x)\bar{W}_1^2\big) = \int_0^1 E'\big(\phi''_a(X_t + \nu\sqrt{h}\bar{W}_1 - x)\bar{W}_1^2\big)\,d\nu.$$

In our case, applying the integration by parts formula (2.3), we obtain

$$\begin{aligned}&E'\big(\phi''_a(\xi_t - x)\bar{W}_1^2\big)\\ &\quad = \int_0^1 E'\big(\phi''_a(X_t + \nu\sqrt{h}\bar{W}_1 - x)\bar{W}_1^2\big)\,d\nu\\ &\quad = \int_0^1 E'\big(\Phi_a(X_t + \nu\sqrt{h}\bar{W}_1 - x)H_3(X_t + \nu\sqrt{h}\bar{W}_1 - x, \bar{W}_1^2)\big)\,d\nu,\end{aligned}$$

where by Φ_a we mean the Gaussian distribution function with density ϕ_a. By definition, H is independent of x and $0 \le \Phi_a \le 1$, so from (2.4) for some constants k, b, b', q and q' we may conclude that

$$\begin{aligned}&\big|E'\big(\phi''_a(\xi_t - x)\bar{W}_1^2\big)\big|\\ &\quad \le \int_0^1 E'\big(\Phi_a(X_t + \nu\sqrt{h}\bar{W}_1 - x)|H_3(X_t + \nu\sqrt{h}\bar{W}_1, \bar{W}_1^2)|\big)\,d\nu\\ &\quad \le C\int_0^1 \|\gamma^{-1}_{X_t+\nu\sqrt{h}\bar{W}_1}\|_k^q \|X_t + \nu\sqrt{h}\bar{W}_1\|_{4,b}^{q'}\|\bar{W}_1^2\|_{3,b'}\,d\nu.\end{aligned}$$

By Theorem 2.1 and Lemma 3.3, $\|\gamma^{-1}_{X_t+\nu\sqrt{h}\bar{W}_1}\|_k \le \|\gamma^{-1}_{X_t}\|_k$ is bounded uniformly in ν and h. Moreover $\|\bar{W}_1^2\|_{3,b'} < \infty$ and $\|X_t + \nu\sqrt{h}\bar{W}_1\|_{4,b} \le \|X_t\|_{4,b} + \|\nu\sqrt{h}\bar{W}_1\|_{4,b}$ and the two terms are bounded, the first because of Lemma 3.2, the second can be bounded independently of ν and h, if we assume without loss of generality that $h \le 1$.

Consequently we may conclude that there exists a constant C independent of h, a and x such that $|E'(\phi''_a(\xi_t - x)\bar{W}_1^2)| \le C$, that implies

$$\left|p_t(x) - E\big(\phi_h(X_t - x)\big)\right| \le \tfrac{1}{2}Ch,$$

and concludes the proof of (3.9).

It remains to show inequality (3.10). We have

$$\begin{aligned}
&\int_{\mathbb{R}} \left|p_t(x) - E\big(\phi_h(X_t - x)\big)\right| dx \\
&\quad = \int_{\mathbb{R}} \left|\lim_{a\to 0} \frac{h}{2} E'\big(\phi''_a(\xi_t - x)\bar{W}_1^2\big)\right| dx \\
&\quad = \frac{h}{2}\int_{\mathbb{R}} \left|\lim_{a\to 0} E'\left(\int_0^1 \phi''_a(X_t + \nu\sqrt{h}\bar{W}_1 - x)\, d\nu\, \bar{W}_1^2\right)\right| dx \\
&\quad \le \frac{h}{2}\int_{\mathbb{R}} \lim_{a\to 0}\int_0^1 \left|E'\big(\phi''_a(X_t + \nu\sqrt{h}\bar{W}_1 - x)\bar{W}_1^2\big)\right| d\nu\, dx \\
&\quad = \frac{h}{2}\lim_{a\to 0}\int_{\mathbb{R}}\int_0^1 \left|E'\big(\phi_a(X_t + \nu\sqrt{h}\bar{W}_1 - x) H_2(X_t + \nu\sqrt{h}\bar{W}_1, \bar{W}_1^2)\big)\right| d\nu\, dx \\
&\quad \le \frac{h}{2}\int_0^1 E'\big(|H_2(X_t + \nu\sqrt{h}\bar{W}_1, \bar{W}_1^2)|\big)\, d\nu.
\end{aligned}$$

In order to assure the interchange between the limit and the integral in the fourth passage, we are going to show that the family of functions is uniformly integrable. This will conclude the proof of (3.10).

Uniform square integrability suffices, so we want to prove that

$$\sup_{a\in(0,1]}\int_{\mathbb{R}}\int_0^1 \left|E'\big(\phi_a(X_t + \nu\sqrt{h}\bar{W}_1 - x) H_2(X_t + \nu\sqrt{h}\bar{W}_1, \bar{W}_1^2)\big)\right|^2 d\nu\, dx < \infty,$$

by exploiting the classical estimates on the exponential tails of the Gaussian density. For fixed $K \in \mathbb{R}^+$, let us divide the integral into two pieces,

$$\int_{\mathbb{R}} = \int_{\{|x|\le K\}} + \int_{\{|x|>K\}} = I_1 + I_2.$$

Using the same proof as for (3.9) we have that $\sup_{a\in(0,1]} I_1 < 2KC_1^2$. For I_2, let us consider $A = \{|X_t + \nu\sqrt{h}\bar{W}_1| < \frac{|x|}{2}\}$ and let us notice that if we consider the function $\Psi_a(x) = -(1 - \Phi_a(x))1_{\{x>0\}} + \Phi_a(x)1_{\{x\le 0\}}$, then $\Psi'_a(x) = \phi_a(x)$, hence by applying the integration by parts, I_2 can be rewritten as follows:

$$I_2 = \int_{\{|x|>K\}} \int_0^1 \big|E'\big(\Psi_a(X_t + \nu\sqrt{h}\bar{W}_1 - x) \times H_3(X_t + \nu\sqrt{h}\bar{W}_1, \bar{W}_1^2)(1_A + 1_{A^c})\big)\big|^2 \, d\nu\, dx$$

$$\le 2\int_{\{|x|>K\}} \int_0^1 \big|E'\big(\Psi_a(X_t + \nu\sqrt{h}\bar{W}_1 - x) \times H_3(X_t + \nu\sqrt{h}\bar{W}_1, \bar{W}_1^2)1_A\big)\big|^2 \, d\nu\, dx$$

$$+ 2\int_{\{|x|>K\}} \int_0^1 \big|E'\big(\Psi_a(X_t + \nu\sqrt{h}\bar{W}_1 - x) \times H_3(X_t + \nu\sqrt{h}\bar{W}_1, \bar{W}_1^2)1_{A^c}\big)\big|^2 \, d\nu\, dx.$$

On A we have that $|X_t + \nu\sqrt{h}\bar{W}_1 - x| > \frac{|x|}{2}$, thus for $|x|$ large enough, we can use the estimate

$$\Psi_a(X_t + \nu\sqrt{h}\bar{W}_1 - x) \le \exp\left(-\frac{x^2}{8a}\right),$$

so that

$$\int_{\{|x|>K\}} \int_0^1 \big|E'\big(\Psi_a(X_t + \nu\sqrt{h}\bar{W}_1 - x) H_3(X_t + \nu\sqrt{h}\bar{W}_1, \bar{W}_1^2)1_A\big)\big|^2 \, d\nu\, dx$$

$$\le \int_{\{|x|>K\}} \exp\left(-\frac{x^2}{4a}\right) \int_0^1 E'\big(|H_3(X_t + \nu\sqrt{h}\bar{W}_1, \bar{W}_1^2)|^2\big)\, d\nu\, dx \le C < \infty \qquad \forall\, a \in (0, 1].$$

On A^c, it is enough to apply Chebyshev's inequality to obtain that

$$\int_{\{|x|>K\}} \int_0^1 \big|E'\big(\Psi_a(X_t + \nu\sqrt{h}\bar{W}_1 - x) H_3(X_t + \nu\sqrt{h}\bar{W}_1, \bar{W}_1^2)1_{A^c}\big)\big|^2 \, d\nu\, dx$$

$$\le \int_{\{|x|>K\}} \int_0^1 E'\big(|H_3(X_t + \nu\sqrt{h}\bar{W}_1, \bar{W}_1^2)|^2\big) P(A^c)\, d\nu\, dx$$

$$\le \int_{\{|x|>K\}} \int_0^1 E'\big(|H_3(X_t + \nu\sqrt{h}\bar{W}_1, \bar{W}_1^2)|^2\big) \times \frac{2^k}{|x|^k} E(|X_t + \nu\sqrt{h}\bar{W}_1|^k)\, d\nu\, dx < C < \infty,$$

for $k > 1$ and all $a \in (0, 1]$. $\square$

We can now begin the second step of our procedure. This is rather more complicated than the first and it needs several lemmas for its proof. The main ingredient is a cumbersome integration by parts result which is given in Lemma A.1 in the Appendix. The main result for the second step is summarized as follows.

THEOREM 3.5. *Under the same hypotheses as in Theorem* 3.1, *the following holds*:

$$\sup_{x\in\mathbb{R}}\left|E\big(\phi_h(X_t-x)-\phi_h(Y_t-x)\big)\right|\le Ch \tag{3.12}$$

with C independent of h.

PROOF. By the mean value theorem, we have

$$\begin{aligned}
&E\big(\phi_h(X_t-x)-\phi_h(Y_t-x)\big)\\
&\qquad=E'\big(\phi_{h/2}(X_t+\sqrt{h}\bar{W}_{1/2}-x)-\phi_{h/2}(Y_t+\sqrt{h}\bar{W}_{1/2}-x)\big)\\
&\qquad=E'\big(\phi'_{h/2}(\xi_t^1+\sqrt{h}\bar{W}_{1/2}-x)(X_t-Y_t)\big),
\end{aligned}$$

where ξ_t^1 is a random midpoint between X_t and Y_t. We will repeat this procedure a few times, keeping in mind that the coefficients F and G are smooth as they inherit this property from the kernels f and g. Therefore the idea is to add and subtract the proper terms and apply the mean value theorem to each difference in the expression for $X-Y$. We obtain

$$\begin{aligned}
X_t-Y_t=&\int_0^t\big\{a_x\big(\xi_s^2,F(X_s;\mu_s)\big)(X_s-Y_s)\\
&\qquad+a_y(Y_s,\eta_s^1)[F(X_s;\mu_s)-F(Y_s;\upsilon_s)]\big\}\,ds\\
&+\int_0^t\big\{a_x\big(\zeta_s^1,F(Y_s;\upsilon_s)\big)(Y_s-Y_{\eta(s)})\\
&\qquad+a_y(Y_{\eta(s)},\theta_s^1)[F(Y_s;\upsilon_s)-F(Y_{\eta(s)};\upsilon_{\eta(s)})]\big\}\,ds\\
&+\int_0^t\big\{b_x\big(\xi_s^3,G(X_s;\mu_s)\big)(X_s-Y_s)\\
&\qquad+b_y(Y_s,\eta_s^2)[G(X_s;\mu_s)-G(Y_s;\upsilon_s)]\big\}\,dW_s\\
&+\int_0^t\big\{b_x\big(\zeta_s^2,G(Y_s;\upsilon_s)\big)(Y_s-Y_{\eta(s)})\\
&\qquad+b_y(Y_{\eta(s)},\theta_s^2)[G(Y_s;\upsilon_s)-G(Y_{\eta(s)};\upsilon_{\eta(s)})]\big\}\,dW_s
\end{aligned}$$

with midpoints [intended in the sense of formula (3.11)],

$$\xi_s^2, \xi_s^3 \in [X_s; Y_s], \qquad \zeta_s^1, \zeta_s^2 \in [Y_s; Y_{\eta(s)}],$$
$$\eta_s^1 \in [F(X_s; \mu_s); F(Y_s; \upsilon_s)], \qquad \eta_s^2 \in [G(X_s; \mu_s); G(Y_s; \upsilon_s)],$$
$$\theta_s^1 \in [F(Y_s; \upsilon_s); F(Y_{\eta(s)}, \upsilon_{\eta(s)})], \qquad \theta_s^2 \in [G(Y_s; \upsilon_s); G(Y_{\eta(s)}, \upsilon_{\eta(s)})],$$

where we adopted $[V; Z]$ as standard notation to indicate the interval with random variables Z, V as endpoints. By adding and subtracting $F(Y_s; \mu_s)$ in the second term of the first time integral, $G(Y_s; \mu_s)$ in the second term of the first Brownian integral, and applying once again the mean value theorem to those, we get

$$\begin{aligned}
X_t - Y_t = &\int_0^t \bigl[a_x\bigl(\xi_s^2, F(X_s; \mu_s)\bigr) + a_y(Y_s, \eta_s^1)F'(\xi_s^4; \mu_s)\bigr](X_s - Y_s)\,ds \\
&+ \int_0^t \bigl\{a_y(Y_s, \eta_s^1)[F(Y_s; \mu_s) - F(Y_s; \upsilon_s)]\bigr\}\,ds \\
&+ \int_0^t \bigl\{a_x\bigl(\zeta_s^1, F(Y_s; \upsilon_s)\bigr)(Y_s - Y_{\eta(s)}) \\
&\qquad + a_y(Y_{\eta(s)}, \theta_s^1)[F(Y_s; \upsilon_s) - F(Y_{\eta(s)}; \upsilon_{\eta(s)})]\bigr\}\,ds \\
&+ \int_0^t \bigl[b_x\bigl(\xi_s^3, G(X_s; \mu_s)\bigr) + b_y(Y_s, \eta_s^2)G'(\xi_s^5; \mu_s)\bigr](X_s - Y_s)\,dW_s \\
&+ \int_0^t \bigl\{b_y(Y_s, \eta_s^2)[G(Y_s; \mu_s) - G(Y_s; \upsilon_s)]\bigr\}\,dW_s \\
&+ \int_0^t \bigl\{b_x\bigl(\zeta_s^2, G(Y_s; \upsilon_s)\bigr)(Y_s - Y_{\eta(s)}) \\
&\qquad + b_y(Y_{\eta(s)}, \theta_s^2)[G(Y_s; \upsilon_s) - G(Y_{\eta(s)}; \upsilon_{\eta(s)})]\bigr\}\,dW_s,
\end{aligned}$$

with $\xi_s^4, \xi_s^5 \in [X_s; Y_s]$. For simplicity of notation, from now on we set

$$\begin{aligned}
\alpha_s &= \bigl[a_x\bigl(\xi_s^2, F(X_s; \mu_s)\bigr) + a_y(Y_s, \eta_s^1)F'(\xi_s^4; \mu_s)\bigr], \\
\beta_s &= \bigl[b_x\bigl(\xi_s^3, G(X_s; \mu_s)\bigr) + b_y(Y_s, \eta_s^2)G'(\xi_s^5; \mu_s)\bigr], \\
H_t &= \int_0^t \bigl\{a_y(Y_s, \eta_s^1)[F(Y_s; \mu_s) - F(Y_s; \upsilon_s)] \\
&\qquad + a_y(Y_{\eta(s)}, \theta_s^1)[F(Y_s; \upsilon_s) - F(Y_{\eta(s)}; \upsilon_{\eta(s)})]\bigr\}\,ds \\
&\quad + \int_0^t \bigl\{b_y(Y_s, \eta_s^2)[G(Y_s; \mu_s) - G(Y_s; \upsilon_s)] \\
&\qquad + b_y(Y_{\eta(s)}, \theta_s^2)[G(Y_s; \upsilon_s) - G(Y_{\eta(s)}; \upsilon_{\eta(s)})]\bigr\}\,dW_s, \\
dK_s &= a_x\bigl(\zeta_s^1, F(Y_s; \upsilon_s)\bigr)\,ds + b_x\bigl(\zeta_s^2, G(Y_s; \upsilon_s)\bigr)\,dW_s, \qquad K_0 = 0.
\end{aligned}$$

With this new notation, the above equation becomes

$$X_t - Y_t = \int_0^t (X_s - Y_s)(\alpha_s\,ds + \beta_s\,dW_s) + H_t + \int_0^t (Y_s - Y_{\eta(s)})\,dK_s,$$

whose explicit solution is given by

$$X_t - Y_t = \mathcal{E}_t \int_0^t \mathcal{E}_s^{-1}\biggl\{dH_s + (Y_s - Y_{\eta(s)})\,dK_s - d\biggl[\int_0^\bullet \beta_r\,dW_r, H + (Y - Y_\eta)\cdot K\biggr]_s\biggr\}, \tag{3.13}$$

where $\mathcal{E}_t$ denotes

$$\exp\biggl(\int_0^t (\alpha_s - \beta_s^2/2)\,ds + \int_0^t \beta_s\,dW_s\biggr).$$

To simplify further and to regroup the terms in ds and dW_s, we consider the process $U_t = \mathcal{E}_t^{-1}(X_t - Y_t)$. With a few computations, from the definition of H, (3.13) can be rewritten as

$$\begin{aligned}
U_t = {} & \int_0^t \mathcal{E}_s^{-1}(Y_s - Y_{\eta(s)})\bigl[dK_s - \beta_s b_x\bigl(\zeta_s^2, G(Y_s; \upsilon_s)\bigr)\,ds\bigr] \\
& + \int_0^t \mathcal{E}_s^{-1}\bigl\{a_y(Y_s, \eta_s^1)[F(Y_s; \mu_s) - F(Y_s; \upsilon_s)] \\
& \qquad + a_y(Y_{\eta(s)}, \theta_s^1)[F(Y_s; \upsilon_s) - F(Y_{\eta(s)}; \upsilon_{\eta(s)})] \\
& \qquad - \beta_s\bigl(b_y(Y_s, \eta_s^2)[G(Y_s; \mu_s) - G(Y_s; \upsilon_s)] \\
& \qquad + b_y(Y_{\eta(s)}, \theta_s^2)[G(Y_s; \upsilon_s) - G(Y_{\eta(s)}; \upsilon_{\eta(s)})]\bigr)\bigr\}\,ds \\
& + \int_0^t \mathcal{E}_s^{-1}\bigl\{b_y(Y_s, \eta_s^2)[G(Y_s; \mu_s) - G(Y_s; \upsilon_s)] \\
& \qquad + b_y(Y_{\eta(s)}, \theta_s^2)[G(Y_s; \upsilon_s) - G(Y_{\eta(s)}; \upsilon_{\eta(s)})]\bigr\}\,dW_s.
\end{aligned}$$

The differences in F and G can be reformulated making use of their respective kernels. Indeed, if we introduce independent copies of X and Y, say $\tilde{X}$ and $\tilde{Y}$ and the canonical space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ where they live, we can look at those differences in the following manner:

$$\begin{aligned}
F(Y_s; \mu_s) - F(Y_s; \upsilon_s) &= \tilde{E}\bigl(f(Y_s, \tilde{X}_s)\bigr) - \tilde{E}\bigl(f(Y_s, \tilde{Y}_s)\bigr) \\
&= \tilde{E}\bigl(f_y(Y_s, \tilde{\xi}_s^1)(\tilde{X}_s - \tilde{Y}_s)\bigr), \\
G(Y_s; \mu_s) - G(Y_s; \upsilon_s) &= \tilde{E}\bigl(g_y(Y_s, \tilde{\xi}_s^2)(\tilde{X}_s - \tilde{Y}_s)\bigr), \\
F(Y_s; \upsilon_s) - F(Y_{\eta(s)}; \upsilon_{\eta(s)}) &= \tilde{E}\bigl(f_x(\zeta_s^3, \tilde{Y}_{\eta(s)})(Y_s - Y_{\eta(s)}) \\
&\qquad + f_y(Y_s, \tilde{\zeta}_s^1)(Y_s - Y_{\eta(s)})\bigr),
\end{aligned}$$

$$G(Y_s;v_s)-G(Y_{\eta(s)};v_{\eta(s)})=\tilde E\big(g_x(\zeta_s^4,\tilde Y_{\eta(s)})(Y_s-Y_{\eta(s)})\\ +g_y(Y_s,\tilde\zeta_s^2)(Y_s-Y_{\eta(s)})\big),$$

where $\tilde E$ denotes the expectation in $(\tilde\Omega,\tilde{\mathcal F},\tilde P)$ and where we used once again the mean value theorem, with $\tilde\xi_s^1,\tilde\xi_s^2\in[\tilde X_s;\tilde Y_s]$ and $\zeta_s^3,\zeta_s^4\in[Y_{\eta(s)};Y_s]$ and $\tilde\zeta_s^1,\tilde\zeta_s^2\in[\tilde Y_{\eta(s)};\tilde Y_s]$. Similarly, if we take an independent copy of $\mathcal E$, say $\tilde{\mathcal E}$, the above equation for U_t is transformed into

$$\begin{aligned}U_t&=\int_0^t\mathcal E_s^{-1}(Y_s-Y_{\eta(s)})\big[dK_s-\beta_sb_x\big(\zeta_s^2,G(Y_s;v_s)\big)\,ds\big]\\&\quad+\int_0^t\mathcal E_s^{-1}\big[a_y(Y_s,\eta_s^1)\tilde E\big(f_y(Y_s,\tilde\xi_s^1)\tilde{\mathcal E}_s\tilde U_s\big)\\&\qquad\qquad-\beta_sb_y(Y_s,\eta_s^2)\tilde E\big(g_y(Y_s,\tilde\xi_s^2)\tilde{\mathcal E}_s\tilde U_s\big)\big]\,ds\\&\quad+\int_0^t\mathcal E_s^{-1}(Y_s-Y_{\eta(s)})\big[a_y(Y_{\eta(s)},\theta_s^1)\tilde E\big(f_x(\zeta_s^3,\tilde Y_{\eta(s)})\big)\\&\qquad\qquad-\beta_sb_y(Y_{\eta(s)},\theta_s^2)\tilde E\big(g_x(\zeta_s^4,\tilde Y_{\eta(s)})\big)\big]\,ds\\&\quad+\int_0^t\mathcal E_s^{-1}\big[a_y(Y_{\eta(s)},\theta_s^1)\tilde E\big(f_y(Y_s,\tilde\zeta_s^1)(\tilde Y_{\eta(s)}-\tilde Y_s)\big)\\&\qquad\qquad-\beta_sb_y(Y_{\eta(s)},\theta_s^2)\tilde E\big(g_y(Y_s,\tilde\zeta_s^2)(\tilde Y_{\eta(s)}-\tilde Y_s)\big)\big]\,ds\\&\quad+\int_0^t\mathcal E_s^{-1}b_y(Y_s,\eta_s^2)\tilde E\big(g_y(Y_s,\tilde\xi_s^2)\tilde{\mathcal E}_s\tilde U_s\big)\,dW_s\\&\quad+\int_0^t\mathcal E_s^{-1}b_y(Y_{\eta(s)},\theta_s^2)\big[(Y_s-Y_{\eta(s)})\tilde E\big(g_x(\zeta_s^4,\tilde Y_{\eta(s)})\big)\\&\qquad\qquad+\tilde E\big(g_y(Y_s,\tilde\zeta_s^2)(\tilde Y_{\eta(s)}-\tilde Y_s)\big)\big]\,dW_s.\end{aligned}$$

We are finally in position to rearrange the terms and obtain a simpler form for (3.13),

$$\begin{aligned}U_t&=\int_0^t\mathcal E_s^{-1}\big[a_y(Y_s,\eta_s^1)\tilde E\big(f_y(Y_s,\tilde\xi_s^1)\tilde{\mathcal E}_s\tilde U_s\big)\\&\qquad-\beta_sb_y(Y_s,\eta_s^2)\tilde E\big(g_y(Y_s,\tilde\xi_s^2)\tilde{\mathcal E}_s\tilde U_s\big)\big]\,ds\\&\quad+\int_0^t\mathcal E_s^{-1}b_y(Y_s,\eta_s^2)\tilde E\big(g_y(Y_s,\tilde\xi_s^2)\tilde{\mathcal E}_s\tilde U_s\big)\,dW_s+\int_0^t\mathcal E_s^{-1}\,dZ_s,\end{aligned}\tag{3.14}$$

where we set

$$\begin{aligned}dZ_s&=(Y_s-Y_{\eta(s)})(A_s\,ds+B_sdW_s)\\&\qquad+\tilde E\big((\tilde Y_s-\tilde Y_{\eta(s)})\tilde A_s\big)\,ds+\tilde E\big((\tilde Y_s-\tilde Y_{\eta(s)})\tilde B_s\big)\,dW_s,\\B_s&=b_x\big(\zeta_s^2,G(Y_s;v_s)\big)+b_y(Y_{\eta(s)},\theta_s^2)\tilde E\big(g_x(\zeta_s^4,\tilde Y_{\eta(s)})\big),\end{aligned}$$

$$A_s = a_x(\zeta_s^1, F(Y_s; v_s)) + a_y(Y_{\eta(s)}, \theta_s^1)\tilde{E}(f_x(\zeta_s^3, \tilde{Y}_{\eta(s)})) - \beta_s B_s,$$
$$\tilde{B}_s = b_y(Y_{\eta(s)}, \theta_s^2) g_y(Y_s, \tilde{\zeta}_s^2),$$
$$\tilde{A}_s = a_y(Y_{\eta(s)}, \theta_s^1) f_y(Y_s, \tilde{\zeta}_s^1) - \beta_s \tilde{B}_s.$$

It is easy to show that (3.14) has a unique solution and that the sequence of iterates defined as

$$\begin{aligned} U_k(t) = \int_0^t &\mathcal{E}_s^{-1}\tilde{E}\Big(\big[a_y(Y_s,\eta_s^1) f_y(Y_s,\tilde{\xi}_s^1) \\ &\qquad - \beta_s b_y(Y_s,\eta_s^2) g_y(Y_s,\tilde{\xi}_s^2)\big]\tilde{\mathcal{E}}_s \tilde{U}_{k-1}(s)\Big)\,ds \\ &+ \int_0^t \mathcal{E}_s^{-1} b_y(Y_s,\eta_s^2)\tilde{E}\big(g_y(Y_s,\tilde{\xi}_s^2)\tilde{\mathcal{E}}_s\tilde{U}_{k-1}(s)\big)\,dW_s + U_0(t), \\ U_0(t) = \int_0^t &\mathcal{E}_s^{-1}\,dZ_s \end{aligned} \tag{3.15}$$

for $k = 1, \dots,$ converges to the solution (see [7]).

In Lemma B.2 in Appendix B it is proved that there exists a constant R independent of k, x and t such that

$$\big|E'\big(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U_k(t)\big)\big| \le h\sum_{j=1}^{k}\frac{(Rt)^j}{j!}. \tag{3.16}$$

Then by the dominated convergence theorem, this implies that

$$\begin{aligned} &\big|E'\big(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U(t)\big)\big| \\ &\qquad = \lim_{k\to\infty}\big|E'\big(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U_k(t)\big)\big| \le h\mathrm{e}^{RT} \end{aligned}$$

and the theorem is proved. □

We now want to establish the same result as Theorem 3.5, for the L^1 norm.

THEOREM 3.6. *Under the same hypotheses as Theorem* 3.1, *the following inequality holds*:

$$\int_{\mathbb{R}}\big|E\big(\phi_h(X_t - x) - \phi_h(Y_t - x)\big)\big|\,dx \le Ch,$$

with C independent of h.

PROOF. Since the proof is a slight modification of that of Theorem 3.5, we are going to sketch it only. By following exactly the same steps as before, we have by dominated convergence theorem that

$$\begin{aligned}
&\int_{\mathbb{R}} \big| E\big(\phi_h(X_t - x) - \phi_h(Y_t - x)\big)\big|\, dx \\
&\quad = \int_{\mathbb{R}} \big| E'\big(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)(X_t - Y_t)\big)\big|\, dx \\
&\quad = \int_{\mathbb{R}} \big| E'\big(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U_t\big)\big|\, dx \\
&\quad = \lim_{k\to\infty} \int_{\mathbb{R}} \big| E'\big(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U_k(t)\big)\big|\, dx.
\end{aligned}$$

One starts by dividing the above integral in two regions,

$$\int_{\mathbb{R}} (1_{\{|x|\le K\}} + 1_{\{|x|>K\}}) \big| E'\big(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U_k(t)\big)\big|\, dx.$$

Interpreting the midpoint ξ^1 in the notation of (3.11), from now on, we use the notation $Z_t^{\nu,\bar{W}} = X_t + \nu(Y_t - X_t) + \sqrt{h}\bar{W}_{1/2}$.

For the region $|x| \le K$ one uses the previous results. For the complementary region one uses the result explained in Remark A.2(c) of Appendix A. In fact one can explain the result in the remark by noting that

$$\big| E'\big(\phi'_{h/2}(Z_t^{\nu,\bar{W}} - x)\mathcal{E}_t U_0(t) 1_B\big)\big| \le At \exp\left(-\frac{x^2}{4h}\right),$$

where A is a constant depending on M and T and $B = \{|Z_t^{\nu,\bar{W}} - x| > \frac{|x|}{2}\}$. Similarly, one also obtains that for fixed $k \in \mathbb{N}$,

$$\big| E'\big(\phi'_{h/2}(Z_t^{\nu,\bar{W}} - x)\mathcal{E}_t U_0(t) 1_B\big)\big| \le \frac{At}{|x|^k}.$$

Then we may conclude that

$$\int_{\mathbb{R}} \big| E'\big(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_1 - x)\mathcal{E}_t U_0(t)\big)\big|\, dx \le Aht,$$

with A a constant independent of t, h and U_0. We then proceed by induction on k. That is, using the definition of U_k in (3.15), we arrive at the following inequality:

$$\begin{aligned}
&\int_{\mathbb{R}} \big| E'\big(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U_k(t)\big)\big|\, dx \\
&\quad \le Aht + \int_{\mathbb{R}} \int_0^1 \int_0^1 \int_0^1 \int_0^t \bigg| E'''\bigg[\phi'_{h/2}(Z_t^{\nu,\bar{W}} - x) \\
&\qquad\qquad \times \sum_{i=1}^4 Z_{k-1}^i(t,s)\bigg]\, ds\bigg|\, d\nu\, d\mu_1\, d\mu_2\, dx,
\end{aligned}$$

where, following the same calculation as in the previous lemmas, we set with a somewhat simplified notation,

$$Z^1_{k-1}(t,s) = \mathcal{E}_t\mathcal{E}_s^{-1}a_y(Y_s, F_s^{\mu_2})\tilde{E}\big(f_y(Y_s,\tilde{Z}_s^{\mu_1})\tilde{\mathcal{E}}_s\tilde{U}_{k-1}(s)\big),$$
$$Z^2_{k-1}(t,s) = \mathcal{E}_t\mathcal{E}_s^{-1}\beta_s b_y(Y_s, G_s^{\mu_2})\tilde{E}\big(g_y(Y_s,\tilde{Z}_s^{\mu_1})\tilde{\mathcal{E}}_s\tilde{U}_{k-1}(s)\big),$$
$$Z^3_{k-1}(t,s) = D_s\mathcal{E}_t\mathcal{E}_s^{-1}b_y(Y_s, G_s^{\mu_2})\tilde{E}\big(g_y(Y_s,\tilde{Z}_s^{\mu_1})\tilde{\mathcal{E}}_s\tilde{U}_{k-1}(s)\big),$$
$$Z^4_{k-1}(t,s) = D_s Z_t^{\nu,\bar{W}}\mathcal{E}_t\mathcal{E}_s^{-1}b_y(Y_s, G_s^{\mu_2})\tilde{E}\big(g_y(Y_s,\tilde{Z}_s^{\mu_1})\tilde{\mathcal{E}}_s\tilde{U}_{k-1}(s)\big).$$

Then for each of the four terms above one has to apply Lemma B.2 in the case that $|x| \le K$. For the case $|x| > K$ one has to apply the result in Remark A.2(c) of Appendix A. The proof is complete by checking that the constant is uniformly bounded, therefore allowing the definition of R as in the proof of Theorem 3.5. □

We can pass to the next step in our procedure and consider the difference

$$E\left(\left|E\big(\phi_h(Y_t - x)\big) - \frac{1}{n}\sum_{j=1}^n \phi_h(Y_t^j - x)\right|\right),$$

where the Y^j are independent copies of Y. By using the strong law of large numbers we have that the difference converges to zero almost surely as $n \to \infty$ for fixed h. Moreover, we can find the rate of convergence in $L^1(P)$; in fact,

$$\left|E\big(\phi_h(Y_t - x)\big) - \frac{1}{n}\sum_{j=1}^n \phi_h(Y_t^j - x)\right| = \left|\frac{1}{n}\sum_{j=1}^n\big[E\big(\phi_h(Y_t - x)\big) - \phi_h(Y_t^j - x)\big]\right|$$

thus, by taking into account the independence of the copies, formula (3.7), Lemmas 3.2 and 3.3 and the boundedness of Φ_h, we obtain

$$\begin{aligned}
&E\left[\left(\frac{1}{n}\sum_{j=1}^n\big[E\big(\phi_h(Y_t - x)\big) - \phi_h(Y_t^j - x)\big]\right)^2\right] \\
&\quad \le \frac{1}{n^2}\sum_{j=1}^n E\big(\phi_h(Y_t - x)\big)^2 = \frac{E'(\phi_{h/4}(Y_t + \sqrt{h}\bar{W}_{1/4} - x))}{2\sqrt{\pi h}n} \\
&\quad = \frac{C}{\sqrt{h}n}\big|E'\big(\Phi_{h/4}(Y_t + \sqrt{h}\bar{W}_{1/4} - x)H(Y_t + \sqrt{h}\bar{W}_{1/4}, 1)\big)\big| \\
&\quad \le \frac{C}{\sqrt{h}n}\|\gamma^{-1}_{Y_t + \sqrt{h}\bar{W}_{1/4}}\|_a\|Y_t + \sqrt{h}\bar{W}_{1/4}\|_{1,b} \le \frac{C}{\sqrt{h}n}
\end{aligned}$$

for some a, b positive constants and for all $x \in \mathbb{R}$. Consequently,

$$\sup_{x\in\mathbb{R}} E\left(\left|E\bigl(\phi_h(Y_t - x)\bigr) - \frac{1}{n}\sum_{j=1}^{n}\phi_h(Y_t^j - x)\right|\right) \le \frac{C}{\sqrt{\sqrt{h}n}}. \tag{3.17}$$

Also we have, employing the same argument as the one at the end of the proof of Lemma 3.4,

$$\begin{aligned}&\int_{\mathbb{R}} E\left(\left|E\bigl(\phi_h(Y_t - x)\bigr) - \frac{1}{n}\sum_{j=1}^{n}\phi_h(Y_t^j - x)\right|\right) dx \\ &\quad \le \left(\frac{1}{2\sqrt{\pi h n}}\right)^{1/2} \int_{\mathbb{R}} \Bigl(E'\bigl(\phi_{h/4}(Y_t + \sqrt{h}\bar{W}_{1/4} - x)\bigr)\Bigr)^{1/2} dx \le \frac{C}{\sqrt{\sqrt{h}n}}.\end{aligned} \tag{3.18}$$

Indeed, for fixed $K > 0$, by integration by parts

$$\begin{aligned}&\int_{\{|x|\le K\}} \bigl(E'\bigl(\phi_{h/4}(Y_t + \sqrt{h}\bar{W}_{1/4} - x)\bigr)\bigr)^{1/2} dx \\ &\qquad = \int_{\{|x|\le K\}} \bigl|E'\bigl(\Phi_{h/4}(Y_t + \sqrt{h}\bar{W}_{1/4} - x) H(Y_t + \sqrt{h}\bar{W}_{1/4}), 1\bigr)\bigr|^{1/2} dx \\ &\qquad \le \int_{\{|x|\le K\}} E'\bigl(|H(Y_t + \sqrt{h}\bar{W}_{1/4}, 1)|\bigr)^{1/2} dx \le CK\end{aligned}$$

for some constant C, by virtue of Lemmas 3.2 and 3.3. Consider the set $A = \{|X_t + \nu\sqrt{h}\bar{W}_1| < \frac{|x|}{2}\}$. Employing considerations analogous to those at the end of the proof of Lemma 3.4 by the exponential decay of the function $\Psi_{h/4}(x) = -(1 - \Phi_{h/4}(x))1_{\{x>0\}} + \Phi_{h/4}(x)1_{\{x\le 0\}}$, we can prove

$$\begin{aligned}&\int_{\{|x|>K\}} \bigl(E'\bigl(\phi_{h/4}(Y_t + \sqrt{h}\bar{W}_{1/4} - x)\bigr)\bigr)^{1/2} dx \\ &\qquad \le \int_{\{|x|>K\}} \int_0^1 \bigl|E'\bigl(\Psi_{h/4}(Y_t + \sqrt{h}\bar{W}_{1/4} - x) \\ &\qquad\qquad \times H(Y_t + \sqrt{h}\bar{W}_{1/4}, 1)(1_A + 1_{A^c})\bigr)\bigr|^{1/2} dx < C < \infty\end{aligned}$$

and hence we can obtain (3.18).

We are ready to proceed with our last step.

THEOREM 3.7. *Under the same hypotheses of Theorem* 3.1, *for each $p > 1$, there exist positive constants C_p and C, independent of x, t and h, such that*

$$\sup_{x\in\mathbb{R}} E\left(\left|\frac{1}{n}\sum_{j=1}^{n}\phi_h(Y_t^j - x) - \frac{1}{n}\sum_{j=1}^{n}\phi_h(X_t^j - x)\right|\right) \le C_p \frac{1}{h^{1-1/2p}\sqrt{n}} \tag{3.19}$$

for $n = O(\frac{1}{h})^k$ *for some* $k > 0$,

$$\int_{\mathbb{R}} E\left(\left|\frac{1}{n}\sum_{j=1}^{n}\phi_h(Y_t^j - x) - \frac{1}{n}\sum_{j=1}^{n}\phi_h(X_t^j - x)\right|\right) dx \le C\frac{1}{\sqrt{hn}}. \tag{3.20}$$

PROOF. As usual, by applying the mean value theorem we may write

$$\phi_h(Y_t^j - x) - \phi_h(X_t^j - x) = \phi_h'(\rho_t^j - x)(Y_t^j - X_t^j),$$

with $\rho_t^j \in [Y_t^j; X_t^j]$. Following the same procedure as before, it is clear that the difference in (3.19) becomes

$$\begin{aligned}\frac{1}{n}E\left(\left|\sum_{j=1}^{n}[\phi_h(Y_t^j - x) - \phi_h(X_t^j - x)]\right|\right) &= \frac{1}{n}E\left(\left|\sum_{j=1}^{n}\phi_h'(\rho_t^j - x)(Y_t^j - X_t^j)\right|\right)\\ &\le \frac{1}{n}\sum_{j=1}^{n}E\big(|\phi_h'(\rho_t^j - x)||Y_t^j - X_t^j|\big).\end{aligned}$$

In the case of (3.20) one can easily see that

$$\begin{aligned}\int_{\mathbb{R}} E\big(|\phi_h'(\rho_t^j - x)||Y_t^j - X_t^j|\big)\,dx &= E\left(\int_{\mathbb{R}}\frac{1}{h}|\rho_t^j - x|\phi_h(\rho_t^j - x)\,dx|Y_t^j - X_t^j|\right)\\ &= \sqrt{\frac{2}{\pi h}}E(|Y_t^j - X_t^j|).\end{aligned}$$

In the case of (3.19), with analogous notation as before, for $Z_t^{\nu,j} = (1-\nu)X_t^j + \nu Y_t^j$ we have

$$E\big(|\phi_h'(\rho_t^j - x)||Y_t^j - X_t^j|\big) = \int_0^1 E\big(|\phi_h'(Z_t^{\nu,j} - x)||Y_t^j - X_t^j|\big)\,d\nu. \tag{3.21}$$

Therefore, choosing $\frac{1}{p} + \frac{1}{q} = 1$, by Hölder's inequality, the integrand can be dominated as

$$E\big(|\phi_h'(Z_t^{\nu,j} - x)||Y_t^j - X_t^j|\big) \le \|\phi_h'(Z_t^{\nu,j} - x)\|_p\|Y_t^j - X_t^j\|_q. \tag{3.22}$$

Furthermore, by the properties of the Gaussian density,

$$\begin{aligned}E\big(|\phi_h'(Z_t^{\nu,j} - x)|^p\big) &= \frac{1}{\sqrt{2\pi}^{p-1}}\frac{1}{\sqrt{p}h^{3p/2-1/2}}E\big(|Z_t^{\nu,j} - x|^p\phi_{h/p}(Z_t^{\nu,j} - x)\big)\\ &= \frac{1}{\sqrt{2\pi}^{p-1}}\frac{1}{\sqrt{p}h^{3p/2-1/2}}\int_{\mathbb{R}}\int_{\mathbb{R}}\phi_{h/2p}(y - x)p_t^{\nu,j}(y)\\ &\qquad\times|y - z - x|^p\phi_{h/2p}(z)\,dv\,du\end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}^{p-1}} \frac{1}{\sqrt{p} h^{p-1/2}} \int_{\mathbb{R}} \int_{\mathbb{R}} \phi_{1/2p}(u) p_t^{\nu,j}(u\sqrt{h}+x) \times |u-v|^p \phi_{1/2p}(v)\, dv\, du$$

$$\leq \frac{1}{h^{p-1/2}} \sum_{i=0}^{p} C(p,i,t) \int_{\mathbb{R}} |u|^i \phi_{1/2p}(u) p_t^{\nu,j}(\sqrt{h}u+x)\, du,$$

where $p_t^{\nu,j}(u)$ denotes the density function of $V_t^{\nu,j} = Z_t^{\nu,j} + \sqrt{h/2p}\bar{W}_1$. The proof of (3.19) is finished once we prove that $\int_{\mathbb{R}} |y|^p \phi_{1/2p}(y) p_t^{\nu,j}(\sqrt{h}y+x)\, dy$ is bounded, for which it is enough to show the boundedness of $p_t^{\nu,j}(y)$. The link described at the beginning of Section 2 between the density of a random variable and its Malliavin derivative (we are now considering the space $\Omega \times \bar{\Omega}$ with derivatives $D, \bar{D}$) can be applied here and we have that there exist positive constants a and b such that

$$p_t^{\nu,j}(y) = E'\Big(1_{\{V_t^{\nu,j} > x\}} H(V_t^{\nu,j}, 1)\Big) \leq \|\gamma_{V_t^{\nu,j}}^{-1}\|_a \|V_t^{\nu,j}\|_{1,b} < \infty. \tag{3.23}$$

By definition, it is clear that

$$\|V_t^{\nu,j}\|_{1,b} \leq \|Y_t^j\|_{1,b} + \|X_t^j - Y_t^j\|_{1,b} + \left\|\sqrt{\frac{h}{2p}}\bar{W}_1\right\|_{1,b}.$$

Since by Lemma 3.2 we know that $\|Y_t^j\|_{1,b}$ is finite, the whole question is reduced to evaluating $\|Y_t^j - X_t^j\|_b$, $\|X_t^j - Y_t^j\|_{1,b}$, for $b > 1$ and $\|\gamma_{V_t^{\nu,j}}^{-1}\|_a$. The first ones are proved in the next lemma, while the second is shown in Lemma 3.9 (where we use the hypothesis $n = O(\frac{1}{h})^k$ for some $k > 0$). Applying these results to (3.21), (3.22) and (3.23), we obtain our thesis. □

LEMMA 3.8. *For any $p > 1$, we have*

$$E(|Y_t^j - X_t^j|^p)^{1/p} \leq C\frac{1}{\sqrt{n}}, \qquad \|Y_t^j - X_t^j\|_{1,p} \leq C\frac{1}{\sqrt{n}}.$$

PROOF. We will only prove the first assertion for $p = 2$. The proofs of the second inequality and of the general case are similar. The difference $Y_t^i - X_t^i$ verifies the following equation:

$$\begin{aligned} Y_t^i - X_t^i &= Y_{\eta(t)}^i - X_{\eta(t)}^i \\ &\quad + \big[a\big(Y_{\eta(t)}^i, F(Y_{\eta(t)}^i; \upsilon_{\eta(t)})\big) - a\big(X_{\eta(t)}^i, F(X_{\eta(t)}^i; \bar{\mu}_{\eta(t)})\big)\big](t - \eta(t)) \\ &\quad + \big[b\big(Y_{\eta(t)}^i, G(Y_{\eta(t)}^i; \upsilon_{\eta(t)})\big) - b\big(X_{\eta(t)}^i, G(X_{\eta(t)}^i; \bar{\mu}_{\eta(t)})\big)\big](W_t^i - W_{\eta(t)}^i). \end{aligned}$$

We want to show that $Y_t^i - X_t^i$ is uniformly bounded in the L^2 norm. In order to show this, by virtue of the mean value theorem, we linearize the above equation. From now on, we denote that by $Z^i = Y^i - X^i$,

$$
\begin{aligned}
Z_t^i = {} & Z_{\eta(t)}^i + a_x(\xi_{\eta(t)}^1(i), F(Y_{\eta(t)}^i; v_{\eta(t)}))Z_{\eta(t)}^i(t-\eta(t)) \\
& + a_y(X_{\eta(t)}^i, \theta_{\eta(t)}^1(i))[F(Y_{\eta(t)}^i; v_{\eta(t)}) - F(X_{\eta(t)}^i; \bar{\mu}_{\eta(t)})](t-\eta(t)) \\
& + b_x(\xi_{\eta(t)}^2(i), G(Y_{\eta(t)}^i; v_{\eta(t)}))Z_{\eta(t)}^i(W_t^i - W_{\eta(t)}^i) \\
& + b_y(X_{\eta(t)}^i, \theta_{\eta(t)}^2(i))[G(Y_{\eta(t)}^i; v_{\eta(t)}) - G(X_{\eta(t)}^i; \bar{\mu}_{\eta(t)})](W_t^i - W_{\eta(t)}^i),
\end{aligned}
\tag{3.24}
$$

with $\theta_{\eta(t)}^1(i) \in [F(Y_{\eta(t)}^i; v_{\eta(t)}); F(X_{\eta(t)}^i; \bar{\mu}_{\eta(t)})]$, $\theta_{\eta(t)}^2(i) \in [G(Y_{\eta(t)}^i; v_{\eta(t)}); G(X_{\eta(t)}^i; \bar{\mu}_{\eta(t)})]$ and $\xi_{\eta(t)}^1(i), \xi_{\eta(t)}^2(i) \in [Y_{\eta(t)}^i; X_{\eta(t)}^i]$. By recalling the definition of F and G, and keeping in mind that the copies of X and those of Y are, respectively, identically distributed, we can write

$$
\begin{aligned}
& F(Y_{\eta(t)}^i; v_{\eta(t)}) - F(X_{\eta(t)}^i; \bar{\mu}_{\eta(t)}) \\
& \quad = \int_{\mathbb{R}} f(Y_{\eta(t)}^i, y) v_{\eta(t)}(dy) - \frac{1}{n} \sum_{j=1}^n f(X_{\eta(t)}^i, X_{\eta(t)}^j) \\
& \quad = \frac{1}{n} \sum_{j=1}^n [E^j(f(Y_{\eta(t)}^i, Y_{\eta(t)}^j)) - f(Y_{\eta(t)}^i, Y_{\eta(t)}^j)] \\
& \qquad + \frac{1}{n} \sum_{j=1}^n [f(Y_{\eta(t)}^i, Y_{\eta(t)}^j) - f(X_{\eta(t)}^i, X_{\eta(t)}^j)] \\
& \quad = \frac{1}{n} \sum_{j=1}^n [E^j(f(Y_{\eta(t)}^i, Y_{\eta(t)}^j)) - f(Y_{\eta(t)}^i, Y_{\eta(t)}^j)] \\
& \qquad + \frac{1}{n} \sum_{j=1}^n \{f_x(\lambda_{\eta(t)}^{1,i}, Y_{\eta(t)}^j) Z_{\eta(t)}^i + f_y(X_{\eta(t)}^i, \lambda_{\eta(t)}^{2,j}) Z_{\eta(t)}^j\},
\end{aligned}
$$

with E^j denoting the expectation relative to W^j; similarly for the terms in G,

$$
\begin{aligned}
& G(Y_{\eta(t)}^i; v_{\eta(t)}) - G(X_{\eta(t)}^i; \bar{\mu}_{\eta(t)}) \\
& \quad = \frac{1}{n} \sum_{j=1}^n [E^j(g(Y_{\eta(t)}^i, Y_{\eta(t)}^j)) - g(Y_{\eta(t)}^i, Y_{\eta(t)}^j)] \\
& \qquad + \frac{1}{n} \sum_{j=1}^n \{g_x(\lambda_{\eta(t)}^{3,i}, Y_{\eta(t)}^j) Z_{\eta(t)}^i + g_y(X_{\eta(t)}^i, \lambda_{\eta(t)}^{4,j}) Z_{\eta(t)}^j\},
\end{aligned}
$$

where $\lambda_{\eta(t)}^{1,i}, \lambda_{\eta(t)}^{3,i} \in [Y_{\eta(t)}^i; X_{\eta(t)}^i]$ and $\lambda_{\eta(t)}^{2,j}, \lambda_{\eta(t)}^{4,j} \in [Y_{\eta(t)}^j; X_{\eta(t)}^j]$.

Hence (3.24) becomes

$$Z_t^i = \int_0^t \left[A_{\eta(s)}^{i,i} Z_{\eta(s)}^i + \sum_{j\neq i}^n A_{\eta(s)}^{i,j} Z_{\eta(s)}^j \right] ds$$

(3.25)

$$+ \int_0^t \left[B_{\eta(s)}^{i,i} Z_{\eta(s)}^i + \sum_{j\neq i}^n B_{\eta(s)}^{i,j} Z_{\eta(s)}^j \right] dW_s^i + \int_0^t C_{\eta(s)}^i \, ds + \int_0^t J_{\eta(s)}^i \, dW_s^i,$$

where for $i, j = 1, \dots, n$,

$$A_\bullet^{i,i} = a_x(\xi_\bullet^1(i), F(Y_\bullet^i; v_\bullet))$$
$$+ a_y(X_\bullet^i, \theta_\bullet^1(i)) \left(\frac{1}{n} \sum_{j=1}^n f_x(\lambda_\bullet^{1,i}, Y_\bullet^j) + \frac{1}{n} f_y(X_\bullet^i, \lambda_\bullet^{2,i}) \right),$$

$$A_\bullet^{i,j} = a_y(X_\bullet^i, \theta_\bullet^1(i)) \frac{1}{n} f_y(X_\bullet^i, \lambda_\bullet^{2,j}), \qquad i \neq j,$$

$$B_\bullet^{i,i} = b_x(\xi_\bullet^2(i), G(Y_\bullet^i; v_\bullet))$$
$$+ b_y(X_\bullet^i, \theta_\bullet^2(i)) \left(\frac{1}{n} \sum_{j=1} g_x(\lambda_\bullet^{3,i}, Y_\bullet^j) + g_y(X_\bullet^i, \lambda_\bullet^{4,i}) \right) k,$$

$$B_\bullet^{i,j} = -b_y(X_\bullet^i, \theta_\bullet^2(i)) \frac{1}{n} g_y(X_\bullet^i, \lambda_\bullet^{4,j}), \qquad i \neq j,$$

$$C_\bullet^i = a_y(X_\bullet^i, \theta_\bullet^1(i)) \frac{1}{n} \sum_{j=1}^n [E^j(f(Y_\bullet^i, Y_\bullet^j)) - f(Y_\bullet^i, Y_\bullet^j)],$$

$$J_\bullet^i = b_y(X_\bullet^i, \theta_\bullet^2(i)) \frac{1}{n} \sum_{j=1}^n [E^j(g(Y_\bullet^i, Y_\bullet^j)) - g(Y_\bullet^i, Y_\bullet^j)],$$

form the entries of the matrices that we denote by A and B and of the vectors C and J. So (3.25) can be written in vector form as

$$Z_t^* = H_t^* + \int_0^t Z_{\eta(s)}^* \, dN_s^*, \tag{3.26}$$

where we are using * to denote the transpose of a matrix and $dN_s^{i,j} = A_s^{i,j} \, ds + B_s^{i,j} \, dW_s^i$ and $dH_s^* = (C_{\eta(s)}^1 \, ds + J_{\eta(s)}^1 \, dW_s^1, \dots, C_{\eta(s)}^n \, ds + J_{\eta(s)}^n \, dW_s^n)$. At the points of the partition, the process Z is given by $Z_{t_m}^* = \sum_{k=0}^{m-1} Z_{t_k}^* (N_{t_{k+1}}^* - N_{t_k}^*) + H_{t_m}^*$, which has a unique solution ([13], page 271),

$$Z_{t_m}^* = U_{t_m}^* \sum_{k=0}^{m-1} (U^*)_{t_k}^{-1} [(H_{t_{k+1}}^* - H_{t_k}^*) - ([H^*, N^*]_{t_{k+1}} - [H^*, N^*]_{t_k})],$$

where U^* and $(U^*)^{-1}$ are, respectively, the unique solutions of the matrix equations,

$$U_t^* = I + \int_0^t U_{\eta(s)}^* \, dN_s^*, \qquad (U^*)_t^{-1} = I - \int_0^t \big(d(N^* - [N^*, N^*])_s\big)(U^*)_{\eta(s)}^{-1}.$$

Let us remark that the entries of the matrices A and B are uniformly bounded; namely, it is immediate to see that

$$|A^{i,i}|, |B^{i,i}| \le M^2 + M \quad \text{and} \quad |A^{i,j}|, |B^{i,j}| \le \frac{M^2}{n} \quad \text{for } i \ne j.$$

From (3.25), keeping in mind that $(\sum_{i=1}^n x_i)^2 \le n \sum_{i=1}^n x_i^2$ and Jensen's inequality, we get

$$\begin{aligned}|Z_t^i|^2 \le 6\Biggl\{ & T \int_0^t \Bigl[|A_{\eta(s)}^{i,i}|^2 |Z_{\eta(s)}^i|^2 + (n-1) \sum_{j \ne i} |A_{\eta(s)}^{i,j}|^2 |Z_{\eta(s)}^j|^2 + |C_{\eta(s)}^i|^2 \Bigr] ds \\ & + \left| \int_0^t B_{\eta(s)}^{i,i} Z_{\eta(s)}^i \, dW_s^i \right|^2 + \left| \int_0^t \sum_{j \ne i} B_{\eta(s)}^{i,j} Z_{\eta(s)}^j \, dW_s^i \right|^2 + \left| \int_0^t J_{\eta(s)}^i dW_s^i \right|^2 \Biggr\}.\end{aligned}$$

Taking the supremum over $[0,t]$ and the expectation, by employing Doob's inequality for martingales we finally obtain

$$\begin{aligned}E\Bigl(\sup_{0\le s\le t} |Z_s^i|^2\Bigr) \le {} & 6T\, E\biggl(\int_0^t \Bigl[(M+M^2)^2 \sup_{0\le r\le s} |Z_r^i|^2 \\ & \qquad + \frac{M^4}{n} \sum_{j\ne i} \sup_{0\le r\le s} |Z_r^j|^2 + |C_{\eta(s)}^i|^2 \Bigr] ds \biggr) \\ & + 24 E\biggl(\int_0^t \Bigl[(M+M^2)^2 \sup_{0\le r\le s} |Z_r^i|^2 \\ & \qquad + \frac{M^4}{n} \sum_{j\ne i} \sup_{0\le r\le s} |Z_r^j|^2 + |J_{\eta(s)}^i|^2 \Bigr] ds \biggr)\end{aligned}$$

summarized into

$$\vartheta_i(t) \le \int_0^t \biggl(K_1 \vartheta_i(s) + \frac{K_2}{n} \sum_{j\ne i} \vartheta_j(s) + K_3 x_i(s) \biggr) ds,$$

where $\vartheta_i(t) = E(\sup_{0\le s\le t} |Z_s^i|^2)$, $x_i(s) = E(|J_{\eta(s)}^i|^2 + |C_{\eta(s)}^i|^2)$, $K_3 = 6T + 24$, $K_2 = K_3 M^4$ and $K_1 = K_3(M+M^2)^2$. Gronwall's inequality then implies

$$\begin{aligned}E\Bigl(\sup_{0\le s\le t} |Z_s^i|^2\Bigr) \le {} & \exp(K_1 T) \int_0^t E\biggl[\frac{K_2}{n} \sum_{j\ne i} \sup_{0\le r\le s} |Z_r^j|^2 \\ & \qquad + K_3(|J_{\eta(s)}^i|^2 + |C_{\eta(s)}^i|^2) \biggr] ds, \end{aligned} \tag{3.27}$$

$$\sum_{i=1}^n E\Bigl(\sup_{0\le s\le t} |Z_t^i|^2\Bigr) \le \exp(K_4 T) K_5 \int_0^t \sum_{i=1}^n E(|J_{\eta(s)}^i|^2 + |C_{\eta(s)}^i|^2)\, ds$$

with $K_4 = \exp(K_1 T)K_2$ and $K_5 = \exp(K_1 T)K_3$. Consequently the problem is reduced to analyzing the vectors C and J. We can evaluate the last two expectations by the propagation of chaos. We focus our attention only on $E(\|C_{t_k}\|^2)$ ($\|\cdot\|$ here means the Euclidean norm), as the other case is similarly carried out.

Since the sequence Y^i is formed by independent copies of the original process Y, also the processes $f(Y^i_\bullet, Y^j_\bullet)$ and $f(Y^i_\bullet, Y^l_\bullet)$ become conditionally independent, given Y^i, provided $j \neq l$, so for each i and each $r = t_k$, the following inequality is fulfilled:

$$\begin{aligned}
E(|C^i_r|^2) &= E\left\{\left|a_y(X^i_{t_k}, \theta^1_r(i))\frac{1}{n}\sum_{j=1}^n \big[E^j\big(f(Y^i_r - Y^j_r)\big), f(Y^i_r, Y^j_r)\big]\right|^2\right\} \\
&\le \frac{M^2}{n^2}\Bigg\{2E\Bigg[\sum_{\substack{j,l\\ l<j}} \big[E^j\big(f(Y^i_r, Y^j_r)\big) - f(Y^i_r, Y^j_r)\big] \\
&\qquad \times \big[E^l\big(f(Y^i_r, Y^l_r)\big) - f(Y^i_r, Y^l_r)\big]\Bigg] + \sum_{j=1}^n \operatorname{Var}\big(f(Y^i_r, Y^j_r)\big)\Bigg\} \\
&\le \frac{M^2}{n^2} 2E\Bigg[E\Bigg(\sum_{\substack{j,l\\ l<j}} \big[E^j\big(f(Y^i_r, Y^j_r)\big) - f(Y^i_r, Y^j_r)\big] \\
&\qquad \times \big[E^l\big(f(Y^i_r, Y^l_r)\big) - f(Y^i_r, Y^l_r)\big]\Big| Y^i_r\Bigg)\Bigg] + \frac{4M^4}{n} \\
&\le \frac{M^2}{n^2} 2E\Bigg\{\sum_{\substack{j,l\\ l<j}} E\big[E^j\big(f(Y^i_r, Y^j_r)\big) - f(Y^i_r, Y^j_r)\big| Y^i_r\big] \\
&\qquad \times E\big[E^l\big(f(Y^i_r, Y^l_r)\big) - f(Y^i_r, Y^l_r)\big| Y^i_r\big]\Bigg\} + \frac{4M^4}{n} = 0 + \frac{4M^4}{n}.
\end{aligned}$$

Substituting in (3.27), we finally obtain

$$E\left(\sum_i \sup_{0\le t\le T} |Z^i_t|^2\right) \le \exp(K_4 T) K_5 \Gamma \sum_{k=0}^{n-1} (t_{k+1} - t_k)\frac{M^4}{n} = \frac{\Lambda}{n} t_m \le \frac{\Lambda}{n} T$$

for appropriately chosen constants Γ, Λ and, of course, the same inequality holds for each component.

To prove the second statement, it remains to show that for all j,

$$\sum_{i=1}^n \int_0^T E\big(|D^i_s(Y^j_t - X^j_t)|^2\big)\, ds \le C\frac{1}{\sqrt{n}}.$$

Starting again from (3.26), it is possible to show that for each i the matrix process $(D_s^i Z_t^j) = X_{(i)}^j(s,t)$ verifies the linear matrix SDE,

$$X_{(i)}^*(s,t) = \tilde{K}_{(i)}(s,t) + \int_s^t X_{(i)}^*(s,r)\, d\tilde{N}_{(i)}^*(s,r),$$

where the matrices are given by

$$\begin{aligned} \big(d\tilde{N}_{(i)}^*(s,r)\big)_{jk} &= D_s^i A_r^{j,k}\, dr \quad \text{for } j \neq i, \\ \big(d\tilde{N}_{(i)}^*(s,r)\big)_{i,k} &= D_s^i A_r^{i,k}\, dr + D_s^i B_r^{i,k}\, dW_r^i, \\ \tilde{K}_{(i)}^k(s,t) &= D_s^i H_t^k \quad \text{for } k \neq i, \\ \tilde{K}_{(i)}^i(s,t) &= D_s^i H_t^i + (Z_s^* B_s^*)^i. \end{aligned}$$

With computations similar to those shown before, it is possible to deduce an inequality analogous to (3.27) with the coefficients of $K_{(i)}$ in place of those of H, from which will follow the result by propagation of chaos and so we conclude the proof. $\square$

We would like to remark that when we apply the inequality of Lemma 3.8 to our terms in Theorem 3.7 we have

$$\frac{1}{n}\frac{1}{\sqrt{2\pi}h}\sum_{j=1}^n E(|Y_t^j - X_t^j|^2)^{1/2} \leq \frac{1}{\sqrt{2\pi}h}(\Lambda T)^{1/2}\frac{1}{\sqrt{n}}\frac{1}{n}n \leq C\frac{1}{\sqrt{n}h},$$

giving the right order of convergence.

It remains to check the boundedness of the last factor.

LEMMA 3.9. *Let* $V_t^{\nu,j} = Z_t^{\nu,j} + \sqrt{\frac{h}{2p}}\tilde{W}_1$ (*where* $Z_t^{\nu,j} = (1-\nu)X_t^j + \nu Y_t^j$) *and* $n = O(\frac{1}{h})^k$ *for some* $k > 0$; *then the following holds*:

$$\sup_{h\in(0,1]}\sup_{\nu\in[0,1]} \|\gamma_{V_t^{\nu,j}}^{-1}\|_p < \infty \qquad \text{for all } p \in \mathbb{N} \text{ and } t \in (0,T].$$

PROOF. Let $\hat{X}^j$ denote the unique strong solution to (2.1) when the stochastic equation is driven by W^j. That is, $\hat{X}^1, \dots, \hat{X}^n$ are n independent copies of X. The proof goes along the same lines of the proof of Lemma 3.3 so we only sketch it. The three main points that one needs to check in order to prove the boundedness of the Malliavin covariance matrix are:

(i) $\sup_{\nu\in[0,1]}\|V_t^{\nu,j} - \hat{X}_t^j\|_{1,p} \leq C(\frac{1}{\sqrt{n}} + \sqrt{h})$.

(ii) $\|\gamma_{\hat{X}_t^j}^{-1}\|_p < \infty$ for all $p \in \mathbb{N}$.

(iii) $\|\gamma_{V_t^{\nu,j}}^{-1}\|_p \leq Ch^{-1}$.

The first inequality uses Lemma 3.2 and the same lines of proof as Proposition 4.1 in [7]. The third inequality is direct. The second one was proved in Lemma 2.2 for the process X, but clearly the same is true for the copies. □

4. Proof of Theorem 3.1. This brief section is intended to gather all the results that we presented previously and to finally obtain the proof of Theorem 3.1.

The statement of Theorem 3.1 deals with both density and distribution functions, but as we announced, we focus our attention only on the proof for the first ones, for which we have laid out all the necessary results. To get the same conclusion in the case of distributions, the whole procedure should be reconstructed, but we will just describe it briefly.

For the densities, we first consider the L^1 norm (3.4):

$$\begin{aligned}
&\int_{\mathbb{R}} E\left(\left|p_t(x) - \frac{1}{n}\sum_{j=1}^n \phi_h(X_t^j - x)\right|\right) dx \\
&\quad \le \int_{\mathbb{R}} \left|p_t(x) - E\big(\phi_h(X_t - x)\big)\right| dx \\
&\qquad + \int_{\mathbb{R}} \left|E\big(\phi_h(X_t - x) - \phi_h(Y_t - x)\big)\right| dx \\
&\qquad + \int_{\mathbb{R}} E\left[\left|E\big(\phi_h(Y_t - x)\big) - \frac{1}{n}\sum_{j=1}^n \phi_h(Y_t^j - x)\right|\right] dx \\
&\qquad + \int_{\mathbb{R}} E\left[\left|\frac{1}{n}\sum_{j=1}^n \phi_h(Y_t^j - x) - \frac{1}{n}\sum_{j=1}^n \phi_h(X_t^j - x)\right|\right] dx \\
&\quad \le C\left(h + \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}h^{1/4}}\right).
\end{aligned}$$

The above bounds follow from Lemma 3.4, Theorem 3.6, (3.18) and Theorem 3.7. The analogous result (3.6), when adopting the norm of the supremum follows by applying instead Lemma 3.4, Theorem 3.5, (3.17) and Theorem 3.7.

Consider now the proofs for distribution functions (3.3):

$$\begin{aligned}
&\int_{\mathbb{R}} E\left[\left|u(t,x) - \frac{1}{n}\sum_{j=1}^n 1_{\{X_t^j \le x\}}\right|\right] dx \\
&\quad \le \int_{\mathbb{R}} |E(1_{\{X_t \le x\}} - 1_{\{Y_t \le x\}})|\, dx \\
&\qquad + \int_{\mathbb{R}} E\left|E(1_{\{Y_t \le x\}}) - \frac{1}{n}\sum_{j=1}^n 1_{\{Y_t^j \le x\}}\right| dx
\end{aligned}$$

$$+\int_{\mathbb{R}} E\left[\left|\frac{1}{n}\sum_{j=1}^{n} 1_{\{Y_t^j \le x\}} - \frac{1}{n}\sum_{j=1}^{n} 1_{\{X_t^j \le x\}}\right|\right] dx$$
$$= A_1 + A_2 + A_3.$$

Let us consider the quantity A_1, for this one has to prove that there exits a positive constant C independent of $\varepsilon \in (0, 1]$ and h such that

$$\left|E\left(1_{\{X_t \le x\}} - \Phi_\varepsilon(x - X_t)\right)\right| \le C\varepsilon,$$
$$\left|E\left(\Phi_\varepsilon(x - X_t) - \Phi_\varepsilon(x - Y_t)\right)\right| \le Ch,$$
$$\left|E\left(\Phi_\varepsilon(x - Y_t) - 1_{\{Y_t \le x\}}\right)\right| \le C\varepsilon.$$

The first and third assertions are proved by the same argument as in Lemma 3.4 [note that we know that $\gamma^{-1}_{Y_t+\sqrt{\varepsilon}\bar{W}_{1/2}} \in \bigcap_{p>1} L^p(\Omega)$ by taking $\nu = 0$ in Lemma 3.3] while the second one is proved along the lines of the proof of Theorem 3.5.

The quantity A_2 can be analyzed in the same way as we showed in (3.18), while for A_3 we have

$$\int_{\mathbb{R}} E\left(\left|\frac{1}{n}\sum_{j=1}^{n} 1_{\{Y_t^j \le x\}} - \frac{1}{n}\sum_{j=1}^{n} 1_{\{X_t^j \le x\}}\right|\right) dx$$
$$\le \frac{1}{n}\sum_{j=1}^{n} E\left(\int_{\mathbb{R}} \left|1_{\{Y_t^j \le x\}} - 1_{\{X_t^j \le x\}}\right| dx\right)$$
$$\le \frac{1}{n}\sum_{j=1}^{n} E(|Y_t^j - X_t^j|) \le \frac{C}{\sqrt{n}}.$$

This completes the proof of (3.3). For (3.5) the proof is similar to that for (3.4). □

5. Conclusions. In this work, we analyzed the rate of convergence of a particle method introduced by Bossy and Talay in order to approximate the solution to the McKean–Vlasov equation and we showed that it is faster than the one obtained in [1]. Besides, the rate of convergence here obtained seems to match their simulations run in the particular case of the Burgers equation.

We also analyzed the rate of convergence when approximating the marginal densities of the solution. In order to carry out the necessary calculations we had to study the existence and smoothness of these densities.

The problem of obtaining the optimal rate of convergence for the Burgers equations is still open and the authors hope the method developed here might apply, if properly adapted, also to this case. The rate in (3.3) is definitely optimal. In the other results the rate is probably not optimal but the authors believe that a modification of the present technique should give the optimal rate. In this presentation we strived for a unified presentation and not for a case-by-case study.

Some straightforward generalizations of the above results were not included in our exposition for reasons of space. For instance, it is not difficult to consider the case when also the initial random variable has to be approximated or when the measurement of the error is done through the variances [i.e., $L^2(\Omega)$] rather than through the expectations. Yet another generalization is to consider approximations of the type ϕ_ε rather than ϕ_h; if $\varepsilon = O(h^r)$ for some $r > 0$ a similar analysis can be carried out. For $r \geq 1/2$ the rates are the same as the ones obtained here. Otherwise h^{2r} becomes the dominant rate.

The condition that all coefficients of the equation have derivatives bounded by M can be further weakened if the arguments given here are analyzed closely. We assumed boundedness to shorten the length of the paper.

Finally we remark that the condition $n = O(\frac{1}{h})^k$ for some $k > 0$ in Theorem 3.1 (used to obtain Lemma 3.9) is merely technical rather than restrictive, since k can be chosen freely.

APPENDIX A

Here we prove an important ingredient of the proofs of Theorems 3.5 and 3.6 which consists of a complex integration by parts formula that measures with some accuracy the effect of each component.

It looks a little cumbersome, but we need to state it in this generality to be able to apply it to Lemmas B.1 and B.2, which provide the steps of the induction invoked in the proof of the main Theorem 3.5. The statement is divided into two parts, the first regards bounded functions, while the second gives the analogous estimate for the the approximations $\phi_{h/2}$, that clearly are not bounded as $h \to 0$. The second part requires the introduction of yet another independent Brownian motion to exploit the integration by parts formula.

LEMMA A.1. *Let W and $\tilde{W}$ be two independent Brownian motions, so that* (2.1) *and* (3.1), *defining X and Y, are driven by W, while independent copies of those, $\tilde{X}$ and $\tilde{Y}$, are driven by $\tilde{W}$. $E'' = E \times \tilde{E}$ denotes the expectation on the canonical product space $\Omega \times \tilde{\Omega}$. Let V^h, Z^h be two sequences of processes adapted to the filtration generated by W, such that*

$$\begin{aligned} &\sup_{s_1,\dots,s_n \leq T} E''\Big[\sup_{t\leq T} |D_{s_1}\cdots D_{s_n} V_t^h|^{2q}\Big]^{1/2q} \leq C_V \equiv C_V(T), \\ &\sup_{s_1,\dots,s_n \leq T} E''\Big[\sup_{t\leq T} |D_{s_1}\cdots D_{s_n} Z_t^h|^{2q}\Big]^{1/2q} \leq C_Z \equiv C_Z(T) \end{aligned} \tag{A.1}$$

for some constants $C_V, C_Z > 0$, some $q \geq 4$ and for all $n = 0, 1, \dots, 4$.

Moreover let $\alpha\colon \mathbb{R}^4 \to \mathbb{R}$, $\gamma\colon \mathbb{R}^+ \times \mathbb{R}^4 \to \mathbb{R}$, $\beta\colon \mathbb{R}^+ \times \mathbb{R}^8 \to \mathbb{R}$ *be differentiable real valued functions, for which exist positive constants* C_α, C_β, C_γ, *such that*

$$\begin{aligned} &\|\alpha^{(i)}\|_\infty, |\alpha(0,0)| \le C_\alpha, \\ \text{(A.2)} \qquad &\sup\nolimits_{s\in[0,T]} \|\beta_s^{(i)}\|_\infty, \sup\nolimits_{s\in[0,T]} |\beta_s(0,0)| \le C_\beta, \\ &\sup\nolimits_{s\in[0,T]} \|\gamma_s^{(i)}\|_\infty, \sup\nolimits_{s\in[0,T]} |\gamma_s(0,0)| \le C_\gamma, \end{aligned}$$

for all $i = 1, \dots, 4$ ($f^{(i)}$ *denotes any partial derivative or order* i).

Then, if we set $\underline{U}_s = (\underline{U}_s^1, \underline{U}_s^2) = ((X_{\eta(s)}, Y_{\eta(s)}, \tilde{X}_{\eta(s)}, \tilde{Y}_{\eta(s)}), (X_s, Y_s, \tilde{X}_s, \tilde{Y}_s))$, *we have*

$$\text{(A.3)} \qquad \begin{aligned} &\left| E''\left[V_t^h \alpha(\underline{U}_t^2) \int_0^t Z_s^h \beta_s(\underline{U}_s) \int_{\eta(s)}^s \gamma_{\eta(r)}(\underline{U}_r^1)\, dW_r^{j_1}\, dW_s^{j_2} \right] \right| \\ &\quad \le C_V C_Z C_\alpha C_\beta C_\gamma C h t, \end{aligned}$$

where $dW_s^0 = ds$, *and* $(W^1, W^2) = (W, \tilde{W})$, $j_1, j_2 = 0, 1, 2$ *and* $C > 0$ *depends only on the constant appearing in Lemma* 3.2 *and it is independent of* h *and* C_V, C_Z, C_α, C_β, C_γ.

Let $\bar{W}$ *be a Wiener process independent of* W *and* $\tilde{W}$ *and let* $E''' = E \times \tilde{E} \times \bar{E}$ *denote the expectation in the cross product space supporting all three independent processes. Let us take* $\alpha(\underline{U}_t^2) = \alpha(X_t, Y_t) = \phi_{h/2}^{(p)}(X_t + \nu(Y_t - X_t) + \sqrt{h}\bar{W}_{1/2} - x)$, *for* $p \in \{0, 1\}$. *Then, if* (A.1) *holds with* $q \ge 32$ *and* β *and* γ *verify* (A.2) *for* $i = 1, \dots, p+3$, *we have*

$$\text{(A.4)} \qquad \begin{aligned} &\left| E'''\left[V_t^h \alpha(\underline{U}_t^2) \int_0^t Z_s^h \beta_s(\underline{U}_s) \int_{\eta(s)}^s \gamma_{\eta(r)}(\underline{U}_r^1) dW_r^{j_1} dW_s^{j_2} \right] \right| \\ &\quad \le C_V C_Z C_\beta C_\gamma C h t, \end{aligned}$$

uniformly for $\nu \in [0, 1]$.

The constant C in (A.4) is not the same as in (A.3) and in (A.4) clearly α is no longer assumed bounded. From now on, we use the notation $Z_t^{\nu,\bar{W}} = X_t + \nu(Y_t - X_t) + \sqrt{h}\bar{W}_{1/2}$.

PROOF. We prove (A.3) only when $j_1 = 1$, $j_2 = 1$, which is computationally the most cumbersome case; all the others can be treated similarly by applying the integration by parts one or two times less. Indeed, we have to use integration by parts one time less each time we have a Lebesgue integral instead of a stochastic one. Also, when $j_1 = 1$ and $j_2 = 2$ (or vice versa), which corresponds to double integrals with respect to the two independent Brownian motions, the procedure we are going to describe simplifies, as some terms will become zero, because the processes V^h and Z^h depend only upon W. To simplify notation, we are going to omit the arguments of the functions.

By the integration by parts formula of Malliavin calculus with respect to W, we have

$$\begin{aligned}
&\left|E''\left[V_t^h\alpha\int_0^t Z_s^h\beta_s\int_{\eta(s)}^s \gamma_{\eta(r)}\,dW_r\,dW_s\right]\right|\\
&\quad=\left|E''\left[\int_0^t D_s\{V_t^h\alpha\}Z_s^h\beta_s\int_{\eta(s)}^s \gamma_{\eta(r)}\,dW_r\,ds\right]\right|\\
&\quad=\left|E''\left[\int_0^t\int_{\eta(s)}^s D_r[D_s\{V_t^h\alpha\}Z_s^h\beta_s]\gamma_{\eta(r)}\,dr\,ds\right]\right|\\
&\quad\le\int_0^t\int_{\eta(s)}^s\bigl|E''\bigl[D_r[D_s\{V_t^h\alpha\}Z_s^h\beta_s]\gamma_{\eta(r)}\bigr]\bigr|\,dr\,ds.
\end{aligned}$$

It is then clear that to obtain (A.3), it suffices to show that

$$\sup_{\substack{s\in[0,t]\\ r\in(\eta(s),s]}}\bigl|E''\bigl[D_r[D_s\{V_t^h\alpha\}Z_s^h\beta_s]\gamma_{\eta(r)}\bigr]\bigr|\le C_VC_ZC_\alpha C_\beta C_\gamma C,$$

where C is a positive constant that depends only on T and the constant appearing in Lemma 3.2. Applying assumption (A.1) and Hölder's inequality, we get

$$\begin{aligned}
&\bigl|E''\bigl(D_r\{D_s\{V_t^h\alpha\}Z_s^h\beta_s\}\gamma_{\eta(r)}\bigr)\bigr|\\
&\quad\le E''\bigl[|D_rD_s\{V_t^h\alpha\}Z_s^h\beta_s\gamma_{\eta(r)}|\\
&\qquad\quad+|D_s\{V_t^h\alpha\}D_rZ_s^h\beta_s\gamma_{\eta(r)}|+|D_s\{V_t^h\alpha\}Z_s^hD_r\beta_s\gamma_{\eta(r)}|\bigr]\\
&\quad\le\|\gamma_{\eta(r)}\|_4\bigl\{\|Z_s^h\|_4\bigl(\|\beta_s\|_4\|D_rD_s\{V_t^h\alpha\}\|_4+\|D_r\beta_s\|_4\|D_s\{V_t^h\alpha\}\|_4\bigr)\\
&\qquad\qquad\qquad\qquad+\|D_rZ_s^h\|_4\|\beta_s\|_4\|D_s\{V_t^h\alpha\}\|_4\bigr\}.
\end{aligned}$$

From now on, we denote each component of $\underline{U}$ by U^i for $i=1,\dots,8$. We are going to analyze each single term. By using assumptions (A.1), (A.2) and Hölder's inequality, the following hold for any $r,s,t\in[0,T]$:

(A) $\qquad\|Z_s^h\|_4\le C_Z,\qquad\|D_rZ_s^h\|_4\le C_Z;$

(B) $\qquad\|\gamma_{\eta(r)}\|_4\le\left\|\sum_{i=1}^4\frac{\partial\gamma_{\eta(r)}}{\partial x_i}U_r^i\right\|_4+\|\gamma_{\eta(r)}(\underline{0})\|_4\le C_\gamma\left(\sum_{i=1}^4\|U_r^i\|_4+1\right);$

(C) $\qquad\|\beta_s\|_4\le\left\|\sum_{i=1}^8\frac{\partial\beta_s}{\partial x_i}U_s^i\right\|_4+\|\beta_s(\underline{0})\|_4\le C_\beta\left(\sum_{i=1}^8\|U_s^i\|_4+1\right);$

(D) $\qquad\|D_r\beta_s\|_4\le\left\|\frac{\partial\beta_s}{\partial x_5}D_rU_s^5+\frac{\partial\beta_s}{\partial x_6}D_rU_s^6\right\|_4\le C_\beta\bigl(\|D_r\tilde{X}_s\|_4+\|D_r\tilde{Y}_s\|_4\bigr);$

$$
\begin{aligned}
&\|D_s(V_t^h\alpha)\|_4\\
&\quad\le \left\|D_sV_t^h\left(\sum_{i=1}^4\frac{\partial\alpha}{\partial x_i}U_t^{i+4}+\alpha(\underline{0})\right)\right\|_4+\left\|V_t^h\left(\frac{\partial\alpha}{\partial x_1}D_sX_t+\frac{\partial\alpha}{\partial x_2}D_sY_t\right)\right\|_4\\
&\quad\le C_\alpha\left\{\|D_sV_t^h\|_8\left(1+\sum_{i=5}^8\|U_t^i\|_8\right)+\|V_t^h\|_8(\|D_sX_t\|_8+\|D_sY_t\|_8)\right\}\qquad\text{(E)}\\
&\quad\le C_VC_\alpha\left[\sum_{i=5}^8\|U_t^i\|_8+1+\|D_sX_t\|_8+\|D_sY_t\|_8\right];
\end{aligned}
$$

$$
\begin{aligned}
&\|D_rD_s\{V_t^h\alpha\}\|_4\\
&\quad\le \|D_rD_sV_t^h\|_8\left(\sum_{i=1}^4\left\|\frac{\partial\alpha}{\partial x_i}U_t^{i+4}\right\|_8+\|\alpha(\underline{0})\|_8\right)\\
&\qquad+\|D_sV_t^h\|_8\left\|\sum_{i=1}^2\frac{\partial\alpha}{\partial x_i}D_rU_t^{i+4}\right\|_8+\|D_rV_t^h\|_8\left\|\sum_{i=1}^2\frac{\partial\alpha}{\partial x_i}D_sU_t^{i+4}\right\|_8\\
&\qquad+\|V_t^h\|_8\left[\left\|\sum_{i,j=1,2}\frac{\partial^2\alpha}{\partial x_i\partial x_j}D_sU_t^{i+4}D_rU_t^{j+4}\right\|_8\right.\qquad\text{(F)}\\
&\qquad\qquad\left.+\left\|\sum_{i,j=1,2}\frac{\partial\alpha}{\partial x_i}D_rD_sU_t^{i+4}\right\|_8\right]\\
&\quad\le C_VC_\alpha\left\{\sum_{i=5}^8\|U_t^i\|_8+1+2\big(\|D_sU_t^5\|_8+\|D_sU_t^6\|_8\big)\right.\\
&\qquad\left.+\big(\|D_sU_t^5\|_{16}+\|D_sU_t^6\|_{16}\big)^2+\|D_rD_sU_t^5\|_8+\|D_rD_sU_t^6\|_8\right\}.
\end{aligned}
$$

By virtue of all the previous estimates and using Lemma 3.2, we may conclude that X and Y together with their Malliavin derivatives are bounded in the L^p norms ($p\le 16$) uniformly in t, let us say by a common constant C, so we finally get

$$
\begin{aligned}
&\sup_{\substack{s\in[0,t]\\ r\in[\eta(s),s]}}\big|E''\big[D_r[D_s\{V_t^h\alpha\}Z_s^h\beta_s]\gamma_{\eta(r)}\big]\big|\\
&\qquad\le C_VC_ZC_\alpha C_\beta C_\gamma(4C+1)(8C+1)(4C^2+20C+2).
\end{aligned}
$$

To prove the second result in the statement, we restrict to the case $j_1=2$, $j_2=1$ (also to deal with a different case). Even if we do not have uniform bounds on the derivatives of α, a double application of integration by parts will help us. Again, by integration by parts, the problem is reduced to showing that $|E'''[\tilde D_r\{D_s[V_t^h\phi_{h/2}^{(p)}(Z_t^{\nu,\bar W}-x)]Z_s^h\beta_s\}\gamma_{\eta(r)}]|$ is bounded uniformly in s, r and ν.

Carrying out the calculations, we get

$$\begin{aligned}
&\big|E'''\big[\tilde{D}_r\big\{D_s[V_t^h\phi_{h/2}^{(p)}(Z_t^{\nu,\bar{W}}-x)]Z_s^h\beta_s\big\}\gamma_{\eta(r)}\big]\big|\\
&\quad\le \big|E'''\big[\phi_{h/2}^{(p)}(Z_t^{\nu,\bar{W}}-x)\gamma_{\eta(r)}D_sV_t^hZ_s^h\tilde{D}_r\beta_s\big]\big|\\
&\qquad+\big|E'''\big[\phi_{h/2}^{(p+1)}(Z_t^{\nu,\bar{W}}-x)D_sZ_t^{\nu,\bar{W}}\gamma_{\eta(r)}V_t^hZ_s^h\tilde{D}_r\beta_s\big]\big|\\
&\quad= \big|E'''\big[\Phi_{h/2}(Z_t^{\nu,\bar{W}}-x)H_{p+1}(Z_t^{\nu,\bar{W}},N^1)\big]\big|\\
&\qquad+\big|E'''\big[\Phi_{h/2}(Z_t^{\nu,\bar{W}}-x)H_{p+2}(Z_t^{\nu,\bar{W}},N^2)\big]\big|,
\end{aligned}$$

where N^1 and N^2 are obviously defined through the above equation. By applying (2.4) to the above terms we may conclude that for $q_3\ge 4$,

$$\begin{aligned}
&\big|E'''\big[\tilde{D}_r\big\{D_s[V_t^h\phi_{h/2}^{(p)}(Z_t^{\nu,\bar{W}}-x)]Z_s^h\beta_s\big\}\gamma_{\eta(r)}\big]\big|\\
&\quad\le C_{p+1}\|\gamma^{-1}_{Z_t^{\nu,\bar{W}}}\|_{q_1}^{m_1}\|Z_t^{\nu,\bar{W}}\|^{m_2}_{p+2,q_2}\|N^1\|_{p+1,q_3}\\
&\qquad+C_{p+2}\|\gamma^{-1}_{Z_t^{\nu,\bar{W}}}\|_{d_1}^{n_1}\|Z_t^{\nu,\bar{W}}\|^{n_2}_{p+3,d_2}\|N^2\|_{p+2,d_3},
\end{aligned}$$

but $\|\gamma^{-1}_{Z_t^{\nu,\bar{W}}}\|_{q_1}$, $\|\gamma^{-1}_{Z_t^{\nu,\bar{W}}}\|_{d_1}$ are bounded by virtue of Lemma 3.3. Moreover, we know that $\|Z_t^{\nu,\bar{W}}\|_{p+3,q_2}\le\|X_t\|_{p+3,q_2}+\|Y_t\|_{p+3,q_2}<+\infty$ and the increasingness of the Sobolev norms implies that also the term $\|Z_t^{\nu,\bar{W}}\|_{p+2,d_2}$ is bounded.

It remains to evaluate $\|N^1\|_{p+1,q_3}$ and $\|N^2\|_{p+2,d_3}$. We will show the boundedness only of the first term, as the proof is the same for both.

If we apply Hölder's inequality for Sobolev norms, we obtain

$$\begin{aligned}
\|N^1\|_{p+1,q_3}&\le\|\gamma_{\eta(r)}D_sV_t^hZ_s^h\tilde{D}_r\beta_s\|_{p+1,q_3}\\
&\le\|\gamma_{\eta(r)}\|_{p+1,b_1}\|D_sV_t^h\|_{p+1,b_2}\|Z_s^h\|_{p+1,b_3}\|\tilde{D}_r\beta_s\|_{p+1,b_4}\\
&\le C_VC_Z\|\gamma_{\eta(r)}\|_{p+1,b_1}\|\tilde{D}_r\beta_s\|_{p+1,b_4},
\end{aligned}$$

where $\frac{1}{b_1}+\frac{1}{b_2}+\frac{1}{b_3}+\frac{1}{b_4}=\frac{1}{q_3}$ (this is the reason why one requires $q\ge 32$ in the statement). On the other hand, it is easy to prove that, if f is a smooth function with its derivatives and $|f(0)|$ all bounded by a constant A and G is a random variable, then

$$\|f(G)\|_{p+1,q}\le\Lambda A\|G\|_{p+1,nq},$$

for appropriate Λ and n. Consequently, in our case we have

$$\begin{aligned}
&\|\gamma_{\eta(r)}\|_{p+1,b_1}\le\Lambda_1C_\gamma\big(\|X_{\eta(r)}\|_{p+1,nb_1}+\|Y_{\eta(r)}\|_{p+1,nb_1}\big),\\
&\|\tilde{D}_r\beta_s\|_{p+1,b_4}\le\Lambda_2C_\beta\rho(C),
\end{aligned}$$

for some fixed polynomial function ρ, constants Λ_1, Λ_2 and integers m, n, which concludes the proof. $\square$

REMARK A.2.

(a) The same technique applies also to prove

$$\left|E'\left(\phi_{h/2}^{(p)}(Z_t^{\nu,\bar{W}}-x)V_t^h\int_0^t Z_s^h\beta_s\,dW_s^j\right)\right|\le\rho(C)C_V C_\beta\int_0^t C_Z(s)\,ds,\qquad j=0,1,\ p=0,1$$

for some properly chosen polynomial ρ, when β depends only on X, Y and verifies (A.2). Here $C_Z(s)$ stands for the bound in (A.1) where instead of T one has s, for $s\in[0,T]$. For example, when the inner integral is stochastic, we have

$$\begin{aligned}
&\left|E'\left(\phi_{h/2}^{(p)}(Z_t^{\nu,\bar{W}}-x)V_t^h\int_0^t Z_s^h\beta_s\,dW_s\right)\right|\\
&\quad\le\int_0^t\big[\|\gamma_{Z_t^{\nu,\bar{W}}}^{-1}\|_{q_1}^{m_1}\|Z_t^{\nu,\bar{W}}\|_{p+2,q_2}^{m_2}\|D_sV_t^hZ_s^h\beta_s\|_{p+1,q_3}\\
&\qquad+\|\gamma_{Z_t^{\nu,\bar{W}}}^{-1}\|_{d_1}^{n_1}\|Z_t^{\nu,\bar{W}}\|_{p+3,d_2}^{n_2}\|D_sZ_t^{\nu,\bar{W}}V_t^hZ_s^h\beta_s\|_{p+2,d_3}\big]\,ds
\end{aligned}$$

and we may proceed as before.

(b) The previous lemma holds even when α is a smooth random function independent of W, $\tilde{W}$, with the bounds in (A.2) holding for every ω. In this case, the same proof goes through, since the Malliavin derivatives of α do not intervene in the computation; only the spatial ones intervene. Therefore the right-hand side of (A.3) will yield a random function bounded by the same constant uniformly in ω.

(c) Combining the above proof with the end of the proof of Lemma 3.4, under similar conditions to (A.4), one can obtain that for any $k\in\mathbb{N}$ and $|x|>K$ for a positive constant K,

$$\begin{aligned}
&\left|E'''\left[V_t^h\alpha(\underline{U}_t^2)\int_0^t Z_s^h\beta_s(\underline{U}_s)\int_{\eta(s)}^s\gamma_{\eta(r)}(\underline{U}_r^1)\,dW_r^{j_1}\,dW_s^{j_2}\right]\right|\\
&\quad\le C_VC_ZC_\beta C_\gamma Cht\left(\exp\left(-\frac{x^2}{4h}\right)+\frac{1}{|x^k|}\right).
\end{aligned}$$

(d) Finally, we remark that one might assume a lower degree of integrability in (A.1), if one chooses to penalize the other terms more when applying Hölder's inequality.

APPENDIX B

In order to prove (3.16), we proceed by induction, the first step being carried out in the next lemma and the general case in Lemma B.2. Here we use the notation established before Theorem 3.5.

LEMMA B.1. *Let ξ_t^1 be a random point between X_t and Y_t in the sense of* (3.11) *and $U_0(t) = \int_0^t \mathcal{E}_s^{-1}\,dZ_s$. Then there exists a deterministic constant A depending on M, but independent of t, x, U_0, such that the following hold*:

$$
\begin{aligned}
&\bigl|E'\bigl(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U_0(t)\bigr)\bigr| \le Ath,\\
&\qquad\qquad |\tilde{E}(u(\tilde{X}_t, \tilde{Y}_t)\tilde{\mathcal{E}}_t \tilde{U}_0(t))| \le Ath.
\end{aligned}
\tag{B.1}
$$

Here u: $\Omega \times \mathbb{R}^2 \to \mathbb{R}$ is any smooth random measurable function with its first four derivatives bounded by M uniformly in Ω, independent of $\tilde{W}$. $(\tilde{X}, \tilde{Y}, \tilde{U})$ is an independent copy of (X, Y, U).

PROOF. Recalling the definition of Z, we can rewrite

$$
\begin{aligned}
&\bigl|E'\bigl(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U_0(t)\bigr)\bigr|\\
&\quad\le \biggl|E'\biggl[\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t \int_0^t \mathcal{E}_s^{-1}\bigl\{(Y_s - Y_{\eta(s)})A_s\\
&\qquad\qquad + \tilde{E}\bigl((\tilde{Y}_s - \tilde{Y}_{\eta(s)})\tilde{A}_s\bigr)\bigr\}\,ds\biggr]\biggr|\\
&\qquad + \biggl|E'\biggl[\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t \int_0^t \mathcal{E}_s^{-1}\bigl\{(Y_s - Y_{\eta(s)})B_s\\
&\qquad\qquad + \tilde{E}\bigl((\tilde{Y}_s - \tilde{Y}_{\eta(s)})\tilde{B}_s\bigr)\bigr\}\,dW_s\biggr]\biggr|.
\end{aligned}
$$

The four terms in the above expression are qualitatively very similar (due to the hypotheses on the coefficients) and they differ basically only in the integrators. We show inequality (B.1) only for the fourth one; this is the most complicated term since it contains both Brownian motions. The proof of all the other terms runs along similar lines. As an independent copy of Y, $\tilde{Y}$ must verify an analogous equation,

$$
\begin{aligned}
\tilde{Y}_s - \tilde{Y}_{\eta(s)} &= a\bigl(\tilde{Y}_{\eta(s)}, F(\tilde{Y}_{\eta(s)}; v_{\eta(s)})\bigr)\bigl(s - \eta(s)\bigr)\\
&\quad + b\bigl(\tilde{Y}_{\eta(s)}, G(\tilde{Y}_{\eta(s)}; v_{\eta(s)})\bigr)(\tilde{W}_s - \tilde{W}_{\eta(s)}),
\end{aligned}
$$

so substituting the latter in the fourth term of the previous inequality, this term becomes

$$
\begin{aligned}
&E'\biggl(\phi'_{h/2}(\xi_t^1+\sqrt{h}\bar{W}_{1/2}-x)\\
&\qquad\times\mathcal{E}_t\int_0^t\mathcal{E}_s^{-1}\tilde{E}\biggl[\tilde{B}_s a(\tilde{Y}_{\eta(s)},F(\tilde{Y}_{\eta(s)};v_{\eta(s)}))\int_{\eta(s)}^s dr\biggr]dW_s\biggr)\\
&+E'\biggl(\phi'_{h/2}(\xi_t^1+\sqrt{h}\bar{W}_{1/2}-x)\\
&\qquad\times\mathcal{E}_t\int_0^t\mathcal{E}_s^{-1}\tilde{E}\biggl[\tilde{B}_s b(\tilde{Y}_{\eta(s)},G(\tilde{Y}_{\eta(s)};v_{\eta(s)}))\int_{\eta(s)}^s d\tilde{W}_r\biggr]dW_s\biggr).
\end{aligned}
$$

Again we look only at the last term, as the other can be treated similarly. As we already mentioned, the midpoint ξ_t^1 is to be understood in the sense of expression (3.11), so recalling the definition of $Z_t^{\nu,\bar{W}}$, under the expectation E''' on $\Omega\times\bar{\Omega}\times\tilde{\Omega}$ (therefore $E'''=E'\times\tilde{E}$), we have

$$
\begin{aligned}
&E'\biggl(\int_0^1\phi'_{h/2}(Z_t^{\nu,\bar{W}}-x)\,d\nu\,\mathcal{E}_t\\
&\qquad\times\int_0^t\mathcal{E}_s^{-1}\tilde{E}\biggl[\tilde{B}_s b(\tilde{Y}_{\eta(s)},G(\tilde{Y}_{\eta(s)};v_{\eta(s)}))\int_{\eta(s)}^s d\tilde{W}_r\biggr]dW_s\biggr)\\
&\quad=\int_0^1E'''\biggl(\phi'_{h/2}(Z_t^{\nu,\bar{W}}-x)\mathcal{E}_t\\
&\qquad\qquad\times\int_0^t\mathcal{E}_s^{-1}\tilde{B}_s b(\tilde{Y}_{\eta(s)},G(\tilde{Y}_{\eta(s)};v_{\eta(s)}))\int_{\eta(s)}^s d\tilde{W}_r\,dW_s\biggr)d\nu,
\end{aligned}
\tag{B.2}
$$

and we are in condition to apply Lemma A.1. If we recall the definition of $\tilde{B}$ and we translate the midpoints θ_s^2 and $\tilde{\xi}_s^2$ appearing there in the notation (3.11), we obtain that this last term can be actually expressed as

$$
\begin{aligned}
&\int_0^1E'''\biggl(\phi'_{h/2}(Z_t^{\nu,\bar{W}}-x)\mathcal{E}_t\int_0^t\mathcal{E}_s^{-1}b_y(Y_{\eta(s)},\theta_s^2)g_y(Y_s,\tilde{\xi}_s^2)\\
&\qquad\qquad\times\int_{\eta(s)}^s b(\tilde{Y}_{\eta(s)},G(\tilde{Y}_{\eta(s)};v_{\eta(s)}))\,d\tilde{W}_r\,dW_s\biggr)d\nu\\
&\quad=\int_0^1\int_0^1\int_0^1E'''\biggl[\phi'_{h/2}(Z_t^{\nu,\bar{W}}-x)\mathcal{E}_t\\
&\qquad\qquad\times\int_0^t\mathcal{E}_s^{-1}b_y\bigl(Y_{\eta(s)},(1-\varepsilon)G(Y_{\eta(s)};v_{\eta(s)})+\varepsilon G(Y_s;v_s)\bigr)\\
&\qquad\qquad\times g_y\bigl(Y_s,(1-\mu)\tilde{Y}_{\eta(s)}+\mu\tilde{Y}_s\bigr)\\
&\qquad\qquad\times\int_{\eta(s)}^s b(\tilde{Y}_{\eta(s)},G(\tilde{Y}_{\eta(s)};v_{\eta(s)}))\,d\tilde{W}_r\,dW_s\biggr]d\mu\,d\varepsilon\,d\nu.
\end{aligned}
$$

Indeed, by virtue of Hypothesis (H0), the functions

$$
\gamma(x_1)=b\biggl(x_1,\int_{\mathbb{R}}g(x_1,z)dv_{\eta(s)}\biggr),
$$

$$\beta(x_1,x_2,x_3,x_4)=b_y\biggl(x_1,(1-\varepsilon)\int_{\mathbb{R}}g(x_1,z)\,dv_{\eta(s)}+\varepsilon\int_{\mathbb{R}}g(x_1,z)\,dv_s\biggr)$$
$$\times g_y\bigl(x_3,(1-\mu)x_2+\mu x_4\bigr),$$

respectively, applied to $Y_{\eta(s)}$ and $(Y_{\eta(s)},\tilde{Y}_{\eta(s)},Y_s,\tilde{Y}_s)$, verify condition (A.2). They are differentiable four times and the derivative of order i of each of them is bounded by $C_2=2^{2(i+1)}M^{2(i+2)}$, in the worst of cases. Besides, $\mathcal{E}_t$ and its inverse are solutions to SDEs with coefficients with bounded spatial derivatives. Therefore it is not difficult to prove that they satisfy for $n=0,1,\dots,4$ and $q\in\mathbb{N}$ (see [10], Theorem 2.2.2),

$$\text{(B.3)}\qquad \sup_{s_1,\dots,s_n\le T}E\Bigl[\sup_{t\le T}|D_{s_1}\cdots D_{s_n}\mathcal{E}_t|^{2q}\Bigr] + \sup_{s_1,\dots,s_n\le T}E\Bigl[\sup_{t\le T}|D_{s_1}\cdots D_{s_n}\mathcal{E}_t^{-1}|^{2q}\Bigr]\le C,$$

for some positive constant C independent of h which without loss of generality we assume is the same as the one appearing in Lemma 3.2.

So we can take $Z_s^h=\mathcal{E}_s^{-1}$, $V_t^h=\mathcal{E}_t$, $p=1$, γ and β as above, satisfying (A.1) and (A.2). From here, we conclude that (B.2) is bounded by some constant $A_1>0$ and

$$\biggl|E'\biggl(\phi'_{h/2}(\xi_t^1+\sqrt{h}\bar{W}_{1/2}-x)\mathcal{E}_t \times\int_0^t\mathcal{E}_s^{-1}\tilde{E}\biggl[\tilde{B}_s b\bigl(\tilde{Y}_{\eta(s)},G(\tilde{Y}_{\eta(s)};v_{\eta(s)})\bigr)\int_{\eta(s)}^s d\tilde{W}_r\biggr]dW_s\biggr)\biggr|\le A_1th.$$

Repeating the same argument with all the other terms, we can find a proper constant A such that the thesis is satisfied. The proof for the case $|\tilde{E}(u(X_t,Y_t,\tilde{X}_t,\tilde{Y}_t)\tilde{\mathcal{E}}_t\tilde{U}_0(t))|\le Ath$ is similar. Indeed, as before, we may decompose $\tilde{\mathcal{E}}_t\tilde{U}_0(t)$ into its integral terms and apply the first part of Lemma A.1 with a uniformly bounded smooth random function, as we observed in Remark A.2(b). □

We now prove the second step of the induction in the lemma that follows.

LEMMA B.2. *There exists a constant $R>0$, independent of t, h, x and k such that*

$$\bigl|E'\bigl(\phi'_{h/2}(\xi_t^1+\sqrt{h}\bar{W}_{1/2}-x)\mathcal{E}_tU_k(t)\bigr)\bigr|\le h\sum_{j=1}^{k+1}\frac{(Rt)^j}{j!},$$
$$\bigl|\tilde{E}\bigl(u(\tilde{X}_t,\tilde{Y}_t)\tilde{\mathcal{E}}_t\tilde{U}_k(t)\bigr)\bigr|\le h\sum_{j=1}^{k+1}\frac{(Rt)^j}{j!}.$$

Here $u\colon\Omega\times\mathbb{R}^2\to\mathbb{R}$ is any smooth random measurable function with its first four derivatives bounded by M uniformly in Ω.

PROOF. We proceed by induction. Having proved the case $k = 0$ in the previous lemma, let us assume the result true for k and prove it for $k+1$. From the proof we will determine the constant R. Using (3.15), we have

$$\begin{aligned}
&\big|E'(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U_{k+1}(t))\big| \\
&\quad \le \big|E'(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U_0(t))\big| \\
&\qquad + \bigg|E'\bigg(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t \\
&\qquad\qquad \times \int_0^t \mathcal{E}_s^{-1}\tilde{E}\big(a_y(Y_s, \eta_s^1) f_y(Y_s, \tilde{\xi}_s^1)\tilde{\mathcal{E}}_s\tilde{U}_k(s)\big)\,ds\bigg)\bigg| \\
&\qquad + \bigg|E'\bigg(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t \\
&\qquad\qquad \times \int_0^t \mathcal{E}_s^{-1}\tilde{E}\big(\beta_s b_y(Y_s, \eta_s^2) g_y(Y_s, \tilde{\xi}_s^2)\tilde{\mathcal{E}}_s\tilde{U}_k(s)\big)\,ds\bigg)\bigg| \\
&\qquad + \bigg|E'\bigg(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t \\
&\qquad\qquad \times \int_0^t \mathcal{E}_s^{-1} b_y(Y_s, \eta_s^2)\tilde{E}\big(g_y(Y_s, \tilde{\xi}_s^2)\tilde{\mathcal{E}}_s\tilde{U}_k(s)\big)\,dW_s\bigg)\bigg|.
\end{aligned}$$

By the previous lemma, the first term in the right-hand side of the inequality is certainly less than or equal to Aht; hence let us focus our attention on the other two terms.

First we rewrite the above inequality, by using the midpoint notation (3.11), for β_s, ξ_s^1, ξ_s^3, ξ_s^5, η_s^1, and η_s^2, so we have

$$\begin{aligned}
&\big|E'(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U_{k+1}(t))\big| \\
&\quad \le Aht + \int_0^1\int_0^1 \bigg|E'\bigg(\phi'_{h/2}(Z_t^{\nu,\bar{W}} - x)\mathcal{E}_t \\
&\qquad\qquad \times \int_0^t \mathcal{E}_s^{-1} a_y(Y_s, F_s^{\varepsilon})\tilde{E}\big(f_y(Y_s, \tilde{\xi}_s^1)\tilde{\mathcal{E}}_s\tilde{U}_k(s)\big)\,ds\bigg)\bigg|\,d\varepsilon\,d\nu \\
&\qquad + \int_0^1\int_0^1\int_0^1 \bigg|E'\bigg(\phi'_{h/2}(Z_t^{\nu,\bar{W}} - x)\mathcal{E}_t \int_0^t \mathcal{E}_s^{-1} b_x\big(Z_s^{\lambda}, G(X_s; u_s)\big) \\
&\qquad\qquad \times b_y(Y_s, G_s^{\varepsilon})\tilde{E}\big(g_y(Y_s, \tilde{\xi}_s^2)\tilde{\mathcal{E}}_s\tilde{U}_k(s)\big)\,ds\bigg)\bigg|\,d\lambda\,d\varepsilon\,d\nu
\end{aligned}$$

$$+\int_0^1\int_0^1\int_0^1\bigg|E'\bigg(\phi'_{h/2}(Z_t^{\nu,\bar W}-x)\mathcal{E}_t\int_0^t\mathcal{E}_s^{-1}b_y^2(Y_s,G_s^\varepsilon)G'(Z_s^\rho;u_s)$$

$$\times\tilde E\big(g_y(Y_s,\tilde\xi_s^2)\tilde{\mathcal{E}}_s\tilde U_k(s)\big)\,ds\bigg)\bigg|\,d\lambda\,d\varepsilon\,d\nu$$

$$+\int_0^1\int_0^1\bigg|E'\bigg(\phi'_{h/2}(Z_t^{\nu,\bar W}-x)\mathcal{E}_t$$

$$\times\int_0^t\mathcal{E}_s^{-1}b_y(Y_s,G_s^\varepsilon)\tilde E\big(g_y(Y_s,\tilde\xi_s^2)\tilde{\mathcal{E}}_s\tilde U_k(s)\big)\,dW_s\bigg)\bigg|\,d\varepsilon\,d\nu,$$

where we set $F_s^\varepsilon=(1-\varepsilon)F(X_s;u_s)+\varepsilon F(Y_s;v_s)$, $G_s^\varepsilon=(1-\varepsilon)G(X_s;u_s)+\varepsilon G(Y_s;v_s)$ and $Z^\lambda=X_t+\lambda(Y_t-X_t)$, for $0\le\lambda,\nu,\varepsilon\le 1$.

Let us notice that the functions

$$\beta_1(x_5,x_6)=a_y\bigg(x_6,(1-\varepsilon)\int_{\mathbb R}f(x_5,z)\,d\mu_s(z)+\varepsilon\int_{\mathbb R}f(x_6,z)\,dv_s(z)\bigg),$$

$$\beta_2(x_5,x_6)=b_y\bigg(x_6,(1-\varepsilon)\int_{\mathbb R}g(x_5,z)\,d\mu_s(z)+\varepsilon\int_{\mathbb R}g(x_6,z)\,dv_s(z)\bigg),$$

$$\beta_3(x_5,x_6)=b_x\bigg((1-\lambda)x_5+\lambda x_6,\int_{\mathbb R}g(x_5,z)\,d\mu_s(z)\bigg),$$

$$\beta_4(x_5,x_6)=\int_{\mathbb R}g_x\big((1-\lambda)x_5+\lambda x_6,z\big)\,d\mu_s(z),$$

$$\beta_5(x_5,x_6)=\beta_2(x_5,x_6)\beta_3(x_5,x_6)+\beta_2^2(x_5,x_6)\beta_4(x_5,x_6),$$

all have derivatives up to order 4, uniformly bounded by a fixed constant depending on M, that we will denote with C_M.

We now want to apply Remark A.2(a), taking $V_t^h=\mathcal{E}_t$, $Z_s^h=\mathcal{E}_s^{-1}\tilde E(f_y(Y_s,\tilde\xi_s^1)\times\tilde{\mathcal{E}}_s\tilde U_k(s))$ or $Z_s^h=\mathcal{E}_s^{-1}\tilde E(g_y(Y_s,\tilde\xi_s^2)\tilde{\mathcal{E}}_s\tilde U_k(s))$ and $\beta_s(x_1,\dots,x_8)=\beta_i(x_1,x_2)$, $i=1,2,5$, so we have to verify that the hypotheses of Lemma A.1 are satisfied. We have to find a bound for $\|Z_s^h\|_{n,q}$, for q large enough and $n\le 4$. For this, first note that $\mathcal{E}_t$ and $\mathcal{E}_s^{-1}$ verify (B.3), in the sense of Sobolev norms with respect only to W, just as in Remark A.2(a).

Using the usual midpoint notation, our task is made equivalent to finding a bound for $\|\mathcal{E}_s^{-1}\tilde E(u_y(Y_s,\tilde Z_s^\tau)\tilde{\mathcal{E}}_s\tilde U_k(s))\|_{n,q}$, where $Z^\tau=X_t+\tau(Y_t-X_t)$ for $0\le\tau\le 1$ and u is either f or g. It is important to observe that in this case $u_y(Y_s,\cdot)$ is a smooth random function independent of U_k and of $\tau\in[0,1]$.

By Hölder's inequality we have for $n\le 4$,

$$\begin{aligned}\|\mathcal{E}_s^{-1}\tilde E\big(u_y(Y_s,\tilde Z_s^\tau)\tilde{\mathcal{E}}_s\tilde U_k(s)\big)\|_{n,q}&\le\|\mathcal{E}_s^{-1}\|_{n,q_1}\|\tilde E\big(u_y(Y_s,\tilde Z_s^\tau)\tilde{\mathcal{E}}_s\tilde U_k(s)\big)\|_{n,q_2}\\&\le C_1\|\tilde E\big(u_y(Y_s,\tilde Z_s^\tau)\tilde{\mathcal{E}}_s\tilde U_k(s)\big)\|_{n,q_2},\end{aligned}$$

with $\frac{1}{q_1} + \frac{1}{q_2} = \frac{1}{q}$. For example, consider the case when $n = 2$. We derive our estimate only in this case, to keep the computations more understandable. For $n = 3, 4$ it is just a matter of considering also the Malliavin derivatives of order 3 and 4. By differentiating we obtain

$$D_r \tilde{E}\big(u_y(Y_s, \tilde{Z}_s^\tau)\tilde{\mathcal{E}}_s \tilde{U}_k(s)\big) = D_r Y_s \tilde{E}\big(u_{yx}(Y_s, \tilde{Z}_s^\tau)\tilde{\mathcal{E}}_s \tilde{U}_k(s)\big),$$
$$D_r D_u \tilde{E}\big(u_y(Y_s, \tilde{Z}_s^\tau)\tilde{\mathcal{E}}_s \tilde{U}_k(s)\big) = D_r D_u Y_s \tilde{E}\big(u_{yxx}(Y_s, \tilde{Z}_s^\tau)\tilde{\mathcal{E}}_s \tilde{U}_k(s)\big)$$

and consequently we have that

$$\begin{aligned}
&\|\tilde{E}\big(u_y(Y_s, \tilde{Z}_s^\tau)\tilde{\mathcal{E}}_s \tilde{U}_k(s)\big)\|_{2,q_2}^{q_2} \\
&\quad\le E\big(\big|\tilde{E}\big(u_y(Y_s, \tilde{Z}_s^\tau)\tilde{\mathcal{E}}_s \tilde{U}_k(s)\big)\big|^{q_2}\big) \\
&\qquad + E\bigg[\bigg(\int_0^T |D_r Y_s|^2 \big|\tilde{E}\big(u_{yx}(Y_s, \tilde{Z}_s^\tau)\tilde{\mathcal{E}}_s \tilde{U}_k(s)\big)\big|^2\, dr \\
&\qquad\qquad + \int_0^T \int_0^T |D_r D_u Y_s|^2 \big|\tilde{E}\big(u_{yxx}(Y_s, \tilde{Z}_s^\tau)\tilde{\mathcal{E}}_s \tilde{U}_k(s)\big)\big|^2\, du\, dr\bigg)^{q_2/2}\bigg].
\end{aligned}$$

But $u_y(Y_s, \cdot)$ and $u_{yxx}(Y_s, \cdot)$ have derivatives uniformly bounded by M for all ω; therefore we see that the inductive hypotheses can be applied to conclude that

$$\big|\tilde{E}\big(\varphi(Y_s, \tilde{Z}_s^\tau)\tilde{\mathcal{E}}_s \tilde{U}_k(s)\big)\big| \le h \sum_{j=1}^{k+1} \frac{(Rt)^j}{j!} \quad \text{for } \varphi = u_y, u_{yx}, u_{yxx},$$

which implies

$$\|\tilde{E}\big(u_y(Y_s, \tilde{Z}_s^\tau)\tilde{\mathcal{E}}_s \tilde{U}_k(s)\big)\|_{2,q_2} \le h \sum_{j=1}^{k+1} \frac{(Rs)^j}{j!}\big(1 + \|Y_s\|_{2,q_2}\big) \le Ch \sum_{j=1}^{k+1} \frac{(Rs)^j}{j!},$$

because of Lemma 3.2. Summarizing, it is possible to find a constant $\bar{C}$, independent of all the parameters, that depends polynomially on the constants M, T and the constant C in Lemma 3.2 such that

$$\begin{aligned}
\big|E'\big(\phi'_{h/2}(\xi_t^1 + \sqrt{h}\bar{W}_{1/2} - x)\mathcal{E}_t U_{k+1}(t)\big)\big| &\le Aht + \bar{C} C_1 \int_0^t h \sum_{j=1}^{k+1} \frac{(Rs)^j}{j!}\, ds \\
&\le h\bigg(Rt + \sum_{j=1}^{k+1} \frac{(Rt)^{j+1}}{(j+1)!}\bigg)
\end{aligned}$$

choosing $R \ge \max(A, \bar{C} C_1)$.

Similarly as in Lemma B.1 one may prove that for a random function, independent of $\tilde{W}$ $u\colon \Omega \times \mathbb{R}^2 \to \mathbb{R}$ with derivatives bounded by M uniformly in $\Omega \times \mathbb{R}^2$ one has that

$$\left|\tilde{E}\big(u(\tilde{Y}_t, \tilde{X}_t)\tilde{\mathcal{E}}_t \tilde{U}_{k+1}(t)\big)\right| \le h\left(Rt + \sum_{j=1}^{k+1} \frac{(Rt)^{j+1}}{(j+1)!}\right).$$

This concludes the proof. □

Acknowledgments. This work was begun with the first author's visit at Purdue University and he thanks the Mathematics Department for their hospitality. The research of the second author was completed while visiting the Department of Mathematics of the University of Tokyo. He thanks them for their hospitality. The authors also thank the referee for suggesting changes that led to several improvements in the presentation.

REFERENCES

[1] BOSSY, M. and TALAY, D. (1997). A stochastic particle method for the McKean–Vlasov and the Burgers equation. *Math. Comp.* **66** 157–192.

[2] BOSSY, M. and TALAY, D. (1996). Convergence rate for the approximation of the limit law of weakly interacting particles: Application to the Burgers equation. *Ann. Appl. Probab.* **6** 818–861.

[3] CHORIN, A. J. (1993). Vortex methods and vortex statistics. In *Lecture for Les Houches Summer School of Theoretical Physics*. Lawrence Berkeley Laboratory Prepublications.

[4] GÄRTNER, J. (1988). On the McKean–Vlasov limit for interacting diffusions. *Math. Nachr.* **137** 197–248.

[5] HALD, O. H. (1981). Convergence of random methods for a reaction diffusion equation. *SIAM J. Sci. Statist. Comput.* **2** 85–94.

[6] HALD, O. H. (1986). Convergence of a random method with creation of vorticity. *SIAM J. Sci. Statist. Comput.* **7** 1373–1386.

[7] KOHATSU-HIGA, A. and OGAWA, S. (1997). Weak rate of convergence for an Euler scheme of nonlinear SDEs. *Monte Carlo Method Appl.* **3** 327–345.

[8] KLOEDEN, P. E. and PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin.

[9] MELEARD, S. (1996). Asymptotic behaviour of some particle systems: McKean–Vlasov and Boltzmann models. *Probabilistic Models for Nonlinear Partial Differential Equations*. *Lecture Notes in Math.* **1627** 42–95. Springer, Berlin.

[10] NUALART, D. (1996). *Malliavin Calculus and Related Topics*. *Probability and Its Applications*. Springer, Berlin.

[11] NUALART, D. (1995). Analysis on Wiener space and anticipating stochastic calculus. *Ecole d'été de Probabilités de Saint-Flour XXV*. *Lecture Notes in Math.* **1690** 123–227. Springer, Berlin.

[12] NUALART, D. and MORET, S. (2001). Generalization of Itô's formula for smooth nondegenerate martingales. *Stochastic Process. Appl.* **91** 115–149.

[13] PROTTER, P. (1990). *Stochastic Integration and Differential Equations: A New Approach*. Springer, Berlin.

[14] ROBERTS, S. (1989). Convergence of a random walk method for the Burgers equation. *Math. Comp.* **52** 647–673.

[15] STROOCK, D. W. (1983). Some applications of stochastic calculus to partial differential equations. *Ecole d'été de Probabilités de Saint-Flour XI*. *Lecture Notes in Math.* **976** 268–382. Springer, Berlin.

[16] TALAY, D. (1996). Probabilistic numerical methods for partial differential equations: Elements of analysis. *Probabilistic Models for Nonlinear Partial Differential Equations*. *Lecture Notes in Math.* **1627** 148–196. Springer, Berlin.

[17] TANIGUCHI, S. (1985). Applications of Malliavin calculus to time-dependent systems of heat equations. *Osaka J. Math.* **22** 307–320.

UNIVERSITÀ DI CHIETI
DIPARTIMENTO DI SCIENZE
VIALE PINDARO, 42
65127 PESCARA
ITALY
E-MAIL: antonf@sci.unich.it

UNIVERSITAT POMPEU FABRA
DEPARTAMENT D'ECONOMIA
RAMÓN TRIAS FARGAS 25-27
BARCELONA 08005
SPAIN
E-MAIL: kohatsu@upf.es