Weak rate of convergence for an Euler scheme of nonlinear SDE's

Short title: Weak approximations for nonlinear diffusions

Arturo Kohatsu-Higa *

Universitat Pompeu Fabra and Kyoto University

Shigeyoshi Ogawa †

Kyoto Institute of Technology

Abstract

We define an Euler type weak approximation for solutions of nonlinear diffusion processes. We find the rate of convergence of this scheme in weak form. As in the diffusion case the weak rate of convergence is better than the strong one. The proof uses the integration by parts formula of Malliavin calculus.

Key words: diffusion process, weak approximation, numerical analysis.

Mathematics Subject Classification (1991): 60H99, 65N12

---

*Research partially supported by DGICYT grants No PB 930052 and PB 950978.
This article was partially written while the author was visiting the Department of Mathematics at Kyoto University with a JSPS fellowship. Address: Universitat Pompeu Fabra. Departament d'Economia. Ramón Trias Fargas 25-27. 08005 Barcelona. Spain. E-mail: kohatsu@upf.es

†Address: Kyoto Institute of Technology. Matsugasaki, Sakyo-ku, Kyoto 606, Japan.

1 Introduction

In this article we concentrate on measuring the error for weak approximations of the nonlinear diffusion process defined through an equation of the type:

$$\begin{aligned} dX_t &= a(X_t, F(X_t; u_t))dt + b^j(X_t, G(X_t; u_t))dW^j(t) \qquad (1) \\ X_0 &= \xi, \text{ for } 0 \le t \le T, \end{aligned}$$

where we assume the convention of summation over repeated indices. $\xi(\omega)$ is an $\mathbb{R}^d$-valued square integrable $\mathcal{F}_0$ measurable random variable with $\xi \in L^p(\Omega)$ for some $p \ge 32$ and

$$\{W(t,\omega) = (W^1(t,\omega), ..., W^k(t,\omega));\ 0 \le t \le T\},$$

is the standard k- dimensional Brownian motion on a fixed probability space $(\Omega, \mathcal{F}, P)$ and

$$\begin{aligned} u_t(dx) &= P(X_t \in (x, x+dx]) \\ H(x;u) &= \int_{-\infty}^{+\infty} h(x-y)u(dy) \text{ for } h = f,\ g \text{ and } u \text{ a measure.} \qquad (2) \end{aligned}$$

We also assume that $a,\ b^j\ : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}^d,\quad f,\ g : \mathbb{R}^d \to \mathbb{R}^p$ are all smooth with bounded derivatives. It is proven in [8] that under the appropiate Lipschitz conditions there is a unique strong solution to (1). These conditions are satisfied in our case due to the smoothness assumed on the functions a, b, f and g. That is,

THEOREM 1.1 *Let a, b, f and g be Lipschitz functions with at most linear growth at infinity. Then there exists a unique strong solution to (1).*

In order to define an approximation to (1) let $\pi = \{0 = t_0 < ... < t_N = T\}$ be a partition of $[0,T]$. Define the shift operator

$$\eta(t) = t_k \qquad \text{if } t_k \le t < t_{k+1},$$

where t_k and t_{k+1} are in π. We set $\eta(0) = 0$. Also suppose that $|\pi| := \max\{t_{k+1} - t_k;\ k = 0, ..., N-1\} \le h$.

We will now introduce a stochastic particle method to approximate $E\varphi(X_t)$ for $\varphi \in C_p^\infty$ (smooth functions whose derivatives have polynomial growth at infinity). Let $\{W_i(t);\ 0 \le t \le T\},\ i = 1, ..., n$, be n independent k-dimensional Wiener processes.

Without loss of generality we assume that the original probability space $(\Omega, \mathcal{F}, P)$ contains these independent copies. Then define the approximation $\overline{X}^i$ for $i = 1, ..., n$ as the solution of the equation

$$\begin{aligned} d\overline{X}_t^i &= a(\overline{X}_{\eta(t)}^i, F(\overline{X}_{\eta(t)}^i; \overline{u}_{\eta(t)}))dt + b^j(\overline{X}_{\eta(t)}^i, G(\overline{X}_{\eta(t)}^i; \overline{u}_{\eta(t)}))dW_i^j(t), \\ X_0^i &= \xi^i \\ \overline{u}_t(dx) &= \frac{1}{n}\sum_{j=1}^n \delta_{\overline{X}_{\eta(t)}^j}(dx) \end{aligned} \tag{3}$$

where $\xi^1, ..., \xi^n$ are n independent copies of ξ and δ_x denotes the point measure at x. With these definitions we intend to prove

Theorem 1.2 *For $\varphi \in C_p^\infty$, we have*

$$|E[\varphi(X_t) - \varphi(\overline{X}_t^1)]| \leq C(\frac{1}{\sqrt{n}} + h). \tag{4}$$

C is a positive constant that depends on the bounds of the growth of φ at infinity but is independent of n, h and the partition π.

The proof of the above fact will be divided into two parts and that can be seen in the way the bound is written. The first entails the error of approximation for the measure $u_t(dx)$ which is being approximated in (3) through a point measure. This study is usually comprised under the name of propagation of chaos. Therefore we will use the aforementioned technique heavily.

The second term appears due to the approximation of the process at time t by the value at time $\eta(t)$. This is the classical rate of weak convergence for Euler schemes, which still holds here. This part is more difficult to prove. In fact, the classical proof (which can be found in e.g. Kloeden and Platen's book [4]) does not work here. The main reason being that in order to measure the weak approximation error using the classical technique, one has to know the weak approximation type error incurred in the measure approximation inside the coefficients of the equation. This obviously is the same type of problem that one wants to solve initially.

As a consequence of the above result we will obtain the following corollary. This result that maybe of interest when actual simulations using Monte Carlo methods are carried out. That is, when one is interested in measuring the error in absolute value.

Corollary 1.1 *For $\varphi \in C_p^\infty$ we have*

$$E|\frac{1}{n}\sum_{i=1}^{n}\varphi(\overline{X}_t^i) - E\varphi(X_t)| \leq C(\frac{1}{\sqrt{n}} + h).$$

C is a constant independent of π but depends on φ and is uniformly bounded for t in bounded intervals.

Bossy and Talay [3] worked on this type of problem where instead of a smooth function φ, they consider the indicator function and an L^1 norm instead of the absolute value. They use a technique based on estimates for the probability densities of the solution process. Instead of h on the right of (4) they obtain $\sqrt{h}$.

In simulation experiments for the Burgers equation (i.e., f is an indicator function and g is constant), they report that the right rate of convergence seems to be h rather than $\sqrt{h}$. Our intention is to prove that at least in the case that all coefficients are smooth the rate of convergence is h.

An important ingredient to achieve this goal is a different method of proof from the one used in [3]. Some ideas developed here can be used in a general non-Markovian setting where smoothness of the solution process is available.

The smoothness is required to be able to make use of the integration by parts formula of Malliavin calculus in order to avoid the cumbersome technique of using estimates for the probability densities of the solution process (see eg. [3]).

The authors believe that the methodology presented here can be adapted in order to consider instead of φ, indicator functions and delta functions using the appropiate Malliavin calculus techniques. We hope to address this issue in future publications. In the course of the proof of Theorem 1.2 we will make use of the integration by parts formula of Malliavin calculus in order to compute expectations of products of smooth random variables with stochastic integrals. This technique allows us to write explicit expressions for the difference $\varphi(X_t) - \varphi(\overline{X}_t^1)$ that could be possibly exploited later for other type of results such as the development of the error in powers of the step size, etc. In order to present our proof in an understandable form, we will present first how the second term on the right of (4) appears.

Now we introduce some of the notation to be used in the article. A function $A(t, x_1, ..., x_n)$, will be sometimes denoted by $A_t(x_1, ..., x_n)$ to differentiate between the time variable t and the space variables $(x_1, ..., x_n)$. A_{x_i} will denote the derivative of A with respect to x_i. $A^{(p)}(x)$ denotes the total p-th order derivative of A. Constants are usually denoted by C that may change from one line to the next unless explicitely stated otherwise. In order to avoid further complications, we will always assume that these constants are bigger than 1. All the constants will depend on T and have the property that they are uniformly bounded when T varies on compact sets.

2 The integration by parts formula of Malliavin calculus

In this section we will introduce a formula that will allow us an easier proof of the Theorem 1.2. For this, we will work on the canonical m-dimensional Wiener space without loss of generality. In fact, to introduce the initial random variable ξ we only need to add an aditional independent σ-field to our canonical Wiener space and the definitions and properties to follow go through without change. From now on, we assume that all the processes under study are in such a probability space.

Now we introduce some basic tools from Malliavin calculus that will be used troughout the text. For further reference see [7]. Let $C_b^\infty(\mathbb{R}^{nm})$ be the set of C^∞ functions $f : \mathbb{R}^{nm} \to \mathbb{R}$ which are bounded and have bounded derivatives of all orders. The class of real random variables of the form $f(W_{t_1}, \ldots, W_{t_n})$, $f \in C_b^\infty(\mathbb{R}^{nm})$, (the class of infinitely differentiable functions with bounded derivatives) is denoted by $\mathcal{S}$. $\mathbb{D}^{1,p}$ designates the Banach space which is the completion of $\mathcal{S}$ with respect to the norm:

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

where

$$D_s^j F = \sum_{i=1}^{n} \frac{\partial f}{\partial x_{ji}}(W_{t_1}, \ldots, W_{t_m})\ \mathbb{I}_{[0,t_i]}(s).$$

$\mathbb{D}^{\alpha,p}$ is defined analogously and its associated norm is denoted by $\|\cdot\|_{\alpha,p}$. That is,

$$\|F\|_{\alpha,p} = \{E|F|^p\}^{1/p} + (\sum_{j_1,\ldots,j_\alpha=1}^{m} E[\{\int_0^T \ldots \int_0^T |D_{s_\alpha}^{j_\alpha} \ldots D_{s_1}^{j_1} F|^2\, ds_\alpha \ldots ds_1\}^{p/2}])^{1/p}.$$

Also, let $\mathbb{D}^\infty = \cap_{p\geq 1} \cap_{\alpha \geq 1} \mathbb{D}^{\alpha,p}$.

The adjoint of the closed unbounded operator $D^j : L^2(\Omega) \to L^2([0,T] \times \Omega)$ is usually denoted by δ^j and is called the Skorohod integral. The calculus concerning this integral was investigated in earlier times also by japanese school around one of the authors. [1] The domain of δ^j is the set of processes $u \in L^2([0,T]\times\Omega)$ such that

$$|E(\int_0^T D_t^j F u_t dt)| \le C\|F\|_2,$$

for all $F \in \mathcal{S}$. Here, C is a constant that may depend on u. In the case that u is in the domain of δ^j then $\delta^j(u)$ is the square integrable random variable determined by the duality relation

$$E(\delta^j(u)F) = E(\int_0^1 D_t^j F u_t dt), \tag{5}$$

for all $F \in \mathbb{D}^{1,2}$. $\delta^j(u)$ turns out to be an extension of the classical Itô integral and it allows the integration of processes that are not necessarily adapted. For this reason one also writes $\delta^j(u) = \int_0^T u_s dW_s^j$.

3 The weak approximation bound

In this section we will consider an approximation different from $\overline{X}$. Define Y^i, for $i = 1, ..., n$, as the i-th copy of the Euler approximation for (1). That is, Y^i is the solution of the equation:

$$\begin{aligned} dY_t^i &= a(Y_{\eta(t)}^i, F(Y_{\eta(t)}^i;\tilde{u}_{\eta(t)}))dt + b^j(Y_{\eta(t)}^i, G(Y_{\eta(t)}^i;\tilde{u}_{\eta(t)}))dW_i^j(t),\ t\in[0,T] \quad (6)\\ X_0^i &= \xi^i.\\ \tilde{u}_t(dx) &= P(Y_t^i \in (x, x+dx]). \quad (7)\end{aligned}$$

The above equation has a unique strong solution because it has been defined recursively. It is clear from the definition of $\{Y^i;\ i=1,...,n\}$ that they are independent identically distributed adapted continuous stochastic processes and that furthermore these are approximations that are not easy to simulate. This is because writing the law of Y_t^i (which is independent of i) involves the use of iterated integrals of order $\frac{\eta(t)}{h}$.

Since all the Y^i, $i = 1, ..., n$ have the same distribution we will write Y when such distinction is not needed. That is, Y is defined through (6) using W as the driving noise process.

[1] see for example the articles, written in japanese, in *Proceedings of The Workshop on Noncausal Stochastic Calculus, (edtd by S.Ogawa)*, RIMS Kyoto University 1984.

In this section we intend to prove the following proposition

PROPOSITION 3.1 *Let* $\varphi \in C_p^\infty$ *then*

$$|E[\varphi(X_t)-\varphi(Y_t)]| \leq Ch.$$

Here the constant C *is independent of* π *but it depends on* φ.

In the next section we will prove that there exists a constant independent of π and i such that

$$E|[\varphi(Y_t^i)-\varphi(\overline{X}_t^i)]| \leq C\frac{1}{\sqrt{n}}.$$

First we prove the following smoothness properties for X and Y.

LEMMA 3.1 $X_t,\ Y_t \in (\mathbb{D}^\infty)^d$ *for any* $t \in [0,T]$ *and furthermore we have for* $j_1, ..., j_n \in \{1, ..., k\}$, $q \geq 1$

$$\sup_{s_1,...,s_n \leq T} E[\sup_{t\leq T} |D_{s_1}^{j_1}...D_{s_n}^{j_n} X_t|^{2q}] + \sup_{s_1,...,s_n \leq T} E[\sup_{t\leq T} |D_{s_1}^{j_1}...D_{s_n}^{j_n} Y_t|^{2q}] \leq C.$$

$$\sup_{s_1,...,s_n \leq T} E[\sup_{t\leq T} |D_{s_1}^{j_1}...D_{s_n}^{j_n} (X_t-Y_t)|^{2q}] \leq Ch^q.$$

Here, C *is a positive constant that depends on* q *and* T.

The method to prove this Proposition is an application of the classical methodology as explained in, for example, Nualart [7], Section 2.2.2.

This Proposition together with the integration by parts formula will allow us the use of the mean value theorem at many stages and to reduce our consideration to the relevant terms that will provide the weak rate of convergence.

Proof: To give an idea of the proof we will show that $X_t, Y_t \in \mathbb{D}^{1,\infty}$ and that

$$\sup_{s\leq T} E[\sup_{t\leq T} |D_s(X_t-Y_t)|^{2q}] \leq Ch^q, \tag{8}$$

in the case $k=1$ and $a \equiv 0$. One starts by defining the Picard iteration sequence for $l = 0, 1, ...$

$$\begin{aligned} dX_t^{l+1} &= b(X_t^l, F(X_t^l; u_t^l))dW_t, \\ dY_t^{l+1} &= b(Y_{\eta(t)}^l, F(Y_{\eta(t)}^l; \tilde{u}_{\eta(t)}^l))dW_t. \end{aligned}$$

$$X_t^0 = Y_t^0 = \xi, \quad u_t^l(dx) = P(X_t^l \in dx), \quad \tilde{u}_t^l(dx) = P(Y_t^l \in dx).$$

Then one proves by induction with the usual techniques (see eg. [8]) that

$$E(\sup_{t\in[0,T]}|X_t^{l+1}-X_t^l|^{2q}) \leq C\int_0^T E(\sup_{s\in[0,v]}|X_s^l-X_s^{l-1}|^{2q})+E(|F(X_v^l;u_v^l)-F(X_v^l;u_v^{l-1})|^{2q})dv.$$

Noting that regardless of the value of x we have

$$|F(x;u_v^l)-F(x;u_v^{l-1})| \leq CE(|X_v^l-X_v^{l-1}|),$$

one obtains the usual inequality that shows the convergence of the Picard sequence to a process X solution of (1). The same method can be applied to Y which gives

$$E(\sup_{t\in[0,T]}|X_t^l-X_t|^{2q})+E(\sup_{t\in[0,T]}|Y_t^l-Y_t|^{2q}) \leq \sum_{j=l+1}^{\infty}\frac{C(q)^j}{j!}, \quad q>1.$$

A similar calculation gives for $S \leq T$,

$$E(\sup_{t\in[0,S]}|X_t^{l+1}-Y_t^{l+1}|^{2q}) \leq C\int_0^S E(\sup_{s\in[0,v]}|X_s^l-Y_s^l|^{2q})+E(|Y_v^l-Y_{\eta(v)}^l|^{2q})dv.$$

From here one obtains by induction that

$$E(\sup_{t\in[0,T]}|X_t^{l+1}-Y_t^{l+1}|^{2q}) \leq h^q\sum_{j=1}^{l+1}\frac{C(q)^j}{j!}, \quad q>1.$$

Here the positive constant $C(q)$ is independent of l. Therefore one can take limits with respect to l above to obtain

$$E(\sup_{t\in[0,T]}|X_t-Y_t|^{2q}) \leq h^qC(q) \quad q>1.$$

Now we compute the stochastic derivatives of X^l and Y^l using their definition and basic rules of derivation. This computation gives

$$\begin{aligned}
D_sX_t^{l+1} &= b(X_s^l,G(X_s^l;u_s^l))+\int_s^t b_x(X_v^l,G(X_v^l;u_v^l))D_sX_v^l dW_v \\
&\quad + \int_s^t b_y(X_v^l,G(X_v^l;u_v^l))G_x(X_v^l;u_v^l)D_sX_v^l dW_v, \quad s\leq t, \\
D_sY_t^{l+1} &= b(Y_{\eta(s)}^l,G(Y_{\eta(s)}^l;\tilde{u}_{\eta(s)}^l))+1[\mu(s)\leq t]\Bigg(\int_{\mu(s)}^t b_x(Y_{\eta(v)}^l,G(Y_{\eta(v)}^l;\tilde{u}_{\eta(v)}^l))D_sY_{\eta(v)}^l dW_v \\
&\quad + \int_{\mu(s)}^t b_y(Y_{\eta(v)}^l,G(Y_{\eta(v)}^l;\tilde{u}_{\eta(v)}^l))G_x(Y_{\eta(v)}^l;\tilde{u}_{\eta(v)}^l)D_sY_{\eta(v)}^l dW_v\Bigg), \quad s\leq t.
\end{aligned}$$

$$D_sX_t^{l+1}=D_sY_t^{l+1}=0, \text{ for } s>t, \quad \mu(s)=\inf\{t_i\in\pi;\ t_i\geq s\}.$$

Then one proves using induction that

$$E(\sup_{s\leq t\leq T}|D_sX_t^{l+1}-D_sX_t^l|^{2q})+E(\sup_{s\leq t\leq T}|D_sY_t^{l+1}-D_sY_t^l|^{2q})\leq\frac{C(q)^{l+1}}{l!},\quad q>1,$$

$$E(\sup_{s\leq t\leq T}|D_sX_t^{l+1}-D_sY_t^{l+1}|^{2q})\leq h^q\sum_{j=1}^{l+1}\frac{C(q)^j}{j!},\quad q>1.$$

The above constants do depend on T and q but not on s or l. Therefore $X_t, Y_t \in \mathbb{D}^{1,\infty}$ because D is a closed operator and furthermore one can take limits with respect to l in the above inequality and obtain (8). $\square$

Now we give a Lemma that will be used when introducing smooth random variables inside integrals.

LEMMA 3.2 *Let* $\{\ F_s^\epsilon(\omega),\ G_s^\epsilon(\omega);\ \omega\in\Omega,\ s\in[0,\infty)\}$ *be continuous adapted random processes where* $\epsilon\in[0,1]$ *is a parameter. Also let* $\alpha\ (\cdot):\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$*,* $\beta(\cdot):[0,\infty)\times(\mathbb{R}^d)^4\to\mathbb{R}$ *and* $\gamma(\cdot):[0,\infty)\times(\mathbb{R}^d)^2\to\mathbb{R}$ *be continuous real valued functions. Suppose that the above random processes satisfy for positive constants* C_1 *and* C_2

$$\sup_{\epsilon\in[0,1]}\sup_{s\in[0,T]}(\|F_s^\epsilon\|_{p,q}+\|G_s^\epsilon\|_{p,q})\leq C_1 \text{ for all } (p,q) \text{ such that } p\leq 2,\ q\leq 2^5, \tag{9}$$

and that $\|\alpha^{(i)}\|_\infty+\sup_{s\in[0,T]}\|\beta_s^{(i)}\|_\infty+\sup_{s\in[0,T]}\|\gamma_s^{(i)}\|_\infty\leq C_2$*, for* $i=1,2$*. Then we have*

$$|E[F_t^\epsilon\alpha(X_t,Y_t)\int_0^t G_s^\epsilon\beta_s(X_s,Y_s,X_{\eta(s)},Y_{\eta(s)})\int_{\eta(s)}^s\gamma_{\eta(s)}(X_{\eta(s)},Y_{\eta(s)})dW^{j_1}(v)dW^{j_2}(s)]|\leq Kh.$$

where $dW^0(s)=ds$*,* $j_1,\ j_2=0,...,k$*.* K *is a positive constants that depend only on* T*,* C_1 *and* C_2*.*

In order to make future calculations easier we will assume without loss of generality that the constant K is chosen so that $(C_1C_2)^5\leq K$.

Proof: The proof is carried out by cases. For example, assume the case $j_1\neq 0$, $j_2=0$. Then the above expectation becomes

$$A=\int_0^t E[F_t^\epsilon\alpha(X_t,Y_t)G_s^\epsilon\beta_s(X_s,Y_s,X_{\eta(s)},Y_{\eta(s)})\int_{\eta(s)}^s\gamma_{\eta(s)}(X_{\eta(s)},Y_{\eta(s)})dW^{j_1}(v)]ds.$$

The Fubini theorem can be applied as the integrands are in $L^2(\Omega)$ uniformly in t. Now we verify that the conditions for (5) are satisfied.

First, $F_t^\epsilon\alpha(X_t,Y_t)G_s^\epsilon\beta_s(X_s,Y_s,X_{\eta(s)},Y_{\eta(s)})\in\mathbb{D}^{1,2}$ by hypothesis and $\gamma_{\eta(s)}(X_{\eta(s)},Y_{\eta(s)})1_{[\eta(s),s]}(\cdot)$

is obviously in the domain of δ^{j_1}. Furthermore the multiplication rule can be applied to explicitely compute

$$D_v^{j_1}[F_t^\epsilon \alpha(X_t, Y_t) G_s^\epsilon \beta_s(X_s, Y_s, X_{\eta(s)}, Y_{\eta(s)})].$$

The proof finishes by noting that due to (9) and the Cauchy-Schwartz inequality

$$|E[D_v^{j_1}[F_t^\epsilon \alpha(X_t, Y_t) G_s^\epsilon \beta_s(X_s, Y_s, X_{\eta(s)}, Y_{\eta(s)})] \gamma_{\eta(s)}(X_{\eta(s)}, Y_{\eta(s)})]| \leq C.$$

Here C is a positive constant independent of π, h, s, t and v. As usual C depends on T, C_1 and C_2. The other cases for j_1, j_2 are solved similarly. Condition (9) is fully used in the case $j_1, j_2 \neq 0$. □

There are other situations where the integration by parts formula has been successfully used to study the approximations to density and distribution functions for diffusion processes. See, Bally and Talay [1], [2] and Kohatsu-Higa [5]. In these articles the Malliavin covariance matrix plays an important role due to the non-smoothness of the function φ.

The proof of Proposition 3.1 is long and will be divided in a series of lemmas.

Proof of Proposition 3.1

Due to the mean value theorem we have that

$$\varphi(X_t) - \varphi(Y_t) = \int_0^1 \varphi_x(Y_t + v(X_t - Y_t)) dv (X_t - Y_t) = \varphi_x(\xi_t^1)(X_t - Y_t),$$

where ξ_t^1 is a random point in the line joining X_t and Y_t. From now on we will denote this by $\xi_t^1 \sim [\varphi(X_t); \varphi(Y_t)]$.

This situation of using the mean value theorem for random quantities will arise often in the rest of the article so we will adopt some conventions. When proving properties of $\varphi_x(\xi_t^1)$, we will always use the expression $\int_0^1 \varphi_x(Y_t + v(X_t - Y_t)) dv$ which is longer to write but easier to handle. Nevertheless in order to simplify notation we will keep using the notation $\varphi_x(\xi_t^1)$ always.

Going back to the proof, we have that due to Lemma 3.2 it would be enough to prove that $X_t - Y_t$ can be written in terms of double integrals where the inner integral has an interval of integration of size less than h like in the statement of Lemma 3.2. This idea can not be carried out directly. Instead, we will have to use a sequence of processes that converge to $X_t - Y_t$. This sequence will have the desired properties of being able to be written in terms of double integrals as established above. Finally we will find that the constant obtained from Lemma 3.2 is uniform along the sequence and therefore the result will follow.

We start by obtaining the stochastic differential equation satisfied by $X_t - Y_t$. This is done by taking the difference between (1) and (6). We then have

$$\begin{aligned} d(X_t - Y_t) &= \{a(X_t, F(X_t; u_t)) - a(Y_{\eta(t)}, F(Y_{\eta(t)}; \widetilde{u}_{\eta(t)}))\}dt \\ &\quad + \{b^j(X_t, G(X_t; u_t)) - b^j(Y_{\eta(t)}, G(Y_{\eta(t)}; \widetilde{u}_{\eta(t)}))\}dW^j(t); \\ X_0 - Y_0 &= 0. \end{aligned}$$

Using the mean value theorem repeatedly one obtains the following linear stochastic differential equation for $Z_t := X_t - Y_t$:

$$\begin{aligned} dZ_t &= \alpha_t Z_t dt + \beta_t^j Z_t dW^j(t) + dH(t) + a_y(Y_t, \xi_t^3)(F(Y_t; u_t) - F(Y_t; \tilde{u}_t))dt \\ &\quad + b_y^j(Y_t, \eta_t^3)(G(Y_t; u_t) - G(Y_t; \tilde{u}_t))dW^j(t), \end{aligned} \tag{10}$$

where

$$\begin{aligned} \alpha_t &= a_x(\xi_t^2, F(X_t; u_t)) + a_y(Y_t, \xi_t^3) F_x(\xi_t^6; u_t), \\ \beta_t^j &= b_x^j(\eta_t^2, G(X_t; u_t)) + b_y^j(Y_t, \eta_t^3) G_x(\eta_t^6; u_t), \\ dH_t &= \Big\{ a_x(\xi_t^4, F(Y_t; \tilde{u}_t))(Y_t - Y_{\eta(t)}) + a_y(Y_t, \xi_t^5)(F(Y_t; \tilde{u}_t) - F(Y_{\eta(t)}; \tilde{u}_{\eta(t)})) \Big\} dt \\ &\quad + \Big\{ b_x^j(\eta_t^4, G(Y_t; \tilde{u}_t))(Y_t - Y_{\eta(t)}) + b_y^j(Y_t, \eta_t^5)(G(Y_t; \tilde{u}_t) - G(Y_{\eta(t)}; \tilde{u}_{\eta(t)})) \Big\} dW^j(t) \\ \xi_t^2 &\sim [a(X_t, F(X_t; u_t)); a(Y_t, F(X_t; u_t))], \\ \xi_t^3 &\sim [a(Y_t, F(X_t; u_t)); a(Y_t, F(Y_t; \tilde{u}_t))], \\ \xi_t^4 &\sim [a(Y_t, F(Y_t; \tilde{u}_t)); a(Y_{\eta(t)}, F(Y_t; \tilde{u}_t))], \\ \xi_t^5 &\sim [a(Y_{\eta(t)}; F(Y_t; \tilde{u}_t)); a(Y_{\eta(t)}, F(Y_{\eta(t)}; \tilde{u}_{\eta(t)}))], \\ \xi_t^6 &\sim [F(X_t; u_t); F(Y_t; u_t)]. \end{aligned}$$

η^2,..., η^6 are analogously defined except that one changes a and f by b^j and g. These quantities also depend on j but we do not write this dependence explicitely to avoid cumbersome notation.

Now solving the equation (10) (see, eg. Protter [9], Section V.9) we have:

$$\mathcal{E}(t)^{-1} Z_t = \int_0^t \mathcal{E}(s)^{-1} dH^1(s) + \int_0^t \mathcal{E}(s)^{-1} a_y(Y_s, \xi_s^3) \Big(E' f(Y_s - X_s') - E' f(Y_s - Y_s') \Big) ds$$

$$
\begin{aligned}
&+\int_0^t \mathcal{E}(s)^{-1} b_y^j(Y_s,\eta_s^3)\Big(E'g(Y_s-X'_s)-E'g(Y_s-Y'_s)\Big)dW^j(s)\\
&-\int_0^t \mathcal{E}(s)^{-1}\beta_s^j b_y^j(Y_s,\eta_s^3)\Big(E'g(Y_s-X'_s)-E'g(Y_s-Y'_s)\Big)ds, \qquad (11)
\end{aligned}
$$

where

$$
dH_s^1 \quad = dH_s-\beta_s^j\Big\{b_x^j(\eta_s^4,G(Y_s;\tilde u_s))(Y_s-Y_{\eta(s)})+b_y^j(Y_s,\eta_s^5)\Big(G(Y_s;\tilde u_s)-G(Y_{\eta(s)};\tilde u_{\eta(s)})\Big)\Big\}ds,
$$

$\mathcal{E}(t)$ denotes the stochastic exponential, unique strong invertible $d\times d$ matrix solution of the following linear stochastic differential equation

$$
\mathcal{E}(t)=I+\int_0^t \alpha_s\mathcal{E}(s)dt+\int_0^t \beta_s^j\mathcal{E}(s)dW^j(s).
$$

E' denotes the expectation in an independent copy of $(\Omega,\mathcal{F},P)$, denoted by $(\Omega',\mathcal{F}',P')$. In this space we also construct the appropiate copies of X and Y, denoted by X' and Y'.

From (11), using again the mean value theorem we have that $U(t)=\mathcal{E}(t)^{-1}Z_t$ satisfies the following equation:

$$
\begin{aligned}
U(t) \quad = \quad &\int_0^t \mathcal{E}(s)^{-1}dH^1(s)+\int_0^t \mathcal{E}(s)^{-1}a_y(Y_s,\xi_s^3)E'[f_x(Y_s-\xi_s^{7'})\mathcal{E}'(s)U'(s)]ds\\
&+\int_0^t \mathcal{E}(s)^{-1}b_y^j(Y_s,\eta_s^3)E'[g_x(Y_s-\eta_s^{7'})\mathcal{E}'(s)U'(s)]dW^j(s)\\
&-\int_0^t \mathcal{E}(s)^{-1}\beta_s^j b_y^j(Y_s,\eta_s^3)E'[g_x(Y_s-\eta_s^{7'})\mathcal{E}'(s)U'(s)]ds, \qquad (12)
\end{aligned}
$$

where $\xi_s^{7'}\sim[f(Y_s-X'_s);f(Y_s-Y'_s)]$ (note that $\xi_s^{7'}$ depends on $(\omega,\omega')\in\Omega\times\Omega'$). $\eta_s^{7'}$ is defined analogously changing f by g. By Lemma 5.1 in the Appendix we have that (12) has a unique strong solution in $L^p(\Omega)$ with the sup norm on compact intervals for $p\geq 2$ and that furthermore the Picard iteration sequence $\{U_k(s);\ s\in[0,\infty),\ k=0,...\}$ converges to this solution. U_k is defined by iteration as follows, $U_{-1}(t)\equiv 0$:

$$
\begin{aligned}
U_{k+1}(t) \quad = \quad &\int_0^t \mathcal{E}(s)^{-1}dH^1(s)+\int_0^t \mathcal{E}(s)^{-1}a_y(Y_s,\xi_s^3)E'[f_x(Y_s-\xi_s^{7'})\mathcal{E}'(s)U'_k(s)]ds\\
&+\int_0^t \mathcal{E}(s)^{-1}b_y^j(Y_s,\eta_s^3)E'[g_x(Y_s-\eta_s^{7'})\mathcal{E}'(s)U'_k(s)]dW^j(s)\\
&-\int_0^t \mathcal{E}(s)^{-1}\beta_s^j b_y^j(Y_s,\eta_s^3)E'[g_x(Y_s-\eta_s^{7'})\mathcal{E}'(s)U'_k(s)]ds. \qquad (13)
\end{aligned}
$$

The rate of convergence of the Picard iteration is given by

$$
E[\sup_{t\in[0,T]}|U_k(t)-U(t)|^p]\leq\frac{C^k}{k!},
$$

for any $p \geq 2$. The positive constant C depends on T and p. Therefore we have that for any $t \in [0,T]$ and $p \geq 2$,

$$L^p - \lim_{k\to\infty} \varphi_x(\xi_t^1)\mathcal{E}(t)U_k(t) = \varphi(X_t) - \varphi(Y_t).$$

From here it is obvious that the proof will be finished with the following result. □

LEMMA 3.3 *For* $t \in [0,T]$

$$|E[\varphi_x(\xi_t^1)\mathcal{E}(t)U_k(t)]| \leq hC_t, \tag{14}$$

where C_t *is a positive constant independent of* k *and* π *which is uniformly bounded for* t *in compact intervals.*

Proof: The proof goes by induction. For $k = 0$ we have to compute

$$\begin{aligned} E[\varphi_x(\xi_t^1)\mathcal{E}(t)\int_0^t \mathcal{E}(s)^{-1}dH^1(s)] = E\Big[\varphi_x(\xi_t^1)\mathcal{E}(t)\int_0^t \mathcal{E}(s)^{-1}\Big\{&[a_x(\xi_s^4, F(Y_s;\tilde{u}_s))(Y_s - Y_{\eta(s)}) \\ &+a_y(Y_s,\xi_s^5)(F(Y_s;\tilde{u}_s) - F(Y_{\eta(s)};\tilde{u}_{\eta(s)}))]ds + [b_x^j(\eta_s^4, G(Y_s;\tilde{u}_s)) \\ &(Y_s - Y_{\eta(s)}) + b_y^j(Y_s,\eta_s^5)(G(Y_s;\tilde{u}_s) - G(Y_{\eta(s)};\tilde{u}_{\eta(s)}))]dW^j(s) \\ &-\beta_s^j[b_x^j(\eta_s^4, G(Y_s;\tilde{u}_s))(Y_s - Y_{\eta(s)}) + b_y^j(Y_s,\eta_s^5)(G(Y_s;\tilde{u}_s) - G(Y_{\eta(s)};\tilde{u}_{\eta(s)}))]ds\Big\}\Big]. \end{aligned} \tag{15}$$

We consider one of the terms that appears above. The other terms can be handled similarly. This is

$$E[\varphi_x(\xi_t^1)\mathcal{E}(t)\int_0^t \mathcal{E}(s)^{-1}\beta_s^j b_x^j(\eta_s^4, G(Y_s;\tilde{u}_s))(Y_s - Y_{\eta(s)})ds]. \tag{16}$$

We want to prove that the expectation above can be bounded by a positive constant times h. Note that

$$Y_s - Y_{\eta(s)} = \int_{\eta(s)}^s a(Y_{\eta(s)}, F(Y_{\eta(s)};\tilde{u}_{\eta(s)}))dv + \int_{\eta(s)}^s b^j(Y_{\eta(s)}, G(Y_{\eta(s)};\tilde{u}_{\eta(s)}))dW_v^j.$$

Now we apply Lemma 3.2 to each of the terms above. For the first integral in the difference $Y_s - Y_{\eta(s)}$, we have $j_1 = 0$, $j_2 = 0$, $\alpha \equiv 1$, $\gamma_t(x,y) = a(y, F(y;\tilde{u}_t))$, $G_t^h = \mathcal{E}(t)^{-1}$,

$$F_t^h = \varphi_x(\xi_t^1)\mathcal{E}(t) = \int_0^1 \varphi_x(Y_t + v(X_t - Y_t))dv\mathcal{E}(t),$$

$$\beta_t(w,x,y,z) = \Big(\int_0^1 b_x^j(x + v(w-x), G(w;u_t))dv + \int_0^1 b_y^j(x, F(x;\tilde{u}_t)$$

$$+v(F(w;u_t)-F(x;\tilde{u}_t)))dv\int_0^1 G_x(x+v(w-x);u_t)dv\Big)\Big(\int_0^1 b_x^j(z+v(x-z),G(x;\tilde{u}_t))dv\Big).$$

It is not difficult to verify that the conditions in Lemma 3.2 are satisfied. In particular, the constant C_2 in Lemma 3.2 is obtained from the bounds for the derivatives of the functions a, f, b and g. The constant C_1 comes from the bounds for the function $\varphi \in C_p^\infty$ and for the random processes $\mathcal{E}(s)$, $\mathcal{E}(s)^{-1}$, $X(s)$ and $Y(s)$ in the appropiate $\| \, \|_{p,q}$ norms.

It then follows that there exists a positive constant K such that the expectation in (16) is bounded by Kh. The same procedure has to be carried out with each term that appears in (15).

Each time that we apply Lemma 3.2, we generate a set of constants C_1, C_2 and K. To simplify future notation we will denote by C_1, C_2 and K the maximum of these constants in each class.

Without loss of generality we will assume that the bounds for the function φ are also satisfied by f and g. In fact, f and g are smooth with bounded derivatives.

For the rest of the proof this constant K will be fixed. Using the same procedure for each term in (15) one obtains that there exists a constant $\overline{K}$ (the maximum of all the K obtained trough the application of Lemma 3.2 in (15)) such that

$$|E[\varphi_x(\xi_t^1)\mathcal{E}(t)U_0(t)]| = |E[\varphi_x(\xi_t^1)\mathcal{E}(t)\int_0^t \mathcal{E}(s)^{-1}dH^1(s)]| \leq a\overline{K}h,$$

for some fixed natural number a (say 18). Furthermore $\overline{K} \geq (\overline{C}_1\overline{C}_2)^5$, where $\overline{C}_1$ and $\overline{C}_2$ are the respective maxima of the constants C_1 and C_2 used in the applications of Lemma 3.2 in (15). For the rest of the proof, the constants $\overline{K}$, $\overline{C}_1$ and $\overline{C}_2$ are fixed constants.

Note here that $\varphi \in C_p^\infty$ and therefore in particular we also have

$$|E'[f_x(Y_s-\xi_s^{7'})\mathcal{E}'(s)U_0'(s)]| \leq a\overline{K}h,$$

for all $\omega \in \Omega$. This statement is satisfied because the function $\alpha(x,y) = \int_0^1 f_x(Y_s + y + v(x-y))dv$ and its derivatives are bounded independently from the value of ω. Therefore the same procedure as used in the case of $\varphi_x(\xi_t^1)$ can be used for $f_x(Y_s-\xi_s^{7'}) = -\alpha(-X_t', -Y_t')$.

The same property is satisfied for g and therefore,

$$|E'[g_x(Y_s-\eta_s^{7'})\mathcal{E}'(s)U_0'(s)]| \leq a\overline{K}h.$$

Using (13) and the integration by parts formula one has that

$$\begin{aligned}
|E[\varphi_x(\xi_t^1)\mathcal{E}(t)U_1(t)]| &\leq |E[\varphi_x(\xi_t^1)\mathcal{E}(t)U_0(t)]| + 2\overline{C}_2^2 E\Bigg[|\varphi_x(\xi_t^1)\mathcal{E}(t)|| \int_0^t \mathcal{E}(s)^{-1} \\
&\quad (E'[f_x(Y_s-\xi_s^{7\prime})\mathcal{E}'(s)U_0'(s)] + E'[g_x(Y_s-\eta_s^{7\prime})\mathcal{E}'(s)U_0'(s)])ds|\Bigg] \\
&\quad +|E[\int_0^t \mathcal{E}(s)^{-1} b_y^j(Y_s,\eta_s^3)E'[g_x(Y_s-\eta_s^{7\prime})\mathcal{E}'(s)U_0'(s)]D_s^j(\varphi_x(\xi_t^1)\mathcal{E}(t))ds]| \\
&\leq a\overline{K}h + (4\overline{C}_2^4\overline{C}_1^2 + \overline{C}_2^4\overline{C}_1^4)a\overline{K}ht \\
&\leq a\overline{K}h + (a\overline{K})^2 ht.
\end{aligned}$$

In the last inequality we have used that $(\overline{C}_1\overline{C}_2)^4 \leq \overline{K}$ and that a is bigger than 5. Repeating the same procedure as before we have that

$$|E[\varphi_x(\xi_t^1)\mathcal{E}(t)U_k(t)]| \leq \sum_{j=0}^{k} \frac{(a\overline{K}t)^{k+1}}{k!}h.$$

From here the result follows. □

Remark. The proof of Proposition 3.1 can be used with some variations in other settings where the solution does not possess the Markov property. See, e.g. [6].

4 The approximation of the measure

Here we establish the estimate of the error incurred in approximating the measure u_t.

PROPOSITION 4.1 *Let $\varphi \in C_p^\infty$ then*

$$E[|\varphi(Y_t^i) - \varphi(\overline{X}_t^i)|] \leq C\frac{1}{\sqrt{n}}. \tag{17}$$

Here the constant C is independent of π and i but it depends on φ and t. Furthermore it is uniformly bounded for t in compact intervals.

Proof: To simplify the proof we will assume that $a \equiv 0$ and $d = k = 1$. The general case is dealt with similar techniques. We start proving that

$$E(\sup_{t\leq T} |\overline{X}_t^i|^2) \leq C \tag{18}$$

where the positive constant C is independent of h, i and the partition π.

First apply Itô's formula to obtain the semimartingale representation for $\frac{1}{n}\sum_{i=1}^{n}|\overline{X}_t^i|^2$. One obtains:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}|\overline{X}_t^i|^2 &= |\xi|^2+\int_0^t \frac{1}{n}\sum_{i=1}^{n}2\overline{X}_s^i b(\overline{X}_{\eta(s)}^i,\frac{1}{n}\sum_{j=1}^{n}g(\overline{X}_{\eta(s)}^i-\overline{X}_{\eta(s)}^j))dW_i(s)\\ &\quad +\int_0^t \frac{1}{n}\sum_{i=1}^{n}b(\overline{X}_{\eta(s)}^i,\frac{1}{n}\sum_{j=1}^{n}g(\overline{X}_{\eta(s)}^i-\overline{X}_{\eta(s)}^j))^2 ds.\end{aligned}$$

Using the Lipschitz properties of b and g, Burkholder, Cauchy-Schwartz inequalities and the standard inequality $\sum_{i,j=1}^{n}|x_i-x_j|^2\le 2n\sum_{i=1}^{n}|x_i|^2$, one obtains that

$$E(\sup_{t\le T}\frac{1}{n}\sum_{i=1}^{n}|\overline{X}_t^i|^2)\le C\Big(T+\int_0^T E(\sup_{t\le s}\frac{1}{n}\sum_{i=1}^{n}|\overline{X}_t^i|^2)ds\Big).$$

Therefore using Gronwall inequality we have that $E(\sup_{t\le T}\frac{1}{n}\sum_{i=1}^{n}|\overline{X}_t^i|^2)\le C$. Now we use the Itô formula for $|\overline{X}_t^i|^2$ and apply arguments as shown above we obtain (18).

Now we apply a modification of the above method to obtain the inequality (17). As in the proof of the Proposition 3.1, we have that there exists $\lambda_t^1(i)\sim[\varphi(\overline{X}_t^i);\varphi(Y_t^i)]$ such that

$$\varphi(\overline{X}_t^i)-\varphi(Y_t^i)=\int_0^1\varphi_x(Y_t^i+v(\overline{X}_t^i-Y_t^i))dv=\varphi_x(\lambda_t^1(i))(\overline{X}_t^i-Y_t^i).$$

Here we find the stochastic differential equation satisfied by $Z_t^i=\overline{X}_t^i-Y_t^i$ as we did in (10).

$$\begin{aligned}dZ_t^i &= \Big\{b_x(\lambda_{\eta(t)}^2(i),G(Y_{\eta(t)}^i;\tilde{u}_{\eta(t)}))Z_{\eta(t)}^i+b_y(\overline{X}_{\eta(t)}^i,\lambda_{\eta(t)}^3(i)) \\ &\quad \Big[\frac{1}{n}\sum_{l=1}^{n}[g_x(\lambda_{\eta(t)}^4(i,l))(Z_{\eta(t)}^i-Z_{\eta(t)}^l)]+\frac{1}{n}\sum_{l=1}^{n}[g(Y_{\eta(t)}^i-Y_{\eta(t)}^l)-G(Y_{\eta(t)}^i;\tilde{u}_{\eta(t)})]\Big]\Big\}dW_i(t).\end{aligned}\tag{19}$$

Here

$$\begin{aligned}\lambda_t^2(i) &\sim [b(\overline{X}_t^i,G(Y_t^i;\tilde{u}_t));b(Y_t^i,G(Y_t^i;\tilde{u}_t))],\\ \lambda_t^3(i) &\sim [b(\overline{X}_t^i,G(\overline{X}_t^i;\overline{u}_t));b(\overline{X}_t^i,G(Y_t^i;\tilde{u}_t))],\\ \lambda_t^4(i,l) &\sim [g(\overline{X}_t^i-\overline{X}_t^l);g(Y_t^i-Y_t^l)].\end{aligned}$$

Note that due to the independence of Y^i for $i=1,...,n$ we have that

$$E\left(\left|\frac{1}{n}\sum_{l=1}^{n}[g(Y_{\eta(t)}^l-Y_{\eta(t)}^i)-G(Y_{\eta(t)}^i;\tilde{u}_{\eta(t)})]\right|^2\right)\le\frac{C\exp(Ct)}{n}.$$

As in the deduction of (18) we have that using the above inequality and (19) one obtains

$$E(\sup_{s\le t}\frac{1}{n}\sum_{i=1}^{n}|Z_s^i|^2) \le C(\int_0^t E(\sup_{s\le v}\frac{1}{n}\sum_{i=1}^{n}|Z_s^i|^2)dv+\frac{t}{n}).$$

Applying Gronwall inequality one obtains that

$$E(\sup_{s\le t}\frac{1}{n}\sum_{i=1}^{n}|Z_s^i|^2)\le \frac{Ct\exp(Ct)}{n}.$$

To finish the proof is enough to consider the quantity

$$\begin{aligned} E\sup_{s\le t}|Z_s^i|^2 &\le C(\int_0^t E|Z_u^i|^2+E(\sup_{s\le v}\frac{1}{n}\sum_{l=1}^{n}|Z_s^l|^2)dv+\frac{t}{n}) \\ &\le C(\int_0^t E|Z_v^i|^2dv+\frac{t\exp(Ct)}{n}). \end{aligned}$$

Therefore the result follows by Gronwall inequality. □

Now we give the proof of Corollary 1.1.

Proof of Corollary 1.1: By Theorem 1.2 it is enough to analyze the quantity

$$\begin{aligned} E|\frac{1}{n}\sum_{i=1}^{n}\varphi(\overline{X}_t^i)-E\varphi(\overline{X}_t^i)| &\le C\{\frac{1}{n}\sum_{i=1}^{n}E|\varphi(\overline{X}_t^i)-\varphi(Y_t^i)|+\frac{1}{n}(\sum_{i=1}^{n}E(\varphi(Y_t^i)-E\varphi(Y_t^i))^2)^{1/2} \\ &\quad +|E[\varphi(Y_t^i)-\varphi(\overline{X}_t^i)]| \\ &\le C(\frac{1}{\sqrt{n}}+h). \end{aligned}$$

Here we have used Proposition 4.1 and the fact that $E(\varphi(Y_t^i))^2<C$, where C is a positive constant independent of π, h and $t\in[0,T]$. □

This result hints that the rates found in Bossy and Talay [3] may be improved. We hope to address this issue in future publications.

5 Appendix

Here we prove th existence and uniqueness of the solution to Eq. (12), used in Section 3.

LEMMA 5.1 *(12) has a unique strong solution in $L^p(\Omega)$ with the sup norm on compact intervals and furthermore the Picard iteration sequence defined in (13) converges to this solution so that*

$$E[\sup_{t\in[0,T]}|U_k(t)-U(t)|^p]\le\frac{C^k}{k!},$$

The positive constant C depends only on $p\ge 2$ and $T>0$.

Proof: The proof of this Lemma is a variation of the classical proof for diffusion processes. For this reason we will only prove uniqueness for $p = 2$. The proofs of the other facts require similar techniques.

Suppose that (12) has 2 solutions U and $\overline{U}$ in $L^2(\Omega)$ with the sup norm on compact intervals. Then take the difference between these solutions to obtain:

$$\begin{aligned} U(t) - \overline{U}(t) &= \int_0^t \mathcal{E}(s)^{-1} a_y(Y_s, \xi_s^3) E'[f_x(Y_s - \xi_s^{7'})\mathcal{E}'(s)(U'(s) - \overline{U}'(s))]ds \\ &+ \int_0^t \mathcal{E}(s)^{-1} b_y^j(Y_s, \eta_s^3) E'[g_x(Y_s - \eta_s^{7'})\mathcal{E}'(s)(U'(s) - \overline{U}'(s))]dW^j(s) \\ &- \int_0^t \mathcal{E}(s)^{-1} \beta_s^j b_y^j(Y_s, \eta_s^3) E'[g_x(Y_s - \eta_s^{7'})\mathcal{E}'(s)(U'(s) - \overline{U}'(s))]ds. \end{aligned}$$

Then take the supremum of the square of the norm and using Burkholder inequality one has:

$$E[\sup_{t\in[0,T]} |U(t) - \overline{U}(t)|^2] \leq C \int_0^T E[|\mathcal{E}(s)^{-1}|^2] E[|\mathcal{E}(s)|^2] E[\sup_{v\in[0,s]} |U(v) - \overline{U}(v)|^2] ds.$$

Here we have used the fact that the derivatives of a, b, f and g are bounded. The proof finishes by using Gronwall inequality with the fact that $\sup_{s\in[0,T]} E(|\mathcal{E}(s)|^2) < \infty$ and $\sup_{s\in[0,T]} E(|\mathcal{E}(s)^{-1}|^2) < \infty$. $\square$

References

[1] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations I: convergence rate of the distribution function. Probability Theory and Related Fields, 104, 43-60 (1996).

[2] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations: II. Convergence rate of the density. Monte Carlo Methods and Its Applications, 2, 93-128 (1996).

[3] M. Bossy and D. Talay. A Stochastic Particle Method for the McKean-Vlasov and Burgers equations. Math. Comp., 66 (217), 157-192 (1997).

[4] Kloeden P.E., Platen E.: *Numerical Solution of Stochastic Differential Equations,* Applications of Mathematics 23, Springer, New York, 1992.

[5] A. Kohatsu-Higa. High Itô-Taylor approximations to heat kernels . Journal of Mathematics of Kyoto University, 37, 129-150 (1997).

[6] A. Kohatsu-Higa. Weak approximations for stochastic differential equations with boundary conditions. Preprint (1997).

[7] D. Nualart. *The Malliavin Calculus and Related Topics.* Springer Verlag 1995.

[8] S. Ogawa. Some problems in the simulation of nonlinear diffusion processes. Mathematics and Computers in Simulation, 38, 217-223 (1995).

[9] P. Protter. *Stochastic Integration and Differential Equations.* Applications of Mathematics 21. Springer-Verlag 1992.