J. Math. Kyoto Univ. (JMKYAZ)
36-3 (1996) 499-518

Donsker's delta functions and approximation of heat kernels by the time discretization methods

By

Yaozhong Hu* and Shinzo Watanabe

Introduction

Time discretization approximation schemes for solutions of stochastic differential equations have been studied by many people and are treated, e.g., in the book of Kloeden-Platen [Kl-Pl92]. Since heat kernels are the probability densities of the law of solutions, it might be worth-while to ask if these approximation schemes provide a natural scheme of approximation for heat kernels. Purpose of this paper is to propose one of such schemes with a help of Malliavin calculus.

In section 1, we introduce the notion of Donsker's delta functions as a class of generalized Wiener functionals on Wiener space. In section 2, we obtain a general approximation result for Donsker's delta functions. In section 3, we consider the case of Wiener functionals given by solutions to stochastic differential equations. An Itô-Taylor approximation scheme of order γ for the solution has been introduced by Kloeden and Platen [Kl-Pl95]. Here we improve their result of the strong convergence in the L_2-norm to the strong convergence in every Sobolev norm in the Malliavin calculus (Theorem 3.1). This is a main result of this paper and its proof is given in section 4. This result, combined with general results in section 2, yields some strong approximation scheme for Donsker's delta functions and thereby an approximation result for the heat kernel in the form of Theorem 3.2. However, it should be remarked that the heat kernel is given by a generalized expectation of Donsker's delta function and therefore, what is involved in this problem is essentially an weak approximation. The rate of convergence in Theorem 3.2 is that of the strong approximation and it can be improved to the rate of weak convergence. For such improvements, we refer to the recent works by Bally and Talay [B-T95] and Kohatsu-Higa [Ko95].

1. Malliavin calculus and Donsker's delta functions

Let (W, H, P) be a (classical or abstract) Wiener space, where H is the Cameron-Martin Hilbert space and P is the Wiener measure. Let $F\colon W \to \mathbf{R}^d$

---

* Supported by NAVF researchship. Department of Mathematics, University of Oslo, N-0316, Oslo, Norway
Received September 2, 1995

be a d-dimensional Wiener functional, i.e., a P-measurable mapping (more precisely, an equivalent class of P-measurable mappings) from W to $\mathbf{R}^d$. Let $T \in \mathscr{S}'(\mathbf{R}^d)$ be a Schwartz tempered distribution on $\mathbf{R}^d$. We want to give a sense to the composite $T \circ F$. This should be naturally defined as a limit of Wiener functionals $\phi_n \cdot F$, where $\phi_n \in \mathscr{S}(\mathbf{R}^d)$ ($\equiv$ the space of test functions on $\mathbf{R}^d$) such that $\phi_n \to T$ in $\mathscr{S}'(\mathbf{R}^d)$ and so we need a framework to realize this convergence and to identify the limit. Such a framework is provided by the Malliavin calculus as follows (cf. [Wa84]).

Starting with the family of L^p spaces $\mathbf{L}^p$ over the Wiener space (W, H, P) and the basic differential operators D (the gradient), D^* (the dual of D, the divergence) and $L = -D^*D$ (the Ornstein-Uhlenbeck operator), a family of Sobolev spaces $\mathbf{D}^p_\alpha$, $1 < p < \infty$, $\alpha \in \mathbf{R}$, can be introduced as

$$\mathbf{D}^p_\alpha = (I - L)^{-\alpha/2}(\mathbf{L}^p)$$

with the norm

$$\|F\|_{p,\alpha} = \|(I - L)^{\alpha/2}F\|_p\,, \quad F \in \mathbf{D}^p_\alpha,$$

where $\|\cdot\|_p$ is the L_p norm on $\mathbf{L}^p$. Then clearly,

$$\mathbf{D}^p_0 = \mathbf{L}^p,\ \mathbf{D}^p_\alpha \subset \mathbf{D}^{p'}_{\alpha'} \qquad \text{if } p \geq p' \quad \text{and} \quad \alpha \geq \alpha'$$

and

$$(\mathbf{D}^p_\alpha)' = \mathbf{D}^{p/(p-1)}_{-\alpha}.$$

As in the Schwartz theory of distributions, we introduce the space of test Wiener functionals by

$$\mathbf{D}^{\infty-}_\infty \equiv \textstyle\bigcap_{\alpha > 0} \bigcap_{1 < p < \infty} \mathbf{D}^p_\alpha$$

and its dual, the space of generalized Wiener functionals, by

$$\mathbf{D}^{1+}_{-\infty} \equiv \textstyle\bigcup_{\alpha < 0} \bigcup_{1 < p < \infty} \mathbf{D}^p_\alpha.$$

When we consider, more generally, E-valued functional, E being a separable real Hilbert space, the corresponding Sobolev spaces are denoted by $\mathbf{D}^p_\alpha[E]$.

The natural coupling of $G \in \mathbf{D}^p_\alpha$ and $\Phi \in (\mathbf{D}^p_\alpha)' = \mathbf{D}^{p/(p-1)}_{-\alpha}$ or that of $G \in \mathbf{D}^{\infty-}_\infty$ and $\Phi \in \mathbf{D}^{1+}_{-\infty}$ is denoted by $E(G \cdot \Phi)$; in particular, when $G = \mathbf{1}$, the Wiener functional identically equal to one, $E(\mathbf{1} \cdot \Phi)$ is simply denoted by $E(\Phi)$ and is called the generalized expectation of $\Phi \in \mathbf{D}^{1+}_{-\infty}$. $E(\Phi)$ coincides with the usual expectation $\int_W \Phi(w)P(dw)$ when Φ is given by an integrable random variable.

Let $\hat{C}(\mathbf{R}^d) = \{f\colon \mathbf{R}^d \to \mathbf{R},$ continuous, $\lim_{|x| \to \infty} f(x) = 0\}$ and define a family of Banach subspaces of $\mathscr{S}'(\mathbf{R}^d)$ by

$$\mathscr{S}_{2n} = (1 + |x|^2 - \Delta)^{-n}(\hat{C}(\mathbf{R}^d))\,, \qquad n = 0,\ \pm 1,\ \pm 2,\ \cdots$$

with the norm

$$\|u\|_{S_{2n}} = \sup_{x \in \mathbf{R}^d} |[(1 + |x|^2 - \Delta)^n u](x)|$$

so that

$$\mathscr{S}(\mathbf{R}^d) \subset \cdots \subset \mathscr{S}_2 \subset \mathscr{S}_0 = \hat{C}(\mathbf{R}^d) \subset \mathscr{S}_{-2} \subset \cdots \subset \mathscr{S}'(\mathbf{R}^d)$$

and

$$\mathscr{S}(\mathbf{R}^d) = \cap_{n=1}^{\infty} \mathscr{S}_{2n}, \quad \mathscr{S}'(\mathbf{R}^d) = \cup_{n=1}^{\infty} \mathscr{S}_{-2n}.$$

Definition 1.1. *Given a d-dimensional Wiener functional* $F\colon W \to \mathbf{R}^d$ *and* $T \in \mathscr{S}'(\mathbf{R}^d)$, *we say that* $T \circ F$ *is defined in* $\mathbf{D}_{-\infty}^{1+}$ *and* $T \circ F = \Phi$ *for some* $\Phi \in \mathbf{D}_{-\infty}^{1+}$ *if there exist* $n \in \mathbf{Z}^+$ *(the set of nonnegative integers),* $1 < p < \infty$, $\alpha > 0$ *such that* $T \in \mathscr{S}_{-2n}$, $\Phi \in \mathbf{D}_{-\alpha}^p$ *and the following holds: For every sequence* $\{\phi_k\colon k = 1, \ , \dots\} \subset \mathscr{S}(\mathbf{R}^d)$ *such that* $\|\phi_k - T\|_{S_{-2n}} \to 0$ *as* $k \to \infty$, *it holds that* $\|\phi_k \circ F - \Phi\|_{p,-\alpha} \to 0$ *as* $k \to \infty$ *(Note that* $\phi_k \circ F \in \mathbf{L}^\infty \subset \mathbf{D}_{-\alpha}^p$ *for all k).*

Clearly Φ is uniquely determined from F and T.

It is well-known [Wa84] that if F is smooth in the sense that $F \in \mathbf{D}_\infty^{\infty-}[\mathbf{R}^d]$, i.e., $F = (F^1, \dots, F^d)$ with $F^i \in \mathbf{D}_\infty^{\infty-}$, and non-degenerate in the sense that

$$(\det \sigma_F)^{-1} \in \mathbf{L}^{\infty-} \quad (\equiv \cap_{1<p<\infty} \mathbf{L}^p,)$$

where $\sigma_F^{ij} = \langle DF^i, DF^j \rangle_H$ is the Malliavin covariance of F, then for every $T \in \mathscr{S}'(\mathbf{R}^d)$, $T \circ F$ can be defined in $\mathbf{D}_{-\infty}^{1+}$ and it holds that

$$T \circ F \in \cup_{\alpha>0} \cap_{1<p<\infty} \mathbf{D}_{-\alpha}^p.$$

In the particular case of $T = (1 - \Delta)^{\beta/2} \delta_y$, $\beta \geq 0$, $y \in \mathbf{R}^d$, where $\delta_y(\cdot) = \delta_0(\cdot - y)$ is the Dirac delta function at y and Δ is the Laplacian on $\mathbf{R}^d$, it is known [Wa93] that

$$T \circ F = (1 - \Delta)^{\beta/2} \delta_y \circ F \in \cap_{1<p<\infty} \mathbf{D}_{-\alpha}^p, \qquad \text{if } \alpha > \beta + d,$$

more precisely,

$$(1.1) \qquad (1 - \Delta)^{\beta/2} \delta_y \circ F \in \mathbf{D}_{-\alpha}^p \qquad \text{if } \alpha > \beta + \frac{d}{q}, \quad \frac{1}{p} + \frac{1}{q} = 1.$$

In particular, when $\beta = 0$,

$$\delta_y \circ F \in \mathbf{D}_{-\alpha}^p \qquad \text{if } \alpha > \frac{d}{q}, \quad \frac{1}{p} + \frac{1}{q} = 1.$$

From this we can see that $\delta_y \circ F$ is arbitrarily close to $\mathbf{L}^1$ space. In particular, we can conclude that if $G \in \mathbf{D}_\alpha^q$, $q > 1$, $\alpha > \frac{d}{q}$, then $u(y) = E[G \cdot \delta_y \circ F] \in C^\beta(\mathbf{R}^d)$ for every $0 \leq \beta < \alpha - \frac{d}{q}$, where $C^\beta(\mathbf{R}^d)$ is the closure of $\mathscr{S}(\mathbf{R}^d)$ with respect to the norm

$$\|\phi\|_\beta = \sum_{m;|m| \leq [\beta]} \|\partial^m \phi\|_\infty + \sum_{m;|m|=[\beta]} \sup_{x \neq y} \frac{|\partial^m \phi(x) - \partial^m \phi(y)|}{|x - y|^{\beta - [\beta]}}.$$

$\delta_y \circ F$ is called a Donsker's delta function and was studied first by H-H. Kuo in the frame of white noise analysis. Note that

$$E[G \cdot \delta_y \circ F] = E[G|F = y]p_F(y),$$

where $p_F(y)$ is the density at y of the probability law of F. Thus we see that these notions are important in the study of the densities and conditional expectations for Wiener functionals.

2. An approximation theorem for Donsker's delta functions

Suppose that a d-dimensional Wiener functional $F: W \to \mathbf{R}^d$ be given which is smooth and non-degenerate in the sense explained above. Then, for every $\beta > 0$ and $y \in \mathbf{R}^d$, $(1 - \Delta)^{\beta/2}\delta_y \circ F$ is defined in $\mathbf{D}_{-\infty}^{1+}$. Suppose that a sequence $\{F_n: n = 1, 2, \dots\}$ of d-dimensional smooth Wiener functionals converge to F as $n \to \infty$ in a certain sense. We want to obtain an approximation of $(1 - \Delta)^{\beta/2}\delta_y \circ F$ in terms of F_n.

Theorem 2.1. *Let F_n, $n = 1, 2, \dots$ and F be smooth d-dimensional Wiener functionals, i.e.,*

$$F_n\,, \quad F \in \mathbf{D}_\infty^{\infty-}[\mathbf{R}^d].$$

We suppose that F and F_n satisfy the following:

i) *F_n approximates F in $\mathbf{D}_\infty^{\infty-}[\mathbf{R}^d]$ with order γ $(\gamma > 0)$ in the sense that for every $1 < p < \infty$ and $\alpha > 0$,*

$$\|F_n - F\|_{p,\alpha} = O(n^{-\gamma}) \quad \text{as} \quad n \to \infty. \tag{2.1}$$

ii) *F is non-degenerate, i.e.,*

$$(\det \sigma_F)^{-1} \in \mathbf{L}^{\infty-}. \tag{2.2}$$

Then for every $\alpha > 0$, $\beta \geq 0$, $\delta > 0$ and $1 < p < \infty$ such that $\alpha > \beta + d/q + 1$, $1/p + 1/q = 1$, it holds that

$$\sup_{y \in \mathbf{R}^d} \|[(1 - \Delta)^{\beta/2}\phi_{n^{-\delta}}](F_n - y) - (1 - \Delta)^{\beta/2}\delta_y \circ F\|_{p,-\alpha} = O(n^{-\gamma \wedge \delta}) \tag{2.3}$$

as $n \to \infty$, where

$$\phi_\rho(x) = (2\pi\rho^2)^{-d/2}e^{-|x|^2/2\rho^2}\,, \quad x \in \mathbf{R}^d\,, \quad \rho > 0. \tag{2.4}$$

Remark. Obviously, $(1 - \Delta)^{\beta/2}\phi_{n^{-\delta}} \in \mathscr{S}(\mathbf{R}^d)$ and hence $[(1 - \Delta)^{\beta/2}\phi_{n^{-\delta}}](F_n - y) \in \mathbf{D}_\infty^{\infty-}$ for each n and $y \in \mathbf{R}^d$.

Corollary 2.1. *Let $\beta \geq 0$, $1 < q < \infty$ and $\alpha > \beta + d/q + 1$ and G_n, $G \in \mathbf{D}_q^\alpha$, $n = 1, 2, \dots$ such that*

$$\|G_n - G\|_{q,\alpha} = O\left(\frac{1}{n^\varepsilon}\right), \quad \textit{as } n \to \infty \tag{2.5}$$

for some $\varepsilon > 0$. *Then it holds that*

(2.6) $$\sup_{y \in \mathbf{R}^d} |(1 - \Delta)_y^{\beta/2} E[G_n \cdot \phi_{n^{-\delta}}(F_n - y)] - (1 - \Delta)_y^{\beta/2} E[G \cdot \delta_y(F)]| = O(n^{-\gamma \wedge \delta \wedge \varepsilon}) \quad \text{as } n \to \infty.$$

In particular, setting

$$u_n(y) = E[G_n \cdot \phi_{n^{-\delta}}(F_n - y)] \quad \textit{and} \quad u(y) = E[G \cdot \delta_y(F)],$$

we have

(2.7) $$\|u_n - u\|_\beta = O(n^{-\gamma \wedge \delta \wedge \varepsilon}) \qquad \textit{as } n \to \infty.$$

We remark that we are not assuming that F_n is non-degenerate so that we do not even know that $\delta_y \circ F_n = \delta_0(F_n - y)$ can be defined. A key point in the proof of Theorem 2.1 is to use an idea of Léandre ([Lé87]) which we formulate in the following:

Lemma 2.1. *Let* H_n, $H \in \mathbf{D}_1^{\infty-}$ *satisfy the following*:

(i) *There exists* $\gamma > 0$ *such that for any* $1 < p < \infty$,

(2.8) $$\|H_n - H\|_{p,1} = O(n^{-\gamma}), \qquad \text{as } n \to \infty.$$

(ii) $(\det \sigma_H)^{-1} \in \mathbf{L}^{\infty-}$.

(iii) *For any* $1 \leq p < \infty$, *there is* $v(p) > 0$ *such that* $\|(\det \sigma_{H_n})^{-1}\|_p = O(n^{v(p)})$ *as* $n \to \infty$.

Then, for any $1 \leq p < \infty$, *we have*

(2.9) $$\sup_n \|(\det \sigma_{H_n})^{-1}\|_p < \infty.$$

Proof. In the following, $C_1(p)$, $C_2(p)$, ... are positive constants depending only on $1 < p < \infty$. Writing $\det \sigma_{H_n} = \tau_n$ and $\det \sigma_H = \tau$, for simplicity, we have for a given $1 < p < \infty$,

$$E(|\tau_n|^{-p}) = E\left(|\tau_n|^{-p}: \left|\frac{\tau_n}{\tau} - 1\right| \leq \frac{1}{2}\right) + E\left(|\tau_n|^{-p}: \left|\frac{\tau_n}{\tau} - 1\right| > \frac{1}{2}\right) := I_1 + I_2.$$

If $\left|\frac{\tau_n}{\tau} - 1\right| \leq \frac{1}{2}$, then $\tau_n \geq \frac{1}{2}\tau$ and hence $\tau_n^{-1} \leq 2\tau^{-1}$. Consequently, by the assumption (ii),

$$\begin{aligned} I_1 &\leq 2^p E\left(|\tau|^{-p}: \left|\frac{\tau_n}{\tau} - 1\right| \leq \frac{1}{2}\right) \\ &\leq 2^p E|\tau|^{-p} := C_1(p) < \infty. \end{aligned}$$

As for I_2, we can estimate by the Schwartz inequality

$$I_2 \leq E(|\tau_n|^{-2p})^{1/2} P\left(\left|\frac{\tau_n - \tau}{\tau}\right| > \frac{1}{2}\right)^{1/2}.$$

By the assumption (iii),

$$E(|\tau_n|^{-2p}) \le C_2(p)\cdot n^{\nu(2p)\cdot 2p}$$

and by (2.8) it is easy to deduce that

$$E(|\tau_n - \tau|^{2k}) \le C_3(k)\cdot n^{-2k\gamma}$$

for every $k > 1$. Combining this with (ii),

$$E\left(\left|\frac{\tau_n - \tau}{\tau}\right|^{2k}\right) \le C_4(k)\cdot n^{-2k\gamma}$$

and hence

$$P\left(\left|\frac{\tau_n - \tau}{\tau}\right| > \frac{1}{2}\right) \le 2^{2k} E\left(\left|\frac{\tau_n - \tau}{\tau}\right|^{2k}\right)$$
$$\le C_5(k)\cdot n^{-2k\gamma}.$$

Consequently

$$I_2 \le \sqrt{C_2(p)\cdot C_5(k)\cdot n^{\nu(2p)\cdot 2p - 2k\gamma}}.$$

By taking k large enough so that $\nu(2p)\cdot 2p < 2k\gamma$, we can conclude (2.9).

Remark. In the above proof, the assumption (i) can be replaced by the following weaker assumption:

(i)′ For every $1 < p < \infty$, there exists a $\mu(p) > 0$ such that $\mu(p)\uparrow\infty$ as $p\uparrow\infty$ and

$$E(|H_n - H|^p + |DH_n - DH|^p) = O(n^{-\mu(p)}) \qquad \text{as } n \to \infty.$$

Proof of theorem 2.1. We take a product Wiener space ($\overline{W} = W \times W'$, $\overline{P} = P \times P'$), where (W', P') is a classical d-dimensional standard Wiener space and consider on $(\overline{W}, \overline{P})$ the following Wiener functionals

$$H_n(\overline{w}) = F_n(w) + \frac{1}{n^\delta} w'(1), \qquad H(\overline{w}) = F(w)$$

for $\overline{w} = (w, w') \in W \times W'$. Then we can apply the above lemma 2.1 to obtain

$$\sup_n \|(\det \sigma_{H_n})^{-1}\|_p < \infty.$$

Set further $K_{n,t}(\overline{w}) = (1 - t)F(w) + tH_n(\overline{w})$ for $t \in [0, 1]$. Then,

$$K_{n,t}(\overline{w}) - F(w) = t(H_n(\overline{w}) - F(w)) \quad \text{and} \quad \|(\det \sigma_{K_{n,t}})^{-1}\|_p = O(n^{2\delta d} t^{-2d}).$$

Applying Lemma 2.1 by fixing n, we deduce that, for each $p > 1$, $\mu > 0$ exists such that

$$\sup_{t\in[0,1]} \|(\det \sigma_{K_{n,t}})^{-1}\|_p = O(n^\mu).$$

Then applying Lemma 2.1 again, we can conclude that, for each $p > 1$,

$$\sup_{t,n} \|(\det \sigma_{K_{n,t}})^{-1}\|_p < \infty.$$

Set $T = (1 - \Delta)^{\beta/2}\delta_y$. Then

$$T \circ H_n - T \circ F = \int_0^1 \partial T \circ K_{n,t} \cdot (H_n - F)dt$$

and, if $\alpha > (\beta + 1) + d/q$, $1/p + 1/q = 1$, then by (1.1) we can choose $\bar{p} > p$ such that

$$\sup_{t,n} \|\partial T \circ K_{n,t}\|_{\bar{p},-\alpha} < \infty.$$

Since $\|H_n - F\|_{\bar{q},\alpha} = O(n^{-\gamma\wedge\delta})$ for $\bar{q} > 1$ such that $1/\bar{p} + 1/\bar{q} \leq 1/p$, we deduce that

$$\begin{aligned}\|T \circ H_n - T \circ F\|_{p,-\alpha} &\leq \int_0^1 \|\partial T \circ K_{n,t} \cdot (H_n - F)\|_{p,-\alpha} dt \\ &\leq \text{const} \int_0^1 \|\partial T \circ K_{n,t}\|_{\bar{p},-\alpha} \|H_n - F\|_{\bar{q},\alpha} dt \\ &= O(n^{-\gamma\wedge\delta}).\end{aligned}$$

From this, we can conclude (2.3) by noting

$$\int_{W'} [(1 - \Delta)^{\beta/2}\delta_y](H_n(w, w'))P'(dw') = [(1 - \Delta)^{\beta/2}\phi_{n^{-\delta}}](F_n(w) - y)$$

and a general fact that the map

$$\Phi(\bar{w}) = \Phi(w, w') \to \hat{\Phi}(w) = \int_{W'} \Phi(w, w')P'(dw')$$

is a contraction from $\mathbf{D}^p_{-\alpha}(\overline{W})$ to $\mathbf{D}^p_{-\alpha}(W)$ for every $1 < p < \infty$ and $\alpha \in \mathbf{R}$. This follows at once from the following relation which can be easily verified by the Wiener chaos expansion:

$$[(1 - \bar{L})^{-\alpha}\Phi]^\wedge = (1 - L)^{-\alpha}\hat{\Phi},$$

$\bar{L}$ and L being the Ornstein-Uhlenbeck operator on $\overline{W}$ and W, respectively.

Corollary 2.2. *If F_n in Theorem 2.1 is uniformly non-degenerate, i.e.*

$$\sup_n \|(\det \sigma_{F_n})^{-1}\|_p < \infty,$$

then, $[(1 - \Delta)^{\beta/2}\phi_{n^{-\delta}}](F_n - y)$ and $O(n^{-\gamma\wedge\delta})$ in (2.3), and $\phi_{n^{-\delta}}(F_n - y)$ and $O(n^{-\gamma\wedge\delta\wedge\varepsilon})$ *in (2.6), may be replaced by $(1 - \Delta)^{\beta/2}\delta_y \circ F_n$ and $O(n^{-\gamma})$, and $\delta_y \circ F_n$ and $O(n^{-\gamma\wedge\varepsilon})$, respectively.*

Indeed, if $T = (1 - \Delta)^{\beta/2}\delta_y$, we can deduce by the same proof that

$$\|T(H_n) - T(F_n)\|_{p,-\alpha} = O(n^{-\delta}).$$

Also we may take δ as large as we want.

3. Approximate heat kernels by time discretization scheme

We consider a typical example of applications of the above theorem.

Let $W = W_0^r$ be the classical r-dimensional Wiener space:

$$W_0^r = \{w \in C([0, T] \to \mathbf{R}^r), w(0) = 0\}$$

and P be the standard Wiener measure on W_0^r. Then $w(t) = (w^1(t), \ldots, w^r(t))$ for $w \in W$ is a realization of r-dimensional Brownian motion on W. Also we write w_t^i for $w^i(t)$.

Fix a $T > 0$. Consider the following stochastic differential equation:

$$X_t = x + \sum_{j=0}^{r} \int_0^t b_j(s, X_s) dw_s^j, \; t \in [0, T], \; x \in \mathbf{R}^d, \tag{3.1}$$

where b_j, $j = 0, 1, \ldots, r$ are given smooth functions from $[0, T] \times \mathbf{R}^d$ to $\mathbf{R}^d$ with bounded derivatives, and we use the convention $dw_s^0 = ds$ to simplify notation. The unique solution is denoted by $X_t = X(t, x)$. Then $X(t, x) \in \mathbf{D}_\infty^{\infty -}[\mathbf{R}^d]$ for every $t \geq 0$ and $x \in \mathbf{R}^d$.

To describe the time discretization schemes, we introduce the following operators on functions f: $[0, T] \times \mathbf{R}^d \to \mathbf{R}$

$$L^j f(s, x) = \sum_{k=1}^{d} b_j^k(s, x) \frac{\partial f}{\partial x^k}(s, x), \quad j = 1, \ldots, r, \tag{3.2}$$

(3.3)

$$L^0 f(s, x) = \frac{\partial f}{\partial s}(s, x) + \sum_{k=1}^{d} b_0^k(s, x) \frac{\partial f}{\partial x^k}(s, x) + \frac{1}{2} \sum_{k,l=1}^{d} \sum_{j=1}^{r} b_j^k(s, x) b_j^l(s, x) \frac{\partial^2 f}{\partial x^k \partial x^l}(s, x),$$

where b_j^k is the k-th component of the vector b_j $(k = 1, \ldots, d)$.

Consider a multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_l)$ with $\alpha_i \in \{0, 1, \ldots, r\}$ and write $l = l(\alpha)$. If $l(\alpha) \geq 2$, we set $-\alpha = (\alpha_2, \ldots, \alpha_l)$ and $\alpha- = (\alpha_1, \ldots, \alpha_{l-1})$. Given α, we define an $\mathbf{R}^d$-valued function $f_\alpha(s, x) = (f_\alpha^k(s, x))$ on $[0, T] \times \mathbf{R}^d$ recursively by

$$\begin{aligned} f_\alpha^k(s, x) &= b_{\alpha_1}^k(s, x), && \text{if } l(\alpha) = 1 \text{ and } \alpha = (\alpha_1), \\ &= (L^{\alpha_1} f_{-\alpha}^k)(s, x), && \text{if } l(\alpha) \geq 2 \text{ and } \alpha = (\alpha_1, -\alpha) \end{aligned}$$

so that

$$f_\alpha^k(s, x) = L^{\alpha_1} L^{\alpha_2} \ldots L^{\alpha_{l-1}}(b_{\alpha_l}^k)(s, x) \tag{3.4}$$

if $\alpha = (\alpha_1, \ldots, \alpha_l)$. Also we define, for α and $0 \leq s < t \leq T$, a Wiener functional $I_{\alpha,s,t}$ recursively by

$$\begin{aligned} I_{\alpha,s,t} &= w_t^{\alpha_1} - w_s^{\alpha_1}, && \text{if } l(\alpha) = 1 \text{ and } \alpha = (\alpha_1), \\ &= \int_s^t I_{\alpha-,s,u} dw_u^{\alpha_l}, && \text{if } l(\alpha) \geq 2 \text{ and } \alpha = (\alpha-, \alpha_l) \end{aligned}$$

so that

$$(3.5)\qquad I_{\alpha,s,t} = \int\cdots\int_{s\le u_1<\cdots<u_l\le t} dw_{u_1}^{\alpha_1} dw_{u_2}^{\alpha_2}\cdots dw_{u_l}^{\alpha_l} \qquad \text{if } \alpha = (\alpha_1,\cdots,\alpha_l).$$

Then by the recursive application of Itô's formula, we obtain the following Itô-Taylor expansion for the solution X_t to (3.1) (cf. [Kl-Pl92]) for every $0 \le s < t \le T$ and a non-empty set Γ of multi-indices with the property that if $\alpha \in \Gamma$, then $-\alpha \in \Gamma$,

$$(3.6)\qquad X_t = X_s + \sum_{\alpha\in\Gamma} f_\alpha(s, X_s) I_{\alpha,s,t} + \sum_{\alpha\in\mathscr{B}_\Gamma} I_\alpha(f_\alpha(\cdot, X.))_{s,t},$$

where

$$(3.7)\qquad \mathscr{B}_\Gamma = \{\alpha; \alpha\notin\Gamma \quad\text{and}\quad -\alpha\in\Gamma\}$$

and

$$(3.8)\qquad I_\alpha(f_\alpha(\cdot, X.))_{s,t} = \int\cdots\int_{s\le u_1<\cdots<u_l\le t} f_\alpha(u_1, X_{u_1}) dw_{u_1}^{\alpha_1} dw_{u_2}^{\alpha_2}\dots dw_{u_l}^{\alpha_l}$$

if $\alpha = (\alpha_1, \dots, \alpha_l)$.

In the sequel, we consider the following case of class Γ of multi-indices exclusively: For an integer or half-integer $\gamma > 0$,

$$(3.9)\qquad \Gamma = \Gamma_\gamma := \{\alpha; l(\alpha) + n(\alpha) \le 2\gamma \quad\text{or}\quad l(\alpha) = n(\alpha) = \gamma + 1/2\},$$

where $n(\alpha) = \#\{k; \alpha_k = 0\}$ for $\alpha = (\alpha_1, \dots, \alpha_l)$.

Consider a partition π of the interval $[0, T]$, $\pi: 0 = t_0 < t_1 < \cdots < t_n = T$ and put $|\pi| = \sup_i(t_{i+1} - t_i)$, the step of the partition. A strong Itô-Taylor approximation scheme (X_t^π) of order γ for the solution X_t to (3.1) is defined by

$$(3.10)\qquad \begin{aligned} &X_0^\pi = x \\ &X_t = X_{t_m}^\pi + \sum_{\alpha\in\Gamma} f_\alpha(t_m, X_{t_m}^\pi) I_{\alpha,t_m,t} \\ &\qquad \text{for } t\in[t_m, t_{m+1}], \qquad m = 0, 1, \dots, n-1, \end{aligned}$$

where $\Gamma = \Gamma_\gamma$. This recursive formula lends itself to explicit computations (the multiple integrals can be even handled by a computer). Examples are the case $\gamma = \frac{1}{2}$: Euler-Maruyama scheme, with

$$\Gamma = \{\alpha = (\alpha), \alpha = 0, 1, \dots, r\}$$

and the case $\gamma = 1$: Milstein schemes, with

$$\Gamma = \{\alpha = (\alpha), \alpha = 0, 1, \dots, r; \quad\text{or}\quad \alpha = (\alpha_1, \alpha_2), 1 \le \alpha_1, \alpha_2 \le r\}.$$

The following estimate was proved in [Kl-Pl92] when $p = 2$ but this can be obtained by the same proof for general p if we use Lemma 4.1 given in the next section:

$$(3.11)\qquad \left\|\left(\sup_{0\le t\le T} |X_t - X_t^\pi|\right)\right\|_p = O(|\pi|^\gamma)$$

for any $1 < p < \infty$. Let $x \in \mathbf{R}^d$ be fixed and set

$$F(w) = X(T, x); \qquad F_\pi = X^\pi(T, x). \tag{3.12}$$

Then F, $F_\pi \in \mathbf{D}_\infty^{\infty-}[\mathbf{R}^d]$. (3.11) implies, in particular, that

$$\|F_\pi - F\|_p = O(|\pi|^\gamma) \qquad \text{as } |\pi| \to 0$$

and in the next section we will improve this result to obtain

Theorem 3.1. *Suppose that the coefficients b_j, $j = 0, 1, \ldots, r$, are in C^∞ with bounded derivatives and further $f_\alpha^k(s, x)$ for every multi-index α and $k = 1, \ldots, d$ have bounded derivatives. Then, for every $p > 1$ and $\beta > 0$,*

$$\|F_\pi - F\|_{p,\beta} = O(|\pi|^\gamma) \qquad \text{as } |\pi| \to 0. \tag{3.13}$$

Let π_n be the equal partition of $[0, T]$: $\pi_n : 0 = t_0 < t_1 < \ldots < t_n = T$ where $t_k = \frac{k}{n}T$. We write $F_{\pi_n} = F_n$. Then F_n approximates F in $\mathbf{D}_\infty^{\infty-}$ with order γ. Now we suppose that F is non-degenerate:

$$(\det \sigma_F)^{-1} \in L^{\infty-}. \tag{3.14}$$

A sufficient condition has been studied, just as the beginning of the Malliavin calculus, by Malliavin [Ma78] and then completely by Kusuoka-Stroock [Ku-St85]. The non-degeneracy of the diffusion coefficients $a^{kl}(x) = \sum_{\alpha=1}^r b_\alpha^k(x) b_\alpha^l(x)$ at x is sufficient: However, a much weaker sufficient condition can be given as a Hörmander type condition at x for vector fields b_α involving the Poisson brackets. We do not intend to state explicitly here (c.f. [Ku-St85]).

Now F_n and F satisfy both conditions (i) and (ii) of the above theorem 2.1 and hence we can conclude that (2.3) holds. Thus we could obtain an approximation scheme for the Donsker's delta function $\delta_y(X(t, x))$ in terms of the Itô-Taylor scheme (3.10) for the solution $X(t, x)$ and thereby for the probability density of $X(t, x)$ and the conditional expectation given $X(t, x)$. In particular, we can summarize Corollary 2.1 in the following

Theorem 3.2. *Let $F_n = X^{\pi_n}(T, x)$ and $F = X(T, x)$ for fixed $x \in \mathbf{R}^d$ and $T > 0$ and assume that equation* (3.1) *satisfy the non-degeneracy condition* (3.14). *For every $\beta > 0$ and $1 < q < \infty$, if G_n, $G \in \mathbf{D}_q^\alpha$ for $\alpha > d/q + \beta + 1$ and*

$$\|G_n - G\|_{q,\alpha} = O(n^{-\varepsilon}) \qquad \textit{for some } \varepsilon > 0 \qquad \text{as } n \to \infty,$$

then setting

$$u_n(y) = E[G_n \cdot \phi_{n^{-\delta}}(F_n - y)] = E[G_n \cdot \phi_{n^{-\delta}}(X^{\pi_n}(T, x) - y)]$$

and

$$u(y) = E[G \cdot \delta_y(F)] = E[G \cdot \delta_y(X(T, x))],$$

we have

(3.15) $$\|u_n - u\|_\beta = O(n^{-\gamma\wedge\delta\wedge\varepsilon}) \qquad \textit{as } n \to \infty.$$

Note that

$$u(y) = P(T, x, y)E[G|F = y]$$

where $P(T, x, y) = E[\delta_y(X(T, x))]$ is the probability density of the solution $X(T, x)$. We know that $P(T, x, y)$ coincides with the heat kernel for the heat operator L^0.

If the non-degeneracy condition (2.10) is satisfied for F_n, i.e.,

$$\sup_n \|(\det \sigma_{F_n})^{-1}\|_p < \infty,$$

then, by Cor 3.2, u_n in (3.15) may be replaced by

$$u_n(y) = E[G_n \cdot \delta_y(F_n)] = P_n(T, x, y)E[G_n|F_n = y],$$

where $P_n(T, x, y) = E[\delta_y(X^{\pi_n}(T, x))]$ and it holds that

$$\|u_n - u\|_\beta = O(n^{-\gamma\wedge\varepsilon}) \qquad \text{as } n \to \infty.$$

In particular, for any $\beta > 0$,

$$\|P_n(T, x, \cdot) - P(T, x, \cdot)\|_\beta = O(n^{-\gamma}) \qquad \text{as } n \to \infty.$$

For example, suppose that the coefficients of SDE (3.1) satisfies the uniform ellipticity condition:

$$\inf_{(s,x)\in[0,T]\times\mathbf{R}^d} \det[a(s, x)] > 0,$$

where $a^{kl}(s, x) = \sum_{j=1}^r b_j^k(s, x)b_j^l(s, x)$, $k, l = 1, \ldots, d$ and consider the Euler-Maruyama scheme:

$$X_{t_{k+1}}^\pi = X_{t_k}^\pi + \sum_{j=1}^r b_j(t_k, X_{t_k}^\pi)[w_{t_{k+1}}^j - w_{t_k}^j] + b_0(t_k, X_{t_k}^\pi)(t_{k+1} - t_k).$$

Then, for any $\beta > 0$,

$$\|P^\pi(T, x, \cdot) - P(T, x, \cdot)\|_\beta = O(|\pi|^{1/2})$$

as $|\pi| = \max_k(t_{k+1} - t_k) \to 0$ and $P^\pi(T, x, y) = E[\delta_y(X_T^\pi)]$ is given explicitly by

$$P^\pi(T, x, y) = \int_{\mathbf{R}^d} \cdots \int_{\mathbf{R}^d} \prod_{k=0}^{n-1} \Lambda(t_k, x_k; t_{k+1}, x_{k+1})dx_1 \ldots dx_{n-1},$$

where $\pi: 0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$, $x_0 = x$, $x_n = y$ and, in general

$$\begin{aligned}
&\Lambda(\sigma, \xi; \tau, \eta) \\
&\quad = [(2\pi(\tau - \sigma))^d \det[a(\sigma, \xi)]]^{-1/2} \\
&\qquad \times \exp[-\tfrac{1}{2} < a^{-1}(\sigma, \xi)(\eta - \xi - b_0(\sigma, \xi)(\tau - \sigma)), \eta - \xi - b_0(\sigma, \xi)(\tau - \sigma) >], \\
&\qquad 0 \le \sigma < \tau \le T, \qquad \xi, \eta \in \mathbf{R}^d.
\end{aligned}$$

4. Convergence in $\mathbf{D}_\alpha^p$ of the Itô-Taylor schemes

In this section, we prove Theorem 3.1. For this we need some key estimates as formulated in Lemma 4.1 below which is a natural extension of Lemma 5.7.3 and Lemma 10.8.1 of [Kl-Pl92]. For a given finite $T > 0$, let

$$\Delta = \{(s_1, s_2, \ldots, s_l) \in \mathbf{R}^l | 0 \le s_1 < s_2, \ldots < s_l \le T\}$$

and $L^2(\Delta)$ be the usual L^2-space of real square-integrable functions on Δ. Let $f(s_1, \ldots, s_l)$ be an $L^2(\Delta)$-valued Wiener functional on W such that $f(s_1, \ldots, s_l)$ is $\mathscr{B}_{s_1}$-measurable for each fixed $s_1 < \ldots < s_l$. Define for a multi-index $\alpha = (\alpha_1, \ldots, \alpha_l)$ and $0 \le u \le v \le T$,

$$I_\alpha(f)_{u,v} = \int_{u < s_1 < \cdots < s_l < v} f(s_1, \ldots, s_l) dw_{s_1}^{\alpha_1} \ldots dw_{s_l}^{\alpha_l}$$

by iterated Itô's stochastic integrals. In particular, if $f \equiv 1$, then $I_\alpha(f)_{u,v}$ coincides with $I_{\alpha,u,v}$ defined by (3.5) and, if $f(s_1, \ldots, s_l) = f_\alpha(s_1, X_{s_1})$, then $I_\alpha(f)_{u,v}$ coincides with $I_\alpha(f_\alpha(\cdot, X.))_{u,v}$ defined by (3.8). Set, for $0 \le u < s \le T$,

$$\begin{aligned} \|f\|_u(s) &= |f(s)| && \text{if } l = 1 \\ &= \sup_{u < s_1 < \cdots < s_{l-1} < s} |f(s_1, \ldots, s_{l-1}, s)| && \text{if } l > 1. \end{aligned}$$

Lemma 4.1. (1) *For* $p \ge 1$ *and* $0 \le u < v \le T$,

$$E\left[\sup_{u \le t \le v} |I_\alpha(f)_{u,t}|^{2p}\right] \le C(u - v)^{p[l(\alpha) + n(\alpha)] - 1} \int_u^v E[\|f\|_u(t)^{2p}] dt \tag{4.1}$$

(2) *Let* $\pi : 0 = t_0 < t_1 < \cdots < t_n = T$ *be a partition of* $[0, T]$. Set $|\pi| = \sup_i (t_{i+1} - t_i)$ *and* $m(s) = m$ *if* $t_m \le s < t_{m+1}$. *Consider the following expectation for each* $0 \le t \le T$:

$$F_t^\alpha = E\left(\sup_{0 \le s \le t} \left| \sum_{m=0}^{m(s)-1} I_\alpha(f)_{t_m, t_{m+1}} + I_\alpha(f)_{t_{m(s)}, s} \right|^{2p}\right).$$

Then, for $p \ge 1$ *and* $0 \le t \le T$,

$$\begin{aligned} F_t^\alpha &\le C|\pi|^{2p[l(\alpha) - 1]} \int_0^t E(\|f\|_{t_{m(s)}}(s)^{2p}) ds && \textit{if } l(\alpha) = n(\alpha), \\ &\le C|\pi|^{p[l(\alpha) + n(\alpha) - 1]} \int_0^t E(\|f\|_{t_{m(s)}}(s)^{2p}) ds && \textit{if } l(\alpha) \ne n(\alpha). \end{aligned} \tag{4.2}$$

Here, C are positive constants depending on T, p and α *which may vary from lines to lines.*

Proof: (1) can be proved by induction on the length $l(\alpha)$ of α: If $l(\alpha) > 1$,

$$I_\alpha(f)_{u,t} = \int_u^t I_{\alpha-}(f^s)_{u,s} dw_s^{\alpha_l}$$

where $f^s(s_1, \dots, s_{l-1}) = f(s_1, \dots, s_{l-1}, s)$. If $\alpha_l = 0$, then

$$E\Big[\sup_{u\le t\le v} |I_\alpha(f)_{u,t}|^{2p}\Big] \le E\left\{\left[\int_u^v |I_{\alpha-}(f^s)_{u,s}| ds\right]^{2p}\right\}$$
$$\le (v-u)^{2p-1}\int_u^v E[|I_{\alpha-}(f^s)_{u,s}|^{2p}]ds$$

by the Hölder inequality. If $\alpha_l \ne 0$, then

$$E\Big[\sup_{u\le t\le v} |I_\alpha(f)_{u,t}|^{2p}\Big] \le C\cdot E\left\{\left[\int_u^v |I_{\alpha-}(f^s)_{u,s}|^2 ds\right]^{p}\right\}$$
$$\le C(v-u)^{p-1}\int_u^v E[|I_{\alpha-}(f^s)_{u,s}|^{2p}]ds$$

by a standard martingale inequality of the Burkholder-Davis-Gundy type for stochastic integrals (cf. [IW89], p. 110) and the Hölder inequality. The inequality (4.1) for the case $l(\alpha) = 1$ can be obtained by the same estimates. Then we can conclude the proof by induction if we note the following facts: $l(\alpha-) = l(\alpha) - 1$, $n(\alpha-) = n(\alpha)$ or $n(\alpha) - 1$ according as $\alpha_l \ne 0$ or $\alpha_l = 0$ and $\|f^s\|_u(t) \le \|f\|_u(s)$ if $t \le s$.

Next, we prove (2). We note that

$$\Xi(t) := \sum_{m=0}^{m(t)-1} I_\alpha(f)_{t_m, t_{m+1}} + I_\alpha(f)_{t_{m(t)}, t} = \int_0^t I_{\alpha-}(f^s)_{t_{m(s)}, s} dw_s^{\alpha_l}.$$

If $n(\alpha) = l(\alpha)$, then

$$E\Big[\sup_{0\le s\le t} |\Xi(s)|^{2p}\Big] \le E\left\{\left[\int_0^t |I_{\alpha-}(f^s)_{t_{m(s)},s}| ds\right]^{2p}\right\}$$
$$\le C\cdot\int_0^t E\{|I_{\alpha-}(f^s)_{t_{m(s)},s}|^{2p}\}ds$$
$$\le C\cdot\sum_{m=0}^{m(t)}\int_{t_m}^{t_{m+1}\wedge t} E\{|I_{\alpha-}(f^s)_{t_m,s}|^{2p}\}ds.$$

By the estimate (4.1), this is dominated by

$$C|\pi|^{2pl(\alpha-)-1}\cdot\sum_{m=0}^{m(t)}\int_{t_m}^{t_{m+1}\wedge t} ds\int_{t_m}^{s} E[\|f^s\|_{t_m}(\tau)^{2p}]d\tau \le C|\pi|^{2pl(\alpha-)}\int_0^t E[\|f\|_{t_{m(s)}}(s)^{2p}]ds.$$

Since $l(\alpha-) = l(\alpha) - 1$, (4.2) is obtained in this case.

If $n(\alpha) \ne l(\alpha)$, and $\alpha_l \ne 0$, then by the Burkholder-Davis-Gundy inequality applied to stochastic integral $\Xi(s)$, we have

$$E\Big[\sup_{0\le s\le t} |\Xi(s)|^{2p}\Big] \le C\cdot E\left\{\left[\int_0^t |I_{\alpha-}(f^s)_{t_{m(s)},s}|^2 ds\right]^{p}\right\}$$
$$\le C\cdot\int_0^t E\{|I_{\alpha-}(f^s)_{t_{m(s)},s}|^{2p}\}ds.$$

By the same estimate as above using (4.1), this is further dominated by

$$C|\pi|^{p[l(\alpha-)+n(\alpha-)]}\int_0^t E[\|f\|_{t_{m(s)}}(s)^{2p}]ds.$$

Since $\alpha_l \neq 0$, we have $l(\alpha-)+n(\alpha-)=l(\alpha)+n(\alpha)-1$ and hence (4.2) is obtained.

Finally we consider the case $n(\alpha)\neq l(\alpha)$ and $\alpha_l=0$. We have

$$F_t^\alpha \le C\cdot E\left\{\sup_{0\le s\le t}\left|\sum_{m=0}^{m(s)-1} I_\alpha(f)_{t_m,t_{m+1}}\right|^{2p}\right\}+C\cdot E\left\{\sup_{0\le s\le t}|I_\alpha(f)_{t_{m(s)},s}|^{2p}\right\}$$

$$:= I_1 + I_2$$

and estimate these two terms separately. We first note that

$$I_1 = C\cdot E\left\{\sup_{0\le k\le m(t)-1}\left|\sum_{m=0}^{k} I_\alpha(f)_{t_m,t_{m+1}}\right|^{2p}\right\}$$

and $S_k=\sum_{m=0}^k I_\alpha(f)_{t_m,t_{m+1}}$ forms a discrete time martingale. Then we can apply the Burkholder-Davis-Gundy inequality for the discrete time martingale (cf. [Ga73]) to obtain that

$$I_1 \le C\cdot E\left\{\left[\sum_{m=0}^{m(t)-1}|I_\alpha(f)_{t_m,t_{m+1}}|^2\right]^p\right\}$$

$$= C\cdot E\left\{\left[\sum_{m=0}^{m(t)-1}\left|\int_{t_m}^{t_{m+1}} I_{\alpha-}(f^s)_{t_m,s}ds\right|^2\right]^p\right\}.$$

Since

$$\left|\int_{t_m}^{t_{m+1}} I_{\alpha-}(f^s)_{t_m,s}ds\right|^2 \le (t_{m+1}-t_m)\cdot\int_{t_m}^{t_{m+1}}|I_{\alpha-}(f^s)_{t_m,s}|^2ds,$$

this is further dominated by

$$C|\pi|^p\cdot E\left\{\left[\sum_{m=0}^{m(t)-1}\int_{t_m}^{t_{m+1}}|I_{\alpha-}(f^s)_{t_m,s}|^2ds\right]^p\right\}\le C|\pi|^p\cdot E\left\{\left[\int_0^t|I_{\alpha-}(f^s)_{t_{m(s)},s}|^2ds\right]^p\right\}$$

$$\le C|\pi|^p\cdot\int_0^t E\{|I_{\alpha-}(f^s)_{t_{m(s)},s}|^{2p}\}ds.$$

Then by the same estimate as above using (4.1), we deduce that the above is dominated by

$$C|\pi|^{p[l(\alpha-)+n(\alpha-)+1]}\int_0^t E(\|f\|_{t_{m(s)}}(s)^{2p})ds.$$

Since $l(\alpha-)+n(\alpha-)+1=l(\alpha)+n(\alpha)-1$, I_1 is now dominated as desired.

As for I_2, we have,

$$E\left[\sup_{0\le s\le t}|I_\alpha(f)_{t_{m(s)},s}|^{2p}\right]=E\left[\sup_{0\le s\le t}\left|\int_{m(s)}^{s}I_{\alpha-}(f^\tau)_{t_{m(s)},\tau}d\tau\right|^{2p}\right]$$
$$\le E\left\{\sup_{0\le s\le t}\left[(s-t_{m(s)})^{2p-1}\int_{m(s)}^{s}|I_{\alpha-}(f^\tau)_{t_{m(s)},\tau}|^{2p}d\tau\right]\right\}$$
$$\le|\pi|^{2p-1}E\left[\sum_{m=0}^{m(t)}\int_{t_m}^{t_{m+1}\wedge t}|I_{\alpha-}(f^s)_{t_m,s}|^{2p}ds\right].$$

By using (4.1), this can be dominated by

$$C|\pi|^{2p-1}\cdot|\pi|^{p[l(\alpha-)+n(\alpha-)]}\int_0^t E(\|f\|_{t_{m(s)}}(s)^{2p})ds$$
$$=C|\pi|^{p-1}\cdot|\pi|^{p[l(\alpha)+n(\alpha)-1]}\int_0^t E(\|f\|_{t_{m(s)}}(s)^{2p})ds.$$

Since $p\ge 1$, we obtained the desired estimate for I_2 and the proof of (4.2) is now complete.

If f is deterministic, i.e., $f\in L^2(\Delta)$, then $I_\alpha(f)_{u,v}\in\mathbf{D}_\infty^{\infty-}$ and we have, for each $s\in[0,T]$ and $k=1,\dots,r$,

(4.3)

$$D_s^kI_\alpha(f)_{u,v}=\sum_{i:\alpha_i=k}\int_{u<s_1<\cdots<\hat{s}_i<\cdots<s_l<v}f(s_1,\dots,s_l)|_{s_i=s}1_{[s_{i-1},s_{i+1}]}(s)dw_{s_1}^{\alpha_1}\dots d\hat{w}_{s_i}^{\alpha_i}\dots dw_{s_l}^{\alpha_l},$$
$$\text{if } l(\alpha)>1$$
$$=1_{[u,v]}(s)\cdot\delta_{k,\alpha}\cdot f(s),\qquad\text{if } l(\alpha)=1$$

where the symbol ^ means to discard the element which bears it. Here, generally, DF for $F\in\mathbf{D}_1^2$ denotes the gradient, i.e. H-derivative, of F and define $D_sF=(D_s^kF)_{k=1,\dots,r}$ by

$$\langle DF,h\rangle_H=\int_0^T\langle D_sF,\dot{h}_s\rangle_{\mathbf{R}^r}ds$$
$$=\sum_{k=1}^r\int_0^T D_s^kF\cdot\dot{h}_s^kds,\qquad\text{for all } h\in H$$

where $H\subset W_0^r$ is the Cameron-Martin subspace:

$$H=\left\{h\in W_0^r;\,h_t=(h_t^k),\,t\to h_t^k\quad\text{is absolutely continuous}\right.$$
$$\left.\text{and } |h|_H^2=\sum_{k=1}^r\int_0^T|\dot{h}_t^k|^2dt<\infty\right\}.$$

Set, for fixed $1\le i\le l$ and $s\in[0,T]$,

(4.4) $$\alpha_{\setminus i}=(\alpha_1,\dots,\hat{\alpha}_i,\dots,\alpha_l)$$

and

(4.5) $$f_{\setminus i}[s](s_1, \ldots, s_{l-1}) = f(s_1, \ldots s_{i-1}, s, s_i, \ldots, s_{l-1}) 1_{[s_{i-1}, s_i]}(s) \qquad \text{if } l > 1.$$

Then we may write (4.3) in the following form:

(4.6) $$\begin{aligned} D_s^k I_\alpha(f)_{u,v} &= \sum_{i:\alpha_i = k} I_{\alpha\setminus i}(f_{\setminus i}[s])_{u,v}, \qquad &&\text{if } l(=l(\alpha)) > 1 \\ &= f(s) \cdot 1_{[u,v]}(s) \cdot \delta_{k,\alpha}, &&\text{if } l = 1. \end{aligned}$$

Note that the right-hand side of (4.6) is well-defined for almost all s even if f is not deterministic and so we define generally, for each $s \in [0, T]$ and $k = 1, \ldots, r$,

(4.7) $$\begin{aligned} \tilde{D}_s^k I_\alpha(f)_{u,v} &= \sum_{i:\alpha_i = k} I_{\alpha\setminus i}(f_{\setminus i}[s])_{u,v}, \qquad &&\text{if } l(=l(\alpha)) > 1 \\ &= f(s) \cdot 1_{[u,v]}(s) \cdot \delta_{k,\alpha}, &&\text{if } l = 1. \end{aligned}$$

Then, we have generally

(4.8) $$D_s^k I_\alpha(f)_{u,v} = \tilde{D}_s^k I_\alpha(f)_{u,v} + I_\alpha(e_s)_{u,v},$$

where $e_s(s_1, \ldots, s_l) = D_s^k[f(s_1, \ldots, s_l)]$.

Consider the solution $X_t = X(t, x)$ of the SDE (3.1) and suppose that coefficients satisfy the same assumptions as in Theorem 3.1.

Lemma 4.2. *We have*

(4.9) $$\sup_{s \in [0,T]} \sup_{1 \le k \le r} \int_0^T E\left[\sup_{0 \le u \le t} |D_s^k X_u|^p\right] dt < \infty$$

for all $p > 1$.

Proof. Taking D_s^k in the both sides of equation (3.1) we obtain, for $0 \le s \le t$,

$$D_s^k X_t = \sum_{j=0}^r \int_0^t \nabla b_j(u, X_u) D_s^k X_u dw_u^j + b_k(s, X_s).$$

So there are two constants C_1 and C_2 such that

$$E\left[\sup_{0 \le u \le t} |D_s^k X_u|^p\right] \le C_1 \sum_{j=0}^r \int_0^t E\left[\sup_{0 \le u \le v} |D_s^k X_u|^p\right] dv + C_2.$$

This shows the boundedness of (4.9) by the Gronwall lemma.

Now we prove Theorem 3.1. We may assume $p \ge 2$ and hence write it as $2p$ for $p \ge 1$. We give a proof in the case of $\beta = 1$: The proof for $\beta = 2, 3, \ldots$ can be given in a similar way. We note that the estimate

$$\|F_\pi - F\|_{2p,1} = O(|\pi|^\gamma) \qquad \text{as } |\pi| \to 0$$

is a consequence of a more sharp estimate

$$(4.10)\qquad \sup_{0\le s\le T} E\left[\sup_{0\le u\le T} |D_s^k X_u^\pi - D_s^k X_u|^{2p}\right] = O(|\pi|^{2p\gamma}) \qquad \text{as } |\pi|\to 0.$$

So we prove (4.10). Taking the derivative in both sides of equations (3.10) and (3.6), we have for $t\in[t_m, t_{m+1}]$,

$$(4.11)\quad D_s^k X_t^\pi = D_s^k X_{t_m}^\pi + \sum_{\alpha\in\Gamma} \nabla f_\alpha(t_m, X_{t_m}^\pi) D_s^k X_{t_m}^\pi I_{\alpha,t_m,t} + \sum_{\alpha\in\Gamma} f_\alpha(t_m, X_{t_m}^\pi) D_s^k I_{\alpha,t_m,t}$$

and

$$(4.12)\qquad \begin{aligned} D_s^k X_t = D_s^k X_{t_m} &+ \sum_{\alpha\in\Gamma} \nabla f_\alpha(t_m, X_{t_m}) D_s^k X_{t_m} I_{\alpha,t_m,t} \\ &+ \sum_{\alpha\in\Gamma} f_\alpha(t_m, X_{t_m}) D_s^k I_{\alpha,t_m,t} \\ &+ \sum_{\alpha\in\mathscr{B}_\Gamma} I_\alpha[\nabla f_\alpha(\cdot, X_\cdot) D_s^k X_\cdot]_{t_m,t} \\ &+ \sum_{\alpha\in\mathscr{B}_\Gamma} \tilde{D}_s^k I_\alpha[f_\alpha(\cdot, X_\cdot)]_{t_m,t} \end{aligned}$$

where $\nabla f(s,x) = \left(\dfrac{\partial f}{\partial x_j}(s,x)\right)_{1\le j\le d}$.

Now substract equation (4.11) from (4.12):

$$(4.13)\quad \begin{aligned} D_s^k X_t - D_s^k X_t^\pi &= D_s^k X_{t_m} - D_s^k X_{t_m}^\pi \\ &\quad + \sum_{\alpha\in\Gamma} [\nabla f_\alpha(t_m, X_{t_m}) D_s^k X_{t_m} I_{\alpha,t_m,t} - \nabla f_\alpha(t_m, X_{t_m}^\pi) D_s^k X_{t_m}^\pi I_{\alpha,t_m,t}] \\ &\quad + \sum_{\alpha\in\Gamma} [f_\alpha(t_m, X_{t_m}) - f_\alpha(t_m, X_{t_m}^\pi)] D_s^k I_{\alpha,t_m,t} \\ &\quad + \sum_{\alpha\in\mathscr{B}_\Gamma} I_\alpha[\nabla f_\alpha(\cdot, X_\cdot) D_s^k X_\cdot]_{t_m,t} + \sum_{\alpha\in\mathscr{B}_\Gamma} \tilde{D}_s^k I_\alpha[f_\alpha(\cdot, X_\cdot)]_{t_m,t} \\ &= D_s^k X_{t_m} - D_s^k X_{t_m}^\pi + \sum_{\alpha\in\Gamma} \nabla f_\alpha(t_m, X_{t_m}^\pi)[D_s^k X_{t_m} - D_s^k X_{t_m}^\pi] I_{\alpha,t_m,t} \\ &\quad + R_{t_m,t}, \end{aligned}$$

where we denote by $R_{t_m,t}(s)$

$$(4.14)\qquad \begin{aligned} R_{t_m,t}(s) &= \sum_{\alpha\in\Gamma} [\nabla f_\alpha(t_m, X_{t_m}) - \nabla f_\alpha(t_m, X_{t_m}^\pi)] D_s^k X_{t_m} I_{\alpha,t_m,t} \\ &\quad + \sum_{\alpha\in\Gamma} [f_\alpha(t_m, X_{t_m}) - f_\alpha(t_m, X_{t_m}^\pi)] D_s^k I_{\alpha,t_m,t} \\ &\quad + \sum_{\alpha\in\mathscr{B}_\Gamma} I_\alpha[\nabla f_\alpha(\cdot, X_\cdot) D_s^k X_\cdot]_{t_m,t} + \sum_{\alpha\in\mathscr{B}_\Gamma} \tilde{D}_s^k I_\alpha[f_\alpha(\cdot, X_\cdot)]_{t_m,t} \\ &:= R_{t_m,t}^1(s) + R_{t_m,t}^2(s) + R_{t_m,t}^3(s) + R_{t_m,t}^4(s). \end{aligned}$$

Repeatedly using the formula (4.13) for $t = t_m$ ($m = m(t), m(t)-1, \ldots, 1$), we have

$$D_s^k X_t - D_s^k X_t^\pi = \sum_{\alpha\in\Gamma}\sum_{m=0}^{m(t)} \nabla f_\alpha(t_m, X_{t_m}^\pi)[D_s^k X_{t_m} - D_s^k X_{t_m}^\pi] I_{\alpha,t_m,t_{m+1}\wedge t} + \sum_{m=0}^{m(t)} R_{t_m,t_{m+1}\wedge t}(s).$$

Now fix s. By Lemma 4.1 (2), we obtain

$$(4.15)\qquad E\sup_{0\le v\le t}|D_s^kX_v - D_s^kX_v^\pi|^{2p} \le C\sum_{\alpha\in\Gamma}\int_0^t E\sup_{0\le u\le v}|D_s^kX_u - D_s^kX_u^\pi|^{2p}dv + E\sup_{0\le u\le t}\left|\sum_{m=0}^{m(u)} R_{t_m,t_{m+1}\wedge u}(s)\right|^{2p}.$$

Now we estimate the last expectation of the above formula:

$$(4.16)\qquad \text{the last term of (4.15)} \le C(R_1(s) + R_2(s) + R_3(s) + R_4(s))$$

where

$$(4.17)\qquad R_i(s) = E\sup_{0\le u\le t}\left|\sum_{m=0}^{m(u)} R^i_{t_m,t_{m+1}\wedge u}(s)\right|^{2p}.$$

Denoting $t_{m(u)}$ by ϕ_u, for simplicity, $R_1(s)$ can be estimated by Lemma 4.1 (2) as

$$R_1(s) \le \sum_{\alpha\in\Gamma}\int_0^t E|[\nabla f_\alpha(\phi_u, X_{\phi_u}) - \nabla f_\alpha(\phi_u, X^\pi_{\phi_u})]D_s^kX_{\phi_u}|^{2p}du$$
$$\le C\int_0^T\left\{E\sup_{0\le u\le v}|D_s^kX_u|^{4p}\right\}^{1/2}\cdot\left\{E\sup_{0\le u\le v}|X_u - X_u^\pi|^{4p}\right\}^{1/2}dv.$$

By Lemma 4.2 and (3.11), this is further dominated by

$$(4.18)\qquad R_1(s) \le C|\pi|^{2p\gamma}.$$

Next, we estimate $R_2(s)$. It is easy to see that

$$D_s^kI_{\alpha,t_m,t_{m+1}} = 1_{[t_m,t_{m+1}]}(s)\cdot D_s^kI_{\alpha,t_m,t_{m+1}} = \delta_{m,m(s)}\cdot D_s^kI_{\alpha,t_{m(s)},t_{m(s)+1}}$$

and

$$\sup_\pi\sup_{0\le s\le T} E|D_s^kI_{\alpha,t_{m(s)},t_{m(s)+1}}|^q < \infty$$

for every $q > 1$. Then,

(4.19)
$$R_2(s) \le C\sum_{\alpha\in\Gamma} E\sup_{0\le u\le t}\left|\sum_{m=0}^{m(u)}[f_\alpha(t_m, X_{t_m}) - f_\alpha(t_m, X^\pi_{t_m})]D_s^kI_{\alpha,t_m,t_{m+1}\wedge u}\right|^{2p}$$
$$\le C\sum_{\alpha\in\Gamma}\{E|f_\alpha(t_{m(s)}, X_{t_{m(s)}}) - f_\alpha(t_{m(s)}, X^\pi_{t_{m(s)}})|^{4p}\}^{1/2}\cdot\{E|D_s^kI_{\alpha,t_{m(s)},t_{m(s)+1}\wedge t}|^{4p}\}^{1/2}$$
$$\le C\left[E\sup_{0\le u\le T}|X_u - X_u^\pi|^{4p}\right]^{1/2} \le C|\pi|^{2p\gamma}.$$

To estimate $R_3(s)$, we first note that if $\alpha\in\mathscr{B}_\Gamma$, then

$$2l(\alpha) - 2 \ge 2\gamma \qquad \text{when } l(\alpha) = n(\alpha)$$

and

$$l(\alpha) + n(\alpha) - 1 \geq 2\gamma \qquad \text{when } l(\alpha) \neq n(\alpha).$$

Then using Lemma 4.1 (2),

$$R_3(s) \leq C|\pi|^{2\gamma p} \sum_{\alpha \in \mathscr{B}_\Gamma} \int_0^T E \sup_{0 \leq u \leq v} |\nabla f_\alpha(u, X_u) D_s^k X_u|^{2p} dv \tag{4.20}$$

$$\leq C|\pi|^{2\gamma p} \int_0^T E \sup_{0 \leq u \leq v} |D_s^k X_u|^{2p} dv$$

$$\leq C|\pi|^{2\gamma p}.$$

Finally we estimate $R_4(s)$. Put $f(s_1, \dots, s_l) = f_\alpha(s_1, X_{s_1})$. Since

$$I_{\alpha\setminus i}(f_{\setminus i}[s])_{t_m, t_{m+1}} = 1_{[t_m, t_{m+1}]}(s) \cdot I_{\alpha\setminus i}(f_{\setminus i}[s])_{t_m, t_{m+1}},$$

$$R_4(s) = C \sum_{\alpha \in \mathscr{B}_\Gamma} \sum_{\alpha_i = k} E\left\{ \sup_{0 \leq u \leq t} \left| \sum_{m=0}^{m(u)} I_{\alpha\setminus i}[f_{\setminus i}[s]]_{t_m, t_{m+1} \wedge u} \right|^{2p} \right\} \tag{4.21}$$

$$= C \sum_{\alpha \in \mathscr{B}_\Gamma} \sum_{\alpha_i \neq 0} E\left\{ \sup_{t_{m(s)} \leq u \leq t_{m(s)+1} \wedge t} |I_{\alpha\setminus i}[f_{\setminus i}[s]]_{t_{m(s)}, u}|^{2p} \right\},$$

and by Lemma 4.1 (1), this is dominated by

$$C \sum_{\alpha \in \mathscr{B}_\Gamma} \sum_{\alpha_i \neq 0} |\pi|^{p[l(\alpha_{\setminus i}) + n(\alpha_{\setminus i})] - 1} \int_{t_{m(s)}}^{t_{m(s)+1} \wedge t} E[\|f_{\setminus i}[s]\|_{t_{m(s)}}(u)^{2p}] du \leq C|\pi|^{p[l(\alpha) + n(\alpha) - 1]}.$$

Since $l(\alpha) + n(\alpha) \geq 2\gamma + 1$ if $\alpha \in \mathscr{B}_\Gamma$, $l(\alpha) + n(\alpha) - 1 \geq 2\gamma$ and hence

$$R_4(s) \leq C|\pi|^{2\gamma p}.$$

Thus we have shown

$$E \sup_{0 \leq u \leq t} |D_s^k X_u - D_s^k X_u^\pi|^{2p} \leq C_1 \int_0^t E \sup_{0 \leq u \leq v} |D_s^k X_u - D_s^k X_u^\pi|^{2p} dv + C_2 |\pi|^{2\gamma p}$$

with positive constants C_1 and C_2 independent of $s \in [0, T]$. Now (4.10) follows from the Gronwall lemma.

INSTITUTE OF MATHEMATICAL SCIENCES
ACADEMIA SINICA

DEPARTMENT OF MATHEMATICS
KYOTO UNIVERSITY

References

[B-T95] V. Bally and D. Tallay, The law of the Euler scheme for stochastic differential equations (II): convergence rate of the density, preprint (1995).

[Ga73] A. M. Garsia, Martingale inequalities, Mathematics Lecture Note Series, Benjamin, 1973.

[Hu96] Y. Z. Hu, Strong and weak order of time discretization schemes of stochastic differential equations, Sém. Prob. XXI, LNM **1626**, Springer (1996), 218–227.

[IW89] N. Ikeda and S. Watanabe, Stochastic differential equations and diffusion processes, second edition, North Holland, 1989.

[Kl-Pl92] P. E. Kloeden and E. Platen, Numerical solutions of stochastic differential equations, Springer, 1992.

[Ko95] A. Kohatsu-Higa, High order Itô-Taylor approximations to the density of a diffusion, Preprint (1995).

[Ku-St85] S. Kusuoka an D. W. Stroock, Applications of Malliavin calculus, part II, J. Fac. Sci. Univ. Tokyo, Sect. IA, Math., **32** (1985), 1–76.

[Lé87] R. Léandre, Minoration en temps petit de la densité d'une diffusion dégénérée, Journ. Funct. Anal., **74** (1987), 399–414.

[Ma78] P. Malliavin, C^k-hypoellicipticity with degeneracy, Stochastic Analysis (ed. A. Friedman and M. Pinsky), Academic Press (1978), 199–214, 327–340.

[Wa84] S. Watanabe, Stochastic differential equation and Malliavin calculus, Tata Institute of Fundamental Research, Springer, 1984.

[Wa93] S. Watanabe, Fractional order Sobolev spaces on Wiener space, Probab. Theory Relat. Fields, **95** (1993), 175–198.