2004

Discrete-time Approximations of Stochastic Delay Equations: The Milstein Scheme

Yaozhong Hu
University of Kansas Main Campus

Salah-Eldin A. Mohammed
Southern Illinois University Carbondale, salah@sfde.math.siu.edu

Feng Yan
Williams Energy

Published in *Annals of Probability*, 32(1A), 265-314.

The Annals of Probability
2004, Vol. 32, No. 1A, 265–314

DISCRETE-TIME APPROXIMATIONS OF STOCHASTIC DELAY EQUATIONS: THE MILSTEIN SCHEME

BY YAOZHONG HU,[1] SALAH-ELDIN A. MOHAMMED[2] AND FENG YAN

University of Kansas, *Southern Illinois University and Williams Energy*

In this paper, we develop a strong Milstein approximation scheme for solving stochastic delay differential equations (SDDEs). The scheme has convergence order 1. In order to establish the scheme, we prove an infinite-dimensional Itô formula for "tame" functions acting on the segment process of the solution of an SDDE. It is interesting to note that the presence of the memory in the SDDE requires the use of the Malliavin calculus and the anticipating stochastic analysis of Nualart and Pardoux. Given the *nonanticipating* nature of the SDDE, the use of *anticipating* calculus methods in the context of strong approximation schemes appears to be novel.

1. Introduction. Discrete-time strong approximation schemes for stochastic ordinary differential equations (SODEs) are well developed. For an extensive study of these numerical schemes, one may refer to [17], [18] and [19], Chapters 5 and 6. Some basic ideas of strong and weak orders of convergence are illustrated in [13].

If the rate of change of a physical system depends only on its present state and some noisy input, then the system can often be described by a stochastic ordinary differential equation (SODE). However, in many physical situations the rate of change of the state depends not only on the present but also on the past states of the system. In such cases, *stochastic delay differential equations* (SDDEs) or *stochastic functional differential equations* (SFDEs) provide important tools to describe and analyze these systems. For various aspects of the qualitative theory of SFDEs the reader may refer to [20, 21] and the references therein.

SDDEs and SFDEs arising in many applications cannot be solved explicitly. Hence, one needs to develop effective numerical techniques for such systems. Depending on the particular physical model, it may be necessary to design strong L^p (or almost sure) numerical schemes for pathwise solutions of the underlying SFDE. Strong approximation schemes for SFDEs may be used to

Received April 2001; revised December 2002.

[1]Supported in part by NSF Grants EPS-98-74732, DMS-02-04613 and General Research Fund, University of Kansas.

[2]Supported in part by NSF Grants DMS-97-03596, DMS-99-75462 and DMS-02-03368.

AMS 2000 subject classifications. Primary 34K50, 60H07, 60H35; secondary 60C30, 60H10, 37H10, 34K28.

Key words and phrases. Milstein scheme, Itô's formula, tame functions, anticipating calculus, Malliavin calculus, weak derivatives.

simulate directly the a.s. stochastic dynamics of their trajectories or their random attractors. SFDEs are used to model population growth with incubation/gestation period [21]. In such models, one is often interested in estimating the actual population rather than its distribution and hence the need for strong approximation schemes.

In this article, we will not consider the order of convergence of *weak* numerical schemes, although such schemes are useful for some applications of SODEs (see [13, 17] and the references therein). In this connection, it is important to note that stochastic systems with memory *do not correspond to deterministic PDEs* (in finitely many space variables) [20, 21]. Typically, a stochastic system with memory corresponds to an *infinite-dimensional* Feller diffusion whose principal coefficient degenerates on a hypersurface with *finite-codimension* ([20], Chapter IV, Theorem 3.2 and [21], Theorem II.3). This aspect of SFDEs is in sharp contrast with the theory of SODEs where the latter theory has traditional ties to diffusions in Euclidean space. In a sense, the numerics of stochastic systems with memory resemble those of SPDEs in one space dimension.

A strong Cauchy–Maruyama scheme for a class of SFDEs with continuous memory, in the context of the Delfour–Mitter state space $\mathbf{R}^m \times L^2([-\tau, 0], \mathbf{R}^m)$, was developed by Ahmed, Elsanousi and Mohammed [1]. See also [20], page 227, [15] and [4]. As in the case of SODEs, the Cauchy–Maruyama scheme for SFDEs has order of convergence $\frac{1}{2}$ ([20], page 227, [15, 4, 8, 14]).

In Sections 2–5, we establish the strong Milstein scheme for SDDEs with several delays. This scheme has a higher strong order of convergence 1 when compared with the Euler scheme which, as indicated above, has the strong order of convergence 0.5. Furthermore, when simulating the whole solution path $\{X(t), t \in [0, a]\}$, the Milstein schemes for SDDEs and SODEs have the same complexity, even when one accounts for the simulation of the iterated stochastic integrals in the scheme. (See Appendix B and the remarks therein.) Although the solution of the SDDE is *adapted* to the (lagged) filtration of the driving noise, methods from *anticipating* stochastic analysis and the Malliavin calculus are necessary in order to derive an Itô formula for the segment of the solution process. The Itô formula is essential for the development of the Milstein scheme.

In order to put our analysis in proper perspective, we highlight its essential features: (a) The dynamics and the coefficients of the SDDEs are adapted, in fact, driven by Itô integrals; (b) the formulation and implementation of the Milstein scheme do not require anticipating calculus ideas; (c) the proof of convergence of the Milstein scheme as well as the Itô formula employ anticipating calculus techniques; (d) anticipating calculus methods are used in the context of strong approximation schemes rather than weak ones (where the Feynman–Kac formula lends itself naturally to the use of Malliavin calculus methods); (e) the application of anticipating calculus methods seems unavoidable as soon as one seeks higher-order approximation results.

In an essentially nonadapted setting, anticipating calculus methods have been used by Pardoux and Protter to study stochastic Volterra equations with anticipating coefficients. See [24] and the references therein. See also [7].

In order to describe our set-up, we need the following notation.

Let $\mathbf{R}^m$ be m-dimensional Euclidean space with the Euclidean norm $|x| := \sqrt{x_1^2 + \cdots + x_m^2}$, $x = (x_1, \ldots, x_m) \in \mathbf{R}^m$. Denote $T := [0, a]$, $J := [-\tau, 0]$, $C := C(J; \mathbf{R}^m)$, where m is a positive integer, $\tau > 0$ is a fixed delay [as in (1.6)] and $a > 0$. Furnish C with the supremum norm $\|\eta\|_C := \sup_{-\tau \le s \le 0} |\eta(s)|$ for all $\eta \in C$.

Define the projection $\Pi : C \to \mathbf{R}^{mk}$ associated with $s_1, \ldots, s_k \in [-\tau, 0]$ by

$$\Pi(\eta) := \big(\eta(s_1), \ldots, \eta(s_k)\big) \in \mathbf{R}^{mk} \tag{1.1}$$

for all $\eta \in C$.

DEFINITION 1.1. A function $\Phi \in C(T \times C(J; \mathbf{R}^m); \mathbf{R})$ is *tame* if there exist $\phi \in C(T \times \mathbf{R}^{mk}, \mathbf{R})$ and a projection $\Pi : C \to \mathbf{R}^{mk}$ such that

$$\Phi(t, \eta) = \phi\big(t, \Pi(\eta)\big) \tag{1.2}$$

for all $t \in T$ and $\eta \in C$.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. For any continuous m-dimensional process $X : [-\tau, a] \times \Omega \to \mathbf{R}^m$, define the *segment process* X_t, $t \in [0, a]$, by

$$X_t(u) := X(t + u), \qquad t \in [0, a],\ u \in [-\tau, 0]. \tag{1.3}$$

Observe that $\{X_t\}$ may be considered as a C-valued or $L^2(J; \mathbf{R}^m)$-valued process.

It is important that one should distinguish between the finite-dimensional current state $X(t)$ and the infinite-dimensional segment X_t, $t \in [0, a]$.

Assume that $g : T \times \mathbf{R}^{mk_1} \to L(\mathbf{R}^d; \mathbf{R}^m)$ and $h : T \times \mathbf{R}^{mk_2} \to \mathbf{R}^m$ satisfy the following Lipschitz condition:

$$\begin{aligned} |g(t, x) - g(t, y)| &\le L|x - y|, \\ |h(t, z) - h(t, w)| &\le L|z - w| \end{aligned} \tag{1.4}$$

for all $t \in T$, $x, y \in \mathbf{R}^{mk_1}$ and $z, w \in \mathbf{R}^{mk_2}$, where $L > 0$ is a constant, together with the boundedness condition

$$\sup_{0 \le t \le a} \big[|g(t, 0)| + |h(t, 0)|\big] < \infty. \tag{1.5}$$

Let Π_1 and Π_2 be two projections associated with two sets of points $s_{1,1}, \ldots, s_{1,k_1} \in [-\tau, 0]$ and $s_{2,1}, \ldots, s_{2,k_2} \in [-\tau, 0]$, respectively. Suppose

$\{W(t) := (W^1(t), \ldots, W^d(t)) : t \geq 0\}$ is a d-dimensional standard Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. Let $\eta : \Omega \to C([-\tau, 0]; \mathbf{R}^m)$ be an $\mathcal{F}_0$-measurable initial process.

We will first consider the following class of Itô SDDEs:

$$X(t) = \begin{cases} \eta(0) + \int_0^t g\big(s, \Pi_1(X_s)\big)\, dW(s) + \int_0^t h\big(s, \Pi_2(X_s)\big)\, ds, & t \geq 0, \\ \eta(t), & -\tau \leq t < 0. \end{cases} \tag{1.6}$$

Under conditions (1.4) and (1.5), the SDDE (1.6) has a unique strong solution (cf. [20], Theorem II.2.1, page 36, and Theorem V.4.3, pages 151 and 152). To see this, let $G(t, \eta) := g(t, \Pi_1(\eta))$ and $H(t, \eta) := h(t, \Pi_2(\eta))$ for $t \in [0, a]$, $\eta \in C$. It is easy to check that G and H satisfy the Lipschitz and local boundedness conditions (with respect to the supremum norm on C) of Theorems II.2.1 and V.4.3 of [20]. Therefore, for each $p \geq 1$, there exists a constant $C = C(p, L, a) > 0$ such that

$$E\|X_t\|_C^{2p} \leq C\big(1 + E\|\eta\|_C^{2p}\big) \tag{1.7}$$

for all $\eta \in C$, $t \in [0, a]$.

For any integers $n, l \geq 1$, let $\pi : t_{-l} < t_{-l+1} < \cdots < 0 = t_0 < t_1 < t_2 < \cdots < t_n$ be a partition of $[-\tau, a]$. Denote by $|\pi| := \max_{-l\leq i\leq n-1}(t_{i+1} - t_i)$, the mesh of π. We now introduce the following Milstein scheme for the SDDE (1.6):

$$\begin{aligned} X^{i,\pi}(t) = X^{i,\pi}(t_k) &+ h^i\big(t_k, \Pi_2\big(X^\pi_{t_k}\big)\big)(t - t_k) \\ &+ g^{ij}\big(t_k, \Pi_1\big(X^\pi_{t_k}\big)\big)\big(W^j(t) - W^j(t_k)\big) \\ &+ \frac{\partial g^{ij}}{\partial x_{i_1 j_1}}\big(t_k, \Pi_1\big(X^\pi_{t_k}\big)\big) u^{i_1 j_1,\pi}(t_k + s_{1,j_1}) \\ &\quad \times I_{j,j_1}\big(t_k + s_{1,j_1}, t + s_{1,j_1}; s_{1,j_1}\big) \end{aligned} \tag{1.8}$$

for $t_k < t \leq t_{k+1}$, and

$$X^\pi(t) := \eta^\pi(t), \qquad t \in [-\tau, 0],$$

where

$$u^{i_1 j_1,\pi}(t) := \begin{cases} g^{i_1 j_1}\big(t, \Pi_1(X^\pi_t)\big), & t \geq 0, \\ 0, & -\tau \leq t < 0, \end{cases} \tag{1.9}$$

$$\begin{aligned} I_{j,j_1}(t_0 + s, t + s; s) := \int_{t_0}^t \int_{t_0+s}^{t_1+s} \circ\, dW^j(t_2) \circ dW^{j_1}(t_1), \\ t \geq t_0 \geq 0,\ s \in [-\tau, 0], \end{aligned} \tag{1.10}$$

and the starting path $\eta^\pi \in C(J, \mathbf{R}^m)$ is prescribed (e.g., a piecewise linear approximation of η using the partition points $\{t_{-l}, \ldots, t_0\}$). In (1.8), X^i, h^i and g^{ij}

denote coordinate representations of X, h and g with respect to standard bases in the underlying Euclidean spaces, and the Einstein summation convention is used for repeated indices.

In order to establish strong convergence of the above Milstein scheme for the SDDE (1.6), it turns out, surprisingly, that one requires the use of anticipating calculus techniques developed by Nualart and Pardoux [23]. In particular, one needs to develop an infinite-dimensional Itô formula for "tame" functions acting on the segment X_t of the solution X of (1.6). Such an Itô formula is given in Section 2, Theorem 2.3. The formula is proved via anticipating calculus methods [23]. To understand the *need* for *anticipating* calculus in such an *intrinsically adapted* setting, it is instructive to look at the following simple one-dimensional SDDE:

$$
\begin{aligned}
dX(t) &= g\bigl(X(t-1), X(t)\bigr)\, dW(t), \qquad t \geq 0,\\
X(t) &= W(t), \qquad -1 \leq t < 0,
\end{aligned}
$$

where $g : \mathbf{R}^2 \to \mathbf{R}$ is a smooth function and $W(t), t \geq -1$, is a one-dimensional Brownian motion. For a second-order scheme, we formally seek a stochastic differential of the coefficient $g(X(t-1), X(t))$ on the right-hand side of the above SDDE. For $t \in (0, 1]$, this gives formally

$$
\begin{aligned}
&d\bigl\{g\bigl(X(t-1), X(t)\bigr)\bigr\}\\
&\quad = d\bigl\{g\bigl(W(t-1), X(t)\bigr)\bigr\}\\
&\quad = \frac{\partial g}{\partial x}\bigl(W(t-1), X(t)\bigr)\, dW(t-1)\\
&\qquad + \frac{\partial g}{\partial y}\bigl(W(t-1), X(t)\bigr) g\bigl(X(t-1), X(t)\bigr)\, dW(t)\\
&\qquad + \text{second-order terms.}
\end{aligned}
$$

Note that although the coefficient $g(X(t-1), X(t))$ is $\mathcal{F}_t$-measurable, the first term $\frac{\partial g}{\partial x}(W(t-1), X(t))\, dW(t-1)$ in the right-hand side of the last equality is an anticipating differential. Furthermore, it appears that the $(\mathcal{F}_t)_{0 \leq t \leq 1}$-adapted process $[0, 1] \ni t \to (X(t-1), X(t)) \in \mathbf{R}^2$ is not a semimartingale with respect to any natural filtration. In addition to this difficulty, the components $X(t-1)$ and $X(t)$ are not independent, so the existing anticipating versions of Itô's formula do not apply (cf. [2, 3] and [23]); hence the need for a new Itô formula for tame functions in order to justify the above computation. In Theorem 2.1 in the next section we establish such a formula.

Using the above-mentioned Itô formula and appropriate estimates on the weak Cameron–Martin derivatives of X, it is shown in Section 5 (Theorem 5.2) that, under suitable regularity conditions on the coefficients of (1.6), one gets the following global error estimate for the Milstein approximations:

$$
E \sup_{0 \leq t \leq a} \|X_t^{\pi} - X_t\|_C^q \leq C(q)|\pi|^q \tag{1.11}
$$

for any $q \geq 1$. This says that the Milstein scheme has strong order of convergence 1.

2. Itô's formula for "tame" functions. In order to derive higher-order numerical schemes for SDDEs, we shall first prove an Itô formula for "tame" functions on $C(J, \mathbf{R}^m)$ (Definition 1.1).

Suppose that $W(t) := (W^1(t), \dots, W^d(t))$, $t \geq 0$, is d-dimensional standard Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. Denote by $D = (D_1, \dots, D_d)$ the Malliavin differentiation operator associated with $\{W(t) : t \geq 0\}$. Assume

$$X(t) = \begin{cases} \eta(0) + \int_0^t u(s)\, dW(s) + \int_0^t v(s)\, ds, & t > 0, \\ \eta(t), & -\tau \leq t \leq 0, \end{cases} \tag{2.1}$$

where η belongs to C and is of bounded variation, $u = (u^1, \dots, u^m)^T$, $u^i \in \mathbb{L}^{2,4}_{d,\mathrm{loc}}$, $v = (v^1, \dots, v^m)^T$, and $v^i \in \mathbb{L}^{1,4}_{\mathrm{loc}}$. One can refer to ([22], pages 61, 151 and 161), for the definition of $\mathbb{L}^{k,p}_d$. Note that the processes u and v may not be adapted to the Brownian filtration $(\mathcal{F}_t)_{t\geq 0}$. For convenience, we define $u(t) = 0$ for $t < 0$ or $t > a$,

$$v(t) = \begin{cases} 0, & t > a, \\ \eta'(t), & -\tau \leq t \leq 0. \end{cases}$$

We also set $W(t) = 0$ if $t < 0$ or $t > a$, and denote

$$\begin{aligned} U(t) &:= \int_0^t u(s)\, dW(s), \\ V(t) &:= \begin{cases} \eta(0) + \int_0^t v(s)\, ds, & t > 0, \\ \eta(t), & -\tau \leq t \leq 0. \end{cases} \end{aligned} \tag{2.2}$$

If $u \in \mathbb{L}^{2,p}_{\mathrm{loc}}$ for some $p > 4$, then the indefinite Skorohod integral $\int_0^t u(s)\, dW(s)$ has a continuous version. Hence, we may assume that the process $X(t)$, $t \geq -\tau$, is sample continuous.

Let $T = [0, a]$, $J = [-\tau, 0]$, $C = C(J, \mathbf{R}^m)$ be as before, and let Π be the projection associated with $s_1, \dots, s_k \in J$. Although there is a multidimensional Itô formula for $\phi(t, X(t))$ ([2, 3] and [22]), we cannot apply it to $\phi(t, \Pi(X_t))$ because $\Pi(U_t)$ is of the form

$$\left(\int_0^t u(s + s_1)\, dW(s + s_1), \dots, \int_0^t u(s + s_k)\, dW(s + s_k) \right), \qquad t > 0 \tag{2.3}$$

and the components of the dk-dimensional process $(W(t + s_1), \dots, W(t + s_k))$ are not independent. However, the ideas in [23], Section 6, and in [22], page 161, can be used to derive an Itô formula for $\phi(t, \Pi(X_t))$. See [28] for further details.

We denote by

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \tag{2.4}$$

the Kronecker delta.

For any process $X(t)$, $t \in [-\tau, a]$, denote its (delayed) increments by

$$\Delta_{li} X := X(t_l + s_i) - X(t_{l-1} + s_i), \qquad 1 \leq l \leq n,\ i = 1, 2, \dots, k. \tag{2.5}$$

Assume that $\phi \in C^{1,2}(T \times \mathbf{R}^{mk}, \mathbf{R})$, and write

$$\phi(t, \vec{x}) := \phi(t, \vec{x}_1, \dots, \vec{x}_m) \tag{2.6}$$

where $\vec{x} := (\vec{x}_1, \dots, \vec{x}_m)$, $\vec{x}_i := (x_{i1}, \dots, x_{ik}) \in \mathbf{R}^k$, $1 \leq i \leq m$.

We now state an Itô formula for "tame" functions.

THEOREM 2.1. *Assume that X is a continuous process defined by* (2.1), *where $\eta : J \to \mathbf{R}^m$ is of bounded variation, $u = (u^1, \dots, u^m)^T$, $u^i \in \mathbb{L}^{2,4}_{d,\mathrm{loc}}$, $v = (v^1, \dots, v^m)^T$ and $v^i \in \mathbb{L}^{1,4}_{\mathrm{loc}}$. Suppose $\phi \in C^{1,2}(T \times \mathbf{R}^{mk}, \mathbf{R})$. Then*

$$\begin{aligned} &\phi\big(t, \Pi(X_t)\big) - \phi\big(0, \Pi(X_0)\big) \\ &\quad = \int_0^t \frac{\partial \phi}{\partial s}\big(s, \Pi(X_s)\big)\, ds + \int_0^t \frac{\partial \phi}{\partial \vec{x}}\big(s, \Pi(X_s)\big)\, d(\Pi(X_s)) \\ &\qquad + \frac{1}{2} \sum_{i,j=1}^{k} \sum_{i_1, j_1 = 1}^{m} \int_0^t \frac{\partial^2 \phi}{\partial x_{i_1 i}\, \partial x_{j_1 j}}(s, \Pi(X_s)) u^{i_1}(s + s_i) \\ &\qquad\qquad \times D_{s+s_i} X^{j_1}(s + s_j)\, ds. \end{aligned} \tag{2.7}$$

REMARKS.

1. The Itô formula (2.7) may also be expressed in the form

$$\begin{aligned} &\phi\big(t, \Pi(X_t)\big) - \phi\big(0, \Pi(X_0)\big) \\ &\quad = \int_0^t \frac{\partial \phi}{\partial s}\big(s, \Pi(X_s)\big)\, ds + \int_0^t \frac{\partial \phi}{\partial \vec{x}}\big(s, \Pi(X_s)\big)\, d(\Pi(X_s)) \\ &\qquad + \frac{1}{2} \sum_{i,j=1}^{k} \int_0^t Tr\bigg[\frac{\partial^2 \phi}{\partial \vec{x}_i\, \partial \vec{x}_j}\big(s, \Pi(X_s)\big)\big(\Theta_s(s_i, s_j)\big)\bigg]\, ds, \end{aligned} \tag{2.8}$$

where

$$\Theta_s(\alpha, \beta) := \tfrac{1}{2}\big\{(u\Lambda)_s X_s(\alpha, \beta) + (u\Lambda)_s X_s(\beta, \alpha)\big\}, \qquad \alpha, \beta \in [-\tau, 0]$$

and the two-parameter process $(u\Lambda)_s X_s : \Omega \times J^2 \to L(\mathbf{R}^m; \mathbf{R}^m)$ is defined by

$$(u\Lambda)_s X_s(\alpha, \beta) := I_{\{0 \le s+\alpha\wedge\beta\}} u(s+\alpha) \times \left[u^T(s+\alpha) I_{\{0\le s+\alpha \le s+\beta\}} + \int_0^{s+\beta} D_{s+\alpha} u(r)\, dW(r) + \int_0^{s+\beta} D_{s+\alpha} v(r)\, dr \right]$$

for all $\alpha, \beta \in [-\tau, 0]$.

2. Suppose $d = m = 1$. Let us define a trace operator $\triangledown$. For $1 \le i, j \le k$, define

$$\triangledown^{\pm}_{s_i, s_j} X(s) := \lim_{\varepsilon \downarrow 0} \big(D_{s+s_i} X(s + s_j + \varepsilon) \pm D_{s+s_i} X(s + s_j - \varepsilon) \big) \in \mathbf{R} \tag{2.9}$$

and $\triangledown^{\pm}_{s_i} X(s) := (\triangledown^{\pm}_{s_i,s_1} X(s), \ldots, \triangledown^{\pm}_{s_i,s_k} X(s)) \in \mathbf{R}^k$. Then the Itô formula for "tame" functions can be written as

$$\begin{aligned} &\phi(t, \Pi(X_t)) - \phi(0, \Pi(X_0)) \\ &\quad = \int_0^t \frac{\partial \phi}{\partial s}(s, \Pi(X_s))\, ds + \int_0^t \frac{\partial \phi}{\partial \vec{x}}(s, \Pi(X_s))\, d\Pi(W_s) \\ &\qquad + \frac{1}{2} \sum_{i=1}^k \int_0^t \left\langle \frac{\partial^2 \phi}{\partial x_i^2}(s, \Pi(X_s))\, \triangledown^{+}_{s_i} X(s), \triangledown^{-}_{s_i} X(s) \right\rangle_{\mathbf{R}^d} ds \end{aligned} \tag{2.10}$$

a.s. for all $t \in T$, where $\vec{x} := (x_1, \ldots, x_k)$ and $\langle \cdot, \cdot \rangle_{\mathbf{R}^d}$ denotes the Euclidean inner product on $\mathbf{R}^d$. See [23], Remark 7.6.

3. The Itô formula (2.7) still holds if the initial path is an $\mathcal{F}_0$-measurable process $\eta : \Omega \to C$ with a.a. sample paths of bounded variation. A similar remark also holds for Theorem 5.2 of Section 5.

For simplicity, we shall prove the Itô formula for the case $d = m = 1$. We thus assume in what follows that $d = m = 1$.

PROOF OF THEOREM 2.1. For any integer $n \ge 1$, let $\{\pi_n : 0 = t_0 < t_1 < \cdots < t_n = a\}$ be a partition of $[0, a]$. Then by Taylor's theorem, we may write

$$\begin{aligned} &\phi(t, \Pi(X_t)) - \phi(0, \Pi(X_0)) \\ &\quad = \sum_{l=1}^n [\phi(t_l, \Pi(X_{t_l})) - \phi(t_{l-1}, \Pi(X_{t_l}))] \\ &\qquad + [\phi(t_{l-1}, \Pi(X_{t_l})) - \phi(t_{l-1}, \Pi(X_{t_{l-1}}))] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{l=1}^{n} \frac{\partial \phi}{\partial s}(\hat{t}_l, \Pi(X_{t_l}))\Delta t_l \\
&\quad + \sum_{l=1}^{n} \Biggl\{ \sum_{i=1}^{k} \frac{\partial \phi}{\partial x_i}(t_{l-1}, \Pi(X_{t_{l-1}}))\Delta_{li} X \\
&\qquad + \frac{1}{2} \sum_{i,j=1}^{k} \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}(t_{l-1}, \Pi(\bar{X}_{t_l}))\Delta_{li} X \Delta_{lj} X \Biggr\}, \qquad t \in T,
\end{aligned}
$$

where

$$
\Delta t_l := t_l - t_{l-1}, \qquad \bar{X}_{t_l} = X_{t_{l-1}} + \alpha_l (X_{t_l} - X_{t_{l-1}}), \qquad \hat{t}_l = t_{l-1} + \gamma_l \Delta t_l
$$

for some random variables $0 \le \alpha_l, \gamma_l \le 1$, $l = 1, \dots, n$. The Itô formula (2.10) will then follow from Propositions 2.3 and 2.4. □

The rest of this section is devoted to the proofs of Propositions 2.2–2.4.

PROPOSITION 2.2. *Suppose that $W(t)$ is a one-dimensional Brownian motion. Let $u \in \mathbb{L}^{1,2}_{\text{loc}}$ be such that $u(t) = 0$ if $t > a$ or $t < 0$. Assume that $-\tau \le s_1$, $s_2 \le 0$ and let $\pi_n : 0 = t_0 < t_1 < \cdots < t_n = a$ be a family of partitions of $T = [0, a]$, with $|\pi_n| \to 0$ as $n \to \infty$. Then*

$$
\lim_{n\to\infty} \left[\sum_{l=1}^{n} \int_{t_{l-1}+s_1}^{t_l+s_1} u(s)\, dW(s) \right]^2 = \int_0^{a+s_1} u^2(s)\, ds \tag{2.11}
$$

in probability. If $s_1 \ne s_2$, then

$$
\lim_{n\to\infty} \sum_{l=1}^{n} \int_{t_{l-1}+s_1}^{t_l+s_1} u(s)\, dW(s) \int_{t_{l-1}+s_2}^{t_l+s_2} u(s)\, dW(s) = 0 \tag{2.12}
$$

in probability. Furthermore, if $u \in \mathbb{L}^{1,2}$, then the above convergences are in $L^1(\Omega, \mathbf{R})$.

PROOF. We prove the proposition for $u \in \mathbb{L}^{1,2}$. The general case $u \in \mathbb{L}^{1,2}_{\text{loc}}$ follows by a standard localization argument [22].

If $u_i, u_j, v_i, v_j \in \mathbb{L}^{1,2}$ with $u_i(t) = v_i(t) = 0$ if $t < 0$ or $t > a + s_i$ and $u_j(t) = v_j(t) = 0$ if $t < 0$ or $t > a + s_j$. Set

$$
\begin{aligned}
U_i(t) &:= \int_0^t u_i(s)\, dW(s), \qquad V_i(t) := \int_0^t v_i(s)\, dW(s), \\
U_j(t) &:= \int_0^t u_j(s)\, dW(s), \qquad V_j(t) := \int_0^t v_j(s)\, dW(s).
\end{aligned} \tag{2.13}
$$

Then

$$\begin{aligned}
&E\left|\sum_{l=1}^{n}\Delta_{li}U_i\Delta_{lj}U_j-\sum_{l=1}^{n}\Delta_{li}V_i\Delta_{lj}V_j\right|\\
&\quad=E\left|\sum_{l=1}^{n}\Delta_{li}(U_i-V_i)\Delta_{lj}U_j+\sum_{l=1}^{n}\Delta_{li}V_i\Delta_{lj}(U_j-V_j)\right|\\
&\quad\le E\left|\sum_{l=1}^{n}\Delta_{li}(U_i-V_i)\Delta_{lj}U_j\right|+E\left|\sum_{l=1}^{n}\Delta_{li}V_i\Delta_{lj}(U_j-V_j)\right|\\
&\quad\le\left(E\sum_{l=1}^{n}|\Delta_{li}(U_i-V_i)|^2\right)^{1/2}\left(E\sum_{l=1}^{n}|\Delta_{lj}(U_j)|^2\right)^{1/2}\\
&\qquad+\left(E\sum_{l=1}^{n}|\Delta_{li}(V_i)|^2\right)^{1/2}\left(E\sum_{l=1}^{n}|\Delta_{lj}(U_j-V_j)|^2\right)^{1/2}.
\end{aligned}$$

By an L^p estimate of the Skorohod integral ([23], Proposition 3.5, and [22], page 158), we have

$$\begin{aligned}
&E\sum_{l=1}^{n}|\Delta_{lj}U_j|^2\\
&\quad=E\sum_{l=1}^{n}\left|\int_{t_{l-1}+s_j}^{t_l+s_j}u_j(s)\,dW(s)\right|^2\\
&\quad=E\sum_{l=1}^{n}\left|\int_0^a I_{(t_{l-1}+s_j,t_l+s_j]}(s)u_j(s)\,dW(s)\right|^2\\
&\quad\le\sum_{l=1}^{n}\int_0^a I_{(t_{l-1}+s_j,t_l+s_j]}(s)Eu_j^2(s)\,ds\\
&\qquad+\sum_{l=1}^{n}\int_0^a\int_0^a I_{(t_{l-1}+s_j,t_l+s_j]}(s)E\big(D_tu_j(s)\big)^2\,ds\,dt\\
&\quad=\int_0^a Eu_j^2(s)\,ds+\int_0^a\int_0^a E\big(D_tu_j(s)\big)^2\,ds\,dt\\
&\quad=\|u_j\|_{1,2}^2.
\end{aligned}$$

Hence we obtain the following inequality:

$$\begin{aligned}
&E\left|\sum_{l=1}^{n}\Delta_{li}U_i\Delta_{lj}U_j-\sum_{l=1}^{n}\Delta_{li}V_i\Delta_{lj}V_j\right|\\
&\quad\le\|u_i-v_i\|_{1,2}\|u_j\|_{1,2}+\|v_i\|_{1,2}\|u_j-v_j\|_{1,2}.
\end{aligned}\tag{2.14}$$

Since $\mathbb{L}^{1,2} \cap L^4(\Omega \times [0,a])$ is dense in $\mathbb{L}^{1,2}$, it suffices to prove (2.12) for the case $u \in \mathbb{L}^{1,2} \cap L^4(\Omega \times [0,a])$. Set

$$u_i(t) := \begin{cases} u(t), & 0 \le t \le a + s_i, \\ 0, & t < 0 \text{ or } t > a + s_i. \end{cases} \tag{2.15}$$

Define

$$u_i^n(t) := \sum_{l=1}^{n} \frac{I_{(t_{l-1}+s_i, t_l+s_i]}(t)}{t_l - t_{l-1}} \int_{t_{l-1}+s_i}^{t_l+s_i} u(s)\, ds \tag{2.16}$$

and u_j^n similarly. Let

$$\begin{aligned} U_i(t) &:= \int_0^t u_i(s)\, dW(s), \qquad U_i^n(t) := \int_0^t u_i^n(s)\, dW(s), \\ U_j(t) &:= \int_0^t u_j(s)\, dW(s), \qquad V_j^n(t) := \int_0^t u_j^n(s)\, dW(s). \end{aligned} \tag{2.17}$$

Using (2.14) it is easy to check that

$$\lim_{n\to\infty} E\left| \sum_{l=1}^{n} \Delta_{li} U_i^n \Delta_{lj} U_j^n - \sum_{l=1}^{n} \Delta_{li} U_i \Delta_{lj} U_j \right| = 0. \tag{2.18}$$

By the formula for the Skorohod integral of a process multiplied by a random variable ([23], Theorem 3.2), we get

$$\begin{aligned} \Delta_{li} U_i^n &= \int_{t_{l-1}+s_i}^{t_l+s_i} \sum_{k=1}^{n} \frac{I_{(t_{k-1}+s_i, t_k+s_i]}(t)}{t_k - t_{k-1}} \int_{t_{k-1}+s_i}^{t_k+s_i} u_i(s)\, ds\, dW(t) \\ &= \frac{1}{t_l - t_{l-1}} \int_{t_{l-1}+s_i}^{t_l+s_i} u_i(s)\, ds [W(t_l + s_i) - W(t_{l-1} + s_i)] \\ &\quad + \frac{1}{t_l - t_{l-1}} \int_{t_{l-1}+s_i}^{t_l+s_i} \int_{t_{l-1}+s_i}^{t_l+s_i} D_t u_i(s)\, ds\, dt \\ &= P_{li} \Delta_{li} W + Q_{li}, \end{aligned}$$

where

$$\begin{aligned} P_{li} &:= \frac{1}{t_l - t_{l-1}} \int_{t_{l-1}+s_i}^{t_l+s_i} u_i(s)\, ds, \\ Q_{li} &:= \frac{1}{t_l - t_{l-1}} \int_{t_{l-1}+s_i}^{t_l+s_i} \int_{t_{l-1}+s_i}^{t_l+s_i} D_t u_i(s)\, ds\, dt. \end{aligned}$$

Therefore,

$$\sum_{l=1}^{n}\Delta_{li}U_i^n\Delta_{lj}U_j^n=\sum_{l=1}^{n}(P_{li}\Delta_{li}W+Q_{li})(P_{lj}\Delta_{lj}W+Q_{lj})$$
$$=\sum_{l=1}^{n}(P_{li}P_{lj})(\Delta_{li}W\Delta_{lj}W)+\sum_{l=1}^{n}(P_{li}Q_{lj})\Delta_{li}W$$
$$+\sum_{l=1}^{n}(P_{lj}Q_{li})\Delta_{lj}W+\sum_{l=1}^{n}Q_{li}Q_{lj}.$$

By Hölder's inequality,

$$\sum_{l=1}^{n}Q_{li}^2\le\sum_{l=1}^{n}\int_{t_{l-1}+s_i}^{t_l+s_i}\int_{t_{l-1}+s_i}^{t_l+s_i}|D_tu_i(s)|^2\,ds\,dt. \tag{2.19}$$

Thus $\lim_{n\to\infty}E\sum_{l=1}^{n}Q_{li}^2=0$. Now

$$\sum_{l=1}^{n}(P_{li}\Delta_{li}W)^2=\sum_{l=1}^{n}\frac{(\Delta_{li}W)^2}{(t_l-t_{l-1})^2}\left(\int_{t_{l-1}+s_i}^{t_l+s_i}u_i(s)\,ds\right)^2$$
$$=\sum_{l=1}^{n}\frac{(\Delta_{li}W)^2}{t_l-t_{l-1}}\int_{t_{l-1}+s_i}^{t_l+s_i}(u_i^n(s))^2\,ds.$$

It is easy to check that $E\|(u_i^n)^2\|_{L^2([0,a+s_i])}\le E\|u_i^2\|_{L^2([0,a+s_i])}$ and

$$\lim_{n\to\infty}E\|(u_i^n)^2-u_i^2\|_{L^2([0,a+s_i])}=0. \tag{2.20}$$

By an argument similar to the one used in the proof of Lemma A.2, we can show that $\{\sum_{l=1}^{n}(P_{li}\Delta_{li}W)^2,\ n\ge 1\}$ is uniformly integrable. Applying Lemma A.2, we have

$$\lim_{n\to\infty}E\left|\sum_{l=1}^{n}(P_{li}\Delta_{li}W)^2-\int_0^{a+s_i}u_i^2(s)\,ds\right|=0. \tag{2.21}$$

The Cauchy–Schwarz-type inequality

$$E\left|\sum_{l=1}^{n}(P_{li}\Delta_{li}W)Q_{li}\right|\le\sqrt{E\sum_{l=1}^{n}(P_{li}\Delta_{li}W)^2E\sum_{l=1}^{n}Q_{li}^2} \tag{2.22}$$

together with (2.19) and (2.21) implies that $\lim_{n\to\infty}E|\sum_{l=1}^{n}(P_{li}\Delta_{li}W)Q_{li}|=0$.

Now consider the case $i\ne j$. The Cauchy–Schwarz inequality implies

$$E\left|\sum_{l=1}^{n}Q_{lj}Q_{li}\right|\le\sqrt{E\sum_{l=1}^{n}Q_{lj}^2E\sum_{l=1}^{n}Q_{li}^2}. \tag{2.23}$$

We may write

$$\begin{aligned}
&\sum_{l=1}^{n}(P_{li}P_{lj})(\Delta_{li}W\Delta_{lj}W)\\
&\quad=\sum_{l=1}^{n}\frac{\Delta_{li}W\Delta_{lj}W}{(t_l-t_{l-1})^2}\int_{t_{l-1}+s_i}^{t_l+s_i}u_i(s)\,ds\int_{t_{l-1}+s_j}^{t_l+s_j}u_j(s)\,ds\\
&\quad=\sum_{l=1}^{n}\frac{\Delta_{li}W\Delta_{lj}W}{t_l-t_{l-1}}\int_{t_{l-1}+s_i}^{t_l+s_i}u_i^n(s)\hat{u}_j^n(s)\,ds,
\end{aligned}\tag{2.24}$$

where

$$\hat{u}_j^n(s)=\sum_{l=1}^{m}\frac{I_{(t_{l-1}+s_i,t_l+s_i]}(s)}{t_l-t_{l-1}}\int_{t_{l-1}+s_i}^{t_l+s_i}u_j(s'+s_j-s_i)\,ds'.\tag{2.25}$$

Similar to the case $i=j$, we have

$$\lim_{n\to\infty}E\left|\sum_{l=1}^{n}(P_{li}P_{lj})(\Delta_{li}W\Delta_{lj}W)\right|=0.\tag{2.26}$$

This completes the proof of the proposition. □

Suppose that

$$\bar{X}_{t_l}=X_{t_{l-1}}+\alpha_l\big(X_{t_l}-X_{t_{l-1}}\big)$$

for some random variables $0\le\alpha_l\le 1$, $l=1,\dots,n$. Denote

$$\Delta\big(\Pi(X_{t_l})\big)=\big(\Pi(\Delta X_{t_l})\big)=\Pi(X_{t_l})-\Pi(X_{t_{l-1}}),\tag{2.27}$$

$$\Pi(\bar{X}_{t_l})=\Pi(X_{t_{l-1}})+\alpha_l\Delta\Pi(X_{t_l}),\tag{2.28}$$

$$\Delta_{li}X=X(t_l+s_i)-X(t_{l-1}+s_i),\qquad\text{for } 1\le i\le k \text{ and } 1\le l\le n.\tag{2.29}$$

PROPOSITION 2.3. *Suppose that $\phi\in C^{1,2}(T\times\mathbf{R}^k,\mathbf{R})$, and let $1\le i,\ j\le k$. Under the hypotheses of Proposition 2.2, we have*

$$\begin{aligned}
&\sum_{l=1}^{n}\frac{\partial^2\phi}{\partial x_i\,\partial x_j}\big(t_{l-1},\Pi(\bar{X}_{t_l})\big)\Delta_{li}X\Delta_{lj}X\\
&\qquad\to\begin{cases}\displaystyle\int_0^{t+s_i}\frac{\partial^2\phi}{\partial x_i^2}\big(s,\Pi(X_s)\big)u^2(s)\,ds, & i=j,\\ 0, & i\ne j\end{cases}
\end{aligned}\tag{2.30}$$

as $n\to\infty$, in probability.

PROOF. For $1 \le i,\ j \le k$,

$$\begin{aligned}\Delta_{li} X \Delta_{lj} X &= (\Delta_{li} U + \Delta_{li} V)(\Delta_{lj} U + \Delta_{lj} V) \\ &= \Delta_{li} U \Delta_{lj} U + \Delta_{li} U \Delta_{lj} V \\ &\quad + \Delta_{li} V \Delta_{lj} U + \Delta_{li} V \Delta_{lj} V,\end{aligned} \tag{2.31}$$

where U, V are defined by (2.2). Since U, V are continuous and $V|[0,a]$ is of bounded variation, it follows that

$$\begin{aligned}&\lim_{n\to\infty} \sum_{l=1}^{n} \Delta_{li} U \Delta_{lj} V = 0, \\ &\lim_{n\to\infty} \sum_{l=1}^{n} \Delta_{li} V \Delta_{lj} U = 0, \\ &\lim_{n\to\infty} \sum_{l=1}^{n} \Delta_{li} V \Delta_{lj} V = 0,\end{aligned} \tag{2.32}$$

in probability, for all $0 \le i,\ j \le n$. To handle the term $\sum_{l=1}^{n} \Delta_{li} U \Delta_{lj} U$, we adapt an approach by Nualart and Pardoux (cf. [23], Theorem 3.4, or [22], Theorem 3.2.1).

Set

$$Y(s) := \frac{\partial^2 \phi}{\partial x_i^2}\big(s, \Pi(X_s)\big) I_{[0,t]}(s)$$

and

$$Y^n(s) := Y(0) I_{\{0\}}(s) + \sum_{l=1}^{n} \frac{\partial^2 \phi}{\partial x_i^2}\big(t_{l-1}, \Pi(\bar{X}_{t_l})\big) I_{(t_{l-1}, t_l]}(s). \tag{2.33}$$

Then $Y^n(s) \to Y(s)$ as $n \to \infty$, uniformly in $s \in [0,t]$. Applying Proposition 2.2 and Lemma A.3, we get

$$\begin{aligned}&\sum_{l=1}^{n} \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\big(t_{l-1}, \Pi(\bar{X}_{t_l})\big) \Delta_{li} X \Delta_{lj} X \\ &\qquad \to \delta_{ij} \int_0^{t+s_i} \frac{\partial^2 \phi}{\partial x_i^2}\big(s, \Pi(X_s)\big) u^2(s)\, ds\end{aligned} \tag{2.34}$$

in probability as $n \to \infty$. □

PROPOSITION 2.4. *Suppose that* $\phi \in C^{1,2}(T \times \mathbf{R}^k)$ *and let* $X(t)$ *be a continuous stochastic process defined by* (2.1), *where* $u \in \mathbb{L}^{2,4}_{\mathrm{loc}}$, $v \in \mathbb{L}^{1,4}_{\mathrm{loc}}$ *and* $\eta \in C([-\tau, 0], \mathbf{R}^m)$ *is of bounded variation. Assume that* $\pi_n : -\tau = s_0 < \cdots < s_n = 0$

are partitions of $[-\tau, 0]$ *such that* $|\pi_n| \to 0$ *as* $n \to \infty$. *Then, for each* $1 \le i \le k$ *and each* $t \in T$, *we have*

$$\begin{aligned}
\lim_{n\to\infty} \sum_{l=1}^{n} \frac{\partial \phi}{\partial x_i}\bigl(t_{l-1}, \Pi\bigl(X_{t_{l-1}}\bigr)\bigr)\Delta_{li} X
&= \int_0^t \frac{\partial \phi}{\partial x_i}\bigl(s, \Pi(X_s)\bigr)\, dX(s+s_i) \\
&\quad + \sum_{j=i+1}^{k} \int_0^t \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\bigl(s, \Pi(X_s)\bigr) u^2(s+s_i)\, ds \\
&\quad + \sum_{j=1}^{k} \int_0^t \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\bigl(s, \Pi(X_s)\bigr) \\
&\qquad \times \biggl[\int_0^{s+s_j} D_{s+s_i} u(r)\, dW(r) \\
&\qquad\qquad + \int_0^{s+s_j} D_{s+s_i} v(r)\, dr\biggr] u(s+s_i)\, ds
\end{aligned} \tag{2.35}$$

in probability.

PROOF. By a localization argument, we may assume that $\phi \in C_b^{1,2}(T \times \mathbf{R}^k, \mathbf{R})$. Let $|\pi_n| < \min_{\{1 \le i \le k\}} |s_i - s_{i-1}|$. Fix $1 \le i \le k$, $1 \le l \le n$, and set

$$F_l := \frac{\partial \phi}{\partial x_i}\bigl(t_{l-1}, \Pi\bigl(X_{t_{l-1}}\bigr)\bigr). \tag{2.36}$$

By property of the Skorohod integral ([23], Theorem 3.2), it follows that

$$F_l \Delta_{li} U = \int_{t_{l-1}+s_i}^{t_l+s_i} u(s) F_l\, dW(s) + \int_{t_{l-1}+s_i}^{t_l+s_i} D_r(F_l) u(r)\, dr, \tag{2.37}$$

where U is defined by (2.2). The chain rule (for weak derivatives) yields

$$D_r(F_l) = \sum_{j=1}^{k} \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\bigl(t_{l-1}, \Pi\bigl(X_{t_{l-1}}\bigr)\bigr) D_r X(t_{l-1} + s_j). \tag{2.38}$$

Now, taking the Malliavin derivative D_r in (2.1) gives

$$D_r X(t) = u(r) I_{\{r \le t\}} + \int_0^t D_r u(s)\, dW(s) + \int_0^t D_r v(s)\, ds. \tag{2.39}$$

Consequently,

$$\sum_{l=1}^{n} \frac{\partial \phi}{\partial x_i}\bigl(t_{l-1}, \Pi\bigl(X_{t_{l-1}}\bigr)\bigr)\Delta_{li} U = c_1 + c_2 + c_3 + c_4,$$

where

$$
\begin{aligned}
c_1 &:= \sum_{l=1}^{n} \int_{t_{l-1}+s_i}^{t_l+s_i} \frac{\partial \phi}{\partial x_i}\big(t_{l-1}, \Pi(X_{t_{l-1}})\big) u(s)\, dW(s), \\
c_2 &:= \sum_{l=1}^{n} \int_{t_{l-1}+s_i}^{t_l+s_i} \sum_{j=1}^{k} \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\big(t_{l-1}, \Pi(X_{t_{l-1}})\big) I_{\{r \le t_{l-1}+s_j\}} u^2(r)\, dr, \\
c_3 &:= \sum_{l=1}^{n} \int_{t_{l-1}+s_i}^{t_l+s_i} \sum_{j=1}^{k} \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\big(t_{l-1}, \Pi(X_{t_{l-1}})\big) \\
&\qquad \times \int_0^{t_{l-1}+s_j} D_r u(s)\, dW(s)\, u(r)\, dr, \\
c_4 &:= \sum_{l=1}^{n} \int_{t_{l-1}+s_i}^{t_l+s_i} \sum_{j=1}^{k} \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\big(t_{l-1}, \Pi(X_{t_{l-1}})\big) \\
&\qquad \times \int_0^{t_{l-1}+s_j} D_r v(s)\, ds\, u(r)\, dr.
\end{aligned}
\tag{2.40}
$$

We will study the limits of the above expressions as $n \to \infty$.

Step 1. First we show that the limit of c_2 is given by

$$
c_2 \to \sum_{j=i+1}^{k} \int_0^{t+s_i} \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\big(r - s_i, \Pi(X_{r-s_i})\big) u^2(r)\, dr \qquad \text{a.s.} \tag{2.41}
$$

If $j \le i$, then $t_{l-1} + s_i \ge t_{l-1} + s_j$. So when $t_{l-1} + s_i < r < t_l + s_i$, $I_{\{r \le t_{l-1}+s_j\}} = 0$. We have

$$
\begin{aligned}
c_2 &= \sum_{j=i+1}^{k} \sum_{l=1}^{n} \int_{t_{l-1}+s_i}^{t_l+s_i} \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\big(t_{l-1}, \Pi(X_{t_{l-1}})\big) I_{\{r \le t_{l-1}+s_j\}} u^2(r)\, dr \\
&\to \sum_{j=i+1}^{k} \int_0^{t+s_i} \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\big(r - s_i, \Pi(X_{r-s_i})\big) u^2(r)\, dr
\end{aligned}
$$

a.s. as $n \to \infty$.

Step 2. Next we study the limit of c_3 as $n \to \infty$. We claim that

$$
\begin{aligned}
c_3 \to \sum_{j=1}^{k} \int_0^{t+s_i} & \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\big(r - s_i, \Pi(X_{r-s_i})\big) \\
&\times \int_0^{r-s_i+s_j} D_r u(s)\, dW(s)\, u(r)\, dr
\end{aligned}
\tag{2.42}
$$

as $k \to \infty$ in probability. In fact,

$$
\begin{aligned}
T_j^n :=& \Biggl|\sum_{l=1}^n \int_{t_{l-1}+s_i}^{t_l+s_i} \biggl[\frac{\partial^2\phi}{\partial x_i\,\partial x_j}\bigl(t_{l-1},\Pi\bigl(X_{t_{l-1}}\bigr)\bigr)\int_0^{t_{l-1}+s_j} D_r u(s)\,dW(s) \\
&\qquad - \int_0^{t+s_i} \frac{\partial^2\phi}{\partial x_i\,\partial x_j}\bigl(r-s_i,\Pi\bigl(X_{r-s_i}\bigr)\bigr) \\
&\qquad\qquad \times \int_0^{r-s_i+s_j} D_r u(s)\,dW(s)\biggr] u(r)\,dr\Biggr| \\
\le& \Biggl|\sum_{l=1}^n \int_{t_{l-1}+s_i}^{t_l+s_i} \frac{\partial^2\phi}{\partial x_i\,\partial x_j}\bigl(t_{l-1},\Pi\bigl(X_{t_{l-1}}\bigr)\bigr)\int_{t_{l-1}+s_j}^{r+s_j-s_i} D_r u(s)\,dW(s)\,u(r)\,dr\Biggr| \\
&+ \Biggl|\sum_{l=1}^n \int_{t_{l-1}+s_i}^{t_l+s_i}\biggl[\frac{\partial^2\phi}{\partial x_i\,\partial x_j}\bigl(t_{l-1},\Pi\bigl(X_{t_{l-1}}\bigr)\bigr) - \frac{\partial^2\phi}{\partial x_i\,\partial x_j}\bigl(r-s_i,\Pi\bigl(X_{r-s_i}\bigr)\bigr)\biggr] \\
&\qquad\qquad \times \int_0^{r-s_i+s_j} D_r u(s)\,dW(s)\,u(r)\,dr\Biggr| \\
\le& \biggl\|\frac{\partial^2\phi}{\partial x_i\,\partial x_j}\biggr\|_\infty \sum_{l=1}^n \int_{t_{l-1}+s_i}^{t_l+s_i}\biggl|\int_{t_{l-1}+s_j}^{r+s_j-s_i} D_r u(s)\,dW(s)\biggr| |u(r)|\,dr \\
&+ \sup_{1\le l\le n}\ \sup_{r\in[t_{l-1}+s_i,t_l+s_i]}\biggl|\frac{\partial^2\phi}{\partial x_i\,\partial x_j}\bigl(t_{l-1},\Pi\bigl(X_{t_{l-1}}\bigr)\bigr) \\
&\qquad\qquad - \frac{\partial^2\phi}{\partial x_i\,\partial x_j}\bigl(r-s_i,\Pi\bigl(X_{r-s_i}\bigr)\bigr)\biggr| \\
&\qquad\qquad \times \int_0^{t_l+s_i}\biggl|\int_0^{r-s_i+s_j} D_r u(s)\,dW(s) u(r)\biggr|\,dr \\
=& T_{j1}^n + T_{j2}^n,
\end{aligned}
$$

where T_{j1}^n and T_{j2}^n denote the first and second term on the right-hand side of the last inequality. Using the Cauchy–Schwarz inequality and the L^p inequality for the Skorohod integral ([23], Proposition 3.5, and [22], page 158), we have

$$
\begin{aligned}
ET_{j1}^n \le& \biggl\|\frac{\partial^2\phi}{\partial x_i\,\partial x_j}\biggr\|_\infty \biggl(E\int_0^{a+s_i} u^2(r)\,dr\biggr)^{1/2} \\
&\times \biggl\{E\sum_{l=1}^n \int_{t_{l-1}+s_i}^{t_l+s_i}\int_{t_{l-1}+s_j}^{r+s_j-s_i} |D_r u(s)|^2\,ds\,dr \\
&\qquad + E\sum_{l=1}^n \int_{t_{l-1}+s_i}^{t_l+s_i}\int_{t_{l-1}+s_j}^{r+s_j-s_i}\int_0^a \bigl|D_\theta\bigl(D_r u(s)\bigr)\bigr|^2\,d\theta\,ds\,dr\biggr\}^{1/2} \to 0
\end{aligned}
$$

as $n \to \infty$. The uniform continuity of $\frac{\partial^2 \phi}{\partial x_i \, \partial x_j}$ implies $T^n_{j2} \to 0$ a.s. So as $n \to \infty$, $T^n_j \to 0$ in probability.

Step 3. Now we will show that

$$c_4 \to \sum_{j=1}^{k} \int_0^{t+s_i} \frac{\partial^2 \phi}{\partial x_i \, \partial x_j}\big(r - s_i, \Pi(X_{r-s_i})\big) \times \int_0^{r+s_j-s_i} D_r v(s)\, ds\, u(r)\, dr \qquad \text{a.s.} \tag{2.43}$$

As in Step 2, we have

$$\begin{aligned}
&\Bigg|\sum_{l=1}^{n} \int_{t_{l-1}+s_i}^{t_l+s_i} \Bigg[\frac{\partial^2 \phi}{\partial x_i \, \partial x_j}\big(t_{l-1}, \Pi(X_{t_{l-1}})\big) \int_0^{t_{l-1}+s_j} D_r v(s)\, ds \\
&\qquad - \frac{\partial^2 \phi}{\partial x_i \, \partial x_j}\big(r - s_i, \Pi(X_{r-s_i})\big) \int_0^{r-s_i+s_j} D_r v(s)\, ds\Bigg] u(r)\, dr\Bigg| \\
&\quad \le \left\| \frac{\partial^2 \phi}{\partial x_i \, \partial x_j} \right\|_\infty \sum_{l=1}^{n} \int_{t_{l-1}+s_i}^{t_l+s_i} \left| \int_{t_{l-1}+s_j}^{r+s_j-s_i} D_r v(s)\, ds \right| |u(r)|\, dr \\
&\qquad + \sup_{1 \le l \le n} \sup_{r \in [t_{l-1}+s_i, t_l+s_i]} \Bigg| \frac{\partial^2 \phi}{\partial x_i \, \partial x_j}\big(t_{l-1}, \Pi(X_{t_{l-1}})\big) \\
&\qquad\qquad - \frac{\partial^2 \phi}{\partial x_i \, \partial x_j}\big(r - s_i, \Pi(X_{r-s_i})\big)\Bigg| \\
&\qquad\qquad \times \int_0^{t_l+s_i} \left| \int_0^{r-s_i+s_j} D_r v(s)\, ds \right| |u(r)|\, dr \\
&\quad \to 0 \qquad \text{a.s. as } n \to \infty.
\end{aligned}$$

Step 4. Finally, we study the limit of c_1 as $n \to \infty$. We shall show that

$$c_1 \to \int_0^{t+s_i} \frac{\partial \phi}{\partial x_i}\big(s - s_i, \Pi(X_{s-s_i})\big) u(s)\, dW(s) \tag{2.44}$$

in $L^2(\Omega, \mathbf{R})$ as $n \to \infty$. To see this, define

$$u^n(s) := u(s) \sum_{l=1}^{n} \frac{\partial \phi}{\partial x_i}\big(t_{l-1}, \Pi(X_{t_{l-1}})\big) I_{(t_{l-1}+s_i, t_l+s_i]}(s). \tag{2.45}$$

It suffices to show that

$$u^n(s) \to \frac{\partial \phi}{\partial x_i}\big(s - s_i, \Pi(X_{t_{s-s_i}})\big) u(s) I_{(0,t+s_i]}(s) \tag{2.46}$$

in $\mathbb{L}^{1,2}$ as $n \to \infty$. It is clear that the sequence $\{u^n(s)\}$ converges to $\frac{\partial\phi}{\partial x_i}(s-s_i, \Pi(X_{s-s_i}))u(s)I_{(0,t+s_i]}(s)$ in $L^2(\Omega \times T, \mathbf{R})$. It remains to show that the sequence $\{D_r u^n(s)\}_{n=1}^{\infty}$, $r, s \in T$, converges in $L^2(\Omega \times T^2, \mathbf{R})$ to $D_r[\frac{\partial\phi}{\partial x_i}(s-s_i, \Pi(X_{s-s_i}))u(s)I_{(0,t+s_i]}(s)]$. Now

$$
\begin{aligned}
D_r u^n(s) &= D_r u(s) \sum_{l=1}^{n} \frac{\partial\phi}{\partial x_i}\big(t_{l-1}, \Pi\big(X_{t_{l-1}}\big)\big) I_{(t_{l-1}+s_i, t_l+s_i]}(s) \\
&\quad + u(s) \sum_{l=1}^{n} \left[\sum_{j=1}^{k} \frac{\partial^2\phi}{\partial x_i\, \partial x_j}\big(t_{l-1}, \Pi\big(X_{t_{l-1}}\big)\big) \int_0^{t_{l-1}+s_j} D_r u(s')\, dW(s') \right] \\
&\qquad \times I_{(t_{l-1}+s_i, t_l+s_i]}(s) \\
&\quad + u(s) \sum_{l=1}^{n} \left[\sum_{j=1}^{k} \frac{\partial^2\phi}{\partial x_i\, \partial x_j}\big(t_{l-1}, \Pi\big(X_{t_{l-1}}\big)\big) \int_0^{t_{l-1}+s_j} D_r v(s')\, ds' \right] \\
&\qquad \times I_{(t_{l-1}+s_i, t_l+s_i]}(s) \\
&\quad + u(s) \sum_{l=1}^{n} \left[\sum_{j=1}^{k} \frac{\partial^2\phi}{\partial x_i\, \partial x_j}\big(t_{l-1}, \Pi\big(X_{t_{l-1}}\big)\big) u(r) I_{[0, t_{l-1}+s_j]}(r) \right] \\
&\qquad \times I_{(t_{l-1}+s_i, t_l+s_i]}(s) \\
&= d_1 + d_2 + d_3 + d_4,
\end{aligned}
$$

where d_1, d_2, d_3 and d_4 stand for the first, second, third and fourth terms on the right-hand side of the above equality, respectively. It is easy to see that

$$
d_1 \to D_r u(s) \frac{\partial\phi}{\partial x_i}\big(\Pi\big(s-s_i, X_{s-s_i}\big)\big) I_{(0,t+s_i]}(s)
$$

in $L^2(\Omega, \mathbf{R})$. Since for all $1 \le j \le k$, $u(s)\int_0^{s+s_j-s_i} D_r v(\theta)\, d\theta$ belongs to $L^2(\Omega \times T^2, \mathbf{R})$, then by Lebesgue's dominated convergence theorem, the $L^2(\Omega \times T^2, \mathbf{R})$ limit of the function $q_3(s, r)$ defined by

$$
q_3 := u(s) \sum_{j=1}^{k} \sum_{l=1}^{n} \left[\frac{\partial^2\phi}{\partial x_i\, \partial x_j}\big(t_{l-1}, \Pi\big(X_{t_{l-1}}\big)\big) \int_0^{s+s_j-s_i} D_r v(\theta)\, d\theta \right] I_{(t_{l-1}+s_i, t_l+s_i]}(s)
$$

is

$$
\sum_{j=1}^{k} u(s) \left[\frac{\partial^2\phi}{\partial x_i\, \partial x_j}\big(s-s_i, \Pi\big(X_{s-s_i}\big)\big) \int_0^{s+s_j-s_i} D_r v(\theta)\, d\theta \right] I_{(0,t+s_i]}(s).
$$

Since $v \in \mathbb{L}^{1,4}$ and $u \in L^4(\Omega \times T, \mathbf{R})$, the following argument shows that the difference between d_3 and q_3 converges to 0 as $n \to \infty$ in $L^1(\Omega, \mathbf{R})$:

$$\sum_{l=1}^{n} \int_{t_{l-1}+s_i}^{t_l+s_i} \int_0^a u^2(s)\left[\frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\big(t_{l-1}, \Pi(X_{t_{l-1}})\big)\right]^2 \left[\int_{t_{l-1}+s_j}^{s+s_j-s_i} D_r v(\theta)\, d\theta\right]^2 dr\, ds$$

$$\le |\pi_n| \left\| \frac{\partial^2 \phi}{\partial x_i\, \partial x_j} \right\|_\infty^2 \int_0^a u^2(s)\, ds \int_0^a \int_0^a \big(D_r v(\theta)\big)^2\, dr\, d\theta$$

$$\to 0.$$

Hence, the $L^2(\Omega \times T^2, \mathbf{R})$ limit of d_3 is the same as that of q_3, namely,

$$\sum_{j=1}^{k} u(s)\left[\frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\big(s - s_i, \Pi(X_{s-s_i})\big) \int_0^{s+s_j-s_i} D_r v(\theta)\, d\theta\right] I_{(0,t+s_i]}(s)$$

in $L^2(\Omega \times T^2, \mathbf{R})$. To find the limit of d_2, we need to check that for all j, the two-parameter process $(u(s)\int_0^{s+s_j-s_i} D_r u(\theta)\, dW(\theta)\, , 0 \le s, r \le a)$ belongs to $L^2(\Omega \times T^2, \mathbf{R})$. This follows from the following estimates:

$$E\int_0^a \int_0^a u^2(s)\left[\int_0^{s+s_j-s_i} D_r u(\theta)\, dW(\theta)\right]^2 ds\, dr$$

$$\le \left\{ E\int_0^a u^4(s)\, ds\, E\int_0^a \left(\int_0^a \left[\int_0^{s+s_j-s_i} D_r u(\theta)\, dW(\theta)\right]^2 dr\right)^2 ds \right\}^{1/2}$$

$$\le C\left\{ E\int_0^a u^4(s)\, ds \left[E\left(\int_0^a \int_0^a |D_r u(\theta)|^2\, d\theta\, dr\right)^2 \right.\right.$$

$$\left.\left. + E\left(\int_0^a \int_0^a \int_0^a D_\alpha\big(D_r u(\theta)\big)\, d\theta\, dr\, d\alpha\right)^2 \right]\right\}^{1/2}.$$

Here we have used a slight modification of the L^p estimate of the Skorohod integral for $p = 4$ (cf. [23], Exercise 3.2.7). Using similar L^p estimates to the above, we obtain

$$\begin{aligned} &\sum_{l=1}^{n} \int_{t_{l-1}+s_i}^{t_l+s_i} \int_0^a u^2(s)\left[\frac{\partial^2 \phi}{\partial x_i\, \partial x_j}\big(t_{l-1}, \Pi(X_{t_{l-1}})\big)\right]^2 \\ &\qquad \times \left[\int_{t_{l-1}+s_j}^{s+s_j-s_i} D_r u(\theta)\, dW(\theta)\right]^2 dr\, ds \\ &\quad \le \left\| \frac{\partial^2 \phi}{\partial x_i\, \partial x_j} \right\|_\infty^2 \left(\int_0^a E u^4(s)\, ds\right)^{1/2} \\ &\qquad \times \left\{ \sum_{l=1}^{n} E \int_{t_{l-1}+s_i}^{t_l+s_i} \left[\int_0^a \left(\int_{t_{l-1}+s_j}^{s+s_j-s_i} D_r v(\theta)\, d\theta\right)^2 dr\right]^2 ds \right\}^{1/2}. \end{aligned} \tag{2.47}$$

Note that the right-hand side of the above inequality tends to zero as $n \to \infty$. Thus

$$d_2 \to \sum_{j=1}^{k} u(s)\biggl[\frac{\partial^2 \phi}{\partial x_i\, \partial x_j}(s - s_i, \Pi(X_{s-s_i})) \times \int_0^{s+s_j-s_i} D_r u(\theta)\, dW(\theta)\biggr] I_{(0,t+s_i]}(s) \tag{2.48}$$

in $L^2(\Omega \times T^2, \mathbf{R})$ as $n \to \infty$.

It is easy to check that

$$d_4 \to \sum_{j=1}^{k} u(s) \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}(s - s_i, \Pi(X_{s-s_i}))u(r) I_{[0,s+s_j-s_i]}(r) I_{(0,t+s_i]}(s)$$

as $n \to \infty$ in $L^2(\Omega, \mathbf{R})$. Therefore,

$$D_r u^n(s) \to D_r\biggl[u(s) \frac{\partial^2 \phi}{\partial x_i\, \partial x_j}(s - s_i, \Pi(X_{s-s_i})) I_{(0,t+s_i]}(s)\biggr] \tag{2.49}$$

in $L^2(\Omega \times T^2, \mathbf{R})$. Finally it is easy to see that

$$c_1 \to \int_0^{t+s_i} \frac{\partial \phi}{\partial x_i}(s - s_i, \Pi(X_{s-s_i}))u(s)\, dW(s)$$

in $L^2(\Omega, \mathbf{R})$ as $n \to \infty$.

Step 5. The convergence

$$\sum_{l=1}^{n} \frac{\partial \phi}{\partial x_i}(t_{l-1}, \Pi(X_{t_{l-1}}))\Delta_{li} V \to \int_{s_i}^{t+s_i} \frac{\partial \phi}{\partial x_i}(s - s_i, \Pi(X_{s-s_i}))\, dV(s) \qquad \text{a.s.} \tag{2.50}$$

as $n \to \infty$, is easy to verify. □

We complete the section by giving a Stratonovich version of the Itô formula (2.7).

Suppose that $k \geq 1$ and $p \geq 2$. The set $\mathbb{L}_{d,C}^{k,p}$ (cf. [23], Definition 7.2, and [22], page 167) is the class of processes $u \in L_d^{k,p}$ such that the mappings $s \hookrightarrow D_{s\wedge t} u(s \vee t)$ and $s \hookrightarrow D_{s\vee t} u(s \wedge t)$ are continuous in $L^p(\Omega)$, uniformly in $t \in T$, and $\sup_{s,t\in T} E(|D_s u(t)|^p) < \infty$.

The space $\mathbb{L}^{1,2}_{d,C,\mathrm{loc}}$ is the class of processes that are locally in $\mathbb{L}^{1,2}_{d,C}$. For any $u \in \mathbb{L}^{1,2}_{d,C}$, the following limits,

$$D_t^+ u(t) = \lim_{\varepsilon \downarrow 0} \sum_{i=1}^d D_t^i u^i(t+\varepsilon), \tag{2.51}$$
$$D_t^- u(t) = \lim_{\varepsilon \downarrow 0} \sum_{i=1}^d D_t^i u^i(t-\varepsilon),$$

exist in $L^2(\Omega)$ uniformly in t, we set $\triangledown = D^+ + D^-$, that is, $(\triangledown u)(t) = D_t^+ u(t) + D_t^- u(t)$.

Consider the process

$$X(t) = \begin{cases} \eta(0) + \int_0^t u(s) \circ dW(s) + \int_0^t v(s)\, ds, & t > 0, \\ \eta(t), & -\tau \le t \le 0, \end{cases} \tag{2.52}$$

where η belongs to C and is of bounded variation, $u = (u^1, \ldots, u^m)^T$, $u^i \in \mathbb{L}^{2,4}_{d,C,\mathrm{loc}}$, $(\triangledown u) \in \mathbb{L}^{1,4}_{\mathrm{loc}}$, $v = (v^1, \ldots, v^m)^T$, $v^i \in \mathbb{L}^{1,4}_{\mathrm{loc}}$, and the stochastic integral is a Stratonovich one. Assume also that the process X is continuous.

Using the relationship between the Skorohod and Stratonovich integrals ([23], Theorem 7.3, and [22], Theorem 3.11) and Theorem 2.3, we can easily obtain the following Stratonovich version of Itô's formula for the segment process X_t (cf. [28]).

COROLLARY 2.5. *Suppose that the process $X(t)$ is defined by* (2.52), *and let $\phi \in C^{1,2}(T \times \mathbf{R}^{mk}, \mathbf{R})$. Then*

$$\begin{aligned} &\phi(t, \Pi(X_t)) - \phi(0, \Pi(X_0)) \\ &\quad = \int_0^t \frac{\partial \phi}{\partial s}(s, \Pi(X_s))\, ds \\ &\qquad + \sum_{i=1}^k \int_0^t \frac{\partial \phi}{\partial \vec{x}^i}(s, \Pi(X_s)) u(s+s_i) \circ dW(s+s_i) \\ &\qquad + \sum_{i=1}^k \int_0^t \frac{\partial \phi}{\partial \vec{x}^i}(s, \Pi(X_s)) v(s+s_i)\, ds \end{aligned} \tag{2.53}$$

for all $t \in T$ a.s.

3. Weak differentiability of solutions of SDDEs. In this section, we will study the weak differentiability of the solution of the Itô SDDE (1.6). Bell and Mohammed [6] have applied the Malliavin calculus to study regularity of solutions of SDDEs with a single delay in the noise term. Their analysis relies on weak differentiability of the solution of the SDDE. In Section 5 of this article, the weak differentiability of the solution to the SDDE (1.6) together with the Itô formula (2.10) are used to develop higher-order numerical schemes for solving the SDDE. The next three results (Proposition 3.1, Lemma 3.2 and Proposition 3.3) are analogous to those in [22], Theorem 2.2.1, Lemma 2.2.2 and Theorem 2.2.2. Denote $\mathbb{D}_m^{k,\infty} := \bigcap_{p\geq 2} \mathbb{D}_m^{k,p}$, for $k \in N$. Recall that D_r^l, $1 \leq l \leq d$, stand for weak differentiation with respect to the lth component of W.

PROPOSITION 3.1 (cf. [22], Proposition 1.2.3). *In the Itô SDDE* (1.6), *assume that* $g \in C_b^{0,1}(T \times \mathbf{R}^{k_1 m}, L(\mathbf{R}^d, \mathbf{R}^m))$ *and* $h \in C_b^{0,1}(T \times \mathbf{R}^{k_2 m}, \mathbf{R}^m)$. *Let* X *be the solution of* (1.6). *Then* $X(t) \in \mathbb{D}_m^{1,\infty}$ *for all* $t \in T$, *and*

$$\sup_{0\leq r\leq a} E\Big(\sup_{r\leq s\leq a} |D_r X(s)|^p\Big) < \infty \tag{3.1}$$

for all $p \geq 2$. *Furthermore, the "partial" weak derivatives* $D_r^l X^j(t)$ *with respect to the* l*th coordinate of* W *satisfy the following linear SDDEs a.s.*:

$$D_r^l X^j(t) = \begin{cases} g^{jl}\big(r, \Pi_1(X_r^j)\big) \\ \quad + \displaystyle\int_r^t \sum_{i=1}^{k_1} \frac{\partial g^{jl}}{\partial \vec{x}_i}\big(s, \Pi_1(X_s)\big) D_r^l X^j(s + s_{1,i})\, dW^l(s) \\ \quad + \displaystyle\int_0^t \sum_{i=1}^{k_2} \frac{\partial h^j}{\partial \vec{x}_i}\big(s, \Pi_2(X_s)\big) D_r^l X^j(s + s_{2,i})\, ds, & t \geq r, \\ 0, & t < r, \end{cases} \tag{3.2}$$

for $l = 1, \ldots, d$, $j = 1, \ldots, m$. *In* (3.2), g^{jl} *is the* (j, l) *entry of the* $m \times d$ *matrix* g, *and* h^j *is the* j*th coordinate of* h.

PROOF. For simplicity, we will only consider the one-dimensional case $d = m = 1$

$$X^0(t) = \begin{cases} \eta(0), & t \geq 0, \\ \eta(t), & -\tau \leq t < 0, \end{cases}$$

$$X^{n+1}(t) = \eta(0) + \int_0^t g\big(s, \Pi_1(X_s^n)\big)\, dW(s) + \int_0^t h\big(s, \Pi_2(X_s^n)\big)\, ds. \tag{3.3}$$

It is easy to see that

$$\begin{aligned} &D_r\Big(\int_0^t g\big(s, \Pi_1(X_s^n)\big)\, dW(s)\Big) \\ &\quad = g\big(r, \Pi_1(X_r^n)\big) + \int_{r-s_{1,k_1}}^t D_r\big(g\big(s, \Pi_1(X_s^n)\big)\big)\, dW(s) \end{aligned} \tag{3.4}$$

and

$$D_r\left(\int_0^t h\big(s,\Pi_2(X_s^n)\big)\,ds\right)=\int_{r-s_{2,k_2}}^t D_r\big(h\big(s,\Pi_2(X_s^n)\big)\big)\,ds. \tag{3.5}$$

Since g and h have bounded space derivatives, it is easy to see that there is a positive constant K such that

$$\begin{aligned}\big\|D_r\big(g\big(s,\Pi_1(X_s^n)\big)\big)\big\| &\le K\sup_{r\le u\le s}|D_rX^n(u)|,\\ \big|D_r\big(h\big(s,\Pi_2(X_s^n)\big)\big)\big| &\le K\sup_{r\le u\le s}|D_rX^n(u)|,\end{aligned} \tag{3.6}$$

almost surely. From the Burkholder–Davis–Gundy inequality and (3.3)–(3.6), it follows that $X^n(t)\in\mathbb{D}^{1,\infty}$ for all $t\in[0,a]$, and there are positive constants C_1, C_2 such that

$$\begin{aligned}&E\Big(\sup_{r\le u\le t}|D_rX^{n+1}(u)|^p\Big)\\ &\quad\le C_1\big(1+E\|X_r^n\|_C^p\big)+C_2\int_r^t E\Big(\sup_{r\le u\le s}|D_rX^n(u)|^p\Big)\,ds.\end{aligned} \tag{3.7}$$

By induction on n, the above inequality implies that $E(\sup_{r\le s\le a}|D_rX^n(s)|^p)$ are uniformly bounded in n for all $p\ge 2$. By [22], Proposition 1.5.5, it follows that $X(t)\in\mathbb{D}^{1,\infty}$ for all t. Applying the operator D to (1.6) (and using [22], Proposition 1.2.3), we obtain the linear SDDE (3.2) for the weak derivative of $X(t)$. The estimate (3.1) follows from (3.2), Burkholder–Davis–Gundy's inequality and Gronwall's lemma. $\square$

The following lemma may be proved using similar ideas. Its proof is left to the reader.

LEMMA 3.2. *Suppose that the real-valued process $\alpha=\{\alpha(r,t):t\in[r,a]\}$ is adapted and continuous. Assume that the processes $a(t)=(a_1(t),\dots,a_{k_1}(t))\in\mathbf{R}^{k_1}$ and $b(t)=(b_1(t),\dots,b_{k_2}(t))\in\mathbf{R}^{k_2}$ are adapted, continuous and uniformly bounded. Furthermore, suppose that the random variables $\alpha(r,t)$, $a(t)$ and $b(t)$ belong to $\mathbb{D}^{1,\infty}$ and satisfy the conditions*

$$\begin{aligned}&\sup_{0\le r\le a}E\Big(\sup_{r\le t\le a}|\alpha(r,t)|^p\Big)+\sup_{0\le r,s\le a}E\Big(\sup_{s\le t\le a}|D_s\alpha(r,t)|^p\Big)<\infty,\\ &\sup_{0\le s\le a}\Big\{E\Big(\sup_{s\le t\le a}|a(t)|^p\Big)+E\Big(\sup_{s\le t\le a}|D_sa(t)|^p\Big)\Big\}<\infty,\\ &\sup_{0\le s\le a}\Big\{E\Big(\sup_{s\le t\le a}|b(t)|^p\Big)+E\Big(\sup_{s\le t\le a}|D_sb(t)|^p\Big)\Big\}<\infty\end{aligned} \tag{3.8}$$

for all $p \geq 2$. *Let* $Y = \{Y(t) : t \in [0, a]\}$ *be the solution of the linear SDDE*

$$
(3.9) \quad Y(t) = \begin{cases} \alpha(r,t) + \int_r^t \langle a(s), \Pi_1(Y_s)\rangle_{\mathbf{R}^{k_1}} \, dW(s) + \int_r^t \langle b(s), \Pi_2(Y_s)\rangle_{\mathbf{R}^{k_2}} \, ds, & t \geq r, \\ 0, \quad 0 \leq t \leq r. \end{cases}
$$

Then $Y(t)$ *belongs to* $\mathbb{D}^{1,\infty}$, *and for all integers* $p \geq 2$, *we have*

$$
(3.10) \quad \begin{aligned} \sup_{0 \leq s \leq a} E\Big(\sup_{s \leq t \leq a} |D_s Y(t)|^p\Big) &< \infty, \\ \sup_{0 \leq s \leq a} E\Big(\sup_{s \leq t \leq a} |Y(t)|^p\Big) &< \infty. \end{aligned}
$$

Furthermore, *the weak derivative* $D_s Y(t)$ *of* $Y(t)$ *satisfies the linear SDDE*

$$
(3.11) \quad \begin{aligned} D_s Y(t) &= D_s \alpha(r,t) + \langle a(s), \Pi_1(Y_s)\rangle_{\mathbf{R}^{k_1}} I_{\{r \leq s \leq t\}} \\ &\quad + \int_r^t \big[\langle D_s a(v), \Pi_1(Y_v)\rangle_{\mathbf{R}^{k_1}} + \langle a(v), \Pi_1(D_s Y_v)\rangle_{\mathbf{R}^{k_1}}\big] \, dW(v) \\ &\quad + \int_r^t \big[\langle D_s b(v), \Pi_2(Y_v)\rangle_{\mathbf{R}^{k_2}} + \langle b(v), \Pi_2(D_s Y_v)\rangle_{\mathbf{R}^{k_2}}\big] \, dv, \qquad s < t. \end{aligned}
$$

The next proposition follows from Proposition 3.1 and Lemma 3.2.

PROPOSITION 3.3. *Let* $X = \{X(t) : t \in T = [0, a]\}$ *be the solution of the SDDE* (1.6), *where* $g \in C_b^{0,2}(T \times \mathbf{R}^{k_1 m}, L(\mathbf{R}^d, \mathbf{R}^m))$, $h \in C_b^{0,2}(T \times \mathbf{R}^{k_2 m}, \mathbf{R}^m)$ *have bounded first and second partial derivatives in the space variables*. *Then* $X(t) \in \mathbb{D}_m^{2,\infty}$ *for all* $t \in T$, *and*

$$
(3.12) \quad \sup_{0 \leq r_1, r_2 \leq a} E\Big(\sup_{r_1 \vee r_2 \leq s \leq a} \big|D_{r_1}^{l_1} D_{r_2}^{l_2} X(s)\big|^p\Big) < \infty
$$

for $l_1, l_2 = 1, \ldots, d$, *and all* $p \geq 2$.

4. Strong approximation of multiple Stratonovich integrals. The following iterated Stratonovich integrals are used in the Milstein scheme for the SDDE (1.6):

$$
(4.1) \quad J_{i,j}(t_0, t_1; -b) := \int_{t_0+b}^{t_1+b} \int_{t_0}^{s-b} \circ\, dW^i(v) \circ dW^j(s),
$$

where $0 < t_0 < t_1$, $b \geq 0$.

We will adopt the discretization scheme in [17], Section 5.8, in order to handle the above double stochastic integral. For alternative discretization approaches to iterated stochastic integrals, see [11] and [26].

Set

$$J(t_0, t_1; -b) := J_{1,1}(t_0, t_1; -b), \tag{4.2}$$

$t := t_1 - t_0$ and $r := 2\pi/t$. We choose a complete orthonormal basis of $L^2[0, t]$ as

$$\left\{\frac{1}{\sqrt{t}}, \sqrt{\frac{2}{t}} \sin nrs, \sqrt{\frac{2}{t}} \cos nrs : n = 1, 2, \ldots, 0 \le s \le t\right\}. \tag{4.3}$$

Set $\bar{W}^i(s) := W^i(s + t_0) - W^i(t_0)$ and $\bar{B}^j(s) := \bar{W}^j(s + b) - \bar{W}^j(b)$, $s \ge 0$, $1 \le i, j \le d$. Using the Kahunen–Loève expansion technique, we have

$$\bar{W}^i(s) - \frac{s}{t}\bar{W}^i(t) = \frac{a_0^i(t)}{2} + \sum_{n=1}^{\infty} [a_n^i(t) \cos nrs + b_n^i(t) \sin nrs] \tag{4.4}$$

and

$$\bar{B}^j(s) - \frac{s}{t}\bar{B}^j(t) = \frac{a_0^{j,b}(t)}{2} + \sum_{n=1}^{\infty} [a_n^{j,b}(t) \cos nrs + b_n^{j,b}(t) \sin nrs] \tag{4.5}$$

where

$$\begin{aligned} a_n^i(t) &= \frac{2}{t} \int_0^t \left(\bar{W}^i(s) - \frac{s}{t}\bar{W}^i(t)\right) \cos nrs\, ds, \\ b_n^i(t) &= \frac{2}{t} \int_0^t \left(\bar{W}^i(s) - \frac{s}{t}\bar{W}^i(t)\right) \sin nrs\, ds \end{aligned} \tag{4.6}$$

and

$$\begin{aligned} a_n^{j,b}(t) &= \frac{2}{t} \int_0^t \left(\bar{B}^j(s) - \frac{s}{t}\bar{B}^j(t)\right) \cos nrs\, ds, \\ b_n^{j,b}(t) &= \frac{2}{t} \int_0^t \left(\bar{B}^j(s) - \frac{s}{t}\bar{B}^j(t)\right) \sin nrs\, ds \end{aligned} \tag{4.7}$$

for $n \ge 1$. The convergences in (4.4) and (4.5) are in $L^2(\Omega \times [0, t])$. It is easy to see that if $n \ge 1$, $a_n^i(t)$, $b_n^i(t)$, $a_n^{j,b}(t)$ and $b_n^{j,b}(t)$ are normally distributed with mean 0 and variance $t/2\pi^2 n^2$ ([17], page 198). Furthermore, $\{a_n^i(t), b_n^i(t)\}$ and $\{a_n^{j,b}(t), b_n^{j,b}(t)\}$ are pairwise independent ([17], page 198). One can use well-known random number generators to simulate these random coefficients (cf. [12], Section 3.1.2, [17], Section 1.3, and [18], Section 1.2).

LEMMA 4.1. *Let $t_0, t \ge 0$. Then*

$$\begin{aligned} &J_{i,j}(t_0, t_0 + t; -b) \\ &\quad = \tfrac{1}{2}(\bar{W}^i(t)\bar{B}^j(t)) - \tfrac{1}{2}(\bar{W}^i(t)a_0^{j,b}(t_0) - \bar{B}^j(t)a_0^i(t_0)) \\ &\qquad + \pi \sum_{n=1}^{\infty} n[a_n^i(t_0)b_n^{j,b}(t_0) - b_n^i(t_0)a_n^{j,b}(t_0)], \qquad 1 \le i, j \le d, \end{aligned} \tag{4.8}$$

where the infinite series converges in $L^2(\Omega, \mathbf{R})$.

PROOF. It suffices to show (4.8) for $t_0 = 0$. Fix $t > 0$. For simplicity of notation, we write

$$a_n^j := a_n^j(0), \qquad b_n^j := b_n^j(0), \qquad a_n^{j,b} := a_n^{j,b}(0), \qquad b_n^{j,b} := b_n^{j,b}(0) \tag{4.9}$$

and

$$W_N^i(s) := \frac{s}{t} W^i(t) + \frac{a_0^i}{2} + \sum_{n=1}^{N} (a_n^i \cos nrs + b_n^i \sin nrs). \tag{4.10}$$

It is easy to check that

$$\int_b^{t+b} \int_0^{s-b} \circ\, dW_N^i(v) \circ dW^j(s) \to \int_b^{t+b} \int_0^{s-b} \circ\, dW^i(v) \circ dW^j(s) \tag{4.11}$$

in $L^2(\Omega)$ as $N \to \infty$. Then we may write

$$\begin{aligned} J_{i,j}(0, t; -b) &= \int_b^{t+b} W^i(s-b) \circ dW^j(s) \\ &= \int_b^{t+b} \frac{(s-b)}{t} W^i(t) \circ dW^j(s) + \frac{a_0^i}{2} \bar{B}^j(t) \\ &\quad + \sum_{n=1}^{\infty} \biggl[a_n^i \int_b^{t+b} \cos nr(s-b)\, dW^j(s) \\ &\qquad + b_n^i \int_b^{t+b} \sin nr(s-b)\, dW^j(s) \biggr]. \end{aligned}$$

For any $n \geq 1$, we have

$$\begin{aligned} &\int_b^{t+b} \cos nr(s-b)\, dW^j(s) \\ &\quad = \int_0^t \cos nrs\, d\bar{B}^j(s) \\ &\quad = \int_0^t \cos nrs\, d\biggl(\bar{B}^j(s) - \frac{s}{t}\bar{B}^j(t)\biggr) + \int_0^t \cos nrs\, d\biggl(\frac{s}{t}\bar{B}^j(t)\biggr) \\ &\quad = \cos nrs \biggl(\bar{B}^j(s) - \frac{s}{t}\bar{B}^j(t)\biggr)\bigg|_0^t \\ &\qquad + nr \int_0^t \biggl(\bar{B}^j(s) - \frac{s}{t}\bar{B}^j(t)\biggr) \sin nrs\, ds + \frac{\bar{B}^j(t)}{t} \int_0^t \cos nrs\, ds \\ &\quad = \frac{t}{2} nr b_n^{j,b}. \end{aligned}$$

Similarly, we have

$$\int_b^{t+b} \sin nr(s-b)\, dW^j(s) = -\frac{t}{2} nr a_n^{j,b}. \tag{4.12}$$

So

$$\begin{aligned} J_{i,j}(0,t;-b) &= \frac{W^i(t)}{t}\int_0^t s\, d\bar{B}^j(s) \\ &\quad + \frac{a_0^i}{2}\bar{B}^j(t) + \pi\sum_{n=1}^{\infty} n(a_n^i b_n^{j,b} - b_n^i a_n^{j,b}). \end{aligned} \tag{4.13}$$

Now,

$$\begin{aligned} \int_0^t s\, d\bar{B}^j(s) &= t\bar{B}^j(t) - \int_0^t \bar{B}^j(s)\, ds \\ &= \frac{t}{2}\bar{B}^j(t) - \int_0^t \left(\bar{B}^j(s) - \frac{s}{t}\bar{B}^j(t)\right) ds \\ &= \frac{t}{2}\big(\bar{B}^j(t) - a_0^{j,b}\big). \end{aligned}$$

Therefore,

$$\begin{aligned} J_{i,j}(0,t;-b) &= \tfrac{1}{2}W^i(t)\bar{B}^j(t) - \tfrac{1}{2}\big(W^i(t)a_0^{j,b} - \bar{B}^j(t)a_0^i\big) \\ &\quad + \pi\sum_{n=1}^{\infty} n(a_n^i b_n^{j,b} - b_n^i a_n^{j,b}). \end{aligned} \tag{4.14}$$

□

The expansion of $J_{i,j}(0,t;-b)$ is a generalization of the expansion of

$$\begin{aligned} &\int_0^t\int_0^s \circ\, dW^i(v) \circ dW^j(s) \\ &\quad = \tfrac{1}{2}\big(W^i(t)W^j(t)\big) - \tfrac{1}{2}\big[W^i(t)a_0^{j,b} - W^j(t)a_0^i\big] \\ &\quad\quad + \pi\sum_{n=1}^{\infty} n(a_n^i b_n^j - b_n^i a_n^j) \end{aligned} \tag{4.15}$$

(see [11, 17] and [18]). Set

$$\begin{aligned} &J_{i,j}^p(t_0, t_0+t; -b) \\ &\quad := \tfrac{1}{2}\big(\bar{W}^i(t)\bar{B}^j(t)\big) - \tfrac{1}{2}\big[\bar{W}^i(t)a_0^{j,b}(t_0) - \bar{B}^j(t)a_0^i(t_0)\big] \\ &\quad\quad + \pi\sum_{n=1}^{p} n\big[a_n^i(t_0)b_n^{j,b}(t_0) - b_n^i(t_0)a_n^{j,b}(t_0)\big]. \end{aligned} \tag{4.16}$$

Then $J_{i,j}^p(t_0, t_0+t; -b)$ can be used to approximate $J_{i,j}(t_0, t_0+t; -b)$ in the mean square. The rate of convergence is given in Lemma 4.2.

LEMMA 4.2. *For any integer* $p \geq 1$ *and* $t > 0$, *we have*

$$E\left|J_{i,j}^{p}(0,t;-b) - J_{i,j}(0,t;-b)\right|^2 \leq \frac{t^2}{2\pi^2 p}. \tag{4.17}$$

PROOF. Let $p \geq 1$ be any integer. Then

$$\sum_{n=p+1}^{\infty} \frac{1}{n^2} \leq \int_p^{\infty} \frac{1}{u^2}\,du = \frac{1}{p}. \tag{4.18}$$

Since a_n^i and b_n^i are independent, $E(a_n^i b_n^i) = 0$ and $E(a_n^{j,b} b_n^{j,b}) = 0$, we have

$$\begin{aligned}
E&\left|J_{i,j}^{p}(0,t;-b) - J_{i,j}(0,t;-b)\right|^2 \\
&= \pi^2 \sum_{n=p+1}^{\infty} n^2 E(a_n^i b_n^{j,b} - b_n^i a_n^{j,b})^2 \\
&= \pi^2 \sum_{n=p+1}^{\infty} n^2 \bigl[E(a_n^i b_n^{j,b})^2 + E(b_n^i a_n^{j,b})^2\bigr] \\
&= \frac{t^2}{2\pi^2} \sum_{n=p+1}^{\infty} \frac{1}{n^2} \\
&\leq \frac{t^2}{2\pi^2 p}.
\end{aligned}$$

□

5. The strong Milstein scheme. In this section we construct a strong Milstein scheme of order 1 for the SDDE (1.6). Our construction relies heavily on the Itô formula for "tame" functions (Theorem 2.1).

Throughout this section, we assume that in (1.6) the coefficients $g \in C^{1,2}(T \times \mathbf{R}^{k_1 m}, L(\mathbf{R}^d, \mathbf{R}^m))$ and $h \in C^{1,2}(T \times \mathbf{R}^{k_2 m}, \mathbf{R}^m)$. For convenience, set $W(s) = W(0) = 0$, for all $s \leq 0$. We also define

$$\begin{aligned}
u(t) &:= \begin{cases} g\bigl(t, \Pi_1(X_t)\bigr), & 0 \leq t \leq a, \\ 0, & t < 0, \end{cases} \\
v(t) &:= \begin{cases} h\bigl(t, \Pi_2(X_t)\bigr), & 0 \leq t \leq a, \\ \eta(t), & t < 0. \end{cases}
\end{aligned} \tag{5.1}$$

We first derive the Milstein scheme for the case $d = m = 1$.

5.1. *Itô–Taylor expansion.* Assume that $0 < t_0 < t$, and $\vec{x} = (x_1, \dots, x_{k_1}) \in \mathbf{R}^{k_1}$. Applying the Itô formula (2.10), we have

$$
\begin{aligned}
&g\big(t, \Pi_1(X_t)\big) - g\big(t_0, \Pi_1\big(X_{t_0}\big)\big) \\
&\qquad = \int_{t_0}^{t} \frac{\partial g}{\partial s}\big(s, \Pi_1(X_s)\big)\, ds \\
&\qquad\quad + \sum_{i=1}^{k_1} \int_{t_0+s_{1,i}}^{t+s_{1,i}} \frac{\partial g}{\partial x_i}\big(s - s_{1,i}, \Pi_1\big(X_{s-s_{1,i}}\big)\big) u(s)\, dW(s) \\
&\qquad\quad + \sum_{i=1}^{k_1} \int_{t_0}^{t} \bigg[\frac{\partial g}{\partial x_i}\big(s, \Pi_1(X_s)\big) v(s + s_{1,i}) \\
&\qquad\qquad + \frac{1}{2} \bigg\langle \frac{\partial^2 g}{\partial x_i^2}\big(s, \Pi_1(X_s)\big) \nabla^{+}_{s_{1,i}} X(s), \nabla^{-}_{s_{1,i}} X(s) \bigg\rangle \bigg] ds,
\end{aligned}
\tag{5.2}
$$

where $\nabla^{\pm}_{s_{1,i}} X(s)$ are defined by (2.9). Applying the Itô formula (2.10) again and using similar notations for h, we obtain

$$
\begin{aligned}
&h\big(t, \Pi_2(X_t)\big) - h\big(t_0, \Pi_2\big(X_{t_0}\big)\big) \\
&\qquad = \int_{t_0}^{t} \frac{\partial h}{\partial s}\big(s, \Pi_2(X_s)\big)\, ds \\
&\qquad\quad + \sum_{i=1}^{k_2} \int_{t_0+s_{2,i}}^{t+s_{2,i}} \frac{\partial h}{\partial x_i}\big(s - s_{2,i}, \Pi_2\big(X_{s-s_{2,i}}\big)\big)\, u(s)\, dW(s) \\
&\qquad\quad + \frac{1}{2} \sum_{i=1}^{k_2} \int_{t_0}^{t} \bigg[\frac{\partial h}{\partial x_i}\big(s, \Pi_2(X_s)\big) v(s + s_{2,i}) \\
&\qquad\qquad + \frac{1}{2} \bigg\langle \frac{\partial^2 g}{\partial x_i^2}\big(s, \Pi_2(X_s)\big) \nabla^{+}_{s_{2,i}} X(s), \nabla^{-}_{s_{2,i}} X(s) \bigg\rangle \bigg] ds.
\end{aligned}
\tag{5.3}
$$

Substituting (5.2) and (5.3) into (1.6), we get the following approximate (Itô–Taylor) expansion of (1.6):

$$
\begin{aligned}
X(t) &= X(t_0) + g\big(t_0, \Pi_1\big(X_{t_0}\big)\big)[W(t) - W(t_0)] + h\big(t_0, \Pi_2\big(X_{t_0}\big)\big)(t - t_0) \\
&\quad + \sum_{i=1}^{k_1} \frac{\partial g}{\partial x_i}\big(t_0, \Pi_1\big(X_{t_0}\big)\big) u(t_0 + s_{1,i}) \\
&\qquad \times \int_{t_0}^{t} \int_{t_0+s_{1,i}}^{t_1+s_{1,i}} dW(t_2)\, dW(t_1) + R(t_0, t),
\end{aligned}
\tag{5.4}
$$

where

$$
\begin{aligned}
R(t_0,t) := \sum_{i=1}^{k_1}\bigg\{\int_{t_0}^{t}\int_{t_0+s_{1,i}}^{t_1+s_{1,i}}\bigg[\frac{\partial g}{\partial x_i}(t_2-s_{1,i},\Pi_1(X_{t_2-s_{1,i}}))u(t_2) \\
-\frac{\partial g}{\partial x_i}(t_0,\Pi_1(X_{t_0}))u(t_0+s_{1,i})\bigg]\,dW(t_2)\,dW(t_1)\bigg\} \\
+\int_{t_0}^{t}\int_{t_0}^{t_1}\sum_{i=1}^{k_1}\bigg[\frac{\partial g}{\partial x_i}(t_2,\Pi_1(X_{t_2}))v(t_2+s_{1,i}) \\
+\frac{1}{2}\bigg\langle\frac{\partial^2 g}{\partial x_i^2}(t_2,\Pi_1(X_{t_2}))\,\nabla_{s_{1,i}}^{+}X_{t_2},\nabla_{s_{1,i}}^{-}X_{t_2}\bigg\rangle\bigg]\,dt_2\,dW(t_1) \\
+\sum_{i=1}^{k_2}\int_{t_0}^{t}\int_{t_0+s_{2,i}}^{t_1+s_{2,i}}\frac{\partial h}{\partial x_i}(t_2-s_{2,i},\Pi_2(X_{t_2-s_{2,i}}))u(t_2)\,dW(t_2)\,dt_1 \\
+\int_{t_0}^{t}\int_{t_0}^{t_1}\sum_{i=1}^{k_2}\bigg[\frac{\partial h}{\partial x_i}(t_2,\Pi_2(X_{t_2}))v(t_2+s_{2,i}) \\
+\frac{1}{2}\bigg\langle\frac{\partial^2 h}{\partial x_i^2}(t_2,\Pi_2(X_{t_2}))\,\nabla_{s_{2,i}}^{+}X_{t_2},\nabla_{s_{2,i}}^{-}X_{t_2}\bigg\rangle\bigg]\,dt_2\,dt_1 \\
+\int_{t_0}^{t}\int_{t_0}^{t_1}\bigg[\frac{\partial g}{\partial t_2}(t_2,\Pi_1(X_{t_2}))+\frac{\partial h}{\partial t_2}(t_2,\Pi_2(X_{t_2}))\bigg]\,dt_2\,dt_1.
\end{aligned}
\tag{5.5}
$$

In the above expression, the stochastic integrals

$$\int_{t_0+s_{1,i}}^{t_1+s_{1,i}}\frac{\partial g}{\partial x_i}(t_2-s_{1,i},\Pi_1(X_{t_2-s_{1,i}}))u(t_2)\,dW(t_2)$$

and

$$\int_{t_0+s_{2,i}}^{t_1+s_{2,i}}\frac{\partial h}{\partial x_i}(t_2-s_{2,i},\Pi_2(X_{t_2-s_{2,i}}))u(t_2)\,dW(t_2)$$

are Skorohod integrals. Define

$$I(t_0+s_{i,j},t+s_{i,j};s_{i,j}) := \int_{t_0}^{t}\int_{t_0+s_{i,j}}^{t_1+s_{i,j}} dW(t_2)\,dW(t_1), \tag{5.6}$$

for $i=1,2$ and $j=1,\dots,k_i$. Recall the definition of $J(t_0+s_{i,j},t+s_{i,j};s_{i,j})$ in (4.1). Note that if $s_{i,j}<0$, then

$$I(t_0+s_{i,j},t+s_{i,j};s_{i,j}) = \int_{t_0}^{t}\int_{t_0+s_{i,j}}^{t_1+s_{i,j}} \circ\, dW(t_2)\circ dW(t_1); \tag{5.7}$$

if $s_{i,j}=0$, then

$$
\begin{aligned}
I(t_0+s_{i,j},t+s_{i,j};s_{i,j}) &= \int_{t_0}^{t}[W(t_1)-W(t_0)]\,dW(t_1) \\
&= \frac{(W(t)-W(t_0))^2}{2}-\frac{t-t_0}{2}.
\end{aligned} \tag{5.8}
$$

5.2. *The one-dimensional Milstein scheme* ($d=m=1$). Assume $d=m=1$. Let $\pi:-\tau=t_{-l}<\cdots<t_0=0<\cdots<t_n=a$ be a partition of $[-\tau,a]$. We introduce the *Milstein* scheme for the SDDE (1.6) as follows:

$$
\begin{aligned}
X^\pi(t) &= X^\pi(t_k)+h\big(t_k,\Pi_2\big(X^\pi_{t_k}\big)\big)(t-t_k)+g\big(t_k,\Pi_1\big(X^\pi_{t_k}\big)\big)\big(W(t)-W(t_k)\big) \\
&\quad+\sum_{i=1}^{k_1}\frac{\partial g}{\partial x_i}\big(t_k,\Pi_1\big(X^\pi_{t_k}\big)\big)u^\pi(t_k+s_{1,i})I(t_k+s_{1,i},t+s_{1,i};s_{1,i})
\end{aligned} \tag{5.9}
$$

for $t_k<t\le t_{k+1}$, where

$$
u^\pi(t)=\begin{cases} g\big(t,\Pi_1(X^\pi_t)\big), & t\ge 0,\\ 0, & -\tau\le t<0,\end{cases}
$$

and

$$
I(t_k+s_{1,i},t+s_{1,i};s_{1,i})=\int_{t_k}^{t}\int_{t_k+s_{1,i}}^{t_1+s_{1,i}}\circ\, dW(t_2)\circ dW(t_1).
$$

Recall the notation

$$
\lfloor s\rfloor:=\begin{cases} t_k, & t_k\le s<t_{k+1},\\ t_{n_t}, & t_{n_t}\le s\le t,\end{cases}
$$

and introduce the following notation:

$$
\lceil s\rceil=\begin{cases} t_{k+1}, & t_k<s\le t_{k+1},\\ t, & t_{n_t}<s\le t.\end{cases}
$$

In view of (5.7) and Lemma 4.2, we will use $J^p(t_i,t;s_{1,i})$ to approximate $I(t_i,t;s_{1,i})$.

Denote by

$$
Z^\pi(t):=X^\pi(t)-X(t),\qquad t\in[-\tau,a]
$$

the global truncation error for the Milstein scheme, with X the unique solution of the SDDE (1.6).

LEMMA 5.1. *In the SDDE* (1.6) (*with* $d=m=1$), *suppose that* $g\in C_b^2(\mathbf{R}^{k_1},\mathbf{R})$, $h\in C_b^2(\mathbf{R}^{k_2},\mathbf{R})$, *have bounded first and second derivatives. Then for* *each integer* $p \geq 1$, *there exists a constant* $K(p) > 0$ *such that*

$$E\left|\left\langle \frac{\partial^2 g}{\partial \vec{x}^2}\big(s, \Pi_1(X_s)\big|\, \nabla^+_{s_{1,i}} X_s, \nabla^-_{s_{1,i}} X_s \right\rangle\right|^p \leq K(p), \tag{5.10}$$
$$E\left|\left\langle \frac{\partial^2 h}{\partial \vec{x}^2}\big(s, \Pi_2(X_s)\big)\, \nabla^+_{s_{2,i}} X_s, \nabla^-_{s_{2,i}} X_s \right\rangle\right|^p \leq K(p),$$

for all $t \in [0, a]$.

PROOF. By the definition of $\nabla^{\pm}_{s_{2,i}} X(s)$ [see (2.9)], we have

$$\begin{aligned}\nabla^+_{s_{1,i},s_{1,j}} X(s) &= 2u(s+s_{1,i}) I_{\{s_{1,i}<s_{1,j}\}} + u(s+s_{1,i})\delta_{ij} \\ &\quad + 2\int_0^{s+s_{1,j}} D_{s+s_{1,i}} u(r)\, dW(r) + 2\int_0^{s+s_{1,j}} D_{s+s_{1,i}} v(r)\, dr\end{aligned} \tag{5.11}$$

and

$$\nabla^-_{s_{1,i},s_{1,j}} X(s) = u(s+s_{1,i})\delta_{ij}. \tag{5.12}$$

Therefore,

$$\begin{aligned}&\left\langle \frac{\partial^2 g}{\partial \vec{x}^2}\big(s, \Pi_1(X_s)\big)\, \nabla^+_{s_{1,i}} X(s), \nabla^-_{s_{1,i}} X(s) \right\rangle \\ &\quad = 2\sum_{i=1}^{k_1} \bigg\{ \frac{\partial^2 g}{\partial x_i\, \partial x_j}\big(s, \Pi_1(X_s)\big) u(s+s_{1,i}) \\ &\qquad \times \bigg[u(s+s_{1,i}) I_{\{s_{1,i}<s_{1,j}\}} + \frac{1}{2} u(s+s_{1,i})\delta_{ij} \\ &\qquad\quad + \int_0^{s+s_{1,j}} D_{s+s_{1,i}} u(r)\, dW(r) \\ &\qquad\quad + \int_0^{s+s_{1,j}} D_{s+s_{1,i}} v(r)\, dr \bigg]\bigg\}.\end{aligned} \tag{5.13}$$

If $r > 0$, then

$$\begin{aligned}D_s u(r) &= D_s g\big(\Pi_1(X_r)\big) \\ &= \sum_{i=1}^{k_1} \frac{\partial g}{\partial x_i}\big(r, \Pi_1(X_r)\big) D_s X(r+s_{1,i})\end{aligned} \tag{5.14}$$

and

$$\begin{aligned}D_t D_s u(r) &= \sum_{i,j=1}^{k_1} \frac{\partial^2 g}{\partial x_i\, \partial x_j}\big(r, \Pi_1(X_r)\big) D_s X(r+s_{1,i}) D_t X(r+s_{1,j}) \\ &\quad + \sum_{i=1}^{k_1} \frac{\partial g}{\partial x_i}\big(r, \Pi_1(X_r)\big) D_t D_s X(r+s_{1,i}).\end{aligned} \tag{5.15}$$

By Proposition 3.1 and Proposition 3.3, there exists a constant $C_1 > 0$ such that

$$\sup_{0\le s\le a} E\Big(\sup_{s\le r\le a} |D_s X(r)|^2\Big) \le C_1,$$

$$\sup_{0\le s,t\le a} E\Big(\sup_{s\vee t\le r\le a} |D_t D_s X(r)|^2\Big) \le C_1.$$

Since g has bounded first and second derivatives, then there is a positive constant C_2 such that

$$\begin{aligned}
&\sup_{0\le s\le a} E\Big(\sup_{s\le r\le a} |D_s u(r)|^2\Big)\\
&\qquad \le C_2 k_1 \sup_{0\le s\le a} E\Big(\sup_{s\le r\le a} |D_s X(r)|^2\Big) \le C_1 C_2 k_1
\end{aligned}$$

and

$$\sup_{0\le s,t\le a} E\Big(\sup_{s\vee t\le r\le a} |D_t D_s u(r)|^2\Big) \le C_1^2 C_2^2 k_1 + C_1 C_2 k_1.$$

If $r < s + s_{1,i}$, then

$$D_{s+s_{1,i}} u(r) = 0,$$
$$D_{s+s_{1,i}} v(r) = 0.$$

Therefore,

$$\begin{aligned}
&E\bigg(\int_{t+s_{1,i}}^{t+s_{1,j}} D_{t+s_{1,i}} u(r)\, dW(r)\bigg)^2\\
&\quad \le \int_{t+s_{1,i}}^{t+s_{1,j}} \int_{t+s_{1,i}}^{t+s_{1,j}} E\big(D_s D_{t+s_{1,i}} u(r)\big)^2\, dr\, ds + \int_{t+s_{1,i}}^{t+s_{1,j}} E\big(D_{t+s_{1,i}} u(r)\big)^2\, dr\\
&\quad \le C_2 k_1^2 C_1^2 + 2C_2 k_1 C_1\\
&\quad =: K_1.
\end{aligned}$$

Similarly, there exists a constant $K_2 > 0$ such that

$$E\bigg(\int_{t+s_{1,i}}^{t+s_{1,j}} D_{t+s_{1,i}} v(r)\, dr\bigg)^2 \le K_2.$$

So the first inequality of (5.10) follows from the above two inequalities and the Lipschitz and bounded conditions on h, g [(1.4) and (1.5)]. The second estimate of (5.10) is proved by a similar argument. □

THEOREM 5.2. *Consider the Milstein scheme* (5.9) *for the SDDE* (1.6). *Recall that* $Z^\pi := X^\pi - X$ *is the global truncation error for any partition* π *of* $[-\tau, a]$.

Let $0 < \gamma \le 1$. *Suppose that* $\eta : [-\tau, 0] \to \mathbf{R}^m$ *is of bounded variation and is* $(\frac{\gamma}{2})$*-Hölder continuous. Let* $g \in C^{1,2}(T \times \mathbf{R}^{k_1}, \mathbf{R})$, $h \in C^{1,2}(T \times \mathbf{R}^{k_2}, \mathbf{R})$ *have bounded first and second spatial derivatives. Assume that*

$$\sup_{-\tau \le s \le 0} |Z^\pi(s)| \le C' |\pi|^\gamma$$

for some positive constant C'. *Then there exists a constant* $C > 0$ (*depending on* a *and independent of* π) *such that*

$$\sup_{-\tau \le s \le a} E|Z^\pi(s)|^2 \le C|\pi|^{2\gamma}.$$

PROOF. We express the global error in the form

$$Z^\pi(t) = Z^\pi(0) + I^\pi(t) - R^\pi(t),$$

where

$$\begin{aligned} I^\pi(t) = {} & \sum_{i=1}^{n_t} \big[h\big(t_{i-1}, \Pi_2\big(X^\pi_{t_{i-1}}\big)\big) - h\big(t_{i-1}, \Pi_2\big(X_{t_{i-1}}\big)\big)\big](t_i - t_{i-1}) \\ & + \sum_{i=1}^{n_t} \big[g\big(t_{i-1}, \Pi_1\big(X^\pi_{t_{i-1}}\big)\big) - g\big(t_{i-1}, \Pi_1\big(X_{t_{i-1}}\big)\big)\big](W_{t_i} - W_{t_{i-1}}) \\ & + \Big[h\big(t_{n_t}, \Pi_2\big(X^\pi_{t_{n_t}}\big)\big) - h\big(t_{n_t}, \Pi_2\big(X_{t_{n_t}}\big)\big)\Big](t - t_{n_t}) \\ & + \Big[g\big(t_{n_t}, \Pi_1\big(X^\pi_{t_{n_t}}\big)\big) - g\big(t_{n_t}, \Pi_1\big(X_{t_{n_t}}\big)\big)\Big]\big(W(t) - W\big(t_{n_t}\big)\big) \\ & + \sum_{i=1}^{n_t} \sum_{j=1}^{k_1} \Bigg\{ I(t_{i-1}, t_i; s_{1,j}) \bigg[\frac{\partial g}{\partial x_j}\big(t_{i-1}, \Pi_1\big(X^\pi_{t_{i-1}}\big)\big) u^\pi(t_{i-1} + s_{1,j}) \\ & \qquad - \frac{\partial g}{\partial x_j}\big(t_{i-1}, \Pi_1\big(X_{t_{i-1}}\big)\big) u(t_{i-1} + s_{1,j}) \bigg] \Bigg\} \\ & + \sum_{j=1}^{k_1} \Bigg\{ I\big(t_{n_t}, t; s_{1,j}\big) \bigg[\frac{\partial g}{\partial x_j}\big(t_{n_t}, \Pi_1\big(X^\pi_{t_{n_t}}\big)\big) u^\pi\big(t_{n_t} + s_{1,j}\big) \\ & \qquad - \frac{\partial g}{\partial x_j}\big(t_{n_t}, \Pi_1\big(X_{t_{n_t}}\big)\big) u\big(t_{n_t} + s_{1,j}\big) \bigg] \Bigg\} \end{aligned}$$

and

$$R^\pi(t) = \sum_{i=1}^{n_t} R(t_{i-1}, t_i) + R\big(t_{n_t}, t\big).$$

We shall decompose $R^\pi(t)$ into five parts:

$$R^\pi(t) = R^\pi_1(t) + R^\pi_2(t) + R^\pi_3(t) + R^\pi_4(t) + R^\pi_5(t),$$

where

$$
\begin{aligned}
R_1^\pi(t) := \sum_{i=1}^{n_t}\sum_{j=1}^{k_1}\bigg\{\int_{t_{i-1}}^{t_i}\int_{t_{i-1}+s_{1,j}}^{s+s_{1,j}}\bigg[&\frac{\partial g}{\partial x_j}(r-s_{1,j},\Pi_1(X_{r-s_{1,j}}))u(r) \\
&-\frac{\partial g}{\partial x_j}(t_{i-1},\Pi_1(X_{t_{i-1}})) \\
&\times u(t_{i-1}+s_{1,j})\bigg]\,dW(r)\,dW(s)\bigg\} \\
+\sum_{j=1}^{k_1}\bigg\{\int_{t_{n_t}}^{t}\int_{t_{n_t}+s_{1,j}}^{s+s_{1,j}}\bigg[&\frac{\partial g}{\partial x_j}(r-s_{1,j},\Pi_1(X_{r-s_{1,j}}))u(r) \\
&-\frac{\partial g}{\partial x_j}(t_{n_t},\Pi_1(X_{t_{n_t}}))u(t_{n_t}+s_{1,j})\bigg]\,dW(r)\,dW(s)\bigg\} \\
=\sum_{j=1}^{k_1}\bigg\{\int_0^t\int_{\lfloor s\rfloor+s_{1,j}}^{s+s_{1,j}}\bigg[&\frac{\partial g}{\partial x_j}(r-s_{1,j},\Pi_1(X_{r-s_{1,j}}))u(r) \\
&-\frac{\partial g}{\partial x_j}(\lfloor s\rfloor,\Pi_1(X_{\lfloor s\rfloor}))u(\lfloor s\rfloor+s_{1,j})\bigg]\,dW(r)\,dW(s)\bigg\},
\end{aligned}
$$

$$
\begin{aligned}
R_2^\pi(t) := \sum_{j=1}^{k_1}\int_0^t\int_{\lfloor s\rfloor}^{s}\bigg[&\frac{\partial g}{\partial x_j}(r,\Pi_1(X_r))v(r+s_{1,j}) \\
&+\frac{1}{2}\Big\langle\frac{\partial^2 g}{\partial \vec{x}^2}(r,\Pi_1(X_r))\,\nabla^+_{s_{1,j}}X_r,\nabla^-_{s_{1,j}}X_r\Big\rangle\bigg]\,dr\,dW(s),
\end{aligned}
$$

$$
R_3^\pi(t) := \sum_{j=1}^{k_2}\int_0^t\int_{\lfloor s\rfloor+s_{2,j}}^{s+s_{2,j}}\frac{\partial h}{\partial x_j}(r-s_{2,j},\Pi_2(X_{r-s_{2,j}}))u(r)\,dW(r)\,ds,
$$

$$
\begin{aligned}
R_4^\pi(t) := \sum_{j=1}^{k_2}\int_0^t\int_{\lfloor s\rfloor}^{s}\bigg[&\frac{\partial h}{\partial x_j}(r,\Pi_2(X_r))v(r+s_{2,j}) \\
&+\frac{1}{2}\Big\langle\frac{\partial^2 h}{\partial \vec{x}^2}(r,\Pi_2(X_r))\,\nabla^+_{s_{2,j}}X_r,\nabla^-_{s_{2,j}}X_r\Big\rangle\bigg]\,dr\,ds
\end{aligned}
$$

and

$$
R_5^\pi(t) := \int_0^t\int_{\lfloor s\rfloor}^{s}\bigg\{\frac{\partial h}{\partial r}(r,\Pi_2(X_r))+\frac{\partial g}{\partial r}(r,\Pi_1(X_r))\bigg\}\,dr\,ds.
$$

By the Itô isometry and the formula for the covariance between two Skorohod integrals ([22], Section 1.3.1), we have

$$\begin{aligned}
&\sup_{0\le s\le t} E|R_1^\pi(s)|^2 \\
&\quad \le k_1 \sum_{j=1}^{k_1} E\int_0^t \biggl\{\int_{\lfloor s\rfloor+s_{1,j}}^{s+s_{1,j}} \biggl[\frac{\partial g}{\partial x_j}\bigl(r-s_{1,j},\Pi_1(X_{r-s_{1,j}})\bigr)u(r) \\
&\qquad\qquad - \frac{\partial g}{\partial x_j}\bigl(\lfloor s\rfloor,\Pi_1(X_{\lfloor s\rfloor})\bigr)u(\lfloor s\rfloor+s_{1,j})\biggr]\,dW(r)\biggr\}^2 ds \\
&\quad \le k_1 \sum_{j=1}^{k_1} \int_0^t\int_{\lfloor s\rfloor+s_{1,j}}^{s+s_{1,j}} E\biggl[\frac{\partial g}{\partial x_j}\bigl(r-s_{1,j},\Pi_1(X_{r-s_{1,j}})\bigr)u(r) \\
&\qquad\qquad - \frac{\partial g}{\partial x_j}\bigl(\lfloor s\rfloor,\Pi_1(X_{\lfloor s\rfloor})\bigr)u(\lfloor s\rfloor+s_{1,j})\biggr]^2 dr\,ds \\
&\qquad + k_1 \sum_{j=1}^{k_1} \int_0^t\int_{\lfloor s\rfloor+s_{1,j}}^{s+s_{1,j}}\int_{\lfloor s\rfloor+s_{1,j}}^{s+s_{1,j}} E\biggl\{D_{r_1}\biggl[\frac{\partial g}{\partial x_j}\bigl(r-s_{1,j}, \\
&\qquad\qquad \Pi_1(X_{r-s_{1,j}})\bigr)u(r) \\
&\qquad\qquad - \frac{\partial g}{\partial x_j}\bigl(\lfloor s\rfloor,\Pi_1(X_{\lfloor s\rfloor})\bigr) \\
&\qquad\qquad \times u(\lfloor s\rfloor+s_{1,j})\biggr]\biggr\}^2 dr_1\,dr_2\,ds \\
&\quad = k_1 R_{11}^\pi(t) + k_1 R_{12}^\pi(t).
\end{aligned}$$

By assumption, the function

$$G_j(s,x,z) = \frac{\partial g}{\partial x_j}(s,x)g(s+s_{1,j},z), \qquad x\in\mathbf{R}^{k_1} \quad\text{and}\quad z\in\mathbf{R}^{k_1}$$

is Lipschitz; that is, there exists a constant $L_1>0$ such that

$$|G_j(s,z)-G_j(s,w)| \le L_1|z-w| \qquad \forall\,(z,w)\in\mathbf{R}^{2k_1} \quad\text{and}\quad 1\le j\le k_1.$$

Using

$$u(r) = \begin{cases} g\bigl(r,\Pi_1(X_r)\bigr), & r\ge 0,\\ 0, & r<0,\end{cases}$$

and

$$\sup_{-\tau\le r_1\le\alpha<\beta\le r_2\le a} E|X(\beta)-X(\alpha)|^2 \le C_2|r_2-r_1|^\gamma,$$

it follows that

$$
\begin{aligned}
R_{11}^{\pi}(t) \le \sum_{j=1}^{k_1} \int_0^t \int_{\lfloor s\rfloor+s_{1,j}}^{s+s_{1,j}} E\Big[&G_j\big(r-s_{1,j}, \Pi_1\big(X_{r-s_{1,j}}\big), \Pi_1(X_r)\big) \\
&- G_j\big(\lfloor s\rfloor, \Pi_1\big(X_{\lfloor s\rfloor}\big), \Pi_1\big(X_{\lfloor s\rfloor+s_{1,j}}\big)\big)\Big]^2 \\
&\times I_{\{\lfloor s\rfloor+s_{1,j}\ge 0\}}\, dr\, ds \\
\le 2k_1 L_1^2 \sum_{j=1}^{k_1} \int_0^t \int_{\lfloor s\rfloor+s_{1,j}}^{s+s_{1,j}} & \sup_{-1\le r_1<r_2\le a,\, |r_2-r_1|\le|\pi|} E|X(r_2)-X(r_1)|^2\, dr\, ds \\
\le 2(a+1)k_1^2 L_1^2 C_2 |\pi|^{2\gamma}&.
\end{aligned}
$$

Now for all $r \ge 0$ and $1 \le j \le k_1$,

$$
\begin{aligned}
&D_s\left(\frac{\partial g}{\partial x_j}\big(r-s_{1,j}, \Pi_1\big(X_{r-s_{1,j}}\big)\big)u(r)\right) \\
&\quad = g\big(r, \Pi_1(X_r)\big) \sum_{i=1}^{k_1} \frac{\partial^2 g}{\partial x_j\, \partial x_i}\big(r-s_{1,j}, \Pi_1\big(X_{r-s_{1,j}}\big)\big) D_s X(r-s_{1,j}+s_{1,i}) \\
&\qquad + \frac{\partial g}{\partial x_j}\big(r-s_{1,j}, \Pi_1\big(X_{r-s_{1,j}}\big)\big) \sum_{i=1}^{k_1} \frac{\partial g}{\partial x_i}\big(r, \Pi_1(X_r)\big) D_s X(r+s_{1,i}).
\end{aligned}
$$

By Proposition 3.1, there exists a constant $C_3 > 0$ such that

$$
\sup_{0\le r\le a} E\left(\sup_{0\le s\le a} |D_r X(s)|^2\right) \le C_3.
$$

By (1.8), (1.10) and boundedness of the spatial derivatives of g, there exists a constant $C_4 > 0$ such that

$$
\sup_{0\le r\le a} \sup_{0\le s\le a} E\left(\left|D_s\left(\frac{\partial g}{\partial x_j}\big(r-s_{1,j}, \Pi_1\big(X_{r-s_{1,j}}\big)\big)u(r)\right)\right|^2\right) \le 2C_4 k_1^2.
$$

Therefore

$$
\begin{aligned}
R_{12}^{\pi}(t) &\le k_1 \sum_{j=1}^{k_1} \int_0^t \int_{\lfloor s\rfloor+s_{1,j}}^{s+s_{1,j}} \int_{\lfloor s\rfloor+s_{1,j}}^{s+s_{1,j}} 4C_4 k_1^2\, dr_1\, dr_2\, ds \\
&\le 4(a+1) C_4 k_1^4 |\pi|^2.
\end{aligned}
$$

Hence there is a constant $C_5 > 0$ such that

$$
\sup_{0\le s\le t} E|R_1^{\pi}(s)|^2 \le C_5 |\pi|^{2\gamma}. \tag{5.16}
$$

Applying Fubini's theorem, we can rewrite $R_3^\pi(t)$ as

$$R_3^\pi(t) = \sum_{i=1}^{n_t}\sum_{j=1}^{k_2}\int_{t_{i-1}}^{t_i}\int_{t_{i-1}+s_{2,j}}^{s+s_{2,j}} \frac{\partial h}{\partial x_j}(r - s_{2,j}, \Pi_2(X_{r-s_{2,j}}))u(r)\,dW(r)\,ds$$
$$+ \sum_{j=1}^{k_2}\int_{t_{n_t}}^{t}\int_{t_{n_t}+s_{2,1}}^{s+s_{2,j}} \frac{\partial h}{\partial x_j}(r - s_{2,j}, \Pi_2(X_{r-s_{2,j}}))u(r)\,dW(r)\,ds.$$

So we have

$$\begin{aligned}
R_3^\pi(t) &= \sum_{i=1}^{n_t}\sum_{j=1}^{k_2}\int_{t_{i-1}+s_{2,j}}^{t_i+s_{2,j}}\int_{r-s_{2,j}}^{t_i} \frac{\partial h}{\partial x_j}(r - s_{2,j}, \Pi_2(X_{r-s_{2,j}}))u(r)\,ds\,dW(r) \\
&\quad + \sum_{j=1}^{k_2}\int_{t_{n_t}+s_{2,j}}^{t+s_{2,j}}\int_{r-s_{2,j}}^{t} \frac{\partial h}{\partial x_j}(r - s_{2,j}, \Pi_2(X_{r-s_{2,j}}))u(r)\,ds\,dW(r) \\
&= \sum_{i=1}^{n_t}\sum_{j=1}^{k_2}\int_{t_{i-1}+s_{2,j}}^{t_i+s_{2,j}} (t_i + s_{2,j} - r)\frac{\partial h}{\partial x_j}(r - s_{2,j}, \Pi_2(X_{r-s_{2,j}}))u(r)\,dW(r) \\
&\quad + \sum_{j=1}^{k_2}\int_{t_{n_t}+s_{2,j}}^{t+s_{2,j}} (t + s_{2,j} - r)\frac{\partial h}{\partial x_j}(r - s_{2,j}, \Pi_2(X_{r-s_{2,j}}))u(r)\,dW(r) \\
&= \sum_{j=1}^{k_2}\int_{s_{2,j}}^{t+s_{2,j}} (\lceil r - s_{2,j}\rceil + s_{2,j} - r)\frac{\partial h}{\partial x_j}(r - s_{2,j}, \Pi_2(X_{r-s_{2,j}}))u(r)\,dW(r).
\end{aligned}$$

Applying the formula for covariance between two Skorohod integrals ([22], Section 1.3.1) and Proposition 3.1, we can show that there exists a constant $C_6 > 0$ such that

$$\sup_{0\le s\le t} E|R_3^\pi(s)|^2 \le C_6|\pi|^2. \tag{5.17}$$

Similarly, by Lemma 5.1, we can easily show that there exist $C_7 > 0$ such that

$$\begin{aligned}
&\sup_{0\le s\le t} E|R_2^\pi(s)|^2 \le C_7|\pi|^2, \\
&\sup_{0\le s\le t} E|R_4^\pi(s)|^2 \le C_7|\pi|^2, \qquad (5.18) \\
&\sup_{0\le s\le t} E|R_5^\pi(s)|^2 \le C_7|\pi|^2.
\end{aligned}$$

By arguments similar to the ones used in the proof of Theorem 3.4 in [15], we obtain the following inequality:

$$\sup_{0\le u\le t} E|I^\pi(u)|^2 \le C_1\int_0^t \sup_{-\tau\le u\le s} E(|Z^\pi(u)|^2)\,ds \tag{5.19}$$

for some constant $C_1 > 0$. From (5.16)–(5.19), there exist $C_8 > 0$ and $C_9 > 0$ such that

$$\sup_{0\le u\le t} E|Z^\pi(u)|^2 \le E|Z^\pi(0)|^2 + C_8|\pi|^{2\gamma} + C_9\int_0^t \sup_{-\tau\le u\le s} E|Z^\pi(u)|^2\,ds. \tag{5.20}$$

So

$$\sup_{-\tau\le u\le t} E|Z^\pi(u)|^2 \le (2C' + C_8)|\pi|^{2\gamma} + C_9\int_0^t \sup_{-\tau\le u\le s} E|Z^\pi(u)|^2\,ds. \tag{5.21}$$

By Gronwall's lemma, there exists a constant $C > 0$ such that

$$E\sup_{-\tau\le s\le t}|Z^\pi(s)|^2 \le C|\pi|^{2\gamma}.$$

□

REMARKS.

1. Let us consider a particular case when g and h are of the form

$$\begin{aligned} g\big(s,\Pi_1(X_s)\big) &= \sum_{i=1}^{k_1} a_i\big(s, X_s(s_{1,i})\big),\\ h\big(s,\Pi_2(X_s)\big) &= \sum_{j=1}^{k_2} b_j\big(s, X_s(s_{2,j})\big), \end{aligned} \tag{5.22}$$

   where $a_i, b_j \in C_b^{1,2}(T\times R)$ for $1\le i\le k_1$ and $1\le j\le k_2$. In this case, one may also apply the nonanticipating Itô formula to

$$a_i\big(t+s_{1,1}, X(t+s_{1,1})\big) - a_i\big(t_0+s_{1,1}, X(t_0+s_{1,1})\big),$$
$$a_i\big(t+s_{1,1}, X(t+s_{1,1})\big) - a_i\big(t_0+s_{1,1}, X(t_0+s_{1,1})\big)$$

   to prove Theorem 5.2 (cf. [28]).
2. One can allow the initial process η to be a sample continuous semimartingale in the following way. Replace W by an extended Brownian motion $W(t)$, $t\ge -\tau$, with the associated Brownian filtration $(\mathcal{F}_t)_{-\tau\le t\le a}$. Assume that $\eta(t)\in\mathbb{D}_m^{1,\infty}$ for all $t\in[-\tau,0]$, and η is an $(\mathcal{F}_t)_{-\tau\le t\le 0}$-continuous semimartingale satisfying

$$\begin{aligned} \sup_{-\tau\le\alpha<\beta\le 0} E|\eta(\beta)-\eta(\alpha)|^2 &\le C_2|\beta-\alpha|^\gamma,\\ \sup_{-\tau\le s\le 0} E\big(|Z^\pi(s)|^2\big) &\le C'|\pi|^{2\gamma} \end{aligned} \tag{5.23}$$

   for some positive constants C_2 and C'. The arguments in Section 2 and the proof of Theorem 5.2 may be adapted to include this generalization.

We can rewrite the SDDE (1.6) in Stratonovich form, namely, if $t \geq 0$,

$$X(t) = \eta(0) + \int_0^t g\big(s, \Pi_1(X_s)\big) \circ dW(s) + \int_0^t \Big[h\big(s, \Pi_2(X_s)\big) - \frac{1}{2}\frac{\partial g}{\partial x_{k_1}}\big(s, \Pi_1(X_s)\big)g\big(s, \Pi_1(X_s)\big)\Big]\,ds, \tag{5.24}$$

if $s_{k_1} = 0$. If $s_{k_1} < 0$, then the SDDE is of the same form as (1.6) except the Itô integral is replaced by a Stratonovich integral, that is,

$$X(t) = \eta(0) + \int_0^t g\big(s, \Pi_1(X_s)\big) \circ dW(s) + \int_0^t h\big(s, \Pi_2(X_s)\big)\,ds.$$

Bell and Mohammed ([5, 6]) derived a similar result in the case of a single delay. From Corollary 2.5, we can obtain the following Stratonovich–Taylor expansion of $X(t)$ (cf. [28]):

$$\begin{aligned} X(t) = {} & X(t_0) + g\big(t_0, \Pi_1(X_{t_0})\big)[W(t) - W(t_0)] \\ & + \bar{h}\big(t_0, \Pi_2(X_{t_0})\big)(t - t_0) \\ & + \sum_{i=1}^{k_1} \frac{\partial g}{\partial x_i}\big(t_0, \Pi_1(X_{t_0})\big)u(t_0 + s_{1,i}) \\ & \times \int_{t_0}^t \int_{t_0+s_{1,i}}^{t_1+s_{1,i}} \circ\, dW(t_2) \circ dW(t_1) + \bar{R}(t_0, t), \end{aligned} \tag{5.25}$$

where

$$\begin{aligned} & \bar{R}(t_0, t) \\ & = \sum_{i=1}^{k_1} \Big\{ \int_{t_0}^t \int_{t_0+s_{1,i}}^{t_1+s_{1,i}} \Big[\frac{\partial g}{\partial x_i}\big(t_2 - s_{1,i}, \Pi_1(X_{t_2-s_{1,i}})\big)u(t_2) \\ & \qquad - \frac{\partial g}{\partial x_i}\big(t_0, \Pi_1(X_{t_0})\big)u(t_0 + s_{1,i})\Big] \circ dW(t_2) \circ dW(t_1) \Big\} \\ & \quad + \int_{t_0}^t \int_{t_0}^{t_1} \sum_{i=1}^{k_1} \frac{\partial g}{\partial x_i}\big(t_2, \Pi_1(X_{t_2})\big)\bar{v}(t_2 + s_{1,i})\,dt_2 \circ dW(t_1) \\ & \quad + \sum_{i=1}^{k_2} \int_{t_0}^t \int_{t_0+s_{2,i}}^{t_1+s_{2,i}} \frac{\partial \bar{h}}{\partial \vec{x}_i}\big(t_2 - s_{2,i}, \Pi_2(X_{t_2-s_{2,i}})\big)u(t_2) \circ dW(t_2)\,dt_1 \\ & \quad + \int_{t_0}^t \int_{t_0}^{t_1} \sum_{i=1}^{k_2} \frac{\partial \bar{h}}{\partial \vec{x}_i}\big(t_2, \Pi_2(X_{t_2})\big)\bar{v}(t_2 + s_{2,i})\,dt_2\,dt_1 \end{aligned} \tag{5.26}$$

and

$$\bar{h} := h - \tfrac{1}{2} g_{k_1} g, \qquad \bar{v}(t) := \begin{cases} \bar{h}\big(t, \Pi_2(X_t)\big), & 0 \leq t \leq a, \\ \eta(t), & t < 0. \end{cases} \tag{5.27}$$

One can also derive the Milstein scheme for (5.24) using the Stratonovich–Taylor expansion (5.25) of $X(t)$ as follows. Let $t_k < t \le t_{k+1}$. Then

$$
\begin{aligned}
X^\pi(t) &= X^\pi(t_k) + \bar{h}\big(t_k, \Pi_2\big(X^\pi_{t_k}\big)\big)(t - t_k) \\
&\quad + g\big(t_k, \Pi_1\big(X^\pi_{t_k}\big)\big)\big(W(t) - W(t_k)\big) \\
&\quad + \sum_{i=1}^{k_1} \frac{\partial g}{\partial x_i}\big(t_k, \Pi_1\big(X^\pi_{t_k}\big)\big)u^\pi(t_k + s_{1,i})J(t_k + s_{1,i}, t + s_{1,i}; s_{1,i}),
\end{aligned}
\tag{5.28}
$$

where

$$
u^\pi(t) = \begin{cases} g\big(t, \Pi_1(X^\pi_t)\big), & t \ge 0, \\ 0, & -\tau \le t < 0. \end{cases}
$$

5.3. *The multidimensional Milstein scheme.* Write $h(s, x) = (h^1(s, x), \dots, h^m(s, x))^T$, $\vec{x} \in \mathbf{R}^{mk_1}$,

$$
\vec{x} = \begin{pmatrix} x_{11}, & \dots, & x_{1k_1} \\ \vdots & \dots & \vdots \\ x_{m1}, & \dots, & x_{mk_1} \end{pmatrix}.
$$

Denote by $g^{jl}(s, \vec{x})$ the (j, l) element of the $m \times d$ matrix $g(s, \vec{x})$. To simplify notation, we use below the summation convention on repeated indices. Recall the notations for the partition $-\tau = t_{-y} < \cdots < t_0 = 0 < \cdots < t_n = t$ introduced in Section 2. We formulate the Milstein scheme for the SDDE (1.6) as follows: if $t_k < t \le t_{k+1}$, the ith coordinate $X^i(t)$ of $X(t) = (X^1(t), \dots, X^m(t))^T$ is approximated by

$$
\begin{aligned}
X^{i,\pi}(t) &= X^{i,\pi}(t_k) + h^i\big(t_k, \Pi_2\big(X^\pi_{t_k}\big)\big)(t - t_k) \\
&\quad + g^{ij}\big(t_k, \Pi_1\big(X^\pi_{t_k}\big)\big)\big(W^j(t) - W^j(t_k)\big) \\
&\quad + \frac{\partial g^{ij}}{\partial x_{i_1 j_1}}\big(t_k, \Pi_1\big(X^\pi_{t_k}\big)\big)u^{i_1 j_1, \pi}\big(t_k + s_{1,j_1}\big) \\
&\qquad \times I_{j,j_1}\big(t_k + s_{1,j_1}, t + s_{1,j_1}; s_{1,j_1}\big),
\end{aligned}
\tag{5.29}
$$

where

$$
u^{i_1 j_1, \pi}(t) = \begin{cases} g^{i_1 j_1}\big(t, \Pi_1(X^\pi_t)\big), & t \ge 0, \\ 0, & -\tau \le t < 0. \end{cases}
$$

As in the SODE case [17, 18] and in view of Lemma 4.2, it is possible to further discretize the double stochastic integral $I_{j,j_1}(t_k + s_{1,j_1}, t + s_{1,j_1}; s_{1,j_1})$ in (5.29) to obtain a modified Milstein scheme for the SDDE (1.6) with strong order of convergence 1. More details are given in Appendix B.

REMARK. One may check that Lemma 5.1 and Theorems 5.2 also hold in the multidimensional case. In fact, it is easy to extend these results to the multidimensional case, thanks to the weak differentiability results (Proposition 3.1, Lemma 3.2 and Proposition 3.3) and the results concerning strong approximation of double Stratonovich integrals (Lemmas 4.1 and 4.2).

Unlike the SODE case, it seems very difficult to develop higher-order strong approximation schemes for the SDDE (1.6). One may try to avoid involving the *differential operator* D and the *trace operator* $\bigtriangledown$ in the numerical scheme by attempting to employ multiple Stratonovich integrals instead of multiple Skorohod integrals. The idea is to use Stratonovich–Taylor expansions of the coefficients in the SDDE (1.6) [cf. (5.3) and (5.4)] instead of Itô–Taylor expansions. However, this is difficult, because it is hard to estimate the order of the error via the remainder term. This is because a multiple (anticipating) Stratonovich integral can not be expressed in terms of multiple (nonanticipating) Itô integrals. The Hu–Meyer formula gives the relationship between multiple Stratonovich and Skorohod integrals ([9], Theorem 3.1 (with nondeterministic kernels), [29], Theorem 3.1, and [27], Theorem 3.4 (with deterministic kernels)) (cf. [25, 29] and [27]). However, the formula still involves the differential operator D and the trace operator $\bigtriangledown$, and hence it is hard to estimate the remainder term.

One may refer to Jolis and Sanz [16], Delgado and Sanz [9], Solé and Utzet [27] and Zakai [29] for multiple Skorohod and multiple Stratonovich integrals.

APPENDIX A

The lemma below follows from the independent increments property of Brownian motion. It is needed in the proof of the Itô formula for tame functions (Theorem 2.1).

LEMMA A.1. *Assume that* $\{\pi_n : 0 = t_0 < t_1 < \cdots < t_n = a\}$ *is a family of partitions of* $[0, a]$, *with* $\lim_{n\to\infty} |\pi_n| = 0$. *Let* $-\tau \le s_1 \le s_2 \le 0$ *and denote by* $\Delta_{lk} W^i := W^i(t_l + s_k) - W^i(t_{l-1} + s_k)$, $1 \le i \le d$, $1 \le l \le n$, $k = 1, 2$, *the increments of Brownian motion. Then*

$$\lim_{n\to\infty} \sum_{l=1}^{n} \Delta_{l1} W^i \Delta_{l2} W^j = \begin{cases} a + s_1, & \text{if } i = j \text{ and } s_1 = s_2, \\ 0, & \text{otherwise}, \end{cases} \tag{A.1}$$

in $L^2(\Omega, \mathbf{R})$.

PROOF. We only need to consider the cases $s_1 < s_2$ and $i = j$. Now

$$\left[\sum_{l=1}^{n} \Delta_{l1} W^i \Delta_{l2} W^i\right]^2 = \sum_{l=1}^{n} (\Delta_{l1} W^i)^2 (\Delta_{l2} W^i)^2 + 2 \sum_{l_1<l_2} \Delta_{l_1 1} W^i \Delta_{l_1 2} W^i \Delta_{l_2 1} W^i \Delta_{l_2 2} W^i.$$

If n is sufficiently large, then $|\pi_n| < s_2 - s_1$. Hence $\Delta_{l_2 2} W^i$ is independent of $\Delta_{l_1 1} W^i \Delta_{l_1 2} W^i \Delta_{l_2 1} W^i$. Taking expectations in the above equality gives

$$E\left[\sum_{l=1}^{n} \Delta_{l1} W^i \Delta_{l2} W^i\right]^2 \le \sum_{l=1}^{n} (t_l - t_{l-1})^2 \le |\pi_n| a$$

for sufficiently large n. Note that $a + s_1$ is the correct limit in (2.6) because of the convention that $W(t) = 0$ for $t < 0$. This completes the proof of the lemma. □

The following lemma extends a result by Nualart and Pardoux ([23], Lemma C1).

LEMMA A.2. *Suppose that $x = \{x(t) : t \in [0, a]\}$ is a measurable real-valued process, $x(t) = 0$ if $t > a$ or $t < 0$, and $x \in L^p([0, a], \mathbf{R})$ a.s., $p > 1$. Assume that $\{\pi_n : 0 = t_0 < t_1 < \cdots < t_n = a\}$ is a family of partitions of $[0, a]$, with $\lim_{n\to\infty} |\pi_n| = 0$, and $-\tau \le s_1, s_2 \le 0$. Then*

$$\lim_{n\to\infty} \sum_{l=1}^{n} \frac{\Delta_{l1} W \Delta_{l2} W}{t_l - t_{l-1}} \int_{t_{l-1}+s_1}^{t_l+s_1} x(s)\, ds = \begin{cases} \int_0^{a+s_1} x(s)\, ds, & s_1 = s_2, \\ 0, & s_1 \ne s_2, \end{cases} \tag{A.2}$$

in probability. Moreover, if $x \in L^p(\Omega \times [0, a], \mathbf{R})$, then the above convergences hold in $L^1(\Omega, \mathbf{R})$.

PROOF. It clearly suffices to show that (A.2) holds in $L^1(\Omega, \mathbf{R})$ whenever $x \in L^p(\Omega \times [0, a], \mathbf{R})$. Fix $m \ge 1$, define

$$x^m := \sum_{l=1}^{m} \frac{I_{(t_{l-1}+s_1, t_l+s_1]}}{t_l - t_{l-1}} \int_{t_{l-1}+s_1}^{t_l+s_1} x(s)\, ds.$$

For $n \ge 1$, define

$$\alpha_n(x) := \sum_{l=1}^{n} \frac{\Delta_{l1} W \Delta_{l2} W}{t_l - t_{l-1}} \int_{t_{l-1}+s_1}^{t_l+s_1} x(s)\, ds.$$

Define $\alpha_n(X_m)$ similarly. It follows from Hölder's inequality that if $1/p + 1/q = 1$, then

$$E|\alpha_n(x)| \le \left\{E \sum_{l=1}^{n} \frac{|\Delta_{l1} W \Delta_{l2} W|^q}{(t_l - t_{l-1})^{q-1}}\right\}^{1/q} \left\{E \sum_{l=1}^{n} \frac{(\int_{t_{l-1}+s_1}^{t_l+s_1} |x(s)|\, ds)^p}{(t_l - t_{l-1})^{p/q}}\right\}^{1/p}, \tag{A.3}$$

that is

$$\|\alpha_n(x)\|_{L^1(\Omega)} \le C_p\|x\|_{L^p(\Omega\times[0,a+s_1])} \le C_p\|x\|_{L^p(\Omega\times[0,a])}.$$

Therefore,

$$\begin{aligned} & E\left|\alpha_n(x) - \int_0^{a+s_1} x(s)\,ds\right| \\ & \quad \le E|\alpha_n(x-x^m)| + E\left|\alpha_n(x^m) - \int_0^{a+s_1} x(s)\,ds\right| \\ & \quad \le E\left|\alpha_n(x^m) - \int_0^{a+s_1} x(s)\,ds\right| + C_p\|x-x^m\|_{L^p(\Omega\times[0,a+s_1])}, \end{aligned} \tag{A.4}$$

since

$$\begin{aligned} \alpha_n(x^m) &= \sum_{i=1}^m \left\{ \sum_{(t_{l-1},t_l]\subseteq(t_{i-1},t_i],\,1\le l\le n} \int_{t_{i-1}+s_1}^{t_i+s_1} \frac{I_{(t_{i-1}+s_1,t_i+s_1]}(t)}{t_l - t_{l-1}}\,dt\,\Delta_{l1}W\Delta_{l2}W \right\} \\ &\qquad \times \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}+s_1}^{t_i+s_1} x(s)\,ds \\ &= \sum_{i=1}^m \left\{ \sum_{(t_{l-1},t_l]\subseteq(t_{i-1},t_i],\,1\le l\le n} \Delta_{l1}W\Delta_{l2}W \right\} \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}+s_1}^{t_i+s_1} x(s)\,ds. \end{aligned}$$

Let k_m be the index such that $t_{k_m-1} + s_1 < 0 \le t_{k_m} + s_1$. If $s_1 = s_2$, then by Lemma A.1, the following limit exists in probability:

$$\begin{aligned} \lim_{n\to\infty} \alpha_n(x^m) &= \sum_{i=1}^m \big[(t_i + s_1)\wedge 0 - (t_{i-1}+s_1)\vee 0\big] \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}+s_1}^{t_i+s_1} x(s)\,ds \\ &= \sum_{i=k_m+1}^m \int_{t_{i-1}+s_1}^{t_i+s_1} x(s)\,ds + \frac{t_{k_m}+s_1}{t_{k_m} - t_{k_m-1}} \int_0^{t_{k_m}+s_1} x(s)\,ds \\ &= \int_0^{a+s_1} x(s)\,ds + \frac{t_{k_m}+s_1}{t_{k_m} - t_{k_m-1}} \int_0^{t_{k_m}+s_1} x(s)\,ds. \end{aligned}$$

Equivalently,

$$\alpha_n(x^m) - \int_0^{a+s_1} x(s)\,ds - \frac{t_{k_m}+s_1}{t_{k_m} - t_{k_m-1}} \int_0^{t_{k_m}+s_1} x(s)\,ds \to 0$$

as $n \to \infty$ in probability.

A slight modification in the proof of (A.3) yields the estimate

$$\|\alpha_n(x^m)\|_{L^{p'}(\Omega)} \le C(p, p')\|x^m\|_{L^p(\Omega\times[0,a+s_1])},$$

for all $p' \in (1, p)$. Therefore, the family $\{\alpha_n(x^m): n \geq 1\}$ is uniformly integrable. From (A.4) we have

$$\begin{aligned}
\lim_{n\to\infty} E\left|\alpha_n(x) - \int_0^{a+s_1} x(s)\,ds\right| &\leq E\left|\frac{t_{k_m}+s_1}{t_{k_m - t_{k_m-1}}} \int_0^{t_{k_m}+s_1} x(s)\,ds\right| + C_p\|x - x^m\|_{L^p(\Omega\times[0,a+s_1])} \\
&\leq E\int_0^{t_{k_m}+s_1} |x(s)|\,ds + C_p\|x - x^m\|_{L^p(\Omega\times[0,a+s_1])}.
\end{aligned}$$

Clearly, $x^m \to x$ in $L^p(\Omega \times [0, a + s_1], \mathbf{R})$ and $E\int_0^{t_{k_m}+s_1} |x(s)|\,ds \to 0$ as $m \to \infty$. So

$$\lim_{n\to\infty} E\left|\alpha_n(x^m) - \int_0^{a+s_1} x(s)\,ds\right| = 0.$$

Now consider the case $s_1 \neq s_2$. Since

$$\begin{aligned}
E|\alpha_n(x)| &\leq E|\alpha_n(x^m)| + E|\alpha_n(x - x^m)| \\
&\leq E|\alpha_n(x^m)| + C_p\|x - x^m\|_{L^p(\Omega\times[0,a+s_1])},
\end{aligned}$$

a similar argument gives $\lim_{n\to\infty} E|\alpha_n(x)| = 0$. $\square$

The following useful result is due to Föllmer [10], and Nualart and Pardoux ([23], Lemma C.2).

LEMMA A.3. *Let $x^i(t)$, $0 \leq t \leq a$, $i = 1, 2$, be two-continuous processes, and $\{\pi_n : 0 = t_0 < t_1 < \cdots < t_n = a\}$ a family of partitions of $[0, a]$, with $\lim_{n\to\infty} |\pi_n| = 0$. For each n and $l = 1, \dots, n$, let $x^i_{t_l,n}$ denote $x^i(t_l)$. Assume that*

$$\sum_{l=1}^{n} (x^i_{t_l,n} - x^i_{t_{l-1},n})(x^j_{t_l,n} - x^j_{t_{l-1},n}) \to \int_0^a a^{ij}(s)\,ds \tag{A.5}$$

in probability as $n \to \infty$, where $\{a^{ij}(t) : 0 \leq t \leq a; i, j = 1, 2\}$ are measurable processes such that a.s.

$$\int_0^a |a^{ij}(s)|\,ds < \infty, \qquad i, j = 1, 2. \tag{A.6}$$

Let $\{Y(t) : 0 \leq t \leq a\}$ be a continuous process, and $\{Y^n(t) : 0 \leq t \leq a\}_{n=1}^{\infty}$ be measurable processes which converge a.s. to $\{Y(t)\}$ as $n \to \infty$, uniformly with respect to $t \in [0, a]$. Then

$$\sum_{l=1}^{n} Y^n(t_{l-1})(x^i_{t_l,n} - x^i_{t_{l-1},n})(x^j_{t_l,n} - x^j_{t_{l-1},n}) \to \int_0^a a^{ij}(s)Y(s)\,ds \tag{A.7}$$

in probability as $n \to \infty$, for $i = 1, 2$.

APPENDIX B

Simulating a double stochastic integral. The following scheme is adapted from Kloeden and Platen ([17], page 202, and [18], page 82).

In view of Lemma 4.2, we can use the truncated sums

$$\tag{B.1} \begin{aligned} J^p_{i,j}(0,t;-b) &= \tfrac{1}{2}\big(W^i(t)B^j(t)\big) - \tfrac{1}{2}\big(W^i(t)a_0^{j,b} - B^j(t)a_0^i\big) \\ &\quad + \pi \sum_{n=1}^{p} n\big(a_n^i b_n^{j,b} - b_n^i a_n^{j,b}\big), \qquad t \ge 0, p \ge 1 \end{aligned}$$

to simulate the double Stratonovich integral

$$\tag{B.2} J_{i,j}(0,t;-b) = \int_b^{t+b} \int_0^{s-b} \circ\, dW^i(v) \circ dW^j(s).$$

Consider the Milstein scheme (5.9). Given an error bound $\delta = O(|\pi|^2)$, we choose an integer $p \ge 1$ such that

$$\frac{1}{p} \le \delta \wedge \min\{|s_{1,i}| : 1 \le i \le k_1\}.$$

We define for all such integers p,

$$\tag{B.3} I^p(t_k + s_{1,i}, t + s_{1,i}; s_{1,i}) := \begin{cases} J^p_{ij}(t_k + s_{1,i}, t + s_{1,i}; s_{1,i}), & s_{1,i} < 0, \\ \dfrac{(W(t) - W(t_0))^2}{2} - \dfrac{t - t_0}{2}, & s_{1,i} = 0. \end{cases}$$

By Lemma 4.2, the following modification of the Milstein scheme:

$$\tag{B.4} \begin{aligned} X^\pi(t) &= X^\pi(t_k) + h\big(t_k, \Pi_2\big(X^\pi_{t_k}\big)\big)(t - t_k) + g\big(t_k, \Pi_1\big(X^\pi_{t_k}\big)\big)\big(W(t) - W(t_k)\big) \\ &\quad + \sum_{i=1}^{k_1} \frac{\partial g}{\partial x_i}\big(t_k, \Pi_1\big(X^\pi_{t_k}\big)\big) u^\pi(t_k + s_{1,i}) I^p(t_k + s_{1,i}, t + s_{1,i}; s_{1,i}), \\ &\qquad t_k < t \le t_{k+1}, \end{aligned}$$

has strong order of convergence 1 (cf. Theorem 5.2).

Suppose that we use the family of partitions: $\pi : -1 = t_{-l} < \cdots < t_0 = 0 < \cdots < t_n = a$, $\operatorname{mesh} \pi = |\pi|$, to calculate the solution of the SDDE (1.6) (with $\tau = 1$) by applying the Milstein scheme. There are some issues we need to consider concerning simulating the family

$$S = \big\{J(t_{k-1}, t_k; s_{1,i}) : k = 1, \dots, n, i = 1, \dots, k_1\big\}.$$

If $k_1 \ne k_2$ or $i_1 \ne i_2$, by the Itô isometry, $J(t_{k_1-1}, t_{k_1}; s_{1,i_1})$ and $J(t_{k_2-1}, t_{k_2}; s_{1,i_2})$ are independent. But the family S may not be independent. The reason is that they come from the same Brownian motion. We can make S independent by choosing appropriate mesh points so that $t_k + s_{1,i} \in \pi$, for all $k \ge 0$, $1 \le i \le k_1$; that is,

$t_k + s_{1,i}$ is also a mesh point. In order to see this, set

(B.5) $$V_k = \{(a_n(t_{k-1}), b_n(t_{k-1})) : n = 0, 1, \dots\},$$

where $a_n(t_{k-1})$ and $b_n(t_{k-1})$ are defined by (4.6). Then for all $1 \le i \le k_1$, $1 \le k \le n$, the set

(B.6) $$V_k(s_{1,i}) = \{(a_n^{-s_{1,i}}(t_{k-1}), b_n^{-s_{1,i}}(t_{k-1})) : n = 0, 1, \dots\}$$

belongs to the family $\{V_k : k = 1, \dots, n\}$, where $a_n^{-s_{1,i}}(t_{k-1})$ and $b_n^{-s_{1,i}}(t_{k-1})$ are defined by (4.7). Indeed, similar to the approximation scheme of multiple Stratonovich integrals ([17], (5.8.10) and (5.8.12), [18], (2.3.30) and (2.3.32)), we have the following approximation scheme of $\{J^p(t_{k-1}, t_k; s_{1,i}) : k = 1, \dots, n, i = 1, \dots, k_1\}$, $p \ge 1$.

For each $k = 1, \dots, n$, and $h = 1, \dots, p$, with $p \ge 1$, we define ρ_p and independent $N(0, 1)$ random variables $\xi(s)$, $\mu_p(s)$, $\zeta_h(s)$, $\eta_h(s)$, $s \in \{t_0, \dots, t_{n-1}\}$, by

(B.7)
$$\begin{aligned}
&\xi(s) = \frac{1}{|\pi|}(W(|\pi| + s) - W(s)),\\
&\zeta_h(s) = \sqrt{\frac{2}{|\pi|}} h\pi a_h(s), \qquad \eta_h(s) = \sqrt{\frac{2}{|\pi|}} h\pi b_h(s),\\
&\rho_p = \frac{1}{12} - \frac{1}{2\pi^2}\sum_{h=1}^{p}\frac{1}{h^2}, \qquad \mu_p(s) = \frac{1}{\sqrt{|\pi|\rho_p}}\sum_{h=p+1}^{\infty} a_h(s),\\
&a_0(s) = -\frac{1}{\pi}\sqrt{2|\pi|}\sum_{h=1}^{p}\frac{1}{h}\zeta_h(s) - 2\sqrt{|\pi|\rho_p}\mu_p(s).
\end{aligned}$$

If $t_{k-1} + s_{1,i} \ge 0$, then

(B.8)
$$\begin{aligned}
&J^p(t_{k-1} + s_{1,i}, t_k + s_{1,i}; s_{1,i})\\
&\quad = \frac{1}{2}|\pi|\xi(t_{k-1} + s_{1,i})\xi(t_{k-1})\\
&\qquad - \frac{1}{2}\sqrt{|\pi|}\,[\xi(t_{k-1} + s_{1,i})a_0(t_{k-1}) - \xi(t_{k-1})a_0(t_{k-1} + s_{1,i})]\\
&\qquad + \frac{|\pi|}{2\pi}\sum_{h=1}^{p}\frac{1}{h}[\zeta_h(t_{k-1} + s_{1,i})\eta_h(t_{k-1}) - \zeta_h(t_{k-1})\eta_h(t_{k-1} + s_{1,i})].
\end{aligned}$$

REMARKS.

1. The space complexity of the Milstein scheme for an SDDE is $O(ma/|\pi|)$ if we only want to simulate the end point $X(a)$ (or the end segment X_a). The space

complexity of the Milstein scheme for an m-dimensional SODE is $O(ma)$ if we only want to simulate the end point $X(a)$. If we want to simulate the whole path $\{X(t): t \in [0, a]\}$, then both schemes have the same space complexity $O(ma/|\pi|)$.

2. Roughly speaking, the time complexity of the Milstein scheme for a multidimensional SDDE ($m > 1$) is K times the time complexity of the corresponding scheme for an SODE, where $K := k_1 + k_2$ is the total number of delays. If $m = 1$, we can directly simulate the double stochastic integral in the Milstein scheme using (B.3).
3. In view of (B.7) and (B.8), we do not need to simulate the joint law of multivariate normal variables for multidimensional SDDEs and SODEs. If m is not very large, simulating the joint law is not a prohibitive task by using Cholesky's decomposition.

Acknowledgments. The authors are very grateful to the anonymous referees and to the editors of the *Annals of Probability* for their helpful suggestions and comments on an earlier version of the manuscript.

REFERENCES

[1] AHMED, T. A. (1983). Stochastic functional differential equations with discontinuous initial data. M.Sc. thesis, Univ. Khartoum, Sudan.
[2] ALÒS, E. and NUALART, D. (1998). An extension of Itô's formula for anticipating processes. *J. Theoret. Probab.* **2** 493–514.
[3] ASCH, J. and POTTHOFF, J. (1991). Itô's lemma without non-anticipatory conditions. *Probab. Theory Related Fields* **88** 17–46.
[4] BAKER, C. T. H. and BUCKWAR, E. (2000). Numerical analysis of explicit one-step methods for stochastic delay differential equations. *LMS J. Comput. Math.* **3** 315–335 (electronic). Available at www.lms.ac.uk/jcm.
[5] BELL, D. and MOHAMMED, S.-E. A. (1989). On the solution of stochastic ordinary differential equations via small delays. *Stochastics Stochastics Rep.* **28** 293–299.
[6] BELL, D. and MOHAMMED, S.-E. A. (1991). The Malliavin calculus and stochastic delay equations. *J. Funct. Anal.* **99** 75–99.
[7] BERGER, M. and MIZEL, V. J. (1982). An extension of the stochastic integral. *Ann. Probab.* **10** 435–450.
[8] CAMBANIS, S. and HU, Y. (1996). The exact convergence rate of Euler–Maruyama scheme and application to sample design. *Stochastics Stochastics Rep.* **59** 211–240.
[9] DELGADO, R. and SANZ, M. (1992). The Hu–Meyer formula for non-deterministic kernels. *Stochastics Stochastics Rep.* **38** 149–158.
[10] FÖLLMER, H. (1981). Calcul d'Itô sans probabilités. *Seminar en Probability XV. Lecture Notes in Math.* **850** 143–150. Springer, New York.
[11] GAINES, J. G. and LYONS, T. J. (1994). Random generation of stochastic area integrals. *SIAM J. Appl. Math.* **54** 1132–1146.
[12] GENTLE, J. (1998). *Random Number Generation and Monte Carlo Methods*. Springer, New York.
[13] HU, Y. (1996). Strong and weak order of time discretization schemes of stochastic differential equations. *Séminaire de Probabilités XXX. Lecture Notes in Math.* **1626** 218–227. Springer, Berlin.

[14] HU, Y. (2000). Optimal times to observe in the Kalman–Bucy model. *Stochastics Stochastic Rep.* **69** 123–140.

[15] HU, Y., MOHAMMED, S.-E. A. and YAN, F. (2001). Discrete-time approximations of stochastic differential systems with memory. Dept. Mathematics, Southern Illinois Univ., Carbondale. Available at http://sfde.math.siu.edu/recentpub.html.

[16] JOLIS, M. and SANZ, M. (1988). On generalized multiple stochastic integrals and multiparameter anticipative calculus. *Stochastic Analysis and Related Topics II. Lecture Notes in Math.* **1444** 141–182. Springer, New York.

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

[18] KLOEDEN, P., PLATEN, E. and SCHURZ, H. (1994). *Numerical Solution of SDE Through Computer Experiments*. Springer, Berlin.

[19] MCSHANE, E. J. (1974). *Stochastic Calculus and Stochastic Models*. Academic Press, New York.

[20] MOHAMMED, S.-E. A. (1984). *Stochastic Functional Differential Equations*. *Research Notes in Mathematics* **99**. Pitman Books, London.

[21] MOHAMMED, S.-E. A. (1998). Stochastic differential systems with memory: Theory, examples and applications. In *Stochastic Analysis and Related Topics VI* (L. Decreusefond, J. Gjerde, B. Øksendal and A. S. Ustunel, eds.) 1–77. Birkhäuser, Boston.

[22] NUALART, D. (1995). *The Malliavin Calculus and Related Topics*. Springer, New York.

[23] NUALART, D. and PARDOUX, E. (1988). Stochastic calculus with anticipating integrands. *Probab. Theory Related Fields* **78** 535–581.

[24] PARDOUX, E. and PROTTER, P. (1990). Stochastic Volterra equations with anticipating coefficients. *Ann. Probab.* **18** 1635–1655.

[25] RUSSO, F. and VALLOIS, P. (1993). Forward, backward and symmetric stochastic integration. *Probab. Theory Related Fields* **97** 403–421.

[26] RYDEN, T. and WIKTORSSON, M. (2001). On the simulation of iterated Itô integrals. *Stochastic Process. Appl.* **91** 151–168.

[27] SOLÉ, J. and UTZET, F. (1990). Stratonovich integral and trace. *Stochastics Stochastics Rep.* **29** 203–220.

[28] YAN, F. (1999). Topics on stochastic delay equations. Ph.D. dissertation, Southern Illinois Univ., Carbondale, Illinois.

[29] ZAKAI, M. (1990). Stochastic integration, trace and the skeleton of Wiener functionals. *Stochastics Stochastics Rep.* **32** 93–108.

Y. HU
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66045-2142
USA
E-MAIL: hu@math.ukans.edu

S.-E. A. MOHAMMED
DEPARTMENT OF MATHEMATICS
SOUTHERN ILLINOIS UNIVERSITY
CARBONDALE, ILLINOIS 62901
USA
E-MAIL: salah@sfde.math.siu.edu

F. YAN
WILLIAMS ENERGY MARKETING
AND TRADING
ONE WILLIAMS CENTER, WRC2-4
TULSA, OKLAHOMA 74119
USA
E-MAIL: fyan1@yahoo.com