SIAM J. SCI. COMPUT.
Vol. 37, No. 1, pp. A295–A318

NUMERICAL METHODS FOR STOCHASTIC DELAY DIFFERENTIAL EQUATIONS VIA THE WONG–ZAKAI APPROXIMATION*

WANRONG CAO†, ZHONGQIANG ZHANG‡, AND GEORGE EM KARNIADAKIS§

Abstract. We use the Wong–Zakai approximation as an intermediate step to derive numerical schemes for stochastic delay differential equations. By approximating the Brownian motion with its truncated spectral expansion and then using different discretizations in time, we present three schemes: a predictor-corrector scheme, a midpoint scheme, and a Milstein-like scheme. We prove that the predictor-corrector scheme converges with order half in the mean-square sense while the Milstein-like scheme converges with order one. Numerical tests confirm the theoretical prediction and demonstrate that the midpoint scheme is of half-order convergence. Numerical results also show that the predictor-corrector and midpoint schemes can be of first-order convergence under commutative noises when there is no delay in the diffusion coefficients.

Key words. predictor-corrector scheme, midpoint scheme, Milstein scheme, Stratonovich formulation

AMS subject classifications. Primary, 60H35; Secondary, 34K50, 65C30

DOI. 10.1137/130942024

1. Introduction.

Numerical solution of stochastic delay differential equations (SDDEs) has attracted increasing interest recently, as memory effects in the presence of noise are modeled with SDDEs in engineering and finance, e.g., [10, 13, 34, 37, 43]. Most numerical methods for SDDEs have focused on the convergence and stability of time-discretization schemes since the early works [38, 39]. Currently, several time-discretization schemes have been well studied: the Euler-type schemes (the forward Euler scheme [1, 21] and the drift-implicit Euler scheme [16, 23, 48]), the Milstein schemes [3, 14, 15, 20], the split-step schemes [11, 44, 49], and also some multistep schemes [4, 5, 6, 7].

Although SDDEs can be thought as a special class of stochastic differential equations (SDEs), the extension of numerical methods for SDEs to SDDEs is nontrivial especially since the delay may induce instabilities in the underlying SDDEs while the corresponding SDEs are stable; see, e.g., [16, 26]. Also, the formulation of appropriate numerical methods requires a somewhat different calculus because of the delay nature of SDDEs (see the Ito–Taylor expansion, e.g., [20, 35]), or anticipative calculus (see, e.g., [15]). Further, the presence of time delay affects the convergence order and computational complexity of numerical methods, as will be shown in section 3.

*Submitted to the journal's Methods and Algorithms for Scientific Computing section October 21, 2013; accepted for publication (in revised form) November 24, 2014; published electronically February 3, 2015. This work was supported partially by an OSD/MURI grant FA9550-09-1-0613, by NSF/DMS grant DMS-1216437, and also by the Collaboratory on Mathematics for Mesoscopic Modeling of Materials (CM4) which is sponsored by DOE.

http://www.siam.org/journals/sisc/37-1/94202.html

†Department of Mathematics, Southeast University, Nanjing 210096, People's Republic of China (wrcao@seu.edu.cn). This author's work was also supported partially by NSF of China 11271068 and 10901036.

‡Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609, and Division of Applied Mathematics, Brown University, Providence, RI 02912 (zzhang7@wpi.edu; zhongqiang_zhang@brown.edu).

§Corresponding author. Division of Applied Mathematics, Brown University, Providence, RI 02912 (george_karniadakis@brown.edu).

Here we employ a different approach, the so-called Wong–Zakai (WZ) approximation; see, e.g., [24, 40, 42]. The difference between the WZ approximation and the aforementioned schemes is that in WZ we first approximate the Brownian motion with an absolute continuous process and then apply proper time-discretization schemes for the resulting equation while the aforementioned schemes are ready for simulation without any further time discretization. The WZ approximation thus can be viewed as an intermediate step for deriving numerical schemes and can provide more flexibility of discretization of Brownian motion before performing any time discretization. In this paper, we show the flexibility of this approach and derive three numerical schemes for SDDEs using the Stratonovich formulation with a spectral truncation of Brownian motion.

In the literature of SDEs and SDDEs, the WZ approximation has been well known for a long time; see, e.g., [46, 47] for SDEs and [24, 40] for SDDEs. However, there is no systematic investigation of designing numerical schemes based on WZ approximation. To the best of our knowledge, the only work addressing full discretization based on WZ approximation is [32] for SDEs using an approximation of Brownian motion with mollifying. To derive consistent numerical schemes based on WZ, we require additional consistency on further time-discretization, i.e., half-order discretization for the diffusion coefficients. Once this requirement is satisfied, we have consistent numerical schemes for the underlying SDDEs, according to Theorem 2.2 and the rule of thumb in section 2.1. Here we will show that further time discretization is crucial to the design of numerical schemes since it will determine the convergence orders of schemes: both half-order schemes and first-order schemes can be derived. We will present three schemes and illustrate how the WZ approximation and time-discretization work on them; see section 2 for details.

In this paper, we employ the classical piecewise linear interpolation of Brownian motion in, e.g., [24, 40, 42] and a Fourier approximation of Brownian motion. Specifically, we will derive three distinct schemes using different time-discretization techniques. After approximating the Brownian motion by a spectral expansion, we then use the trapezoidal rule and the predictor-corrector strategy to obtain a predictor-corrector scheme and prove its convergence in the mean-square sense. We also use the midpoint rule within the WZ approximation to derive a fully implicit scheme (implicit in both drift and diffusion coefficients). These two schemes are convergent with strong order half for SDDEs, as shown numerically in section 3.

If no delay arises, the predictor-corrector scheme and the midpoint scheme coincide with those for SDEs without delay. The predictor-corrector scheme degenerates into a family of the predictor-corrector scheme in [2], which were proposed in order to overcome numerical stability introduced by the Euler scheme and other one-step explicit schemes. Without delay, our midpoint scheme becomes one of the symplectic-preserving schemes in [27] for stochastic Hamiltonian systems. Though we will only focus on the convergence of these schemes and check their numerical performance, we expect that these schemes have larger stability regions than the Euler scheme for SDDEs as in the cases without delay.

Based on Taylor expansion of the diffusion coefficients, we also derive a first-order scheme (called Milstein-like), which is similar to the Milstein scheme [14, 15, 20]. The Milstein-like scheme we propose here can be readily used in routine simulation unlike the Milstein scheme [14, 15, 20] which requires additional approximation of the double integrals. Specifically, the double integrals are approximated with spectral truncation using truncation parameters reciprocal to the time step size to achieve first-order convergence, which will be shown both in theory and in computation. The spectral truncations we use are from the piecewise linear interpolation and a Fourier expansion. Comparison between these two truncations will be presented for a specific numerical example in section 3, where it is shown that the Fourier approach is faster than the piecewise constant approach. It is worth noting that the approximation of double integrals in the present context is similar to those using numerical integration techniques which has been long explored; see, e.g., [19, 29].

In general, the first-order schemes such as the Milstein scheme is not as popular as half-order schemes because of the high cost of simulating double integrals. However, in certain cases the first-order scheme is preferred, e.g., when a commutative condition is satisfied and the double integrals can be represented by the increments of Brownian motions. Also, when the coefficients in front of noises are small, we can achieve satisfactory accuracy with low computational cost since the cost of simulating double integrals can be low; see, e.g., [28] for SDEs with small noise and Remark 2.7. Moreover, the first-order scheme can be used to improve the computational performance of multilevel Monte-Carlo methods; see, e.g., [8].

The rest of the paper is organized as follows. In section 2, we show how to derive our schemes from WZ approximation to the Stratonovich SDDEs. Numerical results will be presented in section 3 to illustrate the convergence of the three schemes and to compare their numerical performance. We will show that the Milstein-like scheme is much slower than the predictor-corrector and midpoint schemes as in each step the evaluation of double integrals is expensive, no matter what approximation for the double integrals is used. Finally, we prove in section 4 that the predictor-corrector scheme is of half-order convergence in the mean-square sense while the Milstein-like scheme is of first-order convergence.

2. Numerical schemes for SDDEs.

Consider the following SDDE with constant delay in Stratonovich form:

$$\mathrm{d}X(t) = f(X(t), X(t-\tau))\mathrm{d}t + \sum_{l=1}^{r} g_l(X(t), X(t-\tau)) \circ \mathrm{d}W_l(t), \quad t \in (0, T],$$

$$X(t) = \phi(t), \quad t \in [-\tau, 0], \tag{2.1}$$

where $\tau > 0$ is a constant, $(W(t), \mathcal{F}_t) = (\{W_l(t),\ 1 \le l \le r\}, \mathcal{F}_t)$ is a system of one-dimensional independent standard Wiener process, the functions $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, $g_l : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, $\phi(t) : [-\tau, 0] \to \mathbb{R}^d$ are continuous with $\mathbb{E}\|\phi\|^2_{L^\infty} < \infty$. We also assume that $\phi(t)$ is $\mathcal{F}_0$-measurable.

For the mean-square stability of (2.1), we assume that f, g_l, $\partial_x g_l g_q$, and $\partial_{x_\tau} g_l g_q$, ($\partial_x$ and ∂_{x_τ} denote the derivatives with respect to the first and second variables, respectively), $l, q = 1, 2, \ldots, r$, in (2.1) satisfy the following Lipschitz conditions:

$$|v(x_1, y_1) - v(x_2, y_2)|^2 \le L_v(|x_1 - x_2|^2 + |y_1 - y_2|^2), \tag{2.2}$$

and the linear growth conditions

$$|v(x_1, y_1)|^2 \le K(1 + |x_1|^2 + |y_1|^2) \tag{2.3}$$

for every x_1, y_1, x_2, $y_2 \in \mathbb{R}^d$, where L_v, K are positive constants, which depend only on v. Under these conditions, (2.1) has a unique sample-continuous and $\mathcal{F}_t$-adapted strong solution $X(t) : [-\tau, +\infty) \to \mathbb{R}^d$; see, e.g., [25, 30].

The WZ approximation (see, e.g., [46, 47]), is a semidiscretization method, where Brownian motion is approximated by finite-dimensional absolute continuous stochastic processes before any discretization in time. There are different types of WZ approximation; see, e.g., [17, 31, 36, 50]. Here we use an orthogonal expansion approach for WZ approximation of Brownian motion:

$$W^{(n)}(t) = \sum_{j=1}^{n} \int_0^t m_j(s)\, ds \int_0^T m_j(t)\, dW, \quad t \in [0,\, T], \tag{2.4}$$

where $\{m_j(t)\}_{j=1}^{\infty}$ is a complete orthonormal system (CONS) in $L^2([0,T])$, and $\xi_j =: \int_0^T m_j(t)\, dW$, are mutually independent standard Gaussian random variables. In this paper, we will use a piecewise version of spectral expansion (2.4) by taking a partition $0 = \mathbf{t}_0 < \mathbf{t}_1 < \cdots < \mathbf{t}_{N_\Delta - 1} < \mathbf{t}_{N_\Delta} = T$ and choosing a truncated CONS, $\{m_j^{(n)}(t)\}_{j=1}^{N_h}$ in $L^2([\mathbf{t}_n, \mathbf{t}_{n+1}])$ for $n = 0, \ldots, N_\Delta - 1$:

$$W^{(N_h,n)}(t) = \sum_{n=0}^{N_\Delta - 1} \sum_{j=1}^{N_h} \int_0^t \chi_{[\mathbf{t}_n, \mathbf{t}_{n+1})} m_j^{(n)}(s)\, ds \xi_j^{(n)}, \quad \xi_j^{(n)} = \int_{\mathbf{t}_n}^{\mathbf{t}_{n+1}} m_j^{(n)}(s)\, dW, \tag{2.5}$$

where χ is the indicator function.

Here different choices of CONS lead to different representations. The orthonormal piecewise constant basis over time interval $[\mathbf{t}_n, \mathbf{t}_{n+1})$, with $\Delta' = (\mathbf{t}_{n+1} - \mathbf{t}_n)/N_h$,

$$m_j^{(n)}(t) = \frac{\sqrt{N_h}}{\sqrt{\mathbf{t}_{n+1} - \mathbf{t}_n}} \chi_{[\mathbf{t}_n + (j-1)\Delta', \mathbf{t}_n + j\Delta')}, \quad j = 1, 2, \ldots, N_h, \tag{2.6}$$

gives the classical piecewise linear interpolation (see, e.g., [17, 41, 46, 47]) and if $N_h = 1$,

$$W^{(1,n)}(t) = W(\mathbf{t}_n) + (t - \mathbf{t}_n) \frac{W(\mathbf{t}_{n+1}) - W(\mathbf{t}_n)}{\mathbf{t}_{n+1} - \mathbf{t}_n}, \; t \in [\mathbf{t}_n, \mathbf{t}_{n+1}]. \tag{2.7}$$

The orthonormal Fourier basis gives Wiener's representation (see, e.g., [19, 29, 33]):

$$\begin{aligned} m_1^{(n)}(t) &= \frac{1}{\sqrt{\mathbf{t}_{n+1} - \mathbf{t}_n}}, \quad m_{2k}^{(n)}(t) = \sqrt{\frac{2}{\mathbf{t}_{n+1} - \mathbf{t}_n}} \sin\left(\frac{2k\pi}{\mathbf{t}_{n+1} - \mathbf{t}_n}(t - \mathbf{t}_n) \right), \\ m_{2k+1}^{(n)}(t) &= \sqrt{\frac{2}{\mathbf{t}_{n+1} - \mathbf{t}_n}} \cos\left(\frac{2k\pi}{\mathbf{t}_{n+1} - \mathbf{t}_n}(t - \mathbf{t}_n) \right), \; t \in [\mathbf{t}_n, \mathbf{t}_{n+1}]. \end{aligned} \tag{2.8}$$

Note that taking $N_h = 1$ in (2.8) leads to the piecewise linear interpolation (2.7). Besides, we can also use the wavelet basis, which gives the Levy–Ciesielsky representation [18]. More choices of CONS in (2.4) can be found in [22].

Though any CONS in $L^2([0,T])$ can be used in the spectral approximation (2.4), the CONS we choose here has an important feature: it contains a constant in the basis. Consequently, we have the following relation

$$\int_{t_n}^{t_{n+1}} d\widetilde{W}_l(t) = \Delta W_{l,n}, \quad \Delta W_{l,n} = W_l(t_{n+1}) - W_l(t_n). \tag{2.9}$$

We will show shortly that this relation with certain time discretizations in WZ will lead to the formulation of existing schemes when there is no delay in (2.1).

In this paper, we consider the spectral approximation (2.5) with piecewise constant basis (2.6) and Fourier basis (2.8). With these approximations, we have the following WZ approximation for (2.1):

$$
\begin{aligned}
&\mathrm{d}\tilde{X}(t) = f(\tilde{X}(t), \tilde{X}(t-\tau))\mathrm{d}t + \sum_{l=1}^{r} g_l(\tilde{X}(t), \tilde{X}(t-\tau))\mathrm{d}\widetilde{W}_l(t),\ t \in [0,\, T], \\
&\tilde{X}(t) = \phi(t), t \in (-\tau, 0], \qquad (2.10)
\end{aligned}
$$

where $\widetilde{W}_l(t)$ can be any approximation of $W_l(t)$ described above. For the piecewise linear interpolation (2.7), we have the following consistency of the WZ approximation (2.10) to (2.1).

THEOREM 2.1 (consistency, [40]). *Suppose f and g_l in* (2.1) *are Lipschitz continuous and satisfy conditions* (2.2) *and have second-order continuous and bounded partial derivatives. Suppose also the initial segment $\phi(t), t \in [-\tau, 0]$, to be on the probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{F}_0$-measurable and right continuous, and $\mathbb{E}[\|\phi\|^2_{L^\infty}] < \infty$. For $\tilde{X}(t)$ in* (2.10) *with piecewise linear approximation of Brownian motion* (2.7)*, we have for any $t \in (0,\, T]$,*

$$
\lim_{n\to\infty} \sup_{0\le s\le t} \mathbb{E}[|X(s) - \tilde{X}(s)|^2] = 0. \qquad (2.11)
$$

The consistency of the WZ approximation with spectral approximation (2.5) can be established by the argument of integration by parts as in [12, 17], under similar conditions on the drift and diffusion coefficients.

2.1. Derivation of numerical schemes. We will further discretize (2.10) in time and derive several numerical schemes for (2.1). To this end, we take a uniform time step size h, which satisfies $\tau = mh$ and m is a positive integer; $N_T = T/h$ (T is the final time); $t_n = nh$, $n = 0, 1, \ldots, N_T$. For simplicity, we take the partition for the WZ approximation exactly the same as the time discretization, i.e.,

$$
\mathbf{t}_n = t_n,\ n = 0, 1, \ldots, N_T \ \text{ and } \ \Delta =: \mathbf{t}_n - \mathbf{t}_{n-1} = t_n - t_{n-1} = h.
$$

For (2.10), we have the following integral form over $[t_n, t_{n+1}]$:

$$
\int_{t_n}^{t_{n+1}} \mathrm{d}\tilde{X}(t) = \int_{t_n}^{t_{n+1}} f(\tilde{X}(t), \tilde{X}(t-\tau))\mathrm{d}t + \sum_{l=1}^{r} \int_{t_n}^{t_{n+1}} g_l(\tilde{X}(t), \tilde{X}(t-\tau))\mathrm{d}\widetilde{W}_l(t). \qquad (2.12)
$$

Here we emphasize the following *rule of thumb*: the time discretization for the diffusion coefficients have to be at least half-order. Otherwise, the resulting scheme is not consistent, e.g., Euler-type schemes, in general, converge to the corresponding SDDEs in the Ito sense instead of those in the Stratonovich sense. In fact, if $g_l(\tilde{X}(t), \tilde{X}(t-\tau))$ ($l = 1, \ldots, r$) is approximated by $g_l(\tilde{X}(t_n), \tilde{X}(t_n - \tau))$ in (2.12), then we have, for both Fourier basis (2.8) and piecewise constant basis (2.6),

$$
\int_{t_n}^{t_{n+1}} \mathrm{d}\tilde{X}(t) = \int_{t_n}^{t_{n+1}} f(\tilde{X}(t), \tilde{X}(t-\tau))\mathrm{d}t + \sum_{l=1}^{r} g_l(\tilde{X}(t_n), \tilde{X}(t_n - \tau)\Delta W_{l,n},
$$

where we have used the relation (2.9). This will lead to Euler-type schemes which converge to the following SDDE in the Ito sense (see, e.g., [1, 23], instead of (2.1)):

$$
\mathrm{d}X(t) = f(X(t), X(t-\tau))\mathrm{d}t + \sum_{l=1}^{r} g_l(X(t), X(t-\tau))\mathrm{d}W_l(t).
$$

In the following, three numerical schemes for solving (2.1) are derived using the Taylor expansion and different discretizations in time in (2.12). The first scheme is a *predictor-corrector* scheme. Using the trapezoidal rule to approximate the integrals on the right side of (2.12), we get

$$\begin{aligned} X_{n+1} &= X_n + \frac{h}{2}\big[f(X_n, X_{n-m}) + f(X_{n+1}, X_{n-m+1})\big] \\ &\quad + \frac{1}{2}\sum_{l=1}^{r}\big[g_l(X_n, X_{n-m}) + g_l(X_{n+1}, X_{n-m+1})\big]\Delta W_{l,n}, \end{aligned} \tag{2.13}$$

where X_n is an approximation of $\tilde{X}(t_n)$ (thus an approximation of $X(t_n)$) and we have used the relation (2.9) for both bases (2.6) and (2.8). The initial conditions are $X_n = \phi(nh)$, when $n = -m, -m+1, \ldots, 0$. Note that the scheme (2.13) is fully implicit and is not solvable as $\Delta W_{l,n}$ can take any values in the real line. To resolve this issue, we further apply the left rectangle rule on the right side of (2.12) to obtain a predictor for X_{n+1} in (2.13) so that the resulting scheme is explicit. Taking the relation (2.9) into account, we obtain a predictor-corrector scheme for SDDE (2.1):

$$\begin{aligned} \overline{X}_{n+1} &= X_n + hf(X_n, X_{n-m}) + \sum_{l=1}^{r} g_l(X_n, X_{n-m})\Delta W_{l,n}, \\ X_{n+1} &= X_n + \frac{h}{2}\big[f(X_n, X_{n-m}) + f(\overline{X}_{n+1}, X_{n-m+1})\big] \\ &\quad + \frac{1}{2}\sum_{l=1}^{r}\big[g_l(X_n, X_{n-m}) + g_l(\overline{X}_{n+1}, X_{n-m+1})\big]\Delta W_{l,n}, \\ &\qquad n = 0, 1, \ldots, N_T - 1. \end{aligned} \tag{2.14}$$

THEOREM 2.2. *Assume that f, g_l, $\partial_x g_l g_q$, and $\partial_{x_\tau} g_l g_q$ $(l, q = 1, 2, \ldots, r)$ satisfy the Lipschitz condition* (2.2) *and also the g_l have bounded second-order partial derivatives with respect to all variables. If $\mathbb{E}[\|\phi\|^p_{L^\infty}] < \infty$, $1 \le p \le 4$, then we have for the predictor-corrector scheme* (2.14),

$$\max_{1 \le n \le N_T} \mathbb{E}|X(t_n) - X_n|^2 = O(h). \tag{2.15}$$

The proof will be presented in section 4.

Remark 2.3. When $\tau = 0$ both in drift and diffusion coefficients, the scheme (2.14) degenerates into one family of the predictor-corrector schemes in [2], which can have a larger stability region than the explicit Euler scheme and some other one-step schemes, especially for SDEs with multiplicative noises. Moreover, we will numerically show that if the time delay only exists in the drift term in SDDEs with commutative noise (for the one-dimensional case, i.e., $d = 1$, the commutative condition is $g_l\partial_x g_q - g_q\partial_x g_l = 0$, $1 \le l, q \le r$; see, e.g., [19, p. 348], [29, p. 28]), the proposed predictor-corrector scheme can be convergent with order one in the mean-square sense.

The second scheme is a *midpoint scheme*. Applying the midpoint rule on the right side of (2.12), by $X(t + \frac{h}{2}) \approx \frac{1}{2}(X(t+h) + X(t))$ and (2.9), we obtain the following midpoint scheme:

$$X_{n+1} = X_n + hf\left(\frac{X_n + X_{n+1}}{2}, \frac{X_{n-m} + X_{n-m+1}}{2}\right) + \sum_{l=1}^{r} g_l\left(\frac{X_n + X_{n+1}}{2}, \frac{X_{n-m} + X_{n-m+1}}{2}\right)\overline{\Delta W}_{l,n}, \quad n = 0, 1, \ldots, N_T - 1, \tag{2.16}$$

where we have truncated $\Delta W_{l,n}$ with $\overline{\Delta W}_{l,n}$ so that the solution to (2.16) has finite second-order moment and is solvable (see, e.g., [29, section 1.3]). Here $\overline{\Delta W}_{l,n} = \zeta^{(l,n)}\sqrt{h}$ instead of $\xi^{(l,n)}\sqrt{h}$, where $\zeta^{(l,n)}$ is a truncation of the standard Gaussian random variable $\xi^{(l,n)}$ (see, e.g., [29, p. 39]):

$$\zeta^{(n)} = \xi^{(n)}\chi_{|\xi^{(n)}|\leq A_h} + \mathrm{sgn}(\xi^{(n)})A_h\chi_{|\xi^{(n)}|>A_h}, \quad A_h = \sqrt{4|\log(h)|}. \tag{2.17}$$

This fully implicit midpoint scheme is symplectic if $\tau = 0$ [27], which allows long-time integration for stochastic Hamiltonian systems. As in the case of no delay, the midpoint scheme complies with the Stratonovich calculus without differentiating the diffusion coefficients. Again, it is of first-order convergence for Stratonovich SDEs with commutative noise when no delay arises in the diffusion coefficients. However, it is of half-order convergence once the delay appears in the diffusion coefficients, which will be shown numerically in section 3. The proof of a half-order mean-square convergence is similar to that of Theorem 2.2. Thus, we only present the main idea of the proof here. Using the estimate $\mathbb{E}[(\xi^{(n)})^2 - (\zeta^{(n)})^2] \leq (1 + 4\sqrt{|\ln h|})h^2$ (see [27, Lemma 2.1]) and applying the Taylor expansion for g_l in (2.16), we can prove that the midpoint scheme is of half-order convergence in the mean-square sense as in the proof of Theorem 2.2 for the predictor-corrector scheme.

Remark 2.4. The relation (2.9) is crucial in the derivation of the schemes (2.14) and (2.16). If a CONS contains no constants, e.g., $\{\sqrt{\frac{2}{t_{n+1}-t_n}}\sin(\frac{k\pi(s-t_n)}{t_{n+1}-t_n})\}_{k=1}^{\infty}$, then from (2.12), $\Delta W_{l,n}$ in the scheme (2.14) should be replaced by

$$\int_{t_n}^{t_{n+1}} d\widetilde{W}_l(t) = \sum_{j=1}^{N_h} \sqrt{\frac{2}{t_{n+1} - t_n}} \int_{t_n}^{t_{n+1}} \sin\left(\frac{j\pi(s - t_n)}{t_{n+1} - t_n}\right) ds \xi_{l,j}^{(n)}, \tag{2.18}$$

which will be simulated with independently and identially distributed (i.i.d.) Gaussian random variables with zero mean and variance $\sum_{j=1}^{N_h} \frac{2h}{j^2\pi^2}(1 - (-1)^j)^2$. According to the proof of Theorem 2.2, we require $N_h \sim O(h^{-1})$ so that

$$\mathbb{E}\left[\left|\int_{t_n}^{t_{n+1}} d\widetilde{W}_l(t) - \int_{t_n}^{t_{n+1}} dW_l(t)\right|^2\right] \sim O(h^2) \tag{2.19}$$

to make the corresponding scheme of half-order convergence. Numerical results show that the scheme (2.14) with ΔW_n replaced by (2.18) and $N_h \sim O(h^{-1})$ leads to similar accuracy and the same convergence order with the predictor-corrector scheme (2.14) (numerical results are not present).

The last scheme is a *Milstein-like scheme*. When $s \in [t_n, t_{n+1}]$, we approximate $f(\tilde{X}(s), \tilde{X}(s-\tau))$ by $f(\tilde{X}(t_n), \tilde{X}(t_n-\tau))$, and by the Taylor's expansion we have

$$
\begin{aligned}
g_l(\tilde{X}(s), \tilde{X}(s-\tau)) \approx{} & g_l(\tilde{X}(t_n), \tilde{X}(t_n-\tau)) + \partial_x g_l(\tilde{X}(t_n), \tilde{X}(t_n-\tau))[\tilde{X}(s) - \tilde{X}(t_n)] \\
& + \partial_{x_\tau} g_l(\tilde{X}(t_n), \tilde{X}(t_n-\tau))[\tilde{X}(s-\tau) - \tilde{X}(t_n-\tau)].
\end{aligned} \tag{2.20}
$$

Substituting the above approximations into (2.12) and omitting the terms whose order is higher than one in (2.12), we then obtain the following scheme:

$$
\begin{aligned}
X_{n+1} ={} & X_n + hf(X_n, X_{n-m}) + \sum_{l=1}^{r} g_l(X_n, X_{n-m})\tilde{I}_0 \\
& + \sum_{l=1}^{r}\sum_{q=1}^{r} \partial_x g_l(X_n, X_{n-m}) g_q(X_n, X_{n-m}) \tilde{I}_{q,l,t_n,t_{n+1},0} \\
& + \sum_{l=1}^{r}\sum_{q=1}^{r} \partial_{x_\tau} g_l(X_n, X_{n-m}) g_q(X_{n-m}, X_{n-2m}) \chi_{t_n \ge \tau} \tilde{I}_{q,l,t_n,t_{n+1},\tau}, \\
& \qquad n = 0, 1, \ldots, N_T - 1,
\end{aligned} \tag{2.21}
$$

where

$$
\tilde{I}_0 = \int_{t_n}^{t_{n+1}} \mathrm{d}\widetilde{W}_l(t), \quad \tilde{I}_{q,l,t_n,t_{n+1},0} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{t} \mathrm{d}\widetilde{W}_q(s)\mathrm{d}\widetilde{W}_l(t),\ t_n \ge 0,
$$

$$
\tilde{I}_{q,l,t_n,t_{n+1},\tau} = \int_{t_n}^{t_{n+1}} \int_{t_n-\tau}^{t-\tau} \mathrm{d}\widetilde{W}_q(s)\mathrm{d}\widetilde{W}_l(t),\ t_n \ge \tau.
$$

Using the Fourier basis (2.8), the three stochastic integrals in (2.21) are computed by

$$
\tilde{I}_0^F = \int_{t_n}^{t_{n+1}} m_i^{(1)}(t)\xi_{l,1}^{(n)}\mathrm{d}t = \Delta W_{l,n}, \tag{2.22}
$$

$$
\begin{aligned}
\tilde{I}^F_{q,l,t_n,t_{n+1},0} ={} & \frac{h}{2}\xi_{q,1}^{(n)}\xi_{l,1}^{(n)} - \frac{\sqrt{2}h}{2\pi}\xi_{q,1}^{(n)}\sum_{p=1}^{s}\frac{1}{p}\xi_{l,2p}^{(n)} \\
& + \frac{h}{2\pi}\sum_{p=1}^{s_1}\frac{1}{p}[\xi_{q,2p+1}^{(n)}\xi_{l,2p}^{(n)} - \xi_{q,2p}^{(n)}\xi_{l,2p+1}^{(n)}], \\
\tilde{I}^F_{q,l,t_n,t_{n+1},\tau} ={} & \frac{h}{2}\xi_{q,1}^{(n-m)}\xi_{l,1}^{(n)} - \frac{\sqrt{2}h}{2\pi}\xi_{q,1}^{(n-m)}\sum_{p=1}^{s}\frac{1}{p}\xi_{l,2p}^{(n)} \\
& + \frac{h}{2\pi}\sum_{p=1}^{s_1}\frac{1}{p}[\xi_{q,2p+1}^{(n-m)}\xi_{l,2p}^{(n)} - \xi_{q,2p}^{(n-m)}\xi_{l,2p+1}^{(n)}],
\end{aligned}
$$

where $s = [\frac{N_h}{2}]$ and $s_1 = [\frac{N_h-1}{2}]$. When piecewise constant basis (2.6) is used, these integrals are

$$
\begin{aligned}
\tilde{I}_0^L &= \sum_{j=0}^{N_h-1}\Delta W_{l,n,j} = \Delta W_{l,n}, \\
\tilde{I}^L_{q,l,t_n,t_{n+1},0} &= \sum_{j=0}^{N_h-1}\Delta W_{l,n,j}\left[\frac{\Delta W_{q,n,j}}{2} + \sum_{i=0}^{j-1}\Delta W_{q,n,i}\right], \\
\tilde{I}^L_{q,l,t_n,t_{n+1},\tau} &= \sum_{j=0}^{N_h-1}\Delta W_{l,n,j}\left[\frac{\Delta W_{q,n-m,j}}{2} + \sum_{i=0}^{j-1}\Delta W_{q,n-m,i}\right],
\end{aligned} \tag{2.23}
$$

where $\Delta W_{k,n,j} = W_k(t_n + \frac{(j+1)h}{N_h}) - W_k(t_n + \frac{jh}{N_h})$, $k = 1, \ldots, r$, $j = 0, \ldots, N_h - 1$, and $\Delta W_{k,n,-1} = 0$. In Example 3.3 of section 3, we will show that the piecewise linear interpolation is less efficient than the Fourier approximation for achieving the same order of accuracy.

The scheme (2.21) can be seen as further discretization of the Milstein scheme for Stratonovich SDDEs proposed in [15]:

$$
\begin{aligned}
X_{n+1}^M &= X_n^M + hf(X_n^M, X_{n-m}^M) + \sum_{l=1}^{r} g_l(X_n^M, X_{n-m}^M)\Delta W_{l,n} \\
&\quad + \sum_{l=1}^{r}\sum_{q=1}^{r} \partial_x g_l(X_n^M, X_{n-m}^M) g_q(X_n^M, X_{n-m}^M) I_{q,l,t_n,t_{n+1},0} \\
&\quad + \sum_{l=1}^{r}\sum_{q=1}^{r} \partial_{x_\tau} g_l(X_n^M, X_{n-m}^M) g_q(X_{n-m}^M, X_{n-2m}^M)\chi_{t_n\geq\tau} I_{q,l,t_n,t_{n+1},\tau}, \\
&\qquad n = 0, 1, \ldots, N_T - 1,
\end{aligned}
\tag{2.24}
$$

as the double integrals approximated by either the Fourier expansion or the piecewise linear interpolation, $\tilde{I}_0$, $\tilde{I}_{q,l,t_n,t_{n+1},0}$, and $\tilde{I}_{q,l,t_n,t_{n+1},\tau}$ are, respectively, approximations of the following integrals

$$
I_0 = \int_{t_n}^{t_{n+1}} \circ \mathrm{d}W_l(t), \qquad I_{q,l,t_n,t_{n+1},0} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{t} \circ \mathrm{d}W_q(s) \circ \mathrm{d}W_l(t),\ t_n \geq 0,
$$
$$
I_{q,l,t_n,t_{n+1},\tau} = \int_{t_n}^{t_{n+1}} \int_{t_n-\tau}^{t-\tau} \circ \mathrm{d}W_q(s) \circ \mathrm{d}W_l(t),\ t_n \geq \tau.
$$

In [15], $I_{q,l,t_n,t_{n+1},0}$ and $I_{q,l,t_n,t_{n+1},\tau}$ are approximated in a similar fashion. The Brownian motion W_q therein is approximated by the sum of $(t-t_n)/(t_{n+1}-t_n)W_q(t_n)$ and a truncated Fourier expansion of the Brownian bridge $W_q(t) - (t - t_n)/(t_{n+1} - t_n)W_q(t_n)$ for $t_n \leq t \leq t_{n+1}$; see also [9] and [19, section 5.8]. It can be readily checked that this approximation is equivalent to the Fourier approximation (2.22). In numerical simulations (results are not present), these two approximations lead to a small difference in computational cost and accuracy but the convergence order is the same.

As we note in the beginning of section 2, the choice of complete orthonormal bases is arbitrary. However, the use of general spectral approximation may lead to different accuracy; see, e.g., [22] for a detailed comparison of some spectral approximations of multiple Stratonovich integrals.

In addition to the Fourier approximation, several methods of approximating $I_{q,l,t_n,t_{n+1},0}$ have been proposed: applying the trapezoid rule (see, e.g., [29, section 1.4]) and the modified Fourier approximation (see, e.g., [45]). We note that the use of the trapezoid rule leads to a formula similar to (2.23), which is shown to be less efficient than the Fourier approximation; see Example 3.3 of section 3. In [45], $I_{q,l,t_n,t_{n+1},0}$ is approximated with the sum of a Fourier approximation and a tail process $A_{q,l,t_n,t_{n+1}}$, where the tail $A_{q,l,t_n,t_{n+1}}$ is modeled with the product of $r(r-1)/2$-dimensional i.i.d. Gaussian random variables and a functional of increments of Brownian motion $\Delta W_{l,n}$. It is shown in [9] that the modified Fourier approximation in [45] requires $O(r^4\sqrt{h})$ i.i.d. Gaussian random variables to maintain the first-order convergence while the Fourier approximation requires $O(r^2h^{-1})$ i.i.d. Gaussian random variables. However, it is difficult to extend this approach to approximate

$I_{q,l,t_n,t_{n+1},\tau}$ even when r is small because a tail $A_{q,l,t_n,t_{n+1},0}$ will be correlated with $A_{q,l,t_n,t_{n+1},\tau}$, which is difficult to identify and brings no computational benefits.

To make the scheme (2.21) of first-order convergence, it is important to efficiently compute the double integrals $\tilde{I}_{q,l,t_n,t_{n+1},0}$ and $\tilde{I}_{q,l,t_n,t_{n+1},\tau}$. In the following, we discuss how to choose the truncation parameters for the spectral approximation of Brownian motion.

LEMMA 2.5. *For the Fourier basis* (2.8)*, it holds that*

$$\tilde{I}_0^F = I_0, \tag{2.25}$$

$$\mathbb{E}[(\tilde{I}^F_{q,l,t_n,t_{n+1},0} - I_{q,l,t_n,t_{n+1},0})^2] = \varsigma(N_h)\frac{2\Delta^2}{(N_h\pi)^2} + \sum_{i=M}^{\infty}\frac{\Delta^2}{(i\pi)^2} \leq c\frac{\Delta^2}{\pi^2 M}, \tag{2.26}$$

$$\mathbb{E}[(\tilde{I}^F_{q,l,t_n,t_{n+1},\tau} - I_{q,l,t_n,t_{n+1},\tau})^2] = \varsigma(N_h)\frac{2\Delta^2}{(N_h\pi)^2} + \sum_{i=M}^{\infty}\frac{\Delta^2}{(i\pi)^2} \leq c\frac{\Delta^2}{\pi^2 M}, \tag{2.27}$$

where $\varsigma(N_h) = 0$ if N_h is odd and 1 otherwise, and M is the integer part of $N_h/2+1$.

The proof of this lemma can be found in section 4. With Lemma 2.5, we can show that the Milstein-like scheme (2.21) can be of first-order convergence in the mean-square sense; see section 4.

THEOREM 2.6. *Assume that f, g_l, $\partial_x g_l g_q$, and $\partial_{x_\tau} g_l g_q$ $(l, q = 1, 2, \ldots, r)$ satisfy the Lipschitz condition* (2.2) *and also the g_l have bounded second-order partial derivatives with respect to all variables. If $\mathbb{E}[\|\phi\|^p_{L^\infty}] < \infty$, $1 \leq p \leq 4$, then we have for the Milstein-like scheme* (2.21),

$$\max_{1\leq n\leq N_T} \mathbb{E}|X(t_n) - X_n|^2 = O(h^2), \tag{2.28}$$

when the double integrals $\tilde{I}_{q,l,t_n,t_{n+1},0}$, $\tilde{I}_{q,l,t_n,t_{n+1},\tau}$ are computed by (2.22) *and N_h is of the order of $1/h$.*

When (2.23) is used in the Milstein-like scheme (2.21), the first-order strong convergence can be proved similarly when N_h is of the order of $1/h$. This is similar to the simulation of double integrals using the Fourier approximation in [15, 20].

Remark 2.7. In practice, the cost of simulating double integrals is prohibitively expensive. However, there are cases where we can reduce or even avoid the simulation of double integrals. For example, when the diffusion coefficients are small and of the order ϵ and the coefficients at the double integrals are of order ϵ^2, we may take $\epsilon^2/M \sim O(h)$ to achieve an accuracy of $\mathcal{O}(h)$ in the mean-square sense, according to the proof of Theorem 2.6. Thus, only a small M is required if $\epsilon \sim O(\sqrt{h})$. Also, when the diffusion coefficients contain no delay and satisfy the so-called commutative noises, i.e., $\frac{\partial g_l(x)}{\partial x} g_q(x) = \frac{\partial g_q(x)}{\partial x} g_l(x)$, the Milstein-like scheme can be rewritten as

$$\begin{aligned}
X^M_{n+1} &= X^M_n + hf(X^M_n, X^M_{n-m}) + \sum_{l=1}^{r} g_l(X^M_n)\Delta W_{l,n} \\
&\quad + \frac{1}{2}\sum_{l=1}^{r}\sum_{q=1}^{r}\partial_x g_l(X^M_n) g_q(X^M_n)\Delta W_{l,n}\Delta W_{q,n}, \quad n = 0, 1, \ldots, N_T - 1.
\end{aligned}$$

In this case, only Wiener increments are used and the Milstein scheme is of low cost.

3. Numerical results. In this section, we test the convergence order of the proposed schemes and compare their numerical performance. In the first two examples, we test the predictor-corrector scheme (2.14) and midpoint scheme (2.16) for multiple noises and show both methods are of half-order mean-square convergence. Further,

TABLE 1

Convergence order of the predictor-corrector scheme (left) and the midpoint scheme (right) for Example 3.1 at $T = 20$ with ten white noises using $n_p = 4000$ sample paths.

τ	h	$\rho_{h,T}$	Order	Time (s.)	$\rho_{h,T}$	Order	Time (s.)
	2^{-8}	4.050e-02	0.51	220	6.698e-02	0.61	807
	2^{-9}	2.851e-02	0.56	420	4.378e-02	0.44	1591
$\frac{1}{16}$	2^{-10}	1.936e-02	0.48	818	3.237e-02	0.53	3157
	2^{-11}	1.390e-02	$*^a$	1620	2.239e-02	$*$	6289
	2^{-12}	$*$	$*$	3221	$*$	$*$	12573
	2^{-8}	3.955e-02	0.54	219	5.942e-02	0.51	805
	2^{-9}	2.728e-02	0.55	417	4.189e-02	0.48	1586
$\frac{1}{4}$	2^{-10}	1.861e-02	0.46	816	3.000e-02	0.47	3148
	2^{-11}	1.359e-02	$*$	1615	2.164e-02	$*$	6284
	2^{-12}	$*$	$*$	3215	$*$	$*$	12557
	2^{-8}	3.914e-02	0.53	221	5.725e-02	0.48	805
	2^{-9}	2.711e-02	0.54	422	4.092e-02	0.51	1588
1	2^{-10}	1.873e-02	0.46	820	2.887e-02	0.47	3148
	2^{-11}	1.364e-02	$*$	1627	2.086e-02	$*$	6289
	2^{-12}	$*$	$*$	3236	$*$	$*$	12569

[a] No results from a smaller time step size are available and the convergence order is absent.

we show that both schemes converge with order one in the mean-square sense for an SDDE with single white noise and no time delay in diffusion coefficients. In the last example, we test the Milstein-like scheme (2.21) and show that it is of first-order convergence for SDDEs with multiple white noises.

Throughout this section, the strong error of numerical solutions is defined as

$$\rho_{h,T} = \left(\frac{1}{n_p} \sum_{i=1}^{n_p} |X_h(T,\omega_i) - X_{\frac{h}{2}}(T,\omega_i)|^2 \right)^{1/2},$$

where ω_i denotes the ith single sample path and n_p is the number of paths.

The numerical tests were performed using MATLAB R2012a on a Dell Optiplex 780 computer with CPU (E8500 3.16 GHz). We used the Mersenne twister random generator with seed 1 and took a large number of paths so that the statistical error can be ignored. Newton's method with tolerance $h^2/100$ was used to solve the nonlinear algebraic equations at each step of the implicit schemes.

We first test the convergence order of the predictor-corrector scheme (2.14) and the midpoint scheme (2.16) for an SDDE with several noises.

Example 3.1. Consider (2.1) with the following coefficients:

$$\begin{aligned}
&f = -15X(t) + 2\sin(X(t-\tau)),\\
&g_1 = \sin(X(t)) + 0.5X(t-\tau), \quad g_2 = 0.9X(t), \quad g_3 = 0.2X(t) + 0.2X(t-\tau),\\
&g_4 = 2\sin(X(t)), \quad g_5 = 0.8X(t) + \cos(X(t-\tau)), \quad g_6 = X(t) + 0.5\sin(X(t-\tau)),\\
&g_7 = 2\cos(X(t-\tau)), \quad g_8 = -X(t) + \cos(X(t-\tau)), \quad g_9 = 0.5X(t) - X(t-\tau),\\
&g_{10} = 1.5\cos(X(t-\tau)),
\end{aligned}$$

and the initial function is $\phi(t) = t + 0.2$.

In this example, we test the convergence order of the predictor-corrector and midpoint schemes at $T = 20$ and with different time delays $\tau = 2^{-4}$, 2^{-2}, 1.

In Table 1, we observe that both schemes are convergent with order half in the mean-square sense. Different time delays do not influence the convergence order of these schemes.

The number of operations of the predictor-corrector scheme for (2.1) is $(5r+6)dT/h$. For the midpoint scheme, the number of operations is $6C_I(r+1)dT/h$, where C_I is the maximum number of Newton's iterations in each time step. In Table 1, we observe that for both schemes, the computational cost doubles when step sizes reduce by half. The CPU time of the midpoint scheme is about four times what the predictor-corrector scheme costs, which is consistent with the prediction as the observed C_I is around 4.

We now test the convergence order for the predictor-corrector scheme (2.14) and the midpoint scheme (2.16) for SDDEs with different types of noises: noncommutative noise, single noise. We will show that the time delay in a diffusion coefficient keeps both methods only convergent at half-order, while for the SDDE with single noise, the two schemes can be of first-order accuracy in the mean-square sense if the time delay does not appear explicitly in the diffusion coefficients.

Example 3.2. Consider (2.1) in one-dimension and assume the initial function $\phi(t) = t + 0.2$, with different diffusion coefficients:

- noncommutative white noises without delay in the diffusion coefficients:
$$\mathrm{d}X = [-X(t)+\sin(X(t-\tau))]\,\mathrm{d}t+\sin(X(t))\circ \mathrm{d}W_1(t)+0.5X(t)\circ \mathrm{d}W_2(t), \tag{3.1}$$
where the noises are noncommutative as $\partial_x(sin(x))0.5x-\partial_x(0.5x)sin(x) \neq 0$;
- single white noise without delay in the diffusion coefficients:
$$\mathrm{d}X = [-X(t)+\sin(X(t-\tau))]\,\mathrm{d}t+\sin(X(t))\circ \mathrm{d}W(t); \tag{3.2}$$
- single white noise with delay in the diffusion coefficients:
$$\mathrm{d}X = [-X(t)+\sin(X(t-\tau))]\,\mathrm{d}t+\sin(X(t-\tau))\circ \mathrm{d}W(t). \tag{3.3}$$

From Figure 1(a) (noncommutative noises, (3.1)) and Figure 1(c) (single delayed diffusion, (3.3)), we observe the half-order strong convergence. In contrast, for (3.2) (single noise, nondelayed diffusion) in Figure 1(b), the convergence order of these two schemes becomes one in the mean-square sense.

From this example, we conclude that for the predictor-corrector and midpoint schemes, when the time delay only appears in the drift term, the convergence order is one for the equation with single noise (commutative noises) and half for the one with noncommutative noises. However, when the diffusion coefficients contain time delays, these two schemes are only half-order even for equations with a single white noise; see equation (3.3).

In the last example, we test the Milstein-like scheme (2.21) using different bases, i.e., the piecewise constant basis (2.6) and the Fourier basis (2.8), and compare its numerical performance with the predictor-corrector and midpoint schemes. For the Milstein-like scheme, we show that for multiple noises, the computational cost for achieving the same accuracy is much higher than the other two schemes, while for single noise, the computational cost for the same accuracy is lower.

Example 3.3. We consider the Milstein-like scheme (2.21) for

$$\begin{aligned}\mathrm{d}X(t) &= [-9X(t)+\sin(X(t-\tau))]\mathrm{d}t+[\sin(X(t))+X(t-\tau)]\circ \mathrm{d}W_1(t)\\ &\quad+[X(t)+\cos(0.5X(t-\tau))]\circ \mathrm{d}W_2(t),\ t\in(0,T],\\ X(t) &= t+\tau+0.1,\ t\in[-\tau,0]\end{aligned} \tag{3.4}$$

and

$$\begin{aligned}\mathrm{d}X(t) &= [-2X(t)+2X(t-\tau)]\mathrm{d}t+[\sin(X(t))+X(t-\tau)]\circ \mathrm{d}W(t),\ t\in(0,T],\\ X(t) &= t+\tau,\ t\in[-\tau,0].\end{aligned} \tag{3.5}$$

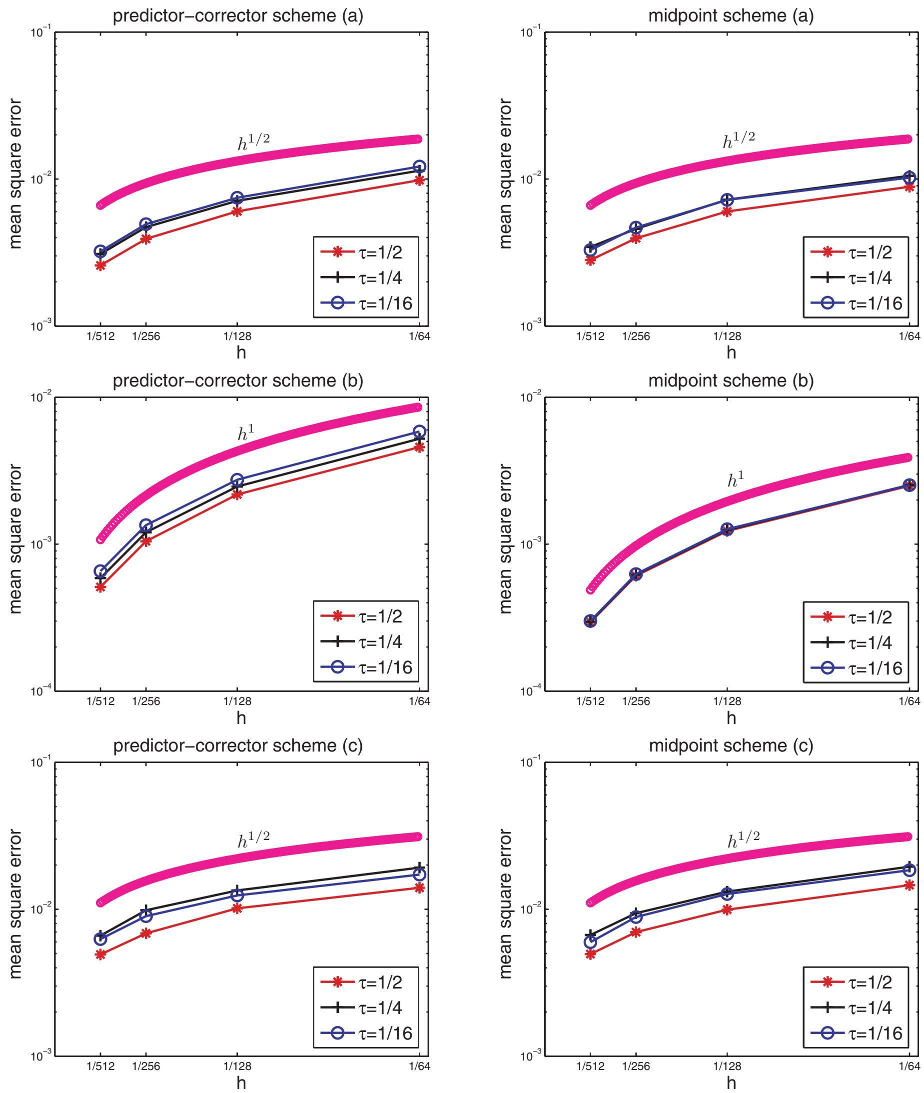

FIG. 1. *Mean-square convergence test of the predictor-corrector (left column) and midpoint schemes (right column) on Example* 3.2 *at* $T = 5$ *with different* τ *using* $n_p = 10000$ *sample paths.* (a) *multi-white-noises with nondelayed diffusion coefficient;* (b) *single white noise with nondelayed diffusion coefficient;* (c) *single white noise with delayed diffusion coefficient.*

To reduce the computational cost, the double integrals are computed by the Fourier expansion approximation (2.22) and the following relation

$$\tilde{I}_{q,l,t_n,t_{n+1},0} = \Delta W_{l,n}\Delta W_{q,n} - \tilde{I}_{l,q,t_n,t_{n+1},0}, \quad \tilde{I}_{l,l,t_n,t_{n+1},0} = \frac{(\Delta W_{l,n})^2}{2}. \tag{3.6}$$

We also use the following relations:

$$\tilde{I}_{q,l,t_n,t_n+ph,0} = \sum_{j=0}^{p-1}\left[\tilde{I}_{q,l,t_n+jh,t_n+(j+1)h,0} + \Delta W_{l,n+j}\chi_{j\geq 1}\sum_{i=0}^{j-1}\Delta W_{q,n+i}\right],$$

$$\tilde{I}_{q,l,t_n,t_n+ph,\tau} = \sum_{j=0}^{p-1}\left[\tilde{I}_{q,l,t_n+jh,t_n+(j+1)h,\tau} + \Delta W_{l,n+j}\chi_{j\geq 1}\sum_{i=0}^{j-1}\Delta W_{q,n-m+i}\right].$$

TABLE 2

Convergence order of the Milstein-like scheme (left) for (3.4) *at* $T = 1$ *and comparison with the convergence order of the predictor-corrector scheme (middle) and the midpoint scheme (right) using* $n_p = 4000$ *sample paths. The upper rows are with* $\tau = 1/16$ *and the lower are with* $\tau = 1/4$.

h	$\rho_{h,T}$	Order	Time (s.)	$\rho_{h,T}$	Order	Time (s.)	$\rho_{h,T}$	Order	Time (s.)
2^{-4}	9.832e-02	1.27	0.72	7.164e-02	0.94	0.05	5.000e-02	0.60	0.16
2^{-5}	4.090e-02	1.09	1.0	3.734e-02	0.69	0.10	3.304e-02	0.55	0.29
2^{-6}	1.921e-02	0.99	1.7	2.308e-02	0.51	0.12	2.263e-02	0.51	0.41
2^{-7}	9.703e-03	$*^a$	3.3	1.616e-02	$*$	0.25	1.590e-02	$*$	0.79
2^{-8}	$*$	$*$	6.4	$*$	$*$	0.40	$*$	$*$	1.54
2^{-4}	9.307e-02	1.28	0.56	6.956e-02	0.96	0.04	5.050e-02	0.68	0.11
2^{-5}	3.824e-02	1.08	0.93	3.582e-02	0.70	0.10	3.155e-02	0.56	0.22
2^{-6}	1.804e-02	0.99	1.6	2.205e-02	0.62	0.17	2.133e-02	0.58	0.39
2^{-7}	9.069e-03	$*$	2.8	1.434e-02	$*$	0.26	1.425e-02	$*$	0.78
2^{-8}	$*$	$*$	5.5	$*$	$*$	0.45	$*$	$*$	1.59

[a]No results from a smaller time step size are available and the convergence order is absent.

TABLE 3

Convergence order of the Milstein-like scheme (left) for (3.5) *(single white noise) at* $T = 1$ *and comparison with the convergence order of the predictor-corrector scheme (middle) and midpoint scheme (right) using* $n_p = 4000$ *sample paths. The delay* τ *is taken as* $1/4$.

h	$\rho_{h,T}$	Order	Time (s.)	h	$\rho_{h,T}$	Order	Time (s.)	$\rho_{h,T}$	Order	Time (s.)
2^{-4}	3.164e-02	0.91	0.19	2^{-7}	1.252e-02	0.44	0.18	1.263e-02	0.45	0.59
2^{-5}	1.688e-02	0.99	0.28	2^{-8}	9.219e-03	0.51	0.37	9.246e-03	0.51	1.09
2^{-6}	8.499e-03	0.90	0.46	2^{-9}	6.462e-03	0.49	0.56	6.471e-03	0.48	2.05
2^{-7}	4.570e-03	$*^a$	0.79	2^{-10}	4.617e-03	$*$	1.03	4.627e-03	$*$	3.97
2^{-8}	$*$	$*$	1.40	2^{-11}	$*$	$*$	1.91	$*$	$*$	7.58

[a]No results from a smaller time step size are available and the convergence order is absent.

In Table 2, we show that for (3.4), the Milstein-like scheme (2.21) converges with order one in the mean-square sense. Compared to the predictor-corrector scheme or the midpoint scheme, when the time step sizes are the same, the computational cost for the Milstein-like scheme (2.21) is several times higher. In fact, in the Milstein-like scheme, the extra computational cost comes from evaluating the double integrals $\tilde{I}^F_{q,l,t_n,t_{n+1},0}$ and $\tilde{I}^F_{q,l,t_n,t_{n+1},\tau}$ at each time step, which requires $7/(2h)(3r^2 - r)/2$ operations when we take the relation (3.6) into account.

We also test the Milstein-like scheme (2.21) using the piecewise constant basis (2.6). The computational cost is even higher than that of using the Fourier basis for the same time step size. Actually, the number of operations for evaluating double integrals using (2.23) is $(1/(2h^2)+5/(2h)-1)(3r^2-r)/2$, which is $O(1/h^2)$, much higher than that of using the Fourier basis, $O(1/h)$. Our numerical tests (not presented here) confirmed the fast increase in the number of operations.

However, the number of operations of the Milstein-like scheme can be significantly reduced when there is just a single diffusion coefficient. In Table 3, we observe that the Milstein-like scheme for (3.5) is still of first-order convergence but the predictor-corrector scheme and the midpoint scheme are only of half-order convergence. For the same accuracy, the computational cost for the Milstein-like scheme using the Fourier basis is less than that for the other two schemes. In fact, for single noise, we only need to compute one double integral $\tilde{I}_{1,1,t_n,t_{n+1},\tau}$. Moreover, when the coefficients of the diffusion coefficient are small, a small number of Fourier modes is required for large time step sizes, i.e., N_h can be $O(1)$ instead of $O(h^{-1})$. The computational cost can thus be reduced somewhat; see, e.g., [28] and [29, Chapter 3] for such a discussion for equations with small noises without delay.

In summary, the proposed predictor-corrector scheme and midpoint scheme are convergent with half-order in the mean-square sense; see Example 3.1. We also show that these two schemes can be of first-order in the mean-square sense if the underlying SDDEs with single noise (commutative noise) and the time delay are only in the drift coefficients; see Example 3.2. In Example 3.3 the numerical tests show that our proposed Milstein-like scheme is of first-order in the mean-square sense for SDDEs with noncommutative noise wherever the time delay appears, i.e., in the drift and/or diffusion coefficients. Compared to the other two schemes, the Milstein-like scheme is more accurate but is more expensive as it requires evaluations of double integrals, with cost inversely proportional to the time step size and proportional to the square of the number of noises. However, for SDDEs with single noise, the Milstein-like scheme (with the Fourier basis) can be superior to the predictor-corrector scheme and the midpoint scheme in terms of both accuracy and computational cost.

4. Proofs. In this section, we prove Theorems 2.2 and 2.6 and Lemma 2.5. While proofs of Theorems 2.2 and 2.6 are presented only for the one-dimensional problem (2.1) ($d = 1$), they can be extended to the multidimensional case $d > 1$ without difficulty.

Proof of Theorem 2.2. We recall that for the Milstein scheme (2.24) (see [15]), $\max_{1\le n\le N_T} \mathbb{E}|X(t_n) - X_n^M|^2 = O(h^2)$. Then by the triangle inequality, it suffices to prove

$$\max_{1\le n\le N_T} \mathbb{E}|X_n^M - X_n|^2 = O(h). \tag{4.1}$$

We denote that $f_n = f(X_n, X_{n-m})$ and $g_{l,n} = g_l(X_n, X_{n-m})$ and also

$$\begin{aligned} \rho_{f_n} &= f(\overline{X}_{n+1}, X_{n-m+1}) - f_n, \\ \rho_{g_{l,n}} &= g_l(\overline{X}_{n+1}, X_{n-m+1}) \\ &\quad - \left[g_{l,n} + \partial_x g_{l,n} \sum_{q=1}^r g_{q,n}\Delta W_{q,n} + \partial_{x_\tau} g_{l,n} \sum_{q=1}^r g_{q,n-m}\Delta W_{q,n-m} \right]. \end{aligned} \tag{4.2}$$

With (4.2), we can rewrite (2.14) as follows:

$$\begin{aligned} X_{n+1} = X_n + hf_n + \sum_{l=1}^r g_{l,n}\Delta W_{l,n} + \frac{1}{2}\sum_{l=1}^r\sum_{q=1}^r \partial_x g_{l,n}\Delta W_{q,n}\Delta W_{l,n} \\ + \frac{1}{2}\sum_{l=1}^r\sum_{q=1}^r \partial_{x_\tau} g_{l,n}\; g_{q,n-m}\Delta W_{q,n-m}\Delta W_{l,n} + \rho_n, \end{aligned} \tag{4.3}$$

where $\rho_n = h\rho_{f_n} + \frac{1}{2}\sum_{l=1}^r \rho_{g_{l,n}}\Delta W_{l,n}$.

It can be readily checked that if f, g_l satisfy the Lipschitz condition (2.2), and g_l has bounded second-order derivatives ($l = 1, \ldots, r$), then by the predictor-corrector scheme (2.14) and Taylor's expansion of $g_l(\overline{X}_{n+1}, X_{n-m+1})$, we have $h^2\mathbb{E}[\rho_{f_n}^2] \le Ch^3$, $\mathbb{E}[(\rho_{g_{l,n}}\Delta W_{l,n})^2] \le Ch^3$, and thus by the triangle inequality,

$$\mathbb{E}[\rho_n^2] \le Ch^3, \tag{4.4}$$

where the constant C depends on r and Lipschitz constants, but is independent of h.

Subtracting (4.3) from (2.24) and taking expectation after squaring over both sides, we have

$$\begin{aligned}\mathbb{E}[(X_{n+1}^M - X_{n+1})^2] = \mathbb{E}[(X_n^M - X_n)^2] + 2\mathbb{E}\left[(X_n^M - X_n)\left(\sum_{i=0}^{4} R_i - \rho_n\right)\right] \\ - 2\sum_{i=0}^{4}\mathbb{E}[\rho_n R_i] + \sum_{i,j=0}^{4}\mathbb{E}[R_i R_j] + \mathbb{E}[\rho_n^2],\end{aligned} \tag{4.5}$$

where we denote $f_n^M = f(X_n^M, X_{n-m}^M)$ and $g_{l,n}^M = g_l(X_n^M, X_{n-m}^M)$ and

$$\begin{aligned}
R_0 &= h(f_n^M - f_n) + \sum_{l=1}^{r}(g_{l,n}^M - g_{l,n})\Delta W_{l,n},\\
R_1 &= \sum_{l=1}^{r}\sum_{q=1}^{r}\left[\partial_x g_{l,n}^M g_{q,n}^M - \partial_x g_{l,n} g_{q,n}\right]\frac{\Delta W_{q,n}\Delta W_{l,n}}{2},\\
R_2 &= \sum_{l=1}^{r}\sum_{q=1}^{r}\left[\partial_{x_\tau} g_{l,n}^M g_{q,n-m}^M - \partial_{x_\tau} g_{l,n} g_{q,n-m}\right]\frac{\Delta W_{q,n-m}\Delta W_{l,n}}{2},\\
R_3 &= \sum_{l=1}^{r}\sum_{q=1}^{r}\partial_x g_{l,n}^M g_{q,n}^M\left(I_{q,l,t_n,t_{n+1},0} - \frac{\Delta W_{q,n}\Delta W_{l,n}}{2}\right),\\
R_4 &= \sum_{l=1}^{r}\sum_{q=1}^{r}\partial_{x_\tau} g_{l,n}^M g_{q,n-m}^M\left(I_{q,l,t_n,t_{n+1},\tau} - \frac{\Delta W_{q,n-m}\Delta W_{l,n}}{2}\right).
\end{aligned}$$

By the Lipschitz condition for f and g_l, and adaptedness of X_n, X_n^M, we have

$$\mathbb{E}[R_0^2] \le C(h^2 + h)(\mathbb{E}[(X_n^M - X_n)^2] + \mathbb{E}[(X_{n-m}^M - X_{n-m})^2]). \tag{4.6}$$

To bound $\mathbb{E}[R_i^2]$ $(i = 1, 2, 3, 4)$, we require that X_n and X_n^M have bounded moments of up to fourth order, which can be readily checked using the predictor-corrector scheme (2.14) and the milstein scheme (2.24) under our assumptions. By the Lipschitz condition of g_l and $\partial_{x_\tau} g_l g_q$, we have

$$\mathbb{E}[R_2^2] \le C \max_{1\le l,q\le r}\mathbb{E}[(\left|X_n^M - X_n\right| + \left|X_{n-m}^M - X_{n-m}\right|)^2(\Delta W_{q,n-m}\Delta W_{l,n})^2],$$

whence by the Cauchy inequality and the boundedness of $\mathbb{E}[X_n^4]$ and $\mathbb{E}[(X_n^M)^4]$, we have $\mathbb{E}[R_2^2] \le Ch^2$. Similarly, we have $\mathbb{E}[R_1^2] \le Ch^2$. By Lemma 2.5, and linear growth condition (2.3) for $\partial_{x_\tau} g_l g_q$, we obtain

$$\begin{aligned}\mathbb{E}[R_4^2] &\le C \max_{1\le l<q\le r}\mathbb{E}\left[(1 + \left|X_n^M\right|^2 + \left|X_{n-m}^M\right|^2)\left(I_{q,l,t_n,t_{n+1},\tau} - \frac{\Delta W_{q,n-m}\Delta W_{l,n}}{2}\right)^2\right]\\ &\le Ch^2,\end{aligned}$$

since X_n^M, X_{n-m}^M have bounded fourth-order moments and by the Burkholder–Davis–

Gundy inequality, it holds that for $l \neq q$

$$\begin{aligned}
&\mathbb{E}\left[\left(I_{q,l,t_n,t_{n+1},\tau} - \frac{\Delta W_{q,n-m}\Delta W_{l,n}}{2}\right)^4\right] \\
&= \mathbb{E}\left[\left(\int_{t_n}^{t_{n+1}}\left(W_q(t-\tau) - \frac{W_q(t_{n+1}-\tau)+W_q(t_n-\tau)}{2}\right)\circ dW_l\right)^4\right] \\
&\le C\left(\mathbb{E}\left[\int_{t_n}^{t_{n+1}}\left(W_q(t-\tau) - \frac{W_q(t_{n+1}-\tau)+W_q(t_n-\tau)}{2}\right)^2 ds\right]\right)^2 \le Ch^4.
\end{aligned}$$

Similarly, we have $\mathbb{E}[R_3^2] \le Ch^2$. Thus we have proved that

$$\mathbb{E}[R_i^2] \le Ch^2, \quad i = 1,2,3,4. \tag{4.7}$$

By the basic inequality $2ab \le a^2 + b^2$, we have

$$2\left|\mathbb{E}[(X_n^M - X_n)\rho_n]\right| \le h\mathbb{E}[(X_n^M - X_n)^2] + h^{-1}\mathbb{E}[\rho_n^2]. \tag{4.8}$$

By the fact that X_n and X_n^M are $\mathcal{F}_{t_n}$-measurable and the Lipschitz condition for f,

$$\begin{aligned}
2\mathbb{E}[(X_n^M - X_n)R_0] &= 2h\mathbb{E}[(X_n^M - X_n)(f_n - f_{n-m})] \\
&\le Ch(\mathbb{E}[(X_n^M - X_n)^2] + \mathbb{E}[(X_{n-m}^M - X_{n-m})^2]).
\end{aligned} \tag{4.9}$$

Further, by the Lipschitz condition (2.2) for $\partial_x g_l g_l$, we have

$$\begin{aligned}
2\mathbb{E}[(X_n^M - X_n)R_1] &= \sum_{l=1}^{r}\mathbb{E}[(X_n^M - X_n)\partial_x g_{l,n}^M g_{l,n}^M - \partial_x g_{l,n} g_{l,n})]\mathbb{E}[(\Delta W_{l,n})^2] \\
&\le Ch(\mathbb{E}[(X_n^M - X_n)^2] + \mathbb{E}[(X_{n-m}^M - X_{n-m})^2]).
\end{aligned} \tag{4.10}$$

By the adaptedness of X_n, X_n^M, and $\mathbb{E}[\Delta W_{l,n}] = \mathbb{E}[(I_{q,l,t_n,t_{n+1},0} - \frac{\Delta W_{q,n}\Delta W_{l,n}}{2})] = 0$, we have

$$\mathbb{E}[(X_n^M - X_n)R_i] = 0, \quad i = 2,3. \tag{4.11}$$

Again by the adaptedness of X_n and X_n^M, we can have

$$\mathbb{E}[(X_n^M - X_n)R_4] = 0. \tag{4.12}$$

In fact, by Lemma 2.5, we can represent $I_{q,l,t_n,t_{n+1},\tau}$ as

$$I_{q,l,t_n,t_{n+1},\tau} = \frac{h}{2}\xi_{q,1}^{(n-m)}\xi_{l,1}^{(n)} + \frac{h}{2\pi}\sum_{p=1}^{\infty}\frac{1}{p}[\xi_{q,2p+1}^{(n)}\xi_{l,2p}^{(n-m)} - \xi_{q,2p}^{(n-m)}\xi_{l,2p+1}^{(n)} - \sqrt{2}\xi_{q,1}^{(n-m)}\xi_{l,2p}^{(n)}]. \tag{4.13}$$

Then by the facts $\mathbb{E}[\left|(X_n^M - X_n)R_4\right|] \le (\mathbb{E}[(X_n^M - X_n)^2])^{1/2}(\mathbb{E}[R_4^2])^{1/2} \le Ch$ and $\mathbb{E}[(X_n^M - X_n)\xi_{l,k}^{(n)}] = 0$ for any $k \ge 1$, we obtain (4.12) from Lebesgue's dominated convergence theorem.

By (4.11)–(4.12) and the Cauchy inequality, from (4.5) we have, for $n \geq m$,

$$\begin{aligned}&\mathbb{E}[(X_{n+1}^M - X_{n+1})^2]\\ &\quad\leq \mathbb{E}[(X_n^M - X_n)^2] + 2\mathbb{E}[(X_n^M - X_n)(R_0 + R_1 - \rho_n)] + C\sum_{i=0}^{4}\mathbb{E}[R_i^2] + C\mathbb{E}[\rho_n^2]\end{aligned}$$

and further by (4.4), (4.6)–(4.8), and (4.9)–(4.10), we obtain, for $n \geq m$,

$$\begin{aligned}&\mathbb{E}[X_{n+1}^M - X_{n+1}]^2\\ &\quad\leq (1+Ch)\mathbb{E}[(X_n^M - X_n)^2] + Ch\mathbb{E}[(X_{n-m}^M - X_{n-m})^2])\\ &\qquad + (C + h^{-1})\mathbb{E}[\rho_n^2] + C\sum_{i=0}^{4}\mathbb{E}[R_i^2]\\ &\quad\leq (1+Ch)\mathbb{E}[(X_n^M - X_n)^2] + Ch\mathbb{E}[(X_{n-m}^M - X_{n-m})^2] + Ch^2,\end{aligned} \tag{4.14}$$

where C is independent of h. Similarly, we can obtain that (4.14) holds for $n = 1, \ldots, m-1$. Taking the maximum over both sides of (4.14) and noting that $X_i^M - X_i = 0$ for $-m \leq i \leq 0$, we have

$$\max_{1\leq i\leq n+1}\mathbb{E}[(X_i^M - X_i)^2] \leq (1+Ch)\max_{1\leq i\leq n}\mathbb{E}[(X_i^M - X_i)^2] + Ch^2.$$

Then (4.1) follows from the discrete Gronwall inequality. $\square$

Proof of Lemma 2.5. From (2.22), the formula (2.25) can be readily obtained. Now we consider (2.26). For $l = q$, it holds that

$$\tilde{I}_{l,l,t_n,t_{n+1},0} = I_{l,l,t_n,t_{n+1},0} = (\Delta W_{l,n})^2/2$$

if (2.5) with either piecewise constant basis (2.6) or Fourier basis (2.8) is used. For any orthogonal expansion (2.4), we have

$$\mathbb{E}[\int_{t_n}^{t_{n+1}}(\widetilde{W}_q(s) - W_q(s))\,dW_l\int_{t_n}^{t_{n+1}}\widetilde{W}_q(s)\,d(\widetilde{W}_l - W_l)] = 0$$

and thus by $W_q(t_n) = \widetilde{W}_q(t_n)$, Ito's isometry, and integration by parts, we have, when $l \neq q$,

$$\begin{aligned}&\mathbb{E}[(\tilde{I}_{q,l,t_n,t_{n+1},0} - I_{q,l,t_n,t_{n+1},0})^2]\\ &\quad= \mathbb{E}\left[\left(\int_{t_n}^{t_{n+1}}[\widetilde{W}_q(s) - W_q(s)]\circ\, dW_l + \int_{t_n}^{t_{n+1}}\widetilde{W}_q(s)\,d[\widetilde{W}_l - W_l]\right)^2\right]\\ &\quad= \mathbb{E}\left[\left(\int_{t_n}^{t_{n+1}}[\widetilde{W}_q(s) - W_q(s)]\,dW_l\right)^2\right] + \mathbb{E}\left[\left(\int_{t_n}^{t_{n+1}}\widetilde{W}_q(s)\,d[\widetilde{W}_l - W_l]\right)^2\right]\\ &\quad= \int_{t_n}^{t_{n+1}}\mathbb{E}\left[[\widetilde{W}_q(s) - W_q(s)]^2\right]\,ds + \mathbb{E}\left[\left(-\int_{t_n}^{t_{n+1}}[\widetilde{W}_l - W_l]\,d\widetilde{W}_q(s)\right)^2\right].\end{aligned} \tag{4.15}$$

Then by the mutual independence of all Gaussian random variables $\xi_{q,i}^{(n)}$, $i = 1, 2, \ldots$, $q = 1, 2, \ldots, r$, we obtain $\mathbb{E}[[\widetilde{W}_q(s) - W_q(s)]^2] = \sum_{i=N_h+1}^{\infty} M_i^2(s)$, where $M_i(s) =$

$\int_{t_n}^{s} m_i(\theta)\, d\theta$ and for $l \neq q$,

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_{t_n}^{t_{n+1}} [\widetilde{W}_l(s) - W_l(s)]\, d\widetilde{W}_q\right)^2\right] \\
&\quad = \mathbb{E}\left[\left(\sum_{i=N_h+1}^{\infty} \sum_{j=1}^{N_h} \int_{t_n}^{t_{n+1}} M_i(s) m_j(s)\, ds \xi_{l,i}^{(n)} \xi_{q,j}^{(n)}\right)^2\right] \\
&\quad = \sum_{i=N_h+1}^{\infty} \sum_{j=1}^{N_h} \left(\int_{t_n}^{t_{n+1}} M_i(s) m_j(s)\, ds\right)^2.
\end{aligned}$$

Then by (4.15), we have

$$\begin{aligned}
&\mathbb{E}\left[(\tilde{I}_{q,l,t_n,t_{n+1},0} - I_{q,l,t_n,t_{n+1},0})^2\right] \\
&\quad = \sum_{i=N_h+1}^{\infty} \int_{t_n}^{t_{n+1}} M_i^2(s)\, ds + \sum_{i=N_h+1}^{\infty} \sum_{j=1}^{N_h} \left(\int_{t_n}^{t_{n+1}} M_i(s) m_j(s)\, ds\right)^2.
\end{aligned} \tag{4.16}$$

In (4.16), we consider the Fourier basis (2.8). Then it can readily checked that

$$\sum_{i=N_h+1}^{\infty} \sum_{j=1}^{N_h} \left(\int_{t_n}^{t_{n+1}} M_i(s) m_j(s)\, ds\right)^2 = \left(\int_{t_n}^{t_{n+1}} M_{N_h+1}(s) m_{N_h}(s)\, ds\right)^2 \tag{4.17}$$

when N_h is even and $\sum_{i=N_h+1}^{\infty} \sum_{j=1}^{N_h} (\int_{t_n}^{t_{n+1}} M_i(s) m_j(s)\, ds)^2 = 0$ when N_h is odd. Moreover, for $i \geq 2$, it holds from simple calculations that

$$\int_{t_n}^{t_{n+1}} M_i^2(s)\, ds = \frac{3\Delta^2}{(2\lfloor i/2 \rfloor \pi)^2} \text{ if } i \text{ is even and } \frac{\Delta^2}{(2\lfloor i/2 \rfloor \pi)^2} \text{ otherwise.} \tag{4.18}$$

Then by (4.16), (4.17), we have

$$\begin{aligned}
&\mathbb{E}[(\tilde{I}^F_{q,l,t_n,t_{n+1},0} - I_{q,l,t_n,t_{n+1},0})^2] \\
&\quad = \sum_{i=N_h+1}^{\infty} \int_{t_n}^{t_{n+1}} M_i^2(s)\, ds + \sum_{i=N_h+1}^{\infty} \sum_{j=1}^{N_h} \left(\int_{t_n}^{t_{n+1}} M_i(s) m_j(s)\, ds\right)^2 \\
&\quad = \varsigma(N_h) \frac{\Delta^2}{(N_h \pi)^2} + \sum_{i=N_h+1}^{\infty} \frac{3^{\varsigma(i)} \Delta^2}{(2[i/2]\pi)^2} = \varsigma(N_h) \frac{2\Delta^2}{(N_h \pi)^2} + \sum_{i=M}^{\infty} \frac{\Delta^2}{(i\pi)^2}.
\end{aligned}$$

Hence, we arrive at (2.26) by the fact $\sum_{i=M}^{\infty} \frac{1}{i^2} \leq \frac{1}{M}$. Similarly, we can obtain (2.27). ∎

Proof of Theorem 2.6. Subtracting (2.21) from (2.24) and taking expectation after squaring over both sides, we have

$$\mathbb{E}[(X_{n+1}^M - X_{n+1})^2] = \mathbb{E}[(X_n^M - X_n)^2] + 2 \sum_{i=0}^{4} \mathbb{E}[(X_n^M - X_n) R_i] + \sum_{i,j=0}^{4} \mathbb{E}[R_i R_j],$$

where we denote $f_n^M = f(X_n^M, X_{n-m}^M)$ and $g_{l,n}^M = g_l(X_n^M, X_{n-m}^M)$ and

$$
\begin{aligned}
R_0 &= h(f_n^M - f_n) + \sum_{l=1}^{r}(g_{l,n}^M - g_{l,n})\Delta W_{l,n},\\
R_1 &= \sum_{l=1}^{r}\sum_{q=1}^{r}\left[\partial_x g_{l,n}^M g_{q,n}^M - \partial_x g_{l,n} g_{q,n}\right]\tilde{I}^F_{q,l,t_n,t_{n+1},0},\\
R_2 &= \sum_{l=1}^{r}\sum_{q=1}^{r}\left[\partial_{x_\tau} g_{l,n}^M g_{q,n-m}^M - \partial_{x_\tau} g_{l,n} g_{q,n-m}\right]\tilde{I}^F_{q,l,t_n,t_{n+1},\tau},\\
R_3 &= \sum_{l=1}^{r}\sum_{q=1}^{r}\partial_x g_{l,n}^M g_{q,n}^M (I_{q,l,t_n,t_{n+1},0} - \tilde{I}^F_{q,l,t_n,t_{n+1},0}),\\
R_4 &= \sum_{l=1}^{r}\sum_{q=1}^{r}\partial_{x_\tau} g_{l,n}^M g_{q,n-m}^M (I_{q,l,t_n,t_{n+1},\tau} - \tilde{I}^F_{q,l,t_n,t_{n+1},\tau}).
\end{aligned}
$$

Similarly to the proof of Theorem 2.2, we have

$$
\begin{aligned}
&\mathbb{E}[R_0^2] \le C(h^2+h)(\mathbb{E}[(X_n^M - X_n)^2] + \mathbb{E}[(X_{n-m}^M - X_{n-m})^2]),\\
&\mathbb{E}[R_1^2] \le C\max_{1\le l,q\le r}\mathbb{E}[(\left|X_n^M - X_n\right|^2 + \left|X_{n-m}^M - X_{n-m}\right|^2)]\mathbb{E}[(\tilde{I}^F_{q,l,t_n,t_{n+1},0})^2],\\
&\mathbb{E}[R_2^2] \le C\max_{1\le l,q\le r}\mathbb{E}[(\left|X_n^M - X_n\right|^2 + \left|X_{n-m}^M - X_{n-m}\right|^2)(\tilde{I}^F_{q,l,t_n,t_{n+1},\tau})^2],\\
&\mathbb{E}[R_3^2] \le C\max_{1\le l<q\le r}\mathbb{E}[(1+\left|X_n^M\right|^2 + \left|X_{n-m}^M\right|^2)]\mathbb{E}[(I_{q,l,t_n,t_{n+1},0} - \tilde{I}^F_{q,l,t_n,t_{n+1},0})^2],\\
&\mathbb{E}[R_4^2] \le C\max_{1\le l<q\le r}\mathbb{E}[(1+\left|X_n^M\right|^2 + \left|X_{n-m}^M\right|^2)(I_{q,l,t_n,t_{n+1},\tau} - \tilde{I}^F_{q,l,t_n,t_{n+1},\tau})^2].
\end{aligned}
\tag{4.19}
$$

First, we establish the following estimations:

$$
\mathbb{E}[R_i^2] \le Ch^3, \quad i = 3, 4. \tag{4.20}
$$

The case for $i = 3$ follows directly from Lemma 2.5 and boundedness of moments of X_n and X_n^M. By Lemma 2.5 and (2.22), we have

$$
\begin{aligned}
&\mathbb{E}\left[(I_{q,l,t_n,t_{n+1},\tau} - \tilde{I}^F_{q,l,t_n,t_{n+1},\tau})^4\right]\\
&\quad = \mathbb{E}\left[\left(-\frac{\sqrt{2}h}{2\pi}\xi_{q,1}^{(n-m)}\sum_{p=s+1}^{\infty}\frac{1}{p}\xi_{l,2p}^{(n)} + \frac{h}{2\pi}\sum_{p=s_1+1}^{\infty}\frac{1}{p}[\xi_{q,2p+1}^{(n-m)}\xi_{l,2p}^{(n)} - \xi_{q,2p}^{(n-m)}\xi_{l,2p+1}^{(n)}]\right)^4\right]\\
&\quad \le Ch^4\left[\left(\sum_{p=s+1}^{\infty}\frac{1}{p^2}\right)^2 + \left(\sum_{p=s_1+1}^{\infty}\frac{1}{p^2}\right)^2\right] \le C\frac{h^4}{N_h^2},
\end{aligned}
$$

where $s = [\frac{N_h}{2}]$ and $s_1 = [\frac{N_h-1}{2}]$. As N_h is of the order of h^{-1}, we have

$$
\mathbb{E}[(I_{q,l,t_n,t_{n+1},\tau} - \tilde{I}^F_{q,l,t_n,t_{n+1},\tau})^4] \le Ch^6. \tag{4.21}
$$

Then by the fact that X_n and X_n^M have bounded fourth-order moments, the Cauchy inequality, and (4.21), we reach (4.20) when $i = 4$.

Second, we estimate $\mathbb{E}[R_i^2]$, $i = 1, 2$. By (2.22), the Lipschitz condition (2.2), and N_h is of the order of h^{-1}, we have

$$\mathbb{E}[R_1^2] \le Ch(\mathbb{E}[(X_n^M - X_n)^2] + \mathbb{E}[(X_{n-m}^M - X_{n-m})^2]). \tag{4.22}$$

Now we require an estimation of $\mathbb{E}[R_2^2]$. By the Lipschitz condition (2.2), the adaptedness of X_{n-m}^M and X_{n-m}, and the Cauchy inequality (twice), we have

$$\begin{aligned}\mathbb{E}[R_2^2] \le C \max_{1\le l,q\le r} \Big\{ & \mathbb{E}\left[\left|X_n^M - X_n\right|^2 (\tilde{I}^F_{q,l,t_n,t_{n+1},\tau})^2\right] \\ & + \mathbb{E}\left[\left|X_{n-m}^M - X_{n-m}\right|^2 (\tilde{I}^F_{q,l,t_n,t_{n+1},\tau})^2\right]\Big\}, \\ \le C \max_{1\le l,q\le r} & \left(\mathbb{E}\left[\left|X_n^M - X_n\right|^4\right]\right)^{1/4} \left(\mathbb{E}\left[(\tilde{I}^F_{q,l,t_n,t_{n+1},\tau})^8\right]\right)^{1/4} \left(\mathbb{E}\left[\left|X_n^M - X_n\right|^2\right]\right)^{1/2} \\ + Ch^2 & \mathbb{E}\left[(X_{n-m}^M - X_{n-m})^2\right].\end{aligned}$$

It can be readily checked from (2.22) that $\mathbb{E}[(\tilde{I}^F_{q,l,t_n,t_{n+1},\tau})^8] \le Ch^8$. Hence, from the boundedness of moments, we have

$$\mathbb{E}[R_2^2] \le Ch^2(\mathbb{E}[(X_n^M - X_n)^2])^{1/2} + Ch^2\mathbb{E}[(X_{n-m}^M - X_{n-m})^2]. \tag{4.23}$$

Now estimate $\mathbb{E}[(X_n^M - X_n)R_i]$, $i = 0, 1, 2, 3, 4$. By the adaptedness of X_n and the Lipschitz condition of f, we have

$$\mathbb{E}[(X_n^M - X_n)R_0] \le Ch\mathbb{E}[(\left|X_n^M - X_n\right|^2 + \left|X_{n-m}^M - X_{n-m}\right|^2)]. \tag{4.24}$$

By the adaptedness of X_n and $\mathbb{E}[\tilde{I}_{q,l,t_n,t_{n+1},0}] = \delta_{q,l}h/2$ ($\delta_{q,l}$ is the Kronecker delta) and the Lipschitz condition of $\partial_x g_l g_q$, we have

$$\mathbb{E}[(X_n^M - X_n)R_1] \le Ch\mathbb{E}[(\left|X_n^M - X_n\right|^2 + \left|X_{n-m}^M - X_{n-m}\right|^2)]. \tag{4.25}$$

By the adaptedness of X_n and $\mathbb{E}[\tilde{I}^F_{q,l,t_n,t_{n+1},0} - I_{q,l,t_n,t_{n+1},0}] = 0$, we have

$$\mathbb{E}[(X_n^M - X_n)R_3] = 0. \tag{4.26}$$

Similarly to the proof of (4.12), we have

$$\mathbb{E}[(X_n^M - X_n)R_4] = 0. \tag{4.27}$$

Then by (4.19), (4.20)–(4.23), (4.24)–(4.27), and the Cauchy inequality, we have

$$\begin{aligned}\mathbb{E}[(X_{n+1}^M - X_{n+1})^2] \le (1 + Ch)\mathbb{E}[(X_n^M - X_n)^2] + Ch\mathbb{E}[(X_{n-m}^M - X_{n-m})^2] \\ + Ch^2(\mathbb{E}[(X_n^M - X_n)^2])^{1/2} + Ch^3,\end{aligned} \tag{4.28}$$

where $n \ge m$. Similarly, we have that (4.28) holds also for $1 \le n \le m - 1$. From here and by the nonlinear Gronwall inequality, we reach the conclusion (2.28). ∎

5. Conclusion. Using the WZ approximation as an intermediate step, we have presented three numerical schemes for SDDEs: a predictor-corrector scheme, a midpoint scheme, and a Milstein-like scheme. The first two schemes are of half order convergence in the mean-square sense while both schemes are of first order in the mean-square sense if the underlying SDDEs with single noise (commutative noise) and the time delay are only in the drift coefficients. In the Milstein-like scheme, a relatively simple algorithm for approximating the stochastic double integrals with and without time delay has been given by a truncated spectral expansion of Brownian motion. With a great enough number of modes in the spectral expansion of Brownian motion, the Milstein-like scheme is shown theoretically and numerically to be of first-order mean-square convergence. Though the Milstein-like scheme is more expensive than the other two schemes in general, the Milstein-like scheme (with the Fourier approximation of Brownian motion) is superior to the predictor-corrector scheme and the midpoint scheme in terms of accuracy and computational cost for SDDEs with a single noise.

Acknowledgments. The authors thank the referees for their valuable comments. The authors would like to thank Professor Michael V. Tretyakov at University of Nottingham for helpful discussion on Wong–Zakai approximation and simulating the Milstein scheme when he was visiting ICERM (Brown University, Providence, Rhode Island).

REFERENCES

[1] C. T. H. Baker and E. Buckwar, *Numerical analysis of explicit one-step methods for stochastic delay differential equations*, LMS J. Comput. Math., 3 (2000), pp. 315–335.

[2] N. Bruti-Liberati and E. Platen, *Strong predictor-corrector Euler methods for stochastic differential equations*, Stochastics Dynam., 8 (2008), pp. 561–581.

[3] E. Buckwar and T. Sickenberger, *A comparative linear mean-square stability analysis of Maruyama- and Milstein-type methods*, Math. Comput. Simulation, 81 (2011), pp. 1110–1127.

[4] E. Buckwar and R. Winkler, *Multistep methods for SDEs and their application to problems with small noise*, SIAM J. Numer. Anal., 44 (2006), pp. 779–803.

[5] E. Buckwar and R. Winkler, *Multi-step Maruyama methods for stochastic delay differential equations*, Stochastic Anal. Appl., 25 (2007), pp. 933–959.

[6] W. Cao and Z. Zhang, *Simulations of two-step Maruyama methods for nonlinear stochastic delay differential equations*, Adv. Appl. Math. Mech., 4 (2012), pp. 821–832.

[7] W. Cao and Z. Zhang, *On exponential mean-square stability of two-step Maruyama methods for stochastic delay differential equations*, J. Comput. Appl. Math., 245 (2013), pp. 182–193.

[8] M. Giles, *Improved multilevel Monte Carlo convergence using the Milstein scheme*, in Monte Carlo and Quasi-Monte Carlo Methods 2006, A. Keller, S. Heinrich, and H. Niederreiter, eds., Springer, Berlin, 2008, pp. 343–358.

[9] H. Gilsing and T. Shardlow, *SDELab: A package for solving stochastic differential equations in MATLAB*, J. Comput. Appl. Math., 205 (2007), pp. 1002–1018.

[10] M. Grigoriu, *Control of time delay linear systems with Gaussian white noise*, Probab. Engrg. Mech., 12 (1997), pp. 89–96.

[11] Q. Guo, W. Xie, and T. Mitsui, *Convergence and stability of the split-step θ-Milstein method for stochastic delay Hopfield neural networks*, Abstr. Appl. Anal., 2013 (2013), 169214.

[12] I. Gyöngy and A. Shmatkov, *Rate of convergence of Wong-Zakai approximations for stochastic partial differential equations*, Appl. Math. Optim., 54 (2006), pp. 315–341.

[13] D. G. Hobson and L. C. G. Rogers, *Complete models with stochastic volatility*, Math. Finance, 8 (1998), pp. 27–48.

[14] N. Hofmann and T. Müller-Gronbach, *A modified Milstein scheme for approximation of stochastic delay differential equations with constant time lag*, J. Comput. Appl. Math., 197 (2006), pp. 89–121.

[15] Y. Hu, S.-E. A. Mohammed, and F. Yan, *Discrete-time approximations of stochastic delay equations: The Milstein scheme*, Ann. Probab., 32 (2004), pp. 265–314.

[16] C. Huang, S. Gan, and D. Wang, *Delay-dependent stability analysis of numerical methods for stochastic delay differential equations*, J. Comput. Appl. Math., 236 (2012), pp. 3514–3527.

[17] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland, Amsterdam, 1981.

[18] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer-Verlag, New York, 1991.

[19] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin, 1992.

[20] P. E. Kloeden and T. Shardlow, *The Milstein scheme for stochastic delay differential equations without using anticipative calculus*, Stochastic Anal. Appl., 30 (2012), pp. 181–202.

[21] U. Küchler and E. Platen, *Strong discrete time approximation of stochastic differential equations with time delay*, Math. Comput. Simulation, 54 (2000), pp. 189–205.

[22] D. F. Kuznetsov, *Strong Approximation of Multiple Ito and Stratonovich Stochastic Integrals: Multiple Fourier Series Approach*, St. Petersburg State Polytechnic University, St. Petersburg, 2011, (in Russian).

[23] M. Liu, W. Cao, and Z. Fan, *Convergence and stability of the semi-implicit Euler method for a linear stochastic differential delay equation*, J. Comput. Appl. Math., 170 (2004), pp. 255–268.

[24] V. Mackevičius, *On approximation of stochastic differential equations with coefficients depending on the past*, Liet. Mat. Rink., 32 (1992), pp. 285–298.

[25] X. Mao, *Stochastic differential equations and their applications*, Horwood Ser. Math. Appl., Horwood, Chichester, 1997.

[26] X. Mao and L. Shaikhet, *Delay-dependent stability criteria for stochastic differential delay equations with Markovian switching*, Stab. Control Theory Appl., 3 (2000), pp. 88–102.

[27] G. N. Milstein, Yu. M. Repin, and M. V. Tretyakov, *Numerical methods for stochastic systems preserving symplectic structure*, SIAM J. Numer. Anal., 40 (2002), pp. 1583–1604.

[28] G. N. Milstein and M. V. Tret'yakov, *Mean-square numerical methods for stochastic differential equations with small noises*, SIAM J. Sci. Comput., 18 (1997), pp. 1067–1087.

[29] G. N. Milstein and M. V. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Springer-Verlag, Berlin, 2004.

[30] S. E. A. Mohammed, *Stochastic Functional Differential Equations*, Pitman, Boston, 1984.

[31] S. Ogawa, *Quelques propriétés de l'intégrale stochastique du type noncausal*, Japan J. Appl. Math., 1 (1984), pp. 405–416.

[32] S. Ogawa, *On a deterministic approach to the numerical solution of the SDE*, Math. Comput. Simulation, 55 (2001), pp. 209–214.

[33] R. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, AMS, Providence, RI, 1934.

[34] M. D. Paola and A. Pirrotta, *Time delay induced effects on control of linear systems under random excitation*, Probab. Engrg. Mech., 16 (2001), pp. 43–51.

[35] N. Rosli, A. Bahar, S. H. Yeak, and X. Mao, *A systematic derivation of stochastic Taylor methods for stochastic delay differential equations*, Bull. Malays. Math. Soc. (2), 36 (2013), pp. 555–576.

[36] H. J. Sussmann, *On the gap between deterministic and stochastic ordinary differential equations*, Ann. Probab., 6 (1978), pp. 19–41.

[37] L. S. Tsimring and A. Pikovsky, *Noise-induced dynamics in bistable systems with delay*, Phys. Rev. Lett., 87 (2001), 250602.

[38] C. Tudor and M. Tudor, *On approximation of solutions for stochastic delay equations*, Stud. Cerc. Mat., 39 (1987), pp. 265–274.

[39] M. Tudor, *Approximation schemes for stochastic equations with hereditary argument*, Stud. Cerc. Mat., 44 (1992), pp. 73–85.

[40] K. Twardowska, *On the approximation theorem of the Wong-Zakai type for the functional stochastic differential equations*, Probab. Math. Statist., 12 (1991), pp. 319–334.

[41] K. Twardowska, *Wong-Zakai approximations for stochastic differential equations*, Acta Appl. Math., 43 (1996), pp. 317–359.

[42] K. Twardowska, T. Marnik, and M. Pasławaska-Południak, *Approximation of the Zakai equation in a nonlinear filtering problem with delay*, Int. J. Appl. Math. Comput. Sci., 13 (2003), pp. 151–160.

[43] K. Vasilakos and A. Beuter, *Effects of noise on a delayed visual feedback system*, J. Theor. Biol., 165 (1993), pp. 389–407.

[44] X. Wang and S. Gan, *B-convergence of split-step one-leg theta methods for stochastic differential equations*, J. Appl. Math. Comput., 38 (2012), pp. 489–503.

[45] M. Wiktorsson, *Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions*, Ann. Appl. Probab., 11 (2001), pp. 470–487.

[46] E. Wong and M. Zakai, *On the convergence of ordinary integrals to stochastic integrals*, Ann. Math. Statist., 36 (1965), pp. 1560–1564.

[47] E. Wong and M. Zakai, *On the relation between ordinary and stochastic differential equations*, Internat. J. Engrg. Sci., 3 (1965), pp. 213–229.

[48] F. Wu, X. Mao, and L. Szpruch, *Almost sure exponential stability of numerical solutions for stochastic delay differential equations*, Numer. Math., 115 (2010), pp. 681–697.

[49] H. Zhang, S. Gan, and L. Hu, *The split-step backward Euler method for linear stochastic delay differential equations*, J. Comput. Appl. Math., 225 (2009), pp. 558–568.

[50] B. Zhibaĭtis and V. Matskyavichyus, *Gaussian approximations of Brownian motion in a stochastic integral*, Liet. Mat. Rink., 33 (1993), pp. 508–526.