THE MILSTEIN SCHEME FOR STOCHASTIC DELAY DIFFERENTIAL EQUATIONS WITHOUT ANTICIPATIVE CALCULUS

Kloeden, P.E. and Shardlow, T.

2010

MIMS EPrint: **2010.77**

ISSN 1749-9097

THE MILSTEIN SCHEME FOR STOCHASTIC DELAY DIFFERENTIAL EQUATIONS WITHOUT ANTICIPATIVE CALCULUS*

P.E. KLOEDEN† AND T. SHARDLOW‡

Abstract. The Milstein scheme is the simplest nontrivial numerical scheme for stochastic differential equations with a strong order of convergence one. The scheme has been extended to the stochastic delay differential equations but the analysis of the convergence is technically complicated due to anticipative integrals in the remainder terms. This paper employs an elementary method to derive the Milstein scheme and its first order strong rate of convergence for stochastic delay differential equations.

AMS subject classifications. 35K90, 41A58, 65C30, 65M99, 60K35

Key words. Taylor expansions, stochastic differential equations, delay equations, strong convergence, SDDE, Milstein method

1. Introduction. The Milstein scheme is the simplest nontrivial numerical scheme for stochastic ordinary differential equations that achieves a strong order of convergence higher than that of the Euler-Maruyama scheme. It was first derived by Milstein, who used the Itô formula to expand an integrand involving the solution in one of the error terms of the Euler-Maruyama scheme. The iterative repetition of this idea underlies the systematic derivation of stochastic Taylor expansions and numerical schemes of arbitrarily high strong and weak orders, as expounded in Kloeden & Platen [9], see also Milstein [12].

An analogue of the Milstein scheme has been derived in a similar way for stochastic delay differential equations (SDDEs), see Hu *et al.* [4]. However, the proofs of convergence are technically complicated due to the presence of anticipative integrals in the remainder terms.

Here we use an elementary method to derive the Milstein scheme for stochastic delay differential equations that does not involve anticipative integrals and anticipative stochastic calculus. Following the approach used by Jentzen & Kloeden for random ordinary differential equations [6, 8] and stochastic partial differential equations [7], we use deterministic Taylor expansions of the coefficient functions, with lower order expansions being inserted into the right hand side of higher order ones to give a closed form for the expansion. (A similar idea, without the final insertion, was also considered in [5, 11]). The Itô formula is not used at all and our proofs are much simpler than in [4].

The paper is organised as follows. §2 gives the precise stochastic delay differential equation that we study and some background. §3 derives the Milstein method (see (3.7)) by use of Taylor expansions and calculates the local truncation error by collecting the Taylor remainder terms. §4 gives the proof of convergence and the main results in Theorems 4.2 and 4.5 . We must restrict the class of SDDEs to have

*Partially supported by the ARC (UK) and DAAD (Germany). The coauthors thanks the Isaac Newton Institute for Mathematical Sciences of the University of Cambridge, where they wrote part of this paper.

†Institut für Mathematik, Goethe-Universität Frankfurt, D-60325 Frankfurt a.M., Germany, kloeden@math.uni-frankfurt.de

‡School of Mathematics, University of Manchester M13 9PL, Manchester, UK. shardlow@maths.man.ac.uk. Partially supported by the ARC (UK) and DAAD (Germany)

finitely many discrete delays and this allows proof of a mean square first order rate of convergence subject to continuity of the drift, diffusion, and initial data.

2. Background. Consider $\mathbf{R}^d$ with the norm $\|x\|_{\mathbf{R}^d} = \langle x, x\rangle^{1/2}$ for $x \in \mathbf{R}^d$. Denote the bounded linear operators between Banach spaces X and Y by $\mathcal{L}(X,Y)$ and $\|L\|_{\text{op}}$ the induced operator norm for $L \in \mathcal{L}(X,Y)$. Let $C^n(X,Y)$ denote the space of functions from X to Y with n uniformly bounded Frechet derivatives $D^j f \in \mathcal{L}(X, \dots \mathcal{L}(X,Y))$ for $j = 0,\dots,n$ and $f \in C^n(X,Y)$.

Denote by C the Banach space $C = C([-\tau,0],\mathbf{R}^d)$ for $\tau > 0$ with norm $\|\eta\|_\infty = \sup_{-\tau\le\theta\le 0}\|\eta(\theta)\|_{\mathbf{R}^d}$ for $\eta \in C$. Let $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbf{P})$ denote a standard filtered probability space. Consider the Itô SDDE on $\mathbf{R}^d$ on the time interval $[0,T]$ with delay $\tau > 0$ given by

$$dx(t) = f(x_t,t)\,dt + g(x_t,t)\,d\beta(t), \tag{2.1}$$

subject to initial data $x_0 = \xi \in C$, where $\beta(t)$ is a $\mathbf{R}^m$ Brownian motion adapted to $\mathcal{F}_t$,

$$f\colon C \times \mathbf{R}^+ \to \mathbf{R}^d, \qquad g\colon C \times \mathbf{R}^+ \to \mathbf{R}^{d\times m},$$

and we use x_t to denote the segment $\{x(t+\theta)\colon \theta \in [-\tau,0]\}$. We assume there exists a unique solution $x(t)$ to (2.1) that is adapted to $\mathcal{F}_t$ and such that $x_t \in C$ for $t \ge 0$ (see [10, 13]). Denote by $x(t;s,\eta)$ the solution of (2.1) with initial condition $x_s = \eta \in C$ at time $t = s$ and the corresponding segment by $x_t(s,\eta) \in C$. Then $x(t) = x(t;0,\xi)$ and $x_t = x_t(0,\xi)$.

Throughout K is a generic constant that varies from one place to another and depends on f, g, the initial data ξ, the interval of integration $[0,T]$, but is independent of the discretisation parameter. The notation $\mathcal{O}(n)$ is used to denote a quantity bounded by Kn.

We make use of the following inequalities: for any $a_1, a_2, \dots, a_N$ and $p > 1$,

$$\Big(\sum_{i=1}^N a_i\Big)^p \le N^{p-1}\sum_{i=1}^N a_i^p. \tag{2.2}$$

Burkholder-Gundy-Davis inequality: for any $\mathbf{R}^{d\times m}$ valued adapted process $z(s)$ and for $p \ge 2$, there exists $C_p, \hat{C}_p$ such that

$$\begin{aligned}\mathbf{E}\Big[\sup_{0\le t\le T}\Big\|\int_0^t z(s)\,d\beta(s)\Big\|_{\mathbf{R}^d}^p\Big] \le& C_p\Big(\int_0^T \mathbf{E}\|z(s)\|_F^2\,ds\Big)^{p/2}\\ \le& \hat{C}_p\int_0^T \mathbf{E}\|z(s)\|_F^p\,ds\end{aligned} \tag{2.3}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

We assume the following regularity throughout.

ASSUMPTION 2.1. *$f \in C^3(C\times\mathbf{R}^+,\mathbf{R}^d)$, $g \in C^3(C\times\mathbf{R}^+,\mathbf{R}^{d\times m})$, and ξ is uniformly Lipschitz continuous from $[-\tau,0]$ to $\mathbf{R}^d$.*

We make use of the following results.

THEOREM 2.1. *Suppose that $0 \le s,t \le T$ and η is a $\mathcal{F}_s$ measurable C valued* *random variable such that* $\mathbf{E}\|\eta(\theta_1)-\eta(\theta_2)\|^p_{\mathbf{R}^d} \le K|\theta_1-\theta_2|^{p/2}$. *Then*

$$\mathbf{E}\Big[\sup_{s-\tau\le t\le T}\|x(t;s,\eta)\|^p_{\mathbf{R}^d}\Big] \le K \tag{2.4}$$

$$\mathbf{E}\Big[\sup_{s\le r\le t}\|x_r(s,\eta)-\eta\|^p_\infty\Big|\mathcal{F}_s\Big] \le K|t-s|^{p/2} \tag{2.5}$$

$$\mathbf{E}\Big[\Big\|x_t(s,\eta)-\mathbf{E}\Big[x_t(s,\eta)|\mathcal{F}_s\Big]\Big\|^p_\infty\Big] \le K|t-s|^{p/2}. \tag{2.6}$$

Proof. The first inequality is a consequence of the Gronwall inequality and a detailed proof is found in [10, 15]. The second follows easily from the integral form for the solution: for $r+\theta\ge s$,

$$\begin{aligned} x(r+\theta;s,\eta) =& x(s+\theta) \\ &+\int_{s+\theta}^{r+\theta} f(x_q(s,\eta),q)\,dq+\int_{s+\theta}^{r+\theta} g(x_q(s,\eta),q)\,d\beta(q). \end{aligned}$$

Assume for notational simplicity that η is non-random,

$$\begin{aligned} &\mathbf{E}\Big[\sup_{s\le r\le t}\|x(r+\theta;s,\eta)-\eta(\theta)\|^p_{\mathbf{R}^d}\Big] \\ \le& K|r-s|^p+K\Big(\int_{s+\theta}^{t+\theta}\mathbf{E}\|g(x(q;s,\eta))\|^2_F\,dq\Big)^{p/2} \\ \le& K|t-s|^p+K|t-s|^{p/2}\le K|t-s|^{p/2}. \end{aligned}$$

For $r+\theta\le s$, $\|x(r+\theta;s,\eta)-\eta(\theta)\|_{\mathbf{R}^d}=\|\eta(r-s+\theta)-\eta(\theta)\|_{\mathbf{R}^d}\le K|r-s|$. Finally,

$$\begin{aligned} &\mathbf{E}\Big[\sup_{s\le r\le t}\|x_r(s,\eta)-\eta\|^p_\infty\Big|\mathcal{F}_s\Big] \\ \le& \mathbf{E}\Big[\sup_{s\le r\le t}\sup_{\theta\in[-\tau,0]}\|x(r+\theta;s,\eta)-\eta(r+\theta)\|^p_{\mathbf{R}^d}\Big|\mathcal{F}_s\Big]\le K|t-s|^{p/2}. \end{aligned}$$

The last is found from

$$\Big\|x_t(s,\eta)-\mathbf{E}\Big[x_t(s,\eta)|\mathcal{F}_s\Big]\Big\|_\infty \le \|x_t(s,\eta)-\eta\|_\infty+\Big\|\mathbf{E}\Big[x_t(s,\eta)-\eta\Big|\mathcal{F}_s\Big]\Big\|_\infty.$$

□

THEOREM 2.2 (derivative in initial condition). *Suppose that* $0\le s,t\le T$ *and* η *is a* $\mathcal{F}_s$ *measurable* C *valued random variable such that* $\mathbf{E}\|\eta(\theta_1)-\eta(\theta_2)\|^p_{\mathbf{R}^d}\le K|\theta_1-\theta_2|^{p/2}$. *For* $h\in C$, *let* y^h_t *denote the solution to*

$$dy^h(t)=Df(x_t(s,\eta),t)y^h_t\,dt+Dg(x_t(s,\eta),t)y^h_t\,d\beta(t),\qquad y^h_s=h \tag{2.7}$$

Then $y^h(t)$ *is the* L^2 *Frechet derivative of* $x(t;s,\eta)$ *with respect to* η:

$$\sup_{\|h\|_\infty<1}\mathbf{E}\Big[\sup_{0\le t\le T}\Big|\frac{x(t;s,\eta+\epsilon h)-x(t;s,\eta)}{\epsilon}-y^h(t)\Big|^2\Big|\mathcal{F}_s\Big]\to 0\qquad \epsilon\to 0. \tag{2.8}$$

We write $\frac{\partial x(t;s,\eta)}{\partial\eta}h$ *for* $y^h(t)$ *and note that* $\frac{\partial x(t;s,\eta)}{\partial\eta}\in\mathcal{L}(C,L^2(\Omega,\mathbf{R}^d))$. *Further, for* $0\le s\le t\le T$ *and* $\eta\in C$,

$$\mathbf{E}\Big[\Big\|\frac{\partial x(t;s,\eta)}{\partial\eta}\Big\|^2_{\text{op}}\Big|\mathcal{F}_s\Big] \le K \tag{2.9}$$

$$\mathbf{E}\Big[\Big\|\frac{\partial x(t;s,\eta)}{\partial\eta}-\mathbf{E}\Big[\frac{\partial x(t;s,\eta)}{\partial\eta}\Big|\mathcal{F}_s\Big]\Big\|^p_{\text{op}}\Big|\mathcal{F}_s\Big] \le K|t-s|^{p/2}. \tag{2.10}$$

Proof. Equations (2.7)–(2.8) are given in [15]. Equations (2.9)–(2.10) follow by applying standard techniques to (2.7) as in Theorem 2.1. For example, the integral form gives

$$y^h(t) = h(0) + \int_0^t Df(x_r(s,\eta), r) y_r^h \, dr + \int_0^t Dg(x_r(s,\eta), t) y_r^h \, d\beta(r).$$

which implies

$$\mathbf{E}\Big[\|y^h(t)\|_{\mathbf{R}^d}^2\Big] \le 3\|h\|_\infty^2 + K\int_0^t \mathbf{E}\Big[\|y_r^h\|_\infty^2\Big]\, dr + K\int_0^t \mathbf{E}\Big[\|y_r^h\|_\infty^2\Big]\, dr.$$

so that Gronwall's inequality provides $\mathbf{E}\|y^h(t)\|_{\mathbf{R}^d}^2 \le K\|h\|_\infty^2$ for $0 \le t \le T$. Hence y^h is a bounded linear operator from C to $L^2(\Omega, \mathbf{R}^d)$. ∎

THEOREM 2.3 (second derivative in initial condition). *Under the assumptions of Theorem 2.2, there exists a second L^2 Frechet derivative for $x(t;s,\eta)$, which we denote by $\frac{\partial^2}{\partial\eta^2}x(t;s,\eta)$. Further, $\frac{\partial^2}{\partial\eta^2}x(t;x,\eta)$ belongs to $\mathcal{L}(C, \mathcal{L}(C, L^2(\Omega, \mathbf{R}^d)))$ and*

$$\mathbf{E}\Big[\Big\|\frac{\partial^2 x(t;s,\eta)}{\partial\eta^2}\Big\|_{\text{op}}^2\Big] \le K. \tag{2.11}$$

Proof. Similar to Theorem 2.2. ∎

COROLLARY 2.4. *If $h(\eta) = \mathbf{E}\Psi(\eta, x(t;s,\eta))$ for $t \ge s$, $\eta \in C$, and some $\Psi \in C^2(C \times \mathbf{R}^d, \mathbf{R})$, then $h \in C^2(C, \mathbf{R})$.*

If

$$h(\eta_1, \dots, \eta_4) = \mathbf{E}\Psi(\eta_1, \dots, \eta_4, x(t_1; s, \eta_1), \dots, x(t_J; s, \eta_4))$$

where $t_1, \dots, t_J \ge s$, $\eta_1, \dots, \eta_4 \in C$, and $\Psi \in C^2(C \times C \times C \times C \times \mathbf{R}^{4dJ}, \mathbf{R}^{m\times m})$, then $h \in C^2(C \times C \times C \times C, \mathbf{R}^{m\times m})$.

Proof. This follows from Theorem 2.2–2.3. For example, with $x(t) = x(t;s,\eta)$ and $\xi_1, \xi_2 \in C$,

$$Dh(\eta)(\xi_1) = \mathbf{E}\Big[D\Psi(\eta, x(t))(\xi_1, \frac{\partial x(t)}{\partial\eta}\xi_1)\Big]$$

$$\begin{aligned} D^2h(\eta)(\xi_1,\xi_2) =& \mathbf{E}\Big[D^2\Psi(\eta, x(t))\left((\xi_1, \frac{\partial x(t)}{\partial\eta}\xi_1), (\xi_2, \frac{\partial x(t)}{\partial\eta}\xi_2)\right)\Big] \\ &+ \mathbf{E}\Big[D\Psi(\eta, x(t))\left(0, \frac{\partial^2 x(t)}{\partial\eta^2}(\xi_1,\xi_2)\right)\Big] \end{aligned}$$

and

$$\begin{aligned} \|D^2h(\eta)\|_{\text{op}} \le& \mathbf{E}\Big[\|D^2\Psi(x(t))\|_{\text{op}}\left(1 + \Big\|\frac{\partial x(t)}{\partial\eta}\Big\|_{\text{op}}\right)^2\Big] \\ &+ \mathbf{E}\Big[\|D\Psi(x(t))\|_{\text{op}}\Big\|\frac{\partial^2 x(t)}{\partial\eta^2}\Big\|_{\text{op}}\Big]. \end{aligned}$$

As $\Psi \in C^2$,

$$\|D^2h(\eta)\|_{\text{op}} \le K\Big[1 + \mathbf{E}\Big[\Big\|\frac{\partial x(t)}{\partial\eta}\Big\|_{\text{op}}^2\Big]\Big] + K\Big[\mathbf{E}\Big[\Big\|\frac{\partial^2 x(t)}{\partial\eta^2}\Big\|_{\text{op}}^2\Big]\Big]^{1/2}.$$

which is uniformly bounded using (2.9) and (2.11)

The second statement is an elementary extension of the first. ∎

3. Derivation of Milstein method and remainders. Taylor's theorem on Banach spaces [1] gives for $0 \le s \le r$

$$
\begin{aligned}
f(x_r, r) =& f(x_s, s) + Df(x_s, s)(x_r - x_s, r - s) \\
&+ \int_0^1 (1-h) D^2 f\Big(x_s + h(x_r - x_s), s + h(r-s)\Big)\Big(x_r - x_s, r - s\Big)^2 \, dh \\
=& f(x_s, s) + R_f(r; s, x_s) && (3.1) \\
g(x_r, r) =& g(x_s, s) + Dg(x_s, s)(x_r - x_s, r - s) \\
&+ \int_0^1 (1-h) D^2 g\Big(x_s + h(x_r - x_s), s + h(r-s)\Big)\Big(x_r - x_s, r - s\Big)^2 \, dh \\
=& g(x_s, s) + Dg(x_s, s)(x_r - x_s, r - s) + R_g(r; s, x_s). && (3.2)
\end{aligned}
$$

Lemma 3.1.

$$
\begin{aligned}
\mathbf{E}\Big[\sup_{0 \le s \le r \le t} \|R_f(r; s, x_s)\|^p_{\mathbf{R}^d}\Big] \le& K|t-s|^{p/2} \\
\mathbf{E}\Big[\sup_{0 \le s \le r \le t} \|R_g(r; s, x_s)\|^p_{F}\Big] \le& K|t-s|^{p}.
\end{aligned}
$$

Proof. Using (2.4) and the regularity of f, g, both inequalities follow from the definitions in (3.1)–(3.2). ∎

Substitute the expansions for f and g into the integral form of (2.1) on $[s, t]$,

$$
x(t) = x(s) + \int_s^t f(x_r, r) \, dr + \int_s^t g(x_r, r) \, d\beta(r),
$$

to find

$$
\begin{aligned}
x(t) =& x(s) + f(x_s, s)(t-s) + g(x_s, s) \int_s^t d\beta(r) \\
&+ \int_s^t Dg(x_s, s)(x_r - x_s, r - s) \, d\beta(r) + R_1(t; s, x_s),
\end{aligned}
$$

where the remainder $R_1(t; s, x_s)$ is $\mathcal{F}_t$ measurable and defined by

$$
R_1(t; s, x_s) = \int_s^t R_f(r; s, x_s) \, dr + \int_s^t R_g(r; s, x_s) \, d\beta(r). \qquad (3.3)
$$

Lemma 3.2. *For $0 \le s \le t \le T$,*

$$
\mathbf{E}\Big[\sup_{s \le r \le t} \|R_1(r; s, x_s)\|^2_{\mathbf{R}^d}\Big] \le K|t-s|^3.
$$

Proof. Applying (2.2) and (2.3),

$$
\begin{aligned}
&\mathbf{E}\Big[\sup_{s \le r \le t} \|R_1(r; s, x_s)\|^2_{\mathbf{R}^d}\Big] \\
&\qquad \le 2\Big(\mathbf{E}\Big[\sup_{s \le r \le t} \|R_f(r; s, x_s)\|^2_{\mathbf{R}^d} |t-s|^2\Big] + \hat{C}_2 \int_s^t \mathbf{E}\Big[\|R_g(r; s, x_r)\|^2_F\Big] \, dr\Big).
\end{aligned}
$$

To finish, apply Lemma 3.1. ∎

Let $I(s,t)=\int_s^t d\beta(r)=\beta(t)-\beta(s)$. Then, for $s\geq 0$

$$\begin{aligned}x(t;s,x_s)=&x(s)+f(x_s,s)(t-s)+g(x_s,s)I(s,t)\\&+\int_s^t Dg(x_s,s)(x_r(s,x_s)-x_{s,r-s})\,d\beta(r)+R_1(t;s,x_s)\end{aligned}$$

and

$$x(t;s,x_s)=x(s)+g(x_s,s)I(s,t)+R_2(t;s,x_s), \tag{3.4}$$

where $R_2(t;s,x_s)$ is $\mathcal{F}_t$ measurable and given by

$$\begin{aligned}R_2(t;s,x_s)=&f(x_s,s)(t-s)\\&+\int_s^t Dg(x_s,s)\Big(x_r(s,x_s)-x_{s,r-s}\Big)\,d\beta(r)+R_1(t;s,x_s).\end{aligned}$$

Lemma 3.3. *For* $\Delta t>0$,

$$\mathbf{E}\Big[\sup_{s\leq r\leq s+\Delta t\leq T}\|R_2(r;s,x_s)\|^2_{\mathbf{R}^d}\Big]\leq K\Delta t^2.$$

Proof. By (2.2) and (2.3),

$$\begin{aligned}&\mathbf{E}\Big[\sup_{s\leq r\leq s+\Delta t\leq T}\|R_2(r;s,x_s)\|^2_{\mathbf{R}^d}\Big]\\
\leq&3\Big(K|t-s|^2+\hat{C}_2\sup_{s\leq r\leq s+\Delta t\leq T}\int_s^r\mathbf{E}\Big[\|Dg(x_s,s)(x_p(s,x_s)-x_{s,p-s})\|_F^2\Big]\,dp\\
&+\mathbf{E}\sup_{s\leq r\leq s+\Delta t\leq T}\|R_1(r;s,x_s)\|^2_{\mathbf{R}^d}\Big)\\
\leq&K\Big(\Delta t^2+\Delta t\Big(\mathbf{E}\Big[\sup_{s\leq r\leq s+\Delta t\leq T}\|x_r(s,x_s)-x_s\|^2_\infty\Big]+\Delta t^2\Big)\\
&+\mathbf{E}\sup_{s\leq r\leq s+\Delta t\leq T}\|R_1(r;s,x_s)\|^2_{\mathbf{R}^d}\Big).\end{aligned}$$

By Lemma 3.1 and (2.5), this is $\mathcal{O}(\Delta t^2)$ as required. ∎

For $\theta\in[-\tau,0]$ and $s+\theta\leq 0\leq t+\theta$,

$$x(t+\theta;s,x_s)=x(s+\theta)+\Big(x(t+\theta;0,x_0)-x(0)\Big)+\Big(x(0)-x(s+\theta)\Big)$$

so by (3.4)

$$\begin{aligned}x(t+\theta;s,x_s)=&x(s+\theta)+\Big(g(x_0,0)I(0,t+\theta)+R_2(t+\theta;0,x_0)\Big)\\&+\Big(\xi(0)-\xi(s+\theta)\Big).\end{aligned}\tag{3.5}$$

For $t+\theta<0$,

$$x(t+\theta;s,x_s)=\xi(t+\theta)=\xi(s+\theta)+(\xi(t+\theta)-\xi(s+\theta)).\tag{3.6}$$

We combine (3.4)–(3.6), to get an expression for the segment x_t as a perturbation of x_s for any $0 \le s \le t$.

$$x_t(s, x_s)(\theta) = x_s(\theta) + \begin{cases} \text{(see (3.4))}, & 0 \le s+\theta, \\ \text{(see (3.5))}, & s+\theta \le 0 \le t+\theta, \\ \text{(see (3.6))}, & t+\theta < 0, \end{cases}$$

As $x_s(\theta) = x(s+\theta)$, the correction term for (3.4) depends on $\hat{x}_s \in C([-2\tau, 0], \mathbf{R}^d)$, which we define for $\theta \in [-2\tau, 0]$ by

$$\hat{x}_s(\theta) = \begin{cases} x(s+\theta), & 0 < s+\theta, \\ \xi(s+\theta), & -\tau < s+\theta \le 0, \\ \xi(-\tau), & s \le -\tau. \end{cases}$$

The choice of constant for $s \le -\tau$ ensures continuity. We further require the following notations:

1. $G\colon C([-2\tau, 0], \mathbf{R}^d) \times \mathbf{R}^+ \to \mathcal{L}(C, C)$, defined by

$$G(\zeta, s)\eta(\theta) = \begin{cases} g(\pi_\theta \zeta, s+\theta)\eta(\theta), & s+\theta > 0, \\ g(\xi, 0)\eta(\theta), & s+\theta \le 0, \end{cases}$$

   where $\zeta \in C([-2\tau, 0], \mathbf{R}^d)$, $s \in \mathbf{R}^+$, $\eta \in C$, $\theta \in [-\tau, 0]$, and $\pi_\theta : C([-2\tau, 0], \mathbf{R}^d) \to C$ is defined by $\pi_\theta \zeta(\phi) = \zeta(\phi + \theta)$ for $\phi \in [-\tau, 0]$.
2. define $I_t(s) \in C$ by

$$I_t(s)(\theta) = \begin{cases} I(s+\theta, t+\theta), & 0 \le s+\theta \le t+\theta, \\ I(0, t+\theta), & -\tau \le s+\theta \le 0 \le t+\theta, \\ 0, & \text{otherwise}. \end{cases}$$

3. For $s \le t$,

$$\delta_t(s)(\theta) = \begin{cases} 0, & 0 \le s+\theta, \\ \xi(0) - \xi(s+\theta), & s+\theta \le 0 \le t+\theta, \\ \xi(t+\theta) - \xi(s+\theta), & t+\theta \le 0. \end{cases}$$

Together we have

$$x_t(s, x_s) = x_s + G(\hat{x}_s, s) I_t(s) + \delta_t(s) + R_{2,t}(s, \hat{x}_s),$$

with $R_{2,t}(s, \hat{x}_s) \in C$ defined by

$$R_{2,t}(s, \hat{x}_s)(\theta) = \begin{cases} R_2(t+\theta; s+\theta, \pi_\theta \hat{x}_s), & 0 \le s+\theta, \\ R_2(t+\theta; 0, \xi), & s+\theta < 0 \le t+\theta, \\ 0, & t+\theta \le 0. \end{cases}$$

LEMMA 3.4. *For* $0 \le s \le T$,

$$\mathbf{E} \sup_{s \le r \le s+\Delta t} \|R_{2,r}(s, \hat{x}_s)\|_\infty^2 \le K \Delta t^2.$$

Proof. This is a consequence of Lemma 3.3. □

If

$$R(t;s,x_s) = R_1(t;s,x_s) + \int_s^t Dg(x_s,s)(R_{2,r}(s,\hat{x}_s),0)\,d\beta(r),$$

then finally,

$$\begin{aligned} x(t;s,x_s) =& x(s) + f(x_s,s)(t-s) + g(x_s,s)I(s,t) \\ &+ \int_s^t Dg(x_s,s)\Big(G(\hat{x}_s,s)I_r(s) + \delta_r(s), r-s\Big)\,d\beta(r) + R(t;s,x_s). \end{aligned}$$

Let $t_k = k\Delta t$ with $\Delta t = T/n$. Let $x^n(t)$ solve for $t_k < t \le t_{k+1}$.

$$\begin{aligned} x^n(t) =& x^n(t_k) + f(x^n_{t_k},t_k)(t-t_k) + g(x^n_{t_k},t_k)\int_{t_k}^t d\beta(r) \\ &+ \int_{t_k}^t Dg(x^n_{t_k},t_k)\Big(G(\hat{x}^n_{t_k},t_k)I_r(t_k) + \delta_r(t_k), r-t_k\Big)\,d\beta(r). \end{aligned} \tag{3.7}$$

This is Milstein's method for (2.1) and our main result (Theorem 4.2) concludes that $x^n(t)$ is a first order approximation in Δt to $x(t)$ over the interval $[0,T]$.

4. Main result. Before proving Theorem 4.2, we give some preliminary results on the size of the remainder terms.

LEMMA 4.1. *For $t \le T$,*

$$\mathbf{E}\Big[\sup_{t_k\le t}\Big\|\sum_{j=0}^{k-1} R(t_{j+1};t_j,x_{t_j})\Big\|^2_{\mathbf{R}^d}\Big] \le K\Delta t^2 + \sup_{t_k\le t} 2\mathbf{E}\Big[\sum_{i,j=0}^{k-1}\langle R_X(t_i), R_X(t_j)\rangle\Big],$$

where

$$R_X(t_i) = \int_{t_i}^{t_{i+1}} Df(x_{t_i},t_i)(G(\hat{x}_{t_i},t_i)I_r(t_i),0)\,dr. \tag{4.1}$$

Proof. If $S_k = \sum_{j=0}^{k-1} r_{j+1}$, where r_k are $\mathbf{R}^d$ valued $\mathcal{F}_{t_k}$ measurable random variables, then $S_k - \mathbf{E}S_k$ is a discrete martingale and Doob's maximal inequality gives $\mathbf{E}\sup_{k\le n}\|S_k - \mathbf{E}S_k\|^2_{\mathbf{R}^d} \le 2\mathbf{E}\|S_n - \mathbf{E}S_n\|^2_{\mathbf{R}^d} \le 4\mathbf{E}\|S_n\|^2_{\mathbf{R}^d} + 4\|\mathbf{E}S_n\|^2_{\mathbf{R}^d}$. Hence,

$$\begin{aligned} \mathbf{E}\sup_{k\le n}\|S_k\|^2_{\mathbf{R}^d} \le& 8\mathbf{E}\|S_n\|^2_{\mathbf{R}^d} + 10\sup_{k\le n}\|\mathbf{E}S_k\|^2_{\mathbf{R}^d} \\ \le& 8\mathbf{E}\|S_n\|^2_{\mathbf{R}^d} + 10\sup_{k\le n}\mathbf{E}\|S_k\|^2_{\mathbf{R}^d} \le 18\sup_{k\le n}\mathbf{E}\|S_k\|^2_{\mathbf{R}^d}, \end{aligned}$$

because $\|\mathbf{E}X\| \le \mathbf{E}\|X\| \le (\mathbf{E}\|X\|^2)^{1/2}$. Now

$$\begin{aligned} \mathbf{E}\Big[\sup_{t_k\le t}\Big\|\sum_{j=0}^{k-1} R(t_{j+1};t_j,x_{t_j})\Big\|^2_{\mathbf{R}^d}\Big] \le& 8\sup_{t_k\le t}\mathbf{E}\Big[\Big\|\sum_{j=0}^{k-1} R(t_{j+1};t_j,x_{t_j})\Big\|^2_{\mathbf{R}^d}\Big] \\ &+ 10\sup_{t_k\le t}\Big\|\mathbf{E}\sum_{j=0}^{k-1} R(t_{j+1};t_j,x_{t_j})\Big\|^2_{\mathbf{R}^d} \end{aligned}$$

and

$$R(t;s,x_s) = \int_s^t R_f(r;s,x_s)\,dr + \int_s^t R_g(r;s,x_s)\,d\beta(r) + \int_s^t Dg(x_s,s)(R_{2,r}(s,\hat{x}_s),0)\,d\beta(r).$$

Hence,

$$\mathbf{E}\Big[\Big\|\sum_{j=0}^{k-1} R(t;t_j,x_{t_j})\Big\|_{\mathbf{R}^d}^2\Big] \le 3\mathbf{E}\Big\|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} R_f(r;t_j,x_{t_j})\,dr\Big\|_{\mathbf{R}^d}^2$$
$$+3\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} \mathbf{E}\|R_g(r;t_j,x_{t_j})\|_F^2\,dr$$
$$+3\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} \mathbf{E}\|Dg(x_s,s)(R_{2,r}(s,\hat{x}_s),0)\|_F^2\,dr$$

Lemmas 3.1 and 3.4 provide estimates of $\mathcal{O}(\Delta t^2)$ for the last two terms. For the first term, we further develop R_f from (3.1)

$$\begin{aligned} &R_f(r;s,x_s)\\ &= Df(x_s,s)(x_r-x_s,r-s)\\ &+\int_0^1 (1-h)D^2 f\Big(x_s+h(x_r-x_s),s+h(r-s)\Big)\Big(x_r-x_s,r-s\Big)^2\,dh\\ =&Df(x_s,s)(df(x_s,s)\hat{G}(x_s,s)I_r(s),r-s)+Df(x_s,s)(\delta_r(s)+R_{2,r}(s,\hat{x}_s),0)\\ &+\int_0^1 (1-h)D^2 f\Big(x_s+h(x_r-x_s),s+h(r-s)\Big)\Big(x_r-x_s,r-s\Big)^2\,dh \end{aligned}$$

so that

$$\Big\|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} R_f(r;t_j,x_{t_j})\,dr\Big\|_{\mathbf{R}^d}^2 = 2\Big\|\sum_{j=0}^{k-1} R_X(t_j)\Big\|_{\mathbf{R}^d}^2$$
$$+2\Big\|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} Df(x_{t_j},t_j)(0,r-t_j)$$
$$+Df(x_{t_j},t_j)(\delta_r(t_j)+R_{2,r}(t_j,\hat{x}_{t_j}),0)+\int_0^1(1-h)\times$$
$$\times D^2 f\Big(x_{t_j}+h(x_r-x_{t_j}),t_j+h(r-t_j)\Big)\Big(x_r-x_{t_j},r-t_j\Big)^2\,dh\,dr\Big\|_{\mathbf{R}^d}^2.$$

The second term is $\mathcal{O}(\Delta t^2)$ by applying definition of δ and Lipschitz property of the initial data ξ, Lemma 3.4, and (2.5). We find

$$\mathbf{E}\Big[\Big\|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}} R_f(r;t_j,x_{t_j})\,dr\Big\|_{\mathbf{R}^d}^2\Big] \le 2\mathbf{E}\Big[\Big\|\sum_{j=0}^{k-1} R_X(t_j)\Big\|_{\mathbf{R}^d}^2\Big] + K\Delta t^2$$

and the following observation completes the proof.

$$\Big\|\sum_{j=0}^{k-1} R_X(t_j)\Big\|_{\mathbf{R}^d}^2 = \sum_{i,j=0}^{k} \langle R_X(t_j), R_X(t_i)\rangle$$

□

Assumption 4.1. *For $\mathcal{P}(n) = \{(i,j)\colon i,j = 0,\dots,n\}$, suppose there exists $\mathcal{Q}(n) \subset \mathcal{P}(n)$ such that*

$$\mathbf{E}\,\langle R_X(t_i), R_X(t_j)\rangle \le \begin{cases} K\Delta t^3, & (i,j)\in\mathcal{Q}(n)\\ K\Delta t^4, & (i,j)\in\mathcal{P}(n)-\mathcal{Q}(n). \end{cases}$$

A calculation with (4.1) shows the assumption holds with $\mathcal{Q}(n)=\mathcal{P}(n)$ and in the stochastic ordinary differential equation case with $\mathcal{Q}(n)=\{(i,i)\colon i=1,\dots,n\}$, as

$$R_X(t_i) = df(x(t_i),t_i)\Big(g(x(t_i))\int_{t_i}^{t_{i+1}}\beta(r)\,dr, 0\Big)$$

and $\mathbf{E}\,\langle R_X(t_i), R_X(t_j)\rangle = 0$ for $i\neq j$ (for $i<j$, use the fact that $R_X(t_i)$ is $\mathcal{F}_{t_{i+1}}$ measurable and $\mathbf{E}\Big[R_X(t_j)|\mathcal{F}_{t_j}\Big]=0$). To show that the Milstein method has order one, we need $\mathcal{Q}(n)$ to have $\mathcal{O}(n)$ members, which as we show in Theorem 4.5 is true when f and g have finitely many discrete delays.

Theorem 4.2. *Suppose that $\mathcal{Q}(n)$ has $\mathcal{O}(n)$ members. Then,*

$$\Big(\mathbf{E}\sup_{t\in[-\tau,T]}\|x(t)-x^n(t)\|_{\mathbf{R}^d}^2\Big)^{1/2} \le K\Delta t. \tag{4.2}$$

Proof. For $t_k \le s < t_{k+1}$,

$$\begin{aligned} x^n(s)-x(s) =& x^n(t_k)-x(t_k)\\ &+\int_{t_k}^{\min\{t_{k+1},s\}}\Big(f(x^n_{t_k},t_k)-f(x_{t_k},t_k)\Big)\,dr\\ &+\int_{t_k}^{\min\{t_{k+1},s\}}\Big(g(x^n_{t_k},t_k)-g(x_{t_k},t_k)\Big)d\beta(r)\\ &+\int_{t_k}^{\min\{t_{k+1},s\}}\Big[Dg(x^n_{t_k},t_k)(G(\hat{x}^n_{t_k},t_k)I_{r,t_k}+\delta_r(t_k),r-t_k)\\ &\qquad -Dg(x_{t_k},t_k)(G(\hat{x}_{t_k},t_k)I_{r,t_k}+\delta_r(t_k),r-t_k)\Big]\,d\beta(r)\\ &+R(s;t_k,x_{t_k}). \end{aligned}$$

For $t_k \le s < t_{k+1}$, let

$$\begin{aligned} D(s) =& f(x^n_{t_k},t_k)-f(x_{t_k},t_k)\\ M(s) =& \Big[g(x^n_{t_k},t_k)-g(x_{t_k},t_k)\Big]\\ &+\Big[Dg(x^n_{t_k},t_k)(G(\hat{x}^n_{t_k},t_k)I_{s,t_k}+\delta_s(t_k),s-t_k)\\ &\qquad -Dg(x_{t_k},t_k)(G(\hat{x}_{t_k},t_k)I_{s,t_k}+\delta_s(t_k),s-t_k)\Big]. \end{aligned}$$

Then,

$$x^n(s) - x(s) = \int_0^s D(r)\,dr + \int_0^s M(r)\,d\beta(r)$$
$$+ \sum_{j=0}^{k-1} R(t_{j+1}; t_j, x_{t_j}) + R(s, t_k, x_{t_k}).$$

For $t_k \le t < t_{k+1}$, let $e(t) = \mathbf{E}\sup_{s\le t} \|x^n(s) - x(s)\|^2_{\mathbf{R}^d}$.

$$e(t) \le 4\mathbf{E}\Big[\sup_{s\le t}\Big\|\int_0^s D(r)\,dr\Big\|^2_{\mathbf{R}^d}\Big] + 4\mathbf{E}\Big[\sup_{s\le t}\Big\|\int_0^s M(r)\,d\beta(r)\Big\|^2_{\mathbf{R}^d}\Big]$$
$$+ 4\mathbf{E}\Big[\Big\|\sum_{j=0}^{k-1} R(t_{j+1}; t_j, x_{t_j})\Big\|^2_{\mathbf{R}^d}\Big] + 4\mathbf{E}\Big[\sup_{t_k\le s\le t} \|R(s; t_k, x_{t_k})\|^2_{\mathbf{R}^d}\Big].$$

1. For $t_k \le t < t_{k+1}$,

$$\mathbf{E}\Big[\sup_{s\le t}\Big\|\int_0^s D(r)\,dr\Big\|^2_{\mathbf{R}^d}\Big] \le K\int_0^t \mathbf{E}\Big[\sup_{s\le r}\|D(s)\|^2_{\mathbf{R}^d}\Big]\,dr$$
$$\le K\int_0^t \mathbf{E}\sup_{s\le r}\|f(x^n_s, s) - f(x_s, s)\|^2_{\mathbf{R}^d}\,dr$$
$$\le K\int_0^t e(r)\,dr.$$

2. By (2.3),

$$\mathbf{E}\Big[\sup_{s\le t}\Big\|\int_0^s M(r)\,d\beta(r)\Big\|^2_{\mathbf{R}^d}\Big] \le K\sup_{s\le t}\int_0^s \mathbf{E}\|M(r)\|^2_F\,dr.$$

Now, for $t_k \le s < t_{k+1}$,

$$\mathbf{E}\Big[\|M(s)\|^2_F\Big] \le 2\mathbf{E}\Big[\|g(x^n_{t_k}, t_k) - g(x_{t_k}, t_k)\|^2_F\Big]$$
$$+ 2\mathbf{E}\Big[\Big\|Dg(x^n_{t_k}, t_k)(G(\hat{x}^n_{t_k}, t_k)I_{s,t_k} + \delta_s(t_k), s - t_k)$$
$$- Dg(x_{t_k}, t_k)(G(\hat{x}_{t_k}, t_k)I_{s,t_k} + \delta_s(t_k), s - t_k)\Big\|^2_F\Big]$$
$$\le K\,e(s).$$

So

$$\mathbf{E}\Big[\sup_{s\le t}\Big\|\int_0^s M(r)\,d\beta(r)\Big\|^2_{\mathbf{R}^d}\Big] \le K\int_0^t e(s)\,ds.$$

3. Under Assumption 4.1 with the condition on $\mathcal{Q}(n)$ and Lemma 4.1,

$$\mathbf{E}\Big[\sup_{t_k\le t}\Big\|\sum_{j=0}^{k-1} R(t_{j+1}; t_j, x_{t_j})\Big\|^2_{\mathbf{R}^d}\Big] \le K\Delta t^2.$$

and we can also show

$$\mathbf{E}\Big[\sup_{t_k\le s\le t}\|R(s; t_k, x_{t_k})\|^2_{\mathbf{R}^d}\Big] \le K\Delta t^2.$$

Putting the estimates together, we have

$$e(t) \leq K \int_0^t e(s)\, ds + K\Delta t^2$$

and an application of Gronwall's inequality completes the proof. □

Assumption 4.2. *For $0 = \tau_1 < \tau_2 < \cdots < \tau_J \leq \tau$, suppose that*

$$f(\eta, t) = F(\eta(-\tau_1), \ldots, \eta(-\tau_J)), \qquad g(\eta, t) = G(\eta(-\tau_1), \ldots, \eta(-\tau_J)),$$

for $\eta \in C$. Denote $\partial_j f = \partial F(x_1, \ldots, x_J)/\partial x_j \in \mathbf{R}^{d\times d}$.

Lemma 4.3. *Suppose that Assumption 4.2 holds. For $0 \leq r \leq t$ and $s \geq 0$, let $\Phi(\eta) = x_t(r, \eta) - \mathbf{E}\Big[x_t(r,\eta)|\mathcal{F}_s\Big]$. Then $\Phi \in C^1(C, C)$ and*

$$\mathbf{E}\Big[\|D\Phi(\eta)\|_{\text{op}}^2\Big] \leq K|t-s|.$$

Proof. For $t + \theta \leq s$, $\Phi(\eta)(\theta) = 0$. For $t + \theta > s$ and $s \geq r$,

$$\begin{aligned}
\Phi(\eta)(\theta) =& x(t+\theta; r, \eta) - \mathbf{E}\Big[x(t+\theta; r,\eta)|\mathcal{F}_s\Big]\\
=& x(t+\theta; s, x_s(r,\eta)) - \mathbf{E}\Big[x(t+\theta; s, x_s(r,\eta))|\mathcal{F}_s\Big]\\
=& \int_s^{t+\theta} f(x_p(s, x_s(r,\eta)))\, dp + \int_s^{t+\theta} g(x_p(s, x_s(r,\eta)))\, d\beta(p)\\
&- \int_s^{t+\theta} \mathbf{E}\Big[f(x_p(s, x_s(r,\eta)))|\mathcal{F}_s\Big]\, dp\\
=& \int_s^{t+\theta} f(x_p(r,\eta)))\, dp + \int_s^{t+\theta} g(x_p(s, x_s(r,\eta)))\, d\beta(p)\\
&- \int_s^{t+\theta} \mathbf{E}\Big[f(x_p(r,\eta))|\mathcal{F}_s\Big]\, dp.
\end{aligned}$$

Under Assumption 4.2, $f(x_p(r,\eta))$ and $g(x_p(r,\eta))$ are mean square differentiable in η with uniformly bounded derivative:

$$\mathbf{E}\Big\|\frac{\partial f(x_p(r,\eta))}{\partial\eta}\Big\|_{\text{op}}^2 = \mathbf{E}\Big\|\sum_{j=1}^J \partial_j f(x_p(r,\eta))\frac{\partial x(p-\tau_j; r,\eta)}{\partial\eta}\Big\|_{\text{op}}^2 \leq K,$$

as $f \in C^1$, $\|\partial x(p-\tau_j; r, \eta)/\partial\eta\|_{\text{op}} \leq 1$ for $p - \tau_j \leq r$ and $\mathbf{E}\|\partial x(p-\tau_j; r,\eta)/\partial\eta\|_{\text{op}}^2 \leq K$ for $p - \tau_j \geq r$ from (2.9) of Theorem 2.2. A similar equation holds for the partial derivatives in g and we conclude that

$$\mathbf{E}\Big[\sup_{\theta\in[-\tau,0]} \Big\|D\Phi(\eta)(\theta)\Big\|_{\text{op}}^2\Big] \leq K|t-s|.$$

□

We will often work on the p times product space $C \times \cdots \times C$ and let

$$\|(w_1, w_2, \ldots, w_p)\| = \max\{\|w_1\|_\infty, \|w_2\|_\infty, \ldots, \|w_p\|_\infty\},$$

for $(w_1, w_2, \ldots, w_p) \in C \times \cdots \times C$.

Lemma 4.4. *Let Assumption 4.2 hold. Suppose that $p+\Delta t \le r \le s$ and that Δ_p, Δ_s are mean zero $\mathbf{R}^m$ random variables such that*

$$\Delta_p \text{ is } \mathcal{F}_{p+\Delta t} \text{ measurable and independent of } \mathcal{F}_p,$$
$$\Delta_s \text{ is independent of } \mathcal{F}_s.$$

Consider $h \in C^2(C\times C\times C\times C, \mathbf{R}^{m\times m})$. Then

$$\mathbf{E}\Big[\,\langle \Delta_p, h(x_p, x_r, x_s, x_{s+\Delta t})\Delta_s\rangle\Big] \le K\Delta t\Big(\mathbf{E}\|\Delta_p\|^4_{\mathbf{R}^m}\mathbf{E}\|\Delta_s\|^4_{\mathbf{R}^m}\Big)^{1/4}. \tag{4.3}$$

and, if Δ_p, Δ_s are $N(0,\sigma^2 I)$,

$$\mathbf{E}\Big[\,\langle \Delta_p, h(x_p, x_r, x_s, x_{s+\Delta t})\Delta_s\rangle\Big] \le K\Delta t\,\sigma^2.$$

Proof. Let $y=[x_p, x_r, x_s, x_{s+\Delta t}]$. By (2.6) of Theorem 2.1,

$$\mathbf{E}\|\bar y - y\|^p_\infty \le K\Delta t^{p/2}, \tag{4.4}$$

where $\bar y = \mathbf{E}[y|\mathcal{F}_s]$. Taylor's theorem provides

$$h(y) = h(\bar y) + Dh(\bar y)\,(y-\bar y) + R_h,$$

where the remainder satisfies $\|R_h\|_{op} \le \frac{1}{2}\|D^2 h\|_{\mathrm{op}}\,\|y-\bar y\|^2_\infty$ and, as the second derivative of h is uniformly bounded,

$$\|R_h\Delta_s\|_{\mathbf{R}^m} \le \|R_h\|_{\mathrm{op}}\|\Delta_s\|_{\mathbf{R}^m} \le K\|y-\bar y\|^2_\infty\|\Delta_s\|_{\mathbf{R}^m}.$$

The Cauchy-Schwarz inequality gives

$$\mathbf{E}\Big[\|R_h\Delta_s\|^2_{\mathbf{R}^m}\Big] \le K\Big(\mathbf{E}\Big[\|y-\bar y\|^8_\infty\Big]\mathbf{E}\Big[\|\Delta_s\|^4_{\mathbf{R}^m}\Big]\Big)^{1/2} \le K\Delta t^2\,\mathbf{E}\Big[\|\Delta_s\|^4_{\mathbf{R}^m}\Big]^{1/2}.$$

As Δ_s has mean zero and is independent of $\bar y$ and hence $h(\bar y)$,

$$\begin{aligned}\mathbf{E}\Big[h(y)\Delta_s|\mathcal{F}_{p+\Delta t}\Big] &= \mathbf{E}\Big[Dh(\bar y)(y-\bar y)\Delta_s|\mathcal{F}_{p+\Delta t}\Big] + \mathbf{E}\Big[R_h\Delta_s|\mathcal{F}_{p+\Delta t}\Big]\\ &= a(z) + \mathbf{E}\Big[R_h\Delta_s|\mathcal{F}_{p+\Delta t}\Big],\end{aligned}$$

where $z=[x_p, x_{p+\Delta t}]$ and

$$a(w) = \mathbf{E}\Big[Dh(\bar Y(w))\Big(Y(w)-\bar Y(w)\Big)\Delta_s\Big], \qquad w \in C\times C$$

where $Y(w) = [w_1, x_r(p+\Delta t, w_2), x_s(p+\Delta t, w_2), x_{s+\Delta t}(p+\Delta t, w_2)]$ and $\bar Y(w) = \mathbf{E}\Big[Y(w)|\mathcal{F}_s\Big]$ for $w=[w_1,w_2]\in C\times C$.

For $w, v \in C\times C$,

$$\begin{aligned}&|a(w)-a(v)|\\ \le&\,\mathbf{E}\Big[\Big\|Dh(\bar Y(w))(Y(w)-\bar Y(w)) - Dh(\bar Y(v))(Y(v)-\bar Y(v))\Big\|^2_{\mathrm{op}}\\ &\qquad\times\|\Delta_s\|^2_{\mathbf{R}^m}\Big]^{1/2}\\ \le&\,\mathbf{E}\Big[\Big\|Dh(\bar Y(w))(Y(w)-\bar Y(w)) - Dh(\bar Y(v))(Y(v)-\bar Y(v))\Big\|^2_{\mathrm{op}}\Big]^{1/2}\\ &\quad\times\mathbf{E}\Big[\|\Delta_s\|^2_{\mathbf{R}^m}\Big]^{1/2}.\end{aligned}$$

As $h \in C^2$,

$$\begin{aligned}
&\Big\|Dh(\bar{Y}(w))(Y(w)-\bar{Y}(w)) - Dh(\bar{Y}(v))(Y(v)-\bar{Y}(v))\Big\|_{\mathrm{op}} \\
\leq &\Big\|(Dh(\bar{Y}(w)-Dh(\bar{Y}(v)))(Y(w)-\bar{Y}(w))\Big\|_{\mathrm{op}} \\
+&\Big\|Dh(\bar{Y}(v))(Y(w)-\bar{Y}(w)-Y(v)+\bar{Y}(v))\Big\|_{\mathrm{op}} \\
\leq &K\Big\|\bar{Y}(w)-\bar{Y}(v)\Big\|_{\infty}\Big\|Y(w)-\bar{Y}(w)\Big\|_{\infty} \\
&+K\Big\|Y(w)-\bar{Y}(w)-Y(v)+\bar{Y}(v)\Big\|_{\infty}.
\end{aligned}$$

By (2.5) and (2.9),

$$\begin{aligned}
\mathbf{E}\Big[\|Y(w)-\bar{Y}(w)\|_{\infty}^{p}\Big] \leq &K\Delta t^{p/2}, \\
\mathbf{E}\Big[\|D\bar{Y}\|_{\mathrm{op}}^{p}\Big] \leq &K.
\end{aligned}$$

and by Lemma 4.3

$$\mathbf{E}\Big[\|Y(w)-\bar{Y}(w)-Y(v)+\bar{Y}(v)\|_{\infty}^{2}\Big] \leq K\Delta t\|w-v\|_{\infty}^{2}.$$

Hence,

$$\begin{aligned}
&\mathbf{E}\Big[\Big\|Dh(\bar{Y}(w))(Y(w)-\bar{Y}(w)) - Dh(\bar{Y}(v))(Y(v)-\bar{Y}(v))\Big\|_{\mathrm{op}}^{2}\Big] \\
\leq &K\|w-v\|_{\infty}^{2}\Delta t
\end{aligned}$$

or

$$|a(w)-a(v)| \leq K\Delta t^{1/2}\mathbf{E}\Big[\|\Delta_s\|_{\mathbf{R}^d}^{2}\Big]^{1/2}\|w-v\|_{\infty}.$$

Let $\bar{z} = \Big[x_p, \mathbf{E}\Big[x_{p+\Delta t}|\mathcal{F}_p\Big]\Big]$, which is independent of Δ_p, so that

$$\mathbf{E}\Big[\,\langle \Delta_p, a(z)\rangle\,\Big] = \mathbf{E}\Big[\,\langle \Delta_p, a(z)-a(\bar{z})\rangle\,\Big].$$

By (2.6),

$$\mathbf{E}\|\bar{z}-z\|_{\infty}^{p} \leq K\Delta t^{p/2}$$

and

$$\begin{aligned}
\mathbf{E}\|a(z)-a(\bar{z})\|_{\mathbf{R}^m}^{2} \leq &K\Delta t\mathbf{E}\Big[\|\Delta_s\|_{\mathbf{R}^m}^{2}\Big]\mathbf{E}\|z-\bar{z}\|_{\infty}^{2} \\
\leq &K\Delta t^{2}\mathbf{E}\Big[\|\Delta_s\|_{\mathbf{R}^m}^{2}\Big].
\end{aligned}$$

so that

$$\begin{aligned}
\mathbf{E}\Big[\,\langle \Delta_p, a(z)\rangle\,\Big] \leq &\mathbf{E}\Big[\|\Delta_p\|_{\mathbf{R}^m}\,\|a(z)-a(\bar{z})\|_{\mathbf{R}^m}\Big] \\
\leq &\Big[\mathbf{E}\|\Delta_p\|_{\mathbf{R}^m}^{2}\,\mathbf{E}\Big[\|a(z)-a(\bar{z})\|_{\mathbf{R}^m}^{2}\Big]\Big]^{1/2} \\
\leq &K\Delta t\mathbf{E}\Big[\|\Delta_p\|_{\mathbf{R}^m}^{2}\Big]^{1/2}\,\mathbf{E}\Big[\|\Delta_s\|_{\mathbf{R}^m}^{2}\Big]^{1/2}.
\end{aligned}$$

Putting everything together, as Δ_p is $\mathcal{F}_{p+h}$ measurable,

$$\mathbf{E}\Big[\,\langle \Delta_p, h(x_p,x_r,x_s,x_{s+\Delta t})\Delta_s\rangle\,\Big]=\mathbf{E}\Big[\,\Big\langle \Delta_p, \mathbf{E}\Big[h(y)\Delta_s|\mathcal{F}_{p+\Delta t}\Big]\Big\rangle\,\Big]$$

and the right hand side

$$\begin{aligned}
&\mathbf{E}\Big[\,\Big\langle \Delta_p, \mathbf{E}\Big[h(y)\Delta_s|\mathcal{F}_{p+\Delta t}\Big]\Big\rangle\,\Big]\\
=&\mathbf{E}\Big[\,\langle \Delta_p, a(z)\rangle+\Big\langle \Delta_p, \mathbf{E}\Big[R_h\Delta_s|\mathcal{F}_{p+\Delta t}\Big]\Big\rangle\,\Big]\\
\le& K\Delta t\mathbf{E}\Big[\|\Delta_p\|^2_{\mathbf{R}^m}\Big]^{1/2}\mathbf{E}\Big[\|\Delta_s\|^2_{\mathbf{R}^m}\Big]^{1/2}+\Big(\mathbf{E}\Big[\|\Delta_p\|^2_{\mathbf{R}^m}\Big]\mathbf{E}\Big[\|R_h\Delta_s\|^2_{\mathbf{R}^m}\Big]\Big)^{1/2}\\
\le& K\Delta t\mathbf{E}\Big[\|\Delta_p\|^2_{\mathbf{R}^m}\Big]^{1/2}\mathbf{E}\Big[\|\Delta_s\|^2_{\mathbf{R}^m}\Big]^{1/2}+K\Delta t\Big(\mathbf{E}\Big[\|\Delta_p\|^2_{\mathbf{R}^m}\Big]\mathbf{E}\Big[\|\Delta_s\|^4_{\mathbf{R}^m}\Big]\Big)^{1/4}.
\end{aligned}$$

The inequality (4.3) follows as $\big(\mathbf{E}\|\Delta\|^2_{\mathbf{R}^m}\big)^2\le\mathbf{E}\|\Delta\|^4_{\mathbf{R}^m}$.

To complete the proof, recall that for mean zero Gaussian random variables the fourth moment is proportional to the second moment squared. ∎

The final theorem says that under the assumption of discrete delays, Assumption 4.2, we can show that $\mathcal{Q}(n)$ has $\mathcal{O}(n)$ members and hence that Milstein method converges in mean square with order one.

THEOREM 4.5. *Suppose that Assumption 4.2 holds. If $\Delta t\le\tau_2$ then Assumption 4.1 holds where $\mathcal{Q}(n)$ has $\mathcal{O}(n)$ members. In particular, the error estimate (4.2) holds and the Milstein method converges with one.*

Proof. With $\Delta(s,t)=\int_s^t I(s,r)\,dr$,

$$R_X(t_i)=\sum_{a=1}^{J}\partial_a f(x_{t_i})g(x_{t_i-\tau_a})\Delta(t_i-\tau_a,t_{i+1}-\tau_a)$$

and for $t_i<t_j$

$$\begin{aligned}
&\mathbf{E}\Big[\,\langle R_X(t_i),R_X(t_j)\rangle\,\Big]\\
&=\sum_{a,b=1}^{J}\mathbf{E}\Big[\Big\langle\partial_a f(x_{t_i})g(x_{t_i-\tau_a})\Delta(t_i-\tau_a,t_{i+1}-\tau_a),\\
&\qquad\qquad\partial_b f(x_{t_j})g(x_{t_j-\tau_b})\Delta(t_j-\tau_b,t_{j+1}-\tau_b)\Big\rangle\Big]\\
&=\sum_{a,b=1}^{J}\mathbf{E}\Big[\Big\langle\Upsilon\Delta(t_i-\tau_a,t_{i+1}-\tau_a),\Delta(t_j-\tau_b,t_{j+1}-\tau_b)\Big\rangle\Big],
\end{aligned}$$

where we treat Υ in the following cases

1. for $t_i-\tau_a\le t_i\le t_{i+1}-\tau_a\le t_j-\tau_b\le t_j$
2. for $t_{i+1}-\tau_a\le t_i\le t_j-\tau_b\le t_j$
3. for $t_{i+1}-\tau_a\le t_j-\tau_b\le t_j$ and $t_i\ge t_j-\tau_b$
4. for $t_{j+1}-\tau_b\le t_i-\tau_a\le t_i<t_j$
5. $t_i-\tau_a\le t_{j+1}-\tau_b$ and $t_j-\tau_b\le t_{i+1}-\tau_a$.

The last case only occurs for $\mathcal{O}(n)$ pairs (i,j). For the first four cases, we give definitions of Υ such that Lemma 4.4 applies.

1. As $\Delta t < \tau_2$, it must be that $\tau_a = \tau_1 = 0$ and hence $t_i < t_{i+1} \le t_j - \tau_b \le t_j$.

$$\begin{aligned}\Upsilon =& h_{a,b}(x_{t_i}, x_{t_j-\tau_b}, x_{t_j-\tau_b}, x_{t_{j+1}-\tau_b})\\ =&\mathbf{E}\Big[g(x_{t_i})^T \partial_a f(x_{t_i})^T \partial_b f(x_{t_j}) g(x_{t_j-\tau_b}) | \mathcal{F}_{t_{j+1}-\tau_b}\Big]\\ h_{a,b}(\eta_1,\eta_2,\eta_3,\eta_4) =& \mathbf{E}\Big[g(\eta_1)^T \partial_a f(\eta_1)^T \partial_b f(x_{t_j}(t_{j+1}-\tau_b,\eta_4)) g(\eta_3))\Big]\end{aligned}$$

for $\tau_b \ge \Delta t$ and

$$h_{a,b}(\eta_1,\eta_2,\eta_3,\eta_4) = \mathbf{E}\Big[g(\eta_1)^T \partial_a f(\eta_1)^T \partial_b f(\eta_3) g(\eta_3)\Big]$$

for $\tau_b = 0$ (the second argument is included to make h have four arguments). By Corollary 2.4, $h_{a,b} \in C^2(C \times \cdots \times C, \mathbf{R}^{m\times m})$ and Lemma 4.4 applies with $p = t_i$, $r = t_j - \tau_b$, $s = t_j - \tau_b$.

2. for $t_{i+1} - \tau_a \le t_i \le t_j - \tau_b$,

$$\begin{aligned}\Upsilon =& h_{a,b}(x_{t_i-\tau_a}, x_{t_i}, x_{t_j-\tau_b}, x_{t_{j+1}-\tau_b})\\ =&\mathbf{E}\Big[g(x_{t_i-\tau_a})^T \partial_a f(x_{t_i})^T \partial_b f(x_{t_j}) g(x_{t_j-\tau_b}) | \mathcal{F}_{t_{j+1}-\tau_b}\Big]\\ h_{a,b} =& \mathbf{E}\Big[g(\eta_1)^T \partial_a f(\eta_2)^T \partial_b f(x_{t_j}(t_{j+1}-\tau_b,\eta_4)) g(\eta_3)\Big]\end{aligned}$$

for $\tau_b \ge \Delta t$ and

$$h_{a,b} = \mathbf{E}\Big[g(\eta_1)^T \partial_a f(\eta_2)^T \partial_b f(\eta_3) g(\eta_3)\Big]$$

for $\tau_b = 0$. Lemma 4.4 applies with $p = t_i - \tau_a$, $r = t_i$, $s = t_j - \tau_b$.

3. for $t_{i+1} - \tau_a \le t_j - \tau_b \le t_j$ and $t_i \ge t_j - \tau_b$

$$\begin{aligned}\Upsilon =& h_{a,b}(x_{t_i-\tau_a}, x_{t_{i+1}-\tau_a}, x_{t_j-\tau_b}, x_{t_{j+1}-\tau_b}),\\ =&\mathbf{E}\Big[g(x_{t_i-\tau_a})^T \partial_a f(x_{t_i})^T \partial_b f(x_{t_j}) g(x_{t_j-\tau_b}) | \mathcal{F}_{t_{j+1}-\tau_b}\Big]\\ h_{a,b} =& \mathbf{E}\Big[g(\eta_1)^T \partial_a f(x_{t_i}(x_{t_{j+1}-\tau_b},\eta_4))^T \partial_b f(x_{t_j}(t_{j+1-\tau_b},\eta_4)) g(\eta_3)\Big]\end{aligned}$$

for $\tau_b \ge \Delta t$ and

$$h_{a,b} = \mathbf{E}\Big[g(\eta_1)^T \partial_a f(x_{t_i}(x_{t_{j+1}-\tau_b},\eta_4))^T \partial_b f(\eta_3) g(\eta_3)\Big]$$

for $\tau_b = 0$ (the second argument is included to make h have four arguments). Lemma 4.4 applies with $p = t_i - \tau_a$, $r = t_{i+1} - \tau_a$, $s = t_j - \tau_b$.

4. for $t_{j+1} - \tau_b \le t_i - \tau_a < t_i < t_j$

$$\begin{aligned}\Upsilon =& h_{a,b}(x_{t_j-\tau_b}, x_{t_{j+1}-\tau_b}, x_{t_i-\tau_a}, x_{t_{i+1}-\tau_a})\\ =&\mathbf{E}\Big[g(x_{t_i-\tau_a})^T \partial_a f(x_{t_i})^T \partial_b f(x_{t_j}) g(x_{t_j-\tau_b}) | \mathcal{F}_{t_{i+1}-\tau_a}\Big].\\ h_{a,b} =& \mathbf{E}\Big[g(\eta_3)^T \partial_a f(x_{t_i}(t_{i+1}-\tau_a,\eta_4))^T \partial_b f(x_{t_j}(t_{i+1}-\tau_a,\eta_4)) g(\eta_1)\Big]\end{aligned}$$

for $\tau_a \ge \Delta t$ and

$$h_{a,b} = \mathbf{E}\Big[g(\eta_3)^T \partial_a f(\eta_3)^T \partial_b f(x_{t_j}(t_{i+1}-\tau_a,\eta_4)) g(\eta_1)\Big].$$

for $\tau_a = 0$. Lemma 4.4 applies with $p = t_j - \tau_b$, $r = t_{j+1} - \tau_b$, $s = t_i - \tau_a$.

If $t_j - \tau_a \neq t_i - \tau_b$ for any $a, b = 1, \ldots, J$, Lemma 4.4 applies with $\Delta_p = I(p, p + \Delta t)$ and $\Delta_s = I(s, s + \Delta t)$. Because Δ_p, Δ_s are $N(0, \sigma^2 I)$ with

$$\sigma^2 = \int_0^{\Delta t} \int_0^{\Delta t} \mathbf{E}[\beta_1(r)\beta_1(s)]\, dr\, ds \leq \Delta t^3,$$

we find

$$\mathbf{E}\Big[\, \langle R_X(t_i), R_X(t_j) \rangle \Big] \leq K \Delta t^4.$$

Hence $\mathcal{Q}(n)$ has $\mathcal{O}(n)$ members and Theorem 4.2 applies to complete the proof. □

References.

[1] A. Ambrosetti and G. Prodi, A Primer of Nonlinear Analysis, *Cambridge studies in advanced mathematics 34*, (1993), CUP.

[2] M. Arriojas, Y. Hu, Yaozhong, S.-E. Mohammed and G. Pap, A delayed Black and Scholes formula, *Stoch. Anal. Appl.* **25** (2007), no. 2, 471–492.

[3] E. Buckwar, R. Kuske, S.-E. Mohammed and T. Shardlow, Weak convergence of the Euler scheme for stochastic differential delay equations. *LMS J. Comput. Math.* **11** (2008), 60–99.

[4] Y. Hu, S.-E. Mohammed and F. Yan, Discrete-time approximations of stochastic delay equations: the Milstein scheme. *Ann. Probab.* **32** (2004), no. 1A, 265–314.

[5] S. Janković and D. Ilić, An analytical approximate method for solving stochastic integro differential equations, *J. Math. Anal. Appl.*, **320** (2006), 230–245.

[6] A. Jentzen and P.E. Kloeden, Taylor-expansions of solutions of stochastic partial differential equations with additive noise, *Annals of Probab.,* **38**, No. 2., (2010), 532–569.

[7] A. Jentzen and P. E. Kloeden, Pathwise Taylor schemes for random ordinary differential equations, *BIT* **49** (1) (2009), 113–140.

[8] P.E. Kloeden and A. Jentzen, Pathwise convergent higher order numerical schemes for random ordinary differential equations, *Proc. Roy. Soc. London A* **463** (2007), 2929–2944.

[9] P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Heidelberg, 1992; second revised printing 1999.

[10] X. Mao, *Stochastic Differential Equations and Applications*, Horwood, 1997.

[11] M. Milosević, M. Jovanović and S. Janković, An approximate method via Taylor series for stochastic functional differential equations, *J. Math. Anal. Appl.*, (submitted, 2009)

[12] G.N. Milstein, *Numerical Integration of Stochastic Differential Equations*, Kluwer, Dordrecht, 1995.

[13] S.-E. Mohammed, Stochastic Functional Differential Equations, Research Notes in Mathematics No. 99, Pitman Books Ltd., London (1984).

[14] S.-E. Mohammed and T. Zhang, Anticipating stochastic differential systems with memory. *Stochastic Process. Appl.* **119** (2009), no. 9, 2773–2802.

[15] M. Reiss, Unpublished lecture notes, Humboldt University, 2010.