Available online at www.sciencedirect.com
SCIENCE DIRECT®
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
Journal of Computational and Applied Mathematics 197 (2006) 89–121
www.elsevier.com/locate/cam

A modified Milstein scheme for approximation of stochastic delay differential equations with constant time lag

Norbert Hofmann, Thomas Müller-Gronbach*

Department of Mathematics, University of Magdeburg, Postfach 4120, Magdeburg, Germany

Received 19 October 2004; received in revised form 10 August 2005

Abstract

We introduce a modified Milstein scheme for pathwise approximation of scalar stochastic delay differential equations with constant time lag on a fixed finite time interval. Our algorithm is based on equidistant evaluation of the driving Brownian motion and is simply obtained by replacing iterated Itô-integrals by products of appropriate Brownian increments in the definition of the Milstein scheme. We prove that the piecewise linear interpolation of the modified Milstein scheme is asymptotically optimal with respect to the mean square L_2-error within the class of all pathwise approximations that use observations of the driving Brownian motion at equidistant points. Moreover, for a large class of equations our scheme is also asymptotically optimal for mean square approximation of the solution at the final time point. Our asymptotic optimality results are complemented by a comparison with the Euler scheme based on exact error formulas for a linear test equation. This comparison demonstrates the superiority of the modified Milstein scheme even for a very small number of discretization points. Finally, we provide a generalization of our approach to the case of a system of SDDEs with an arbitrary finite number of constant delays. We conjecture that the above optimality results carry over to the generalized scheme.

MSC: primary: 65C30; secondary: 60H10

Keywords: Stochastic delay differential equations; Pathwise approximation; Asymptotic optimality; Minimal errors; Exact error formulas

1. Introduction

In this paper we study pathwise approximation of a scalar stochastic delay differential equation (SDDE) in the Itô-sense

$$\begin{aligned} \mathrm{d}X(t) &= a(t, X(t), X(t-r))\,\mathrm{d}t + \sigma(t, X(t), X(t-r))\,\mathrm{d}W(t), \quad 0 \leqslant t \leqslant T, \\ X(t) &= \eta(t), \quad -r \leqslant t \leqslant 0, \end{aligned} \tag{1}$$

with constant time lag $r > 0$, initial path η, drift coefficient a, diffusion coefficient σ and a one-dimensional driving Brownian motion W. For the theory of SDDEs, or more general of stochastic functional differential equations, we refer to, e.g., [12,10].

* Corresponding author. Tel.: +49 391 6712445; fax: +49 391 6718651.
E-mail address: gronbach@mail.math.uni-magdeburg.de (T. Müller-Gronbach).

0377-0427/$ - see front matter
doi:10.1016/j.cam.2005.10.027

While the numerical analysis of stochastic ordinary differential equations (SODEs) is well studied, see, e.g., [11,7,18,16,14,17] for results and further references, much less is known for SDDEs. Numerical solutions of SDDE's are first studied in [19,20]. Convergence properties of the equidistant Euler–Maruyama scheme are established in [1,8,5]. A semi-implicit equidistant Euler method is studied in [9]. The Milstein scheme is analyzed in [6]. All of the above authors provide upper bounds with unspecified constants for the (mean square) error of the respective schemes at the discretization points. Under appropriate conditions on the initial path, the drift coefficient and the diffusion coefficient, the order of convergence of these methods (in terms of the number of equidistant discretization points) turns out to be the same as for SODEs, i.e., $\frac{1}{2}$ for the Euler scheme and 1 for the Milstein scheme. The questions of lower bounds and optimal methods have not been addressed so far.

In the present paper we introduce a new and easily applied method $\widehat{X}_N$ for pathwise approximation of the solution of Eq. (1). The basic idea is simply to replace iterated Itô-integrals that appear in the definition of the Milstein scheme by products of appropriate increments of the driving Brownian motion W. The resulting modified Milstein scheme is based only on N evaluations of W at equidistant points and it coincides with the Milstein scheme if no delay term is present in the diffusion coefficient. The approximation $\widehat{X}_N$ is defined by piecewise linear interpolation.

We prove that $\widehat{X}_N$ is asymptotically optimal with respect to the mean square L_2-error within the class of all pathwise approximations that are based on evaluations of W at equidistant discretization points, see Theorem 2. The resulting optimal order of convergence is $N^{-1/2}$ and the asymptotic constant is determined by smoothness properties of the solution of (1) in the mean square sense.

Furthermore, we provide an asymptotic first order expansion of the mean square error of $\widehat{X}_N$ at the final time point $t = T$. If the corresponding asymptotic constant is not zero then the proposed method is also asymptotically optimal for mean square approximation of $X(T)$ based on equidistant evaluation of W, see Theorem 1. In this case the resulting optimal order of convergence is $N^{-1/2}$, which is in sharp contrast to the optimal order N^{-1} for SODEs, see [15]. For a zero constant the order of convergence of the mean square error of $\widehat{X}_N(T)$ is at least $N^{-3/4}$.

We also address the question whether higher orders of convergence can be achieved if a non-equidistant discretization is used. The answer turns out to be negative in general. For any pathwise approximation that uses evaluations of the driving Brownian motion at N fixed points in the time interval, the order of convergence is at most $N^{-1/2}$ for the mean square L_2-error and also for the mean square error at $t = T$ if the corresponding constant is non-zero, see Theorem 3.

Our results on asymptotic optimality of the modified Milstein scheme are complemented by a comparison of $\widehat{X}_N$ with the equidistant Euler scheme on the basis of exact error formulas for a linear test equation. This comparison clearly demonstrates the superiority of $\widehat{X}_N$ with respect to both error criteria even for a very small number N of discretization points.

The paper is organized as follows. In Section 2 we state our assumptions on Eq. (1). The proposed method $\widehat{X}_N$ is motivated and defined in Section 3. Section 4 contains the error analysis of $\widehat{X}_N$ as well as the analysis of minimal errors for arbitrary discretizations. The comparison of the new algorithm with the Euler scheme is carried out in Section 5. In Section 6 we describe how to generalize our approach to the case of a system of SDDEs with an arbitrary finite number of constant delays. Proofs are postponed to Section 7 and the Appendix.

2. Assumptions

For convenience, we assume that the length of the time interval $[0, T]$ is a multiple of the time lag r,

$$T = m \cdot r,$$

with $m \in \mathbb{N}\backslash\{1\}$.

Furthermore, we impose the following conditions on the initial path η, the drift coefficient a and the diffusion coefficient σ:

(A) The initial path η is deterministic and Hölder continuous of order $\frac{1}{2}$.
(B) The drift coefficient a and the diffusion coefficient σ satisfy

$$a, \sigma \in C^{1,2}([0, T] \times \mathbb{R}^2)$$

with bounded spatial derivatives. Moreover, there exists $K > 0$ such that

$$|a^{(1,0,0)}(t, x_1, x_2)| + |\sigma^{(1,0,0)}(t, x_1, x_2)| \leqslant K \cdot (1 + |x_1| + |x_2|).$$

(C) Eq. (1) is non-deterministic, i.e.,

$$\int_0^T E(\sigma^2(t, X(t), X(t-r))\, \mathrm{d}t > 0.$$

Properties (A) and (B) imply that a pathwise unique strong solution of Eq. (1) with initial condition $X(t) = \eta(t)$ for $-r \leqslant t \leqslant 0$ exists and satisfies

$$E\left(\sup_{-r \leqslant t \leqslant T} |X(t)|^q\right) < \infty \tag{2}$$

as well as

$$E|X(s) - X(t)|^q \leqslant c \cdot |s-t|^{q/2}, \tag{3}$$

for every $q \geqslant 1$. Here, the constant $c > 0$ only depends on η, a, σ and the parameter q, see [5]. The assumption of a deterministic initial path η is for convenience only, see Remark 4.

Throughout the following we formally put

$$W(t) = 0, \quad t < 0,$$

$$X(t) = \eta(t) = 0, \quad t < -r,$$

as well as

$$\sigma(t, x_1, x_2) = 0, \quad t < 0, \quad x_1, x_2 \in \mathbb{R}.$$

3. The algorithm

Let $N \in \mathbb{N}$ with

$$N/m \in \mathbb{N}$$

and consider the resulting equidistant discretization

$$t_\ell = \ell \cdot T/N, \quad \ell = 0, 1, \ldots, N. \tag{4}$$

Note that discretization (4) contains the points $t_\ell - r = t_{\ell - N/m}$ for $\ell \geqslant N/m$. In particular, the points $r, 2\cdot r, \ldots, m \cdot r = T$ are included in the mesh.

We define a corresponding scheme $\widehat{X}_N$ by

$$\widehat{X}_N(-t_\ell) = \eta(-t_\ell), \quad \ell = 0, 1, \ldots, 2N/m,$$

and

$$\begin{aligned}
\widehat{X}_N(t_{\ell+1}) = \widehat{X}_N(t_\ell) &+ a(t_\ell, \widehat{X}_N(t_\ell), \widehat{X}_N(t_\ell - r)) \cdot (t_{\ell+1} - t_\ell) \\
&+ \sigma(t_\ell, \widehat{X}_N(t_\ell), \widehat{X}_N(t_\ell - r)) \cdot (W(t_{\ell+1}) - W(t_\ell)) \\
&+ \tfrac{1}{2} \cdot (\sigma\sigma^{(0,1,0)})(t_\ell, \widehat{X}_N(t_\ell), \widehat{X}_N(t_\ell - r)) \cdot ((W(t_{\ell+1}) - W(t_\ell))^2 - (t_{\ell+1} - t_\ell)) \\
&+ \tfrac{1}{2} \cdot \sigma(t_\ell - r, \widehat{X}_N(t_\ell - r), \widehat{X}_N(t_\ell - 2r)) \cdot \sigma^{(0,0,1)}(t_\ell, \widehat{X}_N(t_\ell), \widehat{X}_N(t_\ell - r)) \\
&\times (W(t_{\ell+1} - r) - W(t_\ell - r)) \cdot (W(t_{\ell+1}) - W(t_\ell)),
\end{aligned}$$

for $\ell = 0, 1, \ldots, N-1$. Piecewise linear interpolation yields a global approximation

$$\widehat{X}_N(t) = \frac{t_{\ell+1} - t}{t_{\ell+1} - t_\ell} \cdot \widehat{X}_N(t_\ell) + \frac{t - t_\ell}{t_{\ell+1} - t_\ell} \cdot \widehat{X}_N(t_{\ell+1}), \quad t \in [t_\ell, t_{\ell+1}],$$

on the time interval $[0, T]$. Note that $\widehat{X}_N$ is solely based on the equidistant Brownian increments $W(t_{\ell+1}) - W(t_\ell)$ and only the $N/m + 1$ values $\eta(0), \eta(-T/N), \ldots, \eta(-r)$ of the initial function η are needed for the computation of the approximation on the interval $[0, T]$.

Remark 1. The definition of the method $\widehat{X}_N$ is motivated by the Milstein scheme $\widehat{X}_N^{\mathrm{M}}$ corresponding to the discretization (4). Put

$$J_N(t) = \int_{t_\ell}^{t} (W(s-r) - W(t_\ell - r))\,\mathrm{d}W(s), \quad t \in]t_\ell, t_{\ell+1}].$$

Then $\widehat{X}_N^{\mathrm{M}}$ is given by

$$\widehat{X}_N^{\mathrm{M}}(-t_\ell) = \eta(-t_\ell), \quad \ell = 0, 1, \ldots, 2N/m,$$

and

$$\begin{aligned}
\widehat{X}_N^{\mathrm{M}}(t_{\ell+1}) = {} & \widehat{X}_N^{\mathrm{M}}(t_\ell) + a(t_\ell, \widehat{X}_N^{\mathrm{M}}(t_\ell), \widehat{X}_N^{\mathrm{M}}(t_\ell - r)) \cdot (t_{\ell+1} - t_\ell) \\
& + \sigma(t_\ell, \widehat{X}_N^{\mathrm{M}}(t_\ell), \widehat{X}_N^{\mathrm{M}}(t_\ell - r)) \cdot (W(t_{\ell+1}) - W(t_\ell)) \\
& + \tfrac{1}{2} \cdot (\sigma\sigma^{(0,1,0)})(t_\ell, \widehat{X}_N^{\mathrm{M}}(t_\ell), \widehat{X}_N^{\mathrm{M}}(t_\ell - r)) \cdot ((W(t_{\ell+1}) - W(t_\ell))^2 - (t_{\ell+1} - t_\ell)) \\
& + \sigma(t_\ell - r, \widehat{X}_N^{\mathrm{M}}(t_\ell - r), \widehat{X}_N^{\mathrm{M}}(t_\ell - 2r))\sigma^{(0,0,1)}(t_\ell, \widehat{X}_N^{\mathrm{M}}(t_\ell), \widehat{X}_N^{\mathrm{M}}(t_\ell - r)) \cdot J_N(t_{\ell+1}),
\end{aligned}$$

for $\ell = 0, 1, \ldots, N-1$. Under assumptions (A) and (B) this scheme satisfies

$$\max_{\ell=1,\ldots,N} E|X(t_\ell) - \widehat{X}_N^{\mathrm{M}}(t_\ell)|^2 \leqslant c \cdot 1/N^2,$$

where the constant $c > 0$ only depends on η, a and σ, see [6]. However, note that $\widehat{X}_N^{\mathrm{M}}$ does not only use Brownian increments but also the iterated Itô-integrals $J_N(t_{\ell+1})$. Due to Lemma 5 we have

$$\begin{aligned}
& E(J_N(t_{\ell+1}) \mid W(t_1), \ldots, W(t_N)) \\
& \quad = \tfrac{1}{2} \cdot (W(t_{\ell+1} - r) - W(t_\ell - r)) \cdot (W(t_{\ell+1}) - W(t_\ell)).
\end{aligned}$$

Replacing $J_N(t_{\ell+1})$ by this conditional expectation in the definition of $\widehat{X}_N^{\mathrm{M}}$ yields the scheme $\widehat{X}_N$. Clearly, $\widehat{X}_N$ coincides with the Milstein scheme $\widehat{X}_N^{\mathrm{M}}$ if $\sigma^{(0,0,1)} = 0$, i.e., if the diffusion coefficient $\sigma(t, x_1, x_2)$ does not depend on the delay state variable x_2.

4. Asymptotic error analysis

We formally introduce the error criteria that are used in this paper for the analysis of pathwise approximation of Eq. (1). Consider an arbitrary approximation

$$\overline{X} = \varphi(W(\tau_1), \ldots, W(\tau_N)),$$

where $\tau_1 \ldots \tau_N \in [0, T]$ and $\varphi : \mathbb{R}^N \to L_2([0, T])$ is measurable. The mean square L_2-error of $\overline{X}$ is defined by

$$e(\overline{X}) = \left(E \int_0^T (X(t) - \overline{X}(t))^2 \,\mathrm{d}t \right)^{1/2},$$

i.e., the approximation $\overline{X}$ is compared to the solution X globally on the time interval $[0, T]$. For the comparison of $\overline{X}$ with X at the final time point $t = T$ we define

$$e(\overline{X}, T) = (E|X(T) - \overline{X}(T)|^2)^{1/2}.$$

We compare the method $\widehat{X}_N$ with the best mean square approximation

$$\widehat{X}^*_N(t) = E(X(t) \mid W(t_1), \ldots, W(t_N)), \quad 0 \leqslant t \leqslant T,$$

based on the evaluation of W at discretization points (4).

First, we analyze the error $e(\widehat{X}_N, T)$ of the modified Milstein scheme $\widehat{X}_N$ at the final time point T. Consider the random field $\Phi(t,s)$ on $[0,T]^2$ given by

$$\Phi(t,s) = \begin{cases} 0 & \text{if } s < t, \\ 1 & \text{if } s = t, \end{cases}$$

and

$$\begin{aligned} \mathrm{d}\Phi(t,s) &= a^{(0,1,0)}(s, X(s), X(s-r)) \cdot \Phi(t,s)\,\mathrm{d}s + a^{(0,0,1)}(s, X(s), X(s-r)) \cdot \Phi(t, s-r)\,\mathrm{d}s \\ &\quad + \sigma^{(0,1,0)}(s, X(s), X(s-r)) \cdot \Phi(t,s)\,\mathrm{d}W(s) + \sigma^{(0,0,1)}(s, X(s), X(s-r)) \cdot \Phi(t, s-r)\,\mathrm{d}W(s), \end{aligned}$$

for $s \geqslant t$. Roughly speaking, $\Phi(t,\cdot)$ is the mean square derivative of the solution with respect to its state at time t, see Remark 5. Next, define a field $\vartheta(t,s)$ on $[0,T]^2$ by

$$\vartheta(t,s) = \Phi(t,s) \cdot \sigma(t-r, X(t-r), X(t-2r)) \cdot \sigma^{(0,0,1)}(t, X(t), X(t-r)),$$

and note that $\vartheta(t,s) = 0$ for $t < r$. Finally, put

$$\mathscr{C} = \left(T/4 \cdot \int_r^T E(\vartheta^2(t,T))\,\mathrm{d}t \right)^{1/2}.$$

Theorem 1. *The approximations $\widehat{X}_N$ and $\widehat{X}^*_N$ satisfy*[1]

$$e(\widehat{X}_N, T) \approx e(\widehat{X}^*_N, T) \approx \mathscr{C} \cdot N^{-1/2},$$

if $\mathscr{C} > 0$. If $\mathscr{C} = 0$ then

$$e(\widehat{X}_N, T) \leqslant c \cdot N^{-3/4},$$

where the constant $c > 0$ only depends on η, a and σ.

For the analysis of the mean square L_2-error $e(\widehat{X}_N)$ of the method $\widehat{X}_N$ we put

$$\overline{\mathscr{C}} = \left(T/6 \cdot \int_0^T E(\sigma^2(t, X(t), X(t-r)))\,\mathrm{d}t + T/4 \cdot \int_r^T \int_t^T E(\vartheta^2(t,s))\,\mathrm{d}s\,\mathrm{d}t \right)^{1/2}.$$

Note that $\overline{\mathscr{C}} > 0$ due to assumption (C).

Theorem 2. *The approximations $\widehat{X}_N$ and $\widehat{X}^*_N$ satisfy*

$$e(\widehat{X}_N) \approx e(\widehat{X}^*_N) \approx \overline{\mathscr{C}} \cdot N^{-1/2}.$$

By Theorem 2 the method $\widehat{X}_N$ is asymptotically optimal with respect to the mean square L_2-error on the time interval $[0,T]$. If the constant $\mathscr{C}$ is positive then, by Theorem 1, the method $\widehat{X}_N$ is also asymptotically optimal with respect to the error at the final time point T. Moreover, in this case the order of convergence of the respective minimal errors $e(\widehat{X}^*_N, T)$ and $e(\widehat{X}^*_N)$ is $N^{-1/2}$. See Remark 2 for a discussion of the case $\mathscr{C} = 0$.

We show that in general the order $N^{-1/2}$ cannot be improved if a non-equidistant discretization is used instead of (4). Consider an arbitrary discretization

$$0 < \tau_1 < \cdots < \tau_N \leqslant T$$

[1] We use $\approx$ to denote the strong asymptotic equivalence of sequences of real numbers, i.e., $a_n \approx b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$.

of the time interval $[0, T]$. Clearly, the conditional expectation

$$\widehat{X}^*_{\tau_1,\dots,\tau_N}(t) = E(X(t) \mid W(\tau_1), \dots, W(\tau_N)), \quad t \in [0, T],$$

is the optimal pathwise approximation of X based on the evaluations $W(\tau_1), \dots, W(\tau_N)$ with respect to both, the mean square L_2-error on $[0, T]$ and the mean square error at the final time point T. Put

$$e(N, T) = \inf\{e(\widehat{X}^*_{\tau_1,\dots,\tau_N}, T) : 0 < \tau_1 < \dots < \tau_N \leqslant T\}$$

as well as

$$e(N) = \inf\{e(\widehat{X}^*_{\tau_1,\dots,\tau_N}) : 0 < \tau_1 < \dots < \tau_N \leqslant T\}.$$

The quantities $e(N, T)$ and $e(N)$ are the minimal error at the point T and the minimal mean square L_2-error on the interval $[0, T]$, respectively, that can be achieved if the driving Brownian motion may be evaluated at N fixed time points in the interval $[0, T]$.

Theorem 3. *If $\mathscr{C} > 0$ then*[2]

$$e(N, T) \asymp 1/N^{1/2}.$$

Furthermore,

$$e(N) \asymp 1/N^{1/2}.$$

Remark 2. If the constant $\mathscr{C}$ is zero then, by Theorem 1, the order of convergence of $e(\widehat{X}_N, T)$ is at least $N^{-3/4}$. In particular, we have $\mathscr{C} = 0$ if the diffusion coefficient $\sigma(t, x_1, x_2)$ does not depend on the delay state variable x_2, i.e., if $\sigma^{(0,0,1)} = 0$. In this case, however, the modified Milstein scheme $\widehat{X}_N$ coincides with the Milstein scheme $\widehat{X}^{\mathrm{M}}_N$ and the order of convergence of $e(\widehat{X}_N, T)$ is at least N^{-1}, see Remark 1. On the other hand, if $\mathscr{C} = 0$ then the scheme $\widehat{X}_N$ may loose the property of being asymptotically optimal with respect to the error at the final time point T. A simple example is provided by the SODE

$$\mathrm{d}X(t) = t\,\mathrm{d}W(t), \quad 0 \leqslant t \leqslant T,$$

with initial value $X(0) = 0$. For this equation it is straightforward to check that

$$e(\widehat{X}_N, T) = e(\widehat{X}^{\mathrm{M}}_N, T) \approx \tfrac{1}{\sqrt{3}} \cdot T^{3/2} \cdot 1/N \approx 2 \cdot e(\widehat{X}^*_N, T).$$

We add that optimal pathwise approximation of scalar SODEs at the final time point based on a finite number of evaluations of the driving Brownian motion is studied in full detail in [15].

Remark 3. Optimal L_2-approximation of a scalar SODE

$$\mathrm{d}X(t) = a(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t), \quad 0 \leqslant t \leqslant T,$$

based on a finite number of evaluations of W is studied in [4]. In particular, these authors prove that

$$e(\widehat{X}^{\mathrm{M}}_N) \approx e(\widehat{X}^*_N) \approx 1/N^{1/2} \cdot (T/6)^{1/2} \cdot \left(\int_0^T E(\sigma^2(t, X(t))\,\mathrm{d}t)\right)^{1/2} \tag{5}$$

and

$$e(N) \approx 1/N^{1/2} \cdot (T/6)^{1/2} \cdot \int_0^T (E(\sigma^2(t, X(t))))^{1/2}\,\mathrm{d}t \tag{6}$$

hold under mild regularity conditions on the initial value $X(0)$, the drift coefficient a and the diffusion coefficient σ. Clearly, under assumptions (A) and (B), (5) is a consequence of Theorem 2. Furthermore, (6) implies $e(N) \asymp N^{-1/2}$, which is a particular instance of Theorem 3.

[2] By $\asymp$ we denote the weak asymptotic equivalence of sequences, i.e., $a_n \asymp b_n$ if $c_1 \leqslant a_n/b_n \leqslant c_2$ for sufficiently large n with positive constants c_1, c_2.

Remark 4. By a slight modification of our proofs we can cover the case of a stochastic initial segment η, too. Assume that η is a.s. continuous and independent of W, and that it satisfies

$$E\left(\sup_{-r\leqslant t\leqslant 0}|\eta(t)|^q\right)<\infty,$$

for a sufficiently large $q>0$ as well as

$$E|\eta(s)-\eta(t)|^2\leqslant c\cdot|s-t|,$$

for all $s,t\in[-r,0]$. Then Theorems 1–3 still hold.

For specific linear equations the respective asymptotic constants $\mathscr{C}$ and $\overline{\mathscr{C}}$ can be computed explicitly. We illustrate this fact by the following two examples.

Example 1. Consider the equation

$$\mathrm{d}X(t)=(1+\beta\cdot X(t-r))\,\mathrm{d}W(t),\quad 0\leqslant t\leqslant 3r,$$

with the constant initial path $\eta=1$. The solution is given by

$$X(t)=1+(1+\beta)\cdot W(t)+\beta\cdot(1+\beta)\cdot\int_r^t W(s-r)\,\mathrm{d}W(s)$$
$$+\beta^2\cdot(1+\beta)\cdot\int_{2r}^t\int_r^{s-r}W(u-r)\,\mathrm{d}W(u)\,\mathrm{d}W(s).$$

For $t\geqslant r$ we have

$$\Phi(t,s)=\begin{cases}1 & \text{if } s\in[t,t+r],\\ 1+\beta\cdot(W(s)-W(t+r)) & \text{if } s\in[t+r,3r],\end{cases}$$

which yields

$$\vartheta(t,s)=\begin{cases}\beta\cdot(1+\beta) & \text{if } t\in[r,2r],\quad s\in[t,t+r],\\ \beta\cdot(1+\beta)\cdot(1+\beta\cdot(W(s)-W(t+r))) & \text{if } t\in[r,2r],\quad s\in[t+r,3r],\\ \beta\cdot(1+\beta)\cdot(1+\beta\cdot W(t-2r)) & \text{if } t\in[2r,3r],\quad s\in[t,3r].\end{cases}$$

By straightforward calculations,

$$\int_r^{3r}E(\vartheta^2(t,3r))\,\mathrm{d}t=\beta^2\cdot(1+\beta)^2\cdot(2r+\beta^2\cdot r^2)\tag{7}$$

and

$$\int_r^{3r}\int_t^{3r}E(\vartheta^2(t,s))\,\mathrm{d}s\,\mathrm{d}t=\beta^2\cdot(1+\beta)^2\cdot(2r^2+\beta^2/3\cdot r^3).\tag{8}$$

Next, observe

$$\sigma(t,X(t),X(t-r))=1+\beta\cdot X(t-r),$$

to derive

$$E(\sigma^2(t,X(t),X(t-r)))=\begin{cases}(1+\beta)^2 & \text{if } t\in[0,r],\\ (1+\beta)^2\cdot(1+\beta^2\cdot(t-r)) & \text{if } t\in[r,2r],\\ (1+\beta)^2\cdot(1+\beta^2\cdot(t-r)+\beta^4/2\cdot(t-2r)^2) & \text{if } t\in[2r,3r],\end{cases}$$

and

$$\int_0^{3r} E(\sigma^2(t, X(t), X(t-r)))\,\mathrm{d}t = (1+\beta)^2 \cdot (3r + 2\beta^2 \cdot r^2 + \beta^4/6 \cdot r^3). \tag{9}$$

Using (7)–(9) we obtain

$$\mathscr{C} = \tfrac{\sqrt{3}}{2} \cdot |\beta \cdot (1+\beta)| \cdot r \cdot (2 + \beta^2 \cdot r)^{1/2},$$

$$\overline{\mathscr{C}} = \tfrac{1}{\sqrt{6}} \cdot |1+\beta| \cdot r \cdot (9 + 15\beta^2 \cdot r + 2\beta^4 \cdot r^2)^{1/2}.$$

Note that the case $\beta = -1$ is excluded by Condition (C).

Example 2. Consider the equation

$$\mathrm{d}X(t) = (\alpha_1 \cdot X(t) + \alpha_2 \cdot X(t-r))\,\mathrm{d}t + (\beta_1 \cdot X(t) + \beta_2 \cdot X(t-r))\,\mathrm{d}W(t), \quad 0 \leqslant t \leqslant 2r,$$

with $\beta_1 \neq 0$, $\alpha_1 \cdot \beta_2 = \alpha_2 \cdot \beta_1$ and constant initial path $\eta = 1$. We have

$$\Phi(t, s) = e^{(\alpha_1 - \beta_1^2/2)\cdot(s-t) + \beta_1 \cdot (W(s) - W(t))},$$

for $t \leqslant s \leqslant t + r$, see Remark 5, and

$$X(t) = (1 + \beta_2/\beta_1) \cdot \Phi(0, t) - \beta_2/\beta_1,$$

for $0 \leqslant t \leqslant r$. Thus

$$\vartheta(t, s) = \beta_2 \cdot (\beta_1 + \beta_2) \cdot \Phi(t, s) \cdot \Phi(0, t - r),$$

for $r \leqslant t \leqslant s \leqslant 2r$, which yields

$$\mathscr{C} = \tfrac{1}{\sqrt{2}} \cdot |\beta_2 \cdot (\beta_1 + \beta_2)| \cdot r \cdot e^{(\alpha_1 + \beta_1^2/2)\cdot r},$$

for the asymptotic constant from Theorem 1. Note that the case $\beta_1 = -\beta_2$ is excluded by assumption (C).

For the computation of the constant $\overline{\mathscr{C}}$ from Theorem 2 we additionally assume $\alpha_1 = \alpha_2 = 0$. Then it is easy to check that

$$\int_r^{2r} \int_t^{2r} E(\vartheta^2(t, s))\,\mathrm{d}s\,\mathrm{d}t = \beta_2^2/\beta_1^2 \cdot (1 + \beta_2/\beta_1)^2 \cdot (\beta_1^2 \cdot r \cdot e^{\beta_1^2 \cdot r} - e^{\beta_1^2 \cdot r} + 1). \tag{10}$$

By straightforward calculations,

$$E(\sigma^2(t, X(t), X(t-r))) = E((\beta_1 \cdot X(t) + \beta_2)^2) = (\beta_1 + \beta_2)^2 \cdot e^{\beta_1^2 \cdot t},$$

for $0 \leqslant t \leqslant r$. Moreover

$$E(X^2(t)) = \begin{cases} (1+\widetilde{\beta})^2 \cdot e^{\beta_1^2 \cdot t} - \widetilde{\beta}(2+\widetilde{\beta}) & \text{if } t \in [0, r], \\ (1+\widetilde{\beta})^2 \cdot e^{\beta_1^2 \cdot t} + \widetilde{\beta}^2(2+\widetilde{\beta})^2 & \\ \quad + \widetilde{\beta}(2+\widetilde{\beta})(1+\widetilde{\beta})^2 \cdot (\beta_1^2 \cdot (t-r) - 1) \cdot e^{\beta_1^2 \cdot (t-r)} & \text{if } t \in [r, 2r], \end{cases}$$

with $\widetilde{\beta} = \beta_2/\beta_1$, see [3, Lemma 15]. Hence

$$\begin{aligned} E(\sigma^2(t, X(t), X(t-r))) &= E((\beta_1 \cdot X(t) + \beta_2 \cdot X(t-r))^2) \\ &= (\beta_1 + \beta_2)^2 \cdot (\beta_2 \cdot (2\beta_1 + \beta_2) \cdot (t-r) \cdot e^{\beta_1^2 \cdot (t-r)} + e^{\beta_1^2 \cdot t}), \end{aligned}$$

for $r \leqslant t \leqslant 2r$. It follows:

$$\int_0^{2r} E(\sigma^2(t, X(t), X(t-r)))\,\mathrm{d}t$$
$$= (1+\beta_2/\beta_1)^2 \cdot (e^{2\beta_1^2 \cdot r} - 1) + (1+\beta_2/\beta_1)^2 \cdot \beta_2/\beta_1 \cdot (2+\beta_2/\beta_1) \cdot (\beta_1^2 \cdot r \cdot e^{\beta_1^2 \cdot r} - e^{\beta_1^2 \cdot r} + 1). \quad (11)$$

Combine (10) and (11) to obtain

$$\overline{\mathscr{C}} = \frac{1}{\sqrt{6}} \cdot |1+\beta_2/\beta_1| \cdot r^{1/2} \cdot (2(e^{2\beta_1^2 \cdot r} - 1) + \beta_2/\beta_1 \cdot (5\beta_2/\beta_1 + 4) \cdot (\beta_1^2 \cdot r \cdot e^{\beta_1^2 \cdot r} - e^{\beta_1^2 \cdot r} + 1))^{1/2},$$

for the asymptotic constant from Theorem 2.

Remark 5. Existence and uniqueness of the field Φ follows from general results of [12] on stochastic functional differential equations. In particular, for $0 \leqslant t \leqslant s \leqslant \min(t+r, T)$ we have

$$\Phi(t,s) = \exp\left(\int_t^s (a^{(0,1,0)} - \frac{1}{2} \cdot (\sigma^{(0,1,0)})^2)(u, X(u), X(u-r))\,\mathrm{d}u + \int_t^s \sigma^{(0,1,0)}(u, X(u), X(u-r))\,\mathrm{d}W(u)\right),$$

see, e.g., [21].

The field Φ may be interpreted in the following way. Fix $t \in [0,T]$ and consider, for every $h \in C([t-r,t])$, the solution X_h of the stochastic delay differential equation

$$\mathrm{d}X_h(s) = a(s, X_h(s), X_h(s-r))\,\mathrm{d}s + \sigma(s, X_h(s), X_h(s-r))\,\mathrm{d}W(s), \quad t \leqslant s \leqslant T,$$

with initial condition

$$X_h(s) = h(s), \quad t-r \leqslant s \leqslant t.$$

The distribution of the process X_h on $C([t,T])$ coincides with the conditional distribution of the solution $(X(s))_{t \leqslant s \leqslant T}$ given $(X(s))_{t-r \leqslant s \leqslant t} = h$, see [13]. As a consequence of condition (B), for every $s \geqslant t$ there exists the L_2-derivative $X'_h(s)$ of $X_h(s)$ with respect to its state at time t, i.e.,

$$\lim_{n\to\infty} E(1/h_n(t) \cdot (X_{h+h_n}(s) - X_h(s)) - X'_h(s))^2 = 0,$$

for every sequence $h_n \in C([t-r,t])$ with $h_n(t) > 0$ and $\lim_{n\to\infty}\|h_n\|_\infty = 0$. Moreover the process X'_h is the unique solution of the stochastic delay differential equation

$$\begin{aligned}\mathrm{d}X'_h(s) &= a^{(0,1,0)}(s, X_h(s), X_h(s-r)) \cdot X'_h(s)\,\mathrm{d}s + a^{(0,0,1)}(s, X_h(s), X_h(s-r)) \cdot X'_h(s-r)\,\mathrm{d}s \\ &\quad + \sigma^{(0,1,0)}(s, X_h(s), X_h(s-r)) \cdot X'_h(s)\,\mathrm{d}W(s) \\ &\quad + \sigma^{(0,0,1)}(s, X_h(s), X_h(s-r)) \cdot X'_h(s-r)\,\mathrm{d}W(s), \quad t \leqslant s \leqslant T,\end{aligned}$$

with initial condition $X'_h(s) = 1_{\{t\}}(s)$ for $t-r \leqslant s \leqslant t$. The latter two facts are straightforward extensions of the respective results for SODEs, see, e.g., [2]. Replacing X_h by the solution X yields the defining equation for the process $\Phi(t,\cdot)$.

We briefly comment on mean square smoothness properties of the fields Φ and ϑ. Using (A) and (B) as well as (2) and (3) it is easy to check that

$$E\left(\sup_{0 \leqslant t,s \leqslant T} |\Phi(t,s)|^q\right) + E\left(\sup_{0 \leqslant t,s \leqslant T} |\vartheta(t,s)|^q\right) \leqslant c \quad (12)$$

and

$$E|\Phi(t_1,s_1) - \Phi(t_2,s_2)|^q + E|\vartheta(t_1,s_1) - \vartheta(t_2,s_2)|^q \leqslant c \cdot (|t_1-t_2|^{q/2} + |s_1-s_2|^{q/2}), \quad (13)$$

for all $q \geqslant 1$, $r \leqslant t_1 \leqslant s_1 \leqslant T$, $r \leqslant t_2 \leqslant s_2 \leqslant T$, where the positive constant c only depends on a, σ, η and q.

5. A comparison with the Euler scheme based on exact error formulas

We compare the new algorithm $\widehat{X}_N$ with the Euler scheme. Recall discretization (4). The corresponding Euler scheme $\widehat{X}_N^{\mathrm{E}}$ is defined by

$$\widehat{X}_N^{\mathrm{E}}(-t_\ell) = \eta(-t_\ell), \quad \ell = 0, 1, \ldots, N/m,$$

and

$$\widehat{X}_N^{\mathrm{E}}(t_{\ell+1}) = \widehat{X}_N^{\mathrm{E}}(t_\ell) + a(t_\ell, \widehat{X}_N^{\mathrm{E}}(t_\ell), \widehat{X}_N^{\mathrm{E}}(t_\ell - r)) \cdot (t_{\ell+1} - t_\ell) \\ + \sigma(t_\ell, \widehat{X}_N^{\mathrm{E}}(t_\ell), \widehat{X}_N^{\mathrm{E}}(t_\ell - r)) \cdot (W(t_{\ell+1}) - W(t_\ell)),$$

for $\ell = 0, 1, \ldots, N-1$. By piecewise linear interpolation we obtain a global approximation

$$\widehat{X}_N^{\mathrm{E}}(t) = \frac{t_{\ell+1} - t}{t_{\ell+1} - t_\ell} \cdot \widehat{X}_N^{\mathrm{E}}(t_\ell) + \frac{t - t_\ell}{t_{\ell+1} - t_\ell} \cdot \widehat{X}_N^{\mathrm{E}}(t_{\ell+1}), \quad t \in [t_\ell, t_{\ell+1}],$$

on the time interval $[0, T]$.

As a test equation we use

$$\mathrm{d}X(t) = \beta \cdot (X(t) + X(t-1))\,\mathrm{d}W(t), \quad 0 \leqslant t \leqslant 2, \tag{14}$$

with $\beta \neq 0$ and constant initial path $\eta = 1$.

Our comparison of $\widehat{X}_N$ with $\widehat{X}_N^{\mathrm{E}}$ is based on exact formulas for both the mean square error at the final time point $t = 2$ and the mean square L_2-error. For the Euler scheme $\widehat{X}_N^{\mathrm{E}}$ these formulas are proven in [3, Appendix B]. The error formulas for the method $\widehat{X}_N$ can be shown in a similar way.

5.1. The error at the final time point

Put

$$\gamma_N = 1 + 2\beta^2/N, \quad \delta_N = 1 + 2\beta^2/N + 2\beta^4/N^2.$$

Lemma 1. *For Eq.* (14) *the mean square errors at $t = 2$ of the schemes $\widehat{X}_N$ and $\widehat{X}_N^{\mathrm{E}}$ are given by*

(i)

$$e^2(\widehat{X}_N^{\mathrm{E}}, 2) = 1/N \cdot 16\beta^4 \cdot \gamma_N^{N/2-1} + 4(e^{2\beta^2} - \gamma_N^N) + 12(\beta^2 - 1) \cdot (e^{\beta^2} - \gamma_N^{N/2})$$

and

(ii)

$$e^2(\widehat{X}_N, 2) = 1/N \cdot 2\beta^4 \cdot \delta_N^{N/2-1} + 4(e^{2\beta^2} - \delta_N^N) + 12(\beta^2 - 1) \cdot (e^{\beta^2} - \delta_N^{N/2}) \\ + 1/N^2 \cdot 8\beta^6/3 \cdot (5 - \beta^2/N) \cdot \delta_N^{N/2-1}.$$

Remark 6. Note that

$$e^{2\beta^2} - \gamma_N^N \approx 1/N \cdot 2\beta^4 \cdot e^{2\beta^2}. \tag{15}$$

Hence Lemma 1(i) yields

$$e(\widehat{X}_N^{\mathrm{E}}, 2) \approx e_{\mathrm{as}}(\widehat{X}_N^{\mathrm{E}}, 2) := 1/N^{1/2} \cdot 2\beta^2 \cdot e^{\beta^2/2} \cdot (2e^{\beta^2} + 3\beta^2 + 1)^{1/2}, \tag{16}$$

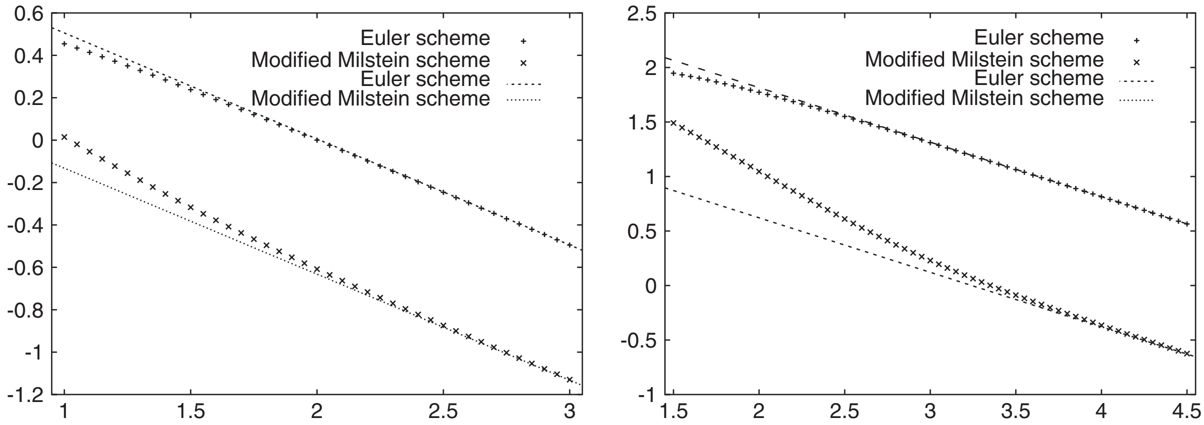

Fig. 1.

for the mean square error at $t = 2$ of the Euler scheme. For the new method $\widehat{X}_N$ we get

$$e(\widehat{X}_N, 2) \approx e_{\mathrm{as}}(\widehat{X}_N, 2) := 1/N^{1/2} \cdot \sqrt{2}\beta^2 \cdot e^{\beta^2/2}, \tag{17}$$

either from Lemma 1(ii) and

$$e^{2\beta^2} - \delta_N^N = \mathrm{o}(1/N), \tag{18}$$

or by taking $r = 1$, $\alpha_1 = \alpha_2 = 0$ and $\beta_1 = \beta_2 = \beta$ in Example 2.

Fig. 1 shows in $\log_{10}$ versus $\log_{10}$ coordinates, for $\beta = 1$ (left) and $\beta = 2$ (right), the dependence of the exact errors $e(\widehat{X}_N^{\mathrm{E}}, 2)$ and $e(\widehat{X}_N, 2)$ (vertical axes) on the number N of evaluations of the driving Brownian motion (horizontal axes). $+$ and $\times$ are computed by means of the exact error formulas from Lemma 1 and lines correspond to the asymptotic formulas (16) and (17). The graphs illustrate the superiority of the modified Milstein scheme $\widehat{X}_N$ over the Euler scheme. We observe a significantly better performance of the method $\widehat{X}_N$ even for very small numbers N. The asymptotic quantities $e_{\mathrm{as}}(\widehat{X}_N^{\mathrm{E}}, 2)$ and $e_{\mathrm{as}}(\widehat{X}_N, 2)$ are quickly approached. For $N \geqslant 1000$ we have

$$\frac{|e(\widehat{X}_N^{\mathrm{E}}, 2) - e_{\mathrm{as}}(\widehat{X}_N^{\mathrm{E}}, 2)|}{e_{\mathrm{as}}(\widehat{X}_N^{\mathrm{E}}, 2)} \leqslant \begin{cases} 0.011 & \text{if } \beta = 2, \\ 0.001 & \text{if } \beta = 1, \end{cases}$$

and

$$\frac{|e(\widehat{X}_N, 2) - e_{\mathrm{as}}(\widehat{X}_N, 2)|}{e_{\mathrm{as}}(\widehat{X}_N, 2)} \leqslant \begin{cases} 0.273 & \text{if } \beta = 2, \\ 0.003 & \text{if } \beta = 1. \end{cases}$$

Fig. 2 provides in $\log_{10}$ versus $\log_{10}$ coordinates, for $\beta = 1$ (left) and $\beta = 2$ (right), the plots of the minimal numbers $N^{\mathrm{E}}(\varepsilon, 2)$ and $N(\varepsilon, 2)$ of discretization points (vertical axes) that are necessary to achieve a mean square error at $t = 2$ of at most $\varepsilon > 0$ (horizontal axes) using the Euler scheme and the modified Milstein scheme, respectively. The number of discretization points needed to obtain a given accuracy is considerably reduced by using the modified Milstein scheme instead of the Euler scheme. For example, in the case $\beta = 2$, an equidistant discretization with $N^{\mathrm{E}}(0.01, 2) = 42346544$ points has to be used to obtain the accuracy $\varepsilon = 0.01$ with the Euler scheme, while the new algorithm only needs $N(0.01, 2) = 175330$ points.

The ratio $N(\varepsilon, 2)/N^{\mathrm{E}}(\varepsilon, 2)$ (vertical axes) versus $\log_{10} \varepsilon$ (horizontal axes) is plotted, for $\beta = 1$ (left) and $\beta = 2$ (right), in Fig. 3 for moderate values of ε. Compared to the Euler scheme the modified Milstein scheme yields a decrease of

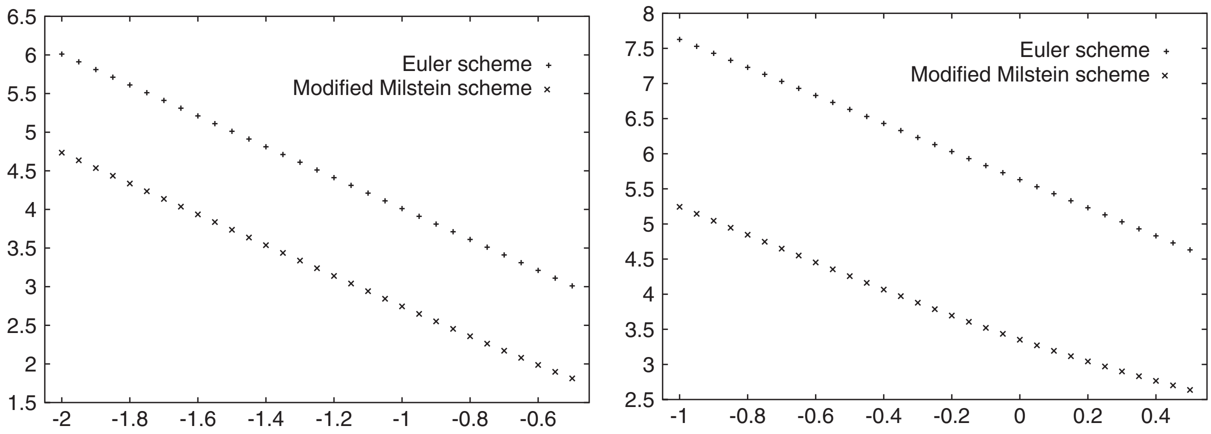

Fig. 2.

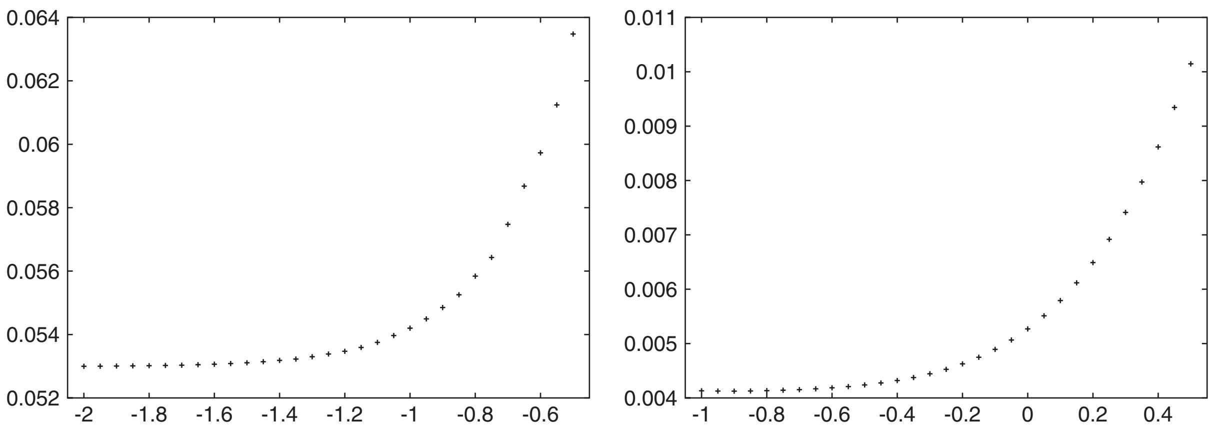

Fig. 3.

computational time (in terms of the number of discretization points) by a factor close to

$$(4e^{\beta^2}+6\beta^2+2)^{-1}=\begin{cases}0.00409\ldots & \text{if } \beta=2,\\ 0.05298\ldots & \text{if } \beta=1,\end{cases}$$

see Remark 6.

5.2. *The mean square L_2-error*

Recall the definition of γ_N and δ_N in Section 5.1 and put

$$\mu_N=(1+\beta^2/N)^{-2}.$$

Lemma 2. *For Eq.* (14) *the mean square L_2-errors of the approximations $\widehat{X}_N$ and $\widehat{X}_N^{\mathrm{E}}$ are given by*

(i)

$$\begin{aligned}e^2(\widehat{X}_N^{\mathrm{E}}) &= 4/\beta^2\cdot(e^{2\beta^2}-\gamma_N^N)+(12-24/\beta^2)\cdot(e^{\beta^2}-\gamma_N^{N/2})\\ &\quad -1/N\cdot\tfrac{8}{3}\cdot(\gamma_N^N-(6+3\beta^2)\cdot\gamma_N^{N/2-1}+5)\\ &\quad +1/N^2\cdot 4\beta^2/3\cdot((25-4\beta^2)\cdot\gamma_N^{N/2-1}-1)+1/N^3\cdot 8\beta^4/3\cdot\gamma_N^{N/2-1}\end{aligned}$$

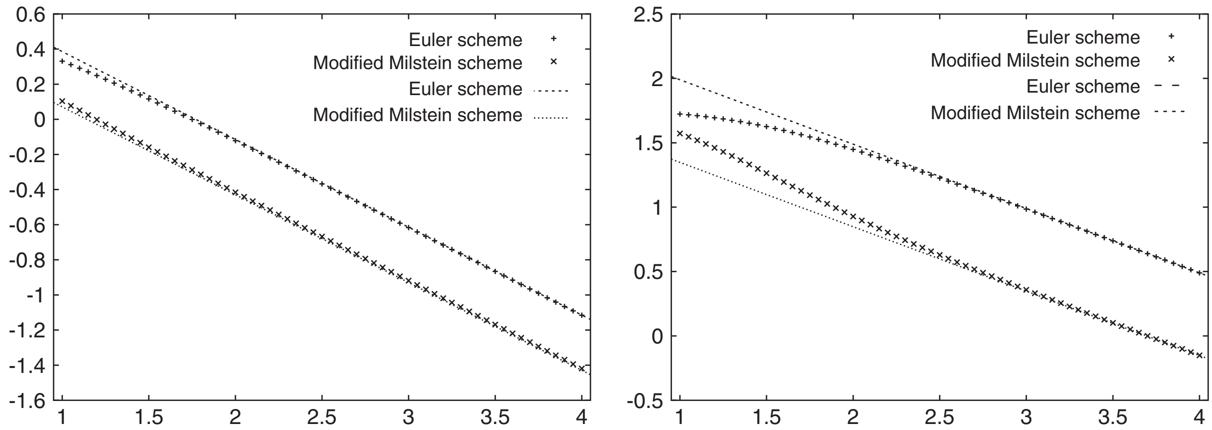

Fig. 4.

and

(ii)

$$
\begin{aligned}
e^2(\widehat{X}_N) = {} & 1/N \cdot 2\mu_N/3 \cdot (2\delta_N^N - (9-9\beta^2)\cdot \delta_N^{N/2-1} + 7) \\
& + 4/\beta^2 \cdot (e^{2\beta^2} - \delta_N^N) + (12 - 24/\beta^2)\cdot (e^{\beta^2} - \delta_N^{N/2}) \\
& - 1/N^2 \cdot 2\beta^2/3 \cdot \mu_N \cdot ((17-37\beta^2)\cdot \delta_N^{N/2-1} + 1) \\
& - 1/N^3 \cdot 4\beta^4/3 \cdot \mu_N \cdot \left(\delta_N^N + \left(\frac{28}{15} - 115\beta^2/6\right)\cdot \delta_N^{N/2-1} + \frac{77}{15}\right) \\
& + 1/N^4 \cdot 4\beta^6/3 \cdot \mu_N \cdot \left(\left(\frac{221}{15} - \beta^2/6\right)\cdot \delta_N^{N/2-1} - \frac{22}{15}\right) \\
& + 1/N^5 \cdot 4\beta^8/3 \cdot \mu_N \cdot \left(\left(\frac{233}{15} - 19\beta^2/3\right)\cdot \delta_N^{N/2-1} - \frac{1}{3}\right) \\
& + 1/N^6 \cdot 4\beta^{10}/3 \cdot \mu_N \cdot \left(\frac{18}{5} - \beta^2\right)\cdot \delta_N^{N/2-1} \\
& + 1/N^7 \cdot 8\beta^{12}/9 \cdot \mu_N \cdot \delta_N^{N/2-1}.
\end{aligned}
$$

Remark 7. From Lemma 2(i) and (15) we obtain

$$
e(\widehat{X}_N^{\mathrm{E}}) \approx e_{\mathrm{as}}(\widehat{X}_N^{\mathrm{E}}) := 1/N^{1/2} \cdot \left(\left(8\beta^2 - \frac{8}{3}\right)\cdot e^{2\beta^2} + (12\beta^4 - 16\beta^2 + 16)\cdot e^{\beta^2} - \frac{40}{3}\right)^{1/2}, \tag{19}
$$

for the piecewise linear interpolated Euler scheme. For the piecewise linear interpolated Milstein scheme $\widehat{X}_N$ we derive

$$
e(\widehat{X}_N) \approx e_{\mathrm{as}}(\widehat{X}_N) := 1/N^{1/2} \cdot \left(\frac{4}{3}\cdot e^{2\beta^2} + (6\beta^2 - 6)\cdot e^{\beta^2} + \frac{14}{3}\right)^{1/2}, \tag{20}
$$

either from Lemma 2(ii) and (18) or by taking $r=1$, $\alpha_1=\alpha_2=0$ and $\beta_1=\beta_2=\beta$ in Example 2.

Log_{10}–$\log_{10}$ plots of the exact errors $e(\widehat{X}_N^{\mathrm{E}})$ and $e(\widehat{X}_N)$ (vertical axes) versus the number N of discretization points (horizontal axes) are given in Fig. 4, for $\beta=1$ (left) and $\beta=2$ (right). $+$ and $\times$ are computed by use of the exact error formulas from Lemma 2 and lines are based on the asymptotic formulas (19) and (20). As in the case of the error at the final time point the new algorithm performs constantly better than the Euler approximation. The asymptotic quantities $e_{\mathrm{as}}(\widehat{X}_N^{\mathrm{E}})$ and $e_{\mathrm{as}}(\widehat{X}_N)$ provide an excellent approximation of the respective exact errors already for a small number of

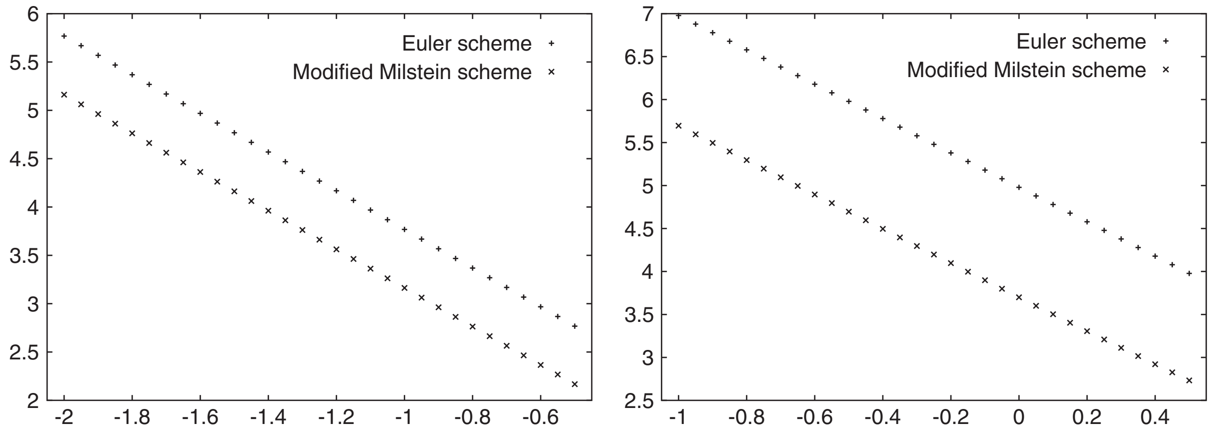

Fig. 5.

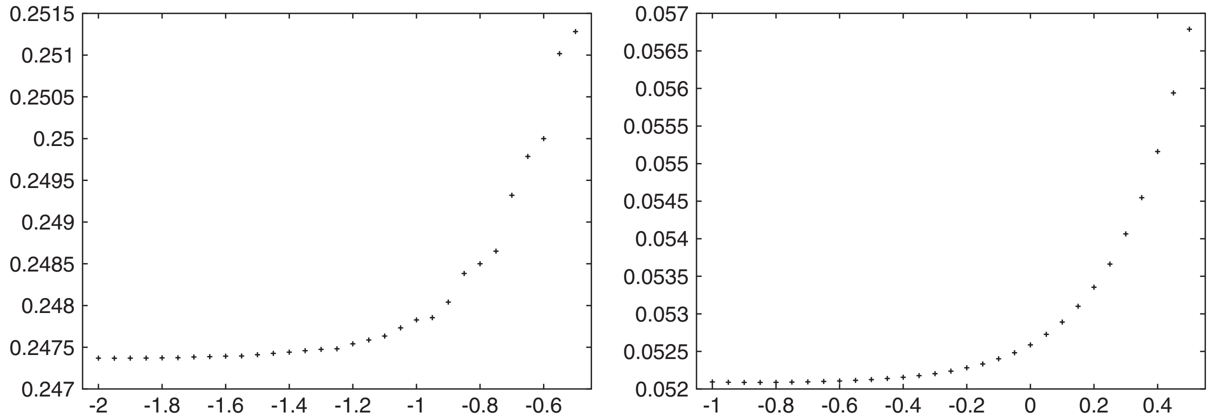

Fig. 6.

discretization points. For $N \geqslant 1000$ we have

$$\frac{|e(\widehat{X}_N^{\mathrm{E}}) - e_{\mathrm{as}}(\widehat{X}_N^{\mathrm{E}})|}{e_{\mathrm{as}}(\widehat{X}_N^{\mathrm{E}})} \leqslant \begin{cases} 0.010 & \text{if } \beta = 2, \\ 0.002 & \text{if } \beta = 1, \end{cases}$$

and

$$\frac{|e(\widehat{X}_N) - e_{\mathrm{as}}(\widehat{X}_N)|}{e_{\mathrm{as}}(\widehat{X}_N)} \leqslant \begin{cases} 0.024 & \text{if } \beta = 2, \\ 0.001 & \text{if } \beta = 1. \end{cases}$$

Fig. 5 shows (in the form of $\log_{10}$–$\log_{10}$ plots again), for $\beta = 1$ (left) and $\beta = 2$ (right), the minimal numbers $N^{\mathrm{E}}(\varepsilon)$ and $N(\varepsilon)$ of discretization points (vertical axes) that are needed to obtain a mean square L_2-error of at most $\varepsilon > 0$ (horizontal axes) using the Euler scheme and the modified Milstein scheme, respectively. The plots indicate once more the increase in performance achieved by the new approach compared to the Euler scheme.

If $\beta = 2$ then $N^{\mathrm{E}}(0.01) = 9525066$ points are necessary to achieve the accuracy $\varepsilon = 0.01$ with the Euler scheme. The piecewise linear interpolation of the modified Milstein scheme yields this accuracy already with $N(0.01) = 496250$ points.

Finally, in Fig. 6 we have plotted, for $\beta = 1$ (left) and $\beta = 2$ (right), the ratio $N(\varepsilon)/N^{\mathrm{E}}(\varepsilon)$ (vertical axes) versus $\log_{10}$ of the accuracy ε (horizontal axes). Using the new algorithm instead of the Euler scheme leads to a decrease of computational time (in terms of the number of discretization points) by a factor close to

$$\frac{2\cdot e^{2\beta^2}+9(\beta^2-1)\cdot e^{\beta^2}+7}{(12\beta^2-4)\cdot e^{2\beta^2}+(18\beta^4-24\beta^2+24)\cdot e^{\beta^2}-20}=\begin{cases}0.05510\ldots & \text{if } \beta=2,\\ 0.24736\ldots & \text{if } \beta=1,\end{cases}$$

see Remark 7.

6. Extensions

We briefly describe how to generalize our approach to the case of a system of SDDEs with an arbitrary finite number of constant delays

$$r_1,\ldots,r_K\in[0,r].$$

More precisely, we consider the d-dimensional system

$$\begin{aligned}&\mathrm{d}X(t)=a(t,X(t-r_1),\ldots,X(t-r_K))\,\mathrm{d}t+\sigma(t,X(t-r_1),\ldots,X(t-r_K))\,\mathrm{d}W(t),\quad 0\leqslant t\leqslant T,\\&X(t)=\eta(t),\quad -r\leqslant t\leqslant 0,\end{aligned}\tag{21}$$

with (for convenience) deterministic initial path

$$\eta=(\eta_1,\ldots,\eta_d)':[-r,0]\to\mathbb{R}^d,$$

m-dimensional driving Brownian motion

$$W=(W_1,\ldots,W_m)',$$

drift coefficient

$$a=(a_1,\ldots,a_d)':[0,T]\times\mathbb{R}^{K\cdot d}\to\mathbb{R}^d$$

and diffusion coefficient

$$\sigma=(\sigma_{i,j})_{\substack{1\leqslant i\leqslant d\\ 1\leqslant j\leqslant m}}:[0,T]\times\mathbb{R}^{K\cdot d}\to\mathbb{R}^{d\times m}.$$

We put

$$\nabla_k\sigma_{i,j}=\left(\frac{\partial}{\partial x_{(k-1)d+1}}\sigma_{i,j},\ldots,\frac{\partial}{\partial x_{kd}}\sigma_{i,j}\right),\quad k=1,\ldots,K,$$

and we use the notation

$$\sigma^{(j)}=\begin{pmatrix}\sigma_{1,j}\\ \vdots\\ \sigma_{d,j}\end{pmatrix},\quad \nabla_k\sigma^{(j)}=\begin{pmatrix}\nabla_k\sigma_{1,j}\\ \vdots\\ \nabla_k\sigma_{d,j}\end{pmatrix},\quad j=1,\ldots,m.$$

Moreover, we let $\eta(t)=0$ for $t<r$ and $W(t)=\sigma(t,x)=0$ for $t<0$ and $x\in\mathbb{R}^{K\cdot d}$.

Consider an equidistant discretizaton

$$t_\ell=\ell\cdot T/N,\quad \ell=0,1,\ldots,N,$$

let $\widetilde{W}$ denote the piecewise linear interpolation of the Brownian motion W at the points $-r,t_0,\ldots,t_N$, and put

$$I^{(\ell)}_{j_1,j_2,k}=\begin{cases}(W_{j_1}(t_{\ell+1})-W_{j_1}(t_\ell))^2-(t_{\ell+1}-t_\ell) & \text{if } j_1=j_2,\ r_k=0,\\ (\widetilde{W}_{j_1}(t_{\ell+1}-r_k)-\widetilde{W}_{j_1}(t_\ell-r_k))\cdot(W_{j_2}(t_{\ell+1})-W_{j_2}(t_\ell)) & \text{otherwise,}\end{cases}$$

for $j_1, j_2 = 1, \dots, m$ and $k = 1, \dots, K$. The modified Milstein scheme $\widehat{X}_N$ is then defined by

$$\widehat{X}_N(-\ell \cdot T/N) = \eta(-\ell \cdot T/N), \quad \ell = 0, 1, \dots, 2N \cdot \lceil r/T \rceil$$

and

$$\begin{aligned}\widehat{X}_N(t_{\ell+1}) = {} & \widehat{X}_N(t_\ell) + a(t_\ell, \widehat{X}_N(t_\ell - r_1), \dots, \widehat{X}_N(t_\ell - r_K)) \cdot (t_{\ell+1} - t_\ell) \\ & + \sigma(t_\ell, \widehat{X}_N(t_\ell - r_1), \dots, \widehat{X}_N(t_\ell - r_K)) \cdot (W(t_{\ell+1}) - W(t_\ell)) \\ & + \frac{1}{2} \cdot \sum_{k=1}^{K} \sum_{j_1, j_2 = 1}^{m} \nabla_k \sigma^{(j_1)}(t_\ell, \widehat{X}_N(t_\ell - r_1), \dots, \widehat{X}_N(t_\ell - r_K)) \\ & \times \sigma^{(j_2)}(t_\ell - r_k, \widehat{X}_N(t_\ell - r_k - r_1), \dots, \widehat{X}_N(t_\ell - r_k - r_K)) \cdot I^{(\ell)}_{j_1, j_2, k},\end{aligned}$$

where the intermediate values $\widehat{X}_N(t_\ell - r_k)$ are computed by piecewise linear interpolation of the scheme up to the point t_ℓ.

The SDDE considered in the preceding sections corresponds to the case $d = m = 1$, $K = 2$, $r_1 = 0$, $r_2 = r$ and $T = m \cdot r$ with $m \in \mathbb{N} \backslash \{1\}$. Then the above definition yields the method introduced in Section 3 if $N/m \in \mathbb{N}$. In this case, the piecewise linear interpolated scheme $\widehat{X}_N$ is asymptotically optimal with respect to the mean square L_2-error in the class of all methods that are based on evaluations of W at equidistant points, see Theorem 2. We conjecture that this remains true for general equation (21) under appropriate smoothness conditions on η, a and σ.

Finally, we stress that the generalized scheme may be used without any restriction on the lags $r_1, \dots, r_K$ or the time horizon T. Moreover, a non-equidistant discretization may be employed as well.

7. Proof of Theorems 1–3

The main idea of the proof is to introduce an appropriate auxiliary process X_N such that, asymptotically, pathwise approximation of the solution X based on N evaluations of W is equivalent to pathwise approximation of X_N with respect to both the mean square L_2-error and the mean square error at $t = T$ (if $\mathscr{C} > 0$).

The process X_N is given by

$$X_N = X_N^{\mathrm{Mt}} + L_N,$$

where X_N^{Mt} denotes a time-continuous truncated Milstein scheme and L_N is an approximation to the difference of X_N^{Mt} and the corresponding time-continuous Milstein scheme. More precisely, X_N^{Mt} is defined by

$$X_N^{\mathrm{Mt}}(t) = \eta(t), \quad t \leqslant 0,$$

and

$$\begin{aligned}X_N^{\mathrm{Mt}}(t) = {} & X_N^{\mathrm{Mt}}(t_\ell) + a(t_\ell, X_N^{\mathrm{Mt}}(t_\ell), X_N^{\mathrm{Mt}}(t_\ell - r)) \cdot (t - t_\ell) \\ & + \sigma(t_\ell, X_N^{\mathrm{Mt}}(t_\ell), X_N^{\mathrm{Mt}}(t_\ell - r)) \cdot (W(t) - W(t_\ell)) \\ & + 1/2 \cdot (\sigma\sigma^{(0,1,0)})(t_\ell, X_N^{\mathrm{Mt}}(t_\ell), X_N^{\mathrm{Mt}}(t_\ell - r)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell)),\end{aligned}$$

for $t \in]t_\ell, t_{\ell+1}]$. Recall the definition of the iterated Itô-integrals $J_N(t)$ in Remark 1. The process L_N is defined by

$$L_N(t) = 0, \quad -r \leqslant t \leqslant r$$

and

$$
\begin{aligned}
L_N(t) = L_N(t_\ell) &+ a^{(0,1,0)}(t_\ell, X_N^{\mathrm{Mt}}(t_\ell), X_N^{\mathrm{Mt}}(t_\ell - r)) \cdot L_N(t_\ell) \cdot (t - t_\ell) \\
&+ a^{(0,0,1)}(t_\ell, X_N^{\mathrm{Mt}}(t_\ell), X_N^{\mathrm{Mt}}(t_\ell - r)) \cdot L_N(t_\ell - r) \cdot (t - t_\ell) \\
&+ \sigma^{(0,1,0)}(t_\ell, X_N^{\mathrm{Mt}}(t_\ell), X_N^{\mathrm{Mt}}(t_\ell - r)) \cdot L_N(t_\ell) \cdot (W(t) - W(t_\ell)) \\
&+ \sigma^{(0,0,1)}(t_\ell, X_N^{\mathrm{Mt}}(t_\ell), X_N^{\mathrm{Mt}}(t_\ell - r)) \cdot L_N(t_\ell - r) \cdot (W(t) - W(t_\ell)) \\
&+ \sigma(t_\ell - r, X_N^{\mathrm{Mt}}(t_\ell - r), X_N^{\mathrm{Mt}}(t_\ell - 2r)) \\
&\times \sigma^{(0,0,1)}(t_\ell, X_N^{\mathrm{Mt}}(t_\ell), X_N^{\mathrm{Mt}}(t_\ell - r)) \cdot J_N(t),
\end{aligned}
$$

for $t \in]t_\ell, t_{\ell+1}]$ with $t_\ell \geqslant r$.

Due to Proposition 2 in the Appendix we have

$$\sup_{t\in[0,T]} E|X(t) - X_N(t)|^2 = \mathrm{O}(N^{-2}). \tag{22}$$

Thus, if $\overline{X}$ is a pathwise approximation of X that is based on evaluations of W at N equidistant points then, asymptotically, it suffices to analyze the difference $X_N - \overline{X}$.

We briefly outline the structure of this section. In Section 7.1 we introduce an approximation to the field ϑ from Section 4 and provide corresponding error bounds. Moreover, we investigate the relationship between the process L_N and the field ϑ. Section 7.2 deals with approximation of iterated Itô-integrals based on Brownian increments. The lower bounds in Theorems 1 and 2 are proven in Section 7.3. The matching upper bounds are shown in Section 7.4. In Section 7.5 we prove Theorem 3.

Throughout the following we use c to denote unspecified positive constants that only depend on r, T, the initial path η, the drift coefficient a, the diffusion coefficient σ and some moment parameter q. Moreover, we will use the notation

$$U(t) = (t, X(t), X(t-r)), \quad U_N^{\mathrm{Mt}}(t) = (t, X_N^{\mathrm{Mt}}(t), X_N^{\mathrm{Mt}}(t-r)).$$

7.1. Approximation of the field ϑ

We first introduce a time-continuous Euler-type approximation $\widehat{\Phi}_N$ of the field Φ based on discretization (4). This approximation is defined by

$$\widehat{\Phi}_N(t_\ell, t) = \begin{cases} 0 & \text{if } t < t_\ell, \\ 1 & \text{if } t = t_\ell, \end{cases}$$

and

$$
\begin{aligned}
\widehat{\Phi}_N(t_\ell, t) = \widehat{\Phi}_N(t_\ell, t_j) &+ a^{(0,1,0)}(U_N^{\mathrm{Mt}}(t_j)) \cdot \widehat{\Phi}_N(t_\ell, t_j) \cdot (t - t_j) \\
&+ a^{(0,0,1)}(U_N^{\mathrm{Mt}}(t_j)) \cdot \widehat{\Phi}_N(t_\ell, t_j - r) \cdot (t - t_j) \\
&+ \sigma^{(0,1,0)}(U_N^{\mathrm{Mt}}(t_j)) \cdot \widehat{\Phi}_N(t_\ell, t_j) \cdot (W(t) - W(t_j)) \\
&+ \sigma^{(0,0,1)}(U_N^{\mathrm{Mt}}(t_j)) \cdot \widehat{\Phi}_N(t_\ell, t_j - r) \cdot (W(t) - W(t_j)),
\end{aligned}
$$

for $t_\ell \leqslant t_j < t \leqslant t_{j+1}$. We use $\widehat{\Phi}_N$ to derive an approximation of the field ϑ. Define

$$\widehat{\vartheta}_N(t_\ell, t) = \begin{cases} 0 & \text{if } t < t_{\ell+1}, \\ \widehat{\Phi}_N(t_{\ell+1}, t) \cdot \sigma(U_N^{\mathrm{Mt}}(t_\ell - r)) \cdot \sigma^{(0,0,1)}(U_N^{\mathrm{Mt}}(t_\ell)) & \text{if } t \geqslant t_{\ell+1}, \end{cases}$$

for $r \leqslant t_\ell \leqslant t_{N-1}$.

Lemma 3. *The approximation $\widehat{\vartheta}_N$ satisfies*

(i)

$$E|\widehat{\vartheta}_N(t_\ell, t)|^q \leqslant c,$$

(ii)

$$E|\vartheta(t_\ell, t) - \widehat{\vartheta}_N(t_\ell, t)|^q \leqslant c \cdot N^{-q/2}.$$

for $t \geqslant t_{\ell+1} > r$ *and every* $q \geqslant 1$.

See [3] for a proof of Lemma 3. Finally, we provide a representation of the process L_N at the discretization points t_ℓ in terms of the approximation $\widehat{\vartheta}_N$ and the process J_N.

Lemma 4. *The process* L_N *satisfies*

$$L_N(t_\ell) = \sum_{r \leqslant t_j < t_\ell} \widehat{\vartheta}_N(t_j, t_\ell) \cdot J_N(t_{j+1}),$$

for every t_ℓ.

Proof. The proof is straightforward by induction on t_ℓ. □

7.2. Approximation of iterated Itô-integral based on Brownian increments

Consider a discretization

$$0 = \tau_0 < \tau_1 < \cdots < \tau_K = T$$

that satisfies

$$\min(\tau_k + r, T), \max(\tau_k - r, 0) \in \{\tau_0, \ldots, \tau_K\}, \tag{23}$$

for every $k = 0, 1, \ldots, K$, and put

$$\delta_k = \tau_{k+1} - \tau_k.$$

Fix $a_1, b_1, a_2, b_2 \in \{\tau_1, \ldots, \tau_K\}$ with

$$r \leqslant a_1 < b_1 \leqslant a_2 < b_2, \quad \max(b_1 - a_1, b_2 - a_2) \leqslant r$$

and put

$$I_j = \int_{a_j}^{b_j} (W(s-r) - W(a_j - r))\, \mathrm{d}W(s), \quad j = 1, 2.$$

Finally, let $\mathfrak{B}$ denote the σ-algebra that is generated by $W(\tau_0), W(\tau_1), \ldots, W(\tau_K)$.

Lemma 5. *The conditional expectation* $E(I_1 \mid \mathfrak{B})$ *satisfies*

$$E(I_1 \mid \mathfrak{B}) = \frac{1}{2} \cdot \sum_{a_1 \leqslant \tau_k < b_1} (W(\tau_k - r) + W(\tau_{k+1} - r) - 2W(a_1 - r)) \cdot (W(\tau_{k+1}) - W(\tau_k)),$$

as well as

$$E((I_1 - E(I_1 \mid \mathfrak{B}))^2 | \mathfrak{B}) = \frac{1}{12} \cdot \sum_{a_1 \leqslant \tau_k < b_1} (\delta_k^2 + \delta_k \cdot ((W(\tau_{k+1} - r) - W(\tau_k - r))^2 + (W(\tau_{k+1}) - W(\tau_k))^2))$$

and

$$E((W(t) - E(W(t) \mid \mathfrak{B})) \cdot (I_1 - E(I_1 \mid \mathfrak{B})) \mid \mathfrak{B}) = 0,$$

for $t \geqslant b_1$. *Furthermore,*

$$E((I_1 - E(I_1 \mid \mathfrak{B})) \cdot (I_2 - E(I_2 \mid \mathfrak{B})) \mid \mathfrak{B}) = 0.$$

Proof. We prove the above statements under the assumption

$$a_j = \tau_{k_j}, \quad b_j = \tau_{k_j+1}$$

with $k_2 > k_1$ and $\tau_{k_1} \geqslant r$. The general case may then easily be handled using the decomposition

$$I_j = \sum_{a_j \leqslant \tau_k < b_j} \int_{\tau_k}^{\tau_{k+1}} (W(s-r) - W(\tau_k - r)) \, \mathrm{d}W(s) + \sum_{a_j \leqslant \tau_k < b_j} (W(\tau_k - r) - W(a_j - r)) \cdot (W(\tau_{k+1}) - W(\tau_k)).$$

Let $n \in \mathbb{N}$, put

$$s_i = \tau_{k_1} + i/n \cdot \delta_{k_1}, \quad i = 0, 1, \ldots, n,$$

and consider the random variable

$$I_{1,n} = \sum_{i=1}^{n-1} (W(s_i - r) - W(\tau_{k_1} - r)) \cdot (W(s_{i+1}) - W(s_i)).$$

Clearly,

$$\lim_{n \to \infty} E(I_1 - I_{1,n})^2 = 0, \tag{24}$$

which implies

$$\lim_{n \to \infty} E(E(I_1 \mid \mathfrak{B}) - E(I_{1,n} \mid \mathfrak{B}))^2 = 0. \tag{25}$$

Let $Z = W - \widetilde{W}$, where $\widetilde{W} = E(W \mid \mathfrak{B})$ denotes the piecewise linear interpolation of W at the points τ_k. By (23),

$$W(s_i - r) - W(\tau_{k_1} - r) = Z(s_i - r) + \frac{s_i - \tau_{k_1}}{\delta_{k_1}} \cdot (W(\tau_{k_1+1} - r) - W(\tau_{k_1} - r)).$$

Furthermore,

$$W(s_{i+1}) - W(s_i) = Z(s_{i+1}) - Z(s_i) + \frac{s_{i+1} - s_i}{\delta_{k_1}} \cdot (W(\tau_{k_1+1}) - W(\tau_{k_1})).$$

Note that

$$E(Z(s) \cdot Z(t)) = \varepsilon(s, t) \tag{26}$$

with

$$\varepsilon(s, t) = \sum_{k=0}^{K-1} \frac{(\tau_{k+1} - \max(s, t))(\min(s, t) - \tau_k)}{\delta_k} \cdot 1_{[\tau_k, \tau_{k+1}]^2}(s, t).$$

Since $s_i - r \leqslant \tau_{k_1}$ it follows:

$$E(Z(s_i - r) \cdot (Z(s_{i+1}) - Z(s_i)) \mid \mathfrak{B}) = E(Z(s_i - r) \cdot (Z(s_{i+1}) - Z(s_i))) = 0.$$

Moreover,

$$E(Z(s_i - r) \cdot (W(\tau_{k_1+1}) - W(\tau_{k_1})) \mid \mathfrak{B}) = (W(\tau_{k_1+1}) - W(\tau_{k_1})) \cdot E(Z(s_i - r)) = 0$$

and similarly,

$$E((W(\tau_{k_1+1}-r)-W(\tau_{k_1}-r))\cdot(Z(s_{i+1})-Z(s_i))\,|\,\mathfrak{B})=0.$$

Summarizing we obtain

$$\begin{aligned}E(I_{1,n}\,|\,\mathfrak{B}) &= \sum_{i=1}^{n-1} E((W(s_i-r)-W(\tau_{k_1}-r))\cdot(W(s_{i+1})-W(s_i))\,|\,\mathfrak{B})\\ &= (W(\tau_{k_1+1}-r)-W(\tau_{k_1}-r))\cdot(W(\tau_{k_1+1})-W(\tau_{k_1}))\\ &\quad\times\sum_{i=1}^{n-1}\frac{(s_i-\tau_{k_1})\cdot(s_{i+1}-s_i)}{\delta_{k_1}^2}.\end{aligned} \tag{27}$$

Hence

$$\lim_{n\to\infty} E(I_{1,n}\,|\,\mathfrak{B})=\tfrac{1}{2}\cdot(W(\tau_{k_1+1}-r)-W(\tau_{k_1}-r))\cdot(W(\tau_{k_1+1})-W(\tau_{k_1})).$$

Now, use (25) to complete the proof of the first equality.

In order to prove the second equality note that (24) together with (25) implies

$$\lim_{n\to\infty} E|E((I_1-E(I_1\,|\,\mathfrak{B}))^2\,|\,\mathfrak{B})-E((I_{1,n}-E(I_{1,n}\,|\,\mathfrak{B}))^2\,|\,\mathfrak{B})|=0. \tag{28}$$

Observing (27) we can write

$$E(I_{1,n}\,|\,\mathfrak{B})=\sum_{i=0}^{n-1}(\widetilde{W}(s_i-r)-W(\tau_{k_1}-r))\cdot(\widetilde{W}(s_{i+1})-\widetilde{W}(s_i)).$$

Thus

$$I_{1,n}-E(I_{1,n}\,|\,\mathfrak{B})=\sum_{i=0}^{n-1}(A_{1,i}+B_{1,i}+C_{1,i}), \tag{29}$$

where

$$A_{1,i}=Z(s_i-r)\cdot(Z(s_{i+1})-Z(s_i)),$$

$$B_{1,i}=Z(s_i-r)\cdot(\widetilde{W}(s_{i+1})-\widetilde{W}(s_i)),$$

$$C_{1,i}=(\widetilde{W}(s_i-r)-W(\tau_{k_1}-r))\cdot(Z(s_{i+1})-Z(s_i)).$$

Due to (26) we have

$$E(A_{1,i}\cdot B_{1,j}\,|\,\mathfrak{B})=E(A_{1,i}\cdot C_{1,j}\,|\,\mathfrak{B})=E(B_{1,i}\cdot C_{1,j}\,|\,\mathfrak{B})=0,$$

which implies

$$E((I_{1,n}-E(I_{1,n}\,|\,\mathfrak{B}))^2\,|\,\mathfrak{B})=E\left(\left(\sum_{i=0}^{n-1}A_{1,i}\right)^2\middle|\,\mathfrak{B}\right)+E\left(\left(\sum_{i=0}^{n-1}B_{1,i}\right)^2\middle|\,\mathfrak{B}\right)+E\left(\left(\sum_{i=0}^{n-1}C_{1,i}\right)^2\middle|\,\mathfrak{B}\right).$$

Put

$$\alpha_{i,j}=\begin{cases} s_{i+1}-s_i-(s_{i+1}-s_i)^2\cdot\delta_{k_1}^{-1} & \text{if } i=j,\\ -(s_{i+1}-s_i)\cdot(s_{j+1}-s_j)\cdot\delta_{k_1}^{-1} & \text{if } i\neq j.\end{cases}$$

Straightforward calculations yield

$$E(A_{1,i} \cdot A_{1,j} \mid \mathfrak{B}) = \varepsilon(s_i, s_j) \cdot \alpha_{i,j},$$

$$E(B_{1,i} \cdot B_{1,j} \mid \mathfrak{B}) = \varepsilon(s_i, s_j) \cdot (W(\tau_{k_1+1}) - W(\tau_{k_1}))^2 \cdot (s_{i+1} - s_i) \cdot (s_{j+1} - s_j) \cdot \delta_{k_1}^{-2},$$

$$E(C_{1,i} \cdot C_{1,j} \mid \mathfrak{B}) = (W(\tau_{k_1+1} - r) - W(\tau_{k_1} - r))^2 \cdot (s_i - \tau_{k_1}) \cdot (s_j - \tau_{k_1}) \cdot \delta_{k_1}^{-2} \cdot \alpha_{i,j}.$$

Hence

$$\lim_{n\to\infty} E\left(\left(\sum_{i=0}^{n-1} A_{1,i}\right)^2 \middle| \mathfrak{B}\right) = \int_{\tau_{k_1}}^{\tau_{k_1+1}} \varepsilon(s,s)\,\mathrm{d}s - \delta_{k_1}^{-1} \cdot \int_{\tau_{k_1}}^{\tau_{k_1+1}} \int_{\tau_{k_1}}^{\tau_{k_1+1}} \varepsilon(s,t)\,\mathrm{d}s\,\mathrm{d}t = \delta_{k_1}^2/12.$$

Furthermore,

$$\begin{aligned}\lim_{n\to\infty} E\left(\left(\sum_{i=0}^{n-1} B_{1,i}\right)^2 \middle| \mathfrak{B}\right) &= (W(\tau_{k_1+1}) - W(\tau_{k_1}))^2 \cdot \delta_{k_1}^{-2} \cdot \int_{\tau_{k_1}}^{\tau_{k_1+1}} \int_{\tau_{k_1}}^{\tau_{k_1+1}} \varepsilon(s,t)\,\mathrm{d}s\,\mathrm{d}t \\ &= (W(\tau_{k_1+1}) - W(\tau_{k_1}))^2 \cdot \delta_{k_1}/12\end{aligned}$$

and

$$\begin{aligned}\lim_{n\to\infty} E\left(\left(\sum_{i=0}^{n-1} C_{1,i}\right)^2 \middle| \mathfrak{B}\right) &= (W(\tau_{k_1+1} - r) - W(\tau_{k_1} - r))^2 \cdot \delta_{k_1}^{-2} \cdot \int_{\tau_{k_1}}^{\tau_{k_1+1}} (s - \tau_{k_1})^2\,\mathrm{d}s \\ &\quad - (W(\tau_{k_1+1} - r) - W(\tau_{k_1} - r))^2 \cdot \delta_{k_1}^{-3} \cdot \left(\int_{\tau_{k_1}}^{\tau_{k_1+1}} (s - \tau_{k_1})\,\mathrm{d}s\right)^2 \\ &= (W(\tau_{k_1+1} - r) - W(\tau_{k_1} - r))^2 \cdot \delta_{k_1}/12.\end{aligned}$$

We conclude that

$$\begin{aligned}&\lim_{n\to\infty} E((I_{1,n} - E(I_{1,n} \mid \mathfrak{B}))^2 \mid \mathfrak{B}) \\ &\quad = \tfrac{1}{12} \cdot (\delta_{k_1}^2 + \delta_{k_1} \cdot ((W(\tau_{k_1+1} - r) - W(\tau_{k_1} - r))^2 + (W(\tau_{k_1+1}) - W(\tau_{k_1}))^2)).\end{aligned}$$

In view of (28) this finishes the proof of the second equality.

Next, let $t > \tau_{k_1+1}$. Use (26) and (29) to obtain

$$E(Z(t) \cdot (I_{1,n} - E(I_{1,n} \mid \mathfrak{B})) \mid \mathfrak{B}) = 0.$$

This establishes the third equality since (24) together with (25) implies

$$\lim_{n\to\infty} E \mid E(Z(t) \cdot ((I_1 - E(I_1 \mid \mathfrak{B})) - (I_{1,n} - E(I_{1,n} \mid \mathfrak{B}))) \mid \mathfrak{B})| = 0.$$

Finally, define $I_{2,n}$ analogously to $I_{1,n}$ and use decomposition (29) to obtain

$$E((I_{1,n} - E(I_{1,n} \mid \mathfrak{B})) \cdot (I_{2,n} - E(I_{2,n} \mid \mathfrak{B})) \mid \mathfrak{B}) = 0.$$

Clearly, by (24) and (25),

$$\begin{aligned}&\lim_{n\to\infty} E|E(((I_1 - E(I_1 \mid \mathfrak{B})) \cdot ((I_2 - E(I_2 \mid \mathfrak{B}))|\mathfrak{B}) \\ &\quad - E((I_{1,n} - E(I_{1,n} \mid \mathfrak{B})) \cdot (I_{2,n} - E(I_{2,n} \mid \mathfrak{B})) \mid \mathfrak{B})| = 0,\end{aligned}$$

which completes the proof. □

7.3. Proof of the lower bounds in Theorems 1 and 2

Let $\mathfrak{B}_N$ denote the σ-algebra that is generated by $W(t_1), \ldots, W(t_N)$ and recall that $\widehat{X}_N^* = E(X \mid \mathfrak{B}_N)$. Put $Z_N = W - E(W \mid \mathfrak{B}_N)$ as well as

$$d_\ell = T/N + (W(t_{\ell+1} - r) - W(t_\ell - r))^2 + (W(t_{\ell+1}) - W(t_\ell))^2,$$

and define

$$A_N = \frac{T}{6N} \cdot \sum_{\ell=0}^{N-1} \sigma^2(U(t_\ell)),$$

$$B_N = \frac{T}{12N} \cdot \sum_{\ell=0}^{N-2} \left(\sum_{j=\ell+1}^{N-1} \vartheta^2(t_\ell, t_j) \right) \cdot d_\ell.$$

Lemma 6.

$$e(\widehat{X}_N^*) \geqslant (T/N)^{1/2} \cdot (E(A_N + B_N))^{1/2} - c \cdot N^{-3/4}.$$

Proof. Due to (22) we have

$$e(\widehat{X}_N^*) \geqslant \left(\int_0^T E|\widehat{X}_N^*(t) - X_N(t)|^2 \, \mathrm{d}t \right)^{1/2} - c/N. \tag{30}$$

Fix ℓ, let $t \in]t_\ell, t_{\ell+1}]$ and put

$$\gamma(t) = \tfrac{1}{2} \cdot (\sigma\sigma^{(0,1,0)})(t_\ell, X_N^{\mathrm{Mt}}(t_\ell), X_N^{\mathrm{Mt}}(t_\ell - r)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell)).$$

Assumption (B) together with Lemma 9 from the Appendix implies

$$E|\gamma(t)|^2 \leqslant c/N^2. \tag{31}$$

Use (31) and Lemma 10 from the Appendix to derive

$$\begin{aligned}(E|X_N(t) - \widehat{X}_N^*(t)|^2)^{1/2} &= (E|X_N^{\mathrm{Mt}}(t) + L_N(t) - \widehat{X}_N^*(t)|^2)^{1/2} \\ &\geqslant (E|X_N^{\mathrm{Mt}}(t) - \gamma(t) + L_N(t_\ell) - \widehat{X}_N^*(t)|^2)^{1/2} - c/N \\ &\geqslant (E|X_N^{\mathrm{Mt}}(t) - \gamma(t) + L_N(t_\ell) - E(X_N^{\mathrm{Mt}}(t) - \gamma(t) + L_N(t_\ell) \mid \mathfrak{B}_N)|^2)^{1/2} - c/N \\ &= (E|\sigma(U_N^{\mathrm{Mt}}(t_\ell)) \cdot Z_N(t) + L_N(t_\ell) - E(L_N(t_\ell) \mid \mathfrak{B}_N)|^2)^{1/2} - c/N. \end{aligned} \tag{32}$$

Lemma 4 yields

$$L_N(t_\ell) - E(L_N(t_\ell) \mid \mathfrak{B}_N) = \sum_{r \leqslant t_j < t_\ell} \widehat{\vartheta}_N(t_j, t_\ell) \cdot (J_N(t_{j+1}) - E(J_N(t_{j+1}) \mid \mathfrak{B}_N)).$$

Thus, by Lemma 5,

$$\begin{aligned} &E(|\sigma(U_N^{\mathrm{Mt}}(t_\ell)) \cdot Z_N(t) + L_N(t_\ell) - E(L_N(t_\ell) \mid \mathfrak{B}_N)|^2 \mid \mathfrak{B}_N) \\ &\quad = \sigma^2(U_N^{\mathrm{Mt}}(t_\ell)) \cdot E(Z_N^2(t) \mid \mathfrak{B}_N) + \sum_{r \leqslant t_j < t_\ell} (\widehat{\vartheta}_N(t_j, t_\ell))^2 \cdot E((J_N(t_{j+1}) - E(J_N(t_{j+1}) \mid \mathfrak{B}_N))^2 \mid \mathfrak{B}_N) \\ &\quad = \sigma^2(U_N^{\mathrm{Mt}}(t_\ell)) \cdot (t_{\ell+1} - t) \cdot (t - t_\ell) \cdot N/T + T/(12N) \cdot \sum_{r \leqslant t_j < t_\ell} (\widehat{\vartheta}_N(t_j, t_\ell))^2 \cdot d_j, \end{aligned}$$

which implies

$$\int_{t_\ell}^{t_{\ell+1}} E|\sigma(U_N^{\mathrm{Mt}}(t_\ell)) \cdot Z_N(t) + L_N(t_\ell) - E(L_N(t_\ell) \,|\, \mathfrak{B}_N)|^2 \,\mathrm{d}t$$

$$= T^2/(6N^2) \cdot E(\sigma^2(U_N^{\mathrm{Mt}}(t_\ell))) + T^2/(12N^2) \cdot E\left(\sum_{r \leqslant t_j < t_\ell} (\widehat{\vartheta}_N(t_j, t_\ell))^2 \cdot d_j\right). \tag{33}$$

Due to (B), (2), and Lemma 9 and Proposition 1 from the Appendix,

$$E|\sigma^2(U_N^{\mathrm{Mt}}(t_\ell)) - \sigma^2(U(t_\ell))| \leqslant c/N^{1/2}. \tag{34}$$

Moreover, employing Lemma 3 it is easy to see that

$$E|((\widehat{\vartheta}_N(t_j, t_\ell))^2 - (\vartheta(t_j, t_\ell))^2) \cdot d_j| \leqslant c/N^{3/2}. \tag{35}$$

Use (33) and observe (34) as well as (35) to derive

$$\sum_{\ell=0}^{N-1} \int_{t_\ell}^{t_{\ell+1}} E|\sigma(U_N^{\mathrm{Mt}}(t_\ell)) \cdot Z_N(t) + L_N(t_\ell) - E(L_N(t_\ell) \,|\, \mathfrak{B}_N)|^2 \,\mathrm{d}t$$

$$\geqslant \frac{T^2}{6N^2} \cdot \sum_{\ell=0}^{N-1} E(\sigma^2(U(t_\ell))) + \frac{T^2}{12N^2} \cdot E\left(\sum_{r < t_\ell < T} \sum_{r \leqslant t_j < t_\ell} \vartheta^2(t_j, t_\ell) \cdot d_j\right) - c/N^{3/2}$$

$$= T/N \cdot E(A_N + B_N) - c/N^{3/2}. \tag{36}$$

Finally, combine (30) with (32) and (36) to complete the proof. □

Next, we turn to the error of $\widehat{X}_N^*$ at the final time point T. Define

$$C_N = \frac{1}{12} \cdot \sum_{\ell=0}^{N-1} \vartheta^2(t_\ell, T) \cdot d_\ell.$$

Lemma 7.

$$e(\widehat{X}_N^*, T) \geqslant (T/N)^{1/2} \cdot (E(C_N))^{1/2} - c \cdot N^{-3/4}.$$

Proof. Employing (22) we have

$$\begin{aligned} e(\widehat{X}_N^*, T) &\geqslant (E|X_N(T) - \widehat{X}_N^*(T)|^2)^{1/2} - c/N \\ &\geqslant (E|X_N(T) - E(X_N(T) \,|\, \mathfrak{B}_N)|^2)^{1/2} - c/N. \end{aligned} \tag{37}$$

By definition of X_N and Lemma 4,

$$\begin{aligned} X_N(T) - E(X_N(T) \,|\, \mathfrak{B}_N) &= L_N(T) - E(L_N(T) \,|\, \mathfrak{B}_N) \\ &= \sum_{r \leqslant t_\ell < T} \widehat{\vartheta}_N(t_\ell, T) \cdot (J_N(t_{\ell+1}) - E(J_N(t_{\ell+1}) \,|\, \mathfrak{B}_N)). \end{aligned}$$

Use Lemma 5 as well as (35) to obtain

$$\begin{aligned} E|X_N(T) - E(X_N(T) \,|\, \mathfrak{B}_N)|^2 &= \frac{T}{12N} \cdot E\left(\sum_{r \leqslant t_\ell < T} (\widehat{\vartheta}_N(t_\ell, T))^2 \cdot d_\ell\right) \\ &\geqslant \frac{T}{12N} \cdot E\left(\sum_{r \leqslant t_\ell < T} \vartheta^2(t_\ell, T) \cdot d_\ell\right) - c/N^{3/2}. \end{aligned} \tag{38}$$

Combining (37) with (38) finishes the proof. □

It remains to analyze the asymptotic behavior of A_N, B_N and C_N.

Lemma 8.

(i)

$$\lim_{N\to\infty} E(A_N) = \frac{1}{6} \cdot \int_0^T E(\sigma^2(U(t))\,\mathrm{d}t,$$

(ii)

$$\lim_{N\to\infty} E(B_N) = \frac{1}{4} \cdot \int_r^T \int_t^T E(\vartheta^2(t,s))\,\mathrm{d}s\,\mathrm{d}t,$$

(iii)

$$\lim_{N\to\infty} E(C_N) = \frac{1}{4} \cdot \int_r^T E(\vartheta^2(t,T))\,\mathrm{d}t.$$

Proof. Note that (B) together with (2) and (3) implies the continuity of the function $t \mapsto E(\sigma^2(U(t))$, which immediately yields (i).

In order to prove (ii) fix $M \in \mathbb{N}$, put

$$s_i = r + i \cdot (T-r)/M, \quad i = 0, 1, \ldots, M,$$

and define

$$B_{M,N} = \frac{T-r}{12M} \cdot \sum_{i=0}^{M-1} \sum_{k=i+1}^{M} \vartheta^2(s_i, s_k) \cdot \sum_{s_i \leqslant t_\ell < s_{i+1}} d_\ell,$$

for $N \geqslant M$. Observing (12) and (13) it is straightforward to check that

$$|E(B_N - B_{M,N})| \leqslant c/M^{1/2}. \tag{39}$$

Fix i and k. With the properties of the quadratic variation of W we get

$$\lim_{N\to\infty} E\left(\sum_{s_i \leqslant t_\ell < s_{i+1}} d_\ell - 3(T-r)/M\right)^2 = 0.$$

Using again (12) we conclude that

$$\lim_{N\to\infty} E\left|\vartheta^2(s_i, s_k) \cdot \left(\sum_{s_i \leqslant t_\ell < s_{i+1}} d_\ell - 3(T-r)/M\right)\right| = 0. \tag{40}$$

Combine (39) with (40) to obtain

$$\lim_{N\to\infty} E(B_N) = \lim_{M\to\infty} \frac{(T-r)^2}{4M^2} \cdot \sum_{i=0}^{M-1} \sum_{k=i+1}^{M} E(\vartheta^2(s_i, s_k)).$$

Finally, observe that (12) together with (13) implies the continuity of the mapping $(t,s) \mapsto E(\vartheta^2(t,s))$, $r \leqslant t \leqslant s \leqslant T$, which finishes the proof of (ii). The proof of (iii) may be handled in much the same way. □

Clearly, the lower bound in Theorem 1 follows from Lemma 7 and Lemma 8(iii). Similarly, Lemma 6 together with Lemma 8(i) and (ii) implies the lower bound in Theorem 2.

7.4. Proof of the upper bounds in Theorems 1 and 2

Consider the discretization (4) and define a corresponding scheme $\overline{X}_N$ by

$$\overline{X}_N(t_\ell) = X_N^{\mathrm{Mt}}(t_\ell) + \overline{L}_N(t_\ell), \quad \ell = 0, 1, \ldots, N,$$

where the process $\overline{L}_N$ is given by

$$\overline{L}_N(t_\ell) = 0, \quad -r \leqslant t \leqslant r,$$

and

$$\begin{aligned}\overline{L}_N(t) = \overline{L}_N(t_\ell) &+ a^{(0,1,0)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot \overline{L}_N(t_\ell) \cdot (t - t_\ell)\\ &+ a^{(0,0,1)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot \overline{L}_N(t_\ell - r) \cdot (t - t_\ell)\\ &+ \sigma^{(0,1,0)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot \overline{L}_N(t_\ell) \cdot (W(t) - W(t_\ell))\\ &+ \sigma^{(0,0,1)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot \overline{L}_N(t_\ell - r) \cdot (W(t) - W(t_\ell))\\ &+ \sigma(U_N^{\mathrm{Mt}}(t_\ell - r)) \cdot \sigma^{(0,0,1)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot E(J_N(t) \,|\, \mathfrak{B}_N),\end{aligned}$$

for $t \in]t_\ell, t_{\ell+1}]$. By piecewise linear interpolation

$$\overline{X}_N(t) = \frac{t_{\ell+1} - t}{t_{\ell+1} - t_\ell} \cdot \overline{X}_N(t_\ell) + \frac{t - t_\ell}{t_{\ell+1} - t_\ell} \cdot \overline{X}_N(t_{\ell+1}), \quad t \in [t_\ell, t_{\ell+1}],$$

we obtain a global approximation on the time interval $[0, T]$.

Note that, similar to the process L_N the process $\overline{L}_N$ can be rewritten in terms of the approximation $\widehat{\vartheta}_N$, i.e.,

$$\overline{L}_N(t_\ell) = \sum_{r \leqslant t_j < t_\ell} \widehat{\vartheta}_K(t_j, t_\ell) \cdot E(J_N(t_{j+1}) \,|\, \mathfrak{B}_N),$$

see Lemma 4, which in particular yields

$$\overline{L}_N(t_\ell) = E(L_N(t_\ell) \,|\, \mathfrak{B}_N). \tag{41}$$

Due to Lemma 13 from the Appendix we have

$$\sup_{0 \leqslant t \leqslant T} E|\widehat{X}_N(t) - \overline{X}_N(t)|^2 \leqslant c/N^2. \tag{42}$$

Combine (42) with (22) to conclude that

$$(E|X(t) - \widehat{X}_N(t)|^2)^{1/2} \leqslant (E|X_N(t) - \overline{X}_N(t)|^2)^{1/2} + c/N, \tag{43}$$

for every $t \in [0, T]$. Fix ℓ, let $t \in]t_\ell, t_{\ell+1}]$ and recall the definition of $\gamma(t)$ in the proof of Lemma 6. By definition,

$$\begin{aligned}X_N(t) - \overline{X}_N(t) = \sigma(U_N^{\mathrm{Mt}}(t_\ell)) \cdot Z_N(t) + \gamma(t) - (t - t_\ell) \cdot N \cdot \gamma(t_{\ell+1})\\ + L_N(t) - \overline{L}_N(t_\ell) - (t - t_\ell) \cdot N \cdot (\overline{L}_N(t_{\ell+1}) - \overline{L}_N(t_\ell)).\end{aligned} \tag{44}$$

Use (31) as well as Lemma 10(ii) from the Appendix to derive

$$\begin{aligned}&(E|X(t) - \widehat{X}_N(t)|^2)^{1/2}\\ &\quad \leqslant (E|\sigma(U_N^{\mathrm{Mt}}(t_\ell)) \cdot Z_N(t) + L_N(t_\ell) - E(L_N(t_\ell) \,|\, \mathfrak{B}_N)|^2)^{1/2} + c/N\end{aligned}$$

from (43) and (44). Next, observe (34) as well as (35), and proceed similar to the proof of Lemma 6 in Section 7.3 to obtain

$$e(\widehat{X}_N) \leqslant (T/N)^{1/2} \cdot (E(A_N + B_N))^{1/2} + c \cdot N^{-3/4}.$$

Now, the upper bound in Theorem 2 follows from Lemma 8(i) and (ii).

We turn to the error of $\widehat{X}_N$ at the final time point *T*. Use (43) and (44) to get

$$(E|X(T) - \widehat{X}_N(T)|^2)^{1/2} \leqslant (E|L_N(T) - E(L_N(T) \,|\, \mathfrak{B}_N)|^2)^{1/2} + c/N.$$

Proceed similar to the proof of Lemma 7 to conclude that

$$e(\widehat{X}_N, T) \leqslant (T/N)^{1/2} \cdot (E(C_N))^{1/2} + c \cdot N^{-3/4}. \tag{45}$$

Finally, observe that $\mathscr{C} = 0$ implies $E(C_N) = 0$ and apply Lemma 8(iii) in the case $\mathscr{C} > 0$ to obtain the upper bound in Theorem 1.

7.5. Proof of Theorem 3

Let $N \in \mathbb{N}$, consider an arbitrary discretization

$$0 = s_0 < s_1 < \cdots < s_N = T$$

of the time interval $[0, T]$ and let

$$Y_N = E(X \,|\, W(s_1), \ldots, W(s_N))$$

denote the corresponding optimal approximation of *X*. Put

$$M_N = T/r \cdot N$$

and

$$t_\ell = \ell \cdot T/M_N, \quad \ell = 0, 1, \ldots, M_N.$$

Next, define a discretization

$$0 = \tau_0 < \tau_1 < \cdots < \tau_{K_N} = T$$

by

$$\{\tau_1, \ldots, \tau_{K_N}\} = \{t_1, \ldots, t_{M_N}\} \cup (\{s_i + j \cdot r : \ j = -T/r, \ldots, T/r, \ i = 1, \ldots, N\} \cap [0, T]).$$

Note that this discretization satisfies assumption (23) of Lemma 5 and

$$K_N \leqslant 3T/r \cdot N. \tag{46}$$

Let $\mathfrak{C}_{K_N}$ denote the σ-algebra that is generated by $W(\tau_1), \ldots, W(\tau_{K_N})$ and put $Z_{K_N} = W - E(W \,|\, \mathfrak{C}_{K_N})$. Due to (22) we have

$$\begin{aligned}(E|X(t) - Y_N(t)|^2)^{1/2} &\geqslant (E|X_{M_N}(t) - Y_N(t)|^2)^{1/2} - (E|X(t) - X_{M_N}(t)|^2)^{1/2} \\ &\geqslant (E|X_{M_N}(t) - E(X_{M_N}(t) \,|\, \mathfrak{C}_{K_N})|^2)^{1/2} - c/N,\end{aligned} \tag{47}$$

for every $t \in [0, T]$.

Fix $\ell \in \{0, 1, \ldots, M_N - 1\}$ and let $t \in]t_\ell, t_{\ell+1}]$. Similar to the proof of Lemma 6 we obtain

$$\begin{aligned}&(E|X_{M_N}(t) - E(X_{M_N}(t) \,|\, \mathfrak{C}_{K_N})|^2)^{1/2} \\ &\quad \geqslant (E|\sigma(U_{M_N}^{\mathrm{Mt}}(t_\ell) \cdot Z_{K_N}(t) + L_{M_N}(t_\ell) - E(L_{M_N}(t_\ell) \,|\, \mathfrak{C}_{K_N})|^2)^{1/2} - c/N.\end{aligned} \tag{48}$$

Employing Lemma 4 as well as Lemma 5 we get

$$
\begin{aligned}
&E(|\sigma(U^{\mathrm{Mt}}_{M_N}(t_\ell)\cdot Z_{K_N}(t)+L_{M_N}(t_\ell)-E(L_{M_N}(t_\ell)\,|\,\mathfrak{C}_{K_N})|^2\,|\,\mathfrak{C}_{K_N}))\\
&\quad=\sigma^2(U^{\mathrm{Mt}}_{M_N}(t_\ell)\cdot E(Z^2_{K_N}(t)\,|\,\mathfrak{C}_{K_N})+\sum_{r\leqslant t_j<t_\ell}(\widehat{\vartheta}_{M_n}(t_j,t_\ell))^2\cdot E((J_{M_N}(t_{j+1})\\
&\qquad-E(J_{M_N}(t_{j+1})\,|\,\mathfrak{C}_{K_N}))^2\,|\,\mathfrak{C}_{K_N}).
\end{aligned}
$$

Put

$$
n_\ell=\#\{\tau_k:\ t_\ell\leqslant\tau_k<t_{\ell+1}\}.
$$

By the Hölder inequality,

$$
\int_{t_\ell}^{t_{\ell+1}}E(Z^2_{K_N}(t)\,|\,\mathfrak{C}_{K_N})\,\mathrm{d}t=\frac{1}{6}\sum_{t_\ell\leqslant\tau_k<t_{\ell+1}}(\tau_{k+1}-\tau_k)^2\geqslant T^2/(6M_N^2\cdot n_\ell).
$$

Furthermore, by Lemma 5,

$$
E((J_{M_N}(t_{j+1})-E(J_{M_N}(t_{j+1})\,|\,\mathfrak{C}_{K_N}))^2\,|\,\mathfrak{C}_{K_N})\geqslant\frac{1}{12}\sum_{t_j\leqslant\tau_k<t_{j+1}}(\tau_{k+1}-\tau_k)^2\geqslant T^2/(12M_N^2\cdot n_j).
$$

Consequently, observing (34) and Lemma 3,

$$
\begin{aligned}
&\int_{t_\ell}^{t_{\ell+1}}E(|\sigma(U^{\mathrm{Mt}}_{M_N}(t_\ell)\cdot Z_{K_N}(t)+L_{M_N}(t_\ell)-E(L_{M_N}(t_\ell)\,|\,\mathfrak{C}_{K_N})|^2\,\mathrm{d}t\\
&\quad\geqslant T^2/(6M_N^2\cdot n_\ell)\cdot E(\sigma^2(U^{\mathrm{Mt}}_{M_N}(t_\ell)))+T^3/(12M_N^3)\cdot\sum_{r\leqslant t_j<t_\ell}E((\widehat{\vartheta}_{M_n}(t_j,t_\ell))^2)\cdot 1/n_j\\
&\quad\geqslant T^2/(6M_N^2\cdot n_\ell)\cdot E(\sigma^2(U(t_\ell)))+T^3/(12M_N^3)\cdot\sum_{j=0}^{\ell-1}E(\vartheta^2(t_j,t_\ell))\cdot 1/n_j-c/M_N^{5/2}.
\end{aligned}
$$

Summing up over all subintervals $[t_\ell, t_{\ell+1}]$ and employing the Hölder inequality again we derive

$$
\begin{aligned}
&\sum_{\ell=0}^{M_N-1}\int_{t_\ell}^{t_{\ell+1}}E(|\sigma(U^{\mathrm{Mt}}_{M_N}(t_\ell))\cdot Z_{K_N}(t)+L_{M_N}(t_\ell)-E(L_{M_N}(t_\ell)\,|\,\mathfrak{C}_{K_N})|^2\,\mathrm{d}t\\
&\quad\geqslant T^2/(6M_N^2)\cdot\sum_{\ell=0}^{M_N-1}E(\sigma^2(U(t_\ell)))\cdot 1/n_\ell\\
&\qquad+T^3/(12M_N^3)\cdot\sum_{j=0}^{M_N-2}\sum_{\ell=j+1}^{M_N-1}E(\vartheta^2(t_j,t_\ell))\cdot 1/n_j-c/M_N^{3/2}\\
&\quad\geqslant T^2/(6M_N^2)\cdot 1/K_N\cdot\left(\sum_{\ell=0}^{M_N-1}(E(\sigma^2(U(t_\ell))))^{1/2}\right)^2\\
&\qquad+T^3/(12M_N^3)\cdot 1/K_N\cdot\left(\sum_{j=0}^{M_N-2}\left(\sum_{\ell=j+1}^{M_N-1}E(\vartheta^2(t_j,t_\ell))\right)^{1/2}\right)^2-c/M_N^{3/2}.
\end{aligned}
\tag{49}
$$

Combine (47) with (48) and (49), and observe (46) to conclude that

$$\begin{aligned}\liminf_{N\to\infty} N\cdot e^2(Y_N) &\geqslant \frac{r}{18T}\cdot\liminf_{N\to\infty}\left(\frac{T}{M_N}\cdot\sum_{\ell=0}^{M_N-1}(E(\sigma^2(U(t_\ell))))^{1/2}\right)^2\\ &\quad+\frac{r}{36T}\cdot\liminf_{N\to\infty}\left(\frac{T}{M_N}\cdot\sum_{j=0}^{M_N-2}\left(\frac{T}{M_N}\cdot\sum_{\ell=j+1}^{M_N-1}E(\vartheta^2(t_j,t_\ell))\right)^{1/2}\right)^2\\ &=\frac{r}{18T}\cdot\left(\int_0^T(E(\sigma^2(U(t))))^{1/2}\,\mathrm{d}t\right)^2\\ &\quad+\frac{r}{36T}\cdot\left(\int_r^T\left(\int_t^T E(\vartheta^2(t,s))\,\mathrm{d}s\right)^{1/2}\mathrm{d}t\right)^2.\end{aligned}$$

Due to Assumption (C) the last sum above is positive, which finishes the proof of the lower bound for the minimal mean square L_2-error $e(N)$ in Theorem 3. Clearly, the corresponding matching upper bound is a consequence of Theorem 2.

In order to prove the part on $e(N,T)$ in Theorem 3 use (47), (48) and Lemma 10 from the Appendix to obtain

$$(E|X(T)-Y_N(T)|^2)^{1/2}\geqslant(E|L_{M_N}(T)-E(L_{M_N}(T)\,|\,\mathfrak{C}_{K_N})|^2)^{1/2}-c/N. \tag{50}$$

Next, proceed as above to derive

$$\begin{aligned}E|L_{M_N}(T)-E(L_{M_N}(T)\,|\,\mathfrak{C}_{K_N})|^2 &\geqslant T^2/(12M_N^2)\cdot\sum_{j=0}^{M_N-1}E(\vartheta^2(t_j,T))\cdot 1/n_j - c/M_N^{3/2}\\ &\geqslant T^2/(12M_N^2)\cdot 1/K_N\cdot\left(\sum_{j=0}^{M_N-1}(E(\vartheta^2(t_j,T)))^{1/2}\right)^2-c/M_N^{3/2}.\end{aligned} \tag{51}$$

Finally, combine (50) with (51) and observe (46) to conclude that

$$\begin{aligned}\liminf_{N\to\infty} N\cdot e^2(Y_N,T) &\geqslant \frac{r}{36T}\cdot\liminf_{N\to\infty}\left(\frac{T}{M_N}\cdot\sum_{j=0}^{M_N-1}(E(\vartheta^2(t_j,T)))^{1/2}\right)^2\\ &=\frac{r}{36T}\cdot\left(\int_r^T(E(\vartheta^2(t,T)))^{1/2}\,\mathrm{d}t\right)^2.\end{aligned}$$

Clearly, the last integral is positive if and only if the asymptotic constant $\mathscr{C}$ is positive, which completes the proof of the lower bound for the minimal square error at the final time point $e(N,T)$ in Theorem 3. The corresponding matching upper bound is a consequence of the first statement in Theorem 1.

Acknowledgements

We thank two anonymous referees for a number of constructive suggestions that helped to improve the presentation of the paper.

Appendix

The main goal of this section is to establish the estimates (22) and (42) for the auxiliary processes X_N and $\overline{X}_N$, respectively, from Section 7.

Throughout, we consider discretization (4) from Section 3. Recall the definition of the corresponding time-continuous truncated Milstein scheme X_N^{Mt} from Section 7. We begin with error bounds for X_N^{Mt} and the time-continuous versions X_N^{E} and X_N^{M} of the Euler scheme and the Milstein scheme, respectively. The latter two processes are defined by

$$X_N^{\mathrm{E}}(t) = X_N^{\mathrm{M}}(t) = \eta(t), \quad t \leqslant 0,$$

and

$$X_N^{\mathrm{E}}(t) = X_N^{\mathrm{E}}(t_\ell) + a(t_\ell, X_N^{\mathrm{E}}(t_\ell), X_N^{\mathrm{E}}(t_\ell - r)) \cdot (t - t_\ell) + \sigma(t_\ell, X_N^{\mathrm{E}}(t_\ell), X_N^{\mathrm{E}}(t_\ell - r)) \cdot (W(t) - W(t_\ell)),$$

$$\begin{aligned} X_N^{\mathrm{M}}(t) = {} & X_N^{\mathrm{M}}(t_\ell) + a(t_\ell, X_N^{\mathrm{M}}(t_\ell), X_N^{\mathrm{M}}(t_\ell - r)) \cdot (t - t_\ell) + \sigma(t_\ell, X_N^{\mathrm{M}}(t_\ell), X_N^{\mathrm{M}}(t_\ell - r)) \cdot (W(t) - W(t_\ell)) \\ & + \tfrac{1}{2} \cdot (\sigma\sigma^{(0,1,0)})(t_\ell, X_N^{\mathrm{M}}(t_\ell), X_N^{\mathrm{M}}(t_\ell - r)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell)) \\ & + \sigma(t_\ell - r, X_N^{\mathrm{M}}(t_\ell - r), X_N^{\mathrm{M}}(t_\ell - 2r)) \cdot \sigma^{(0,0,1)}(t_\ell, X_N^{\mathrm{M}}(t_\ell), X_N^{\mathrm{M}}(t_\ell - r)) \cdot J_N(t), \end{aligned}$$

for $t \in]t_\ell, t_{\ell+1}]$.

To shorten notation we define

$$U_N^{\mathrm{M}}(t) = (t, X_N^{\mathrm{M}}(t), X_N^{\mathrm{M}}(t - r)), \quad U_N^{\mathrm{Mt}}(t) = (t, X_N^{\mathrm{Mt}}(t), X_N^{\mathrm{Mt}}(t - r)).$$

Moreover, we put $\|Y\|_q = (E|Y|^q)^{1/q}$ for a random variable Y and $q \geqslant 1$. Finally, we use c to denote unspecified positive constants that only depend on a, σ, η and q.

Lemma 9. *The processes X_N^{E}, X_N^{Mt} and X_N^{M} satisfy*

(i)

$$\sup_{t \in [0,T]} (\|X_N^{\mathrm{M}}(t)\|_q + \|X_N^{\mathrm{Mt}}(t)\|_q + \|X_N^{\mathrm{E}}(t)\|_q) \leqslant c,$$

(ii)

$$\sup_{t \in [0,T]} (\|X_N^{\mathrm{Mt}}(t) - X_N^{\mathrm{E}}(t)\|_q + \|X_N^{\mathrm{M}}(t) - X_N^{\mathrm{Mt}}(t)\|_q) \leqslant c \cdot N^{-1/2},$$

for every $q \geqslant 1$.

Proof. These estimates are obtained by using standard techniques. See [3] for details. □

Proposition 1. *The processes X_N^{E}, X_N^{Mt} and X_N^{M} satisfy*

(i)

$$\sup_{t \in [0,T]} \|X(t) - X_N^{\mathrm{E}}(t)\|_q \leqslant c \cdot N^{-1/2},$$

(ii)

$$\sup_{t \in [0,T]} \|X(t) - X_N^{\mathrm{Mt}}(t)\|_q \leqslant c \cdot N^{-1/2},$$

for every $q \geqslant 1$, and

(iii)

$$\sup_{t \in [0,T]} \|X(t) - X_N^{\mathrm{M}}(t)\|_2 \leqslant c \cdot N^{-1}.$$

Parts (i) and (iii) are consequences of more general results in [5, Theorem 3.1] and [6, Theorem 5.2], respectively. Part (ii) immediately follows from (i) and Lemma 9(ii).

Next, we analyze the processes L_N and $\overline{L}_N$, see Section 7.

Lemma 10. *The processes* L_N *and* $\overline{L}_N$ *satisfy*

$$\sup_{t\in[0,T]} (\|L_N(t)\|_q + \|\overline{L}_N(t)\|_q) \leqslant c \cdot N^{-1/2}$$

and

$$\sup_{t\in[t_\ell,t_{\ell+1}]} (\|L_N(t) - L_N(t_\ell)\|_q + \|\overline{L}_N(t) - \overline{L}_N(t_\ell)\|_q) \leqslant c \cdot N^{-1}.$$

for every $q \geqslant 1$.

Proof. For convenience we restrict to the case $q = 2$ and we only consider the process L_N. The same arguments also apply to the process $\overline{L}_N$. Put

$$g(t) = \sup_{0\leqslant s\leqslant t} \|L_N(s)\|_2.$$

Let $t \in [t_\ell, t_{\ell+1}]$ with $t_\ell \geqslant r$. Write

$$L_N(t) = A(t) + B(t),$$

with

$$A(t) = L_N(t_\ell) + a^{(0,1,0)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot L_N(t_\ell) \cdot (t - t_\ell) + a^{(0,0,1)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot L_N(t_\ell - r) \cdot (t - t_\ell).$$

Clearly, by the boundedness of the spatial derivatives,

$$\|A(t)\|_2^2 \leqslant g^2(t_\ell) \cdot (1 + c \cdot (t - t_\ell)).$$

Moreover, using (B) and Lemma 9,

$$\begin{aligned}\|B(t)\|_2^2 &\leqslant c \cdot \|(L_N(t_\ell) + L_N(t_\ell - r)) \cdot (W(t) - W(t_\ell))\|_2^2 + c \cdot \|\sigma(U_N^{\mathrm{Mt}}(t_\ell - r)) \cdot J_N(t)\|_2^2 \\ &\leqslant c \cdot g^2(t_\ell) \cdot (t - t_\ell) + c \cdot (t - t_\ell)^2.\end{aligned}$$

Observe that $E(A(t) \cdot B(t)) = 0$. Summarizing we get

$$g^2(t) \leqslant g^2(t_\ell) \cdot (1 + c/N) + c/N^2.$$

The first estimate thus follows from Gronwall's Lemma.

Finally, by (i),

$$\begin{aligned}\|L_N(t) - L_N(t_\ell)\|_2^2 &\leqslant c \cdot \|L_N(t_\ell) + L_N(t_\ell - r)\|_2^2 \cdot (t - t_\ell)^2 + c \cdot \|B(t)\|_2^2 \\ &\leqslant c \cdot g^2(t_\ell) \cdot (t - t_\ell)^2 + c \cdot g^2(t_\ell) \cdot (t - t_\ell) + c \cdot (t - t_\ell)^2 \\ &\leqslant c/N^2,\end{aligned}$$

which finishes the proof. □

We proceed with a comparison of X_N^{M} and $X_N^{\mathrm{Mt}} + L_N$.

Lemma 11. *For every* $q \geqslant 1$,

$$\sup_{t\in[0,T]} \|X_N^{\mathrm{M}}(t) - X_N^{\mathrm{Mt}}(t) - L_N(t)\|_q \leqslant c \cdot N^{-1}.$$

Proof. As in the proof of Lemma 10 we assume $q = 2$. The case $q > 2$ may be handled in a similar way. Define

$$g(t) = \sup_{0\leqslant s\leqslant t} \|X_N^{\mathrm{M}}(s) - X_N^{\mathrm{Mt}}(s) - L_N(s)\|_2.$$

Let $t \in [t_\ell, t_{\ell+1}]$ and define the quantities

$$A = a(U_N^{\mathrm{M}}(t_\ell)) - a(U_N^{\mathrm{Mt}}(t_\ell)) - a^{(0,1,0)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot (X_N^{\mathrm{M}}(t_\ell) - X_N^{\mathrm{Mt}}(t_\ell)) \\ - a^{(0,0,1)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot (X_N^{\mathrm{M}}(t_\ell - r) - X_N^{\mathrm{Mt}}(t_\ell - r))$$

and

$$\Sigma = \sigma(U_N^{\mathrm{M}}(t_\ell)) - \sigma(U_N^{\mathrm{Mt}}(t_\ell)) - \sigma^{(0,1,0)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot (X_N^{\mathrm{M}}(t_\ell) - X_N^{\mathrm{Mt}}(t_\ell)) \\ - \sigma^{(0,0,1)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot (X_N^{\mathrm{M}}(t_\ell - r) - X_N^{\mathrm{Mt}}(t_\ell - r)).$$

Then

$$X_N^{\mathrm{M}}(t) - X_N^{\mathrm{Mt}}(t) - L_N(t) = Y + A \cdot (t - t_\ell) + B + C + D,$$

where

$$Y = X_N^{\mathrm{M}}(t_\ell) - X_N^{\mathrm{Mt}}(t_\ell) - L_N(t_\ell) + a^{(0,1,0)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot (X_N^{\mathrm{M}}(t_\ell) - X_N^{\mathrm{Mt}}(t_\ell) - L_N(t_\ell)) \cdot (t - t_\ell) \\ + a^{(0,0,1)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot (X_N^{\mathrm{M}}(t_\ell - r) - X_N^{\mathrm{Mt}}(t_\ell - r) - L_N(t_\ell - r)) \cdot (t - t_\ell),$$

$$B = \sigma^{(0,1,0)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot (X_N^{\mathrm{M}}(t_\ell) - X_N^{\mathrm{Mt}}(t_\ell) - L_N(t_\ell)) \cdot (W(t) - W(t_\ell)) \\ + \sigma^{(0,0,1)}(U_N^{\mathrm{Mt}}(t_\ell)) \cdot (X_N^{\mathrm{M}}(t_\ell - r) - X_N^{\mathrm{Mt}}(t_\ell - r) - L_N(t_\ell - r)) \cdot (W(t) - W(t_\ell)) \\ + \Sigma \cdot (W(t) - W(t_\ell)),$$

$$C = \tfrac{1}{2} \cdot ((\sigma \cdot \sigma^{(0,1,0)})(U_N^{\mathrm{M}}(t_\ell)) - (\sigma \cdot \sigma^{(0,1,0)})(U_N^{\mathrm{Mt}}(t_\ell))) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell))$$

and

$$D = (\sigma(U_N^{\mathrm{M}}(t_\ell - r)) \cdot \sigma^{(0,0,1)}(U_N^{\mathrm{M}}(t_\ell)) - \sigma(U_N^{\mathrm{Mt}}(t_\ell - r)) \cdot \sigma^{(0,0,1)}(U_N^{\mathrm{Mt}}(t_\ell))) \cdot J_N(t).$$

Clearly,

$$\|Y\|_2^2 \leqslant g^2(t_\ell) \cdot (1 + c \cdot (t - t_\ell)).$$

Furthermore, by Lemma 9,

$$\|C\|_2^2 + \|D\|_2^2 \leqslant c \cdot (\|X_N^{\mathrm{M}}(t_\ell) - X_N^{\mathrm{Mt}}(t_\ell)\|_4^2 + \|X_N^{\mathrm{M}}(t_\ell - r) - X_N^{\mathrm{Mt}}(t_\ell - r)\|_4^2) \cdot (t - t_\ell)^2 \\ \leqslant c/N \cdot (t - t_\ell)^2$$

as well as

$$\|A\|_2^2 + \|\Sigma\|_2^2 \leqslant c \cdot (\|X_N^{\mathrm{M}}(t_\ell) - X_N^{\mathrm{Mt}}(t_\ell)\|_4^4 + \|X_N^{\mathrm{M}}(t_\ell - r) - X_N^{\mathrm{Mt}}(t_\ell - r)\|_4^4) \leqslant c/N^2.$$

The latter inequality yields

$$\|B\|_2^2 \leqslant c \cdot (g^2(t_\ell) + \|\Sigma\|_2^2) \cdot (t - t_\ell) \leqslant c \cdot (g^2(t_\ell) + 1/N^2) \cdot (t - t_\ell).$$

Note that

$$E(Y \cdot B) = E(Y \cdot C) = E(Y \cdot D) = 0.$$

Hence

$$|E(Y \cdot (A \cdot (t - t_\ell) + B + C + D))| = |E(Y \cdot A)| \cdot (t - t_\ell) \leqslant c \cdot g(t_\ell) \cdot 1/N \cdot (t - t_\ell).$$

Summarizing we obtain

$$g^2(t) \leqslant g^2(t_\ell) + c \cdot (t - t_\ell) \cdot (g^2(t_\ell) + g(t_\ell)/N + 1/N^2) \\ \leqslant g^2(t_\ell) \cdot (1 + c/N) + c/N^3.$$

Now, use Gronwall's Lemma to complete the proof. □

As a consequence of Proposition 1 and Lemma 11 we obtain

Proposition 2. *The process X_N satisfies*

$$\sup_{t\in[0,T]} \|X(t) - X_N(t)\|_2 \leqslant c \cdot N^{-1}.$$

Finally, we consider a time-continuous version $\widetilde{X}_N$ of the approximation $\widehat{X}_N$ introduced in Section 3. Put

$$\widetilde{J}_N(t) = \tfrac{1}{2} \cdot (W(t-r) - W(t_\ell - r)) \cdot (W(t) - W(t_\ell)),$$

for $t \in]t_\ell, t_{\ell+1}]$ and define the process $\widetilde{X}_N$ by

$$\widetilde{X}_N(t) = \eta(t), \quad t \leqslant 0,$$

and

$$\begin{aligned}\widetilde{X}_N(t) = \widetilde{X}_N(t_\ell) &+ a(t_\ell, \widetilde{X}_N(t_\ell), \widetilde{X}_N(t_\ell - r)) \cdot (t - t_\ell) + \sigma(t_\ell, \widetilde{X}_N(t_\ell), \widetilde{X}_N(t_\ell - r)) \cdot (W(t) - W(t_\ell)) \\ &+ \tfrac{1}{2} \cdot (\sigma\sigma^{(0,1,0)})(t_\ell, \widetilde{X}_N(t_\ell), \widetilde{X}_N(t_\ell - r)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell)) \\ &+ \sigma(t_\ell - r, \widetilde{X}_N(t_\ell - r), \widetilde{X}_N(t_\ell - 2r)) \cdot \sigma^{(0,0,1)}(t_\ell, \widetilde{X}_N(t_\ell), \widetilde{X}_N(t_\ell - r)) \cdot \widetilde{J}_N(t),\end{aligned}$$

for $t \in]t_\ell, t_{\ell+1}]$. Thus $\widetilde{X}_N$ is defined as the time-continuous Milstein scheme X_N^{M} with J_N replaced by the approximation $\widetilde{J}_N$.

We compare $\widetilde{X}_N$ with the processes X_N^{Mt} and $X_N^{\mathrm{Mt}} + \overline{L}_N$.

Lemma 12. *The process $\widetilde{X}_N$ satisfies*

(i)

$$\sup_{t\in[0,T]} \|\widetilde{X}_N(t) - X_N^{\mathrm{Mt}}(t)\|_q \leqslant c \cdot N^{-1/2},$$

(ii)

$$\sup_{t\in[0,T]} \|\widetilde{X}_N(t) - X_N^{\mathrm{Mt}}(t) - \overline{L}_N(t)\|_q \leqslant c \cdot N^{-1}$$

for every $q \geqslant 1$.

These estimates may be derived in the same way as the corresponding estimates for the time-continuous Milstein scheme X_N^{M} in Lemma 9(ii) and Lemma 11.

Note that, at the discretization points t_ℓ the processes $\widetilde{X}_N$ and $X_N^{\mathrm{Mt}} + \overline{L}_N$ coincide with the approximations $\widehat{X}_N$ and $\overline{X}_N$, respectively. Therefore, Lemma 12 implies the following estimate of the difference $\widehat{X}_N - \overline{X}_N$.

Lemma 13. *The approximations $\widehat{X}_N$ and $\overline{X}_N$ satisfy*

$$\sup_{t\in[0,T]} \|\widehat{X}_N(t) - \overline{X}_N(t)\|_2 \leqslant c \cdot N^{-1}.$$

References

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

[2] A. Friedman, Stochastic Differential Equations and Applications, Academic Press, New York, 1975.

[3] N. Hofmann, T. Müller-Gronbach, An optimal algorithm for strong approximation of scalar SDDE's with constant delay based on equidistant Brownian evaluation, Fakultät für Mathematik, Otto-von-Guericke-Universität Magdeburg, 2004, preprint 04-22.

[4] N. Hofmann, T. Müller-Gronbach, K. Ritter, The optimal discretization of stochastic differential equations, J. Complexity 17 (2001) 117–153.

[5] Y. Hu, S.-E.A. Mohammed, F. Yan, Numerical solution of stochastic differential systems with memory, Department of Mathematics, Southern Illinois University, Carbondale, 2001, preprint.
[6] Y. Hu, S.-E.A. Mohammed, F. Yan, Discrete-time approximation of stochastic delay equations: the Milstein scheme, Ann. Probab. 32 (1A) (2004) 265–314.
[7] P. Kloeden, E. Platen, Numerical Solution of Stochastic Differential Equations, Springer, Berlin, 1995.
[8] U. Küchler, E. Platen, Strong discrete time approximation of stochastic differential equations with time delay, Math. Comput. Simulation 54 (2000) 189–205.
[9] M. Liu, W. Cao, Z. Fan, Convergence and stability of the semi-implicit Euler method for a linear stochastic differential delay equation, J. Comput. Appl. Math. 170 (2004) 255–268.
[10] X. Mao, Stochastic Differential Equations and Applications, Horwood Publishing, Chichester, 1997.
[11] G.N. Milstein, Numerical Integration of Stochastic Differential Equations, Kluwer Academic, Dordrecht, 1995.
[12] S.-E.A. Mohammed, Stochastic Functional Differential Equations, Research Notes in Mathematics, vol. 99, Pitman, London, 1984.
[13] S.-E.A. Mohammed, Stochastic differential systems with memory: theory, examples and applications, in: L. Decreusefond, J. Gjerde, B. Oksendal, A.S. Ustunel (Eds.), Stochastic Analysis and Related Topics VI. The Geilo Workshop, 1996, Progress in Probability, Birkhäuser, Basel, 1998.
[14] T. Müller-Gronbach, Strong Approximation of Systems of Stochastic Differential Equations, Habilitationsschrift, Technische Universität Darmstadt, 2002.
[15] T. Müller-Gronbach, Optimal pointwise approximation of SDE's based on Brownian motion at discrete points, Ann. Appl. Probab. 14 (4) (2004) 1605–1642.
[16] E. Platen, An introduction to numerical methods for stochastic differential equations, Acta Numer. 8 (1999) 195–244.
[17] H. Schurz, Numerical analysis of SDEs without tears, in: D. Kannan, V. Lakshmikantham (Eds.), Handbook of Stochastic Analysis and Applications, Marcel Dekker, Basel, 2002.
[18] D. Talay, Simulation of stochastic differential systems, in: P. Krée, W. Wedig (Eds.), Probabilistic Methods in Applied Physics, Lecture Notes in Physics, vol. 451, Springer, Berlin, 1995.
[19] M. Tudor, Approximation schemes for stochastic equations with hereditary argument, Stud. Cerc. Mat. 44 (1992) 73–85.
[20] C. Tudor, M. Tudor, On approximation of solutions for stochastic delay equations, Stud. Cerc. Mat. 39 (1987) 265–274.
[21] I. Karatzas, S.E. Shreve, Brownian Motion and Stochastic Calculus, Springer, Berlin, 1991.