BIT Numerical Mathematics (2023) 63:33
https://doi.org/10.1007/s10543-023-00969-9

BIT

Strong convergence of an adaptive time-stepping Milstein method for SDEs with monotone coefficients

Cónall Kelly[1] · Gabriel J. Lord[2] · Fandi Sun[3]

Received: 23 July 2021 / Accepted: 24 March 2023 / Published online: 22 May 2023
© The Author(s) 2023

Abstract
We introduce an explicit adaptive Milstein method for stochastic differential equations with no commutativity condition. The drift and diffusion are separately locally Lipschitz and together satisfy a monotone condition. This method relies on a class of path-bounded time-stepping strategies which work by reducing the stepsize as solutions approach the boundary of a sphere, invoking a backstop method in the event that the timestep becomes too small. We prove that such schemes are strongly L_2 convergent of order one. This order is inherited by an explicit adaptive Euler–Maruyama scheme in the additive noise case. Moreover we show that the probability of using the backstop method at any step can be made arbitrarily small. We compare our method to other fixed-step Milstein variants on a range of test problems.

Keywords Stochastic differential equations · Adaptive time-stepping · Milstein method · Non-globally Lipschitz coefficients · Strong convergence

Mathematics Subject Classification 60H10 · 60H35 · 65C30

Communicated by David Cohen.

✉ Fandi Sun
fs30@hw.ac.uk

Cónall Kelly
conall.kelly@ucc.ie

Gabriel J. Lord
gabriel.lord@ru.nl

1 School of Mathematical Sciences, University College Cork, Cork, Ireland

2 Mathematics, IMAPP, Radboud University, Nijmegen, The Netherlands

3 Maxwell Institute, Department of Mathematics, MACS, Heriot-Watt University, Edinburgh, UK

1 Introduction

We investigate the use of adaptive time-stepping strategies in the construction of a strongly convergent explicit Milstein-type numerical scheme for a d-dimensional stochastic differential equation (SDE) of Itô-type on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$,

$$X(t) = X(0) + \int_0^t f(X(r))dr + \sum_{i=1}^m \int_0^t g_i(X(r))dW_i(r), \tag{1.1}$$

for $t \in [0, T]$, $T \geq 0$ and $i = 1, \ldots, m \in \mathbb{N}$, where $W = [W_1, \cdots, W_m]^T$ is an m-dimensional Wiener process, the drift coefficient $f : \mathbb{R}^d \to \mathbb{R}^d$ and the diffusion coefficient $g : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ each satisfy a local Lipschitz condition along with a polynomial growth condition and, together, a monotone condition. Both are twice continuously differentiable; see Assumptions 2.1 and 2.2. Throughout, we take the initial vector $X(0) = X_0 \in \mathbb{R}^d$ to be deterministic.

It was pointed out in [34] that, because the Euler–Maruyama and Euler–Milstein methods coincide in the additive noise case, and as a consequence of the analysis in [15], an explicit Milstein scheme over a uniform mesh cannot converge in L_p to solutions of (1.1). We propose here an adaptive variant of the explicit Milstein method that achieves strong L_2 convergence of order one to solutions of (1.1). As an immediate consequence of this, in the case of additive noise an adaptive Euler–Maruyama method also has L_2 convergence of order one. To prove our convergence result it is essential to introduce a new variant of the admissible class of time-stepping strategies introduced in [17, 18], which we call path-bounded strategies.

Several variants on the fixed-step Milstein method have been proposed, see for example the tamed Milstein [20, 34], projected and split-step backward Milstein [1], truncated Milstein [10], implicit Milstein methods [13, 35] and a recent tamed stochastic Runge–Kutta (of order one) method of [8], all designed to converge strongly to solutions of SDEs with more general drift and diffusions, such as in (1.1). However, with few exceptions (see [1, 20]) explicit methods of this kind have only examined the case where the diffusion coefficients g_i satisfies a commutativity condition. We do not impose a commutativity restriction and hence must consider the associated Lévy areas (see Lemma 2.2).

A review of methods that adapt the timestep in order to control local error may be found in the introduction to [17]; we cite here [2, 7, 16, 21, 28, 31] and remark that our purpose is instead to handle the nonlinear response of the discrete system see also [5, 6] and discussion in [17, 18]. A common feature of the adaptivity is the use of both a minimum and maximum time step where the magnitude of the minimum step is controlled by a free parameter which requires some a-priori knowledge on the part of the user. The approach of [5, 6] was recently extended to McKean–Vlasov equations in [30] and include a Milstein approximation. In addition we note the fully adaptive Milstein method proposed in [14] for a scalar SDE with light constraints on the coefficients. There the authors stated that such a method was easy to implement but hard to analyse and as a result considered a different, but related method.

Our framework for adaptivity was introduced in [17] for an explicit Euler–Maruyama method, and has since been extended to SDE systems with monotone coefficients in [18] and to SPDE methods in [3]. These methods all use a backstop method when the chosen strategy attempts to select a stepsize below the minimum step. We demonstrate here, for a path-bounded strategy, that the probability of using the backstop can be made arbitrarily small by choosing an appropriately large ρ, and an appropriately small $h_{\max}$. This is consistent with observation, and with the intuitive notion that the use of the backstop should be rare in practice.

The structure of the article is as follows. Mathematical preliminaries are considered in Sect. 2, including precise specifications of the conditions imposed on each f and g_i, and the characterisation of an explicit Milstein method on an arbitrary mesh. The construction and result of the adaptive time-stepping strategy is outlined in Sect. 3, where we formulate the adaptive Milstein scheme with backstop which will be the subject of our main theorem. Both main results: on strong L_2 convergence and on the probability of using the backstop method, are stated in Sect. 4; we defer their proofs to Sect. 7. In Sect. 5 we compare the adaptive scheme numerically to other fixed step methods and illustrate both convergence and efficiency. The proof of Lemma 2.2 is in Appendix A.

2 Mathematical preliminaries

We consider the d-dimensional Itô-type SDE (1.1) and for the remainder of the article let $(\mathcal{F}_t)_{t\geq 0}$ be the natural filtration of W. For all $x \in \mathbb{R}^d$ and for all $\phi(x) \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^d)$, the Jacobian matrix of $\phi(x)$ is denoted $\mathbf{D}\phi(x) \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)$; the second derivative of $\phi(x)$ with respect to a vector x forms a 3-tensor and is denoted $\mathbf{D}^2\phi(x) \in \mathcal{L}(\mathbb{R}^{d\times d}, \mathbb{R}^d)$; and $[x]^2 := x \otimes x$ stands for the outer product of x and itself. Furthermore, let $\|\cdot\|$ denote the standard l^2 norm in $\mathbb{R}^d$, $\|\cdot\|_{\mathbf{F}(a\times b)}$ the Frobenious norm of the matrix in $\mathbb{R}^{a\times b}$; for simplicity we write $\|\cdot\|_{\mathbf{F}}$ as the Frobenious norm of the matrix in $\mathbb{R}^{d\times d}$. $\|\cdot\|_{\mathbf{T}_3}$ denotes the induced tensor norm (spectral norm) of the 3-tensor in $\mathbb{R}^{d\times d\times d}$ and it is defined as $\left\|\cdot\right\|_{\mathbf{T}_3} := \sup_{h_1,h_2\in\mathbb{R}^d, \|h_1\|,\|h_2\|\leq 1} \left\|\cdot(h_1\otimes h_2)\right\|$. For $a, b \in \mathbb{R}$, $a \vee b$ denotes $\max\{a, b\}$ and $a \wedge b$ denotes $\min\{a, b\}$. We frequently make use of the elementary inequality

$$2ab \leq a^2 + b^2, \quad a, b \in \mathbb{R}, \tag{2.1}$$

and of the following two standard extensions of Jensen's inequality (see [23, Corollary A.10]). For $f \in L^1$, if $p \geq 1$,

$$\left|\int_0^t f(s)ds\right|^p \leq t^{p-1}\int_0^t |f(s)|^p ds, \quad t \geq 0. \tag{2.2}$$

For $a_i \in \mathbb{R}$ and $p \geq 1$,

$$\left|\sum_{i=1}^n a_i\right|^p \leq n^{p-1}\sum_{i=1}^n |a_i|^p, \quad n \in \mathbb{N}\backslash\{0\}. \tag{2.3}$$

We now present our assumptions on f and g_i in (1.1).

Assumption 2.1 Let $f \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^d)$ and $g \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^{d\times m})$ with $g_i(x) = [g_{1,i}(x), \ldots, g_{d,i}(x)]^T \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^d)$. For each $\varkappa \geq 1$ there exist $L_\varkappa > 0$ such that

$$\left\| f(x) - f(y) \right\|^2 + \left\| g(x) - g(y) \right\|^2_{\mathbf{F}(d\times m)} \leq L_\varkappa \left\| x - y \right\|^2, \tag{2.4}$$

for $x, y \in \mathbb{R}^d$ with $\|x\| \vee \|y\| \leq \varkappa$, and there exists $c \geq 0$ such that for some $\eta \geq 2$

$$\left\langle x - y, f(x) - f(y) \right\rangle + \frac{\eta - 1}{2} \left\| g(x) - g(y) \right\|^2_{\mathbf{F}(d\times m)} \leq c \left\| x - y \right\|^2. \tag{2.5}$$

In addition, for some constants $c_{3,4,5,6}, q_1, q_2 \geq 0$; $i = 1, \ldots, m$, we have

$$\left\| \mathbf{D} f(x) \right\|_{\mathbf{F}} \leq c_3(1 + \|x\|^{q_1+1}), \qquad \left\| \mathbf{D} g_i(x) \right\|_{\mathbf{F}} \leq c_4(1 + \|x\|^{q_2+1}), \tag{2.6}$$

$$\left\| f(x) \right\| \leq c_5(1 + \|x\|^{q_1+2}), \qquad \left\| g(x) \right\|_{\mathbf{F}(d\times m)} \leq c_6(1 + \|x\|^{q_2+2}). \tag{2.7}$$

Furthermore, for some $c_{1,2} \geq 0$; $i = 1, \ldots, m$, we have

$$\left\| \mathbf{D}^2 f(x) \right\|_{\mathbf{T}_3} \leq c_1(1 + \|x\|^{q_1}), \quad \left\| \mathbf{D}^2 g_i(x) \right\|_{\mathbf{T}_3} \leq c_2(1 + \|x\|^{q_2}). \tag{2.8}$$

Under (2.4) and (2.5), the SDE (1.1) has a unique strong solution on any interval $[0, T]$, where $T < \infty$ on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, see [11], [25] and [33].

Assumption 2.2 Suppose that (2.5) in Assumption 2.1 holds with

$$\eta \geq 4q + 2q_2 + 10,$$

where $q := q_1 \vee q_2$, q_1 and q_2 are from (2.7) in Assumption 2.1.

We now give the following Lemma on moments of the solution.

Lemma 2.1 *[26, Lem. 4.2] Let f and g satisfy (2.4) and (2.5), and suppose that Assumption 2.2 holds. If g further satisfies (2.7), then there is a constant $C_X > 0$ such that the solution of (1.1) satisfies*

$$\mathbb{E}\left[\sup_{s\in[0,T]} \|X(s)\|^{\eta - 2q_2 - 2} \right] \leq C_X. \tag{2.9}$$

Next we present the fixed-step Milstein method (see [19, Sec. 10.3]) that is the basis of the adaptive method presented in this article.

Definition 2.1 (Milstein method) For $n \in \mathbb{N}$, $s \in [t_n, t_{n+1}]$ and given $Y(t_n)$, the fixed-step Milstein scheme for (1.1), interpolated over the interval $[t_n, t_{n+1}]$, is given by

$$Y(s) := Y(t_n) + f\big(Y(t_n)\big)|s - t_n| + \sum_{i=1}^{m} g_i\big(Y(t_n)\big) I_i^{t_n,s} + \sum_{i,j=1}^{m} \mathbf{D}g_i\big(Y(t_n)\big) g_j\big(Y(t_n)\big) I_{j,i}^{t_n,s}, \tag{2.10}$$

where following [1, 34], the stochastic integral and the iterated stochastic integral are defined as

$$I_i^{t_n,s} := \int_{t_n}^{s} dW_i(r), \qquad I_{j,i}^{t_n,s} := \int_{t_n}^{s}\int_{t_n}^{r} dW_j(p)dW_i(r). \tag{2.11}$$

Expanding the last term in (2.10) we have that

$$\begin{aligned}
&\sum_{i,j=1}^{m} \mathbf{D}g_i\big(Y(t_n)\big) g_j\big(Y(t_n)\big) I_{j,i}^{t_n,s} \\
&= \frac{1}{2}\sum_{i=1}^{m} \mathbf{D}g_i\big(Y(t_n)\big) g_i\big(Y(t_n)\big)\left(\big(I_i^{t_n,s}\big)^2 - |s - t_n|\right) \\
&\quad + \frac{1}{2}\sum_{\substack{i,j=1\\ i<j}}^{m} \Big(\mathbf{D}g_i\big(Y(t_n)\big) g_j\big(Y(t_n)\big) + \mathbf{D}g_j\big(Y(t_n)\big) g_i\big(Y(t_n)\big)\Big) I_i^{t_n,s} I_j^{t_n,s} \\
&\quad + \sum_{\substack{i,j=1\\ i<j}}^{m} \Big(\mathbf{D}g_i\big(Y(t_n)\big) g_j\big(Y(t_n)\big) - \mathbf{D}g_j\big(Y(t_n)\big) g_i\big(Y(t_n)\big)\Big) A_{ij}^{t_n,s},
\end{aligned} \tag{2.12}$$

where the term $A_{ij}^{t_n,s}$ is the Lévy area (see for example [22, Eq. (1.2.2)]) defined by

$$A_{ij}^{t_n,s} := \frac{1}{2}\left(I_{i,j}^{t_n,s} - I_{j,i}^{t_n,s}\right), \tag{2.13}$$

and we have used the relations $I_{i,i}^{t_n,s} = \frac{1}{2}\big((I_i^{t_n,s})^2 - |t-s|\big)$ and $I_{i,j}^{t_n,s} + I_{j,i}^{t_n,s} = I_i^{t_n,s} I_j^{t_n,s}$. As mentioned in the introduction many authors assume the following commutativity condition: suppose that $\mathbf{D}g_i(y)g_j(y) = \mathbf{D}g_j(y)g_i(y)$ for all $i, j = 1, \ldots, m$ and $y \in \mathbb{R}^d$. When this holds, the last term in (2.12) vanishes, avoiding the need for any analysis of $A_{ij}^{t_n,s}$ defined in (2.13). We do not impose such a condition in this paper, and therefore make use of the following conditional moment bounds on the Lévy areas.

Lemma 2.2 *(`Lévy Area`) For all $i, j = 1, \ldots, m$, $0 \le t_n \le s < T$ and for a pair of Wiener process $(W_i(r), W_j(r))^T$ where $r \in [t_n, s]$ and the Lévy area $A_{ij}^{t_n,s}$ defined*

in (2.13)*, there exists a finite constant* C_{LA} *whose explicit form is in* (A.1) *such that for* $k \geq 1$

$$\mathbb{E}\left[\left| A_{ij}^{t_n,s} \right|^k \middle| \mathcal{F}_{t_n} \right] \leq C_{LA}(k) \; |s - t_n|^k, \quad a.s. \tag{2.14}$$

For proof see Appendix A.

3 Adaptive time-stepping strategies

To deal with the extra terms that arise from Milstein over Euler–Maruyama type discretisations, we introduce a new class of time-stepping strategies in Definition 3.5. Let $\{h_{n+1}\}_{n\in\mathbb{N}}$ be a sequence of strictly positive random timesteps with corresponding random times $\{t_n := \sum_{i=1}^{n} h_i\}_{n\in\mathbb{N}\setminus\{0\}}$, where $t_0 = 0$.

Definition 3.1 Suppose that each member of $\{t_n\}_{n\in\mathbb{N}\setminus\{0\}}$ is an $\mathcal{F}_t$-stopping time: i.e. $\{t_n \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$, where $(\mathcal{F}_t)_{t\geq 0}$ is the natural filtration of W. If τ is any $(\mathcal{F}_t)$-stopping time, then (see [27, p. 14])

$$\mathcal{F}_\tau := \left\{ A \in \mathcal{F} : \; A \cap \{\tau \leq t\} \in \mathcal{F}_t, \;\; \text{for all } \; t \geq 0 \right\}. \tag{3.1}$$

In particular this allows us to condition on $\mathcal{F}_{t_n}$ at any point on the random time-set $\{t_n\}_{n\in\mathbb{N}}$.

Assumption 3.1 For the sequence of random timesteps $\{h_{n+1}\}_{n\in\mathbb{N}}$, there are constant values $h_{\max} > h_{\min} > 0$, $\rho > 1$ such that $h_{\max} = \rho h_{\min}$, and

$$0 < h_{\min} \leq h_{n+1} \leq h_{\max} \leq 1. \tag{3.2}$$

In addition, we assume each h_{n+1} is $\mathcal{F}_{t_n}$-measurable.

Definition 3.2 Let $N^{(t)}$ be a random integer such that

$$N^{(t)} := \max \left\{ n \in \mathbb{N}\setminus\{0\} : t_{n-1} < t \right\}, \tag{3.3}$$

and let $N = N^{(T)}$ and $t_N = T$, so that T is always the last point on the mesh. Note that $N^{(t)}$ indicates the step number such that $t \in \left[t_{N^{(t)}-1}, \, t_{N^{(t)}} \right]$. Furthermore, by Assumption 3.1, $N^{(t)}$ only takes values in the finite set $\{N_{\min}^{(t)}, \ldots, N_{\max}^{(t)}\}$, where $N_{\min}^{(t)} := \lfloor t/h_{\max} \rfloor$ and $N_{\max}^{(t)} := \lceil t/h_{\min} \rceil$.

In Assumption 3.1, the lower bound $h_{\min}$ given by (3.2) ensures that a simulation over the interval $[0, T]$ can be completed in a finite number of time steps. In the event that at time t_n our strategy attempts to select a stepsize $h_{n+1} \leq h_{\min}$, we instead apply a single step of a backstop method (φ in Definition 3.3 below), a known method that satisfies a mean-square consistency requirement with deterministic step $h_{n+1} = h_{\min}$ (see also discussion in Remarks 3.1 and 5.1).

First we recall the Milstein method expressed as a map. Over each step $[t_n, t_{n+1}]$ the Milstein map $\theta : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^d$ is defined as

$$\theta\big(x, t_n, s - t_n\big) := x + (s - t_n) f(x) + \sum_{i=1}^{m} g_i(x) I_i^{t_n,s} + \sum_{i,j=1}^{m} \mathbf{D} g_i(x) g_j(x) I_{j,i}^{t_n,s}. \tag{3.4}$$

Following [18, Def. 9], we now define an adaptive Milstein scheme combining the Milstein method and a backstop method.

Definition 3.3 (Adaptive Milstein Scheme) Let $\{h_{n+1}\}_{n\in\mathbb{N}}$ satisfy Assumption 3.1. Using indicator functions to distinguish the backstop case when $h_{n+1} = h_{\min}$ (and allowing for the possibility that the final step taken to time T is smaller than $h_{\min}$, in which case the backstop is also used), we define the continuous form of an *adaptive Milstein scheme* associated with a particular time-stepping strategy $\{h_{n+1}\}_{n\in\mathbb{N}}$ as

$$\widetilde{Y}(s) := \theta\left(\widetilde{Y}(t_n),\ t_n,\ s - t_n\right) \cdot \mathbf{1}_{\{h_{\min} < h_{n+1} \le h_{\max}\}} + \varphi\left(\widetilde{Y}(t_n),\ t_n,\ s - t_n\right) \cdot \mathbf{1}_{\{h_{n+1} \le h_{\min}\}}, \tag{3.5}$$

for $s \in [t_n, t_{n+1}]$, $n \in \mathbb{N}$, $\widetilde{Y}(0) = X(0)$, and θ is as given in (3.4). Thus the scheme is characterised by the sequence of tuples, $\left\{\left(\widetilde{Y}(s)\right)_{s\in[t_n,t_{n+1}]}, h_{n+1}\right\}_{n\in\mathbb{N}}$. The backstop map $\varphi : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^d$ in (3.5) satisfies for each $n \in \mathbb{N}$

$$\mathbb{E}\left[\left\|X(s) - \varphi\left(\widetilde{Y}(t_n),\ t_n,\ s - t_n\right)\right\|^2 \middle| \mathcal{F}_{t_n}\right] \le \left\|X(t_n) - \widetilde{Y}(t_n)\right\|^2 + C_{B_1} \int_{t_n}^{s} \mathbb{E}\left[\left\|X(r) - \varphi\left(\widetilde{Y}(t_n),\ t_n,\ r - t_n\right)\right\|^2 \middle| \mathcal{F}_{t_n}\right] dr + C_{B_2} h_{\min}^3, \tag{3.6}$$

a.s, for positive constants C_{B_1} and C_{B_2}.

Throughout the article it is notationally convenient to make the following definition.

Definition 3.4 Let $\widetilde{Y}$ be as given in Definition 3.3 and define for each $n \in \mathbb{N}$

$$Y_\theta(s) := \theta\Big(\widetilde{Y}(t_n), t_n, s - t_n\Big), \quad s \in [t_n, t_{n+1}]. \tag{3.7}$$

Remark 3.1 The upper bound $h_{\max}$ prevents step sizes from becoming too large and allows us to examine strong convergence of the adaptive Milstein method (3.5) to solutions of (1.1) as $h_{\max} \to 0$ (and hence as $h_{\min} \to 0$). Note that φ satisfies (3.6) if the backstop method satisfies a mean-square consistency requirement. In practice, instead of testing (3.6), we choose a backstop method that is strongly convergent with rate 1.

Remark 3.2 For all $i, j = 1, 2, \ldots, m$, $I_i^{t_n,t_{n+1}}$ in (2.11) is a Wiener increment taken over a random step of length h_{n+1}, which itself may depend on $\widetilde{Y}(t_n)$ and therefore is not necessarily independent and normally distributed. However, since h_{n+1} is $\mathcal{F}_{t_n}$-measurable, then $I_i^{t_n,t_{n+1}}$ is $\mathcal{F}_{t_n}$-conditionally normally distributed and by the Optional Sampling Theorem (see for example [32]), for all $p = 0, 1, 2, \ldots$

$$\mathbb{E}\left[I_i^{t_n,t_{n+1}} \middle| \mathcal{F}_{t_n} \right] = 0, \quad a.s.; \tag{3.8}$$

$$\mathbb{E}\left[\left| I_i^{t_n,t_{n+1}} \right|^2 \middle| \mathcal{F}_{t_n} \right] = h_{n+1}, \quad a.s.; \tag{3.9}$$

$$\mathbb{E}\left[\left| I_i^{t_n,s} \right|^p \middle| \mathcal{F}_{t_n} \right] = \boldsymbol{\gamma}_p |s - t_n|^{\frac{p}{2}}, \quad a.s.; \tag{3.10}$$

where $\boldsymbol{\gamma}_p := 2^{p/2}\Gamma\left((p+1)/2\right)\pi^{-1/2}$, and Γ is the Gamma function (see for example [29, p. 148]). In implementation, it is sufficient to replace the sequence of Wiener increments with i.i.d. $\mathcal{N}(0, 1)$ random variables scaled at each step by the $\mathcal{F}_{t_n}$-measurable random variable $\sqrt{h_{n+1}}$.

We now provide a specific example of a time-stepping strategy that we use in Sect. 5 and that satisfies the assumptions for our convergence proof in Theorem 4.1. Suppose that for each $n = 0, \ldots, N-1$ and some fixed constant $\kappa > 0$, we choose constant values $h_{\max} > h_{\min} > 0$, $\rho > 1$ such that $h_{\max} = \rho h_{\min}$ and

$$h_{n+1} = h_{\min} \vee \left(h_{\max} \wedge \frac{h_{\max}}{\left\| \widetilde{Y}(t_n) \right\|^{1/\kappa}} \right). \tag{3.11}$$

Then (3.2) in Assumption 3.1 holds for (3.11). Notice also that, from (3.11), the following bound applies on the event $\{h_{\min} < h_{n+1} \leq h_{\max}\}$:

$$0 \leq \left\| \widetilde{Y}(t_n) \right\| < \left(\frac{h_{\max}}{h_{\min}} \right)^{\kappa} = \rho^{\kappa}.$$

The strategy given by (3.11) is admissible in the sense given in [17, 18]. However, it also motivates the following class of time-stepping strategies to which our convergence analysis applies.

Definition 3.5 (Path-bounded time-stepping strategies) Let $\left\{ \widetilde{Y}(t_n), h_{n+1} \right\}_{n \in \mathbb{N}}$ be a numerical approximation for (1.1) given by (3.5), associated with a timestep sequence $\{h_{n+1}\}_{n \in \mathbb{N}}$ satisfying Assumption 3.1. We say that $\{h_{n+1}\}_{n \in \mathbb{N}}$ is a path-bounded time-stepping strategy for (3.5) if there exist real non-negative constants $0 \leq Q < R$ (where R may be infinite if $Q \neq 0$) such that on the event $\{h_{\min} < h_{n+1} \leq h_{\max}\}$,

$$Q \leq \left\| \widetilde{Y}(t_n) \right\| < R, \quad n = 0, \ldots, N-1. \tag{3.12}$$

Note that throughout this paper we use a strategy where $Q = 0$ and $R < \infty$. As we will see in Sect. 5.2, a careful choice of the parameter κ can be used to minimise invocations of the backstop method when ρ is fixed.

4 Main results

Our first main result shows strong convergence with order 1 of solutions of (3.5) to solutions of (1.1) when $\{h_{n+1}\}_{n\in\mathbb{N}}$ is a path-bounded time-stepping strategy ensuring that (3.12) holds.

Theorem 4.1 *(Strong Convergence) Let $(X(t))_{t\in[0,T]}$ be a solution of (1.1) with initial value $X(0) = X_0 \in \mathbb{R}^d$. Suppose that the conditions of Assumptions 2.1 and 2.2 hold. Let $\left\{\left(\widetilde{Y}(s)\right)_{s\in[t_n,t_{n+1}]}, h_{n+1}\right\}_{n\in\mathbb{N}}$ be the adaptive Milstein scheme given in Definition 3.3 with initial value for the first component $\widetilde{Y}_0 = X_0$ and path-bounded time-stepping strategy $\{h_{n+1}\}_{n\in\mathbb{N}}$ satisfying the conditions of Definition 3.5 for some $R < \infty$. Then there exists a constant $C(R, \rho, T) > 0$ such that*

$$\max_{t\in[0,T]} \left(\mathbb{E}\left[\|X(t) - \widetilde{Y}(t)\|^2\right]\right)^{1/2} \leq C(R, \rho, T)\, h_{\max}. \tag{4.1}$$

Furthermore,

$$\lim_{\rho\to\infty} C(R, \rho, T) = \infty. \tag{4.2}$$

The proof of Theorem 4.1, which is given in Sect. 7.2, accounts for the properties of the random sequences $\{t_n\}_{n\in\mathbb{N}}$ and $\{h_{n+1}\}_{n\in\mathbb{N}}$ and uses (3.12) to compensate for the non-Lipschitz drift and diffusion.

Our second main result shows that for the specific strategy given by (3.11) the probability of needing a backstop method can be made arbitrarily small by taking ρ sufficiently large for fixed κ.

Theorem 4.2 ***(Probability of Backstop)*** *Let all the conditions of Theorem 4.1 hold, and suppose that the path-bounded time-stepping strategy $\{h_{n+1}\}_{n\in\mathbb{N}}$ also satisfies (3.11). Let $C(R, \rho, T)$ be the error constant in estimate (4.1) from the statement of Theorem 4.1.*

For any fixed $\kappa \geq 1$ there exists a constant $C_{prob} = C_{prob}(T, R, h_{\max})$ such that, for $h_{\max} \leq 1/C(R, \rho, T)$,

$$\mathbb{P}\left[h_{n+1} = h_{\min}\right] \leq C_{prob}\ \rho^{1-2\kappa}. \tag{4.3}$$

Further for arbitrarily small tolerance $\varepsilon \in (0, 1)$, there exists $\rho > 0$ such that

$$\mathbb{P}\left[h_{n+1} = h_{\min}\right] < \varepsilon, \quad n \in \mathbb{N}.$$

For proof see Sect. 7.3.

5 Numerical examples

Remark 5.1 We use the adaptive strategy in (3.11). We ensure that we reach the final time by taking $h_N = T - t_{N-1}$ as our final step, and in a situation where this is smaller than $h_{\min}$ we use the backstop method (this is compatible with the proofs below).

In the numerical experiments below, we set the *adaptive Milstein scheme* (`AMil`) as in (3.5) with (3.11) as the choice of h_{n+1}. *Projected Milstein* (`PMil`) in [1, Eq. (24)] is set to be the backstop method of `AMil` and the reference method of all models. Then we compare the strong convergence,looking at the root mean square (RMS) error, and efficiency, by comparing the CPU time, of `AMil` and `PMil`, *Split-Step Backward Milstein* method (`SSBM`) [1, Eq. (25)], the *new variant of Milstein* (`TMil`) in [20], and the *Tamed Stochastic Runge–Kutta of order* 1.0 (`TSRK1`) method [8, Eq. (3.8) (3.9)]. For the non-adaptive schemes, to examine strong convergence, we take as the fixed step h_{mean} the average of all time steps over each path and each Monte Carlo realization $m = 1, \ldots, M$ so that

$$h_{\text{mean}} := \frac{1}{M} \sum_{m=1}^{M} \frac{T}{N_m},$$

where N_m denotes the number of steps taken on the mth sample path to reach T.

5.1 One-dimensional test equations with multiplicative and additive noise

In order to demonstrate strong convergence of order one for a scalar test equation with non-globally Lipschitz drift, consider

$$dX(t) = \big(X(t) - 3X(t)^3\big)dt + G(X(t))dW(t), \quad t \in [0, 1]. \tag{5.1}$$

For illustrating both the multiplicative and additive noise cases, we estimate the RMS error by a Monte Carlo method using $M = 1000$ trajectories for $h_{\max} = [2^{-14}, 2^{-12}, 2^{-10}, 2^{-8}, 2^{-6}]$, $\rho = 2^2$, $\kappa = 1$, and use as a reference solution `PMil` over a mesh with uniform step sizes $h_{\text{ref}} = 2^{-18}$.

For additive noise we set $G(x) = \sigma$ in (5.1), and for multiplicative noise we set $G(x) = \sigma(1 - x^2)$ with $\sigma = 0.2$ and $X(0) = 11$ in both cases. Strong convergence of order one is displayed by all methods in Fig. 1 part (a) and (c) for the additive and multiplicative cases respectively, with the efficiency displayed in parts (b) and (d).

Finally, consider Theorem 4.2. We illustrate that the probability of our time-stepping strategy selecting $h_{\min}$, and therefore triggering an application of the backstop method, can be made arbitrarily small at every step by an appropriate choice of ρ (with fixed $\kappa = 1$). Consider (5.1) again with $G(x) = \sigma(1 - x^2)$, this time with $X(0) = 100$, $\kappa, T = 1$, $h_{\max} = 2^{-20}$ and $\rho = [2, 4, \ldots, 16]$. In Fig. 1e, we plot two paths of h when $\rho = 2, 6$. Observe that when $\rho = 2$ the backstop is triggered only for the first 10^5 steps approximately, whereas once ρ is increase to 6 this is reduced to the first 2×10^4 steps approximately. Estimated probabilities of using $h_{\min}$ are plotted on a

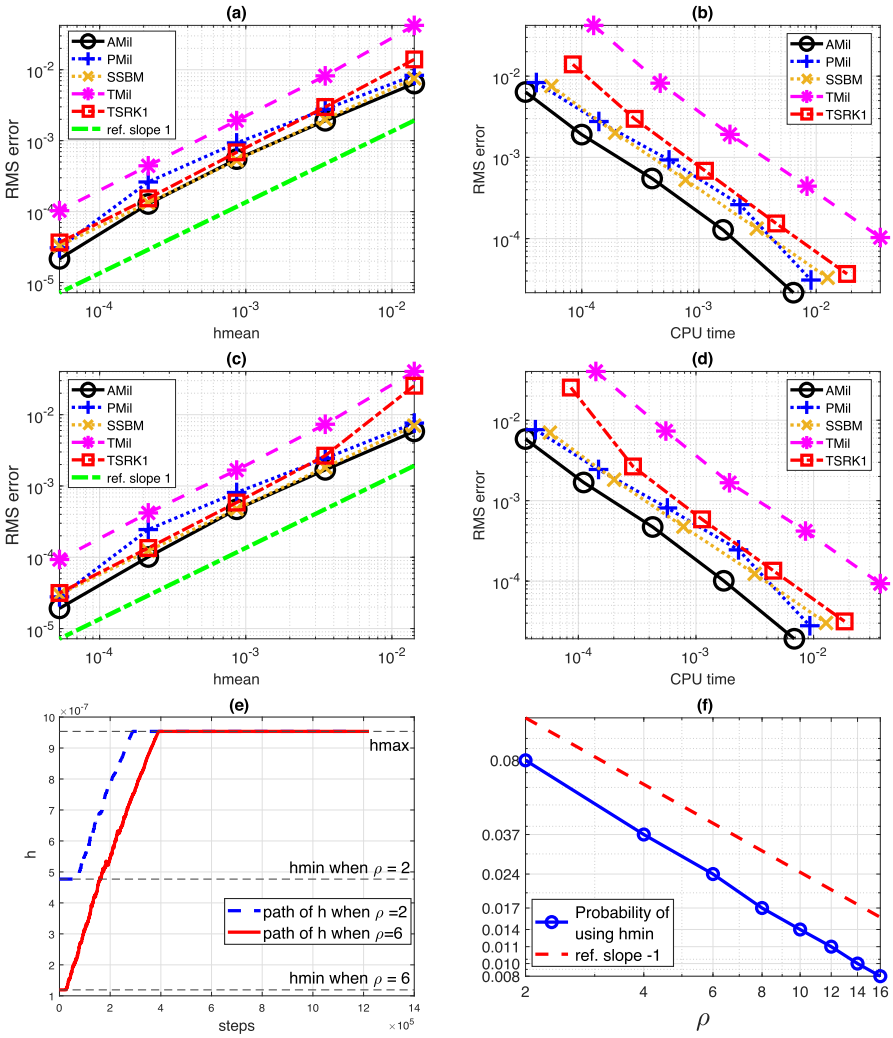

Fig. 1 Strong convergence and efficiency of model (5.1) with **a** and **b** for additive noise; **c** and **d** for multiplicative noise. **e** Two paths of timestep h for $\rho = 2, 6$ and in **f** the estimated probability of using $h_{\min}$ for the multiplicative noise model with $M = 100$ realizations

log-log scale as a function of ρ in Fig. 1f (with $M = 100$ realizations). The estimated probability of using $h_{\min}$ declines to zero as ρ increases. We observe a rate close to -1, matching that in (4.3) with $\kappa = 1$.

5.2 One-dimensional model of telomere shortening

The following one-dimensional SDE model was given in [9, Eq. (A6)] for modelling the shortening over time of telomere length L in DNA replication

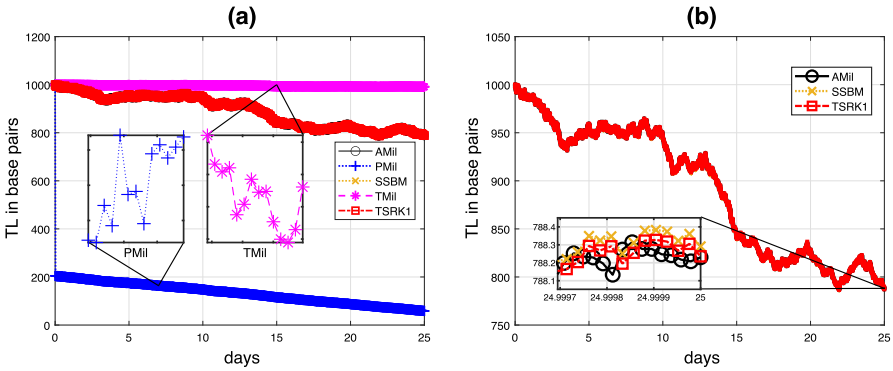

Fig. 2 Single paths of the Telomere length SDE (5.2) solved over 25 days. **b** shows a detailed plot from (**a**)

$$dL(t) = -\big(c + aL(t)^2\big)dt + \sqrt{\frac{1}{3}aL(t)^3}dW(t). \tag{5.2}$$

The parameter c determines the underlying decay rate of the length and a controls the intensity at which random breaks occur in the telomere; we take $(a, c) = (0.41 \times 10^{-6}, 7.5)$ as in [9]. In this example we fix $\rho = 4$, instead adjusting the parameter κ in (3.11) to control use of the backstop method. Individual paths are shown in Fig. 2 where we take $h_{\max} = 2^{-18}$, and $h = 2^{-20}$ for the fixed step methods.

We set $L(0) = 1000$, noting from [9] that initial values could be as high as (say) $L(0) = 6000$ and remain physically realistic. The end of the interval of valid simulation is determined by the first time at which trajectories reach zero, and is therefore random. However this is not observed to occur in the timescale (25 days) we consider here.

By design `PMil` projects the data onto a ball of radius determined in part by the growth of the drift term. We see in Fig. 2a that (`PMil`) immediately is reduced to approximately 200.

Contrarily, the design of `TMil` scales both drift and diffusion terms by $1/(1+h|L|^2)$ for this model. When $h|L|^2$ is large this scaling can damp out changes from step to step, and in Fig. 2a we see that `TMil` shows as (spuriously) almost constant. The paths of the other methods, `AMil`, `SSBM` and `TSRK1` are close together as shown in Fig. 2a and in high detail in (b).

Notice that we used $\kappa = 8$ in (3.11) for `AMil` method to reduce the chance of requiring the backstop method `PMil` while keeping $\rho = 4$. We avoid setting $\kappa = 1$ in this case because $L(0) = 1000$ and so the adaptive step h_{n+1} would too frequently require the backstop method.

5.3 Two-dimensional test systems

We now consider three ($i = 1, 2, 3$) different SDEs:

$$dX(t) = F(X(t))dt + G_i(X(t))dW(t), \quad t \in [0, 1], \quad X(0) = [7, 9]^T, \tag{5.3}$$

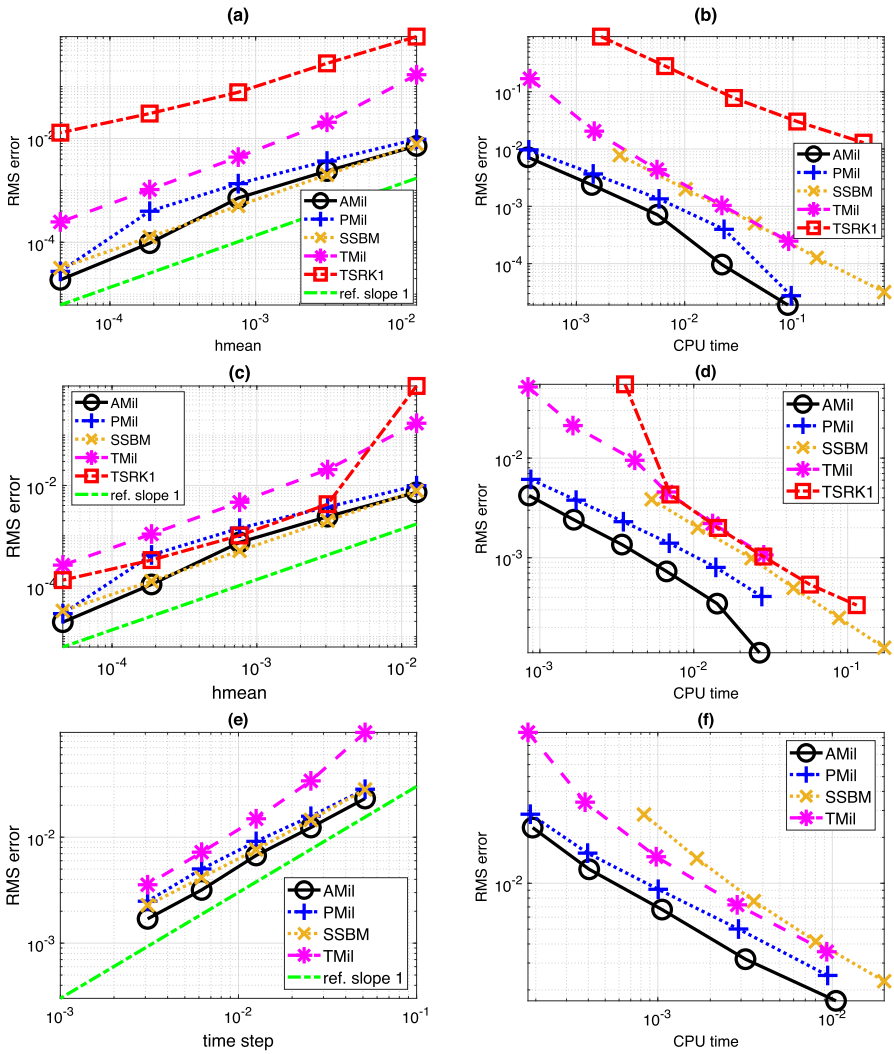

Fig. 3 Two-dimensional system (5.3). **a** and **b** show the strong convergence and efficiency for diagonal noise, **c** and **d** with commutative and **e** and **f** for non-commutative noise. We choose $a = 3$, $\sigma = 0.2$ and $b = 1.5$

with $W(t) = [W_1(t), W_2(t)]^T$, where W_1 and W_2 are independent scalar Wiener processes, $X(t) = [X_1(t), X_2(t)]^T$, $F(x) = [x_2 - 3x_1^3, x_1 - 3x_2^3]^T$, and

$$G_1(x) = \sigma \begin{pmatrix} x_1^2 & 0 \\ 0 & x_2^2 \end{pmatrix}, \; G_2(x) = \sigma \begin{pmatrix} x_2^2 & x_2^2 \\ x_1^2 & x_1^2 \end{pmatrix}, \; G_3(x) = \sigma \begin{pmatrix} 1.5x_1^2 & x_2 \\ x_2^2 & 1.5x_1 \end{pmatrix}.$$

G_1 is an example of diagonal noise, G_2 commutative noise, and G_3 non-commutative noise.

For G_1 and G_2 we use $h_{\max} = [2^{-14}, 2^{-12}, 2^{-10}, 2^{-8}, 2^{-6}]$, $h_{\text{ref}} = 2^{-18}$, $\rho = 4$ and $\kappa = 1$. In Fig. 3a, c, we see order one strong convergence for all methods. Parts (b) and (d) show the efficiency of the adaptive method.

For $i = 3$, the non-commutative noise case, take $h_{\max} = [2^{-8}, 2^{-7}, 2^{-6}, 2^{-5}, 2^{-4}]$, $h_{\text{ref}} = 2^{-11}$, $\rho = 2^2$ and $X(0) = [3, 4]^T$. To simulate the Lévy areas we follow the method in [12, Sec. 4.3], which is based on the Euler approximation of a system of SDEs. Again, we observe order one convergence for all methods in Fig. 3e and that `AMil` is the most efficient in (f). Note that as `TSRK1` is only supported theoretically for commutative noise we do not consider it here.

6 Preliminary lemmas

We present five lemmas necessary for the proof of Theorem 4.1 and Theorem 4.2. Throughout this section we assume that f and g satisfy Assumptions 2.1 and (except for Lemma 6.4) that we are on the event $\{h_{\min} < h_{n+1} \leq h_{\max}\}$ so that (3.12) holds of Definition 3.5. We use (2.6), (2.8) and (2.7) to define some bounded constant coefficients depending on $R < \infty$. The constants in (6.1) are then used in the development of a one-step error bound for the adaptive part of the scheme.

$$\begin{aligned} \left\| f\left(\widetilde{Y}(t_n)\right) \right\| &\leq c_5(1 + R^{q_1+2}) =: C_f; \\ \left\| \mathbf{D} f\left(\widetilde{Y}(t_n)\right) \right\|_{\mathbf{F}} &\leq c_3(1 + R^{q_1+1}) =: C_{Df}; \\ \left\| g_i\left(\widetilde{Y}(t_n)\right) \right\| \leq \left\| g\left(\widetilde{Y}(t_n)\right) \right\|_{\mathbf{F}(d\times m)} &\leq c_6(1 + R^{q_2+2}) =: C_{g_i}; \\ \left\| \mathbf{D} g_i\left(\widetilde{Y}(t_n)\right) \right\|_{\mathbf{F}} &\leq c_4(1 + R^{q_2+1}) =: C_{Dg_i}. \end{aligned} \tag{6.1}$$

The following lemma provides a bound for the even conditional moments of the iterated stochastic integral in (2.11).

Lemma 6.1 *(`Iterated Stochastic Integral`) Let* $\left\{\left(\widetilde{Y}(s)\right)_{s\in[t_n,t_{n+1}]}, h_{n+1}\right\}_{n\in\mathbb{N}}$ *be the adaptive Milstein scheme given in Definitions 3.3 and 3.5. Then there exists a constant* C_{ISI} *such that for* $k \geq 1$, $n \in \mathbb{N}$ *and* $s \in [t_n, t_{n+1}]$, *on the event* $\{h_{\min} < h_{n+1} \leq h_{\max}\}$

$$\mathbb{E}\left[\left\| \sum_{i,j=1}^{m} \mathbf{D} g_i\left(\widetilde{Y}(t_n)\right) g_j\left(\widetilde{Y}(t_n)\right) I_{j,i}^{t_n,s} \right\|^{2k} \middle| \mathcal{F}_{t_n}\right] \leq C_{ISI}(k, R)\, |s - t_n|^{2k}, \tag{6.2}$$

where

$$C_{ISI}(k, R) := 3^{2k} m^{4k} C_{Dg_i}^{2k} C_{g_i}^{2k} \left(\boldsymbol{\gamma}_{4k} + 1 + \boldsymbol{\gamma}_{2k}^2 + C_{LA}(2k)\right). \tag{6.3}$$

Here, $\boldsymbol{\gamma}_p$ *is from (3.10),* $C_{LA}(2k)$ *is from Lemma 2.2 with explicit form given in (A.1), and the* R *dependence in* $C_{ISI}(k, R)$ *arises from (6.1).*

Proof First of all, for convenience we set

$$G_{\mathtt{ISI}}(s) := \left\| \sum_{i,j=1}^{m} \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_{i,j}^{t_n,s} \right\|^{2k}.$$

By (2.12) and (2.3), we have, for $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$,

$$\begin{aligned} G_{\mathtt{ISI}}(s) \leq 3^{2k-1}\Bigg(& \left\| \frac{1}{2}\sum_{i=1}^{m} \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)\Big(\big(I_i^{t_n,s}\big)^2 - |s-t_n|\Big) \right\|^{2k} \\ & + \left\| \frac{1}{2}\sum_{\substack{i,j=1\\ i<j}}^{m} \Big(\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big) + \mathbf{D}g_j\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)\Big)I_i^{t_n,s}I_j^{t_n,s} \right\|^{2k} \\ & + \left\| \sum_{\substack{i,j=1\\ i<j}}^{m} \Big(\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big) - \mathbf{D}g_j\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)\Big)A_{ij}^{t_n,s} \right\|^{2k} \Bigg). \end{aligned}$$

Applying (2.3) again and by submultiplicativity of the Euclidean norm and the fact that the induced matrix 2-norm is bounded above by the Frobenius norm, for $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$, we get

$$\begin{aligned} G_{\mathtt{ISI}}(s) \leq 3^{2k-1}\Bigg(& \frac{m^{2k-1}}{2^{2k}}\sum_{i=1}^{m} \big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}^{2k}\big\|g_i\big(\widetilde{Y}(t_n)\big)\big\|^{2k}\Big(\big(I_i^{t_n,s}\big)^2 + |s-t_n|\Big)^{2k} \\ & + \left(\frac{m(m-1)}{2}\right)^{2k-1} \sum_{\substack{i,j=1\\ i<j}}^{m} \left\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big) + \mathbf{D}g_j\big(\widetilde{Y}(t_n)\big)g_i(Y(t_n)\right\|^{2k} \\ & \times\left(\frac{1}{2^{2k}}\left|I_i^{t_n,s}I_j^{t_n,s}\right|^{2k} + \left|A_{ij}^{t_n,s}\right|^{2k}\right)\Bigg). \end{aligned}$$

Applying conditional expectations on both sides, together with the pairwise conditional independence of $I_i^{t_n,s}$ and $I_j^{t_n,s}$ for $i \neq j$, (2.6) and (6.1), we have for $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$

$$\begin{aligned} \mathbb{E}\Big[G_{\mathtt{ISI}}(s)\Big|\mathcal{F}_{t_n}\Big] \leq 3^{2k}\Bigg(& m^{2k-1}C_{Dg_i}^{2k}C_{g_i}^{2k}\sum_{i=1}^{m}\left(\mathbb{E}\left[\left|I_i^{t_n,s}\right|^{4k}\Big|\mathcal{F}_{t_n}\right] + |s-t_n|^{2k}\right) \\ & + \left(\frac{m(m-1)}{2}\right)^{2k-1} C_{Dg_i}^{2k}C_{g_i}^{2k}\sum_{\substack{i,j=1\\ i<j}}^{m}\Bigg(\mathbb{E}\left[\left|I_i^{t_n,s}\right|^{2k}\Big|\mathcal{F}_{t_n}\right]\mathbb{E}\left[\left|I_j^{t_n,s}\right|^{2k}\Big|\mathcal{F}_{t_n}\right] \\ & + \mathbb{E}\left[\left|A_{ij}^{t_n,s}\right|^{2k}\Big|\mathcal{F}_{t_n}\right]\Bigg)\Bigg). \end{aligned}$$

Using (3.9), (3.10) and (2.14) we have

$$\mathbb{E}\Big[G_{\mathtt{ISI}}(s)\Big|\mathcal{F}_{t_n}\Big] \leq C_{\mathtt{ISI}}\,(k,\,R)\,|s-t_n|^{2k},$$

where $C_{\mathtt{ISI}}\,(k,\,R)$ is in (6.3). □

The following lemma provides a bound on the conditional moments of the adaptive Milstein scheme in (3.5) over one step, in the case where the method applies the map θ.

Lemma 6.2 *Consider* $\left\{\left(\widetilde{Y}(s)\right)_{s\in[t_n,t_{n+1}]},\,h_{n+1}\right\}_{n\in\mathbb{N}}$ *from Definitions 3.3 and 3.5, and let* $(Y_\theta(s))_{s\in(t_n,t_{n+1}]}$ *be as defined in Definition 3.4. Then there exists a constant* $C_{Y_\theta}>0$ *such that for* $k\geq 1$, $n\in\mathbb{N}$ *and* $s\in(t_n,t_{n+1}]$, *on the event* $\{h_{\min}<h_{n+1}\leq h_{\max}\}$,

$$\mathbb{E}\Big[\left\|Y_\theta(s)\right\|^k\Big|\mathcal{F}_{t_n}\Big] \leq C_{Y_\theta}\big(k,\,R\big), \tag{6.4}$$

where

$$C_{Y_\theta}\big(k,\,R\big) := \;4^{k-1}\Big(R^k + C_f^k + m^k C_{g_i}^k\;\boldsymbol{\gamma}_k + C_{\mathtt{ISI}}\,(2k)^{1/2}\Big), \tag{6.5}$$

with the constant $C_{\mathtt{ISI}}$ *from Lemma 6.1.*

Proof By (3.7), (3.4) and (2.3), we have, for $s\in(t_n,t_{n+1}]$ and $n\in\mathbb{N}$,

$$\begin{aligned}\left\|Y_\theta(s)\right\|^k &= \left\|\theta\left(\widetilde{Y}(t_n),\;t_n,\;s-t_n\right)\right\|^k\\ &\leq 4^{k-1}\Bigg(\left\|\widetilde{Y}(t_n)\right\|^k + \left\|f\big(\widetilde{Y}(t_n)\big)\right\|^k|s-t_n|^k + \left\|\sum_{i=1}^m g_i\big(\widetilde{Y}(t_n)\big)I_i^{t_n,s}\right\|^k\\ &\quad+\left(\left\|\sum_{i,j=1}^m \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_{j,i}^{t_n,s}\right\|^{2k}\right)^{1/2}\Bigg).\end{aligned}$$

Applying (2.3), (3.12) and (6.1) for $s\in(t_n,t_{n+1}]$ and $n\in\mathbb{N}$, it yields

$$\begin{aligned}\left\|Y_\theta(s)\right\|^k \leq 4^{k-1}\Bigg(&R^k + C_f^k|s-t_n|^k + m^{k-1}C_{g_i}^k\sum_{i=1}^m\left|I_i^{t_n,s}\right|^k\\ &+\left(\left\|\sum_{i,j=1}^m \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_{j,i}^{t_n,s}\right\|^{2k}\right)^{1/2}\Bigg).\end{aligned}$$

Taking conditional expectation on both sides, with Jensen's inequality on the last term we have for $s\in(t_n,t_{n+1}]$ and $n\in\mathbb{N}$

$$\mathbb{E}\Big[\left\|Y_\theta(s)\right\|^k\Big|\mathcal{F}_{t_n}\Big] \leq 4^{k-1}\Bigg(R^k + C_f^k|s-t_n|^k + m^{k-1}C_{g_i}^k\sum_{i=1}^m\mathbb{E}\Big[\left|I_i^{t_n,s}\right|^k\Big|\mathcal{F}_{t_n}\Big]$$

$$+\left(\mathbb{E}\left[\left\|\sum_{i,j=1}^{m}\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_{j,i}^{t_n,s}\right\|^{2k}\middle|\mathcal{F}_{t_n}\right]\right)^{1/2}\Bigg).$$

Using (3.9), (6.2) from Lemma 6.1 and since $|s-t_n|\le h_{\max}\le 1$ (3.2) we have

$$\mathbb{E}\Big[\|Y_\theta(s)\|^k\Big|\mathcal{F}_{t_n}\Big]\le C_{Y_\theta}(k,R),$$

where $C_{Y_\theta}(k,R)$ is in (6.5). □

The following lemma proves regularity in time of the adaptive Milstein scheme in (3.5) when applying the map θ.

Lemma 6.3 *(Scheme Regularity) Consider* $\big\{\big(\widetilde{Y}(s)\big)_{s\in[t_n,t_{n+1}]},h_{n+1}\big\}_{n\in\mathbb{N}}$ *in Definitions 3.3 and 3.5, and let* $(Y_\theta(s))_{s\in(t_n,t_{n+1}]}$ *be as defined in Definition 3.4. Then there exists a constant* C_{SR} *such that for* $k\ge 1$, $n\in\mathbb{N}$ *and* $s\in(t_n,t_{n+1}]$, *on the event* $\{h_{\min}<h_{n+1}\le h_{\max}\}$

$$\mathbb{E}\Big[\big\|Y_\theta(s)-\widetilde{Y}(t_n)\big\|^{2k}\Big|\mathcal{F}_{t_n}\Big]\le C_{SR}(k,R)\,|s-t_n|^k, \tag{6.6}$$

where

$$C_{SR}(k,R):=\ 3^{2k-1}\Big(C_f^{2k}+m^{2k}C_{g_i}^{2k}\,\boldsymbol{\gamma}_{2k}+C_{ISI}(2k)\Big), \tag{6.7}$$

with the constant C_{ISI} *from Lemma 6.1.*

Proof The method of proof is similar to the proof of Lemma 6.2. □

Remark 6.1 Our analysis requires a certain number of finite moments for the SDE (1.1), and it is necessary to track exactly what those are in order to see that the conditions of Assumption 2.2 are not violated. To this end, we introduce a superscript notation for random variables appearing as conditional expectations at this point. The notation should be interpreted according to the following example: in (6.9) below the random variable $C_{\mathrm{PR}}^{\{2k(q+2)\}}$ requires $2k(q+2)$ finite moments of the SDE (1.1) to have finite expectation.

The following lemma examines the regularity of solutions of the SDE (1.1).

Lemma 6.4 *(Path Regularity) Let* f, g *also satisfy Assumption 2.2, and let* $(X(s))_{s\in[t_n,t_{n+1}]}$ *be a solution of* (1.1). *Then there exists an* $\mathcal{F}_{t_n}$*-measurable random variable* $\overline{C}_{PR}^{\{2k(q+2)\}}$ *such that for* $k\ge 1$, $n\in\mathbb{N}$ *and* $s\in[t_n,t_{n+1}]$ *a.s.*

$$\mathbb{E}\Big[\|X(s)-X(t_n)\|^{2k}\Big|\mathcal{F}_{t_n}\Big]\ \le\ \overline{C}_{PR}^{\{2k(q+2)\}}\ |s-t_n|^k, \tag{6.8}$$

where $q=q_1\vee q_2$ *is as defined in Assumption 2.2. Where a.s.*

$$\overline{C}_{PR}^{\{2k(q+2)\}} = 2^{4k-2}c_5^{2k}\left(1+\mathbb{E}\left[\sup_{p\in[t_n,t_{n+1}]}\|X(p)\|^{2k(q_1+2)}\middle|\mathcal{F}_{t_n}\right]\right) + 2^{4k-2}(k(2k-1))^k c_6^{2k}\left(1+\mathbb{E}\left[\sup_{p\in[t_n,t_{n+1}]}\|X(p)\|^{2k(q_2+2)}\middle|\mathcal{F}_{t_n}\right]\right), \tag{6.9}$$

where the expectation of $\overline{C}_{PR}^{\{2k(q+2)\}}$ *is denoted* $C_{PR}(k)$*, given by*

$$C_{PR}(k) := \mathbb{E}\left[\overline{C}_{PR}^{\{2k(q+2)\}}\right] \leq 2^{4k-2}(1+C_X)\left(c_5^{2k}+(k(2k-1))^k c_6^{2k}\right). \tag{6.10}$$

Proof The method of proof follows that of [25, Thm. 7.1]. The bound (6.10) follows from (2.9) and Assumption 2.2. □

The following lemma provides a bound on the even conditional moments of the remainder term from a Taylor expansion of either the drift f or diffusion g, around $\widetilde{Y}(t_n)$.

Lemma 6.5 *(Taylor Error) Consider* $\left\{\left(\widetilde{Y}(s)\right)_{s\in[t_n,t_{n+1}]}, h_{n+1}\right\}_{n\in\mathbb{N}}$ *from Definitions 3.3 and 3.5, and let* $(Y_\theta(s))_{s\in[t_n,t_{n+1}]}$ *be as defined in Definition 3.4. Let* $u \in \{f, g\}$ *and set* $c_{\mathbf{D2}} := c_1 \vee c_2$*. Then there exists a constant* C_{TE} *such that for* $k \geq 1$*,* $n \in \mathbb{N}$ *and* $s \in [t_n, t_{n+1}]$*, on the event* $\{h_{\min} < h_{n+1} \leq h_{\max}\}$*,*

$$\mathbb{E}\left[\left\|\int_0^1(1-\epsilon)\mathbf{D}^2u\left(\widetilde{Y}(t_n)-\epsilon\left(Y_\theta(s)-\widetilde{Y}(t_n)\right)\right)d\epsilon\right\|_{\mathbf{T_3}}^{2k}\middle|\mathcal{F}_{t_n}\right] \leq C_{TE}(k,R), \tag{6.11}$$

where $C_{TE}(k,R) := c_{\mathbf{D2}}^{2k}\left(1+3^{2kq+1}\left(R^{2kq}+C_{Y_\theta}(k,R)\right)\right)$*, where* $C_{Y_\theta}(k,R)$ *is from Lemma 6.2.*

Proof By using (2.2), (2.3), (2.8), Lemma 6.2, (3.12) and since $c_{\mathbf{D2}} = c_1 \vee c_2$, $q = q_1 \vee q_2$ we have

$$\begin{aligned}
&\mathbb{E}\left[\left\|\int_0^1(1-\epsilon)\mathbf{D}^2u\left(Y(t_n)-\epsilon\left(Y_\theta(s)-\widetilde{Y}(t_n)\right)\right)d\epsilon\right\|_{\mathbf{T_3}}^{2k}\middle|\mathcal{F}_{t_n}\right]\\
&\leq \mathbb{E}\left[\int_0^1(1-\epsilon)^{2k}\left\|\mathbf{D}^2u\left(\widetilde{Y}(t_n)-\epsilon\left(Y_\theta(s)-\widetilde{Y}(t_n)\right)\right)\right\|_{\mathbf{T_3}}^{2k}d\epsilon\middle|\mathcal{F}_{t_n}\right]\\
&\leq c_{\mathbf{D2}}^{2k}\mathbb{E}\left[\int_0^1(1-\epsilon)^{2k}\left(1+\left\|\widetilde{Y}(t_n)-\epsilon\cdot\left(Y_\theta(s)-\widetilde{Y}(t_n)\right)\right\|^{2kq}\right)d\epsilon\middle|\mathcal{F}_{t_n}\right]\\
&\leq c_{\mathbf{D2}}^{2k}\mathbb{E}\Big[1+3^{2kq-1}\left\|\widetilde{Y}(t_n)\right\|^{2kq}\\
&\quad+\int_0^1(1-\epsilon)^{2k}\epsilon^{2kq}3^{2kq}\left(\|Y_\theta(s)\|^{2kq}+\left\|\widetilde{Y}(t_n)\right\|^{2kq}\right)d\epsilon\middle|\mathcal{F}_{t_n}\Big]\\
&\leq c_{\mathbf{D2}}^{2k}\left(1+3^{2kq+1}\left\|\widetilde{Y}(t_n)\right\|^{2kq}+3^{2kq}\mathbb{E}\left[\|Y_\theta(s)\|^{2kq}\middle|\mathcal{F}_{t_n}\right]\right)
\end{aligned}$$

$$= C_{\mathrm{TE}}(k, R),$$

where $(1-\epsilon)^{2k}\,\epsilon^{2kq} \le 1$ for $k, q \ge 1$ and $\epsilon \in [0, 1]$. □

7 Proof of main theorems

In this section we prove the strong convergence result of Theorem 4.1 and Theorem 4.2 on the probability of using the backstop and the role of ρ.

7.1 Setting up the error function

Notice that $\widetilde{Y}(s)$, from the explicit adaptive Milstein scheme (3.5), takes either the Milstein map θ in (3.4) or the backstop map φ in (3.6) depending on the value of h_{n+1}. Thus, we define the error by

$$\widetilde{E}(s) := X(s) - \widetilde{Y}(s) = E_\theta(s) + E_\varphi(s), \tag{7.1}$$

for $s \in [t_n, t_{n+1}]$ and $n \in \mathbb{N}$. Here

$$E_\varphi(s) := \left(X(s) - \varphi\left(\widetilde{Y}(t_n),\, t_n,\, s - t_n\right)\right) \mathbf{1}_{\{h_{n+1} \le h_{\min}\}}, \tag{7.2}$$

and $Y_\theta(s)$ is as defined in Definition 3.4 and

$$\begin{aligned} E_\theta(s) &:= \left(X(s) - Y_\theta(s)\right) \mathbf{1}_{\{h_{\min} < h_{n+1} \le h_{\max}\}} \\ &= \left(\widetilde{E}(t_n) + \int_{t_n}^{s} \Delta f\left(X(r), \widetilde{Y}(t_n)\right) dr \right. \\ &\quad \left. + \sum_{i=1}^{m} \int_{t_n}^{s} \Delta g_i\left(r, X(r), \widetilde{Y}(t_n)\right) dW_i(r)\right) \mathbf{1}_{\{h_{\min} < h_{n+1} \le h_{\max}\}}, \end{aligned} \tag{7.3}$$

with

$$\Delta f\left(X(r), \widetilde{Y}(t_n)\right) := f(X(r)) - f\left(\widetilde{Y}(t_n)\right); \tag{7.4}$$

$$\Delta g_i\left(r, X(r), \widetilde{Y}(t_n)\right) := g_i(X(r)) - g_i\left(\widetilde{Y}(t_n)\right) - \sum_{j=1}^{m} \mathbf{D} g_i\left(\widetilde{Y}(t_n)\right) g_j\left(\widetilde{Y}(t_n)\right) I_j^{t_n, r}. \tag{7.5}$$

To simplify the proof of Theorem 4.1 and Theorem 4.2, we require two lemmas. First, we find the second-moment bound of Δg_i in (7.5) on the event $\{h_{\min} < h_{n+1} \le h_{\max}\}$ (so that (3.12) holds).

Lemma 7.1 *Let g satisfy Assumption 2.1 and Δg_i be as in (7.5). Take $s \in [t_n, t_{n+1}]$, let $X(s)$ be a solution of (1.1), consider $\left(\widetilde{Y}(s), h_{n+1}\right)$ from Definitions 3.3 and 3.5,*

and let $Y_\theta(s)$ be as defined in Definition 3.4. In this case there exists a constant C_G such that, on the event $\{h_{\min} < h_{n+1} \le h_{\max}\}$,

$$\begin{aligned}&\mathbb{E}\left[\left\|\Delta g_i\big(s, X(s), \widetilde{Y}(t_n)\big)\right\|^2 \middle| \mathcal{F}_{t_n}\right]\\ &\le 2\mathbb{E}\Big[\left\|g(X(s)) - g\big(Y_\theta(s)\big)\right\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big] + C_G(R)|s-t_n|^2,\end{aligned} \tag{7.6}$$

where

$$C_G(R) := 8C^2_{Dg_i}\big(C_f^2 + C_{ISI}(1,R)\big) + 4C_{TE}(2,R)^{1/2}\,C_{SR}(4,R)^{1/2}, \tag{7.7}$$

and C_{ISI}, C_{TE} and C_{SR} are from Lemmas 6.1, 6.5 and 6.3, respectively.

Proof Substitute Δg_i by (7.5) in the LHS of (7.6), add and subtract $g_i\big(Y_\theta(s)\big)$, and use (2.3) to get

$$\begin{aligned}&\mathbb{E}\left[\left\|\Delta g_i\big(s, X(s), \widetilde{Y}(t_n)\big)\right\|^2 \middle| \mathcal{F}_{t_n}\right] \le 2\mathbb{E}\Big[\left\|g_i(X(s)) - g_i\big(Y_\theta(s)\big)\right\|^2\Big|\mathcal{F}_{t_n}\Big]\\ &\underbrace{+2\mathbb{E}\Big[\left\|g_i\big(Y_\theta(s)\big) - g_i\big(\widetilde{Y}(t_n)\big) - \sum_{j=1}^m \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_j^{t_n,s}\right\|^2\Big|\mathcal{F}_{t_n}\Big]}_{=:G_1}.\end{aligned} \tag{7.8}$$

To analyse G_1, we expand $g_i(Y_\theta(s))$ using Taylor's theorem (see for example [23, A.1]) around $g_i\big(\widetilde{Y}(t_n)\big)$ to get

$$\begin{aligned}&g_i\big(Y_\theta(s)\big) - g_i\big(\widetilde{Y}(t_n)\big) = \mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\big(Y_\theta(s) - \widetilde{Y}(t_n)\big)\\ &+\int_0^1 (1-\epsilon)\mathbf{D}^2 g_i\Big(\widetilde{Y}(t_n) - \epsilon\big(Y_\theta(s) - \widetilde{Y}(t_n)\big)\Big)\Big[Y_\theta(s) - \widetilde{Y}(t_n)\Big]^2 d\epsilon,\end{aligned} \tag{7.9}$$

where we recall from Sect. 2 that $[\cdot]^2$ represents the outer product of a vector with itself. Substituting (7.9) into G_1 in (7.8), taking out $\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)$ as a common factor, and applying (2.3) gives

$$\begin{aligned}G_1 &\le 4\mathbb{E}\Big[\Big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\Big(Y_\theta(s) - \widetilde{Y}(t_n) - \sum_{j=1}^m g_j\big(\widetilde{Y}(t_n)\big)I_j^{t_n,s}\Big)\Big\|^2\Big|\mathcal{F}_{t_n}\Big]\\ &\quad + 4\mathbb{E}\Big[\Big\|\int_0^1 (1-\epsilon)\mathbf{D}^2 g_i\Big(\widetilde{Y}(t_n) - \epsilon\big(Y_\theta(s) - \widetilde{Y}(t_n)\big)\Big)\\ &\qquad\qquad \times\Big[Y_\theta(s) - \widetilde{Y}(t_n)\Big]^2 d\epsilon\Big\|^2\Big|\mathcal{F}_{t_n}\Big]\\ &=: G_{1.1} + G_{1.2}.\end{aligned} \tag{7.10}$$

For $G_{1.1}$ in (7.10), by submultiplicativity of the Euclidean norm and the fact that the induced matrix 2-norm is bounded above by the Frobenius norm; by (3.7), (6.1) and (6.2) in the statement of Lemma 6.1 with $k = 1$, we have

$$\begin{aligned} G_{1.1} &\leq 8\left\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\right\|_{\mathbf{F}}^2\Bigg(\left\|f\big(\widetilde{Y}(t_n)\big)\right\|^2|s-t_n|^2 \\ &\quad + \mathbb{E}\Bigg[\left\|\sum_{i,j=1}^{m}\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)I_{j,i}^{t_n,s}\right\|^2\Bigg|\mathcal{F}_{t_n}\Bigg]\Bigg) \\ &\leq 8C_{Dg_i}^2\big(C_f^2 + C_{\mathtt{ISI}}(1,R)\big)|s-t_n|^2. \end{aligned} \tag{7.11}$$

For $G_{1.2}$ in (7.10), we apply (2.2), the Cauchy-Schwarz inequality, then using (6.11) in Lemma 6.5 with $k = 2$ and (6.6) in Lemma 6.3 with $k = 4$ we get

$$\begin{aligned} G_{1.2} &\leq 4\int_0^1\left(\mathbb{E}\Bigg[\left\|(1-\epsilon)\mathbf{D}^2g_i\Big(\widetilde{Y}(t_n)-\epsilon\big(Y_\theta(s)-\widetilde{Y}(t_n)\big)\Big)\right\|_{\mathbf{T_3}}^4\Bigg|\mathcal{F}_{t_n}\Bigg]\right)^{1/2}d\epsilon \\ &\quad\times\left(\mathbb{E}\Bigg[\left\|Y_\theta(s)-\widetilde{Y}(t_n)\right\|^8\Bigg|\mathcal{F}_{t_n}\Bigg]\right)^{1/2} \\ &\leq 4C_{\mathtt{TE}}(2,R)^{1/2}\,C_{\mathtt{SR}}(4,R)^{1/2}\,|s-t_n|^2. \end{aligned} \tag{7.12}$$

Substituting the bounds (7.11) and (7.12) back to (7.10) before bringing together the terms in (7.8), we have

$$\begin{aligned} \mathbb{E}\left[\left\|\Delta g_i\big(s,X(s),\widetilde{Y}(t_n)\big)\right\|^2\Bigg|\mathcal{F}_{t_n}\right] &\leq 2\mathbb{E}\Bigg[\left\|g_i(X(s))-g_i\big(Y_\theta(s)\big)\right\|^2\Bigg|\mathcal{F}_{t_n}\Bigg] \\ &\quad + C_G(R)|s-t_n|^2, \end{aligned}$$

where $C_G(R)$ is given in (7.7). By bounding $\|g_i\|^2$ with $\|g\|^2_{\mathbf{F}(d\times m)}$, the statement of Lemma 7.1 follows. □

The second lemma in the following gives the conditional second-moment bound of $E_\theta(s)$ as in (7.3), which is the first part of the one-step error in (7.1).

Lemma 7.2 *Let f, g satisfy Assumption 2.1 and 2.2. Let $X(s)$ be a solution of (1.1) and $\widetilde{E}(s)$ be given by (7.1) with $E_\theta(s)$ defined in (7.3), with $s \in [t_n, t_{n+1}]$, $n \in \mathbb{N}$. In this case there exists a constant C_E and an $\mathcal{F}_{t_n}$-measurable random variable $\overline{C}_M^{\{4(q+2)\}}$ such that*

$$\begin{aligned} \mathbb{E}\Bigg[\left\|E_\theta(t_{n+1})\right\|^2\Bigg|\mathcal{F}_{t_n}\Bigg] &\leq \left\|\widetilde{E}(t_n)\right\|^2 + C_E(R)\int_{t_n}^{t_{n+1}}\mathbb{E}\Bigg[\left\|E_\theta(r)\right\|^2\Bigg|\mathcal{F}_{t_n}\Bigg]dr \\ &\quad + \overline{C}_M^{\{4(q+2)\}}(R)\,h_{n+1}^3, \quad a.s. \end{aligned} \tag{7.13}$$

where

$$C_E(R) := 2K_1(R) + 2c, \tag{7.14}$$

with constant K_1 as defined in (7.40). The $\mathcal{F}_{t_n}$-measurable random variable $\overline{C}_M^{\{4(q+2)\}}$ is given by

$$\overline{C}_M^{\{4(q+2)\}}(R) := m^4 C_{Df}^2 C_{g_i}^2 + 2\overline{K}_2^{\{4(q+2)\}} + mC_G(R), \tag{7.15}$$

with the $\mathcal{F}_{t_n}$-measurable random variable $\overline{K}_2^{\{4(q+2)\}}$ in (7.41), constant C_G in Lemma 7.1. Denote $\mathbb{E}\left[\overline{C}_M^{\{4(q+2)\}}(R)\right] =: C_M(R)$, the finiteness of which is ensured in (7.44).

We recall that the superscript notation in (7.15) follows the convention introduced in the statement of Lemma 6.4 and indicates the number of finite moments required of the SDE solution.

Proof Throughout the proof, we restrict attention to trajectories on the event $\{h_{\min} < h_{n+1} \leq h_{\max}\}$, since by (7.3), $E_\theta(s)$ is only nonzero on this event, otherwise (7.13) holds trivially. Applying the stopping time variant of Itô formula (see Mao & Yuan [27]) to (7.3), we have,

$$\begin{aligned}\|E_\theta(t_{n+1})\|^2 &= \|\widetilde{E}(t_n)\|^2 + 2\int_{t_n}^{t_{n+1}} \underbrace{\left\langle E_\theta(r), \Delta f\left(X(r), \widetilde{Y}(t_n)\right)\right\rangle}_{=:J_f} dr \\ &+ \sum_{i=1}^m \int_{t_n}^{t_{n+1}} \Big\| \underbrace{\Delta g_i\left(r, X(r), \widetilde{Y}(t_n)\right)}_{=:J_{g_i}} \Big\|^2 dr + 2\sum_{i=1}^m \int_{t_n}^{t_{n+1}} \langle E_\theta(r), J_{g_i}\rangle dW_i(r).\end{aligned}$$

Take expectations on both sides conditional upon $\mathcal{F}_{t_n}$, and since $\int_{t_n}^{t_{n+1}} |J_f| dr$ has finite expectation (by the boundedness of $\widetilde{Y}(t_n)$ in (3.12) and the finiteness of absolute moments of $X(r)$ see (2.9)), using Fubini's Theorem (see for example [4, Proposition 12.10]) and (3.8) we have,

$$\begin{aligned}\mathbb{E}\left[\|E_\theta(t_{n+1})\|^2 \Big| \mathcal{F}_{t_n}\right] &= \|\widetilde{E}(t_n)\|^2 + 2\int_{t_n}^{t_{n+1}} \mathbb{E}\left[J_f \big| \mathcal{F}_{t_n}\right] dr \\ &+ \sum_{i=1}^m \int_{t_n}^{t_{n+1}} \mathbb{E}\left[\|J_{g_i}\|^2 \big| \mathcal{F}_{t_n}\right] dr,\end{aligned} \tag{7.16}$$

By Lemma 7.1, we have the bound of $\|J_{g_i}\|^2$ in (7.16) as

$$\begin{aligned}\mathbb{E}\left[\|J_{g_i}\|^2 \big| \mathcal{F}_{t_n}\right] \leq\ & 2\mathbb{E}\left[\|g(X(r)) - g(Y_\theta(r))\|^2_{\mathbf{F}(d\times m)} \Big| \mathcal{F}_{t_n}\right] \\ & + C_G(R)|r - t_n|^2.\end{aligned} \tag{7.17}$$

For J_f, by substituting Δf with (7.4) with adding in and subtracting out $f(Y_\theta(r))$, we have

$$J_f = \left\langle E_\theta(r), f(X(r)) - f(Y_\theta(r))\right\rangle + \underbrace{\left\langle E_\theta(r), f(Y_\theta(r)) - f\left(\widetilde{Y}(t_n)\right)\right\rangle}_{=:H}. \tag{7.18}$$

Substituting (7.18) and (7.17) back into (7.16), we have

$$\mathbb{E}\Big[\big\|E_\theta(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] \le \big\|\widetilde{E}(t_n)\big\|^2 + mC_G(R)h_{n+1}^3$$
$$+2\int_{t_n}^{t_{n+1}} \mathbb{E}\big[J_{f,g}\big|\mathcal{F}_{t_n}\big]dr + 2\int_{t_n}^{t_{n+1}} \mathbb{E}\big[H\big|\mathcal{F}_{t_n}\big]dr, \tag{7.19}$$

where

$$J_{f,g} := \big\langle E_\theta(r),\, f(X(r)) - f(Y_\theta(r))\big\rangle + \big\|g(X(r)) - g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}. \tag{7.20}$$

For H in (7.18), and in a similar way to (7.9), we expand $f(Y_\theta(r))$ using Taylor's theorem around $\widetilde{Y}(t_n)$ to have

$$f(Y_\theta(r)) - f\big(\widetilde{Y}(t_n)\big) = \mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big(Y_\theta(r) - \widetilde{Y}(t_n)\big)$$
$$+\int_0^1 (1-\epsilon)\mathbf{D}^2 f\Big(\widetilde{Y}(t_n) - \epsilon\cdot\big(Y_\theta(r) - \widetilde{Y}(t_n)\big)\Big)\Big[Y_\theta(r) - \widetilde{Y}(t_n)\Big]^2 d\epsilon. \tag{7.21}$$

Then we substitute $Y_\theta(r)$ in the first term on the RHS of (7.21) with (3.4) where we use the expanded form of the map as characterised in (2.12) for $s = r$. Therefore, for the last term on the RHS of (7.19), we have

$$\mathbb{E}\big[H\big|\mathcal{F}_{t_n}\big] \le H_1 + H_2 + H_3 + H_4 + H_5 + H_6, \tag{7.22}$$

where

$$H_1 := \mathbb{E}\Big[\big\langle E_\theta(r),\, \mathbf{D}f\big(\widetilde{Y}(t_n)\big)|r - t_n|f\big(\widetilde{Y}(t_n)\big)\big\rangle\Big|\mathcal{F}_{t_n}\Big];$$
$$H_2 := \mathbb{E}\Big[\Big\langle E_\theta(r),\, \underbrace{\sum_{i=1}^m \mathbf{D}f\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)I_i^{t_n,r}}_{=:H_{2R}}\Big\rangle\Big|\mathcal{F}_{t_n}\Big];$$
$$H_3 := \mathbb{E}\Big[\Big\langle E_\theta(r),\, \frac{1}{2}\sum_{i=1}^m \mathbf{D}f\big(\widetilde{Y}(t_n)\big)\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)$$
$$\times\Big(\big(I_i^{t_n,r}\big)^2 - |r - t_n|\Big)\Big\rangle\Big|\mathcal{F}_{t_n}\Big];$$
$$H_4 := \mathbb{E}\Big[\Big\langle E_\theta(r),\, \frac{1}{2}\sum_{\substack{i,j=1\\ i<j}}^m \mathbf{D}f\big(\widetilde{Y}(t_n)\big)\Big(\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big)$$
$$+\mathbf{D}g_j\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)\Big)I_i^{t_n,r}I_j^{t_n,r}\Big\rangle\Big|\mathcal{F}_{t_n}\Big];$$

$$H_5 := \mathbb{E}\Bigg[\Bigg\langle E_\theta(r), \sum_{\substack{i,j=1\\ i<j}}^{m} \mathbf{D}f\big(\widetilde{Y}(t_n)\big)\Big(\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)g_j\big(\widetilde{Y}(t_n)\big) - \mathbf{D}g_j\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)\Big)A_{ij}(t_n,r)\Bigg\rangle\Bigg|\mathcal{F}_{t_n}\Bigg];$$

$$H_6 := \mathbb{E}\Bigg[\Bigg\langle E_\theta(r), \int_0^1 (1-\epsilon)\mathbf{D}^2 f\Big(\widetilde{Y}(t_n) - \epsilon\cdot\big(Y_\theta(r) - \widetilde{Y}(t_n)\big)\Big) \times \Big[Y_\theta(r) - \widetilde{Y}(t_n)\Big]^2 d\epsilon\Bigg\rangle\Bigg|\mathcal{F}_{t_n}\Bigg].$$

We will now determine suitable upper bounds for each of H_1, H_2, H_3, H_4, H_5, and H_6 in turn. For H_1 in (7.22), by the Cauchy-Schwarz inequality, (2.1), and (6.1), we have

$$\begin{aligned} H_1 &\le \mathbb{E}\Big[\|E_\theta(r)\|\,\big\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}\,\big\|f\big(\widetilde{Y}(t_n)\big)\big\|\,|r-t_n|\,\Big|\mathcal{F}_{t_n}\Big] \\ &\le \mathbb{E}\left[\frac{1}{2}\big\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}^2\,\big\|f\big(\widetilde{Y}(t_n)\big)\big\|^2\|E_\theta(r)\|^2 + \frac{1}{2}|r-t_n|^2\,\Big|\mathcal{F}_{t_n}\right] \\ &\le \frac{1}{2}C_{Df}^2 C_f^2\ \mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big] + \frac{1}{2}|r-t_n|^2. \end{aligned} \tag{7.23}$$

Next, for the analysis of H_2 in (7.22), by (3.8), we firstly have

$$\mathbb{E}\big[H_{2R}\big|\mathcal{F}_{t_n}\big] = \sum_{i=1}^{m}\mathbf{D}f\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)\mathbb{E}\Big[I_i^{t_n,r}\Big|\mathcal{F}_{t_n}\Big] = 0. \tag{7.24}$$

By (2.3), the Cauchy-Schwarz inequality, (6.1) and (3.9) we also have

$$\begin{aligned} \mathbb{E}\Big[\big\|H_{2R}\big\|^2\Big|\mathcal{F}_{t_n}\Big] &\le m\sum_{i=1}^{m}\big\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}^2\big\|g_i\big(\widetilde{Y}(t_n)\big)\big\|^2\mathbb{E}\left[\big|I_i^{t_n,r}\big|^2\Big|\mathcal{F}_{t_n}\right] \\ &\le m^2 C_{Df}^2 C_{g_i}^2\,|r-t_n|. \end{aligned} \tag{7.25}$$

Then, for H_2 in (7.22) we firstly expand $E_\theta(r)$ using (7.3) to have

$$
\begin{aligned}
H_2 =& \mathbb{E}\left[\left\langle \widetilde{E}(t_n),\ H_{2R}\right\rangle \middle| \mathcal{F}_{t_n}\right] + \mathbb{E}\left[\left\langle \int_{t_n}^{r} \Delta f(X(p), Y(t_n))dp,\ H_{2R}\right\rangle \middle| \mathcal{F}_{t_n}\right] \\
&+ \mathbb{E}\left[\left\langle \sum_{i=1}^{m} \int_{t_n}^{r} \Delta g_i\big(p, X(p), \widetilde{Y}(t_n)\big) dW_i(p),\ H_{2R}\right\rangle \middle| \mathcal{F}_{t_n}\right] \\
=:& H_{2.1} + H_{2.2} + H_{2.3}.
\end{aligned} \tag{7.26}
$$

For $H_{2.1}$ in (7.26), by (7.24) we have

$$
H_{2.1} = \left\langle \widetilde{E}(t_n),\ \mathbb{E}\big[H_{2R}\big|\mathcal{F}_{t_n}\big]\right\rangle = 0. \tag{7.27}
$$

For $H_{2.2}$ in (7.26), by adding in and subtracting out $f(X(t_n))$ in Δf in (7.4):

$$
\begin{aligned}
H_{2.2} =& \ \mathbb{E}\left[\left\langle \int_{t_n}^{r} f(X(r)) - f(X(t_n))dp,\ H_{2R}\right\rangle \middle| \mathcal{F}_{t_n}\right] \\
&+ \mathbb{E}\left[\left\langle \int_{t_n}^{r} f(X(t_n)) - f\big(\widetilde{Y}(t_n)\big)dp,\ H_{2R}\right\rangle \middle| \mathcal{F}_{t_n}\right] \\
=:& \ H_{2.21} + H_{2.22}.
\end{aligned} \tag{7.28}
$$

Similar to $H_{2.1}$ in (7.27), we have $H_{2.22} = 0$. For $H_{2.21}$ in (7.28), using the Cauchy-Schwarz inequality and (7.25) we have

$$
\begin{aligned}
H_{2.21} \le& \ \mathbb{E}\left[\left\| \int_{t_n}^{r} f(X(p)) - f(X(t_n))dp \right\| \left\| H_{2R} \right\| \middle| \mathcal{F}_{t_n}\right] \\
\le& \left(|r - t_n| \int_{t_n}^{r} \mathbb{E}\left[\| f(X(p)) - f(X(t_n)) \|^2 \middle| \mathcal{F}_{t_n}\right] dp \ \mathbb{E}\left[\| H_{2R} \|^2 \middle| \mathcal{F}_{t_n}\right]\right)^{1/2} \\
\le& \ m C_{Df} C_{g_i} |r - t_n| \left(\int_{t_n}^{r} \mathbb{E}\left[\| f(X(p)) - f(X(t_n)) \|^2 \middle| \mathcal{F}_{t_n}\right] dp \right)^{1/2}.
\end{aligned} \tag{7.29}
$$

By Taylor expansion of $f(X(p))$ around $f(X(t_n))$ to first order, and using (2.6), the Cauchy-Schwarz inequality, Lemma 6.4 with $k = 2$ and (2.3):

$$
\begin{aligned}
&\mathbb{E}\left[\| f(X(p)) - f(X(t_n)) \|^2 \middle| \mathcal{F}_{t_n}\right] \\
&= \ \mathbb{E}\left[\left\| \int_0^1 \mathbf{D} f(X(t_n) - \epsilon \cdot (X(p) - X(t_n))(X(p) - X(t_n)) d\epsilon \right\|^2 \middle| \mathcal{F}_{t_n}\right] \\
&\le \left(\mathbb{E}\left[\| X(p) - X(t_n) \|^4 \middle| \mathcal{F}_{t_n}\right]\right)^{1/2} \\
&\qquad \times \left(\mathbb{E}\left[\left\| \int_0^1 \mathbf{D} f\big(X(t_n) - \epsilon \cdot (X(p) - X(t_n)\big) d\epsilon \right\|_{\mathbf{F}}^4 \middle| \mathcal{F}_{t_n}\right]\right)^{1/2} \\
&\le \ \overline{C}_{H2.21}^{\{4(q+2)\}} |p - t_n|,
\end{aligned} \tag{7.30}
$$

where

$$\overline{C}_{H2.21}^{\{4(q+2)\}} := \left(\overline{C}_{\mathrm{PR}}^{\{4(q+2)\}}\right)^{1/2} \times c_3^2\Big(1 + 3^{4q_1+4}\mathbb{E}\Big[\sup_{p\in[t_n,t_{n+1}]}\|X(p)\|^{4q_1+4}\Big|\mathcal{F}_{t_n}\Big]\Big)^{1/2}. \tag{7.31}$$

Substituting (7.30) back to (7.29) and using that $H_{2.22} = 0$, we have

$$H_{2.2} \le mC_{Df}C_{g_i}\left(\overline{C}_{H2.21}^{\{4(q+2)\}}\right)^{1/2}|r-t_n|^2. \tag{7.32}$$

For $H_{2.3}$ as in (7.26), using the Cauchy-Schwarz inequality, (2.3), (6.1), (3.9) and Itô's isometry we have

$$\begin{aligned} H_{2.3} &\le \left(\mathbb{E}\left[\left\|\sum_{i=1}^m\int_{t_n}^r \Delta g_i\big(p, X(p), \widetilde{Y}(t_n)\big)dW_i(p)\right\|^2\Big|\mathcal{F}_{t_n}\right]\right)^{1/2} \\ &\qquad\times\left(\mathbb{E}\left[\left\|\sum_{i=1}^m \mathbf{D}f\big(\widetilde{Y}(t_n)\big)g_i\big(\widetilde{Y}(t_n)\big)I_i^{t_n,r}\right\|^2\Big|\mathcal{F}_{t_n}\right]\right)^{1/2} \\ &\le \left(m\sum_{i=1}^m\int_{t_n}^r \mathbb{E}\left[\|\Delta g_i\big(p, X(p), \widetilde{Y}(t_n)\big)\|^2\Big|\mathcal{F}_{t_n}\right]dp\right)^{1/2} \\ &\qquad\times mC_{Df}C_{g_i}|r-t_n|^{1/2}. \end{aligned}$$

Then, by Lemma 7.1 we have

$$\begin{aligned} H_{2.3} \le \Bigg(&2m^2\int_{t_n}^r \mathbb{E}\Big[\|g(X(p)) - g(Y_\theta(p))\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big]dp \\ &+ C_G(R)|r-t_n|^3\Bigg)^{1/2} mC_{Df}C_{g_i}|r-t_n|^{1/2}. \end{aligned}$$

Since the integrand $\mathbb{E}\Big[\|g(X(p)) - g(Y_\theta(p))\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big]$ is non-negative for all $p \in [t_n, t_{n+1}]$, we can replace the upper limit of integration with t_{n+1}. With $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$, we have

$$\begin{aligned} H_{2.3} \le\ &\sqrt{2}m^2C_{Df}C_{g_i}|r-t_n|^{1/2} \\ &\times\left(\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\|g(X(r)) - g(Y_\theta(r))\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big]dr\right)^{1/2} \\ &+mC_{Df}C_{g_i}C_G(R)^{1/2}|r-t_n|^2. \end{aligned} \tag{7.33}$$

Notice that we changed the variable of integration from p back to r for consistency. Substituting (7.27), (7.32) and (7.33) back into (7.26), we have

$$\begin{aligned} H_2 \leq\ & mC_{Df}C_{g_i}\left(\left(\overline{C}_{H2.21}^{\{4(q+2)\}}\right)^{1/2} + C_G(R)^{1/2}\right)|r-t_n|^2 \\ & + \sqrt{2}m^2 C_{Df}C_{g_i}|r-t_n|^{1/2} \\ & \times \left(\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\big\|g(X(r)) - g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big]dr\right)^{1/2}. \end{aligned} \tag{7.34}$$

For H_3 in (7.22), by the Cauchy-Schwarz inequality, triangle inequality, (2.3), (2.1), (3.10), (2.6) and (6.1) we have

$$\begin{aligned} H_3 \leq\ & \mathbb{E}\Bigg[\frac{1}{4}\sum_{i=1}^{m}\Bigg(\big\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big\|^2_{\mathbf{F}}\,\big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\big\|^2_{\mathbf{F}}\,\big\|g_i\big(\widetilde{Y}(t_n)\big)\big\|^2\|E_\theta(r)\|^2 \\ & + 2\left|I_i^{t_n,r}\right|^4 + 2|r-t_n|^2\Bigg)\Bigg|\mathcal{F}_{t_n}\Bigg] \\ \leq\ & \frac{m}{4}C_{Df}^2C_{Dg_i}^2C_{g_i}^2\ \mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big] + \frac{(\boldsymbol{\gamma}_4+1)m}{2}|r-t_n|^2. \end{aligned} \tag{7.35}$$

For H_4 in (7.22), by the Cauchy-Schwarz inequality, conditional independence of the Itô integrals, (3.9), triangle inequality, (2.1), Itô's isometry, (2.6), and (6.1), we have

$$\begin{aligned} H_4 \leq\ & \mathbb{E}\Bigg[\frac{1}{2}\sum_{\substack{i,j=1\\ i<j}}^{m}\|E_\theta(r)\|\big\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}\Big(\big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}\big\|g_j\big(\widetilde{Y}(t_n)\big)\big\| \\ & + \big\|\mathbf{D}g_j\big(\widetilde{Y}(t_n)\big)\big\|_{\mathbf{F}}\big\|g_i\big(\widetilde{Y}(t_n)\big)\big\|\Big)\left|I_i^{t_n,r}\right|\left|I_j^{t_n,r}\right|\Bigg|\mathcal{F}_{t_n}\Bigg] \\ \leq\ & \frac{1}{4}m(m-1)C_{Df}^2C_{Dg_i}^2C_{g_i}^2\mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big] + \frac{1}{8}m(m-1)|r-t_n|^2. \end{aligned} \tag{7.36}$$

For H_5 in (7.22), by the Cauchy-Schwarz inequality, triangle inequality, (2.1), (6.1), (2.6), and Lemma 2.2 with $b=2$, we have

$$\begin{aligned} H_5 \leq\ & \mathbb{E}\Bigg[\frac{1}{2}\sum_{\substack{i,j=1\\ i<j}}^{m}\Bigg(\big\|\mathbf{D}f\big(\widetilde{Y}(t_n)\big)\big\|^2_{\mathbf{F}}\|E_\theta(r)\|^2\Big(\big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\big\|^2_{\mathbf{F}}\big\|g_j\big(\widetilde{Y}(t_n)\big)\big\|^2 \\ & + \big\|\mathbf{D}g_i\big(\widetilde{Y}(t_n)\big)\big\|^2_{\mathbf{F}}\big\|g_j\big(\widetilde{Y}(t_n)\big)\big\|^2\Big) + \left|A_{ij}(t_n,r)\right|^2\Bigg)\Bigg|\mathcal{F}_{t_n}\Bigg] \\ \leq\ & \frac{1}{2}m(m-1)C_{Df}^2C_{Dg_i}^2C_{g_i}^2\mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big] \\ & + \frac{1}{4}m(m-1)(C_{\mathrm{LA}}(2))^2|r-t_n|^2. \end{aligned} \tag{7.37}$$

For H_6 in (7.22), by the Cauchy-Schwarz inequality, triangle inequality, and (2.1) we have (noting that $\|[\cdot]^2\|_{\mathbf{F}} = \|\cdot\|^2$)

$$\begin{aligned} H_6 \leq & \mathbb{E}\Bigg[\|E_\theta(r)\| \left\|Y_\theta(r) - \widetilde{Y}(t_n)\right\|^2 \\ & \times \left\| \int_0^1 (1-\epsilon)\mathbf{D}^2 f\Big(\widetilde{Y}(t_n) - \epsilon \cdot \big(Y_\theta(r) - \widetilde{Y}(t_n)\big)\Big) d\epsilon \right\|_{\mathbf{T}_3} \Bigg| \mathcal{F}_{t_n} \Bigg] \\ \leq & \frac{1}{2}\mathbb{E}\Big[\|E_\theta(r)\|^2 \Big| \mathcal{F}_{t_n}\Big] + \frac{1}{2} \underbrace{\sqrt{\mathbb{E}\left[\left\|Y_\theta(r) - \widetilde{Y}(t_n)\right\|^8 \Big| \mathcal{F}_{t_n}\right]}}_{H_{6.1}} \\ & \times \underbrace{\sqrt{\mathbb{E}\left[\left\| \int_0^1 (1-\epsilon)\mathbf{D}^2 f\Big(\widetilde{Y}(t_n) - \epsilon \cdot \big(Y_\theta(r) - \widetilde{Y}(t_n)\big)\Big) d\epsilon \right\|_{\mathbf{T}_3}^4 \Bigg| \mathcal{F}_{t_n}\right]}}_{H_{6.2}}. \end{aligned}$$

From (6.6) in Lemma 6.3 with $k = 4$, we have $H_{6.1} \leq C_{\mathrm{SR}}(4, R)^{1/2} |r - t_n|^2$. From (6.11) in Lemma 6.5 with $k = 2$, we have $H_{6.2} \leq C_{\mathrm{TE}}(2, R)^{1/2}$. Therefore, H_6 in (7.22) becomes

$$H_6 \leq \frac{1}{2}\mathbb{E}\Big[\|E_\theta(r)\|^2 \Big| \mathcal{F}_{t_n}\Big] + \frac{1}{2} C_{\mathrm{SR}}(4, R)^{1/2}\, C_{\mathrm{TE}}(2, R)^{1/2}\, |r - t_n|^2. \quad (7.38)$$

Substituting (7.23), (7.34), (7.35), (7.36), (7.37) and (7.38) back into (7.22) for H, we have

$$\begin{aligned} \mathbb{E}[H|\mathcal{F}_{t_n}] \leq & K_1(R)\mathbb{E}\Big[\|E_\theta(r)\|^2 \Big| \mathcal{F}_{t_n}\Big] + \overline{K}_2^{\{4(q+2)\}}(R)|r - t_n|^2 \\ & + \sqrt{2} m^2 C_{Df} C_{g_i} |r - t_n|^{1/2} \\ & \times \left(\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\left\|g(X(r)) - g(Y_\theta(r))\right\|_{\mathbf{F}(d\times m)}^2 \Big| \mathcal{F}_{t_n}\Big] dr \right)^{1/2}, \end{aligned} \quad (7.39)$$

where

$$K_1(R) := \frac{1}{2} + \frac{1}{2} C_{Df}^2 C_f^2 + m(m-1) C_{Df}^2 C_{Dg_i}^2 C_{g_i}^2, \quad (7.40)$$

and with $\overline{C}_{H2.21}^{\{4(q+2)\}}$ from (7.31)

$$\begin{aligned} & \overline{K}_2^{\{4(q+2)\}}(R) \\ & := \frac{1}{2} + m C_{Df} C_{g_i} \left(\left(\overline{C}_{H2.21}^{\{4(q+2)\}}\right)^{1/2} + \frac{1}{2}(\boldsymbol{\gamma}_4 + 1) m \right. \\ & \left. + \frac{1}{4} m(m-1)\big(1 + \big(C_{\mathrm{LA}}(2)\big)^2\big) + \frac{1}{2} C_{\mathrm{SR}}(4, R)^{1/2}\, C_{\mathrm{TE}}(2, R)^{1/2}. \right. \end{aligned} \quad (7.41)$$

Substituting $\mathbb{E}[H|\mathcal{F}_{t_n}]$ from (7.39) back into (7.19), we have

$$\begin{aligned}\mathbb{E}\Big[\big\| E_\theta(t_{n+1}) \big\|^2\Big|\mathcal{F}_{t_n}\Big] \le \big\|\widetilde{E}(t_n)\big\|^2 &+ 2K_1(R)\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big]dr \\ &+ mC_G(R)h_{n+1}^3 + \overline{K}_2^{\{4(q+2)\}}(R)h_{n+1}^3 \\ &+ 2\int_{t_n}^{t_{n+1}} \mathbb{E}\big[J_{f,g}\big|\mathcal{F}_{t_n}\big]dr + \sqrt{2}m^2 C_{Df}C_{g_i}h_{n+1}^{3/2} \\ &\times\left(\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\big\|g(X(r)) - g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big]dr\right)^{1/2}. \quad (7.42)\end{aligned}$$

Using (2.1) on the last term on the RHS of (7.42), we have

$$\begin{aligned}\mathbb{E}\Big[\big\| E_\theta\ (t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] \le\ \big\|\widetilde{E}(t_n)\big\|^2 &+ 2K_1(R)\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big]dr \\ &+ \overline{C}_M^{\{4(q+2)\}}(R)\, h_{n+1}^3 \\ + 2\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[J_{f,g} &+ \frac{1}{2}\big\|g(X(r)) - g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}\Big|\mathcal{F}_{t_n}\Big]dr, \quad (7.43)\end{aligned}$$

where $\overline{C}_M^{\{4(q+2)\}}$ is as defined in (7.15). Recall $J_{f,g}$ is given in (7.20) so that

$$\begin{aligned}&J_{f,g} + \frac{1}{2}\big\|g(X(r)) - g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)} \\ &= \Big\langle E_\theta(r),\, f(X(r)) - f(Y_\theta(r))\Big\rangle + \frac{3}{2}\big\|g(X(r)) - g(Y_\theta(r))\big\|^2_{\mathbf{F}(d\times m)}.\end{aligned}$$

By Assumption 2.2 we can apply the monotone condition (2.5):

$$\begin{aligned}\mathbb{E}\Big[\big\| E_\theta(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big] \le\ \big\|\widetilde{E}(t_n)\big\|^2 &+ C_E(R)\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\|E_\theta(r)\|^2\Big|\mathcal{F}_{t_n}\Big]dr \\ &+ \overline{C}_M^{\{4(q+2)\}}(R)\, h_{n+1}^3,\end{aligned}$$

where $C_E(R)$ is in (7.14).

To obtain the the final estimate on $C_M(R)$ in the Lemma we use the explicit form of $\overline{C}_M^{\{4(q+2)\}}$, $\overline{K}_2^{\{4(q+2)\}}$, $\overline{C}_{H2.21}^{\{4(q+2)\}}$, given by (7.15), (7.41), and (7.31) respectively, (6.10) in the statement of Lemma 6.5, (2.9), and Assumption 2.2 we bound the expectation of $\overline{C}_M^{\{4(q+2)\}}$ as follows,

$$\begin{aligned}C_M(R) &:= \mathbb{E}\Big[\overline{C}_M^{\{4(q+2)\}}(R)\Big] \\ &\le m^4 C_{Df}^2 C_{g_i}^2 + 2mC_{Df}C_{g_i}\Big(c_3 C_{\mathrm{PR}}\,(2)^{1/4}\left(1 + 3^{q_1+1}C_{\mathrm{X}}\right) + C_G(R)^{1/2}\Big)\end{aligned}$$

$$+\frac{1}{2}(\boldsymbol{\gamma}_4+1)m+\frac{1}{2}m(m-1)\big(1+(C_{\mathrm{LA}}(2))^2\big)$$
$$+C_{\mathrm{SR}}(4,R)^{1/2}C_{\mathrm{TE}}(2,R)^{1/2}+mC_G(R)+1. \tag{7.44}$$

□

7.2 Proof of Theorem 4.1 on strong convergence

Proof Firstly, by (7.1) we have the conditional second-moment bound of the one-step error as

$$\mathbb{E}\Big[\big\|\widetilde{E}(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big]=\mathbb{E}\Big[\big\|E_\theta(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big]+\mathbb{E}\Big[\big\|E_\varphi(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big], \tag{7.45}$$

where by (3.6) and (7.2), the one-step error bound of the backstop map yields

$$\mathbb{E}\Big[\big\|E_\varphi(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big]\le\big\|\widetilde{E}(t_n)\big\|^2+C_{B_1}\int_{t_n}^{t_{n+1}}\mathbb{E}\Big[\|E_\varphi(r)\|^2\Big|\mathcal{F}_{t_n}\Big]dr$$
$$+C_{B_2}h_{n+1}^3,\quad a.s. \tag{7.46}$$

Therefore, by substituting (7.13) and (7.46) into (7.45), and recalling (7.1) we have for any h_{n+1} that satisfies Assumption 3.1,

$$\mathbb{E}\Big[\big\|\widetilde{E}(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big]\le\big\|\widetilde{E}(t_n)\big\|^2+\Gamma_1(R)\int_{t_n}^{t_{n+1}}\mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_n}\Big]dr$$
$$+\overline{\Gamma}_2^{\{4(q+2)\}}\big(R\big)h_{n+1}^3,\quad a.s. \tag{7.47}$$

where we define Γ_1, $\overline{\Gamma}_2$ and by (7.44) its expected form as

$$\Gamma_1(R):=C_E(R)+C_{B_1};$$
$$\overline{\Gamma}_2^{\{4(q+2)\}}\big(R\big):=\overline{C}_M^{\{4(q+2)\}}(R)+C_{B_2};$$
$$\Gamma_2(R):=\mathbb{E}\Big[\overline{\Gamma}_2^{\{4(q+2)\}}\big(R\big)\Big]\le C_M(R)+C_{B_2}. \tag{7.48}$$

For a fixed $t>0$, let $N^{(t)}$ be as in Definition 3.2, we multiply both sides of (7.47) with the indicator function $\mathbf{1}_{\{N^{(t)}>n+1\}}$ and sum up the steps excluding the last step $N^{(t)}$ to have

$$\sum_{n=0}^{N^{(t)}-2}\mathbb{E}\Big[\big\|\widetilde{E}(t_{n+1})\big\|^2\Big|\mathcal{F}_{t_n}\Big]\mathbf{1}_{\{N^{(t)}>n+1\}}\le\sum_{n=0}^{N^{(t)}-2}\big\|\widetilde{E}(t_n)\big\|^2\mathbf{1}_{\{N^{(t)}>n+1\}}$$
$$+\Gamma_1(R)\sum_{n=0}^{N^{(t)}-2}\int_{t_n}^{t_{n+1}}\mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_n}\Big]\mathbf{1}_{\{N^{(t)}>n+1\}}dr$$
$$+\overline{\Gamma}_2^{\{4(q+2)\}}(R)\sum_{n=0}^{N^{(t)}-2}h_{n+1}^3\mathbf{1}_{\{N^{(t)}>n+1\}}. \tag{7.49}$$

Since $t \in \left[t_{N^{(t)}-1}, t_{N^{(t)}}\right]$, we use (7.47) to express the last step, noting that it holds when t_n, t_{n+1} are replaced by $t_{N^{(t)}-1}$ and t respectively:

$$\mathbb{E}\Big[\|\widetilde{E}(t)\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big] \le \|\widetilde{E}(t_{N^{(t)}-1})\|^2 + \Gamma_1(R)\int_{t_{N^{(t)}-1}}^{t} \mathbb{E}\Big[\|\widetilde{E}(r)\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big]dr + \overline{\Gamma}_2^{\{4(q+2)\}}(R)\big|t - t_{N^{(t)}-1}\big|^3. \tag{7.50}$$

By adding the both sides of (7.49) and (7.50), and taking an expectation:

$$\begin{aligned}
&\mathbb{E}\Bigg[\sum_{n=0}^{N^{(t)}-2}\Big(\mathbb{E}\Big[\|\widetilde{E}(t_{n+1})\|^2\Big|\mathcal{F}_{t_n}\Big] - \|\widetilde{E}(t_n)\|^2\Big)\mathbf{1}_{\{N^{(t)}>n+1\}} \\
&\qquad + \mathbb{E}\Big[\|\widetilde{E}(t)\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big] - \|\widetilde{E}(t_{N^{(t)}-1})\|^2\Bigg] =: \mathrm{LHS} \\
&\le \Gamma_1(R)\mathbb{E}\Bigg[\sum_{n=0}^{N^{(t)}-2}\int_{t_n}^{t_{n+1}} \mathbb{E}\Big[\|\widetilde{E}(r)\|^2\Big|\mathcal{F}_{t_n}\Big]\mathbf{1}_{\{N^{(t)}>n+1\}}dr \\
&\qquad + \int_{t_{N^{(t)}-1}}^{t} \mathbb{E}\Big[\|\widetilde{E}(r)\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big]dr\Bigg] =: \mathrm{R}_1 \\
&\quad + \mathbb{E}\Bigg[\overline{\Gamma}_2^{\{4(q+2)\}}(R)\Bigg(\sum_{n=0}^{N^{(t)}-2} h_{n+1}^3\mathbf{1}_{\{N^{(t)}>n+1\}} + \big|t - t_{N^{(t)}-1}\big|^3\Bigg)\Bigg] =: \mathrm{R}_2
\end{aligned} \tag{7.51}$$

where we analyse (7.51) (LHS $\le \mathrm{R}_1 + \mathrm{R}_2$) below. For the LHS in (7.51), $N^{(t)}$ is a random number taking value from $N_{\min}^{(t)}$ to $N_{\max}^{(t)}$, and $\mathbf{1}_{\{N^{(t)}>n+1\}}$ is an $\mathcal{F}_{t_n}$-measurable random variable. Therefore it is useful decompose the range of n into three parts on each trajectory. First, when $n < N^{(t)} - 1$, then $1_{\{N^{(t)}>n+1\}} = 1_{\{N^{(t)}>n\}} = 1$. Second, when $n = N^{(t)} - 1$, then $1_{\{N^{(t)}>n+1\}} = 0$ and $1_{\{N^{(t)}>n\}} = 1$. Finally, when $n > N^{(t)} - 1$, then $1_{\{N^{(t)}>n+1\}} = 1_{\{N^{(t)}>n\}} = 0$. Hence we obtain a telescoping sum with the appropriate cancellation that terminates at $\mathbb{E}\big[\|\widetilde{E}(t_{N^{(t)}-1})\|^2\, 1_{\{N^{(t)}>N^{(t)}-1\}}\big] = \mathbb{E}\big[\|\widetilde{E}(t_{N^{(t)}-1})\|^2\big]$. Applying this with the tower property for conditional expectations, and using the fact that $\|\widetilde{E}(t_0)\|^2 = 0$, we have

$$\begin{aligned}
\mathrm{LHS} &= \sum_{n=0}^{N_{\max}^{(t)}-2} \mathbb{E}\Big[\|\widetilde{E}(t_{n+1})\|^2\mathbf{1}_{\{N^{(t)}>n+1\}} - \|\widetilde{E}(t_n)\|^2\mathbf{1}_{\{N^{(t)}>n+1\}}\Big] \\
&\qquad + \mathbb{E}\Big[\mathbb{E}\Big[\|\widetilde{E}(t)\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big] - \|\widetilde{E}(t_{N^{(t)}-1})\|^2\Big] \\
&= \mathbb{E}\Big[\|\widetilde{E}(t_{N^{(t)}-1})\|^2\Big] - \mathbb{E}\Big[\|\widetilde{E}(t_0)\|^2\Big] + \mathbb{E}\big[\|\widetilde{E}(t)\|^2\big] - \mathbb{E}\Big[\|\widetilde{E}(t_{N^{(t)}-1})\|^2\Big] \\
&= \mathbb{E}\Big[\|\widetilde{E}(t)\|^2\Big].
\end{aligned} \tag{7.52}$$

Consider the term R_1 on the RHS of (7.51). By Definition 3.2 we have each $n = N^{(r)} - 1$ for $r \in [t_n, t_{n+1}]$. So we restate $\mathcal{F}_{t_n}$ as $\mathcal{F}_{t_{N^{(r)}-1}}$, and the indicator function as $\mathbf{1}_{\{N^{(t)} > N^{(r)}\}}$. Summing up all the steps results in an integral from 0 to $t_{N^{(t)}-1}$ that

$$\begin{aligned}\mathrm{R}_1 = \; & \Gamma_1(R)\mathbb{E}\Bigg[\int_0^{t_{N^{(t)}-1}} \mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2 \mathbf{1}_{\{N^{(t)} > N^{(r)}\}}\Big|\mathcal{F}_{t_{N^{(r)}-1}}\Big]dr \\ & + \int_{t_{N^{(t)}-1}}^{t} \mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_{N^{(t)}-1}}\Big]dr\Bigg] \\ \leq \; & \Gamma_1(R)\int_0^t \mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big]dr. \end{aligned} \tag{7.53}$$

For R_2 in (7.51), by (7.48), Definition 3.2 and $\rho h_{\min} = h_{\max}$ we have

$$\mathrm{R}_2 \leq \Gamma_2(R) N_{\max}^{(t)} h_{\max}^3 \leq \Gamma_2(R)\,(\rho t + 1)\, h_{\max}^2. \tag{7.54}$$

We see that $4(q+2)$ is the minimum number of finite SDE moments required for a finite R_2, and this is guaranteed by Assumption 2.2. Combining (7.52), (7.53) and (7.54) back into (7.51), for all $t \in [0, T]$ we have

$$\mathbb{E}\Big[\big\|\widetilde{E}(t)\big\|^2\Big] \;\leq\; \Gamma_1(R)\int_0^t \mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big]dr + \Gamma_2(R)\,(\rho t + 1)\, h_{\max}^2.$$

By Gronwall's inequality (see [25, Thm. 8.1]), we have for all $t \in [0, T]$

$$\Big(\mathbb{E}\Big[\big\|\widetilde{E}(t)\big\|^2\Big]\Big)^{\frac{1}{2}} \leq C(R, \rho, t)\, h_{\max}. \tag{7.55}$$

Taking the maximum over t on the both sides, the proof follows with

$$C(R, \rho, t) := \sqrt{\big(C_M(R) + C_{B_2}\big)\,(\rho t + 1)\exp\Big(t\big(C_E(R) + C_{B_1}\big)\Big)}.$$

□

7.3 Proof of Theorem 4.2 on the probability of using the backstop

Proof By (3.11) and by the Markov inequality we have

$$\mathbb{P}\big[h_{n+1} = h_{\min}\big] = \mathbb{P}\left[\frac{h_{\max}}{\big\|\widetilde{Y}(t_n)\big\|^{1/\kappa}} \leq h_{\min}\right] \leq \frac{\mathbb{E}\Big[\big\|\widetilde{Y}(t_n)\big\|^2\Big]}{\rho^{2\kappa}}. \tag{7.56}$$

By adding in and subtracting out $X(t_n)$ together with the tower property of conditional expectation, (2.3), (7.1) and (2.9) we have

$$\begin{aligned}\mathbb{E}\Big[\big\|\widetilde{Y}(t_n)\big\|^2\Big] &\le 2\mathbb{E}\Big[\big\|X(t_n)-\widetilde{Y}(t_n)\big\|^2\Big]+2\mathbb{E}\big[\|X(t_n)\|^2\big]\\ &\le 2\mathbb{E}\Big[\mathbb{E}\Big[\big\|X(t_n)-\widetilde{Y}(t_n)\big\|^2\Big|\mathcal{F}_{t_{n-1}}\Big]\Big]+2\mathbb{E}\Big[\sup_{t_n\in[0,T]}\|X(t_n)\|^2\Big]\\ &\le 2\mathbb{E}\Big[\mathbb{E}\Big[\big\|\widetilde{E}(t_n)\big\|^2\Big|\mathcal{F}_{t_{n-1}}\Big]\Big]+2C_{\mathrm{X}}. \end{aligned}\tag{7.57}$$

Next, we repeatedly substitute (7.47) for decreasing values of n into the RHS of (7.57) until $n=0$. Then with tower property, Definition 3.2, (3.11) and (7.48) we have

$$\begin{aligned}\mathbb{E}\Big[\big\|\widetilde{Y}(t_n)\big\|^2\Big] &\le 2\mathbb{E}\Big[\big\|\widetilde{E}(t_{n-1})\big\|^2\Big]+2\Gamma_1(R)\mathbb{E}\Big[\int_{t_{n-1}}^{t_n}\mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_{n-1}}\Big]dr\Big]\\ &\quad+2\mathbb{E}\Big[\overline{\Gamma}_2^{\{4(q+2)\}}(R)h_n^3\Big]+2C_{\mathrm{X}}\\ &\le 2\mathbb{E}\Big[\big\|\widetilde{E}(t_0)\big\|^2\Big]+2\Gamma_1(R)\mathbb{E}\Big[\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_{j-1}}\Big]dr\Big]\\ &\quad+2N_{\max}^{(T)}\Gamma_2(R)h_{\max}^3+2C_{\mathrm{X}}\\ &\le 2\Gamma_1(R)\mathbb{E}\Big[\int_0^{t_n}\mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_{N(r)-1}}\Big]dr\Big]\\ &\quad+2\,(\rho\,T+1)\,\Gamma_2(R)h_{\max}^3+2C_{\mathrm{X}}. \end{aligned}\tag{7.58}$$

Since the integrand $\mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_{N(r)-1}}\Big]$ in the second term on the RHS of (7.58) is almost surely non-negative for all $r\in[0,T]$, we can replace the upper limit of integration with T. Using $\widetilde{E}(t_0)=0$, (3.11), the tower property of conditional expectation, and (7.55) from Theorem 4.1, we have

$$\begin{aligned}&\mathbb{E}\Big[\int_0^{t_n}\mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big|\mathcal{F}_{t_{N(r)-1}}\Big]dr\Big]\\ &\le\int_0^T\mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big]dr\le T\max_{r\in[0,T]}\mathbb{E}\Big[\big\|\widetilde{E}(r)\big\|^2\Big]\le T\,C^2(R,\rho,T)h_{\max}^2. \end{aligned}\tag{7.59}$$

By choosing $h_{\max}\le 1/C(R,\rho,T)$, we substitute (7.59) into (7.58) and then (7.56) to get

$$\mathbb{P}\big[h_{n+1}=h_{\min}\big]\le\frac{2\Big(\Gamma_1(R)+(T+1)\,\Gamma_2(R)h_{\max}^2+C_{\mathrm{X}}\Big)}{\rho^{2\kappa-1}}=:\frac{C_{\mathrm{prob}}}{\rho^{2\kappa-1}},\tag{7.60}$$

and the rest of the proof follows. □

Decalartions

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix A: Proof of Lemma 2.2 (Lévy Area)

Proof Set $\mathtt{i}^2 = -1$. Since the pair of Wiener processes $(W_i(r), W_j(r))^T$, $r \in [t_n, s]$, are mutually independent, by [22, Eq. (1.3.5)] the characteristic function of the Lévy area (2.13) is given by $\phi(\lambda) = (\cosh\left(\frac{1}{2}|s - t_n|\lambda\right))^{-1}$. This was applied in the context of numerical methods for SDEs in [24]. The Taylor expansion of the function $\cosh\left(\frac{1}{2}|s - t_n|\lambda\right)$ around 0 gives

$$\phi(\lambda) = \sum_{N=0}^{\infty} \frac{\mathbf{E}_{2N}}{(2N)!} \left(\frac{1}{2}|s - t_n|\right)^{2N} \lambda^{2N}, \quad \left|\frac{1}{2}|s - t_n|\lambda\right| < \frac{\pi}{2},$$

where $\mathbf{E}_{2N}$ stands for the $2N^{\text{th}}$ Euler number, which may be expressed as

$$\mathbf{E}_{2N} = \mathtt{i} \sum_{b=1}^{2N+1} \sum_{j=0}^{b} \binom{j}{b} \frac{(-1)^j (b - 2j)^{2N+1}}{2^b\, \mathtt{i}^b\, b}, \quad N = 0, 1, 2, 3, \ldots.$$

All odd Euler numbers are zero. The kth derivative of the characteristic function with respect to λ is

$$\phi(\lambda)_{\lambda}^{(k)} = \sum_{N=\lceil \frac{k}{2} \rceil}^{\infty} \left(\prod_{B=0}^{k-1} (2N - B)\right) \frac{\mathbf{E}_{2N}}{(2N)!} \left(\frac{1}{2}|s - t_n|\right)^{2N} \lambda^{2N-k}.$$

As $\lambda \to 0$, since all terms vanish unless $k = 2N$, we have

$$\lim_{\lambda \to 0} \phi(\lambda)_{\lambda}^{(k)} = \begin{cases} \left(\prod_{B=0}^{k-1} (k - B)\right) \frac{\mathbf{E}_k}{(k)!} \left(\frac{1}{2}|s - t_n|\right)^{k}, & k \text{ even;} \\ 0, & k \text{ odd.} \end{cases}$$

In the calculation of expectations, we make use of the mutual independence, conditional upon $\mathcal{F}_{t_n}$, of the pair of Brownian increments $(W_i(t), W_j(t))^T$. Therefore, the

k^{th} conditional moment of $A_{ij}^{t_n,s}$ is

$$\mathbb{E}\left[\left(A_{ij}^{t_n,s}\right)^k \middle| \mathcal{F}_{t_n}\right] = L_k \left|s - t_n\right|^k,$$

where for all $a = 1, 2, 3, \ldots$

$$\begin{aligned} L_k &= \left(\prod_{B=0}^{k-1}(k-B)\right) \frac{\mathbf{E}_k}{(k)!}\left(-\frac{1}{2}\,\mathtt{i}\right)^k \\ &:= \begin{cases} \left(\prod_{B=0}^{k-1}(k-B)\right) \frac{\mathbf{E}_k}{(k)!}\left(\frac{1}{2}\right)^k, & k = 4a = 4, 8, 12, \ldots \\ -\left(\prod_{B=0}^{k-1}(k-B)\right) \frac{\mathbf{E}_k}{(k)!}\left(\frac{1}{2}\right)^k, & k = 4a - 2 = 2, 6, 10 \ldots \\ 0, & k = 2a - 1 = 1, 3, 5, \ldots \end{cases} \end{aligned}$$

which is finite, as a finite product of finite factors. When k is even, we have

$$\mathbb{E}\left[\left|A_{ij}^{t_n,s}\right|^k \middle| \mathcal{F}_{t_n}\right] = \mathbb{E}\left[\left(A_{ij}^{t_n,s}\right)^k \middle| \mathcal{F}_{t_n}\right] = L_k \left|s - t_n\right|^k, \quad a.s.$$

When k is odd, i.e. $k = 2c + 1$ for all $c = 0, 1, 2, \ldots$, we have a.s.

$$\begin{aligned} \mathbb{E}\left[\left|A_{ij}^{t_n,s}\right|^k \middle| \mathcal{F}_{t_n}\right] &= \mathbb{E}\left[\left|A_{ij}^{t_n,s}\right|^{2c+1} \middle| \mathcal{F}_{t_n}\right] \\ &\leq \sqrt{\mathbb{E}\left[\left(A_{ij}^{t_n,s}\right)^{4c} \middle| \mathcal{F}_{t_n}\right] \mathbb{E}\left[\left(A_{ij}^{t_n,s}\right)^{2} \middle| \mathcal{F}_{t_n}\right]} \\ &= \begin{cases} \sqrt{L_2}\,\left|s - t_n\right|, & c = 0; \\ \sqrt{L_{4c} \cdot L_2}\left|s - t_n\right|^{2c+1}, & c = 1, 2, 3, \ldots \end{cases} \\ &= \begin{cases} \sqrt{L_2}\,\left|s - t_n\right|, & k = 1; \\ \sqrt{L_{2k-2} \cdot L_2}\left|s - t_n\right|^{k}, & k = 3, 5, 7, \ldots. \end{cases} \end{aligned}$$

Therefore, in conclusion we have

$$\mathbb{E}\left[\left|A_{ij}^{t_n,s}\right|^k \middle| \mathcal{F}_{t_n}\right] \leq C_{\mathrm{LA}}(k)\left|s - t_n\right|^k, \quad a.s.$$

where

$$C_{\mathrm{LA}}(k) = \begin{cases} \sqrt{L_2}, & k = 1; \\ \sqrt{L_{2k-2} \cdot L_2}, & k = 3, 5, 7, \ldots; \\ L_k, & k = 2, 4, 6, \ldots \end{cases} \tag{A.1}$$

□

References

1. Beyn, W.J., Isaak, E., Kruse, R.: Stochastic C-stability and B-consistency of explicit and implicit Milstein-type schemes. J. Sci. Comput. **70**(3), 1042–1077 (2017)
2. Burrage, P.M., Herdiana, R., Burrage, K.: Adaptive stepsize based on control theory for stochastic differential equations. J. Comput. Appl. Math. **171**(1–2), 317–336 (2004). https://doi.org/10.1016/j.cam.2004.01.027
3. Campbell, S., Lord, G.: Adaptive time-stepping for stochastic partial differential equations with non-Lipschitz drift. arXiv preprint arXiv:1812.09036 (2018)
4. Dineen, S.: Probability Theory in Finance: a mathematical guide to the Black–Scholes Formula. Graduate studies in mathematics; v.70. American Mathematical Society, Universities Press (2011)
5. Fang, W., Giles, M.B.: Adaptive Euler–Maruyama method for SDEs with non-globally Lipschitz drift. In: International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, pp. 217–234. Springer (2016)
6. Fang, W., Giles, M.B.: Adaptive Euler–Maruyama method for SDEs with nonglobally Lipschitz drift. Ann. Appl. Probab. **30**(2), 526–560 (2020)
7. Gaines, J., Lyons, T.: Variable step size control in the numerical solution of stochastic differential equations. SIAM J. Appl. Math. **57**(5), 1455–1484 (1997). https://doi.org/10.1137/S0036139995286515
8. Gan, S., He, Y., Wang, X.: Tamed Runge–Kutta methods for SDEs with super-linearly growing drift and diffusion coefficients. Appl. Numer. Math. **152**, 379–402 (2020)
9. Grasman, J., Salomons, H., Verhulst, S.: Stochastic modeling of length-dependent telomere shortening in Corvus monedula. J. Theor. Biol. **282**(1), 1–6 (2011)
10. Guo, Q., Liu, W., Mao, X., Yue, R.: The truncated Milstein method for stochastic differential equations with commutative noise. J. Comput. Appl. Math. **338**, 298–310 (2018). https://doi.org/10.1016/j.cam.2018.01.014
11. Hasminskii, R.: Stochastic Stability of Differential Equations. Sijthoff & Noordhoff (1980)
12. Higham, D.J., Kloeden, P.E.: Maple and MATLAB for stochastic differential equations in finance. In: Programming Languages and Systems in Computational Economics and Finance, pp. 233–269. Springer (2002)
13. Higham, D.J., Mao, X., Szpruch, L.: Convergence, non-negativity and stability of a new Milstein scheme with applications to finance. Discrete and Continuous Dynamical Systems B, pp. 2083–2100 (2013)
14. Hofmann, N., Müller-Gronbach, T., Ritter, K.: The optimal discretization of stochastic differential equations. J. Complex. **17**(1), 117–153 (2001). https://doi.org/10.1006/jcom.2000.0570
15. Hutzenthaler, M., Jentzen, A., Kloeden, P.E.: Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. Proc. R. Soc. Lond. A Math. Phys. Eng. Sci. **467**(2130), 1563–1576 (2011)
16. Ilie, S., Jackson, K.R., Enright, W.H.: Adaptive time-stepping for the strong numerical solution of stochastic differential equations. Numer. Algorithms **68**(4), 791–812 (2015). https://doi.org/10.1007/s11075-014-9872-6
17. Kelly, C., Lord, G.J.: Adaptive time-stepping strategies for nonlinear stochastic systems. IMA J. Numer. Anal. **38**(3), 1523–1549 (2018)
18. Kelly, C., Lord, G.J.: Adaptive Euler methods for stochastic systems with non-globally Lipschitz coefficients. Numerical Algorithms pp. 1–27 (2021)
19. Kloeden, P., Platen, E.: Numerical methods for stochastic differential equations. Stoch. Hydrol. Hydraul. **5**(2), 172–172 (1991)
20. Kumar, C., Sabanis, S.: On Milstein approximations with varying coefficients: the case of super-linear diffusion coefficients. BIT Numer. Math. **59**(4), 929–968 (2019). https://doi.org/10.1007/s10543-019-00756-5
21. Lamba, H., Mattingly, J.C., Stuart, A.M.: An adaptive Euler-Maruyama scheme for SDEs: convergence and stability. IMA J. Numer. Anal. **27**(3), 479–506 (2007). https://doi.org/10.1093/imanum/drl032
22. Lévy, P.: Wiener's random function, and other Laplacian random functions. In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. The Regents of the University of California (1951)
23. Lord, G.J., Powell, C.E., Shardlow, T.: An Introduction to Computational Stochastic PDEs. Cambridge University Press, Cambridge (2014)
24. Malham, S., Wiese, A.: Efficient almost-exact Lévy area sampling. Stat. Probab. Lett. **88**, 50–55 (2014)

25. Mao, X.: Stochastic Differential Equations and Applications, 2nd edn. Woodhead Publishing, Cambridge (2007)
26. Mao, X.: The truncated Euler-Maruyama method for stochastic differential equations. J. Comput. Appl. Math. **290**, 370–384 (2015)
27. Mao, X., Yuan, C.: Stochastic Differential Equations with Markovian Switching. Imperial College Press, London (2006)
28. Mauthner, S.: Step size control in the numerical solution of stochastic differential equations. J. Comput. Appl. Math. **100**(1), 93–109 (1998). https://doi.org/10.1016/S0377-0427(98)00139-3
29. Papoulis, A., Pillai, S.U.: Probability, Random Variables, and Stochastic Processes, 4th edn. McGraw-Hill, New York (2002)
30. Reisinger, C., Stockinger, W.: An adaptive Euler–Maruyama scheme for Mckean–Vlasov SDEs with super-linear growth and application to the mean-field FitzHugh–Nagumo model. arXiv preprint arXiv:2005.06034 (2020)
31. Shardlow, T., Taylor, P.: On the pathwise approximation of stochastic differential equations. BIT Numer. Math. **56**(3), 1101–1129 (2016)
32. Shiryaev, A.: Probability, 2nd edn. Springer, Berlin (1996)
33. Tretyakov, M.V., Zhang, Z.: A fundamental mean-square convergence theorem for SDEs with locally Lipschitz coefficients and its applications. SIAM J. Numer. Anal. **51**(6), 3135–3162 (2013)
34. Wang, X., Gan, S.: The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients. J. Differ. Equ. Appl. **19**(3), 466–490 (2013)
35. Yao, J., Gan, S.: Stability of the drift-implicit and double-implicit Milstein schemes for nonlinear SDEs. Appl. Math. Comput. **339**, 294–301 (2018). https://doi.org/10.1016/j.amc.2018.07.026

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.