On the pathwise approximation of stochastic differential equations

Tony Shardlow[*]   Phillip Taylor

November 24, 2015

Abstract

We consider one-step methods for integrating stochastic differential equations and prove pathwise convergence using ideas from rough path theory. In contrast to alternative theories of pathwise convergence, no knowledge is required of convergence in pth mean and the analysis starts from a pathwise bound on the sum of the truncation errors. We show how the theory is applied to the Euler–Maruyama method with fixed and adaptive time-stepping strategies. The assumption on the truncation errors suggests an error-control strategy and we implement this as an adaptive time-stepping Euler–Maruyama method using bounded diffusions. We prove the adaptive method converges and show some computational experiments.

1   Introduction

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space and consider independent $\mathcal{F}_t$-Brownian motions $W^j(t)$ for $j = 1, \ldots, m$. We study the following Itô stochastic differential equation (SDE) in $\mathbb{R}^d$:

$$dy(t) = g_0(y(t))\, dt + \sum_{j=1}^{m} g_j(y(t))\, dW^j(t) \tag{1.1}$$

where $g_j \colon \mathbb{R}^d \to \mathbb{R}^d$ for $j = 0, \ldots, m$. We assume that g_j are sufficiently regular and there exists a stochastic process $y(t)$ that satisfies this equation on a time interval $[0, T]$ and, if an initial condition is specified, the solution is unique (in the pathwise sense). We denote by $y(t; s, z)$ the solution of Eq. (1.1) for $t \in [s, T]$ with initial condition $y(s) = z \in \mathbb{R}^d$. In general, exact solutions $y(t)$ are not known and numerical integrators are required to determine quantities of interest, such as averages, sample paths, or exit times. In this paper, we look at one-step methods for approximating sample paths of $y(t)$ and analyse the pathwise error using techniques from rough path theory (Davie, 2007; Friz & Victoir, 2010). In dynamical system, we are often interested in how sample paths of SDEs change with model parameters and it is important to compute sample paths reliably. The main result is Theorem 3.5. It gives pathwise convergence of the one-step method at a polynomial rate subject to a regularity condition on the sample paths (Assumptions 2.5 and 3.1) and a bound on the sum of the truncation errors (Assumption 3.2). As well as identifying the rate of convergence in terms of the bound on the truncation-error sum, we identify the constant explicitly in terms of those appearing in the assumptions.

Pathwise error analysis is normally performed (Gyöngy, 1998; Kloeden & Neuenkirch, 2007) by showing the pth mean error converges at a polynomial rate and applying the Borel–Cantelli lemma. Theorem 3.5 predicts the same rates of convergence, for example, for the fixed time-stepping Euler–Maruyama or Milstein method. However, it does not use pth mean error estimates.

---

[*]University of Bath `t.shardlow@bath.ac.uk`

We take particular interest in the so called *bounded-diffusion* time-stepping strategy of (Milstein & Tretyakov, 1999). Instead of taking uniformly spaced times and sampling the Brownian increments from the Gaussian distribution, we choose random times such that the Brownian increment is bounded. Specifically, we define a cuboid $[0, a_0] \times [-a_1, a_1] \times \cdots \times [-a_m, a_m]$ and choose the first exit time τ of the process $(t, W^1(t), \dots, W^m(t))$ from the cuboid. This defines a stopping time τ and associated exit points $W^i(\tau)$ that can be used for the time step and Brownian increments in a numerical integrator for Eq. (1.1). We will use the convergence criterion developed for Theorem 3.5 to choose a_i adaptively and thereby implement an error-control strategy for the Euler–Maruyama method. The adaptivity leads to an improvement in the constant in the theoretical error bound, compared to fixed time-stepping. The constant depends on the inherent exponential divergence of sample paths of the SDE with different initial data (see Assumption 3.1) and the constant in the local truncation error (see Assumption 3.2). The adaptive strategy is able to control the second source of error, not the first.

The paper is organised as follows. §2 gives background on the time-stepping methods of interest. Working pathwise from the start, §3 provides the statement and proof of the main result Theorem 3.5. In §4, we give preliminary lemmas that provide pathwise bounds on the sum of the truncation errors, which help in establishing Assumption 3.2, and show that the pathwise-convergence theorem applies to the Euler–Maruyama method with fixed time-steps. In §5, we introduce two adaptive time-stepping strategies based on bounded diffusions and present convergence theory and numerical experiments. An appendix reviews some useful results.

2 Background

Our pathwise convergence theory applies to one-step methods for Eq. (1.1) in the case of variable and random time-steps. We work on the time interval $[0, T]$ and consider partitions $\mathcal{T}$ of $[0, T]$.

Definition 2.1 (partitions). Let $\mathcal{T}$ denote the set of partitions $0 = \tau_0 < \tau_1 < \cdots < \tau_N = T$ consisting of $[0, T]$-valued random variables. Let $\mathcal{T}_{\text{stop}}$ be the subset of $\mathcal{T}$ consisting of stopping times (i.e., if $(\tau_0, \dots, \tau_N) \in \mathcal{T}_{\text{stop}}$ then τ_j is $\mathcal{F}_{\tau_j}$-measurable for $j = 0, \dots, N$). Let $\hat{s} = \tau_n$ if $\tau_n \leq s < \tau_{n+1}$.

We generate approximations $y_k = S_{0k}(z)$ to $y(\tau_k; \tau_0, z)$ at times τ_k in a partition $(\tau_0, \dots, \tau_n) \in \mathcal{T}$ using one-step methods.

Definition 2.2 (one-step method). Given $\tau_n < \tau_{n+1}$, a *one-step method* $S_{n,n+1}$ is a map from $\mathbb{R}^d$ to the set of $\mathbb{R}^d$-valued random variables. For a given partition $\mathcal{T}$ and $0 \leq k < n \leq N$, we use the notation $S((\tau_k, \dots, \tau_n), z)$ or the abbreviation $S_{kn}(z)$ to denote $S_{n-1,n} \circ \cdots \circ S_{k,k+1}(z)$, the action of applying the one-step method successively over the time steps $[\tau_k, \tau_{k+1}], \dots, [\tau_{n-1}, \tau_n]$.

The simplest useful example is the Euler–Maruyama method with fixed time-step h given by $S_{n,n+1}(y_n) = y_{n+1}$ for times $\tau_n = \tau_0 + nh$ and

$$y_{n+1} = y_n + g_0(y_n)(\tau_{n+1} - \tau_n) + \sum_{j=1}^{m} g_j(y_n)\big(W^j(\tau_{n+1}) - W^j(\tau_n)\big). \qquad (2.1)$$

The implicit Euler–Maruyama method, given by

$$y_{n+1} = y_n + g_0(y_{n+1})(\tau_{n+1} - \tau_n) + \sum_{j=1}^{m} g_j(y_n)\big(W^j(\tau_{n+1}) - W^j(\tau_n)\big), \qquad (2.2)$$

is included if the nonlinear equations can be solved for any y_n to define y_{n+1} uniquely. We will introduce an example of random times τ_n in §5.

Key to the analysis of convergence of one-step methods is the local truncation error.

Definition 2.3 (local truncation error). For $\tau_n < \tau_{n+1}$, the *local truncation error* at $z \in \mathbb{R}^d$ of a one-step method $S_{n,n+1}(z)$ is

$$\delta(\tau_n, \tau_{n+1}, z) := y(\tau_{n+1}; \tau_n, z) - S_{n,n+1}(z). \tag{2.3}$$

For the Euler–Maruyama method, writing $dW^0(t) = dt$, the local truncation error is

$$\delta(\tau_k, \tau_{k+1}, z) = \sum_{i,j=0}^{m} \int_{\tau_k}^{\tau_{k+1}} \int_{\tau_k}^{s} q_{ij}\big(y(r; \tau_k, z)\big)\, dW^i(r)\, dW^j(s), \tag{2.4}$$

where

$$q_{ij}(y) := \begin{cases} Dg_j(y)g_0(y) + \dfrac{1}{2} \sum_{k=1}^{m} D^2 g_j(y)(g_k(y), g_k(y)), & i = 0, \\[2mm] Dg_j(y)g_i(y), & i \neq 0. \end{cases} \tag{2.5}$$

See for example (Kloeden & Platen, 1992; Milstein, 1995). Under regularity assumptions on g_j, we can estimate the pth moment by using the Burkholder–Davis–Gundy inequality and find $\mathbb{E}[\|\delta\|^p] = \mathcal{O}(h^p)$ for $p \geq 2$ and then show that the local truncation error $\delta(\tau_k, \tau_{k+1}, z) = \mathcal{O}(h^{\gamma+1/2-\epsilon})$ with $\gamma = 1/2$ and $\epsilon > 0$. Note the use of ϵ in writing the condition on the truncation error, similar to saying Brownian motion is Hölder continuous with exponent $1/2 - \epsilon$, any $\epsilon > 0$. We make the following assumption on the truncation error.

Assumption 2.4. *For a partition $(\tau_0, \ldots, \tau_N) \in \mathcal{T}$, an $\epsilon \in (0,1)$, and a random variable $K_{\mathrm{TE}} > 0$, it holds that*

$$\big\|\delta(\tau_k, \tau_{k+1}, z)\big\| \leq K_{\mathrm{TE}} |\tau_{k+1} - \tau_k|^{\gamma+1/2-\epsilon}, \qquad \forall z \in \mathbb{R}^d,\ k = 0, \ldots, N-1. \tag{2.6}$$

In general, we want to refine the partition $\mathcal{T}$ and keep K_{TE} and ϵ constant in (2.6), in which case γ indicates the pathwise order of convergence with respect to the mesh width $h := \max_k |\tau_{k+1} - \tau_k|$ (as we show in Theorem 3.5). We will denote such methods by S_{kn}^γ. Before Theorem 3.5, we review a simpler result from (Gaines & Lyons, 1997) that gives order $\gamma - 1/2$ convergence. We assume the following continuity with respect to initial data property, which is satisfied by $y(t; s, z)$ if $g_j \in C_{\mathrm{b}}^3(\mathbb{R}^d, \mathbb{R}^d)$ (Friz & Victoir, 2010, Theorem 10.26). Here, $C_{\mathrm{b}}^r(\mathbb{R}^d, \mathbb{R}^d)$ is the set of functions $g: \mathbb{R}^d \to \mathbb{R}^d$ with $r \in \mathbb{N}$ uniformly bounded and continuous derivatives and norm $\|g\|_{C_{\mathrm{b}}^r} := \sup_{|\alpha| \leq r} \sup_{x \in \mathbb{R}^d} \|D^\alpha g(x)\|$.

Assumption 2.5. *For a random variable $L_{\mathrm{SDE}} > 0$,*

$$\big\|y(t; s, z^1) - y(t; s, z^2)\big\| \leq L_{\mathrm{SDE}} \big\|z^1 - z^2\big\| \tag{2.7}$$

for $0 \leq s \leq t \leq T$ and $z^1, z^2 \in \mathbb{R}^d$.

We define $h := \max_k |\tau_{k+1} - \tau_k|$ and will use h to measure rates of convergence.

Theorem 2.6 (convergence rate $\gamma - 1/2$). *Let Assumption 2.5 hold for Eq. (1.1) and Assumption 2.4 hold for S_{kn}^γ. Then, for any $y_0 \in \mathbb{R}^d$,*

$$\max_{n=0,\ldots,N} \big\|S_{0n}^\gamma(y_0) - y(\tau_n; 0, y_0)\big\| \leq L_{\mathrm{SDE}}\, K_{\mathrm{TE}}\, T\, h^{\gamma-\epsilon-1/2},$$

where ϵ, K_{TE}, and L_{SDE} are given in (2.6) and (2.7).

Proof. Let $y_n := S_{0n}^\gamma(y_0)$ and let $y^k(t) := y(t; \tau_k, y_k)$ for $t \in [\tau_k, T]$. Then,

$$\big\|y^k(\tau_{k+1}) - y_{k+1}\big\| = \big\|y(\tau_{k+1}; \tau_k, y_k) - S_{k,k+1}^\gamma(y_k)\big\|.$$

By Definition 2.3, this is bounded by $K_{\mathrm{TE}} |\tau_{k+1} - \tau_k|^{\gamma+1/2-\epsilon}$. By Assumption 2.5,

$$\big\|y^k(\tau_n) - y^{k+1}(\tau_n)\big\| \leq L_{\mathrm{SDE}} \big\|y^k(\tau_{k+1}) - y^{k+1}(\tau_{k+1})\big\| = L_{\mathrm{SDE}} \big\|y^k(\tau_{k+1}) - y_{k+1}\big\|$$
$$\leq L_{\mathrm{SDE}} \big(K_{\mathrm{TE}} |\tau_{k+1} - \tau_k|^{\gamma-\epsilon+1/2}\big).$$

If we now sum over $k = 0, \ldots, n-1$, the triangle inequality applies and we obtain

$$\left\| y(\tau_n; 0, y_0) - S_{0n}^{\gamma}(y_0) \right\| = \left\| y^0(\tau_n) - y^n(\tau_n) \right\| \leq L_{\text{SDE}} K_{\text{TE}} \sum_{k=0}^{n-1} |\tau_{k+1} - \tau_k|^{\gamma - \epsilon + 1/2},$$

since $y^0(\tau_n) = y(\tau_n; 0, y_0)$ and $y^n(\tau_n) = S_{0n}^{\gamma}(y_0)$. Finally, $\tau_n \leq T$, so that

$$\left\| y(\tau_n; 0, y_0) - S_{0n}^{\gamma}(y_0) \right\| \leq L_{\text{SDE}} K_{\text{TE}} T h^{\gamma - \epsilon - 1/2}.$$

$\square$

Suppose that $\mathcal{T}_h$ is a family of partitions with $h = \max_k |\tau_{k+1} - \tau_k|$ for $(\tau_0, \ldots, \tau_N) \in \mathcal{T}_h$. When Assumption 2.5 holds with the same K_{TE} for each $\mathcal{T}_h$, this theorem provides convergence of the numerical approximation $S_{0n}^{\gamma}(y_0)$ on the partition $\mathcal{T}_h$ to the true solution in the pathwise sense in the limit $h \to 0$ and the pathwise error is $\mathcal{O}(h^{\gamma - \epsilon - 1/2})$, for $\epsilon > 0$. The conditions are general (e.g., allowing time steps that are not uniformly spaced or are not stopping times). However the theorem does not imply convergence of the Euler–Maruyama method and is not optimal for the Milstein method where the error is $\mathcal{O}(h^{1-\epsilon})$ not $\mathcal{O}(h^{1/2-\epsilon})$ for fixed time-steps.

3   Main result on pathwise convergence

To achieve a rate of convergence higher than the rate in Theorem 2.6, we introduce two more assumptions. The first like Assumption 2.5 concerns the solution of the SDE itself. Again, the condition holds if $g_j \in \mathrm{C}_{\mathrm{b}}^3(\mathbb{R}^d, \mathbb{R}^d)$ (see Lemma A.1).

Assumption 3.1. *For an $0 < \epsilon < 1/2$ and a random variable $K_{\text{SDE}} > 0$,*

$$\left\| \left(y(t; s, z^1) - z^1 \right) - \left(y(t; s, z^2) - z^2 \right) \right\| \leq K_{\text{SDE}} (t-s)^{1/2 - \epsilon/3} \| z^1 - z^2 \|, \tag{3.1}$$

for all $0 \leq s < t \leq T$ and $z^1, z^2 \in \mathbb{R}^d$.

The second assumption is a pathwise version of the independent-increment property of Brownian motion and sets a bound on the sum of truncation errors along each path. We will verify this assumption in the case $(\tau_0, \ldots, \tau_N) \in \mathcal{T}_{\text{stop}} \subset \mathcal{T}$, so that the τ_j are stopping times. Indeed, (Gaines & Lyons, 1997) provides an example of a time-stepping strategy where the Euler–Maruyama method fails when τ_j are not adapted. Consider an integrator S_{kn}^{γ} with respect to $(\tau_0, \ldots, \tau_N) \in \mathcal{T}$ and recall $h := \max_{k=0, \ldots, N-1} |\tau_{k+1} - \tau_k|$.

Assumption 3.2 (truncation-error sum)**.** *For an initial condition $y_0 \in \mathbb{R}^d$, an $\epsilon \in (0, 1/2)$, and random variable $K_{\text{TES}} > 0$, it holds that*

$$\left\| X_{kn}(y_0) \right\| \leq K_{\text{TES}} (\tau_n - \tau_k)^{(1+\epsilon)/2} h^{\gamma - \epsilon}, \quad 0 \leq k < n \leq N, \tag{3.2}$$

where $X_{kn}(y_0) := \sum_{j=k}^{n-1} \delta(\tau_j, \tau_{j+1}, y_j)$ and $y_n = S_{0n}^{\gamma}(y_0)$.

In general, we want to refine the partition $\mathcal{T}$ and take limits as $h \downarrow 0$ while keeping K_{TES} and ϵ constant.

For the Euler–Maruyama method (2.1) with fixed time-step $\tau_{n+1} - \tau_n = h$, (3.2) follows because

$$X_{kn}(y_0) = X(\tau_k, \tau_n), \quad \text{for } X(r,t) := \sum_{i,j=0}^{m} \int_r^t \int_{\hat{s}}^s q_{ij}\big(y(r; \hat{r}, \hat{y}(r))\big) \, dW^i(r) \, dW^j(s),$$

by writing Eqs. (2.4) and (2.5) with the notation $\hat{y}(r) := y_n$ for $\tau_n \leq r < \tau_{n+1}$. Then, if the q_{ij} are well-behaved, we can show that $\mathbb{E}[\|X(r,t)\|^p] = \mathcal{O}(h^{p/2}(t-r)^{p/2})$ and we expect that $\|X_{kn}(y_0)\| \leq K_{\text{TES}}(\tau_n - \tau_k)^{(1/2 - \epsilon/4)} h^{1/2 - \epsilon/4}$ for some $K_{\text{TES}} > 0$ independent of h. A further manipulation using $|\tau_n - \tau_k| \geq h$ then yields (3.2) with $\gamma = 1/2$. Thus, there are two steps to verifying (3.2): first, derive pathwise estimates of certain stochastic integrals and, second, show that the time step $\tau_{j+1} - \tau_j$ is not too small relative to h. We verify Assumption 3.2 for the Euler–Maruyama method with different time-stepping methods in the proofs of Theorem 4.2 (fixed time-stepping) and Theorems 5.2 and 5.4 (adaptive bounded diffusions).

3.1 Preliminary lemma

Before giving the main convergence result in Theorem 3.5, we give two lemmas required for its proof. The following result plays the role normally assumed by Gronwall's inequality in proving Theorem 3.5: the local truncation error will determine C_2 and the J_{kn} is one of two terms that control the global error $y_n - y(\tau_n; \tau_k, y_k)$. Then, (3.3) describes how the local truncation error affects the global error. We derive (3.4), which shows J_{kn} is proportional to L and hence C_2. That is, roughly, the local truncation error C_2 controls the global error J_{kn}. The result is adapted from (Davie, 2007).

Lemma 3.3. *For a sequence* $0 = t_0 < t_1 < t_2 < \cdots$, *consider a set of vectors* J_{kn} *indexed by* $k, n \in \mathbb{N}$. *For constants* $C_1, C_2, \gamma > 0$, $\alpha > 1$, *suppose that*

$$\left\| J_{kn} \right\| \leq \left\| J_{k\ell} \right\| \left[1 + C_1 |t_n - t_k|^\gamma \right] + \left\| J_{\ell+1, n} \right\| + C_2 |t_n - t_k|^\alpha \tag{3.3}$$

if $|t_k - t_\ell| \leq \frac{1}{2}|t_n - t_k|$ *and* $|t_{\ell+1} - t_n| \leq \frac{1}{2}|t_n - t_k|$ *and that*

$$J_{kk} = 0, \qquad \left\| J_{k,k+1} \right\| \leq L|t_{k+1} - t_k|^\alpha \quad \text{for } L := \frac{2C_2}{1 - 2^{1-\alpha}}.$$

Then

$$\left\| J_{kn} \right\| \leq L|t_n - t_k|^\alpha \quad \text{if } |t_n - t_k| \leq \delta \tag{3.4}$$

for $\delta > 0$ *such that* $2^{1-\alpha} C_1 \delta^\gamma = 1 - 2^{1-\alpha}$.

Proof. By the assumption on J_{kk} and $J_{k,k+1}$, (3.4) holds for $|k - n| < 2$. We complete the proof by induction on $n - k$ by showing (3.4) given

$$\left\| J_{kn'} \right\| \leq L|t_{n'} - t_k|^\alpha \quad \text{for all } k \leq n' < n. \tag{3.5}$$

Let ℓ be the largest integer satisfying $k \leq \ell < n$ and $|t_\ell - t_k| \leq \frac{1}{2}|t_n - t_k|$. Then, also $|t_n - t_{\ell+1}| \leq \frac{1}{2}|t_n - t_k|$ and, by (3.5),

$$\left\| J_{k\ell} \right\| \leq L|t_\ell - t_k|^\alpha \leq L2^{-\alpha}|t_n - t_k|^\alpha, \qquad \left\| J_{\ell+1,n} \right\| \leq L|t_n - t_{\ell+1}|^\alpha \leq L2^{-\alpha}|t_n - t_k|^\alpha.$$

From (3.3),

$$\begin{aligned}
\left\| J_{kn} \right\| &\leq \left\| J_{k\ell} \right\| \left[1 + C_1 |t_n - t_k|^\gamma \right] + \left\| J_{\ell+1,n} \right\| + C_2 |t_n - t_k|^\alpha \\
&\leq L2^{-\alpha}|t_n - t_k|^\alpha \left[1 + C_1 |t_n - t_\ell|^\gamma \right] + L2^{-\alpha}|t_n - t_k|^\alpha + C_2 |t_n - t_k|^\alpha \\
&\leq \left[L2^{1-\alpha} + L2^{-\alpha} C_1 |t_n - t_\ell|^\gamma + C_2 \right] |t_n - t_k|^\alpha.
\end{aligned}$$

Therefore, $\|J_{kn}\| \leq L|t_n - t_k|^\alpha$ provided that

$$L2^{1-\alpha} + L2^{-\alpha} C_1 |t_n - t_\ell|^\gamma + C_2 \leq L.$$

By choice of L, this means

$$2^{-\alpha} C_1 |t_n - t_\ell|^\gamma \leq (1 - 2^{1-\alpha}) - \frac{C_2}{L} = \frac{1 - 2^{1-\alpha}}{2}. \tag{3.6}$$

Thus, the proof is complete as long as we choose $|t_n - t_\ell|$ to satisfy (3.6). This is guaranteed if $|t_n - t_\ell| \leq \delta$ as $|t_n - t_\ell| \leq |t_n - t_k| \leq \delta$. $\qquad \square$

We now remove the assumption that $|t_n - t_k| \leq \delta$ and allow $|t_n - t_k| \leq T$.

Corollary 3.4. *Suppose the conditions of Lemma 3.3 hold. Then, if* $0 \leq t_k \leq t_n \leq T$,

$$\left\| J_{kn} \right\| \leq L'|t_n - t_k|^\alpha, \qquad L' = L \left\lceil \frac{T}{\delta} \right\rceil.$$

Proof. If $\delta \geq T$, this is implied by Lemma 3.3. Assume $\delta < T$. Divide the interval $[t_k, t_n]$ into the partition $t_k = t_{j_0} < t_{j_1} < \cdots < t_{j_{k_{\max}}} = t_n$, where each $|t_{j_{k+1}} - t_{j_k}| \leq \delta$ or $j_{k+1} - j_k = 1$. Notice that, by choice of the partition, we can ensure that $k_{\max} \leq \lceil T/\delta \rceil$. By Lemma 3.3, we must have $\|J_{j_k, j_{k+1}}\| \leq L|t_{j_{k+1}} - t_{j_k}|^\alpha$. Finally, by (3.3) with $k = \ell = j_k$ and $\ell + 1 = j_{k+1}$ and $n = j_{k+2}$,

$$\|J_{j_k, j_{k+2}}\| \leq \|J_{j_{k+1}, j_{k+2}}\| + C_2 |t_{j_{k+2}} - t_{j_k}|^\alpha.$$

As $C_2 \leq L = 2C_2/(1 - 2^{1-\alpha})$,

$$\|J_{j_k, j_{k+2}}\| \leq 2L \, |t_{j_{k+2}} - t_{j_k}|^\alpha.$$

Thus, the result now holds with $L' = 2L$ for $|t_n - t_k| \leq 2\delta$. This argument can be repeated $k_{\max}$ times to gain $\|J_{kn}\| \leq L \lceil T/\delta \rceil |t_n - t_k|^\alpha$. $\qquad \square$

3.2   Main result

We now use the two assumptions to show pathwise convergence.

Theorem 3.5 (convergence rate γ). *Let Assumptions 2.5 and 3.1 hold for the SDE* (1.1) *and Assumption 3.2 hold for S_{kn}^γ with the partition $(\tau_0, \ldots, \tau_N) \in \mathcal{T}$. Then,*

$$\max_{n=0,\ldots,N} \big\| S_{0n}^\gamma(y_0) - y(\tau_n; 0, y_0) \big\| \leq K_{\text{TES}} \, L_{\text{SDE}}(\epsilon, T) \, h^{\gamma-\epsilon}, \tag{3.7}$$

where

$$L_{\text{SDE}}(\epsilon, T) := T^{(1+\epsilon)/2}$$
$$+ \frac{K_{\text{SDE}}(L_{\text{SDE}} + 2)}{(1 - 2^{-\epsilon/6})} \left(T^{1+\epsilon/6} + \frac{T^{2+\epsilon/6}(K_{\text{SDE}}(L_{\text{SDE}} + 1))^{2/(1-2\epsilon)}}{(2^{\epsilon/6} - 1)^{2/(1-2\epsilon)}} \right).$$

Proof. Let $y_n := S_{0n}^\gamma(y_0)$ and consider $u_{kn} := y_n - y(\tau_n; \tau_k, y_k) + X_{kn}(y_0)$, for $X_{kn}(y_0)$ given in Assumption 3.2. Note that $u_{k,k+1} = 0$ as

$$X_{k,k+1} = \delta(\tau_k, \tau_{k+1}, y_k) = y(\tau_{k+1}; \tau_k, y_k) - y_{k+1}$$

by Eq. (2.3). Further, u_{kn} captures the difference between the integrator S_{kn}^γ and the true solution corrected by X_{kn} and we expect it to be small. Rearranging, we have

$$y_n = y(\tau_n; \tau_k, y_k) - X_{kn}(y_0) + u_{kn}. \tag{3.8}$$

We prove the result by estimating the terms in Eq. (3.8) with $k = 0$,

$$\big\| y_n - y(\tau_n; 0, y_0) \big\| \leq \big\| X_{0n}(y_0) \big\| + \big\| u_{0n} \big\|. \tag{3.9}$$

The first term satisfies $\|X_{0n}(y_0)\| \leq K_{\text{TES}} T^{(1+\epsilon)/2} h^{\gamma-\epsilon}$ by Assumption 3.2, which is consistent with (3.7). We will bound the second term by using Corollary 3.4 and hence show (3.7).

We will develop the inequality required to apply Corollary 3.4. We first define two useful quantities $W_{k\ell}$ and $v_{k\ell n}$ and derive simple bounds on their magnitudes:

First let

$$z^1 := y(\tau_{\ell+1}; \tau_\ell, y(\tau_\ell; \tau_k, y_k) - X_{k\ell}(y_0) + u_{k\ell}) - X_{\ell,\ell+1}(y_0) \quad \text{and} \quad z^2 := y(\tau_{\ell+1}; \tau_k, y_k);$$

then define

$$v_{k\ell n} := \big(y\big(\tau_n; \tau_{\ell+1}, z^1\big) - z^1 \big) - \big(y\big(\tau_n; \tau_{\ell+1}, z^2\big) - z^2 \big). \tag{3.10}$$

6

This quantity is small when $|\tau_n - \tau_{l+1}|$ is small because, applying (3.1), we have

$$
\begin{aligned}
&\|v_{k\ell n}\| \\
&\leq K_{\mathrm{SDE}}\big|\tau_n - \tau_{\ell+1}\big|^{1/2-\epsilon/3} \\
&\quad \times \big\|y\big(\tau_{\ell+1}; \tau_\ell, y(\tau_\ell; \tau_k, y_k) - X_{k\ell}(y_0) + u_{k\ell}\big) - X_{\ell,\ell+1}(y_0) - y(\tau_{\ell+1}; \tau_k, y_k)\big\| \\
&= K_{\mathrm{SDE}}\big|\tau_n - \tau_{\ell+1}\big|^{1/2-\epsilon/3} \\
&\quad \times \big\|y\big(\tau_{\ell+1}; \tau_\ell, y(\tau_\ell; \tau_k, y_k) - X_{k\ell}(y_0) + u_{k\ell}\big) - X_{\ell,\ell+1}(y_0) \\
&\quad\quad - y\big(\tau_{\ell+1}, \tau_\ell, y(\tau_\ell; \tau_k, y_k)\big)\big\| \\
&\leq K_{\mathrm{SDE}}\big|\tau_n - \tau_{\ell+1}\big|^{1/2-\epsilon/3}\big[L_{\mathrm{SDE}}\big\|X_{k\ell}(y_0) - u_{k\ell}\big\| + \big\|X_{\ell,\ell+1}(y_0)\big\|\big].
\end{aligned}
\tag{3.11}
$$

Here, we use (2.7) in the last step.

Let

$$
\begin{aligned}
W_{k\ell} :=\; & y\big(\tau_{\ell+1}; \tau_\ell, y(\tau_\ell; \tau_k, y_k) - X_{k\ell}(y_0) + u_{k\ell}\big) - y\big(\tau_{\ell+1}; \tau_\ell, y(\tau_\ell; \tau_k, y_k)\big) \\
& + X_{k\ell}(y_0) - u_{k\ell}.
\end{aligned}
\tag{3.12}
$$

As with $v_{k\ell n}$, this is bound by applying (3.1),

$$
\big\|W_{k\ell}\big\| \leq K_{\mathrm{SDE}}|\tau_{\ell+1} - \tau_\ell|^{1/2-\epsilon/3}\big\|X_{k\ell}(y_0) - u_{k\ell}\big\|.
\tag{3.13}
$$

We now aim to apply Corollary 3.4 to u_{kn}. We write u_{kn} in terms of $v_{k\ell n}$ and $W_{k\ell}$, and use the above bounds to derive an inequality (3.3) for u_{kn}. Start by applying Eq. (3.8) three times, for $k \leq \ell < n$,

$$
\begin{aligned}
y_n &= y\big(\tau_n; \tau_{\ell+1}, y_{\ell+1}\big) - X_{\ell+1,n}(y_0) + u_{\ell+1,n} \\
&= y\big(\tau_n; \tau_{\ell+1}, y(\tau_{\ell+1}; \tau_\ell, y_\ell) - X_{\ell,\ell+1}(y_0)\big) - X_{\ell+1,n}(y_0) + u_{\ell+1,n} \quad \text{as } u_{\ell,\ell+1} = 0 \\
&= y\big(\tau_n; \tau_{\ell+1}, \underbrace{y\big(\tau_{\ell+1}; \tau_\ell, y(\tau_\ell; \tau_k, y_k) - X_{k\ell}(y_0) + u_{k\ell}\big) - X_{\ell,\ell+1}(y_0)}_{=z^1}\big) \\
&\quad - X_{\ell+1,n}(y_0) + u_{\ell+1,n}.
\end{aligned}
$$

As $y(\tau_n; \tau_\ell, y(\tau_\ell; \tau_k, y_k)) = y(\tau_n; \tau_k, y_k)$, Eq. (3.10) gives

$$
\begin{aligned}
v_{k\ell n} =\; & \big(y_n + X_{\ell+1,n}(y_0) - u_{\ell+1,n}\big) - y(\tau_n; \tau_k, y_k) \\
& - \big(\underbrace{y\big(\tau_{\ell+1}; \tau_\ell, y(\tau_\ell, \tau_k, y_k) - X_{k\ell}(y_0) + u_{k\ell}\big) - X_{\ell,\ell+1}(y_0)}_{=z^1}\big) + y(\tau_{\ell+1}, \tau_k, y_k).
\end{aligned}
$$

Now, substitute Eq. (3.12),

$$
\begin{aligned}
v_{k\ell n} =\; & \big(y_n + X_{\ell+1,n}(y_0) - u_{\ell+1,n}\big) + X_{\ell,\ell+1}(y_0) - y(\tau_n; \tau_k, y_k) \\
& - \big(W_{k\ell} - X_{k\ell}(y_0) + u_{k\ell}\big).
\end{aligned}
\tag{3.14}
$$

Rearranging, and using $X_{k\ell}(y_0) + X_{\ell,\ell+1}(y_0) + X_{\ell+1,n}(y_0) = X_{kn}(y_0)$,

$$
v_{k\ell n} + u_{\ell+1,n} - X_{kn}(y_0) + u_{k\ell} + W_{k\ell} = y_n - y(\tau_n; \tau_k, y_k).
$$

Starting from Eq. (3.8),

$$
\begin{aligned}
u_{kn} &= \big(y_n - y(\tau_n; \tau_k, y_k)\big) + X_{kn}(y_0) \\
&= \big(v_{k\ell n} + u_{\ell+1,n} - X_{kn}(y_0) + u_{k\ell} + W_{k\ell}\big) + X_{kn}(y_0) \\
&= v_{k\ell n} + u_{\ell+1,n} + u_{k\ell} + W_{k\ell}.
\end{aligned}
\tag{3.15}
$$

Apply (3.11) and (3.13) to Eq. (3.15),

$$
\begin{aligned}
\big\|u_{kn}\big\| \leq\; & K_{\mathrm{SDE}}\big|\tau_n - \tau_{\ell+1}\big|^{1/2-\epsilon/3}\big[L_{\mathrm{SDE}}\big\|X_{k\ell}(y_0) - u_{k\ell}\big\| + \big\|X_{\ell,\ell+1}(y_0)\big\|\big] \\
& + \big\|u_{\ell+1,n}\big\| + \big\|u_{k\ell}\big\| \\
& + K_{\mathrm{SDE}}\big|\tau_{\ell+1} - \tau_\ell\big|^{1/2-\epsilon/3}\big\|X_{k\ell}(y_0) - u_{k\ell}\big\|.
\end{aligned}
$$

Then, with Assumption 3.2,

$$
\begin{aligned}
\|u_{kn}\| \leq \|u_{k\ell}\| &\Big[1 + K_{\mathrm{SDE}}\big(L_{\mathrm{SDE}}|\tau_n - \tau_{\ell+1}|^{1/2-\epsilon/3} + |\tau_{\ell+1} - \tau_\ell|^{1/2-\epsilon/3}\big)\Big] \\
&+ \|u_{\ell+1,n}\| \\
&+ K_{\mathrm{SDE}}K_{\mathrm{TES}}|\tau_\ell - \tau_k|^{(1+\epsilon)/2}h^{\gamma-\epsilon}\big[L_{\mathrm{SDE}}|\tau_n - \tau_{\ell+1}|^{1/2-\epsilon/3} + |\tau_{\ell+1} - \tau_\ell|^{1/2-\epsilon/3}\big] \\
&+ K_{\mathrm{SDE}}|\tau_n - \tau_{\ell+1}|^{1/2-\epsilon/3}K_{\mathrm{TES}}|\tau_\ell - \tau_{\ell+1}|^{\gamma+1/2-\epsilon/2}.
\end{aligned}
$$

Choose ℓ so that $|\tau_k - \tau_\ell|, |\tau_n - \tau_{\ell+1}| \leq |\tau_n - \tau_k|/2$. Then, we have (dropping all the $1/2^\alpha < 1$ factors)

$$
\begin{aligned}
\|u_{kn}\| \leq \|u_{k\ell}\| &\Big[1 + K_{\mathrm{SDE}}\big(L_{\mathrm{SDE}}|\tau_n - \tau_k|^{1/2-\epsilon/3} + |\tau_n - \tau_k|^{1/2-\epsilon/3}\big)\Big] \\
&+ \|u_{\ell+1,n}\| \\
&+ K_{\mathrm{SDE}}K_{\mathrm{TES}}|\tau_n - \tau_k|^{(1+\epsilon)/2}h^{\gamma-\epsilon}\big[L_{\mathrm{SDE}}|\tau_n - \tau_k|^{1/2-\epsilon/3} + |\tau_n - \tau_k|^{1/2-\epsilon/3}\big] \\
&+ K_{\mathrm{SDE}}|\tau_n - \tau_k|^{1+\epsilon/6}K_{\mathrm{TES}}h^{\gamma-\epsilon}.
\end{aligned}
$$

Simplifying, we get

$$
\begin{aligned}
\|u_{kn}\| \leq \|u_{k\ell}\|&\Big[1 + K_{\mathrm{SDE}}(L_{\mathrm{SDE}} + 1)|\tau_n - \tau_k|^{1/2-\epsilon}\Big] + \|u_{\ell+1,n}\| \\
&+ K_{\mathrm{TES}}|\tau_n - \tau_k|^{1+\epsilon/6}h^{\gamma-\epsilon}K_{\mathrm{SDE}}(L_{\mathrm{SDE}} + 2).
\end{aligned}
$$

Thus, we have shown that (3.3) holds for $J_{kn} = u_{kn}$ with

$$
\gamma = 1/2 - \epsilon, \quad \alpha = 1 + \epsilon/6, \quad C_1 = K_{\mathrm{SDE}}(L_{\mathrm{SDE}} + 1),
$$

and

$$
C_2 = K_{\mathrm{TES}}h^{\gamma-\epsilon}K_{\mathrm{SDE}}(L_{\mathrm{SDE}} + 2).
$$

Corollary 3.4 implies that

$$
\|u_{kn}\| \leq L\lceil T/\delta\rceil T^{1+\epsilon/6} \leq L\left(T^{1+\epsilon/6} + \frac{T^{2+\epsilon/6}}{\delta}\right),
$$

for $L = 2C_2/(1 - 2^{-\epsilon/6})$ and δ such that $2^{-\epsilon/6}C_1\delta^{1/2-\epsilon} = 1 - 2^{-\epsilon/6}$. Rearranging the equation for δ, we have

$$
\delta^{1/2-\epsilon} = \frac{2^{\epsilon/6} - 1}{C_1} \quad \Rightarrow \quad \delta = \frac{(2^{\epsilon/6} - 1)^{2/(1-2\epsilon)}}{C_1^{2/(1-2\epsilon)}}.
$$

Returning to (3.9), we see that

$$
\|y_n - y(\tau_n; 0, y_0)\| \leq K_{\mathrm{TES}}T^{(1+\epsilon)/2}h^{\gamma-\epsilon} + L\left(T^{1+\epsilon/6} + \frac{T^{2+\epsilon/6}}{\delta}\right).
$$

Define

$$
L_{\mathrm{SDE}}(\epsilon, T) := T^{(1+\epsilon)/2} + \frac{2C_2/(K_{\mathrm{TES}}h^{\gamma-\epsilon})}{1 - 2^{-\epsilon/6}}\left(T^{1+\epsilon/6} + \frac{T^{2+\epsilon/6}C_1^{2/(1-2\epsilon)}}{(2^{\epsilon/6} - 1)^{2/(1-2\epsilon)}}\right).
$$

Note that $L_{\mathrm{SDE}}(\epsilon, T)$ is independent of h and K_{TES} (by definition of C_2). We have

$$
\|y_n - y(\tau_n; 0, y_0)\| \leq K_{\mathrm{TES}}L_{\mathrm{SDE}}(\epsilon, T)h^{\gamma-\epsilon}
$$

and (3.7) is now proved. $\qquad\square$

4 Fixed time-steps

To demonstrate the theory, we consider the case of fixed time-steps with the Euler–Maruyama method. In this case, Assumption 2.4 holds with $\gamma = 1/2$ with K_{TE} uniform in the step size (Friz & Victoir, 2010, Corollary 10.17). This means Theorem 2.6 does not prove convergence in any sense and we need Theorem 3.5 even to prove convergence, as well as to establish the correct rate of convergence. The following lemma is key to establishing the bound on the truncation-error sum necessary for Theorem 3.5. For $p \geq 1$, let $L^p(\Omega)$ denote the Banach space of real-valued random variables X with finite pth moments and norm $\|X\|_{L^p(\Omega)} := (\mathbb{E}[|X|^p])^{1/p}$.

Lemma 4.1. *Consider predictable processes $\{\phi_N(s) \colon s \in [0,T]\}$ for $N \in \mathbb{N}$. Fix $p \geq 1$ and suppose that there exists $K_p^* > 0$ such that*

$$\left\|\bar{\phi}_N^2\right\|_{L^p(\Omega)} \leq K_p^*, \qquad \text{for } \bar{\phi}_N := \sup_{0 \leq s \leq T} \left|\phi_N(s)\right|. \tag{4.1}$$

Choose $\lambda \geq 0$ and $i \in \{0, 1, \ldots, m\}$. Define

$$X^N(t) := \frac{1}{N^\lambda} \int_0^t \phi_N(s)\, dW^i(s).$$

Then, for all $\epsilon > 0$, there exists $C \in L^p(\Omega)$ such that

$$\left\|X^N(s) - X^N(t)\right\| \leq C \frac{1}{N^{\lambda - \epsilon/2}} \times \begin{cases} |s - t|, & i = 0, \\ |s - t|^{1/2 - \epsilon}, & i = 1, \ldots, m, \end{cases} \tag{4.2}$$

for $N \geq 1$ and $0 \leq s < t \leq T$. Further, we can choose C so that $\|C\|_{L^p(\Omega)} \leq \bar{C}(\epsilon, p, T, K_p^*)$ for some $\bar{C}(\epsilon, p, T, K_p^*)$ independent of ϕ_N.

Proof. Note that $\phi_N(s)$ is predictable and $\int_r^t \phi_N(s)^2\, ds \leq \bar{\phi}_N^2 |t - r|$. In the cases that $i \in \{1, \ldots, m\}$, Proposition A.2 (with $\xi = \bar{\phi}_N^2$ and $\mu = p/2$) shows that the modulus of continuity ω_Y of $Y := N^\lambda X^N$ satisfies

$$\omega_Y(h) = \sup_{t,s \in [0,T],\ |t-s| \leq h} \left\|Y(t) - Y(s)\right\| \leq C_p h^{1/2 - \epsilon}, \qquad 0 < h \leq T,$$

for a constant $C_p \in L^p(\Omega)$ with $\|C_p\|_{L^p(\Omega)} \leq \bar{C}_p$, where $\bar{C}_p$ depends only on ϵ, p, T, and K_p^*. Here the constant C_p may depend on N and the bound on ω_Y does not imply (4.2). For $i = 0$, $\|Y(t) - Y(s)\| \leq |t - s| \bar{\phi}_N$, which means $\omega_Y(h) \leq C_p h$ for $C_p := \bar{\phi}_N$.

It remains to show that C_p can be chosen uniformly in N. For the case $i = 1, \ldots, m$, let

$$A^N := \sup_{0 \leq s,t \leq T} \frac{\|X^N(t) - X^N(s)\|\, N^{\lambda - \epsilon/2}}{|t - s|^{1/2 - \epsilon}}.$$

If $A^N \leq C$, then (4.2) holds for $i = 1, \ldots, m$. As $X^N = N^{-\lambda} Y$, we see that $A^N = \sup_{0 \leq s,t \leq T} \|Y(t) - Y(s)\| N^{-\epsilon/2} / |s - t|^{1/2 - \epsilon}$ and $\mathbb{E}[|A^N|^p] \leq \bar{C}_p^p N^{-\epsilon p/2}$. For any $\delta > 0$, Chebyshev's inequality gives

$$\sum_{N=1}^\infty \mathbb{P}\left(A^N > \delta\right) \leq \sum_{N=1}^\infty \frac{\bar{C}_p^p}{N^{\epsilon p/2} \delta^p}.$$

If $p > 2/\epsilon$, the sum converges and the Borel–Cantelli lemma implies that $A^N \to 0$ almost surely. Let $C := \sup_{N \geq 1} A^N$ and note that

$$\|C\|_{L^p(\Omega)} \leq \left(\sum_{N=1}^\infty \|A^N\|_{L^p(\Omega)}^p\right)^{1/p} \leq \bar{C}_p \left(\sum_{N=1}^\infty \frac{1}{N^{\epsilon p/2}}\right)^{1/p} < \infty.$$

Thus, $C \in L^p(\Omega)$ for p large and this extends to any $p \geq 1$ by Jensen's inequality. We have shown that (4.2) holds for a constant C independent of N. A similar argument applies for the case $i = 0$. $\qquad \square$

We now prove convergence of the Euler–Maruyama method with fixed time-steps. Similar results are given in (Gyöngy, 1998; Kloeden & Neuenkirch, 2007) and our result gives more details about the constants.

Theorem 4.2 (fixed time-steps). *Let Assumptions 2.5 and 3.1 hold for the SDE* (1.1) *and let $g_j \in \mathrm{C}_b^2(\mathbb{R}^d, \mathbb{R}^d)$ for $j = 0, \ldots, m$. Let y_n be the fixed time-stepping Euler–Maruyama approximation* (2.1) *at uniformly spaced times $t_n = nh$ for $h = T/N$ and some $N > 0$, with initial condition $y_0 \in \mathbb{R}^d$. Then, for all $0 < \epsilon < 1/2$, there exists a random variable K_{F} such that, almost surely,*

$$\sup_{n=0,1\ldots,N} \big\| y_n - y(t_n; 0, y_0) \big\| \leq L_{\mathrm{SDE}}(\epsilon, T) K_{\mathrm{F}} \frac{1}{N^{1/2-\epsilon}}, \qquad N \geq 1,$$

and $\|K_{\mathrm{F}}\|_{L^p(\Omega)} \leq \max_j \|g_j\|_{\mathrm{C}_b^2}^2 K(\epsilon, p, T)$ for a constant $K(\epsilon, p, T)$ independent of g_j and N.

Proof. Fix $\epsilon \in (0, 1/2)$. From Eq. (2.4),

$$\delta(t_n, t_{n+1}, y_n) = \sum_{i,j} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} q_{ij}(y(r; t_n, y_n)) \, dW^i(r) \, dW^j(s),$$

where q_{ij} is defined in Eq. (2.5) and

$$X_{kn}(y_0) = \sum_{i,j=0}^{m} \int_{t_k}^{t_n} \int_{\hat{s}}^{s} q_{ij}(y(r; \hat{r}, \hat{y}(r))) \, dW^i(r) \, dW^j(s), \qquad \hat{y}(r) = y_k \text{ if } \hat{r} = t_k. \qquad (4.3)$$

We now establish Assumption 3.2 in order to apply Theorem 3.5. Thus, we seek K_{TES} and γ such that

$$\big\| X_{kn}(y_0) \big\| \leq K_{\mathrm{TES}}(t_n - t_k)^{(1+\epsilon)/2} h^{\gamma-\epsilon}. \qquad (4.4)$$

Let $\bar{g} := \max_j \|g_j\|_{\mathrm{C}_b^2}$ and

$$\psi_{ij}(s) := \frac{1}{\bar{g}^2} \int_0^s q_{ij}(r; \hat{r}, \hat{y}(r)) \, dW^i(r), \qquad 0 \leq s \leq T,$$

so that $X_{kn}(y_0) = \bar{g}^2 \sum_{i,j=0}^{m} \int_{t_k}^{t_n} (\psi_{ij}(s) - \psi_{ij}(\hat{s})) \, dW^j(s)$. Now $\psi_{ij}(s)$ is a predictable process and $\|q_{ij}\|$ is bounded by $\bar{g}^2$. Then Lemma 4.1 applies with $\epsilon \mapsto \epsilon/12$, $\lambda = 0$, and $X^N = \psi_{ij}$, so that there exists $C^1 > 0$ such that

$$\sup_{0 \leq s \leq T} \big\| \psi_{ij}(s) - \psi_{ij}(\hat{s}) \big\| \leq C^1 N^{-\epsilon/24} \times \begin{cases} |s - \hat{s}|, & i = 0, \\ |s - \hat{s}|^{1/2-\epsilon/12}, & i = 1, \ldots, m. \end{cases} \qquad (4.5)$$

Further, $C^1 \in L^p(\Omega)$ and $\|C^1\|_{L^p(\Omega)} \leq \bar{C}^1$ for a constant $\bar{C}^1$ independent of g_j.

As $|s - \hat{s}| \leq h = T/N$, (4.5) implies that (4.1) holds with $\phi_N(s) = (\psi(s) - \psi(\hat{s}))N^\lambda$ and $\lambda = 1/2 - \epsilon/8$ for $i = 1, \ldots, m$, and also for $i = 0$ as $1 - \epsilon/24 > 1/2 - \epsilon/8$ (as $\epsilon \in (0, 1/2)$). As $\hat{s} < s$, $\phi_N(s)$ is again a predictable process. Applying Lemma 4.1 once more with $\epsilon \mapsto \epsilon/4$, we find a $C^2 \in L^p(\Omega)$ such that for $N \in \mathbb{N}$

$$\left\| \int_{t_k}^{t_n} (\psi_{ij}(s) - \psi_{ij}(\hat{s})) \, dW^j(s) \right\| \leq C^2 \frac{1}{N^{\lambda-\epsilon/8}} \times \begin{cases} |t_n - t_k|, & j = 0, \\ |t_n - t_k|^{1/2-\epsilon/4}, & j = 1, \ldots, m. \end{cases}$$

Then, summing over $i, j = 0, \ldots, m$, we find $K_{\mathrm{TES}} \in L^p(\Omega)$ such that

$$\big\| X_{kn}(y_0) \big\| \leq K_{\mathrm{TES}} h^{1/2-\epsilon/4} |t_n - t_k|^{1/2-\epsilon/4}.$$

Further $K_{\mathrm{TES}} = \bar{g}^2 K(\epsilon, p)$ for a random variable $K(\epsilon, p)$ such that $\|K(\epsilon, p)\|_{L^p(\Omega)} \leq \overline{K}(\epsilon, p, T)$ for a $\overline{K}(\epsilon, p, T)$ independent of g_j.

Finally, $X_{nn} = 0$ and $h \le |t_n - t_k|$ for $n \ne k$, so that

$$\left\| X_{kn}(y_0) \right\| \le K_{\mathrm{TES}} h^{1/2-\epsilon} |t_n - t_k|^{(1+\epsilon)/2}.$$

This gives the required bound in (4.4) with $\gamma = 1/2$. Thus, we have found a constant $K_{\mathrm{TES}} \in L^p(\Omega)$ such that (4.4) holds. Theorem 3.5 now applies to complete the proof. $\qquad\square$

5   Adaptive time-stepping with bounded diffusions

To demonstrate the theory for random times, we introduce an adaptive time-stepping method based on the method of bounded diffusions (Milstein & Tretyakov, 1999). For Euler–Maruyama, the local truncation error

$$\delta(\tau_n, \tau_{n+1}, y_n) = \sum_{i,j=0}^{m} \int_{\tau_n}^{\tau_{n+1}} \int_{\tau_n}^{s} q_{ij}(y(r; \tau_n, y_n)) \, dW^i(r) \, dW^j(s)$$

and we control this error by selecting the time step $\tau_{n+1} - \tau_n$ as follows. First, fix $\alpha > 0$ as a parameter and denote the maximum time-step by $h = T/N$ for a discretisation parameter $N \in \mathbb{N}$. Suppose that y_n is a given approximation at a stopping time τ_n. We consider two schemes for choosing τ_{n+1}.

Adaptive-I Choose τ_{n+1} to be the largest $t \in [\tau_n, \tau_n + h] \cap [0, T]$ such that for $i = 1, \dots, m$

$$\max_{j=1,\dots,m} \left\| q_{ij}(y_n) \right\|^{1/2} \left| W^i(t) - W^i(\tau_n) \right| \le \alpha h^{1/2}. \tag{5.1}$$

Adaptive-II Choose τ_{n+1} to be the largest $t \in [\tau_n, \tau_n + h] \cap [0, T]$ such that

$$\max_{i,j=1,\dots,m} \left\| q_{ij}(y_n) \right\| \left| \int_{\tau_n}^{\tau_{n+1}} \int_{\tau_n}^{s} dW^i(r) \, dW^j(s) \right| \le \frac{1}{2} \alpha^2 h. \tag{5.2}$$

Notice that τ_{n+1} is also a stopping time. We define the next approximation y_{n+1} at time τ_{n+1} by Eq. (2.1). Given y_0 and $\tau_0 = 0$, this rule defines an approximation y_n at stopping times τ_n for all $n = 0, \dots, N$ where $\tau_N = T$. That is, $(\tau_0, \dots, \tau_N) \in \mathcal{T}_{\mathrm{stop}}$. The first method of choosing the time step $\tau_{n+1} - \tau_n$ is equivalent to finding the first exit time t of $(t, W^1(t), \dots, W^m(t))$ from a cuboid $[0, a_0] \times [-a_1, a_1] \times \cdots \times [-a_m, a_m]$ with

$$a_0 := \min\{h, T - \tau_n\}, \qquad a_i := \min\left\{ h^{1/2} \frac{\alpha}{\|q_{ij}(y_n)\|^{1/2}} : j = 1, \dots, m \right\}, \quad i = 1, \dots, m.$$

Milstein & Tretyakov (1999) give an algorithm for sampling a time step from this distribution, which is used for the experiments in §5.3.

For Adaptive-II, we replace (5.1) with a term involving a double stochastic integral to get (5.2). In the case of diagonal noise, (5.2) simplifies to

$$\left\| q_{jj}(y_n) \right\| \left| \left(W^j(\tau_{n+1}) - W^j(\tau_n) \right)^2 - \left(\tau_{n+1} - \tau_n \right) \right| \le \alpha^2 h. \tag{5.3}$$

In the case $\alpha^2 > \|q_{jj}\|$, this condition holds automatically if (5.1) holds and allows for longer time-steps to be taken. See Figure 1. In general, it is not clear how to sample such a time step and an approximate method is utilised in §5.3 for an example in $m = 1$ dimensions.

5.1   Pathwise convergence for Adaptive-I

There are two parts to the proof of convergence: first, assuming smoothness of g_j, we establish that the time steps are not too small in Lemma 5.1 and then, in Theorem 5.2, we show the conditions of Theorem 3.5 hold.
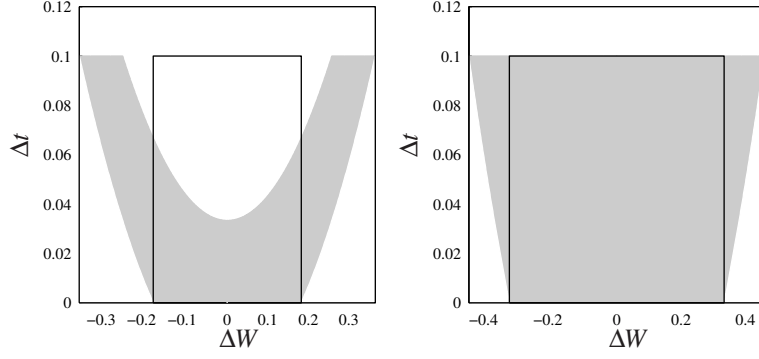
Figure 1: The plots show the regions in $(\Delta t, \Delta W) = (\tau_{n+1} - \tau_n, W^j(\tau_{n+1}) - W^j(\tau_n))$ where $\tau_{n+1} - \tau_n \in [0, h]$ and (5.1) holds (rectangular box) and (5.3) holds (grey region) hold for $\alpha = 1$, $h = 0.1$ and (left) $\|q_{jj}(y_n)\| = 3$ where $\alpha^2 < \|q_{jj}(y_n)\|$ and (right) $\|q_{jj}(y_n)\| = 0.9$ where $\alpha^2 > \|q_{jj}(y_n)\|$.

Lemma 5.1. *Suppose that $g_j \in \mathrm{C}_\mathrm{b}^2(\mathbb{R}^d, \mathbb{R}^d)$ for $j = 0, \ldots, m$. Choose $(\tau_0, \ldots, \tau_N) \in \mathfrak{T}$ such that (5.1) holds for parameters $\alpha, h > 0$. For all $0 < \delta < 1$, there exists a random variable $C_\delta > 0$ (independent of g_j and y_0 and α) such that*

$$h \leq K_{\mathrm{BI}}|\tau_{n+1} - \tau_n|^{1-\delta} \qquad \text{for } n = 0, \ldots, N-2, \tag{5.4}$$

where $K_{\mathrm{BI}} := T^\delta + C_\delta^2 \max_j \|g_j\|_{\mathrm{C}_\mathrm{b}^2}^2 / \alpha^2$.

Proof. Choose a random variable C_δ such that $|W^i(t) - W^i(s)| \leq C_\delta |t-s|^{(1-\delta)/2}$ for $i = 1, \ldots, m$ and $0 \leq s, t \leq T$.

If $\tau_{n+1} - \tau_n = h$, we have $h \leq T^\delta |\tau_{n+1} - \tau_n|^{1-\delta}$ and (5.4) holds. Alternatively, $\tau_{n+1} - \tau_n < h$ and, as $n \leq N - 2$, we know that $\tau_{n+1} < T$. In this case, we must have that τ_{n+1} satisfies (5.1) with equality at some specific $i, j = 1, \ldots, m$. That is,

$$\|q_{ij}(y_n)\|^{1/2} |W^i(\tau_{n+1}) - W^i(\tau_n)| = \alpha h^{1/2}.$$

Then,

$$C_\delta |\tau_{n+1} - \tau_n|^{(1-\delta)/2} \geq |W^i(\tau_{n+1}) - W^i(\tau_n)| = \frac{\alpha h^{1/2}}{\|q_{ij}(y_n)\|^{1/2}}.$$

In other words,

$$h \leq \frac{C_\delta^2 \|q_{ij}(y_n)\|}{\alpha^2} |\tau_{n+1} - \tau_n|^{1-\delta} \leq K_{\mathrm{BI}}|\tau_{n+1} - \tau_n|^{1-\delta}$$

for the given K_{BI}, as required. □

The next theorem describes an error bound for Adaptive-I. To leading order, the constant K_A in the error bound scales like $\bar{g}^{1+\epsilon}$ for $\bar{g} = \max_j \|g_j\|_{\mathrm{C}_\mathrm{b}^2}$, whilst the corresponding constant K_F in Theorem 4.2 scales like $\bar{g}^2$. This is a more favourable scaling of the error estimate and says the error bound scales nearly linearly with the magnitude of the vector fields g_j.

Theorem 5.2 (adaptive-I)**.** *Let Assumptions 2.5 and 3.1 hold for the SDE (1.1) and suppose that $g_j \in \mathrm{C}_\mathrm{b}^3(\mathbb{R}^d, \mathbb{R}^d)$. Let y_n denote the Euler–Maruyama approximation (2.1) at times τ_n given by Adaptive-I with $h = T/N$, some $N \in \mathbb{N}$. Then, for $0 < \epsilon < 1/2$ and $y_0 \in \mathbb{R}^d$, there exists a random variable $K_\mathrm{A} > 0$ such that, almost surely,*

$$\sup_{0 \leq \tau_n \leq T} \|y_n - y(\tau_n; 0, y_0)\| \leq L_{\mathrm{SDE}}(\epsilon, T) \, K_\mathrm{A} \frac{1}{N^{1/2-\epsilon}}, \qquad N \geq 1,$$

where $\|K_\mathrm{A}\|_{L^p(\Omega)} \leq K_1(\epsilon, p, \alpha, T) \max_j \|g_j\|_{\mathrm{C}_\mathrm{b}^2}^{1+2\epsilon} + K_2(\epsilon, p, \|g_j\|_{\mathrm{C}_\mathrm{b}^3}, \alpha, T)/N^{1/2}$ for some $K_1(\epsilon, p, \alpha, T)$ independent of g_j and N and $K_2(\epsilon, p, \|g_j\|_{\mathrm{C}_\mathrm{b}^3}, \alpha, T)$ independent of N.

Proof. We have as in Eq. (4.3)

$$X_{kn}(y_0) = \sum_{i,j=0}^{m} \int_{t_k}^{t_n} \int_{\hat{s}}^{s} q_{ij}\big(y(r;\hat{r},\hat{y}(r))\big)\, dW^i(r)\, dW^j(s), \qquad \hat{y}(r) = y_k \text{ if } \hat{r} = \tau_k.$$

Expanding the terms $i, j = 1, \ldots, m$ again using Itô's formula

$$\begin{aligned}
X_{kn}(y_0) &= \sum_{i,j=1}^{m} \int_{t_k}^{t_n} \int_{\hat{s}}^{s} q_{ij}\big(y(\hat{r};\hat{r},\hat{y}(r))\big)\, dW^i(r)\, dW^j(s) \\
&+ \sum_{j=0}^{m} \int_{t_k}^{t_n} \int_{\hat{s}}^{s} q_{0j}\big(y(r;\hat{r},\hat{y}(r))\big)\, dr\, dW^j(s) \\
&+ \sum_{i=1}^{m} \int_{t_k}^{t_n} \int_{\hat{s}}^{s} q_{i0}\big(y(r;\hat{r},\hat{y}(r))\big)\, dW^i(r)\, ds \\
&+ \sum_{i,j=1,\ell=0}^{m} \int_{t_k}^{t_n} \int_{\hat{s}}^{s} \int_{\hat{s}}^{r} q_{\ell ij}\big(y(u;\hat{u},\hat{y}(u))\big)\, dW^\ell(u)\, dW^i(r)\, dW^j(s),
\end{aligned} \tag{5.5}$$

where $q_{\ell ij} := \mathcal{L}^\ell q_{ij}$ for $\mathcal{L}^0 \phi := D\phi g_0 + \frac{1}{2} \sum_{j=1}^{m} D^2\phi(g_j, g_j)$ and $\mathcal{L}^\ell \phi := D\phi g_\ell$. Fix $\epsilon \in (0, 1/2)$. Following the same arguments as in the proof of Theorem 4.2, the last three terms here are bounded by $K_{\mathrm{TES}}^2 h^{1-\epsilon} |t_n - t_k|^{(1+\epsilon)/2}$ for a constant K_{TES}^2 that satisfies $\|K_{\mathrm{TES}}^2\|_{L^p(\Omega)} \leq K\left(\epsilon, \alpha, p, \|g_j\|_{\mathrm{C}_b^3}\right)$ for some $K\left(\epsilon, \alpha, p, \|g_j\|_{\mathrm{C}_b^3}\right)$ independent of N.

To bound the first term, we use a more refined argument. Let $\bar{g} := \max_j \|g_j\|_{\mathrm{C}_b^2}$ and

$$\phi_N(s) := \frac{1}{\bar{g} h^{1/2}} \int_{\hat{s}}^{s} q_{ij}\big(y(\hat{r};\hat{r},\hat{y}(r))\big)\, dW^i(r) = \frac{1}{\bar{g} h^{1/2}} \int_{\hat{s}}^{s} q_{ij}\big(\hat{y}(r)\big)\, dW^i(r)$$

for $\tau_k = \hat{r} \leq r < \tau_{k+1}$. Here $\phi_N(s)$ is continuous and adapted (as $\hat{s}$ is a stopping time and $\hat{s} < s$) and hence predictable. By (5.1),

$$\left\| q_{ij}\big(y(\hat{r};\hat{r},\hat{y}(r))\big) \right\|^{1/2} \left| W^i(\tau_{n+1}) - W^i(\tau_n) \right| \leq \alpha h^{1/2}, \qquad j = 1, \ldots, m.$$

As $\|q_{ij}\| \leq \bar{g}^2$, this gives

$$\sup_{0 \leq s \leq T} \left\| \phi_N(s) \right\|^2 = \frac{1}{\bar{g}^2 h} \sup_{0 \leq s \leq T} \left\| \int_{\hat{s}}^{s} q_{ij}\big(y(\hat{r};\hat{r},\hat{y}(r))\big)\, dW^i(r) \right\|^2 \leq \alpha^2.$$

Hence, (4.1) holds and Lemma 4.1 applies with $\lambda = 1/2$ and

$$X^N(t) = h^{1/2} \int_0^t \phi_N(s)\, dW^i(s).$$

This gives a $c_{ij} \in L^p(\Omega)$ such that

$$\frac{1}{\bar{g}} \left\| \int_{\tau_k}^{\tau_n} \int_{\hat{s}}^{s} q_{ij}\big(y(\hat{r};\hat{r},\hat{y}(r))\big)\, dW^i(r)\, dW^j(s) \right\| \leq c_{ij} |\tau_n - \tau_k|^{1/2 - \epsilon/4}\, h^{1/2 - \epsilon/8}.$$

We may choose c_{ij} so that its $L^p(\Omega)$ norm is independent of h and $\bar{g}$. This provides the necessary bound on the first term of Eq. (5.5).

Taken all the terms in Eq. (5.5) together, we can find K_{TES}^1 such that

$$X_{kn}(y_0) \leq K_{\mathrm{TES}}^1 |\tau_n - \tau_k|^{1/2 - \epsilon/4} h^{1/2 - \epsilon/8} + K_{\mathrm{TES}}^2 h^{1-\epsilon} |\tau_n - \tau_k|^{(1+\epsilon)/2}.$$

Here $\|K_{\mathrm{TES}}^1\|_{L^p(\Omega)} \leq \bar{g} K(\epsilon, p, \alpha, T)$ for some $K(\epsilon, p, \alpha, T)$ independent of N and g_j.

Using Lemma 5.1 with $\delta = 1/7$, we have the lower bound $h \leq K_{\text{BI}}|\tau_n - \tau_k|^{1-\delta} = K_{\text{BI}}|\tau_n - \tau_k|^{6/7}$ on the time step for $k, n = 0, \ldots, N - 1$. Hence,

$$
\begin{aligned}
|\tau_n - \tau_k|^{1/2-\epsilon/4}h^{1/2-\epsilon/8} &\leq |\tau_n - \tau_k|^{1/2-\epsilon/4}h^{1/2-\epsilon}h^{7\epsilon/8} \\
&\leq |\tau_n - \tau_k|^{1/2-\epsilon/4}h^{1/2-\epsilon}K_{\text{BI}}^{7\epsilon/8}|\tau_n - \tau_k|^{\epsilon 3/4} \\
&= |\tau_n - \tau_k|^{1/2+\epsilon/2}h^{1/2-\epsilon}K_{\text{BI}}^{7\epsilon/8}.
\end{aligned}
$$

Consequently,

$$
\left\|X_{kn}(y_0)\right\| \leq K_{\text{TES}}^1 K_{\text{BI}}^\epsilon |\tau_n - \tau_k|^{(1+\epsilon)/2}h^{1/2-\epsilon} + K_{\text{TES}}^2|\tau_n - \tau_k|^{(1+\epsilon)/2}h^{1-\epsilon}.
$$

We have all the conditions of Theorem 3.5, which gives the desired error bound for $n = 0, \ldots, N - 1$. The final step from τ_{N-1} to τ_N may equal $|T - \tau_{N-1}|$, which may be very small and does not yield to Lemma 5.1. However, the error resulting from this step can be added into the K_{TES}^2 term by utilising the bound in (2.6). $\qquad\square$

5.2 Analysis of Adaptive-II

The convergence result (Theorem 5.4 below) for Euler–Maruyama with Adaptive-II is similar to Theorem 5.2 and the new time-stepping strategy behaves like Adaptive-I. However, the method of proof is different and we will make use of Azuma's inequality. First, we show the time step does not become too small, similarly to Lemma 5.1.

Lemma 5.3. *Suppose that $g_j \in \mathrm{C}_{\mathrm{b}}^2(\mathbb{R}^d, \mathbb{R}^d)$ for $j = 0, \ldots, m$. Choose $(\tau_0, \ldots, \tau_N) \in \mathcal{T}$ such that (5.2) holds for parameters $\alpha, h > 0$. For $0 < \delta < 1$, there exists a random variable $C_\delta > 0$ (independent of α, h, g_j, and y_0) such that*

$$
h \leq K_{\text{BI}}|\tau_{k+1} - \tau_k|^{1-\delta}, \quad for \ k = 0, \ldots, N - 2, \tag{5.6}
$$

where $K_{\text{BI}} = T^\delta + 2C_\delta \max_j \|g_j\|_{\mathrm{C}_{\mathrm{b}}^2}^2/\alpha^2$.

Proof. Choose a random variable C_δ such that $|\int_s^t W^i(r)\,dW^j(r)| \leq C_\delta|t - s|^{1-\delta}$ for $i, j = 1, \ldots, m$ and $0 \leq s, t \leq T$.

If $\tau_{k+1} - \tau_k = h$, then $h \leq T^\delta|\tau_{k+1} - \tau_k|^{1-\delta}$ and (5.6) holds. Otherwise, we must have $\tau_{k+1} - \tau_k < h$ and $\tau_{k+1} < T$. Then τ_{k+1} satisfies (5.2) with equality for some $i, j = 1, \ldots, m$:

$$
\left\|q_{ij}(y_k)\right\| \left|\int_{\tau_k}^{\tau_{k+1}} \int_{\tau_k}^s dW^i(r)\,dW^j(s)\right| = \frac{1}{2}\alpha^2 h.
$$

Then,

$$
C_\delta|\tau_{k+1} - \tau_k|^{1-\delta} \geq \left|\int_{\tau_k}^{\tau_{k+1}} \int_{\tau_k}^s W^i(r)\,dW^j(s)\right| \geq \frac{\alpha^2 h}{2\|q_{ij}(y_k)\|}.
$$

In other words,

$$
h \leq \frac{2C_\delta\|q_{ij}(y_k)\|}{\alpha^2}|\tau_{k+1} - \tau_k|^{1-\delta} \leq K_{\text{BI}}|\tau_{k+1} - \tau_k|^{1-\delta}.
$$

Thus, (5.6) holds and the proof is complete. $\qquad\square$

The error bound for Adaptive-II found in the next theorem scales (in terms of g_j) like the one for Adaptive-I.

Theorem 5.4 (adaptive-II). *Let Assumptions 2.5 and 3.1 hold for the SDE (1.1) and suppose that $g_j \in \mathrm{C}_{\mathrm{b}}^3(\mathbb{R}^d, \mathbb{R}^d)$. Let y_n denote the Euler–Maruyama approximation at times τ_n with adaptive increments given by (5.2) and Eq. (2.1). Then, for $0 < \epsilon < 1/2$ and $y_0 \in \mathbb{R}^d$, there exists a random variable $K_{\text{A}} > 0$ such that, almost surely,*

$$
\sup_{0 \leq \tau_n \leq T}\left\|y_n - y(\tau_n; 0, y_0)\right\| \leq L_{\text{SDE}}(\epsilon, T)\,K_{\text{A}}\frac{1}{N^{1/2-\epsilon}}, \qquad N \geq 1,
$$

where $\|K_{\text{A}}\|_{L^p(\Omega)} \leq K_1(\epsilon, p, \alpha, T)\|g_j\|_{\mathrm{C}_{\mathrm{b}}^2}^{1+\epsilon} + K_2(\epsilon, p, \alpha, \|g_j\|_{\mathrm{C}_{\mathrm{b}}^3}, T)/N^{1/2}$ for some $K_1(\epsilon, p, \alpha, T)$ independent of N and g_j, and some $K_2(\epsilon, p, \alpha, \|g_j\|_{\mathrm{C}_{\mathrm{b}}^3}, T)$ independent of N.

Proof. Following the proof of Theorem 5.2, it is enough to treat the term

$$S_{kn} := \sum_{j=k}^{n-1} S_j, \qquad \text{for } S_k := \sum_{i,j=1}^{m} \int_{\tau_k}^{\tau_{k+1}} \int_{\tau_k}^{s} q_{ij}(y_k)\, dW^i(r)\, dW^j(s) \text{ and } 0 \le k < n \le N,$$

and show it satisfies the condition on X_{kn} in Assumption 3.2. By the optional stopping theorem, each S_k has mean zero and, by (5.2), $\|S_k\| \le m^2\alpha^2 h/2$ for $h = T/N$. Then, S_{kn} is a sum independent random variables, each with mean zero and bounded by $m^2\alpha^2 h/2$. Azuma's inequality (see Lemma A.3 with $\lambda = (n-k)^{1/2} h^{1-\epsilon/4}$) gives

$$\mathbb{P}\big(\|S_{kn}\| \ge \lambda\big) \le 2\exp\left(\frac{-4\lambda^2}{2(n-k)m^4\alpha^4 h^2}\right) = 2\exp\left(\frac{-2}{m^4\alpha^4 h^{\epsilon/2}}\right).$$

Let $F_N := \{\omega \in \Omega \colon \sup_{0 \le k < n \le N}\|S_{kn}(\omega)\|/(n-k)^{1/2} \ge h^{1-\epsilon/4}\}$. Using $h = T/N$,

$$\mathbb{P}\big(F_N\big) \le \sum_{0 \le k < n \le N} \mathbb{P}\big(\|S_{kn}\| \ge \lambda\big) \le (N+1)(N+2)\exp\left(\frac{-2N^{\epsilon/2}}{m^4\alpha^4 T^{\epsilon/2}}\right).$$

Choose C_ϵ so that $(N+1)(N+2)\exp(-N^{\epsilon/2}/m^4\alpha^4 T^{\epsilon/2}) \le C_\epsilon$ and

$$\mathbb{P}\big(F_N\big) \le C_\epsilon \exp\left(\frac{-N^{\epsilon/2}}{m^4\alpha^4 T^{\epsilon/2}}\right) \qquad \text{for } N \in \mathbb{N} .$$

Then $\sum_{N=1}^{\infty} \mathbb{P}(F_N) < \infty$ and the Borel–Cantelli lemma applies, to give

$$\sup_{0 \le k < n \le N} \frac{\|S_{kn}\|N^{1-\epsilon/4}}{(n-k)^{1/2}} \le C, \tag{5.7}$$

almost surely, for some random variable C.

By Lemma 5.3, for each $\delta > 0$, there is a K_{BI} such that

$$\left(\frac{1}{K_{\mathrm{BI}}}\frac{T}{N}\right)^{1/(1-\delta)} \le |\tau_{k+1} - \tau_k|$$

and, summing $|\tau_{k+1} - \tau_k|, \ldots, |\tau_n - \tau_{n-1}|$,

$$(n-k)\left(\frac{1}{K_{\mathrm{BI}}}\frac{T}{N}\right)^{1/(1-\delta)} \le |\tau_n - \tau_k|.$$

Choose δ so that $(1+\epsilon)/2(1-\delta) = 1/2 + 3\epsilon/4$. Then,

$$\frac{N^{1/2-\epsilon}}{|\tau_n - \tau_k|^{(1+\epsilon)/2}} \le \left(\frac{K_{\mathrm{BI}}}{T}\right)^{(1+\epsilon)/2(1-\delta)} \frac{N^{1/2-\epsilon+(1+\epsilon)/2(1-\delta)}}{|n-k|^{(1+\epsilon)/2}} \le \left(\frac{K_{\mathrm{BI}}}{T}\right)^{1/2+3\epsilon/4} \frac{N^{1-\epsilon/4}}{|n-k|^{1/2}}.$$

Now use (5.7) to gain

$$\sup_{0 \le k < n \le N} \frac{\|S_{kn}\|N^{1/2-\epsilon}}{|\tau_n - \tau_k|^{(1+\epsilon)/2}} \le \left(\frac{K_{\mathrm{BI}}}{T}\right)^{1/2+3\epsilon/4} \sup_{0 \le k < n \le N} \frac{\|S_{kn}\|N^{1-\epsilon/4}}{(n-k)^{1/2}} \le C\left(\frac{K_{\mathrm{BI}}}{T}\right)^{1/2+3\epsilon/4}.$$

Thus, $\|S_{kn}\| \le \tilde{C}|\tau_n - \tau_k|^{(1+\epsilon)/2} h^{1/2-\epsilon}$ for a constant $\tilde{C}$. The remainder of the argument is the same as in the proof of Theorem 5.2. In this case, the dependence on $\|g_j\|_{C_b^2}$ comes from K_{BI}. $\qquad\square$
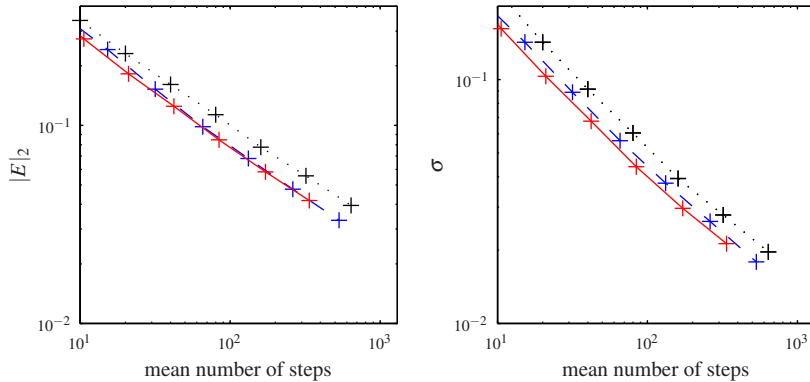
15

Figure 2: For Eq. (5.9), (left) the sample mean of the maximum relative error $|E|_2$ and (right) standard deviation σ of the error E (see Eq. (5.10)) against the mean number of steps taken (based on 5000 samples). We plot (dashed) Adaptive-I with $\alpha = 0.5$, (solid) Adaptive-II with $\alpha = 0.9$, and (dotted) fixed time-stepping Euler–Maruyama method.

5.3 Experiments with adaptive algorithms

We now test Adaptive-I and Adaptive-II with the Euler–Maruyama method using the following initial value problems:

$$dy = 0.1y\, dt + 1.2y\, dW(t), \qquad y(0) = 1 \tag{5.8}$$

and

$$dy = 1.5y\, dt + 2.4y\, dW(t), \qquad y(0) = 1. \tag{5.9}$$

Here, $d = m = 1$ and $W(t)$ is a standard Brownian motion. We integrate both equations numerically with the Euler–Maruyama method on the interval $[0, T] = [0, 1]$ and compare the result with the exact solution (geometric Brownian motion). In this example, the drift and diffusion functions $g_j(y)$ are linear and unbounded. In the experiments, we replace $\|q_{ij}(y_k)\|$ in (5.1) and (5.2) with $\min\{\|q_{ij}(y_k)\|, 100\}$; this ensures the time steps do not become too small.

For Adaptive-I, time steps and Brownian increments are generated using the method of (Milstein & Tretyakov, 1999). For Adaptive-II, we'd like to sample from an exit time problem on the domain shown in Figure 1. We use the following approximate algorithm to sample from this distribution: Let R denote the shaded region in Figure 1 for a given α and h. Choose a parameter $\beta > 0$.

1) Let $(\tau^0, x^0) = (0, 0)$ and $k = 0$.

2) Choose the largest a_1 such that $\{\tau^k\} \times [x^k - a_1, x^k + a_1] \subset R$, and then the largest a_0 such that $[\tau^k, \tau^k + a_0] \times [x^k - a_1, x^k + a_1] \subset R$.

3) Use the algorithm of (Milstein & Tretyakov, 1999) to find the first exit point $(\tau, W(\tau))$ from $[0, a_0] \times [-a_1, a_1]$ of the process $(t, W(t))$.

4) Let $\tau^{k+1} = \tau^k + \tau$ and $x^{k+1} = x^k + W(\tau)$. If $\tau < \beta h$, stop and output (τ^{k+1}, x^{k+1}). Otherwise, increase k and go to 2).

The output can be used as the time step and Brownian increment, which always belongs to the region R but is unlikely to be an exit point. We apply this method with $\beta = 1/10$ to implement Adaptive-II and the resulting distribution of steps taken is shown in Figure 5.

We compute the maximum relative error on the partition $\mathcal{T}$, defined by

$$E := \frac{1}{Z} \max_{\tau_n \in \mathcal{T}} |y_n - y(\tau_n)|, \qquad \text{where } Z := \max_{\tau_n \in \mathcal{T}} |y(\tau_n)|. \tag{5.10}$$
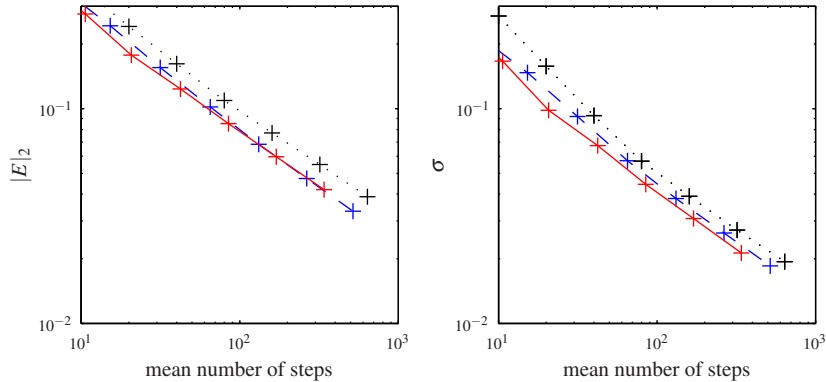
16

Figure 3: For Eq. (5.8), (left) $|E|_2$ and (right) standard deviation σ of E against the mean number of steps taken. We plot (dashed) Adaptive-I with $\alpha = 0.5$, (solid) Adaptive-II with $\alpha = 0.9$, and (dotted) fixed time-stepping Euler–Maruyama method.

For M *iid* samples $E_1, \ldots, E_M$ of E, we plot

$$|E|_2 := \left(\sum_{j=1}^{M} E_j^2 \right)^{1/2}$$

and the standard deviation σ of $E_1, \ldots, E_M$ against the mean number of steps, using $M = 5000$ samples in Figures 2 and 3. The adaptive methods outperform the fixed time-stepping method in terms of mean error (when taken with the same mean number of steps) and a 20% reduction in error is observed using either adaptive algorithm. Furthermore, the adaptive methods, especially Adaptive-II, produce a narrower range of errors, as we find the standard deviation of the errors is smaller. Whilst this is encouraging, we are discounting the extra time involved in sampling the bounded diffusion and, as this algorithm is slow compared to sampling a Gaussian increment, the adaptive methods are not yet fully practical.

Two further plots are shown for Eq. (5.8). Figure 4 shows a scatter plot of the errors E against the number of steps taken for the three methods. Figure 5 shows the mean and standard deviation of the number of steps used in the experiments for a set of h values.

6   Conclusion

We presented a new proof of pathwise convergence for numerical approximation of SDEs, which avoids the need to prove the pth-mean error converges at a polynomial rate. This requires two pathwise assumptions: first, a Lipschitz type assumption of the deviation of the sample path of the solution from its initial data and, second, a pathwise bound on the sum of the truncation errors. The proof of the pathwise-convergence theorem uses no probabilistic arguments. We showed how to apply the theorem to the Euler–Maruyama method with fixed time-stepping. We also introduced two adaptive time-stepping methods, motivated by the truncation error sum condition. To complete these proofs and verify the assumptions of the pathwise-convergence theorem, we do use probabilistic arguments and estimate moments and apply the Borel–Cantelli lemma. The main advantage of this approach is that more detailed constants are gained and we are able to find a tighter error bound for the adaptive methods. Computations with SDEs usually work by computing paths, even if the quantity of interest is an average or exit time, and the pathwise condition on the truncation error sum is a convenient framework for studying adaptive methods.

Some experiments were shown with geometric Brownian motion using a method of (Milstein & Tretyakov, 1999) for bounded diffusions to generate the appropriate time steps and Brownian increments. For a given mean number of time steps, the errors are smaller and show less
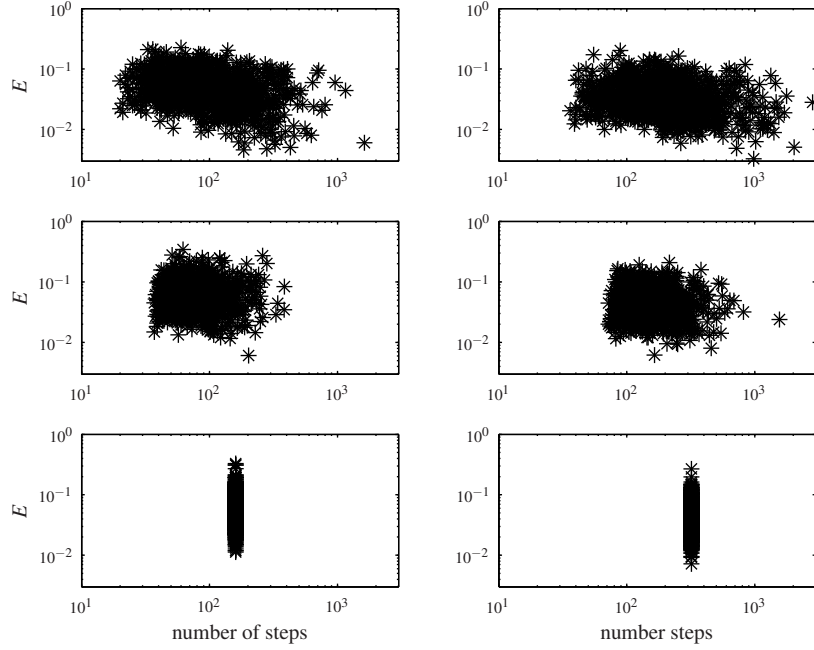
Figure 4: Plots of the relative error E against number of steps taken N for 100 samples of the approximation to Eq. (5.8) for (top) Adaptive-I with $\alpha = 0.5$, (middle) Adaptive-II with $\alpha = 0.9$, and (bottom) fixed time-stepping Euler–Maruyama.
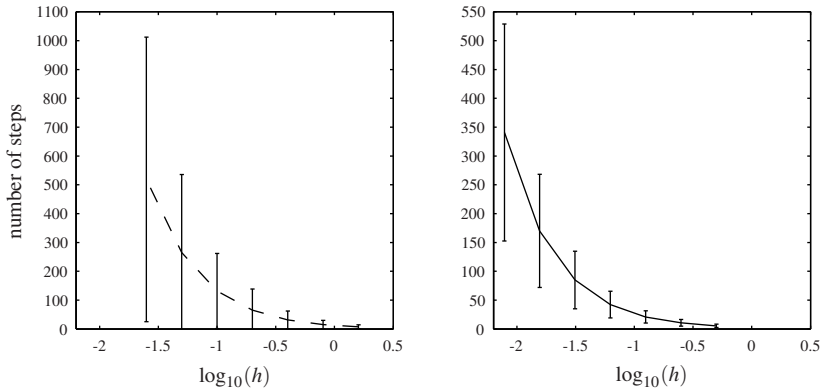

Figure 5: Mean number of steps with error bars indicating one standard deviation for (left) Adaptive-I with $\alpha = 0.5$ and (right) Adaptive-II with $\alpha = 0.9$ applied to Eq. (5.8).

variation than for fixed time-stepping. This extra accuracy, which does not change the rate of convergence, requires sampling bounded diffusions rather than diffusions over fixed time-steps (*iid* Gaussian random variables) and this makes the algorithm expensive to implement. Finding a fast method for sampling bounded diffusions is a key point to be addressed in future research.

A Useful results

The first result concerns the regularity of sample paths and gives a large class of SDEs where Assumption 3.1 holds.

Lemma A.1. *If $g_j \in C_b^3(\mathbb{R}^d, \mathbb{R}^d)$, then the solution $y(t; s, z)$ of Eq. (1.1) satisfies Assumption 3.1.*

Proof. Fix $\epsilon > 0$. Let $x(t; s, z) := y(t; s, z) - z$ for $z \in \mathbb{R}^d$ and $0 \le s, t \le T$. Then, $x(t; s, z)$ satisfies

$$dx = dy = \sum_{j=0}^m g_j(y) \, dW^j(t) = \sum_{j=0}^m g_j(x+z) \, dW^j(t), \qquad x(s; s, z) = 0$$

and $x(t; s, z^i)$ for $i = 1, 2$ are two solutions of an SDE with modified vector fields $V_j^i(q) = g_j(q + z^i)$. Then, (Friz & Victoir, 2010, Theorem 10.26) gives

$$\left\| x(t; s, z^1) - x(t; s, z^2) \right\| \le C |t - s|^{1/2 - \epsilon} \max_j \left\| V_j^1 - V_j^2 \right\|_{C_b^2}$$

for some C dependent on the sample path of $W^j(t)$. Because $g_j \in C_b^3(\mathbb{R}^d, \mathbb{R}^d)$, it is clear that $\| V_j^1 - V_j^2 \|_{C_b^2} \le \| g_j \|_{C_b^3} \| z^1 - z^2 \|$ and hence

$$\left\| \left(y(t; s, z^1) - z^1 \right) - \left(y(t; s, z^2) - z^2 \right) \right\| \le C |t - s|^{1/2 - \epsilon} \left\| z^1 - z^2 \right\|,$$

for all $z^1, z^2 \in \mathbb{R}^d$ and $0 \le s, t \le T$. This is Assumption 3.1. $\qquad\square$

For a stochastic process $Y(t)$, the modulus of continuity

$$\omega_Y(h) := \sup_{t, s \in [0, T], \, |t-s| \le h} \left\| Y(t) - Y(s) \right\|, \qquad h > 0.$$

The following result examines the modulus of continuity when Y is an Itô integral and is applied to the truncation error in order to derive Assumption 3.2.

Proposition A.2. *Let*

$$Y(t) = \int_0^t q_0(s) \, ds + \sum_{j=1}^m \int_0^t q_j(s) \, dW^j(s)$$

for predictable processes $\{q_j(s) \colon s \in [0, T]\}$. Consider random variables ζ, ξ such that, for some $p \ge 1$ and $K, \mu > 0$,

$$\mathbb{E}\left[|\zeta|^p \right] \le K \quad and \quad \mathbb{E}\left[|\xi|^{p/2 + \mu} \right] \le K,$$

and such that, for $0 \le s < t \le T$,

$$\int_s^t \left\| q_0(r) \right\| dr \le \zeta \sqrt{|t - s| \log(2T / |t - s|)}, \qquad \sum_{j=1}^m \int_s^t \left\| q_j(r) \right\|^2 dr \le \xi \, |t - s|.$$

Then, for any $0 < \epsilon < 1/2$, there exists a random variable $C_\epsilon \in L^p(\Omega)$ such that

$$\omega_Y(h) \le C_\epsilon h^{1/2 - \epsilon} \qquad for \ all \ s, t \in [0, T]$$

and $\| C_\epsilon \|_{L^p(\Omega)}$ depends only on ϵ, μ, p, K, T and is independent of q_j.

Proof. Under these conditions, Fischer & Nappo (2009) tell us there exists $C > 0$ depending only on p, μ, and K such that

$$\left\| \omega_Y(h) \right\|_{L^p(\Omega)} \leq C \sqrt{h \log(2T/h)}.$$

Then, for $\epsilon > 0$, there exists a $C_\epsilon \in L^p(\Omega)$ such that $\omega_Y(h) \leq C_\epsilon h^{1/2-\epsilon}$ and hence

$$|Y(s) - Y(t)| \leq C_\epsilon |s - t|^{1/2-\epsilon}, \qquad s, t \in [0, T].$$

The $L^p(\Omega)$ norm of C_ϵ is bounded uniformly in ϵ, μ, and K, as required. $\qquad\square$

The following is a consequence of the well-known Azuma inequality (Azuma, 1967; DasGupta, 2011).

Lemma A.3. *Let Y_i be a sequence of scalar random variables with $|Y_i| \leq \overline{Y}$ a.s for a constant $\overline{Y}$. Suppose that $\mathbb{E}[Y_i | Y_0, \ldots, Y_{i-1}] = 0$ a.s. for $i = 1, \ldots, n$. Let $S_n = Y_1 + \cdots + Y_n$. Then, for $\lambda > 0$,*

$$\mathbb{P}\big(|S_n - S_k| \geq \lambda\big) \leq 2 \exp\left(-\frac{\lambda^2}{2(n-k)\overline{Y}^2} \right).$$

References

Azuma, Kazuoki. 1967. Weighted sums of certain dependent random variables. *Tôhoku Math. J. (2)*, **19**, 357–367.

DasGupta, Anirban. 2011. *Probability for Statistics and Machine Learning.* Springer Texts in Statistics. Springer.

Davie, A. M. 2007. Differential equations driven by rough paths: an approach via discrete approximation. *Appl. Math. Res. Express.*, **2**, 40.

Fischer, M., & Nappo, G. 2009. On the moments of the modulus of continuity of Itô processes. *Stochastic Analysis and Applications*, **28**(1), 103–122.

Friz, Peter K., & Victoir, Nicolas B. 2010. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications.* Cambridge University Press.

Gaines, J. G., & Lyons, T. J. 1997. Variable step size control in the numerical solution of stochastic differential equations. *SIAM J. Appl. Math.*, **57**(5), 1455–1484.

Gyöngy, István. 1998. A note on Euler's approximations. *Potential Anal.*, **8**(3), 205–216.

Kloeden, P. E., & Neuenkirch, A. 2007. The pathwise convergence of approximation schemes for stochastic differential equations. *LMS J. Comput. Math.*, **10**, 235–253.

Kloeden, Peter E., & Platen, Eckhard. 1992. *Numerical Solution of Stochastic Differential Equations.* Applications of Mathematics, no. 23.

Milstein, G. N. 1995. *The Numerical Integration of Stochastic Differential Equations.* Kluwer.

Milstein, G. N., & Tretyakov, M. V. 1999. Simulation of a space-time bounded diffusion. *Ann. Appl. Probab.*, **9**(3), 732–779.