Adaptive timestepping strategies for nonlinear stochastic systems
Kelly, Conall; Lord, Gabriel James

Published in:
IMA Journal of Numerical Analysis

DOI:
10.1093/imanum/dr036

Publication date:
2017

Document Version
Peer reviewed version

Link to publication in Heriot-Watt University Research Portal

Citation for published version (APA):
Adaptive timestepping strategies for nonlinear stochastic systems

Cónall Kelly†
Department of Mathematics, The University of the West Indies, Mona, Kingston 7, Jamaica

AND

Gabriel J. Lord‡
Maxwell Institute, Department of Mathematics, MACS, Heriot-Watt University, Edinburgh, EH14 4AS, UK

[Received on 8 November 2016; revised on XX April 2017]

We introduce a class of adaptive timestepping strategies for stochastic differential equations with non-Lipschitz drift coefficients. These strategies work by controlling potential unbounded growth in solutions of a numerical scheme due to the drift. We prove that the Euler-Maruyama scheme with an adaptive timestepping strategy in this class is strongly convergent. Specific strategies falling into this class are presented and demonstrated on a selection of numerical test problems. We observe that this approach is broadly applicable, can provide more dynamically accurate solutions than a drift-tamed scheme with fixed stepsize, and can improve MLMC simulations.

Keywords: Stochastic differential equations; Adaptive timestepping; Euler-Maruyama method; Locally Lipschitz drift coefficient; Strong convergence.

1. Introduction

We investigate adaptive timestepping for the numerical approximation of a $d$-dimensional stochastic differential equation (SDE) of Itô type

$$
\begin{align*}
\dot{X}(t) &= f(X(t))dt + g(X(t))dW(t), \quad t > 0, \\
X(0) &\in \mathbb{R}^d,
\end{align*}
$$

(1.1)

where $W$ is an $m$-dimensional Wiener process and the drift coefficient $f$ is not globally Lipschitz continuous, but rather satisfies a one-sided Lipschitz condition and a polynomial growth condition.

Since it was pointed out in ? that the Euler-Maruyama method fails to converge in the strong sense for such equations, there has been much interest in tamed numerical methods, the first of which was presented in ? (see (1.8) in Section 1.1). We also refer the reader to the variant presented in ?, and to the related class of truncated methods which may be found in, for example, ?. Generally, speaking, these methods work by enforcing a higher order modification to the drift and (if necessary) diffusion coefficients in order to control unbounded growth permitted by non-globally Lipschitz coefficients. The idea has been extended to higher order schemes [?], to SDEs with Lévy noise [?], and to stochastic partial differential equations (SPDEs) [?].

†Corresponding author. Email conall.kelly@gmail.com. Supported by a SQuaRE activity entitled “Stochastic stabilisation of limit-cycle dynamics in ecology and neuroscience” funded by the American Institute of Mathematics.

‡Email: g.j.lord@hw.ac.uk

© The author 2017. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.
However, as noted in \cite{KELLY2018}, (fully) tamed methods can lead to dynamically inaccurate results for even moderately small step-sizes, due at least in part to the perturbation of the flow that results from modifying the coefficients. We illustrate this further in Section 3 when we show, for example, that the drift-tamed Euler-Maruyama method does not give a good approximation of the period for the stochastic Van der Pol oscillator. Moreover, while implicit schemes like the backward Euler method can still be strongly convergent in this setting (see for example \cite{KELLY2018}) their implementation requires the solution of a nonlinear algebraic equation at each timestep, and can therefore be highly inefficient. This is particularly true for large systems of SDEs arising, as in Section 3.5, from space discretisation of SPDEs.

In this article, we propose an alternative approach to the control of growth arising from a non globally-Lipschitz drift coefficient. Rather than modifying the drift directly, we adjust the length of the timestep taken at each iteration in order to control the norm of the drift response. In spirit this idea is closer to the projected Euler and Milstein methods given in \cite{KELLY2018}, where solutions are prevented from leaving a ball, the radius of which is dependent on the step-size. We also point out \cite{KELLY2018}, where adaptive timestepping was used to control solution dynamics, and in particular to preserve the positivity of solutions of the numerical discretisation of nonlinear SDEs. The recent preprint of \cite{KELLY2018} takes a related approach; see Remark 2.2 in Section 2.4 for a comparative discussion.

Otherwise, adaptive timestepping for SDEs has tended to concentrate on local error control; see for example \cite{KELLY2018}. A serious drawback of using adaptive methods for SDEs is the potential requirement to interpolate the Brownian path in the case that a timestep is rejected. This is not necessary for the method we propose here, as long as the diffusion coefficient satisfies a global Lipschitz condition.

The structure of the article is as follows. The remainder of the introduction lays out the mathematical framework for the article, and summarises relevant results from the literature. In Section 2 we describe the Euler-type discretisation with random stepsize that forms the basis of our scheme. We demonstrate how stepsize controls can be motivated, either by ensuring that the discretised drift coefficient responds similarly to that of a scheme which is known to converge strongly (e.g. tamed Euler), or by examining the dynamics of the discrete drift map. Finally, we define a class of admissible timestepping strategies for (1.1), provide examples, and state the strong convergence theorem that is our main result.

In Section 3 we investigate our methods with two adaptive timestepping strategies and compare their performance to a tamed Euler method with fixed stepsize for eight test problems, illustrating convergence and reporting the details of stepsizes chosen by each strategy. In particular for the stochastic Van der Pol oscillator we see that the fixed-step tamed Euler method consistently underestimates the period but that adaptive methods give a better approximation. We also examine a multi-level Monte Carlo (MLMC) approximation with adaptive timestepping and observe that this approach reduces the variance on each level, leading to fewer realisations and hence reducing the computational cost. In Section 4 we provide the proof of our main result. Our conclusions and a short discussion of possible future directions for this work are in Section 5.

1.1 Mathematical preliminaries

Consider the $d$-dimensional Itô-type SDE (1.1). For the remainder of the article we let $(\mathcal{F}_t)_{t\geq0}$ be the natural filtration of $W$. Suppose $f: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is continuously differentiable with derivative that grows at most polynomially: for some $c \in (0, \infty)$

$$\|Df(x)\| \leq c(1 + \|x\|^c);$$

and satisfies a one-sided Lipschitz condition with constant $\alpha > 0$:

$$\langle f(x) - f(y), x - y \rangle \leq \alpha \|x - y\|^2.$$
Suppose also that \( g : \mathbb{R}^d \to \mathbb{R}^{d \times m} \) is continuously differentiable and satisfies a global Lipschitz condition with constant \( \kappa > 0 \):

\[
\|g(x) - g(y)\|_F \leq \kappa \|x - y\|. \tag{1.4}
\]

Under conditions (1.2)–(1.4), (1.1) has a unique strong solution on any interval \([0, T]\), where \( T < \infty \) on the filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\). Moreover, the following moment bounds apply over any finite interval \([0, T]\):

**Lemma 1.1** Let \( f, g \) be \( C^1 \) functions satisfying (1.3) and (1.4) respectively. Then for each \( p > 0 \) there is \( C = C(p, T, X(0)) > 0 \) such that

\[
\mathbb{E} \left[ \sup_{s \in [0,T]} \|X(s)\|^p \right] \leq C. \tag{1.5}
\]

This was proved as Lemma 3.2 in \cite{??} for \( p > 2 \), which can be extended to include \( 0 < p \leq 2 \) via Jensen’s inequality.

The following bound is used to develop timestepping strategies in Section 2.4, and in the proof of our main theorem.

**Lemma 1.2** The polynomial bound on the derivative of \( f \) given by (1.2) implies

\[
\|f(x)\| \leq c_1 (1 + \|x\|^{(c+1)}) \tag{1.6}
\]

where \( c_1 := 2c + \|f(0)\| \).

**Proof.** See, for example, Lemma 3.1 in \cite{??}.

The Euler-Maruyama method and the notion of strong convergence used in this article are as follows.

**Definition 1.1** Fix \( T < \infty \) and \( N \in \mathbb{N} \), and define \( h = T/N \). The Euler-Maruyama discretisation of (1.1) over the interval \([0, T]\) with \( N \) steps is given by

\[
X^n_{n+1} = X^n_n + hf(X^n_n) + g(X^n_n)(W((n+1)h) - W(nh)), \quad X_0 = X(0), \quad n = 0, \ldots, N, \tag{1.7}
\]

**Definition 1.2** If there exists \( p \in [1, \infty) \) and constants \( C_p, \beta > 0 \) such that

\[
\left( \mathbb{E} \left[ \|X(T) - X_T^N\|^p \right] \right)^{1/p} \leq C_p h^\beta,
\]

then the Euler-Maruyama method given by (1.7) is said to converge strongly with order \( \beta \) in \( L_p \) to solutions of (1.1) at time \( T \).

In the scalar single noise case, \cite{??}, Theorem 1, showed that the Euler-Maruyama method given in (1.7) cannot converge strongly if at least one of the coefficients grows superlinearly. We restate their result here:

**Theorem 1.3** Let \( d = m = 1, \) and let \( C \geq 1, \beta > \alpha > 1 \) be constants such that

\[
\max \{ |f(x)|, |g(x)| \} \geq |x|^\beta / C \quad \text{and} \quad \min \{ |f(x)|, |g(x)| \} \leq C|x|^{\alpha}
\]

for all \( |x| \geq C \). If the exact solution of (1.1) satisfies \( \mathbb{E} [X(T)]^p < \infty \) for some \( p \in [1, \infty) \), then

\[
\lim_{N \to \infty} \mathbb{E} [X(T) - X_T^N]^p = \infty \quad \text{and} \quad \lim_{N \to \infty} \mathbb{E} [X(T)]^p - \mathbb{E} [X_T^N]^p = \infty.
\]
The drift-tamed Euler–Maruyama method given by
\[ Y_{n+1}^N = Y_n^N + \frac{h f(Y_n^N)}{1 + h \| f(Y_n^N) \|} + g(Y_n^N)(W((n+1)h) - W(nh)), \quad n = 0, \ldots, N, \tag{1.8} \]
was introduced in \cite{2} to provide an explicit numerical method that would display strong convergence in circumstances where the Euler-Maruyama method does not. In fact, strong convergence was proved under Conditions (1.2)–(1.4). The following theorem states two key results from that article: the first on boundedness of moments, the second on strong convergence of a sequence of continuous time interpolants \{\tilde{Y}^N\} to the true solution. For reasons of space, we refer readers to \cite{2} for the specific construction of these interpolants.

**Theorem 1.4** \cite{2} Let \(X(t)\) be a solution of (1.1), where \(f\) and \(g\) satisfy Conditions (1.2)–(1.4). Let \(\{Y_n^N\}\) be a solution of (1.8). Then
\[
\sup_{n \in \mathbb{N}} \sup_{n \in \{0, 1, \ldots, N\}} \mathbb{E}[\| Y_n^N \|^p] < \infty. \tag{1.9}
\]

Let \(\{T_N\}\) be a particular sequence of continuous time interpolants of the time discrete approximation (1.8). There exists a family \(C_p (p \in [1, \infty))\) of real numbers such that
\[
\left( \mathbb{E} \left[ \sup_{t \in [0,T]} \| X(t) - \tilde{Y}_t^N \|^p \right] \right)^{1/p} \leq C_p h^{1/2},
\]
for all \(N \in \mathbb{N}\) and all \(p \in [1, \infty)\).

\cite{2} showed that the Euler-Maruyama scheme (1.7) is strongly convergent over the interval \([0, T]\) if its moments are bounded in the sense of (1.9). In Section 2 we show how stepsize control can be used to bound the drift response pathwise, sufficient to ensure strong convergence.

### 2. Adaptive timestepping strategies

#### 2.1 Euler-type schemes with random timesteps

Consider the following Euler-type method for (1.1) over a random mesh \(\{t_n\}_{n \in \mathbb{N}}\) on the interval \([0, T]\) given by
\[
Y_{n+1} = Y_n + h f(Y_n) + g(Y_n) (W(t_{n+1}) - W(t_n)), \quad Y_0 = X_0, \quad n < N, \tag{2.1}
\]
where \(\{h_n\}_{n \in \mathbb{N}}\) is a sequence of random timesteps, and \(t_N = \sum_{i=1}^{N} h_i\) with \(t_0 = 0\). The random time step \(h_{n+1}\) (and the corresponding point on the random mesh \(t_{n+1}\)) is determined by \(Y_n\).

**Definition 2.1** Suppose that each member of the sequence \(\{t_n\}_{n \in \mathbb{N}}\) 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 \(W\). We may then define a discrete-time filtration \(\mathcal{G}_n\) by
\[
\mathcal{G}_n = \{ A \in \mathcal{F} : A \cap \{ t_n \leq t \} \in \mathcal{F}_t \}, \quad n \in \mathbb{N}.
\]

**Assumption 2.2** Suppose that each \(h_n\) is \(\mathcal{G}_{n-1}\)-measurable, and let \(N\) be a random integer such that
\[
N := \max\{ n \in \mathbb{N} : t_{n-1} < T\} \quad \text{and} \quad t_N = T.
\]
In addition let \(h_n\) satisfy the following constraint: minimum and maximum stepsizes \(h_{\min}\) and \(h_{\max}\) are imposed in a fixed ratio \(0 < \rho \in \mathbb{R}\) so that
\[
h_{\max} = \rho h_{\min}. \tag{2.2}
\]
In Assumption 2.2, the lower bound \( h_{\min} \) ensures that a simulation over the interval \([0, T]\) can be completed in a finite number of timesteps. In the event that at time \( t_n \) we compute \( h_{n+1} = h_{\min} \), we apply a single step of the drift-tamed Euler method (1.8) over a timestep of length \( h = h_{\min} \), rather than (2.1). Therefore the adaptive timestepping scheme under investigation in this article is

\[
Y_{n+1} = Y_n + h_{n+1} \left[ f(Y_n) \mathcal{J}_{(h_{n+1}>h_{\min})} + \frac{f(Y_n)}{1+h_{\min} \|f(Y_n)\|} \mathcal{J}_{(h_{n+1}=h_{\min})} \right] + g(Y_n) (W(t_{n+1}) - W(t_n)), \quad n = 0, \ldots, N - 1. \tag{2.3}
\]

The upper bound \( h_{\max} \) prevents stepsizes from becoming too large and allows us to examine the strong convergence of the adaptive method (2.3) to solutions of (1.1) as \( h_{\max} \to 0 \) (and hence as \( h_{\min} \to 0 \)).

**Remark 2.1** In (2.3), note that each \( W(t_{n+1}) - W(t_n) \) is a Wiener increment taken over a random step of length \( h_{n+1} \) which itself may depend on \( Y_n \), and therefore is not necessarily normally distributed. However, if \( h_{n+1} \) is an \( \mathcal{F}_n \)-stopping time then \( W(t_{n+1}) - W(t_n) \) is \( \mathcal{F}_n \)-conditionally normally distributed with, almost surely (a.s.),

\[
\mathbb{E} \left[ \|W(t_{n+1}) - W(t_n)\| \mid \mathcal{F}_n \right] = 0, \quad \mathbb{E} \left[ \|W(t_{n+1}) - W(t_n)\|^2 \mid \mathcal{F}_n \right] = h_{n+1}.
\]

In practice therefore, we can replace the sequence of Wiener increments with i.i.d. \( \mathcal{N}(0,1) \) random variables denoted \( \{\xi_n\}_{n=1}^N \), scaled at each step by the \( \mathcal{F}_n \)-measurable random variable \( \sqrt{h_{n+1}} \).

In Sections 2.2 and 2.3 we provide two motivating discussions, each illustrating how a timestepping strategy can be designed. The first focuses on the properties of the drift-tamed Euler method (1.8), the second on the local dynamics of polynomial maps. In Section 2.4 we set out a sufficient set of conditions for such strategies to ensure that solutions of (2.3) converge strongly to those of (1.1).

### 2.2 Stepsize selection via the drift-tamed Euler map

For the stochastic differential equation (1.1) the explicit Euler and drift-tamed Euler maps associated with the drift coefficient \( f \) are

\[
F_h(y) = y + hf(y) \quad \text{and} \quad \bar{F}_h(y) = y + \frac{hf(y)}{1+h\|f(y)\|}
\]

respectively. At each timestep we choose \( h(y) \) so that

\[
\|F_h(y) - \bar{F}_h(y)\| = \frac{h^2 \|f(y)\|^2}{1+h\|f(y)\|} < \varepsilon \tag{2.4}
\]

for some tolerance \( \varepsilon > 0 \). Equivalently we have \( h^2 \|f(y)\|^2 - \varepsilon h\|f(y)\| - \varepsilon < 0 \) and so require \( h \) such that

\[
\frac{\varepsilon - \sqrt{\varepsilon^2 + 4\varepsilon}}{2\|f(y)\|} < h < \frac{\varepsilon + \sqrt{\varepsilon^2 + 4\varepsilon}}{2\|f(y)\|}.
\]

Since \( \varepsilon - \sqrt{\varepsilon^2 + 4\varepsilon} < 0 \) for all \( \varepsilon > 0 \) we are left with the requirement that

\[
0 < h < \frac{1}{\|f(y)\|} \left[ \frac{\varepsilon + \sqrt{\varepsilon^2 + 4\varepsilon}}{2} \right],
\]
for (2.4) to hold. This leads to an adaptive strategy

\[ h_{n+1}(Y_n) = \max \left\{ h_{\text{min}}, \min \left\{ h_{\text{max}} \frac{1}{\|f(Y_n)\|} \left[ \varepsilon + \sqrt{\varepsilon^2 + \frac{4\varepsilon}{2}} \right] \right\} \right\}. \] (2.5)

By construction, each term in the sequence \( \{h_n\}_{n \in \mathbb{N}} \) is an \( \mathcal{G}_{n-1} \)-measurable random variable, and Assumption 2.2 holds.

2.3 Step size selection via local dynamics

Consider the drift coefficient function

\[ f(x) = -\gamma |x|^\nu, \quad x \in \mathbb{R}, \] (2.6)

where \( \gamma, \nu > 0 \). The associated Euler map with step size \( h \) is given by the function

\[ F_h(x) = x - h\gamma |x|^\nu, \quad x \in \mathbb{R}. \] (2.7)

Consider the discrete-time dynamics of the map given by (2.7) (a more detailed analysis may be found in ?). The difference equation \( x_{n+1} = F_h(x_n) \) has a stable equilibrium solution at zero and an unstable two-cycle at \( \pm \sqrt{2/(h\gamma)} \). So the basin of attraction of the zero solution is \( |x_0| < \sqrt{2/(h\gamma)} \). For fixed \( \gamma \), we can increase the size of the basin of attraction arbitrarily by choosing \( h \) sufficiently small. Moreover, the derivatives are

\[ F_h'(x) = \begin{cases} 1 - h\gamma (v + 1)|x|^v, & x \geq 0, \\ 1 + h\gamma (v + 1)|x|^v, & x < 0, \end{cases} \]

and so outside of the basin of attraction, repeated applications of the map induce oscillations that grow rapidly at a rate determined by \( \nu \). At each iteration, a stochastic perturbation with non-compact support can move trajectories outside the basin of attraction and into a regime characterised by rapidly growing oscillation.

This suggests an adaptive timestepping strategy motivated by the control of stability. Our approach is as follows. For (1.1) with drift coefficient given by (2.6), we select each stepsize to be (for an appropriately chosen \( \delta > 0 \)),

\[ h_{n+1} = \max \left\{ h_{\text{min}}, \min \left\{ h_{\text{max}}, \frac{\delta}{\gamma |Y_n|^\nu} \right\} \right\}. \] (2.9)

This ensures that if the solution moves out of the basin of attraction of the unperturbed equation then the stepsize is decreased so that it is included once again.

This strategy can be extended to equations with a drift coefficient satisfying the polynomial bound

\[ \|f(x)\| \geq \|x\|^{\beta}/C, \] (2.8)

for \( C \geq 1, \beta > 1 \) and all \( \|x\| \geq C \), by considering the basin of attraction of the Euler map corresponding to the polynomial bound on \( f \). This suggests the following adaptation strategy:

\[ h_{n+1} = \max \left\{ h_{\text{min}}, \min \left\{ h_{\text{max}}, \frac{\delta}{\|Y_n\|^\beta - 1} \right\} \right\}. \] (2.9)
for equations with drift satisfying (2.8) with $\beta$ an odd integer and an appropriately chosen $\delta \leq h_{\max}$. More generally, if we consider the growth over a single step of a perturbation $v$

\[ v_{\text{new}} = (I + hDf)v \]

so that

\[ h = \frac{v_{\text{new}} - v}{Dfv}. \]

then the following strategy is indicated: for some $\delta \leq h_{\max}$, let

\[ h_{n+1} = \max \left\{ h_{\min}, \min \left\{ h_{\max}, \delta \left\| Df(Y_n) \right\| \right\} \right\}. \tag{2.10} \]

2.4 Strong convergence of adaptive timestepping methods

We begin by defining a class of timestepping strategies that guarantee the strong convergence of solutions of (2.3) to solutions of (1.1) by ensuring that, at each step of the discretisation, the norm of the drift response has a pathwise linear bound.

**Definition 2.3** Let \( \{Y_n\}_{n \in \mathbb{N}} \) be a solution of (2.3) where $f$ satisfies (1.2)-(1.3) and $g$ satisfies (1.4). We say that \( \{h_n\}_{n \in \mathbb{N}} \) is an admissible timestepping strategy for (2.3) if Assumption 2.2 is satisfied and there exist real non-negative constants $R_1, R_2 < \infty$ such that whenever $h_{\min} < h_n < h_{\max}$,

\[ \|f(Y_n)\|^2 \leq R_1 + R_2 \|Y_n\|^2, \quad n = 0, \ldots, N - 1. \tag{2.11} \]

In the next Lemma we provide specific examples of admissible timestepping schemes.

**Lemma 2.1** Let \( \{Y_n\}_{n \in \mathbb{N}} \) be a solution of (2.3), let $\delta \leq h_{\max}$, and let $c$ be the constant in (1.6). Let \( \{h_n\}_{n \in \mathbb{N}} \) be a timestepping strategy that satisfies Assumption 2.2. \( \{h_n\}_{n \in \mathbb{N}} \) is admissible for (2.3) if, for each $n = 0, \ldots, N - 1$, one of the following holds

(i) $h_{n+1} \leq \delta/(\|f(Y_n)\|)$,

(ii) $h_{n+1} \leq \delta/(1 + \|Y_n\|^{1+c})$,

(iii) $h_{n+1} \leq \delta \|Y_n\|/(\|f(Y_n)\|)$,

(iv) $h_{n+1} \leq \delta \|Y_n\|/(1 + \|Y_n\|^{1+c})$,

whenever $h_{\min} < h_n < h_{\max}$.

**Proof.** For Part (i) we can apply (2.2):

\[ \|f(Y_n)\|^2 \leq \left( \frac{\delta}{h_{n+1}} \right)^2 \leq \frac{h_{\max}^2}{h_{\min}^2} = \rho^2, \]

and so (2.11) is satisfied with $R_1 = \rho^2$ and $R_2 = 0$.

For Part (ii), by (1.6) and (2.2) we have

\[ \|f(Y_n)\|^2 \leq (2c + \|f(0)\|)^2(1 + \|Y_n\|^{1+c})^2 \leq (2c + \|f(0)\|)^2 \frac{h_{\max}^2}{h_{n+1}^2} \leq (2c + \|f(0)\|)^2 \rho^2, \]

and so (2.11) is satisfied with $R_1 = (2c + \|f(0)\|)^2\rho^2$ and $R_2 = 0$.

For Parts (iii) and (iv) similar arguments give the bounds $\|f(Y_n)\|^2 \leq \rho^2 \|Y_n\|^2$ and $\|f(Y_n)\|^2 \leq (2c + \|f(0)\|)^2 \rho^2 \|Y_n\|^2$ respectively, so (2.11) is satisfied with $R_1 = 0, R_2 = \rho^2$ for Part (iii), and $R_2 = (2c + \|f(0)\|)^2\rho^2$ for Part (iv).

Our main result shows strong convergence in $\mathcal{L}_2$ with order 1/2 of solutions of (2.3) to solutions of (1.1) when \( \{h_n\}_{n \in \mathbb{N}} \) is an admissible timestepping strategy.
Theorem 2.4 Let \((X(t))_{t \in [0,T]}\) be a solution of (1.1) with initial value \(X(0) = X_0\). Let \(\{Y_n\}_{n \in \mathbb{N}}\) be a solution of (2.3) with initial value \(Y_0 = X_0\) and admissible timestepping strategy \(\{h_n\}_{n \in \mathbb{N}}\) satisfying the conditions of Definition 2.3. Then for some \(C > 0\), independent of \(h_{\text{max}}\),

\[
(\mathbb{E} \left[ \|X(T) - Y_N\|^2 \right])^{1/2} \leq C h_{\text{max}}^{1/2}.
\]

The proof of Theorem 2.4 is a modification of a standard Euler-Maruyama convergence argument accounting for the properties of the random sequences \(\{t_n\}_{n \in \mathbb{N}}\) and \(\{h_n\}_{n \in \mathbb{N}}\), and using (2.11) to compensate for the non-Lipschitz drift. It is presented in Section 4.

It is possible to link the notion of admissibility to the strategies developed via taming and local dynamics in Sections 2.2 and 2.3 as follows. The adaptive timestepping strategy given by (2.5) is admissible for an appropriate choice of tolerance \(\varepsilon\): to see this, let

\[
0 \leq \varepsilon < \frac{h_{\text{max}}^2}{1 + h_{\text{max}}}.
\]

Then (2.5) is equivalent to the strategy given in Part (i) of Lemma 2.1, with

\[
\delta := \frac{\varepsilon + \sqrt{\varepsilon^2 + 4\varepsilon}}{2} < h_{\text{max}}.
\]

We investigate the performance of (2.5) numerically in Section 3.

The adaptive timestepping strategy given by (2.9) in Section 2.3 is equivalent to that given in Part (iii) of Lemma 2.1 when the drift coefficient is precisely the polynomial expression on the right hand side of (2.8), and (2.9) is therefore admissible in that case. For more general drift coefficients the closest correspondence is with Part (iv) of Lemma 2.1, for which a priori knowledge of the polynomial bound parameter \(c\) is needed; in practice this may be difficult to determine. The variant given by (2.10), which uses the norm of the Jacobian of \(f\), is not known to be admissible but neither does it require precise knowledge of \(c\), and we investigate it numerically in Section 3.

Remark 2.2 In [2], an adaptive timestepping strategy is presented which satisfies

\[
\langle Y_n, f(Y_n) \rangle + \frac{1}{2} h_{n+1} \| f(Y_n) \|^2 \leq \alpha \| Y_n \|^2 + \beta, \quad n = 0, \ldots, N - 1,
\]

where the one sided linear bound \(\langle x, f(x) \rangle \leq \alpha \| x \|^2 + \beta\), for \(\alpha, \beta > 0\), has been imposed upon the drift coefficient \(f\). With additional upper and lower bounds on each timestep, and the introduction of a convergence parameter \(\delta \leq 1\), the authors show that the Euler-Maruyama scheme is strongly convergent with order 1/2.

We note that, in Section 3.1 of [2], specific timestepping rules are proposed for two scalar equations with drift satisfying a polynomial bound of the form (2.8) for large arguments: the stochastic Ginzburg Landau equation and the stochastic Verhulst equation. These rules are consistent with the adaptive timestepping strategy given by (2.9). Similarly, in Section 3.2, two specific timestepping rules for multi-dimensional SDEs are proposed, the first of which, within our framework, corresponds to Part (iii) of Lemma 2.1. The second of those rules, within our framework, corresponds to

\[
h_{n+1} \leq \delta \frac{\| Y_n \|^2}{\| f(Y_n) \|^2}.
\]
If we suppose that $\delta \leq h_{\text{max}}$ then we have

$$\|f(Y_n)\|^2 \leq \frac{\delta}{h_{n+1}}\|Y_n\|^2 \leq \rho \|Y_n\|^2,$$

which is admissible for (2.3).

3. Numerical examples

In the numerical experiments below we compare two different adaptive time-stepping strategies for (2.3) with the fixed step drift-tamed Euler-Maruyama scheme (1.8). For the latter we take as the fixed step $h_{\text{mean}}$ the average of all timesteps $h_n^{(m)}$ over each path and each realisation $m = 0, 1, \ldots, M$ so that

$$h_{\text{mean}} = \frac{1}{M} \sum_{m=1}^M \frac{1}{N^{(m)}} \sum_{n=1}^{N^{(m)}} h_n^{(m)}.$$

Thus we are comparing to a fixed step scheme of similar average cost. We solve (2.3) with the taming inspired adaptive timestepping strategy (2.5) and denote this AT. The fixed step comparison using $h_{\text{mean}}$ computed from AT is denoted FT. Similarly, we solve (2.3) with the local dynamics inspired adaptive timestepping scheme (2.10) which we denote ALD and the fixed step comparison is denoted FLD.

3.1 A stochastic Ginzburg Landau equation

This equation arises from the theory of superconductivity and takes the form

$$dX(t) = \left( (\eta + \frac{1}{2} \sigma^2)X(t) - \lambda X(t)^3 \right)dt + \sigma G(X(t))dW(t), \quad X(0) = x_0 > 0,$$

for $t \geq 0$, and where $\eta \geq 0$ and $\lambda, \sigma > 0$. When $G(X) = X$, the explicit form of the solution over $[0, \infty)$, provided by ?, is

$$X(t) = \frac{x_0 \exp(\eta t + \sigma W(t))}{\sqrt{1 + 2\sigma \sqrt{\lambda} \int_0^t \exp(2\eta s + 2\sigma W(s))ds}}, \quad t \geq 0. \quad (3.2)$$

We use this exact solution to illustrate numerically the strong convergence result of Theorem 2.4, see Figure 1, computing to a final time of $T = 2$ with 100 realisations. We compare in Figure 1 (a) all four methods AT, FT, ALD and FLD and show reference lines of slope 1 and $1/2$. Note that the global error of the adaptive methods at time $T$ is close to that computed with the mean step $h_{\text{mean}}$ by the fixed step method. In (b) we show comparison of estimated rates of strong convergence and root mean square error (RMS) error against the CPU time between the adaptive methods AT, ALD and the fixed step tamed Euler methods FT, FLD. We see there is a slight computational overhead in performing the adaptive step, as expected. In Figure 2 we examine convergence for (3.1) with additive noise (taking $G(X) = 1$). As we do not have an exact solution we use a reference solution computed with (1.8) using $h = 10^{-5}$. We observe, as for a standard Euler-Maruyama method, an improvement in the rate of convergence to order 1 for the adaptive methods AT, ALD as well as the fixed step schemes FT and FLD (see also Remark 4.1). Comparing with $h_{\text{mean}}$ leads to similar errors and we again note a slight computational overhead to account for the adaptive step in the algorithm.
(a) (b)

<table>
<thead>
<tr>
<th>Mean Timestep</th>
<th>RMS Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>10^{-4}</td>
<td>AT</td>
</tr>
<tr>
<td>10^{-3}</td>
<td>ALD</td>
</tr>
<tr>
<td>10^{-2}</td>
<td>FT</td>
</tr>
<tr>
<td>10^{-1}</td>
<td>FLD</td>
</tr>
</tbody>
</table>

**Fig. 1.** (a): A numerical demonstration of strong convergence for multiplicative noise as the mean stepsize decreases for the adaptive methods AT, ALD and fixed step methods FT, FLD applied to Eq. (3.1) with parameters $\eta = 0.1$, $\lambda = 2$ and $\sigma = 0.5$, $T = 2$. (b) plot showing the efficiency and reduction in root mean square (RMS) error as the CPU time (s) increases and $h_{\text{max}}$ decreases. For each $h_{\text{max}}$ value $\rho = 100$.

(a) (b)

<table>
<thead>
<tr>
<th>Mean Timestep</th>
<th>RMS Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>10^{-4}</td>
<td>AT</td>
</tr>
<tr>
<td>10^{-3}</td>
<td>ALD</td>
</tr>
<tr>
<td>10^{-2}</td>
<td>FT</td>
</tr>
<tr>
<td>10^{-1}</td>
<td>FLD</td>
</tr>
</tbody>
</table>

**Fig. 2.** (a): A numerical demonstration of strong convergence for additive noise as the mean stepsize decreases for the adaptive methods AT, ALD and fixed step methods FT, FLD applied to Eq. (3.1) with parameters $\eta = 0.1$, $\lambda = 2$ and $\sigma = 0.5$, $T = 2$. (b) plot showing the efficiency and reduction in root mean square (RMS) error as the CPU time (s) increases and $h_{\text{max}}$ decreases. For each $h_{\text{max}}$ value $\rho = 100$. 
3.2 The stochastic Van der Pol oscillator

This is a stochastic additive noise version of the van der Pol oscillator, which describes the effect of external noise on stable oscillations, and takes the form

\[
d \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} X_2(t) \\ (1 - (X_1(t))^2)X_2(t) - X_1(t) \end{pmatrix} dt + \begin{pmatrix} 0 \\ dW(t) \end{pmatrix}. \tag{3.3}
\]

In Figure 3 we show two realisations for (3.3) obtained using the drift-tamed Euler–Maruyama scheme (1.8) with \( h = 10^{-4} \). We clearly see periodic behaviour over the interval \([0, T]\). We ask how well the period is captured by the adaptive methods AT and ALD and by the fixed step methods FT and FLD. Figure 4 compares two realisations computed using the same paths for \( W(t) \), so that the path in (a) is the same as that in (b) (similarly for (c) and (d). The fixed step methods FT and FLD in (b) and (d) appear to have fewer oscillations than the adaptive simulations in (a) and (c) (and Figure 3 (a)).

This is borne out in Table 1 which compares data on the estimated mean period and variance from 100 realisations of (3.3) for \( t \in [0, 100] \). We also include maximum and minimum periods observed. The adaptive methods AT and ALD both give a better estimate of the period than the equivalent fixed step methods and have a smaller relative error. We also note that AT uses, on average, smaller steps than ALD and has a smaller relative error. For the equivalent fixed step schemes FT and FLD the error for these different timesteps are similar.

In Table 2 we examine for \( T = 200 \) the timesteps \( h_n \) taken by AT and ALD for different values of \( \rho \) with \( h_{\text{max}} = 2 \). We report \( h_{\text{mean}} \), along with the timestep variance, the minimum and maximum timesteps, the computational time taken, and the percentage of timesteps taken at the minimum \( h_{\text{min}} \). We see that for \( \rho \) large enough \( h_{\text{min}} \) is not reached often and the frequency with which this occurs for \( \rho = 100 \) (where \( h_{\text{min}} = 0.02 \)) is similar to that for \( \rho = 1000 \) (where \( h_{\text{min}} = 0.002 \)).
Fig. 4. Sample realisations to (3.3) obtained using AT (a) compared to $h = 0.0838$ for FT in (b). In (c) we use ALD and compare to FLD with $h = 0.1269$ in (d). Here $\rho = 100$ and $h_{\text{max}} = 1$. Note that the paths in (a) and (b) (and (c) and (d)) are the same and the difference arises from the timestepping.

<table>
<thead>
<tr>
<th></th>
<th>Rel. Error</th>
<th>Mean Period</th>
<th>Var</th>
<th>Min</th>
<th>Max</th>
<th>$h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE (1.8)</td>
<td>0.089037</td>
<td>6.684832</td>
<td>0.294930</td>
<td>5.555556</td>
<td>9.090909</td>
<td>0.0005</td>
</tr>
<tr>
<td>AT</td>
<td>0.213953</td>
<td>8.543185</td>
<td>0.484346</td>
<td>5.882353</td>
<td>9.090909</td>
<td>0.1269</td>
</tr>
<tr>
<td>FT</td>
<td>0.213953</td>
<td>8.543185</td>
<td>0.484346</td>
<td>5.882353</td>
<td>9.090909</td>
<td>0.1269</td>
</tr>
<tr>
<td>TE (1.8)</td>
<td>0.183946</td>
<td>6.725343</td>
<td>0.250395</td>
<td>5.555556</td>
<td>8.333333</td>
<td>0.0005</td>
</tr>
<tr>
<td>ALD</td>
<td>0.279599</td>
<td>9.394958</td>
<td>1.132636</td>
<td>7.142857</td>
<td>14.285714</td>
<td>0.120965</td>
</tr>
<tr>
<td>FLD</td>
<td>0.279599</td>
<td>9.394958</td>
<td>1.132636</td>
<td>7.142857</td>
<td>14.285714</td>
<td>0.120965</td>
</tr>
</tbody>
</table>

Table 1. Comparison for the van der Pol equation (3.3) of estimated mean period, variance, minimum period and maximum period based on 100 realisations with $\rho = 100$, $h_{\text{max}} = 1$ and $T = 100$. We also report an estimate of the relative error in the mean period.
Table 2. Comparison of step size data $h_n$ for the stochastic Van der Pol equation (3.3) with additive noise based on 100 realisations. See Table 5 for an example with multiplicative noise.

<table>
<thead>
<tr>
<th>Problem</th>
<th>%</th>
<th>RMS AT (var SE)</th>
<th>RMS FT</th>
<th>$h_{\text{mean}}$ AT</th>
</tr>
</thead>
<tbody>
<tr>
<td>VdP</td>
<td>61%</td>
<td>0.34694 (0.034781)</td>
<td>0.89270 (0.93181)</td>
<td>0.10705 (5.578 × $10^{-4}$)</td>
</tr>
<tr>
<td>SIR</td>
<td>45%</td>
<td>0.66983 (0.88807)</td>
<td>1.22421 (1.1375)</td>
<td>0.15140 (0.0085793)</td>
</tr>
<tr>
<td>LV</td>
<td>90%</td>
<td>0.070716 (2.6658 × $10^{-8}$)</td>
<td>0.675299 (3.8455 × $10^{-4}$)</td>
<td>0.032249 (9.6585 × $10^{-8}$)</td>
</tr>
<tr>
<td>CIR</td>
<td>0%</td>
<td>0.08222</td>
<td>0.082304</td>
<td>1.0000</td>
</tr>
<tr>
<td>PK</td>
<td>-12%</td>
<td>0.055995</td>
<td>0.049778</td>
<td>0.73995</td>
</tr>
<tr>
<td>2D</td>
<td>-12%</td>
<td>0.125057</td>
<td>0.09722</td>
<td>0.34302</td>
</tr>
<tr>
<td>SGLA</td>
<td>-11%</td>
<td>0.26514 (0.0160980)</td>
<td>0.23868 (0.0113570)</td>
<td>0.80255 (0.06370)</td>
</tr>
<tr>
<td>Lang</td>
<td>-42%</td>
<td>0.58753 (1.3885252)</td>
<td>0.41315 (0.088238)</td>
<td>0.086018 (0.0022911)</td>
</tr>
</tbody>
</table>

Table 3. Percentage improvement in RMS error using AT over FT. The numbers in bold show where the fixed step method was more accurate. For half problems there was no advantage, see Table 4 for ALD. In all problems we have 1000 samples apart from for SIR where there are 918 samples for AT and 733 for FT.

<table>
<thead>
<tr>
<th>Problem</th>
<th>%</th>
<th>RMS ALD</th>
<th>RMS FLD</th>
<th>$h_{\text{mean}}$ ALD</th>
</tr>
</thead>
<tbody>
<tr>
<td>VdP</td>
<td>26%</td>
<td>0.256707 (2.107959)</td>
<td>0.034781</td>
<td>0.10705 (5.578 × $10^{-4}$)</td>
</tr>
<tr>
<td>SIR</td>
<td>65%</td>
<td>0.41617 (0.15571)</td>
<td>0.15140 (0.0085793)</td>
<td></td>
</tr>
<tr>
<td>LV</td>
<td>75%</td>
<td>0.170979 (8.1571 × $10^{-7}$)</td>
<td>0.15039 × $10^{-6}$)</td>
<td></td>
</tr>
<tr>
<td>CIR</td>
<td>1%</td>
<td>0.084989</td>
<td>0.085862</td>
<td>1.0000</td>
</tr>
<tr>
<td>PK</td>
<td>49%</td>
<td>0.029680</td>
<td>0.058361</td>
<td>0.36984</td>
</tr>
<tr>
<td>2D</td>
<td>-177%</td>
<td>0.297763</td>
<td>0.107374</td>
<td>0.76021</td>
</tr>
<tr>
<td>SGLA</td>
<td>27%</td>
<td>0.16440 (0.0019026)</td>
<td>0.22668 (0.0074636)</td>
<td>0.46193 (0.044552)</td>
</tr>
<tr>
<td>Lang</td>
<td>76%</td>
<td>0.16387 (0.0039904)</td>
<td>0.69239 (4.604639)</td>
<td>0.013474 (8.6690 × $10^{-5}$)</td>
</tr>
</tbody>
</table>
3.3 A Langevin equation

The following example is taken from \cite{ref}: \[
\begin{align*}
\frac{dX_1(t)}{dt} &= X_2(t) \\
\frac{dX_2(t)}{dt} &= -\frac{1}{2}X_2(t) \left(\frac{4(5X_1(t)^2 + 1)}{5(X_1(t)^2 + 1)}\right)^{2} dt + \frac{4(5X_1(t)^2 + 1)}{5(X_1(t)^2 + 1)} dW(t).
\end{align*}
\] (3.4)

We take \(X(0) = [1, 1]^T\) and solve to \(T = 20\) with \(h_{\text{max}} = 2\). We now examine the choice of \(\rho\). In Table 5 we give the mean step \(h_{\text{mean}}\), variance, minimum and maximum step, computational time and the percentage of steps that were at \(h_{\text{min}}\). Note that both \(h_{\text{max}}\) and \(h_{\text{mean}}\) are larger for AT than for ALD (and a smaller computational time is observed). In Figure 5 we plot the percentage of steps taken at the minimum step size as \(\rho\) is increased for AT and ALD. We see that for small \(\rho\) the minimum step is reached with a high probability (1 when \(\rho = 1\)). As \(\rho\) is increased the minimum step is no longer reached for either scheme (at \(\rho = 10^3\) for (2.10) and \(\rho = 10^4\) for (2.4)). This illustrates that the adaptivity is actively controlling the dynamics (so the scheme is not simply selecting the minimum stepsize at each iteration). Although it is not clear from Figure 5 which of AT or ALD takes larger steps we can see that the step size choice is different and that the variance is smaller for ALD. In some situations it may be preferable not to have large switches in stepsize.

3.4 Comparison of 8 different problems

Both AT and ALD control growth associated with a non-globally-Lipschitz drift term. We now compare the stepsize selection made by each of these strategies over a range of different problems. Mean stepsizes with associated variances are presented in Figure 3.4 for \(\rho = 100\) (a) and \(\rho = 1000\) in (b). We include the stochastic Van der Pol oscillator (3.3) (VdP), the Langevin equation (3.4) (Lang), and the Stochastic Ginzburg-Landau equation (3.1) with \(G(X) = 1\) (additive noise) (SGLA). The other models that we examine can also be found, for example, in \cite{ref}. Note that for certain of these models the coefficients change randomly on each realisation. Since conditions (1.2)–(1.4) on \(f\) and \(g\) fail to hold for all the models, realisations were terminated when the \(L^2\) norm of the approximate solution exceeded a threshold of 10,000 in order to avoid blow-up. The number of times this occurred is indicated in each case below.

In Tables 3 and 4 we examine the root mean square error (RMS) of the adaptive scheme AT and ALD compared to the tamed (1.8) for the eight problems. We have taken \(\rho = 1000\), \(h_{\text{max}} = 1\) and in each case used a reference solution computed \(h = 10^{-4}\) as a true solution and solved to time \(T = 2\). We
observe that the adaptive time stepping methods can have more accurate solutions than the fixed step equivalent, but that the choice of adaptive timestepping strategy is important: ALD seems to perform better than AT with a clear improvement in the error in 6 out of the 8 problems.

SIR: Simulation of the stochastic Susceptible, Infected, Recovered (SIR) model

\[
\begin{align*}
    dX_1(t) &= [-\alpha X_1(t)X_2(t) - \delta X_1(t) + \delta]dt + [-\beta X_1(t)X_2(t)]dW_1(t), \\
    dX_2(t) &= [\alpha X_1(t)X_2(t) - (\gamma + \delta)X_2(t)]dt + [\beta X_1(t)X_2(t)]dW_2(t), \\
    dX_3(t) &= [\gamma X_2(t) - \delta X_3(t)]dt,
\end{align*}
\]

over the simulation interval \([0,T]\), \(T = 2\) with initial data \(X(0) = [0.5; 0.3; 0.2]\). For each simulation we take \(\alpha, \beta, \gamma, \delta \sim U[0, 10]\). For \(\rho = 1000\), 14 of 100 realisations of AT, and 19 of 100 realisations of ALD, were terminated when they exceeded the threshold in \(L_2\) (through becoming negative). For \(\rho = 100\), 17 of 100 realisations of AT, and 2 of 100 realisations of ALD, were terminated when they exceeded the threshold in \(L^2\).

LV: Simulation of the stochastic Lokta-Volterra (LV) model in the well stirred sense

\[
\begin{align*}
    dX_1(t) &= [X_1(t)(\alpha - \beta X_2(t))]dt + \sigma_1X_1(t)dW_1(t), \\
    dX_2(t) &= [X_2(t)(\gamma X_1(t) - \delta)]dt + \sigma_2X_2(t)dW_2(t),
\end{align*}
\]

over the simulation interval \([0,T]\) with \(T = 20\) and initial value \(X(0) = [5, 10]^T\). The parameters \(\alpha, \beta, \gamma, \delta \sim U[0, 1]\) for each realisation and \(\sigma_1 = \sigma_2 = 0.01\). No realisations were terminated for exceeding the threshold in the \(L^2\) norm.

PK: Simulation of a Proto-Kinetics (PK) model. Here \(X\) represents the proportion of one form of a certain protein and therefore should be constrained to the interval \([0, 1]\) and can be modelled by the following SDE

\[
dX(t) = \left[ \frac{1}{2} - X(t) + X(t)(1-X(t)) + \frac{1}{2} X(t)(1-X(t))(1-2X(t)) \right] dt + [X(t)(1-X(t))]dW(t).
\]
We take the simulation step size to be \([0, T], T = 100\). For \(\rho = 1000\), 2 of 100 realisations of AT, and 0 of 100 realisations of ALD, were terminated when they exceeded the threshold in \(L^2\) norm. For \(\rho = 100\), 1 of 100 realisations of AT, and 0 of 100 realisations of ALD, were terminated when they exceeded the threshold in the \(L^2\) norm through becoming negative.

2D: Simulations of the polynomial type SDE

\[
dX(t) = (AX(t) - \beta X(t)|X(t)|^\nu)dt + GdW(t),
\]

where \(A, \beta, G \in \mathbb{R}^{J \times J}\). The simulation interval is \([0, T]\) with \(T = 100\), \(\nu = 2\) and

\[
A = \begin{pmatrix}
0.807019 & 0.589848 \\
0.080506 & 0.477723
\end{pmatrix}, \quad \beta = \begin{pmatrix}
0.99133 & 0.60672 \\
0.29234 & 0.96434
\end{pmatrix}, \quad G = \begin{pmatrix}
0.5 & 0 \\
0 & 0.5
\end{pmatrix}.
\]

No realisations were terminated for exceeding the threshold in \(L^2\). 

CIR: Simulation of special case of the stochastic Cox-Ingersoll-Ross (CIR) model

\[
X(t) = \kappa(\theta - X)dt + \sigma \sqrt{|X|}dW, \quad X(0) = 1
\]

over \([0, T]\) with \(T = 200\) with \(\kappa = 0.1\), \(\theta = 0.5\) and \(\sigma = 0.5\). No realisations were terminated for exceeding the threshold in the \(L^2\) norm.

### 3.5 Stochastic Allen-Cahn SPDE

To investigate adaptive timestepping for a large system of SDEs we consider the discretisation of the Allen-Cahn SPDE

\[
du = [Du_{xx} + u - u^3] dt + \sigma dW,
\]
with \( x \in [0, 1] \), periodic boundary conditions, and initial data \( u(0, x) = \sin(2\pi x) \). The \( Q \)-Wiener process \( W \) is white in time and takes values in \( H^1_{\text{per}} (0, 1) \). We take \( D = 0.01 \) and \( \sigma = 0.5 \) and discretise in space by a spectral Galerkin approximation \( \mathcal{G} \) to get an SDE system in \( \mathbb{R}^{100} \). We take \( h_{\text{max}} = 0.05 \) and \( \rho = 100 \). We show in Figure 3.5 (a) the \( L^2(0, 1) \) norm of one realisation as we solve over \( t \in [0, 10] \) using AT and in Figure 3.5 (b) we plot the corresponding timestep \( h_n \). Note that where the \( L^2(0, 1) \) of the solution becomes small in (a), and hence the non-linearity becomes small, larger steps are taken.

### 3.6 An application to multi-level Monte-Carlo simulation

One major motivation in \cite{HJ2009} for looking at the non-convergence of the Euler-Maruyama method was the recent interest in multi-level Monte-Carlo (MLMC) methods for SDEs; see for example \cite{GJ2008} and \cite{SV2016}. In its basic form the idea is to use a telescoping sum over different numerical approximations (levels) as a form of variance reduction. If we seek to estimate some (Lipschitz) quantity of interest \( Q \) of the solution \( X(T) \) to the SDE we can use approximations with a hierarchy of accuracies from most accurate \( L \) to least accurate \( L_0 \) and compute

\[
\mathbb{E}[Q(X_L)] = \mathbb{E}[Q(X_{L_0})] + \sum_{j=L_0}^{L-1} \mathbb{E}[Q(X_{j+1}) - Q(X_j)].
\]

We can estimate each expectation on the right hand side with a different number of realisations determined according to the method described in \cite{HJ2009}. As \( j \) increases we would expect to take fewer realisations.

We implemented the MLMC method for AT and illustrate results below for the Stochastic Ginzburg-Landau equation with additive noise, i.e. (3.1) with \( G(X) = 1 \). In our implementation we formed each level by imposing a level dependent \( h_{\text{max}}^\ell = h_{\text{max}}^0 k^{-\ell} \), with \( h_{\text{max}}^0 = 1 \) and \( k = 4 \). We compare the number of realisations (and hence computational cost) to those required for the drift-tamed Euler-Maruyama method (1.8). Figure 8 (a) shows a variance reduction at each level when adaptive timestepping is used compared to taking fixed steps. Hence the number of samples required at each level is also reduced in (b). This is consistent with other adaptive timestepping results \cite{HJ2009}.
4. Proof of main result

Lemma 4.1 Let $(X(t))_{t \in [0,T]}$ be a solution of (1.1) with initial value $X(0) = X_0$, and with $f$ and $g$ satisfying conditions (1.2)–(1.4). Let $\{t_n\}_{n \in \mathbb{N}}$ arise from an adaptive timestepping strategy for (2.3) satisfying the conditions of Assumption 2.2. Consider the Taylor expansions of $F$ where the remainders $R$ and $z$ can be taken to read either $f$ or $g$. Then there are a.s. finite and $\mathcal{G}_n$-measurable random variables $\bar{K}_1, \bar{K}_2 > 0$, and constants $K_1, K_2, K_3 < \infty$, the latter three independent of $h_{n+1}$, such that

\begin{align*}
(i) \quad & \mathbb{E} \left[ \left\| \int_{t_n}^{t_{n+1}} R_c(s,t_n,X(t_n)) ds \right\|_{\mathcal{G}_n} \right] \leq \bar{K}_1 h_{n+1}^{3/2}, \quad \text{a.s.;} \\
(ii) \quad & \mathbb{E} \left[ \left\| \int_{t_n}^{t_{n+1}} R_c(s,t_n,X(t_n)) ds \right\|_{\mathcal{G}_n}^2 \right] \leq \bar{K}_2 h_{n+1}^2, \quad \text{a.s.;} \\
(iii) \quad & \mathbb{E}[\bar{K}_1] \leq K_1, \quad \text{and} \quad \mathbb{E}[\bar{K}_2] \leq K_2; \\
(iv) \quad & \mathbb{E} \left[ \bar{K}_1 \left( R_1 + 2R_2 \|X(t_n)\|^2 c_1^2 \left( 1 + 2\|X(t_n)\|^{c+1} + \|X(t_n)\|^{2(c+1)} \right) \right) \right] \leq K_3.
\end{align*}

Proof. Let $t_n$ be a term of $\{t_n\}_{n \in \mathbb{N}}$, and suppose that $t_n < s \leq T$. Then

$$X(s) - X(t_n) = \int_{t_n}^{s} f(X(r)) dr + \int_{t_n}^{s} g(X(r)) dW(r).$$
By the triangle inequality, Jensen’s inequality, and the conditional form of the Itô isometry (see, for
example, Theorem 5.21 in ?),
\[
\mathbb{E} \left[ \|X(s) - X(t_n)\|^2 | \mathcal{G}_n \right] 
\leq 2 \mathbb{E} \left[ \left\| \int_{t_n}^s f(X(r)) \, dr \right\|^2 | \mathcal{G}_n \right] + 2 \mathbb{E} \left[ \int_{t_n}^s \|g(X(r))\|^2 \, dr | \mathcal{G}_n \right] 
\leq 2 \int_{t_n}^s \mathbb{E} \left[ \|f(X(r))\|^2 | \mathcal{G}_n \right] \, dr + 2 \int_{t_n}^s \mathbb{E} \left[ \|g(X(r))\|^2 | \mathcal{G}_n \right] \, dr, \quad \text{a.s.}
\]

Next, we apply (1.4) and (1.6) to get
\[
\mathbb{E} \left[ \|X(s) - X(t_n)\|^2 | \mathcal{G}_n \right] 
\leq 4 \int_{t_n}^s \mathbb{E} \left[ c_1^2 (1 + \|X(r)\|^{2c+2}) | \mathcal{G}_n \right] \, dr + 2 \kappa^2 \int_{t_n}^s \mathbb{E} \left[ \|X(r)\|^2 | \mathcal{G}_n \right] \, dr 
\leq \left( 4 \mathbb{E} \left[ c_1^2 \left( 1 + \sup_{u \in [0,T]} \|X(u)\|^{2c+2} \right) | \mathcal{G}_n \right] + 2 \kappa^2 \mathbb{E} \left[ \sup_{u \in [0,T]} \|X(u)\|^2 | \mathcal{G}_n \right] \right) |s - t_n|, \quad \text{a.s.}
\]

Therefore, by (1.5) in the statement of Lemma 1.1, we can define an a.s. finite and \( \mathcal{G}_n \)-measurable random variable
\[
\tilde{L}_n := \left( 4 \mathbb{E} \left[ c_1^2 \left( 1 + \sup_{u \in [0,T]} \|X(u)\|^{2c+2} \right) | \mathcal{G}_n \right] + 2 \kappa^2 \mathbb{E} \left[ \sup_{u \in [0,T]} \|X(u)\|^2 | \mathcal{G}_n \right] \right), \quad (4.1)
\]
so that
\[
\mathbb{E} \left[ \|X(s) - X(t_n)\|^2 | \mathcal{G}_n \right] \leq \tilde{L}_n |s - t_n|, \quad \text{a.s.} \quad (4.2)
\]

Now consider Part (i) with \( R_f \). By (1.2), and the Cauchy-Schwarz inequality
\[
\mathbb{E} \left[ \|R_f(s,t_n,X(t_n))\| | \mathcal{G}_n \right] 
\leq c_1 \mathbb{E} \left[ \int_0^1 (1 + \|X(t_n) + \tau(X(s) - X(t_n))\|^{c}) \|X(s) - X(t_n)\| \, d\tau | \mathcal{G}_n \right] 
\leq c_1 \sqrt{\mathbb{E} \left[ \|X(s) - X(t_n)\|^2 | \mathcal{G}_n \right]} \sqrt{\mathbb{E} \left[ \int_0^1 (1 + \|X(t_n) + \tau(X(s) - X(t_n))\|^{c})^2 \, d\tau | \mathcal{G}_n \right]}, \quad \text{a.s.}
\]

By (1.5) in the statement of Lemma 1.1 we can define an a.s. finite and \( \mathcal{G}_n \)-measurable random variable
\[
\tilde{M}_n := \mathbb{E} \left[ 2c_1^2 + 18c_1^2 \sup_{u \in [0,T]} \|X(u)\|^{2c} | \mathcal{G}_n \right]
\]
and so, by (4.2),
\[
\mathbb{E} \left[ \|R_f(s,t_n,X(t_n))\| | \mathcal{G}_n \right] \leq \sqrt{\tilde{M}_n \tilde{L}_n} \sqrt{|s - t_n|}, \quad \text{a.s.}
\]
Since \( t_{n+1} \) is an \( \mathcal{G}_n \)-measurable random variable, there is an a.s. finite and \( \mathcal{G}_n \)-measurable random variable \( 0 < K_1 := \frac{2}{3} \sqrt{M_n L_n} \) such that
\[
\mathbb{E} \left[ \left\| \int_{t_n}^{t_{n+1}} R_f(s, t_n, X(t_n)) ds \right\|_{\mathcal{G}_n} \right] \leq \int_{t_n}^{t_{n+1}} \mathbb{E} \left[ \| R_f(s, t_n, X(t_n)) \|_{\mathcal{G}_n} \right] ds \\
\leq \sqrt{M_n L_n} \int_{t_n}^{t_{n+1}} \sqrt{|s-t_n|} ds \leq K_1 h_{n+1}^{3/2}, \text{ a.s.}
\]

For Part (i) with \( R_g \), the same approach using the global Lipschitz condition (1.4) instead of (1.3) yields the result.

Now consider Part (ii) with \( R_f \). We have by (1.5) and the Cauchy-Schwarz inequality
\[
\mathbb{E} \left[ \left\| \int_{t_n}^{t_{n+1}} R_f(s, t_n, X(t_n)) ds \right\|_{\mathcal{G}_n} \right] \leq c \mathbb{E} \left[ \left\| \int_{t_n}^{t_{n+1}} \int_0^1 Df((X(t_n) + \tau(X(s) - X(t_n)))(X(s) - X(t_n)) d\tau ds \right\|_{\mathcal{G}_n} \right] \\
\leq c \mathbb{E} \left[ \left\| \int_{t_n}^{t_{n+1}} ds \right\|_{\mathcal{G}_n} \sup_{u \in [t_n, t_{n+1}]} \int_0^1 (1 + \|X(t_n) + \tau(X(u) - X(t_n))\| \|X(u) - X(t_n)\|^2 d\tau \right\|_{\mathcal{G}_n} \right] \\
\leq c \mathbb{E} \left[ \left\| \int_{t_n}^{t_{n+1}} ds \right\|_{\mathcal{G}_n} \sup_{u \in [0, T]} 4 \int_0^1 (1 + 1 + 2\tau)^c \|X(u)\|^2 d\tau \right\|_{\mathcal{G}_n} \right] \\
\leq K_2 h_{n+1}^2, \text{ a.s.,}
\]
where \( K_2 \) is the a.s. finite and \( \mathcal{G}_n \)-measurable random variable
\[
\bar{K}_2 := \sqrt{128c^2 \mathbb{E} \left[ \sup_{u \in [0, T]} (\|X(u)\|^4 + 34c^4 \|X(u)\|^{4c+4}) \right]}.
\]

A similar approach for \( R_g \) using the global Lipschitz condition (1.4) completes Part (ii).

Part (iii) follows from the construction of \( \bar{K}_1 \) and \( \bar{K}_2 \) as follows. An application of Cauchy Schwarz and (1.5) in the statement of Lemma 1.1 gives that there exists \( K_2 < \infty \), independent of \( h_{n+1} \), such that
\[
\mathbb{E} \left[ \bar{K}_1 \right] = \mathbb{E} \left[ \frac{2}{3} \sqrt{M_n \sqrt{L_n}} \right] \leq \frac{2}{3} \sqrt{\mathbb{E}[M_n] \sqrt{\mathbb{E}[L_n]}} =: K_1.
\]

A similar argument using Jensen’s inequality shows that there exists \( K_2 < \infty \), independent of \( h_{n+1} \), such that \( \mathbb{E}[\bar{K}_2] \leq K_2 \).

Finally, for Part (iv), define the a.s. finite and \( \mathcal{G}_n \)-measurable random variable
\[
P(\|X(t_n)\|) := R_1 + 2R_2 \|X(t_n)\|^2 c_1^2 \left( 1 + 2 \|X(t_n)\|^{c+1} + \|X(t_n)\|^{2(c+1)} \right).
\]

Then, by Cauchy-Schwarz and (1.5) in the statement of Lemma 1.1, we have that there exists \( K_3 < \infty \), independent of \( h_{n+1} \), such that
\[
\mathbb{E} \left[ \bar{K}_1 P(\|X(t_n)\|) \right] = \mathbb{E} \left[ \sqrt{M_n \sqrt{L_n} P(\|X(t_n)\|)^2} \right] \leq \mathbb{E} \left[ \sqrt{M_n \sqrt{L_n}} \right] = K_3.
\]
where, by (4.1), we have

\[ \hat{L}_n = 4E \left[ \sum_{k=1}^{n} \Delta X_k^2 \right] \]

This completes the proof.

**Proof of Theorem 2.4.** By Theorem 1.4 it is sufficient to consider only the event that \( h_{\text{min}} < h_n < h_{\text{max}} \) for all \( n = 0, \ldots, N - 1 \). Define the error sequence \( \{E_n\}_{n \in \mathbb{N}} \) by

\[ E_{n+1} = Y_{n+1} - X(t_{n+1}) = Y_n - X(t_n) + \int_{t_n}^{t_{n+1}} \left[ f(Y_n) - f(X(t_n)) \right] ds + \int_{t_n}^{t_{n+1}} \left[ g(Y_n) - g(X(t_n)) \right] dW(s). \]

Expand \( f \) and \( g \) as Taylor series around \( X(t_n) \) over the interval of integration. As in Lemma 4.1 we get

\[ E_{n+1} = E_n + \int_{t_n}^{t_{n+1}} \left[ f(Y_n) - f(X(t_n)) \right] ds + \int_{t_n}^{t_{n+1}} \left[ g(Y_n) - g(X(t_n)) \right] dW(s) + \int_{t_n}^{t_{n+1}} R_f(s, t_n, X(t_n)) ds + \int_{t_n}^{t_{n+1}} R_g(s, t_n, X(t_n)) dW(s) \]

which may be rewritten using the notation \( \triangle W_{n+1} = W(t_{n+1}) - W(t_n) \) as

\[ E_{n+1} = E_n + h_{n+1} \left[ f(Y_n) - f(X(t_n)) \right] + \left[ [g(Y_n) - g(X(t_n))] \triangle W_{n+1} + \hat{R}_f(t_n, X(t_n)) + \hat{R}_g(t_n, X(t_n)) \right]. \]

Next we develop appropriate bounds on

\[ E \left[ \|E_{n+1}\|^2 \right] = E \left[ \left\{ E_{n+1}, E_{n+1} \right\} | \mathcal{F}_n \right]. \tag{4.3} \]

Note that

\[ \|E_{n+1}\|^2 = \langle E_n, E_{n+1} \rangle + h_{n+1} \left[ f(Y_n) - f(X(t_n)) \right]^2 + \langle \triangle [g(Y_n) - g(X(t_n))] \triangle W_{n+1}, E_{n+1} \rangle + \langle \hat{R}_f(t_n, X(t_n)) + \hat{R}_g(t_n, X(t_n)), E_{n+1} \rangle. \]

Then, since \( \langle E_n, E_{n+1} \rangle \leq \frac{1}{2} (\|E_n\|^2 + \|E_{n+1}\|^2) \), we have \( \|E_{n+1}\|^2 = \|E_n\|^2 + 2A_n + 2B_n + 2C_n \). Next we omit the arguments from \( \hat{R}_f, \hat{R}_g \) and write

\[ A_n = h_{n+1} \left[ f(Y_n) - f(X(t_n)) \right]^2 + h_{n+1} \left[ f(Y_n) - f(X(t_n)) \right]^2 + h_{n+1} \left[ g(Y_n) - g(X(t_n)) \right]^2 \]

\[ B_n = \langle [g(Y_n) - g(X(t_n))] \triangle W_{n+1}, E_{n+1} \rangle + \langle [g(Y_n) - g(X(t_n))] \triangle W_{n+1}, f(Y_n) - f(X(t_n)) \rangle \]

\[ C_n = \langle \hat{R}_f(t_n, X(t_n)) + \hat{R}_g(t_n, X(t_n)), E_{n+1} \rangle. \]
By Remark 2.1, and applying the Lipschitz bounds (1.3) and (1.4), we may now estimate (4.3) as

\[ \mathbb{E} \left[ \| E_{n+1} \|^2 \big| \mathcal{G}_n \right] \leq \| E_n \|^2 + h_{n+1} (2 \alpha + 2 \kappa^2) \| E_n \|^2 + 2 h_{n+1}^2 \| f(Y_n) - f(X(t_n)) \|^2 \\
+ 4 h_{n+1} \mathbb{E} \left[ (f(Y_n) - f(X(t_n)), \tilde{R}_f + \tilde{R}_g) \big| \mathcal{G}_n \right] + 4 \mathbb{E} \left[ (|g(Y_n) - g(X(t_n))|, \triangle W_{n+1}, \tilde{R}_f + \tilde{R}_g) \big| \mathcal{G}_n \right] \\
+ 2 \mathbb{E} \left[ (\tilde{R}_f + \tilde{R}_g, E_n) \big| \mathcal{G}_n \right] + 2 \mathbb{E} \left[ \| \tilde{R}_f + \tilde{R}_g \|^2 \big| \mathcal{G}_n \right], \quad \text{a.s.} \]

\[ \tilde{A}_n = \mathbb{E} \left[ \| E_{n+1} \|^2 \big| \mathcal{G}_n \right] - \mathbb{E} \left[ E_n \right]^2 \leq h_{n+1} (2 \alpha + 2 \kappa^2 + 8 R_2) \| E_n \|^2 \\
+ 4 h_{n+1}^2 \left[ \tilde{R}_1 + 2 R_2 \| X(t_n) \|^2 \right] + c_1^2 \left( 1 + 2 \| X(t_n) \|^c + \| X(t_n) \|^2 \right) \\
+ \tilde{A}_n + \tilde{B}_n + \tilde{C}_n + \tilde{D}_n, \quad \text{a.s.} \]

Now we write

\[ \mathbb{E} \left[ \| E_{n+1} \|^2 \big| \mathcal{G}_n \right] - \mathbb{E} \left[ E_n \right]^2 \leq h_{n+1} (2 \alpha + 2 \kappa^2 + 8 R_2) \| E_n \|^2 \\
+ 4 h_{n+1}^2 \left[ \tilde{R}_1 + 2 R_2 \| X(t_n) \|^2 \right] + c_1^2 \left( 1 + 2 \| X(t_n) \|^c + \| X(t_n) \|^2 \right) \\
+ \tilde{A}_n + \tilde{B}_n + \tilde{C}_n + \tilde{D}_n, \quad \text{a.s.} \]

Next we must consider the terms \( \tilde{A}_n, \tilde{B}_n, \tilde{C}_n, \) and \( \tilde{D}_n. \) After an application of the triangle inequality, it immediately follows from Part (ii) of Lemma 4.1 that

\[ \tilde{D}_n = 2 \mathbb{E} \left[ \| \tilde{R}_f + \tilde{R}_g \|^2 \big| \mathcal{G}_n \right] \leq 8 \tilde{K}_2 h_{n+1}^2, \quad \text{a.s.} \]

This estimate, along with the conditional second moment of \( \triangle W_{n+1} \) provided in Remark 2.1, and additionally applying two variants of the Cauchy-Schwarz inequality, first to the inner product and second to the conditional expectation, gives us

\[ \tilde{B}_n = 4 \mathbb{E} \left[ (|g(Y_n) - g(X(t_n))|, \triangle W_{n+1}, \tilde{R}_f + \tilde{R}_g) \big| \mathcal{G}_n \right] \\
\leq \kappa \| E_n \| \mathbb{E} \left[ \| \triangle W_{n+1} \| \| \tilde{R}_f + \tilde{R}_g \| \big| \mathcal{G}_n \right] \\
\leq \kappa \| E_n \| \sqrt{ \mathbb{E} \left[ \| \triangle W_{n+1} \|^2 \big| \mathcal{G}_n \right] } \sqrt{ \mathbb{E} \left[ \| \tilde{R}_f + \tilde{R}_g \|^2 \big| \mathcal{G}_n \right] } \\
\leq 2 \kappa \sqrt{ \tilde{K}_2 } \| E_n \| h_{n+1}^{3/2} \\
\leq \frac{1}{2} \| E_n \|^2 h_{n+1} + \kappa^2 \tilde{K}_2 h_{n+1}^2, \quad \text{a.s.} \]
Part (i) of Lemma 4.1 yields
\[ \tilde{C}_n = 2 \mathbb{E} \left[ \langle \tilde{R}_f + \tilde{R}_g, E_n \rangle \middle| \mathcal{G}_n \right] \leq 2 \mathbb{E} \left[ \| \tilde{R}_f \| \| E_n \| \middle| \mathcal{G}_n \right] + 2 \mathbb{E} \left[ \| \tilde{R}_g \| \| E_n \| \middle| \mathcal{G}_n \right] \]
\[ \leq 4 \bar{K}_1 \| E_n \| h_n^{3/2} \leq 2 \bar{K}_1 h_n + 2 \bar{K}_1 h_n^2, \quad \text{a.s.} \]

Finally,
\[ \bar{A}_n = 4 h_{n+1} \mathbb{E} \left[ (f(Y_n) - f(X(t_n)), \tilde{R}_f) \middle| \mathcal{G}_n \right] \]
\[ = 4 h_{n+1} \mathbb{E} \left[ (f(Y_n) - f(X(t_n)), \tilde{R}_f) \middle| \mathcal{G}_n \right] + 4 h_{n+1} \mathbb{E} \left[ (f(Y_n) - f(X(t_n)), \tilde{R}_g) \middle| \mathcal{G}_n \right]. \]

Moreover we have
\[ \mathbb{E} \left[ (f(Y_n) - f(X(t_n)), \tilde{R}_f) \middle| \mathcal{G}_n \right] \leq \mathbb{E} \left[ \| f(Y_n) - f(X(t_n)) \| \| \tilde{R}_f \| \middle| \mathcal{G}_n \right] \]
\[ = \| f(Y_n) - f(X(t_n)) \| \mathbb{E} \left[ \| \tilde{R}_f \| \middle| \mathcal{G}_n \right] \quad \text{a.s.} \]

A similar bound holds for \( \mathbb{E} \left[ (f(Y_n) - f(X(t_n)), \tilde{R}_g) \middle| \mathcal{G}_n \right] \), and therefore, by Part (i) of Lemma 4.1,
\[ \bar{A}_n \leq 8 \| f(Y_n) - f(X(t_n)) \| \bar{K}_1 \bar{h}_{n+1}^{5/2} \leq 4 \bar{K}_1 \bar{h}_{n+1}^2 + 4 \bar{K}_1 \| f(Y_n) - f(X(t_n)) \|^2 \bar{h}_{n+1}^3, \quad \text{a.s.} \]

Applying these bounds to (4.5), along with (4.4) and noting that \( h_{\text{max}} \leq 1 \), yields
\[ \mathbb{E} \left[ \| E_{n+1} \|^2 \middle| \mathcal{G}_n \right] - \| E_n \|^2 \leq h_{n+1}^2 \bar{I}_2 \| E_n \|^2 \]
\[ + \left[ 6 \bar{K}_1 + (8 + \kappa^2) \bar{K}_2 + 4(1 + \bar{K}_1) \left( \bar{R}_1 + 2 \bar{R}_2 \| X(t_n) \|^2 \right) \right. \]
\[ + \bar{c}_1^2 \left( 1 + 2 \| X(t_n) \| c + \| X(t_n) \|^{2(c+1)} \right)] \bar{h}_{n+1}^2, \quad \text{a.s.} \quad (4.6) \]

where, recalling \( K_1 \) as defined in Part (iii) of the statement of Lemma 4.1,
\[ \bar{I}_2 := 2(\alpha + \kappa^2) + 1/2 + 2 \bar{K}_1 + 24 \bar{R}_2. \]

Summing both sides of (4.6) over \( n \) from 0 to \( N - 1 \) and taking expectations yields
\[ \mathbb{E} [\| E_N \|^2] \leq \bar{I}_1 T h_{\text{max}} + \bar{I}_2 h_{\text{max}} \sum_{n=0}^{N-1} \mathbb{E} [\| E_n \|^2], \quad (4.7) \]

where, recalling \( K_2 \) and \( K_3 \) as defined in Parts (iii) and (iv) of the statement of Lemma 4.1, we have defined the constant \( \bar{I}_1 \) as
\[ \bar{I}_1 := 6 \bar{K}_1 + (8 + \kappa^2) \bar{K}_2 + 4 \left[ \bar{R}_1 + 2 \bar{R}_2 \mathbb{E} \left[ \sup_{u \in [0,T]} \| X(u) \|^2 \right] \right. \]
\[ + \bar{c}_1^2 \left( 1 + 2 \mathbb{E} \left[ \sup_{u \in [0,T]} \| X(u) \| c + 1 \right] + \mathbb{E} \left[ \sup_{u \in [0,T]} \| X(u) \|^{2(c+1)} \right] \right) \] + 4 \bar{K}_3.
The discrete Gronwall inequality (see for example \cite{ Reference1 }), (2.2), and the fact that \( Nh_{\min} \leq T \), may now be applied to (4.7), giving the statement of the theorem:

\[
E [ \| E_N \|^2 ] \leq \Gamma_1 T h_{\max} \exp \{ Nh_{\max} \Gamma_2 \} = \Gamma_1 T h_{\max} \exp \{ \rho Nh_{\min} \Gamma_2 \} \leq \Gamma_1 T h_{\max} \exp \{ \rho T \Gamma_2 \}.
\]

\[ \Box \]

**Remark 4.1** The numerical demonstration of strong convergence for the stochastic Van der Pol equation with additive noise (3.3) in Figure 3 leads us to conjecture that, when the diffusion coefficient \( g \) is constant, the rate of strong convergence is 1. However, we cannot prove this as a straightforward corollary of Theorem 2.4: the inequality (4.4), which appears as a result of the loss of a global Lipschitz bound on \( f \) and which requires the use of adaptive timestepping, contributes terms of insufficiently high order to the first term on the right hand side of (4.7) to give the conjectured rate of strong convergence. A different approach will be required.

5. **Conclusions and future work**

We introduced a class of adaptive timestepping strategies for SDEs with non-Lipschitz drift coefficients and proved strong convergence without the need to prove additional moment bounds on the numerical method, since such bounds are naturally enforced by the timestepping strategy.

Our numerical results using the stochastic Van der Pol equation indicate that adaptive timestepping may lead to dynamically more accurate solutions than those from a fixed step tamed scheme where the drift is perturbed, this was not only true for the two adaptive schemes we presented here but also for other admissible schemes. From the suite of problems that we examined, the method ALD seems to lead to a smaller variance in the timestep selection. We see from the numerical experiments that the parameter \( \rho \) required in the analysis does not impose a hard restriction in the simulations.

Adaptive timestepping strategies are readily applicable to large scale systems, such as the Allen-Cahn SPDE. We observe that when applied in a MLMC context adaptivity can lead to more efficient computation. It has already been noted by a number of authors \cite{ Reference2, Reference3 } that adaptivity maybe useful for Langevin sampling dynamics and our analysis offers techniques suitable for equations with non-Lipschitz drift terms that may arise, for example, in image processing. These could be combined with error control timestepping strategies.

Possible future work includes extending the analysis to include to SDEs with non-Lipschitz diffusion coefficients, to SDEs with Lévy noise, to SPDEs and to other forms of explicit methods.

6. **Acknowledgement**

The authors thank Prof. Alexandra Rodkina (UWI) for her feedback on an earlier draft, and the anonymous referees for their constructive comments.