UNIVERSITY OF BATH

On the convergence of adaptive approximations for SDEs

James Foster

University of Bath

Outline

Introduction

Consider the following Itô SDE on $[0, T]$:

$$dy_t = f(y_t)\,dt + \sum_{i=1}^{d} g_i(y_t)\,dW_t^i\,, \tag{1}$$

where $y_0 \in \mathbb{R}^e$, $W = (W^1, \cdots, W^d)$ is a d-dimensional Brownian motion and $f, g_i : \mathbb{R}^e \to \mathbb{R}^e$ are bounded and smooth with bounded derivatives.

Equivalently, we can write the SDE (1) in Stratonovich form:

$$dy_t = \widetilde{f}(y_t)\,dt + \sum_{i=1}^{d} g_i(y_t) \circ dW_t^i\,, \tag{2}$$

where $\widetilde{f}(y) := f(y) - \frac{1}{2}\sum_{i=1}^{d} g_i'(y)g_i(y)$.

In practice, it is often necessary to numerically approximate SDEs [1, 2].

Introduction

Consider the following Itô SDE on $[0, T]$:

$$dy_t = W_t \, dW_t. \tag{3}$$

Then, we know the solution is given by $y_t = \int_0^t W_s \, dW_s = \frac{1}{2}\big((W_t)^2 - t\big)$.

We can also approximate (3) using the Euler-Maruyama method:

$$\begin{aligned} Y_{k+1} &:= Y_k + W_{t_k}\big(W_{t_{k+1}} - W_{t_k}\big), \\ Y_0 &:= y_0, \end{aligned}$$

where $t_k := kh$ and $h = \frac{T}{K}$ for $k \in \{0, 1, \cdots, K\}$. It is then easy to show

$$\mathbb{E}\Big[\big(Y_K - y(T)\big)^2\Big] = \frac{1}{2} hT,$$

which converges to zero as $h \to 0$ (or, equivalently, as $K \to \infty$).

Introduction

What if we make the step size adaptive?
(which is popular in ODEs numerics)

For example, given a fixed λ, we can consider a condition of the form:

$$\left|W_{t_{k+1}} - W_{t_k}\right| \leq \lambda\sqrt{h}, \tag{4}$$

to help reduce errors when W has large fluctuations. In [3], they define

$$Y_{k+1} := \begin{cases} Y_k + W_{t_k}\big(W_{t_{k+1}} - W_{t_k}\big), & \text{if (4) holds,} \\ Y_k + W_{t_k}\big(W_{t_{k+\frac{1}{2}}} - W_{t_k}\big) + W_{t_{k+\frac{1}{2}}}\big(W_{t_{k+1}} - W_{t_{k+\frac{1}{2}}}\big), & \text{otherwise.} \end{cases}$$

Surprisingly however, it was shown in [3, Section 4.1] that this adaptive Euler method **fails to converge to the Itô solution!** (as $h \to 0$).

Introduction

Consider the following SDE:

$$dy_t = W_t^1 \, dW_t^2 \,, \tag{5}$$

where W^1 and W^2 denote two independent Brownian motions.

We can approximate (5) using Euler-Maruyama or a "trapezium" rule:

$$\begin{aligned} Y_{k+1} &:= Y_k + \frac{1}{2}\big(W_{t_k}^1 + W_{t_{k+1}}^1\big)\big(W_{t_{k+1}}^2 - W_{t_k}^2\big), \\ Y_0 &:= y_0 \,. \end{aligned}$$

where $k \in \{0, 1, \cdots, K\}$. By Itô's isometry, it is straightforward to show

$$\mathbb{E}\Big[\big(Y_K - y(T)\big)^2\Big] = \begin{cases} \frac{1}{2}hT & \text{if Euler-Maruyama is used} \\ \frac{1}{4}hT & \text{if the trapezium rule is used} \end{cases},$$

where $T = Kh$. Hence, we see that the trapezium rule is more accurate.

Introduction

However, consider the following (less natural) adaptive step size:

We choose either h (i.e. 1 step) or $\frac{1}{2}h$ (i.e. 2 half-steps) to maximise Y.

$$Y_{k+1} = \max\Big\{ Y_k + \frac{1}{2}\big(W^1_{t_k} + W^1_{t_{k+1}}\big)\big(W^2_{t_{k+1}} - W^2_{t_k}\big),$$
$$Y_k + \frac{1}{2}\big(W^1_{t_k} + W^1_{t_{k+\frac{1}{2}}}\big)W^2_{t_k, t_{k+\frac{1}{2}}} + \frac{1}{2}\big(W^1_{t_{k+\frac{1}{2}}} + W^1_{t_{k+1}}\big)W^2_{t_{k+\frac{1}{2}}, t_{k+1}}\Big\},$$

where $W^i_{s,t} := W^i_t - W^i_s$. Then, for any $h > 0$,

$$\mathbb{E}\big[Y_K\big] = \frac{1}{8}T,$$

whereas $\mathbb{E}\big[y_T\big] = 0$. So, once again, ***Y* does not converge to the SDE!**

Introduction

The proof follows from Brownian scaling and the following theorem:

Theorem (L^1 norm of the determinant of a 2×2 Gaussian matrix)

Let $A, B, C, D \sim \mathcal{N}(0,1)$ be independent random variables. Then

$$\mathbb{E}\big[|AD - BC|\big] = 1.$$

Proof

Can be established by a long calculation [4, Appendix C]. If you know of this result in the random matrix literature – please let me know! □

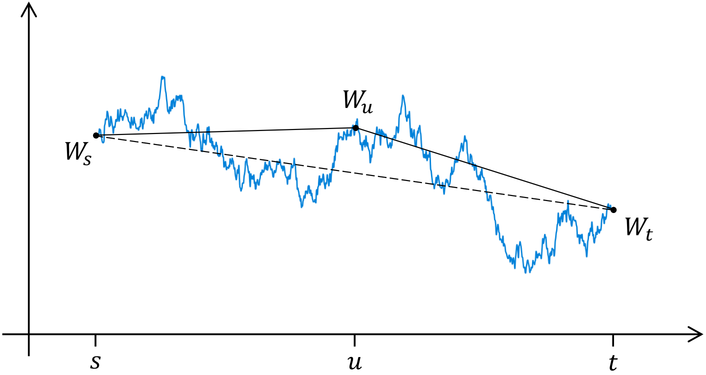
This leads to some natural questions...

Question

Do adaptive numerical methods for SDEs converge? If so, when?

Lévy's construction of Brownian motion

How can we generate Brownian motion after we halve the step sizes?

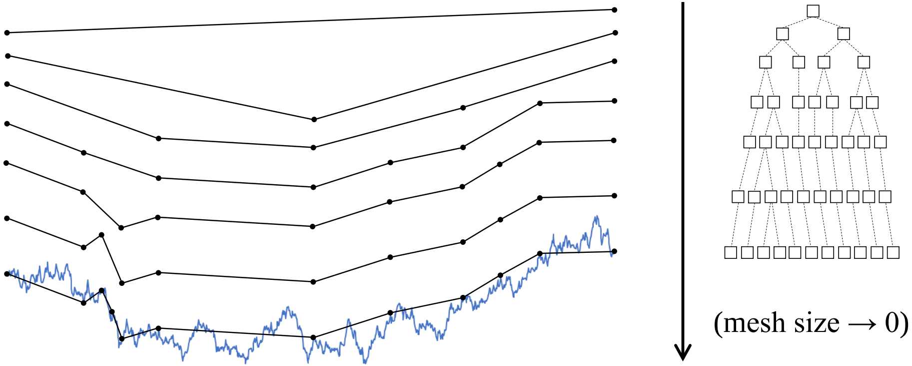


Using the notation $W_{a,b} := W_b - W_a$, we can generate W_u after W_t as

$$W_{s,t} \sim \mathcal{N}\big(0, (t-s)I_d\big), \qquad W_{s,u} \,|\, W_{s,t} \sim \mathcal{N}\bigg(\frac{1}{2}W_{s,t}, \frac{1}{4}(t-s)I_d\bigg).$$

The Brownian tree

By recursively applying Lévy's construction, we can construct a tree:



This is known as the Brownian tree (introduced in [3]) and also gives a natural data structure for storing sample paths of Brownian motion [5].

A “Brownian tree” condition

In our second counterexample, we could “ignore” information about the Brownian path – as the following update is decided using $W_{t_{k+\frac{1}{2}}}$:

$$Y_{k+1} = Y_k + \frac{1}{2}\big(W_{t_k}^1 + W_{t_{k+1}}^1\big)\big(W_{t_{k+1}}^2 - W_{t_k}^2\big)$$

but then does not use the value of $W_{t_{k+\frac{1}{2}}}$ in the approximation itself.

Hence, this goes against the natural direction of the Brownian tree (indicated by the downwards arrow).

First important condition

If information about the Brownian motion is generated, it <u>must be used</u> “correctly” (to be explained in condition 2). Equivalently, the numerical approximation uses all the information at the lowest level of the tree.

Outline

Stochastic Taylor expansions

Consider the Stratonovich SDE ($y_t \in \mathbb{R}^e$ and $f, g_i : \mathbb{R}^e \to \mathbb{R}^e$ are smooth)

$$dy_t = f(y_t)\,dt + \sum_{i=1}^{d} g_i(y_t) \circ dW_t^i\,, \tag{6}$$

A very useful tool in SDE numerical analysis is the Taylor expansion:

Theorem (Stratonovich-Taylor expansion [2, Thm 5.6.1])

For times $0 \leq s \leq t \leq T$, the solution of the SDE (6) can be expanded as

$$y_t = y_s + f(y_s)h + \sum_{i=1}^{d} g_i(y_s)W_{s,t}^i + \sum_{i,j=1}^{d} g_j'(y_s)g_i(y_s)\int_s^t W_{s,u}^i \circ dW_u^j + R,$$

where $h := t - s$ and there exists $C > 0$ such that $\mathbb{E}\big[\|R\|_2^2\big]^{\frac{1}{2}} \leq Ch^{\frac{3}{2}}$.

Non-Gaussian integrals involving Brownian motion

The stochastic integrals $\{\int_s^t W_{s,u}^i \circ dW_u^j\}_{1\leq i,j\leq d}$ are non-Gaussian and an algorithm for exact simulation has only been found when $d = 2$ [6].

Recently, when $d = 3, 4$, a neural network called "LévyGAN" has been trained to simulate these integrals [7] (conditional on the increment).

Whilst it empirically outperforms traditional approximations, such as Fourier series, this neural network does not have a Lévy construction.

Thus, we use the following approximation for these Brownian integrals:

$$\mathbb{E}\left[\int_s^t W_{s,u}^i \circ dW_u^j \,\middle|\, W_{s,t}\right] = \frac{1}{2} W_{s,t}^i W_{s,t}^j \,. \tag{7}$$

Among the $W_{s,t}$-measurable estimators, this minimises the $L^2(\mathbb{P})$ error.

An "integral" condition

Second important condition

The numerical method for the Stratonovich SDE (6) must satisfy

$$Y_{k+1} = Y_k + f(Y_k)h + \sum_{i=1}^{d} g_i(Y_k)W_k^i + \frac{1}{2}\sum_{i,j=1}^{d} g_j'(Y_k)g_i(Y_k)W_k^i W_k^j + R,$$

where $h := t_{k+1} - t_k$, $W_k := W_{t_{k+1}} - W_{t_k}$ and $R \sim o(h)$ almost surely.

More generally, if the numerical approximation uses certain Gaussian integrals $\mathcal{W}_k$ generated over the interval $[t_k, t_{k+1}]$, then we require:

$$Y_{k+1} = Y_k + f(Y_k)h + \sum_{i=1}^{d} g_i(Y_k)W_k^i$$
$$+ \sum_{i,j=1}^{d} g_j'(Y_k)g_i(Y_k)\mathbb{E}\left[\int_{t_k}^{t_{k+1}} W_{t_k,t}^i \circ dW_t^j \,\middle|\, \mathcal{W}_k\right] + o(h).$$

Examples of methods satisfying the integral condition

Milstein's method*

*using $q+1$ integrals of Brownian motion (which are Gaussian [8, 9])

$$Y_{k+1} := Y_k + f(Y_k)h + \sum_{i=1}^{d} g_i(Y_k)W_k^i$$
$$+ \sum_{i,j=1}^{d} g_j'(Y_k)g_i(Y_k)\,\mathbb{E}\left[\int_{t_k}^{t_{k+1}} W_{t_k,t}^i \circ dW_t^j \;\Big|\; \left\{\int_{t_k}^{t_{k+1}} \left(\tfrac{t-t_k}{h}\right)^m dW_t\right\}_{0\leq m\leq q}\right].$$

Heun's method (expanding will give $\frac{1}{2}W_{s,t}^i W_{s,t}^j$ instead of $\int_s^t W_{s,u}^i \circ dW_u^j$)

$$\widetilde{Y}_{k+1} = Y_k + f(Y_k)h + \sum_{i=1}^{d} g_i(Y_k)W_k^i\,,$$
$$Y_{k+1} = Y_k + \frac{1}{2}\big(f(Y_k) + f(\widetilde{Y}_{k+1})\big)h + \frac{1}{2}\sum_{i=1}^{d}\big(g_i(Y_k) + g_i(\widetilde{Y}_{k+1})\big)W_k^i\,.$$

Stochastic Runge-Kutta method (based on the "$q=1$" approximation)

Outline

Main convergence theorem

Theorem (Convergence of adaptive methods [4, Theorem 2.19])

Let $\{Y^n\}$ be a sequence of numerical solutions to (6) computed at times $D_n = \{0 = t_0^n < t_1^n < \cdots < t_{K_n}^n = T\}$ so that D_{n+1} is determined by D_n and

$$\mathcal{W}_k^n := \left\{ \int_{t_k^n}^{t_{k+1}^n} \left(\frac{t - t_k^n}{h_k^n}\right)^m dW_t \right\}_{0 \leq m \leq q}.$$

Suppose $D_{n+1} \subseteq D_n$ and $\mathrm{mesh}(D_n) \to 0$ almost surely (condition 1) and

$$\left\| Y_{k+1}^n - \widetilde{Y}_{k+1}^n \right\|_2 \sim o(h_k^n),$$

where $h_k^n := t_{k+1}^n - t_k^n$ and

$$\widetilde{Y}_{k+1} := Y_k^n + f(Y_k^n)h_k^n + \sum_{i=1}^{d} g_i(Y_k^n) W^i_{t_k^n, t_{k+1}^n} \qquad \textit{(condition 2)}$$

$$+ \sum_{i,j=1}^{d} g_j'(Y_k^n) g_i(Y_k^n) \mathbb{E}\left[\int_{t_k^n}^{t_{k+1}^n} W^i_{t_k^n, t} \circ dW_t^j \,\middle|\, \mathcal{W}_k^n \right].$$

Main convergence theorem

Theorem (Convergence of adaptive methods [4], continued)

We assume $Y_0^n = y_0$ and $f, \{g_i\}$ are bounded twice differentiable vector fields with α-Hölder continuous second derivatives for some $\alpha \in (0,1)$.

More precisely, we assume that

$$\left\| Y_{k+1}^n - \widetilde{Y}_{k+1}^n \right\|_2 \leq w(t_k^n, t_{k+1}^n),$$

where

$$\sum_{k=0}^{K_n - 1} w(t_k^n, t_{k+1}^n) \to 0,$$

almost surely. Then the approximations $\{Y^n\}$ converge pathwise. That is

$$\sup_{0 \leq k \leq K_n} \left\| Y_k^n - y_{t_k^n} \right\|_2 \to 0,$$

as $n \to \infty$ almost surely.

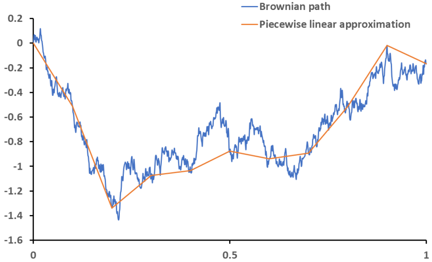
Ideas in the proof

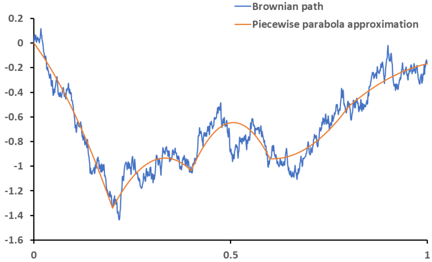
We first note that the expectation of Brownian motion conditional on $\{\mathcal{W}_k^n\}_{0\leq k\leq K_n-1}$ is the unique piecewise degree $q+1$ polynomial which, on each $[t_k^n, t_{k+1}^n]$, matches the increment and q integrals of W [10, 11].

$$\int_{t_k^n}^{t_{k+1}^n}\left(\frac{t-t_k^n}{h_k^n}\right)^m d\widetilde{W}_t^n = \int_{t_k^n}^{t_{k+1}^n}\left(\frac{t-t_k^n}{h_k^n}\right)^m dW_t\,,$$

for $0\leq m\leq q$, where

$$\widetilde{W}_t^n := \mathbb{E}\big[W_t\,\big|\,\{\mathcal{W}_k^n\}_{0\leq k\leq K_n-1}\big].$$

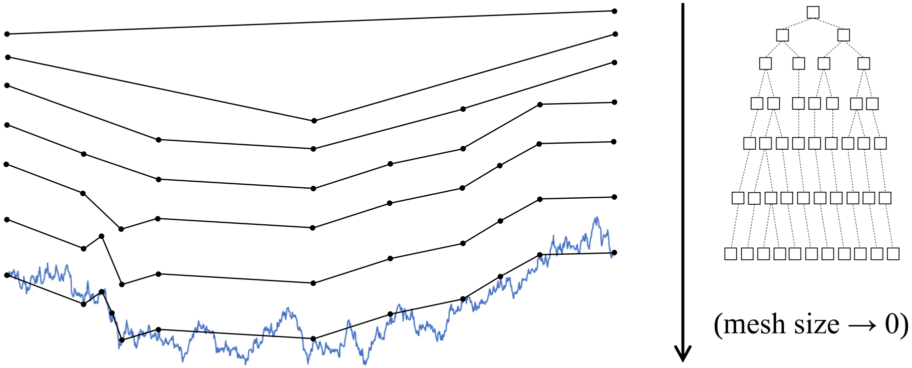




Ideas in the proof

Lemma

Define a sequence of σ-algebras $\{\mathcal{F}_n\}_{n\geq 0}$, *by* $\mathcal{F}_0 := \sigma\big(\{\mathcal{W}_0^n\} \cup D_0\big)$ *and* $\mathcal{F}_{n+1} := \sigma\big(\mathcal{F}_n \cup \{\mathcal{W}_k^n\} \cup D_n\big)$. *By the assumptions in the theorem,* $\{\mathcal{F}_n\}$ *is a filtration and* $\widetilde{W}^n = \mathbb{E}\big[W \,|\, \mathcal{F}_n\big]$ *is a square-integrable martingale.*



Ideas in the proof

Using Doob's martingale convergence theorem and maximal inequality, we can show that

$$d_{p\text{-}var;[0,T]}\big(\widetilde{\boldsymbol{W}}^n, \boldsymbol{W}\big) \to 0,$$

as $n \to \infty$ almost surely, where $p \in (2,3)$ and

- $\widetilde{\boldsymbol{W}}^n$ is the piecewise polynomial "lifted" to a "p-rough path"
- $\boldsymbol{W}$ is "Stratonovich enhanced" Brownian motion (p-rough path)
- $d_{p\text{-}var;[0,T]}(\boldsymbol{X}, \boldsymbol{Y})$ is the p-variation between p-rough paths $\boldsymbol{X}$ and $\boldsymbol{Y}$

It is not clear how to prove "rough path" convergence for $\{\widetilde{W}^n\}$ without using the martingale property coming from the nested property of $\{D_n\}$.

Ideas in the proof

By the Universal Limit Theorem (a key result in rough path theory [12]), it immediately follows that

$$d_{p\text{-}var;[0,T]}\big(\widetilde{\boldsymbol{y}}^{\,n}, \boldsymbol{y}\big) \to 0,$$

as $n \to \infty$ almost surely, where

- $\widetilde{\boldsymbol{y}}^{\,n}$ is the solution of the rough differential equation (RDE):

  $$d\widetilde{\boldsymbol{y}}_t^{\,n} = f\big(\widetilde{\boldsymbol{y}}_t^{\,n}\big)dt + g\big(\widetilde{\boldsymbol{y}}_t^{\,n}\big)\, d\widetilde{\boldsymbol{W}}_t^{\,n},$$

  with $g(y) := \big(g_1(y), \cdots, g_d(y)\big)$ and initial condition $y_0 \in \mathbb{R}^e$.

- $\boldsymbol{y}$ is the solution of the rough differential equation (RDE):

  $$d\boldsymbol{y}_t = f\big(\boldsymbol{y}_t\big)dt + g\big(\boldsymbol{y}_t\big)\, d\boldsymbol{W}_t\,,$$

  with initial condition $y_0 \in \mathbb{R}^e$.

Ideas in the proof

Final piece of the puzzle...

$$d\widetilde{y}_t^n = f(\widetilde{y}_t^n)dt + g(\widetilde{y}_t^n)\, d\widetilde{W}_t^n, \tag{8}$$

Theorem (Controlled Taylor expansion)

For times $0 \leq s \leq t \leq T$, the solution of the CDE (8) can be expanded as

$$\begin{aligned}\widetilde{y}_t^n = \widetilde{y}_s^n + f(\widetilde{y}_s^n)h + \sum_{i=1}^{d} g_i(\widetilde{y}_s^n)\big(\widetilde{W}_{s,t}^n\big)^i \\ + \sum_{i,j=1}^{d} g_j'(\widetilde{y}_s^n)g_i(\widetilde{y}_s^n)\int_s^t \big(\widetilde{W}_{s,u}^n\big)^i \circ d\big(\widetilde{W}_u^n\big)^j + o(h),\end{aligned} \tag{9}$$

where $h := t - s$.

Ideas in the proof

Theorem (Polynomials optimally approximate W and its integrals)

$$W_{t_k^n, t_{k+1}^n} = \widetilde{W}^n_{t_k^n, t_{k+1}^n}, \quad \mathbb{E}\bigg[\int_{t_k^n}^{t_{k+1}^n} W^i_{t_k^n, t} \circ dW^j_t \,\bigg|\, \mathcal{W}_k^n\bigg] = \int_{t_k^n}^{t_{k+1}^n} \big(\widetilde{W}^n_{t_k^n, t}\big)^i \circ d\big(\widetilde{W}^n_t\big)^j.$$

Theorem (Polynomial-driven CDEs have the desired expansions)

To simply notation, we let $\widetilde{y}_k^n$ denote $\widetilde{y}_{t_k^n}^n$ and suppose $\widetilde{Y}_{k+1}^n$ satisfies

$$\widetilde{Y}_{k+1}^n = \widetilde{y}_k^n + f(\widetilde{y}_k^n)h_k^n + \sum_{i=1}^d g_i(\widetilde{y}_k^n) W^i_{t_k^n, t_{k+1}^n} + \sum_{i,j=1}^d g_j'(\widetilde{y}_k^n) g_i(\widetilde{y}_k^n)\, \mathbb{E}\bigg[\int_{t_k^n}^{t_{k+1}^n} W^i_{t_k^n, t} \circ dW^j_t \,\bigg|\, \mathcal{W}_k^n\bigg] + o(h_k^n). \tag{10}$$

Then

$$\big\|\widetilde{Y}_{k+1}^n - \widetilde{y}_{k+1}^n\big\|_2 = o(h_k^n). \tag{11}$$

Outline

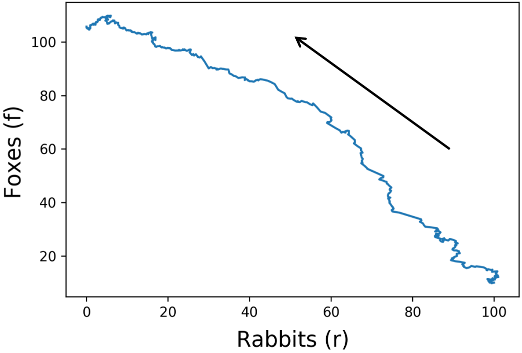
Numerical experiment

We consider a stochastic Lotka-Volterra predator-prey model [13]:

$$dr_t = r_t(2 - f_t)\,dt + \sqrt{r_t(2 + f_t)}\,dW_t^1,$$

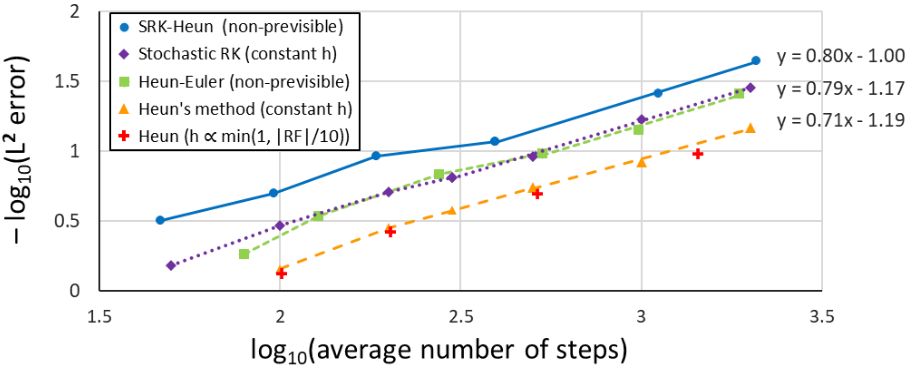
$$df_t = f_t(r_t - 1)\,dt + \sqrt{f_t(r_t + 1)}\,dW_t^2,$$

where $(r_0, f_0) = (100, 10)$. We estimate the $L^2(\mathbb{P})$ error at $T = 0.5$.



Numerical experiment

We compare a variety of methods and strategies for choosing step sizes (non-previsible means it depends on future values of Brownian motion)



However, methods with adaptive steps (whilst slightly more accurate) are much slower to run!

Outline

Conclusion and future work

Conclusion

- Numerical methods that use adaptive step sizes are popular for ODEs, but can experience convergence issues in the SDE setting.
- Using rough paths, we showed that convergence occurs for a large class of adaptive methods (including Milstein and Heun schemes).
- The main idea is that whenever information about W is generated, it must be used (condition 1) in a "correct way" (condition 2).

Future work

- Can we establish explicit convergence rates for adaptive methods?
- When is it acceptable for step sizes to "skip" information about W?
- Faster implementations of adaptive methods (e.g. using Diffrax [5])

Thank you
for your attention!

and the preprint can be found at:

J. Foster. *On the convergence of adaptive approximations for stochastic differential equations*, arxiv:2311.14201, 2023.

Outline

References I

- D. Higham and P. Kloeden. *An Introduction to the Numerical Simulation of Stochastic Differential Equations*. Society for Industrial and Applied Mathematics, 2021.
- P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, Springer, 1992.
- J. G. Gaines and T. J. Lyons. *Variable Step Size Control in the Numerical Solution of Stochastic Differential Equations*. SIAM Journal on Applied Mathematics, vol. 57, no. 5, 1997.
- J. Foster. *On the convergence of adaptive approximations for stochastic differential equations*. arxiv.org/abs/2311.14201, 2023.
- P. Kidger. *On Neural Differential Equations*. University of Oxford, 2022. Software available at github.com/patrick-kidger/diffrax.

References II

J. G. Gaines and T. J. Lyons. *Random Generation of Stochastic Area Integrals*. SIAM Journal on Applied Mathematics, vol. 54, no. 4, 1994.

A. Jelinčič, J. Tao, W. F. Turner, T. Cass, J. Foster and H. Ni. *Generative Modelling of Lévy Area for High Order SDE Simulation*. arxiv.org/abs/2308.02452, 2023.

J. Foster and K. Habermann. *Brownian bridge expansions for Lévy area approximations and particular values of the Riemann zeta function*. Combinatorics, Probability and Computing, vol. 32, no. 3, pp. 370-397, 2023.

J. Foster. *Numerical approximations for stochastic differential equations*. University of Oxford, 2020.

References III

- J. Foster, T. Lyons and H. Oberhauser. *An optimal polynomial approximation of Brownian motion*. SIAM Journal on Numerical Analysis, vol. 58, no. 3, pp. 1393–1421, 2020.

- K. Habermann. *A semicircle law and decorrelation phenomena for iterated Kolmogorov loops*. Journal of the London Mathematical Society, vol. 103, no. 2, pp. 558–586, 2021.

- T. J. Lyons, M. Caruana, T. Lévy, and J. Picard. *Differential Equations Driven by Rough Paths*. Ecole d'Eté de Probabilités de Saint-Flour. vol. 34, 2004.

- F. Vadillo. *Comparing stochastic Lotka-Volterra predator-prey models*. Applied Mathematics and Computation, vol. 360, pp. 181–189, 2019.