High order splitting methods for stochastic differential equations satisfying a commutitivity condition

James Foster

University of Bath

joint with Gonçalo dos Reis and Calum Strange (Edinburgh)
and ongoing with Andraž Jelinčič (Bath)

IMA Leslie Fox Prize Meeting
University of Strathclyde, 23rd June 2025

Outline

Introduction

Ordinary Differential Equations (ODEs) have seen widespread use for modelling continuous-time systems. However, they are deterministic.

So for modelling random phenomena, noise terms can be incorporated into ODEs. This leads us to Stochastic Differential Equations (or SDEs).

In this talk, we will consider SDEs defined by Stratonovich integration[1],

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

where $f, g_i : \mathbb{R}^e \to \mathbb{R}^e$ are smooth and bounded vector fields on $\mathbb{R}^e$ and each $W^i = \{W_t^i\}_{t\geq 0}$ denotes an independent Brownian motion.

[1]The two most common types of stochastic integration are "Itô" and "Stratonovich". The good news is that we can convert between them by changing the drift function [1].

Introduction

For many SDEs in applications, such as finance [2], statistical physics and machine learning [3], we use numerical methods to simulate (1).

This is often done using Monte Carlo simulation. That is, we first sample our noise (which is Gaussian) and then discretise the SDE.

When the noise due to Brownian motion is included, the standard Euler's method for ODEs becomes the Euler-Maruyama method:

$$Y_{n+1} := Y_n + f(Y_n)h_n + \sum_{i:=1}^{d} g_i(Y_n)W_n^i\,,$$

where $Y_0 := y_0$ is the initial value, $h_n := t_{n+1} - t_n$ is the step size and

$$W_n^i := W_{t_{n+1}}^i - W_{t_n}^i \sim \mathcal{N}(0, h_n)$$

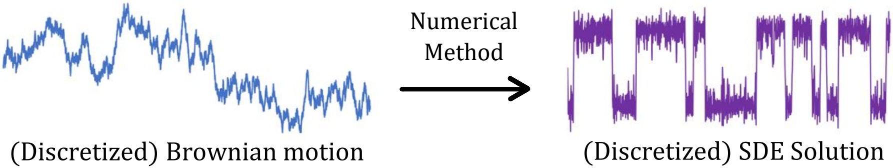
is an independent Gaussian random variable. We expect $Y_n \approx y(t_n)$.

Introduction

The numerical methods we propose in [4] come from classical ideas.

Idea 1:

SDEs and their numerical methods can be viewed as functions on paths.



Idea 2:

Noise terms are often "tractable" (e.g. affine noise $g_i(y) = A_i y + B_i$).

That is, without any drift vector field, we can solve the system exactly (or approximate it very well). This leads us to study splitting methods.

A motivating example: the CIR Model

The Cox-Ingersoll-Ross (CIR) model [2] is defined by the following SDE:

$$dy_t = a(b - y_t)\,dt + \sigma\sqrt{y_t}\,dW_t, \tag{2}$$

with the following parameters

- Mean reversion speed: $a > 0$
- Mean reversion level: $b > 0$
- Volatility: $\sigma > 0$

This diffusion is commonly used as a one-factor short rate model in mathematical finance for modelling interest rates and volatilities [5].

Note the ODEs, $\frac{dy}{dt} = a(b - y)$ and $\frac{dy}{dt} = \sigma\sqrt{y}$, can be solved analytically!

A motivating example: the CIR Model

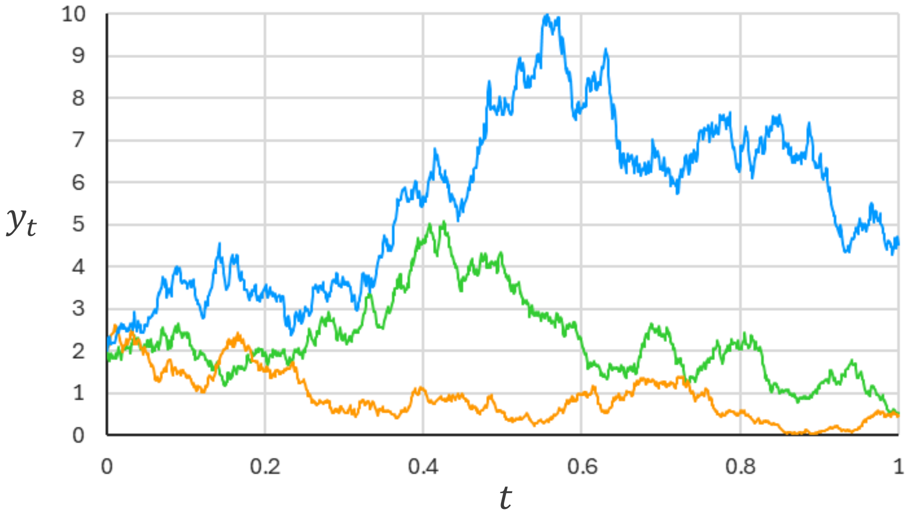


Figure: Sample paths of the CIR model with $a = b = 1$ and $\sigma = 2$.

A motivating example: the CIR Model

Lie-Trotter splitting

$$Y_n^{(1)} := \mathrm{Drift}(Y_n, h) = e^{-ah}Y_n + \widetilde{b}\big(1 - e^{-ah}\big),$$

$$Y_{n+1} := \mathrm{Noise}(Y_n^{(1)}, W_n) = \bigg(\sqrt{Y_n^{(1)}} + \frac{1}{2}\sigma W_n\bigg)^2,$$

where $h > 0$ denotes the step size and $W_n := W_{t_{n+1}} - W_{t_n} \sim \mathcal{N}(0, h)$ is the increment of the Brownian motion. This has $O(h)$ convergence.

Strang splitting

$$Y_n^{(1)} := \mathrm{Drift}\Big(Y_n, \frac{1}{2}h\Big),$$

$$Y_n^{(2)} := \mathrm{Noise}\big(Y_n^{(1)}, W_n\big),$$

$$Y_{n+1} := \mathrm{Drift}\Big(Y_n^{(2)}, \frac{1}{2}h\Big),$$

has $O(h)$ strong convergence, but $O(h^2)$ weak convergence (see [6]).

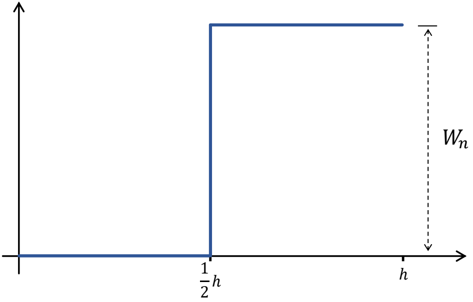
Strang splitting as a piecewise linear path

More generally, we can define a Strang splitting for Stratonovich SDEs as

$$Y_{n+1} := \exp\left(\frac{1}{2}f(\cdot)h\right)\exp\left(\sum_{i=1}^{d} g_i(\cdot)W_n^i\right)\exp\left(\frac{1}{2}f(\cdot)h\right)Y_n\,,$$

where $\exp(V)x$ is the solution $z(1)$ at $u=1$ of $z' = V(z)$ with $z(0)=x$.

Key idea: This is just solving (1) with the Brownian motion $\{(t, W_t)\}_{t\geq 0}$, replaced by the following piecewise linear path $X = \{(X^\tau, X^\omega)\}$ in $\mathbb{R}^{1+d}$,



Stochastic Taylor expansion

Informal Theorem (Taylor expansion for additive noise SDEs)

Let y be the unique solution to (1), with $g(y) = \sigma \in \mathbb{R}$. Then for $s < t$,

$$y_t = y_s + \int_s^t f(y_s)\,du + \int_s^t \big(f(y_u) - f(y_s)\big)\,du + \int_s^t \sigma \circ dW_u\,,$$

$$= y_s + f(y_s)(t-s) + \int_s^t \int_s^u f'(y_v) \circ \underbrace{dy_v}_{= f(y_v)\,dv + \sigma \circ dW_v} du + \sigma W_{s,t}\,,$$

$$\vdots$$

$$y_t = y_s + f(y_s)h + \sigma W_{s,t} + f'(y_s)\,\sigma \int_s^t W_{s,u}\,du + \frac{1}{2} f'(y_s)\,f(y_s)\,h^2$$

$$+ \frac{1}{2} f''(y_s)\,\sigma^2 \int_s^t (W_{s,u})^2\,du + R_{s,t}\,,$$

where $h = t - s$ and $W_{s,u} := W_u - W_s$.

In addition, the L2 norm of the remainder $R_{s,t}$ is $\mathbb{E}\big[(R_{s,t})^2\big]^{1/2} = O\big(h^{\frac{5}{2}}\big)$.

Outline

A path-based approach to splitting methods

Informal Theorem (Stochastic Taylor expansion [7, Thm 5.6.1])

The solution of the SDE (1) can be expressed as

$$y_t \approx y_s + f(y_s)h + \sum_{i=1}^{d} g_i(y_s)W_{s,t}^i + \sum_{i,j=1}^{d} (\cdots) \int_s^t W_{s,u}^i \circ dW_u^j \qquad (3)$$

$$+ (\cdots) \int_s^t W_{s,u}\, du + (\cdots) \int_s^t (u-s)\, dW_u + (\cdots) h^2$$

$$+ (\cdots)(\text{“third” iterated integrals of } \{t, W_t\})$$

$$+ (\cdots)(\text{“fourth” iterated integrals of } W),$$

where $h = t - s$ and $W_{s,u} := W_u - W_s$.

Observation from rough path theory

The Taylor expansion (3) can be extended beyond Brownian motion.

A path-based approach to splitting methods

By replacing (t, W_t) with a path $X = (X^\tau, X^\omega) : [0,1] \to \mathbb{R}^{1+d}$, we obtain

$$dY_t = f(Y_t)\, dX_t^\tau + \sum_{i=1}^{d} g_i(Y_t)\, d\big(X_t^\omega\big)^i. \tag{4}$$

Informal Theorem (Rough Taylor expansion [4, Proposition 3.2])

The solution of the controlled differential equation (4) is expressible as

$$\begin{aligned} Y_1 \approx Y_0 + f(Y_0)X_1^\tau + \sum_{i=1}^{d} g_i(Y_0)\big(X_1^\omega\big)^i + \sum_{i,j=1}^{d} (\cdots) \int_0^1 \big(X_t^\omega\big)^i d\big(X_t^\omega\big)^j \\ + (\cdots)\int_0^1 X_t^\omega\, dX_t^\tau + (\cdots)\int_0^1 X_t^\tau\, dX_t^\omega + (\cdots)\big(X_1^\tau\big)^2 \\ + (\cdots)\big(\text{“third” iterated integrals of } \{X^\tau, X^\omega\}\big) \\ + (\cdots)\big(\text{“fourth” iterated integrals of } X^\omega\big). \end{aligned}$$

A path-based approach to splitting methods

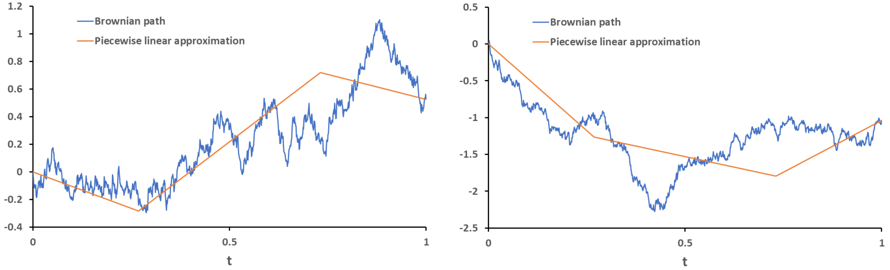
For Y to accurately approximate y, we will construct the path X so that

$$X_1 = \left(h, W_{s,t}\right),$$

$$\int_0^1 X_t^\omega \, dX_t^\tau = \int_s^t W_{s,u} \, du,$$

$$\mathbb{E}\left[\int_0^1 \left(X_t^\omega\right)^2 dX_t^\tau\right] = \mathbb{E}\left[\int_s^t (W_{s,u})^2 \, du\right].$$

Two examples of such piecewise linear paths X are illustrated below:



Outline

Establishing moment bounds for the approximation

Key assumption (Brownian-like scaling)

Let $X = (X^\tau, X^\omega)^\top : [0,1] \to \mathbb{R}^{1+d}$ be a piecewise linear path with $m \in \mathbb{N}$ components of a.s. finite length. Suppose each piece, $X_{r_i, r_{i+1}}$, satisfies

- X^τ is deterministic and scales with the step size h, i.e. $X^\tau_{r_i, r_{i+1}} = O(h)$
- Even moments of X^ω scale with h, i.e. $\mathbb{E}\big[\|X^\omega_{r_i, r_{i+1}}\|^{2k}\big] = O(h^k)$

Theorem (Moment bounds for the system (4) driven by X)

Suppose that $\mathbb{E}\big[\|Y_0\|^4\big] < \infty$ and the vector fields f, g have linear growth:

$$\|f(Y)\| \leq C(1 + \|Y\|), \qquad \|g(Y)\| \leq C(1 + \|Y\|),$$

with $\mathbb{E}\big[\exp\big(16C\int_0^1 |dX_u|\big)\big] < \infty$. Then there exists $\widetilde{C}$ so that for $r \in [0,1]$

$$\mathbb{E}\big[\|Y_r - Y_0\|^4\big] \leq \widetilde{C}h^2\big(1 + \mathbb{E}\big[\|Y_0\|^4\big]\big). \tag{5}$$

Main result

Theorem (Convergence of path-based splitting [4, Thm 3.9])

Consider the Stratonovich SDE on $[0, T]$,

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

where $f \in \mathcal{C}^2_{\mathrm{Lip}}(\mathbb{R}^e)$, $g_i \in \mathcal{C}^3_{\mathrm{Lip}}(\mathbb{R}^e)$ *have Lipschitz continuous derivatives and*

$$g_i'(y)g_j(y) = g_j'(y)g_i(y), \qquad \forall y \in \mathbb{R}^e. \tag{7}$$

Let φ *be a map on the space of continuous* $\mathbb{R}^{1+d}$*-valued paths such that* $X = \varphi\big(\{(u, W_u)\}_{u\in[s,t]}\big)$ *is a piecewise linear path on* $[0,1]$ *which satisfies*

$$X_{0,1} = (h, W_{s,t}), \qquad \int_0^1 X_{0,r}^{\omega}\, dX_r^{\tau} = \int_s^t W_{s,u}\, du,$$
$$\mathbb{E}\Big[\int_0^1 (X_{0,r}^{\omega})^{\otimes 2}\, dX_r^{\tau}\Big] = \frac{1}{2}h^2 I_d, \qquad \text{“Brownian-like scaling”}, \tag{8}$$

almost surely, where “Brownian-like scaling” refers to the previous slide.

Main result

Theorem (Convergence of path-based splitting [4, Thm 3.9])

We define a numerical solution Y by $Y_0 := y_0$ and for $n \in \{0, \cdots, N-1\}$,

$$Y_{n+1} := \exp\Big(f(\cdot)X^{\tau}_{1,r_m} + g(\cdot)X^{\omega}_{1,r_m}\Big) \cdots \exp\Big(f(\cdot)X^{\tau}_{0,r_1} + g(\cdot)X^{\omega}_{0,r_1}\Big) Y_n, \quad (9)$$

where each piecewise linear path X has m joints at $\{r_1 < \cdots < r_m\}$ and is defined by $\{(t, W_t)\}_{t \in [t_n, t_{n+1}]}$ using a step size of $h = \frac{T}{N}$ with $t_n := nh$.

In (9), $\exp(V)x$ is the solution at time $u = 1$ of

$$z' = V(z),$$
$$z(0) = x.$$

Then there exists constants $C_Y, h_{\max} > 0$, not depending on N, such that

$$\mathbb{E}\Big[\|Y_n - y_{t_n}\|_2^2\Big]^{\frac{1}{2}} \leq C_Y h^{\frac{3}{2}}, \quad (10)$$

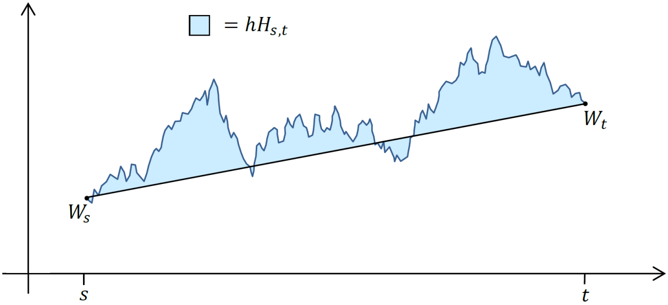
for $n \in \{1, \cdots, N\}$, provided that $h \leq h_{\max}$.

Generating Brownian increments and integrals

Definition (Space-time Lévy area of Brownian motion)

We define (rescaled) space-time Lévy area of W over an interval $[s,t]$ as

$$H_{s,t} := \frac{1}{h}\int_s^t W_{s,u}\,du - \frac{1}{2}W_{s,t}, \qquad (\text{where } h = t - s).$$

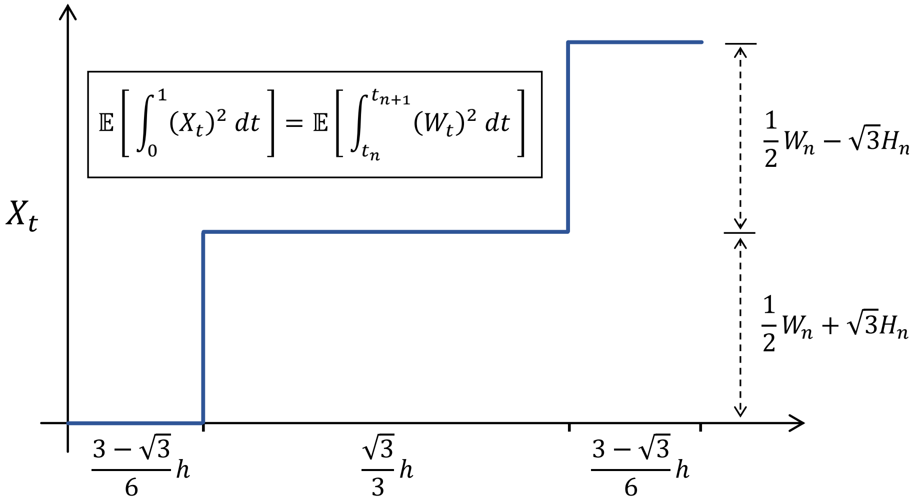


Theorem (Brownian increments and space-time Lévy areas [8])

The vectors $W_{s,t} \sim \mathcal{N}(0, hI_d)$ and $H_{s,t} \sim \mathcal{N}(0, \frac{1}{12}hI_d)$ are independent.

A higher order Strang splitting

We replace the Brownian motion with this piecewise linear path in $\mathbb{R}^2$:



Example: CIR Model

In Stratonovich form, the CIR model (2) becomes

$$dy_t = a(\widetilde{b} - y_t)\,dt + \sigma\sqrt{y_t} \circ dW_t, \tag{11}$$

where $\widetilde{b} := b - \frac{1}{4a}\sigma^2$. Thus, our splitting requires $\sigma^2 \leq 4ab$ and becomes

$$\begin{aligned}
Y_n^{(1)} &:= e^{-\frac{3-\sqrt{3}}{6}ah}Y_n + \widetilde{b}\big(1 - e^{-\frac{3-\sqrt{3}}{6}ah}\big),\\
Y_n^{(2)} &:= \left(\sqrt{Y_n^{(1)}} + \frac{\sigma}{2}\left(\frac{1}{2}W_n + \sqrt{3}H_n\right)\right)^2,\\
Y_n^{(3)} &:= e^{-\frac{\sqrt{3}}{3}ah}Y_n^{(2)} + \widetilde{b}\big(1 - e^{-\frac{\sqrt{3}}{3}ah}\big),\\
Y_n^{(4)} &:= \left(\sqrt{Y_n^{(3)}} + \frac{\sigma}{2}\left(\frac{1}{2}W_n - \sqrt{3}H_n\right)\right)^2,\\
Y_{n+1} &:= e^{-\frac{3-\sqrt{3}}{6}ah}Y_n^{(4)} + \widetilde{b}\big(1 - e^{-\frac{3-\sqrt{3}}{6}ah}\big).
\end{aligned} \tag{12}$$

Example: CIR Model (all parameters set to 1)

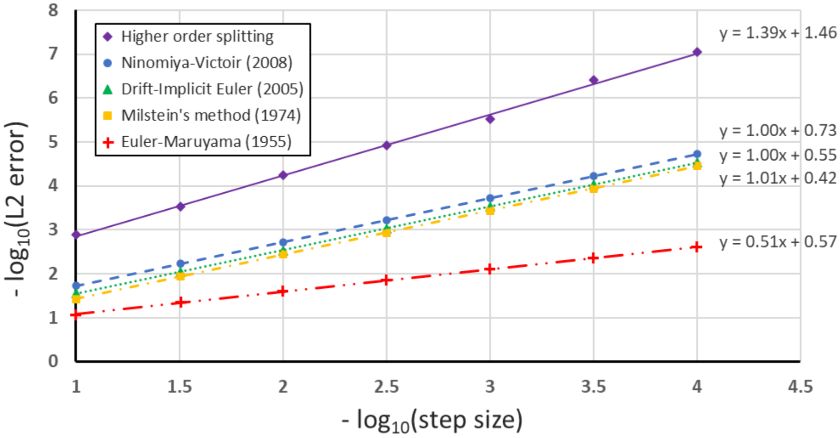


Table: Estimated time to produce 10^6 paths with a RMSE of 10^{-3} (seconds)

Splitting	Ninomiya-Victoir	Drift-Implicit Euler	Milstein	Euler
0.27	1.99	4.17	3.69	490

Example: Underdamped Langevin Dynamics

The underdamped Langevin diffusion (ULD) is a model for molecular dynamics and is given by the stochastic differential equation (SDE):

$$
\begin{aligned}
dx_t &= v_t\,dt, \\
dv_t &= -\gamma v_t\,dt - \nabla f(x_t)\,dt + \sqrt{2\gamma}\,dW_t,
\end{aligned} \tag{13}
$$

where

- $x, v \in \mathbb{R}^d$ will represent the *position* and *momentum* of a particle.
- $f : \mathbb{R}^d \to \mathbb{R}$ is a *scalar potential* that the particle moves around in.
- $\gamma > 0$ is the *friction* coefficient (we will use $\gamma = 1$).
- $W = (W^1, \cdots, W^d)$ is a d-dimensional Brownian motion,

$$W_t^i - W_s^i \sim \mathcal{N}(0, t - s).$$

Example: Underdamped Langevin Dynamics

"ULD = Newton's second law + frictional forces + stochastic forces"

$$dx_t = v_t\,dt,$$

$$dv_t = \underbrace{-\gamma v_t\,dt}_{\text{friction}} - \underbrace{\nabla f(x_t)\,dt}_{\substack{\text{gradient of the}\\ \text{potential / target}}} + \underbrace{\sqrt{2\gamma}\,dW_t}_{\text{noise}},$$

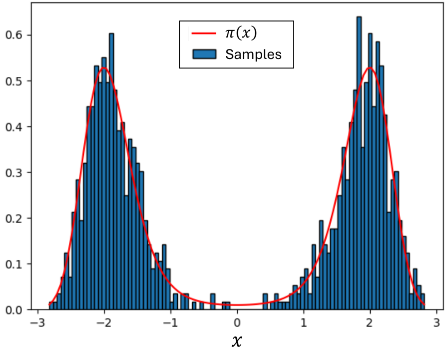
Under mild assumptions on f, the SDE admits a unique strong solution that is ergodic with stationary distribution $\pi(x,v) \propto e^{-f(x)}e^{-\frac{1}{2}\|v\|^2}$ [10].

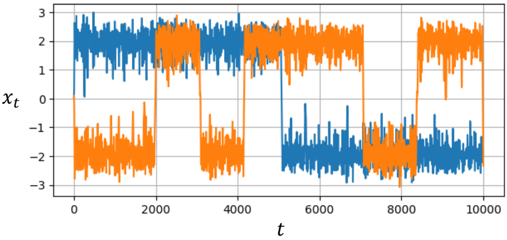
So, as well as physics [11], ULD can be applied to (high-dimensional) sampling problems in data science [12], as simulating ULD for long times produces approximate samples from the target distribution:

$$\pi(x) \propto e^{-f(x)}.$$

Thus, solving (13) gives a Markov Chain Monte Carlo (MCMC) algorithm.

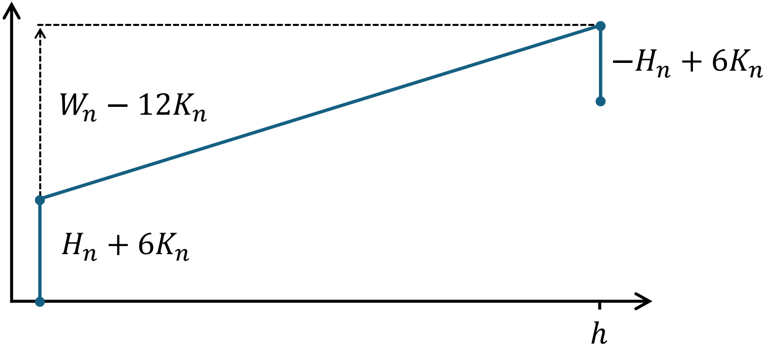
Underdamped Langevin Dynamics, $f(x) = \frac{1}{4}(x^2 - 4)^2$





Example: Underdamped Langevin Dynamics

For simulating this specific SDE, we use another piecewise linear path:



where $K_n \sim \mathcal{N}\big(0, \frac{1}{720}hI_d\big)$ is independent of (W_n, H_n) and defined as

$$K_n := \frac{1}{h^2}\int_{t_n}^{t_{n+1}}\left(W_{t_n,t} - \frac{t-t_n}{h}W_n\right)\left(\frac{1}{2}h - (t-t_n)\right)dt.$$

Example: Underdamped Langevin Dynamics

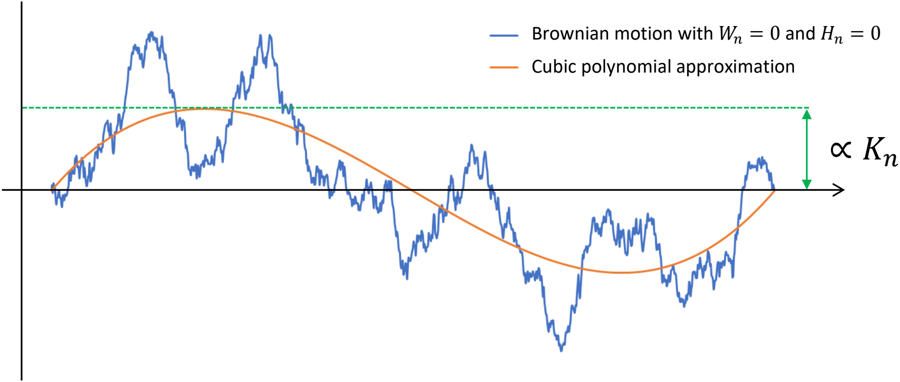


Figure: Space-time-time Lévy area gives the "skew" of the Brownian path [13].

Example: Underdamped Langevin Dynamics

Remarkably, due to ULD's structure, this has third order convergence!

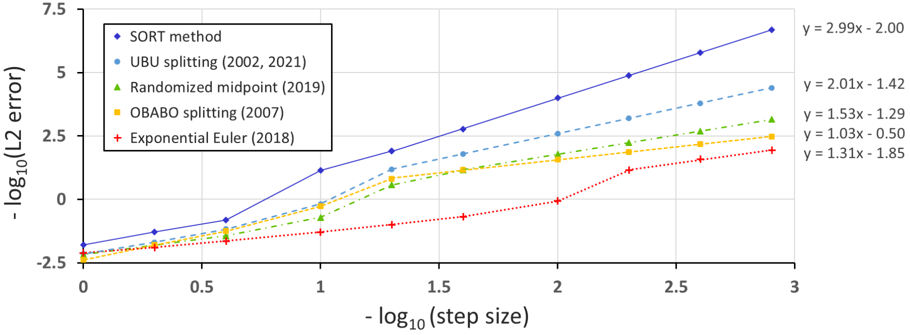


Figure: Estimated convergence rates for different ULD numerical methods [4]. Here, the L2 error (or root mean square error) is estimated at time $T = 1000$.

SORT[2] is a discretisation of the SDE driven by the new piecewise linear path.

[2] **S**hifted **O**DE with **R**unge-Kutta **T**hree

Example: Underdamped Langevin Dynamics

An $O(h^3)$ error bound (that is independent of T) was shown for the ODE splitting [13]. But we were unable to extend this to the SORT method.

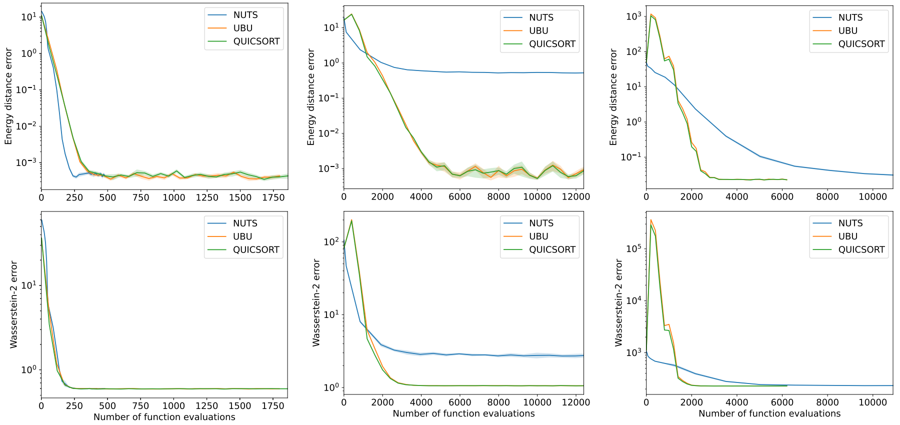
Recently, we made progress on QUICSORT[3] (to appear on arxiv soon):

$$
\begin{aligned}
V_n^{(1)} &:= V_n + \sqrt{2\gamma}\,(H_n + 6K_n),\\
X_n^{(1)} &:= X_n + \frac{1-e^{-a\gamma h}}{\gamma}V_n^{(1)} + \frac{e^{-a\gamma h} + a\gamma h - 1}{\gamma^2 h}\,C_n\,,\\
X_n^{(2)} &:= X_n + \frac{1-e^{-b\gamma h}}{\gamma}V_n^{(1)} - \frac{1-e^{-\frac{1}{3}\gamma h}}{\gamma}\nabla f\Big(X_n^{(1)}\Big)h + \frac{e^{-b\gamma h} + b\gamma h - 1}{\gamma^2 h}\,C_n\,,\\
V_n^{(2)} &:= e^{-\gamma h_n}V_n^{(1)} - \frac{1}{2}e^{-b\gamma h}\nabla f\Big(X_n^{(1)}\Big)h - \frac{1}{2}e^{-a\gamma h}\nabla f\Big(X_n^{(2)}\Big)h + \frac{1-e^{-\gamma h}}{\gamma h}\,C_n\,,\\
X_{n+1} &:= X_n + \frac{1-e^{-\gamma h}}{\gamma}V_n^{(1)} - \frac{1-e^{-b\gamma h}}{2\gamma}\nabla f\Big(X_n^{(1)}\Big)h - \frac{1-e^{-a\gamma h}}{2\gamma}\nabla f\Big(X_n^{(2)}\Big)h + \frac{e^{-\gamma h} + \gamma h - 1}{\gamma^2 h}\,C_n\,,\\
V_{n+1} &:= V_n^{(2)} - \sqrt{2\gamma}\,(H_n - 6K_n),
\end{aligned}
$$

where $a = \frac{3-\sqrt{3}}{6}$, $b = \frac{3+\sqrt{3}}{6}$ and $C_n = \sqrt{2\gamma}\,(W_n - 12K_n)$.

[3] **QU**adrature **I**nspired and **C**ontractive **S**hifted **O**DE with **R**unge-Kutta **T**hree

Example: Underdamped Langevin Dynamics



(a) Waveform ($d = 22$) (b) Splice ($d = 61$) (c) Isolet ($d = 617$)

Figure: Here, we performed Bayesian logistic regression across 14 datasets, with the ULD methods achieving faster convergence compared to the popular "No U-Turn Sampler" (NUTS) [18] in the two highest dimensional datasets.

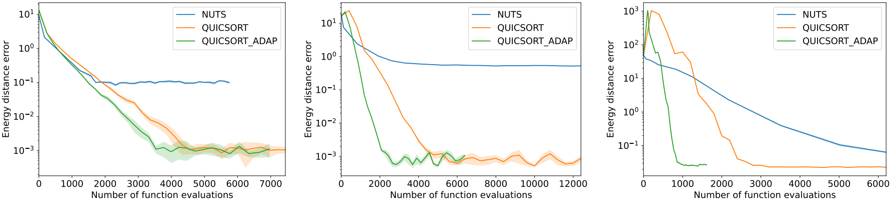
Outline

Conclusion and future work

Conclusion

- Path-based framework for developing high order splitting methods
- Effective for the CIR model and underdamped Langevin dynamics

Future work



(a) Flare Solar ($d = 10$) (b) Splice ($d = 61$) (c) Isolet ($d = 617$)

Figure: The QUICSORT method was recently implemented in Diffrax [19, 20], which is a high-performance Python package for simulating ODEs and SDEs. As a consequence, we can now perform QUICSORT with adaptive step sizes.

Thank you
for your attention!

and, if you are interested, our paper can be found at

J. Foster, G. dos Reis and C. Strange, *High order splitting methods for SDEs satisfying a commutativity condition*, SIAM Journal on Numerical Analysis, 2024. (Preprint is available at arxiv.org/abs/2210.17543).

Python code for our ongoing research into Langevin MCMC is also available at

github.com/andyElking/ThirdOrderLMC
github.com/andyElking/Single-seed_BrownianMotion (adaptive step sizes)

Outline

References I

L. Smith. *Itô and Stratonovich; a guide for the perplexed*, oatml.cs.ox.ac.uk/blog/2022/03/22/ito-strat.html, 2022.

J. Cox, J. Ingersoll and S. Ross. *A Theory of the Term Structure of Interest Rates*, Econometrica, vol. 53, no. 2, pp. 385–407, 1985.

M. Welling and Y. W. Teh. *Bayesian Learning via Stochastic Gradient Langevin Dynamics*, Proceedings of the 28th International Conference on Machine Learning, 2011.

J. Foster, G. dos Reis and C. Strange. *High Order Splitting Methods for SDEs Satisfying a Commutativity Condition*, SIAM Journal on Numerical Analysis, vol. 62, no. 1, pp. 500–532, 2024.

D. Brigo and F. Mercurio, *Interest Rate Models: Theory and Practice*, Springer, 2001.

References II

- S. Ninomiya and N. Victoir. *Weak Approximation of Stochastic Differential Equations and Application to Derivative Pricing*, Applied Mathematical Finance, vol. 15, no. 2, pp. 107–121, 2008.
- P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, Springer, 1992.
- 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.
- A. Alfonsi. *On the discretization schemes for the CIR (and Bessel squared) processes*, Monte Carlo Methods and Applications, 2005.
- G. A. Pavliotis. *Stochastic Processes and Applications*, Springer, New York, 2014.

References III

- W. T. Coffey, Y. P. Kalmykov and J. T. Waldron. *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*, World Scientific, 2012.
- T. Chen, E. B. Fox and C. Guestrin. *Stochastic Gradient Hamiltonian Monte Carlo*, Proceedings of the 31st International Conference on Machine Learning, 2014.
- J. Foster, T. Lyons and H. Oberhauser. *The shifted ODE method for underdamped Langevin MCMC*, arxiv:2101.03446, 2021.
- J. M. Sanz-Serna and K. C. Zygalakis. *Wasserstein distance estimates for the distributions of numerical approximations to ergodic stochastic differential equations*. Journal of Machine Learning Research, vol. 22, no. 242, pp. 1–37, 2021.

References IV

- R. Shen and Y. T. Lee. *The Randomized Midpoint Method for Log-Concave Sampling*. Advances in Neural Information Processing Systems, vol. 32, 2019.
- P. Monmarché. *High-dimensional MCMC with a standard splitting scheme for the underdamped Langevin diffusion*. Electronic Journal of Statistics, vol. 15, no. 2, pp. 4117–4166, 2021.
- X. Cheng, N. Chatterji, P. L. Bartlett and M. I. Jordan. *Underdamped Langevin MCMC: A non-asymptotic analysis*. Proceedings of the 31st Conference On Learning Theory, vol. 75, pp. 300–323, 2018.
- M. Hoffman and A. Gelman. *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*, Journal of Machine Learning Research, vol. 15, no. 47, pp. 1593–1623, 2014.

References V

- P. Kidger. *On Neural Differential Equations*, Doctoral Thesis, University of Oxford, arxiv:2202.02435, 2021.
  The Diffrax package is available at github.com/patrick-kidger/diffrax.
- P. Kidger and A. Jelinčič. *Underdamped Langevin Diffusion simulation*, docs.kidger.site/diffrax/examples/underdamped_langevin_example, 2024.
- E. Buckwar, A. Samsonand, M. Tamborrino and I. Tubikanec, *Splitting methods for SDEs with locally Lipschitz drift. An illustration on the FitzHugh-Nagumo model*, Applied Numerical Mathematics, vol. 179, 2022.
- M. Hutzenthaler, A. Jentzen and P. E. Kloeden, *Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients*, Annals of Applied Probability, 2012.

Additional information on the CIR example

Table: Computer time to simulate 100,000 paths with 100 steps (seconds)

Splitting	Ninomiya-Victoir	Drift-Implicit Euler	Milstein	Euler
2.13	1.07	1.42	1.01	0.86

Since $\frac{1}{2}W_n + \sqrt{3}H_n$ and $\frac{1}{2}W_n - \sqrt{3}H_n$ are independent, we can prove

Theorem (High order weak approximation of the CIR model)

The numerical solution given by (12) has the following moments:

$$\mathbb{E}[Y_{n+1}|Y_n] = e^{-ah}Y_n + b\big(1 - e^{-ah}\big) + O(h^5),$$
$$\mathrm{Var}(Y_{n+1}|Y_n) = \frac{\sigma^2}{a}\big(e^{-ah} - e^{-2ah}\big)Y_n + \frac{b\sigma^2}{2a}\big(1 - e^{-ah}\big)^2 + O(h^5).$$

Ignoring the $O(h^5)$ remainder terms, the above formulae are precisely the conditional mean and variance of the CIR model (started at Y_n).

Example: FitzHugh-Nagumo Model

The stochastic FitzHugh-Nagumo (FHN) model [21] is given by the SDE:

$$d\begin{pmatrix} V_t \\ U_t \end{pmatrix} = \begin{pmatrix} \frac{1}{\epsilon}\left(V_t - V_t^3 - U_t\right) \\ \gamma V_t - U_t + \beta \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} dW_t. \qquad (14)$$

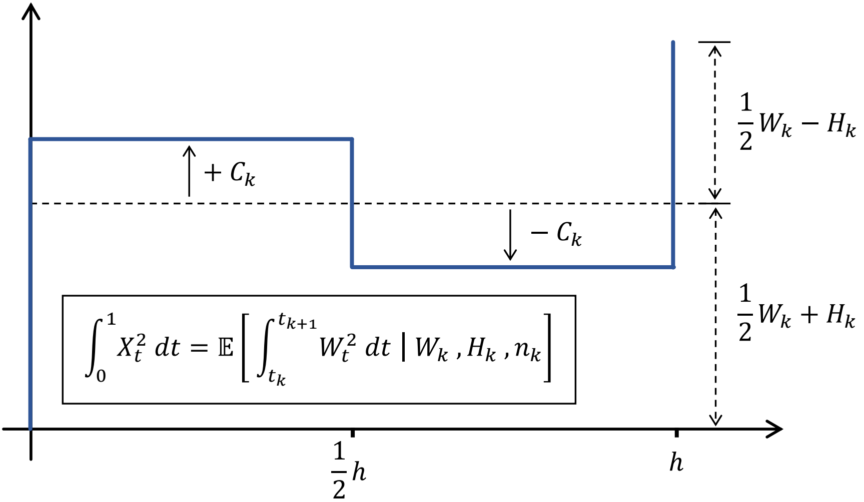
with the following parameters

- Time scale separation: $\epsilon > 0$
- Position parameter of an excitation: $\beta \geq 0$
- Duration parameter of an excitation: $\gamma > 0$
- Noise parameters: $\sigma_1, \sigma_2 \geq 0$

The FHN model is used to describe the firing activity of single neurons. The first component V describes the membrane voltage of the neuron, whilst the second component U can be viewed as a recovery variable.

Example: FitzHugh-Nagumo Model

We replace each Brownian motion by the following piecewise linear path:

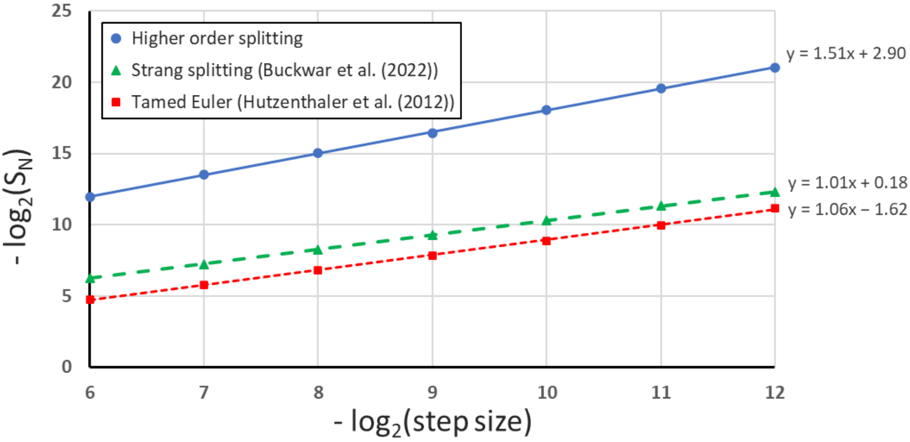


$$\int_0^1 X_t^2\,dt = \mathbb{E}\left[\int_{t_k}^{t_{k+1}} W_t^2\,dt \,\middle|\, W_k\,, H_k\,, n_k\right]$$

($n_k \in \{-1, 1\}$ is independent and gives the half-interval with largest H)

FitzHugh-Nagumo Model (parameters set to 1, $T = 5$)

The system cannot be exactly solved along the “horizontal” pieces, so we apply a further Strang splitting to approximate the resulting ODEs.



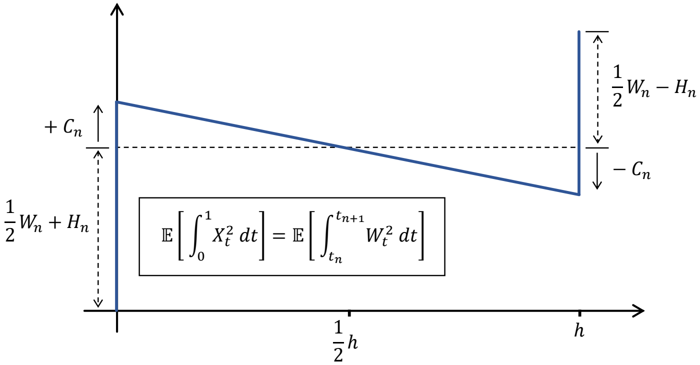
With 640 steps, we’re as accurate as Strang splitting with 10,240 steps!

Example: Additive-noise SDEs

In the paper, we also consider a stochastic FitzHugh-Nagumo model and additive-noise SDEs:

$$dy_t = f(y_t)\,dt + \sigma\,dW_t\,,$$

where $f : \mathbb{R}^e \to \mathbb{R}^e$ denotes a vector field on $\mathbb{R}^e$ and $\sigma \in \mathbb{R}^{e\times d}$ is a matrix. For a general additive-noise SDE, we propose a path with three pieces:



Example: Additive-noise SDEs

We need to discretize the “diagonal ODE” defined by the middle piece. Runge-Kutta methods may be a good choice of solver, but which one?

$$\int_0^1 (A + Bt)^{\otimes 2}\,dt = \frac{1}{4}A^{\otimes 2} + \frac{3}{4}\Big(A + \frac{2}{3}B\Big)^{\otimes 2} \Rightarrow \text{Ralston's method is ideal!}$$

(i.e. f is evaluated at 0 and $\frac{2}{3}$)

Thus, we propose

$$\begin{aligned}
Y_n^{(1)} &:= Y_n + \sigma\bigg(\frac{1}{2}W_n + H_n + C_n\bigg),\\
Y_{n+\frac{2}{3}}^{(1)} &:= Y_n^{(1)} + \frac{2}{3}\Big(f\big(Y_n^{(1)}\big)h - 2\sigma C_n\Big),\\
Y_{n+1}^{(2)} &:= Y_n^{(1)} + \frac{1}{4}f\big(Y_n^{(1)}\big)h + \frac{3}{4}f\big(Y_{n+\frac{2}{3}}^{(1)}\big)h - 2\sigma C_n\,,\\
Y_{n+1} &:= Y_{n+1}^{(2)} + \sigma\bigg(\frac{1}{2}W_n - H_n + C_n\bigg).
\end{aligned}$$

Example: Additive-noise SDEs

In our experiment, we compare against Rößler's strong 1.5 scheme [?]

$$\widetilde{Y}_{n+\frac{3}{4}} := Y_n + \frac{3}{4}f(Y_n)h + \frac{3}{4}\sigma W_n + \frac{3}{2}\sigma H_n\,,$$
$$Y_{n+1} := Y_n + \frac{1}{3}f(Y_n)h + \frac{2}{3}f\big(\widetilde{Y}_{n+\frac{3}{4}}\big)h + \sigma W_n\,.$$

Note that these methods have essentially the same computational cost (two Gaussian random variables, two vector field evaluations per step).

If we set $C_n = 0$ and solve the ODE with Euler instead of Ralston, we get

$$Y_{n+1} := Y_n + f\Big(Y_n + \frac{1}{2}\sigma W_n + \sigma H_n\Big)h + \sigma W_n\,.$$

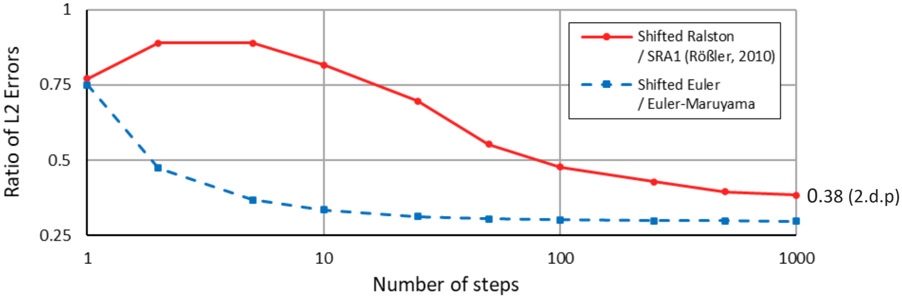
We expect this to be more accurate than the Euler-Maruyama method (though still first order convergent).

Example: Additive-noise SDEs

We test these methods on the following scalar anharmonic oscillator:

$$dy_t = \sin(y_t)\,dt + dW_t, \qquad (y_0 = 1, \quad T = 1).$$

All methods exhibit their expected strong and weak convergence rates, though the proposed schemes are more accurate (in line with theory).



$$\frac{\left\| \int_s^t W_{s,u}^2 du - \mathbb{E}\left[\int_s^t W_{s,u}^2 du \,\middle|\, W_{s,t}, H_{s,t}, n_{s,t}\right] \right\|_{L^2(\mathbb{P})}}{\left\| \int_s^t W_{s,u}^2 du - \frac{3}{2}\left(\frac{1}{2}hW_{s,t} + hH_{s,t}\right)^2 \right\|_{L^2(\mathbb{P})}} = \left(\frac{7}{30} - \frac{5}{16\pi}\right)^{\frac{1}{2}} \approx 0.37 \;\; (2.\text{d.p})$$