High order splitting methods for stochastic differential equations

James Foster

University of Bath

(joint with Gonçalo dos Reis and Calum Strange)

Outline

Introduction

Consider the following (Stratonovich) stochastic differential equation,

$$dy_t = f_0(y_t)\,dt + \sum_{i=1}^{d} f_i(y_t) \circ dW_t^i, \tag{1}$$

where each $f_i : \mathbb{R}^n \to \mathbb{R}^n$ is a smooth and bounded vector field on $\mathbb{R}^n$ and $W = \{W_t\}$ denotes a standard d-dimensional Brownian motion.

SDEs can model random time-evolving systems and have applications ranging from finance [1] to statistical physics and machine learning [2].

Just as for ODEs, numerical methods for solving SDEs are designed and analysed using their Taylor expansions.

Introduction

Informal Theorem (Stochastic Taylor expansion [3])

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

$$
\begin{aligned}
y_t \approx y_s + f_0(y_s)h + \sum_{i=1}^{d} f_i(y_s)W_{s,t}^i + \sum_{i,j=1}^{d}(\cdots)\int_s^t W_{s,u}^i \circ dW_u^j \\
+ (\cdots)\int_s^t W_{s,u}\,du + (\cdots)\int_s^t (u-s)\,dW_u + (\cdots)h^2 \\
+ (\cdots)(\text{“third” iterated integrals of } \{t, W_t\}) \\
+ (\cdots)(\text{“fourth” iterated integrals of } W),
\end{aligned}
\tag{2}
$$

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

Introduction

Informal Theorem (Stochastic Taylor expansion [3])

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

$$
\begin{aligned}
y_t \approx y_s + f_0(y_s)h + \sum_{i=1}^{d} f_i(y_s)W_{s,t}^i + \sum_{i,j=1}^{d} (\cdots) \int_s^t W_{s,u}^i \circ dW_u^j \qquad (2) \\
+ (\cdots) \int_s^t W_{s,u}\, du + (\cdots) \int_s^t (u-s)\, dW_u + (\cdots)h^2 \\
+ (\cdots)(\text{“third” iterated integrals of } \{t, W_t\}) \\
+ (\cdots)(\text{“fourth” iterated integrals of } W),
\end{aligned}
$$

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

Observation from rough path theory

The Taylor expansion (2) can be extended beyond Brownian motion.

A “rough path” approach to SDE numerics

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

$$dY_t = f_0(Y_t)\, dX_t^\tau + \sum_{i=1}^{d} f_i(Y_t) d\big(X_t^\omega\big)^i, \tag{3}$$

A "rough path" approach to SDE numerics

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

$$dY_t = f_0(Y_t)\, dX_t^\tau + \sum_{i=1}^{d} f_i(Y_t) d\big(X_t^\omega\big)^i, \tag{3}$$

Informal Theorem (Rough Taylor expansion [4])

The solution of the controlled differential equation (3) is expressible as

$$\begin{aligned} Y_1 \approx Y_0 + f_0(Y_0)X_1^\tau + \sum_{i=1}^{d} f_i(Y_0)\big(X_1^\omega\big)^i + \sum_{i,j=1}^{d} (\cdots) \int_0^1 \big(X_t^\omega\big)^i d\big(X_t^\omega\big)^j \\ + (\cdots)\int_0^1 X_t^\omega\, dX_t^\tau + (\cdots)\int_0^1 X_t^\tau\, dX_t^\omega + (\cdots)\big(X_1^\tau\big)^2 \\ + (\cdots)\big(\text{“third” iterated integrals of } \{X^\tau, X^\omega\}\big) \\ + (\cdots)\big(\text{“fourth” iterated integrals of } X^\omega\big). \end{aligned}$$

A “rough path” approach to SDE numerics

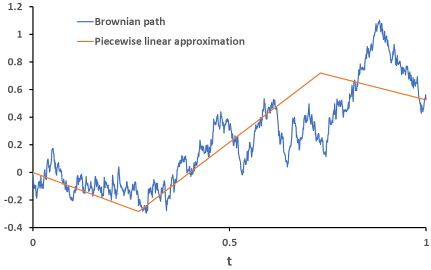
For Y to accurately approximate y, we will construct the path X so that

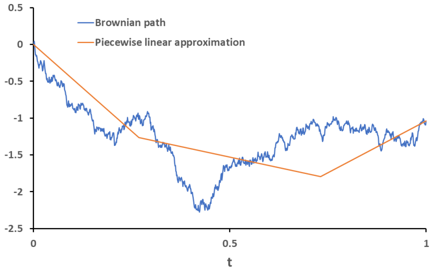
$$X_1 = \big(h, W_{s,t}\big)\,, \tag{4}$$

$$\int_0^1 X_t^{\omega}\, dX_t^{\tau} = \int_s^t W_{s,u}\, du\,, \tag{5}$$

$$\mathbb{E}\bigg[\int_0^1 \big(X_t^{\omega}\big)^{\otimes 2}\, dX_t^{\tau}\bigg] = \mathbb{E}\bigg[\int_s^t W_{s,u}^{\otimes 2}\, du\bigg] = \frac{1}{2}h^2 I_d\,. \tag{6}$$

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





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. $\mathbb{E}\big[\|X^\omega_{r_i, r_{i+1}}\|^{2k}\big] = O(h^k)$

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 on the CDE solution)

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* $r \in [0,1]$

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

Outline

Example: Strang splitting

The well-known Strang splitting for ODEs can be extended to SDEs [6] as

$$Y_{k+1} := \exp\left(\frac{1}{2}f_0(\cdot)h\right)\exp\left(\sum_{i=1}^{d} f_i(\cdot)W_k\right)\exp\left(\frac{1}{2}f_0(\cdot)h\right)Y_k,$$

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

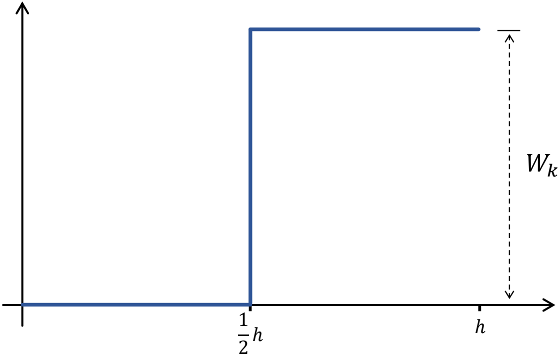
Example: Strang splitting

The well-known Strang splitting for ODEs can be extended to SDEs [6] as

$$Y_{k+1} := \exp\left(\frac{1}{2}f_0(\cdot)h\right)\exp\left(\sum_{i=1}^{d} f_i(\cdot)W_k\right)\exp\left(\frac{1}{2}f_0(\cdot)h\right)Y_k,$$

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

However, this is simply driving the SDE with the piecewise linear path:

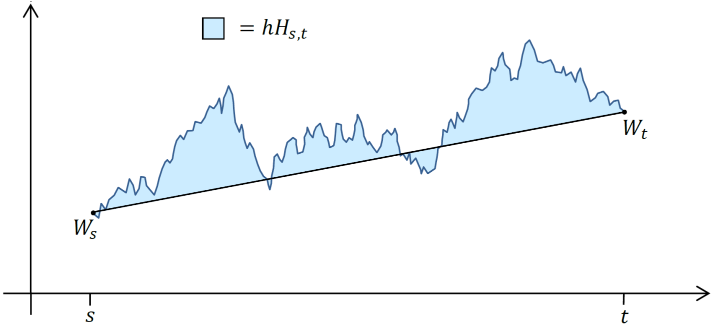


Why does Strang have low order strong convergence?

Definition (Space-time Lévy area of Brownian motion)

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

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



Theorem (Distribution of increments and space-time Lévy areas)

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.

Example: CIR Model

The Cox-Ingersoll-Ross (CIR) model [1] is defined by the following SDE:

$$dy_t = a(b - y_t)\,dt + \sigma\sqrt{y_t}\,dW_t, \tag{7}$$

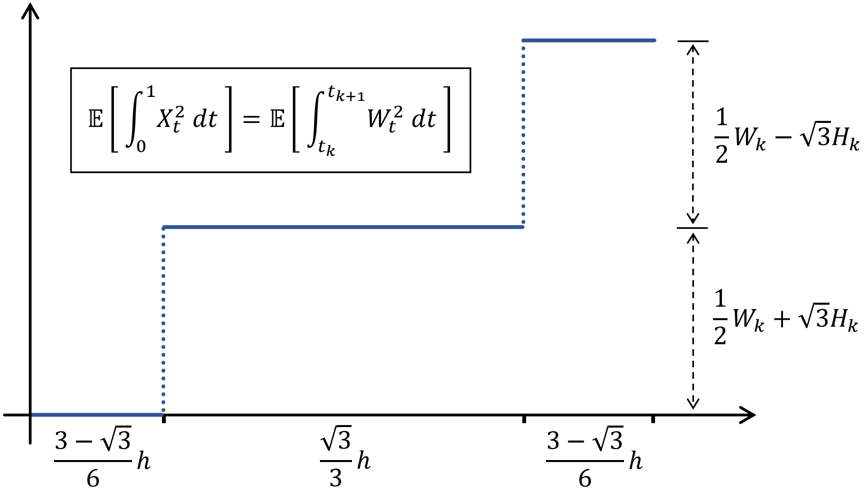
with the following parameters

- 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].

Example: CIR Model

We replace the Brownian motion with the following piecewise linear path:



Example: CIR Model

In Stratonovich form, the CIR model becomes

$$dy_t = a(\widetilde{b} - y_t)\,dt + \sigma\sqrt{y_t} \circ dW_t, \tag{8}$$

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

$$\begin{aligned}
Y_k^{(1)} &:= e^{-\frac{3-\sqrt{3}}{6}ah}Y_k + \widetilde{b}\big(1 - e^{-\frac{3-\sqrt{3}}{6}ah}\big),\\
Y_k^{(2)} &:= \left(\sqrt{Y_k^{(1)}} + \frac{\sigma}{2}\left(\frac{1}{2}W_k + \sqrt{3}H_k\right)\right)^2,\\
Y_k^{(3)} &:= e^{-\frac{\sqrt{3}}{3}ah}Y_k^{(2)} + \widetilde{b}\big(1 - e^{-\frac{\sqrt{3}}{3}ah}\big),\\
Y_k^{(4)} &:= \left(\sqrt{Y_k^{(3)}} + \frac{\sigma}{2}\left(\frac{1}{2}W_k - \sqrt{3}H_k\right)\right)^2,\\
Y_{k+1} &:= e^{-\frac{3-\sqrt{3}}{6}ah}Y_k^{(4)} + \widetilde{b}\big(1 - e^{-\frac{3-\sqrt{3}}{6}ah}\big).
\end{aligned} \tag{9}$$

Example: CIR Model (all parameters set to 1)

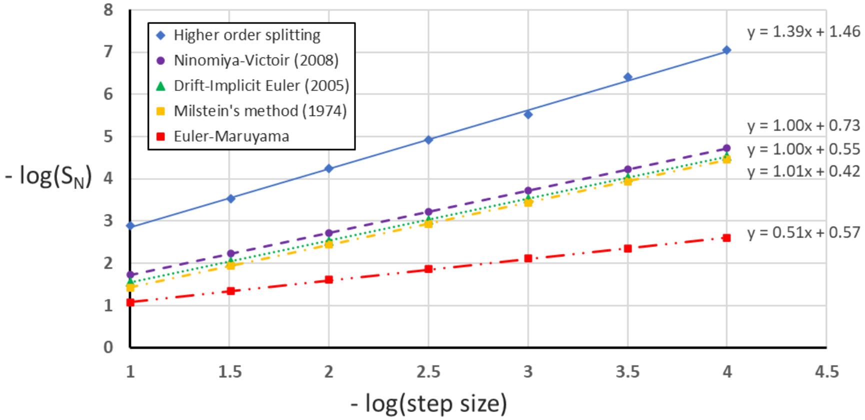


Table: Computer time to simulate 100,000 paths with 100 steps (seconds)

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

Example: CIR Model (all parameters set to 1)

Hence, the proposed splitting method is significantly more accurate!

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: CIR Model (all parameters set to 1)

Hence, the proposed splitting method is significantly more accurate!

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

Moreover, as $\frac{1}{2}W_k + \sqrt{3}H_k$ and $\frac{1}{2}W_k - \sqrt{3}H_k$ are independent, we have

Theorem

The numerical solution given by (9) has the following moments:

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

Example: FitzHugh-Nagumo Model

The stochastic FitzHugh-Nagumo (FHN) model [8] is given by the SDE:

$$d\begin{pmatrix} V_t \\ U_t \end{pmatrix} = \begin{pmatrix} \frac{1}{\epsilon}\left(V_t - V_t^3 - U_t\right) \\ \gamma V_t - U_t + \beta \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} dW_t, \tag{10}$$

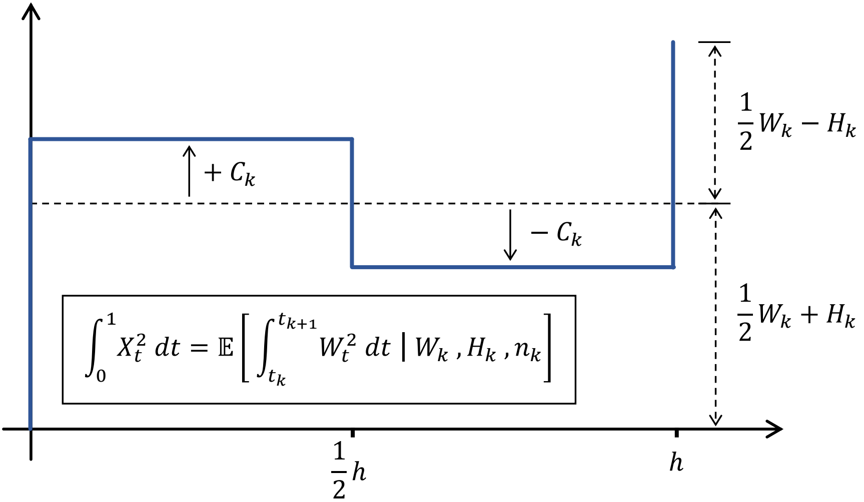
with the following parameters

- 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

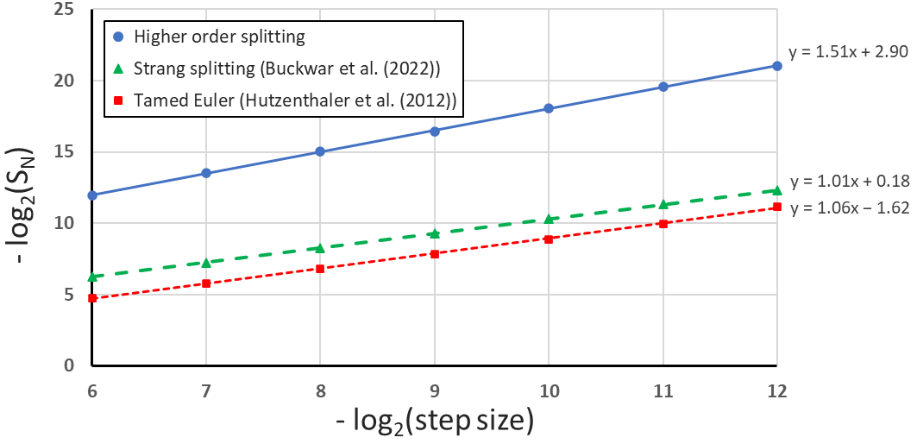
We replace each Brownian motion by the following piecewise linear path:



($n_k \in \{-1, 1\}$ 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.



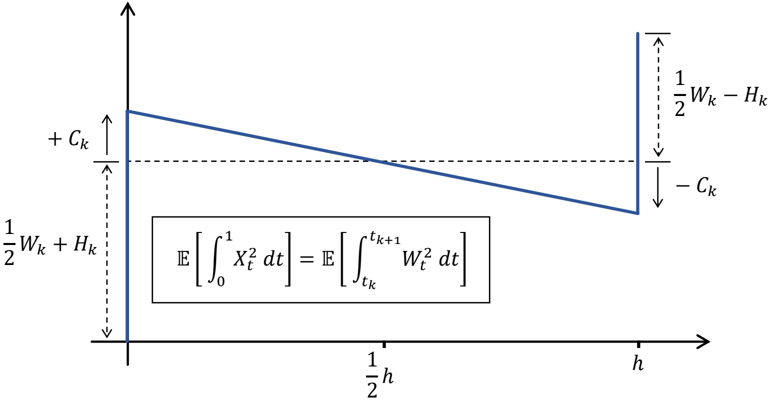
With 640 steps, we're as accurate as Strang splitting with 10,240 steps!

Example: Additive-noise SDEs

The FitzHugh-Nagumo model is an example of an additive noise SDE:

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

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 more general SDEs with additive noise, we propose using the path:



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?

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?

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

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

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

$$\begin{aligned}
Y_k^{(1)} &:= Y_k + \sigma\bigg(\frac{1}{2}W_k + H_k + C_k\bigg),\\
Y_{k+\frac{2}{3}}^{(1)} &:= Y_k^{(1)} + \frac{2}{3}\Big(f\big(Y_k^{(1)}\big)h - 2\sigma C_k\Big),\\
Y_{k+1}^{(2)} &:= Y_k^{(1)} + \frac{1}{4}f\big(Y_k^{(1)}\big)h + \frac{3}{4}f\big(Y_{k+\frac{2}{3}}^{(1)}\big)h - 2\sigma C_k,\\
Y_{k+1} &:= Y_{k+1}^{(2)} + \sigma\bigg(\frac{1}{2}W_k - H_k + C_k\bigg).
\end{aligned}$$

Example: Additive-noise SDEs

In our experiment, we compare with Rößler's strong 1.5 scheme [12],

$$\widetilde{Y}_{k+\frac{3}{4}} := Y_k + \frac{3}{4} f(Y_k)h + \frac{3}{4}\sigma W_k + \frac{3}{2}\sigma H_k,$$
$$Y_{k+1} := Y_k + \frac{1}{3} f(Y_k)h + \frac{2}{3} f\big(\widetilde{Y}_{k+\frac{3}{4}}\big)h + \sigma W_k.$$

Note that these numerical methods have a similar computational cost (two Gaussian random variables, two vector field evaluations per step).

Example: Additive-noise SDEs

In our experiment, we compare with Rößler's strong 1.5 scheme [12],

$$\widetilde{Y}_{k+\frac{3}{4}} := Y_k + \frac{3}{4}f(Y_k)h + \frac{3}{4}\sigma W_k + \frac{3}{2}\sigma H_k,$$
$$Y_{k+1} := Y_k + \frac{1}{3}f(Y_k)h + \frac{2}{3}f\big(\widetilde{Y}_{k+\frac{3}{4}}\big)h + \sigma W_k.$$

Note that these numerical methods have a similar computational cost (two Gaussian random variables, two vector field evaluations per step).

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

$$Y_{k+1} := Y_k + f\Big(Y_k + \frac{1}{2}\sigma W_k + \sigma H_k\Big)h + \sigma W_k.$$

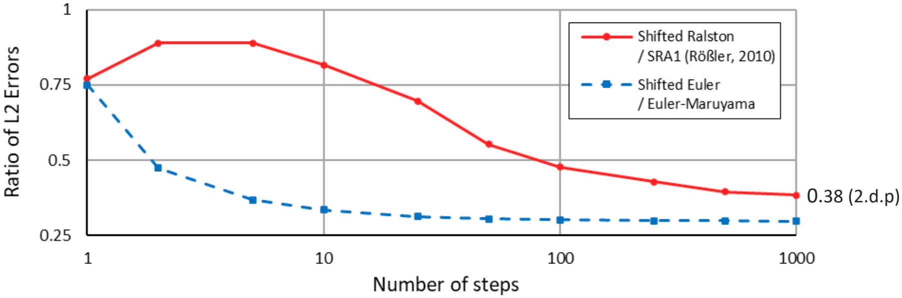
We expect this to be more accurate than the Euler-Maruyama method (though still first order convergent).

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).

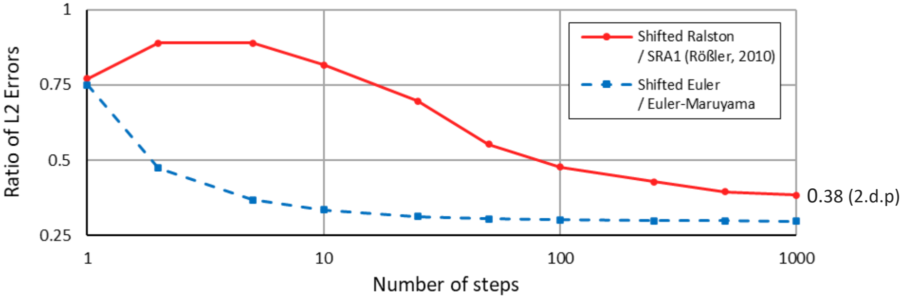


Example: Additive-noise SDEs

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).



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

Outline

Conclusion and future work

Conclusion

- Path-based framework for developing high order splitting methods
- Flexible and can exploit new approximation theory for SDEs [4, 10]
- Able to produce methods with state-of-the-art convergence rates

Future work

- Application to high-dimensional SDEs used in machine learning (such as Langevin dynamics [2, 11])
- Application to more general SDEs (i.e. not additive or scalar noise)
- Incorporating $(W, H, \cdot\,)$-based methods into Multilevel Monte Carlo

Thank you
for your attention!

Outline

References I

- 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 ICML, 2011.

- P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, Springer, 1992.

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

- 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.

A. Alfonsi. *On the discretization schemes for the CIR (and Bessel squared) processes*, Monte Carlo Methods and Applications, 2005.

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.

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.

J. Foster, T. Lyons and H. Oberhauser. *The shifted ODE method for underdamped Langevin MCMC*, arxiv:2101.03446, 2021.

A. Rößler. *Runge–Kutta Methods for the Strong Approximation of Solutions of Stochastic Differential Equations*, SIAM Journal on Numerical Analysis, vol. 8, no. 3, pp. 922–952, 2010.