UNIVERSITY OF BATH

High order splitting methods for stochastic differential equations*

James Foster

University of Bath

joint with Gonçalo dos Reis and Calum Strange (Edinburgh)

SPA invited session on “Numerical methods for SDEs: standard, mean-field and forward-backward” – Lisbon, 24 July 2023

*satisfying a commutativity condition

Outline

Introduction

Consider the following (Stratonovich) stochastic differential equation,

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

where $f, g_i : \mathbb{R}^n \to \mathbb{R}^n$ are smooth and bounded vector fields 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].

The noise terms are often "tractable" (e.g. affine noise $g_i(y) = A_i y + B_i$). As we expect numerical error to primary come from these noise terms, this provides good motivation to investigate *splitting methods for SDEs*.

A simple example: the CIR Model

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

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

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 [3].

A simple example: the CIR Model

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

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

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 [3].

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

A simple example: the CIR Model

Lie-Trotter splitting

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

$$Y_{k+1} := \text{Noise}(Y_k^{(1)}, W_k) = \left(\sqrt{Y_k^{(1)}} + \frac{1}{2}\sigma W_k\right)^2,$$

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

A simple example: the CIR Model

Lie-Trotter splitting

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

$$Y_{k+1} := \text{Noise}(Y_k^{(1)}, W_k) = \left(\sqrt{Y_k^{(1)}} + \frac{1}{2}\sigma W_k\right)^2,$$

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

Strang splitting

$$Y_k^{(1)} := \text{Drift}\left(Y_k, \frac{1}{2}h\right),$$

$$Y_k^{(2)} := \text{Noise}\big(Y_k^{(1)}, W_k\big),$$

$$Y_{k+1} := \text{Drift}\left(Y_k^{(2)}, \frac{1}{2}h\right),$$

has $O(h)$ strong convergence, but $O(h^2)$ weak convergence (see [4]).

Strang splitting

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

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

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

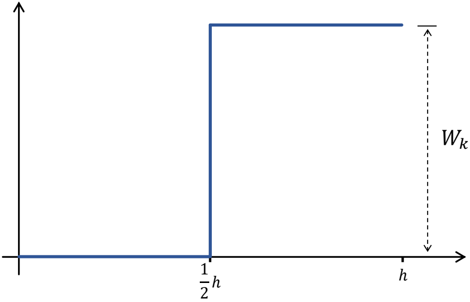
Strang splitting as a piecewise linear path

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

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

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

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 $s < t$,

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

where $h = t - s$ and the $O(h^{\frac{5}{2}})$ term is understood in an $L^2(\mathbb{P})$ sense.

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 $s < t$,

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

where $h = t - s$ and the $O(h^{\frac{3}{2}})$ term is understood in an $L^2(\mathbb{P})$ sense.

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 $s < t$,

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

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

where $h = t - s$ and the $O(h^{\frac{3}{2}})$ term is understood in an $L^2(\mathbb{P})$ sense.

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 $s < t$,

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

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

$$\vdots$$

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

$$+ \frac{1}{2}f''(y_s)\,\sigma^2 \int_s^t W_{s,u}^{\otimes 2}\,du + O\big(h^{\frac{5}{2}}\big),$$

where $h = t - s$ and the $O\big(h^{\frac{5}{2}}\big)$ term is understood in an $L^2(\mathbb{P})$ sense.

Outline

A “rough path” approach to splitting methods

Informal Theorem (Stochastic Taylor expansion [5, Thm 5.6.1])

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

$$y_t \approx y_s + f(y_s)h + \sum_{i=1}^{d} g_i(y_s)W_{s,t}^i + \sum_{i,j=1}^{d} (\cdots) \int_s^t W_{s,u}^i \circ dW_u^j \tag{3}$$
$$+ (\cdots) \int_s^t W_{s,u}\, du + (\cdots) \int_s^t (u-s)\, dW_u + (\cdots) h^2$$
$$+ (\cdots)(\text{“third” iterated integrals of } \{t, W_t\})$$
$$+ (\cdots)(\text{“fourth” iterated integrals of } W),$$

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

A "rough path" approach to splitting methods

Informal Theorem (Stochastic Taylor expansion [5, Thm 5.6.1])

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

$$
\begin{aligned}
y_t \approx y_s + f(y_s)h + \sum_{i=1}^{d} g_i(y_s) W_{s,t}^i + \sum_{i,j=1}^{d} (\cdots) \int_s^t W_{s,u}^i \circ dW_u^j \\
+ (\cdots) \int_s^t W_{s,u}\, du + (\cdots) \int_s^t (u-s)\, dW_u + (\cdots) h^2 \\
+ (\cdots)(\text{"third" iterated integrals of } \{t, W_t\}) \\
+ (\cdots)(\text{"fourth" iterated integrals of } W),
\end{aligned}
\tag{3}
$$

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 "rough path" 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 [6, Proposition 3.2])

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

$$Y_1 \approx Y_0 + f(Y_0)X_1^\tau + \sum_{i=1}^{d} g_i(Y_0)\big(X_1^\omega\big)^i + \sum_{i,j=1}^{d} (\cdots) \int_0^1 \big(X_t^\omega\big)^i d\big(X_t^\omega\big)^j$$

$$+ (\cdots) \int_0^1 X_t^\omega\, dX_t^\tau + (\cdots) \int_0^1 X_t^\tau\, dX_t^\omega + (\cdots)\big(X_1^\tau\big)^2$$

$$+ (\cdots)\big(\text{“third” iterated integrals of } \{X^\tau, X^\omega\}\big)$$

$$+ (\cdots)\big(\text{“fourth” iterated integrals of } X^\omega\big).$$

A “rough path” approach to splitting methods

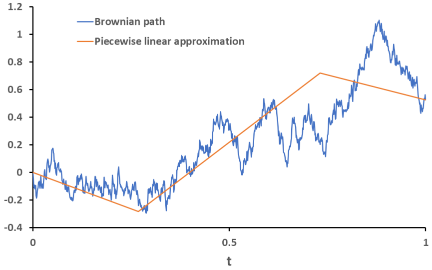
For Y to accurately approximate y, we will construct the path X so that

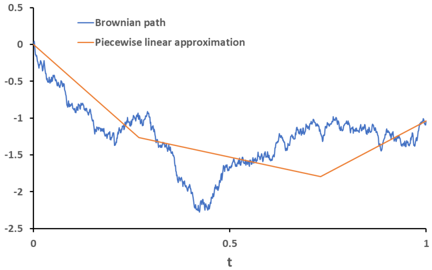
$$X_1 = \left(h, W_{s,t}\right),$$

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

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

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

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* $r \in [0,1]$

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

Main result: Error analysis for path-based splitting

Theorem (Convergence rates of splitting schemes [6, 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 $f \in \mathcal{C}^2(\mathbb{R}^n)$, $g_i \in \mathcal{C}^3(\mathbb{R}^n)$ *have Lipschitz continuous derivatives and*

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

Let φ *denote 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* $[0,1]$ *satisfying*

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

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

Main result: Error analysis for path-based splitting

Theorem (Convergence rates of splitting schemes, continued)

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

$$Y_{k+1} := \exp\Big(f(\cdot)X^{\tau}_{1,t_m} + g(\cdot)X^{\omega}_{1,t_m}\Big)\cdots\exp\Big(f(\cdot)X^{\tau}_{0,t_1} + g(\cdot)X^{\omega}_{0,t_1}\Big)Y_k, \quad (9)$$

where each piecewise linear path X has $m \geq 1$ joints at $\{t_1 < \cdots < t_m\}$ and is constructed from $\{(t, W_t)\}_{t\in[kh,(k+1)h]}$ using a step size of $h = \frac{T}{N}$. In (9), $\exp(V)x$ denotes 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

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

for $k \in \{1, \cdots, N\}$ whenever $h \leq h_{\max}$.

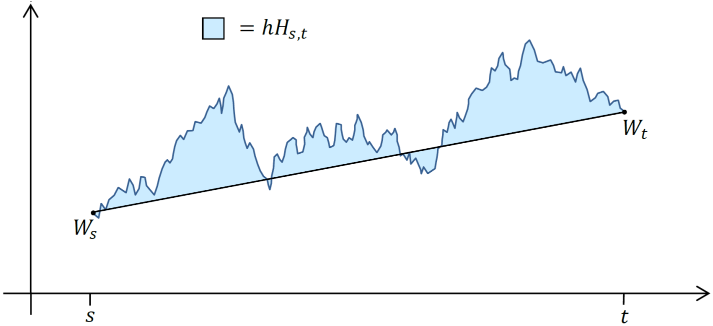
Outline

A higher order Strang splitting

Definition (Space-time Lévy area of Brownian motion)

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

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

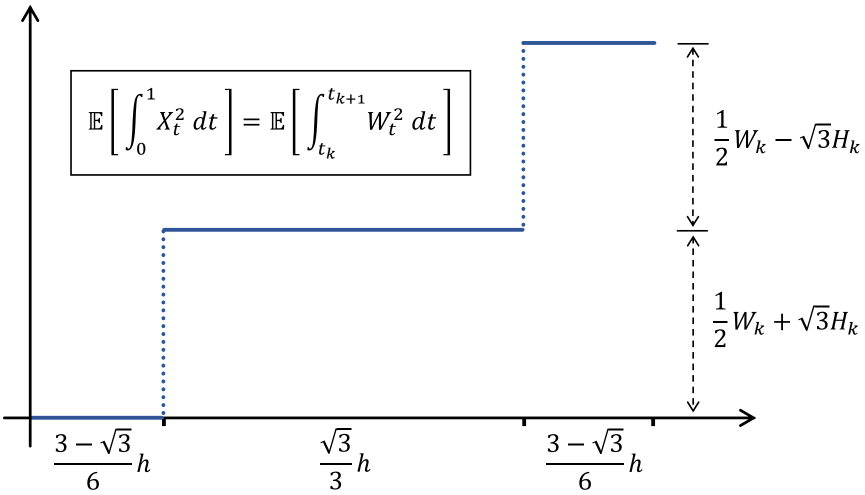


Theorem (Distribution of increments and space-time Lévy areas)

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.

A higher order Strang splitting

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



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

$$\begin{aligned}
Y_k^{(1)} &:= e^{-\frac{3-\sqrt{3}}{6}ah}Y_k + \widetilde{b}\big(1 - e^{-\frac{3-\sqrt{3}}{6}ah}\big),\\
Y_k^{(2)} &:= \left(\sqrt{Y_k^{(1)}} + \frac{\sigma}{2}\Big(\frac{1}{2}W_k + \sqrt{3}H_k\Big)\right)^2,\\
Y_k^{(3)} &:= e^{-\frac{\sqrt{3}}{3}ah}Y_k^{(2)} + \widetilde{b}\big(1 - e^{-\frac{\sqrt{3}}{3}ah}\big),\\
Y_k^{(4)} &:= \left(\sqrt{Y_k^{(3)}} + \frac{\sigma}{2}\Big(\frac{1}{2}W_k - \sqrt{3}H_k\Big)\right)^2,\\
Y_{k+1} &:= e^{-\frac{3-\sqrt{3}}{6}ah}Y_k^{(4)} + \widetilde{b}\big(1 - e^{-\frac{3-\sqrt{3}}{6}ah}\big).
\end{aligned} \tag{12}$$

Example: CIR Model (all parameters set to 1)

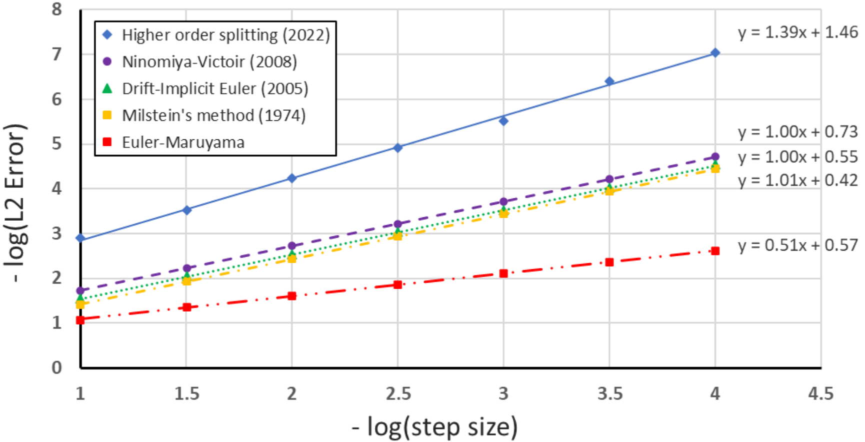


Table: Computer time to simulate 100,000 paths with 100 steps (seconds)

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 (12) 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 [9] is given by the SDE:

$$d\begin{pmatrix} V_t \\ U_t \end{pmatrix} = \begin{pmatrix} \frac{1}{\epsilon}\left(V_t - V_t^3 - U_t\right) \\ \gamma V_t - U_t + \beta \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} dW_t. \tag{13}$$

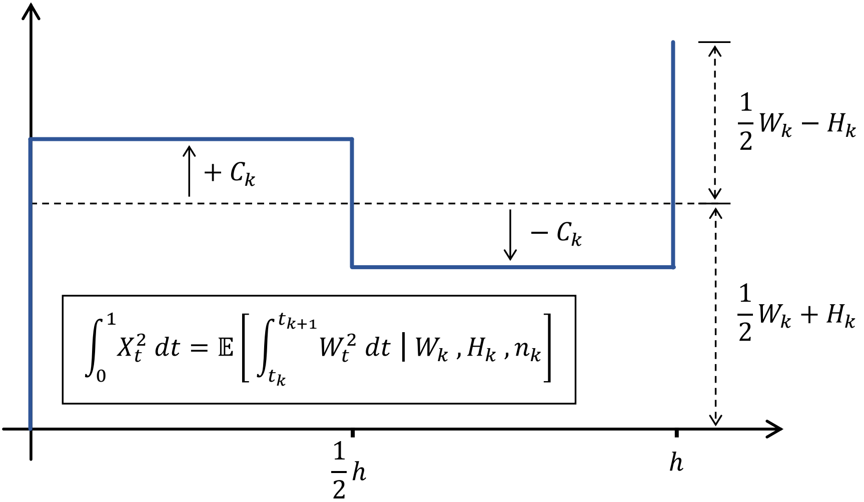
with the following parameters

- 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:

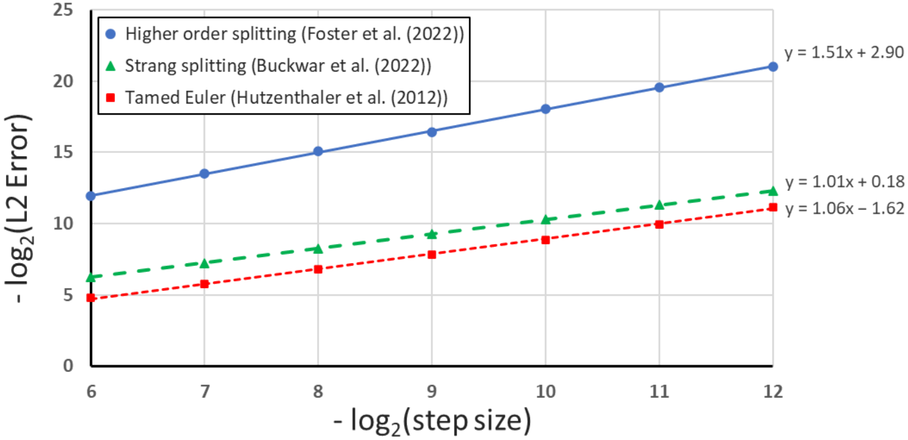


$$\int_0^1 X_t^2 \, dt = \mathbb{E}\left[\int_{t_k}^{t_{k+1}} W_t^2 \, dt \,\middle|\, W_k, H_k, n_k \right]$$

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

Outline

Conclusion and future work

Conclusion

- Path-based framework for developing high order splitting methods
- Flexible and can exploit new approximation theory for SDEs [7, 11]
- 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, 12])
- 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!

and our preprint can be found at:

J. Foster, G. dos Reis and C. Strange, *High order splitting methods for SDEs satisfying a commutativity condition*, arxiv:2210.17543, 2022.

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

References II

- J. Foster, G. dos Reis and C. Strange, *High order splitting methods for SDEs satisfying a commutativity condition*, arxiv:2210.17543, 2022.
- J. Foster. *Numerical approximations for stochastic differential equations*, University of Oxford, 2020.
- 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.

References III

- 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.

- 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.