Improving Heun's method for SDEs with additive noise

James Foster

Joint work with Terry Lyons and Harald Oberhauser

University of Oxford

2021

Introduction

We consider stochastic differential equations (SDEs) with the form:

$$dy_t = f(y_t)\,dt + \sigma\,dW_t, \tag{1}$$

where the solution $y = \{y_t\}_{t\in[0,T]}$ takes its values in $\mathbb{R}^d$, the function (also known as the "vector field") $f:\mathbb{R}^d \to \mathbb{R}^d$ is sufficiently regular (e.g. globally Lipschitz continuous and with linear growth), $\sigma \in \mathbb{R}^{d\times w}$ is a $d\times w$ matrix and W denotes a w-dimensional Brownian motion.

- Stochastic gradient system [1] with potential energy $F:\mathbb{R}^d \to \mathbb{R}$

$$dy_t = -\nabla F(y_t)\,dt + \sigma\,dW_t.$$

- Lotka-Volterra model with diffusion [2] (a predator-prey model)

$$dx_t = (b_1 - a_1 y_t)x_t\,dt + \sigma_1\,dW_t^{(1)},$$

$$dy_t = (b_2 - a_2 x_t)y_t\,dt + \sigma_2\,dW_t^{(2)}.$$

Introduction

In general, (1) cannot be solved exactly and must be approximated.

The two most popular methods for solving additive noise SDEs are:

1. The Euler-Maruyama method [3, 4, 5]. For additive noise SDEs, this method coincides with the higher order Milstein's method.

$$Y_{n+1}^{\mathsf{EM}} := Y_n^{\mathsf{EM}} + f\big(Y_n^{\mathsf{EM}}\big)(t_{n+1} - t_n) + \sigma\big(W_{t_{n+1}} - W_{t_n}\big). \tag{2}$$

2. Heun's method [1, 4, 6, 7], which involves two evaluations of f.

$$\begin{aligned}\widetilde{Y}_{n+1}^{\mathsf{H}} &:= Y_n^{\mathsf{H}} + f\big(Y_n^{\mathsf{H}}\big)(t_{n+1} - t_n) + \sigma\big(W_{t_{n+1}} - W_{t_n}\big),\\ Y_{n+1}^{\mathsf{H}} &:= Y_n^{\mathsf{H}} + \frac{1}{2}\Big(f\big(Y_n^{\mathsf{H}}\big) + f\big(\widetilde{Y}_{n+1}^{\mathsf{H}}\big)\Big)(t_{n+1} - t_n) + \sigma\big(W_{t_{n+1}} - W_{t_n}\big).\end{aligned} \tag{3}$$

We note that each increment of the Brownian motion over $[t_n, t_{n+1}]$ is independent and distributed as $\big(W_{t_{n+1}} - W_{t_n}\big) \sim N\big(0, (t_{n+1} - t_n)I_w\big)$.

Introduction

Definition (Strong convergence)

A numerical solution Y for (1) is said to converge in a strong sense with order α if there exists a constant $C > 0$ such that

$$\left\| Y_N - y_T \right\|_{L^2(\mathbb{P})} \leq C h^{\alpha}, \tag{4}$$

for all sufficiently small step sizes $h = \frac{T}{N}$, where we define $\| \cdot \|_{L^2(\mathbb{P})}$ as

$$\| \cdot \|_{L^2(\mathbb{P})} := \sqrt{\mathbb{E}\left[\| \cdot \|_2^2 \right]}.$$

Definition (Weak convergence)

A numerical solution Y for (1) is said to converge in a weak sense with order β if for any polynomial p there exists $C_p > 0$ such that

$$\left| \mathbb{E}\left[p(Y_N) \right] - \mathbb{E}\left[p(y_T) \right] \right| \leq C_p h^{\beta}, \tag{5}$$

for all sufficiently small step sizes $h = \frac{T}{N}$.

A Heun-based stochastic Runge-Kutta method

The following stochastic Runge-Kutta (SRK) method given in [6] achieves both strong order 1.5 and weak order 2.0 convergence.

$$\widetilde{Y}_{n+1}^{\mathsf{RK}} := Y_n^{\mathsf{RK}} + \frac{3+\sqrt{6}}{6}\sigma W_n + \sigma H_n,$$

$$\widehat{Y}_{n+1}^{\mathsf{RK}} := Y_n^{\mathsf{RK}} + f\big(Y_n^{\mathsf{RK}}\big)h + \frac{3-\sqrt{6}}{6}\sigma W_n + \sigma H_n,$$

$$Y_{n+1}^{\mathsf{RK}} := Y_n^{\mathsf{RK}} + \frac{1}{2}\Big(f\big(\widetilde{Y}_{n+1}^{\mathsf{RK}}\big) + f\big(\widehat{Y}_{n+1}^{\mathsf{RK}}\big)\Big)h + \sigma W_n, \tag{6}$$

where $h = \frac{T}{N}$ and $H_n \sim N\big(0, \frac{1}{12}hI_w\big)$ is independent of $W_n \sim N\big(0, hI_w\big)$. Moreover, H_n can be defined from $\{W_t\}_{t\in[t_n,t_{n+1}]}$ (see [8] for details).

$$W_n := W_{t_{n+1}} - W_{t_n},$$

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

Introduction

Table: Convergence rates for numerical methods in the additive noise case

	Number of vector field evaluations per step	Type of convergence		
		Strong	Weak	$\sigma = 0$
Euler-Maruyama	1	$O(h)$	$O(h)$	$O(h)$
Heun	2	$O(h)$	$O(h^2)$	$O(h^2)$
SRK	3	$O(h^{1.5})$	$O(h^2)$	$O(h^2)$

We note that for Heun's method and SRK to achieve higher order convergence, we require additional smoothness assumptions on f (for example, f' and f'' to be bounded and Lipschitz continuous).

The non-Markov Euler-Maruyama method

The following non-Markov version of the Euler-Maruyama method is considered in [1, 9] for when $f = -\nabla F$ and the SDE (1) is ergodic.

$$Y_{n+1}^{\mathsf{NEM}} := Y_n^{\mathsf{NEM}} + f\big(Y_n^{\mathsf{NEM}}\big)h + \frac{1}{\sqrt{2}}\sigma\big(\xi_n + \xi_{n+1}\big), \tag{7}$$

where $h = \frac{T}{N}$ and $\{\xi_n\}$ are iid random variables with $\xi_n \sim N\big(0, hI_w\big)$.

Whilst this method has first order weak convergence on a finite time horizon, it achieves second order convergence in the limit as $t \to \infty$.

Moreover, this non-Markovian Euler-Maruyama method has shown superior performance at targeting the SDE's stationary distribution than Heun's method.

AdHoc method (Additive-noise Heun with one computation)

Instead of adapting the Euler-Maruyama method to be high order, we slightly alter Heun's method by reusing vector field evaluations.

Definition (The AdHoc method)

We construct a numerical solution $Y = \{Y_n\}_{n\geq 0}$ for the SDE (1) by setting $Y_0 = \widetilde{Y}_0 = y_0$ and for $n \geq 0$ defining Y_{n+1} from $(Y_n, f(\widetilde{Y}_n))$ as

$$\widetilde{Y}_{n+1} := Y_n + f(\widetilde{Y}_n)h_n + \sigma W_n,$$

$$Y_{n+1} := Y_n + \frac{1}{2}\big(f(\widetilde{Y}_n) + f(\widetilde{Y}_{n+1})\big)h_n + \sigma W_n, \tag{8}$$

where $h_n := t_{n+1} - t_n$ and $W_n := W_{t_{n+1}} - W_{t_n} \sim N(0, h_n I_w)$.

This numerical method was inspired by the asynchronous leapfrog (ALF) solver [10, 11] which uses just one evaluation of f per step.

AdHoc method (Additive-noise Heun with one computation)

Since we store $f(\widetilde{Y}_n)$, we only have to evaluate $f(\widetilde{Y}_{n+1})$ in each step.

Therefore the AdHoc method evaluates $f(\cdot)$ once per step and thus has the same computational cost as the Euler-Maruyama method.

Of course, this comes at a price and the resulting approximation is likely to be less accurate than Heun's method (for a given step size).

Based on our intuition and some numerical evidence, we conjecture that the AdHoc method will generally outperform Euler-Maruyama.

The AdHoc method (strong convergence)

For the error analysis, it suffices to compare against Heun's method.

We assume f is twice continuously differentiable with its derivatives f' and f'' globally bounded. We shall define $M := \sup\limits_{y\in\mathbb{R}^d} \|f'(y)\|_2 < \infty$.

Theorem (Local error estimate for AdHoc and Heun methods)

Let $h_{\max} \in (0,1)$ be fixed. Then there exists a constant C_1 such that

$$\begin{aligned}\|Y_{n+1} - Y_{n+1}^H\|_{L^2(\mathbb{P})}^2 &\leq \left(1 + \frac{1}{2}\big(1+4M^2\big)h_n + M^2 h_n^2\right)\|Y_n - Y_n^H\|_{L^2(\mathbb{P})}^2\\ &\quad + \frac{1}{2}M^2\big(1+2M^2h_n^2\big)\big(2h_n + h_n^2\big)\|\widetilde{Y}_n - Y_n^H\|_{L^2(\mathbb{P})}^2,\end{aligned}$$

$$\begin{aligned}\|\widetilde{Y}_{n+1} - Y_{n+1}^H\|_{L^2(\mathbb{P})}^2 &\leq \left(1 + \frac{1}{2}h_n\right)\|Y_n - Y_n^H\|_{L^2(\mathbb{P})}^2\\ &\quad + 3M^2\big(h_n + h_n^2\big)\|\widetilde{Y}_n - Y_n^H\|_{L^2(\mathbb{P})}^2 + C_1 h_n^3,\end{aligned}$$

for $h_n = t_{n+1} - t_n \leq h_{\max}$.

The AdHoc method (strong convergence)

Sketch Proof.

The result follows by a direct calculation that uses certain moment bounds for Heun's method (i.e. $\sup_{n\geq 0}\mathbb{E}\big[\|Y_n^{\mathsf{H}}\|_2^2\big]<\infty$). We start with

$$
\begin{aligned}
&\|Y_{n+1}-Y_{n+1}^{\mathsf{H}}\|_{L^2(\mathbb{P})}^2\\
&\quad=\|Y_n-Y_n^{\mathsf{H}}\|_{L^2(\mathbb{P})}^2+\left\|\frac{1}{2}\Big(f\big(\widetilde{Y}_n\big)-f\big(Y_n^{\mathsf{H}}\big)+f\big(\widetilde{Y}_{n+1}\big)-f\big(\widetilde{Y}_{n+1}^{\mathsf{H}}\big)\Big)h_n\right\|_{L^2(\mathbb{P})}^2\\
&\qquad\quad+\mathbb{E}\Big[\Big\langle Y_n-Y_n^{\mathsf{H}},\Big(f\big(\widetilde{Y}_n\big)-f\big(Y_n^{\mathsf{H}}\big)+f\big(\widetilde{Y}_{n+1}\big)-f\big(\widetilde{Y}_{n+1}^{\mathsf{H}}\big)\Big)h_n\Big\rangle\Big]\\
&\quad\leq\Big(1+\frac{1}{2}h_n\Big)\|Y_n-Y_n^{\mathsf{H}}\|_{L^2(\mathbb{P})}^2\\
&\qquad\quad+\frac{1}{4}\Big\|\Big(f\big(\widetilde{Y}_n\big)-f\big(Y_n^{\mathsf{H}}\big)+f\big(\widetilde{Y}_{n+1}\big)-f\big(\widetilde{Y}_{n+1}^{\mathsf{H}}\big)\Big)\Big\|_{L^2(\mathbb{P})}^2\big(2h_n+h_n^2\big),
\end{aligned}
$$

where the last line follows by Youngs's inequality. We then apply the triangle inequality before using the Lipschitz continuity of f.

The AdHoc method (strong convergence)

Sketch Proof. (continued)

The second inequality can be similarly shown, expect we now take an $\mathscr{F}_{t_n}$-conditional expectation within the inner product and apply

$$\begin{aligned} f\big(\widetilde{Y}^{\mathsf{H}}_{n+1}\big) &= f\big(Y^{\mathsf{H}}_n + f\big(Y^{\mathsf{H}}_n\big)h_n + \sigma W_n\big) \\ &= f\big(Y^{\mathsf{H}}_n\big) + f'\big(Y^{\mathsf{H}}_n\big)\big(f\big(Y^{\mathsf{H}}_n\big)h_n + \sigma W_n\big) + R^{\mathsf{H}}_n, \end{aligned}$$

where the remainder term R^{H}_n is given by

$$R^{\mathsf{H}}_n = \int_0^1 (1-r) f''\big(Y^{\mathsf{H}}_n + r\big(f\big(Y^{\mathsf{H}}_n\big)h_n + \sigma W_n\big)\big)\,dr\,\big(f\big(Y^{\mathsf{H}}_n\big)h_n + \sigma W_n\big)^{\otimes 2},$$

which can be estimated since f'' and the second moment of Y^{H}_n are bounded. Since we take an $\mathscr{F}_{t_n}$-conditional expectation of $f\big(\widetilde{Y}^{\mathsf{H}}_{n+1}\big)$, the W_n term will disappear and we can apply the boundedness of f'.

The AdHoc method (strong convergence)

From these estimates, it follows that there exists $C_0 > 0$ such that

$$\begin{aligned}\| Y_{n+1} - Y^{\mathsf{H}}_{n+1} \|^2_{L^2(\mathbb{P})} &+ \| \widetilde{Y}_{n+1} - Y^{\mathsf{H}}_{n+1} \|^2_{L^2(\mathbb{P})} \\ &\leq (1 + C_0 h_n)\Big(\| Y_n - Y^{\mathsf{H}}_n \|^2_{L^2(\mathbb{P})} + \| \widetilde{Y}_n - Y^{\mathsf{H}}_n \|^2_{L^2(\mathbb{P})} \Big) + C_1 h_n^3.\end{aligned}$$

Since $Y_1 = Y^{\mathsf{H}}_1$ and $\| \widetilde{Y}_1 - Y^{\mathsf{H}}_1 \|^2_{L^2(\mathbb{P})} \sim O(h^3)$, it is now easy to show:

Theorem (Global error estimate for AdHoc and Heun methods)

Let $h_{\max} \in (0,1)$ be fixed. Then there exists a constant C such that

$$\| Y_n - Y^H_n \|^2_{L^2(\mathbb{P})} + \| \widetilde{Y}_n - Y^H_n \|^2_{L^2(\mathbb{P})} \leq C h^2.$$

for $0 \leq n \leq N$, where Y, Y^H are computed with a step size $h \leq h_{\max}$.

The AdHoc method (weak convergence)

So by the triangle inequality, we obtain the following error estimate

$$\| Y_n - y_{t_n} \|_{L^2(\mathbb{P})} \leq \| Y_n^{\mathsf{H}} - y_{t_n} \|_{L^2(\mathbb{P})} + O(h).$$

Since Heun's method converges strongly for (1) with first order [4], it follows that the AdHoc method also has $O(h)$ strong convergence.

However, it is ongoing research to quantify the weak convergence of the method. In our numerical experiment, we will see that the AdHoc method can demonstrate second order weak convergence!

Open question

What is the weak convergence rate of the AdHoc method?

The ASH method (Additive-noise Shifted Heun)

Definition (The ASH method)

We construct a numerical solution $Y^{\mathsf{S}} = \{Y_n^{\mathsf{S}}\}_{n\geq 0}$ for the SDE (1) by setting $Y_0^{\mathsf{S}} := y_0$ and for each $n \geq 0$ defining Y_{n+1}^{S} from Y_n^{S} as

$$\begin{aligned}
\widetilde{Y}_n^{\mathsf{S}} &:= Y_n^{\mathsf{S}} + \frac{3-\sqrt{6}}{6}\sigma W_n + \sigma H_n, \\
\widehat{Y}_{n+1}^{\mathsf{S}} &:= \widetilde{Y}_n^{\mathsf{S}} + f\big(\widetilde{Y}_n^{\mathsf{S}}\big)h_n + \frac{\sqrt{6}}{3}\sigma W_n, \\
Y_{n+1}^{\mathsf{S}} &:= Y_n^{\mathsf{S}} + \frac{1}{2}\Big(f\big(\widetilde{Y}_n^{\mathsf{S}}\big) + f\big(\widehat{Y}_{n+1}^{\mathsf{S}}\big)\Big)h_n + \sigma W_n, \qquad (9)
\end{aligned}$$

where $h_n = t_{n+1} - t_n$ and (W_n, H_n) are independent random vectors:

$$W_n \sim N\big(0, h_n I_w\big),$$

$$H_n \sim N\Big(0, \frac{1}{12}h_n I_w\Big).$$

The ASH method (Additive-noise Shifted Heun)

The word "shifted" refers to us modifying Y_n^{S} before evaluating $f(\cdot)$.

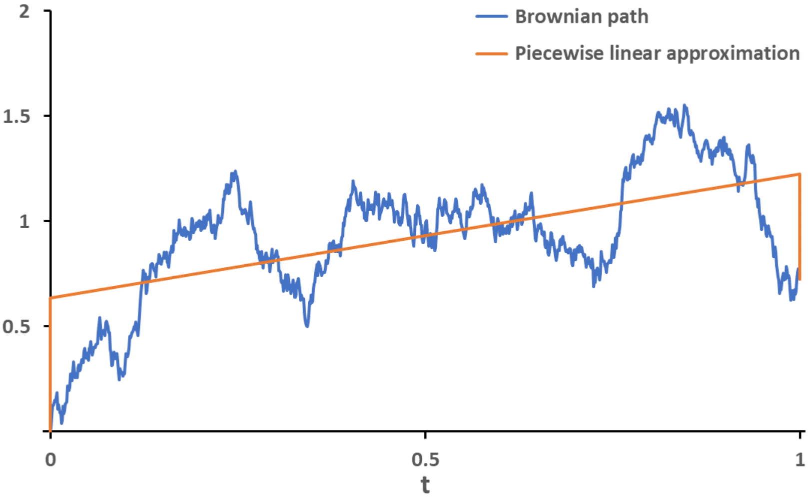

Figure: The ASH method is based on driving the SDE by a certain path.

The ASH method (Additive-noise Shifted Heun)

Unsurprisingly, the ASH method gives the same Taylor expansion as the SRK method (6) when we exclude the $O(h^{2.5})$ remainder terms.

$$\begin{aligned} Y_{n+1}^{\mathsf{S}} &\approx Y_n^{\mathsf{S}} + f(Y_n^{\mathsf{S}})h_n + \sigma W_n \\ &\quad + \sigma f'(Y_n^{\mathsf{S}})\left(\frac{1}{2}h_n W_n + h_n H_n\right) + \frac{1}{2}f'(Y_n^{\mathsf{S}})f(Y_n^{\mathsf{S}})h_n^2 \\ &\quad + \sigma^2 f''(Y_n^{\mathsf{S}})\left(\frac{5}{24}W_n^{\otimes 2} + \frac{1}{4}W_n\otimes H_n + \frac{1}{4}H_n\otimes W_n + \frac{1}{2}H_n^{\otimes 2}\right)h_n. \end{aligned}$$

Therefore it also converges with strong order 1.5 and weak order 2.0.

However the new method uses two evaluations of f instead of three!

Improving Heun for underdamped Langevin dynamics (ULD)

Consider the following SDE on $\mathbb{R}^{2d}$,

$$dx_t = v_t\,dt, \tag{10}$$

$$dv_t = -\gamma v_t\,dt - u\nabla F(x_t)\,dt + \sqrt{2\gamma u}\,dW_t, \tag{11}$$

where $\gamma, u > 0$ denote the friction and gradient coefficients.

Under mild assumptions on F, the SDE admits a unique strong solution that is ergodic with stationary distribution $\pi(x,v) \propto e^{-F(x)+\frac{1}{2u}\|v\|^2}$ [12].

Hence by simulating ULD, we can generate samples from $\pi(x) \propto e^{-F(x)}$. (technically, samples are "close" to π in an optimal transport sense [13])

To simulate (10), we employ techniques inspired by the AdHoc method. Due to the structure of ULD, we conjecture our method is second order.

The LIGHT method (Langevin with Interpolated Gradients and Heun-like Time-stepping)

1. Evaluate ∇F at X_n
2. Compute X_{n+1} by solving the following SDE on $[t_n, t_{n+1}]$:

$$\begin{aligned} dx_t &= v_t\, dt, \\ dv_t &= -\lambda v_t\, dt - u\nabla F(X_n)\, dt + \sqrt{2\gamma u}\, dW_t, \end{aligned} \tag{12}$$

   with initial value (X_n, V_n).

   Solving both parts of (12) gives the exponential Euler method [13].

3. Evaluate ∇F at X_{n+1}
4. Compute V_{n+1} by solving the following SDE on $[t_n, t_{n+1}]$:

$$dv_t = -\lambda v_t\, dt - u\left(\nabla F(X_n) + \frac{t - t_n}{h_n}\left(\nabla F(X_{n+1}) - \nabla F(X_n)\right)\right)dt + \sqrt{2\gamma u}\, dW_t.$$

   with initial value V_n.

The LIGHT method (Langevin with Interpolated Gradients and Heun-like Time-stepping)

Like AdHoc, this only requires one extra gradient evaluation per step.

Definition (The LIGHT method)

We construct a numerical solution $\{X_n, V_n\}_{n\geq 0}$ for the SDE (10) by setting $V_0 = v_0$, sampling X_0 and for $n \geq 0$ defining (X_{n+1}, V_{n+1}) as

$$X_{n+1} := X_n + \Big(\frac{1-e^{-\gamma h_n}}{\gamma}\Big)V_n - \Big(\frac{e^{-\gamma h_n}+\gamma h_n - 1}{\gamma^2}\Big)u\nabla F(X_n) \qquad (13)$$
$$+ \sqrt{2\gamma u}\int_{t_n}^{t_{n+1}}\int_{t_n}^{t} e^{-\gamma(t-s)}\,dW_s\,dt,$$

$$V_{n+1} := e^{-\gamma h_n}V_n - \Big(\frac{1-(1+\gamma h_n)e^{-\gamma h_n}}{\gamma^2 h_n}\Big)u\nabla F(X_n) \qquad (14)$$
$$- \Big(\frac{e^{-\gamma h_n}+\gamma h_n - 1}{\gamma^2 h_n}\Big)u\nabla F(X_{n+1}) + \sqrt{2\gamma u}\int_{t_n}^{t_{n+1}} e^{-\gamma(t_{n+1}-t)}\,dW_t.$$

Convergence in the 2-Wasserstein metric (constant step size)

Numerical method for ULD	Assumptions on the strongly convex F	Number of steps to achieve an error of $W_2(X_n, e^{-F}) \leq \varepsilon$
Exponential Euler method [13]	Lipschitz gradient	$\mathcal{O}(\sqrt{d}/\varepsilon)$
Second order Kinetic Langevin [14, 15] OBABO splitting [16, 17] **Conjectured:** LIGHT, Strang splitting [18]	Lipschitz gradient + Lipschitz $\nabla^2 F$	$\mathcal{O}(\sqrt{d}/\sqrt{\varepsilon})$ $\mathcal{O}(\sqrt{d}/\varepsilon)$
Randomized midpoint [19, 20]	Lipschitz gradient	$\mathcal{O}(\sqrt[3]{d}/\varepsilon^{\frac{2}{3}})$

Numerical example: Scalar anharmonic oscillator

We consider the following scalar SDE,

$$dy_t = \sin\left(y_t\right)dt + dW_t,$$

with $y_0 = 1$ and define the following error estimators:

$$S_N := \sqrt{\mathbb{E}\left[\left(Y_N - Y_T^{\text{fine}}\right)^2\right]},$$

$$E_N := \left|\mathbb{E}\left[Y_N\right] - \mathbb{E}\left[Y_T^{\text{fine}}\right]\right|,$$

$$V_N := \left|\mathbb{E}\left[Y_N^2\right] - \mathbb{E}\left[\left(Y_T^{\text{fine}}\right)^2\right]\right|,$$

where the expectations are approximated by Monte-Carlo simulation and Y_T^{fine} is the numerical solution of (1) obtained at the time $T = 1$ using Heun's method with a "fine" step size of $\frac{h}{10}$.

We will compute both Y_N and Y_T^{fine} using the same Brownian paths.

Numerical example: Scalar anharmonic oscillator

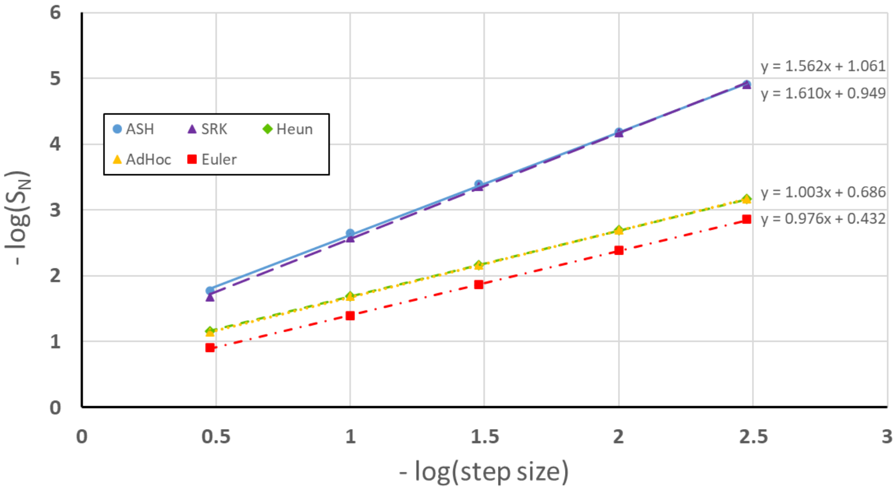

Figure: S_N computed with 1,000,000 sample paths using a fixed step size.

Numerical example: Scalar anharmonic oscillator

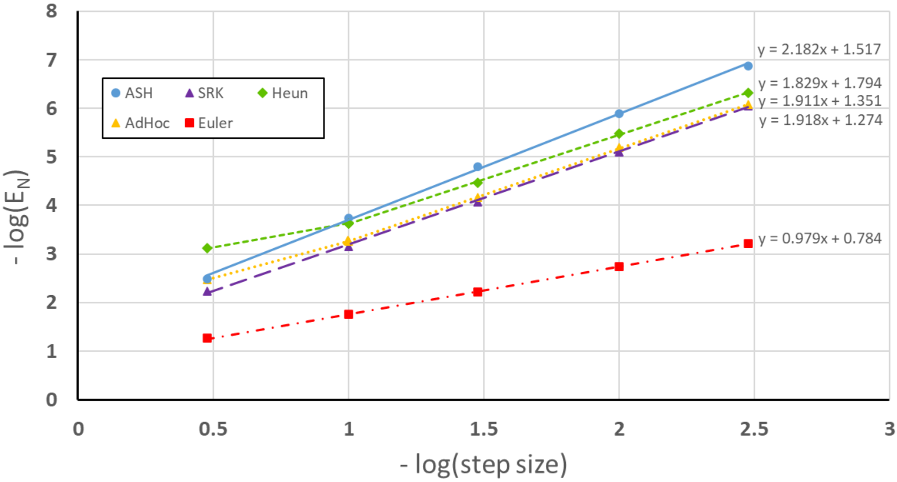

Figure: E_N computed with 1,000,000 sample paths using a fixed step size.

Numerical example: Scalar anharmonic oscillator

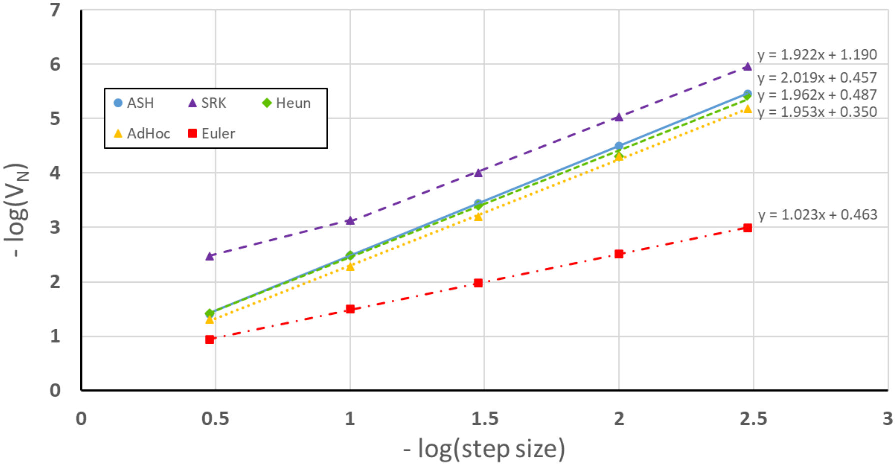

Figure: V_N computed with 1,000,000 sample paths using a fixed step size.

Numerical example: Logistic regression

- The dataset is m pairs of labels $y_i \in \{-1, 1\}$ and features $x_i \in \mathbb{R}^d$.
- Target density $\pi(\theta) \propto \exp(-F(\theta))$ comes from a logistic regression:

$$F(\theta) = \frac{\delta}{2}\|\theta\|_2^2 + \sum_{i=1}^{m} \log\Big(1 + \exp\big(-y_i x_i^{\mathsf{T}}\theta\big)\Big),$$

  where δ is a regularization parameter which we will set to $\delta = 0.1$.

- German credit data from UCI repository [21] $(m = 1000,\ d = 49)$.
- Estimate $L^2(\mathbb{P})$ error by simulating chains with step sizes h and $\frac{1}{2}h$:

$$S_{N,n} := \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\overrightarrow{\theta}_{N,i}^{\,h} - \overrightarrow{\theta}_{N,i}^{\,\frac{1}{2}h}\right\|_2^2},$$

  where we use a fixed time horizon $T = 1000$ and step size $h = \frac{T}{N}$.

Numerical example: Logistic regression

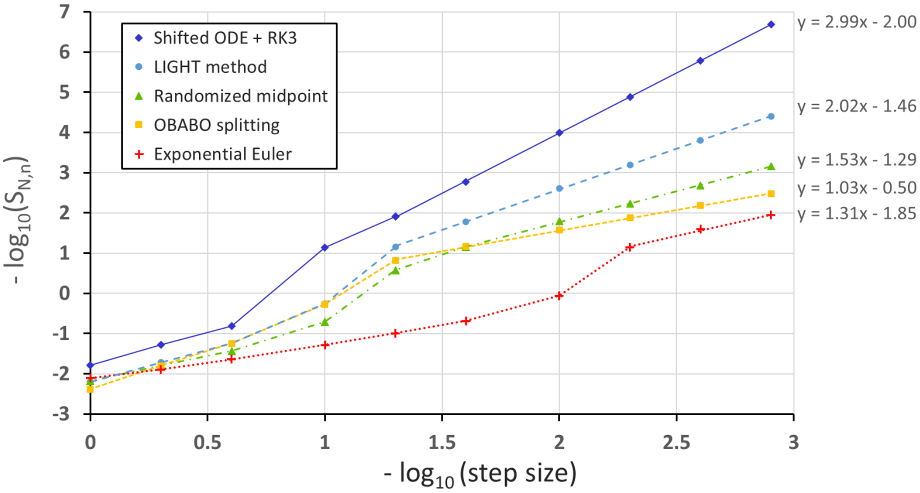

Figure: $S_{N,n}$ computed with $n = 100$ sample paths using a fixed step size.

Future work

- Would one expect AdHoc to outperform the Euler-Maruyama method in high-dimensional settings?

  (for example, in Stochastic gradient Langevin dynamics [23])

- What is the order of weak convergence for the AdHoc method?

- Does the AdHoc method converge if we use variable step sizes?

- Could ASH extend to SDEs with scalar or commutative noise?

- To what extent do noisy gradients reduce accuracy for LIGHT?

Thank you
for your attention!

References I

[1] B. Leimkuhler, C. Matthews and M. V. Tretyakov, *On the long-time integration of stochastic gradient systems*, Proceedings of the Royal Society A, vol. 470, no. 2170, 2014.

[2] M. Arató, *A famous nonlinear stochastic equation (Lotka-Volterra model with diffusion)*, Mathematical and Computer Modelling, vol. 38, no. 7-9, pp. 709–726, 2003.

[3] D. J. Higham, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Review, vol. 43, no. 3, pp. 525–546, 2001.

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

References II

[5] O. Butkovsky, K. Dareiotis, M. Gerencsér, *Approximation of SDEs - a stochastic sewing approach*, https://arxiv.org/abs/1909.07961, 2020.

[6] G. N. Milstein and M. V. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Springer, 2004.

[7] K. Burrage, P. M. Burrage and T. Tian, *Numerical methods for strong solutions of stochastic differential equations: an overview*, Proceedings of the Royal Society A, vol. 460, no. 2041, 2004.

[8] 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.

References III

[9] B. Leimkuhler and C. Matthews, *Rational construction of stochastic numerical methods for molecular sampling*, Applied Mathematics Research eXpress 2013, A, vol 1, pp. 34–56, 2013.

[10] U. Mutze, *An asynchronous leapfrog method*, Mathematical Physics Preprint Archive 2008, no. 197, https://web.ma.utexas.edu/mp_arc/c/08/08-197.pdf, 2008.

[11] U. Mutze, *An asynchronous leapfrog method II*, https://arxiv.org/abs/1311.6602, 2013.

[12] G. A. Pavliotis. *Stochastic Processes and Applications*, Springer, New York, 2014.

References IV

[13] X. Cheng, N. S. Chatterji, P. L. Bartlett and M. I. Jordan. *Underdamped Langevin MCMC: A non-asymptotic analysis*, Proceedings of Machine Learning Research, vol. 75, 2018.

[14] A. S. Dalalyan and L. Riou-Durand. *On sampling from a log-concave density using kinetic Langevin diffusions*, Bernoulli, vol. 26, no.3, pp. 1956–1988, 2020.

[15] A. S. Dalalyan, A. Karagulyan and L. Riou-Durand. *Bounding the error of discretized Langevin algorithms for non-strongly log-concave targets*, https://arxiv.org/abs/1906.08530, 2020.

[16] Z. Song and Z. Tan. *Hamiltonian Assisted Metropolis Sampling*, https://arxiv.org/pdf/2005.08159, 2020.

References V

[17] P. Monmarché. *High-dimensional MCMC with a standard splitting scheme for the underdamped Langevin diffusion*, https://arxiv.org/pdf/2007.05455, 2020.

[18] E. Buckwar, M. Tamborrino and I. Tubikanec. *Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs*, Statistics and Computing, vol. 30, no. 3, pp. 627–648 2020.

[19] R. Shen and Y. T. Lee. *The Randomized Midpoint Method for Log-Concave Sampling*, Advances in Neural Information Processing Systems, 2019.

References VI

[20] Y. He, K. Balasubramanian and M. A. Erdogdu. *On the Ergodicity, Bias and Asymptotic Normality of Randomized Midpoint Sampling Method*, Advances in Neural Information Processing Systems, 2020.

[21] M. Lichman. *UCI machine learning repository*, https://archive.ics.uci.edu/ml, 2013.

[22] J. Foster, T. Lyons and H. Oberhauser, *The shifted ODE method for underdamped Langevin MCMC*, https://arxiv.org/abs/2101.03446, 2021.

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