The Performance of the Euler Scheme for SDEs with Discontinuous Drift

Tim Johnston [1]

[1]School of Mathematics, University of Edinburgh

15 July 2024

Main Takeaway

The accuracy of numerical simulations of SDEs does not **necessarily** depend on the regularity of the coefficients

Strong Convergence

Consider stochastic numerical approximations $X^1, X^2, \ldots$ converging to some object of interest X. In this presentation we are interested in **strong convergence**, in particular **convergence in L^p**. Specifically, X^n converges to X in L^p if

$$(E|X^n - X|^p)^{1/p} \to 0 \tag{1}$$

as $n \to \infty$. This is **pathwise convergence**. We require that for large n

$$|X^n(\omega) - X(\omega)|, \tag{2}$$

is 'small' for 'most' $\omega \in \Omega$. We are interested in strong convergence because:

- strong convergence informs more about qualitative properties of dynamics (for instance, proof of strong solutions in Gyöngy, Krylov 1996)
- allows us to control difference between $\mathcal{L}(X^n)$ and $\mathcal{L}(X)$ in some metrics (Wasserstein distance)
- multi-level monte carlo: one may achieve superior bounds by writing

$$f(X^n) = f(X^1) + \sum_{i=1}^{n-1} f(X^{i+1}) - f(X^i), \tag{3}$$

  but generally requires strong convergence of X^n.

What makes it difficult to simulate an SDE?

What properties of

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{4}$$

mean that the Euler scheme approximation X_t^n given for $t \in [\frac{m}{n}, \frac{m+1}{n}]$ as

$$X_t^n = X_{m/n}^n + (t - m/n)b(X_{m/n}^n) + \sigma(X_{m/n}^n)(W_t - W_{m/n}). \tag{5}$$

is accurate? Likely to be more accurate if the dynamics do not depend 'too sensitively' on the space variable - i.e. if there is some **regularity** to b and σ

What makes it difficult to simulate an SDE?

Indeed, if b and σ obey the Lipschitz assumption

$$|b(x) - b(y)| \leq L|x - y|, \tag{6}$$

$$|\sigma(x) - \sigma(y)| \leq L|x - y|, \tag{7}$$

one has the following (classical) result

Theorem

Suppose b, σ are Lipschitz. Then for every $p, T > 0$ there exists $c > 0$ such that

$$(E \sup_{t \in [0,T]} |X_t^n - X_t|^p)^{1/p} \leq cn^{-1/2}. \tag{8}$$

Now suppose ∇b exists and is also a Lipschitz function, and σ is constant. Then

$$(E \sup_{t \in [0,T]} |X_t^n - X_t|^p)^{1/p} \leq cn^{-1}. \tag{9}$$

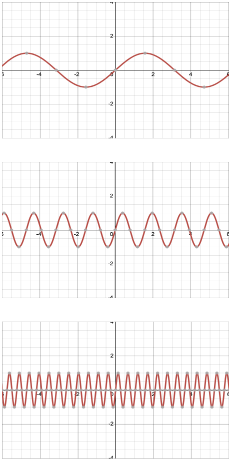
However the constants $c > 0$ depend **exponentially** on the Lipschitz constant $L > 0$.

What makes it difficult to simulate an SDE?

Given this, one expects that the L^p error of the Euler scheme would explode for the SDE

$$dX_t = \sin(\alpha X_t)dt + dW_t, \tag{10}$$

as $\alpha \to \infty$.



However, recent work has shown that this is **not true at all**.

Theorem (Dareiotis, Gerencsér, Lê 2022, Theorem 1.2)

Consider the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{11}$$

on $\mathbb{R}^d$, with diffusion σ that is bounded and twice differentiable with bounded derivatives $\sigma\sigma^T \succeq \lambda I_d$, and the drift b is **bounded and measurable**. *Then for every $\delta > 0$ and $p > 0$ there exists $c > 0$ such that the Euler scheme X^n satisfies*

$$(E \sup_{t\in[0,T]} |X_t^n - X_t|^p)^{1/p} \leq cn^{\delta-1/2}. \tag{12}$$

Furthermore the constant $c = c(d, \sigma, \sup_{x\in\mathbb{R}^d} |b(x)|)$.

This result therefore is **entirely independent of the regularity of** b. In particular, this means the Euler scheme converges even for (very) discontinuous drift coefficients b. One could take for instance

$$b(x) := sign(\sin(\alpha x)), \tag{13}$$

for $\alpha > 0$ very large.

Regularisation by Noise

The key thing here is that the noise has a **regularising effect**. To this end, for b **measurable and bounded**, lets look at how one bounds

$$E[b(W_t) - b(W_{m/n})] \tag{14}$$

where W_t is a Wiener martingale and $t \in [m/n, (m+1)/n]$

- let $p_s(x) = \frac{1}{\sqrt{2\pi s}} e^{\frac{-x^2}{2s}}$ be the density of W_s. Then

$$Eb(W_s) = \int_{\mathbb{R}^d} p_s(x) b(x) dx. \tag{15}$$

- since $p_s(x)$ is differentiable with respect to s, one can easily show

$$\partial_s Eb(W_s) \leq cs^{-1}. \tag{16}$$

- then using the fundamental theorem of calculus one has

$$E(b(W_t) - b(W_{m/n})) = \int_{m/n}^{t} \partial_s Eb(W_s) \leq cn^{-1}(m/n)^{-1}. \tag{17}$$

The study of numerical methods under low regularity assumptions is a very active area of research. In particular we highlight the following additional results

- the result presented prior can be sharpened under some conditions when b is slightly more regular (but possibly still discontinuous), see (Dareiotis, Gerencsér, Lê 2022)
- the Euler scheme converges at rate 1/2 in L^p when the drift coefficient is 'piecewise regular', and otherwise Lipschitz, see (Müller-Gronbach, Yaroslavtseva 2018). This result uses slightly different techniques but allows for growth.
- for discontinuous coefficients no numerical method for SDEs can converge better than rate 3/4 in L^p, see (Müller-Gronbach, Yaroslavtseva 2020) and (Ellinger 2024)

Counterpoint: Bad Convergence

For certain SDEs the Euler scheme (and other methods) can converge **extremely badly**

- in general, for $X^n(W_{t_1}, ..., W_{t_n})$ approximating SDE solution X_t one has

$$E|X^n(W_{t_1}, ..., W_{t_n}) - X_t| \geq cn^{-1}. \tag{18}$$

  and for general non-constant diffusion coefficient (Thomas Muller-Gronbach 2002)

$$E|X^n(W_{t_1}, ..., W_{t_n}) - X_t| \geq cn^{-1/2}. \tag{19}$$

- for $d \geq 2$ one can construct an SDE with bounded coefficients such that for every $\delta > 0$ (Jentzen, Müller-Gronbach, Yaroslavtseva 2015)

$$E|X^n(W_{t_1}, ..., W_{t_n}) - X_t| \geq cn^{-\delta}. \tag{20}$$

- the Cox-Ingersoll-Ross Process

$$dX_t = a - bX_t dt + r\sqrt{X_t} dW_t, \tag{21}$$

  cannot be approximated by any method at rate better than $2a/r^2$ (Hefter, Jentzen 2019)

Relevance to Sampling Algorithms: Proximal Methods

Say one wishes to sample from a measure π on $\mathbb{R}^d$. For many sampling algorithms, theoretical bounds depend on the **regularity of the Lebesgue density** of π, often given as

$$\pi \sim e^{-U}, \tag{22}$$

for some $U : \mathbb{R}^d \to \mathbb{R}$. Theoretical bounds for sampling from U often depend on the **Lipschitz constant of ∇U**. What if ∇U is **irregular**? One may use **proximal methods**. One targets a new density given as

$$\pi_\lambda \sim e^{-U_\lambda}, \tag{23}$$

such that

- π_λ is close to π for λ small
- ∇U_λ is regular (has small Lipschitz constant) for λ large

However calculating $U_\lambda(x)$ for any $x \in \mathbb{R}^d$ involves solving an optimisation problem.

- using proximal methods one recovers **theoretical convergence** of the sampling algorithm
- however one also incurs additional **error** and **computational cost**.
- using ideas from numerical results discussed earlier, we prove theoretical convergence of a method with irregular ∇U **without proximal methods**
- the technical challenge here is that we wish to show **uniform in time bounds**

Convergence of ULA for Discontinuous Gradients

We consider the Unadjusted Langevin Algorithm (ULA), also known as the stochastic Langevin algorithm, with stepsize $\gamma > 0$, i.e.

$$x_{n+1} = x_n - \gamma \nabla U(x_n) + \sqrt{\gamma} z_{n+1}. \tag{24}$$

This is the **Euler scheme discretisation** of a certain SDE that **converges in law to the target π**

Theorem (Johnston, Sabanis, 2024)

Suppose $U : \mathbb{R}^d \to \mathbb{R}$ is μ-strongly convex, differentiable almost everywhere, and ∇U obeys a linear growth bound. Then one has

$$W_p(\pi_\beta, \mathcal{L}(X_t^\gamma)) \leq W_p(\xi, \pi_\beta) e^{-\mu t} + c d^{1/2} \gamma^{1/4}. \tag{25}$$

Furthermore, if ∇U is discontinuous only on C^3 compact hypersurfaces, and Lipschitz otherwise one has

$$W_p(\pi_\beta, \mathcal{L}(X_t^\gamma)) \leq W_p(\xi, \pi_\beta) e^{-\mu t} + c d^{3/2} \gamma^{1/2}. \tag{26}$$

This therefore shows that one does not **neccesarily** have to smooth bad gradients.

To recap:

- the performance of numerical methods for SDE does not **necessarily** depend on the regularity (i.e. Lipschitz constant) of coefficients
- in particular a wide range of positive and negative results have been obtained for numerical methods for **discontinuous** coefficients
- these new methodologies can be applied to the performance of **sampling algorithms**
- many **open problems** in numerics and algorithms - when is convergence retained for 'bad' gradients? When is **smoothing** (i.e. the use of proximal methods) necessary? What about methods not based on the Euler scheme?

Thank You!