Adaptive timestepping for SDEs with non-globally Lipschitz drift

Mike Giles Wei Fang

Mathematical Institute, University of Oxford

Workshop on Numerical Schemes for SDEs and SPDEs
Université Lille

June 1-3, 2016

Outline

- multilevel Monte Carlo and standard SDE analysis
- motivating application: long-chain molecules
- SDEs with non-global Lipschitz drifts
- finite time analysis
  - bounded moments
  - strong error analysis
- infinite time analysis
  - bounded moments
  - strong error analysis
- conclusions

Multilevel Monte Carlo

MLMC is based on the telescoping sum

$$\mathbb{E}[\widehat{P}_L] = \mathbb{E}[\widehat{P}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$$

where $\widehat{P}_\ell$ represents an approximation of some output P on level ℓ.

In SDE applications with uniform timestep $h_\ell = 2^{-\ell}\, h_0$, if the weak convergence is

$$\mathbb{E}[\widehat{P}_\ell - P] = O(2^{-\alpha\,\ell}),$$

and $\widehat{Y}_\ell$ is an unbiased estimator for $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$, based on N_ℓ samples, with variance

$$\mathbb{V}[\widehat{Y}_\ell] = O(N_\ell^{-1}\, 2^{-\beta\,\ell}),$$

and expected cost

$$\mathbb{E}[C_\ell] = O(N_\ell\; 2^{\gamma\,\ell}), \quad \ldots$$

Multilevel Monte Carlo

... then the finest level L and the number of samples N_ℓ on each level can be chosen to achieve an RMS error of ε at an expected cost

$$C = \begin{cases} O\left(\varepsilon^{-2}\right), & \beta > \gamma, \\ O\left(\varepsilon^{-2}(\log \varepsilon)^2\right), & \beta = \gamma, \\ O\left(\varepsilon^{-2-(\gamma-\beta)/\alpha}\right), & 0 < \beta < \gamma. \end{cases}$$

Multilevel Monte Carlo

The standard estimator for SDE applications is

$$\widehat{Y}_\ell = N_\ell^{-1} \sum_{n=0}^{N_\ell} \left(\widehat{P}_\ell(W^{(n)}) - \widehat{P}_{\ell-1}(W^{(n)}) \right)$$

using the same Brownian motion $W^{(n)}$ for the n^{th} sample on the fine and coarse levels.

Uniform timestepping is not required – it is fairly straightforward to implement MLMC using non-nested adaptive timestepping.

(G, Lester, Whittle: MCQMC14 proceedings)

Standard SDE analysis

Given the SDE

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W_t$$

the "standard assumptions" are that f and g are both globally Lipschitz:

$$\exists L: \quad \|f(u)-f(v)\| + \|g(u)-g(v)\| \leq L\,\|u-v\|$$

Under these conditions, the SDE is well-posed, has finite moments for all time, and the Euler-Maruyama method

$$\widehat{X}_{t_{n+1}} = \widehat{X}_{t_n} + f(\widehat{X}_{t_n})\,h + g(\widehat{X}_{t_n})\,\Delta W_n$$

has $O(h^{1/2})$ strong convergence, using an appropriate interpolant:

$$\left(\mathbb{E}\left[\sup_{[0,T]} \|\widehat{X}_t - X_t\|^2\right]\right)^{1/2} \leq c\,h^{1/2}$$

Standard SDE analysis

If the scalar output P is a Lipschitz function of the path X_t, then

$$\mathbb{V}[\widehat{P}-P] \;\leq\; \mathbb{E}[(\widehat{P}-P)^2] \;\leq\; L^2\,\mathbb{E}\left[\sup_{[0,T]} \|\widehat{X}_t-X_t\|^2\right] \;\leq\; c^2\,L^2\,h$$

A triangle inequality for the standard deviation then gives

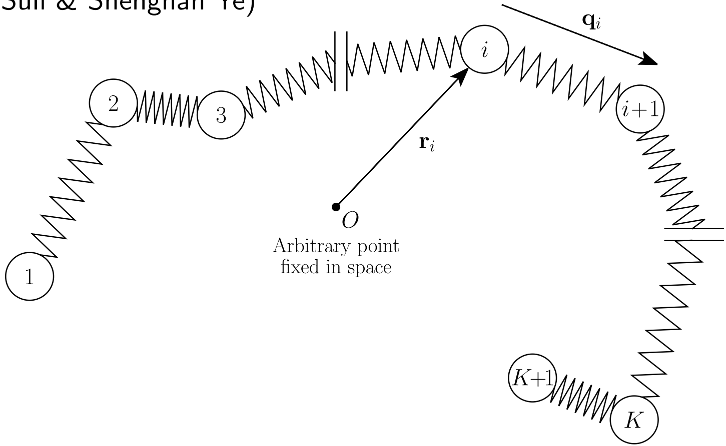
$$\mathbb{V}[\widehat{P}_\ell-\widehat{P}_{\ell-1}] \;\leq\; 4\,c^2 L^2\,h_{\ell-1}$$

and so we get $\beta\!=\!1, \gamma\!=\!1$ in the MLMC theorem.

However, what happens if the “standard assumptions” are not satisfied?

Motivation: long-chain molecules in a fluid

(Endre Süli & Shenghan Ye)



- modelled as ball-and-spring systems, subject to random forcing
- K bonds, $K+1$ "balls", separation q_i will be key variable

Motivation: long-chain molecules in a fluid

- motion of "balls" given by force balance:

$$\text{elastic force} + \text{random force} + \text{viscous drag} \;=\; 0$$

$$-\,\nabla V \;\;+\;\; R \;\;-\;\; k\,(\dot{r}_i - v(r_i)) \;\;=\;\; 0$$

  where

$$V(r) = \sum_{i=1}^{K} U_i(\|q_i\|^2/2)$$

  is the elastic potential, and v is the velocity of the fluid

- shifting to a moving frame of reference, a local Taylor series expansion gives

$$v(x) \approx \kappa\; x$$

  where κ is the local rate-of-strain tensor $\partial v/\partial x$

Motivation: long-chain molecules in a fluid

This leads to a Langevin system of coupled SDEs

$$\mathrm{d}q_i = \left(\kappa\, q_i + U'_{i+1} q_{i+1} - 2\, U'_i q_i + U'_{i-1} q_{i-1}\right)\, \mathrm{d}t + \sqrt{2}\, (\mathrm{d}W_{i+1} - \mathrm{d}W_i)$$

which can be written collectively as

$$\mathrm{d}q = (K\, q - D\, \nabla V)\, \mathrm{d}t + \sqrt{2}\; L\, \mathrm{d}W$$

where $V(q) \equiv \sum_i U_i(\|q_i\|^2/2)$, and K, L and D are of the form

$$K = \begin{pmatrix} \kappa & & \\ & \kappa & \\ & & \kappa \end{pmatrix}, \qquad L = \begin{pmatrix} -I & I & & \\ & -I & I & \\ & & -I & I \end{pmatrix},$$

$$D = \begin{pmatrix} 2I & -I & \\ -I & 2I & -I \\ & -I & 2I \end{pmatrix} = L\, L^T.$$

Motivation: long-chain molecules in a fluid

Modelling problem: with the standard quadratic potential $U_i = \beta \left\| q_i \right\|^2$, large κ leads to $\|q_i\| \to \infty$, as $t \to \infty$.

To avoid this, use stiffening potentials such as the FENE (Finitely Extensible Nonlinear Elastic) model

$$U_i(s) = -\beta \log(1 - \|q_i\|^2).$$

Numerical approximation of this naturally uses adaptive timestepping to try to avoid crossing $\|q_i\| = 1$.

Could also use potentials such as

$$U_i(s) = \beta \left\| q_i \right\|^2 + \gamma \left\| q_i \right\|^4$$

– key point is that ∇U_i is not globally Lipschitz.

Simple example

Cubic drift:

$$\mathrm{d}X_t = -\,X_t^3\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$$

Euler-Maruyama approximation with timestep h:

$$\widehat{X}_{t_{n+1}} = \widehat{X}_{t_n} - \widehat{X}_{t_n}^3\,h + \sigma\,\Delta W_n$$

If $\sigma = 0$, ODE solution converges monotonically – needs $h \leq \widehat{X}_{t_n}^{\;-2}$ for similar monotonic behaviour for approximation, and get wild oscillatory growth if $h > 2\,\widehat{X}_{t_n}^{\;-2}$.

If $\sigma > 0$, $\widehat{X}_{t_1}$ can take any value – always a small probability of strongly nonlinear blow-up. Hence,

$$\mathbb{E}[\,|\widehat{X}_t|^p] \to \infty, \quad \text{as} \quad h \to 0$$

even though $\mathbb{E}[\,|X_t|^p]$ is finite for all p.

Assumptions

Generic SDE:

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W_t$$

We are interested in two situations:

- finite time interval $[0, T]$
  - want to establish stability and strong convergence
- infinite time interval $[0, \infty)$
  - again want to establish stability and strong convergence
  - interested in applications in which there is convergence to an invariant measure, and we want expectations with respect to it

In both cases, we will always assume locally Lipschitz, differentiable f, and globally Lipschitz $g \implies g$ also satisfies linear growth bound

Assumptions

For the finite time interval, will also assume

- one-sided linear growth condition:

$$\langle x, f(x)\rangle \;\leq\; \alpha \, \|x\|^2 + \beta \quad \text{for some } \alpha, \beta > 0, \text{ all } x$$

$$\Longrightarrow \quad \text{finite } \mathbb{E}[\, \|X_t\|^p] \text{ for all } p \geq 2.$$

- global one-sided Lipschitz condition:

$$\langle x-y, f(x)-f(y)\rangle \leq L \, \|x-y\|^2 \quad \text{for some } L > 0, \text{ all } x, y$$

$$\Longleftrightarrow \quad \langle e, e\cdot\nabla f(x)\rangle \leq L \, \|e\|^2 \quad \text{for some } L > 0, \text{ all } e, x$$

- polynomially-bounded derivative:

$$\|\nabla f(x)\| \leq \gamma \, \|x\|^q + \mu \quad \text{for some } \gamma, \mu, q > 0, \text{ all } x$$

Last two needed for strong convergence analysis.

Existing literature

- D.J. Higham, X. Mao, and A.M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. SINUM, 2002.
  - one-sided Lipschitz assumption
  - implicit Euler methods such as

$$\widehat{X}_{t_{n+1}} = \widehat{X}_{t_n} + f(\widehat{X}_{t_{n+1}})\, h + g(\widehat{X}_{t_n})\, \Delta W_n$$

  - emphasises importance of stability – strong convergence then follows

- M. Hutzenthaler, A. Jentzen and P. Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. AAP, 2012.
  - one-sided Lipschitz, polynomially-bounded derivative
  - "tamed" explicit Euler method:

$$\widehat{X}_{t_{n+1}} = \widehat{X}_{t_n} + \frac{f(\widehat{X}_{t_n})\, h}{1 + \|f(\widehat{X}_{t_n})\|\, h} + g(\widehat{X}_{t_n})\, \Delta W_n$$

New analysis – finite time interval

We start by reviewing SDE stability:

$$\begin{aligned} \mathrm{d}X_t &= f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W_t \\ \Longrightarrow \quad \mathrm{d}\left(\tfrac{1}{2}\|X_t\|^2\right) &= \left(\langle X_t, f(X_t)\rangle + \tfrac{1}{2}\|g(X_t)\|^2\right)\mathrm{d}t + \langle X_t, g(X_t)\rangle\,\mathrm{d}W_t \end{aligned}$$

Hence

$$\mathrm{d}\left(\tfrac{1}{2}\mathbb{E}[\,\|X_t\|^2]\right) \leq \left(\alpha\,\mathbb{E}[\,\|X_t\|^2] + \beta\right)\mathrm{d}t$$

and therefore Grönwall's inequality gives finite $\mathbb{E}[\,\|X_t\|^2]$ for any t.

The stability analysis for the numerical approximation $\widehat{X}_t$ follows a similar approach, aiming towards the use of Grönwall's inequality, and along the way using the Burkholder-Davis-Gundy (BDG) inequality.

New analysis – finite time interval

Theorem (stability)

If the SDE satisfies the finite-time assumptions, and the continuous adaptive timestep function $h(x)$ satisfies the constraints

$$\begin{aligned} \langle x, f(x)\rangle + \tfrac{1}{2}\, h(x)\, \|f(x)\|^2 &\leq \alpha\, \|x\|^2 + \beta \\ h(x) &\geq (\xi \|x\|^q + \zeta)^{-1} \end{aligned}$$

for some $\alpha, \beta, \xi, \zeta, q > 0$, then for all finite $T > 0$, and all $p \geq 2$,

$$\mathbb{E}\left[\sup_{[0,T]} \|\widehat{X}_t\|^p \right] < \infty, \qquad \mathbb{E}\left[\, n_T^p \,\right] < \infty$$

Two simple examples:

- scalar, $f(x) = -x^3$: can use $h(x) = 2\, \max(1, |x|) \,/\, \max(1, |f\,|)$
- vector, $\langle x, f(x)\rangle = 0$: can use $h(x) = 2\,\alpha\, \max(1, \|x\|^2) \,/\, \max(1, \|f\,\|^2)$

New analysis – finite time interval

Timestep based on current state: $t_{n+1} = t_n + h(\widehat{X}_{t_n})$

Convenient to define $\underline{t} = \sup_n\{t_n : t_n \leq t\}$, $n_t = \sup_n\{n : t_n \leq t\}$

Standard Euler-Maruyama algorithm:

$$\widehat{X}_t = \widehat{X}_0 + \int_0^t f(\widehat{X}_{\underline{s}})\, \mathrm{d}s + \int_0^t g(\widehat{X}_{\underline{s}})\, \mathrm{d}W_s$$

K-truncated Euler-Maruyama algorithm:

$$\widehat{X}_t^K = P_K\left(\widehat{X}_0 + \int_0^t f(\widehat{X}_{\underline{s}}^K)\, \mathrm{d}s + \int_0^t g(\widehat{X}_{\underline{s}}^K)\, \mathrm{d}W_s\right)$$

where $P_K(Y) \equiv \min(1, K/\|Y\|)\, Y$ so $\|\widehat{X}_t^K\| \leq K$. This is used as a technical tool in the proof – it ends by taking $K \to \infty$.

New analysis – finite time interval

Looking at one timestep,

$$\widehat{X}_{t_{n+1}} = \widehat{X}_{t_n} + f(\widehat{X}_{t_n})\, h_n + g(\widehat{X}_{t_n})\, \Delta W_n$$

so

$$\begin{aligned}
\|\widehat{X}_{t_{n+1}}\|^2 \;&=\; \|\widehat{X}_{t_n}\|^2 + 2\, h_n \left(\langle \widehat{X}_{t_n}, f(\widehat{X}_{t_n}) \rangle + \tfrac{1}{2} h_n \|f(\widehat{X}_{t_n})\|^2 \right) \\
&\quad + 2\, \langle (\widehat{X}_{t_n} + f(\widehat{X}_{t_n})\, h_n), g(\widehat{X}_{t_n})\, \Delta W_n \rangle + \|g(\widehat{X}_{t_n})\, \Delta W_n\|^2 \\
&\leq\; \|\widehat{X}_{t_n}\|^2 + 2\, h_n (\alpha \|\widehat{X}_{t_n}\|^2 + \beta) + \left(\alpha \|\widehat{X}_{t_n}\|^2 + \beta \right)\, \|\Delta W_n\|^2 \\
&\quad + 2\, \langle \widehat{X}_{t_n} + f(\widehat{X}_{t_n})\, h_n, g(\widehat{X}_{t_n})\, \Delta W_n \rangle
\end{aligned}$$

and hence

$$\begin{aligned}
\|\widehat{X}_{t_n}\|^2 \;\leq\; & \|\widehat{X}_0\|^2 + \int_0^{t_n} 2 \left(\alpha \|\widehat{X}_{\underline{t}}\|^2 + \beta \right) \mathrm{d}t \; + \sum_{m<n} \left(\alpha \|\widehat{X}_{t_m}\|^2 + \beta \right)\, \|\Delta W_m\|^2 \\
& + 2 \int_0^{t_n} \langle \widehat{X}_{\underline{t}} + f(\widehat{X}_{\underline{t}})\, h(\widehat{X}_{\underline{t}}), g(\widehat{X}_{\underline{t}})\, \mathrm{d}W_t \rangle
\end{aligned}$$

New analysis – finite time interval

$$
\begin{aligned}
\|\widehat{X}_t\|^2 \;\; & \leq \;\; \|\widehat{X}_0\|^2 + \int_0^t 2\left(\alpha\|\widehat{X}_{\underline{s}}\|^2 + \beta\right) \mathrm{d}s \\
& + \;\; \sum_{m<n_t} \left(\alpha\|\widehat{X}_{t_m}\|^2 + \beta\right) \|\Delta W_m\|^2 + \left(\alpha\|\widehat{X}_{\underline{t}}\|^2 + \beta\right) \|W_t - W_{\underline{t}}\|^2 \\
& + \;\; 2\int_0^t \langle \widehat{X}_{\underline{s}} + f(\widehat{X}_{\underline{s}}) \min\left\{h(\widehat{X}_{\underline{s}}), t - \underline{s}\right\}, g(\widehat{X}_{\underline{s}})\, \mathrm{d}W_s \rangle
\end{aligned}
$$

Can raise to the power $p/2$, use Jensen inequality, take sup over $[0, t]$, then take expectation, and use BDG inequality. Eventually leads to

$$
\mathbb{E}[\widehat{S}_t^p] \leq \|\widehat{X}_0\|^p + \int_0^t (c_1\, \mathbb{E}[\widehat{S}_s^p] + c_2)\, \mathrm{d}s
$$

where $\widehat{S}_t = \sup_{[0,t]} \|\widehat{X}_s\|$. Then Grönwall inequality gives the result.

New analysis – finite time interval

Theorem (strong convergence)

If the SDE satisfies the finite-time assumptions, and the adaptive timestep is

$$h^{\delta}(x) = \delta\, h(x), \quad 0 < \delta < 1$$

where $h(x)$ satisfies the constraints of the previous theorem, then

$$\mathbb{E}\left[\sup_{[0,T]} \|\widehat{X}_t - X_t\|^p\right] = O(\delta^{p/2}), \qquad \mathbb{E}\left[\, n_T^p \right] = O(\delta^{-p}).$$

Comments:

- this is equivalent to the standard $O(h^{1/2})$ strong convergence.
- proof is relatively straightforward, given stability result
- if g is identity matrix, the strong error is first order

Infinite time interval – assumptions

For the infinite time interval will additionally assume

- g is globally bounded
- dissipative condition:

$$\langle x, f(x) \rangle \leq -\alpha \, \|x\|^2 + \beta \quad \text{for some } \alpha, \beta > 0, \text{ all } x$$

  This ensures convergence to an invariant distribution, so for all $p \geq 2$ $\mathbb{E}[\, \|X_t\|^p]$ is uniformly bounded in t.

- contraction property: for some $\alpha > 0, p \geq 2$, and any x, y, e,

$$\langle x-y, f(x)-f(y) \rangle + \tfrac{1}{2} p\,(p-1)\, \|g(x)-g(y)\|^2 \leq -\alpha \, \|x-y\|^2$$

$$\Leftrightarrow \;\; \langle e, e\cdot\nabla f(x) \rangle + \tfrac{1}{2} p\,(p-1)\, \|e\cdot\nabla g(x)\|^2 \leq -\alpha \, \|e\|^2 \quad \text{for some } \alpha > 0$$

  This ensures that $X_t^{(2)} - X_t^{(1)} \rightarrow 0$ if starting from different initial data but driven by same W_t — needed for L_p strong convergence

New analysis – infinite time interval

Theorem (stability)

If the SDE satisfies the infinite-time assumptions, and the adaptive timestep satisfies the constraint

$$\langle x, f(x)\rangle + \tfrac{1}{2}\, h(x)\, \|f(x)\|^2 \;\leq\; -\alpha\, \|x\|^2 + \beta$$

for some $\alpha, \beta > 0$*, then for all* $p \geq 2$*, there exist constants* C_p, c_p *such that for all* $T > 0$

$$\mathbb{E}\left[\, \|\widehat{X}_T\|^p \right] \leq C_p, \qquad \mathbb{E}\left[\, n_T^p \right] \leq c_p\, T^p$$

- analysis is similar to before
- key change is to use $S_t = \sup_{[0,t]} \left\{ \mathrm{e}^{-\gamma(t-s)} \|X_s\| \right\}$ for a suitable γ

New analysis – infinite time interval

Theorem (strong convergence)

If the SDE satisfies the infinite-time assumptions, and the adaptive timestep is again

$$h^\delta(x) = \delta\, h(x), \quad 0 < \delta < 1$$

where $h(x)$ satisfies the timestep constraints, then there exist constants C_p, c_p such that for all $T>0$

$$\mathbb{E}\left[\|\widehat{X}_T - X_T\|^p \right] \leq C_p\, \delta^{p/2}, \qquad \mathbb{E}\left[\, n_T^p \right] \leq c_p\, \delta^{-p}\, T^p.$$

Comments:

- this is again equivalent to the standard $O(h^{1/2})$ strong convergence
- proof is a bit trickier this time, to avoid a bound which increases exponentially in time
- if g is the identity matrix, the strong error is first order

Conclusions

- Euler-Maruyama discretisation with adaptive timesteps is stable for SDEs with non-globally Lipschitz drift
- order of strong convergence same as usual, when viewed as accuracy versus cost
- works as expected within MLMC computation
- also works well for invariant distributions for SDEs with contraction property
- future challenge: ergodic SDEs without contraction property

Webpages:
`http://people.maths.ox.ac.uk/gilesm/mlmc.html`
`http://people.maths.ox.ac.uk/gilesm/mlmc_community.html`