Stochastic Runge-Kutta Accelerates Langevin Monte Carlo and Beyond

Xuechen Li [1,2]

Denny Wu [1,2] Lester Mackey [3] Murat A. Erdogdu [1,2]

[1]University of Toronto [2]Vector Institute [3]Microsoft Research

The Problem and Our Work

Given smooth potential $f : \mathbb{R}^d \to \mathbb{R}$, sample from given density

$$p(x) \propto \exp(-f(x)).$$

- We study both **strongly convex** and **non-convex** potentials.
- Many papers study individual algorithms [1, 2, 3, 4, 5]. However, there has yet to be a unifying theoretical framework.
- We provide a theorem that gives the convergence rate of sampling algorithms obtained by discretizing an *exponentially contracting diffusion* based on **local properties** of the numerical method.
- A direct extension is we obtain faster converging algorithms with the class of *stochastic Runge-Kutta* (SRK) methods.

Exponential W_2-Contraction of Diffusions

Diffusion X_t has exponential W_2-contraction if two instances $X_{t,x}, X_{t,y}$ initiated respectively from x and y satisfy

$$W_2(X_{t,x}, X_{t,y}) \le e^{-\alpha t}\|x - y\|_2, \quad \text{for all } x, y \in \mathbb{R}^d, t \ge 0.$$

Informal: The marginals of the continuous-time diffusion become the same very quickly regardless of the initial state.

Example: When f is strongly convex, the Langevin diffusion characterized by the SDE

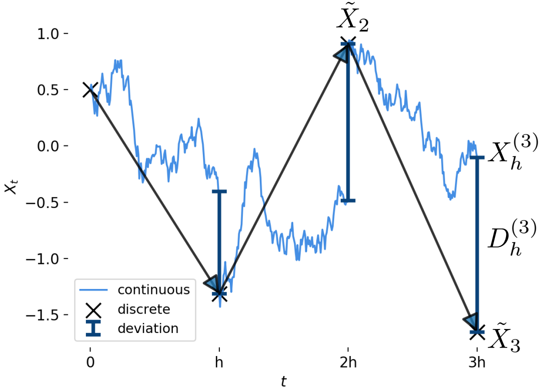
$$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$$

has exponential W_2-contraction.

Local Deviation

Let $\{\tilde{X}_k\}_{k\in\mathbb{N}}$ be a discretization of $\{X_t\}_{t\geq 0}$, and $\{X_s^{(k)}\}_{s\geq 0}$ be another instance of the diffusion starting from $\tilde{X}_{k-1}$ at $s = 0$.

The *local deviation* at iteration k is defined as $D_h^{(k)} = X_h^{(k)} - \tilde{X}_k$.



Uniform Orders of Local Deviation

Recall local deviation $D_h^{(k)} = X_h^{(k)} - \tilde{X}_k$. A numerical scheme has uniform mean-square and mean orders of (p_1, p_2) if for all $k \in \mathbb{N}$

$$\mathcal{E}_k^{(1)} = \mathbb{E}\left[\mathbb{E}\big[\|D_h^{(k)}\|_2^2 | \mathcal{F}_{t_{k-1}}\big]\right] \leq \lambda_1 h^{2p_1}, \quad (1)$$

$$\mathcal{E}_k^{(2)} = \mathbb{E}\left[\big\|\mathbb{E}\big[D_h^{(k)} | \mathcal{F}_{t_{k-1}}\big]\big\|_2^2\right] \leq \lambda_2 h^{2p_2}, \quad (2)$$

for constants λ_1 and λ_2 independent of h.

Remark: Bounds like (1) appeared explicitly in previous works (see e.g. [1]). To the best of our knowledge, (2) did not appear explicitly in previous works.

A General Theorem

Theorem (Informal)

Diffusion has a stationary distribution $p(x) \propto \exp(-f(x))$ *and exhibits exponential* W_2*-contraction. Acting on this diffusion, a numerical discretization with uniform mean-square and mean orders of* (p_1, p_2) *for* $p_2 \geq p_1 + \frac{1}{2}$ *has rate* $\tilde{\mathcal{O}}(\epsilon^{-1/(p_1-1/2)})$ *in* W_2*.*

Remark 1: Connects the numerical SDE and sampling literatures: Take any classical SDE discretization method, instantly know the convergence rate when it's used for sampling!

Remark 2: Can also be used for discretizing the underdamped Langenvin diffusion! Check out our examples in the paper.

Convergence Rates for EM and SRK

Result	Diffusion	Smoothness	Unif. Orders	Rate
EM (Durmus et al.)	Langevin	1st	$(1.0, 1.5)$	$\tilde{\mathcal{O}}(d\epsilon^{-2})$
EM (Ex. 1)	Langevin	1st & 2nd	$(1.5, 2.0)$	$\tilde{\mathcal{O}}(d\epsilon^{-1})$
SRK-LD (This work)	Langevin	1st-3rd	$(2.0, 2.5)$	$\tilde{\mathcal{O}}(d\epsilon^{-2/3})$
EM (Ex. 2)	General	1st	$(1.0, 1.5)$	$\tilde{\mathcal{O}}(d\epsilon^{-2})$
SRK-ID (This work)	General	1st	$(1.5, 2.0)$	$\tilde{\mathcal{O}}(d^{3/4}m^2\epsilon^{-1})$

Table: Convergence rates in W_2, i.e. number of iterations required to reach ϵ accuracy to the target in W_2. Top three for strongly convex f; bottom two for non-convex f that admits uniformly dissipative diffusion.

EM = Euler-Maruyama
SRK = Stochastic Runge-Kutta

Thanks to you and my coauthors:

Denny Wu

Lester Mackey

Murat A. Erdogdu

Our poster: **East Exhibition Hall B + C #162**

[1] Xiang Cheng, Niladri S Chatterji, Peter L Bartlett, and Michael I Jordan. Underdamped Langevin MCMC: A non-asymptotic analysis.

[2] Arnak S Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities.

[3] Alain Durmus, Eric Moulines, et al. Nonasymptotic convergence analysis for the unadjusted langevin algorithm.

[4] Yin Tat Lee, Zhao Song, and Santosh S Vempala. Algorithmic theory of odes and sampling from well-conditioned logconcave densities.

[5] Santosh S Vempala and Andre Wibisono. Rapid convergence of the unadjusted langevin algorithm: Log-sobolev suffices.