Statistical Error of Numerical Integrators for Underdamped Langevin Dynamics with Deterministic And Stochastic Gradients

Xuda Ye

joint work with Zhennan Zhou

arXiv:2405.06871

Peking University

August 8, 2024

Sampling in Statistical Mechanics

- Consider the classical Hamiltonian system

$$H(x,v) = \frac{|v|^2}{2} + U(x), \quad x, v \in \mathbb{R}^d,$$

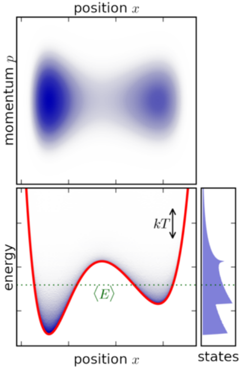
where $U(x)$ is the potential function in $\mathbb{R}^d$.

- The thermal equilibrium at the temperature $T = 1$ is descried by the Boltzmann distribution

$$\pi(x,v) = \frac{1}{Z} e^{-\frac{|v|^2}{2} - U(x)}, \quad Z = \int e^{-\frac{|v|^2}{2} - U(x)} \mathrm{d}x\mathrm{d}v,$$

where Z is the partition function.

Sampling in Statistical Mechanics



(Wikipedia: Canonical ensemble)

The figure shows the density of $\pi(x, v)$ for a double-well potential.

Sampling in Statistical Mechanics

- In computational physics, an important task is to compute the statistical average of a given test function $f(x,v)$:

$$\langle f\rangle := \int f(x,v)\pi(x,v)\mathrm{d}x\mathrm{d}v = \frac{1}{Z}\int f(x,v)e^{-\frac{|v|^2}{2}-U(x)}\mathrm{d}x\mathrm{d}v.$$

- Numerical methods for sampling the Boltzmann distribution $\pi(x,v)$ are the core strategy to compute $\langle f\rangle$.

- Suppose the numerical method produces the sample points $(X_n,V_n)_{n\geqslant 0}$, then $\langle f\rangle$ can be computed from

$$\langle f\rangle \approx \frac{1}{N}\sum_{n=0}^{N-1} f(X_n,V_n).$$

Underdamped Langevin Dynamics

- The underdamped Langevin dynamics for sampling $\pi(x, v)$ is

$$\begin{cases} \dot{x}_t = v_t, \\ \dot{v}_t = -\nabla U(x_t) - \gamma v_t + \sqrt{2\gamma}\dot{B}_t, \end{cases}$$

  where x_t and v_t are position and velocity coordinates in $\mathbb{R}^d$, $\gamma > 0$ is the damping rate, and $(B_t)_{t \geqslant 0}$ is the Brownian motion in $\mathbb{R}^d$.

- The corresponding Fokker–Planck equation is

$$\frac{\partial \rho}{\partial t} + v \cdot \nabla_x \rho - \nabla U \cdot \nabla_v \rho = \gamma \nabla_v \cdot (v\rho + \nabla_v \rho).$$

- The ergodic theory indicates

$$\langle f \rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T f(x_t, v_t) \mathrm{d}t, \quad \text{a.s.},$$

  which means the calculation of $\langle f \rangle$ is exact as long as $T \to \infty$.

Statistical Average & Time Average

- The dynamics needs to discretized by a numerical integrator to produce the numerical solution $(X_n, V_n)_{n\geqslant 0}$.
- For a numerical integrator with time step h, $\langle f \rangle$ is approximated by the time average

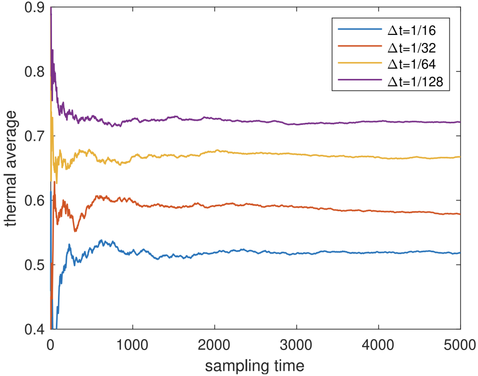
$$\langle f \rangle \approx \langle f \rangle_{N,h} := \frac{1}{N} \sum_{n=0}^{N-1} f(X_n, V_n),$$

- The accuracy of the numerical solution is characterized by

$$e_{N,h} := \langle f \rangle_{N,h} - \langle f \rangle,$$

which depends on $(X_n, V_n)_{n\geqslant 0}$ and is a random variable!

Error Estimate: Statistical Error



(–, Z. Zhou, J. Chem. Phys., 2021)

Time average of the numerical solution with different time steps. The model problem is the path integral molecular dynamics for interacting particle system.

Error Estimate: Statistical Error

- The goal of the error estimate is to quantify the sample qualities of the numerical solution $(X_n, V_n)_{n\geqslant 0}$.
- A common form of the long-time error estimate is

$$\mathcal{W}_p(\nu_n, \pi) \leqslant C_1 e^{-\lambda nh} + C_2 h^\alpha, \quad \forall n \geqslant 0,$$

  where ν_n is the distribution law of (X_n, V_n) and $\mathcal{W}_p(\cdot, \cdot)$ is the Wasserstein distance with $p = 1, 2$.
- In terms of the time average, the result above implies

$$\left|\mathbb{E}[\langle f\rangle_{N,h} - \langle f\rangle]\right| \leqslant \frac{2C_1}{\lambda Nh} + C_2 h^\alpha, \quad \forall N \geqslant 0,$$

  which characterizes the bias of the time average estimator.

Error Estimate: Statistical Error

- Compared to the bias $\mathbb{E}[\langle f\rangle_{N,h} - \langle f\rangle]$, the statistical error

$$\mathrm{SE}_{N,h} := \mathbb{E}\left[(\langle f\rangle_{N,h} - \langle f\rangle)^2\right]$$

  can capture the random fluctuation of the numerical solution.

- How to quantify the statistical error of a given numerical integrator?

Error Estimate: Statistical Error

Our main results on the estimate of the statistical error:

Theorem 1. (convex outside a ball)

Suppose the numerical integrator has strong order p. If $U(x)$ is strongly convex outside a ball, then there exists a $\gamma_0 > 0$ such that when $\gamma \geqslant \gamma_0$,

$$\mathrm{SE}_{N,h} = \mathcal{O}\left(h^{2p-1} + \frac{1}{Nh}\right).$$

Theorem 2. (globally convex)

Suppose the numerical integrator has strong order p. If $U(x)$ is globally convex in $\mathbb{R}^d$, then there exists a $\gamma_0 > 0$ such that when $\gamma \geqslant \gamma_0$,

$$\mathrm{SE}_{N,h} = \mathcal{O}\left(h^{2p} + \frac{1}{Nh}\right).$$

Stochastic Gradient Sampling in Data Science

- In data science, the potential $U(x)$ is computed from a large data set, and using the stochastic gradient reduces the sampling cost.
- Suppose the stochastic gradient is $b(x, \omega)$ with

$$\mathbb{E}^{\omega}[b(x, \omega)] = \nabla U(x).$$

A- numerical integrator with an i.i.d. sequence $(\omega_n)_{n \geqslant 0}$ produces the numerical solution $(X_n, V_n)_{n \geqslant 0}$.

Stochastic Gradient Sampling in Data Science

We introduce two specific examples of the stochastic gradients.

- Large data set. If the potential $U(x)$ is formed as

$$U(x) = \frac{1}{J}\sum_{j=1}^{J} U_j(x),$$

then $b(x,\omega)$ can be chosen as

$$b(x,\omega) = \frac{1}{p}\sum_{j\in\mathcal{C}(\omega)} \nabla U_j(x),$$

where $\mathcal{C}(\omega) \subset \{1,\cdots,J\}$ is the subset of indices.

Stochastic Gradient Sampling in Data Science

We introduce two specific examples of the stochastic gradients.

- Large particle number. If the potential $U(x)$ is formed as

$$U(x) = \sum_{i=1}^{M} V_o(x^i) + \frac{1}{M-1} \sum_{1 \leqslant i < j \leqslant M} V_i(x^i - x^j), \qquad x_1, \cdots, x_M \in \mathbb{R}^d,$$

then $b(x, \omega) = (b^i(x, \omega))_{i=1}^M$ can be chosen as

$$b^i(x, \omega) = \nabla V_o(x^i) + \frac{1}{p-1} \sum_{j \in \mathcal{C}_{\mathcal{B}(i)}} \nabla V(x^i - x^j), \quad i = 1, \cdots, M.$$

- Here, the index set $\{1, \cdots, M\}$ is randomly divided into small batches $\{\mathcal{C}_1, \cdots, \mathcal{C}_q\}$ with the batch size $p = M/q$. For each $i = 1, \cdots, M$, $\mathcal{B}(i)$ is the index of batch which contains i.
- The method above is referred to as the Random Batch Method[1].

[1] S. Jin, L. Li, and J. Liu. Journal of Computational Physics 400 (2020): 108877.

Error Estimate: Statistical Error

Our error estimates also apply to the stochastic gradient case.

Theorem 3. (globally convex + stochastic gradient)

Suppose the stochastic gradient numerical integrator has strong order p. If $U(x)$ is globally convex, then there exists a $\gamma_0 > 0$ such that when $\gamma \geqslant \gamma_0$,

$$\mathrm{SE}_{N,h} = \mathcal{O}\left(h^{\min\{2p,2\}} + \frac{1}{Nh}\right).$$

Comparison with Related Works: Global Contractivity

- The global contractivity of the numerical integrator means that

$$d(\mu_{n+1}, \nu_{n+1}) \leqslant e^{-\lambda h} d(\mu_n, \nu_n)$$

  for any distributions μ_n, ν_n in $\mathbb{R}^d \times \mathbb{R}^d$, where $d(\cdot, \cdot)$ is a distance for distributions in $\mathbb{R}^d \times \mathbb{R}^d$.

- The global contractivity is proved by a specially coupling scheme, and requires strong restriction on the time step h. Our results merely rely on the uniform-in-time moments.

Comparison with Related Works: Global Contractivity

The statistical error for full gradient numerical integrators.

integrator	order	statistical error	global convexity	explicit
gHMC[2][3]	2	$\mathcal{O}(h^4 + \frac{1}{Nh})$	Not required	Yes
UBU[4]	2	$\mathcal{O}(h^4 + \frac{1}{Nh})$	Not required	Yes
general	p	$\mathcal{O}(h^{2p-1} + \frac{1}{Nh})$	Not required	No
general	p	$\mathcal{O}(h^{2p} + \frac{1}{Nh})$	Required	No

Table 1: Comparison of our results for the full gradient integrators and those proved by the global contractivity.

[2] N. Bou-Rabee and K. Schuh. Electronic Journal of Probability 28 (2023): 1-40.
[3] X. Cheng, et al. arXiv:1805.01648 (2018).
[4] K. Schuh, and P. Whalley. arXiv:2405.09992 (2024).

Comparison with Related Works: Global Contractivity

The statistical error for stochastic gradient numerical integrators.

integrator	order	statistical error	global convexity	explicit
SG-gHMC[5][6]	2	$\mathcal{O}(h + \frac{1}{Nh})$	Not required	Yes
SG-general	p	$\mathcal{O}(h^{\min\{2p,2\}} + \frac{1}{Nh})$	Required	No

Table 2: Comparison of our results for the stochastic gradient integrators and those proved by the global contractivity.

Our results showcase better order in the time step h.

[5] N. Gouraud, et al. arXiv preprint arXiv:2202.00977 (2022).

[6] M. Chak, and P. Monmarché. arXiv preprint arXiv:2310.18774 (2023).

Comparison with Related Works: Poisson equation

- The Poisson equation for the test function $f(x, v)$ is

$$-(\mathcal{L}\phi)(x, v) = f(x, v) - \langle f \rangle, \quad x, v \in \mathbb{R}^d,$$

where $\mathcal{L}$ is the generator of the Langevin dynamics:

$$\mathcal{L} = v \cdot \nabla_x - (\nabla U(x) + \gamma v) \cdot \nabla_v + \gamma \Delta_v.$$

- The Itô's calculus implies

$$\mathrm{d}\phi(x_t, v_t) = (\mathcal{L}\phi)(x_t, v_t)\mathrm{d}t + \sqrt{2}\nabla_v \phi(x, v) \cdot \mathrm{d}B_t,$$

hence

$$\begin{aligned} &\frac{1}{T}\int_0^T f(x_t, v_t)\mathrm{d}t - \pi(f) \\ &= \frac{\phi(x_0, v_0) - \phi(x_T, v_T)}{T} + \underbrace{\frac{1}{T}\int_0^T \sqrt{2}\nabla_v \phi(x_t, v_t) \cdot \mathrm{d}B_t}_{\text{mean-zero}}. \end{aligned}$$

Comparison with Related Works: Poisson equation

- The Poisson equation approach relies on the regularity of $\phi(x, v)$, because of the need for the high-order approximation of

$$\phi(X_{n+1}, V_{n+1}) - \phi(X_n, V_n).$$

- However, the proof of regularity can be extremely difficult for the underdamped Langevin dynamics because $\mathcal{L}$ is hypoelliptic.

Proof Strategy: Discrete Poisson Equation

- We employ the discrete Poisson equation to study the statistical errors. For given test function $f(x,v)$, define the function

$$u(x,v,t) = (e^{t\mathcal{L}}f)(x,v) - \langle f \rangle = \mathbb{E}^{x,v}[f(x_t,v_t)] - \langle f \rangle,$$

where $\mathbb{E}^{x,v}$ indicates that the solution (x_t,v_t) starts with (x,v).

- The Poisson solution $\phi(x,v)$ can be interpreted as

$$\phi(x,v) = \int_0^\infty u(x,v,t)\mathrm{d}t.$$

- Given the time step $h > 0$, define the function

$$\phi_h(x,v) = h\sum_{n=0}^\infty u(x,v,nh),$$

then $\phi_h(x,v)$ satisfies the discrete Poisson equation:

$$\frac{1-e^{h\mathcal{L}}}{h}\phi_h(x,v) = f(x,v) - \langle f \rangle.$$

Proof Strategy: Discrete Poisson Equation

- The discrete Poisson solution $\phi_h(x, v)$ provides a natural expression of the time average.

- Let $Z_n = (X_n, V_n)$. Write the difference term $\phi(Z_{n+1}) - \phi(Z_n)$ as

$$\phi_h(Z_{n+1}) - \phi_h(Z_n(h)) + \phi_h(Z_n(h)) - \phi_h(Z_n),$$

  where $Z_n(h)$ is the exact solution in time h with initial state Z_n.

- Define the random variables S_n and T_n by

$$S_n = \frac{\phi_h(Z_{n+1}) - \phi_h(Z_n(h))}{h}, \quad \text{(local error)}$$
$$T_n = \frac{\phi_h(Z_n(h)) - \phi_h(Z_n)}{h} + f(Z_n) - \langle f \rangle, \quad \text{(mean-zero)}$$

- The time average $\langle f \rangle_{N,h}$ can be expressed as

$$\boxed{\langle f \rangle_{N,h} - \langle f \rangle = \frac{\phi_h(Z_0) - \phi_h(Z_N)}{Nh} + \frac{1}{N} \sum_{n=0}^{N-1} (S_n + T_n).}$$

Proof Strategy: Discrete Poisson Equation

- The random variable T_n is mean-zero because

$$\mathbb{E}[T_n|Z_n] = \mathbb{E}\left[\frac{(e^{h\mathcal{L}}\phi_h)(Z_n) - \phi_h(Z_n)}{h}\right] + f(Z_n) - \langle f\rangle.$$

  Furthermore, $\{T_n\}_{n=1}^N$ are mutually independent:

$$\mathbb{E}[T_n T_m] = 0 \text{ for } 0 \leqslant n < m \leqslant N-1.$$

- The statistical error $\mathrm{SE}_{N,h} = \mathbb{E}[(\langle f\rangle_{N,h} - \langle f\rangle)^2]$ has the estimate

$$\mathrm{SE}_{N,h} \lesssim \frac{\mathbb{E}[(\phi_h(Z_0) - \phi_h(Z_n))^2]}{N^2h^2} + \frac{1}{N^2}\underbrace{\mathbb{E}\left[\left(\sum_{n=0}^{N-1} S_n\right)^2\right]}_{\text{core task}} + \frac{1}{N^2}\sum_{n=0}^{N-1}\mathbb{E}[T_n^2]$$

Proof Strategy: Discrete Poisson Equation

- If $U(x)$ is strongly convex outside a ball, then

$$|\nabla\phi_h(x,v)| \lesssim 1 \Longrightarrow \mathbb{E}\left[\left(\sum_{n=0}^{N-1} S_n\right)^2\right] \lesssim N^2h^{2p-1}\mathbb{E}(|Z_0|+1)^{2q}.$$

And we obtain Theorem 1: $\boxed{\mathrm{SE}_{N,h} \lesssim h^{2p-1} + \frac{1}{Nh}}$.

- If $U(x)$ is globally convex outside a ball, then

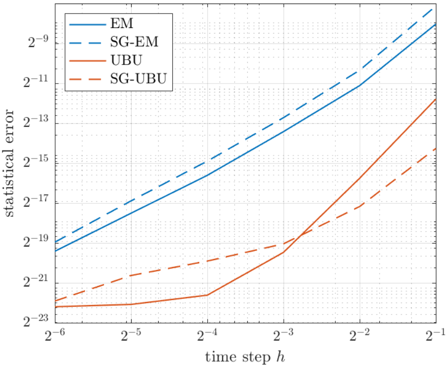
$$|\nabla\phi_h(x,v)|, |\nabla^2\phi_h(x,v)| \lesssim 1 \Longrightarrow$$

$$\mathbb{E}\left[\left(\sum_{n=0}^{N-1} S_n\right)^2\right] \lesssim (N^2h^{2p} + Nh^{2p-1})\mathbb{E}(|Z_0|+1)^{4q}.$$

And we obtain Theorem 2: $\boxed{\mathrm{SE}_{N,h} \lesssim h^{2p} + \frac{1}{Nh}}$.

Numerical Verification

We plot the statistical error of Euler–Maruyama (EM) and UBU integrators and their stochastic gradient versions.



The statistical error of UBU is $\mathcal{O}(h^4 + \frac{1}{Nh})$, while the others are $\mathcal{O}(h^2 + \frac{1}{Nh})$.

Summary

Advantages of our results:

- Applicable to a broad class of numerical integrators.
- Require no explicit restriction on the time step h except for the uniform-in-time moments condition.

Drawbacks of our results:

- The constants are not explicit in the dimension d or the batch size p (stochastic gradient case).
- Requires the potential function $U(x)$ to be globally convex (essentially difficult!)

Future works:

- Make the constants explicit on d and p.
- Statistical error of Stochastic Variance Reduced Gradient (SVRG) type integrators.