Convergence of Unadjusted Langevin in High Dimensions

Delocalization of Bias

Yifan Chen

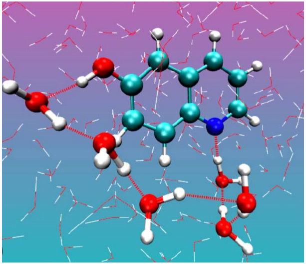
Courant Institute, New York University

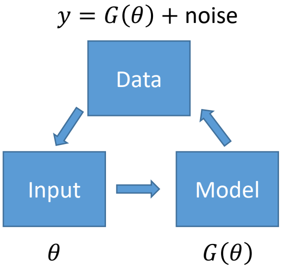
joint work with Xiaoou Cheng, Jonathan Niles-Weed, Jonathan Weare

Context

Classical sampling problem

Goal: draw (approximate) samples from $\pi \propto \exp(-V)$





Applications in molecular dynamics, Bayes inverse problems, ...

- In molecular dynamics: V is the inter-atomic potential
- In Bayes inverse problem: π is posterior distribution

Challenges: High dimensional probability distributions

MCMC algorithm with Langevin's dynamics

Overdamped Langevin's dynamics

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$$

Under mild assumptions, as $t \to \infty$, Law$(X_t) \to \pi \propto \exp(-V)$

MCMC algorithm with Langevin's dynamics

Overdamped Langevin's dynamics

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$$

Under mild assumptions, as $t \to \infty$, Law$(X_t) \to \pi \propto \exp(-V)$

- **Unadjusted Langevin:** Euler–Maruyama scheme

$$X_{(k+1)h} = X_{kh} - h\nabla V(X_{kh}) + \sqrt{2}(W_{(k+1)h} - W_{kh})$$

As $k \to \infty$, Law$(X_{kh}) \to \pi_h$ where hopefully $\pi_h \approx \pi$ (bias)

MCMC algorithm with Langevin's dynamics

Overdamped Langevin's dynamics

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$$

Under mild assumptions, as $t \to \infty$, Law$(X_t) \to \pi \propto \exp(-V)$

- **Unadjusted Langevin:** Euler–Maruyama scheme

$$X_{(k+1)h} = X_{kh} - h\nabla V(X_{kh}) + \sqrt{2}(W_{(k+1)h} - W_{kh})$$

  As $k \to \infty$, Law$(X_{kh}) \to \pi_h$ where hopefully $\pi_h \approx \pi$ (bias)

- How large is the bias? For $V \in C^2$ with $\alpha I \preceq \nabla^2 V \preceq \beta I$:

$$W_2(\pi, \pi_h) = O(\frac{\beta}{\alpha}\sqrt{dh})$$

  [Durmus, Moulines, 2019], etc.

- Implication: $h \sim 1/d$ for bounded bias in any dimension

  Can be improved to $h \sim 1/d^{1/2}$ with more assumptions [Li, Zha, Tao 2022]

Bias can be completely eliminated

Metropolis-adjusted Langevin: accept $X_{(k+1)h}$ w/ probability

$$p_{\text{accept}} = \min\left\{1, \frac{\pi(X_{(k+1)h})q(X_{kh}|X_{(k+1)h})}{\pi(X_{kh})q(X_{(k+1)h}|X_{kh})}\right\}$$

where q is the transition kernel of unadjusted Langevin; otherwise reject and $X_{(k+1)h} = X_{kh}$. There will be no bias

[Rossky, Doll, Friedman 1978], [Roberts, Tweedie 1997]

Bias can be completely eliminated

Metropolis-adjusted Langevin: accept $X_{(k+1)h}$ w/ probability

$$p_{\text{accept}} = \min\left\{1, \frac{\pi(X_{(k+1)h})q(X_{kh}|X_{(k+1)h})}{\pi(X_{kh})q(X_{(k+1)h}|X_{kh})}\right\}$$

where q is the transition kernel of unadjusted Langevin; otherwise reject and $X_{(k+1)h} = X_{kh}$. There will be no bias

[Rossky, Doll, Friedman 1978], [Roberts, Tweedie 1997]

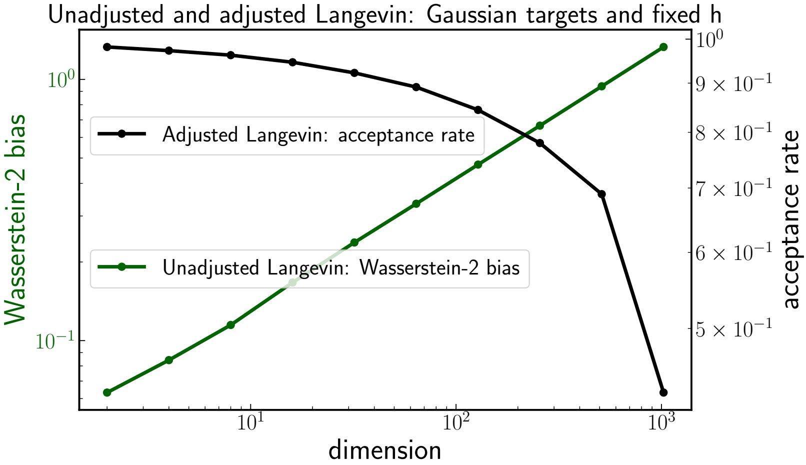
However, for this algorithm, h must be small when d is large

- Existing theory suggests $h \sim 1/d^{1/3}, 1/d^{1/2}, 1/d$ depending on notion of convergence and distribution of X_0

  [Roberts, Rosenthal 1998], [Christensen, Roberts, Rosenthal 2005], [Dwivedi, Chen, Wainwright, Yu 2018], [Chewi, Lu, Ahn, Cheng, Gouic, Rigollet 2021], etc

- This is necessary for non-negligible acceptance rates

Performance illustration: for fixed stepsize h



- Fixed h will fail when d increases
- Is this a full story?

A Closer Look at Existing Theoretical Results in High Dimensions

For MALA: h needs to be small for high acceptance rates

- Theories in the literature suggest $h \sim 1/d^{1/3}$ or $1/\sqrt{d}$ or $1/d$
  [Roberts, Rosenthal 1998], [Christensen, Roberts, Rosenthal 2005], [Dwivedi, Chen, Wainwright, Yu 2018], [Chewi, Lu, Ahn, Cheng, Gouic, Rigollet 2021], etc
- This scaling is not avoidable in general

For unadjusted Langevin: h needs to be small for small bias

- Theories in the literature suggest $h \sim 1/\sqrt{d}$ or $h \sim 1/d$
  [Durmus, Moulines, 2019], [Li, Zha, Tao 2022], etc.
- Such bias is measured in the W_2 distance or other divergence

A Closer Look at Existing Theoretical Results in High Dimensions

For MALA: h needs to be small for high acceptance rates

- Theories in the literature suggest $h \sim 1/d^{1/3}$ or $1/\sqrt{d}$ or $1/d$

  [Roberts, Rosenthal 1998], [Christensen, Roberts, Rosenthal 2005], [Dwivedi, Chen, Wainwright, Yu 2018], [Chewi, Lu, Ahn, Cheng, Gouic, Rigollet 2021], etc
- This scaling is not avoidable in general

For unadjusted Langevin: h needs to be small for small bias

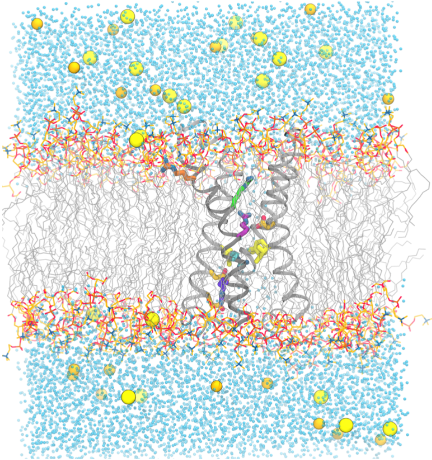
- Theories in the literature suggest $h \sim 1/\sqrt{d}$ or $h \sim 1/d$

  [Durmus, Moulines, 2019], [Li, Zha, Tao 2022], etc.
- Such bias is measured in the W_2 distance or other divergence

Fact: In unadjusted Langevin, $h = O(1)$ could suffice for certain averaged observables, e.g. $f(x) = \frac{1}{d}\sum_{i=1}^{d} x^{(i)}$, which satisfies $|\nabla f(x)|_2 \leq |x|_2/\sqrt{d}$ [Bou-Rabee, Schuh 2023], [Durmus, Eberle 2024]

Which Observables Will Be of Interest?

Often high dimensionality occurs when many nuisance variables are required to accurately describe the remaining variables' distribution



[Thanks to Spencer Guo]

Molecular dynamics (MD) example

- We care about averages with respect to a few atoms in the voltage sensing protein in the middle
- We do not care about averages with respect to atoms in the lipid or water molecules
- We need all the atoms to accurately describe the system

We are interested in *a small part!*

Disclaimer: the potential V in MD is more complex than considered in our analysis

This Work: Accuracy for Low Dimensional Marginal Distributions

For $\pi(x) = \pi(x^{(1)}, ..., x^{(d)})$, a K-dimensional marginal distribution is obtained by marginalizing over the remaining $d - K$ coordinates

Theorem [Chen, Cheng, Niles-Weed, Weare 2024]

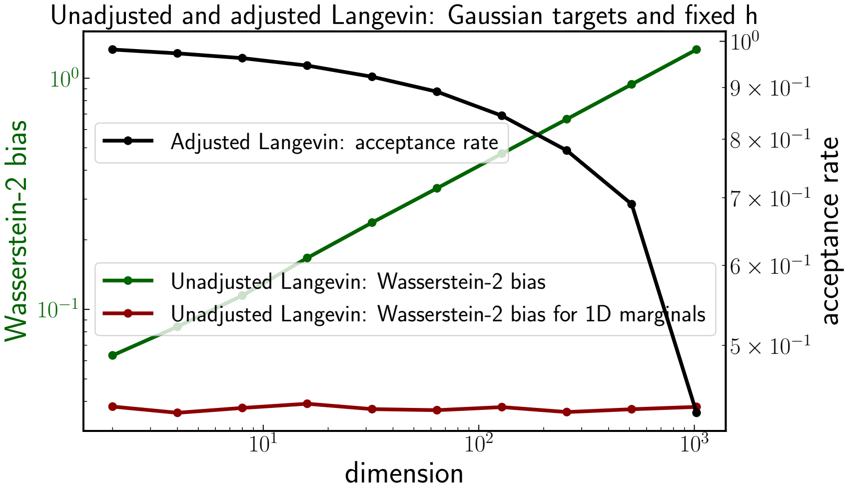
(informal) for unadjusted Langevin in d dimensions, $h = O(1/K)$ could suffice for bounded bias in all K dimensional marginals

- Rigorous results proved under the assumption $\alpha I \preceq \nabla^2 V \preceq \beta I$ and V is Gaussian/"sparse" (and some generalizations)
- Iteration complexity is $O(K)$, nearly independent of d

($\log d$ terms omitted)

Bias in low dimensional marginals can behave much better than in full distribution!

Updated Figure: If Interested in A Small Number of Coordinates



- Same for K-marginals, if K is independent of dimension
  (under the assumption of our theorem)
- Unadjusted approaches can be more scalable than adjusted

Roadmap of this Talk

Roadmap of this Talk

New Metric for Low Dimensional Marginals

Standard W_p metric: ℓ^2 measures full coordinates

$$W_p(\mu,\nu) = \left(\min_{\gamma\in\Pi(\mu,\nu)} \int |x-y|_2^p \, \gamma(\mathrm{d}x,\mathrm{d}y)\right)^{1/p}$$

New W_{p,ℓ^∞} metric: replace ℓ^2 by ℓ^∞

$$W_{p,\ell^\infty}(\mu,\nu) = \left(\min_{\gamma\in\Pi(\mu,\nu)} \int |x-y|_\infty^p \gamma(\mathrm{d}x,\mathrm{d}y)\right)^{1/p}$$

New Metric for Low Dimensional Marginals

Standard W_p metric: ℓ^2 measures full coordinates

$$W_p(\mu,\nu) = \left(\min_{\gamma \in \Pi(\mu,\nu)} \int |x-y|_2^p \, \gamma(\mathrm{d}x, \mathrm{d}y) \right)^{1/p}$$

New W_{p,ℓ^∞} metric: replace ℓ^2 by ℓ^∞

$$W_{p,\ell^\infty}(\mu,\nu) = \left(\min_{\gamma \in \Pi(\mu,\nu)} \int |x-y|_\infty^p \gamma(\mathrm{d}x, \mathrm{d}y) \right)^{1/p}$$

The rationale

- $K|x-y|_\infty^p \geq \sum_{t=1}^{K} |x^{(j_t)} - y^{(j_t)}|^p$ for any $1 \leq j_t \leq d$
- $K^{1/p} \cdot W_{p,\ell^\infty}(\mu,\nu)$ serves as an upper bound for the W_p distance between any K-dimensional marginals of μ and ν

From now on, we consider $p = 2$ and $K = 1$

Roadmap of this Talk

1. A New Metric Designed for Low Dimensional Marginals
2. Delocalization? Product, Gaussian, and Rotations
3. Delocalization: Potentials with Sparse and Local Interactions
4. Generalization with Asymptotic Arguments

Positive Examples: Product Measures

W_{2,ℓ^∞} bias for product measures

Consider $\pi \propto \exp(-V)$ where $V(x) = \sum_{i=1}^{d} V_i(x^{(i)})$ satisfies $\alpha \leq \nabla^2 V_i \leq \beta$. Then, for $h \leq 1/\beta$, it holds that

$$W_{2,\ell^\infty}(\pi_h, \pi) = O\left(\frac{\beta}{\alpha}\sqrt{h \log(2d)}\right)$$

- Thus $W_2(\pi^{(j)}, \pi_h^{(j)}) = O(\frac{\beta}{\alpha}\sqrt{h \log(2d)})$
- In fact 1D marginal $W_2(\pi^{(j)}, \pi_h^{(j)}) = O(\frac{\beta}{\alpha}\sqrt{h})$ dimension free
- In comparison:

$$W_2(\pi, \pi_h) = O(\frac{\beta}{\alpha}\sqrt{dh})$$

[Durmus, Moulines, 2019], etc.

- Thus overall bias nearly delocalized accross all 1D marginals

Positive Examples: Gaussian Measures

W_{2,ℓ^∞} bias for Gaussian measures

Consider $\pi \propto \exp(-V)$ and $V(x) = \frac{1}{2}(x-m)^T\Sigma^{-1}(x-m)$ where $m \in \mathbb{R}^d$ and $\alpha I \preceq \Sigma^{-1} \preceq \beta I$. Then, for $h \leq 1/\beta$, it holds that

$$W_{2,\ell^\infty}(\pi_h, \pi) = O\left(\sqrt{h \log(2d)}\right)$$

- Use explicit formula $\pi_h = \mathcal{N}(0, \Sigma(I - \frac{h}{2}\Sigma^{-1})^{-1})$
- Thus $W_2(\pi^{(j)}, \pi_h^{(j)}) = O\left(\sqrt{h \log(2d)}\right)$ nearly dimension free
- Again, overall bias nearly delocalized accross all 1D marginals

A Negative Example

W_{2,ℓ^∞} bias for rotated product measures

Consider $\pi = \rho^{\otimes d}$ where ρ is a 1D centered distribution, such that the mean of ρ and the biased ρ_h differs by $\delta > 0$.

Let $\tilde{\pi} = Q\#\pi$ where Q is a rotation $(Qx)^{(1)} = \frac{1}{\sqrt{d}} \sum_{i=1}^d x^{(i)}$. Then

$$W_{2,\ell^\infty}(\tilde{\pi}, \tilde{\pi}_h) \geq \sqrt{d}\delta$$

where $\tilde{\pi}_h$ is the corresponding biased distribution for $\tilde{\pi}$

Proof sketch: we have $\tilde{\pi}_h = Q\#\pi_h$

$$\begin{aligned} W_{2,\ell^\infty}(\tilde{\pi}, \tilde{\pi}_h) &\geq W_{1,\ell^\infty}(\tilde{\pi}, \tilde{\pi}_h) \\ &\geq \left| \int x^{(1)}(\tilde{\pi} - \tilde{\pi}_h) \right| \\ &= \left| \int (\frac{1}{\sqrt{d}} \sum_{i=1}^d x^{(i)})(\pi - \pi_h) \right| = \sqrt{d}\delta \end{aligned}$$

Bias can concentrate on one coordinate!

Delocalization of Bias

Observations:

- Positive examples: product measures, Gaussian measures
- Negative examples: some rotated product measures

The negative example is characterized by strong, dense interactions between coordinates after the rotation

Question: To which broader extent that delocalization holds?

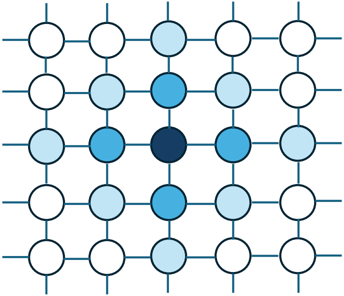
Roadmap of this Talk

Main Results: Sparse and Local Potentials

Theorem: W_{2,ℓ^∞} bias for sparse/local potentials

For $V \in C^2$ with $\alpha I \preceq \nabla^2 V \preceq \beta I$ that satisfies the sparsity condition illustrated in the figure with $s_k \leq C(k+1)^n$, then

$$W_{2,\ell^\infty}(\pi, \pi_h) \leq \sqrt{h \log(2d)} \left(O\left(\frac{\beta}{\alpha} \log(2d)\right) \right)^{\frac{n}{2}+1}$$



Some ith variable $x^{(i)}$

1st layer: $N_1(x^{(i)})$

2nd layer: $N_2(x^{(i)})$

Potential $V(x) = \sum_{i=1}^{d} V_i(x)$ and V_i only depends on $N_1(x^{(i)})$

Sparsity parameter $s_k = \max_i \left|N_k\left(x^{(i)}\right)\right|$. This example: $s_k = O(k^2)$

Sketch of Arguments

- Continuous time $Y_t, t \in [kh, (k+1)h]$ and unadjusted X_{kh}

$$X_{(k+1)h} = X_{kh} - h\nabla V(X_{kh}) + \sqrt{2}(B_{(k+1)h} - B_{kh})$$

coupled with the same B_t

- Define $\overline{Y}_{(k+1)h} = Y_{kh} - h\nabla V(Y_{kh}) + \sqrt{2}(B_{(k+1)h} - B_{kh})$

$$\sqrt{\mathbb{E}[|X_{(k+1)h} - Y_{(k+1)h}|_\infty^2]}$$
$$\leq \underbrace{\sqrt{\mathbb{E}[|X_{(k+1)h} - \overline{Y}_{(k+1)h}|_\infty^2]}}_{\text{(a)}} + \underbrace{\sqrt{\mathbb{E}[|\overline{Y}_{(k+1)h} - Y_{(k+1)h}|_\infty^2]}}_{\text{(b) "discretization error"}}$$

- Part (b): discretization error = $O(\beta h^{3/2}\sqrt{\log(2d)})$

(reminiscent of the fact that $\mathbb{E}[|B_t|_\infty^2] \leq t\log(2d)$)

Sketch of Arguments

- Continuous time $Y_t, t \in [kh, (k+1)h]$ and unadjusted X_{kh}

$$X_{(k+1)h} = X_{kh} - h\nabla V(X_{kh}) + \sqrt{2}(B_{(k+1)h} - B_{kh})$$

coupled with the same B_t

- Define $\overline{Y}_{(k+1)h} = Y_{kh} - h\nabla V(Y_{kh}) + \sqrt{2}(B_{(k+1)h} - B_{kh})$

$$\sqrt{\mathbb{E}[|X_{(k+1)h} - Y_{(k+1)h}|_\infty^2]} \le \underbrace{\sqrt{\mathbb{E}[|X_{(k+1)h} - \overline{Y}_{(k+1)h}|_\infty^2]}}_{\text{(a)}} + \underbrace{\sqrt{\mathbb{E}[|\overline{Y}_{(k+1)h} - Y_{(k+1)h}|_\infty^2]}}_{\text{(b) ``discretization error''}}$$

- Part (b): discretization error = $O(\beta h^{3/2}\sqrt{\log(2d)})$

(reminiscent of the fact that $\mathbb{E}[|B_t|_\infty^2] \le t\log(2d)$)

Sketch of Arguments

- Continuous time $Y_t, t \in [kh, (k+1)h]$ and unadjusted X_{kh}

$$X_{(k+1)h} = X_{kh} - h\nabla V(X_{kh}) + \sqrt{2}(B_{(k+1)h} - B_{kh})$$

coupled with the same B_t

- Define $\overline{Y}_{(k+1)h} = Y_{kh} - h\nabla V(Y_{kh}) + \sqrt{2}(B_{(k+1)h} - B_{kh})$

$$\sqrt{\mathbb{E}[|X_{(k+1)h} - Y_{(k+1)h}|_\infty^2]} \leq \underbrace{\sqrt{\mathbb{E}[|X_{(k+1)h} - \overline{Y}_{(k+1)h}|_\infty^2]}}_{\text{(a)}} + \underbrace{\sqrt{\mathbb{E}[|\overline{Y}_{(k+1)h} - Y_{(k+1)h}|_\infty^2]}}_{\text{(b) “discretization error”}}$$

- Part (b): discretization error = $O(\beta h^{3/2}\sqrt{\log(2d)})$

(reminiscent of the fact that $\mathbb{E}[|B_t|_\infty^2] \leq t\log(2d)$)

- Part (a):

$$\text{(a)} = \sqrt{\mathbb{E}[|X_{kh} - Y_{kh} - h(\nabla V(X_{kh}) - \nabla V(Y_{kh}))|_\infty^2]}$$
$$= \sqrt{\mathbb{E}[|H_k(X_{kh} - Y_{kh})|_\infty^2]}$$

  where $H_k = I - h\int_0^1 \nabla^2 V(uX_{kh} + (1-u)Y_{kh})\mathrm{d}u$

- When $\nabla^2 V$ is diagonal, $|H_k|_\infty = |H_k|_2 \leq 1 - \alpha h \leq \exp(-\alpha h)$ so we get contraction

- In general, H_k is non-diagonal but sparse. We have

$$|H_k|_\infty \leq \sqrt{s_1}|H_k|_2 \leq \sqrt{s_1}\exp(-\alpha h)$$

  Not a one-step contraction in general

Sketch of Arguments: Multiple-step Coupling

- One-step iteration

$$\sqrt{\mathbb{E}[|X_{(k+1)h} - Y_{(k+1)h}|_\infty^2]} \le \sqrt{\mathbb{E}[|H_k(X_{kh} - Y_{kh})|_\infty^2]} + \mathrm{error}(1)$$

Sketch of Arguments: Multiple-step Coupling

- One-step iteration

$$\sqrt{\mathbb{E}[|X_{(k+1)h} - Y_{(k+1)h}|_\infty^2]} \le \sqrt{\mathbb{E}[|H_k(X_{kh} - Y_{kh})|_\infty^2]} + \mathrm{error}(1)$$

- Moving back and two-step iterations

$$\begin{aligned}
&\sqrt{\mathbb{E}[|H_k(X_{kh} - Y_{kh})|_\infty^2]} + \mathrm{error}(1) \\
\le&\sqrt{\mathbb{E}[|H_k(X_{kh} - \overline{Y}_{kh})|_\infty^2]} + \sqrt{\mathbb{E}[|H_k(\overline{Y}_{kh} - Y_{kh})|_\infty^2]} + \mathrm{error}(1) \\
=&\sqrt{\mathbb{E}[|H_k H_{k-1}(X_{(k-1)h} - Y_{(k-1)h})|_\infty^2]} + \mathrm{error}(2)
\end{aligned}$$

Sketch of Arguments: Multiple-step Coupling

- One-step iteration

$$\sqrt{\mathbb{E}[|X_{(k+1)h} - Y_{(k+1)h}|_\infty^2]} \leq \sqrt{\mathbb{E}[|H_k(X_{kh} - Y_{kh})|_\infty^2]} + \mathrm{error}(1)$$

- Moving back and two-step iterations

$$\begin{aligned}
&\sqrt{\mathbb{E}[|H_k(X_{kh} - Y_{kh})|_\infty^2]} + \mathrm{error}(1) \\
\leq &\sqrt{\mathbb{E}[|H_k(X_{kh} - \overline{Y}_{kh})|_\infty^2]} + \sqrt{\mathbb{E}[|H_k(\overline{Y}_{kh} - Y_{kh})|_\infty^2]} + \mathrm{error}(1) \\
= &\sqrt{\mathbb{E}[|H_k H_{k-1}(X_{(k-1)h} - Y_{(k-1)h})|_\infty^2]} + \mathrm{error}(2)
\end{aligned}$$

- N-step iterations

$$\begin{aligned}
&\sqrt{\mathbb{E}[|X_{(k+N)h} - Y_{(k+N)h}|_\infty^2]} \\
\leq &\sqrt{\mathbb{E}[|H_{k+N-1}H_{k+N-2}\cdots H_k(X_{kh} - Y_{kh})|_\infty]} + \mathrm{error}(N) \\
\leq &\exp(-\alpha N h)\sqrt{d}\sqrt{\mathbb{E}[|X_{kh} - Y_{kh}|_\infty^2]} + \mathrm{error}(N)
\end{aligned}$$

Here $N \sim (\log d)/h$ leads to a contraction

Sketch of Arguments: Bound Discretization Errors

How to control $\mathrm{error}(N)$?

- For $N = 1$:

$$
\begin{aligned}
&\mathbb{E}[|\overline{Y}_{(k+1)h} - Y_{(k+1)h}|_\infty^2] \\
=&\mathbb{E}[|\int_{kh}^{(k+1)h} \nabla V(Y_t) - \nabla V(Y_{kh})\mathrm{d}t|_\infty^2] \\
\le& h\int_{kh}^{(k+1)h} \mathbb{E}[|\nabla V(Y_t) - \nabla V(Y_{kh})|_\infty^2]\mathrm{d}t \\
\le& h\int_{kh}^{(k+1)h}\int_0^1 \mathbb{E}[|\nabla^2 V(uY_t + (1-u)Y_{kh})(Y_t - Y_{kh})|_\infty^2]\mathrm{d}u\mathrm{d}t \\
\le& h s_1\beta^2 \int_{kh}^{(k+1)h} \mathbb{E}[|Y_t - Y_{kh}|_\infty^2]\mathrm{d}t = h s_1 \beta^2 \cdot O(h^2\log(2d))
\end{aligned}
$$

Sketch of Arguments: Bound Discretization Errors

How to control $\mathrm{error}(N)$?

- For $N = 2$:

$$
\begin{aligned}
&\mathbb{E}[|H_k(\overline{Y}_{kh} - Y_{kh})|_\infty^2] \\
\leq& h\int_{(k-1)h}^{kh} \mathbb{E}[|H_k(\nabla V(Y_t) - \nabla V(Y_{(k-1)h}))|_\infty^2]\mathrm{d}t \\
\leq& h\int_{(k-1)h}^{kh}\int_0^1 \mathbb{E}[|H_k(\nabla^2 V(uY_t + (1-u)Y_{(k-1)h}))(Y_t - Y_{(k-1)h})|_\infty^2]\mathrm{d}u\mathrm{d}t
\end{aligned}
$$

- How to bound $|H_k(\nabla^2 V(uY_t + (1-u)Y_{(k-1)h}))|_\infty$?

Sketch of Arguments: Bound Discretization Errors

How to control $\mathrm{error}(N)$?

- For $N = 2$:

$$\begin{aligned}
&\mathbb{E}[|H_k(\overline{Y}_{kh} - Y_{kh})|_\infty^2] \\
\leq & h \int_{(k-1)h}^{kh} \mathbb{E}[|H_k(\nabla V(Y_t) - \nabla V(Y_{(k-1)h}))|_\infty^2] \mathrm{d}t \\
\leq & h \int_{(k-1)h}^{kh} \int_0^1 \mathbb{E}[|H_k(\nabla^2 V(uY_t + (1-u)Y_{(k-1)h}))(Y_t - Y_{(k-1)h})|_\infty^2] \mathrm{d}u\mathrm{d}t
\end{aligned}$$

- How to bound $|H_k(\nabla^2 V(uY_t + (1-u)Y_{(k-1)h}))|_\infty$?
- A simple bound

$$|H_k(\nabla^2 V(uY_t + (1-u)Y_{(k-1)h}))|_\infty \leq \sqrt{s_2}\beta \exp(-\alpha h)$$

Sketch of Arguments: Bound Discretization Errors

How to control $\mathrm{error}(N)$?

- For $N = 2$:

$$
\begin{aligned}
&\mathbb{E}[|H_k(\overline{Y}_{kh} - Y_{kh})|_\infty^2] \\
\leq & h \int_{(k-1)h}^{kh} \mathbb{E}[|H_k(\nabla V(Y_t) - \nabla V(Y_{(k-1)h}))|_\infty^2] \mathrm{d}t \\
\leq & h \int_{(k-1)h}^{kh} \int_0^1 \mathbb{E}[|H_k(\nabla^2 V(uY_t + (1-u)Y_{(k-1)h}))(Y_t - Y_{(k-1)h})|_\infty^2] \mathrm{d}u \mathrm{d}t
\end{aligned}
$$

- How to bound $|H_k(\nabla^2 V(uY_t + (1-u)Y_{(k-1)h}))|_\infty$?
- A simple bound

$$
|H_k(\nabla^2 V(uY_t + (1-u)Y_{(k-1)h}))|_\infty \leq \sqrt{s_2}\beta \exp(-\alpha h)
$$

- Issue: although the bound does take into account sparsity, the sparsity growth s_2 does not depend on h

Sketch of Arguments: Sparsity Growth Bound

Consider the general N-case

- Let $J_N = |H_{k+N-1}H_{k+N-2}\cdots H_k(\nabla^2 V(uY_t + (1-u)Y_{(k-1)h})|_\infty$, then simple bound $|J_N|_\infty \leq \beta\sqrt{s_N}\exp(-\alpha N h)$

  The issue again is that s_N does not depend on h

Sketch of Arguments: Sparsity Growth Bound

Consider the general N-case

- Let $J_N = |H_{k+N-1}H_{k+N-2}\cdots H_k(\nabla^2 V(uY_t + (1-u)Y_{(k-1)h})|_\infty$, then simple bound $|J_N|_\infty \le \beta\sqrt{s_N}\exp(-\alpha Nh)$

  The issue again is that s_N does not depend on h

- Improved bound by using sparsity bound for terms involving small powers of h and using maximum bound for terms involving large powers of h

$$|J_N|_\infty \le \beta(\sqrt{s_r}\exp(-\alpha Nh) + \sqrt{d}\exp(-r))$$

  for any $r \ge e^2 Nh\beta$

Sketch of Arguments: Sparsity Growth Bound

Consider the general N-case

- Let $J_N = |H_{k+N-1}H_{k+N-2}\cdots H_k(\nabla^2 V(uY_t + (1-u)Y_{(k-1)h})|_\infty$, then simple bound $|J_N|_\infty \le \beta\sqrt{s_N}\exp(-\alpha Nh)$

  The issue again is that s_N does not depend on h

- Improved bound by using sparsity bound for terms involving small powers of h and using maximum bound for terms involving large powers of h

$$|J_N|_\infty \le \beta(\sqrt{s_r}\exp(-\alpha Nh) + \sqrt{d}\exp(-r))$$

  for any $r \ge e^2 Nh\beta$

- In particular, taking $r_N = \lceil e^2 Nh\beta + \log\sqrt{d}\rceil$ leads to

$$|J_N|_\infty \le 2\beta\sqrt{s_{r_N}}\exp(-\alpha Nh)$$

  Here r_N scales with physical time Nh

Sketch of Arguments: Back to Discretization Errors

Back to the estimate of $\mathrm{error}(N)$

- For $N = 2$:

$$
\begin{aligned}
&\mathbb{E}[|H_k(\overline{Y}_{kh} - Y_{kh})|_\infty^2] \\
\leq & h \int_{(k-1)h}^{kh} \mathbb{E}[|H_k(\nabla V(Y_t) - \nabla V(Y_{(k-1)h}))|_\infty^2] \mathrm{d}t \\
\leq & h \int_{(k-1)h}^{kh} \int_0^1 \mathbb{E}[|H_k(\nabla^2 V(uY_t + (1-u)Y_{(k-1)h}))(Y_t - Y_{(k-1)h})|_\infty^2] \mathrm{d}u\mathrm{d}t \\
\leq & 4h s_{r_2} \beta^2 \exp(-2\alpha h) \int_{(k-1)h}^{kh} \mathbb{E}[|Y_t - Y_{(k-1)h}|_\infty^2] \mathrm{d}t \\
= & 4h s_{r_2} \beta^2 \exp(-2\alpha h) \cdot O(h^2 \log(2d))
\end{aligned}
$$

Sketch of Arguments: Back to Discretization Errors

Putting everything together

- For general N:

$$\mathrm{error}(N) \leq 2\beta \left(\sum_{i=1}^{N} \exp(-\alpha h(i-1)) \sqrt{s_{r_i}} \right) \cdot O\left(h^{3/2} \sqrt{\log(2d)} \right)$$

Sketch of Arguments: Back to Discretization Errors

Putting everything together

- For general N:

$$\mathrm{error}(N) \leq 2\beta \left(\sum_{i=1}^{N} \exp(-\alpha h(i-1)) \sqrt{s_{r_i}} \right) \cdot O\left(h^{3/2} \sqrt{\log(2d)} \right)$$

- Therefore, we get

$$W_{2,\ell^\infty}(\rho_{(k+N)h}, \pi) \leq \exp(-\alpha N h) \sqrt{d} W_{2,\ell^\infty}(\rho_{kh}, \pi) + \mathrm{error}(N)$$

Sketch of Arguments: Back to Discretization Errors

Putting everything together

- For general N:

$$\mathrm{error}(N) \leq 2\beta \left(\sum_{i=1}^{N} \exp(-\alpha h(i-1)) \sqrt{s_{r_i}} \right) \cdot O\left(h^{3/2} \sqrt{\log(2d)} \right)$$

- Therefore, we get

$$W_{2,\ell^\infty}(\rho_{(k+N)h}, \pi) \leq \exp(-\alpha N h)\sqrt{d} W_{2,\ell^\infty}(\rho_{kh}, \pi) + \mathrm{error}(N)$$

- Using $s_k = O((k+1)^n)$ and taking $N = \lceil \frac{\log(2\sqrt{d})}{h\alpha} \rceil$

$$W_{2,\ell^\infty}(\rho_{(k+N)h}, \pi) \leq \frac{1}{2} W_{2,\ell^\infty}(\rho_{kh}, \pi) + \sqrt{h \log(2d)} \left(O\left(\frac{\beta}{\alpha} \log(2d)\right) \right)^{\frac{n}{2}+1}$$

Sketch of Arguments: Back to Discretization Errors

Putting everything together

- For general N:

$$\mathrm{error}(N) \leq 2\beta \left(\sum_{i=1}^{N} \exp(-\alpha h(i-1))\sqrt{s_{r_i}} \right) \cdot O\left(h^{3/2}\sqrt{\log(2d)} \right)$$

- Therefore, we get

$$W_{2,\ell^\infty}(\rho_{(k+N)h}, \pi) \leq \exp(-\alpha N h)\sqrt{d} W_{2,\ell^\infty}(\rho_{kh}, \pi) + \mathrm{error}(N)$$

- Using $s_k = O((k+1)^n)$ and taking $N = \lceil \frac{\log(2\sqrt{d})}{h\alpha} \rceil$

$$W_{2,\ell^\infty}(\rho_{(k+N)h}, \pi) \leq \frac{1}{2} W_{2,\ell^\infty}(\rho_{kh}, \pi) + \sqrt{h \log(2d)} \left(O\left(\frac{\beta}{\alpha} \log(2d)\right) \right)^{\frac{n}{2}+1}$$

- Finally $W_{2,\ell^\infty}(\pi_h, \pi) \leq \sqrt{h \log(2d)} \left(O\left(\frac{\beta}{\alpha} \log(2d)\right) \right)^{\frac{n}{2}+1}$

Roadmap of this Talk

Asymptotic Arguments for the Bias of Observables

Bias of Observables [Chen, Cheng, Niles-Weed, Weare 2024]

Assume f is sufficiently regular and $\int f\pi = 0$. Then, it holds that

$$\int f\pi - \int f\pi_h = \frac{1}{4}h\left(\int(-2\Delta f + |\nabla \log \pi|_2^2 f)\pi\right) + o(h)$$

Moreover, we also have the following formula:

$$\int f\pi - \int f\pi_h = -\frac{1}{4}h\left(\int(\Delta f + f\Delta \log \pi)\pi\right) + o(h)$$

Asymptotic Arguments for the Bias of Observables

Bias of Observables [Chen, Cheng, Niles-Weed, Weare 2024]

Assume f is sufficiently regular and $\int f\pi = 0$. Then, it holds that

$$\int f\pi - \int f\pi_h = \frac{1}{4}h\left(\int(-2\Delta f + |\nabla \log \pi|_2^2 f)\pi\right) + o(h)$$

Moreover, we also have the following formula:

$$\int f\pi - \int f\pi_h = -\frac{1}{4}h\left(\int(\Delta f + f\Delta \log \pi)\pi\right) + o(h)$$

Poisson argument: Let $\mathcal{L}$ and $\mathcal{L}_h$ be the generators of Langevin dynamics and unadjusted Langevin [Mattingly, Stuart, Tretyakov 2010]

- $\mathcal{L}u = \nabla \log \pi \cdot \nabla u + \Delta u$,
  $\mathcal{L}_h u(x) = \frac{1}{h}(\mathbb{E}[u(x + h\nabla \log \pi(x) + \sqrt{2h}\xi)] - u(x))$
- Let $\mathcal{L}u = f$. Then, we get

$$\int f\pi - \int f\pi_h = -\int \mathcal{L}u\pi_h = \int (\mathcal{L}_h u - \mathcal{L}u)\pi_h, \quad \ldots$$

Delocalization of Bias for Observables

Bias of Observables **[Chen, Cheng, Niles-Weed, Weare 2024]**

Assume f is sufficiently regular and $\int f\pi = 0$. Then, it holds that

$$\int f\pi - \int f\pi_h = \frac{1}{4}h\left(\int (-2\Delta f + |\nabla \log \pi|_2^2 f)\pi\right) + o(h)$$

Moreover, we also have the following formula:

$$\int f\pi - \int f\pi_h = -\frac{1}{4}h\left(\int (\Delta f + f\Delta \log \pi)\pi\right) + o(h)$$

- If $\pi(x) = \mathcal{N}(x; m, \Sigma)$, then $\int f(\Delta \log \pi)\pi = 0$. The first order term $\int \pi\Delta f$ only depends on the coordinates that f takes
- This delocalization of observable bias can be generalized to

$$\pi(x) \propto \exp(-V(x)) \propto \mathcal{N}(x; m, \Sigma)\exp(-U(x))$$

i.e., perturbation of Gaussians

Summary

A **“delocalization of bias”** phenomenon for unadjusted Langevin

- Nearly d-independent step size and complexity
- Phenomenon not shared by unbiased schemes
- We prove it for log-concave Gaussians and sparse potentials
- Not hold for some potentials with strong, dense interactions
- Asymptotic arguments for general observables and potentials
(up to first order)

Extension to general dynamics and distributions?