Convergence of Unadjusted Langevin in High Dimensions

A "Delocalization of Bias" Phenomenon

Yifan Chen, NYU Courant

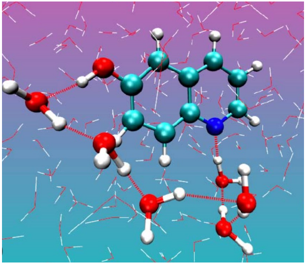
Computational Bayes Statistics Journal Club, Sep 2024

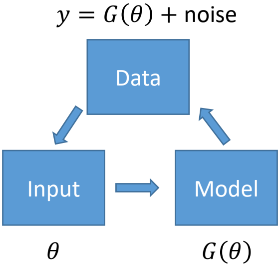
joint work with Xiaoou Cheng, Jonathan Niles-Weed, Jonathan Weare

Context

Classical sampling problem

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





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

Challenges: High dimensional probability distributions

Methodology: Langevin Dynamics

(Overdamped) Langevin's dynamics

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

Under mild assumptions, as $t \to \infty$, $\mathrm{Law}(X_t) \to \pi$

Methodology: Langevin Dynamics

(Overdamped) Langevin's dynamics

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

Under mild assumptions, as $t \to \infty, \mathrm{Law}(X_t) \to \pi$

Biased scheme: unadjusted Langevin, converging to $\pi_h \neq \pi$

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

Unbiased scheme: converging to π

- MALA: accept w/ some probability, otherwise reject
  [Rossky, Doll, Friedman 1978], [Roberts, Tweedie 1997], etc.
- Proximal sampler: Gibbs sampling on $\exp\left(-V(x) - \frac{|x-y|^2}{2h}\right)$
  [Lee, Shen, Tian 2021], [Chen, Chewi, Salim, Wibisono 2022], etc.

Methodology: Langevin Dynamics

(Overdamped) Langevin's dynamics

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

Under mild assumptions, as $t \to \infty$, $\mathrm{Law}(X_t) \to \pi$

Biased scheme: unadjusted Langevin, converging to $\pi_h \neq \pi$

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

Unbiased scheme: converging to π

- MALA: accept w/ some probability, otherwise reject
  [Rossky, Doll, Friedman 1978], [Roberts, Tweedie 1997], etc.
- Proximal sampler: Gibbs sampling on $\exp\left(-V(x) - \frac{|x-y|^2}{2h}\right)$
  [Lee, Shen, Tian 2021], [Chen, Chewi, Salim, Wibisono 2022], etc.

Question: any biased versus unbiased guidance in high dims?

Existing Results on Complexity in High Dimensions

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

Existing Results on Complexity in High Dimensions

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

- Under assumption $\alpha I \preceq \nabla^2 V \preceq \beta I$ where $0 < \alpha \leq \beta < \infty$

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

$\Rightarrow h = O(1/d)$ for bounded bias in any dimension

W_2: [Durmus, Moulines, 2019], etc.

TV or KL: [Dalalyan 2017], [Cheng, Bartlett 2018], etc.

Existing Results on Complexity in High Dimensions

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

- Under assumption $\alpha I \preceq \nabla^2 V \preceq \beta I$ where $0 < \alpha \leq \beta < \infty$

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

$\Rightarrow h = O(1/d)$ for bounded bias in any dimension

W_2: [Durmus, Moulines, 2019], etc.

TV or KL: [Dalalyan 2017], [Cheng, Bartlett 2018], etc.

- Improvement to $h = O(1/\sqrt{d})$ under additional assumption

[Li, Zha, Tao 2022]

Iteration complexity: $O(d)$ or $O(\sqrt{d})$ steps, up to $\log d$ terms

(assuming $\alpha I \preceq \nabla^2 V \preceq \beta I$)

Existing Results on Complexity in High Dimensions

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

- Diffusion limit arguments: $h \sim 1/d^{1/3}$ or $1/\sqrt{d}$

  [Roberts, Rosenthal 1998], [Christensen, Roberts, Rosenthal 2005], etc.

- Non-asymptotic mixing time bounds: $h \sim 1/\sqrt{d}$ from a warm start or $h \sim 1/d$ from feasible start (assume $\alpha I \preceq \nabla^2 V \preceq \beta I$)

  [Dwivedi, Chen, Wainwright, Yu 2018], [Chewi, Lu, Ahn, Cheng, Le Gouic, Rigollet 2021], [Wu, Schmidler, Chen 2022], etc.

Existing Results on Complexity in High Dimensions

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

- Diffusion limit arguments: $h \sim 1/d^{1/3}$ or $1/\sqrt{d}$

  [Roberts, Rosenthal 1998], [Christensen, Roberts, Rosenthal 2005], etc.

- Non-asymptotic mixing time bounds: $h \sim 1/\sqrt{d}$ from a warm start or $h \sim 1/d$ from feasible start (assume $\alpha I \preceq \nabla^2 V \preceq \beta I$)

  [Dwivedi, Chen, Wainwright, Yu 2018], [Chewi, Lu, Ahn, Cheng, Le Gouic, Rigollet 2021], [Wu, Schmidler, Chen 2022], etc.

For proximal samplers: h needs to be small for efficient implementation of restricted Gaussian oracle (RGO)

- $h \sim 1/d$ if the RGO is implemented via rejection sampling; $h \sim 1/\sqrt{d}$ if implemented via approximate rejection sampling (assume $\alpha I \preceq \nabla^2 V \preceq \beta I$)

  [Chen, Chewi, Salim, Wibisono 2022], [Fan, Yuan, Chen 2023], etc.

Existing Results on Complexity in High Dimensions

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

- Diffusion limit arguments: $h \sim 1/d^{1/3}$ or $1/\sqrt{d}$

  [Roberts, Rosenthal 1998], [Christensen, Roberts, Rosenthal 2005], etc.

- Non-asymptotic mixing time bounds: $h \sim 1/\sqrt{d}$ from a warm start or $h \sim 1/d$ from feasible start (assume $\alpha I \preceq \nabla^2 V \preceq \beta I$)

  [Dwivedi, Chen, Wainwright, Yu 2018], [Chewi, Lu, Ahn, Cheng, Le Gouic, Rigollet 2021], [Wu, Schmidler, Chen 2022], etc.

For proximal samplers: h needs to be small for efficient implementation of restricted Gaussian oracle (RGO)

- $h \sim 1/d$ if the RGO is implemented via rejection sampling; $h \sim 1/\sqrt{d}$ if implemented via approximate rejection sampling (assume $\alpha I \preceq \nabla^2 V \preceq \beta I$)

  [Chen, Chewi, Salim, Wibisono 2022], [Fan, Yuan, Chen 2023], etc.

Iteration complexity: again, $O(d)$ or $O(\sqrt{d})$ steps, up to $\log d$ terms

Summary of Existing Results

Constraint on h: Both biased and unbiased require $h = O(d^{-c})$

- for small bias in the biased approach;
- for good acceptance rate / efficiency in the unbiased approach

Summary of Existing Results

Constraint on h: Both biased and unbiased require $h = O(d^{-c})$

- for small bias in the biased approach;
- for good acceptance rate / efficiency in the unbiased approach

Similar c: little theoretical distinction between biased and unbiased

Summary of Existing Results

Constraint on h: Both biased and unbiased require $h = O(d^{-c})$

- for small bias in the biased approach;
- for good acceptance rate / efficiency in the unbiased approach

Similar c: little theoretical distinction between biased and unbiased

- For unbiased approaches, the scaling is not avoidable generally

  Local methods: Metropolis-within-Gibbs [Tong, Morzfeld, Marzouk 2020], etc.

Summary of Existing Results

Constraint on h: Both biased and unbiased require $h = O(d^{-c})$

- for small bias in the biased approach;
- for good acceptance rate / efficiency in the unbiased approach

Similar c: little theoretical distinction between biased and unbiased

- For unbiased approaches, the scaling is not avoidable generally

  Local methods: Metropolis-within-Gibbs [Tong, Morzfeld, Marzouk 2020], etc.

- For biased approaches, these results seem at odds with abundant empirical evidence. For example, large-scale molecular dynamics simulations, using similar integrators, typically employ fixed step sizes of several femtoseconds, regardless of system size [Leimkuhler, Matthews 2015]

Summary of Existing Results

Constraint on h: Both biased and unbiased require $h = O(d^{-c})$

- for small bias in the biased approach;
- for good acceptance rate / efficiency in the unbiased approach

Similar c: little theoretical distinction between biased and unbiased

- For unbiased approaches, the scaling is not avoidable generally

  Local methods: Metropolis-within-Gibbs [Tong, Morzfeld, Marzouk 2020], etc.

- For biased approaches, these results seem at odds with abundant empirical evidence. For example, large-scale molecular dynamics simulations, using similar integrators, typically employ fixed step sizes of several femtoseconds, regardless of system size [Leimkuhler, Matthews 2015]

- In unadjusted schemes, $h = O(1)$ could suffice for accurate averaged observables, e.g. $f(x) = \frac{1}{d}\sum_{i=1}^{d} \Phi(x^{(i)})$ with Lipschitz-Φ, which satisfies $|\nabla f(x)|_2 \leq |x|_2/\sqrt{d}$

  [Bou-Rabee, Schuh 2023], [Durmus, Eberle 2024]

This Work

For unadjusted Langevin [Chen, Cheng, Niles-Weed, Weare 2024]

$h = O(1/K)$ could suffice for desired accuracy in all K-marginals!

- Iteration complexity: $O(K)$
- Results proved under the assumption $\alpha I \preceq \nabla^2 V \preceq \beta I$ and V is Gaussian/"sparse" (and some generalizations)

($\log d$ terms omitted)

This Work

For unadjusted Langevin [Chen, Cheng, Niles-Weed, Weare 2024]

$h = O(1/K)$ could suffice for desired accuracy in all K-marginals!

- Iteration complexity: $O(K)$
- Results proved under the assumption $\alpha I \preceq \nabla^2 V \preceq \beta I$ and V is Gaussian/"sparse" (and some generalizations)

($\log d$ terms omitted)

Bias in each individual coordinate behaves nearly dimension-free!

- We refer to this benign dimension dependence phenomenon as

"Delocalization of Bias"

Roadmap of this Talk

1. A New Metric Designed for Low Dimensional Marginals

2. Delocalization? Product, Gaussian, and Rotations

3. Delocalization: Potentials with Sparse Interactions

4. Generalization with Asymptotic Arguments

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: ℓ^∞ "can" measure a small set of variables

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

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: ℓ^∞ "can" measure a small set of variables

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

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

In this work, we consider $p = 2$

Roadmap of this Talk

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 $\alpha \le \nabla^2 V_i \le \beta$. Then, for $h \le 1/\beta$, it holds that

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

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

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

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

- **Part (a)**:

$$\begin{aligned}(\text{a}) &= \sqrt{\mathbb{E}[|X_{kh} - Y_{kh} - h(\nabla V(X_{kh}) - \nabla V(Y_{kh}))|_\infty^2]} \\ &= \sqrt{\mathbb{E}[|H_k(X_{kh} - Y_{kh})|_\infty^2]} \\ &\le (1-\alpha h)\sqrt{\mathbb{E}[|X_{kh} - Y_{kh}|_\infty^2]} \le \exp(-\alpha h)\sqrt{\mathbb{E}[|X_{kh} - Y_{kh}|_\infty^2]}\end{aligned}$$

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

- Here $|H_k|_2 \le 1 - \alpha h$. When π is a product measure, H_k is a diagonal matrix so $|H_k|_\infty \le 1 - \alpha h$ as well

- Couple X_{kh} and Y_{kh} so $\sqrt{\mathbb{E}[|X_{kh} - Y_{kh}|_\infty^2]} = W_{2,\ell^\infty}(\rho_{kh}, \pi)$, then

$$W_{2,\ell^\infty}(\rho_{(k+1)h}, \pi) \le \exp(-\alpha h) W_{2,\ell^\infty}(\rho_{kh}, \pi) + O(\beta h^{3/2}\sqrt{\log(2d)})$$

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

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

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

- **Part (a)**:

$$
\begin{aligned}
\text{(a)} &= \sqrt{\mathbb{E}[|X_{kh} - Y_{kh} - h(\nabla V(X_{kh}) - \nabla V(Y_{kh}))|_\infty^2]} \\
&= \sqrt{\mathbb{E}[|H_k(X_{kh} - Y_{kh})|_\infty^2]} \\
&\le (1-\alpha h)\sqrt{\mathbb{E}[|X_{kh} - Y_{kh}|_\infty^2]} \le \exp(-\alpha h)\sqrt{\mathbb{E}[|X_{kh} - Y_{kh}|_\infty^2]}
\end{aligned}
$$

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

- Here $|H_k|_2 \le 1-\alpha h$. When π is a product measure, H_k is a diagonal matrix so $|H_k|_\infty \le 1-\alpha h$ as well

- Couple X_{kh} and Y_{kh} so $\sqrt{\mathbb{E}[|X_{kh} - Y_{kh}|_\infty^2]} = W_{2,\ell^\infty}(\rho_{kh}, \pi)$, then

$$W_{2,\ell^\infty}(\rho_{(k+1)h}, \pi) \le \exp(-\alpha h) W_{2,\ell^\infty}(\rho_{kh}, \pi) + O(\beta h^{3/2}\sqrt{\log(2d)})$$

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

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

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

- **Part (a)**:

$$
\begin{aligned}
\text{(a)} &= \sqrt{\mathbb{E}[|X_{kh} - Y_{kh} - h(\nabla V(X_{kh}) - \nabla V(Y_{kh}))|_\infty^2]} \\
&= \sqrt{\mathbb{E}[|H_k(X_{kh} - Y_{kh})|_\infty^2]} \\
&\le (1-\alpha h)\sqrt{\mathbb{E}[|X_{kh} - Y_{kh}|_\infty^2]} \le \exp(-\alpha h)\sqrt{\mathbb{E}[|X_{kh} - Y_{kh}|_\infty^2]}
\end{aligned}
$$

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

- Here $|H_k|_2 \le 1-\alpha h$. When π is a product measure, H_k is a diagonal matrix so $|H_k|_\infty \le 1-\alpha h$ as well

- Couple X_{kh} and Y_{kh} so $\sqrt{\mathbb{E}[|X_{kh} - Y_{kh}|_\infty^2]} = W_{2,\ell^\infty}(\rho_{kh},\pi)$, then

$$W_{2,\ell^\infty}(\rho_{(k+1)h},\pi) \le \exp(-\alpha h) W_{2,\ell^\infty}(\rho_{kh},\pi) + O(\beta h^{3/2}\sqrt{\log(2d)})$$

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

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 $\pi_h \sim \mathcal{N}(0, \Sigma(I - \frac{h}{2}\Sigma^{-1})^{-1})$
- W_2 between K-marginals of π_h and π is $O(\sqrt{Kh \log(2d)})$
- Overall bias nearly delocalized accross all 1D marginals

A Negative Example

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

Consider $\pi = \rho^{\otimes d}$ where ρ is a 1D centered distribution for which the biased distribution ρ_h has a nonzero mean so that their mean differs by $\delta > 0$. Consider the rotation matrix Q which satisfies $(Qx)^{(1)} = \frac{1}{\sqrt{d}} \sum_{i=1}^d x^{(i)}$. Let $\tilde{\pi} = Q\#\pi$. Then

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

A Negative Example

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

Consider $\pi = \rho^{\otimes d}$ where ρ is a 1D centered distribution for which the biased distribution ρ_h has a nonzero mean so that their mean differs by $\delta > 0$. Consider the rotation matrix Q which satisfies $(Qx)^{(1)} = \frac{1}{\sqrt{d}} \sum_{i=1}^d x^{(i)}$. Let $\tilde{\pi} = Q\#\pi$. Then

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

- We have $\tilde{\pi}_h = Q\#\pi_h$, and

$$|\int x^{(1)}(\tilde{\pi} - \tilde{\pi}_h)| = |\int f(\pi - \pi_h)| = \sqrt{d}\delta$$

  where $f(x) = \frac{1}{\sqrt{d}} \sum_{i=1}^d x^{(i)}$
- Thus $W_{2,\ell^\infty}(\tilde{\pi}, \tilde{\pi}_h) \geq W_{1,\ell^\infty}(\tilde{\pi}, \tilde{\pi}_h) \geq \left|\int x^{(1)}(\tilde{\pi} - \tilde{\pi}_h)\right| = \sqrt{d}\delta$

A Negative Example

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

Consider $\pi = \rho^{\otimes d}$ where ρ is a 1D centered distribution for which the biased distribution ρ_h has a nonzero mean so that their mean differs by $\delta > 0$. Consider the rotation matrix Q which satisfies $(Qx)^{(1)} = \frac{1}{\sqrt{d}} \sum_{i=1}^d x^{(i)}$. Let $\tilde{\pi} = Q\#\pi$. Then

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

- We have $\tilde{\pi}_h = Q\#\pi_h$, and

$$|\int x^{(1)}(\tilde{\pi} - \tilde{\pi}_h)| = |\int f(\pi - \pi_h)| = \sqrt{d}\delta$$

  where $f(x) = \frac{1}{\sqrt{d}} \sum_{i=1}^d x^{(i)}$
- Thus $W_{2,\ell^\infty}(\tilde{\pi}, \tilde{\pi}_h) \geq W_{1,\ell^\infty}(\tilde{\pi}, \tilde{\pi}_h) \geq \left|\int x^{(1)}(\tilde{\pi} - \tilde{\pi}_h)\right| = \sqrt{d}\delta$

No delocalization, but concentration 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 Potentials

W_{2,ℓ^∞} bias for some sparse potentials

(informal) suppose V is log-concave and satisfies the sparsity conditions illustrated in the figure with $s_k \le C(k+1)^n$, then

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

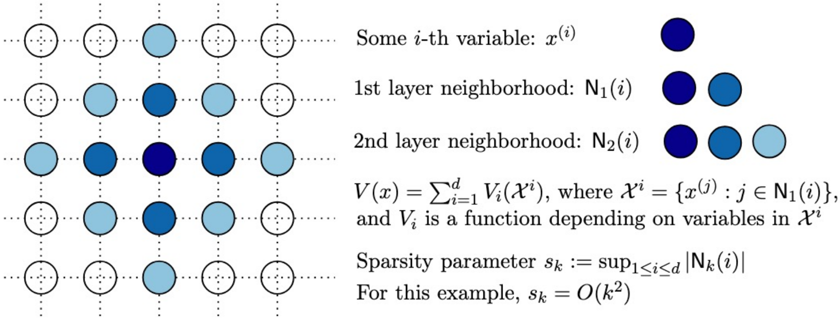


Figure: Illustration of a sparse potential we considered

Examples of Sparse Potentials

Consider the matrix

$$\begin{bmatrix} 2+\lambda(x) & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2+\lambda(x) & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2+\lambda(x) & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 2+\lambda(x) & -1 \\ 0 & 0 & \cdots & 0 & -1 & 2+\lambda(x) \end{bmatrix} \in \mathbb{R}^{d\times d}$$

- Let $\nabla^2 V(x)$ equal to the above matrix and λ satisfies $\min_{x\in\mathbb{R}^d} \lambda(x) = \alpha > 0$, then we have $s_k = \max\{2k+1, d\}$ and $\alpha I \preceq \nabla^2 V(x) \preceq \beta I$ with $\beta = 4 + \max_{x\in\mathbb{R}^d} \lambda(x)$
- $\lambda(x)$ can come from physical restoring force or Bayes priors
- More generally: $\nabla^2 V(x)$ can be a strongly log-concave perturbation of some graph Laplacian matrix

Sketch of Arguments

As before

- 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 (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 $H_k = I - h\int_0^1 \nabla^2 V(uX_{kh} + (1-u)Y_{kh})\mathrm{d}u$ is non-diagonal but sparse. We have $|H_k|_\infty \leq \sqrt{s_1}|H_k|_2 \leq \sqrt{s_1}\exp(-\alpha h)$

Sketch of Arguments

As before

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

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

- Part (a):

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

where $H_k = I - h\int_0^1 \nabla^2 V(uX_{kh} + (1-u)Y_{kh})\mathrm{d}u$ is non-diagonal but sparse. We have $|H_k|_\infty \le \sqrt{s_1}|H_k|_2 \le \sqrt{s_1}\exp(-\alpha h)$

Sketch of Arguments

As before

- 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 $\overline{Y}_{(k+1)h} = Y_{kh} - h\nabla V(Y_{kh}) + \sqrt{2}(B_{(k+1)h} - B_{kh})$
$$\begin{aligned}&\sqrt{\mathbb{E}[|X_{(k+1)h} - Y_{(k+1)h}|_\infty^2]} \\ &\le \underbrace{\sqrt{\mathbb{E}[|X_{(k+1)h} - \overline{Y}_{(k+1)h}|_\infty^2]}}_{\text{(a)}} + \underbrace{\sqrt{\mathbb{E}[|\overline{Y}_{(k+1)h} - Y_{(k+1)h}|_\infty^2]}}_{\text{(b) ``discretization error''}}\end{aligned}$$
- Part (a):
$$\begin{aligned}\text{(a)} &= \sqrt{\mathbb{E}[|X_{kh} - Y_{kh} - h(\nabla V(X_{kh}) - \nabla V(Y_{kh}))|_\infty^2]} \\ &= \sqrt{\mathbb{E}[|H_k(X_{kh} - Y_{kh})|_\infty^2]}\end{aligned}$$
where $H_k = I - h\int_0^1 \nabla^2 V(uX_{kh} + (1-u)Y_{kh})\mathrm{d}u$ is non-diagonal but sparse. We have $|H_k|_\infty \le \sqrt{s_1}|H_k|_2 \le \sqrt{s_1}\exp(-\alpha h)$

Sketch of Arguments: Multiple-step Coupling

- One-step iteration

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

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

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

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

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

- N-step iterations

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

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

Sketch of Arguments: Bound Discretization Errors

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

- For $N = 1$:

$$
\begin{aligned}
&\mathbb{E}[|\overline{Y}_{(k+1)h} - Y_{(k+1)h}|_\infty^2] \\
=&\mathbb{E}[|\int_{kh}^{(k+1)h} \nabla V(Y_t) - \nabla V(Y_{kh})\mathrm{d}t|_\infty^2] \\
\leq& h\int_{kh}^{(k+1)h} \mathbb{E}[|\nabla V(Y_t) - \nabla V(Y_{kh})|_\infty^2]\mathrm{d}t \\
\leq& h\int_{kh}^{(k+1)h} \int_0^1 \mathbb{E}[|\nabla^2 V(uY_t + (1-u)Y_{kh})(Y_t - Y_{kh})|_\infty^2]\mathrm{d}u\mathrm{d}t \\
\leq& hs_1\beta^2 \int_{kh}^{(k+1)h} \mathbb{E}[|Y_t - Y_{kh}|_\infty^2]\mathrm{d}t = hs_1\beta^2 \cdot O(h^2\log(2d))
\end{aligned}
$$

Sketch of Arguments: Bound Discretization Errors

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

- For $N = 2$:

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

Sketch of Arguments: Bound Discretization Errors

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

- For $N = 2$:

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

- Now, 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 $N = 2$:

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

- Now, 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 $N = 2$:

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

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

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

- The bound does take into account sparsity, but 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 $|J_N|_\infty \leq \beta\sqrt{s_N}\exp(-\alpha Nh)$

  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

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

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

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 $N = 2$:

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

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:

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

- Therefore, we get

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

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 $N = \lceil \frac{\log\left(2\sqrt{d}\right)}{h\alpha} \rceil$

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

Sketch of Arguments: Back to Discretization Errors

Putting everything together

- For general N:

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

- Therefore, we get

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

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

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

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

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: highly scalable
- 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 unadjusted schemes and distributions?