Complexity of sampling

Axel Böhm

July 26, 2023

Complexity of sampling

Given query access to a smooth function $V : \mathbb{R}^d \to \mathbb{R}$, what is the minimum **number of queries** required to output an **approximate sample** from the probability density $\pi \propto \exp(-V)$ on $\mathbb{R}^d$. [1]

[1] S. Chewi, *Log-concave sampling* 2022.

Complexity of sampling

Given query access to a smooth function $V : \mathbb{R}^d \to \mathbb{R}$, what is the minimum **number of queries** required to output an **approximate sample** from the probability density $\pi \propto \exp(-V)$ on $\mathbb{R}^d$. [1]

- no interest in **normalizing constant** $Z := \int \exp(-V)$
- it is difficult to compute and not necessary
- assume that V is **strongly convex**
- access to the **gradient** ∇V

[1] S. Chewi, *Log-concave sampling* 2022.

Applications

- **Bayesian statistics**

$$p_{\theta|X}(\theta|X) \propto p_{X|\theta}(X|\theta)p_{\theta}(\theta)$$

- high dimensional integration (via Monte Carlo)
- statistical physics: Boltzmann (Gibbs) distribution is proportional to $\exp(-V/T)$.

HOW?

Langevin diffusion

$$\mathrm{d}Z_t = \underbrace{-\nabla V(Z_t)\mathrm{d}t}_{\text{gradient flow}} + \underbrace{\sqrt{2}\mathrm{d}B_t}_{\text{Brownian motion}}$$

"gradient flow + noise"

HOW?

Langevin diffusion

$$\mathrm{d}Z_t = \underbrace{-\nabla V(Z_t)\mathrm{d}t}_{\text{gradient flow}} + \underbrace{\sqrt{2}\mathrm{d}B_t}_{\text{Brownian motion}}$$

"gradient flow + noise"

Pure **gradient flow**

$$\mathrm{d}Z_t = -\nabla V(Z_t)\mathrm{d}t$$

would just converge to a minimum of V.

Discretizing the Langevin diffusion

Euler-Maruyama

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

Called **Langevin Monte Carlo (LMC)**,
or **unadjusted Langevin algorithm (ULA)**

Discretizing the Langevin diffusion

Euler-Maruyama

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

Called **Langevin Monte Carlo (LMC)**,
or **unadjusted Langevin algorithm (ULA)**

Complexity

Let $X_N \sim \mu_N$. To achieve

$$d(\mu_N, \pi) \le \epsilon$$

requires $N = O(\kappa d/\epsilon^2)$.

Low-accuracy vs high-accuracy

low-accuracy

- discretizations of versions of Langevin diffusion
- are **biased** $\rightarrow$ step size needs to scale with ϵ
- best complexities $O(d^{1/3}\epsilon^{-2/3})$

high-accuracy

- dependence on ϵ is only $O(\log(1/\epsilon))$ [2]
- requires rejecting some points (based on *filters*)

[2] Y. T. Lee *et al.*, "Logsmooth gradient concentration and tighter runtimes for metropolized hamiltonian monte carlo" 2020.

Metropolis–Hastings algorithm

1. Propose a new point $Y_k \sim Q(X_{k-1}, \cdot)$
2. With probability $A(X_{k-1}, Y_k)$, set $X_k := Y_k$; otherwise, $X_k := X_{k-1}$.

A is the **acceptance probability**.

Metropolis–Hastings filter:

$$A(x, y) := 1 \wedge \frac{\pi(y) Q(y, x)}{\pi(x) Q(x, y)}$$

High-Accuracy Samplers

- **Metropolized random walk (MRW).** Take
$$Q(x,\cdot) = \mathcal{N}(x, hI_d).$$
A random walk around the state space

- **Metropolis-adjusted Langevin algorithm (MALA).** A more efficient choice
$$Q(x,\cdot) = \mathcal{N}(x - h\nabla V(x), 2hI_d).$$

- others...

High-Accuracy Samplers

- **Metropolized random walk (MRW).** Take
$$Q(x, \cdot) = \mathcal{N}(x, hI_d).$$
A random walk around the state space

- **Metropolis-adjusted Langevin algorithm (MALA).**
A more efficient choice
$$Q(x, \cdot) = \mathcal{N}(x - h\nabla V(x), 2hI_d).$$

- others...

MALA - the gold standard

Complexity
$$O(\kappa d \,\mathrm{polylog}(1/\epsilon))$$

recap high vs. low-accuracy

$$O(d^{1/3}\epsilon^{-2/3}) \quad vs. \quad O(d \log \epsilon^{-1})$$

What's up with the dependence on d??

Dimension dependence of MALA

- Complexity of MALA can be improved to $O(d^{1/2})$ after a **warmstart** [3] (because of larger steps).
  - **But(!)**, no way to find a starting point in $O(d^{1/2})$.
- Later, a $O(d)$ lower bound was established [4].
- This year, [5]: *underdamped Langevin Monte Carlo* (LMC+momentum) is the answer.

---

[3] S. Chewi *et al.*, "Optimal dimension dependence of the metropolis-adjusted langevin algorithm" 2021.

[4] Y. T. Lee *et al.*, "Lower bounds on Metropolized sampling methods for well-conditioned distributions" 2021.

[5] J. M. Altschuler *et al.*, "Faster High-Accuracy Log-Concave Sampling Via Algorithmic Warm Starts" 2023.

Dimension dependence of MALA

- Complexity of MALA can be improved to $O(d^{1/2})$ after a **warmstart** [3] (because of larger steps).
  - **But(!)**, no way to find a starting point in $O(d^{1/2})$.
- Later, a $O(d)$ lower bound was established [4].
- This year, [5]: *underdamped Langevin Monte Carlo* (LMC+momentum) is the answer.

---

[3] S. Chewi *et al.*, "Optimal dimension dependence of the metropolis-adjusted langevin algorithm" 2021.
[4] Y. T. Lee *et al.*, "Lower bounds on Metropolized sampling methods for well-conditioned distributions" 2021.
[5] J. M. Altschuler *et al.*, "Faster High-Accuracy Log-Concave Sampling Via Algorithmic Warm Starts" 2023.

Dimension dependence of MALA

- Complexity of MALA can be improved to $O(d^{1/2})$ after a **warmstart** [3] (because of larger steps).
  - **But(!)**, no way to find a starting point in $O(d^{1/2})$.
- Later, a $O(d)$ lower bound was established [4].
- This year, [5]: *underdamped Langevin Monte Carlo* (LMC+momentum) is the answer.

[3] S. Chewi *et al.*, "Optimal dimension dependence of the metropolis-adjusted langevin algorithm" 2021.
[4] Y. T. Lee *et al.*, "Lower bounds on Metropolized sampling methods for well-conditioned distributions" 2021.
[5] J. M. Altschuler *et al.*, "Faster High-Accuracy Log-Concave Sampling Via Algorithmic Warm Starts" 2023.

Thank You!