Proceedings of Machine Learning Research vol 291:1–2, 2025 38th Annual Conference on Learning Theory

On the query complexity of sampling from non-log-concave distributions

Yuchen He YUCHEN_HE@SJTU.EDU.CN
Shanghai Jiao Tong University
Chihao Zhang CHIHAO@SJTU.EDU.CN
Shanghai Jiao Tong University

Editors: Nika Haghtalab and Ankur Moitra

We study the problem of sampling from a distribution μ on $\mathbb{R}^d$ with density $p_\mu(x) \propto e^{-f_\mu(x)}$, which does not necessarily satisfy good isoperimetric conditions. To be specific, we only assume the following two properties on the target distribution.

Assumption 1 *The second moment of μ is bounded, i.e.* $\mathbf{E}_{X\sim\mu}\left[\|X\|^2\right] \le M$ *for some* $M < \infty$.

Assumption 2 *The potential function f_μ is differentiable and ∇f_μ is L-Lipschitz.*

Given query access to the value and gradients of the potential function f_μ, the task is to generate a sample from μ within as few queries as possible. We show that to compute a sample whose distribution is within an error of ε in total variation distance to the target distribution, $\left(\frac{LM}{d\varepsilon}\right)^{\Omega(d)}$ queries is inevitable.

Theorem 3 *Let $\varepsilon \in \left(0, \frac{1}{200}\right)$. For any $L, M > 0$ such that $LM \ge d$ and for any $d \ge 5$, if a sampling algorithm $\mathcal{A}$ always terminates within K queries on any target distribution μ under Assumption 1 and 2, and guarantees that the distribution of $\mathcal{A}$'s output, denoted as $\tilde{\mu}$, satisfies $\mathsf{TV}(\tilde{\mu}, \mu) \le \varepsilon$, then $K = \left(\frac{LM}{d\varepsilon}\right)^{\Omega(d)}$.*

We complement the lower bound with an algorithm requiring $\left(\frac{LM}{d\varepsilon}\right)^{O(d)}$ queries, thereby characterizing the tight (up to the constant in the exponent) query complexity for sampling from non-log-concave distributions. The formal statement of the upper bound is as follows.

Theorem 4 *Assume $d \ge 3$. There exists an algorithm $\mathcal{A}$ such that, for any distribution μ with density $p_\mu(x) \propto e^{-f_\mu(x)}$ where $f_\mu \in C^1(\mathbb{R}^d)$, $\nabla f_\mu(0) = 0$, and satisfies Assumption 1 and 2, $\mathcal{A}$ outputs a sample x with distribution $\tilde{\mu}$ satisfying $\mathsf{TV}(\mu, \tilde{\mu}) \le \varepsilon$ within $\left(\frac{LM}{d\varepsilon}\right)^{O(d)} \cdot \mathtt{poly}\left(\varepsilon^{-1}, d, L, M\right)$ queries to f_μ and ∇f_μ, for any $\varepsilon \in (0, 1)$.*

Our results are in sharp contrast with the recent work of Huang et al. (2024), where an algorithm with quasi-polynomial query complexity was proposed for sampling from a non-log-concave distribution when $M = \mathtt{poly}(d)$. Their algorithm works under the stronger condition that all distributions along the trajectory of the Ornstein-Uhlenbeck process, starting from the target distribution, are $O(1)$-log-smooth. We investigate this condition and prove that it is strictly stronger than requiring the target distribution to be $O(1)$-log-smooth. Additionally, we study this condition in the context of mixtures of Gaussians.

Finally, we place our results within the broader theme of "sampling versus optimization", as studied in Ma et al. (2019). We show that for a wide range of parameters, sampling is strictly easier than optimization by a super-exponential factor in the dimension d.[1]

1. Extended abstract. Full version appears as [arXiv:2502.06200,v3].

Acknowledgments

Chihao Zhang is supported by National Natural Science Foundation of China No. 62372289.

References

Xunpeng Huang, Difan Zou, Hanze Dong, Yi-An Ma, and Tong Zhang. Faster sampling without isoperimetry via diffusion-based Monte Carlo. In *The Annual Conference on Learning Theory*, volume 247, pages 2438–2493. PMLR, 2024.

Yi-An Ma, Yuansi Chen, Chi Jin, Nicolas Flammarion, and Michael I Jordan. Sampling can be faster than optimization. *Proceedings of the National Academy of Sciences*, 116(42):20881–20885, 2019.