The query complexity of sampling from strongly log-concave distributions in one dimension

Sinho Chewi SCHEWI@MIT.EDU
Massachusetts Institute of Technology
Patrik Gerber PRGERBER@MIT.EDU
Massachusetts Institute of Technology
Chen Lu CHENL819@MIT.EDU
Massachusetts Institute of Technology
Thibaut Le Gouic THIBAUT.LE_GOUIC@MATH.CNRS.FR
Aix Marseille Univ, Centrale Marseille, CNRS, I2M, Marseille, France
Philippe Rigollet RIGOLLET@MIT.EDU
Massachusetts Institute of Technology

Editors: Po-Ling Loh and Maxim Raginsky

Abstract

We establish the first tight lower bound of $\Omega(\log\log\kappa)$ on the query complexity of sampling from the class of strongly log-concave and log-smooth distributions with condition number κ in one dimension. Whereas existing guarantees for MCMC-based algorithms scale polynomially in κ, we introduce a novel algorithm based on rejection sampling that closes this doubly exponential gap.

Keywords: sampling, query complexity, lower bound, rejection sampling

1. Introduction

The task of sampling from a target probability distribution known up to a normalizing constant is of fundamental importance in fields such as Bayesian statistics, randomized algorithms, and online learning. Recently, there has been a resurgence of interest in sampling and its interplay with the more well-developed field of optimization. On the one hand, the extensive optimization toolkit has inspired the development of novel sampling algorithms (Bernton, 2018; Wibisono, 2018, 2019; Chewi et al., 2020; Salim et al., 2020; Ding et al., 2021a,b; Ma et al., 2021); on the other hand, the theory of optimization has motivated researchers to provide quantitative and non-asymptotic convergence guarantees for sampling methods, which depend on parameters that describe the problem complexity (e.g., the condition number and the dimension) (Durmus and Moulines, 2017; Dalalyan and Karagulyan, 2019; Chewi et al., 2021).

Conspicuously absent from this interplay, however, are lower complexity bounds for sampling, in analogy to the oracle lower bounds initiated in the seminal work by Nemirovsky and Yudin for optimization (Nemirovsky and Yudin, 1983). Besides charting the fundamental limits of optimization, such lower bounds have been instrumental in the development of faster algorithms, most notably Nesterov's acceleration, which was "found mainly because the investigating of complexity enforced to believe that such a method should exist" (Nemirovski, 1994, Ch. 10).

A canonical structured class of distributions is that of strongly log-concave and log-smooth distributions on $\mathbb{R}^d$, i.e., the class of distributions with a density $p \propto \exp(-V)$, where the potential $V : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable, α-strongly convex, and β-smooth. The relevant parameters of this class are the dimension d, as well as the *condition number* $\kappa := \beta/\alpha$, and we seek to understand the number of queries to V (and its derivatives) necessary to generate a sample close in total variation distance to p. We call a solution to this problem a *general sampling lower bound*.

© 2022 S. Chewi, P. Gerber, C. Lu, T.L. Gouic & P. Rigollet.

Related works. Despite several attempts at establishing query complexity lower bounds for sampling, we are not aware of a general sampling lower bound. Whereas sampling upper bounds are derived using techniques that are close to those employed in optimization (Dalalyan, 2017; Durmus et al., 2019), it is unclear how to use lower bound techniques for optimization (Nesterov, 2013) to derive general sampling lower bounds. Note that sampling upper bounds typically assume that the minimizer of V is known a priori; thus, a direct reduction of the sampling task to apply existing optimization lower bounds would likely capture the complexity of finding the mode of V rather than the intrinsic difficulty of the sampling task itself. In lieu of a direct reduction, it is possible to envision, at least in principle, an approach which adapts the optimization lower bound constructions to the sampling setting, but we are not aware of any successful results in this direction.

Another family of approaches is based on information-theoretic ideas which have been highly successful for developing a minimax theory of statistics (Le Cam, 1986; Le Cam and Yang, 2000; Tsybakov, 2009); however, prior works applying these ideas have largely focused on various adjacent questions which do not imply a lower bound for the sampling task itself. A notable example is the estimation of the normalizing constant of a strongly log-concave distribution, for which a lower bound was established in Ge et al. (2020). However, this lower bound does not yield a general sampling lower bound; in fact, the two problems differ in difficulty. Indeed, the randomized midpoint discretization of the underdamped Langevin dynamics (Shen and Lee, 2019) obtains samples in $\mathcal{O}(d^{1/3})$ queries, whereas the lower bound for estimating the normalizing constant in Ge et al. (2020) grows as $\widetilde{\Omega}(d)$. Another example is the paper Chatterji et al. (2022), which studies sampling with access to stochastic gradient queries. However, the resulting lower bound arises primarily out of the need to overcome the noise in the gradient queries, and it again does not yield sampling lower bounds for our setting of precise gradient queries.

To circumvent the difficulties in establishing general sampling lower bounds, various works have focused on establishing lower bounds for specific and popular algorithms such as the Underdamped Langevin Algorithm (ULA) (Cao et al., 2021) and the Metropolis-Adjusted Langevin Algorithm (MALA) (Chewi et al., 2021; Lee et al., 2021; Wu et al., 2021). In particular, the last three papers establish that the minimax query complexity for MALA over the class of strongly log-concave and log-smooth distributions is $\widetilde{\Theta}(d\kappa)$ from a "cold start" and $\Theta(\sqrt{d}\kappa)$ from a "warm start".

A lower complexity bound in one dimension. Recall that for convex optimization, there are two relevant regimes (see, e.g., Bubeck, 2015): (1) the low-dimension regime, in which algorithms such as the cutting plane method achieve the rate $\mathcal{O}(d\log(1/\varepsilon))$ (where ε is the accuracy parameter), and (2) the high-dimensional regime, in which algorithms such as gradient descent achieve dimension-free rates at the cost of inverse polynomial dependency on the accuracy. In this paper, we study the low-dimensional regime for sampling; in particular, we consider $d = 1$.

We prove that for the class of α-strongly log-concave and β-log-smooth distributions in one dimension (with mode at 0), any algorithm, which can produce a sample that is at total variation distance at most $\frac{1}{64}$ from the target distribution p (uniformly over p belonging to the class), must make at least $\Omega(\log\log\kappa)$ queries to V or any of its derivatives. To our knowledge, this is the first lower complexity bound for this problem class.

Achievability of the lower bound. The lower bound of $\Omega(\log\log\kappa)$ is surprisingly small, and existing guarantees for standard algorithms such as the Langevin algorithm (or its variants), the Metropolis-Adjusted Langevin Algorithm, or Hamiltonian Monte Carlo, all have a dependence that

scales polynomially with the condition number κ (see for instance the comparison in Shen and Lee, 2019).

To provide an algorithm which matches the lower bound, we return to the fundamental idea of rejection sampling, developed by Stan Ulam and John von Neumann (von Neumann, 1951; Eckhardt, 1987). We develop an algorithm which uses $\mathcal{O}(\log\log\kappa)$ queries in order to build a proposal distribution. Once the proposal distribution is constructed, new samples which are ε-close to p in total variation distance can be generated using $\mathcal{O}(\log(1/\varepsilon))$ additional queries per sample.

Although our algorithm is tailored to distributions in one dimension, the task of sampling from a one-dimensional log-concave distribution is an important subroutine for higher dimensional algorithms, such as the Hit-and-Run algorithm. We describe the application of our algorithm to Hit-and-Run in Section 4.

2. Lower bound

We begin by formally defining the class of strongly log-concave and log-smooth distributions in one dimension, which is the focus of this paper.

Definition 1 *The class of univariate* α-strongly log-concave and β-log-smooth *distributions, for constants* $0 < \alpha \le \beta$*, is the class of continuous distributions* p *supported on* $\mathbb{R}$*, whose density is of the form* $p(x) = \exp(-V(x))$*, for a potential function* $V : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ *which is twice continuously differentiable and satisfies*

$$\alpha \le V''(x) \le \beta\,, \qquad \forall x \in \mathbb{R}\,. \tag{1}$$

In addition, we always assume[1] *that the mode of the distribution is at* 0*, or equivalently* $V'(0) = 0$*.*

We study the *query complexity* of sampling from this class. Formally, suppose that the target distribution is $p = \exp(-V)$. The sampling algorithm is allowed to make queries to the following oracle: given a point $x \in \mathbb{R}$, the oracle returns some or all of (1) the evaluation of the potential $V(x) + C$ up to a constant C, which is unknown to the algorithm but does not change from query to query; (2) the evaluation of the gradient $V'(x)$; or (3) the evaluation of the Hessian $V''(x)$. Depending on what information the oracle returns, it may be described as providing 0^{th}-, 1^{st}-, or 2^{nd}-order information. For instance, the Langevin algorithm uses 1^{st}-order information, whereas the Metropolis-Adjusted Langevin Algorithm uses both 0^{th}-order and 1^{st}-order information. Our lower bound will in fact apply to the strongest of these oracles, namely the one that returns all three pieces of information.

We now state our lower bound.

Theorem 2 *Consider the class* $\mathcal{P}$ *of univariate* α*-strongly log-concave and* β*-log-smooth distributions as defined in Definition 1, and let* $\kappa := \beta/\alpha$ *denote the condition number. Suppose that an algorithm satisfies the following guarantee: for any* $p \in \mathcal{P}$*, the algorithm makes* n *queries to the oracle providing* 0^{th}*-,* 1^{st}*-, and* 2^{nd}*-order information for* p*, and outputs a random variable whose law is at most* $\frac{1}{64}$ *away from* p *in total variation distance. Then,* $n \gtrsim \log\log\kappa$*.*

1. This localization assumption is common in the sampling literature; without some knowledge of the mode (e.g. that the mode is contained in an interval) it is impossible to even find the mode in the query model.

We now give some intuition for the lower bound construction, and defer the proof to Appendix A. The strategy is to construct a family of distributions $\{p_1 \dots, p_m\}$ which forms a packing of the class $\mathcal{P}$ in total variation distance. Because the family is well-separated, if an algorithm can accurately sample from each p_i, it can also *identify* p_i. We construct the family $\{p_i\}_{i=1}^m$ in such a way that identifying p_i from queries to low-order oracles requires at least $\Omega(\log m)$ queries, e.g., via bisection.

With the strategy in place, we now describe motivation for the construction of the family $\{p_i\}_{i=1}^m$. Suppose that we have a distribution $p \propto \exp(-V)$ which is rescaled to satisfy $1 \le V'' \le \kappa$. The bound $V'' \ge 1$ implies that a substantial fraction of the mass of p is supported on the interval $[-1, 1]$. On the other hand, the bound $V'' \le \kappa$ allows for the density p to suddenly drop from ≈ 1 to nearly 0 over an interval of much smaller length, $\asymp 1/\sqrt{\kappa}$. Hence, as a first approximation, we can imagine dividing the interval $[-1, 1]$ into $\asymp \sqrt{\kappa}$ bins, and thinking of each p_i as piecewise constant on each bin. While keeping the log-concavity constraint in mind, for the purpose of this heuristic discussion we will consider the family $\{p_i\}_{i=1}^m$ of $m \asymp \sqrt{\kappa}$ distributions, where p_i is the uniform distribution on $[-i/\sqrt{\kappa}, i/\sqrt{\kappa}]$; see Figure 1.

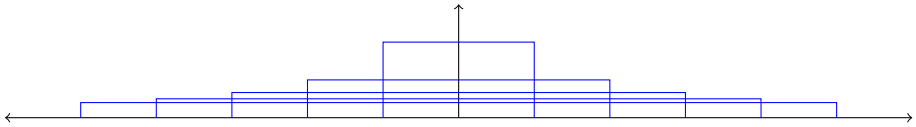

Figure 1: A family of uniform distributions.

However, this family is not well-separated in total variation distance. Indeed, it can be checked that for $i < j$, in order for the total variation distance between p_i and p_j to be appreciable, we require $j \ge 2i$. This motivates us to consider the subfamily $\{p_{2^i}, 1 \le i \le \log_2 \sqrt{\kappa}\}$, of which there are $\mathcal{O}(\log \kappa)$ elements. For this subfamily, we can hope to reduce the task of sampling to that of identifying p_{2^i} via queries, and binary search for this problem requires only $\mathcal{O}(\log \log \kappa)$ queries. This is the basis for our somewhat unusual lower bound.

The uniform distributions involved in this informal discussion do not belong to the class $\mathcal{P}$, as they are neither strongly log-concave nor log-smooth. The main technical challenge in our lower bound is to produce distributions which lie in $\mathcal{P}$ but still behaves similarly to uniform distributions, in the sense of requiring $\Omega(\log \log \kappa)$ oracle queries to identify a distribution via queries. We defer these details to the appendix.

3. Upper bound

In this section, we show that the $\Omega(\log \log \kappa)$ lower bound in the previous section is achievable. Note that the existing guarantees for standard sampling algorithms (c.f. the comparison in Shen and Lee (2019)) usually scale polynomially in the condition number κ, so they are not optimal for our setting.

Moreover, the heuristic discussion of the lower bound construction motivates choosing the query points according to a binary search strategy. In order to implement this idea, we turn towards the classical idea of rejection sampling: first, we make queries in order to construct a *proposal* distribution q. To generate new samples from p, we repeatedly draw samples from q, and each sample

is accepted with a carefully chosen acceptance probability (which can be computed via additional queries to the oracle for the density up to normalization).

Algorithm 1: ENVELOPE

Use binary search to find the first index $i_+ \in \{0, 1, \ldots, \lceil \frac{1}{2} \log_2 \kappa \rceil\}$ with $V(2^{i_+}/\sqrt{\kappa}) \geq \frac{1}{2}$.
Use binary search to find the first index $i_- \in \{0, 1, \ldots, \lceil \frac{1}{2} \log_2 \kappa \rceil\}$ with $V(-2^{i_-}/\sqrt{\kappa}) \geq \frac{1}{2}$.
Set $x_- \coloneqq -2^{i_-}/\sqrt{\kappa}$ and $x_+ \coloneqq 2^{i_+}/\sqrt{\kappa}$.
return

$$\tilde{q}(x) \coloneqq \begin{cases} \exp\Big[-\dfrac{x - x_-}{2x_-} - \dfrac{(x - x_-)^2}{2}\Big], & x \leq x_-\,, \\ 1\,, & x_- \leq x \leq x_+\,, \\ \exp\Big[-\dfrac{x - x_+}{2x_+} - \dfrac{(x - x_+)^2}{2}\Big], & x \geq x_+\,. \end{cases}$$

We give the high-level pseudocode for building an upper envelope in Algorithm 1, and for generating new samples in Algorithm 2. Note that while our lower bound applies to algorithms using 0^{th}-, 1^{st}-, and 2^{nd}-order information, our upper bound algorithm in fact only requires 0^{th}-order information. We next proceed to discuss details of the algorithms.

Before implementing Algorithm 1, we first perform several preprocessing steps. Recall that the mode of the distribution p is assumed to be at 0, and that $p \propto \exp(-V)$. We also assume that $1 \leq V'' \leq \kappa$. To reduce to this case, say we start with $\alpha \leq V'' \leq \beta$, and the bounds α, β are known. Then, observe that the rescaled potential $\bar{V}(x) \coloneqq V(x/\sqrt{\alpha})$ satisfies $1 \leq \bar{V}'' \leq \kappa = \beta/\alpha$. Given access to an oracle for V (up to additive constant), we can simulate an oracle to $\bar{V}$ (up to additive constant) and apply our algorithm to generate a sample $\bar{X}$ from the density $\bar{p} \propto \exp(-\bar{V})$; it can be checked that $\bar{X}/\sqrt{\alpha}$ is a sample from p. Finally, we assume that the oracle, when given a query point x, returns $V(x)$, where V is normalized to satisfy $V(0) = 0$; this is achieved by replacing the output $V(x)$ of the oracle by $V(x) - V(0)$.

Algorithm 2: SAMPLE

Normalize $\tilde{q}$ to form q.
while *sample is not accepted* **do**
  Sample $X \sim q$.
  Accept X w.p. $\tilde{p}(X)/\tilde{q}(X)$.
end
return X

Implementing the first step of Algorithm 1 requires performing binary search over an array of size $\mathcal{O}(\log \kappa)$, which requires only $\mathcal{O}(\log \log \kappa)$ queries; similar comments apply to the second step. We prove in Appendix B that the indices i_- and i_+ always exist under our assumptions. We prove in Appendix B that the output $\tilde{q}$ of Algorithm 1 is an *upper envelope* for the oracle, i.e., $\tilde{q} \geq \exp(-V)$. The upper envelope $\tilde{q}$ constructed in Algorithm 1 is the input to Algorithm 2; see Figure 2.

In Algorithm 2, we normalize $\tilde{q}$ to a probability distribution q, which requires computing a one-dimensional integral for the normalizing constant: $\int_{\mathbb{R}} \tilde{q}$. Once normalized, we must also be able to draw samples from the distribution q. These steps can be implemented with low computational burden, but we do not dwell on this point here because we are primarily interested in the *query complexity* in this work. Note that the steps of normalizing q and drawing new samples from q do not require additional queries to the oracle.

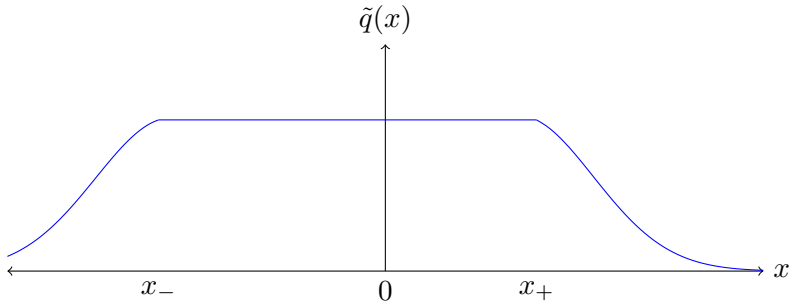

Figure 2: The upper envelope $\tilde{q}$ constructed in Algorithm 1.

The framework of rejection sampling provides a flexible guarantee: if we desire an *exact* sample from p, then we can continue drawing samples from q until one is accepted, yielding an exact sample with a guarantee on the expected total number of queries. On the other hand, if we are content with producing a sample whose law is at a fixed distance ε away from p in total variation distance, then we can force the algorithm to stop after a prespecified number of iterations, declaring failure if no sample from q is accepted, and achieve the total variation guarantee. We describe both of these guarantees in the following theorem, which summarizes the query complexity of our algorithm.

Theorem 3 *Suppose that the target distribution p belongs to the class of univariate strongly log-concave and log-smooth distributions (Theorem 1). Algorithm 1 uses $\mathcal{O}(\log\log\kappa)$ queries to build the upper envelope $\tilde{q}$. Once $\tilde{q}$ is constructed, we can use it for either of the following tasks.*

1. *(exact sampling) Algorithm 2 returns an exact sample from p after an additional $\mathcal{O}(1)$ expected queries to the oracle.*

2. *(approximate sampling) Fix an accuracy parameter $0 < \varepsilon < 1$. If we limit Algorithm 2 to use at most $\mathcal{O}(\log(1/\varepsilon))$ queries, then the output of Algorithm 2 (or 'FAILURE', if Algorithm 2 fails to accept a sample within the allowed number of queries) has a distribution which is at total variation distance at most ε away from p.*

We give the proof in Appendix B.

4. Application to Hit-and-Run

Although our upper bound algorithm applies only to univariate distributions, in this section we demonstrate how to use our algorithm as a subroutine for the well-known Hit-and-Run algorithm for sampling from high-dimensional log-concave distributions.

Hit-and-Run algorithms form a class of Markov chain Monte Carlo methods (Bélisle et al., 1993). Given a target distribution with density $p : \mathbb{R}^d \to \mathbb{R}$ and a distribution ν on the sphere $\mathbb{S}^{d-1}$, one can construct a Markov chain with stationary distribution p as follows. If the current point of the Markov chain is at the point $x_t \in \mathbb{R}^d$ at iteration t, we

1. choose a random direction $u \sim \nu$,

2. choose a random step size λ from the distribution with density $\propto p(x_t + \lambda u)$,

3. and set $x_{t+1} = x_t + \lambda u$.

It is not hard to see that the above dynamics are reversible with respect to p. Lovász and Vempala (Lovász, 1999; Lovász and Vempala, 2003) study this algorithm when p is log-concave, and establish rapid mixing of the chain when the direction distribution is uniform. However, they only partially addressed the issue of implementation, suggesting the use of an approximate algorithm for Step 2. In the case when p is strongly log-concave and log-smooth, our methods allow us to resolve this issue and provide an efficient and exact sampling algorithm.

Proposition 4 *If p is 1-strongly log-concave and κ-log-smooth, then Step 2 can be implemented exactly, with amortized query and time complexity $\mathcal{O}(\log(\kappa d))$.*

We defer the description of the algorithm, and the proof of Proposition 4 to Appendix C. We remark that the query complexity is actually dominated by the cost of finding the maximum of p when restricted to the line, and the sampling step is comparatively cheap.

5. Conclusion and outlook

In this paper, we established the oracle complexity of sampling from the class of univariate strongly log-concave and log-smooth distributions, in analogy with the now pervasive oracle lower bounds for optimization initiated by Nemirovsky and Yudin (Nemirovsky and Yudin, 1983). A clear future direction suggested by this work is to extend this result to higher dimensions, and to ultimately develop a theory of lower complexity bounds and optimal algorithms for sampling.

Recently, an intense amount of research has been devoted to the use of Markov chain Monte Carlo-based methods for sampling, and it may come as a surprise that the complexity lower bound we have proven in this paper is attained by an entirely different type of algorithm, namely rejection sampling. Our result highlights that standard algorithms may not be optimal, and that the search for optimal algorithms goes hand-in-hand with lower bound constructions.

In particular, our work motivates revisiting the idea of rejection sampling through the modern lens of minimax optimality.

Acknowledgments

Sinho Chewi was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. Thibaut Le Gouic was supported by NSF award IIS-1838071. Philippe Rigollet was supported by NSF awards IIS-1838071, DMS-1712596, and DMS-2022448.

References

Claude JP Bélisle, H Edwin Romeijn, and Robert L Smith. Hit-and-run algorithms for generating multivariate distributions. *Mathematics of Operations Research*, 18(2):255–266, 1993.

Espen Bernton. Langevin Monte Carlo and JKO splitting. In *Conference on Learning Theory*, pages 1777–1798. PMLR, 2018.

Sébastien Bubeck. Convex optimization: algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

Yu Cao, Jianfeng Lu, and Lihan Wang. Complexity of randomized algorithms for underdamped Langevin dynamics. *Commun. Math. Sci.*, 19(7):1827–1853, 2021.

Niladri S. Chatterji, Peter L. Bartlett, and Philip M. Long. Oracle lower bounds for stochastic gradient sampling algorithms. *Bernoulli*, 28(2):1074 – 1092, 2022.

Sinho Chewi, Thibaut Le Gouic, Chen Lu, Tyler Maunu, Philippe Rigollet, and Austin J. Stromme. Exponential ergodicity of mirror-Langevin diffusions. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 19573–19585. Curran Associates, Inc., 2020.

Sinho Chewi, Chen Lu, Kwangjun Ahn, Xiang Cheng, Thibaut Le Gouic, and Philippe Rigollet. Optimal dimension dependence of the Metropolis-adjusted Langevin algorithm. In Mikhail Belkin and Samory Kpotufe, editors, *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 1260–1300. PMLR, 15–19 Aug 2021.

Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.

Arnak S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(3):651–676, 2017.

Arnak S. Dalalyan and Avetik Karagulyan. User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Process. Appl.*, 129(12):5278–5311, 2019.

Zhiyan Ding, Qin Li, Jianfeng Lu, and Stephen J. Wright. Random coordinate Langevin Monte Carlo. In Mikhail Belkin and Samory Kpotufe, editors, *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 1683–1710. PMLR, 15–19 Aug 2021a.

Zhiyan Ding, Qin Li, Jianfeng Lu, and Stephen J. Wright. Random coordinate underdamped Langevin Monte Carlo. In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 2701–2709. PMLR, 13–15 Apr 2021b.

Alain Durmus and Éric Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, 27(3):1551–1587, 2017.

Alain Durmus, Szymon Majewski, and Bł ażej Miasojedow. Analysis of Langevin Monte Carlo via convex optimization. *J. Mach. Learn. Res.*, 20:Paper No. 73, 46, 2019.

Roger Eckhardt. Stan Ulam, John von Neumann, and the Monte Carlo method. *Los Alamos Science*, 15:131–136, 1987.

Rong Ge, Holden Lee, and Jianfeng Lu. Estimating normalizing constants for log-concave distributions: Algorithms and lower bounds. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 579–586, 2020.

Robert D. Gordon. Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *The Annals of Mathematical Statistics*, 12(3):364 – 366, 1941.

Lucien Le Cam. *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer-Verlag, New York, 1986.

Lucien Le Cam and Grace Lo Yang. *Asymptotics in statistics*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 2000. Some basic concepts.

Yin Tat Lee, Ruoqi Shen, and Kevin Tian. Lower bounds on Metropolized sampling methods for well-conditioned distributions. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 18812–18824. Curran Associates, Inc., 2021.

László Lovász. Hit-and-run mixes fast. *Mathematical Programming*, 86(3):443–461, 1999.

László Lovász and Santosh Vempala. Hit-and-run is fast and fun. *preprint, Microsoft Research*, 2003.

Yi-An Ma, Niladri S. Chatterji, Xiang Cheng, Nicolas Flammarion, Peter L. Bartlett, and Michael I. Jordan. Is there an analog of Nesterov acceleration for gradient-based MCMC? *Bernoulli*, 27(3): 1942 – 1992, 2021.

Arkadi Nemirovski. Efficient methods in convex programming. Lecture Notes (Tecehnion - Georgia Tech), 1994.

A. S. Nemirovsky and D. B. and Yudin. *Problem complexity and method efficiency in optimization*. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1983. Translated from the Russian and with a preface by E. R. Dawson, Wiley-Interscience Series in Discrete Mathematics.

Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

Adil Salim, Anna Korba, and Giulia Luise. The Wasserstein proximal gradient algorithm. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12356–12366. Curran Associates, Inc., 2020.

Ruoqi Shen and Yin Tat Lee. The randomized midpoint method for log-concave sampling. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

John von Neumann. Various techniques used in connection with random digits. In A. S. Householder, G. E. Forsythe, and H. H. Germond, editors, *Monte Carlo Method*, volume 12 of *National Bureau of Standards Applied Mathematics Series*, chapter 13, pages 36–38. US Government Printing Office, Washington, DC, 1951.

Andre Wibisono. Sampling as optimization in the space of measures: The Langevin dynamics as a composite optimization problem. In Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet, editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 2093–3027. PMLR, 06–09 Jul 2018.

Andre Wibisono. Proximal Langevin algorithm: rapid convergence under isoperimetry. *arXiv e-prints*, art. arXiv:1911.01469, 2019.

Keru Wu, Scott Schmidler, and Yuansi Chen. Minimax mixing time of the Metropolis-adjusted Langevin algorithm for log-concave sampling. *arXiv e-prints*, art. arXiv:2109.13055, 2021.

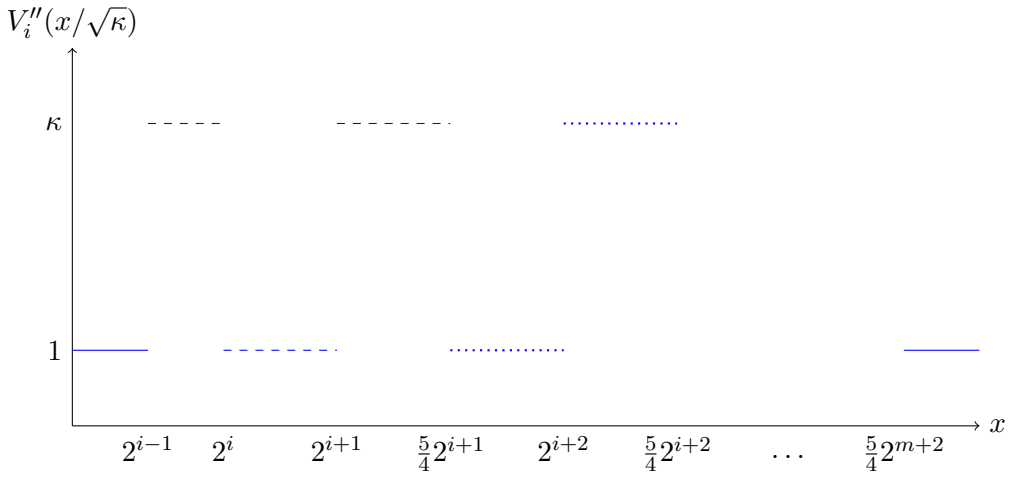

Figure 3: The dashed lines correspond to ϕ and the dotted lines correspond to ψ.

Appendix A. Proof of the lower bound

A.1. The construction

Let m be the largest integer such that

$$\exp\Big(-\frac{2^{2m-2}}{2\kappa}\Big) \geq \frac{1}{2}\,. \tag{2}$$

Define two auxiliary functions

$$\phi(x) \coloneqq \begin{cases} \kappa\,, & 1/2 \leq x < 1\,, \\ 1\,, & 1 \leq x < 2\,, \\ \kappa\,, & 2 \leq x < 5/2\,, \\ 0 & \text{otherwise}\,, \end{cases} \qquad \psi(x) \coloneqq \begin{cases} 1\,, & 5/2 \leq x < 4\,, \\ \kappa\,, & 4 \leq x < 5\,, \\ 0\,, & \text{otherwise}\,. \end{cases}$$

We define a family $(V_i)_{i=1}^m$ of 1-strongly convex and κ-smooth potentials as follows. We require that $V_i(0) = V_i'(0) = 0$ and that V_i be an even function, so it suffices to specify V_i'' on $\mathbb{R}_+$. The

second derivative is given by

$$V_i''(x) \coloneqq \mathbb{1}\{x \le \kappa^{-\frac{1}{2}} 2^{i-1}\} + \phi\big(\frac{x}{\kappa^{-\frac{1}{2}} 2^i}\big) + \sum_{j=i}^{m-1} \psi\big(\frac{x}{\kappa^{-\frac{1}{2}} 2^j}\big) + \mathbb{1}\{x \ge 5\kappa^{-\frac{1}{2}} 2^{m-1}\}\,, \qquad x \ge 0\,.$$

Observe that all of the terms in the above summation have disjoint supports, see Figure 3.

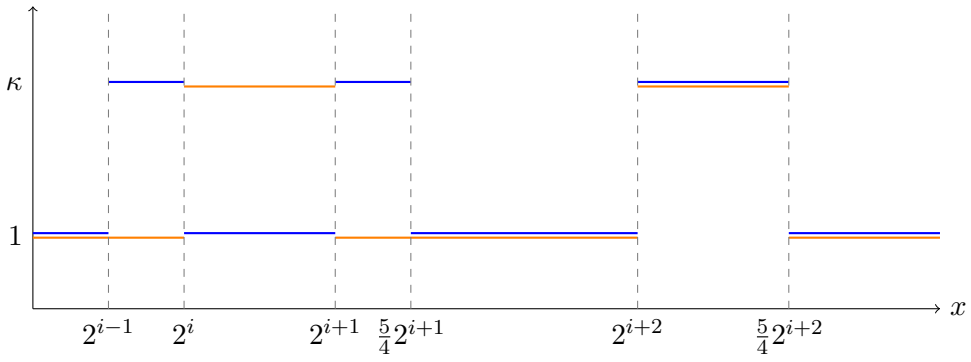

Figure 4: We plot V_i'' (in blue) and V_{i+1}'' (in orange). In this figure, we do not distort the horizontal axis lengths to make it easier to visually compare the relative lengths of intervals on which the second derivatives are constant.

The following lemma provides intuition for the construction.

Lemma 5 *We have the equalities*

$$\begin{aligned} V_i &= V_{i+1}\,, \\ V_i' &= V_{i+1}'\,, \\ V_i'' &= V_{i+1}''\,, \end{aligned}$$

outside of the set $\{x \in \mathbb{R} : \kappa^{-\frac{1}{2}} 2^{i-1} \le |x| \le \frac{5}{4}\kappa^{-\frac{1}{2}} 2^{i+1}\}$.

Proof Refer to Figure 4 for a visual aid for the proof.

Clearly the potentials and derivatives match when $|x| \le \kappa^{-\frac{1}{2}} 2^{i-1}$. Since the second derivatives match when $|x| \ge \frac{5}{4}\kappa^{-\frac{1}{2}} 2^{i+1}$, it suffices to show that $V_i'(\frac{5}{4}\kappa^{-\frac{1}{2}} 2^{i+1}) = V_{i+1}'(\frac{5}{4}\kappa^{-\frac{1}{2}} 2^{i+1})$ and $V_i(\frac{5}{4}\kappa^{-\frac{1}{2}} 2^{i+1}) = V_{i+1}(\frac{5}{4}\kappa^{-\frac{1}{2}} 2^{i+1})$.

To that end, note that for $x \ge 0$,

$$\begin{aligned} V_{i+1}''(x) - V_i''(x) &= \mathbb{1}\{\kappa^{-\frac{1}{2}} 2^{i-1} < x \le \kappa^{-\frac{1}{2}} 2^i\} - \phi\big(\frac{x}{\kappa^{-\frac{1}{2}} 2^i}\big) + \phi\big(\frac{x}{\kappa^{-\frac{1}{2}} 2^{i+1}}\big) - \psi\big(\frac{x}{\kappa^{-\frac{1}{2}} 2^i}\big) \\ &= \begin{cases} -(\kappa - 1)\,, & \kappa^{-\frac{1}{2}} 2^{i-1} \le x \le \kappa^{-\frac{1}{2}} 2^i\,, \\ +(\kappa - 1)\,, & \kappa^{-\frac{1}{2}} 2^i \le x \le \kappa^{-\frac{1}{2}} 2^{i+1}\,, \\ -(\kappa - 1)\,, & \kappa^{-\frac{1}{2}} 2^{i+1} \le x \le \frac{5}{4}\kappa^{-\frac{1}{2}} 2^{i+1}\,, \\ 0\,, & \text{otherwise}\,. \end{cases} \end{aligned}$$

A little algebra shows that the above expression integrates to zero, hence we deduce the equality $V_i'(\frac{5}{4}\kappa^{-\frac{1}{2}}2^{i+1}) = V_{i+1}'(\frac{5}{4}\kappa^{-\frac{1}{2}}2^{i+1})$. Also, by integrating this expression twice, we see that

$$\begin{aligned} V_{i+1}\big(\frac{5}{4}\kappa^{-\frac{1}{2}}2^{i+1}\big) - V_i\big(\frac{5}{4}\kappa^{-\frac{1}{2}}2^{i+1}\big) = \; & \underbrace{-\frac{\kappa-1}{2}\left(\kappa^{-\frac{1}{2}}2^{i-1}\right)^2}_{\text{integral on } [\kappa^{-\frac{1}{2}}2^{i-1},\kappa^{-\frac{1}{2}}2^{i}]} \\ & \underbrace{-(\kappa-1)\kappa^{-\frac{1}{2}}2^{i-1}\,\kappa^{-\frac{1}{2}}2^{i} + \frac{\kappa-1}{2}\left(\kappa^{-\frac{1}{2}}2^{i}\right)^2}_{\text{integral on } [\kappa^{-\frac{1}{2}}2^{i},\kappa^{-\frac{1}{2}}2^{i+1}]} \\ & \underbrace{+(\kappa-1)\kappa^{-\frac{1}{2}}2^{i-1}\,\frac{1}{4}\kappa^{-\frac{1}{2}}2^{i+1} - \frac{\kappa-1}{2}\left(\frac{1}{4}\kappa^{-\frac{1}{2}}2^{i+1}\right)^2}_{\text{integral on } [\kappa^{-\frac{1}{2}}2^{i+1},\frac{5}{4}\kappa^{-\frac{1}{2}}2^{i+1}]} \\ = \; & \frac{\kappa-1}{\kappa}\left\{-2^{2i-3} - 2^{2i-1} + 2^{2i-1} + 2^{2i-2} - 2^{2i-3}\right\} \\ = \; & 0\,, \end{aligned}$$

as desired. ■

We also need a lemma showing that each probability distribution $p_i \propto \exp(-V_i)$ places a substantial amount of mass on the interval $(\kappa^{-\frac{1}{2}}2^{i-2}, \kappa^{-\frac{1}{2}}2^{i-1}]$.

Lemma 6 *For each $i \in [m]$,*

$$p_i\big((\kappa^{-\frac{1}{2}}2^{i-2}, \kappa^{-\frac{1}{2}}2^{i-1}]\big) \geq \frac{1}{32}\,.$$

Proof According to the definition of p_i, we have

$$p_i\big((\kappa^{-\frac{1}{2}}2^{i-2}, \kappa^{-\frac{1}{2}}2^{i-1}]\big) = \frac{\int_{\kappa^{-\frac{1}{2}}2^{i-2}}^{\kappa^{-\frac{1}{2}}2^{i-1}} \exp(-x^2/2)\,\mathrm{d}x}{Z_{p_i}}\,, \qquad Z_{p_i} := \int_{\mathbb{R}} \exp(-V_i)\,.$$

Recalling that m is chosen so that $\exp(-x^2/2) \geq 1/2$ whenever $|x| \leq \kappa^{-\frac{1}{2}}2^{m-1}$ (see (2)), we can conclude that

$$\int_{\kappa^{-\frac{1}{2}}2^{i-2}}^{\kappa^{-\frac{1}{2}}2^{i-1}} \exp\big(-\frac{x^2}{2}\big)\,\mathrm{d}x \geq \frac{1}{2}\,\kappa^{-\frac{1}{2}}2^{i-2}\,.$$

For the normalizing constant, observe that

$$\int_0^\infty \exp(-V_i) = \int_0^{\kappa^{-\frac{1}{2}}2^{i}} \exp(-V_i) + \int_{\kappa^{-\frac{1}{2}}2^{i}}^\infty \exp(-V_i) \leq \kappa^{-\frac{1}{2}}2^{i} + \int_{\kappa^{-\frac{1}{2}}2^{i}}^\infty \exp(-V_i)\,.$$

Since $V_i'' = \kappa$ on $[\kappa^{-\frac{1}{2}}2^{i-1}, \kappa^{-\frac{1}{2}}2^{i}]$, it follows that $V_i'(\kappa^{-\frac{1}{2}}2^{i}) \geq \kappa^{\frac{1}{2}}2^{i-1}$, and so

$$V_i(x) \geq \kappa^{\frac{1}{2}}2^{i-1}\left(x - \kappa^{-\frac{1}{2}}2^{i}\right) + \frac{\left(x - \kappa^{-\frac{1}{2}}2^{i}\right)^2}{2}\,, \qquad x \geq \kappa^{-\frac{1}{2}}2^{i}\,.$$

Therefore,

$$\int_{\kappa^{-\frac{1}{2}}2^i}^{\infty} \exp(-V_i) \leq \int_{\kappa^{-\frac{1}{2}}2^i}^{\infty} \exp\bigl(-\kappa^{\frac{1}{2}}2^{i-1}\,(x-\kappa^{-\frac{1}{2}}2^i) - \frac{(x-\kappa^{-\frac{1}{2}}2^i)^2}{2}\bigr)\,\mathrm{d}x \leq \frac{1}{\kappa^{\frac{1}{2}}2^{i-1}} \leq \frac{1}{\sqrt{\kappa}}\,,$$

where we applied a standard tail estimate for Gaussian densities (Theorem 8). Putting it together,

$$p_i\bigl((\kappa^{-\frac{1}{2}}2^{i-2}, \kappa^{-\frac{1}{2}}2^{i-1}]\bigr) \geq \frac{2^{i-3}}{2\,(2^i+1)} \geq \frac{1}{32}\,,$$

which proves the result. ∎

A.2. Lower bound via Fano's inequality

In this section, we use the densities $\{p_i\}_{i=1}^m$ constructed in the previous section together with Fano's inequality from information theory in order to prove the lower bound.

Proof [Proof of Theorem 2] Let $Z \sim \mathrm{unif}([m])$ be an index chosen uniformly at random. Suppose that an algorithm makes n queries to the oracle for p_Z, and given $Z = i$, outputs a sample Y whose law q_i is at total variation distance at most $\frac{1}{64}$ from p_i. In light of Lemma 6, a good candidate estimator for Z from the observation of Y is given by

$$\widehat{Z} := \{k \in \mathbb{N} : Y \in (\kappa^{-\frac{1}{2}}2^{k-2}, \kappa^{-\frac{1}{2}}2^{k-1}]\}\,.$$

On the one hand, the probability that the estimator is correct is bounded by

$$\begin{aligned}\mathbb{P}\{\widehat{Z} = Z\} &= \frac{1}{m}\sum_{i=1}^m \mathbb{P}\{\widehat{Z} = i \mid Z = i\} = \frac{1}{m}\sum_{i=1}^m \mathbb{P}\{Y \in (\kappa^{-\frac{1}{2}}2^{i-2}, \kappa^{-\frac{1}{2}}2^{i-1}] \mid Z = i\} \\ &= \frac{1}{m}\sum_{i=1}^m q_i\bigl((\kappa^{-\frac{1}{2}}2^{i-2}, \kappa^{-\frac{1}{2}}2^{i-1}]\bigr) \geq \frac{1}{m}\sum_{i=1}^m p_i\bigl((\kappa^{-\frac{1}{2}}2^{i-2}, \kappa^{-\frac{1}{2}}2^{i-1}]\bigr) - \frac{1}{64} \geq \frac{1}{64}\,,\end{aligned} \tag{3}$$

where the last inequality uses Lemma 6.

On the other hand, we can lower bound $\mathbb{P}\{\widehat{Z} \neq Z\}$ using Fano's inequality. Let $x_1, \ldots, x_n$ denote the query points of the algorithm, and let W_i be a shorthand for the triple (V_i, V_i', V_i''). We will first prove the lower bound for deterministic algorithms, i.e., assuming that each query point x_j is a deterministic function of the previous query points and query values. Since

$$Z \to \{x_j, W_Z(x_j),\ j \in [n]\} \to \widehat{Z}$$

forms a Markov chain, Fano's inequality (Cover and Thomas, 2006) yields

$$\mathbb{P}\{\widehat{Z} \neq Z\} \geq 1 - \frac{I(\{x_j, W_Z(x_j)\}_{j\in[n]}; Z) + \log 2}{\log m}\,,$$

where I denotes the mutual information. By the chain rule for mutual information (Cover and Thomas, 2006),

$$I\bigl(\{x_j, W_Z(x_j)\}_{j\in[n]}; Z\bigr) = \sum_{j=1}^n I\bigl(x_j, W_Z(x_j); Z \bigm| x_1, W_Z(x_1), \ldots, x_{j-1}, W_Z(x_{j-1})\bigr)\,.$$

Observe that, conditioned on $\{x_i, W_Z(x_i)\}_{i=1}^{j-1}$, the query point x_j is deterministic. Also, from the construction of the family of potentials, we know that $W_Z(x_j) = W_1(x_j)$ if $x_j \le \kappa^{-\frac{1}{2}} 2^{Z-1}$, and $W_Z(x_j) = W_m(x_j)$ if $x_j \ge \frac{5}{4}\kappa^{-\frac{1}{2}} 2^{Z+1}$. It yields that:

- for $Z \le \log_2(\frac{4}{5}\sqrt{\kappa}x_j) - 1$, $W_Z(x_j)$ takes a unique value given by $W_m(x_j)$,
- for $Z \ge \log_2(\sqrt{\kappa}x_j) + 1$, $W_Z(x_j)$ takes a unique value given by $W_1(x_j)$,

and otherwise, Z lives in an interval of size at most $\log_2(\sqrt{\kappa}x_j) + 1 - (\log_2(\frac{4}{5}\sqrt{\kappa}x_j) - 1) \le 2 + \log_2(5/4)$ which covers at most three integers, say $z_0 - 1, z_0, z_0 + 1$. Hence, the conditional distribution of $W_Z(x_j)$ can be supported on at most 5 points given respectively by

$$W_1(x_j), W_m(x_j), W_{z_0-1}(x_j), W_{z_0}(x_j), \text{ and } W_{z_0+1}(x_j)\,.$$

Since the mutual information is upper bounded by the conditional entropy of $W_Z(x_j)$, we can conclude

$$I\big(\{x_j, W_Z(x_j)\}_{j\in[n]}; Z\big) \le n \log 5\,.$$

Substituting this into Fano's inequality yields

$$\mathbb{P}\{\widehat{Z} \ne Z\} \ge 1 - \frac{n\log 5 + \log 2}{\log m}\,. \tag{4}$$

In general, if the algorithm is randomized, then we can apply the inequality (4) conditioned on the random seed ξ of the algorithm, since ξ is independent of Z. It yields

$$\mathbb{P}\{\widehat{Z} \ne Z \mid \xi\} \ge 1 - \frac{n\log 5 + \log 2}{\log m}\,,$$

and upon taking expectations we see that (4) holds for randomized algorithms as well.

Combined with (3), we obtain $n \gtrsim \log m \gtrsim \log\log\kappa$ as desired. ■

Appendix B. Proof of the upper bound

Let p be the target distribution and let $\tilde{p} = pZ_p$ denote the unnormalized distribution which we access via oracle queries. We recall our preprocessing steps: we assume that the query values take the form $\tilde{p}(x) = \exp(-V(x))$, with $V(0) = V'(0) = 0$ and V satisfying (1). This is without loss of generality because we can query $\tilde{p}(0)$ and replace subsequent queries $\tilde{p}(x)$ with $\tilde{p}(x)/\tilde{p}(0)$, thereby normalizing V to satisfy $V(0) = 0$. By rescaling the distribution, we can assume that $1 \le V'' \le \kappa$. Also, we can assume that the target distribution is only supported on the positive reals $\mathbb{R}_+$, because we can then construct an upper envelope on all of $\mathbb{R}$ by repeating our algorithm on the negative reals, which only doubles the number of queries and does not change the complexity.

Proof [Proof of Theorem 3] Our goal is to use the oracle queries to construct an upper envelope $\tilde{q}$ that satisfies $\tilde{q} \ge \tilde{p}$, and $Z_q \lesssim Z_p$, where

$$Z_p := \int_{\mathbb{R}} \tilde{p}\,, \qquad Z_q := \int_{\mathbb{R}} \tilde{q}$$

are the normalizing constants. The guarantees of Theorem 3 will then follow from standard results on rejection sampling. For completeness, we state and prove the relevant result as Theorem 7 in Appendix D.

Let i_0 denote the smallest integer such that $V(2^{i_0}/\sqrt{\kappa}) \geq 1/2$. Note that $x^2/2 \leq V(x) \leq \kappa x^2/2$ implies that $0 \leq i_0 \leq (\log_2 \kappa)/2$. Using binary search over an array of size $\mathcal{O}(\log \kappa)$, we can find i_0 using only $\mathcal{O}(\log\log \kappa)$ queries to $\tilde{p}$.

Let $x_0 := 2^{i_0}/\sqrt{\kappa}$. We first claim that

$$\int_0^{x_0} \tilde{p} \gtrsim x_0 \,. \tag{5}$$

When $i_0 = 0$, this holds because

$$\int_0^{x_0} \tilde{p} = \int_0^{1/\sqrt{\kappa}} \exp(-V) \geq \int_0^{1/\sqrt{\kappa}} \exp\bigl(-\frac{\kappa x^2}{2}\bigr)\,\mathrm{d}x \geq \frac{1}{3\sqrt{\kappa}} = \frac{x_0}{3}\,.$$

When $i_0 > 0$, this holds because, by definition of i_0, we have $V(x_0/2) \leq 1/2$, and so

$$\int_0^{x_0} \tilde{p} \geq \int_0^{x_0/2} \exp(-V) \gtrsim x_0\,.$$

Next, define the upper envelope as follows:

$$\tilde{q}(x) = \begin{cases} 1\,, & x \leq x_0\,, \\ \exp\{-(x-x_0)/(2x_0) - (x-x_0)^2/2\}\,, & x > x_0\,. \end{cases}$$

To see that $\tilde{q} \geq \tilde{p}$ and hence that $\tilde{q}$ is a valid upper envelope, observe first that since $\tilde{p}(0) = 1$, and $\tilde{p}$ is decreasing, we get that $\tilde{p}(x) \leq 1 = \tilde{q}(x)$ for all $x \in [0, x_0]$.

Next, if $x > x_0$, using the fact that V is convex and $V(x_0) \geq 1/2$ by the definition of x_0,

$$V'(x_0) \geq \frac{V(x_0) - V(0)}{x_0} \geq \frac{1}{2x_0}\,.$$

Hence, for any $x > x_0$ we have

$$\begin{aligned} V(x) &\geq V(x_0) + V'(x_0)\,(x - x_0) + \frac{1}{2}\,(x-x_0)^2 \\ &\geq \frac{1}{2x_0}\,(x-x_0) + \frac{1}{2}\,(x-x_0)^2\,. \end{aligned}$$

It implies that $\tilde{p}(x) \leq \tilde{q}(x)$ also for the tail $x > x_0$.

To complete the proof, we show that $Z_q \lesssim Z_p$. In light of (5) it is sufficient to show that $Z_q \lesssim x_0$. To see this, observe that by Theorem 8, we have

$$Z_q = \int_0^{x_0} \tilde{q} + \int_{x_0}^{\infty} \tilde{q} \leq x_0 + \int_{x_0}^{\infty} \exp\bigl(-\frac{1}{2x_0}\,(x-x_0) - \frac{1}{2}\,(x-x_0)^2\bigr)\,\mathrm{d}x \leq 3x_0\,.$$

This completes the proof. ∎

Appendix C. Proof for the hit-and-run algorithm

In this section, we prove Proposition 4.

Let $p \propto \exp(-V)$ where $V : \mathbb{R}^d \to \mathbb{R}$ is 1-strongly convex and κ-smooth with $V(0) = \nabla V(0) = 0$. We are given a line $\ell = \{x^\star + \lambda u : \lambda \in \mathbb{R}\}$ where $x^\star$ is the closest point of ℓ to the origin and u is a unit vector. Note that

$$1 \le u^\mathsf{T} \nabla^2 V(x^\star + \lambda u) u = \frac{\mathrm{d}^2 V(x^\star + \lambda u)}{\mathrm{d}\lambda^2} \le \kappa$$

so that V restricted to ℓ is also 1-strongly convex and κ-smooth.

First, we need to find the minimum of V on ℓ. To that end, we restrict the line ℓ to the subset of points $x^\star + \lambda u$ with

$$|\lambda| \le 2\kappa \, |x^\star|_2 \,, \tag{6}$$

where we write $|\cdot|_2$ for the Euclidean norm on $\mathbb{R}^d$.

To prove (6), note that by strong convexity of V on ℓ, if $x^\star + \lambda u$ is a minimizer of V on ℓ, then we must have

$$\lambda \nabla V(x^\star)^\mathsf{T} u + \frac{1}{2}\lambda^2 \le V(x^\star + \lambda u) - V(x^\star) \le 0 \,.$$

The above yields $|\lambda| \le 2\,|\nabla V(x^\star)^\mathsf{T} u|$. Moreover, by κ-smoothness of V we have

$$|\nabla V(x^\star)^\mathsf{T} u| \le |\nabla V(x^\star)|_2 \le \kappa |x^\star|_2 \,,$$

which completes the proof of (6).

To ease notation, from here on we write V for its one-dimensional restriction to ℓ. We modify our rejection sampling algorithm from Theorem 3 so that it works when (i) the minimizer $m \in \mathbb{R}$ of V is only known approximately and (ii) the value $V(m)$ is known only approximately. To make this precise, suppose that we know $a, b \in \mathbb{R}$ such that $a \le m \le b$. By (6), such a, b can be found using $\mathcal{O}(\log(\kappa\,|x^\star|_2/(b-a)))$ queries via bisection. Strong convexity and smoothness give us

$$V(m) + \frac{1}{2}(b-m)^2 \le V(b) \le V(m) + \frac{\kappa}{2}(b-m)^2 \,, \tag{7}$$

$$V(m) + \frac{1}{2}(m-a)^2 \le V(a) \le V(m) + \frac{\kappa}{2}(m-a)^2 \,. \tag{8}$$

In particular,

$$V(a) \vee V(b) - \frac{\kappa}{2}(b-a)^2 \le V(m) \le V(a) \wedge V(b) \,.$$

Relabel $V - (V(a) \vee V(b))$ as V, so that we may assume

$$-\frac{\kappa}{2}(b-a)^2 \le V(m) \le 0 \,.$$

Assume from here on that $\kappa\,(b-a)^2/2 = 1$ (this means we used $\mathcal{O}(\log(\kappa\,|x^\star|_2))$ queries). Define

$$i_b \coloneqq \min\big\{ i \in \mathbb{N} : V\big(b + \frac{2^i}{\sqrt{\kappa}}\big) \ge 3 \big\} \,,$$

$$i_a \coloneqq \min\big\{ i \in \mathbb{N} : V\big(a - \frac{2^i}{\sqrt{\kappa}}\big) \ge 3 \big\} \,.$$

Let $x_b = b + 2^{i_b}/\sqrt{\kappa}$ and $x_a = a - 2^{i_a}/\sqrt{\kappa}$. Note that $1 \le i_a, i_b \lesssim \log \kappa$. Indeed, for the lower bound observe that by (7), we have

$$V\big(b + \frac{1}{\sqrt{\kappa}}\big) \le V(m) + \frac{\kappa}{2}\big(b + \frac{1}{\sqrt{\kappa}} - a\big)^2 \le 0 + 1 + \frac{1}{2} + \sqrt{2} < 3\,,$$

so that $i_b \ge 1$.

For the upper bound, we have

$$V\big(b + \frac{2^{\frac{1}{2}(\log_2 \kappa + 3)}}{\sqrt{\kappa}}\big) \ge V(m) + \frac{1}{2}\big(b + \frac{2^{\frac{1}{2}(\log_2 \kappa + 3)}}{\sqrt{\kappa}} - m\big)^2 \ge -1 + \frac{2^{\log_2 \kappa + 2}}{\kappa} = 3\,.$$

It yields $i_b \lesssim \log \kappa$. Similarly (8) yields the desired bounds for i_a.

Thus, i_a, i_b may be found via binary search in $\mathcal{O}(\log\log\kappa)$ queries to $\tilde{p}$.

We now move to the definition of the upper envelope $\tilde{q}$. To that end, note first that

$$V'(x_b) \ge \frac{V(x_b) - V(m)}{x_b - m} \ge \frac{3}{x_b - a}$$

so that for $x \ge x_b$ we can write

$$\begin{aligned} V(x) &\ge V(x_b) + V'(x_b)(x - x_b) + \frac{1}{2}(x - x_b)^2 \\ &\ge 3 + 3\,\frac{x - x_b}{x_b - a} + \frac{1}{2}(x - x_b)^2\,. \end{aligned}$$

Similarly, we also have

$$V(x) \ge 3 + 3\,\frac{x - x_a}{x_a - b} + \frac{1}{2}(x - x_a)^2$$

for $x \le x_a$. We are ready for the definition of $\tilde{q}$.

$$\tilde{q}(x); = \begin{cases} e\,, & \text{for } x \in [x_a, x_b]\,, \\ \exp(-3 - 3\,\frac{x - x_b}{x_b - a} - \frac{1}{2}(x - x_b)^2)\,, & \text{for } x \ge x_b\,, \\ \exp(-3 - 3\,\frac{x - x_a}{x_a - b} - \frac{1}{2}(x - x_a)^2)\,, & \text{for } x \le x_a\,. \end{cases}$$

By our above calculations we know that $\tilde{p} \le \tilde{q}$ on all of $\mathbb{R}$. We further have

$$Z_p \ge \int_{x_a}^{x_b} \tilde{p} \ge \int_{a - 2^{i_a - 1}/\sqrt{\kappa}}^{b + 2^{i_b - 1}/\sqrt{\kappa}} \exp(-V) \ge \frac{\exp(-3)}{2}\,(x_b - x_a).$$

Thus, to conclude, it suffices to show that $Z_q = \int_{\mathbb{R}} \tilde{q} \lesssim x_b - x_a$. This follows by Theorem 8.

Given that the chain is at point x_t at some time t, we have described an algorithm that constructs an efficient rejection envelope with number of queries bounded by

$$\mathcal{O}\big(\log(\kappa\,|x_t|_2) \vee 1 + \log\log\kappa\big)\,.$$

To bound the amortized query complexity, it remains to control the time average of $|x_t|_2$. To that end, recall that the Hit-and-Run chain is geometrically ergodic by Lovász and Vempala (2003) so that,

$$\frac{1}{T}\sum_{t=1}^{T} \log(\kappa\,|x_t|_2) \vee 1 \xrightarrow{\text{a.s.}} \mathbb{E}_{x\sim p} \log(\kappa\,|x|_2) \vee 1.$$

By Jensen's inequality we further have

$$\mathbb{E}_{x\sim p} \log(\kappa\,|x|_2) \vee 1 \lesssim \log(\kappa\,\mathbb{E}_{x\sim p}\,|x|_2) \vee 1 \lesssim \log(\kappa d)\,,$$

since p is 1-sub-Gaussian. Noting that $\log\log\kappa \le \log(\kappa d)$, the result follows.

Appendix D. Auxiliary results

The following result on rejection sampling is standard, and we include it for the sake of completeness.

Theorem 7 *Suppose we have query access to the unnormalized target $\tilde{p} = pZ_p$ supported on $\mathscr{X}$, and that we have an upper envelope $\tilde{q} \ge \tilde{p}$. Let q denote the corresponding normalized probability distribution and write Z_q for the normalizing constant, i.e., $\tilde{q} = qZ_q$. Then, rejection sampling with acceptance probability $\tilde{p}/\tilde{q}$ outputs a point distributed according to p, and the number of samples drawn from q until a sample is accepted follows a geometric distribution with mean Z_q/Z_p.*

Proof Since $\tilde{q}$ is an upper envelope for $\tilde{p}$, then $\tilde{p}(X)/\tilde{q}(X) \le 1$ is a valid acceptance probability. Clearly, the number of rejections follows a geometric distribution. The probability of accepting a sample is given by

$$\mathbb{P}(\text{accept}) = \int_{\mathscr{X}} \frac{\tilde{p}(x)}{\tilde{q}(x)}\, q(\mathrm{d}x) = \frac{Z_p}{Z_q} \int_{\mathscr{X}} p(\mathrm{d}x) = \frac{Z_p}{Z_q}\,.$$

Let $X_1, X_2, X_3 \ldots$ be a sequence of i.i.d. samples from q and let $U_1, U_2, U_3 \ldots$ be i.i.d. $\mathrm{unif}[0,1]$. Let $A \subseteq \mathscr{X}$ be a measurable set, and let X be the output of the rejection sampling algorithm. Partitioning by the number of rejections, we may write

$$\begin{aligned}
\mathbb{P}(X \in A) &= \sum_{n=0}^{\infty} \mathbb{P}\Big(X_{n+1} \in A,\ U_i > \frac{\tilde{p}(X_i)}{\tilde{q}(X_i)}\,\forall\, i \in [n],\ U_{n+1} \le \frac{\tilde{p}(X_{n+1})}{\tilde{q}(X_{n+1})}\Big) \\
&= \sum_{n=0}^{\infty} \mathbb{P}\Big(X_{n+1} \in A,\ U_{n+1} \le \frac{\tilde{p}(X_{n+1})}{\tilde{q}(X_{n+1})}\Big)\, \mathbb{P}\Big(U_1 > \frac{\tilde{p}(X_1)}{\tilde{q}(X_1)}\Big)^n \\
&= \sum_{n=0}^{\infty} \Big(\int_A \frac{\tilde{p}(x)}{\tilde{q}(x)}\, q(\mathrm{d}x)\Big)\Big(\int_{\mathscr{X}} \big(1 - \frac{\tilde{p}(x)}{\tilde{q}(x)}\big)\, q(\mathrm{d}x)\Big)^n \\
&= p(A)\,\frac{Z_p}{Z_q} \sum_{n=0}^{\infty} \big(1 - \frac{Z_p}{Z_q}\big)^n = p(A)\,.
\end{aligned}$$

■

We also use the following elementary lemma about Gaussian integrals.

Lemma 8 *Let $a, x_0 > 0$. Then,*

$$\int_{x_0}^{\infty} \exp\big(-a\,(x - x_0) - \frac{1}{2}\,(x - x_0)^2\big)\,\mathrm{d}x \le \frac{1}{a}\,.$$

Proof Completing the square,

$$\int_{x_0}^{\infty} \exp\bigl(-a\,(x-x_0) - \frac{1}{2}\,(x-x_0)^2\bigr)\,\mathrm{d}x = \int_{0}^{\infty} \exp\bigl(-ax - \frac{1}{2}\,x^2\bigr)\,\mathrm{d}x$$
$$= \sqrt{2\pi}\exp\bigl(\frac{a^2}{2}\bigr)\mathbb{P}(Z > a)\,,$$

where $Z \sim \mathcal{N}(0,1)$. The result follows from the Mills ratio inequality (Gordon, 1941). ■