Towards a Complete Analysis of Langevin Monte Carlo: Beyond Poincaré Inequality

Alireza Mousavi-Hosseini* MOUSAVI@CS.TORONTO.EDU
Department of Computer Science at the University of Toronto, and Vector Institute

Tyler Farghly* FARGHLY@STATS.OX.AC.UK
Department of Statistics, University of Oxford, UK.

Ye He LEOHE@UCDAVIS.EDU
Department of Mathematics, University of California, Davis.

Krishnakumar Balasubramanian KBALA@UCDAVIS.EDU
Department of Statistics, University of California, Davis.

Murat A. Erdogdu ERDOGDU@CS.TORONTO.EDU
Department of Computer Science and Department of Statistical Sciences at the University of Toronto, and Vector Institute.

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

Langevin diffusions are rapidly convergent under appropriate functional inequality assumptions. Hence, it is natural to expect that with additional smoothness conditions to handle the discretization errors, their discretizations like the Langevin Monte Carlo (LMC) converge in a similar fashion. This research program was initiated by Vempala and Wibisono (2019), who established results under log-Sobolev inequalities. Chewi et al. (2022a) extended the results to handle the case of Poincaré inequalities. In this paper, we go beyond Poincaré inequalities, and push this research program to its limit. We do so by establishing upper and lower bounds for Langevin diffusions and LMC under weak Poincaré inequalities that are satisfied by a large class of densities including polynomially-decaying heavy-tailed densities (i.e., Cauchy-type). Our results explicitly quantify the effect of the initializer on the performance of the LMC algorithm. In particular, we show that as the tail goes from sub-Gaussian, to sub-exponential, and finally to Cauchy-like, the dependency on the initial error goes from being logarithmic, to polynomial, and then finally to being exponential. This three-step phase transition is in particular unavoidable as demonstrated by our lower bounds, clearly defining the boundaries of LMC.

Keywords: Langevin Monte Carlo, Langevin diffusion, Complexity of Sampling, Rényi divergence, Weak Poincaré Inequalities, Lower bounds, Heavy-tailed sampling.

1. Introduction

Consider the problem of sampling from a target probability density $\pi \propto \exp(-V)$ on $\mathbb{R}^d$ using the canonical algorithm, Langevin Monte Carlo (LMC). The LMC iterations are given by

$$x_{k+1} = x_k - h\nabla V(x_k) + \sqrt{2h}\xi_k, \tag{LMC}$$

* Both authors contributed equally to this work.

where $h > 0$ is the step size, and $(\xi_k)_{k\in\mathbb{N}}$ is an i.i.d. sequence of standard Gaussian random vectors. This algorithm is based on discretizing the following stochastic differential equation (SDE), often referred to as the (overdamped) Langevin diffusion,

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

where $(B_t)_{t\in\mathbb{R}_+}$ is the d-dimensional standard Brownian motion. When π is (strongly) log-concave and smooth, non-asymptotic convergence of LMC has been extensively studied (Dalalyan and Tsybakov, 2012; Dalalyan, 2017b,a; Durmus and Moulines, 2019; Durmus et al., 2019).

The Langevin diffusion (LD), however, converges under relatively milder functional inequality assumptions which are less restrictive compared to global curvature conditions like log-concavity. Indeed, while log-concavity restricts π to be uni-modal, functional inequality based conditions allow for some degree of multi-modality in π (Chen et al., 2021). Furthermore, functional inequalities characterize a wide range of target densities by capturing the tail behavior of the potential. For example, a target potential with tail growth $V(x) \approx \|x\|^\alpha$ at infinity, would satisfy a logarithmic Sobolev inequality (LSI) when $\alpha = 2$, and satisfies a Poincaré inequality (PI) when $\alpha = 1$. Thus, an LSI induces a faster tail growth and is consequently a stronger condition than a PI.

Motivated by this, the following research program was initiated by Vempala and Wibisono (2019): *Can one provide convergence guarantees for* (LMC) *when the target density satisfies a functional inequality and a smoothness condition?* The authors answered the question in the affirmative, showing that the following two conditions on the target $\pi \propto e^{-V}$ suffice to establish a sharp non-asymptotic guarantee for LMC: (i) π satisfies an LSI and (ii) ∇V is Lipschitz continuous. Chewi et al. (2022a) extended this framework significantly; among other contributions, they also proved that LSI can be replaced with a Latała-Oleszkiewicz inequality (LOI), which can cover a range of tail behavior, i.e. $\alpha \in [1, 2]$, interpolating between the edge cases PI ($\alpha = 1$) and LSI ($\alpha = 2$). These works provide a thorough characterization of the convergence of LMC for at least linearly growing potentials, and to our knowledge, providing the state of the art guarantees under minimal set of conditions for this algorithm. However, it is rather unclear how much further this program can be extended. For example, what is the threshold for the tail behavior α beyond which LMC *fails*, if at all it fails to sample from such *heavy-tailed* targets?

In this paper, we aim to complete the program initiated by Vempala and Wibisono (2019), and push the convergence analysis of LMC to its limits. We study the behavior of LMC for potentials that satisfy a family of weak-Poincaré inequalities (WPI), which are one of the mildest conditions required to prove the ergodicity of the Langevin diffusion (Röckner and Wang, 2001; Bakry et al., 2014). A particularly interesting aspect of WPI is that virtually *any target density* satisfies such an inequality. Thus, by proving a convergence guarantee for LMC under a WPI with explicit rate estimates, we establish its convergence universally for any sufficiently smooth target. Interestingly, for targets with sublinear tails, i.e. $V(x) \approx \|x\|^\alpha$ for $\alpha \in (0, 1)$, our rate is polynomial in the initial error, and smoothly extrapolates the rate derived by Chewi et al. (2022a) which was originally covering the regime $\alpha \in [1, 2]$. In the case α approaches 0, however, the tail is logarithmic $V(x) \approx \ln(\|x\|)$, e.g. for Cauchy-type distributions, and our rate estimates exhibit an exponential dependence on the initial error; thus, when there is no warm-start available, LMC would require exponentially many iterations in the initial error for such targets with extreme heavy tails. We also provide a lower bound for LMC under a general tail-growth condition, proving that for Cauchy-type distributions, there is an initialization such that this exponential dependence is unavoidable. Our main contributions can be summarized as follows.

- For a target $\pi \propto e^{-V}$ satisfying a WPI with a Hölder continuous ∇V, we establish non-asymptotic convergence guarantees for LMC and the Langevin diffusion in Rényi divergence. Since any distribution with a locally bounded potential V satisfies a WPI (Röckner and Wang, 2001), our results provide a convergence guarantee for LMC for any sufficiently smooth target.
- We prove WPIs with explicit dimension dependence for two model examples of heavy-tailed distributions that do not satisfy a Poincaré inequality, hence cannot be covered by the results of Chewi et al. (2022a). First, we consider sub-linearly decaying potentials of the form $V(x) = (1+\|x\|^2)^{\alpha/2}$ and establish a rate, which coincides with the estimates of Chewi et al. (2022a), but also holds for $\alpha \in (0, 1)$, namely beyond a Poincaré inequality. Notably, this rate is polynomial in the initial error for all $\alpha > 0$. We also consider the case of extreme heavy tails, i.e. Cauchy-type potentials with $\nu > 0$ degrees of freedom of the form $V(x) = \frac{d+\nu}{2} \ln(1 + \|x\|^2)$, which does not have moments of order $\geq \nu$ defined. For this class of distributions, we prove that, even though LMC converges in Rényi divergence of any order, the dependence on the initial error may be exponential, which may limit its performance severely.
- Finally, we establish lower bounds for the complexity of LMC as well as the Langevin diffusion in Rényi divergence, under various tail growths in the range $\alpha \in [0, 2]$. Our lower bounds indicate that, as the tail growth becomes heavier, LMC and the diffusion both exhibit a slow start behavior by having a worse dependence on the initial divergence. In the particular case of Cauchy-type targets, the exponential dependence on the initial error for LMC and the diffusion is unavoidable, unless there is a good initialization available.

More related work. There have been innumerable works in the recent past focusing on (strongly) log-concave sampling with LMC, which makes it hard to summarize them here. We refer the interested reader to Chewi (2023) for a detailed exposition. Beyond the log-concave setting, the assumption of dissipativity, which controls the growth order of the potential, is used in a large number of prior works to obtain convergence rates for LMC (Durmus and Moulines, 2017; Raginsky et al., 2017; Erdogdu et al., 2018; Erdogdu and Hosseinzadeh, 2021; Mou et al., 2022; Erdogdu et al., 2022). Additionally, a recent result by Balasubramanian et al. (2022) characterized the performance of (averaged) LMC for target densities that are only Hölder continuous (without any further functional inequality or curvature-based assumptions); however, their guarantees were provided in the relatively weaker Fisher information.

We also remark that in the (strongly) log-concave or light-tailed settings, several non-asymptotic results exist on variants of LMC, including higher order integrators (Shen and Lee, 2019; Li et al., 2019; He et al., 2020), the underdamped Langevin Monte Carlo (Cheng et al., 2018; Eberle et al., 2019; Cao et al., 2019; Dalalyan and Riou-Durand, 2020), and the Metropolis-adjusted Langevin Algorithm (Dwivedi et al., 2018; Lee et al., 2020; Chewi et al., 2021; Wu et al., 2022).

Research on the analysis of heavy-tailed sampling is relatively scarce, especially results that are non-asymptotic in nature. Chandrasekaran et al. (2009) studied the iteration complexity of the Metropolis random walk algorithm for sampling from s-concave distributions. Jarner and Roberts (2007) established polynomial ergodicity results for several sampling algorithms including LMC. Kamatani (2018) developed modifications of the standard Metropolis random walk that are suitable for handling heavy-tailed targets and established asymptotic convergence results. Recently, Andrieu et al. (2022) analyzed Metropolis random walk algorithms under WPI and established rate of convergence results in variance-like metrics. Johnson and Geyer (2012) introduced a variable transformation method for Metropolis random walk algorithms, transforming heavy-tailed densities into light-tailed ones using invertible transformations to benefit from existing light-tailed sampling algorithms. He et al. (2022) looked into ULA on a class of transformed densities and provided non-asymptotic results, mainly focusing on isotropic densities. The transformation approach has been extended in recent works such as Yang et al. (2022), and has also been used to prove asymptotic exponential ergodicity for various sampling algorithms in the heavy-tailed settings (Deligiannidis et al., 2019; Durmus et al., 2020; Bierkens et al., 2019).

While the literature on upper bounds on the complexity of sampling algorithms has seen significant progress, the literature on lower bounds is quite limited. Algorithm-independent query complexity of sampling from strongly log-concave distributions in one dimension was obtained by Chewi et al. (2022c). Li et al. (2022) established lower bounds for LMC for sampling from strongly log-concave distributions. Chatterji et al. (2022) established lower bounds for sampling from strongly log-concave distributions in the stochastic setting, when the gradients are observed with noise. Ge et al. (2020) established lower bounds for the related problem of estimating the normalizing constants of a log-concave density. Lee et al. (2021) and Wu et al. (2022) established lower bounds for the class of metropolized algorithms (including metropolized Langevin and Hamiltonian Monte Carlo methods) for sampling from strongly log-concave distributions. Finally, lower bounds in Fisher information for non-log-concave sampling were obtained in Chewi et al. (2022b).

Notation. Throughout the paper, we will use $\pi \propto \exp(-V)$ to denote the target probability measure with unnormalized potential V, and ρ_t and μ_k to denote the law of the Langevin diffusion at time t, and the law of the LMC at iteration k. $(P_t)_{t\in\mathbb{R}_+}$ will denote the Markov semigroup of the Langevin diffusion. Probability measures we work with in this paper admit densities with respect to the Lebesgue measure, and we will use the same notation for their densities. We will use $\tilde{\Theta}_\Psi$ and $\tilde{O}_\Psi$ to hide polylog factors and constants depending only on the set of variables Ψ. $\Gamma(z) := \int_0^\infty t^{z-1}e^{-t}\mathrm{d}t$ for $z > 0$ denotes the Gamma function, and ω_d is the volume of the unit d-ball.

2. Weak Poincaré Inequalities and Rényi Convergence of the Diffusion

We consider a class of functional inequalities introduced by Röckner and Wang (2001), motivated by the work of Liggett (1991)[1]. Throughout this work, we avoid concerns regarding the domain of the generator for the diffusion (LD) by assuming that the set of infinitely differentiable functions, $\mathcal{C}^\infty(\mathbb{R}^d)$, forms a core for the domain. For example, this is given when V is infinitely differentiable itself (see, *e.g.,* Bakry et al. (2014, Proposition 3.2.1)).

Definition 1 (Weak Poincaré Inequality) *A probability measure π on $\mathbb{R}^d$ satisfies a weak Poincaré inequality (WPI) if there exists non-increasing $\beta_{\mathsf{WPI}} : (0,\infty) \to \mathbb{R}_+$ and $\Phi : L^2(\pi) \to [0,\infty]$ with $\Phi(cf + a) = c^2\Phi(f)$ for every $c, a \in \mathbb{R}$ and $f \in L^2(\pi)$, such that for every $f \in \mathcal{C}^\infty(\mathbb{R}^d)$,*

$$\mathrm{Var}_\pi(f) \le \beta_{\mathsf{WPI}}(r)\,\mathbb{E}_\pi\left[\|\nabla f\|^2\right] + r\,\Phi(f), \quad \forall r > 0. \tag{WPI}$$

We remark that virtually any target measure of interest in sampling satisfies such an inequality. More specifically, Röckner and Wang (2001) showed that $\pi \propto e^{-V}$ satisfies a WPI with $\Phi(\cdot) = \mathrm{Osc}(\cdot)^2 := (\sup f - \inf f)^2$ and for some β_{WPI} as soon as V is locally bounded. In the special

1. We refer the interested reader to Aida (1998) and Mathieu (1998) for other attempts to propose weaker versions of Poincaré inequalities, and to Röckner and Wang (2001) for their relationship with Definition 1.

cases where $\Phi = 0$ or the function β_{WPI} is uniformly bounded, the above inequality reduces to the classical Poincaré inequality which reads, for a constant β_{PI} and for every $f \in \mathcal{C}^\infty(\mathbb{R}^d)$,

$$\mathrm{Var}_\pi(f) \leq \beta_{\mathsf{PI}}\, \mathbb{E}_\pi\left[\|\nabla f\|^2\right]. \tag{PI}$$

The tail properties of the distribution π are captured by the function β_{WPI}, which will essentially determine the convergence rate of LMC. We will present our convergence guarantees under the generic condition (WPI), and for several model examples, we will derive explicit estimates of β_{WPI} to make our results more explicit.

Functional inequalities of the form (WPI) naturally extend Poincaré inequalities (PI) to arbitrary distributions, removing any tail growth requirements. In particular, as PI is equivalent to an exponential L^2 convergence rate for the Markov semigroup, a WPI is equivalent to a subexponential L^2 convergence rate (Röckner and Wang, 2001; Bakry et al., 2014). Similarly, one can also replace the variance term in (WPI) with entropy, in which case the functional inequality is of the form of a weak log-Sobolev inequality (WLSI). As shown by Cattiaux et al. (2007), a WLSI is equivalent to a WPI; thus, we find it sufficient to present our results in terms of the WPI.

Following recent works (see, *e.g.,* Ganesh and Talwar (2020); Erdogdu et al. (2022); Chewi et al. (2022a)), we use Rényi divergence as a measure of distance between two probability distributions. Rényi divergence of order q is defined by

$$R_q(\rho \,\|\, \pi) := \frac{1}{q-1} \ln \left\| \frac{\mathrm{d}\rho}{\mathrm{d}\pi} \right\|^q_{L^q(\pi)} \quad \text{for} \quad 1 < q < \infty, \tag{1}$$

when ρ is absolutely continuous with respect to π, and $+\infty$ otherwise. By Jensen's inequality, $R_q(\rho \,\|\, \pi)$ is non-decreasing in q. If we consider the limits, (i) as $q \downarrow 1$ it reduces to KL divergence, i.e. $\lim_{q\downarrow 1} R_q(\rho\,\|\,\pi) = \mathsf{KL}(\rho\,\|\,\pi)$ and (ii) as $q \to \infty$ it reduces to the L^∞-norm, i.e. $\lim_{q\to\infty} R_q(\rho \,\|\, \pi) = \ln \|\mathrm{d}\rho/\mathrm{d}\pi\|_{L^\infty(\pi)}$. It is also related to the χ^2 divergence via $\chi^2(\rho \,\|\, \pi) + 1 = \exp(R_2(\rho \,\|\, \pi))$.

Providing convergence guarantees in Rényi divergence is of particular interest since it upper bounds many commonly used distance measures. Specifically, by Pinsker's inequality and the monotonicity of Rényi divergence, we have

$$2D_{\mathsf{TV}}(\rho, \pi)^2 \leq \mathsf{KL}(\rho \,\|\, \pi) \leq R_q(\rho \,\|\, \pi) \quad \text{for} \quad q > 1.$$

Notice that comparing the quadratic Wasserstein distance $\mathrm{W}_2^2(\rho, \pi)$ with $R_q(\rho \,\|\, \pi)$ is more subtle. Under a PI with constant β_{PI}, or more broadly under finite fourth moments, one can write

$$\ln\left(1 + \tfrac{1}{2\beta_{\mathsf{PI}}}\, \mathrm{W}_2^2(\rho, \pi)\right) \leq R_2(\rho \,\|\, \pi) \quad \text{and} \quad \ln\left(1 + \frac{\mathrm{W}_2^4(\rho, \pi)}{4\mathbb{E}_\pi\left[\|x\|^4\right]}\right) \leq R_2(\rho \,\|\, \pi),$$

respectively. The first inequality is due to Liu (2020) and the second one can be derived from the weighted total variation control on W_2 (Villani, 2003, Proposition 7.10) along with the Cauchy-Schwartz inequality. Hence, even under a WPI, a bound in Rényi divergence can be translated to a bound in W_2 (when $\mathbb{E}_\pi\left[\|x\|^4\right] < \infty$), possibly at the expense of introducing additional dimension dependency into the bounds. It is worth highlighting that a distribution satisfying (WPI) does not need to have any particular moment defined, in which case W_2 may be undefined, but its Rényi divergence of some order may still be well-defined.

2.1. Rényi Convergence of the Langevin Diffusion

Classically, convergence of the Langevin diffusion under (WPI) is considered only in variance, or equivalently, the χ^2 divergence (see e.g. Wang (2006, Chapter 4) and Bakry et al. (2014, Chapter 7.5)). The following result characterizes its convergence in Rényi divergence which is stronger in the case of $q > 2$.

Theorem 2 *Suppose π satisfies (WPI) for some β_{WPI} and $\Phi(\cdot) = \operatorname{Osc}(\cdot)^2$. For any $2 \le q < q' \le \infty$ such that $R_{q'}(\rho_0 \,\|\, \pi) < \infty$, define $\delta_0 := \exp(qR_{q'}(\rho_0 \,\|\, \pi))$. Then, for any $r > 0$,*

$$R_q(\rho_t \,\|\, \pi) \le \begin{cases} R_q(\rho_0 \,\|\, \pi) - \frac{2-4r\delta_0}{\beta(r)q}t & \text{if} \quad R_q(\rho_0 \,\|\, \pi), R_q(\rho_t \,\|\, \pi) \ge 1 \\ e^{-\frac{2t}{\beta(r)q}}(R_q(\rho_0 \,\|\, \pi) - 2r\delta_0) + 2r\delta_0 & \text{if} \quad R_q(\rho_0 \,\|\, \pi) < 1, \end{cases}$$

where

$$\beta(r) := \begin{cases} \beta_{\mathsf{WPI}}(r) & \text{if} \quad q' = \infty \\ \beta_{\mathsf{WPI}}\Big((r/5)^{\frac{q'}{q'-q}}\Big) \ln\Big((5/r)^{\frac{q'}{q'-q}} \vee 1\Big) & \text{if} \quad q' < \infty. \end{cases} \tag{2}$$

Therefore, we have $R_q(\rho_T \,\|\, \pi) \le \varepsilon$ whenever

$$T \ge q\beta\Big(\frac{1}{4\delta_0}\Big)R_q(\rho_0 \,\|\, \pi) + \frac{q}{2}\beta\Big(\frac{\varepsilon}{4\delta_0}\Big)\ln\Big(\frac{1}{\varepsilon}\Big). \tag{3}$$

We emphasize that while the classical convergence results under (WPI) (Wang, 2006; Bakry et al., 2014) require $R_\infty(\rho_0 \,\|\, \pi) < \infty$ at initialization, our convergence guarantees hold as soon as the initial error satisfies $R_{q'}(\rho_0 \,\|\, \pi) < \infty$ for some $q' > q$. Moreover, in the case where π satisfies a PI, i.e. when β_{WPI} is constant, we can remove the requirement of $R_{q'}(\rho_0 \,\|\, \pi) < \infty$, and the above theorem recovers Vempala and Wibisono (2019, Theorem 3) by defining $\beta(0) := \lim_{r\to 0}\beta_{\mathsf{WPI}}(r)$.

The proof of Theorem 2 is presented in Appendix A.1, and it relies on a proof technique also used in Vempala and Wibisono (2019); Chewi et al. (2022a). However, because of the $\Phi(\cdot) = \operatorname{Osc}(\cdot)^2$ term in (WPI), we additionally need to control oscillations of $\frac{\mathrm{d}\rho_t}{\mathrm{d}\pi}$ uniformly over the process. Via contraction properties for the semigroup, we can reduce such control to $\frac{\mathrm{d}\rho_0}{\mathrm{d}\pi} \in L^\infty(\pi)$ (or equivalently $R_\infty(\rho_0 \,\|\, \pi) < \infty$). To further relax this assumption, one needs to obtain a WPI with a *weaker* Φ. Specifically, given an initial control of the type $\frac{\mathrm{d}\rho_0}{\mathrm{d}\pi} \in L^{q'}(\pi)$ for some $q' < \infty$, it suffices to obtain a WPI with $\Phi(\cdot) = \|\cdot\|^2_{L^u(\pi)}$ (for mean-zero functions), where $u = \frac{2q'}{q}$. The following proposition, which is a consequence of a general L_p decay result due to Cattiaux et al. (2012) and might be of independent interest, is crucial in the proof of Theorem 2. Specifically, it converts a WPI with $\Phi(\cdot) = \operatorname{Osc}(\cdot)^2$ into a different WPI with $\Phi(\cdot) = \|\cdot\|^2_{L^u(\pi)}$; thus, allows for a less restrictive initialization. We defer the proof of this proposition to Appendix A.2.

Proposition 3 *Suppose π satisfies (WPI) with $\Phi(\cdot) = \operatorname{Osc}(f)^2$ and some $\beta_{\mathsf{WPI}}(r)$. Then, for every $u > 2$, π also satisfies (WPI) with weighting β' and regularization function Φ' such that*

$$\beta'_{\mathsf{WPI}}(r) = \beta_{\mathsf{WPI}}\left((r/5)^{\frac{u}{u-2}}\right)\ln\left((5/r)^{\frac{u}{u-2}} \vee 1\right) \quad \text{and} \quad \Phi'(\cdot) = \big\|f - \mathbb{E}_\pi[f]\big\|^2_{L^u(\pi)}. \tag{4}$$

Additionally, provided that π does not satisfy a PI, π cannot satisfy a WPI with $\Phi = \Phi'$ for $u = 2$.

The last statement of the above proposition clarifies why we need to choose $q' > q$. In order to establish guarantees with $q' = q$, one would be required to obtain a WPI with $\Phi(\cdot) = \|\cdot\|^2_{L^2(\pi)}$, which is equivalent to a PI.

3. Langevin Monte Carlo for Heavy-Tailed Targets

In this section, we present our main convergence guarantees for LMC when the target satisfies (WPI) and ∇V is s-Hölder continuous for some $s \in (0, 1]$, that is,

$$\|\nabla V(x) - \nabla V(y)\| \leq L \|x - y\|^s \quad \forall x, y \in \mathbb{R}^d. \qquad (s\text{-Hölder})$$

The case $s = 1$ corresponds to the ubiquitous Lipschitz continuity where the potential is smooth, and the regime where $s < 1$ is often termed as weak smoothness. Since our main focus is potentials that do not satisfy (PI), any order Hölder continuity is feasible.

Below, we state our main convergence result for a generic (WPI) with an unspecified β_{WPI}. We will explicitly derive its implications for specific targets in the subsequent sections.

Theorem 4 *Suppose $\pi \propto e^{-V}$ satisfies (WPI) for some β_{WPI} and $\Phi(\cdot) = \mathrm{Osc}(\cdot)^2$, and ∇V is (s-Hölder) continuous, with $\nabla V(0) = 0$ for simplicity. For any $q \in [2, \infty), q' \in (2q - 1, \infty]$ such that $R_{q'}(\mu_0 \,\|\, \pi) < \infty$, define $\delta_0 := \exp((2q-1)R_{q'}(\mu_0 \,\|\, \pi))$, $m := \frac{1}{2}\inf\{R : \pi(\|x\| \geq R) \leq \frac{1}{2}\}$, and*

$$T := (2q-1)\left\{\beta\left(\frac{1}{4\delta_0}\right) R_{2q-1}(\mu_0 \,\|\, \pi) + \beta\left(\frac{\varepsilon}{8\delta_0}\right) \ln\left(\frac{2}{\varepsilon}\right)\right\},$$

for $\varepsilon \leq q^{-1}$, where β is as in (2). Let $\hat{\pi}$ denote a modified version of π (explicitly defined in (10)) and assume, for simplicity, that $\varepsilon^{-1}, m, L, T, R_2(\mu_0 \,\|\, \hat{\pi}) \geq 1$. Then, for a sufficiently small step size h, denoting by μ_N, the law of the N-th iterate of LMC initialized at ρ_0, after

$$N = \Theta_s\left(\frac{T^{1+1/s} d q^{1/s} L^{2/s}}{\varepsilon^{1/s}} \max\left\{1, \frac{\varepsilon^{1/(2s)} m^s}{L^{1/s-1} T^{1/(2s)} d}, \frac{\varepsilon^{1/(2s)} R_2(\mu_0 \,\|\, \hat{\pi})^{s/2}}{L^{1/s-1} T^{(1-s^2)/(2s)} d} \ln\left(\frac{qTLR_2(\mu_0 \,\|\, \hat{\pi})}{\varepsilon}\right)^{s/2}\right\}\right)$$

iterations of LMC, we obtain $R_q(\mu_N \,\|\, \pi) \leq \varepsilon$.

We make a few remarks. First, it is possible to find an initialization μ_0, e.g. isotropic Gaussian, such that $R_q(\mu_0 \,\|\, \pi), R_q(\mu_0 \,\|\, \hat{\pi}) \leq \tilde{\mathcal{O}}(d)$, with details provided in Lemma 31. In such a scenario, up to log factors, the last term in the maximum will never dominate. Additionally, the middle term will never dominate for sufficiently small ε, and in fact, in our model examples, it will not dominate even for $\varepsilon = 1$ due to proper control on m. Then our convergence rate reads

$$N = \tilde{\Theta}_s\left(\frac{q^{1+2/s} L^{2/s} d\big\{d\beta(\frac{1}{4\delta_0}) + \beta(\frac{\varepsilon}{8\delta_0})\big\}^{1+1/s}}{\varepsilon^{1/s}}\right).$$

In the case where π satisfies (PI), we set $\beta(r) = \beta_{\mathsf{PI}}$ for any $r > 0$, and the above rate reduces to the rate implied by Chewi et al. (2022a, Theorem 7).

The proof of Theorem 4 is given in Appendix A.3, and it is based on Theorem 2 and the Girsanov argument used in Chewi et al. (2022a). Two key distinctions are (i) our analysis is tailored for heavy-tailed targets, hence does not require any moment order to be defined, and also requires a finer control of different error terms to mitigate suboptimal rates, and (ii) it exploits the continuous time convergence under WPI, rather than stronger functional inequalities such as PI.

3.1. Examples

In this section, we focus on various heavy-tailed targets that do not satisfy (PI). In particular, we consider sampling from Cauchy-type measures in Section 3.1.2, which may not even have a defined expectation or any order moment for that matter; yet, we are able to provide convergence guarantees for LMC in Rényi divergence of any finite order.

3.1.1. POTENTIALS WITH SUB-LINEAR TAILS

Consider the measure $\pi_\alpha \propto \exp(-V_\alpha)$ where $V_\alpha(x) = (1 + \|x\|^2)^{\alpha/2}$ with $\alpha \in (0, 1)$. This potential satisfies (s-Hölder) with $s = 1$ and $L = 1$. We analyze this potential as a substitute for $\|x\|^\alpha$ since the latter does not have continuous gradients, while the former still behaves similar to $\|x\|^\alpha$ for large $\|x\|$. In the following Proposition, we present the β_{WPI} estimate for this potential.

Proposition 5 *The measure $\pi_\alpha(x) \propto \exp(-V_\alpha)$ with $\alpha \in (0, 1)$ satisfies (WPI) with*

$$\beta_{\mathsf{WPI}}(r) = \inf_{\gamma\in(0,2\alpha]} C_\alpha \left(d^{\frac{2(2-2\alpha+\gamma)}{\gamma}} + \ln\left(r^{-1}\right)^{\frac{2-2\alpha+\gamma}{\alpha}}\right) \quad \textit{and} \quad \Phi(\cdot) = \mathrm{Osc}(\cdot)^2, \tag{5}$$

where C_α is a constant depending only on α.

It is worth noting that this estimate is polynomial in dimension, improving the implicit and potentially exponential dependency implied by Cattiaux et al. (2010, Proposition 5.6) at the expense of introducing an additional factor of $\ln(1/r)^2$ into the estimate. Specifically, they show that any potential of the form $V(x) = \psi(x)^\alpha$ for some convex non-negative ψ and $\alpha \in (0, 1)$ satisfies a WPI with $\beta_{\mathsf{WPI}}(r) = C_{\mathsf{WPI}}(1 + \ln(r^{-1})^{2(1/\alpha-1)})$ where $C_{\mathsf{WPI}} = C_{\mathsf{WPI}}(d, \alpha)$ is implicit in dimension, and in general it is not known how to achieve a dependency better than exponential via such techniques. Invoking Theorems 2 and 4 with the estimate for β_{WPI} given by Proposition 5, we can establish the following convergence guarantees for LMC and the Langevin diffusion.

Corollary 6 *Consider the setting of Theorem 4 with the target measure $\pi_\alpha \propto \exp(-V_\alpha)$. Denoting the distribution of the Langevin diffusion at time T with ρ_T, we have $R_q(\rho_T \,\|\, \pi_\alpha) \le \varepsilon$ whenever*

$$T \ge C_{\alpha,q} \inf_{\gamma\in(0,2\alpha]}\left\{\left(d^{\frac{2}{\gamma}(2-2\alpha+\gamma)} + R_\infty(\rho_0 \,\|\, \pi)^{\frac{2-2\alpha+\gamma}{\alpha}}\right)\left(R_q(\rho_0 \,\|\, \pi) + \ln(1/\varepsilon)\right) + \ln(1/\varepsilon)^{\frac{2-\alpha+\gamma}{\alpha}}\right\}.$$

Further, denoting the distribution of LMC after N iterations with μ_N, we have $R_q(\mu_N \,\|\, \pi_\alpha) \le \varepsilon$ if

$$N = \tilde{\Theta}_{\alpha,q}\left(\frac{(d^{4/\alpha} + R_\infty(\mu_0 \,\|\, \pi_\alpha)^{4/\alpha})R_{2q-1}(\mu_0 \,\|\, \pi_\alpha)^2 d}{\varepsilon} \max\left\{1, \frac{\sqrt{\varepsilon R_2(\mu_0 \,\|\, \hat{\pi}_\alpha)}}{d}\right\}\right).$$

In particular, with $\rho_0 = \mu_0 = \mathcal{N}(0, I_d)$, and $\gamma = 2\alpha$, and using the bound on the initial Rényi divergence provided by Corollary 24 and Lemma 32, we obtain

$$T \gtrsim C_{\alpha,q}\left(d^{2/\alpha+1} + d^{2/\alpha}\ln(1/\varepsilon) + \ln(1/\varepsilon)^{2/\alpha+1}\right), \quad \text{and} \quad N = \tilde{\Theta}_{\alpha,q}\left(\tfrac{d^{4/\alpha+3}}{\varepsilon}\right),$$

where $\gtrsim$ hides polylog factors in d.

As a final remark, if instead of the β_{WPI} estimate of Proposition 5, we use the estimate from Cattiaux et al. (2010) together with Theorem 4, we can obtain a rate for LMC which reads $N =$

$\tilde{\Theta}_{\alpha,q}\big(C_{\mathsf{WPI}}^2 \frac{d^{4/\alpha-1}}{\varepsilon}\big)$. In particular, this can be seen as a smooth extrapolation of the rate given by Chewi et al. (2022a) for the case of $\alpha \in [1, 2]$. By showing that an LOI holds with some constant C_{LOI}, they obtain rates identical to ours once C_{WPI} is replaced by C_{LOI}, and the two rates match at $\alpha = 1$ as C_{LOI} and C_{WPI} will be equivalent up to an absolute constant in this case. However, when $\alpha < 1$, C_{WPI} of Cattiaux et al. (2010) has an implicit and potentially exponential dependence on dimension. Although our estimate in Proposition 5 improves this to polynomial, due to the additional log factor introduced into the bound, the rate will no longer smoothly extrapolate the rate of Chewi et al. (2022a) to the regime $\alpha < 1$.

3.1.2. POTENTIALS WITH LOGARITHMIC TAILS: CAUCHY-TYPE MEASURES

In this section, we consider Cauchy-type measures of the form $\pi_\nu \propto \exp(-V_\nu)$ where $V_\nu(x) = \frac{d+\nu}{2}\ln(1 + \|x\|^2)$ with $\nu > 0$, which is Hölder continuous with $s = 1$ and $L \leq d + \nu$. In particular, π_ν belongs to the family of d-dimensional Student-t distributions with ν degrees of freedom; see e.g. Jarner and Roberts (2007); Kamatani (2018); He et al. (2022); Yang et al. (2022) for sampling from such distributions. Cauchy-type measures only have finite moments of order less than ν.

The following result presents a sharp estimate of β_{WPI} for π_ν.

Proposition 7 *The measure $\pi_\nu \propto \exp(-V_\nu)$ for $\nu > 0$ satisfies (WPI) with*

$$\beta_{\mathsf{WPI}}(r) = \frac{2}{\nu} + 2\left(\frac{d}{\nu} + 1\right) r^{-2/\nu} \quad \textit{and} \quad \Phi(\cdot) = \operatorname{Osc}(\cdot)^2. \tag{6}$$

Similar to the previous case, the above estimate improves the potentially exponential dimension dependence in Cattiaux et al. (2010, Proposition 5.4) to linear, while keeping r dependency the same. Employing the above estimate in Theorems 2 and 4, we obtain the following rate for LMC and the Langevin diffusion.

Corollary 8 *Consider the setting of Theorem 4 with the target measure $\pi_\nu \propto \exp(-V_\nu)$. Denoting the distribution of the Langevin diffusion at time T with ρ_T, we have $R_q(\rho_T \,\|\, \pi_\nu) \leq \varepsilon$ whenever*

$$T \geq C_\nu q d \exp\left(\frac{2qR_\infty(\rho_0 \,\|\, \pi)}{\nu}\right)\Big(R_q(\rho_0 \,\|\, \pi) + (1/\varepsilon)^{2/\nu}\ln(1/\varepsilon)\Big).$$

Furthermore, let $\nu \leq cd$ for some absolute constant c, and denote the distribution of LMC after N iterations with μ_N. Then, we have $R_q(\mu_N \,\|\, \pi_\nu) \leq \varepsilon$ whenever

$$N = \tilde{\Theta}_{\nu,q}\left(\frac{d^5 e^{\frac{4(2q-1)}{\nu}R_\infty(\mu_0\|\pi)}}{\varepsilon}\left(R_{2q-1}^2(\mu_0\,\|\,\pi) + \left(\frac{1}{\varepsilon}\right)^{4/\nu}\right)\max\left\{1, \frac{\sqrt{\varepsilon R_2(\mu_0\,\|\,\hat{\pi})R_\infty(\mu_0\,\|\,\pi)}}{d}\right\}\right).$$

We highlight that the dependence on the initial error is exponential in contrast to the polynomial dependence in the previous case. In other words, with a naive initialization, in the worst case, this dependency will negatively impact the complexity of sampling from Cauchy-type measures. However, if one initializes with an isotropic Gaussian with an appropriately scaled variance, using the estimates on the initial Rényi divergence in Corollary 23 and Lemma 32, we can obtain

$$T \geq C_{\nu,q} d^{2q/\nu+1}\Big(\ln d + (1/\varepsilon)^{2/\nu}\ln(1/\varepsilon)\Big) \quad \text{and} \quad N = \tilde{\Theta}_{\nu,q}\big(d^{\frac{4(2q-1)}{\nu}+5}/\varepsilon^{4/\nu+1}\big).$$

Finally, we remark that such an initialization may not be available in general; thus, the exponential dependence might be unavoidable.

4. Lower Bounds for LMC via Variance Decay

As demonstrated in the examples, the convergence guarantees given in Section 3 have worse dependence on the initial divergence when the tails of the target are heavier, with a sharp transition at (Cauchy-type) logarithmic tails, in which case the dependence on the initial error becomes exponential. In this section, we show that this is not due to a limitation in the analysis, but is in fact a property of LMC in heavy-tailed settings. We present a method for developing lower bounds for the convergence rate of LMC, with the goal of observing a similar dependence on the initial divergence and in particular, establishing that the exponential dependency for Cauchy-type measures is unavoidable in the worst case. First, we introduce the notation of complexity that we lower bound.

Definition 9 *Let $\mathcal{D}$ and $\mathcal{D}'$ denote two divergences.*

- *The iteration complexity of LMC, denoted by $\mathscr{C}^{LMC}_{\mathcal{D},\mathcal{D}'}(\pi, h, \Delta_0, \varepsilon)$, is the smallest $k \in \mathbb{N}$ such for any μ_0 satisfying $\mathcal{D}'(\mu_0 \,\|\, \pi) \leq \Delta_0$, LMC with step-size h satisfies $\mathcal{D}(\mu_k \,\|\, \pi) \leq \varepsilon$.*
- *Similarly, the time complexity of the Langevin diffusion, $\mathscr{C}^{LD}_{\mathcal{D},\mathcal{D}'}(\pi, \Delta_0, \varepsilon)$, is the smallest $T \in \mathbb{R}_+$ such that for any ρ_0 satisfying $\mathcal{D}'(\rho_0 \| \pi) \leq \Delta_0$, the diffusion satisfies $\mathcal{D}(\rho_T \| \pi) \leq \varepsilon$.*

For example, $\mathcal{D}$ and $\mathcal{D}'$ could be the KL divergence, the order-q Rényi divergence or the p-Wasserstein distance. By having this notion of complexity uniformly over all initial distributions whose initial error does not exceed Δ_0, we mirror the types of bounds given in Section 3 and in the literature on LMC more broadly (see e.g. Chewi et al. (2022b, Definition 5)). Indeed, in Section 3.1, we provided upper bounds for this quantity with $\mathcal{D} = R_q$ and $\mathcal{D}' = R_{q'}$.

Our methodology for developing lower bounds is based on the observation that in order for LMC to be close to the target in some divergence, its moments should match those of the target. We study the second moments of LMC and the Langevin diffusion and find that, by tracking their evolution through differential inequalities, we can relate their convergence to the convergence of processes that are more tractable. To relate this to the Rényi divergence, we rely on the strategy of variational representations; the most famous of these, the Donsker-Varadhan theorem, has the KL divergence represented as a supremum over a functional. We use a similar representation for the q-Rényi divergence with $q > 1$, which is due to Birrell et al. (2021), with a particular choice of test function to obtain

$$R_q(\rho \| \pi) \geq \ln\left(\rho(\|\cdot\|^2)^{\frac{q}{q-1}} / \pi(\|\cdot\|^{\frac{2q}{q-1}})\right) \quad \text{whenever} \quad \pi(\|\cdot\|^{2q/(q-1)}) < \infty.$$

Therefore, a strategy for obtaining lower bounds in the case where $\mathbb{E}\|x_0\|^2$ is large follows: (i) relate the decay of the Rényi divergence to the decay of $\mathbb{E}\|x_k\|^2$, (ii) lower bound the number of iterations required for $\mathbb{E}\|x_k\|^2$ to decay by the number of iterations for a more tractable process (e.g. gradient descent) to converge. To strengthen the lower bounds obtained, we use a similar approach to develop an upper bound on the step size necessary to obtain a given accuracy. Details for the methodology can be found in Appendix B.

4.1. Heavy-Tailed Potentials and Slow Starts

In order to obtain lower bounds, we will make an assumption of the type

$$\|\nabla V(x)\| \lesssim \|x\|^{\alpha-1},$$

with $\alpha \in [0, 2]$, for sufficiently large $\|x\|$, which will capture potentials that grow no faster than the order of $\|x\|^\alpha$ when $\alpha > 0$, and $\ln(\|x\|)$ when $\alpha = 0$. Under such a condition, we show that as $\alpha \to 0$, the dependence on initial error deteriorates. In particular, the dependence is logarithmic with $\alpha = 2$, becomes polynomial with $\alpha \in (0, 2)$, and finally turns exponential with $\alpha = 0$, as outlined below.

Theorem 10 (Three-step Phase Transition) *Let $q \in (1, \infty)$, $q' \in [1, \infty]$ and the moment condition $\pi(\|\cdot\|^{2q/(q-1)}) < \infty$ hold. Suppose that*

$$\|\nabla V(x)\| \le \frac{b\,\|x\|}{(1 + \|x\|^2)^{1-\alpha/2}}, \tag{7}$$

with $\alpha \in [0, 2]$ and $b > 0$, and let $\nu := b - d$. Then for all sufficiently large Δ_0 (see (18), (19), and (20) for respective cases), the time complexity of the Langevin diffusion, and the iteration complexity of LMC, for obtaining an accuracy of 1 in q-Rényi divergence, is lower bounded as follows

$$\alpha = 2: \qquad \mathscr{C}^{LMC}_{R_q,\mathsf{KL}}(\pi, h, \Delta_0, 1) \gtrsim \frac{\ln(\Delta_0)}{h}, \qquad \mathscr{C}^{LD}_{R_q,\mathsf{KL}}(\pi, \Delta_0, 1) \gtrsim \ln(\Delta_0),$$

$$\alpha \in (0,2): \quad \mathscr{C}^{LMC}_{R_q,R_{q'}}(\pi, h, \Delta_0, 1) \gtrsim \frac{d^{1-\alpha/2}\Delta_0^{\frac{(2-\alpha)^2}{2\alpha}}}{h}, \quad \mathscr{C}^{LD}_{R_q,R_{q'}}(\pi, \Delta_0, 1) \gtrsim d^{1-\alpha/2}\Delta_0^{\frac{(2-\alpha)^2}{2\alpha}},$$

$$\alpha = 0: \qquad \mathscr{C}^{LMC}_{R_q,R_{q'}}(\pi, h, \Delta_0, 1) \gtrsim \frac{d}{\nu h}\exp(\Delta_0/\nu), \quad \mathscr{C}^{LD}_{R_q,R_{q'}}(\pi, \Delta_0, 1) \gtrsim \frac{d}{\nu}\exp(\Delta_0/\nu),$$

where $\gtrsim$ hides a constant depending only on b and α. The lower bounds hold for all $h > 0$, except when $\alpha = 2$ where they hold for $h < b^{-1}$.

In Theorem 10, the case of $\alpha = 2$ corresponds to tails no lighter than Gaussian and indeed, the lower bounds that we obtain reproduce the dependence of Δ_0 known in the Gaussian setting (Vempala and Wibisono (2019)). The case of $\alpha \in (0, 2)$ corresponds to potentials with tail growth similar to $\|x\|^\alpha$ and the case of $\alpha = 0$ corresponds to Cauchy-type logarithmic tails. Indeed, the generalized Cauchy potentials V_ν and the sub-linear potentials V_α from Section 3.1 satisfy (7) with $b = d+\nu$ and $b = \alpha$ respectively. For these examples, our lower bounds recover the polynomial and exponential dependence on the initial error given in the upper bounds of Section 3. More generally, we point out that (7) is satisfied for any smooth ∇V with $\nabla V(0) = 0$ and $\|\nabla V(x)\| \lesssim \|x\|^{\alpha-1}$ for large $\|x\|$. Similar to the upper bounds, the implicit assumption $\nabla V(0) = 0$ is made only for the simplicity of presentation. In fact, to achieve the most generality, even this assumption can be relaxed to a bound of the type $\langle \nabla V(x), x\rangle \lesssim \|x\|^\alpha$ for large $\|x\|$ while recovering similar results. One can interpret such a condition as *reversed dissipativity*. While dissipativity is used in the literature to ensure a lower bound on tail growth for obtaining upper bounds on moments (see e.g. Raginsky et al. (2017); Erdogdu et al. (2018, 2022); Farghly and Rebeschini (2021)), the reversed condition imposes an upper bound on the tail growth that leads to obtaining lower bounds on the moments.

The above result highlights one reason why LMC can be seen to perform worse in heavy-tailed settings. As the tail becomes heavier, LMC exhibits a slow start behavior by having a worse dependence on the initial divergence. Showing that this is a property of the Langevin diffusion and not due to discretization motivates using alternative diffusions for sampling from heavy-tailed targets (see, for example, Li et al. (2019); Şimşekli et al. (2020); He et al. (2022); Zhang and Zhang (2022)).

To complete our exposition, we note that while the results of Theorem 10 are stated for any fixed choice of step size, in practice we have to ensure step size is small enough for the discretization to not harm the convergence of LMC. This usually leads to additional dependence on dimension or final accuracy. In a special case where the target potential is radially symmetric, the following proposition provides a general tool for determining suitable ranges of step size for LMC, hence completing the complexity lower bound in conjunction with Theorem 10.

Proposition 11 (Step-size upper bound) *Suppose the potential is radially symmetric with* $V(x) = f(\|x\|^2)$ *and* $g : \mathbb{R}_+ \to \mathbb{R}_+$ *given by* $g(r) = (1 - 2hf'(r))^2 r$ *is convex and non-decreasing. Let* $\varepsilon > 0, q > 1$. *Suppose that* $\mathbb{E}\left[\|x_0\|^2\right] > \sigma_\varepsilon^2$ *where* $\sigma_\varepsilon^2 = e^{\frac{q-1}{q}\varepsilon}\pi\big(\|\cdot\|^{\frac{2q}{q-1}}\big)^{\frac{q-1}{q}}$. *Then it must hold that,*

$$\inf_{k\in\mathbb{N}} R_q(\rho_k\|\pi) < \varepsilon \implies h \le \frac{1}{f'(\sigma_\varepsilon^2)}\left(1 - \frac{d}{2f'(\sigma_\varepsilon^2)\sigma_\varepsilon^2}\right).$$

We conclude our discussion by considering an example that demonstrates an application of the main tools developed in this paper. Specifically, we consider the setting of Section 3.1.2 and recall the Cauchy-type target, π_ν. The following corollary, which gives a sharp characterization of the convergence behavior, is a direct consequence of the lower bound from Theorem 10 and the upper bound from Corollary 8.

Corollary 12 *Let* $q \ge 2$ *and suppose that* $\nu > \frac{2q}{q-1}$. *Then for any* $\Delta_0 \ge C_q(1 + \nu\ln(d+\nu))$, *we have the following bound for the Langevin diffusion,*

$$\frac{d}{4\nu}\exp\left(\frac{\Delta_0}{\nu}\right) \le \mathscr{C}^{LD}_{R_q,R_\infty}(\pi, \Delta_0, 1) \le \frac{d}{\nu}\exp\left(\frac{C_q\Delta_0}{\nu}\right).$$

Further, LMC with an appropriate step size $h > 0$ *satisfies*

$$\frac{d}{4h\nu}\exp\left(\frac{\Delta_0}{\nu}\right) \le \mathscr{C}^{LMC}_{R_q,R_\infty}(\pi, h, \Delta_0, 1) \le \frac{d}{h\nu}\exp\left(\frac{C_q\Delta_0}{\nu}\right).$$

The above result states that the exponential dependence on the initial error for LMC and the diffusion is unavoidable, unless there is a good initialization available. Notice that, by Proposition 11, to obtain $R_q(\mu_N \,\|\, \pi) \le \varepsilon$, step size needs to be sufficiently small. Incorporating this into the above bound, we obtain the following lower bound for the iteration complexity of LMC when $d \ge \nu - p$:

$$\mathscr{C}^{\mathrm{LMC}}_{R_q,R_{q'}}(\pi, h, \Delta_0, \varepsilon) \ge \frac{(d+\nu)^2}{48e\nu}\min\left\{\varepsilon^{-1}, \frac{\nu - p}{p}, \frac{(\nu-p)d}{(2+p)\nu}\right\}\exp\left(\frac{\Delta_0}{\nu}\right),$$

where $p = \frac{2q}{q-1}$. Note that the step-size bound leads to additional dependence on d and ε. However, the dependence on ε only reveals itself when ν and d are relatively large.

5. Conclusion

We provided convergence guarantees for LMC and Langevin diffusion, for target distributions $\pi \propto e^{-V}$ satisfying a WPI. This corresponds to potentials with tails that behave like $V(x) \sim \|x\|^\alpha$ for $\alpha \in (0,1]$, and also covers Cauchy-type densities in the case $\alpha \downarrow 0$. Our results push the program initiated by Vempala and Wibisono (2019) to its limits; by providing explicit WPI constants for specific examples, we obtained guarantees demonstrating that targets with heavier tails lead to a worse dependence on the initial error. Particularly, the dependence on initial error is a polynomial of order $\frac{(2-\alpha)^2}{2\alpha}$ for $\alpha > 0$, with a phase transition at $\alpha = 0$, i.e. Cauchy-type logarithmic tails, for which the dependence becomes exponential. Additionally, we established lower bounds under generic tail growth conditions that asserted such dependence on the initial error is unavoidable unless suitable warm start initializations are available.

Our quantitative results (via upper and lower bounds) highlight the precise limitations of LMC for heavy-tailed sampling and provide further motivations to develop a complete complexity theory of heavy-tailed sampling by discretizing other diffusions like specific Itô diffusions or stable-driven diffusions, an area of research which is still in its infancy.

One limitation of our upper bounds is the fact that the step size needs to be chosen in advance according to the number of iterations, and with a fixed step size, the upper bounds diverge as $N \to \infty$. However, as pointed out by Chewi et al. (2022a), this is a general limitation of any analysis that does not assume the log-Sobolev inequality. We leave the stability of fixed step size LMC in the number of iterations under heavy-tailed targets as an open direction for future research.

Acknowledgments

TF was supported by the Engineering and Physical Sciences Research Council (EP/T5178) and by the DeepMind scholarship. KB was supported in part by NSF grant DMS-2053918. MAE was supported by NSERC Grant [2019-06167], CIFAR AI Chairs program, CIFAR AI Catalyst grant, and Data Sciences Institute at University of Toronto.

References

Shigeki Aida. Uniform positivity improving property, Sobolev inequalities, and spectral gaps. *Journal of functional analysis*, 158(1):152–185, 1998.

Christophe Andrieu, Anthony Lee, Sam Power, and Andi Q Wang. Poincaré inequalities for Markov chains: a meeting with Cheeger, Lyapunov and Metropolis. *arXiv preprint arXiv:2208.05239*, 2022.

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.

Krishnakumar Balasubramanian, Sinho Chewi, Murat A Erdogdu, Adil Salim, and Shunshi Zhang. Towards a theory of non-log-concave sampling: First-order stationarity guarantees for Langevin monte carlo. In *Conference on Learning Theory*, pages 2896–2923. PMLR, 2022.

Joris Bierkens, Gareth O Roberts, and Pierre-André Zitt. Ergodicity of the zigzag process. *The Annals of Applied Probability*, 29(4):2266–2301, 2019.

Jeremiah Birrell, Paul Dupuis, Markos A Katsoulakis, Luc Rey-Bellet, and Jie Wang. Variational representations and neural network estimation of Rényi divergences. *SIAM Journal on Mathematics of Data Science*, 3(4):1093–1116, 2021.

Sergey G Bobkov and Michel Ledoux. Weighted Poincaré-type inequalities for Cauchy and other convex measures. *The Annals of Probability*, 37(2):403–427, 2009.

Yu Cao, Jianfeng Lu, and Lihan Wang. On explicit L^2-convergence rate estimate for underdamped Langevin dynamics. *arXiv preprint arXiv:1908.04746*, 2019.

Patrick Cattiaux, Ivan Gentil, and Arnaud Guillin. Weak logarithmic Sobolev inequalities and entropic convergence. *Probab. Theory Relat. Fields*, 139(3-4):563–603, November 2007.

Patrick Cattiaux, Nathael Gozlan, Arnaud Guillin, and Cyril Roberto. Functional inequalities for heavy tailed distributions and application to isoperimetry. *Electronic Journal of Probability*, 15: 346–385, 2010.

Patrick Cattiaux, Djalil Chafaı, and Arnaud Guillin. Central limit theorems for additive functionals of ergodic Markov diffusions processes. *Latin American Journal of Probability and Mathematical Statistics (ALEA)*, 9(2):337–382, 2012.

Karthekeyan Chandrasekaran, Amit Deshpande, and Santosh Vempala. Sampling s-concave functions: The limit of convexity based isoperimetry. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 420–433. Springer, 2009.

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

Hong-Bin Chen, Sinho Chewi, and Jonathan Niles-Weed. Dimension-free log-sobolev inequalities for mixture distributions. *Journal of Functional Analysis*, 281(11):109236, 2021.

Xiang Cheng, Niladri S Chatterji, Peter L Bartlett, and Michael I Jordan. Underdamped Langevin MCMC: A non-asymptotic analysis. In *Conference on learning theory*, pages 300–323. PMLR, 2018.

Sinho Chewi. *Log-concave sampling*. 2023. Book draft available at https://chewisinho.github.io/.

Sinho Chewi, Chen Lu, Kwangjun Ahn, Xiang Cheng, Thibaut Le Gouic, and Philippe Rigollet. Optimal dimension dependence of the metropolis-adjusted langevin algorithm. In *Conference on Learning Theory*, pages 1260–1300. PMLR, 2021.

Sinho Chewi, Murat A Erdogdu, Mufan Li, Ruoqi Shen, and Shunshi Zhang. Analysis of Langevin Monte Carlo from Poincaré to Log-Sobolev. In *Conference on Learning Theory*, pages 1–2. PMLR, 2022a.

Sinho Chewi, Patrik Gerber, Holden Lee, and Chen Lu. Fisher information lower bounds for sampling. *arXiv preprint arXiv:2210.02482*, 2022b.

Sinho Chewi, Patrik R Gerber, Chen Lu, Thibaut Le Gouic, and Philippe Rigollet. The query complexity of sampling from strongly log-concave distributions in one dimension. In Po-Ling Loh and Maxim Raginsky, editors, *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 2041–2059. PMLR, 2022c.

Arnak Dalalyan. Further and stronger analogy between sampling and optimization: Langevin monte carlo and gradient descent. In *Conference on Learning Theory*, pages 678–689. PMLR, 2017a.

Arnak S Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, pages 651–676, 2017b.

Arnak S Dalalyan and Lionel Riou-Durand. On sampling from a log-concave density using kinetic langevin diffusions. *Bernoulli*, 26(3):1956–1988, 2020.

Arnak S Dalalyan and Alexandre B Tsybakov. Sparse regression learning by aggregation and Langevin Monte-Carlo. *Journal of Computer and System Sciences*, 78(5):1423–1443, 2012.

George Deligiannidis, Alexandre Bouchard-Côté, and Arnaud Doucet. Exponential ergodicity of the bouncy particle sampler. *Annals of Statistics*, 47(3), 2019.

Alain Durmus and Éric Moulines. Nonasymptotic convergence analysis for the unadjusted langevin algorithm. *The Annals of Applied Probability*, 27(3):1551–1587, June 2017.

Alain Durmus and Éric Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli*, 25(4A):2854 – 2882, 2019.

Alain Durmus, Szymon Majewski, and Błażej Miasojedow. Analysis of Langevin Monte Carlo via convex optimization. *The Journal of Machine Learning Research*, 20(1):2666–2711, 2019.

Alain Durmus, Arnaud Guillin, and Pierre Monmarché. Geometric ergodicity of the bouncy particle sampler. *Annals of applied probability*, 30(5):2069–2098, 2020.

Raaz Dwivedi, Yuansi Chen, Martin J Wainwright, and Bin Yu. Log-concave sampling: Metropolis-hastings algorithms are fast! In *Conference on learning theory*, pages 793–797. PMLR, 2018.

Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *The Annals of Probability*, 47(4):1982–2010, 2019.

Murat A Erdogdu and Rasa Hosseinzadeh. On the convergence of Langevin Monte Carlo: The interplay between tail growth and smoothness. In *Conference on Learning Theory*, pages 1776–1822. PMLR, 2021.

Murat A Erdogdu, Lester Mackey, and Ohad Shamir. Global non-convex optimization with discretized diffusions. *Advances in Neural Information Processing Systems*, 31, 2018.

Murat A Erdogdu, Rasa Hosseinzadeh, and Shunshi Zhang. Convergence of Langevin Monte Carlo in chi-squared and Rényi divergence. In *International Conference on Artificial Intelligence and Statistics*, pages 8151–8175. PMLR, 2022.

Tyler Farghly and Patrick Rebeschini. Time-independent generalization bounds for SGLD in non-convex settings. In M Ranzato, A Beygelzimer, Y Dauphin, P S Liang, and J Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 19836–19846. Curran Associates, Inc., 2021.

Arun Ganesh and Kunal Talwar. Faster differentially private samplers via Rényi divergence analysis of discretized Langevin MCMC. *Advances in Neural Information Processing Systems*, 33:7222–7233, 2020.

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.

Ye He, Krishnakumar Balasubramanian, and Murat A Erdogdu. On the ergodicity, bias and asymptotic normality of randomized midpoint sampling method. *Advances in Neural Information Processing Systems*, 33:7366–7376, 2020.

Ye He, Krishnakumar Balasubramanian, and Murat A Erdogdu. Heavy-tailed sampling via transformed unadjusted Langevin algorithm. *arXiv preprint arXiv:2201.08349*, 2022.

Søren F Jarner and Gareth O Roberts. Convergence of Heavy-tailed Monte Carlo Markov Chain Algorithms. *Scandinavian Journal of Statistics*, 34(4):781–815, 2007.

Leif Johnson and Charles Geyer. Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *The Annals of Statistics*, pages 3050–3076, 2012.

Kengo Kamatani. Efficient strategy for the Markov chain Monte Carlo in high-dimension with heavy-tailed target probability distribution. *Bernoulli*, 24(4B):3711–3750, 2018.

Andrea Laforgia and Pierpaolo Natalini. On some inequalities for the Gamma function. *Advances in Dynamical Systems and Applications*, 8(2):261–267, 2013.

Yin Tat Lee, Ruoqi Shen, and Kevin Tian. Logsmooth gradient concentration and tighter runtimes for Metropolized Hamiltonian Monte Carlo. In *Conference on learning theory*, pages 2565–2597. PMLR, 2020.

Yin Tat Lee, Ruoqi Shen, and Kevin Tian. Lower bounds on Metropolized sampling methods for well-conditioned distributions. *Advances in Neural Information Processing Systems*, 34:18812–18824, 2021.

Ruilin Li, Hongyuan Zha, and Molei Tao. Sqrt(d) Dimension Dependence of Langevin Monte Carlo. In *The International Conference on Learning Representations*, 2022.

Xuechen Li, Yi Wu, Lester Mackey, and Murat A Erdogdu. Stochastic Runge-Kutta accelerates Langevin Monte Carlo and beyond. *Advances in neural information processing systems*, 32, 2019.

Thomas M Liggett. L_2 Rates of Convergence for Attractive Reversible Nearest Particle Systems: The Critical Case. *The Annals of Probability*, 19(3):935–959, 1991.

Yuan Liu. The Poincaré inequality and quadratic transportation-variance inequalities. *Electron. J. Probab*, 25(1):1–16, 2020.

P Mathieu. Quand l'inégalité log-Sobolev implique l'inégalité de trou spectral. *Séminaire de probabilités de Strasbourg*, 32:30–35, 1998.

Wenlong Mou, Nicolas Flammarion, Martin J Wainwright, and Peter L Bartlett. Improved bounds for discretization of Langevin diffusions: Near-optimal rates without convexity. *Bernoulli*, 28(3): 1577–1601, 2022.

Maxim Raginsky, Alexander Rakhlin, and Matus Telgarsky. Non-convex learning via stochastic gradient Langevin dynamics: A nonasymptotic analysis. In *Conference on Learning Theory*, pages 1674–1703. PMLR, 2017.

Michael Röckner and Feng-Yu Wang. Weak Poincaré inequalities and L_2 convergence rates of Markov semigroups. *J. Funct. Anal.*, 185(2):564–603, October 2001.

Ruoqi Shen and Yin Tat Lee. The randomized midpoint method for log-concave sampling. *Advances in Neural Information Processing Systems*, 32, 2019.

Umut Şimşekli, Lingjiong Zhu, Yee Whye Teh, and Mert Gurbuzbalaban. Fractional underdamped Langevin dynamics: Retargeting SGD with momentum under heavy-tailed gradient noise. In *International Conference on Machine Learning*, pages 8970–8980, 2020.

Santosh Vempala and Andre Wibisono. Rapid convergence of the Unadjusted Langevin Algorithm: Isoperimetry suffices. *Advances in neural information processing systems*, 32, 2019.

Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2003.

Fengyu Wang. *Functional inequalities Markov semigroups and spectral theory*. Elsevier, 2006.

Keru Wu, Scott Schmidler, and Yuansi Chen. Minimax Mixing Time of the Metropolis-Adjusted Langevin Algorithm for Log-Concave Sampling. *Journal of Machine Learning Research*, 23 (270):1–63, 2022.

Jun Yang, Krzysztof Łatuszyński, and Gareth O Roberts. Stereographic markov chain monte carlo. *arXiv preprint arXiv:2205.12112*, 2022.

Xiaolong Zhang and Xicheng Zhang. Ergodicity of supercritical SDEs driven by α-stable processes and heavy-tailed sampling. *arXiv preprint arXiv:2201.10158*, 2022.

Appendix A. Proofs of Sections 2 and 3

A.1. Proof of Theorem 2

Following Vempala and Wibisono (2019), define the following quantities

$$F_q(\rho \,\|\, \pi) := \mathbb{E}_\pi \left(\frac{\rho}{\pi}\right)^q, \qquad G_q(\rho \,\|\, \pi) := \mathbb{E}_\pi\Big[\left(\frac{\rho}{\pi}\right)^q \left\|\nabla \log \frac{\rho}{\pi}\right\|^2\Big] = \frac{4}{q^2}\mathbb{E}_\pi\left[\left\|\nabla \left(\frac{\rho}{\pi}\right)^{q/2}\right\|^2\right].$$

Then (Vempala and Wibisono, 2019, Lemma 6) shows that along the Langevin diffusion

$$\frac{\partial R_q(\rho_t \,\|\, \pi)}{\partial t} = -\frac{q G_q(\rho_t \,\|\, \pi)}{F_q(\rho_t \,\|\, \pi)}. \tag{8}$$

Furthermore, we have the following estimates on the quantities appearing in our functional inequalities.

Lemma 13 (Vempala and Wibisono (2019)) *Let $f = \left(\frac{\rho_t}{\pi}\right)^{q/2}$. Then, for $q \geq 2$,*

$$\mathbb{E}_\pi\left[\|\nabla f\|^2\right] = \frac{q^2}{4} G_q(\rho_t \,\|\, \pi),$$
$$\mathrm{Var}_\pi(f) \geq F_q(\rho_t \,\|\, \pi)(1 - e^{-R_q(\rho_t\|\pi)}).$$

Proof The equality follows by definition. For the inequality,

$$\begin{aligned}
\mathrm{Var}_\pi(f) &= F_q(\rho_t \,\|\, \pi) - F_{q/2}(\rho_t \,\|\, \pi)^2 \\
&= e^{(q-1)R_q(\rho_t\|\pi)} - e^{(q-2)R_{q/2}(\rho_t\|\pi)} \\
&\geq e^{(q-1)R_q(\rho_t\|\pi)} - e^{(q-2)R_q(\rho_t\|\pi)} \\
&= F_q(\rho_t \,\|\, \pi)\left(1 - e^{-R_q(\rho_t\|\pi)}\right),
\end{aligned}$$

where we used the fact that $q \mapsto R_q(\rho_t \,\|\, \pi)$ is non-decreasing. ∎

Proof of Theorem 2. Suppose π satisfies (WPI) with β_{WPI} and Φ' and assume that $\Phi'\big(\left(\frac{\rho_t}{\pi}\right)^{q/2}\big) \leq \delta_0$ for all $t \in \mathbb{R}_+$. Choosing $f = \left(\frac{\rho_t}{\pi}\right)^{q/2}$, it follows from (WPI) and Lemma 13 that

$$\begin{aligned}
\frac{q^2 G_q(\rho_t \,\|\, \pi)}{4F_q(\rho_t \,\|\, \pi)} &\geq \frac{1 - e^{-R_q(\rho_t\|\pi)}}{\beta_{\mathsf{WPI}}(r)} - \frac{r\Phi(f)}{\beta_{\mathsf{WPI}}(r) F_q(\rho_t \,\|\, \pi)} \\
&\geq \frac{1 - e^{-R_q(\rho_t\|\pi)}}{\beta_{\mathsf{WPI}}(r)} - \frac{r\delta_0}{\beta_{\mathsf{WPI}}(r)}.
\end{aligned}$$

Hence,

$$\frac{\partial R_q(\rho_t \,\|\, \pi)}{\partial t} \leq \frac{-4(1 - e^{-R_q(\rho_t\|\pi)}) + 4r\delta_0}{q\beta_{\mathsf{WPI}}(r)}.$$

Thus with $R_q(\rho_t \,\|\, \pi) \geq 1$ we have

$$\frac{\partial R_q(\rho_t \,\|\, \pi)}{\partial t} \leq \frac{-2 + 4r\delta_0}{q\beta_{\mathsf{WPI}}(r)},$$

and for $R_q(\rho_t \,\|\, \pi) < 1$ we have

$$\frac{\partial R_q(\rho_t \,\|\, \pi)}{\partial t} \leq \frac{-2R_q(\rho_t \,\|\, \pi) + 4r\delta_0}{q\beta_{\mathsf{WPI}}(r)}.$$

Integration and Grönwall's lemma yield

$$R_q(\rho_t \,\|\, \pi) \leq \begin{cases} R_q(\rho_0 \,\|\, \pi) - \frac{2-4r\delta_0}{\beta_{\mathsf{WPI}}(r)q} t, & \text{if} \quad R_q(\rho_0 \,\|\, \pi), R_q(\rho_t \,\|\, \pi) \geq 1 \\ e^{-\frac{2t}{\beta_{\mathsf{WPI}}(r)q}} (R_q(\rho_0 \,\|\, \pi) - 2r\delta_0) + 2r\delta_0, & \text{if} \quad R_q(\rho_0 \,\|\, \pi) < 1. \end{cases}$$

Suppose that $q' = \infty$. In this case, we can choose the original WPI with $\Phi(\cdot) = \mathrm{Osc}(\cdot)^2$. Then, we need to show $\mathrm{Osc}\left(\left(\frac{\rho_t}{\pi}\right)^{q/2}\right)^2 \leq \left\|\frac{\rho_0}{\pi}\right\|^q_{L^\infty(\pi)}$. Notice that $P_t \frac{\rho_0}{\pi} = \frac{\rho_t}{\pi}$ (which one can verify by the tower property of the conditional expectation and the symmetry of the Markov process). Then we have,

$$\mathrm{Osc}\left(\left(P_t \frac{\rho_0}{\pi}\right)^{q/2}\right)^2 \leq \left\|P_t \frac{\rho_0}{\pi}\right\|^q_{L^\infty(\pi)} \leq \left\|\frac{\rho_0}{\pi}\right\|^q_{L^\infty(\pi)},$$

where the second inequality follows from the contraction property of the semigroup. Thus, we can set $\delta_0 = \left\|\frac{\rho_0}{\pi}\right\|^q_{L^\infty(\pi)} = \exp(qR_\infty(\rho_0 \,\|\, \pi))$, which completes the proof for the case of $q' = \infty$.

Now, suppose $q < q' < \infty$. Using $u = \frac{2q'}{q}$ in Proporostion 3, π satisfies a WPI with weighting β and $\Phi'(\cdot) = \left\|\cdot - \mathbb{E}_\pi[\cdot]\right\|^2_{L^{2q'/q}(\pi)}$, where β is given in the statement of Theorem 2. Additionally, by Lemma 34, we have

$$\Phi'\left(\left(\frac{\rho_t}{\pi}\right)^{q/2}\right) \leq \left\|\left(\frac{\rho_t}{\pi}\right)^{q/2}\right\|^2_{L^{2q'/q}(\pi)} = \left\|P_t \frac{\rho_0}{\pi}\right\|^q_{L^{q'}(\pi)} \leq \left\|\frac{\rho_0}{\pi}\right\|^q_{L^{q'}(\pi)} \leq \exp(qR_{q'}(\rho_0 \,\|\, \pi)).$$

Thus in this case, we can choose $\delta_0 = \exp(qR_{q'}(\rho_0 \,\|\, \pi))$, which finishes the proof. ■

A.2. Proof of Proposition 3

By Theorem 2.3 of Röckner and Wang (2001), in order to show that (WPI) holds with Φ' as in (4), it suffices to show that for every mean-zero $f \in \mathcal{C}^\infty(\mathbb{R}^d)$,

$$\|P_t f\|^2_{L^2(\pi)} \leq \xi(t) \|f\|^2_{L^u(\pi)} \tag{9}$$

such that $\xi : \mathbb{R}_+ \to \mathbb{R}_+$ is decreasing and $\lim_{t\to\infty} \xi(t) = 0$. Then, we obtain the (WPI) with

$$\beta'_{\mathsf{WPI}}(r) = 2r \inf_s \frac{1}{s} \xi^{-1}(s \exp(1 - s/r)).$$

To establish (9), a similar argument to that of Cattiaux et al. (2012, Lemma 5.1) shows that given such a function f, there exists a constant $K > 0$ such that,

$$\|P_t f\|^2_{L^2(\pi)} \leq 4^{1+2/u} \|f\|_{L^u(\pi)} \left(\left\|P_t(\tilde{f} - \pi(\tilde{f}))\right\|^2_{L^2(\pi)}\right)^{1-2/u},$$

where $\tilde{f} = K^{-1}(f \wedge K \vee (-K))$. Since $f \in \mathcal{C}^\infty(\mathbb{R}^d)$, $\tilde{f}$ must be infinitely differentiable almost everywhere and thus, in the domain of the generator. Therefore, we may apply Theorem 2.1 of Röckner and Wang (2001) to deduce that

$$\left\|P_t(\tilde{f} - \pi(\tilde{f}))\right\|^2_{L^2(\pi)} \leq \lambda(t)(\Phi(\tilde{f}) + \mathrm{Var}_\pi(\tilde{f})) \leq 5\lambda(t),$$

where

$$\lambda(t) = \inf\{r > 0 \; : \; \frac{1}{2}\beta_{\mathsf{WPI}}(r)\log(1/r) \leq t\}.$$

Hence, (9) holds with $\xi(t) = 5\lambda(t)^{1-2/u}$. By definition, λ and hence ξ are decreasing, and $\lim_{t\to\infty}\xi(t) = 0$. From this, we observe that

$$\beta'_{\mathsf{WPI}}(r) \leq 2\xi^{-1}(r) = 2\lambda^{-1}\left((r/5)^{\frac{u}{u-2}}\right) = \beta_{\mathsf{WPI}}\left((r/5)^{\frac{u}{u-2}}\right)\log\left((5/r)^{\frac{u}{u-2}} \vee 1\right).$$

Finally, suppose π does not satisfy a Poincaré inequality, but satisfies (4) for some β'_{WPI} with $u = 2$. Then, for any $r > 0$ we have

$$(1 - r)\,\mathrm{Var}(f) \leq \beta_{\mathsf{WPI}}(r)\mathbb{E}\left[\|\nabla f\|^2\right].$$

Thus, π satisfies a Poincaré inequality (with a constant at most $2\beta_{\mathsf{WPI}}(1/2)$), which is a contradiction. ■

A.3. Proof of Theorem 4

First, we present the following lemma which enables us to control the discretization error $R_{2q}(\mu_N \,\|\, \rho_T)$. This proposition can be retrieved by a careful evaluation of the terms in the proof of Proposition 25 of Chewi et al. (2022a). We avoid the simplifications made by Chewi et al. (2022a) since that can affect our rate due to $m = \mathcal{O}(d^{1/\alpha})$ for π_α defined in Section 3 with $\alpha \in (0, 1)$, and also T can be exponential in d for specific examples.

Proposition 14 ((Chewi et al., 2022a, Proposition 25)) *Suppose ∇V is s-Hölder continuous with constant L and $\nabla V(0) = 0$. Let $m := \frac{1}{2}\inf\{R : \pi(\|x\| \geq R) \leq \frac{1}{2}\}$. Define*

$$\hat{\pi}(x) \propto \exp\left(-V(x) - \frac{1}{6144T}\max\{\|x\| - 2m, 0\}^2\right). \tag{10}$$

Assume for simplicity $\varepsilon^{-1}, m, L, T, R_2(\rho_0 \,\|\, \hat{\pi}) \geq 1$, Then, for $q \leq \varepsilon^{-1}$, if the step size satisfies

$$h \leq \mathcal{O}_s\left(\frac{\varepsilon^{1/s}}{dq^{1/s}L^{2/s}T^{1/s}}\min\left\{1, \frac{L^{1/s-1}T^{1/(2s)}d}{\varepsilon^{1/(2s)}m^s}, \frac{L^{1/s-1}T^{(1-s^2)/(2s)}d}{\varepsilon^{1/(2s)}R_2(\rho_0 \,\|\, \hat{\pi})^{s/2}\ln(N)^{s/2}}\right\}\right),$$

we have for $T = Nh$

$$R_q(\mu_N \,\|\, \rho_T) \leq \varepsilon.$$

Remark 15 *If the first moment of π is finite, we have $m \leq \mathbb{E}_\pi\big[\|x\|\big]$ by Markov's inequality. In fact, the original result in Chewi et al. (2022a) is presented using $m = \mathbb{E}_\pi\big[\|x\|\big]$. However, the result is still valid with the choice of m in Proposition 14, which is useful for targets with infinite first moment.*

With this proposition in hand, we are ready to state the proof of Theorem 4.
Proof of Theorem 4 By Theorem 2 we have $R_{2q-1}(\rho_T \parallel \pi) \leq \frac{\varepsilon}{2}$. Furthermore, by choosing the step-size as in Proposition 14, we have $R_{2q}(\mu_N \parallel \rho_T) \leq \frac{\varepsilon}{2}$. By the weak triangle inequality of Rényi divergence, we have $R_q(\mu_N \parallel \pi) \leq \varepsilon$ for

$$N = \frac{T}{h} = \Theta_s\left(\frac{dq^{1/s}L^{2/s}T^{1+1/s}}{\varepsilon^{1/s}} \max\left\{1, \frac{\varepsilon^{1/(2s)}m^s}{L^{1/s-1}T^{1/(2s)}d}, \frac{\varepsilon^{1/(2s)}R_2(\rho_0 \parallel \hat{\pi})^{s/2}\ln(N)^{s/2}}{L^{1/s-1}T^{(1-s^2)/(2s)}d}\right\}\right).$$

A further simplification of this result yields the statement of the Theorem. ■

A.4. Computing Weak Poincaré Constants

In this section, we will provide WPI estimates for our model examples $\pi_\nu(x) \propto (1 + \|x\|^2)^{\frac{d+\nu}{2}}$ and $\pi_\alpha(x) \propto \exp(-(1 + \|x\|^2)^{\alpha/2})$.

We will use the following chain of implications to establish WPIs with suitable dimension dependencies:

$$\text{Weighted Poincaré} \quad \Rightarrow \quad \text{Converse Poincaré} \quad \Rightarrow \quad \text{Weak Poincaré}$$

In particular, to obtain a WPI from a converse Poinacré inequality, we can use the following result due to Cattiaux et al. (2010).

Lemma 16 (Theorem 5.1 of Cattiaux et al. (2010)) *Assume π satisfies a converse Poincaré inequality, i.e.*

$$\inf_c \int |f(x) - c|^2\, w(x) d\pi(x) \leq C \int \|\nabla f(x)\|^2\, d\pi(x),$$

for some non-negative weight function w, such that $\int w d\pi < \infty$. Define $G(r) := \inf\{u : \pi(w \leq u) > r\}$. Then, π satisfies a WPI with $\Phi(\cdot) = \mathrm{Osc}(\cdot)^2$ and $\beta_{\mathsf{WPI}}(r) = \frac{C}{G(r)}$.

Hence, our main effort is to establish a converse Poincaré inequality for our model examples. In fact, for generalized Cauchy measures, we can immediately use a result of Bobkov and Ledoux (2009).

Lemma 17 (Corollary 3.2 of Bobkov and Ledoux (2009)) *Let $\pi_\nu(x) \propto (1 + \|x\|^2)^{-(d+\nu)/2}$ with $p > 0$. Then, for any $f \in \mathcal{C}^\infty(\mathbb{R}^d)$, we have the following converse (weighted) Poincaré inequality*

$$\inf_c \int |f(x) - c|^2\, w(x) d\pi_\nu(x) \leq C_{d,\nu} \int \|\nabla f(x)\|^2\, d\pi_\nu(x), \tag{11}$$

for $w(x) = \frac{1}{1+\|x\|^2}$, with

$$C_{d,\nu} := \begin{cases} \frac{1}{d+\nu} & \text{if } \nu \geq d+2 \\ \frac{2}{\nu} & \text{otherwise} \end{cases}.$$

Thus, along with a concentration bound, we can invoke Lemma 16 to estimate β_{WPI} for generalized Cauchy measures.

Proof of Proposition 7 In order to invoke Lemma 16, we need to estimate $\pi(w \le u)$ for the weight function given by Lemma 17 $w(x) = (1 + \|x\|^2)^{-1}$. By Lemma 30, we have

$$\pi(w \le u) = \pi\left(\|x\| \ge \sqrt{u^{-1} - 1}\right) \le (d+\nu)^{\nu/2}(u^{-1}-1)^{-\nu/2}.$$

Choosing $u^{-1} = 1 + (d+\nu)r^{-2/\nu}$, we obtain by invoking Lemma 16

$$\beta_{\mathsf{WPI}}(r) = \frac{2}{\nu} + 2\left(\frac{d}{\nu}+1\right)r^{-2/\nu}.$$

■

Estimating β_{WPI} for π_α is more involved as we do not readily have a suitable converse Poincaré inequality. We will work towards this by first deriving a weighted Poincaré inequality using the perturbation argument.

Lemma 18 *Let* $\pi_\alpha(x) \propto \exp\left(-\left(1+\|x\|^2\right)^{\alpha/2}\right)$ *with* $\alpha \in (0,1)$. *Then for any* $f \in \mathcal{C}^\infty(\mathbb{R}^d)$, *we have the following weighted Poincaré inequality*

$$\mathrm{Var}_{\pi_\alpha}(f) \le eC_{d,\alpha}\int w(x)^2 \left\|\nabla f(x)\right\|^2 d\pi_\alpha(x). \tag{12}$$

for $w(x) = \|x\|^{1-\alpha}$ *with* $C_{d,\alpha}$ *satisfying* (13).

Proof Let $\tilde{\pi}_\alpha(x) \propto \exp\left(-\|x\|^\alpha\right)$. According to (Cattiaux et al., 2010, Proposition 4.7), $\tilde{\pi}_\alpha$ satisfies a weighted Poincaré inequality with weight $w(x)^2 = \|x\|^{2(1-\alpha)}$ and parameter $C_{d,\alpha}$ satisfies

$$\frac{d}{\alpha^3} \le C_{d,\alpha} \le 12\frac{d}{\alpha^3} + \frac{d+\alpha}{\alpha^4} \tag{13}$$

Meanwhile, $\frac{d\pi_\alpha}{d\tilde{\pi}_\alpha} = \exp\left(k(x)\right)$ with $k(x) = -\left(1+\|x\|^2\right)^{\alpha/2} + \|x\|^\alpha + \text{constant}$. Notice that $\mathrm{Osc}(k) = 1$. Then, (12) follows from the perturbation property of the weighted Poincaré inequality. ■

In the next step, by an argument similar to (Bobkov and Ledoux, 2009, Proposition 3.3), we transform the weighted Poincaré inequality for π_α to a converse Poincaré inequality.

Lemma 19 *Let* $\pi_\alpha(x) \propto \exp\left(-\left(1+\|x\|^2\right)^{\alpha/2}\right)$ *with* $\alpha \in (0,1)$. *Let* $\gamma \in (0, 2\alpha]$. *Then for any* $g \in \mathcal{C}^\infty(\mathbb{R}^d)$, *we have the following converse weighted Poincaré inequality*

$$\inf_c \int \frac{|f(x)-c|^2}{\left(a+b\|x\|^2\right)^{1-\alpha+\gamma/2}}d\pi_\alpha(x) \le \left[1-(1-\alpha+\gamma/2)\,b^{\frac{1}{2}}a^{-\frac{1}{2}(\alpha-\gamma/2)}\right]^{-2}\int\left\|\nabla f(x)\right\|^2 d\pi_\alpha(x) \tag{14}$$

with $a = \frac{\gamma(eC_{d,\alpha})^{\frac{2}{\gamma}}}{2(1-\alpha)+\gamma}$, $b = \frac{2(1-\alpha)}{2(1-\alpha)+\gamma}$, *and* $C_{d,\alpha}$ *satisfies* (13)

Proof First we apply Young's inequality to bound the weights in Lemma 18.

$$
\begin{aligned}
eC_{d,\alpha}\|x\|^{2(1-\alpha)} &= \left[(eC_{d,\alpha})^{\frac{2}{2(1-\alpha)+\gamma}}\|x\|^{\frac{4(1-\alpha)}{2(1-\alpha)+\gamma}}\right]^{1-\alpha+\gamma/2} \\
&\le \left[\frac{(eC_{d,\alpha})^{\frac{2}{\gamma}}}{\frac{2(1-\alpha)+\gamma}{\gamma}} + \frac{\|x\|^2}{\frac{2(1-\alpha)+\gamma}{2(1-\alpha)}}\right]^{1-\alpha+\gamma/2} \\
&=: \left(a + b\|x\|^2\right)^{1-\alpha+\gamma/2}
\end{aligned}
$$

Therefore, Lemma 18 yields

$$
\mathrm{Var}_{\pi_\alpha}(g) \le \int \|\nabla g(x)\|^2 \left(a + b\|x\|^2\right)^{1-\alpha+\gamma/2} d\pi_\alpha(x) \tag{15}
$$

The rest of the proof proceeds similarly to that of (Bobkov and Ledoux, 2009, Proposition 3.3). Consider $g(x) = (f(x) - c)\left(a + b\|x\|^2\right)^{-\frac{1-\alpha+\gamma/2}{2}}$, where c is chosen such that g has mean 0. Then we have

$$
\nabla g(x) = \nabla f(x)\left(a + b\|x\|^2\right)^{-\frac{1-\alpha+\gamma/2}{2}} - b\left(1-\alpha+\gamma/2\right)\left(a + b\|x\|^2\right)^{-\frac{1-\alpha+\gamma/2}{2}-1}(f(x)-c)x.
$$

By the elementary inequality $\|u+v\|^2 \le \frac{r}{r-1}\|u\|^2 + r\|v\|^2$ for any $r > 1$, we have

$$
\begin{aligned}
\|\nabla g(x)\|^2 \le{}& \frac{r}{r-1}\|\nabla f(x)\|^2\left(a + b\|x\|^2\right)^{-(1-\alpha+\gamma/2)} \\
&+ rb^2\left(1-\alpha+\gamma/2\right)^2 \frac{\|x\|^2}{a+b\|x\|^2}\frac{1}{\left(a+b\|x\|^2\right)^{\alpha-\gamma/2}}\left(a + b\|x\|^2\right)^{-2(1-\alpha+\gamma/2)}(f(x)-c)^2 \\
\le{}& \frac{r}{r-1}\|\nabla f(x)\|^2\left(a + b\|x\|^2\right)^{-(1-\alpha+\gamma/2)} \\
&+ rb^2\left(1-\alpha+\gamma/2\right)^2 b^{-1}a^{-\alpha+\gamma/2}\left(a + b\|x\|^2\right)^{-2(1-\alpha+\gamma/2)}(f(x)-c)^2
\end{aligned}
$$

Applying (15) to g, we obtain

$$
\begin{aligned}
\int (f(x)-c)^2\left(a + b\|x\|^2\right)^{-(1-\alpha+\gamma/2)} d\pi_\alpha(x) \le{}& \frac{r}{r-1}\int\|\nabla f(x)\|^2 d\pi_\alpha(x) \\
&+ rba^{-\alpha+\gamma/2}\left(1-\alpha+\gamma/2\right)^2\int (f(x)-c)^2\left(a + b\|x\|^2\right)^{-(1-\alpha+\gamma/2)} d\pi_\alpha(x)
\end{aligned}
$$

Since $\gamma \in (0, 2\alpha]$, $b \in (0,1)$, $a > 1$ and $1-\alpha+\gamma/2 \in (1-\alpha, 1)$, we have

$$
\int f(x)^2\left(a + b\|x\|^2\right)^{-(1-\alpha+\gamma/2)} d\pi_\alpha(x) \le \frac{r}{r-1}\frac{1}{1 - rba^{-\alpha+\gamma/2}\left(1-\alpha+\gamma/2\right)^2}\int\|\nabla f(x)\|^2 d\pi_\alpha(x)
$$

Last, (14) follows by choosing the optimal $r^* = b^{-\frac{1}{2}}a^{\frac{1}{2}(\alpha-\gamma/2)}(1-\alpha+\gamma/2)^{-1} > 1$. ■

With this converse Poincaré inequality, we are ready to invoke Lemma 16 and obtain a WPI for π_α.

Proof of Proposition 5 Let a and b be defined according to the statement of Lemma 19. In order to invoke Lemma 16, we need to estimate $\pi_\alpha(w \leq u)$ for the weight $w(x) = \left(a + b\left\|x\right\|^2\right)^{-(1-\alpha+\gamma/2)}$. Using the tail bound of Lemma 29, f

$$\begin{aligned}
\pi_\alpha\left(w(x) \leq u\right) &= \pi_\alpha\left(\|x\| \geq b^{-\frac{1}{2}}\sqrt{u^{-\frac{1}{1-\alpha+\gamma/2}} - a}\right) \\
&\leq e^{\frac{1}{2}} 2^{d/\alpha} \exp\left(-\frac{1}{2}\left(1 + b^{-1}\left(u^{-\frac{1}{1-\alpha+\gamma/2}} - a\right)\right)^{\alpha/2}\right).
\end{aligned}$$

Choosing u such that the above is at most r, we obtain

$$\begin{aligned}
G(r) &:= \inf\left\{u :\ \pi_\alpha\left(w(x) \leq u\right) > r\right\} \\
&\geq \left\{a + b\left[\left(1 + \frac{2d}{\alpha}\ln 2 + 2\ln\left(r^{-1}\right)\right)^{\frac{2}{\alpha}} - 1\right]\right\}^{-(1-\alpha+\gamma/2)} \\
&\geq \left\{a + b\left(1 + \frac{2d}{\alpha}\ln 2 + 2\ln\left(r^{-1}\right)\right)^{\frac{2}{\alpha}}\right\}^{-(1-\alpha+\gamma/2)}.
\end{aligned}$$

Therefore, by Lemma 16 the weak Poincaré constant satisfies

$$\begin{aligned}
\alpha_{\mathsf{WPI}}(r) &\leq \left[1 - (1-\alpha+\gamma/2)\, b^{\frac{1}{2}} a^{-\frac{1}{2}(\alpha-\gamma/2)}\right]^{-2}\left\{a + b\left(1 + \frac{2d}{\alpha}\ln 2 + 2\ln\left(r^{-1}\right)\right)^{\frac{2}{\alpha}}\right\}^{1-\alpha+\gamma/2} \\
&\leq \left(1 - a^{-\frac{1}{2}(\alpha-\gamma/2)}\right)^{-2}\left(a^{1-\alpha+\gamma/2} + \left(1 + \frac{2d}{\alpha}\ln 2 + 2\ln(r^{-1})\right)^{\frac{2(1-\alpha)+\gamma}{\alpha}}\right)
\end{aligned}$$

Recall $a = \frac{\gamma}{2(1-\alpha)+\gamma}(eC_{d,\alpha})^{\frac{2}{\gamma}} < (eC_{d,\alpha})^{\frac{2}{\gamma}}$, $\gamma \in (0, 2\alpha]$, and notice that $\inf_\gamma a > 1$. Therefore,

$$\begin{aligned}
\alpha_{\mathsf{WPI}}(r) &\leq \frac{3^{\frac{2-3\alpha+\gamma}{\alpha}} \vee 1}{\left(1 - a^{-\frac{1}{2}(\alpha-\gamma/2)}\right)^2}\left((eC_{d,\alpha})^{\frac{2(2-2\alpha+\gamma)}{\gamma}} + 1 + \left(\frac{2\ln 2}{\alpha}\right)^{\frac{2-2\alpha+\gamma}{\alpha}} d^{\frac{2-2\alpha+\gamma}{\alpha}} + 2^{\frac{2-2\alpha+\gamma}{\alpha}}\ln\left(r^{-1}\right)^{\frac{2-2\alpha+\gamma}{\alpha}}\right) \\
&\leq C_\alpha\left(d^{\frac{2(2-2\alpha+\gamma)}{\gamma}} + \ln\left(r^{-1}\right)^{\frac{2-2\alpha+\gamma}{\alpha}}\right)
\end{aligned}$$

■

Appendix B. Proofs of Section 4

Our goal in this section is to prove Theorem 10 and Proposition 11. To do so, we begin by introducing a lower bound that compares the second moment of ρ with a certain moment from π in order to lower bound the Rényi divergence between ρ and π, thus allowing us to track the evolution of the second moment of the process rather than the Rényi divergence itself.

Lemma 20 *Let $q > 1$, and ρ and π be a pair of measures with $\rho(\|\cdot\|^2) < \infty$ and $\pi(\|\cdot\|^{\frac{2q}{q-1}}) < \infty$, then the q-Rényi divergence is lower bounded by*

$$R_q(\rho\|\pi) \geq \ln\left(\frac{\rho(\|\cdot\|^2)^{\frac{q}{q-1}}}{\pi(\|\cdot\|^{\frac{2q}{q-1}})}\right).$$

Proof From (Birrell et al., 2021, Theorem 3.1), we have the following variational representation for the Rényi divergence,

$$R_q(\rho \,\|\, \pi) \geq \sup_{\phi}\left\{\frac{q}{q-1}\ln\int \exp((q-1)\phi(x))d\rho(x) - \ln\int \exp(q\phi(x))d\pi(x)\right\},$$

where the supremum runs over all measurable functions. The choice of $\phi(x) = \frac{2}{q-1}\ln(\|x\|)$ proves the statement of the lemma. ■

The following lemma uses the gradient bound condition in (7) to lower bound the decay rate of the second moment for the Langevin diffusion and LMC.

Lemma 21 (Evolution of the Second Moment) *Suppose Eq. (7) holds with $\alpha \in [0, 2]$, and $\mathbb{E}\,\|X_0\|^2 < \infty$. Then,*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\,\|X_t\|^2 \geq 2d - 2b\mathbb{E}\left[\|X_t\|^2\right]^{\alpha/2}.$$

Similarly, LMC satisfies

$$\mathbb{E}\,\|x_{k+1}\|^2 \geq \mathbb{E}\,\|x_k\|^2 - 2bh\mathbb{E}\left[\|x_k\|^2\right]^{\alpha/2} + 2hd.$$

Proof We begin by proving the result for LMC. By the independence of ξ_k and x_k, we have

$$\begin{aligned}\mathbb{E}\,\|x_{k+1}\|^2 &= \mathbb{E}\,\|x_k - h\nabla V(x_k)\|^2 + 2hd \\ &\geq \mathbb{E}\,\|x_k\|^2 - 2h\mathbb{E}\,\langle\nabla V(x_k), x_k\rangle + 2hd \\ &\geq \mathbb{E}\,\|x_k\|^2 - 2bh\mathbb{E}\left[\|x_k\|^2\right]^{\alpha/2} + 2hd,\end{aligned}$$

where the last inequality follows from Cauchy-Schwartz and Jensen's inequalities.

For the diffusion, it follows from Itô's lemma that

$$\|X_t\|^2 = -2\int_0^t \langle\nabla V(X_s), X_s\rangle\,\mathrm{d}s + 2td + 2\sqrt{2}\int_0^t \langle X_s, \mathrm{d}B_s\rangle. \tag{16}$$

We proceed by showing the last term is a martingale and can be removed once taking expectations. For this, it is sufficient to show that X_t is B_t-integrable or equivalently,

$$\mathbb{E}\Big[\int_0^t \|X_s\|^2\mathrm{d}s\Big] < \infty. \tag{17}$$

Indeed, from Itô's lemma, and Tonelli's theorem,

$$\begin{aligned}\mathbb{E}\left\|X_t\right\|^2 &\leq -2\mathbb{E}\left[\int_0^t \langle\nabla V(X_s), X_s\rangle\,\mathrm{d}s\right] + 2td + 2\sqrt{2}\mathbb{E}\left[\left(\int_0^t \langle X_s, \mathrm{d}B_s\rangle\right)^2\right]^{1/2} \\ &\leq 2(b+d)t + 2b\mathbb{E}\left[\int_0^t \left\|X_s\right\|^2 \mathrm{d}s\right] + 2\sqrt{2}\mathbb{E}\left[\int_0^t \left\|X_s\right\|^2 \mathrm{d}s\right]^{1/2} \\ &\leq \sqrt{2} + 2(b+d)t + (2b+\sqrt{2})\int_0^t \mathbb{E}\left\|X_s\right\|^2 \mathrm{d}s.\end{aligned}$$

Define $f(s) := \sup_{r\in[0,s]} \mathbb{E}\left[\left\|X_r\right\|^2\right]$. Then, for any $T \geq t$,

$$f(t) \leq \underbrace{\sqrt{2} + 2(b+d)T}_{=:C_1} + \underbrace{(2b+\sqrt{2})}_{=:C_2}\int_0^t f(s)\mathrm{d}s.$$

Then, by Grönwall's lemma,

$$\int_0^t f(s)ds \leq \frac{C_1}{C_2}(\exp(C_2 t) - 1) < \infty,$$

and consequently (17) is satisfied. Thus, we can remove the last term in (16) after taking expectation, and after taking a time derivative obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left\|X_t\right\|^2 &= -2\mathbb{E}\left[\langle\nabla V(X_t), X_t\rangle\right] + 2d \\ &\geq -2b\mathbb{E}\left\|X_t\right\|^\alpha + 2d \\ &\geq -2b\mathbb{E}\left[\left\|X_t\right\|^2\right]^{\alpha/2} + 2d,\end{aligned}$$

where once again the last inequality is implied by Cauchy-Schwartz and Jensen's inequalities. ■

Another key ingredient of our proof will be controlling the Rényi or KL divergence using the variance of an isotropic Gaussian initialization, which is provided by the following lemmas.

Lemma 22 (Controlling Rényi Divergence by Initial Variance) *Let $\pi \propto e^{-V}$, $Z := \int e^{-V(x)}dx$ and suppose V satisfies (7). Then the following holds:*

1. *If $\alpha = 0$, then for $\sigma^2 \geq (d+\nu)^{-1}$, where we recall $\nu := b - d > 0$,*

$$R_\infty(\mathcal{N}_{\sigma^2 I_d} \,\|\, \pi) \leq \frac{\nu}{2}\ln\sigma^2 + \ln\left\{\left(\frac{Z}{(2\pi)^{d/2}}\right)\left(\frac{d+\nu}{e}\right)^{\frac{d+\nu}{2}}\right\} + \frac{1}{2\sigma^2}.$$

2. *If $\alpha \in (0,2)$, then for $\sigma^2 \geq b^{-1}$,*

$$R_\infty(\mathcal{N}_{\sigma^2 I_d} \,\|\, \pi) \leq \frac{(b\sigma^2)^{\frac{2}{2-\alpha}}}{\alpha} + \ln\left(\frac{Z}{(2\pi\sigma^2)^{d/2}}\right) + \frac{1}{2\sigma^2}.$$

Proof We begin by upper bounding V. Using (7), we have that

$$V(x) \leq \int_0^1 \|\nabla V(tx)\| \, \|x\| \, \mathrm{d}t \leq \begin{cases} \frac{b}{\alpha}\left((1+\|x\|^2)^{\alpha/2}-1\right) & \text{if} \quad \alpha > 0, \\ \frac{b}{2}\ln(1+\|x\|^2) & \text{if} \quad \alpha = 0. \end{cases}$$

In the case that $\alpha = 0$, it follows that

$$R_\infty(\mathcal{N}_{\sigma^2 I_d} \,\|\, \pi) \leq \ln\left(\frac{Z}{(2\pi\sigma^2)^{d/2}}\right) + \sup_x \frac{-\|x\|^2}{2\sigma^2} + \frac{b}{2}\ln(1+\|x\|^2).$$

For $\sigma^2 \geq \frac{1}{a+b}$, the supremum occurs at $\|x^*\|^2 = b\sigma^2 - 1$, and

$$R_\infty(\mathcal{N}_{\sigma^2 I_d} \,\|\, \pi) \leq \frac{b-d}{2}\ln\sigma^2 + \ln\left(\frac{Z}{(2\pi)^{d/2}}\right) + \frac{1}{2\sigma^2} + \frac{b}{2}\ln\frac{b}{e}.$$

Now suppose $\alpha > 0$, then

$$R_\infty(\mathcal{N}_{\sigma^2 I_d} \,\|\, \pi) \leq \ln\left(\frac{Z}{(2\pi\sigma^2)^{d/2}}\right) + \sup_x \frac{-\|x\|^2}{2\sigma^2} + \frac{b}{\alpha}(1+\|x\|^2)^{\alpha/2}.$$

For $\sigma^2 \geq b^{-1}$, the supremum is attained at $\|x^*\|^2 = (b\sigma^2)^{\frac{2}{2-\alpha}} - 1$, and

$$R_\infty(\mathcal{N}_{\sigma^2 I_d} \,\|\, \pi) \leq \frac{b^{\frac{2}{2-\alpha}}}{\alpha}\sigma^{\frac{2\alpha}{2-\alpha}} + \ln\left(\frac{Z}{(2\pi\sigma^2)^{d/2}}\right) + \frac{1}{2\sigma^2}.$$

■

We pause here to state the result of Lemma 22 more explicitly for the generalized Cauchy potential V_ν and the sublinear potential V_α. For the first example, we have the normalizing constant

$$Z_\nu := \frac{\pi^{d/2}\Gamma(\nu/2)}{\Gamma((d+\nu)/2)} \leq \pi^{d/2}\Gamma(\nu/2)\left(\frac{d+\nu}{2e}\right)^{-\frac{d+\nu}{2}+1},$$

where the second inequality holds when $d \geq 2$ using $\Gamma(z) \geq (z/e)^{z-1}$ for $z \geq 1$.

Corollary 23 *Consider the measure $\pi_\nu \propto \exp(-V_\nu)$ with $V_\nu(x) = \frac{d+\nu}{2}\ln(1+\|x\|^2)$. Then, for $d \geq 2$, $\alpha \in (0,2]$, and $\sigma^2 \geq (d+\nu)^{-1}$, we have*

$$R_\infty(\mathcal{N}_{\sigma^2 I_d} \,\|\, \pi_\nu) \leq \frac{\nu}{2}\ln\sigma^2 + \ln\left(2^{\nu/2}\Gamma(\nu/2)\right) + \ln\left(\frac{d+\nu}{2e}\right).$$

The second example, the sub-linear potential $V_\alpha = (1+\|x\|^2)^{\alpha/2} - 1$ with $\alpha \in (0,1)$, satisfies (7) with α, and its normalizing constant can be estimated by

$$Z_\alpha := \int \exp((1+\|x\|^2)^{\alpha/2} - 1)dx \leq \int \exp(\|x\|^\alpha)dx = \frac{\pi^{d/2} d/\alpha \Gamma(d/\alpha)}{\Gamma(d/2+1)} \leq \pi^{d/2}\left(\frac{d}{\alpha}\right)^{d/\alpha - d/2},$$

where we refer to (23) for a proof of the identity, and the second inequality follows from Lemma 33.

Corollary 24 *Consider the measure $\pi_\alpha \propto \exp(-V_\alpha)$ with $V_\alpha = (1+\|x\|^2)^{\alpha/2}$, $\alpha \in (0,1)$, and $d \geq 2$. Then for $\sigma^2 \geq 1/\alpha$,*

$$R_\infty(\mathcal{N}_{\sigma^2 I_d} \,\|\, \pi_\alpha) \leq (\alpha\sigma^2)^{\frac{\alpha}{2-\alpha}} + \left(\frac{d}{\alpha} - \frac{d}{2}\right)\ln(d/\alpha) - \frac{d}{2}\ln(2\sigma^2) + \frac{1}{2\sigma^2}.$$

When working with tail growth of order $\alpha = 2$, a large initial variance can lead to infinity Rényi divergence of any order $q > 1$. Hence, we will instead use the KL divergence as our initial metric in this setting.

Lemma 25 (Controlling KL Divergence by Initial Variance) *Let $Z := \int e^{-V(x)}dx$ and suppose that V satisfies (7) with $\alpha = 2$, then*

$$\mathsf{KL}(\mathcal{N}_{\sigma^2 I_d} \,\|\, \pi) \leq \frac{(b\sigma^2 - 1)d}{2} + \ln\left(\frac{Z}{(2\pi\sigma^2)^{d/2}}\right).$$

Proof First, we upper bound V with

$$V(x) \leq \int_0^1 \|\nabla V(tx)\| \, \|x\| \, \mathrm{d}t \leq \frac{b}{2}\|x\|^2.$$

Thus we have

$$\begin{aligned}\mathsf{KL}(\mathcal{N}_{\sigma^2 I_d} \,\|\, \pi) &= \mathbb{E}_{\mathcal{N}_{\sigma^2 I_d}}\left[\ln\left(\frac{Z}{(2\pi\sigma^2)^{d/2}}\exp\left(\frac{-\|x\|^2}{2\sigma^2} + V(x)\right)\right)\right] \\ &\leq \ln\left(\frac{Z}{(2\pi\sigma^2)^{d/2}}\right) + \frac{(b\sigma^2-1)d}{2}.\end{aligned}$$

■

We are now ready to present the proof of Theorem 10.

Proof of Theorem 10 Notice that to lower bound the time or iteration complexity, it suffices to lower bound the time or iteration complexity for one specific initialization with initial divergence less than Δ_0. Our general strategy will be to use Gaussian initializations with large initial variances, such that Lemma 22 ensures the initial divergence is less than Δ_0, while the time estimate from Lemma 21 provides the lower bound. Throughout this proof, $Z := \int e^{-V}$ denotes the normalizing constant.

1. <u>The case $\alpha = 0$</u>: Suppose that Δ_0 satisfies

$$\Delta_0 \geq \left\{1 + 2\ln\left(\frac{Z((d+\nu)/e)^{\frac{d+\nu}{2}}}{(2\pi)^{d/2}}\right)\right\} \vee \nu\ln\left(\frac{2e\pi(\|\cdot\|^{\frac{2q}{q-1}})^{\frac{q-1}{q}}}{d}\right) \vee \nu. \tag{18}$$

   Choose $\rho_0 = \mathcal{N}(0, \sigma^2 I_d)$ with $\sigma^2 = \exp\left(\frac{\Delta_0}{\nu}\right)$. By Lemma 22 we have

$$R_\infty(\rho_0 \,\|\, \pi) \leq \frac{\Delta_0}{2} + \ln\left\{\left(\frac{Z}{(2\pi)^{d/2}}\right)\left(\frac{d+\nu}{e}\right)^{\frac{d+\nu}{2}}\right\} + \frac{1}{2\sigma^2} \leq \Delta_0.$$

Note that that due to Lemma 20, in order to have $R_q(\rho_T \parallel \pi) \le 1$, T needs to be sufficiently large such that

$$\mathbb{E}\left[\|X_T\|^2\right] \le e^{\frac{q-1}{q}} \pi(\|\cdot\|^{\frac{2q}{q-1}})^{\frac{q-1}{q}}.$$

From Lemma 21 we obtain

$$T \ge \frac{\mathbb{E}\left[\|X_0\|^2\right] - \mathbb{E}\left[\|X_T\|^2\right]}{2\nu},$$

and

$$N \ge \frac{\mathbb{E}\left[\|X_0\|^2\right] - \mathbb{E}\left[\|X_k\|^2\right]}{2h\nu}.$$

Consequently, T needs to satisfy

$$T \ge \frac{d \exp\left(\frac{\Delta_0}{\nu}\right) - e\pi(\|\cdot\|^{\frac{2q}{q-1}})^{\frac{q-1}{q}}}{2\nu} \ge \frac{d \exp\left(\frac{\Delta_0}{\nu}\right)}{4\nu},$$

and N needs to satisfy

$$N \ge \frac{d \exp\left(\frac{\Delta_0}{\nu}\right)}{4h\nu}.$$

2. <u>The case $0 < \alpha < 2$</u>: Suppose that Δ_0 satisfies

$$\Delta_0 \ge \left\{ \frac{b^{\frac{\alpha}{2-\alpha}}}{\alpha} \left(1 \vee \frac{(eZ^2)^{1/d}}{2\pi}\right)^{\frac{\alpha}{2-\alpha}} \right\} \vee \left(\frac{2^{\frac{2}{2-\alpha}} eb\pi(\|\cdot\|^{\frac{2q}{q-1}})^{\frac{q-1}{q}}}{\alpha^{2/\alpha-1} d} \right)^{\frac{\alpha}{2-\alpha}} \vee \frac{1}{\alpha}. \tag{19}$$

This time, we choose $\rho_0 = \mathcal{N}(0, \sigma^2 I_d)$ with $\sigma^2 = \frac{(\alpha\Delta_0)^{\frac{2-\alpha}{\alpha}}}{b}$. Then, (19) ensures $\sigma^2 \ge 1$ and by Lemma 22,

$$R_\infty(\rho_0 \parallel \pi) \le \Delta_0 + \ln\left(\frac{Z}{(2\pi\sigma^2)^{d/2}}\right) + \frac{1}{2\sigma^2} \le \Delta_0.$$

From Lemma 21

$$T \ge \frac{\mathbb{E}\left[\|X_0\|^2\right]^{1-\alpha/2} - \mathbb{E}\left[\|X_T\|^2\right]^{1-\alpha/2}}{b(2-\alpha)},$$

hence we can write

$$\begin{aligned} T &\ge \frac{\left(\frac{\alpha^{2/\alpha-1}}{b}\right)^{1-\alpha/2} d^{1-\alpha/2} \Delta_0^{\frac{(2-\alpha)^2}{2\alpha}} - (e\pi(\|\cdot\|^{\frac{2q}{q-1}})^{\frac{q-1}{q}})^{1-\alpha/2}}{b(2-\alpha)} \\ &\ge \frac{\left(\frac{\alpha^{2/\alpha-1}}{b}\right)^{1-\alpha/2} d^{1-\alpha/2} \Delta_0^{\frac{(2-\alpha)^2}{2\alpha}}}{2(2-\alpha)b}. \end{aligned}$$

For the discrete-time case, from Lemma 21 we obtain

$$\mathbb{E}\left[\|x_{k+1}\|^2\right] \geq \mathbb{E}\left[\|x_k\|^2\right] - 2hb\mathbb{E}\left[\|x_k\|^2\right]^{\alpha/2}.$$

Let $r_k := \mathbb{E}\left[\|x_k\|^2\right]$. Suppose $r_k \geq r_{k+1}$, rearranging the inequality above, we obtain

$$2hb \geq r_k^{1-\alpha/2} - r_{k+1}r_k^{-\alpha/2} \geq r_k^{1-\alpha/2} - r_{k+1}^{1-\alpha/2}.$$

On the other hand, when $r_k < r_{k+1}$,

$$2hb > 0 > r_k^{1-\alpha/2} - r_{k+1}^{1-\alpha/2}.$$

Thus the bound holds in either case, and by iterating it we have

$$N \geq \frac{\mathbb{E}\left[\|X_0\|^2\right]^{1-\alpha/2} - \mathbb{E}\left[\|X_T\|^2\right]^{1-\alpha/2}}{hb(2-\alpha)} \geq \frac{\left(\frac{\alpha^{2/\alpha-1}}{\tilde{a}}\right)^{1-\alpha/2} d^{1-\alpha/2}\Delta_0^{\frac{(2-\alpha)^2}{2\alpha}}}{2(2-\alpha)hb},$$

where the second inequality follows analogously to the continuous-time case.

3. The case $\alpha = 2$: Suppose that Δ_0 satisfies

$$\Delta_0 \geq \frac{bZ^{2/d}e^{2a/d-1}}{4\pi} \vee b\left(e\pi(\|\cdot\|^{\frac{2q}{q-1}})\right)^{\frac{(1+c)(q-1)}{q}}. \tag{20}$$

for any absolute constant $c > 0$. Choose $\rho_0 = \mathcal{N}(0, \sigma^2 I_d)$ with $\sigma^2 = \frac{2\Delta_0}{bd}$. Then, by Lemma 25 we have

$$\mathsf{KL}(\rho_0 \parallel \pi) \leq \Delta_0 - \frac{d}{2} + \ln\left(\frac{Z}{(2\pi\sigma^2)^{d/2}}\right) \leq \Delta_0.$$

Moreover, from Lemma 21, we have

$$T \geq \frac{\ln\left(\mathbb{E}\left[\|X_0\|^2\right]\right) - \ln\left(\mathbb{E}\left[\|X_T\|^2\right]\right)}{2b} \geq \frac{c\ln\left(\frac{\Delta_0}{b}\right)}{2(1+c)b}.$$

Similarly for LMC, when $h < b^{-1}$, we have

$$N \geq \frac{\ln\left(\mathbb{E}\left[\|x_0\|^2\right]\right) - \ln\left(\mathbb{E}\left[\|x_N\|^2\right]\right)}{2hb} \geq \frac{c\ln\left(\frac{\Delta_0}{b}\right)}{2(1+c)b},$$

which completes the proof of the theorem.

■

In order to prove Proposition 11, we need a sharper control on the decay of the second moment of LMC that does not ignore terms of order $O(h^2)$. In the following lemma, we achieve this control in the radially symmetric setting.

Lemma 26 (A Sharper Evolution Inequality for LMC) *Suppose* $\mathbb{E}\left[\|x_0\|^2\right] < \infty$*, the potential is radially symmetric with* $V(x) = f(\|x\|^2)$ *and the function* $g : \mathbb{R}_+ \to \mathbb{R}_+$ *given by* $g(r) = (1 - 2hf'(r))^2 r$ *is convex, then for each* $k \geq 0$*,*

$$\mathbb{E}\|x_{k+1}\|^2 \geq g(\mathbb{E}\left[\|x_k\|^2\right]) + 2hd.$$

Furthermore, if g *is non-decreasing then,* $\mathbb{E}\|x_k\|^2 \geq \|y_k\|^2$ *for each* $k \geq 0$*, where we define* y_k *by*

$$y_{k+1} = y_k - \eta \nabla V(y_k), \quad \|y_0\|^2 = \mathbb{E}\left[\|x_0\|^2\right].$$

Proof Using the independence of the Gaussian perturbations and the fact that $\nabla V(x) = 2f'(\|x\|^2)x$,

$$\mathbb{E}\left[\|x_{k+1}\|^2\right] = \mathbb{E}\|x_k - h\nabla V(x_k)\|^2 + 2hd = \mathbb{E}g(\|x_k\|^2) + 2hd.$$

Using the convexity of g along with Jensen's inequality, we conclude that

$$\mathbb{E}\left[\|x_{k+1}\|^2\right] \geq g(\mathbb{E}\left[\|x_k\|^2\right]) + 2hd.$$

If g is non-decreasing, it follows by comparison that $\mathbb{E}\left[\|x_k\|^2\right] \geq \|y_k\|^2$. ■

Finally, we can present the proof of Proposition 11.

Proof of Proposition 11 First, notice that by Lemma 20, $\inf_k R_q(\mu_k \parallel \pi) < \varepsilon$ is equivalent to $\inf_{k\in\mathbb{N}} \mathbb{E}\left[\|x_k\|^2\right] < \sigma_\varepsilon^2$. Let z_k be the process defined by the update

$$z_{k+1} = g(z_k) + 2hd, \quad z_0 = \mathbb{E}\left[\|x_0\|^2\right]$$

so that, using Lemma 26, we have $\mathbb{E}\left[\|x_k\|^2\right] \geq z_k$. Thus, if it holds that $\inf_{k\in\mathbb{N}} \mathbb{E}\left[\|x_k\|^2\right] < \sigma_\varepsilon^2$ then $\inf_{k\in\mathbb{N}} z_k < \sigma_\varepsilon^2$ must hold also. If this holds, there must be some $k \in \mathbb{N}$ such that $z_{k+1} \leq \sigma_\varepsilon^2$ and $z_k \geq \sigma_\varepsilon^2$. Thus, by the fact that g is non-decreasing,

$$g(\sigma_\varepsilon^2) + 2hd \leq g(z_k) + 2hd = z_{k+1} \leq \sigma_\varepsilon^2.$$

Rearranging this leads to the bound given in the statement. ■

Appendix C. Auxiliary Lemmas

In this section, we prove various moment and tail bounds for generalized Cauchy measures and measures with sublinear potentials, which we use in other proofs of the paper.

Lemma 27 *Consider the measure* $\tilde{\pi}_\alpha(x) \propto \exp(-\lambda\|x\|^\alpha)$ *for* $0 < \alpha \leq d$ *and* $\lambda > 0$*. Then, for any* $p > 0$

$$\mathbb{E}_{\tilde{\pi}_\alpha}\left[\|x\|^p\right] = \lambda^{-p/\alpha}\frac{\Gamma\left(\frac{d+p}{\alpha}\right)}{\Gamma\left(\frac{d}{\alpha}\right)} \leq \lambda^{-p/\alpha}\left(\frac{d+p}{\alpha}\right)^{\frac{p}{\alpha}}. \tag{21}$$

Moreover, for $\alpha \in (0,1)$ *and* $\pi_\alpha(x) \propto \exp(-(1+\lambda^{2/\alpha}\|x\|^2)^{\alpha/2})$*,*

$$\mathbb{E}_{\pi_\alpha}\left[\|x\|^p\right] \leq \lambda^{-p/\alpha}\frac{e\Gamma\left(\frac{d+p}{\alpha}\right)}{\Gamma\left(\frac{d}{\alpha}\right)}. \tag{22}$$

Proof First, consider the case of $\lambda = 1$. We begin by computing the normalizing factor of $\tilde{\pi}_\alpha$. Using the polar coordinates, we have

$$Z := \int \exp(-\|x\|^\alpha)\mathrm{d}x = d\omega_d \int_0^\infty \exp(-r^\alpha) r^{d-1}\mathrm{d}r$$
$$= \frac{d\omega_d}{\alpha}\int_0^\infty \exp(-u)u^{d/\alpha - 1}\mathrm{d}u$$
$$= \frac{d\omega_d}{\alpha}\Gamma\left(\frac{d}{\alpha}\right). \tag{23}$$

Via similar calculations, we obtain

$$\int \exp(-\|x\|^\alpha)\|x\|^p \,\mathrm{d}x = \frac{d\omega_d}{\alpha}\Gamma\left(\frac{d+p}{\alpha}\right).$$

Applying Lemma 33 yields,

$$\mathbb{E}_{\tilde{\pi}_\alpha}\left[\|x\|^p\right] = \frac{\Gamma\left(\frac{d+p}{\alpha}\right)}{\Gamma\left(\frac{d}{\alpha}\right)} \leq \left(\frac{d+p}{\alpha}\right)^{\frac{p}{\alpha}}.$$

Finally, we observe that for $\pi_\alpha(x) \propto (1 + \|x\|^2)^{\alpha/2}$,

$$\mathbb{E}_{\pi_\alpha}\left[\|x\|^p\right] = \frac{\int \exp\left(-(1+\|x\|^2)^{\frac{\alpha}{2}}\right)\|x\|^p \,\mathrm{d}x}{\int \exp\left(-(1+\|x\|^2)^{\frac{\alpha}{2}}\right)\mathrm{d}x} \leq \frac{\int \exp(-\|x\|^\alpha)\|x\|^p\,\mathrm{d}x}{e^{-1}\int \exp(-\|x\|^\alpha)\mathrm{d}x} = \frac{e\Gamma\left(\frac{d+p}{\alpha}\right)}{\Gamma\left(\frac{d}{\alpha}\right)}.$$

For the case of $\lambda > 0$, we use the change of variables formula to show that scaling by $\lambda^{-1/\alpha}$ recovers a random variable with the density given by the case with $\lambda = 1$. ■

Lemma 28 *Consider the measure $\pi_\nu(x) \propto (1 + \|x\|^2)^{-(d+\nu)/2}$ for $\alpha > 0$ with $\nu > p \geq 0$. Then,*

$$\mathbb{E}_{\pi_\nu}\left[\|x\|^p\right] = \frac{d}{d+p}\frac{\Gamma\left(\frac{\nu-p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{\Gamma\left(\frac{d+2+p}{2}\right)}{\Gamma\left(\frac{d+2}{2}\right)} \leq \frac{\Gamma\left(\frac{\nu-p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{d+2+p}{2}\right)^{p/2}.$$

Proof Recall that the normalizing constant of this measure is given by

$$Z_{d,\nu} := \frac{\Gamma\left(\frac{\nu}{2}\right)\pi^{d/2}}{\Gamma\left(\frac{\nu+d}{2}\right)} = \frac{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{d+2}{2}\right)\omega_d}{\Gamma\left(\frac{\nu+d}{2}\right)}.$$

On the other hand, using the polar coordinates, one can observe

$$Z_{d,\nu} = d\omega_d \int (1 + r^2)^{-(\nu+d)/2} r^{d-1}\mathrm{d}r. \tag{24}$$

We proceed to compute the following

$$\mathbb{E}_{\pi_\nu}\left[\|x\|^p\right] = \frac{d\omega_d}{Z_{d,\nu}}\int (1 + r^2)^{-(d+\nu)/2} r^{d+p-1}\mathrm{d}r = \frac{d\omega_d}{Z_{d,\nu}}\frac{Z_{d+p,\nu-p}}{(d+p)\omega_{d+p}},$$

where the second equality follows from a change of variables in (24). The statement of the lemma follows by an application of Lemma 33. ■

Lemma 29 *The measure* $\pi_\alpha \propto \exp\left(-\left(1+\|x\|^2\right)^{\alpha/2}\right)$ *with* $\alpha \in (0,1)$ *satisfies*

$$\pi_\alpha\left(\|x\| \geq R\right) \leq e^{\frac{1}{2}} 2^{d/\alpha} \exp\left(-\frac{1}{2}\left(1+R^2\right)^{\alpha/2}\right). \tag{25}$$

Proof Using the Markov inequality,

$$\begin{aligned}\pi_\alpha\left(\|x\| \geq R\right) &= \pi_\alpha\left\{\exp\left(\frac{1}{2}\left(1+\|x\|^2\right)^{\alpha/2}\right) \geq \exp\left(\frac{1}{2}\left(1+R^2\right)^{\alpha/2}\right)\right\} \\ &\leq \exp\left(-\frac{1}{2}\left(1+R^2\right)^{\alpha/2}\right) \mathbb{E}_{\pi_\alpha}\left[\exp\left(\frac{1}{2}\left(1+\|x\|^2\right)^{\alpha/2}\right)\right].\end{aligned}$$

Using polar coordinates and change of coordinates,

$$\begin{aligned}\mathbb{E}_{\pi_\alpha}\left[\exp\left(\frac{1}{2}\left(1+\|x\|^2\right)^{\alpha/2}\right)\right] &= \frac{\int \exp\left(-\frac{1}{2}\left(1+\|x\|^2\right)^{\alpha/2}\right) \mathrm{d}x}{\int \exp\left(-\left(1+\|x\|^2\right)^{\alpha/2}\right) \mathrm{d}x} \\ &= \frac{\int_0^\infty r^{d-1} \exp\left(-\frac{1}{2}\left(1+r^2\right)^{\alpha/2}\right) \mathrm{d}r}{\int_0^\infty r^{d-1} \exp\left(-\left(1+r^2\right)^{\alpha/2}\right) \mathrm{d}r} \\ &= \frac{e^{-\frac{1}{2}} \int_0^\infty (u+1)^{2/\alpha-1}\left[(u+1)^{2/\alpha}-1\right]^{(d-2)/2} \exp\left(-u/2\right) \mathrm{d}u}{e^{-1} \int_0^\infty (u+1)^{2/\alpha-1}\left[(u+1)^{2/\alpha}-1\right]^{(d-2)/2} \exp\left(-u\right) \mathrm{d}u} \\ &\leq e^{\frac{1}{2}} 2^{d/\alpha},\end{aligned}$$

where we used the change of variables $u = (1+r^2)^{\alpha/2} - 1$. This completes the proof. ■

Lemma 30 *The measure* $\pi_\nu(x) \propto (1+\|x\|^2)^{-(d+\nu)/2}$ *satsifies*

$$\pi_\nu(\|x\| \geq R) \leq (\nu+d)^{\nu/2} R^{-\nu}. \tag{26}$$

Proof Using polar coordinates

$$\pi_\nu(\|x\| \geq R) = \frac{1}{Z}\int_{\|x\|\geq R} (1+\|x\|^2)^{-(d+\nu)/2} \mathrm{d}x = \frac{d\omega_d}{Z}\int_R^\infty (1+r^2)^{-(d+\nu)/2} r^{d-1} \mathrm{d}r \leq \frac{d\omega_d R^{-\nu}}{\nu Z}.$$

And $Z = \frac{\Gamma(\nu/2)\pi^{d/2}}{\Gamma((\nu+d)/2)}$. Hence,

$$\pi_\nu(\|x\| \geq R) \leq \underbrace{\frac{d\Gamma\left(\frac{\nu+d}{2}\right)}{\nu\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{d+2}{2}\right)}}_{=:A_{\nu,d}} R^{-\nu}.$$

Suppose $\nu < 2$. Then by Equation (3.1) of Laforgia and Natalini (2013), we have $\frac{\Gamma((d+\nu)/2)}{\Gamma((d+2)/2)} \leq (d/2)^{\nu/2-1}$. Moreover, when $\nu \geq 2$, using Lemma 33 we have $\frac{\Gamma((d+\nu)/2)}{\Gamma((d+2)/2)} \leq ((d+\nu)/2)^{\nu/2-1}$. Consequeuntly,

$$A_{\nu,d} = \frac{d^{\nu/2}(1/2)^{\nu/2-1}(1 \vee (1+\nu/d)^{\nu/2-1})}{\nu\Gamma(\nu/2)} \leq \frac{(d+\nu)^{\nu/2}}{2^{\nu/2}\Gamma(\nu/2+1)} \leq (d+\nu)^{\nu/2},$$

where we used the fact that $2^{\nu/2}\Gamma(\nu/2+1) \geq \Gamma(1) = 1$. ■

The following Lemma, adapted from Chewi et al. (2022a), shows the existence of isotropic Gaussian initializations such that $R_q(\mu_0 \,\|\, \pi), R_q(\mu_0 \,\|\, \hat{\pi}) = \tilde{O}(d)$.

Lemma 31 *Let $\pi(x) \propto \exp(-V(x))$ such that ∇V is s-Hölder continuous and $\nabla V(0) = 0$. Define $\hat{\mu}$ as in Proposition 14. Let $\mu_0 = \mathcal{N}(0, (2L+1)^{-1} I_d)$. Then,*

$$R_\infty(\mu_0 \,\|\, \pi) \leq 2 + L + V(0) - \min_x V(x) + \frac{d}{2}\ln(12m^2 L), \tag{27}$$

$$R_\infty(\mu_0 \,\|\, \hat{\pi}) \leq 3 + L + V(0) - \min_x V(x) + \frac{d}{2}\ln(12(m+6144T)^2 L). \tag{28}$$

Proof The Lemma is directly based on Lemmas 30 and 31 of Chewi et al. (2022a). ■

The following lemma translates a bound on $R_2(\mu_0 \,\|\, \pi)$ to a bound on $R_2(\mu_0 \,\|\, \hat{\pi})$ when μ_0 is some isotropic Gaussian measure. As we only calculate the former quantity for our model examples, we use this lemma to establish similar bounds for the latter.

Lemma 32 *Suppose $\pi \propto \exp(-V(x))$ and $\hat{\pi} \propto \exp\left(-\hat{V}(x)\right)$ with*

$$\hat{V}(x) = V(x) + \frac{\gamma}{2}\max\{\|x\| - R, 0\}^2$$

for some $\gamma, R > 0$. Then, for any $\sigma^2 \leq \frac{1}{\gamma}$ we have

$$R_2(\mathcal{N}_{\sigma^2 I_d} \,\|\, \hat{\pi}) \leq d\ln 2 + R_2(\mathcal{N}_{2\sigma^2 I_d} \,\|\, \pi).$$

Proof Let $Z := \int \exp(-V(x))\mathrm{d}x$ and $\hat{Z} := \int \exp(-\hat{V}(x))\mathrm{d}x$. Notice that $V \leq \hat{V}$, thus $\hat{Z} \leq Z$. Therefore,

$$\begin{aligned}
R_2(\mathcal{N}_{\sigma^2 I_d} \,\|\, \hat{\pi}) &= \ln\left(\frac{\hat{Z}}{(2\pi\sigma^2)^d}\int \exp\left(\frac{-\|x\|^2}{\sigma^2} + V(x) + \frac{\gamma}{2}\max\{\|x\| - R, 0\}^2\right)\mathrm{d}x\right) \\
&\leq \ln\left(\frac{Z}{(2\pi\sigma^2)^d}\int \exp\left(\frac{-\|x\|^2}{2\sigma^2} + V(x)\right)\mathrm{d}x\right) \\
&= d\ln 2 + R_2(\mathcal{N}_{2\sigma^2 I_d} \,\|\, \pi).
\end{aligned}$$

■

Lemma 33 ((Laforgia and Natalini, 2013, Theorem 3.1)) *Suppose $x \geq y \geq 1$, then*

$$y^{x-y} e^{1-\frac{x}{y}} \leq \frac{\Gamma(x)}{\Gamma(y)} \leq x^{x-y}. \tag{29}$$

Lemma 34 *Suppose $Z \geq 0$ is a non-negative random variable. Then, for any $p \geq 2$,*

$$\mathbb{E}\left[Z^p\right] \geq \mathbb{E}\left[|Z - \mathbb{E}\left[Z\right]|^p\right].$$

Proof Normalize Z such that $\mathbb{E}\left[Z\right] = 1$. Using the inequality $Z^p - |Z-1|^p \geq Z - 1$ for every $Z \geq 0$ and $p \geq 2$ and taking expectations proves the lemma. ■