

LOG-CONCAVE SAMPLING ON COMPACT SUPPORTS: A VERSATILE PROXIMAL FRAMEWORK

Anonymous authors

Paper under double-blind review

ABSTRACT

In this paper, we investigate the theoretical aspects of sampling from strongly log-concave distributions defined on convex and compact supports. We propose a general proximal framework that involves projecting onto the constrained set, which is highly flexible and supports various projection options. Specifically, we consider the cases of Euclidean and Gauge projections, with the latter having the advantage of being performed efficiently using a membership oracle. This framework can be seamlessly integrated with multiple sampling methods. Our analysis focuses on Langevin-type sampling algorithms within the context of constrained sampling. We provide nonasymptotic upper bounds on the W_1 and W_2 errors, offering a detailed comparison of the performance of these methods in constrained sampling.

1 INTRODUCTION

Sampling from probability distributions plays a critical role in various fields of science and engineering, especially when dealing with convex and compact sets (Andrieu et al., 2003; Gelman et al., 1995; Stuart, 2010). In this context, the problem involves sampling from a probability measure ν on such sets, characterized by its density function

$$\nu(x) = \frac{e^{-U(x)}}{\int_{\mathbb{R}^p} e^{-U(s)} ds},$$

Here, $U(x) = f(x) + \ell_{\mathcal{K}}(x)$, where $f(x)$ represents a potential function and $\ell_{\mathcal{K}}(x)$ is an indicator function ensuring x lies within the convex and compact set $\mathcal{K} \subset \mathbb{R}^p$. Specifically, $\ell_{\mathcal{K}}(x)$ takes the form

$$\ell_{\mathcal{K}} := \begin{cases} +\infty & \text{if } x \notin \mathcal{K} \\ 0 & \text{if } x \in \mathcal{K}. \end{cases}$$

Solving this constrained sampling problem is challenging and has garnered considerable interest across various fields, including computer science and statistics. In the realm of computer science, a line of research initiated by Dyer et al. (1991) explored polynomial-time algorithms for uniformly sampling convex bodies. This has been followed by seminal studies on the convergence properties of the Ball Walk and the Hit-and-Run algorithm toward uniform density on a convex body or, more broadly, to log-concave densities (Kannan et al., 1997; Kook et al., 2024; Lovász, 1999; Lovász & Simonovits, 1993; Lovász & Vempala, 2007; Smith, 1984). Other Markov Chain Monte Carlo (MCMC) methods, such as Gibbs sampling (Gelfand et al., 1992) and Hamiltonian Monte Carlo (Brubaker et al., 2012; Gürbüzbalaban et al., 2022; Kook et al., 2022), have also been adapted and enhanced to sample from distributions defined on convex and compact sets.

In recent years, leveraging optimization techniques to facilitate sampling has become a prevalent approach. By formulating the sampling challenge as an optimization problem, methods like projected stochastic gradient descent (Bubeck et al., 2015; 2018; Lamperski, 2021; Lehec, 2023), proximal approaches (Brosse et al., 2017; Durmus et al., 2018; Salim & Richtárik, 2020), particle-based algorithms Li et al. (2022), and mirror descent (Ahn & Chewi, 2021; Chewi et al., 2020; Hsieh et al., 2018; Zhang et al., 2020) have proven effective in navigating the target distribution to generate samples. Further innovations have emerged from the intersection of deep learning and neural networks, leading to the development of novel sampling techniques via generative adversarial networks (Goodfellow et al., 2014) and variational autoencoders (Kingma & Welling, 2013). These advanced

methodologies offer promising pathways for sampling from intricate and high-dimensional distributions, particularly those defined on convex and compact sets (Ortiz-Haro et al., 2022). We will present a detailed discussion comparing our work with previous mentioned studies in Appendix A.

In this work, we tackle the challenge posed by a non-smooth target density ν by employing a proximal method. This method leverages a regularization technique that involves projecting onto set \mathcal{K} , effectively transforming the constrained sampling problem into an unconstrained one. Our framework is versatile, accommodating various projection options, such as the Euclidean projection—which corresponds to the Moreau envelope of the indicator function $\ell_{\mathcal{K}}$ (Rockafellar & Wets, 2009; Durmus et al., 2018; Brosse et al., 2017; Pereyra, 2016)—and the Gauge projection (Lu et al., 2022; Mhammedi, 2022), which can be efficiently performed using a membership oracle. Specifically, we introduce a smooth and strongly convex surrogate distribution that closely mimics the target density. By fine-tuning the regularization parameter, we can significantly reduce approximation errors. The advantageous properties of our surrogate distribution enable the effective application of various sampling techniques. In this work, we specifically explore the vanilla Langevin Monte Carlo (LMC) (Roberts & Tweedie, 1996; Dalalyan, 2017; Durmus & Moulines, 2017; Erdogdu & Hosseinzadeh, 2021; Mousavi-Hosseini et al., 2023; Raginsky et al., 2017; Erdogdu et al., 2018; Mou et al., 2022; Erdogdu et al., 2022), kinetic Langevin Monte Carlo (KLMC) (Cheng et al., 2018; Dalalyan & Riou-Durand, 2020; Shen & Lee, 2019; Ma et al., 2021; Zhang et al., 2023), and the parallelized versions of the midpoint randomization method for these algorithms (Shen & Lee, 2019; Yu & Dalalyan, 2024; He et al., 2020; Yu et al., 2023) (referred to as pRLMC and pRKLMC, respectively), within the context of constrained sampling. We derive the convergence rates for these algorithms in Wasserstein-1 and Wasserstein-2 distances, and provide a detailed comparison of their performance. To this end, we make the following contributions.

- We establish both general upper and lower bounds for the distance between the smooth approximation of the target density and its original form in Wasserstein- q distance for any $q \geq 1$. These bounds are detailed in Proposition 2.1 and Proposition 2.2, respectively.
- In Section 2, we demonstrate that our proposed framework can seamlessly incorporate various projection options, including both Euclidean and Gauge projections.
- In Section 3, we incorporate several MCMC sampling methods, such as (kinetic) Langevin Monte Carlo and the parallelized midpoint randomization method for these algorithms, into our general framework. We present a detailed convergence analysis of these methods in both Wasserstein-1 and Wasserstein-2 distances.

In summary, we develop a comprehensive framework specifically designed for sampling from convex and compact sets, utilizing a regularization technique that involves projecting onto these constrained sets. Our study notably presents an improved error bound for LMC in constrained sampling settings, as well as the first convergence analysis for KLMC, pRLMC, and pRKLMC algorithms in this scenario. We emphasize that the convergence analysis for these three algorithms relies heavily on the smooth properties of the target density—conditions that are not met in the constrained setting we examine. Our proposed framework addresses these challenges and is exceptionally flexible and adaptable, accommodating various projection operators and sampling methods. Additionally, it enables clear comparisons of the behaviors of different sampling methods in constrained sampling scenarios. Overall, this new framework provides valuable theoretical insights into the dynamics of constrained sampling.

Notation. Denote the p -dimensional Euclidean space by \mathbb{R}^p . The letter θ denotes the deterministic vector and its calligraphic counterpart ϑ denotes the random vector. We use \mathbf{I}_p and $\mathbf{0}_p$ to denote, respectively, the $p \times p$ identity and zero matrices. Define the relations $\mathbf{A} \preceq \mathbf{B}$ and $\mathbf{B} \succeq \mathbf{A}$ for two symmetric $p \times p$ matrices \mathbf{A} and \mathbf{B} to mean that $\mathbf{B} - \mathbf{A}$ is positive semi-definite. The gradient and the Hessian of a function $f : \mathbb{R}^p \rightarrow \mathbb{R}$ are denoted by ∇f and $\nabla^2 f$, respectively. Given any pair of measures μ and ν defined on $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$, the Wasserstein- q distance between μ and ν is defined as

$$W_q(\mu, \nu) = \left(\inf_{\varrho \in \Gamma(\mu, \nu)} \int_{\mathbb{R}^p \times \mathbb{R}^p} \|\theta - \theta'\|_2^q d\varrho(\theta, \theta') \right)^{1/q}, \quad q \geq 1,$$

where the infimum is taken over all joint distributions ϱ that have μ and ν as marginals. For a set $\mathcal{K} \subset \mathbb{R}^p$, we use \mathcal{K}^c to denote its complement. The ceiling function maps $x \in \mathbb{R}$ to the smallest integer greater than or equal to x , denoted by $\lceil x \rceil$.

2 SMOOTH APPROXIMATION FOR THE TARGET DENSITY

The absence of smoothness in the target distribution ν presents significant challenges because sampling algorithms often depend on the smoothness properties of the target distribution to effectively explore the space and produce representative samples. Motivated by this, we consider the approximation for $\ell_{\mathcal{K}}$ of the form

$$\ell_{\mathcal{K}}^{\lambda}(\boldsymbol{\theta}) := \frac{1}{2\lambda^2} \|\boldsymbol{\theta} - \mathbf{P}_{\mathcal{K}}(\boldsymbol{\theta})\|_2^2,$$

where $\lambda > 0$ is the tuning parameter and $\mathbf{P}_{\mathcal{K}} : \mathbb{R}^p \rightarrow \mathcal{K}$ denotes a projection operator which projects the vector $x \in \mathbb{R}^p$ onto the set \mathcal{K} . Define $U^{\lambda}(\boldsymbol{\theta}) := f(\boldsymbol{\theta}) + \ell_{\mathcal{K}}^{\lambda}(\boldsymbol{\theta})$, and we define the corresponding surrogate target density ν^{λ}

$$\nu^{\lambda}(\boldsymbol{\theta}) = \frac{e^{-U^{\lambda}(\boldsymbol{\theta})}}{\int_{\mathbb{R}^p} e^{-U^{\lambda}(\boldsymbol{\theta}')} d\boldsymbol{\theta}'}. \quad (2.1)$$

Throughout the paper, we define $\mu_k(\nu) = \int \|\boldsymbol{\theta}\|_2^k \nu(d\boldsymbol{\theta})$ as the k -th moment of the distribution ν ($k \geq 1$), and we assume the convex and compact set \mathcal{K} satisfy the following assumption.

Assumption 2.1. *Given a positive constant $r \in (0, \infty)$, we assume the Euclidean ball centered at the origin with radius r , denoted by $\mathcal{B}_2(r)$, is contained in \mathcal{K}*

$$\mathcal{B}_2(r) = \{\boldsymbol{\theta} \in \mathbb{R}^p : \|\boldsymbol{\theta}\|_2 \leq r\} \subset \mathcal{K}.$$

This assumption has been commonly made in the work of constrained sampling (Lamperski, 2021; Bubeck et al., 2018; Brosse et al., 2017; Gürbüzbalaban et al., 2022). Moreover, we assume that the potential function f satisfies the following assumption.

Assumption 2.2. *The function $f : \mathbb{R}^p \rightarrow \mathbb{R}$ is continuously differentiable, and its Hessian matrix $\nabla^2 f$ satisfies*

$$m\mathbf{I}_p \preceq \nabla^2 f(\boldsymbol{\theta}) \preceq M\mathbf{I}_p, \quad \forall \boldsymbol{\theta} \in \mathbb{R}^p.$$

Moreover, we assume the function f is lower bounded and $\mathbf{0} = \arg \min_{\boldsymbol{\theta} \in \mathcal{K}} f(\boldsymbol{\theta})$.

The assumption that f attains the minimum at the origin simplifies the presentation of our results. All our results will still hold even if this condition is not met. Given the minimizer of the function f over \mathcal{K} , denoted by $\boldsymbol{\theta}_*$, we simply need to shift all the coordinates and consider the constrained set $\mathcal{K} + \boldsymbol{\theta}_*$. Additionally, we note that the approximation U^{λ} inherits the strong convexity of f . Moreover, we require U^{λ} to be integrable, continuously differentiable, and smooth, even if U is not.

Condition 2.1. *The function U^{λ} is m -strongly convex, continuously differentiable, and M^{λ} -smooth. Moreover, it holds that $\int_{\mathbb{R}^p} e^{-U^{\lambda}(\boldsymbol{\theta})} d\boldsymbol{\theta} < \infty$.*

Below, we present two examples of the approximation $\ell_{\mathcal{K}}^{\lambda}$, each utilizing a different, commonly used projection operator. We will demonstrate that these examples effectively approximate the indicator function $\ell_{\mathcal{K}}$ and the resulting potential functions fulfill Condition 2.1.

Example 2.1. [Moreau envelope] *One example of a projection operator is the Euclidean operator, which results in the Moreau envelope of the indicator function $\ell_{\mathcal{K}}$*

$$\ell_{\mathcal{K}}^{E,\lambda}(\boldsymbol{\theta}) = \inf_{\boldsymbol{\theta}' \in \mathbb{R}^p} \left(\ell_{\mathcal{K}}(\boldsymbol{\theta}') + \frac{1}{2\lambda^2} \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2^2 \right) = \frac{1}{2\lambda^2} \|\boldsymbol{\theta} - \mathbf{P}_{\mathcal{K}}^E(\boldsymbol{\theta})\|_2^2,$$

where $\mathbf{P}_{\mathcal{K}}^E$ denotes the Euclidean projection on \mathcal{K} .

Define the corresponding surrogate potential $U^{E,\lambda}(\boldsymbol{\theta}) := f(\boldsymbol{\theta}) + \ell_{\mathcal{K}}^{E,\lambda}(\boldsymbol{\theta})$. This surrogate potential $U^{E,\lambda}$ satisfies the following property.

Lemma 2.1. *Assume Assumptions 2.1 and 2.2 hold, the potential $U^{E,\lambda}$ satisfies Condition 2.1 with $M^{\lambda} = 1/\lambda^2 + M$.*

Example 2.2. [Gauge projection] *Another example of the projection operator is the Gauge projection $\mathbf{P}_{\mathcal{K}}^G : \mathbb{R}^p \rightarrow \mathbb{R}$, which is defined as*

$$\mathbf{P}_{\mathcal{K}}^G(\boldsymbol{\theta}) := \frac{\boldsymbol{\theta}}{g_{\mathcal{K}}(\boldsymbol{\theta})}$$

with a variation of Gauge function (also known as the Minkowski function) $g_{\mathcal{K}} : \mathbb{R}^p \rightarrow \mathbb{R}$ of set \mathcal{K}

$$g_{\mathcal{K}}(\boldsymbol{\theta}) := \inf\{t \geq 1 : \boldsymbol{\theta} \in t\mathcal{K}\}. \quad (2.2)$$

This function encapsulates the scaling required to project $\boldsymbol{\theta}$ onto the set \mathcal{K} .

Set $\ell_{\mathcal{K}}^{G,\lambda}(\boldsymbol{\theta}) = \frac{1}{2\lambda^2} \|\boldsymbol{\theta} - \text{P}_{\mathcal{K}}^G(\boldsymbol{\theta})\|_2^2$. The corresponding surrogate function is defined via $U^{G,\lambda}(\boldsymbol{\theta}) := f(\boldsymbol{\theta}) + \ell_{\mathcal{K}}^{G,\lambda}(\boldsymbol{\theta})$. This surrogate potential $U^{G,\lambda}$ satisfies the following property.

Lemma 2.2. *Assume Assumptions 2.1 and 2.2 hold, the surrogate potential $U^{G,\lambda}$ satisfies Condition 2.1 with $M^\lambda = 1/\lambda^2 + M$.*

In the following, we aim to measure the discrepancy between the surrogate distribution and the target distribution. To this end, we establish an upper bound in W_q distance between the target distribution ν and the approximated distribution ν^λ defined in equation 2.1. Our aim is to demonstrate that as λ approaches zero, the distance between these two distributions converges to zero.

Proposition 2.1. *Under Assumption 2.1, assume that the potential function f is convex with $\boldsymbol{\theta}$ as its minimizer. Then, for any $q \geq 1$ and any $\lambda \in (0, \frac{r}{2(p+q)})$, it holds that*

$$W_q^q(\nu, \nu^\lambda) \leq 3\mu_q(\nu)\lambda(2p+q)/r,$$

where $\mu_q(\nu) = \int \|\boldsymbol{\theta}\|_2^q \nu(d\boldsymbol{\theta})$, $q \geq 1$.

In this proposition, we provide a general and precise quantification of the distance between the approximate distribution ν^λ to the target distribution ν in W_q distance for any $q \geq 1$. When $q = 1$, our result aligns with the bound established for $W_1(\nu, \nu^\lambda)$ as stated in Proposition 5 of Brosse et al. (2017). Notably, our Assumption 2.1 is less restrictive than the assumption outlined in Proposition 5 of that work, as we do not require the domain to be contained within a ball, as specified in Assumption H2 of that paper. In Proposition 2.1, the size of \mathcal{K} is captured in the bounds through the radius r and the q -th moment of the target distribution μ_q . In this context, μ_q plays a role analogous to the radius of the ball containing \mathcal{K} , denoted as R in Brosse et al. (2017), while offering a more precise description of the domain's size.

We also derive the lower bound for the distance between the approximate distribution ν^λ to the target distribution ν in W_q distance, specifically for the scenario where $\mathcal{K} = \mathcal{B}_2(r)$ and the projections discussed in Examples 2.1 and 2.2. In this case, the Gauge projection and Euclidean projection coincide, yielding the following results.

Proposition 2.2. *Let $\mathcal{K} = \mathcal{B}_2(r)$ and assume the potential function f is convex and M -smooth, with $\boldsymbol{0}$ as its minimizer and $f(\boldsymbol{0}) = 0$. Set $\ell_{\mathcal{K}} = \ell_{\mathcal{K}}^{E,\lambda}$. Under Assumptions 2.1, for any $q \geq 1$ and any $\lambda \in (0, \frac{r}{2(p+q)} \wedge \frac{1}{\sqrt{M}})$, it holds that*

$$W_q^q(\nu, \nu^\lambda) \geq \mu_q(\nu) \min \left\{ \left| \left(\sqrt{\frac{\pi}{8}} e^{-\frac{3}{4}Mr^2} \frac{\lambda(p+q)}{r+3\lambda p} + \frac{r}{r+3\lambda p} \right)^{1/q} - 1 \right|^q, \left| \left(\frac{3\lambda(p+q)}{r} + 1 \right)^{1/q} - 1 \right|^q \right\}.$$

This result provides a quantitative assessment of the tightness of the upper bound established in Proposition 2.1. Before we discuss the results of the lower bound, we present the following corollary.

Corollary 2.1. *Under the assumptions stated in Proposition 2.2, when $\lambda = o(r/p)$ and $q = 1$, it holds that*

$$W_1(\nu, \nu^\lambda) \geq C_{M,r} \mu_1(\nu) \frac{p+1}{r} \lambda,$$

where $C_{M,r} > 0$ is a constant that depends on M and r exponentially.

Based on the results outlined in the preceding corollary, when $q = 1$ and with a sufficiently small λ , the lower bound aligns with the rate estimates from Proposition 2.1 up to a constant factor, thereby confirming the optimality of the upper bound for the case of $q = 1$. However, when $q > 1$ and λ is small, the lower bound follows the order $((1 + \lambda)^{1/q} - 1)^q$, which increases at a rate slower than λ . This indicates the suboptimality¹ of the rate stated in Proposition 2.1, highlighting a difference in behavior for higher values of q .

3 PROXIMAL LANGEVIN ALGORITHMS

A central theme of this work is the approximation of the non-smooth potential ν using a carefully crafted smooth surrogate density ν^λ . We note that the convergence analysis for the log-concave

¹We conjecture that this suboptimality arises from the general technique used in our proof. With specific assumptions about the density, a refined lower bound for the Wasserstein distance between ν and ν^λ .

sampling algorithms typically relies heavily on the strongly log-concave and smooth properties of the target density. However, these conditions are not met in the constrained setting considered in this work. Owing to the desirable properties of the surrogate density ν^λ , we now can utilize various sampling methods that are effective for distributions with smooth and strongly convex densities.

In this section, we demonstrate how the proposed scheme can be adapted to various sampling methods that collectively aim to sample from the density ν^λ . We particularly focus on Langevin Monte Carlo (LMC) and its variant, Kinetic Langevin Monte Carlo (KLMC). LMC employs stochastic differential equations to sample effectively from complex distributions, making it a powerful tool in high-dimensional spaces. KLMC enhances efficiency by incorporating kinetic energy dynamics. Additionally, we explore the parallelized randomized midpoint method for these two algorithms, which improves sampling speed by leveraging parallel computations. Due to space constraints, we provide only a brief overview of each algorithm for the convenience of readers in the following subsection. We refer interested readers to the works of Dalalyan (2017); Durmus & Moulines (2017); Shen & Lee (2019); Yu & Dalalyan (2024); He et al. (2020); Yu et al. (2023) for further details on these algorithms.

More specifically, we combine the discretization error with the approximation error analyzed in Section 2 to evaluate errors in Wasserstein-1 and Wasserstein-2 distances between the sample distribution and the target density ν in the context of constrained sampling. Table 1 below offers a comparison of the results for the four sampling methods mentioned above, using the specific projection operators discussed in Example 2.1 and Example 2.2². For simplicity, we omit the constants and logarithmic terms that appear in the bounds. The table illustrates that when the number of parallel steps $R > 1$, the randomized midpoint method markedly enhances the performance of both the vanilla Langevin Monte Carlo and kinetic Langevin Monte Carlo algorithms. When the number of parallel steps $R = 1$, the pRLMC and pRKLMC algorithms correspond exactly to the randomized midpoint method applied to Langevin Monte Carlo (RLMC) and kinetic Langevin Monte Carlo (RKLMC), respectively, as detailed in Yu & Dalalyan (2024). The convergence rate for RLMC is comparable to that of LMC when assessing the error in W_2 distance. Further details are provided following Corollary 3.2.

	LMC	KLMC	pRLMC	pRKLMC
W_1	$\tilde{O}(\varepsilon^{-4})$	$\tilde{O}(\varepsilon^{-4})$	$\tilde{O}(R^{-1/3}\varepsilon^{-10/3})$	$\tilde{O}(R^{-1/3}\varepsilon^{-8/3})$
W_2	$\tilde{O}(\varepsilon^{-6})$	$\tilde{O}(\varepsilon^{-7})$	$\tilde{O}(R^{-1/3}\varepsilon^{-6})$	$\tilde{O}(R^{-1/3}\varepsilon^{-5})$

Table 1: The number of iterations required by $\{L, KL, pRL, pRKL\}$ MC algorithms to achieve an error bounded by $\varepsilon\sqrt{p/m}$ in W_1 distance and W_2 distance. $R \geq 1$ denotes the number of parallel steps.

3.1 LANGEVIN MONTE CARLO (LMC)

Let ϑ_0 be a random vector drawn from a distribution ν^λ on \mathbb{R}^p and let $\mathbf{W} = (\mathbf{W}_t : t \geq 0)$ be a p -dimensional Brownian motion that is independent of ϑ_0 . To sample from the approximation distribution ν^λ , we consider the vanilla Langevin diffusion, which is a strong solution to the stochastic differential equation

$$d\mathbf{L}_t^{\text{LD}} = -\nabla U^\lambda(\mathbf{L}_t^{\text{LD}}) dt + \sqrt{2} d\mathbf{W}_t, \quad t \geq 0, \quad \mathbf{L}_0^{\text{LD}} = \vartheta_0. \quad (3.1)$$

This equation has a unique strong solution, which is a continuous-time Markov process, termed Langevin diffusion. Under the further assumptions on the potential U^λ , such as strong convexity, the Langevin diffusion is ergodic, geometrically mixing and has ν^λ as its unique invariant distribution (Bhattacharya, 1978). Moreover, we can sample from the distribution defined by ν^λ by using a suitable discretization of the Langevin diffusion. LMC algorithm is based on this idea, combining the considerations with the Euler discretization. Specifically, for small values of $h \geq 0$ and

²In this work, we mainly focus on the Wasserstein metric. To our knowledge, the only comparable results are those provided by Brosse et al. (2017) concerning the application of Langevin Monte Carlo. We offer a detailed discussion of this comparison following Corollary 3.1.

$\Delta_h \mathbf{W}_t = \mathbf{W}_{t+h} - \mathbf{W}_t$, the following approximation holds

$$\mathbf{L}_{t+h}^{\text{LD}} = \mathbf{L}_t^{\text{LD}} - \int_0^h \nabla U^\lambda(\mathbf{L}_{t+s}^{\text{LD}}) ds + \sqrt{2} \Delta_h \mathbf{W}_t \approx \mathbf{L}_t^{\text{LD}} - h \nabla U^\lambda(\mathbf{L}_t^{\text{LD}}) + \sqrt{2} \Delta_h \mathbf{W}_t.$$

By repeatedly applying this approximation with a small step-size h , we can construct a Markov chain $(\vartheta_k^{\text{LMC}} : k \in \mathbb{N})$ that converges to the target distribution ν^λ as h goes to zero. More precisely, $\vartheta_k^{\text{LMC}} \approx \mathbf{L}_{kh}^{\text{LD}}$, for $k \in \mathbb{N}$, is given by

$$\vartheta_{k+1}^{\text{LMC}} = \vartheta_k^{\text{LMC}} - h \nabla U^\lambda(\vartheta_k^{\text{LMC}}) + \sqrt{2} (\mathbf{W}_{(k+1)h} - \mathbf{W}_{kh}).$$

We now provide explicit upper bounds for the error of the LMC algorithm in terms of W_1 and W_2 distances within the context of constrained sampling.

Theorem 3.1. *Under Assumptions 2.1 and 2.2, we further assume that the potential U^λ satisfies Condition 2.1. Let the step size $h \in (0, 1/M^\lambda)$ and the tuning parameter $\lambda \in (0, \frac{r}{2(p+2)})$. Then, for every $n \geq 1$, the distribution ν_n^{LMC} of ϑ_n^{LMC} satisfies*

$$W_2(\nu_n^{\text{LMC}}, \nu) \leq e^{-\frac{mnh}{2}} W_2(\nu_0^{\text{LMC}}, \nu) + 2\sqrt{\frac{3\mu_2(\nu)\lambda(2p+2)}{r}} + \sqrt{\frac{2phM^\lambda}{m}}.$$

Moreover, when the initial distribution ν_0^{LMC} is set to be the Dirac measure at the minimizer of f , it holds that

$$W_1(\nu_n^{\text{LMC}}, \nu) \leq e^{-\frac{mnh}{2}} \sqrt{\frac{p}{m}} + \frac{3\mu_1(\nu)\lambda(2p+1)}{r} + \sqrt{\frac{2phM^\lambda}{m}}.$$

The term M^λ introduces an additional factor of λ into the convergence rate. To clarify this dependence on λ , we specify in the following corollary that $M^\lambda = M + 1/\lambda^2$. This specification corresponds to the specific projection operators used in Example 2.1 and Example 2.2. We then optimize λ to obtain the results presented below.

Corollary 3.1. *Let $\varepsilon \in (0, 1)$ be a small number, and $M^\lambda = M + \frac{1}{\lambda^2}$.*

(a) *Set $\lambda = \sqrt{2/3}(ph/m)^{1/4} \sqrt{r/(\mu_1(\nu)(2p+1))}$, choose $h > 0$ and $n \in \mathbb{N}$ so that*

$$h = 2^{-10} 3^{-2} (p/m) r^2 (\mu_1(\nu)(2p+1))^{-2} \varepsilon^4 \quad \text{and} \quad n \geq \frac{2}{mh} \log(2/\varepsilon),$$

then we have $W_1(\nu_n^{\text{LMC}}, \nu) \leq \varepsilon \sqrt{p/m}$.

(b) *Set $\lambda = 2^{-2/3} 3^{-1/3} (rh)^{1/3} (\mu_2(\nu)m)^{-1/3}$, choose $h > 0$ and $n \in \mathbb{N}$ so that*

$$h = 2^{-20} 3^{-2} r^2 (\mu_2(\nu)m)^{-2} \varepsilon^6 \quad \text{and} \quad n \geq \frac{2}{mh} \log(2/\varepsilon),$$

then we have $W_2(\nu_n^{\text{LMC}}, \nu) \leq \varepsilon \sqrt{p/m}$.

According to this corollary, when evaluating the error in W_1 distance, the required sample size n to achieve a prespecified error level is of order $\tilde{O}(\varepsilon^{-4})$. This represents an improvement over the rate of $\tilde{O}(\varepsilon^{-6})$ obtained in Brosse et al. (2017). Additionally, our general framework provides the convergence rate in W_2 for the proximal LMC algorithm, thus complementing the study of the proximal LMC presented in Brosse et al. (2017).

3.2 PARALLELIZED RANDOMIZED MIDPOINT DISCRETIZATION OF VANILLA LANGEVIN DIFFUSION (PRLMC)

As an alternative to the Euler discretization method discussed in Section 3.1 for the stochastic differential equation 3.1, we consider the randomized midpoint method, initially introduced in Shen & Lee (2019), as a different discretization framework. This method enables the discretization and simulation of the Langevin diffusion 3.1, ultimately converging to the distribution ν^λ . Building upon the foundations laid by Shen & Lee (2019); Yu et al. (2023), Yu & Dalalyan (2024) developed a parallel computing scheme for this algorithm, significantly enhancing the efficiency of the sampling process. The formal definition of pRLMC is defined in Algorithm 1. For simplicity, the superscript pRLMC is omitted therein. We now introduce the primary findings in the work, which establish explicit upper bounds for the error associated with the pRLMC algorithm, measured in W_1 and W_2 distances, specifically within the framework of constrained sampling.

Algorithm 1 Parallelized RLMC (pRLMC)

Input: number of parallel steps R , number of sequential iterations Q , step size h , number of iterations K , initial point ϑ_0 .

Output: iterate ϑ_{K+1}

```

1: for  $k = 1$  to  $K$  do
2:   Draw  $U_{kr}$  uniformly from  $[\frac{r-1}{R}, \frac{r}{R}]$ ,  $r = 1, \dots, R$ .
3:   Generate  $\xi_{kr} = \mathbf{W}_{(k+U_{kr})h} - \mathbf{W}_{kh}$ ,  $r = 1, \dots, R$ .
4:   Set  $\vartheta_k^{(0,r)} = \vartheta_k$ ,  $r = 1, \dots, R$ .
5:   for  $q = 1$  to  $Q - 1$  do
6:     for  $r = 1$  to  $R$  in parallel do
7:        $a_{kj} = \min\{\frac{1}{R}, U_{kr} - \frac{j-1}{R}\}$ ,  $j = 1, \dots, r$ 
8:        $\vartheta_k^{(q,r)} = \vartheta_k - h \sum_{j=1}^r a_{kj} \nabla U^\lambda(\vartheta_k^{(q-1,j)}) + \sqrt{2} \xi_{kr}$ .
9:     end for
10:  end for
11:   $\vartheta_{k+1} = \vartheta_k - \frac{h}{R} \sum_{r=1}^R \nabla U^\lambda(\vartheta_k^{(Q-1,r)}) + \sqrt{2}(\mathbf{W}_{(k+1)h} - \mathbf{W}_{kh})$ .
12: end for

```

In parallel

Theorem 3.2. Under Assumptions 2.1 and 2.2, we further assume that the potential U^λ satisfies Condition 2.1. Let the tuning parameter $\lambda \in (0, \frac{r}{2(p+2)})$. Choose the step size h such that $M^\lambda h \leq 1/10$. Then, for every $n \geq 1$, the distribution ν_n^{pRLMC} of $\vartheta_n^{\text{pRLMC}}$ satisfies

$$\begin{aligned} \mathbb{W}_2(\nu_n^{\text{pRLMC}}, \nu) &\leq \left(1 + \sqrt{\kappa M^\lambda h} (0.82(M^\lambda h)^{Q-1} + 0.94M^\lambda h/R)\right) e^{-mnh/2} \mathbb{W}_2(\nu_0^{\text{pRLMC}}, \nu) \\ &\quad + \sqrt{\kappa M^\lambda h} (3.98(M^\lambda h)^{Q-1} + 6.91M^\lambda h/\sqrt{R}) \sqrt{p/m} \\ &\quad \left(2 + \sqrt{\kappa M^\lambda h} (0.82(M^\lambda h)^{Q-1} + 0.94M^\lambda h/R)\right) \sqrt{3\mu_2(\nu)\lambda(2p+2)/r}, \end{aligned}$$

where $\kappa = M^\lambda/m$. Moreover, when the initial distribution ν_0^{pRLMC} is set to be the Dirac measure at the minimizer of f , it holds that

$$\begin{aligned} \mathbb{W}_1(\nu_n^{\text{pRLMC}}, \nu) &\leq \left(1 + \sqrt{\kappa M^\lambda h} (0.82(M^\lambda h)^{Q-1} + 0.94M^\lambda h/R)\right) e^{-mnh/2} \sqrt{p/m} \\ &\quad + \sqrt{\kappa M^\lambda h} (3.98(M^\lambda h)^{Q-1} + 6.91M^\lambda h/\sqrt{R}) \sqrt{p/m} \\ &\quad + 3\mu_1(\nu)\lambda(2p+1)/r. \end{aligned}$$

To our knowledge, this represents the first convergence rate for the pRLMC algorithm within the context of constrained sampling. When $R = 1$, these results recover the convergence rate for the randomized midpoint method applied to Langevin Monte Carlo algorithm (RLMC) (He et al., 2020; Yu et al., 2023; Yu & Dalalyan, 2024) in constrained sampling. Below, we define $M^\lambda = M + 1/\lambda^2$, a setting that corresponds to the projection choices outlined in Examples 2.1 and 2.2. We then optimize λ to derive the following corollary.

Corollary 3.2. Given the number of parallel steps R , we set $Q = \lceil \log R \rceil + 1$. Let $\varepsilon \in (0, 1)$ be a small number, and $M^\lambda = M + \frac{1}{\lambda^2}$.

(a) Set $\lambda = (\frac{2^{4.7}}{3})^{1/5} (\frac{p}{R})^{1/10} (\frac{r}{m\mu_1(\nu)(2p+1)})^{1/5} h^{3/10}$, choose $h > 0$ and $n \in \mathbb{N}$ so that

$$h = 15.6^{-10/3} p^{2/5} m^{-1} \left(\frac{r}{\mu_1(\nu)(2p+1)}\right)^{8/3} R^{1/3} \varepsilon^{10/3} \quad \text{and} \quad n \geq \frac{2}{mh} \log\left(\frac{2.22}{\varepsilon}\right),$$

then we have $\mathbb{W}_1(\nu_n^{\text{pRLMC}}, \nu) \leq \varepsilon \sqrt{p/m}$.

(b) $\lambda = 320^{2/9} 3^{-1/3} (\frac{rp}{m^2\mu_2(\nu)(2p+2)R})^{1/9} h^{1/3}$, choose $h > 0$ and $n \in \mathbb{N}$ so that

$$h = 13^{-6} m^{-7/3} \left(\frac{rp}{\mu_2(\nu)(2p+2)}\right)^{8/3} R^{1/3} \varepsilon^6 \quad \text{and} \quad n \geq \frac{2}{mh} \log\left(\frac{2.22}{\varepsilon}\right),$$

then we have $\mathbb{W}_2(\nu_n^{\text{pRLMC}}, \nu) \leq \varepsilon \sqrt{p/m}$.

Comparing this convergence rate with that obtained for proximal LMC in Corollary 3.1, we observe superior performance of pRLMC over LMC when $R > 1$. When assessing the error in W_2 distance for the case of $R = 1$, the sample complexity $\tilde{O}(\epsilon^{-6})$ aligns with that of LMC. This outcome results from our specific choice of λ during the optimization of the upper bound concerning λ , placing it within a region where the performance of LMC is comparable to that of RLMC.

3.3 KINETIC LANGEVIN MONTE CARLO (KLMC)

In this part, we explore the application of kinetic Langevin diffusion in constrained sampling. Recall that the kinetic Langevin process \mathbf{L}^{KLD} is a solution to a second-order stochastic differential equation that can be informally written as

$$\frac{1}{\gamma} \ddot{\mathbf{L}}_t^{\text{KLD}} + \dot{\mathbf{L}}_t^{\text{KLD}} = -\nabla U^\lambda(\mathbf{L}_t^{\text{KLD}}) + \sqrt{2} \dot{\mathbf{W}}_t, \quad (3.2)$$

with initial conditions $\mathbf{L}_0^{\text{KLD}} = \vartheta_0$ and $\dot{\mathbf{L}}_0^{\text{KLD}} = \mathbf{v}_0$. In equation 3.2, $\gamma > 0$, \mathbf{W} is a standard p -dimensional Brownian motion and dots are used to designate derivatives with respect to time $t \geq 0$. This can be formalized using Itô's calculus and introducing the velocity field \mathbf{V}^{KLD} so that the joint process $(\mathbf{L}^{\text{KLD}}, \mathbf{V}^{\text{KLD}})$ satisfies

$$d\mathbf{L}_t^{\text{KLD}} = \mathbf{V}_t^{\text{KLD}} dt; \quad \frac{1}{\gamma} d\mathbf{V}_t^{\text{KLD}} = -(\mathbf{V}_t^{\text{KLD}} + \nabla U^\lambda(\mathbf{L}_t^{\text{KLD}})) dt + \sqrt{2} d\mathbf{W}_t. \quad (3.3)$$

Similar to the vanilla Langevin diffusion 3.1, the kinetic Langevin diffusion $(\mathbf{L}^{\text{KLD}}, \mathbf{V}^{\text{KLD}})$ is a Markov process that exhibits ergodic properties when the potential U^λ is strongly convex (see Eberle et al. (2019) and references therein). The invariant density of this process is given by

$$p_*(\boldsymbol{\theta}, \mathbf{v}) \propto \exp\{-U^\lambda(\boldsymbol{\theta}) - \frac{1}{2\gamma} \|\mathbf{v}\|^2\}, \quad \text{for all } \boldsymbol{\theta}, \mathbf{v} \in \mathbb{R}^p.$$

Note that the marginal of p_* corresponds to $\boldsymbol{\theta}$ coincides with the target density ν^λ . The kinetic Langevin Monte Carlo (KLMC) algorithm is a discretized version of KLD 3.3, where the term $\nabla U^\lambda(\mathbf{L}_t)$ is replaced by $\nabla U^\lambda(\mathbf{L}_{kh})$ on each interval $[kh, (k+1)h)$. The resulting error bounds are given in the following theorem.

Theorem 3.3. *Under Assumptions 2.1 and 2.2, we further assume that the potential U^λ satisfies Condition 2.1. Let the tuning parameter $\lambda \in (0, \frac{r}{2(p+2)})$. Choose γ and h so that $\gamma \geq 5M^\lambda$ and $\sqrt{\kappa} \gamma h \leq 0.1$, where $\kappa = M^\lambda/m$. Assume that $\vartheta_0^{\text{KLMC}}$ is independent of $\mathbf{v}_0^{\text{KLMC}}$ and that $\mathbf{v}_0^{\text{KLMC}} \sim \mathcal{N}_p(0, \gamma \mathbf{I}_p)$. Then, for any $n \geq 1$, the distribution ν_n^{KLMC} of $\vartheta_n^{\text{KLMC}}$ satisfies*

$$W_2(\nu_n^{\text{KLMC}}, \nu) \leq 2\varrho^n W_2(\nu_0^{\text{KLMC}}, \nu) + 3\sqrt{\frac{3\mu_2(\nu)\lambda(2p+2)}{r}} + 0.05\sqrt{\frac{\varrho^n \mathbb{E}[U^\lambda(\vartheta_0^{\text{KLMC}}) - f(\mathbf{0})]}{m}} + 0.9\gamma h \sqrt{\frac{\kappa p}{m}},$$

where $\varrho = e^{-mh}$. Moreover, let the starting points $\vartheta_0^{\text{KLMC}} = \mathbf{0}$, it holds that

$$W_1(\nu_n^{\text{KLMC}}, \nu) \leq 2\varrho^n \sqrt{\frac{p}{m}} + \frac{3\mu_1(\nu)\lambda(2p+1)}{r} + 0.9\gamma h \sqrt{\frac{\kappa p}{m}}.$$

To our knowledge, these results are the first reported convergence rate for the KLMC algorithm within the context of constrained sampling. Below, we set $M^\lambda = M + 1/\lambda^2$, aligning with the projection methods detailed in Example 2.1 and Example 2.2. We proceed by optimizing λ to establish the subsequent corollary.

Corollary 3.3. *Let $\varepsilon \in (0, 1)$ be a small number. Set $\gamma = 5M^\lambda$, $M^\lambda = M + \frac{1}{\lambda^2}$.*

(a) *Set $\lambda = 3^{3/4} 2^{1/8} r^{1/4} (\mu_1(\nu)(2p+1))^{-1/4} p^{1/8} m^{-1/4} h^{1/4}$, choose $h > 0$ and $n \in \mathbb{N}$ so that*

$$h = 2^{-7/2} r^3 (\mu_1(\nu)(2p+1))^{-3} p^{3/2} m^{-1} \varepsilon^4 \quad \text{and} \quad n \geq \frac{1}{mh} \log(4/\varepsilon),$$

then we have $W_1(\nu_n^{\text{KLMC}}, \nu) \leq \varepsilon \sqrt{p/m}$.

(b) *Set $\lambda = 2^{1/7} 3^{3/7} (r/\mu_2(\nu))^{1/7} (h/m)^{2/7}$, choose $h > 0$ and $n \in \mathbb{N}$ so that*

$$h = 32.26^{-7} m^{-5/2} r^3 \mu_2(\nu)^{-3} \varepsilon^7 \quad \text{and} \quad n \geq \frac{1}{mh} \log(4/\varepsilon),$$

then we have $W_2(\nu_n^{\text{KLMC}}, \nu) \leq \varepsilon \sqrt{p/m}$.

We note that our error bounds depend in part on the synchronous coupling between the KLMC and the KLD 3.3. However, for the vanilla Langevin algorithm, Durmus et al. (2019) have shown that the dependency of the error bound on κ can be improved by employing alternative couplings. We propose that similar improvements might be achievable for the KLMC algorithm in constrained sampling scenarios using non-synchronous coupling. For further insights, interested readers are directed to Yu et al. (2023).

3.4 PARALLELIZED RANDOMIZED MIDPOINT DISCRETIZATION OF KINETIC LANGEVIN DIFFUSION (PRKLMC)

The randomized midpoint method, introduced and studied in Shen & Lee (2019), aims at providing a discretization of the kinetic Langevin process 3.2 that reduces the bias of sampling as compared to more conventional discretizations. The parallel computing of this algorithm is outlined in Yu & Dalalyan (2024). This algorithm is referred to as pRKLMC, and for the convenience of the readers, we restate it in Algorithm 2. To ease the notation, we omit the superscript pRKLMC therein. In

Algorithm 2 Parallelized RKLMC (pRKLMC)

Input: number of parallel steps R , number of sequential iterations Q , step size h , friction coefficient γ , number of iterations K , initial points \mathbf{v}_0 and $\boldsymbol{\vartheta}_0$.

Output: iterates $\boldsymbol{\vartheta}_{K+1}$ and \mathbf{v}_{K+1}

```

1: for  $k = 1$  to  $K$  do
2:   Draw  $U_{k1}, \dots, U_{kR}$  uniformly from  $[0, \frac{1}{R}], \dots, [\frac{R-1}{R}, 1]$ , respectively.
3:   Generate  $\bar{\mathbf{W}}_s = \mathbf{W}_{kh+s} - \mathbf{W}_{kh}$ 
4:   Generate  $\boldsymbol{\xi}_{kr} = \int_0^{U_{kr}h} (1 - e^{-\gamma(U_{kr}h-s)}) d\bar{\mathbf{W}}_s, r = 1, \dots, R.$ 
5:   Set  $\boldsymbol{\vartheta}_k^{(0,r)} = \boldsymbol{\vartheta}_k, r = 1, \dots, R.$ 
6:   for  $q = 1$  to  $Q - 1$  do
7:     for  $r = 1$  to  $R$  in parallel do
8:        $a_{kr} = \frac{1 - e^{-\gamma h U_{kr}}}{\gamma}.$ 
9:        $b_{kj} = \int_{(j-1)h}^{\min(\frac{j}{R}, U_{kr})} (1 - e^{-\gamma(U_{kr}h-s)}) ds, j = 1, \dots, r.$ 
10:       $\boldsymbol{\vartheta}_k^{(q,r)} = \boldsymbol{\vartheta}_k + a_{kr} \mathbf{v}_k - \sum_{j=1}^r b_{kj} \nabla U^\lambda(\boldsymbol{\vartheta}_k^{(q-1,j)}) + \sqrt{2} \boldsymbol{\xi}_{kr}.$ 
11:    end for
12:  end for
13:   $\boldsymbol{\vartheta}_{k+1} = \boldsymbol{\vartheta}_k + \frac{1 - e^{-\gamma h}}{\gamma} \mathbf{v}_k - \sum_{r=1}^R \frac{h}{R} (1 - e^{-\gamma h(1-U_{kr})}) \nabla U^\lambda(\boldsymbol{\vartheta}_k^{(Q-1,r)}) + \sqrt{2} \int_0^h (1 - e^{-\gamma(h-s)}) d\bar{\mathbf{W}}_s.$ 
14:   $\mathbf{v}_{k+1} = e^{-\gamma h} \mathbf{v}_k - \gamma \sum_{r=1}^R \frac{h}{R} e^{-\gamma h(1-U_{kr})} \nabla U^\lambda(\boldsymbol{\vartheta}_k^{(Q-1,r)}) + \sqrt{2} \gamma \int_0^h e^{-\gamma(h-s)} d\bar{\mathbf{W}}_s.$ 
15: end for

```

In parallel

the theorem below, we quantify the error bounds for this algorithm when applied to constrained sampling.

Theorem 3.4. *Under Assumptions 2.1 and 2.2, we further assume that the potential U^λ satisfies Condition 2.1. Let the tuning parameter $\lambda \in (0, \frac{r}{2(p+2)})$. Choose γ and h so that $\gamma \geq 5M^\lambda$ and $\gamma h \leq 0.1\kappa^{-1/6}$, where $\kappa = M^\lambda/m$. Assume that $\boldsymbol{\vartheta}_0^{\text{pRKLMC}}$ is independent of $\mathbf{v}_0^{\text{pRKLMC}}$ and that $\mathbf{v}_0^{\text{pRKLMC}} \sim \mathcal{N}_p(0, \gamma \mathbf{I}_p)$. Then, for any $n \geq 1$, the distribution ν_n^{pRKLMC} of $\boldsymbol{\vartheta}_n^{\text{pRKLMC}}$ satisfies*

$$\begin{aligned}
W_2(\nu_n^{\text{pRKLMC}}, \nu) &\leq 1.8\varrho^n W_2(\nu_0^{\text{pRKLMC}}, \nu) + 2.8\sqrt{\frac{3\mu_2(\nu)\lambda(2p+2)}{r}} + 0.28\sqrt{\frac{\varrho^n \mathbb{E}[U^\lambda(\boldsymbol{\vartheta}_0^{\text{pRKLMC}}) - f(\mathbf{0})]}{m}} \\
&\quad + 44.78\sqrt{\frac{(\gamma M^\lambda h^2)^3}{R^2} + (\gamma M^\lambda h^2)^{2Q-1} \sqrt{\frac{\kappa p}{m}}},
\end{aligned}$$

where $\varrho = e^{-mh}$. Moreover, let the starting points $\boldsymbol{\vartheta}_0^{\text{pRKLMC}} = \mathbf{0}$, it holds that

$$W_1(\nu_n^{\text{pRKLMC}}, \nu) \leq 1.8\varrho^n \sqrt{\frac{p}{m}} + \frac{3\mu_1(\nu)\lambda(2p+1)}{r} + 44.78\sqrt{\frac{(\gamma M^\lambda h^2)^3}{R^2} + (\gamma M^\lambda h^2)^{2Q-1} \sqrt{\frac{\kappa p}{m}}}.$$

To our knowledge, this represents the first convergence analysis for the pRKLMC algorithm within the context of constrained sampling. When $R = 1$, we recover the convergence rate for RKLMC Shen & Lee (2019); Yu et al. (2023); Yu & Dalalyan (2024) when applied in the context of constrained sampling. We define $M^\lambda = M + 1/\lambda^2$, in line with the projection options described in Examples 2.1 and 2.2. We then select the optimal λ to derive the following corollary.

Corollary 3.4. *Given the number of parallel steps R , set $Q = \lceil \log R \rceil + 2$. Let $\varepsilon \in (0, 1)$ be a small number, $\gamma = 5M^\lambda$ with $M^\lambda = M + 1/\lambda^2$.*

(a) *Set $\lambda = (\frac{7.8011}{3})^{1/8} (\frac{rh^3}{R\mu_1(\nu)(2p+1)})^{1/8} (\frac{p}{m^2})^{1/16}$, choose $h > 0$ and $n \in \mathbb{N}$ so that*

$$h = 23.6^{-8/3} p^{7/6} m^{-1} r^{7/3} (\mu_1(\nu)(2p+1))^{-7/3} R^{1/3} \varepsilon^{8/3} \quad \text{and} \quad n \geq \frac{1}{mh} \log\left(\frac{3.6}{\varepsilon}\right),$$

486 then we have $W_1(\nu_n^{\text{pRKLMC}}, \nu) \leq \varepsilon \sqrt{p/m}$.

487
488 (b) Set $\lambda = (\frac{14.8011}{5})^{2/15} (\frac{rp}{m^2 \mu_2(\nu)(2p+2)})^{1/15} (\frac{h^3}{R})^{2/15}$, choose $h > 0$ and $n \in \mathbb{N}$ so that

$$489 \quad h = 20.9^{-5} p^{7/3} m^{-13/6} r^{7/3} (\mu_2(\nu)(2p+2))^{-7/3} R^{1/3} \varepsilon^5 \quad \text{and} \quad n \geq \frac{1}{mh} \log\left(\frac{3.6}{\varepsilon}\right),$$

490
491 then we have $W_2(\nu_n^{\text{pRKLMC}}, \nu) \leq \varepsilon \sqrt{p/m}$.

492
493 The corollary demonstrates that pRKLMC achieves the best performance in terms of sample com-
494 plexity compared to the other three algorithms considered in this work.

495 496 497 4 DISCUSSION

498
499 **Strongly convexity** In this work, we focus on a strongly convex potential function f coupled
500 with specific choices of projection operators, resulting in a strongly log-concave surrogate density
501 ν^λ . However, it is important to note that the approximation error bound between the target density
502 ν and the surrogate density ν^λ , as stated in Proposition 2.1, depends solely on the convexity of f .
503 By relaxing the strong convexity requirement of f and considering alternative projection operators,
504 one can generate a surrogate density ν^λ that is less restrictive than strongly log-concave. This
505 modification enables the extension of these findings to a broader spectrum of sampling methods.

506
507 **Other metrics** The Wasserstein distance we employ is a natural metric for measuring sampling
508 errors due to its relevance to optimal transport theory. However, recent advancements in gradient-
509 based sampling have investigated other metrics, including total variation distance, KL divergence,
510 and χ^2 divergence. An intriguing avenue for future research would be to establish error guarantees
511 for constrained sampling concerning these alternative metrics.

512
513 **Limitation** The primary focus of this work is to provide theoretical insights into the analysis of
514 constrained sampling. We hope that these insights pave the way for empirical evaluations of the
515 performance of Langevin-type algorithms in various settings, as well as for implementations in real
516 data applications, which offer promising directions for future research.

517 REFERENCES

- 518 Kwangjun Ahn and Sinho Chewi. Efficient constrained sampling via the mirror-langevin algorithm.
519 *Advances in Neural Information Processing Systems*, 34:28405–28418, 2021.
- 520 Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to
521 mcmc for machine learning. *Machine learning*, 50:5–43, 2003.
- 522 R. N. Bhattacharya. Criteria for recurrence and existence of invariant measures for multidimensional
523 diffusions. *Ann. Probab.*, 6(4):541–553, 08 1978.
- 524 Nicolas Brosse, Alain Durmus, Éric Moulines, and Marcelo Pereyra. Sampling from a log-concave
525 distribution with compact support with proximal langevin monte carlo. In *Conference on learning
526 theory*, pp. 319–342. PMLR, 2017.
- 527 Marcus Brubaker, Mathieu Salzmann, and Raquel Urtasun. A family of mcmc methods on implicitly
528 defined manifolds. In *Artificial intelligence and statistics*, pp. 161–172. PMLR, 2012.
- 529 Sebastien Bubeck, Ronen Eldan, and Joseph Lehec. Finite-time analysis of projected langevin monte
530 carlo. *Advances in Neural Information Processing Systems*, 28, 2015.
- 531 Sébastien Bubeck, Ronen Eldan, and Joseph Lehec. Sampling from a log-concave distribution with
532 projected langevin monte carlo. *Discrete & Computational Geometry*, 59(4):757–783, 2018.
- 533 Xiang Cheng, Niladri S Chatterji, Peter L Bartlett, and Michael I Jordan. Underdamped Langevin
534 MCMC: A non-asymptotic analysis. In *COLT 2018*, 2018.
- 535 Sinho Chewi, Thibaut Le Gouic, Chen Lu, Tyler Maunu, Philippe Rigollet, and Austin Stromme.
536 Exponential ergodicity of mirror-langevin diffusions. *Advances in Neural Information Processing
537 Systems*, 33:19573–19585, 2020.

- 540 Arnak S. Dalalyan. Theoretical guarantees for approximate sampling from a smooth and log-
541 concave density. *J. R. Stat. Soc. B*, 79:651 – 676, 2017.
- 542
- 543 Arnak S Dalalyan and Lionel Riou-Durand. On sampling from a log-concave density using kinetic
544 Langevin diffusions. *Bernoulli*, 26(3):1956–1988, 2020.
- 545 Alain Durmus and Eric Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin
546 algorithm. *Ann. Appl. Probab.*, 27(3):1551–1587, 06 2017. doi: 10.1214/16-AAP1238. URL
547 <http://dx.doi.org/10.1214/16-AAP1238>.
- 548
- 549 Alain Durmus and Eric Moulines. High-dimensional bayesian inference via the unadjusted langevin
550 algorithm. 2019.
- 551 Alain Durmus, Eric Moulines, and Marcelo Pereyra. Efficient bayesian computation by proximal
552 markov chain monte carlo: when langevin meets moreau. *SIAM Journal on Imaging Sciences*, 11
553 (1):473–506, 2018.
- 554 Alain Durmus, Szymon Majewski, and Błażej Miasojedow. Analysis of langevin monte carlo via
555 convex optimization. *The Journal of Machine Learning Research*, 20(1):2666–2711, 2019.
- 556
- 557 Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approxim-
558 ating the volume of convex bodies. *Journal of the ACM (JACM)*, 38(1):1–17, 1991.
- 559
- 560 Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Couplings and quantitative contraction rates
561 for Langevin dynamics. *Ann. Probab.*, 47(4):1982–2010, 07 2019.
- 562 Murat A Erdogdu and Rasa Hosseinzadeh. On the convergence of langevin monte carlo: The inter-
563 play between tail growth and smoothness. In *COLT 2021*, 2021.
- 564
- 565 Murat A Erdogdu, Lester Mackey, and Ohad Shamir. Global non-convex optimization with dis-
566 cretized diffusions. *NeurIPS 2018*, 2018.
- 567
- 568 Murat A Erdogdu, Rasa Hosseinzadeh, and Shunshi Zhang. Convergence of langevin monte carlo
569 in chi-squared and rényi divergence. In *AISTATS 2022*, 2022.
- 570
- 571 Alan E Gelfand, Adrian FM Smith, and Tai-Ming Lee. Bayesian analysis of constrained parameter
572 and truncated data problems using gibbs sampling. *Journal of the American Statistical Associa-*
573 *tion*, 87(418):523–532, 1992.
- 574
- 575 Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*. Chap-
576 man and Hall/CRC, 1995.
- 577
- 578 Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair,
579 Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information*
580 *processing systems*, 27, 2014.
- 581
- 582 Mert Gürbüzbalaban, Yuanhan Hu, and Lingjiong Zhu. Penalized langevin and hamiltonian monte
583 carlo algorithms for constrained sampling. *arXiv preprint arXiv:2212.00570*, 2022.
- 584
- 585 Ye He, Krishnakumar Balasubramanian, and Murat A Erdogdu. On the ergodicity, bias and asymp-
586 totic normality of randomized midpoint sampling method. *Advances in Neural Information Pro-*
587 *cessing Systems*, 33:7366–7376, 2020.
- 588
- 589 Ya-Ping Hsieh, Ali Kavis, Paul Rolland, and Volkan Cevher. Mirrored langevin dynamics. *Advances*
590 *in Neural Information Processing Systems*, 31, 2018.
- 591
- 592 Ravi Kannan, László Lovász, and Miklós Simonovits. Random walks and an $o^*(n^5)$ volume algo-
593 rithm for convex bodies. *Random Structures & Algorithms*, 11(1):1–50, 1997.

- 594 Yunbum Kook, Santosh S Vempala, and Matthew S Zhang. In-and-out: Algorithmic diffusion for
595 sampling convex bodies. *arXiv preprint arXiv:2405.01425*, 2024.
- 596
- 597 Andrew Lamperski. Projected stochastic gradient langevin algorithms for constrained sampling and
598 non-convex learning. In *Conference on Learning Theory*, pp. 2891–2937. PMLR, 2021.
- 599 Joseph Lehec. The langevin monte carlo algorithm in the non-smooth log-concave case. *arXiv*
600 *preprint arXiv:2101.10695*, 2021.
- 601
- 602 Joseph Lehec. The langevin monte carlo algorithm in the non-smooth log-concave case. *The Annals*
603 *of Applied Probability*, 33(6A):4858–4874, 2023.
- 604 Lingxiao Li, Qiang Liu, Anna Korba, Mikhail Yurochkin, and Justin Solomon. Sampling with
605 mollified interaction energy descent. *arXiv preprint arXiv:2210.13400*, 2022.
- 606
- 607 László Lovász. Hit-and-run mixes fast. *Mathematical programming*, 86:443–461, 1999.
- 608 László Lovász and Miklós Simonovits. Random walks in a convex body and an improved volume
609 algorithm. *Random structures & algorithms*, 4(4):359–412, 1993.
- 610
- 611 László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algo-
612 rithms. *Random Structures & Algorithms*, 30(3):307–358, 2007.
- 613 Zhou Lu, Nataly Brukhim, Paula Gradu, and Elad Hazan. Projection-free adaptive regret with
614 membership oracles. *arXiv preprint arXiv:2211.12638*, 2022.
- 615
- 616 Yi-An Ma, Niladri S. Chatterji, Xiang Cheng, Nicolas Flammarion, Peter L. Bartlett, and Michael I.
617 Jordan. Is there an analog of Nesterov acceleration for gradient-based MCMC? *Bernoulli*, 27(3):
618 1942 – 1992, 2021.
- 619 Zakaria Mhammedi. Efficient projection-free online convex optimization with membership oracle.
620 In *Conference on Learning Theory*, pp. 5314–5390. PMLR, 2022.
- 621
- 622 Wenlong Mou, Nicolas Flammarion, Martin J Wainwright, and Peter L Bartlett. Improved bounds
623 for discretization of Langevin diffusions: Near-optimal rates without convexity. *Bernoulli*, 28(3):
624 1577–1601, 2022.
- 625 Alireza Mousavi-Hosseini, Tyler Farghly, Ye He, Krishnakumar Balasubramanian, and Murat A
626 Erdogdu. Towards a Complete Analysis of Langevin Monte Carlo: Beyond Poincaré Inequality.
627 *arXiv preprint arXiv:2303.03589*, 2023.
- 628
- 629 Joaquim Ortiz-Haro, Jung-Su Ha, Danny Driess, and Marc Toussaint. Structured deep generative
630 models for sampling on constraint manifolds in sequential manipulation. In *Conference on Robot*
631 *Learning*, pp. 213–223. PMLR, 2022.
- 632 Marcelo Pereyra. Proximal markov chain monte carlo algorithms. *Statistics and Computing*, 26:
633 745–760, 2016.
- 634 Maxim Raginsky, Alexander Rakhlin, and Matus Telgarsky. Non-convex learning via stochastic
635 gradient langevin dynamics: a nonasymptotic analysis. In *COLT 2017*, 2017.
- 636
- 637 G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their
638 discrete approximations. *Bernoulli*, 2(4):341–363, 1996.
- 639 R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science &
640 Business Media, 2009.
- 641
- 642 Adil Salim and Peter Richtárik. Primal dual interpretation of the proximal stochastic gradient
643 langevin algorithm. *Advances in Neural Information Processing Systems*, 33:3786–3796, 2020.
- 644 Ruoqi Shen and Yin Tat Lee. The randomized midpoint method for log-concave sampling. *Advances*
645 *in Neural Information Processing Systems*, 32, 2019.
- 646
- 647 Robert L Smith. Efficient monte carlo procedures for generating points uniformly distributed over
bounded regions. *Operations Research*, 32(6):1296–1308, 1984.

648 Andrew M Stuart. Inverse problems: a bayesian perspective. *Acta numerica*, 19:451–559, 2010.
649
650 Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.
651
652 Lu Yu and Arnak Dalalyan. Parallelized midpoint randomization for langevin monte carlo. *arXiv*
653 *preprint arXiv:2402.14434*, 2024.
654
655 Lu Yu, Avetik Karagulyan, and Arnak Dalalyan. Langevin monte carlo for strongly log-concave
656 distributions: Randomized midpoint revisited. *arXiv preprint arXiv:2306.08494*, 2023.
657
658 Kelvin Shuangjian Zhang, Gabriel Peyré, Jalal Fadili, and Marcelo Pereyra. Wasserstein control of
659 mirror langevin monte carlo. In *Conference on Learning Theory*, pp. 3814–3841. PMLR, 2020.
660
661 Matthew Zhang, Sinho Chewi, Mufan Bill Li, Krishnakumar Balasubramanian, and Murat A Er-
662 dogdu. Improved Discretization Analysis for Underdamped Langevin Monte Carlo. *arXiv*
663 *preprint arXiv:2302.08049*, 2023.
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701

A RELATION TO PREVIOUS WORKS ON CONSTRAINED SAMPLING

Sampling from constrained log-concave distributions is a fundamental task across various fields. In computer science, several studies have explored the convergence of the ball walk and hit-and-run algorithms toward the uniform density on a convex body \mathcal{K} , or more generally, a log-concave density (Dyer et al., 1991; Kannan et al., 1997; Kook et al., 2024; Lovász, 1999; Lovász & Simonovits, 1993; Lovász & Vempala, 2007; Smith, 1984). Unlike the first-order sampling methods examined in this work, these algorithms require calls to zero-order oracles at each iteration. Additionally, the performance of these geometric random walks is influenced by the skewed shape of \mathcal{K} , necessitating pre-processing steps to enhance efficiency.

Another category of samplers addresses the geometric structure of convex constraints, including Riemannian Hamiltonian Monte Carlo (Brubaker et al., 2012; Gürbüzbalaban et al., 2022; Kook et al., 2022), Gibbs sampling (Gelfand et al., 1992), and Mirror Langevin methods (Ahn & Chewi, 2021; Chewi et al., 2020; Hsieh et al., 2018; Zhang et al., 2020).

A particularly relevant approach to this work involves diffusion-type samplers (Brosse et al., 2017; Bubeck et al., 2015; 2018; Lamperski, 2021; Lehec, 2021; 2023). These samplers, based on discretizations of Itô diffusions, exhibit rapid convergence to the target density in continuous time and have been extensively employed in unconstrained settings. While the underlying stochastic processes can be generalized to constrained settings, the discretization analysis heavily relies on the smoothness of the target distribution, which is challenging to achieve in the constrained context considered here.

Among these works, Brosse et al. (2017) is the most closely aligned with the spirit of this work. It considers the same setting as the current study, proposing the use of the Moreau envelope (Example 2.1) to derive a smooth surrogate density, followed by applying LMC to sample from this density. Their approach supports our general framework, which integrates Euclidean projection with LMC, and provides convergence bounds in both the W_1 metric and total variation norm. Our work expands on this approach by presenting a general scheme that accommodates a broader range of projection methods and sampling techniques beyond LMC. This enhances the applicability of algorithms such as KLMC, pRLMC, and pRKLMC in constrained settings. Furthermore, we establish a lower bound for the distance between the surrogate distribution and the target distribution in Proposition 2.2, demonstrating the tightness of the bound for the distance between the approximation and the target density established in Proposition 2.1. This aspect has not been addressed in any of the previously mentioned works on the diffusion-type samplers.

B PROOFS OF SECTION 2

In this part, we provide the proof of Lemma 2.1, Lemma 2.2, Proposition 2.1, and Proposition 2.2.

B.1 PROOFS OF LEMMA 2.1 AND LEMMA 2.2

Proof of Lemma 2.1. By Theorem 2.26 in Rockafellar & Wets (2009), the function $\ell_{\mathcal{K}}^{E,\lambda}(x)$ is convex and continuously differentiable, and $\ell_{\mathcal{K}}^{E,\lambda}(x)$ is $\frac{1}{\lambda^2}$ -smooth. The last claim follows from Lemma 1 in Brosse et al. (2017). □

Proof of Lemma 2.2. By the definition of $\ell_{\mathcal{K}}^{G,\lambda}$, we have

$$\ell_{\mathcal{K}}^{G,\lambda} = \frac{1}{2\lambda^2} \left(1 - \frac{1}{g_{\mathcal{K}}(x)}\right)^2 \|x\|_2^2.$$

By the definition of $U^{G,\lambda}$, the function $U^{G,\lambda}$ is m -strongly convex, continuously differentiable. Moreover, it holds that

$$\|\nabla \ell_{\mathcal{K}}^{G,\lambda}(x) - \nabla \ell_{\mathcal{K}}^{G,\lambda}(y)\|_2 \leq \frac{1}{\lambda^2} \|x - y\|_2,$$

This implies $U^{G,\lambda}$ is $\frac{1}{\lambda^2} + M$ -smooth. Now we show that $\int_{\mathbb{R}^p} e^{-U^{G,\lambda}(s)} ds < \infty$. Let $R := \max_{x \in \mathcal{K}} \|x\|_2$, and there is a constant $l \in \mathbb{R}$ such that $f(x) \geq l, \forall x \in \mathbb{R}$. Define $R := \sup_{x \in \mathcal{K}} \|x\|_2$. Note that

$$\int_{\mathbb{R}^p} e^{-U^{G,\lambda}(x)} dx = \int_{\mathcal{K}} e^{-f(x)} dx + \int_{\mathcal{B}_2(0,R) \cap \mathcal{K}^c} e^{-U^{G,\lambda}(x)} dx + \int_{\mathcal{B}_2(0,R)^c} e^{-U^{G,\lambda}(x)} dx.$$

We now derive the bound for the second and the third terms on the right-hand side of the previous display. Since f is convex, there exists $a \in \mathbb{R}$ and $b \in \mathbb{R}^p$ such that

$$f(x) \geq a + \langle b, x \rangle.$$

We then find

$$\int_{\mathcal{B}_2(0,R) \cap \mathcal{K}^c} e^{-U^{G,\lambda}(x)} dx \leq e^{-a+bR} \int_{\mathcal{B}_2(0,R)} dx.$$

Thus,

$$\begin{aligned} \int_{\mathcal{B}_2(0,R) \cap \mathcal{K}^c} e^{-U^{G,\lambda}(x)} dx &\leq e^{-a+bR} \int_{\mathcal{B}_2(0,R)} dx \\ &\leq e^{-a+bR} \int_{\mathcal{B}_2(0,R)} dx \\ &= e^{-a+bR} \frac{\pi^{p/2}}{\Gamma(\frac{d}{2} + 1)} R^p \\ &< \infty. \end{aligned}$$

Note that

$$\int_{\mathcal{B}_2(0,R)^c} e^{-U^{G,\lambda}(x)} dx \leq e^{-l} \int_{\mathcal{B}_2(0,R)^c} e^{-\frac{1}{2\lambda^2}(\|x\|_2 - R)^2} dx$$

By Fubini's theorem and the fact that

$$e^{-\frac{1}{2\lambda^2}(\|x\|_2 - R)^2} = \int_0^\infty \frac{t}{\lambda^2} e^{-\frac{t^2}{2\lambda^2}} \mathbf{1}_{[\|x\|_2 - R, \infty)}(t) dt,$$

we obtain

$$\begin{aligned} \int_{\mathcal{B}_2(0,R)^c} e^{-U^{G,\lambda}(x)} dx &\leq e^{-l} \int_0^\infty \frac{t}{\lambda^2} e^{-\frac{t^2}{2\lambda^2}} \int_{x: \|x\|_2 \geq R} \mathbf{1}_{[\|x\|_2 - R, \infty)}(t) dx dt \\ &\leq e^{-l} \int_0^\infty \frac{t}{\lambda^2} e^{-\frac{t^2}{2\lambda^2}} \int_{\mathcal{B}_2(0,R+t)} dx dt \\ &= e^{-l} \int_0^\infty \frac{t}{\lambda^2} e^{-\frac{t^2}{2\lambda^2}} \frac{\pi^{p/2}}{\Gamma(\frac{p}{2} + 1)} (R+t)^d dt \\ &< \infty. \end{aligned}$$

Collecting pieces gives the desired result. □

B.2 PROOFS OF PROPOSITION 2.1

To prove the Proposition 2.1, we need the following auxiliary lemma.

Lemma B.1. *Assume f is convex and $\mathbf{0} = \arg \min_{\theta \in \mathcal{K}} f(\theta)$. Under Assumption 2.1, it holds for any $\lambda \in (0, \frac{r}{2(p+k)})$ that*

$$\int_{\mathcal{K}^c} \|x\|_2^k \nu^\lambda(x) dx \leq \frac{3\lambda(p+k)}{r} \times \mu_k(\nu),$$

where $\mu_k(\nu) = \int_{\mathbb{R}^p} \|\theta\|_2^k \nu(d\theta)$.

810 *Proof of Lemma B.1.* By Fubini's theorem and the fact that

$$811 \int_0^\infty \frac{t}{\lambda^2} e^{-\frac{t^2}{2\lambda^2}} \mathbf{1}_{\{\|x - \mathbb{P}_{\mathcal{K}}(x)\|_2, +\infty\}}(t) dt = e^{-\frac{1}{2\lambda^2} \|x - \mathbb{P}_{\mathcal{K}}(x)\|_2^2},$$

812 it holds for every $k \geq 0$ that

$$813 \int_{\mathcal{K}^c} \|x\|_2^k e^{-U^\lambda(x)} dx = \int_0^\infty \underbrace{\left(\int_{\mathcal{K}^c} \|x\|_2^k e^{-f(x)} \mathbf{1}_{\{\|x - \mathbb{P}_{\mathcal{K}}(x)\|_2 \leq t\}} dx \right)}_{\psi(t)} \frac{t}{\lambda^2} e^{-\frac{t^2}{2\lambda^2}} dt. \quad (\text{B.1})$$

814 Define the set $\mathcal{K}_t := \{x \in \mathbb{R}^p : \|x - \mathbb{P}_{\mathcal{K}}(x)\|_2 \leq t\}$, we then have

$$815 \psi(t) = \int_{\mathcal{K}_t} \|x\|_2^k e^{-f(x)} dx - \int_{\mathcal{K}} \|x\|_2^k e^{-f(x)} dx. \quad (\text{B.2})$$

816 On the other hand, it holds for every $b > 0$ that

$$817 \int_{\mathcal{K}_t} \|x\|_2^k e^{-f(x)} dx = (1+b)^{p+k} \int_{\mathcal{K}_t/(1+b)} \|y\|_2^k e^{-f((1+b)y)} dy$$

$$818 \leq (1+b)^{p+k} \int_{\mathcal{K}_t/(1+b)} \|y\|_2^k e^{-f(y)} dy$$

819 where we have used the inequality

$$820 f((1+b)y) + bf(0) \geq (1+b)f\left(\frac{1}{1+b}(1+b)y + \frac{b}{b+1}0\right) = (1+b)f(y) \geq f(y) + bf(0),$$

821 implying that $f((1+b)y) \geq f(y)$ for every $b > 0$ and every $y \in \mathbb{R}^p$. In addition, it holds for any $x \in \mathcal{K}_t/(1+b)$ that

$$822 \|(1+b)x - \mathbb{P}_{\mathcal{K}}((1+b)x)\|_2 \leq t$$

823 which is equivalent to

$$824 \frac{1+b}{b} \|x - \frac{1}{1+b} \mathbb{P}_{\mathcal{K}}((1+b)x)\|_2 \leq \frac{t}{b}.$$

825 Note that we can rewrite x as

$$826 x = \frac{1}{1+b} \mathbb{P}_{\mathcal{K}}((1+b)x) + \frac{b}{1+b} \frac{x - \frac{1}{1+b} \mathbb{P}_{\mathcal{K}}((1+b)x)}{\frac{b}{1+b}}.$$

827 Notice that

$$828 \mathbb{P}_{\mathcal{K}}((1+b)x) \in \mathcal{K}$$

829 and when $b = t/r$ we have

$$830 \frac{x - \frac{1}{1+b} \mathbb{P}_{\mathcal{K}}((1+b)x)}{\frac{b}{1+b}} \in \mathcal{K}.$$

831 By the convexity of the set \mathcal{K} , this implies $x \in \mathcal{K}$. Therefore, we obtain

$$832 \int_{\mathcal{K}_t} \|x\|_2^k e^{-f(x)} dx \leq \left(1 + \frac{t}{r}\right)^{p+k} \int_{\mathcal{K}} \|y\|_2^k e^{-f(y)} dy \leq e^{(p+k)t/r} \int_{\mathcal{K}} \|y\|_2^k e^{-f(y)} dy.$$

833 Combining this inequality with display B.2, we get

$$834 \psi(t) \leq (e^{(p+k)t/r} - 1) \int_{\mathcal{K}} \|y\|_2^k e^{-f(y)} dy.$$

Substituting this upper bound in display B.1 leads to

$$\begin{aligned}
\int_{\mathcal{K}^c} \|x\|_2^k e^{-U^\lambda(x)} dx &\leq \int_0^\infty \frac{t}{\lambda^2} e^{-t^2/2\lambda^2} (e^{(p+k)t/r} - 1) dt \times \int_{\mathcal{K}} \|y\|_2^k e^{-f(y)} dy \\
&= \int_0^\infty (e^{(p+k)t/r} - 1) d(-e^{-t^2/2\lambda^2}) \times \int_{\mathcal{K}} \|y\|_2^k e^{-f(y)} dy \\
&= \frac{(p+k)}{r} \int_0^\infty \exp\left\{-\frac{t^2}{2\lambda^2} + \frac{(p+k)t}{r}\right\} dt \times \int_{\mathcal{K}} \|y\|_2^k e^{-f(y)} dy \\
&= \frac{(p+k)}{r} \exp\left\{\frac{\lambda^2(p+k)^2}{2r^2}\right\} \int_0^\infty \exp\left\{-\frac{(t - (\lambda^2/r)(p+k))^2}{2\lambda^2}\right\} dt \times \int_{\mathcal{K}} \|y\|_2^k e^{-f(y)} dy \\
&\leq \frac{\sqrt{2\pi}\lambda(p+k)}{r} \exp\left\{\frac{\lambda^2(p+k)^2}{2r^2}\right\} \times \int_{\mathcal{K}} \|y\|_2^k e^{-f(y)} dy.
\end{aligned}$$

When $\lambda < \frac{r}{2(p+k)}$, it holds that

$$\int_{\mathcal{K}^c} \|x\|_2^k e^{-U^\lambda(x)} dx \leq \frac{3\lambda(p+k)}{r} \times \int_{\mathcal{K}} \|y\|_2^k e^{-f(y)} dy. \quad (\text{B.3})$$

Finally, note that the normalizing constant of ν^λ can be lower bounded as follows

$$\int_{\mathbb{R}^p} e^{-U^\lambda(x)} dx \geq \int_{\mathcal{K}} e^{-U^\lambda(x)} dx = \int_{\mathcal{K}} e^{-f(x)} dx.$$

The desired result follows readily by dividing the two sides of display B.3 by $\int_{\mathbb{R}^p} e^{-U^\lambda(x)} dx$ and employing this inequality. \square

We are now ready to prove Proposition 2.1

Proof of Proposition 2.1. By Theorem 6.15 in Villani (2009), it holds that

$$\mathbb{W}_q^q(\nu, \nu^\lambda) \leq \int_{\mathbb{R}^p} \|x\|_2^q |\nu(x) - \nu^\lambda(x)| dx.$$

We note that

$$\begin{aligned}
\int_{\mathbb{R}^p} \|x\|_2^q |\nu(x) - \nu^\lambda(x)| dx &= \int_{\mathcal{K}} \|x\|_2^q |\nu(x) - \nu^\lambda(x)| dx + \int_{\mathcal{K}^c} \|x\|_2^q \nu^\lambda(x) dx \\
&= \left(1 - \frac{\int_{\mathcal{K}} e^{-f(y)} dy}{\int_{\mathbb{R}^p} e^{-U^\lambda(y)} dy}\right) \int_{\mathcal{K}} \|x\|_2^q \nu(x) dx + \int_{\mathcal{K}^c} \|x\|_2^q \nu^\lambda(x) dx \\
&= \frac{\int_{\mathcal{K}^c} e^{-U^\lambda(y)} dy}{\int_{\mathbb{R}^p} e^{-U^\lambda(y)} dy} \times \int_{\mathcal{K}} \|x\|_2^q \nu(x) dx + \int_{\mathcal{K}^c} \|x\|_2^q \nu^\lambda(x) dx \\
&= \int_{\mathcal{K}^c} \nu^\lambda(y) dy \times \mu_q(\nu) + \int_{\mathcal{K}^c} \|x\|_2^q \nu^\lambda(x) dx.
\end{aligned}$$

When $\lambda \leq \frac{r}{2(p+q)}$, we apply Lemma B.1 with $k = 0$ and $k = q$, which leads us to the inequality

$$\int_{\mathbb{R}^p} \|x\|_2^q |\nu(x) - \nu^\lambda(x)| dx \leq \frac{3\lambda(p\mu_0(\nu) + (p+q)\mu_q(\nu))}{r} = \frac{3\lambda(2p+q)}{r} \times \mu_q(\nu).$$

This completes the proof. \square

918 B.3 PROOF OF PROPOSITION 2.2
919

920 To prove Proposition 2.2, we need the following auxiliary lemma.

921 **Lemma B.2.** *Under the assumptions stated in Proposition 2.2, for any $k \geq 1$ and any $\lambda \in$
922 $(0, \frac{r}{2(p+k)} \wedge \frac{1}{\sqrt{M}})$, it holds that*

923
924
$$\int_{\mathcal{K}^c} \|x\|_2^k \nu^\lambda(dx) / \mu^k(\nu) \geq \frac{1}{2} \sqrt{\frac{\pi}{2}} e^{-\frac{3}{4}Mr^2} \frac{p+k}{r+3\lambda p} \lambda.$$

925
926
927 *Proof of Lemma B.2.* By the definition of ν^λ and $\mu^k(\nu)$, we have

928
929
$$\int_{\mathcal{K}^c} \|x\|_2^k \nu^\lambda(dx) / \mu^k(\nu) = \frac{\int_{\mathcal{K}^c} \|x\|_2^k e^{-f(x) - \frac{1}{2\lambda^2}(\|x\|_2 - r)^2} dx}{\int_{\mathbb{R}^p} e^{-U^\lambda(x)} dx} \cdot \frac{\int_{\mathcal{K}} e^{-f(x)} dx}{\int_{\mathcal{K}} \|x\|_2^k e^{-f(x)} dx}. \quad (\text{B.4})$$

930
931 Note that f is M -smooth and $f(\mathbf{0}) = 0$, it then holds that

932
933
$$f(x) \leq \frac{M}{2} \|x\|_2^2.$$

934
935 This implies

936
937
$$\int_{\mathcal{K}^c} \|x\|_2^k e^{-f(x) - \frac{1}{2\lambda^2}(\|x\|_2 - r)^2} dx \geq \int_{\mathcal{K}^c} \|x\|_2^k e^{-\frac{M}{2}\|x\|_2^2 - \frac{1}{2\lambda^2}(\|x\|_2 - r)^2} dx.$$

938
939 Employing the integration in polar coordinates gives

940
941
$$\int_{\mathcal{K}^c} \|x\|_2^k e^{-\frac{M}{2}\|x\|_2^2 - \frac{1}{2\lambda^2}(\|x\|_2 - r)^2} dx = \omega_{p-1} \int_r^{+\infty} t^{k+p-1} e^{-\frac{M}{2}t^2 - \frac{1}{2\lambda^2}(t-r)^2} dt,$$

942
943 where $\omega_{p-1} = \int_{-\pi}^{\pi} \int_0^{\pi} \cdots \int_0^{\pi} \prod_{j=1}^{p-2} (\sin \theta_j)^{p-j-1} d\theta_1 \cdots d\theta_{p-1}$. Let $t = r + \lambda z$, it then follows that

944
945
$$\begin{aligned} & \int_{\mathcal{K}^c} \|x\|_2^k e^{-f(x) - \frac{1}{2\lambda^2}(\|x\|_2 - r)^2} dx \\ & \geq \lambda \omega_{p-1} \int_0^{+\infty} (r + \lambda z)^{k+p-1} e^{-\frac{M}{2}(r+\lambda z)^2 - \frac{1}{2}z^2} dz \\ & \geq \lambda \omega_{p-1} r^{k+p-1} \int_0^{+\infty} e^{-\frac{M}{2}(r+\lambda z)^2 - \frac{1}{2}z^2} dz \\ & = \lambda \omega_{p-1} r^{k+p-1} \int_0^{+\infty} e^{-\left(\sqrt{\frac{1+M\lambda^2}{2}}x + \frac{M\lambda r}{\sqrt{2(1+M\lambda^2)}}\right)^2} dx e^{-\frac{Mr^2}{2}\left(1 + \frac{M\lambda^2}{1+M\lambda^2}\right)}. \end{aligned}$$

946
947 The last step follows from a change of variables. Using the fact that $\lambda < 1/\sqrt{M}$ and $(x+y)^2 \leq$
948 $2x^2 + 2y^2, \forall x, y \geq 0$, we then obtain

949
950
$$\begin{aligned} & \int_{\mathcal{K}^c} \|x\|_2^k e^{-f(x) - \frac{1}{2\lambda^2}(\|x\|_2 - r)^2} dx \geq \lambda \omega_{p-1} r^{k+p-1} \int_0^{+\infty} e^{-(1+M\lambda^2)x^2} dx e^{-\frac{Mr^2}{2}\left(1 + \frac{M\lambda^2}{1+M\lambda^2}\right)} \\ & = \lambda \omega_{p-1} r^{k+p-1} \frac{1}{2} \sqrt{\frac{\pi}{1+M\lambda^2}} e^{-\frac{Mr^2}{2}\left(1 + \frac{M\lambda^2}{1+M\lambda^2}\right)} \\ & \geq \lambda \omega_{p-1} r^{k+p-1} \frac{1}{2} \sqrt{\frac{\pi}{2}} e^{-\frac{3}{4}Mr^2}. \end{aligned}$$

951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967 Moreover, by display B.3, we find

968
969
$$\begin{aligned} \int_{\mathbb{R}^p} e^{-U^\lambda(x)} dx &= \int_{\mathcal{K}} e^{-U^\lambda(x)} dx + \int_{\mathcal{K}^c} e^{-U^\lambda(x)} dx \\ &\leq \left(1 + \frac{3\lambda p}{r}\right) \int_{\mathcal{K}} e^{-\|x\|_2^2/2} dx. \end{aligned}$$

Furthermore, using the integration in polar coordinates again gives

$$\begin{aligned} \int_{\mathcal{K}} \|x\|_2^k e^{-f(x)} dx &\leq \omega_{p-1} \int_0^r t^{k+p-1} dt \\ &= \omega_{p-1} \frac{r^{k+p}}{k+p}. \end{aligned}$$

Collecting pieces and plugging them into display B.4 gives

$$\int_{\mathcal{K}^c} \|x\|_2^k \nu^\lambda(dx) / \mu^k(\nu) \geq \frac{1}{2} \sqrt{\frac{\pi}{2}} e^{-\frac{3}{4}Mr^2} \frac{p+k}{r+3\lambda p} \lambda$$

as desired. □

We are now ready to prove Proposition 2.2.

Proof of Proposition 2.2. By Proposition 7.29 in Villani (2009), it holds for any $q \geq 1$ that

$$W_q^q(\nu^\lambda, \nu) \geq \left| \left(\int_{\mathbb{R}^p} \|x\|_2^q \nu^\lambda(dx) \right)^{1/q} - \left(\int_{\mathbb{R}^p} \|x\|_2^q \nu(dx) \right)^{1/q} \right|^q.$$

Dividing both sides by $\mu_q(\nu)$ gives

$$\begin{aligned} W_q^q(\nu^\lambda, \nu) / \mu_q(\nu) &\geq \left| \left(\int_{\mathbb{R}^p} \|x\|_2^q \nu^\lambda(dx) / \mu_q(\nu) \right)^{1/q} - 1 \right|^q \\ &= \left| \left(\frac{\int_{\mathcal{K}} \|x\|_2^q \nu^\lambda(dx) + \int_{\mathcal{K}^c} \|x\|_2^q \nu^\lambda(dx)}{\int_{\mathbb{R}^p} \|x\|_2^q \nu(dx)} \right)^{1/q} - 1 \right|^q. \end{aligned}$$

By the definitions of ν and ν^λ , we have

$$\begin{aligned} \frac{\int_{\mathcal{K}} \|x\|_2^q \nu^\lambda(dx)}{\int_{\mathbb{R}^p} \|x\|_2^q \nu(dx)} &= \frac{\int_{\mathcal{K}} \|x\|_2^q e^{-f(x)} dx}{\int_{\mathbb{R}^p} e^{-U^\lambda} dx} \cdot \frac{\int_{\mathcal{K}} e^{-f(x)} dx}{\int_{\mathcal{K}} \|x\|_2^q e^{-f(x)} dx} \\ &= \frac{\int_{\mathcal{K}} e^{-f(x)} dx}{\int_{\mathcal{K}} e^{-f(x)} dx + \int_{\mathcal{K}^c} e^{-f(x)} dx}. \end{aligned}$$

Combining this with display B.3 then yields

$$\frac{\int_{\mathcal{K}} \|x\|_2^q \nu^\lambda(dx)}{\int_{\mathbb{R}^p} \|x\|_2^q \nu(dx)} \geq \frac{r}{r+3\lambda p}.$$

By Lemma B.2 and Lemma B.1, we have

$$\frac{1}{2} \sqrt{\frac{\pi}{2}} e^{-\frac{3}{4}Mr^2} \frac{p+q}{r+3\lambda p} \lambda \leq \frac{\int_{\mathcal{K}^c} \|x\|_2^q \nu^\lambda(dx)}{\int_{\mathbb{R}^p} \|x\|_2^q \nu(dx)} \leq \frac{3\lambda(p+q)}{r}.$$

Collecting pieces gives the desired result. □

C PROOFS OF SECTION 3

Proof of Theorem 3.1. The triangle inequality for the Wasserstein distance provides us with

$$W_2(\nu_n^{\text{LMC}}, \nu) \leq W_2(\nu_n^{\text{LMC}}, \nu^\lambda) + W_2(\nu^\lambda, \nu). \quad (\text{C.1})$$

By Theorem 9 in Durmus et al. (2019), it holds that

$$W_2(\nu_n^{\text{LMC}}, \nu^\lambda) \leq e^{-\frac{mnh}{2}} W_2(\nu_0^{\text{LMC}}, \nu^\lambda) + \sqrt{\frac{2M^\lambda p h}{m}} \quad (\text{C.2})$$

provided that $h \leq 1/M^\lambda$. Using the triangle inequality for the Wasserstein distance again gives

$$W_2(\nu_n^{\text{LMC}}, \nu^\lambda) \leq e^{-\frac{mnh}{2}} W_2(\nu_0^{\text{LMC}}, \nu) + \sqrt{\frac{2M^\lambda ph}{m}} + W_2(\nu, \nu^\lambda).$$

Plugging this back to display C.1 then provides us with

$$W_2(\nu_n^{\text{LMC}}, \nu) \leq e^{-\frac{mnh}{2}} W_2(\nu_0^{\text{LMC}}, \nu) + \sqrt{\frac{2M^\lambda ph}{m}} + 2W_2(\nu, \nu^\lambda).$$

Combining this with Proposition 2.1, for any $\lambda \in (0, \frac{r}{2(p+2)})$, we obtain

$$W_2(\nu_n^{\text{LMC}}, \nu) \leq e^{-\frac{mnh}{2}} W_2(\nu_0^{\text{LMC}}, \nu) + \sqrt{\frac{2M^\lambda ph}{m}} + 2\left(\frac{3\mu_2(\nu)\lambda(2p+2)}{r}\right)^{\frac{1}{2}}.$$

This completes the proof of the first claim.

We then proceed to prove the second claim. By the triangle inequality and the monotonicity of Wasserstein distance, we have

$$\begin{aligned} W_1(\nu_n^{\text{LMC}}, \nu) &\leq W_1(\nu_n^{\text{LMC}}, \nu^\lambda) + W_1(\nu^\lambda, \nu) \\ &\leq W_2(\nu_n^{\text{LMC}}, \nu^\lambda) + W_1(\nu^\lambda, \nu). \end{aligned}$$

When ν_0^{LMC} is the Dirac mass at the minimizer of the function f , by Proposition 1 in Durmus & Moulines (2019), it holds that

$$W_2(\nu_0^{\text{LMC}}, \nu^\lambda) \leq \sqrt{\frac{p}{m}}.$$

Combining this with previous display, display C.2, and Proposition 2.1 then gives

$$\begin{aligned} W_1(\nu_n^{\text{LMC}}, \nu) &\leq e^{-\frac{mnh}{2}} W_2(\nu_0^{\text{LMC}}, \nu^\lambda) + \sqrt{\frac{2M^\lambda ph}{m}} + W_1(\nu^\lambda, \nu) \\ &\leq e^{-\frac{mnh}{2}} \sqrt{\frac{p}{m}} + \sqrt{\frac{2M^\lambda ph}{m}} + \frac{3\mu_1(\nu)\lambda(2p+1)}{r} \end{aligned}$$

as desired. □

Proof of Theorem 3.2. The proof employs a similar technique to that used in the proof of Theorem 3.1. Note that $M^\lambda h \leq 1/10$, by the triangle inequality and Theorem 1 from Yu & Dalalyan (2024), we have

$$\begin{aligned} W_2(\nu_n^{\text{pRLMC}}, \nu) &\leq W_2(\nu^\lambda, \nu) + W_2(\nu_n^{\text{pRLMC}}, \nu^\lambda) \\ &\leq \left(1 + \sqrt{\kappa M^\lambda h} (0.82(M^\lambda h)^{Q-1} + 0.94M^\lambda h/R)\right) e^{-mnh/2} W_2(\nu_0^{\text{pRLMC}}, \nu^\lambda) \\ &\quad + \sqrt{\kappa M^\lambda h} (3.98(M^\lambda h)^{Q-1} + 6.91M^\lambda h/\sqrt{R}) \sqrt{p/m} + W_2(\nu^\lambda, \nu) \\ &\leq \left(1 + \sqrt{\kappa M^\lambda h} (0.82(M^\lambda h)^{Q-1} + 0.94M^\lambda h/R)\right) e^{-mnh/2} (W_2(\nu_0^{\text{pRLMC}}, \nu) + W_2(\nu, \nu^\lambda)) \\ &\quad + \sqrt{\kappa M^\lambda h} (3.98(M^\lambda h)^{Q-1} + 6.91M^\lambda h/\sqrt{R}) \sqrt{p/m} + W_2(\nu^\lambda, \nu). \end{aligned}$$

Combining this with Proposition 2.1, for any $\lambda \in (0, \frac{r}{2(p+2)})$, we obtain

$$\begin{aligned} W_2(\nu_n^{\text{pRLMC}}, \nu) &\leq \left(1 + \sqrt{\kappa M^\lambda h} (0.82(M^\lambda h)^{Q-1} + 0.94M^\lambda h/R)\right) e^{-mnh/2} W_2(\nu_0^{\text{pRLMC}}, \nu) \\ &\quad + \sqrt{\kappa M^\lambda h} (3.98(M^\lambda h)^{Q-1} + 6.91M^\lambda h/\sqrt{R}) \sqrt{p/m} \\ &\quad \left(2 + \sqrt{\kappa M^\lambda h} (0.82(M^\lambda h)^{Q-1} + 0.94M^\lambda h/R)\right) \sqrt{3\mu_2(\nu)\lambda(2p+2)/r} \end{aligned}$$

as desired. Invoking Proposition 2.1 and the monotonicity of the Wasserstein distance, since ν_0^{pRLMC} is the Dirac mass at the minimizer of the function f , it holds that

$$\begin{aligned} W_1(\nu_n^{\text{pRLMC}}, \nu) &\leq W_2(\nu_n^{\text{pRLMC}}, \nu^\lambda) + W_1(\nu^\lambda, \nu) \\ &\leq \left(1 + \sqrt{\kappa M^\lambda h} (0.82(M^\lambda h)^{Q-1} + 0.94M^\lambda h/R)\right) e^{-mnh/2} \sqrt{p/m} \\ &\quad + \sqrt{\kappa M^\lambda h} (3.98(M^\lambda h)^{Q-1} + 6.91M^\lambda h/\sqrt{R}) \sqrt{p/m} \\ &\quad \left(2 + \sqrt{\kappa M^\lambda h} (0.82(M^\lambda h)^{Q-1} + 0.94M^\lambda h/R)\right) \sqrt{3\mu_2(\nu)\lambda(2p+2)/r} \end{aligned}$$

as desired. \square

Proof of Theorem 3.3. The proof relies a similar technique to that used in the proof of Theorem 3.1. The key distinction lies in the utilization of Theorem 3 from Yu et al. (2023) to bound the W_2 distance between ν_n^{KLMC} and ν^λ . By the triangle inequality and Theorem 3 from Yu et al. (2023), we have

$$\begin{aligned} W_2(\nu_n^{\text{KLMC}}, \nu) &\leq W_2(\nu^\lambda, \nu) + W_2(\nu_n^{\text{KLMC}}, \nu^\lambda) \\ &\leq 2\varrho^n W_2(\nu_0^{\text{KLMC}}, \nu^\lambda) + 0.05 \sqrt{\frac{\varrho^n \mathbb{E}[U^\lambda(\vartheta_0^{\text{KLMC}}) - f(\mathbf{0})]}{m}} + 0.9\gamma h \sqrt{\frac{\kappa p}{m}} + W_2(\nu^\lambda, \nu) \\ &\leq 2\varrho^n (W_2(\nu_0^{\text{KLMC}}, \nu) + W_2(\nu, \nu^\lambda)) + 0.05 \sqrt{\frac{\varrho^n \mathbb{E}[U^\lambda(\vartheta_0^{\text{KLMC}}) - f(\mathbf{0})]}{m}} + 0.9\gamma h \sqrt{\frac{\kappa p}{m}} \\ &\quad + W_2(\nu^\lambda, \nu). \end{aligned}$$

Combining this with Proposition 2.1, for any $\lambda \in (0, \frac{r}{2(p+2)})$, we obtain

$$\begin{aligned} W_2(\nu_n^{\text{KLMC}}, \nu) &\leq 2\varrho^n W_2(\nu_0^{\text{KLMC}}, \nu) + 3\sqrt{\frac{3\mu_2(\nu)\lambda(2p+2)}{r}} + 0.05 \sqrt{\frac{\varrho^n \mathbb{E}[U^\lambda(\vartheta_0^{\text{KLMC}}) - f(\mathbf{0})]}{m}} \\ &\quad + 0.9\gamma h \sqrt{\frac{\kappa p}{m}} \end{aligned}$$

as desired.

Invoking Proposition 2.1 and the monotonicity of the Wasserstein distance, since $\vartheta_0^{\text{KLMC}} = \mathbf{0}$, it holds that

$$\begin{aligned} W_1(\nu_n^{\text{KLMC}}, \nu) &\leq W_2(\nu_n^{\text{KLMC}}, \nu^\lambda) + W_1(\nu^\lambda, \nu) \\ &\leq 2\varrho^n \sqrt{\frac{p}{m}} + \frac{3\mu_1(\nu)\lambda(2p+1)}{r} + 0.9\gamma h \sqrt{\frac{\kappa p}{m}} \end{aligned}$$

as desired. \square

Proof of Theorem 3.4. By the triangle inequality and Theorem 2 from Yu & Dalalyan (2024), we have

$$\begin{aligned} W_2(\nu_n^{\text{pRKLMLC}}, \nu) &\leq W_2(\nu^\lambda, \nu) + W_2(\nu_n^{\text{pRKLMLC}}, \nu^\lambda) \\ &\leq 1.8\varrho^n W_2(\nu_0^{\text{pRKLMLC}}, \nu^\lambda) + 0.28 \sqrt{\frac{\varrho^n \mathbb{E}[U^\lambda(\vartheta_0^{\text{pRKLMLC}}) - f(\mathbf{0})]}{m}} \\ &\quad + 44.78 \sqrt{\frac{(\gamma M^\lambda h^2)^3}{R^2} + (\gamma M^\lambda h^2)^{2Q-1}} \sqrt{\frac{\kappa p}{m}} + W_2(\nu^\lambda, \nu) \\ &\leq 1.8\varrho^n (W_2(\nu_0^{\text{pRKLMLC}}, \nu) + W_2(\nu, \nu^\lambda)) + 0.28 \sqrt{\frac{\varrho^n \mathbb{E}[U^\lambda(\vartheta_0^{\text{pRKLMLC}}) - f(\mathbf{0})]}{m}} \\ &\quad + 44.78 \sqrt{\frac{(\gamma M^\lambda h^2)^3}{R^2} + (\gamma M^\lambda h^2)^{2Q-1}} \sqrt{\frac{\kappa p}{m}} + W_2(\nu^\lambda, \nu). \end{aligned}$$

1134 Combining this with Proposition 2.1, for any $\lambda \in (0, \frac{r}{2(p+2)})$, we obtain

$$1135$$

$$1136$$

$$1137 \mathbb{W}_2(\nu_n^{\text{pRKLMC}}, \nu) \leq 1.8\varrho^n \mathbb{W}_2(\nu_0^{\text{pRKLMC}}, \nu) + 0.28 \sqrt{\frac{\varrho^n \mathbb{E}[U^\lambda(\vartheta_0^{\text{pRKLMC}}) - f(\mathbf{0})]}{m}}$$

$$1138$$

$$1139 + 44.78 \sqrt{\frac{(\gamma M^\lambda h^2)^3}{R^2} + (\gamma M^\lambda h^2)^{2Q-1} \sqrt{\frac{\kappa p}{m}}} + 2.8 \sqrt{\frac{3\mu_2(\nu)\lambda(2p+2)}{r}}$$

$$1140$$

$$1141$$

1142 as desired.

1143 Invoking Proposition 2.1 and the monotonicity of the Wasserstein distance, since $\vartheta_0^{\text{pRKLMC}} = \mathbf{0}$, it

1144 holds that

$$1145 \mathbb{W}_1(\nu_n^{\text{pRKLMC}}, \nu) \leq \mathbb{W}_2(\nu_n^{\text{pRKLMC}}, \nu^\lambda) + \mathbb{W}_1(\nu^\lambda, \nu)$$

$$1146$$

$$1147 \leq 1.8\varrho^n \sqrt{\frac{p}{m}} + \frac{3\mu_1(\nu)\lambda(2p+1)}{r} + 44.78 \sqrt{\frac{(\gamma M^\lambda h^2)^3}{R^2} + (\gamma M^\lambda h^2)^{2Q-1} \sqrt{\frac{\kappa p}{m}}}$$

$$1148$$

$$1149$$

1150 as desired.

1151 \square

1152

1153

1154

1155

1156

1157

1158

1159

1160

1161

1162

1163

1164

1165

1166

1167

1168

1169

1170

1171

1172

1173

1174

1175

1176

1177

1178

1179

1180

1181

1182

1183

1184

1185

1186

1187