Proceedings of Machine Learning Research vol 291:1–62, 2025 38th Annual Conference on Learning Theory

# Regularized Dikin Walks for Sampling Truncated Logconcave Measures, Mixed Isoperimetry and Beyond Worst-Case Analysis

**Minhui Jiang** MIJIANG@ETHZ.CH
**Yuansi Chen** YUANSI.CHEN@STAT.MATH.ETHZ.CH
*Seminar for Statistics, ETH Zürich, Switzerland*

**Editors:** Nika Haghtalab and Ankur Moitra

## Abstract

We study sampling from logconcave distributions truncated on polytopes, motivated by Bayesian models with indicator variables. Built on interior point methods and the Dikin walk, we analyze the mixing time of regularized Dikin walks. Our contributions include: (1) proving that the soft-threshold Dikin walk mixes in $\widetilde{O}\left(mn + \kappa n\right)$ iterations for logconcave distributions with condition number $\kappa$, dimension $n$ and $m$ linear constraints, without requiring bounded polytopes. Moreover, we introduce the regularized Dikin walk using Lewis weights and show it mixes in $\widetilde{O}(n^{2.5} + \kappa n)$; (2) extending the above mixing time guarantees to weakly log-concave truncated distributions with finite covariance matrices; and (3) going beyond worst-case mixing time analysis, we show that soft-threshold Dikin walk mixes significantly faster when $O(1)$ number of constraints intersect the high-probability mass of the distribution, improving the $\widetilde{O}(mn + \kappa n)$ upper bound to $\widetilde{O}(m + \kappa n)$. Additionally, we provide practical implementation to generate a warm initialization.

**Keywords:** Dikin walk, interior-point methods, isoperimetric inequality, warm-start generation

## 1. Introduction

Sampling from high-dimensional distributions under constraints is an important challenge across Bayesian statistics, computer science, and systems biology. In Bayesian statistics, posterior distributions are often truncated on a subset of $\mathbb{R}^n$, rendering standard samplers for smooth distributions such as Langevin algorithms or Hamiltonian Monte Carlo (HMC) ineffective. Examples include censored data models, and ordered linear models (Gelfand et al., 1992). In addition, Bayesian probit regression models (Albert and Chib, 1993), isotonic regression (Neelon and Dunson, 2004), and tobit models (Anceschi et al., 2023) all require sampling from truncated normal distributions as a subroutine. In computer science, efficient volume computation of convex bodies relies heavily on sampling from distributions truncated on these bodies (Lovász and Simonovits, 1993; Lovász and Vempala, 2006; Lee and Vempala, 2017; Cousins and Vempala, 2018). In systems biology, constraint-based modeling of human metabolic networks uses polytopes to define solution spaces, making high-dimensional polytope-constrained sampling essential (Wiback et al., 2004; Lewis et al., 2012; Haraldsdóttir et al., 2017; Heirendt et al., 2019).

A particular motivation is the sampling of SUN (unified skew normal) distributions (Anceschi et al., 2023). It serves as conjugate priors/posteriors for various generalized linear models, including probit and tobit models. Anceschi et al. (2023) showed that sampling from SUN distributions can be reduced to sampling from truncated Gaussians. This way of posterior sampling is called “i.i.d. sampling”.

Experiments show that sampling from posterior distributions through the “i.i.d. sampling” significantly outperforms directly sampling via HMC (Anceschi et al., 2023). While the proposed algorithm works well for small sample size $N$ and large feature dimension $p$, it becomes computationally prohibitive for datasets with large sample size $N$. In these scenarios, deterministic approximations of the posterior were considered, such as variational Bayes (VB) (Blei et al., 2017) and expectation-propagation (Minka, 2013). However, these methods can struggle with approximation quality. For example, VB in Fasano and Durante (2022) only performs well when $p > N$. This leaves a gap for determining efficient sampling algorithms in the moderate or large $N$ regime.

The challenge in "i.i.d. sampling" for large $N$ stems from the fact that the truncated Gaussian vector has both its dimension and the number of linear constraints scales linearly in $N$. Specifically, the truncated Gaussian is given by $\pi(x) \propto \mathbb{I}_{\{x \geq 0\}} \exp\left[-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right]$, where $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ is a positive definite (PD) matrix. This leads us to study efficient sampling algorithms for Gaussian and general logconcave distributions truncated on polytopes.

**Assumptions on Target Distributions** For notations, let $K \subseteq \mathbb{R}^n$ denote an open and convex set, $\Pi$ denote our target distribution with density function $\pi$ to be

$$\pi(x) \propto \mathbf{1}_K(x) e^{-f(x)}. \tag{1}$$

In this paper, we focus on when $K$ is a polytope, and we assume $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. Moreover, $f$ is said to be *$\alpha$-convex* if $\nabla^2 f(x) \succeq \alpha I$ for all $x \in \mathbb{R}^n$, and $f$ is *$\beta$-smooth* if $\nabla^2 f(x) \preceq \beta I$ for all $x \in \mathbb{R}^n$. We give the following two assumptions on the target distribution $\Pi$.

**Assumption 1.1 (strong logconcave target)** *$K$ is a polytope given by $\{x | Ax > b\}$ where $A \in \mathbb{R}^{m \times n}$ with nonzero rows $a_i \in \mathbb{R}^n$ for $i = 1, \ldots, m$ and $b \in \mathbb{R}^m$. There exists fixed constants $\beta \geq \alpha > 0$, such that $f$ is $\alpha$-convex and $\beta$-smooth, and we define $\kappa := \beta/\alpha$ to be the condition number of* $\Pi$.

**Assumption 1.2 (weakly logconcave target)** *$K$ is a polytope given by $\{x | Ax > b\}$ where $A \in \mathbb{R}^{m \times n}$ with nonzero rows $a_i \in \mathbb{R}^n$ for $i = 1, \ldots, m$ and $b \in \mathbb{R}^m$. For all $x \in K$, $0 \preceq \nabla^2 f(x) \preceq \beta I$, and the distribution $\Pi$ has a bounded covariance matrix $\Sigma_\pi$ with its largest eigenvalue to be $\eta$.*

In the case of Gaussian sampling truncated on the polytope, we can assume the condition number $\kappa = 1$ without loss of generality after suitable affine transformations, see Appendix A.2.

We say a function $q : \mathbb{R}^n \to [0, \infty)$ is *logconcave* if $\log q(x)$ is a concave function over $\mathbb{R}^n$. We say a probability distribution $\Pi$ on $\mathbb{R}^n$ is *logconcave* if it admits a density function $\pi$ w.r.t. the Lebesgue measure and $\pi$ is a logconcave function.

A probability distribution $\Pi$ on $\mathbb{R}^n$ is said to be *more logconcave than Gaussian with covariance* $\frac{1}{\alpha} I_n$ if its density $\pi$ satisfies

$$\pi(x) \propto \exp\left(-\frac{\alpha}{2} \|x\|_2^2\right) \cdot q(x), \tag{2}$$

for $q$ to be a logconcave function over $\mathbb{R}^n$. As a special case, the target distribution in Eq.(1) with $\alpha$-convex $f$ is more logconcave than Gaussian with covariance $\frac{1}{\alpha} I_n$.

**Markov chain basics** Assume $\mathcal{X}$ is a Borel measurable subset of $\mathbb{R}^n$, a Markov chain on $\mathcal{X}$ is characterized by a *transition kernel* $\mathcal{T} : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to \mathbb{R}^+$, see Appendix A for detailed definition.

In this paper, we write $\mathcal{T}_x(B) := \mathcal{T}(x, B)$ and use $\mathcal{T}_x$ to denote the probability measure at $x$. A transition kernel $\mathcal{T}$ can be seen as an operator on probability measures. For a probability measure $\mu_0$ over $\mathcal{X}$, $\mathcal{T}(\mu_0)$ denotes the probability measure after one step of Markov chain is defined by:

$$\mathcal{T}(\mu_0)(B) := \int_{x \in \mathcal{X}} \mathcal{T}_x(B) \mu_0(dx),$$

for any $B \in \mathcal{B}(\mathcal{X})$. Applying $\mathcal{T}$ recursively on $\mu_0$ gives us the measure after applying $k$ steps of Markov chain $\mathcal{T}^k(\mu_0) := \mathcal{T}\left(\mathcal{T}^{k-1}(\mu_0)\right)$, denoted as $\mu_k$ for short. $\Pi$ is called a *stationary distribution* of $\mathcal{T}$ if $\mathcal{T}(\Pi) = \Pi$.

To quantify how fast $\mathcal{T}^k(\mu_0)$ converges to the target distribution $\Pi$, we use the *total variation distance*, $\|\mathbb{P} - \mathbb{Q}\|_{TV} := \sup_B |\mathbb{P}(B) - \mathbb{Q}(B)|$, where $B$ is taken over Borel sets in $\mathbb{R}^n$.

For an error tolerance $\epsilon > 0$, given a Markov chain with transition kernel $\mathcal{T}$ and stationary distribution $\Pi$, we define its *mixing time* as the number of iterations it takes to be $\epsilon$-close to its stationary distribution. More precisely,

$$T_{\text{mix}}(\epsilon; \mu_0) := \inf \left\{ k \in \mathbb{Z}^+ \middle| \left\| \mathcal{T}^k(\mu_0) - \Pi \right\|_{TV} \le \epsilon \right\}.$$

We wish to prove an upper bound of $T_{\text{mix}}(\epsilon; \mu_0)$, as a function of inputs to our algorithm, such as the dimension of the distribution $n$, the number of constraints of the polytope $m$, the initial distribution $\mu_0$ and the error tolerance $\epsilon$.

We say an initial distribution $\mu_0$ is $M$*-warm* with respect to the target distribution $\Pi$ if for any Borel measurable $B \subseteq \mathbb{R}^n$, we have $\mu_0(B) \le M\Pi(B)$.

**Other notations** Given a vector $v \in \mathbb{R}^k$, we use $\mathrm{Diag}(v)$ to denote the diagonal matrix in $\mathbb{R}^{k\times k}$ that has $v$ as its diagonal elements. Similarly, given a matrix $P \in \mathbb{R}^{k\times k}$, we use $\mathrm{diag}(P)$ to denote the vector in $\mathbb{R}^k$ that contains diagonal elements of $P$.

For $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$ positive definite (PD) matrix, we use the notation $\mathcal{N}(\mu, \Sigma)$ to denote the multivariate Gaussian distribution with mean $\mu$ and covariance $\Sigma$. We also use $\mathcal{N}(z; \mu, \Sigma)$ to denote the probability density of $\mathcal{N}(\mu, \Sigma)$ computed at $z$.

**Regularized Dikin Walks** We define a local metric as a matrix function $H : K \to \mathbb{S}^n_+$, which maps any $x \in K$ to a positive semi-definite (PSD) matrix $H(x)$. For convenience, we use $G$ to denote the regularized version of $H$ for a fixed $\lambda > 0$, $G(x) := H(x) + \lambda I$.

**input :** local metric $H$, regularization size $\lambda > 0$, step-size $r$,
initial distribution $\mu_0$, number of iterations $T$
**output:** $x_T$
draw $x_0 \sim \mu_0$ **for** $t \leftarrow 1$ **to** $T$ **do**
    $x_t \leftarrow x_{t-1}$ draw $U \sim \mathrm{Unif}(0, 1)$ `//lazification step`
    **if** $U < \frac{1}{2}$ **then**
        draw $z \sim \mathcal{N}(x_t, \frac{r^2}{n}(H(x_t) + \lambda I)^{-1})$
        $x_t \leftarrow z$ with probability $\min\left\{1, \mathbf{1}_K(z)\frac{\exp(-f(z))}{\exp(-f(x_t))} \cdot \frac{\mathcal{N}\left(x_t; z, \frac{r^2}{n}(H(z)+\lambda I)^{-1}\right)}{\mathcal{N}\left(z; x_t, \frac{r^2}{n}(H(x_t)+\lambda I)^{-1}\right)}\right\}$
    **end**
**end**

**Algorithm 1:** regularized Dikin walk

For polytope $K = \{x | Ax > b\}$, we define the slackness at $x \in K$ to be $s_x := Ax - b$, $S_x := \mathrm{Diag}(s_x)$ and $A_x := S_x^{-1}A$. Now we introduce two tailored local metrics for regularized Dikin walk on polytopes. The first is the soft-threshold Dikin walk proposed in Mangoubi and Vishnoi (2023), and the second is the regularized Lewis metric designed by us, which is motivated by the properties of Lewis weights in Lee and Sidford (2019); Laddha et al. (2020).

**Definition 1.3 (soft-threshold (Mangoubi and Vishnoi, 2023))** *The soft-threshold metric $G$ is defined as for $x \in K$,*

$$G(x) := H(x) + \lambda I_n, \text{ with } H(x) := A_x^\top A_x.$$

**Definition 1.4 (regularized Lewis metric)** *The regularized Lewis metric $G(x)$ is defined as for $x \in K$,*

$$G(x) := H(x) + \lambda I_n, \text{ with } H(x) := c_1 \sqrt{n} (\log m)^{c_2} A_x^\top W_x A_x,$$

*where $c_1, c_2$ are some absolute constant, $W_x := \mathrm{Diag}(w_x)$ is the Lewis weights of the matrix $A_x$, which is defined by the following optimization problem:*

$$w_x := \arg\max_{w \in \mathbb{R}_+^m} \left[ \log\det\left( A_x^\top W^{c_q} A_x \right) - c_q \sum_{i=1}^m w_i \right], \tag{3}$$

*where $c_q := 1 - \frac{2}{q}$, and we choose $q = \Theta(\log m)$.*

**Main Contributions** Our contributions can be summarized in threefold: First, we improve the analysis on the soft-threshold Dikin walk (Mangoubi and Vishnoi, 2023) and obtain a mixing time of $\widetilde{O}((m + \kappa)n)$ given warmness. For truncated Gaussian sampling specifically, we reduce the upper bound from $\widetilde{O}(n^2)$ in Kook and Vempala (2024) to $\widetilde{O}(n)$ when $m = o(n)$, matching the unconstrained sampling using Random Walk Metropolis. We also explore regularized Dikin walk wtith Lewis metric and proved a mixing time of $\widetilde{O}(n^{2.5} + \kappa n)$. Second, we extend the above mixing time analysis to weakly logconcave distributions with finite covariances. Finally, we demonstrate that soft-threshold Dikin walk achieves a mixing time dependent on the constraints that a high-probability ball intersects instead of the total number of constraints $m$. The core innovation in this paper is a new isoperimetric inequality for truncated logconcave distributions under a mixture of Euclidean and Hilbert metrics.

**Roadmap** The rest of the paper is structured as follows. Section 2 reviews the literature on Markov chains for sampling distributions truncated on a convex set. In Section 3, we summarize our main results stated as theorems, as well as practical implementation. In Section 4, we give a sketch of proof for our results. Section 5 discusses potential directions to extend our work.

## 2. Related Works

Sampling from distributions truncated on convex sets has been widely studied. The simplest scenario is uniform sampling on a convex body, and significant progress has been made. Using the terminology from Kook and Vempala (2024), the so-called "general-purpose" samplers assume having access to the convex body only through a membership oracle which checks whether $x \in K$. Thus the mixing times are measured by how many times the oracle is called. Two famous general-purpose samplers are ball walk and hit-and-run.

For uniform sampling over an isotropic[1] convex body $K$, the mixing times for both ball walk and hit-and-run are $\widetilde{O}(n^2/\psi_n^2)$ under warmness, see Kannan et al. (1997) and Chen and Eldan (2022) respectively, where $\psi_n$ is the Kannan-Lovasz-Simonovits (KLS) constant with a lower bound $\psi_n \gtrsim \log^{-\frac{1}{2}}(n)$ (Klartag, 2023). One drawback of general-purpose samplers is that a preprocessing

1. $K$ is in the isotropic position if $\mathbb{E}_K(x) = 0$ and $\mathbb{E}_K(xx^\top) = I$

Table 1: Mixing time upper bounds of Dikin walks with different weight choices in their local metrics for sampling a logconcave distribution truncated on a polytope from a warm start. $^{\S}f$ is the negative log-density of the target distribution $\pi(x) \propto e^{-f(x)}\mathbf{1}_K(x)dx$. $^{\#}$Logarithmic factors are omitted in the mixing time upper bounds. $^{\dagger}$Mangoubi and Vishnoi (2023) assumed the polytope $K$ is bounded in a ball of radius $R$. $^{\ddagger}$For the special case where $f$ is quadratic, one can always do an affine transformation so that $\kappa = 1$. $^{\star}\eta$ denotes the spectral norm of the covariance matrix of the target distribution

| Weights in Dikin Walks | Assumptions on $f^{\S}$ | Mixing Time$^{\#}$ |
|---|---|---|
| logarithmic Kannan and Narayanan (2012) | $f \equiv 0$ | $mn$ |
| Approximate John Chen et al. (2018) | $f \equiv 0$ | $n^{5/2}$ |
| logarithmic+$\ell_2$-reg Mangoubi and Vishnoi (2023) | $\beta$-smooth $f$ | $mn + n\beta R^{2\,\dagger}$ |
| logarithmic+Gaussian Kook and Vempala (2024) | quadratic $f$ | $(m+n)n$ |
| Lewis+Gaussian Kook and Vempala (2024) | quadratic $f$ | $n^{5/2}$ |
| logarithmic+$\ell_2$-reg (this paper) | $\alpha$-convex & $\beta$-smooth $f$ | $(m+\kappa)n$ |
| logarithmic+$\ell_2$-reg (this paper) | quadratic $f$ | $mn^{\ddagger}$ |
| logarithmic+$\ell_2$-reg (this paper) | convex & $\beta$-smooth $f$ | $(m+\eta\beta)n^{\star}$ |
| Lewis+$\ell_2$-reg (this paper) | $\alpha$-convex & $\beta$-smooth $f$ | $(n^{3/2}+\kappa)n$ |
| Lewis+$\ell_2$-reg (this paper) | convex & $\beta$-smooth $f$ | $(n^{3/2}+\eta\beta)n$ |

step called rounding is needed to bring the convex bodies to near-isotropic positions. The state-of-art rounding algorithm (Jia et al., 2024) has an oracle complexity of $\widetilde{O}(n^{3.5})$.

In practice, we often have richer information of the target distribution than a membership oracle. For example, a class of interior point methods called Dikin walk was proposed to exploit the structure of the linear constraints. In Dikin walk, a local ellipsoid (often defined by the Hessian of a barrier function) is used as the proposal, which automatically adjust its shape based on its distance to the boundary of $K$. This circumvents the rounding procedure for non-isotropic distributions.

For Dikin walk on uniform polytope sampling, Kannan and Narayanan (2012) proved a mixing time of $\widetilde{O}(mn)$. In the $m \gg n$ regime, by putting weights on different constraints, Chen et al. (2018) proved mixing times of $\widetilde{O}(m^{1/2}n^{3/2})$ and $\widetilde{O}(n^{5/2})$ using Vaidya weights and approximate John weights respectively. To further exploit the Riemannian geometry induced by linear constraints, see Lee and Vempala (2017, 2018); Gatmiry et al. (2024) for developments of geodesic walk and Riemannian Hamiltonian Monte Carlo.

Besides uniform sampling, non-uniform sampling truncated on convex sets also attracted tremendous attention as explained in Section 1. Both ball walk and hit-and-run have been extended to sampling from general logconcave distributions. For well-rounded[2] logconcave distributions, both ball walk (Lovász and Vempala, 2007) and hit-and-run (Lovász and Vempala, 2006) mix in $\widetilde{O}(n^3)$ given a warm start. Similarly, rounding is needed for non-isotropic targets, which has a cost of $\widetilde{O}(n^4)$ given knowledge of maximum of the function $f$ (Lovasz and Vempala, 2006).

The Dikin walk was also extended to non-uniform distributions on polytopes. Narayanan and Rakhlin (2017) defined the Dikin ellipsoid by rescaling the logarithmic metric by a constant. How-

2. The mixing time is $\widetilde{O}(n^2\frac{R^2}{r^2})$ where $\mathbb{E}_\pi(\|x - z_f\|_2^2) \leq R^2$ and the level set of $\Pi$ with measure $1/8$ contains a ball of radius $r$. The well-rounded logconcave distribution is defined when $\frac{R}{r} \approx \sqrt{n}$.

ever, their result applied to the uniform case implies a mixing time of $\widetilde{O}(m^2n^3)$, not matching the $\widetilde{O}(mn)$ bound in Kannan and Narayanan (2012). Later Mangoubi and Vishnoi (2023) proposed a soft-threshold version of Dikin walk by adding a Euclidean metric (as a regularizer) to the local metric and the mixing time is $\widetilde{O}\left((mn + n\beta R^2)\right)$ for a polytope bounded in a ball of radius $R$ and $f$ to be $\beta$-smooth. Kook and Vempala (2024) provided a general framework for combining local metrics in Dikin walk. By lifting up the state-space, adding the Hessian of a Gaussian barrier function to the local metric, Kook and Vempala (2024) achieved a mixing time of $\widetilde{O}((m+n)n)$ for sampling Gaussian distributions truncated on polytopes. To facilitate comparison, we summarize related analysis on Dikin walk with various weight choices in Table 1.

Apart from zeroth-order methods, alternative methods that use gradient information were introduced. For instance, reflected Hamiltonian Monte Carlo where the trajectory is reflected after the sampler hits the constraints, see Pakman and Paninski (2014), Afshar and Domke (2015) for algorithmic designs and simulation. Another approach to deal with the non-smoothness of the truncation is to use proximal-based proposals followed by Metropolis-Hasting step, where Lee et al. (2021) proved a mixing time of $\widetilde{O}(\kappa n)$. However, each step of proximal-sampling requires calling a proximal sampling oracle, an efficient subroutine to draw from the proximal distributions. The designs of such oracles are often highly problem-specific (see Mou et al. (2022) for examples). Moreover, Langevin diffusion has been extended to truncated distributions via projection operations Brosse et al. (2017); Bubeck et al. (2018). Specifically, Bubeck et al. (2018) proved a mixing time of $\widetilde{O}\left(\frac{R^6 \max(n, R\beta)^{12}}{\epsilon^{12}}\right)$ for convex body bounded in a ball of radius $R$ and $\beta$-smooth negative logdensity $f$.

Borrowing the idea of mirror gradient descent in optimization, Zhang et al. (2020); Ahn and Chewi (2021); Jiang (2021); Li et al. (2022) developed mirror Langevin algorithm (MLA) based on various discretization schemes of mirror Langevin diffusion and established mixing time guarantees. More recently, Srinivasan et al. (2024) makes the MLA unbiased by adding a Metropolis-Hasting step. However, the mixing time guarantees there do not directly translate to our setting (Assumption 1.1 or 1.2). Since under the typical MLA framework, the negative logdensity $f$ in Eq (1) is assumed to be strongly convex/smooth/Lipschitz *relative to* $\phi$, where $\phi$ is a specific barrier function defined on $K$ with certain properties such as self-concordance. For polytope $K$, $\phi$ is often chosen to be the log-barrier function. However, simple function $f(x) = \frac{x^\top x}{2}$ fails to satisfy the relative smoothness/convexity/Lipschitz with respect to the log-barrier $\phi$, making the MLA framework not applicable.

## 3. Main Results

We now present the main results of this paper. Section 3.1 establishes mixing time upper bounds for regularized Dikin walks applied to $\alpha$-strongly logconcave and $\beta$-log-smooth distributions truncated on a polytope $K$. Section 3.2 extends these results to weakly logconcave distributions. Section 3.3 improves the above worst-case analysis, showing that the mixing time upper bound of soft-threshold Dikin walk is determined by the fraction of constraints that intersect the high-probability mass of the distribution. Section 3.4 discusses per-step complexity and strategies for obtaining warm starts.

### 3.1. Sampling Strongly Logconcave Distributions

Theorem 3.1, Corollary 3.2, and Theorem 3.3 obtain quantitative mixing time upper bounds, when the target distribution satisfies Assumption 1.1 and $G$ is soft-threshold logarithmic metric in Definition 1.3 or soft-threshold Lewis metric in Definition 1.4.

**Theorem 3.1 (soft-threshold metric, strongly logconcave)** *Assume the target distribution with density $\pi(x) \propto \mathbf{1}_K(x)\exp(-f(x))$ satisfies Assumption 1.1. If in Algorithm 1 we provide the local metric $H$ and regularization size $\lambda = \beta$ in Definition 1.3, then there exists step-size $r > 0$ and a universal constant $C > 0$ such that for any error tolerance $\epsilon > 0$ and any $M$-warm initial distribution $\mu_0$, as long as*

$$T \geq C\,(m+\kappa)\,n\log\left(\frac{\sqrt{M}}{\epsilon}\right),$$

*the output distribution satisfies $\|\mu_T - \Pi\|_{TV} \leq \epsilon$.*

Theorem 3.1 shows that soft-threshold Dikin walk mixes in $\widetilde{O}\left((m+\kappa)n\right)$ for truncated log-concave sampling, ignoring constants and logarithmic factors. Since the uniform distribution on a polytope can be approximated as the limit of Gaussian distributions truncated on the same polytope with increasing covariance and $\kappa = 1$, we recover the $O(mn)$ mixing time from a warm start for the standard Dikin walk (Kannan and Narayanan, 2012) in this limit. Compared to previous work on sampling truncated log-concave distributions, Theorem 3.1 has a few improvements. In Mangoubi and Vishnoi (2023), soft-threshold Dikin walk is proved a mixing time of $\widetilde{O}(mn + n\beta R^2)$, where $R$ is the radius of a ball containing the polytope. Using different proof techniques, we establish a mixing time of $\widetilde{O}(mn + \kappa n)$, removing the dependence on $R$.

Furthermore, the regularized Dikin walk only evaluates $f$ when computing acceptance rates, and no first or higher order derivatives of $f$ are required. Thus, we do not impose any restrictions on the analytical forms of $f$ other than it can be evaluated. This flexibility enables wide statistical applications. For example, $f$ has the form $f(\theta) = \sum_i l(\theta; x_i)$, in Bayesian Lasso logistic regression (Mangoubi and Vishnoi, 2023; Tian et al., 2008) and differentially private optimization (Mangoubi and Vishnoi, 2022), where $l$ denotes a loss function and $\{x_i\}$ denote data points. In contrast, the approach to perform non-uniform sampling truncated on $K$ in Kook and Vempala (2024) is to construct a barrier function induced by $f$ and to add the Hessian of the barrier function into the local metric. This approach evaluates the Hessian of $f$ at each step, which can be costly in above applications.

Additionally, for Gaussian sampling truncated on polytopes specifically, Kook and Vempala (2024) introduces a new barrier walk with a mixing time of $\widetilde{O}(mn+n^2)$ by combining barriers from the polytope and the Gaussian distribution. Here, we prove a mixing time of $\widetilde{O}(mn)$ after reducing the condition number $\kappa$ to 1 through an affine transformation. Concretely, this bound allows us to improve the mixing time to be $O(n)$ when $m = o(n)$, which is highlighted in Corollary 3.2.

**Corollary 3.2 (soft-threshold metric, Gaussian)** *Assume the target distribution with density $\pi(x) \propto \mathbf{1}_K(x)\exp(-f(x))$ satisfies Assumption 1.1. If in Algorithm 1 we provide the local metric $H$ and regularization size $\lambda = \beta$ in Definition 1.3. If we further assume $m = o(n)$, then there exists step-size $r > 0$ and a universal constant $C > 0$ such that for any error tolerance $\epsilon > 0$ and any*

*$M$-warm initial distribution $\mu_0$, as long as*

$$T \geq C\kappa n \log\left(\frac{\sqrt{M}}{\epsilon}\right),$$

*the output distribution satisfies $\|\mu_T - \Pi\|_{TV} \leq \epsilon$.*

In uniform sampling of polytopes, it is typically assumed that $m \geq n$ because otherwise the polytope is unbounded, and the target distribution becomes improper. In truncated strongly logconcave sampling, however, the target distribution remains well-defined even when $m$ is as small as 0. Therefore, in cases where $m \ll n$, we expect a better mixing time than $O(n^2)$. This small-$m$ scenario arises in statistical applications like Bayesian linear models with Gaussian priors under linear inequality constraints (Geweke, 1996; Ghosal and Ghosh, 2022), where $m$ reflects prior information and can be arbitrarily small.

Corollary 3.2 provides a mixing time upper bound which matches that of Random Walk Metropolis (RWM) for unconstrained logconcave sampling (Dwivedi et al., 2019; Andrieu et al., 2024). Specifically when $m = 0$, the polytope part $A_x^\top A_x$ in the soft-threshold metric $G$ in Definition 1.3 vanishes, making soft-threshold Dikin walk the same algorithm as RWM, so a same mixing time bound is expected. Our result of $\widetilde{O}(\kappa n)$ indeed matches the state-of-the-art RWM bound (Andrieu et al., 2024).

Furthermore, Corollary 3.2 demonstrates that for truncated Gaussian sampling, the mixing time scales $O(n)$ when $m = o(n)$. In comparison, the upper bound $\widetilde{O}(mn + n^2)$ in Kook and Vempala (2024) is $\widetilde{O}(n^2)$ under the same condition. This extra $n$-factor arises becuase an $n$-scaling is multiplied to the Gaussian barrier to ensure the $(\nu, \bar{\nu})$-Dikin amenability in Kook and Vempala (2024), which forces them to choose a smaller step size. In contrast, we add an $\ell_2$-regularization to our local metric with a dimension-independent scaling. While this adjustment may appear simple, circumventing the $(\nu, \bar{\nu})$-Dikin amenability requires new proof techniques.

Our proof deviates from the existing ones in two key ways. First, when we control the acceptance rate, we extend the coupling argument in Andrieu et al. (2024) to accommodate asymmetric proposals, which improves the radius $R$ dependent control in Mangoubi and Vishnoi (2023). Second, we develop a novel isoperimetric inequality under a mixed metric that combines Euclidean and cross-ratio distances, which allows us to have a direct analysis of regularized Dikin walks.

Our technique can is applied beyond logarithmic metrics. It is well-known in truncated uniform sampling that the Dikin walk can be improved using Lewis weights in the case $m \gg n$ (Laddha et al., 2020). Following this idea, we define the regularized Dikin walk using Lewis metric in Definition 1.4, and we obtain Theorem 3.3.

**Theorem 3.3 (regularized Lewis metric, strongly logconcave)** *Assume the target distribution with density $\pi(x) \propto \mathbf{1}_K(x)\exp(-f(x))$ satisfies Assumption 1.1. If in Algorithm 1 we provide the local metric $H$ and regularization size $\lambda = \beta$ as in Definition 1.4, then there exists step-size $r > 0$ and a universal constant $C_1 > 0, C_2 > 0$ such that for any error tolerance $\epsilon > 0$ and any $M$-warm initial distribution $\mu_0$, as long as*

$$T \geq C_1(\log m)^{C_2}\left(n^{3/2} + \kappa\right) n \log\left(\frac{\sqrt{M}}{\epsilon}\right),$$

*the output distribution satisfies $\|\mu_T - \Pi\|_{TV} \leq \epsilon$.*

### 3.2. Extension to Weakly Logconcave Distributions

We extend the mixing time upper bounds in Section 3.1 to distributions that are not necessarily $\alpha$-strongly logconcave, and instead only have a finite covariance matrix. This extension is summarized as Theorem 3.4 and 3.5 with $G$ being soft-threshold logarithmic metric as in Definition 1.3 or regularized Lewis metric as in Definition 1.4 respectively.

Here we emphasize that the upper bounds for weakly logconcave distributions lose a factor of the KLS constant $\psi_n$. For a distribution $\mu$ over $\mathbb{R}^n$, its isoperimetric constant is

$$\frac{1}{\psi_\mu} := \inf_{A \subseteq \mathbb{R}^n} \left\{ \frac{\mu^+(A)}{\min\left\{\mu(A), 1-\mu(A)\right\}} \right\}, \tag{4}$$

where $\mu^+$ is the boundary measure of $\mu$. The KLS constant $\psi_n$ is defined by taking supremum of $\psi_\mu$ over all isotropic and logconcave measures over $\mathbb{R}^n$. Kannan et al. (1995) famously conjectures that $\psi_n = O(1)$. Klartag (2023) proved $\psi_n = O(\sqrt{\log n})$, thus we lose at most a log-factor in $n$ in the mixing time upper bounds compared to Section 3.1.

**Theorem 3.4 (soft-threshold metric, weakly logconcave)** *Assume the target distribution $\pi(x) \propto \mathbf{1}_K(x)\exp(-f(x))$ satisfies Assumption 1.2. If in Algorithm 1 we set $G$ to be the soft-threshold metric in Definition 1.3 with regularization size $\lambda := \beta$, then there exists step-size $r > 0$ and a universal constant $C > 0$ such that for any error tolerance $\epsilon > 0$ and any $M$-warm initial distribution $\mu_0$, as long as*

$$T \geq C\psi_n^2\left(m + \beta\eta\right) n \log\left(\frac{\sqrt{M}}{\epsilon}\right),$$

*the output distribution satisfies $\|\mu_T - \Pi\|_{TV} \leq \epsilon$.*

It is worth noting that Theorem 3.4 recovers the mixing time of $\widetilde{O}(mn + n\beta R^2)$ in Mangoubi and Vishnoi (2023), since $\Sigma_\pi \preceq \mathbb{E}_{x\sim\pi}\left(xx^\top\right) \preceq \mathbb{E}_{x\sim\pi}\left(\|x\|_2^2\, I\right) \preceq R^2 I$, thus we have $\eta \leq R^2$.

**Theorem 3.5 (regularized Lewis metric, weakly logconcave)** *Assume the target distribution $\pi(x) \propto \mathbf{1}_K(x)\exp(-f(x))$ satisfies Assumption 1.2. If in Algorithm 1 we provide the local metric $H$ and regularization size $\lambda = \beta$ in Definition 1.4, then there exists step-size $r > 0$ and a universal constant $C_1 > 0, C_2 > 0$ such that for any error tolerance $\epsilon > 0$ and any $M$-warm initial distribution $\mu_0$, as long as*

$$T \geq C_1(\log m)^{C_2}\psi_n^2\left(n^{3/2} + \beta\eta\right) n \log\left(\frac{\sqrt{M}}{\epsilon}\right),$$

*the output distribution satisfies $\|\mu_T - \Pi\|_{TV} \leq \epsilon$.*

### 3.3. Beyond Worst-Case Analysis

In Section 3.3, we focus on sampling from the target distribution under Assumption 1.1 and let the local metric $G$ be the soft-threshold metric in Definition 1.3.

Intuitively, when a target logconcave distribution is well concentrated inside the polytope, even if $m$ is large, the soft-threshold Dikin walk should behave similar to RMW and should share the same mixing time of $\widetilde{O}(\kappa n)$ rather than $\widetilde{O}((m+\kappa)n)$. Based on this intuition, in this section, we seek sufficient conditions on the target distribution for the soft-threshold Dikin walk to achieve a mixing time smaller than that in Theorem 3.1. More precisely, we demonstrate that the mixing time of the soft-threshold Dikin walk depends only on a subset of constraints, those intersect a high probability region of the target distribution, rather than all the constraints. To formalize this phenomenon, we first have to define what we mean by a high probability region.

**Definition 3.6** *Let $\Pi$ be a probability distribution with density $\pi(x) \propto \mathbf{1}_K(x)e^{-f(x)}$, where $f$ is twice differentiable and $\alpha$-convex. For all $s \in (0,1)$ we define $R_o$ as the smallest radius (scaled by $\sqrt{n/\alpha}$) such that the ball centered at the mode contains at least $1-s$ probability mass of $\Pi$:*

$$R_o(s) := \inf\left\{R \geq 0 \middle| \Pi\left(B\left(x^\star, R\sqrt{\frac{n}{\alpha}}\right)\right) \geq 1-s\right\},$$

*where $x^\star := \underset{x\in K}{\arg\min} f(x)$ is the mode of $\Pi$.*

In particular, for $\alpha$-convex $f$, $R_o$ is always upper bounded by the simpler function $\widehat{R_o}$, defined as follows (see Lemma 1 in Dwivedi et al. (2019) for a proof).

$$R_o(s) \leq \widehat{R_o}(s) := 2 + 2\max\left\{\frac{1}{n^{0.25}}\log^{0.25}\left(\frac{1}{s}\right), \frac{1}{n^{0.5}}\log^{0.5}\left(\frac{1}{s}\right)\right\}. \tag{5}$$

For $R, \delta > 0$, we use $\mathcal{B}_R$ and $\mathcal{B}_R^\delta$ to denote the following balls,

$$\mathcal{B}_R := \mathbb{B}\left(x^\star, R\sqrt{\frac{n}{\alpha}}\right), \quad \mathcal{B}_R^\delta := \mathbb{B}\left(x^\star, (R+\delta)\sqrt{\frac{n}{\alpha}}\right). \tag{6}$$

Additionally, given the polytope in Eq. (1), we define $\mathcal{M}_R^\delta$ to be the number of linear constraints violated inside the ball $\mathcal{B}_R^\delta$ as follows,

$$\mathcal{M}_R^\delta := \mathrm{Card}\{i \in [m] \,|\, \exists\, x \in \mathcal{B}_R^\delta \text{ s.t. } a_i^\top x - b_i \leq 0\}.$$

Now we are ready to state Theorem 3.7 which establishes the mixing time of the soft-threshold Dikin walk as a function of how $\mathcal{B}_R^\delta$ intersects the polytope.

**Theorem 3.7** *Assume the target distribution satisfies Assumption 1.1. If in Algorithm 1 we provide the local metric $H$ and regularization size $\lambda = \beta$ in Definition 1.3, then there exists step-size $r > 0$ and a universal constant $C > 0$ such that for any error tolerance $\epsilon > 0$ and any $M$-warm initial distribution $\mu_0$, As long as*

$$T \geq C\inf_{\delta>0}\left[\kappa n + \frac{m}{\delta^2} + n\cdot\mathcal{M}_\Upsilon^\delta\right]\log\left(\frac{2M}{\epsilon}\right), \tag{7}$$

*where $\Upsilon := R_o(\frac{\epsilon}{2M})$ denotes the radius function $R_o$ valued at $\frac{\epsilon}{2M}$ (see Definition 3.6), the output distribution satisfies $\|\mu_T - \Pi\|_{TV} \leq \epsilon$.*

*To ease computation, remark that the result also holds if we replace $\Upsilon$ in Eq. (7) with $\widehat{\Upsilon} := \widehat{R_o}(\frac{\epsilon}{2M})$, defined in Eq. (5).*

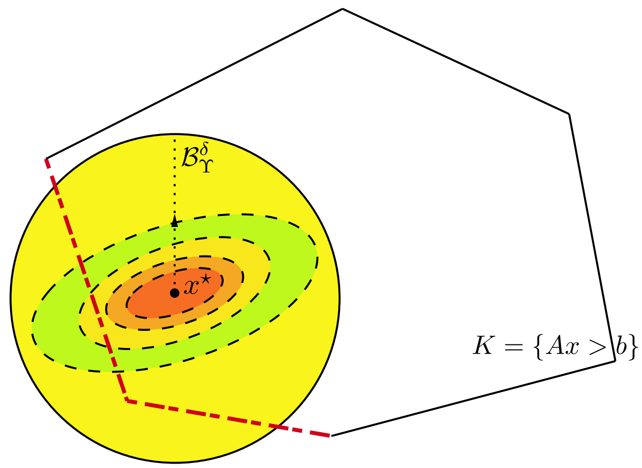


Figure 1: An example of Theorem 3.7 in $\mathbb{R}^2$ ($n = 2$), where $K$ is the polytope and the ball $\mathcal{B}^{\delta}_{\Upsilon}$ refers to the high-probability ball of the truncated distribution $\Pi$. The dashed ellipsoids are contours for the potential $f(x)$ of the distribution. The dashed segments are the constraints that are violated, the solid segments are untouched constraints, so $\mathcal{M}^{\delta}_{\Upsilon} = 2$, $m = 6$.

In Theorem 3.7, note that $\delta$ can be tuned to obtain a more interpretable mixing time bound. If we take $\delta \to \infty$, then $m/\delta^2 \to 0$ and $\mathcal{M}^{\delta}_{\Upsilon} \to m$, we recover the $\widetilde{O}((m + \kappa)n)$ bound in Theorem 3.1.

Figure 1 shows an example where applying Theorem 3.7 can be beneficial than directly applying Theorem 3.1. In general, for a distribution that is well-concentrated inside the polytope, where its high probability mass does not intersect many linear constraints of the polytope, choosing a suitable gap $\delta$ could make the $\mathcal{M}^{\delta}_{\Upsilon}$ increase slower than $m$, as well as ensuring $m/\delta^2$ not increase too much. For the special case where $m = O(n)$, and the number of violated constraints $\mathcal{M}^{\delta}_{\Upsilon} = O(1)$ for some absolute constant $\delta > 0$, we achieved a mixing time in $\widetilde{O}(\kappa n)$, matching the bound for RMW in unconstrained sampling.

### 3.4. Practical Implementations

Finally, since our main results rely on the existence of warm starts, and our mixing time analysis ignores per-iteration cost, we discuss here the per-iteration complexity and warm initialization.

**Per-Iteration Complexity** In our regularized Dikin walk, we assume that in each iteration, we can obtain $G(x)$, $G(x)^{-1}$ and $\det G(x)$ exactly. Our mixing time upper bounds given in Section 3.1 and 3.2 are given without detailing per-step complexity.We show in Appendix G.1 that the number of arithmetic operations needed for soft-threshold metric is $O(\max\{m, n\}\, n^{\omega-1})$ in each iteration. For regularized Dikin walk with regularized Lewis metric, we need to further compute the Lewis weights defined in Eq. (3), which has to be done approximately. The high-accuracy solver for Lewis-weights in Fazel et al. (2022) can be applied with complexity $\widetilde{O}(\max\{m, n\}\, n^{\omega-1})$.

**Feasible $M$-Warm Start** Our mixing time bounds all assume an $M$-warm initial distribution $\mu_0$. Here we propose one initialization approach. Suppose that $K$ is contained within a ball of radius $\widetilde{R}$ and $K$ contains a ball of radius $\widetilde{r}$, then we prove in Appendix G.2 there exists a uniform distribution over a certain ball $\mathbb{B}(x_0, r_0)$ to be $M$-warm where $M$ satisfies

$$\log M \leq 1 + n \log \frac{3\widetilde{R}}{\widetilde{r}} + n \cdot \max \left\{ \frac{1}{2} \log \left( \beta \widetilde{R}^2 \right), \log \left( 2\beta \widetilde{R} \left\| x^\dagger - x^\star \right\|_2 \right) \right\}, \tag{8}$$

where $f$ is $\beta$-smooth, $x^\star$ denotes the global mode and $x^\dagger$ denotes the mode of $\pi(x) \propto e^{-f(x)}$ within the polytope $K$. The ball $\mathbb{B}(x_0, r_0)$ can be computed efficiently by solving $x^\star$ and $x^\dagger$. A $\delta$-approximation of $x^\dagger$ and $x^\star$ can be obtained in $\widetilde{O}(\kappa \log \frac{1}{\delta})$ steps of (projected) gradient descents (Bubeck and others, 2015). Although our mixing time bounds do not assume the $\widetilde{R}$-boundedness and $\widetilde{r}$-inclusion of $K$, these assumptions often arise in sampling literature (Kannan et al., 1997; Lovász and Vempala, 2006; Mangoubi and Vishnoi, 2023).

Note that a factor of $n$ is lost in our mixing time bounds if we plug in the warmness in Eq. (8). Still, we argue that this warmness bound is still reasonable in several aspects. In unconstrained logconcave sampling, a feasible start concentrated around the mode often has a warmness $M = O(\kappa^n)$ (Dwivedi et al., 2019). Moreover, due to the high accuracy nature of Dikin walks, the mixing time from our initialization depends logarithmically on $\frac{\widetilde{R}}{\widetilde{r}}$, $\beta \widetilde{R}^2$ and $\beta \widetilde{R} \left\| x^\dagger - x^\star \right\|_2$. Hence, as long as these terms scale polynomially in $n$, the additional complexity from them remains logarithmic. Reducing this $n$-factor in the bounds under feasible initializations remains an area for future exploration, with techniques such as Gaussian cooling (Kook and Vempala, 2024) and conductance profile (Chen et al., 2020) presenting promising possibilities.

## 4. Sketch of Proofs

We prove the mixing time upper bounds for a general convex body $K$ as long as its associated local metric $G$ satisfies certain sufficient conditions. The mixing time bound for strongly logconcave distribution is summarized in Lemma A.4, where Theorem 3.1, Corollary 3.2 and Theorem 3.3 are its special cases for $K$ being polytope and $G$ taking special forms in Definition 1.3 and Definition 1.4. Similarly, the mixing time bound for weakly logconcave distributions for general $K$ and $G$ is summarized in Lemma A.5, where Theorem 3.4 and 3.5 are its special cases.

To prove the mixing time upper bounds in Lemma A.4 and A.5, we find a lower bound for the conductance $\Phi$ of the transition kernel. To bound the conductance, we first bound the overlap between transition kernels $\|\mathcal{T}_x - \mathcal{T}_y\|_{TV}$ for close enough (measured in the local metric $G$) $x$ and $y$ in $K$. This result is listed as Lemma B.2, where we control the global acceptance rate to be at least around $1/2$ globally, extending the close coupling argument in Andrieu et al. (2024) to our assymetric proposal distributions.

Lemma B.2 shows that two subsets of $K$ with bad conductance $A_1'$ and $A_2'$ are far away in local metric $G$. To continue, we need to show that $K \setminus (A_1' \cup A_2')$ is also large if both $A_1'$ and $A_2'$ are large, so we need a certain isoperimetric inequality under the local metric $G$. Since we need to consider both the logconcave distribution and the convex set $K$, we prove a new isoperimetric inequality (Lemma B.3) under a mixed metric of Euclidean and cross-ratio distances: $d'(x,y) := \max\{d_K(x,y), \log(2)\sqrt{\alpha}\,\|y-x\|_2\}$, and this is also the core innovation behind our mixing time bounds. In the proof, we use a localization argument proposed in Kannan et al. (1995) to reduce the $n$-dimensional integral to 1-dimensional integral.

In order to extend the mixing time bounds to weakly logconcave distributions as in Section 3.2, we extend the mixed isoperimetry to weakly logconcave distributions as well. This extension is listed as Lemma B.4. For technical reasons, we define the mixed metric as $\mathfrak{d}(x,y) =$

$\max\left\{\frac{(\log 2)}{\sqrt{\eta}}\|x-y\|_2, d_K^{\mathcal{H}}(x,y)\right\}$, where $d_K^{\mathcal{H}} := \log(1+d_K)$ is the Hilbert metric on $K$. The proof of Lemma B.4 applies the co-area formula for general metric spaces and the stochastic localization scheme introduced in Eldan (2013). The main idea of stochastic localization is to smoothly multiply the weaky logconcave distribution with a Gaussian part, and the modified distribution is $\alpha$-strongly logconcave.

Going beyond worst-case analysis, we prove Theorem 3.7 using $s$-conductance to cut off the parts that are too close to the boundaries of $K$, see details in Appendix D.3. The proof for warm initialization and per-step complexity is detailed in G.

## 5. Discussions

This work opens several avenues for future extensions. In terms of warm initialization, although we proposed the uniform distribution on a certain ball in Appendix G.2 as the initial distribution, it introduces an extra $n$-factor in the mixing time upper bounds since the warmness $M$ scales exponentially with the dimension $n$. To mitigate this, Gaussian cooling emerges as one promising direction. This method constructs a sequence of distributions to sample from, each serving as a warm start for the next. For truncated sampling, Kook and Vempala (2024) explores a Gaussian cooling procedure under a different framework than ours. Adapting Gaussian cooling techniques to our framework could significantly reduce the warmness dependency. Another potential approach is conductance profiling (Chen et al., 2020).

Moreover, regularized Dikin walk can be generalized to sample distributions on convex bodies defined by nonlinear constraints (see the general mixing time upper bounds in Lemma A.4 and A.5 for details). For instance, we can apply the barrier function for quadratic constraints from Kook and Vempala (2024), and our results yield mixing time bounds for sampling log-concave distributions truncated on ellipsoids.

Finally, our regularized local metric could inspire new mixing time analyses for higher order algorithms such as geodesic walk and RHMC.

## Acknowledgements

Both authors are partially supported by NSF CAREER Award DMS-2237322, Sloan Research Fellowship and Ralph E. Powe Junior Faculty Enhancement Awards.

## References

Hadi Mohasel Afshar and Justin Domke. Reflection, refraction, and Hamiltonian Monte Carlo. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'15, pages 3007–3015, Cambridge, MA, USA, 2015. MIT Press. event-place: Montreal, Canada.

Kwangjun Ahn and Sinho Chewi. Efficient constrained sampling via the mirror-langevin algorithm. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 28405–28418. Curran Associates, Inc., 2021.

James H. Albert and Siddhartha Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, June 1993.

Niccolò Anceschi, Augusto Fasano, Daniele Durante, and Giacomo Zanella. Bayesian conjugacy in probit, tobit, multinomial probit and extensions: A review and new results. *Journal of the American Statistical Association*, 118(542):1451–1469, April 2023.

Christophe Andrieu, Anthony Lee, Sam Power, and Andi Q. Wang. Explicit convergence bounds for Metropolis Markov chains: Isoperimetry, spectral gaps and profiles. *The Annals of Applied Probability*, 34(4), August 2024.

David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.

S. G. Bobkov and C. Houdré. Isoperimetric constants for product probability measures. *The Annals of Probability*, 25(1), January 1997.

Nicolas Brosse, Alain Durmus, Éric Moulines, and Marcelo Pereyra. Sampling from a log-concave distribution with compact support with proximal Langevin Monte Carlo. In Satyen Kale and Ohad Shamir, editors, *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 319–342. PMLR, July 2017.

Sébastien Bubeck and others. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015. Publisher: Now Publishers, Inc.

Sébastien Bubeck, Ronen Eldan, and Joseph Lehec. Sampling from a log-concave distribution with projected Langevin Monte Carlo. *Discrete & Computational Geometry*, 59(4):757–783, June 2018.

Peter Bürgisser, Michael Clausen, and Mohammad Amin Shokrollahi. *Algebraic Complexity Theory*, volume 315 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.

Peter Buser. A note on the isoperimetric constant. In *Annales scientifiques de l'École normale supérieure*, volume 15, pages 213–230, 1982.

Jeff Cheeger. A lower bound for the smallest eigenvalue of the laplacian. *Problems in analysis*, 625 (195-199):110, 1970.

Yuansi Chen. An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture. *Geometric and Functional Analysis*, 31(1):34–61, February 2021.

Yuansi Chen and Ronen Eldan. Hit-and-run mixing via localization schemes, 2022.

Yuansi Chen, Raaz Dwivedi, Martin J Wainwright, and Bin Yu. Fast MCMC sampling algorithms on polytopes. *The Journal of Machine Learning Research*, 19(1):2146–2231, 2018.

Yuansi Chen, Raaz Dwivedi, Martin J. Wainwright, and Bin Yu. Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients. *Journal of Machine Learning Research*, 21 (92):1–72, 2020.

Ben Cousins and Santosh Vempala. Gaussian cooling and $O^*(n^3)$ algorithms for volume and Gaussian volume. *SIAM Journal on Computing*, 47(3):1237–1273, January 2018.

Pierre De La Harpe. On Hibert's metric for simplices. In Graham A. Niblo and Martin A. Roller, editors, *Geometric Group Theory*, pages 97–119. Cambridge University Press, 1 edition, July 1993.

Raaz Dwivedi, Yuansi Chen, Martin J. Wainwright, and Bin Yu. Log-concave sampling: Metropolis-hastings algorithms are fast. *Journal of Machine Learning Research*, 20(183):1–42, 2019.

Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23(2):532–569, April 2013.

Augusto Fasano and Daniele Durante. A class of conjugate priors for multinomial probit models which includes the multivariate normal one. *Journal of Machine Learning Research*, 23(30): 1–26, 2022.

Maryam Fazel, Yin Tat Lee, Swati Padmanabhan, and Aaron Sidford. Computing lewis weights to high precision. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2723–2742, 2022.

Khashayar Gatmiry, Jonathan Kelner, and Santosh S. Vempala. Sampling polytopes with Riemannian HMC: Faster mixing via the Lewis weights barrier. In *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 1796–1881. PMLR, July 2024.

Alan E. Gelfand, Adrian F. M. Smith, and Tai-Ming Lee. Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, 87(418):523–532, June 1992.

John F. Geweke. Bayesian inference for linear models subject to linear inequality constraints. In Jack C. Lee, Wesley O. Johnson, and Arnold Zellner, editors, *Modelling and Prediction Honoring Seymour Geisser*, pages 248–263. Springer New York, New York, NY, 1996.

Rahul Ghosal and Sujit K. Ghosh. Bayesian inference for generalized linear model with linear inequality constraints. *Computational Statistics & Data Analysis*, 166:107335, February 2022.

Hulda S Haraldsdóttir, Ben Cousins, Ines Thiele, Ronan M.T Fleming, and Santosh Vempala. CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*, 33(11):1741–1743, June 2017.

Laurent Heirendt, Sylvain Arreckx, Thomas Pfau, Sebastián N. Mendoza, Anne Richelle, Almut Heinken, Hulda S. Haraldsdóttir, Jacek Wachowiak, Sarah M. Keating, Vanja Vlasov, Stefania Magnusdóttir, Chiam Yu Ng, German Preciat, Alise Žagare, Siu H. J. Chan, Maike K. Aurich, Catherine M. Clancy, Jennifer Modamio, John T. Sauls, Alberto Noronha, Aarash Bordbar, Benjamin Cousins, Diana C. El Assal, Luis V. Valcarcel, Iñigo Apaolaza, Susan Ghaderi, Masoud Ahookhosh, Marouen Ben Guebila, Andrejs Kostromins, Nicolas Sompairac, Hoai M. Le, Ding Ma, Yuekai Sun, Lin Wang, James T. Yurkovich, Miguel A. P. Oliveira, Phan T. Vuong, Lemmer P. El Assal, Inna Kuperstein, Andrei Zinovyev, H. Scott Hinton, William A. Bryant, Francisco J. Aragón Artacho, Francisco J. Planes, Egils Stalidzans, Alejandro Maass, Santosh Vempala, Michael Hucka, Michael A. Saunders, Costas D. Maranas, Nathan E. Lewis, Thomas Sauter, Bernhard Ø. Palsson, Ines Thiele, and Ronan M. T. Fleming. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, 14(3):639–702, March 2019.

Svante Janson. *Gaussian Hilbert Spaces*. Cambridge University Press, 1 edition, June 1997.

He Jia, Aditi Laddha, Yin Tat Lee, and Santosh S. Vempala. Reducing Isotropy and Volume to KLS: Faster Rounding and Volume Algorithms, August 2024. arXiv:2008.02146 [cs].

Qijia Jiang. Mirror langevin monte carlo: the case under isoperimetry. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 715–725. Curran Associates, Inc., 2021.

Ravi Kannan, László Lovász, and Miklós Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13:541–559, 1995.

Ravi Kannan, László Miklós Lovász, and Miklós Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Struct. Algorithms*, 11:1–50, 1997.

Ravindran Kannan and Hariharan Narayanan. Random walks on polytopes and an affine interior point method for linear programming. *Mathematics of Operations Research*, 37(1):1–20, February 2012.

Bo'az Klartag. Logarithmic bounds for isoperimetry and slices of convex sets, June 2023. arXiv:2303.14938 [math].

Bo'az Klartag and Joseph Lehec. Bourgain's slicing problem and KLS isoperimetry up to polylog, April 2022. arXiv:2203.15551 [math].

Yunbum Kook and Santosh S. Vempala. Gaussian cooling and Dikin walks: The interior-point method for logconcave sampling. In Shipra Agrawal and Aaron Roth, editors, *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 3137–3240. PMLR, 30 Jun–03 Jul 2024.

Aditi Laddha, Yin Tat Lee, and Santosh Vempala. Strong self-concordance and sampling. In *Proceedings of the 52nd annual ACM SIGACT symposium on theory of computing*, pages 1212–1222, 2020.

Michel Ledoux. Spectral gap, logarithmic sobolev constant, and geometric bounds. *Surveys in differential geometry*, 9(1):219–240, 2004.

Yin Tat Lee and Aaron Sidford. Solving linear programs with sqrt (rank) linear system solves. *arXiv preprint arXiv:1910.08033*, 2019.

Yin Tat Lee and Santosh S. Vempala. Geodesic walks in polytopes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 927–940, New York, NY, USA, 2017. Association for Computing Machinery.

Yin Tat Lee and Santosh S. Vempala. Convergence rate of Riemannian Hamiltonian Monte Carlo and faster polytope volume computation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 1115–1121, New York, NY, USA, 2018. Association for Computing Machinery.

Yin Tat Lee and Santosh S. Vempala. Eldan's stochastic localization and the KLS conjecture: Isoperimetry, concentration and mixing, January 2019. arXiv:1612.01507 [cs, math].

Yin Tat Lee, Ruoqi Shen, and Kevin Tian. Structured logconcave sampling with a restricted Gaussian oracle. In Mikhail Belkin and Samory Kpotufe, editors, *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 2993–3050. PMLR, August 2021.

Nathan E. Lewis, Harish Nagarajan, and Bernhard O. Palsson. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4):291–305, April 2012.

Ruilin Li, Molei Tao, Santosh S. Vempala, and Andre Wibisono. The Mirror Langevin Algorithm Converges with Vanishing Bias. In Sanjoy Dasgupta and Nika Haghtalab, editors, *Proceedings of The 33rd International Conference on Algorithmic Learning Theory*, volume 167 of *Proceedings of Machine Learning Research*, pages 718–742. PMLR, April 2022.

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

László Lovász and Miklós Simonovits. Random walks in a convex body and an improved volume algorithm. *Random structures & algorithms*, 4(4):359–412, 1993.

László Lovász and Santosh Vempala. Hit-and-run from a corner. *SIAM Journal on Computing*, 35 (4):985–1005, 2006.

Laszlo Lovasz and Santosh Vempala. Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 57–68, Berkeley, CA, USA, 2006. IEEE.

László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms*, 30(3):307–358, 2007.

Oren Mangoubi and Nisheeth Vishnoi. Sampling from log-concave distributions with infinity-distance guarantees. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 12633–12646. Curran Associates, Inc., 2022.

Oren Mangoubi and Nisheeth K. Vishnoi. Sampling from structured log-concave distributions via a soft-threshold Dikin walk. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 31908–31942. Curran Associates, Inc., 2023.

Thomas P Minka. Expectation propagation for approximate bayesian inference. *arXiv preprint arXiv:1301.2294*, 2013.

Wenlong Mou, Nicolas Flammarion, Martin J. Wainwright, and Peter L. Bartlett. An efficient sampling algorithm for non-smooth composite potentials. *Journal of Machine Learning Research*, 23 (233):1–50, 2022.

Hariharan Narayanan and Alexer Rakhlin. Efficient sampling from time-varying log-concave distributions. *Journal of Machine Learning Research*, 18(112):1–29, 2017.

Brian Neelon and David B. Dunson. Bayesian isotonic regression and trend analysis. *Biometrics*, 60(2):398–406, June 2004.

Ari Pakman and Liam Paninski. Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *Journal of Computational and Graphical Statistics*, 23(2):518–542, April 2014.

Sushant Sachdeva and Nisheeth K Vishnoi. The mixing time of the Dikin walk in a polytope—a simple proof. *Operations Research Letters*, 44(5):630–634, 2016.

Vishwak Srinivasan, Andre Wibisono, and Ashia Wilson. Fast sampling from constrained spaces using the Metropolis-adjusted Mirror Langevin algorithm. In Shipra Agrawal and Aaron Roth, editors, *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 4593–4635. PMLR, July 2024.

Guo-Liang Tian, Man-Lai Tang, Hong-Bin Fang, and Ming Tan. Efficient methods for estimating constrained parameters with applications to regularized (lasso) logistic regression. *Computational Statistics & Data Analysis*, 52(7):3528–3542, March 2008.

Sharon J. Wiback, Iman Famili, Harvey J. Greenberg, and Bernhard Ø. Palsson. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *Journal of Theoretical Biology*, 228(4):437–447, June 2004.

Kelvin Shuangjian Zhang, Gabriel Peyré, Jalal Fadili, and Marcelo Pereyra. Wasserstein control of mirror langevin monte carlo. In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 3814–3841. PMLR, 09–12 Jul 2020.

## Appendix A. Definitions, Facts and Sufficient Conditions

The section introduces the notations and definitions used in the proofs. As well as some preliminary facts and necessary conditions for the local metiric $G$ to satisfy.

### A.1. Necessary Definitions

**Definition A.1 (Transition Kernel)** *Assume $\mathcal{X}$ is a Borel measurable subset of $\mathbb{R}^n$, and $\mathcal{B}(\mathcal{X})$ denotes the Borel $\sigma$-algebra over $\mathcal{X}$. Then a Markov chain on $\mathcal{X}$ is characterized by a transition kernel $\mathcal{T}$, which is a function $\mathcal{T} : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to \mathbb{R}^+$ satisfying*

- *For each $x \in \mathcal{X}$, $\mathcal{T}(x, \cdot)$ is a probability measure over $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.*
- *For each $B \in \mathcal{B}(\mathcal{X})$, $\mathcal{T}(\cdot, B)$ is a $\mathcal{B}(\mathcal{X})$-measurable function over $\mathcal{X}$.*

Given a differentiable matrix function $F : \mathbb{R}^n \to \mathbb{R}^{k\times l}$, fix $x \in \mathbb{R}^n$ and $h \in \mathbb{R}^n$, we use $\mathcal{D}F(x)[h]$ to denote the derivative of $F$ at $x$ in the direction $h$:

$$\mathcal{D}F(x)[h] := \lim_{t\to 0} \frac{F(x + th) - F(x)}{t}.$$

We use $\mathbb{B}(x, r)$ as a shorthand for a ball centered at $x \in \mathbb{R}^n$ with radius $r > 0$. In other words,

$$\mathbb{B}(x, r) := \left\{ y \in \mathbb{R}^n \middle| (y - x)^\top (y - x) \le r^2 \right\}.$$

Without ambiguity, $\mathbb{B}_r$ denotes the ball centered at 0 in $\mathbb{R}^n$ with radius $r$. We let $E(x, G(x), r)$ denote the ellipsoid $E(x, G(x), r) := \{z|(z - x)^\top G(x)(z - x) \le r^2\}$.

Now we formally introduce the transition kernels before and after the Metropolis step respectively. We use $\mathcal{P}_x$ to denote the proposal distribution $\mathcal{P}_x := \mathcal{N}(x, \frac{r^2}{n}G(x)^{-1})$ and $\mathcal{T}_x$ to denote the probability distribution after one step of Metropolis filter. In other words,

$$\mathcal{T}_x(B) := [1 - \mathbb{E}_{z\sim\mathcal{P}_x} (\alpha(x, z))]\, \delta_x(B) + \int_{z\in B} \alpha(x, z)\mathcal{N}\left(z; x, \frac{r^2}{n}G(x)^{-1}\right) dz, \tag{9}$$

where $B$ is Borel set in $\mathbb{R}^n$, $\delta_x$ is the Dirac measure at $x$, and $\alpha(x, z)$ is the acceptance rate given by

$$\alpha(x, z) := \min\left\{1, \mathbf{1}_K(z)\frac{\exp(-f(z))}{\exp(-f(x_t))}\frac{\mathcal{N}\left(x_t; z, \frac{r^2}{n}G(z)^{-1}\right)}{\mathcal{N}\left(z; x_t, \frac{r^2}{n}G(x_t)^{-1}\right)}\right\}. \tag{10}$$

It is worth noting that the actual transition kernel for each step in Algorithm 1 is $\overline{\mathcal{T}}$, the lazification of $\mathcal{T}$. In other words,

$$\overline{\mathcal{T}}_x(B) := \frac{1}{2}\delta_x(B) + \frac{1}{2}\mathcal{T}_x(B),$$

and it is easy to verify that the target distribution $\Pi$ in Eq. (1) is the stationary distribution of both $\mathcal{T}$ and $\overline{\mathcal{T}}$.

### A.2. Affine Transformation for Truncated Gaussian Sampling

In the special case of sampling a Gaussian distribution truncated on a polytope $K = \{x|Ax > b\}$, the negative log-density $f(x) := (x-\mu)^\top \Sigma(x-\mu)$, where $\mu$, $\Sigma$ are the mean and covariance of the untruncated Gaussian. Note that without loss of generality, we can assume that the condition number $\kappa$ is 1. This is because one can sample an affine transformed distribution and recover samples of the original distribution by an inverse affine transform. More precisely, consider the affine transformation $\mathcal{A} : x \mapsto \Sigma^{-\frac{1}{2}}(x-\mu)$, which transforms a random variable $X \sim \pi$ to $Y := \mathcal{A}(X)$. The density of $Y$ satisfies

$$\pi_Y(y) \propto \pi_X(\mathcal{A}^{-1}(y)) \propto \exp\left(-\frac{1}{2}y^\top y\right)\mathbb{I}_{\left\{A\Sigma^{\frac{1}{2}}y > b - A\mu\right\}}. \tag{11}$$

Let $\widetilde{A} = A\Sigma^{\frac{1}{2}}$ and $\widetilde{b} = b - A\mu$, we can first sample $\exp\left(-\frac{1}{2}y^\top y\right)$ truncated on $\left\{\widetilde{A}x > \widetilde{b}\right\}$ and apply an inverse affine transform to obtain samples from $\pi$.

### A.3. Sufficient Conditions on Local Metrics

We define a few properties of a local metric $G$ that frequently appear in optimization and sampling literature. Since we do not restrict the convex set upon which the target distribution is truncated to be bounded/compact, we first need to extend the cross-ratio distances defined on convex bodies to unbounded convex sets.

**Definition A.2** *Assume $K \subseteq \mathbb{R}^n$ is an open and convex set. Fix $x, y \in K$, we define the cross-ratio distance $d_K(x,y)$ under three scenarios:*

- *If the line $\overline{xy}$ intersects $\partial K$ (the boundary of $K$) with $p, q$, and in the order $p, x, y, q$, then*

$$d_K(x,y) := \frac{\|p-q\|_2 \|x-y\|_2}{\|p-x\|_2 \|q-y\|_2}.$$

- *If one of $\{p,q\}$ is at infinity, then we delete the two distances involving that point. In other words, if $p = \infty$, then $d_K(x,y) := \frac{\|x-y\|_2}{\|q-y\|_2}$; If $q = \infty$, then $d_K(x,y) := \frac{\|x-y\|_2}{\|p-x\|_2}$.*
- *If both $\{p,q\}$ are at infinity, then $d_K(x,y) := 0$.*

It is worth noting that the above three scenarios can be summarized in one equivalent definition: Fix convex set $K$, fix $x, y \in K$, the cross-ratio distance is defined by $d_K(x,y) = \lim_{R\to\infty} d_{K\cap\mathbb{B}(0,R)}(x,y)$. Moreover, for all $x \in \partial K$, $y \in K$ and $y \neq x$, we have $d_K(x,y) = +\infty$. We just avoid this technical issue by requiring $K$ to be an open subset of $\mathbb{R}^n$ in this paper. Our mixing time result can be extended to non-open $K$ easily, because the Lebesgue measure of the boundary $\partial K$ of a convex set $K$ is 0.

The statement and proof of our main theorems involve a few properties for $G$ to satisfy, following Kook and Vempala (2024).

**Definition A.3 (Laddha et al. (2020); Kook and Vempala (2024))** *A matrix function $G : K \to \mathbb{S}^n_+$ is said to satisfy*

- *Strong self-concordance (SSC) if $G$ is positive-definite on $K$ and*

$$\left\| G(x)^{-\frac{1}{2}}\mathcal{D}G(x)[h]G(x)^{-\frac{1}{2}} \right\|_F \leq 2\left\|h\right\|_{G(x)}, \quad \forall x \in K, h \in \mathbb{R}^n.$$

- *Lower trace self-concordance(LTSC) if $G$ is positive-definite on $K$ and*

$$\operatorname{Tr}\left\{G(x)^{-1}\mathcal{D}^2G(x)[h,h]\right\} \geq -\left\|h\right\|_{G(x)}^2, \quad \forall x \in K, h \in \mathbb{R}^n.$$

  *We say it satisfies strongly lower trace self-concordant (SLTSC) if for any PSD matrix function $\overline{G}$ on $K$ it holds that*

$$\operatorname{Tr}\left\{\left(\overline{G}(x) + G(x)\right)^{-1}\mathcal{D}^2G(x)[h,h]\right\} \geq -\left\|h\right\|_{G(x)}^2, \quad \forall x \in K, h \in \mathbb{R}^n.$$

- *Average self-concordance (ASC) if for any $\epsilon > 0$ there exists $r_\epsilon > 0$ such that for $r \leq r_\epsilon$,*

$$\mathbb{P}_{z\sim\mathcal{N}(x,\frac{r^2}{d}G(x)^{-1})}\left(\left|\left\|z-x\right\|_{G(z)}^2 - \left\|z-x\right\|_{G(x)}^2\right| \leq \frac{2\epsilon r^2}{n}\right) \geq 1-\epsilon.$$

  *We say it satisfies strongly average self-concordant (SASC) if for any $\epsilon > 0$ and any PSD matrix function $\overline{G}$ on $K$ it holds that*

$$\mathbb{P}_{z\sim\mathcal{N}(x,\frac{r^2}{d}[G(x)+\overline{G}(x)]^{-1})}\left(\left|\left\|z-x\right\|_{G(z)}^2 - \left\|z-x\right\|_{G(x)}^2\right| \leq \frac{2\epsilon r^2}{n}\right) \geq 1-\epsilon.$$

Note that we make the ASC definition stricter than it was in Kook and Vempala (2024), by requiring the difference in local norms to be bounded on both sides. This modification does not have a major impact on the results, because the verification of ASC is based on Taylor approximation as in Kook and Vempala (2024), which automatically ensures that both sides are controlled.

Now we are prepared to give Lemma A.4, which establishes sufficient conditions for a regularized Dikin walk with a local metric to mix fast on $\alpha$-strongly logconcave and $\beta$-log-smooth distributions. Theorem 3.1, Corollary 3.2, and Theorem 3.3 are applications of Lemma A.4 to the case when $K$ is a polytope and $G$ is soft-threshold logarithmic metric or regularized Lewis metric respectively.

**Lemma A.4** *Let $\Pi$ be a target distribution with density $\pi(x) \propto \mathbf{1}_K(x)e^{-f(x)}$, where $K \subseteq \mathbb{R}^n$ is an open and convex set, $\alpha I_n \preceq \nabla^2 f \preceq \beta I_n$. If in Algorithm 1 we provide the local metric $H$ and regularization size $\lambda > 0$ such that $G := H + \lambda I$ satisfies*

- *SSC, LTSC and ASC in Definition A.3,*
- *$G(x) \succeq \beta I$ and the ellipsoid $E(x, G(x), 1) \subseteq K$ for all $x \in K$,*
- *there exists $C_K \geq 1$ and $C_E \geq \beta$, such that*

$$\min\left\{\left\|y-x\right\|_{G(x)}^2, \left\|y-x\right\|_{G(y)}^2\right\} \leq C_K \cdot d_K(x,y)^2 + C_E \cdot \left\|y-x\right\|_2^2, \forall x, y \in K,$$

*then there exists a step size $r > 0$ and a universal constant $C > 0$ such that for any error tolerance $\epsilon > 0$ and any $M$-warm initial distribution $\mu_0$, as long as*

$$T \geq C\left(C_K + \frac{C_E}{\alpha}\right) n \log\left(\frac{\sqrt{M}}{\epsilon}\right),$$

*the output distribution satisfies $\|\mu_T - \Pi\|_{TV} \leq \epsilon$.*

We extend the mixing time upper bound (Lemma A.4) to distributions that are not necessarily $\alpha$-strongly logconcave, and instead only has a finite covariance matrix. This extension is summarized as Lemma A.5. Theorems 3.4 and 3.5 are the applications of Lemma A.5 to polytopes, and $G$ being soft-threshold logarithmic metric or regularized Lewis metric respectively.

As we have emphasized in Section 3.2 the upper bounds for weakly logconcave distributions lose a factor of the Kannan-Lovász-Simonovits (KLS) constant $\psi_n$, which appears in a new isoperimetric inequality proved by us under a mixed metric for weakly logconave measures.

**Lemma A.5** *Let $\Pi$ be a target distribution with density $\pi(x) \propto \mathbf{1}_K(x) e^{-f(x)}$, where $K \subseteq \mathbb{R}^n$ is an open and convex set, $0 \preceq \nabla^2 f \preceq \beta I_n$. Suppose the covariance matrix of $\Pi$ is bounded as $\Sigma_\pi \preceq \eta I_n$. If in Algorithm 1 we provide the local metric $H$ and regularization size $\lambda > 0$ such that $G := H + \lambda I$ satisfies*

- *$G$ is SSC, LTSC, ASC*
- *$G(x) \succeq \beta I$ and the ellipsoid $E(x, G(x), 1) \subseteq K$ for all $x \in K$.*
- *There exists $C_K \geq 1$ and $C_E \geq \beta$, such that*

$$\min\left\{\|y - x\|_{G(x)}^2, \|y - x\|_{G(y)}^2\right\} \leq C_K \cdot d_K(x, y)^2 + C_E \cdot \|y - x\|_2^2, \forall x, y \in K,$$

*Then there exists a step-size $r > 0$ and a universal constant $C > 0$ such that for any $M$-warm initial distribution $\mu_0$ and any error tolerance $\epsilon > 0$, as long as*

$$T \geq C\psi_n^2 (C_K + C_E \eta) n \log\left(\frac{\sqrt{M}}{\epsilon}\right),$$

*the output distribution satisfies $\|\mu_T - \Pi\|_{TV} \leq \epsilon$.*

## Appendix B. Main Lemmas and Proof Outline

Now we provide the necessary lemmas for establishing mixing times of regularized Dikin walk for sampling truncated logconcave distributions.

We bound the mixing time by finding a lower bound for the conductance of the transition kernel $\overline{\mathcal{T}}$, which is the transition kernel $\mathcal{T}$ after lazification, and the connection between the mixing time and conductance can be found in Lovász and Simonovits (1993). For the sake of completeness, we list it as Lemma B.1. Since the lazification $\overline{\mathcal{T}}$ at most shrinks the conductance by a factor of 2, we only need to bound the conductance of the transition kernel $\mathcal{T}$ before lazification.

We first introduce the important concepts of ($s$-)conductance of Markov chains. Given the transition kernel $\mathcal{T}$ of a Markov chain, and $\Pi$ be its stationary distribution, we define its conductance $\Phi$ and $s$-conductance $\Phi_s$ to be (we require $0 < s < 1/2$):

$$\Phi := \inf_{0<\Pi(A)\leq\frac{1}{2}} \frac{\int_A \mathcal{T}_u(A^c)\Pi(du)}{\Pi(A)}, \quad \Phi_s := \inf_{s<\Pi(A)\leq\frac{1}{2}} \frac{\int_A \mathcal{T}_u(A^c)\Pi(du)}{\Pi(A)-s}. \tag{12}$$

**Lemma B.1 (Lovász and Simonovits (1993),Lovász (1999))** *For a reversible lazy Markov chain defined by the transition kernel $\widetilde{\mathcal{T}}$, let $\Pi$ denote its unique stationary distribution and $\widetilde{\Phi}$, $\widetilde{\Phi}_s$ denote its conductance, $s$-conductance respectively. Given an $M$-warm start $\mu_0$, the convergence to $\Pi$ can be controlled using the conductance:*

$$\left\|\widetilde{\mathcal{T}}^k(\mu_0) - \Pi\right\|_{TV} \leq \sqrt{M}\left(1 - \frac{1}{2}\widetilde{\Phi}^2\right)^k. \tag{13}$$

*One can achieve a similar result using the notion of $s$-conductance:*

$$\left\|\widetilde{\mathcal{T}}^k(\mu_0) - \Pi\right\|_{TV} \leq Ms + M\left(1 - \frac{\widetilde{\Phi}_s^2}{2}\right)^k \leq Ms + M\exp\left(-\frac{k\widetilde{\Phi}_s^2}{2}\right). \tag{14}$$

Lemma B.1 is well-known and its proof can be found in Lovász and Simonovits (1993), so we omit its proof in this paper.

In controlling the transition overlap in Lemma B.2, the acceptance rate is controlled around $1/2$ globally (we extend the close coupling argument in Andrieu et al. (2024) to our case of assymetrical proposal distributions), we do not need to cut off small probability regions, and using the conductance $\Phi$ is adequate for Lemma A.4 and A.5. However, when showing the mixing time depends on a fraction of linear constraints in Theorem 3.7, we need to cut off the part of $K$ that is too close to the remaining linear constraints, thus the $s$-conductance and Eq. (14) are needed.

As argued in Section 4, we now give Lemma B.2 that controls the transition overlap between $\mathcal{T}_x$ and $\mathcal{T}_y$ to be small when $x$ and $y$ are close measured in the local metric $G$.

**Lemma B.2 (Transition Overlap)** *Let $G$ be a SSC, ASC, LTSC matrix function defined on $K$, and we further assume that $G(x) \succeq \beta I_n$ for all $x \in K$. Let $\mathcal{T}$ be the transition overlap of soft-threshold Dikin walk as defined in Eq. 9, then we can set a step-size $r > 0$ such that for all $x, y$ such that $\|x-y\|_{G(x)} \leq \frac{r}{10\sqrt{n}}$, we have*

$$\|\mathcal{T}_x - \mathcal{T}_y\|_{TV} \leq \frac{4}{5}$$

Note that Lemma B.2 is very similar to the one-step coupling argument in Kook and Vempala (2024), and we can directly apply Lemma B.3 in Kook and Vempala (2024) and get the transition overlap we desire. However, we are motivated by the close coupling argument in Andrieu et al. (2024), and we extend it to our assymetrical proposal distributions, which is a different angle from the direct computation appeared in Kook and Vempala (2024). We also provide a modular and streamlined format of proof, making it convenient for readers and facilitates potential applications. Namely, we separate the proof of Lemma B.2 into three parts: bounding the acceptance rate to be around $1/2$ globally, extending the close coupling argument in Andrieu et al. (2024), and controlling the TV-distance between two proposal distributions. So we still provide our proof of Lemma B.2 in Appendix E.2.

Now we can give the isoperimetric inequality under a combination of Euclidean distance and cross-ratio distance in this extended sense.

**Lemma B.3** *Suppose $\Pi$ is a probability distribution supported on a (possibly unbounded) convex set $K \subseteq \mathbb{R}^n$ and is more logconcave than Gaussian with covariance $\frac{1}{\alpha} I_n$ (see Eq. (2)). Assume $K$ is partitioned into three measurable sets $K = S_1 \sqcup S_2 \sqcup S_3$, then we have*

$$\Pi(S_3) \geq d'(S_1, S_2)\Pi(S_1)\Pi(S_2), \tag{15}$$

*where $d'$ is a mixed distance defined by $d'(x, y) = \max\{d_K(x, y), \log(2)\sqrt{\alpha}\|x - y\|_2\}$ and the corresponding distance between two sets $S_1$ and $S_2$ is defined as: $d'(S_1, S_2) := \inf_{(x,y)\in S_1\times S_2} d'(x, y)$. Note that here $d_K$ refers to the cross-ratio distance defined over $K$ in the extended sense as in Definition A.2.*

The reason we need Lemma B.3 is that, to use Lemma B.2 to bound the conductance, we need a suitable isoperimetric inequality for the local metric $G$. The conventional wisdom is to use the isoperimetric inequality for cross-ratio distance $d_K$ over the polytope $K$ Mangoubi and Vishnoi (2023); Kook and Vempala (2024), and furthermore an upper bound of $G$ by local metric $d_K$ is needed. This upper bound is often found by the $\bar{\nu}$-symmetry of a local metric introduced in Laddha et al. (2020). However, such an upper bound does not exist in our case, since the Euclidean term $\lambda I$ simply can not be bounded by cross-ratio distance. In fact, for unbounded convex set $K$, we can always find some $x, y$ such that the Euclidean distance $\beta\|y - x\|_2$ is of higher order of magnitude than $d_K(x, y)$.

We defer the proof of Lemma B.3 to Section C.1. Isoperimetric inequality under either metric (cross-ratio distance/ Euclidean distance) is already well-known, while their combination is not trivial in that the new distance $d'(x, y) := \max\{d_K(x, y), \sqrt{\alpha}\log 2\|x - y\|_2\}$ for two sets $d'(S_1, S_2)$ may not be achieved when either $d_K(S_1, S_2)$ or $d_{\text{euclid}}(S_1, S_2)$ is achieved.

In order to extend the mixing time results to weakly logconcave distributions as in Lemma A.5, we need to extend this new isoperimetry to weakly logconcave distributions. However, given a convex body $K$, it is well-known that the cross-ratio distance $d_K$ in Definition A.2 does not satisfy triangle inequality. To circumvent this inconvenience, we use the Hilbert metric $d_K^{\mathcal{H}}$ instead:

$$d_K^{\mathcal{H}}(x, y) := \log(1 + d_K(x, y)), \tag{16}$$

where $d_K$ is the cross-ratio distance defined in the extended sense for possibly unbounded convex sets as in Definition *A.2*. Unlike cross-ratio distances, Hilbert metric satisfies the triangle inequality. As we will see, since $d_K^{\mathcal{H}}(x, y) \approx d_K(x, y)$ for close $x, y$ by definition, thus changing $d_K$ to $d_K^{\mathcal{H}}$ would not impact the application of the isoperimetric inequality to bound mixing times. Now we give the new isoperimetry for weakly logconcave distributions.

**Lemma B.4** *Suppose $\Pi$ is a logconcave distribution supported on a open convex set $K \subseteq \mathbb{R}^n$, assume that $\Pi$ has a bounded covariance matrix $\Sigma_\pi$, and let $\eta = \|\Sigma_\pi\|_2$ denote its spectral norm. Then for any Borel measurable decomposition $K = S_1 \sqcup S_2 \sqcup S_3$, we have*

$$\Pi(S_3) \geq \frac{1}{6\max\{1, \psi_n\}} \cdot \mathfrak{d}(S_1, S_2)\Pi(S_1)\Pi(S_2), \tag{17}$$

*where $\mathfrak{d}$ is a mixed metric defined by $\mathfrak{d}(x,y) = \max\left\{\frac{(\log 2)}{\sqrt{\eta}}\left\|x-y\right\|_2, d_K^{\mathcal{H}}(x,y)\right\}$, and $\psi_n$ is the KLS constant.*

Klartag (2023) proved the KLS constant $\psi_n$ satisfies $\psi_n = O(\sqrt{\log n})$, thus we lost at most a logarithmic factor in the mixing time for transitioning to weakly logconcave distributions.

The proof of Lemma B.4 is left to Section C.2. The key technique behind this extension to weakly logconcave distributions is the stochastic localization scheme introduced in Eldan (2013). The main idea of stochastic localization is to smoothly multiply the weaky logconcave distribution with a Gaussian part, and the modified distribution is $\alpha$-strongly logconcave. As a result, we can apply our new isoperimetry (Lemma B.3) on $\alpha$-strongly logconcave distributions to the modified distribution. While stochastic localization has been studied extensively to upper bound the KLS constant $\psi_n$ Lee and Vempala (2019); Chen (2021); Klartag and Lehec (2022); Klartag (2023), there are still several difficulties to apply it to our new isoperimetry. The evolution of the measures of sets of arbitrary initial size (for $S_1, S_2, S_3$ as Lemma B.4) is unclear Chen and Eldan (2022). We circumvent this issue by instead proving a boundary version of the isoperimetric inequality. To do that, we introduce the boundary measures Bobkov and Houdré (1997) for a general metric space, and also the co-area formula Bobkov and Houdré (1997) for a general metric spaces that allows us to convert back to Lemma B.4. To prove the boundary version of isoperimetry, we used an bound on the evolution of $L^2$-functions from Klartag (2023) over the distribution $\Pi$. Finally, we show that the target distribution is compatible with the topology under $\mathfrak{d}$ by proving the open convex set $K$ under the mixed metric $\mathfrak{d}$ has the usual Euclidean topology, so we can apply the co-area formula correctly.

## Appendix C. Combining Euclidean & Hilbert Isoperimetries

In this section, we prove a new isoperimetric inequality which uses a metric that combines the Euclidean metric and the cross-ratio distance/Hilbert metric over the convex set $K$. In Section C.1 we prove Lemma B.3, the isoperimetric inequality for $\alpha$-strongly logconcave distributions truncated on a convex set $K$. In Section C.2, using stochastic localization, we extend the isoperimetric inequality to weakly log-concave distributions truncated on $K$ with finite covariance, and prove Lemma B.4.

### C.1. Isoperimetry for Strongly Logconcave Distributions

For unconstrained logconcave sampling on Euclidean space, an isoperimetric inequality for $\alpha-$strongly logconcave distributions using Euclidean metric is well-studied Cousins and Vempala (2018) (listed as Proposition C.1). While in constrained sampling, an isoperimetric inequality for logconcave distributions using cross-ratio distance is often used instead Lovász and Vempala (2007) (listed as Proposition C.2). This is because one can often use cross-ratio distance to bound the metric induced by the polytope using the concept of $\bar{\nu}-$symmetry introduced in Laddha et al. (2020).

Both our local metrics in Definition 1.3 and 1.4 can be viewed as the summation of a Hessian metric and a Euclidean metric. Since one metric cannot be bounded by the other with dimension-independent factors, the two isoperimetric inequalities above are not useful for obtaining tight mixing time bounds. We propose a new isoperimetric inequality that combines Proposition C.1 and C.2. Following the localization lemma for proving isoperimetric inequalities Lovász and Vempala (2007); Cousins and Vempala (2018), we transform the inequality over $n$-dimensional integrals into an inequality over 1-dimensional integrals. Then we check the 1-dimensional isoperimetric inequality that uses the maximum of two metrics.

**Proposition C.1 (Theorem 5.4 in Cousins and Vempala (2018))** *Assume $\Pi$ is a probability distribution on $\mathbb{R}^n$ and is more logconcave than Gaussian with covariance $\frac{1}{\alpha}I_n$ (see Eq. (2)) then we have*

$$\Pi(S_3) \geq (\log 2)\sqrt{\alpha} \cdot d_{\text{euclid}}(S_1, S_2)\Pi(S_1)\Pi(S_2).$$

**Proposition C.2 (Theorem 2.5 in Lovász and Vempala (2007))** *Let $\Pi$ be a logconcave distribution whose support $K \subseteq \mathbb{R}^n$ is partitioned into three measurable sets: $K = S_1 \sqcup S_2 \sqcup S_3$, then we have*

$$\Pi(S_3) \geq d_K(S_1, S_2)\Pi(S_1)\Pi(S_2).$$

**Proposition C.3 (Integrals on $\mathbb{R}^1$ Lovász and Vempala (2007); Kannan et al. (1995))** *For $a$, $b$, $c$, $d$ in $\mathbb{R}$ such that $a < b < c < d$, and let $g : \mathbb{R} \to \mathbb{R}_{++}$ be a logconcave function , we have*

$$\frac{\int_a^d g(t)dt \int_b^c g(t)dt}{\int_a^b g(t)dt \int_c^d g(t)dt} \geq \frac{(d-a)(c-b)}{(b-a)(d-c)}.$$

*Moreover, for any one-dimensional isotropic function $f$, and any partition $S_1, S_2, S_3$ of the real line,*

$$\Pi_f(S_3) \geq \log(2)d_{\text{euclid}}(S_1, S_2)\Pi_f(S_1)\Pi_f(S_2).$$

Now we are prepared to give the proof of the new isoperimetric inequality (Lemma B.3), and the proof is similar to Theorem 2.5 in Lovász and Vempala (2007).

**Proof** [Proof of Lemma B.3] Let $\pi$ be the (unnormalized) density function of $\Pi$ supported on $K$, then $f$ is continuous over $K$ . Let $h_i$ be the indicator function of $S_i$ for $i = 1, 2, 3$, and let $h_4$ be the indicator function of $K$. It suffices to prove that:

$$d'(S_1, S_2)(\int_{\mathbb{R}^n} \pi h_1)(\int_{\mathbb{R}^n} \pi h_2) \leq (\int_{\mathbb{R}^n} \pi h_3)(\int_{\mathbb{R}^n} \pi h_4). \tag{18}$$

For any $a, b \in K$, any non-negative linear function $l : [a, b] \to \mathbb{R}_{++}$. Set $v(t) = ta + (1-t)b$, and we define the following integral over the needle $N = ([a, b], l)$:

$$J_i = \int_0^1 h_i(v(t))\pi(v(t))l(v(t))^{n-1}dt.$$

According to the localization lemma (the detailed discription and proof appeared as Corollary 2.2 in Kannan et al. (1995)), we can reduce the problem to $n$-dimensional integral inequality to 1-dimensinoal integrals. It suffices to prove that:

$$d'(S_1, S_2)J_1 \cdot J_2 \leq J_3 \cdot J_4. \tag{19}$$

For convenience, we define $I_i := \{t|h_i(v(t)) > 0\}$ for $i = 1, 2, 3, 4$ to be subsets of $[0, 1]$, and $I_4 = [0, 1] = I_1 \sqcup I_2 \sqcup I_3$. We first prove the special case that $I_1, I_2, I_3$ are intervals $[0, u_1], [u_2, 1]$ and $(u_1, u_2)$ respectively.

If $u_1 = 0$ or $u_2 = 1$, then $J_1 = 0$ or $J_2 = 0$, thus Eq. (19) is trivially true. If $u_1 = u_2$, this implies $v(u_1) = v(u_2)$, since we also have $v(u_1) \in S_1, v(u_2) \in S_2$, thus $d_K(v(u_1), v(u_2)) = 0$ and $\|v(u_1) - v(u_2)\|_2 = 0$, this implies $d'(S_1, S_2) \leq d'(v(u_1), v(u_2)) = 0$, again the Inequality (19) is trivially true. So we can assume that $0 < u_1 < u_2 < 1$.

Set $c_i = u_i a + (1 - u_i)b$ for $i = 1, 2$. It is clear that $c_i \in S_i$, thus $d'(c_1, c_2) \geq d'(S_1, S_2)$, that is

$$\max\left\{d_K(c_1, c_2), (\log 2)\sqrt{\alpha}\|c_1 - c_2\|_2\right\} \geq d'(S_1, S_2).$$

So we only need to prove the two following inequalities:

$$d_K(c_1, c_2)J_1 \cdot J_2 \leq J_3 \cdot J_4, \quad (\log 2)\sqrt{\alpha}\,\|c_1 - c_2\|_2\, J_1 J_2 \leq J_3 J_4. \tag{20}$$

We can prove the first inequality in Eq. (20). we have

$$d_K(c_1, c_2) \overset{(i)}{\leq} \frac{\|b - a\|_2\,\|c_2 - c_1\|_2}{\|c_1 - a\|_2\,\|b - c_2\|_2} \overset{(ii)}{\leq} \frac{J_3 \cdot J_4}{J_1 \cdot J_2},$$

where inequality $(i)$ holds because $a, c_1, c_2, b \in K$ are on the same line, and the intersection $p, q$ of $\overline{c_1 c_2}$ (in the order $p, c_1, c_2, q$) with $K$ always satisfies $\|p - c_1\|_2 \geq \|c_1 - a\|_2$ and $\|q - c_2\|_2 \geq \|b - c_2\|_2$ including the case when $p$ or $q$ are at the infinity. Inequality $(ii)$ holds since $\pi(v(t))l(v(t))^{n-1}$ is a logconcave function in $t$ and we use Proposition C.3.

Then we prove the second inequality in Eq. (20), for convenience, we can directly employ Proposition C.1 over the 1-dimensional special case. We observe the integrand of $J_i$ $(i = 1, 2, 3, 4)$ satisfies

$$\begin{aligned}
\pi_{1d}(t) =& h_i(v(t))\pi(v(t))l(v(t))^{n-1} \\
\overset{(i)}{\propto}& h_i(v(t))q(v(t)) \exp\left[-\frac{\alpha}{2}v(t)^\top v(t)\right] l(v(t))^{n-1} \\
\propto& h_i(v(t))q(v(t))l(v(t))^{n-1} \exp\left(-\alpha(a - b)^\top bt\right) \exp\left(-\frac{\alpha}{2}\|a - b\|_2^2 t^2\right),
\end{aligned}$$

where in proportion $(i)$, $q$ is a logconcave function since we assumed $\pi$ to be more logconcave than Gaussian with covariance $\frac{1}{\alpha}I_n$. We notice that if $g : \mathbb{R}^n \to \mathbb{R}^+$ is a logconcave function and $v : \mathbb{R} \to \mathbb{R}^n$ is linear, then $g \circ v$ is logconcave. Since $h_i, f, l$ are logconcave, thus $h_i(v(t)), q(v(t)), l(v(t))^{n-1}$ are logconcave in $t$, and $\exp(-\alpha(a - b)^\top bt)$ is clearly logconcave. Together with the fact that the product of logconcave functions is logconcave, $\pi_{1d}$ is more logconcave than Gaussian with variance $\alpha^{-1}\|a - b\|_2^{-2}$. Viewing $\pi_{1d}$ as a probability density of $\Pi_{1d}$ over $[0, 1]$, we apply Proposition C.1:

$$\Pi_{1d}(I_3) \geq (\log 2)\sqrt{\alpha}\,\|a - b\|_2\,\Pi_{1d}(I_1)\Pi_{1d}(I_2),$$

here $\Pi_{1d}(I_i)$ satisfies $\Pi_{1d}(I_i) = \frac{J_i}{J_4}$ for $i = 1, 2, 3$. Thus we have

$$(\log 2)\sqrt{\alpha}\,\|c_1 - c_2\|_2 \overset{(i)}{\leq} (\log 2)\sqrt{\alpha}\,\|a - b\|_2 \leq \frac{J_3 J_4}{J_1 J_2},$$

where inequality $(i)$ satisfies since $c_1, c_2$ are inside the line segment $\overline{ab}$. So both inequalities in Eq. (20) are proved, thus we proved that $d'(S_1, S_2) \cdot J_1 \cdot J_2 \leq J_3 \cdot J_4$.

Following the same combinatorial argument in Theorem 5.2 from Kannan et al. (1995), we can prove the Eq. (19) for general 1-dimensional measurable sets $I_1, I_2, I_3$, the details are omitted here. As a result of the localization lemma, we have proved the Eq. (18) in $\mathbb{R}^n$ and thus the isoperimetric inequality under $d'$.

■

### C.2. Extension to Weakly Logconcave Distributions

In this section, we prove Lemma B.4, which extends the new isoperimetry in Lemma B.3 to weakly logconcave distributions. Instead of strong logconcavity, we only assume that the distribution $\Pi$ is supported on $K \subseteq \mathbb{R}^n$ an open and convex set, and $\Pi$ has a bounded covariance matrix $\Sigma_\pi$.

Without loss of generality, we can further assume that $\Pi$ is an isotropic logconcave distribution. In other words, $\Sigma_\pi = I_n$. This is because one can apply an affine transformation to $\Pi$ so that $\Sigma_\pi = I_n$, and the effect of this transformation on the Euclidean and Hilbert metrics can be quantified. To be rigorous, we provide a proof in Appendix I.1.

To prove Lemma B.4, it is easier to first prove an isoperimetric inequality with boundary measures, and then use the co-area formula in a metric space to recover the isoperimetric inequality on three sets $K = S_1 \sqcup S_2 \sqcup S_3$. To do this, we first extend the notion of boundary measures to a more general metric than Euclidean metric, and we also introduce gradient modulus in a general metric space. These two definitions are listed in Definition C.4. We also list the co-area formula for a general metric space in Lemma C.5.

**Definition C.4 (boundary measure Bobkov and Houdré (1997))** *Let $(X, d)$ be a separable metric space, and let $\mathcal{B}(d)$ denotes the Borel $\sigma$-algebra induced by the topology of $(X, d)$. Assume $\Pi$ is a probability measure defined over $\mathcal{B}(d)$. If $B \in \mathcal{B}(d)$, then the boundary measure $\Pi^+$ of $B$ is defined to be*

$$\Pi^+(B) = \liminf_{h \to 0^+} \frac{\Pi(B^h) - \Pi(B)}{h},$$

*where $B^h := \{x \in X | \exists a \in B, d(x, a) < h\}$ is the open $h$-neighborhood of $A$. For any function $f : X \to \mathbb{R}$, we define the modulus of its gradient $|\nabla f(x)|$[3] to be:*

$$|\nabla f(x)| = \limsup_{d(x,y) \to 0^+} \frac{|f(x) - f(y)|}{d(x, y)}, \tag{21}$$

*and we set $|\nabla f(x)| = 0$ if $x$ is an isolated point in $X$*

**Lemma C.5 (co-area inequality Bobkov and Houdré (1997))** *Let $(X, d)$ be a separable metric space, and let $\Pi$ denotes a probability measure defined on $\mathcal{B}(d)$. Assume $\rho$ is a function on $X$ with a finite Lipschitz constant, then:*

$$\int_X |\nabla \rho(x)| \, \Pi(dx) \geq \int_{-\infty}^{+\infty} \Pi^+ \{x \in X | \rho(x) > t\} \, dt. \tag{22}$$

The proof of Lemma C.5 can be found in Bobkov and Houdré (1997). One issue to apply the lemma is that our target probability distribution $\Pi$ in Equation (1) is only defined over the Borel $\sigma$-algebra generated by the Euclidean topology, so to use boundary measure and co-area inequality for the new metric, we need to show $\Pi$ is also measurable over the Borel $\sigma$-algebra generated by our new metric $\mathfrak{d}$. Formally, we show that the open convex set $K$ equipped with the mixed metric $\mathfrak{d}(x, y) := \max \left\{ (\log 2) \|x - y\|_2, d_K^{\mathcal{H}}(x, y) \right\}$ satisfies the condition to be a metric space, and the topology induced by $\mathfrak{d}$ is the usual Euclidean topology, and we summarize this in the following Lemma C.6.

3. $|\nabla f(x)|$ is $\mathcal{B}(d)$-measurable if $f$ is continuous

**Lemma C.6** *Let $K$ denotes an open convex subset of $\mathbb{R}^n$, we define $\mathfrak{d}$ to be the following binary function over $K$:*

$$\mathfrak{d}(x, y) := \max \left\{ \log 2 \left\| x - y \right\|_2, d_K^{\mathcal{H}}(x, y) \right\}, \tag{23}$$

*then $(K, \mathfrak{d})$ is a metric space and $\mathfrak{d}$ induces the Euclidean topology over $K$.*

The proof of Lemma C.6 is left to Appendix I.2, where we check the definitions of metric spaces, and verify that the identity map is a homeomorphism between $(K, \mathfrak{d})$ and $(K, \|\cdot\|_2)$. Next, we prove a new isoperimetric inequality with boundary measures for isotropic logconcave distributions.

**Lemma C.7** *Assume $\Pi$ is an isotropic logconcave distribution supported on an open convex set $K \subseteq \mathbb{R}^n$. Then for any Borel measurable set $B \subseteq K$, we have*

$$\Pi^+(B) \geq \frac{1}{6 \max\{1, \psi_n\}} \Pi(B) \left[1 - \Pi(B)\right], \tag{24}$$

*where $\psi_n$ denotes the KLS constant, and the metric $\mathfrak{d}$ in defining boundary measure $\Pi^+$ is the mixed metric given by:*

$$\mathfrak{d}(x, y) := \max \left\{ (\log 2) \left\| x - y \right\|_2, d_K^{\mathcal{H}}(x, y) \right\}$$

It is worth mention that here the KLS constant $\psi_n$ is still under the usual definition by Euclidean metric as in Eq. (4). So we still have $\psi_n = O(\sqrt{\log n})$ by Klartag (2023). However, the boundary measure $\Pi^+$ in Lemma C.7 is defined using our mixed metric $\mathfrak{d}$.

The idea to prove Lemma C.7 is to use stochastic localization, where a family of random measures $(\Pi_t)_t$ is defined for $t \geq 0$ by stochastic differential equations. At time $t > 0$, the measure $\Pi_t$ is $t$-strongly logconcave, thus we can apply the mixed isoperimetry for strongly-logconcave distributions in Lemma B.3. Meanwhile, Taking $t \leq \psi_n^{-2}$ allows us to control the right-hand side via approximate conservation of variance in Lemma C.8.

**Stochastic localization** Now we introduce stochastic localization, following the notation in Chen (2021). Assume $\Pi$ is a logconcave distribution. For $t \geq 0$ and $\theta \in \mathbb{R}^n$, let $\pi_{t,\theta}$ denote the following probability density function

$$\pi_{t,\theta}(x) = \frac{1}{Z(t, \theta)} \exp\left(\theta^\top x - \frac{t}{2} x^\top x\right) \pi(x).$$

The mean of the probability density $\pi_{t,\theta}$ is denoted by

$$a(t, \theta) = \int_{\mathbb{R}^n} x \, \pi_{t,\theta}(x) dx$$

We consider the stochastic evolution of target distribution $\Pi$ by defining the following stochastic differential equation in terms of $\theta$

$$d\theta_t = dW_t + a(t, \theta_t) dt, \quad \theta_0 = 0, \tag{25}$$

where $(W_t)$ is standard Brownian motion. Then the distribution at time $t$ is defined as $\pi_t(x) := \pi_{t,\theta_t}(x)$. The existence and uniqueness of the solution in time $[0, t]$ for Equation (25) can be shown by verifying that $a(t, \theta)$ is a bounded function that is Lipschitz with respect to $t$ and $\theta$. Using Itô formula, we obtain that for $x \in \mathbb{R}^n$:

$$d\pi_t(x) = \pi_t(x) \langle x - a_t, dW_t \rangle .$$

Thus the process $(\pi_t(x))_t$ is a martingale with respect to the filtration induced by the Brownian motion, and we have for $t \geq 0$ and $x \in \mathbb{R}^n$:

$$\mathbb{E}\pi_t(x) = \pi_0(x) = \pi(x).$$

The stochastic localization technique developed in Eldan (2013) has been successful in upper bounding the KLS constant $\psi_n$. A series of works Lee and Vempala (2019); Chen (2021); Klartag and Lehec (2022); Klartag (2023) use stochastic localization to provide tighter upper bounds for the KLS constant $\psi_n$, with the state-of-the-art bound $\psi_n = O(\sqrt{\log n})$ obtained by Klartag (2023). While it has an elaborate proof, for the purpose of our result, we only need the approximate conservation of variance along the stochastic process in Lemma 3.1 in Klartag (2023) which we restate below.

**Lemma C.8 (Approximate conservation of variance, Lemma 3.1 in Klartag (2023))** *For any $t \geq 0$ and $f \in L^2(\Pi)$:*

$$\mathbb{E}\,\mathrm{Var}_{\pi_t}(f) \leq \mathrm{Var}_{\pi_0}(f) \leq \left(2 + \frac{t}{\lambda_0}\right) \mathbb{E}\,\mathrm{Var}_{\pi_t}(f),$$

*where $\lambda_0 = \frac{1}{C_P(\Pi)}$ is the reciprocal of the Poincaré constant of the distribution $\Pi$.*

The proof of Lemma C.8 can be found in Klartag (2023) and is omitted. Now we are ready to prove Lemma C.7.

**Proof** [Proof of Lemma C.7] Given $\Pi$ an isotropic logconcave distribution, then we define the following mixed metric $\mathfrak{d}$ by

$$\mathfrak{d}(x, y) = \max\left\{ d_K^{\mathcal{H}}(x, y), (\log 2) \left\|x - y\right\|_2 \right\}. \tag{26}$$

For any Borel set $B \subseteq K$, $\Pi^+(B), \Pi(B)$ are well-defined quantities according to the homeomorphism between $\mathfrak{d}$ and $\|\cdot\|_2$ (see Lemma C.6 and Definition C.4). We run stochastic localization as defined in Equation (25) until $t = \psi_n^{-2}$, and we have

$$\begin{aligned} \Pi^+(B) &= \liminf_{h \to 0+} \frac{\Pi(B^h) - \Pi(B)}{h} \overset{(i)}{=} \liminf_{h \to 0+} \frac{\mathbb{E}\left\{\Pi_t(B^h) - \Pi_t(B)\right\}}{h} \\ &\overset{(ii)}{\geq} \mathbb{E}\left\{ \liminf_{h \to 0+} \frac{\Pi_t(B^h) - \Pi_t(B)}{h} \right\} = \mathbb{E}\left\{\Pi_t^+(B)\right\}, \end{aligned} \tag{27}$$

where equality $(i)$ used the martingale property of $(\Pi_t)_t$, and inequality $(ii)$ applies Fatou's lemma. In order to lower bound $\Pi_t^+(B)$, we can apply Lemma B.3 to the distribution $\Pi_t$ since it is $t$-strongly logconcave. The distance $d'$ in Lemma B.3 can be lower bounded by $\frac{1}{\max\{1, \psi_n\}}\mathfrak{d}$ by definition in

Eq. (26). To be explicit, for any measurable decomposition $K := S_1 \sqcup S_2 \sqcup S_3$, we have

$$\begin{aligned}
\Pi(S_3) &\overset{(i)}{\geq} \inf_{x\in S_1, y\in S_2} \max\left\{d_K(x,y), (\log 2)\sqrt{t}\left\|x-y\right\|_2\right\}\Pi(S_1)\Pi(S_2)\\
&\overset{(ii)}{\geq} \inf_{x\in S_1, y\in S_2} \max\left\{d_K^{\mathcal{H}}(x,y), (\log 2)\sqrt{t}\left\|x-y\right\|_2\right\}\Pi(S_1)\Pi(S_2)\\
&\overset{(iii)}{\geq} \inf_{x\in S_1, y\in S_2} \frac{1}{\max\left\{1,\psi_n\right\}} \max\left\{d_K^{\mathcal{H}}(x,y), (\log 2)\left\|x-y\right\|_2\right\}\Pi(S_1)\Pi(S_2)\\
&= \frac{1}{\max\left\{1,\psi_n\right\}}\mathfrak{d}(S_1,S_2)\Pi(S_1)\Pi(S_2),
\end{aligned}$$

where inequality $(i)$ follows from Lemma B.3, inequality $(ii)$ follows from the fact $\log(1+x) \leq x$ for all $x > 0$, and inequality $(iii)$ holds since $1 \leq \max\left\{1,\psi_n\right\}$. Then the boundary version of the above isoperimetric inequality is easily implied if we take the limit $S_1 \to B$, $S_2 \to K \setminus B$ and $S_3 \to \partial B$. So at $t = \psi_n^{-2}$ we have

$$\Pi_t^+(B) \geq \frac{1}{\max\left\{1,\psi_n\right\}}\Pi_t(B)\left(1-\Pi_t(B)\right). \tag{28}$$

As a result, we can continue to bound Eq. (27) by

$$\begin{aligned}
\Pi^+(B) \geq \mathbb{E}\left\{\Pi_t^+(B)\right\} &\geq \frac{1}{\max\left\{1,\psi_n\right\}}\mathbb{E}\left\{\Pi_t(B)\left(1-\Pi_t(B)\right)\right\}\\
&= \frac{1}{\max\left\{1,\psi_n\right\}}\mathbb{E}\left\{\mathrm{Var}_{\pi_t}\left(\mathbf{1}_B\right)\right\}\\
&\overset{(i)}{\geq} \frac{1}{\max\left\{1,\psi_n\right\}}\frac{1}{\left[2+C_P(\Pi)\cdot\psi_n^{-2}\right]}\mathrm{Var}_\Pi(\mathbf{1}_B)\\
&\overset{(ii)}{\geq} \frac{1}{6\max\left\{1,\psi_n\right\}}\Pi(B)\left(1-\Pi(B)\right),
\end{aligned} \tag{29}$$

where inequality $(i)$ is an application of Lemma C.8 with $t = \psi_n^{-2}$, and inequality $(ii)$ follows from the fact that Poincaré constant and the isoperimetric constant are closely related due to Cheeger's inequality Cheeger (1970) and Buser and Ledoux Buser (1982); Ledoux (2004). Namely, for any probability measure $\mu$ on $\mathbb{R}^n$, $\frac{1}{4} \leq \psi_\mu^2/C_P(\mu) \leq \pi$. Then $C_P(\Pi) \leq 4\psi_n^2$ since $\Pi$ is isotropic and logconcave. ∎

With Lemma C.6 we know that $(K,\mathfrak{d})$ induces the usual Euclidean topology, and this indicates that the $\sigma$-algebra over which the probability measure $\Pi$ is defined is the same as the Borel $\sigma$-algebra induced by $\mathfrak{d}$-topology. Let $\mathcal{B}(K)$ denotes Borel $\sigma$-algebra induced by Euclidean topology over $K$, assume $\Pi$ is a probability measure supported on $K$.

Having proved the boundary version of the isoperimetric inequality (Lemma C.7), we attempt to prove the corresponding isoperimetric inequality under the decomposition $K = S_1 \sqcup S_2 \sqcup S_3$ as Lemma B.4. For reasons talked above, we only need to prove Lemma B.4 assuming that the distribution $\Pi$ is isotropic.

**Proof** [Proof of Lemma B.4] Based on the discussion in Appendix I.1, it suffices to consider isotropic logconcave $\Pi$. Let $\mathfrak{d}$ denote the following mixed metric

$$\mathfrak{d}(x,y) := \max\left\{(\log 2)\left\|x-y\right\|_2, d_K^{\mathcal{H}}(x,y)\right\}. \tag{30}$$

If $\mathfrak{d}(S_1, S_2) = 0$, then the isoperimetric inequality is trivial. We can assume that $\mathfrak{d}(S_1, S_2) > 0$. Let $\rho$ be defined as

$$\rho(x) := \frac{\mathfrak{d}(x, S_1)}{\mathfrak{d}(x, S_2) + \mathfrak{d}(x, S_1)},$$

where we define $\mathfrak{d}(x, S) := \inf_{z \in S} \mathfrak{d}(x, z)$. We bound $|\nabla\rho(x)|$, the modulus of the gradient of $\rho$, as follows

$$\begin{aligned}
\frac{|\rho(x) - \rho(y)|}{\mathfrak{d}(x, y)} &= \frac{1}{\mathfrak{d}(x, y)} \frac{|\mathfrak{d}(y, S_2)\,[\mathfrak{d}(x, S_1) - \mathfrak{d}(y, S_1)] + \mathfrak{d}(y, S_1)\,[\mathfrak{d}(y, S_2) - \mathfrak{d}(x, S_2)]|}{[\mathfrak{d}(x, S_2) + \mathfrak{d}(x, S_1)]\,[\mathfrak{d}(y, S_2) + \mathfrak{d}(y, S_1)]} \\
&\overset{(i)}{\leq} \frac{1}{\mathfrak{d}(x, y)} \cdot \frac{\mathfrak{d}(y, S_2)\mathfrak{d}(x, y) + \mathfrak{d}(y, S_1)\mathfrak{d}(x, y)}{[\mathfrak{d}(x, S_2) + \mathfrak{d}(x, S_1)]\,[\mathfrak{d}(y, S_2) + \mathfrak{d}(y, S_1)]} \overset{(ii)}{\leq} \frac{1}{\mathfrak{d}(S_1, S_2)}.
\end{aligned}$$

Inequality $(i)$ follows from the fact that $|d(x, S) - d(y, S)| \leq d(x, y)$ for any set $S$, which is a consequence of the triangle inequality $\inf_{z \in S} d(x, z) \leq \inf_{z \in S}\,[d(x, y) + d(y, z)]$. Inequality $(ii)$ follows from $\mathfrak{d}(S_1, S_2) \leq \mathfrak{d}(x, S_1) + \mathfrak{d}(x, S_2)$. As a result, the modulus of the gradient of $\rho$ is bounded uniformly as follows

$$|\nabla\rho(x)| := \limsup_{d(x,y) \to 0^+} \frac{|\rho(x) - \rho(y)|}{d(x, y)} \leq \frac{1}{\mathfrak{d}(S_1, S_2)}, \forall x \in K. \tag{31}$$

Following an argument in the proof of Theorem 2.6 in Lovász and Simonovits (1993), we may assume that $S_1$ and $S_2$ are open without loss of generality. This avoids the necessity to deal to boundaries of $S_1$ and $S_2$. Applying the co-area formula in Lemma C.5 over $(K, \mathfrak{d}, \Pi)$, we obtain:

$$\begin{aligned}
\Pi(S_3) &\overset{(i)}{\geq} \mathfrak{d}(S_1, S_2) \int_{S_3} |\nabla\rho(x)|\,\Pi(dx) \\
&\overset{(ii)}{=} \mathfrak{d}(S_1, S_2) \int_{K} |\nabla\rho(x)|\,\Pi(dx) \\
&\overset{(iii)}{\geq} \mathfrak{d}(S_1, S_2) \int_0^1 \Pi^+ \{x \in X | \rho(x) > t\}\,dt \\
&\overset{(iv)}{\geq} \mathfrak{d}(S_1, S_2) \int_0^1 \frac{1}{6 \max\{1, \psi_n\}} \Pi\{\rho(x) > t\} \cdot \Pi\{\rho(x) \leq t\}\,dt \\
&\overset{(v)}{\geq} \frac{\mathfrak{d}(S_1, S_2)}{6 \max\{1, \psi_n\}} \Pi(S_1)\Pi(S_2).
\end{aligned}$$

Inequality $(i)$ follows from Eq. (31). Equality $(ii)$ follows from $|\nabla\rho(x)| = 0$ on open $S_1, S_2$. Inequality $(iii)$ applies co-area formula and the fact that $\Pi^+(B)$ is 0 for $B = \varnothing, K$. Inequality $(iv)$ applies the isoperimetric inequality with boundary measure in Lemma C.7. Inequality $(v)$ holds since for any $t \in (0, 1)$, we have

$$S_1 \subseteq \{x \in K | \rho(x) \leq t\} \quad \text{and} \quad S_2 \subseteq \{x \in K | \rho(x) > t\}.$$

■

## Appendix D. Proofs of Mixing Time Upper Bounds

In this section, we prove our upper bounds of mixing times of soft-threshold Dikin walk. In Section D.1, we focus on the soft-threshold Dikin walk applied to $\alpha$-strongly logconcave and $\beta$-log-smooth distributions truncated on $K$, and we prove Lemma A.4 with its associated results: Theorem 3.1, Corollary 3.2 and Theorem3.3. In Section D.2, we remove the assumption of $\alpha$-strong logconcaveness, and prove Lemma A.5 with its associated results: Theorems 3.4 and 3.5. Finally, in Section D.3, we go beyond worst-case analysis and prove Theorem 3.7.

### D.1. Mixing for Strongly Logconcave Distributions

We first prove Lemma A.4, where we lower bound the conductance and apply Lemma B.1 to obtain the desired mixing time upper bound. In the proof we use the transition overlap (Lemma B.2) and apply our combined isoperimetry on $\alpha$-strongly logconcave distributions.

**Proof** [proof of Lemma A.4] Set the step-size $r$ indicated in Lemma B.2, we try to bound the conductance of the Markov chain: For any partition $K = A_1 \sqcup A_2$, we prove the following inequality

$$\int_{A_1} \mathcal{T}_u(A_2)\pi(u)du \geq \frac{C}{\sqrt{n\left(C_K + \frac{C_E}{\alpha}\right)}} \min\left\{\Pi(A_1), \Pi(A_2)\right\}, \tag{32}$$

for some absolute constant $C$. To prove Eq. (32). We define two bad conductance subsets of $A_1, A_2$ to be:

$$A_1' := \left\{u \in A_1 \middle| \mathcal{T}_u(A_2) < \frac{1}{10}\right\}, \quad A_2' := \left\{u \in A_2 \middle| \mathcal{T}_u(A_1) < \frac{1}{10}\right\}.$$

We divide the proof into two scenarios. If $\Pi(A_1') \leq \frac{1}{2}\Pi(A_1)$ or $\Pi(A_2') \leq \frac{1}{2}\Pi(A_2)$. Since $\Pi$ is the stationary distribution of $\mathcal{T}$, we have $\int_{A_1} \mathcal{T}_x(A_2)\pi(x)dx = \int_{A_2} \mathcal{T}_x(A_1)\pi(x)dx$. So we may assume $\Pi(A_1') \leq \frac{1}{2}\Pi(A_1)$ without loss of generality, and we have

$$\int_{A_1} \mathcal{T}_x(A_2)\pi(x)dx \geq \int_{A_1\setminus A_1'} \mathcal{T}_x(A_2)\pi(x)dx \geq \frac{1}{10}\Pi(A_1 \setminus A_1') \geq \frac{1}{20}\min\left\{\Pi(A_1), \Pi(A_2)\right\}.$$

Thus Eq. (32) is verified in this scenario.

Now we deal with the case $\Pi(A_1') > \frac{1}{2}\Pi(A_1)$ and $\Pi(A_2') > \frac{1}{2}\Pi(A_2)$. For any $u \in A_1'$, $v \in A_2'$, by definition of $A_1'$ and $A_2'$, we have

$$\begin{aligned} \|\mathcal{T}_u - \mathcal{T}_v\|_{TV} &\geq \mathcal{T}_u(A_1) - \mathcal{T}_v(A_1) = 1 - \mathcal{T}_u(A_2) - \mathcal{T}_v(A_1) \\ &\geq 1 - \frac{1}{10} - \frac{1}{10} > \frac{4}{5}. \end{aligned}$$

Thus, by Lemma B.2 and the symmetry between $x, y$, we have

$$\begin{aligned} &\|x-y\|_{G(x)} > \frac{r}{10\sqrt{n}}, \quad \|x-y\|_{G(y)} > \frac{r}{10\sqrt{n}} \\ \Rightarrow &\frac{r^2}{100n} < \min\left\{\|x-y\|^2_{G(x)}, \|x-y\|^2_{G(y)}\right\} \\ &\frac{r^2}{100n} \overset{(i)}{<} C_K d_K(x,y)^2 + C_E\|x-y\|_2^2 \overset{(ii)}{\leq} \left(C_K + \frac{C_E}{(\log 2)^2\alpha}\right) d'(x,y)^2, \end{aligned}$$

where inequality $(i)$ holds due to our assumption on $G$, and inequality $(ii)$ holds since we define the new metric $d'(u,v) := \max\{d_K(u,v), \log(2)\sqrt{\alpha}\|u-v\|_2\}$. We also have:

$$\begin{aligned}\int_{A_1} \mathcal{T}_u(A_2)\pi(u)du &= \frac{1}{2}\left[\int_{A_1} \mathcal{T}_u(A_2)\pi(u)du + \int_{A_2} \mathcal{T}_u(A_1)\pi(u)du\right] \\ &\geq \frac{1}{2}\left[\int_{A_1\setminus A_1'} \mathcal{T}_u(A_2)\pi(u)du + \int_{A_2\setminus A_2'} \mathcal{T}_u(A_1)\pi(u)du\right] \\ &\geq \frac{1}{20}\Pi(K\setminus(A_1'\cup A_2')).\end{aligned} \tag{33}$$

Since we assumed that $f(x)$ is $\alpha$-strongly convex, $f(x) - \frac{\alpha}{2}\|x-x^\star\|_2^2$ is a convex function where $x^\star$ is any point in $K$. As a result, for $x \in K$, the probability density function of $\Pi$ can be written as: $\pi(x) \propto \exp\left(-\frac{\alpha}{2}\|x-x^\star\|_2^2\right)\phi(x)$, where $\phi(x) = \exp\left(-f(x) + \frac{\alpha}{2}\|x-x^\star\|_2^2\right)$ is logconcave. Thus we can apply the isoperimetric inequality in Lemma B.3, and we have

$$\Pi(K\setminus(A_1'\cap A_2')) \geq d'(A_1', A_2')\Pi(A_1')\Pi(A_2'). \tag{34}$$

Now insert Eq. (34) into Eq. (33), applying the lower bound of $d'(A_1', A_2')$, we deduce that:

$$\begin{aligned}\int_{A_1} \mathcal{T}_u(A_2)\pi(u)du &\geq \frac{1}{20}d'(A_1', A_2')\Pi(A_1')\Pi(A_2') \\ &\overset{(i)}{\geq} \frac{r}{200\times\sqrt{3n(C_K + \frac{C_E}{\alpha})}}\cdot\frac{1}{2}\Pi(A_1)\cdot\frac{1}{2}\Pi(A_2) \\ &\overset{(ii)}{\geq} \frac{r}{800\times\sqrt{3n(C_K + \frac{C_E}{\alpha})}}\cdot\frac{1}{2}\min\{\Pi(A_1), \Pi(A_2)\},\end{aligned}$$

where inequality $(i)$ due to the assumption $\Pi(A_i') \geq \frac{1}{2}\Pi(A_i)$ for $i = 1, 2$. Inequality $(ii)$ holds because $\Pi(A_1) + \Pi(A_2) = 1$.

Combining the two situations, we proved that the conductance for the Markov chain $\mathcal{T}$ satisfies $\Phi \geq c\sqrt{C_K + \frac{C_E}{\alpha}}$ for some absolute constant $c > 0$ since the step-size $r$ is a fixed constant. In order to use Lemma B.1, we need to use the conductance $\overline{\Phi}$ for the Lazy version $\overline{\mathcal{T}}$, and this at most shrinks the conductance by a constant factor of 2.

According to Lemma B.1, there exists an absolute constant $C > 0$ such that for all error tolerance $\epsilon > 0$ and $M$-warm initial distribution $\mu_0$, the mixing time can be bounded as:

$$T_{\text{mix}}(\epsilon, \mu_0) \leq \frac{2}{\overline{\Phi}^2}\log\left(\frac{\sqrt{M}}{\epsilon}\right) \leq \frac{8}{\Phi^2}\log\left(\frac{\sqrt{M}}{\epsilon}\right) \leq C\cdot n\left(C_K + \frac{C_E}{\alpha}\right)\log\left(\frac{\sqrt{M}}{\epsilon}\right).$$

■

Next, we derive its corollaries. For Corollary 3.1 and 3.2, we notice that Corollary 3.2 is directly implied by Corollary 3.1 by letting $m = o(n)$. In order to prove Corollary 3.1, we need to show that the soft-threshold metric $G$ in Definition 1.3 setting the regularization size $\lambda := \beta$ satisfies all the conditions in Lemma A.4. We divide these properties into the following two lemmas, and Corollary 3.1 is directly implied by the following two lemmas.

**Lemma D.1** *Given a (possibly unbounded) polytope $K = \{x|Ax > b\}$ for $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, the soft-threshold metric $G$ in Definition 1.3 is SSC, ASC and LTSC.*

**Lemma D.2** *Given a (possibly unbounded) polytope $K = \{x|Ax > b\}$ for $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, the soft-threshold metric $G$ in Definition 1.3 satisfies: for all $x, y \in K$*

$$\min\left\{\|y-x\|_{G(x)}, \|y-x\|_{G(y)}\right\} \le m d_K(x,y)^2 + \beta\|x-y\|_2^2. \tag{35}$$

*Moreover, we have $E(x, G(x), 1) \subseteq K$.*

The proof of Lemma D.1 is left to Appendix F. If the polytope $K$ is bounded, these properties are already implied in Kook and Vempala (2024) since the unregularized logarithmic metric $H(x) := A_x A_x^\top$ for $K$ is an invertible local metric. Kook and Vempala (2024) proved that $H(x)$ is SSC, SASC and SLTSC, thus $G(x) := H(x) + \lambda I$ is SSC, ASC, LTSC by the additivity of these properties. However, we emphasize that in this paper, we do not require the polytope $K$ to be bounded, and we even allow $m < n$, thus $H(x) = A_x A_x^\top$ may not be invertible, and those self-concordance properties are not well-defined for non-invertible matrices, thus we can not use the additivity to prove those self-concordance properties for $G(x) = H(x) + \lambda I$.

To be rigorous, we provide the proof of SSC, LTSC, and ASC for the soft-threshold metric $G$ defined in Definition 1.3 in Appendix F. The SSC and LTSC of soft-threshold metric is proved by a limit argument: we add artificial constraints so that $H(x) := A_x A_x^\top$ is invertible and let these constraints vanish. We proved ASC of soft-threshold metric by concentration of Gaussian polynomials, which also appears in Sachdeva and Vishnoi (2016), and the intuition is that adding a regularization term $\lambda I$ only makes the Gaussian concentration tighter.

Now we prove Lemma D.2, we introduce the concept of $\bar{\nu}$-symmetry from Laddha et al. (2020) for convenience. The reason we need Definition D.3 is that we can upper bound the $\bar{\nu}$-symmetric local metric $H$ by $\sqrt{\bar{\nu}} d_K$, and this important property is summarized as Proposition D.4.

**Definition D.3 (from Laddha et al. (2020))** *For any convex set $K \subseteq \mathbb{R}^n$, we define a PSD matrix function $H : K \to \mathbb{S}_+^n$, and let $E(x, H(x), r)$ denote the following ellipsoid*

$$E(x, H(x), r) := \{z|(z-x)^\top H(x)(z-x) \le r^2\}.$$

*We define $H$ to be $\bar{\nu}$-symmetric if for any $x \in K$, we have*

$$E(x, H(x), 1) \subseteq K \cap (2x - K) \subseteq E(x, H(x), \sqrt{\bar{\nu}})$$

**Proposition D.4 (revised from Lemma 2.3 in Laddha et al. (2020))** *We use $d_K(x,y)$ to denote the extended cross-ratio distance (as defined in Definition A.2) between $x$ and $y$ in the convex set $K$, assume $H$ is a $\bar{\nu}-$symmetric local metric, then we have*

$$d_K(x,y) \ge \frac{1}{\sqrt{\bar{\nu}}}\min\left\{\|x-y\|_{H(x)}, \|x-y\|_{H(y)}\right\}. \tag{36}$$

The proof of Proposition D.4 is postponed to Appendix H, and it differs from the proof in Laddha et al. (2020) in two important aspects: First, since we defined the cross-ratio distance in the extended sense, we no longer require $K$ to be a compact convex set, and we divide the proof into several cases. Second, we notice some computational problems in the symmetry argument of Laddha et al. (2020) and cannot recover the inequality with only $\|\cdot\|_{H(x)}$ on the RHS of Eq. (36), so we provide an weaker inequality with minimum of $\|\cdot\|_H$ computed at both $x$ and $y$. For the sake of rigor and completeness, we include the proof in Appendix H.

Now we are ready to give the proof of Lemma D.2 using the $\bar{\nu}$-symmetry and Proposition D.4.

**Proof** [proof of Lemma D.2] We first prove that the unregularized logarithmic metric $H(x) := A_x A_x^\top$ is $\bar{\nu}$-symmetric with $\bar{\nu} = m$. For all $z$ such that $z \in K \cap (2x - K)$, this is equivalent to for all $i \in [m]$, we have

$$a_i^\top z - b_i > 0, \quad a_i^\top (2x - z) - b_i > 0,$$

and this translates to $\max_{i\in[m]} \left| s_{x,i}^{-1} \left(a_i^\top (z - x)\right)\right| \le 1$. Then we have the desired bound:

$$\|z - x\|_{H(x)}^2 \le m \left\|S_x^{-1} A(z - x)\right\|_\infty^2 = m \cdot \max_{i\in[m]} \left| s_{x,i}^{-1} \left(a_i^\top (z - x)\right)\right| \le m,$$

On the other hand, if we assume $z \in E(x, H(x), 1)$, then we have

$$\max_{i\in[m]} \left| s_{x,i}^{-1} \left(a_i^\top (z - x)\right)\right| = \left\|S_x^{-1} A(z - x)\right\|_\infty \le \left\|S_x^{-1} A(z - x)\right\|_2 \le 1.$$

Thus, for all $i \in [m]$, $a_i^\top z - b_i > 0$ and $a_i^\top (2x - z) - b_i > 0$. We have $z \in K \cap (2x - K)$. In conclusion, $H$ is $\bar{\nu}$-symmetric with $\bar{\nu} = m$. As a result of Proposition D.4, for any $x, y \in K$, we have

$$\min\left\{\|y - x\|_{H(x)}^2, \|y - x\|_{H(y)}^2\right\} \le m d_K(x, y)^2.$$

By definition of $G$ and setting $\lambda := \beta$, we have

$$\begin{aligned}\min\left\{\|y - x\|_{G(x)}^2, \|y - x\|_{G(y)}^2\right\} &= \min\left\{\|y - x\|_{H(x)}^2, \|y - x\|_{H(y)}^2\right\} + \beta \|x - y\|_2^2 \\ &\le m \cdot d_K(x, y)^2 + \beta \|x - y\|_2^2.\end{aligned}$$

Moreover, since we proved $H$ is $\bar{\nu}$-symmetric, so we have

$$E(x, G(x), 1) \subseteq E(x, H(x), 1) \subseteq K \cap (2x - K) \subseteq K.$$

■

The remaining task in this section is to prove Corollary 3.3. Similarly, we need to verify the conditions in Lemma A.4 for regularized Lewis metric $G$ in Definition 1.4. These properties are nicely implied in Kook and Vempala (2024). Since the unregularized Lewis metric $H(x) := c_1\sqrt{n}(\log m)^{c_2} A_x^\top W_x A_x$ with Lewis weights $w_x$ defined in Eq. (3). We list the properties here in Proposition D.5.

**Proposition D.5 (Lemma E.5, E.7 and E.12 Kook and Vempala (2024))** *There exists positive constants $c_1$ and $c_2$ such that the unregularized Lewis-weights metric $H(x) := \sqrt{n}c_1(\log m)^{c_2} A_x^\top W_x A_x$ is SSC, SLTSC, SASC and $\bar{\nu}$-symmetric with $\bar{\nu} = O((\log m)^{c_2} n^{3/2})$.*

**Proposition D.6 (Lemma D.12 and D.14 Kook and Vempala (2024))** *Given PSD matrix functions $G_i$ on $K_i$ for $i = 1, \ldots, l$, let $G := \sum_i G_i$ be PD on $\cap_i K_i$. We have*

- *if $G_i$ is SLTSC on $K_i$, then $G$ is LTSC on $\cap_i K_i$;*
- *if $l = O(1)$ and $g_i$ is SASC on $K_i$, then $G$ is ASC on $\cap_i K_i$.*

With Proposition D.5 and D.6, we provide the proof of Corollary 3.3 by verifying all the conditions in Lemma A.4 for regularized Lewis metric $G$ with regularization size $\lambda := \beta$ in Definition 1.4.

**Proof** [proof of Corollary 3.3] Define $H(x) := c_1\sqrt{n}(\log m)^{c_2} A_x^\top W_x A_x$ as in Proposition D.5, then the regularized Lewis metric $G$ satisfies $G(x) := H(x) + \beta I$. We first show that $G(x) := H(x) + \beta I$ is SSC, where $H(x)$ is the unregularized Lewis metric. For any direction $h \in \mathbb{R}^n$, we have

$$\begin{aligned}\left\|G(x)^{-\frac{1}{2}}\mathcal{D}G(x)[h]G(x)^{-\frac{1}{2}}\right\|_F &= \left\|G(x)^{-\frac{1}{2}}\mathcal{D}H(x)[h]G(x)^{-\frac{1}{2}}\right\|_F \\ &\overset{(i)}{\leq} \left\|H(x)^{-\frac{1}{2}}\mathcal{D}H(x)[h]H(x)^{-\frac{1}{2}}\right\|_F \overset{(ii)}{\leq} 2\left\|h\right\|_{H(x)} \leq 2\left\|h\right\|_{G(x)},\end{aligned}$$

where inequality $(i)$ holds since $G(x) \succeq H(x)$ (see more details in Lemma F.1), and inequality $(ii)$ follows from Proposition D.5 that $H$ is SSC. Moreover, since $\beta I$ is a constant matrix function defined over $\mathbb{R}^n$, thus $\beta I$ is trivially SASC and SLTSC by definition. By the additivity in Proposition D.6 and $H$ is SASC and SLTSC in Proposition D.5, thus $G(x) := H(x) + \beta I$ is ASC and LTSC.

Next we prove the metric inequality, it is obvious that $G(x) \succeq \beta I$. Since $H$ is $\bar{\nu}$-symmetric with $\bar{\nu} = \widetilde{O}(n^{3/2})$. Thus, by Proposition D.4, we have the following metric inequality

$$\begin{aligned}\min\left\{\left\|y-x\right\|_{G(x)}^2, \left\|y-x\right\|_{G(y)}^2\right\} &= \min\left\{\left\|y-x\right\|_{H(x)}^2, \left\|y-x\right\|_{H(y)}^2\right\}\beta\left\|y-x\right\|_2^2 \\ &\leq C\cdot(\log m)^{c_2}\cdot n^{3/2} d_K(x,y)^2 + \beta\left\|y-x\right\|_2^2,\end{aligned}$$

where $C > 0$ is some absolute constant. Moreover, since $H$ is $\bar{\nu}$-symmetric and $G(x) \succeq H(x)$, we have $E(x, G(x), 1) \subseteq E(x, H(x), 1) \subseteq K$. Thus Corollary 3.3 is proved by inserting $C_E = \beta$ and $C_K = (\log m)^{c_2} n^{3/2}$. ■

### D.2. Extension to Weakly Logconcave Distributions

In this section, we prove Lemma A.5 and its corresponding Corollaries 3.4 and 3.5. Actually, the only task in Section D.2 is to prove Lemma A.5. This is because the required properties of the local metric $G$ in Lemma A.5 are the same as in Lemma A.4. Those properties of both soft-threshold metric (Definition 1.3) and regularized Lewis metric (Definition 1.4) are already proved in Section D.1. Thus Corollaries 3.4 and 3.5 can be derived directly from Lemma A.5. The proof follows similar conductance arguments from the proof of Lemma A.4 in Section D.1, except that we use the isoperimetric inequality over weakly logconcave measures.

**Proof** [proof of Lemma A.5] Set the step-size $r$ indicated in Lemma B.2, we try to bound the conductance of the Markov chain: For any partition $K = A_1 \sqcup A_2$, we will prove the following inequality:

$$\int_{A_1} \mathcal{T}_u(A_2)\pi(u)du \geq \frac{C}{\sqrt{n\left(C_K + \eta C_E\right)}}\min\left\{\Pi(A_1), \Pi(A_2)\right\}, \tag{37}$$

where $C$ is some absolute constant. Let $K = A_1 \sqcup A_2$ be any measurable partition, and we define two bad conductance subsets of $A_1, A_2$ to be:

$$A_1' := \left\{ x \in A_1 \middle| \mathcal{T}_x(A_2) < \frac{1}{10} \right\}, \quad A_2' := \left\{ x \in A_2 \middle| \mathcal{T}_x(A_1) < \frac{1}{10} \right\}.$$

We can assume $\Pi(A_1') > \frac{1}{2}\Pi(A_1)$ and $\Pi(A_2') > \frac{1}{2}\Pi(A_2)$, otherwise Eq. (37) can be proved the same way as in the proof of Lemma A.4. Then for any $x \in A_1'$ and $y \in A_2'$, by definition of $A_1', A_2'$ we have

$$\|\mathcal{T}_x - \mathcal{T}_y\|_{TV} \geq \mathcal{T}_x(A_1) - \mathcal{T}_y(A_1) = 1 - \mathcal{T}_x(A_2) - \mathcal{T}_y(A_1) \geq 1 - \frac{1}{10} - \frac{1}{10} > \frac{4}{5}.$$

So by Lemma B.2 and the symmetry between $x$ and $y$ we have

$$\|y - x\|_{G(x)} > \frac{r}{10\sqrt{n}}, \quad \|y - x\|_{G(y)} > \frac{r}{10\sqrt{n}},$$

thus according to our assumption of $G$, we have

$$\begin{aligned} \frac{r^2}{100n} &< C_K d_K(x,y)^2 + C_E \|y - x\|_2^2 \\ &= C_K \left[ e^{d_K^{\mathcal{H}}(x,y)} - 1 \right]^2 + C_E \|x - y\|_2^2 \\ &\leq C_K \left[ e^{\mathfrak{d}(x,y)} - 1 \right]^2 + C_E \frac{\eta}{(\log 2)^2} \mathfrak{d}(x,y)^2. \end{aligned} \tag{38}$$

Recall that we use the Hilbert metric $d_K^{\mathcal{H}}$ instead of the cross-ratio distance $d_K$ to apply the mixed isoperimetric inequality for weakly logconcave distributions (Lemma B.4). The last inequality holds since we define $\mathfrak{d}(x,y) = \max\left\{ \frac{\log 2}{\sqrt{\eta}} \|y - x\|_2, d_K^{\mathcal{H}}(x,y) \right\}$. Consider the case that $\mathfrak{d}(x,y) \leq 1$, then according to the fact $e^x \leq 1 + 2x$ for $x \in [0,1]$, we have

$$\frac{r^2}{100n} < 4C_K \mathfrak{d}(x,y)^2 + C_E \frac{\eta}{(\log 2)^2} \mathfrak{d}(x,y)^2. \tag{39}$$

As a result, for all $x \in A_1'$ and $y \in A_2'$ we have

$$\mathfrak{d}(x,y) \geq \min\left\{ 1, \frac{r}{20\sqrt{n\,(C_K + C_E\eta)}} \right\} \geq \frac{r}{20\sqrt{n\,(C_K + C_E\eta)}},$$

where the last inequality holds since we assumed $C_K \geq 1$, $n \geq 1$ and we can always set the step-size $r \leq 1$. Now we are ready to control the ergodic flow from $A_1$ to $A_2$, using the same argument in the proof of Lemma A.4 (Eq. (33)) we have

$$\int_{A_1} \mathcal{T}_x(A_2)\pi(x)dx \geq \frac{1}{20}\Pi(K \setminus (A_1' \cup A_2')). \tag{40}$$

Now we apply Lemma B.4 over the target distribution $\Pi$, and we have

$$\Pi(K \setminus (A_1' \cup A_2')) \geq \frac{1}{6\max\{1, \psi_n\}} \mathfrak{d}(A_1', A_2')\Pi(A_1')\Pi(A_2') \tag{41}$$

Now insert Eq. (41) into Eq. (40), and we have

$$\begin{aligned}\int_{A_1} \mathcal{T}_x(A_2)\pi(x)dx &\overset{(i)}{\geq} \frac{\Pi(A_1')\Pi(A_2')}{120\max\{1,\psi_n\}} \cdot \frac{r}{20\sqrt{n\left(C_K+C_E\eta\right)}} \\ &\overset{(ii)}{\geq} \frac{\min\{\Pi(A_1),\Pi(A_2)\}}{8\cdot 120\max\{1,\psi_n\}} \cdot \frac{r}{20\sqrt{n\left(C_K+C_E\eta\right)}},\end{aligned}$$

where inequality $(i)$ follows by the lower bound for $\mathfrak{d}(A_1', A_2')$ in Equation (39), and inequality $(ii)$ holds because we assumed $\Pi(A_i') \geq \frac{1}{2}\Pi(A_i)$ for $i = 1, 2$ and $\Pi(A_1) + \Pi(A_2) = 1$.

Combining the situations, we have proved an lower bound on the conductance $\Phi$ of the transition kernel $\mathcal{T}$. There exists a universal constant $c$, and since the step-size $r$ is also an universal constant, $n \geq 1$ and $C_K \geq 1$ , we have

$$\Phi \geq \frac{c}{\max\{1,\psi_n\}} \frac{1}{\sqrt{n\left(C_K+C_E\eta\right)}} \geq \frac{c}{\psi_n\sqrt{n(C_K+C_E\eta)}}, \tag{42}$$

where the last inequality holds since $\psi_n = \Omega(1)$. Since the conductance $\overline{\Phi}$ for the lazified transition kernel $\overline{\mathcal{T}}$ satisfies $\overline{\Phi} \geq \frac{1}{2}\Phi$. According to Lemma B.1, there exists a universal constant $C > 0$, for any error tolerance $\epsilon > 0$ and $M$-warm initial distribution $\mu_0$, the mixing time satisfies:

$$T_{\text{mix}}(\epsilon;\mu_0) \leq \frac{8}{\Phi^2}\log\left(\frac{\sqrt{M}}{\epsilon}\right) \leq C\psi_n^2\cdot n\left(C_K+C_E\eta\right)\log\left(\frac{\sqrt{M}}{\epsilon}\right).$$

■

### D.3. Violated Constraints in a High Probability Ball

In Section D.3, we prove Theorem 3.7. The idea is similar to prove the conductance lower bound for the Markov chain $\mathcal{T}$ in Lemma A.4. The difference is that we needed to cut off the corner of $K$ that is too close to the remaining $(m - \mathcal{M}_R^\delta)$ linear constraints. Since this corner has a low probability mass, it does not impact the mixing time much. We need the concept of $s$-conductance $\Phi_s$(see Eq. (12)) to enable the cutoff.

We first give the statement and the proof of Lemma D.7, which controls the ergodic flows from $A_1$ to $A_2$.

**Lemma D.7** *Assume our target distribution $\Pi$ with density $\pi(x) \propto \mathbf{1}_K(x)e^{-f(x)}$ with twice differentiable $f$ to be both $\alpha$-convex and $\beta$-smooth as in Eq. (1). Let $\mathcal{T}$ denotes the transition kernel in Eq. (9) defined using the soft-threshold metric in Definition 1.3. Then there exists a step-size $r > 0$ such that for any partition $K = A_1 \sqcup A_2$, for any $R > 0$ and $\delta > 0$, we have*

$$\int_{A_1} \mathcal{T}_u(A_2)\pi(u)du \geq \frac{r}{32000\sqrt{\left[\kappa+\frac{m}{n\delta^2}+\mathcal{M}_R^\delta\right]n}} \min\left\{\Pi(A_1\cap\mathcal{B}_R),\Pi(A_2\cap\mathcal{B}_R)\right\}, \tag{43}$$

*where the definitions of $\mathcal{B}_R$, and $\mathcal{M}_R^\delta$ can be found in Section 3.3.*

**Proof** For any $R > 0$ and $\delta > 0$, recall the definition of $\mathcal{M}_R^\delta$, we denote $[\mathcal{M}_R^\delta]$ to be the following subset of $[m] := \{1, 2, \ldots, m\}$,

$$[\mathcal{M}_R^\delta] := \left\{ i | \, \exists \, x \in \mathcal{B}_R^\delta \text{ such that } a_i^\top x - b_i \le 0 \right\}.$$

The distance of any point $x$ to the $i$-th face $\left\{ z | a_i^\top z - b_i = 0 \right\}$ equals $\frac{|a_i^\top x - b_i|}{\|a_i\|_2}$, thus for all $x \in \mathcal{B}_R$, we have

$$\frac{a_i a_i^\top}{(a_i^\top x - b_i)^2} \preceq \frac{\|a_i\|_2^2}{(a_i^\top x - b_i)^2} I_n \overset{\text{(I)}}{\preceq} \frac{\alpha}{n\delta^2} I_n \text{ for } i \in [m] \setminus [\mathcal{M}_R^\delta],$$

where inequality (I) holds since for any $x \in \mathcal{B}_R$, the distance of $x$ to any face $i \in [m] \setminus [\mathcal{M}_R^\delta]$ is greater than $\delta\sqrt{\frac{n}{\alpha}}$. Otherwise, there exists $z$ such that $\|z - x\|_2 \le \delta\sqrt{\frac{n}{\alpha}}$ satisfying $a_i^\top z - b_i \le 0$ for some $i \in [m] \setminus [\mathcal{M}_R^\delta]$, which implies $z \notin \mathcal{B}_R^\delta$. Meanwhile, we also have $z \in \mathcal{B}_R^\delta$ because $\|z - x^\star\|_2 \le \|x - x^\star\|_2 + \|x - z\|_2 \le (R + \delta)\sqrt{\frac{n}{\alpha}}$. This is a contradiction.

Now we can control the soft-threshold metric $G(x)$ for all $x \in K \cap \mathcal{B}_R$

$$\begin{aligned} G(x) &= \beta I + \sum_{i \in [\mathcal{M}_R^\delta]} \frac{a_i a_i^\top}{(a_i^\top x - b_i)^2} + \sum_{i \in [m] \setminus [\mathcal{M}_R^\delta]} \frac{a_i a_i^\top}{(a_i^\top x - b_i)^2} \\ &\preceq \left[ \beta + \frac{\alpha(m - \mathcal{M}_R^\delta)}{n\delta^2} \right] I + \sum_{i \in [\mathcal{M}_R^\delta]} \frac{a_i a_i^\top}{(a_i^\top x - b_i)^2}. \end{aligned}$$

Then we can control the $\|y - x\|_{G(x)}$ for all $x, y \in \mathcal{B}_R \cap K$:

$$\begin{aligned} \|y - x\|_{G(x)}^2 &\overset{\text{(i)}}{\le} \left[ \beta + \frac{\alpha(m - \mathcal{M}_R^\delta)}{n\delta^2} \right] \|x - y\|_2^2 + \mathcal{M}_R^\delta d_K^2(x, y) \\ &\le \left[ \mathcal{M}_R^\delta + \frac{m}{n\delta^2(\log 2)^2} + \frac{\kappa}{(\log 2)^2} \right] d'(x, y)^2 \\ &\le 4 \left[ \mathcal{M}_R^\delta + \frac{m}{n\delta^2} + \kappa \right] d'(x, y)^2, \end{aligned} \tag{44}$$

where inequality $(i)$ holds due to the $\mathcal{M}_R^\delta$-symmetry of the logarithmic metric $H(x) := \sum_{i \in [\mathcal{M}_R^\delta]} \frac{a_i a_i^\top}{(a_i^\top x - b_i)^2}$ (see the proof of Lemma D.2 in Section D.1). $d'(x, y)$ is the combination of Euclidean and cross-ratio distances defined in Lemma B.3.

As the routine for bounding conductance of a transition kernel, we consider the subsets of $A_1, A_2$ with bad conductance. We define:

$$A_1' := \left\{ u \in A_1 \cap \mathcal{B}_R \middle| \mathcal{T}_u(A_2) < \frac{1}{10} \right\}, \quad A_2' := \left\{ u \in A_2 \cap \mathcal{B}_R \middle| \mathcal{T}_u(A_1) < \frac{1}{10} \right\}.$$

We first consider the case $\Pi(A_i') \le \frac{1}{2}\Pi(A_i \cap \mathcal{B}_R)$ for $i = 1$ or $i = 2$. Since $\Pi$ is the stationary distribution of the transition kernel $\mathcal{T}$, we can assume $\Pi(A_1') \le \frac{1}{2}\Pi(A_1 \cap \mathcal{B}_R)$ without loss of generality. we have

$$\begin{aligned} \int_{A_1} \mathcal{T}_u(A_2)\pi(u)du &\ge \int_{A_1 \setminus A_1'} \mathcal{T}_u(A_2)\pi(u)du \overset{(i)}{\ge} \frac{1}{10}\Pi(A_1 \setminus A_1') \ge \frac{1}{20}\Pi(A_1 \cap \mathcal{B}_R) \\ &\ge \frac{1}{20} \min\left\{ \Pi(A_1 \cap \mathcal{B}_R), \Pi(A_2 \cap \mathcal{B}_R) \right\}, \end{aligned}$$

where inequality $(i)$ holds because we assumed $\Pi(A'_1) \leq \frac{1}{2}\Pi(A_1 \cap \mathcal{B}_R)$. So Eq. (43) is proved where we insert $r = 10^{-5}$, $n \geq 1$ and $\kappa \geq 1$.

Next we deal with the case $\Pi(A'_i) > \frac{1}{2}\Pi(A_i \cap \mathcal{B}_R)$ for both $i = 1$ and $i = 2$. For any $x \in A'_1$ and $y \in A'_2$, we have

$$\begin{aligned}\|\mathcal{T}_x - \mathcal{T}_y\|_{TV} &\geq \mathcal{T}_x(A_1) - \mathcal{T}_y(A_1) \\ &= 1 - \mathcal{T}_x(A_2) - \mathcal{T}_y(A_1) > \frac{4}{5},\end{aligned}$$

According to the transition overlap we proved in Lemma B.2, and by symmetry between $x$ and $y$, we have

$$\|y - x\|_{G(x)} > \frac{r}{100\sqrt{n}} \text{ and } \|y - x\|_{G(y)} > \frac{r}{100\sqrt{n}}.$$

Since we assumed $x, y \in \mathcal{B}_R \cap K$, Eq. (44) holds, thus we have

$$d'(x, y) \geq \frac{\|y - x\|_{G(x)}}{2\sqrt{\mathcal{M}_R^\delta + \frac{m}{n\delta^2} + \kappa}} \geq \frac{r}{200\sqrt{n}\sqrt{\mathcal{M}_R^\delta + \frac{m}{n\delta^2} + \kappa}}. \tag{45}$$

Then the LHS of Eq. (43) can be bounded by:

$$\begin{aligned}\int_{A_1} \mathcal{T}_u(A_2)\pi(u)du &= \frac{1}{2}\left[\int_{A_1} \mathcal{T}_u(A_2)\pi(u)du + \int_{A_2} \mathcal{T}_u(A_1)\pi(u)du\right] \\ &\geq \frac{1}{2}\left[\int_{A_1 \setminus A'_1} \mathcal{T}_u(A_2)\pi(u)du + \int_{A_2 \setminus A'_2} \mathcal{T}_u(A_1)\pi(u)du\right] \\ &\geq \frac{1}{20}\Pi(\mathcal{B}_R \setminus (A'_1 \cap A'_2)).\end{aligned}$$

Now we define $\Pi_R(\cdot)$ to be the probability measure $\Pi(\cdot)$ constrained on $\mathcal{B}_R \cap K$. In other words, $\Pi_R(C) := \frac{\Pi(C)}{\Pi(\mathcal{B}_R \cap K)}$ for all Borel sets $C \subseteq K \cap \mathcal{B}_R$. And we define $d'_R(x, y)$ to be the mixed metric whose cross-ratio metric is (see definition in Lemma B.3) restricted on $\mathcal{B}_R \cap K$:

$$d'_R(x, y) := \max\left\{d_{K \cap \mathcal{B}_R}(x, y), \log 2\sqrt{\alpha}\, \|x - y\|_2\right\}.$$

where we changed the cross-ratio distance from over $K$ to over $K \cap \mathcal{B}_R$. It is easy to verify that $d_{K \cap \mathcal{B}_R}(x, y) \geq d_K(x, y)$ because $K \cap \mathcal{B}_R \subseteq K$, so we also have: $d'_R(x, y) \geq d'(x, y)$. Countinuing to lower bound the LHS of Eq. (43), we apply isoperimetric inequality (Lemma B.3) under the metric $d'_R(x, y)$:

$$
\begin{aligned}
\int_{A_1} \mathcal{T}_u(A_2)\pi(u)du &\geq \frac{1}{20}\Pi(\mathcal{B}_R \setminus (A'_1 \cap A'_2)) = \frac{1}{20}\Pi(\mathcal{B}_R)\Pi_R(\mathcal{B}_R \setminus (A'_1 \cap A'_2)) \\
&\overset{(i)}{\geq} \frac{1}{20}\Pi(\mathcal{B}_R)d'_R(A'_1, A'_2)\Pi_R(A'_1)\Pi_R(A'_2) \\
&\overset{(ii)}{\geq} \frac{1}{20}\Pi(\mathcal{B}_R)d'(A'_1, A'_2)\Pi_R(A'_1)\Pi_R(A'_2) \\
&\overset{(iii)}{\geq} \frac{1}{80\Pi(\mathcal{B}_R)}d'(A'_1, A'_2)\Pi(A_1 \cap \mathcal{B}_R)\Pi(A_2 \cap \mathcal{B}_R) \\
&= \frac{1}{80}d'(A'_1, A'_2)\Pi(\mathcal{B}_R)\Pi_R(A_1 \cap \mathcal{B}_R)\Pi_R(A_2 \cap \mathcal{B}_R) \\
&\overset{(iv)}{\geq} \frac{1}{160}d'(A'_1, A'_2)\Pi(\mathcal{B}_R)\min\left\{\Pi_R(A_1 \cap \mathcal{B}_R), \Pi_R(A_2 \cap \mathcal{B}_R)\right\} \\
&= \frac{1}{160}d'(A'_1, A'_2)\min\left\{\Pi(A_1 \cap \mathcal{B}_R), \Pi(A_2 \cap \mathcal{B}_R)\right\},
\end{aligned}
$$

where inequality $(i)$ is the application of our new isoperimetric inequality (Lemma B.3) over the constrained metric $d'_R(x, y)$, inequality $(ii)$ is due to our argument that $d'(x, y) \leq d'_R(x, y)$, inequality $(iii)$ holds because of the assumptions $\Pi(A'_i) > \frac{1}{2}\Pi(A_i \cap \mathcal{B}_R)$ for $i = 1, 2$, and inequality $(iv)$ is due to the fact that $\Pi_R(A_1 \cap \mathcal{B}_R) + \Pi_R(A_2 \cap \mathcal{B}_R) = 1$.

Now insert the upper bound of $d'(x, y)$ in Eq. (44), we have

$$
\int_{A_1} \mathcal{T}_u(A_2)\pi(u)du \geq \frac{r}{32000\sqrt{n}\sqrt{\mathcal{M}_R^\delta + \frac{m}{n\delta^2} + \kappa}} \min\left\{\Pi(A_1 \cap \mathcal{B}_R), \Pi(A_2 \cap \mathcal{B}_R)\right\}.
$$

■

Now we can give the proof of Theorem 3.7, where we apply Lemma D.7 setting $R$ to be $R_o\left(\frac{\epsilon}{2M}\right)$, and $\delta > 0$ is still an arbitrary parameter that can be tuned by the user:

**Proof** [proof of Theorem 3.7] We set $\Upsilon := R_o\left(\frac{\epsilon}{2M}\right)$, then by definition of $R_o$ (see Definition 3.6), we have

$$
\Pi(\mathcal{B}_\Upsilon) \geq 1 - \frac{\epsilon}{2M}.
$$

For any partition $K = A_1 \sqcup A_2$, we have $\Pi(A_i \cap \mathcal{B}_\Upsilon) \geq \Pi(A_i) - \Pi(\mathcal{B}_\Upsilon)$ for $i = 1, 2$. Apply Lemma D.7 where we set $R := \Upsilon$, we have the result as:

$$
\int_{A_1} \mathcal{T}_u(A_2)\pi(u)du \geq \frac{r}{32000\sqrt{n}\sqrt{\mathcal{M}_\Upsilon^\delta + \frac{m}{n\delta^2} + \kappa}} \min\left\{\Pi(A_1) - \frac{\epsilon}{2M}, \Pi(A_2) - \frac{\epsilon}{2M}\right\}.
$$

We set $s := \frac{\epsilon}{2M}$, then the $s$-conductance of our transition kernel $\mathcal{T}$ before lazification satisfies:

$$
\Phi_s \geq \frac{r}{32000\sqrt{n}\sqrt{\mathcal{M}_\Upsilon^\delta + \frac{m}{n\delta^2} + \kappa}}. \tag{46}
$$

After lazification, the $s$-conductance under our new transition kernel $\widetilde{\mathcal{T}}_x(\cdot) := \frac{1}{2}\delta_x(\cdot) + \frac{1}{2}\mathcal{T}_x(\cdot)$ satisfies: $\widetilde{\Phi}_s = \frac{1}{2}\Phi_s$. Using the mixing time bound for lazy Markov chains in Eq. (14), we obtain

$$\left\| \widetilde{\mathcal{T}}^k(\mu_0) - \Pi \right\|_{TV} \leq Ms + M\left(1 - \frac{\widetilde{\Phi}_s^2}{2}\right)^k \leq \frac{\epsilon}{2} + M\exp\left(-\frac{k\widetilde{\Phi}_s^2}{2}\right).$$

To ensure $\left\| \widetilde{\mathcal{T}}^k(\mu_0) - \Pi \right\|_{TV} \leq \epsilon$, we only need to ensure $M\exp\left(-\frac{k\widetilde{\Phi}_s^2}{2}\right) \leq \frac{\epsilon}{2}$. That is, $k \geq \frac{2}{\widetilde{\Phi}_s^2}\log\left(\frac{2M}{\epsilon}\right)$. Insert the lower bound of $\Phi_s$ as in Eq. (46) and the fact that $\widetilde{\Phi}_s = \frac{1}{2}\Phi_s$, we only need to ensure:

$$k \geq \frac{8 \times 32000^2}{r^2}\log\left(\frac{2M}{\epsilon}\right)\left(\kappa + \frac{m}{n\delta^2} + \mathcal{M}_\Upsilon^\delta\right)n, \tag{47}$$

where the step-size $r$ is also an absolute constant, and the desired mixing time is proved. To prove that Eq. (47) still suffices for $\left\| \widetilde{\mathcal{T}}^k(\mu_0) - \Pi \right\|_{TV} \leq \epsilon$ to hold if we replace $\Upsilon$ with $\widehat{\Upsilon} := \widehat{R_o}\left(\frac{\epsilon}{2M}\right)$, we only need to prove $\mathcal{M}_\Upsilon^\delta \leq \mathcal{M}_{\widehat{\Upsilon}}^\delta$. This is easy since $\mathcal{M}_R^\delta$ is clearly an increasing function in $R$, and we have $\Upsilon \leq \widehat{\Upsilon}$ (see the concentration inequality in Eq. (5)). ■

## Appendix E. Bounding the Transition Overlap

In this section, we prove the upper bound of the TV-distance between transitions (Lemma B.2). In Appendix E.1 we prove the properties for strongly self-concordant local metrics. The proof is similar to Laddha et al. (2020), while we can only recover the computation in Laddha et al. (2020) up to constant 2. Though this does not affect the usefulness, we list our proof to be rigorous. In Appendix E.2, we prove Lemma B.2.

### E.1. Strong Self-Concordance Properties

In Appendix E.1 we prove the properties for strongly self-concordant local metrics (Proposition E.1).

**Proposition E.1 (adapted from Laddha et al. (2020))** *Assume $G$ is SSC over $K$, then for any $x, y \in K$ with $\|x - y\|_{G(x)} < 1$, we have*

$$\left\| G(x)^{-\frac{1}{2}}(G(y) - G(x))G(x)^{-\frac{1}{2}} \right\|_F \leq \frac{2\|x - y\|_{G(x)}}{(1 - \|x - y\|_{G(x)})^2}, \tag{48}$$

*for any $\|x - y\|_{G(x)} \leq \frac{1}{2}$, we have*

$$\det\left[G(x)^{-\frac{1}{2}}G(y)G(x)^{-\frac{1}{2}}\right] \leq \exp\left(8\sqrt{n}\|x - y\|_{G(x)}\right). \tag{49}$$

The upper bound on Frobenius norm of $G(x)^{-\frac{1}{2}}\left[G(y) - G(x)\right]G(x)^{-\frac{1}{2}}$ was proved as Lemma 1.2 in Laddha et al. (2020). However, we are not able to recover this bound following the argument in Laddha et al. (2020). We followed the procedures in Laddha et al. (2020) and found our bound (Eq. 48) differs from Laddha et al. (2020) by a factor of 2. Even though this does not have a major impact over the applications of this lemma, we still list our proof for the sake of completeness.

The upper bound of the determinant of $G(x)^{-\frac{1}{2}}\left[G(y)-G(x)\right]G(x)^{-\frac{1}{2}}$ (Eq. (49)) is an easy application of Cauchy-Schwarz inequality over Eq. (48). Same as in Laddha et al. (2020), we first introduce the weaker notion of self-concordance and a well-known lemma:

**Definition E.2 (Self-Concordance)** *For convex set $K \subseteq \mathbb{R}^n$, we call a local metric $G : K \to \mathbb{R}^{n\times n}$ self-concordant if for any $x \in K$, any direction $h \in \mathbb{R}^n$, the derivative of $G$ along $h$ satisfies:*

$$-2\left\|h\right\|_{G(x)} G(x) \preceq \mathcal{D}G(x)[h] \preceq 2\left\|h\right\|_{G(x)} G(x).$$

It is easy to see from the definition that all strongly self-concordant local metric $G$ is self-concordant. The proof of Proposition E.1 requires the following properties of self-concordant local metrics. For the sake of completeness, we list it as Proposition E.3 without proof.

**Proposition E.3 (Lemma 1.1 from Laddha et al. (2020))** *Given any self-concordant matrix function $G$ on $K \subseteq \mathbb{R}^n$, for any $x, y \in K$ with $\|y-x\|_{G(x)} < 1$, we have*

$$\left[1-\|x-y\|_{G(x)}\right]^2 G(x) \preceq G(y) \preceq \left[1-\|x-y\|_{G(x)}\right]^{-2} G(x).$$

Now we are prepared to give the proof of Proposition E.1:

**Proof** [proof of Proposition E.1] First, we prove Eq. (48). Fix $x, y \in K$, let $h := y - x$. For $t \in [0,1]$, define $x_t := x + th$, then we have

$$\begin{aligned}\left\|G(x)^{-\frac{1}{2}}(G(y)-G(x))G(x)^{-\frac{1}{2}}\right\|_F &= \left\|G(x)^{-\frac{1}{2}}\left[\int_0^1 \frac{d}{dt}G(x_t)dt\right]G(x)^{-\frac{1}{2}}\right\|_F \\ &\le \int_0^1 \left\|G(x)^{-\frac{1}{2}}\frac{d}{dt}G(x_t)G(x)^{-\frac{1}{2}}\right\|_F dt.\end{aligned}$$

Then we can further bound the integrand:

$$\begin{aligned}\left\|G(x)^{-\frac{1}{2}}\frac{d}{dt}G(x_t)G(x)^{-\frac{1}{2}}\right\|_F^2 &= \operatorname{Tr}\left\{G(x)^{-1}\frac{d}{dt}G(x_t)G(x)^{-1}\frac{d}{dt}G(x_t)\right\} \\ &\overset{(i)}{\le} \left[1-\|x_t-x\|_{G(x)}\right]^{-4}\operatorname{Tr}\left\{G(x_t)^{-1}\frac{d}{dt}G(x_t)G(x_t)^{-1}\frac{d}{dt}G(x_t)\right\} \\ &= \left[1-\|x_t-x\|_{G(x)}\right]^{-4}\left\|G(x_t)^{-\frac{1}{2}}\mathcal{D}G(x_t)[h]G(x_t)^{-\frac{1}{2}}\right\|_F^2 \\ &\overset{(ii)}{\le} 4\left[1-\|x_t-x\|_{G(x)}\right]^{-4} h^\top G(x_t)h \\ &\overset{(iii)}{\le} 4\left[1-\|x_t-x\|_{G(x)}\right]^{-4}\left[1-\|x_t-x\|_{G(x)}\right]^{-2} h^\top G(x)h \\ &= 4\left[1-t\left\|h\right\|_{G(x)}\right]^{-6} h^\top G(x)h,\end{aligned}$$

where inequality $(i)$ and $(iii)$ hold because we apply Proposition E.3 at $x_t$, and inequality $(ii)$ holds due to the assumption that $G$ is strongly self-concordant. Using the upper bound of the integrand, we can bound the integral:

$$\left\|G(x)^{-\frac{1}{2}}(G(y)-G(x))G(x)^{-\frac{1}{2}}\right\|_F \le \int_0^1 \frac{2\left\|h\right\|_{G(x)}dt}{(1-t\left\|h\right\|_{G(x)})^3} \overset{(i)}{=} \frac{2\left\|h\right\|_{G(x)}-\left\|h\right\|_{G(x)}^2}{(1-\left\|h\right\|_{G(x)})^2},$$

where equality $(i)$ is a direct calculation of the definite integral. As a result, Eq. (48) is proved since $\|h\|_{G(x)}^2 \geq 0$. However, we notice that the definite integral implies an extra factor of 2 compared to Lemma 1.2 in Laddha et al. (2020).

Next, to prove Eq. (49), we the eigenvalues of $G(x)^{-\frac{1}{2}}G(y)G(x)^{-\frac{1}{2}}$ to be $\lambda_i$ for $i = 1, \ldots, n$. Then the determinant of $G(x)^{-\frac{1}{2}}G(y)G(x)^{-\frac{1}{2}}$ can be written as:

$$\det\left[G(x)^{-\frac{1}{2}}G(y)G(x)^{-\frac{1}{2}}\right] = \prod_{i=1}^{n} \lambda_i \leq \exp\left[\sum_{i=1}^{n}(\lambda_i - 1)\right].$$

To prove Eq. (49), we only need to prove $\sum |\lambda_i - 1| \leq 8\sqrt{n}\, \|y - x\|_{G(x)}$. This bound can be obtained:

$$\begin{aligned}\sum_{i=1}^{n} |\lambda_i - 1| &\leq \sqrt{n}\left[\sum_{i=1}^{n}(\lambda_i - 1)^2\right]^{\frac{1}{2}} = \sqrt{n}\left\|G(x)^{-\frac{1}{2}}G(y)G(x)^{-\frac{1}{2}} - I\right\|_F \\ &\overset{(i)}{\leq} \frac{2\sqrt{n}\,\|y - x\|_{G(x)}}{(1 - \|y - x\|_{G(x)})^2} \overset{(ii)}{\leq} 8\sqrt{n}\,\|y - x\|_{G(x)}\,,\end{aligned}$$

where inequality $(i)$ is due to Eq. (48) for strongly self-concordant $G$, and inequality $(ii)$ holds since we assumed $\|y - x\|_{G(x)} \leq \frac{1}{2}$. ■

### E.2. Bounding Transition Overlap

In this section, we are prove the transition overlap in Lemma B.2. We separate this task into three parts: Lemma E.5 bounds the acceptance rate to be a little less than $1/2$ globally, Lemma E.6 extends the close coupling argument in Andrieu et al. (2024) and circumvent the naive use of triangle inequality $\|\mathcal{T}_x - \mathcal{T}_y\|_{TV} \leq \|\mathcal{P}_x - \mathcal{T}_x\|_{TV} + \|\mathcal{P}_x - \mathcal{P}_y\|_{TV} + \|\mathcal{P}_y - \mathcal{T}_y\|_{TV}$, and Lemma E.7 controls the TV-distance between two proposal distributions using SSC and Pinsker's inequality. Throughout the proof, we need the following two facts.

**Proposition E.4 (Gaussian Concentration)** *Assume $\xi \sim \mathcal{N}(0, I_n)$, then for any $t > 0$, we have*

$$\mathbb{P}\left(\|\xi\|_2 \geq t\sqrt{n}\right) \leq 2\exp\left(-\frac{t^2}{2}\right)$$

**Lemma E.5 (Acceptance Rate Control)** *Let $G$ be a SSC, ASC, LTSC matrix function defined on $K$, and we also assume $G(x) \succeq \beta I$ for $x \in K$, then there exists an absolute constant $r_0 > 0$, such that for all step-size $r < r_0$, we have*

$$\|\mathcal{T}_x - \mathcal{P}_x\|_{TV} \leq 0.6$$

**Proof** Due to Markov inequality, fix any $\gamma \in (0, 1)$, we have

$$\begin{aligned}\|\mathcal{T}_x - \mathcal{P}_x\|_{TV} &= 1 - \mathbb{E}_{z\sim\mathcal{P}_x}\left[\alpha(x, z)\right] \\ &\leq 1 - \gamma\mathbb{P}_{z\sim\mathcal{P}_x}\left[\alpha(x, z) \geq \gamma\right] \\ &\leq 1 - \gamma\mathbb{P}_{z\sim\mathcal{P}_x}\left[\mathbf{1}_K(z)\frac{e^{-f(z)}p_z(x)}{e^{-f(x)}p_x(z)} \geq \gamma\right]\end{aligned} \tag{50}$$

For convenience, we can first assume that $r \leq 10^{-4}$. It is easy to bound the term $\mathbf{1}_K(x)$, using the fact that the ellipsoid $E(x, G(x), 1)$ is contained in $K$, combined with Gaussian concentration properties. To be explicit, we have

$$\begin{aligned}\mathbb{P}\left\{\mathbf{1}_K(z)=1\right\} &\geq \mathbb{P}\left\{\|z-x\|_{G(x)}<1\right\} = \mathbb{P}_{\xi\sim\mathcal{N}(0,I_n)}\left\{\xi^\top\xi<\frac{n}{r^2}\right\}\\ &\overset{(i)}{\geq} 1-2\exp\left(\frac{-1}{2r^2}\right) \overset{(ii)}{\geq} 0.99\end{aligned} \tag{51}$$

where the inequality $(i)$ is a result of Proposition E.4, and the inequality $(ii)$ holds since we set $r \leq 10^{-4}$. Now we bound the term $\frac{p_z(x)}{p_x(z)}$, for convenience we define $g(x) := \log\det G(x)$, and we have

$$\begin{aligned}\log\frac{p_z(x)}{p_x(z)} &= -\frac{n}{2r^2}(z-x)^\top(G(z)-G(x))(z-x)+\frac{1}{2}(g(z)-g(x))\\ &\overset{(i)}{=} \underbrace{-\frac{n}{2r^2}(z-x)^\top(G(z)-G(x))(z-x)}_{\mathrm{I}}+\frac{1}{2}\underbrace{\nabla g(x)^\top(z-x)}_{\mathrm{II}}+\frac{1}{4}\underbrace{\mathcal{D}^2 g(x^*)[z-x,z-x]}_{\mathrm{III}}.\end{aligned}$$

where we used Taylor's theorem, and $x^* \in [x, z]$ denotes some midpoint for which equality $(i)$ holds. We first bound term I with high probability. By the ASC property in Definition A.3, there exists an $r_{\mathrm{ASC}} > 0$ such that the following holds for any $r < r_{\mathrm{ASC}}$

$$\mathbb{P}_{z\sim\mathcal{N}(x,\frac{r^2}{n}G(x))}\left\{\|z-x\|_{G(z)}^2-\|z-x\|_{G(x)}^2 \leq 2\cdot\frac{0.01r^2}{n}\right\} \geq 1-0.01 \tag{52}$$

Then we bound term II with high probability. We notice that term II can be bounded by the concentration of Gaussian variables. Using Proposition E.4 with $t = 4$, then with probability $\geq 0.99$ in $z$, we have

$$\begin{aligned}\mathrm{II} &= \left\langle G(x)^{-\frac{1}{2}}\nabla g(x), G(x)^{\frac{1}{2}}(z-x)\right\rangle \geq -\frac{4r}{\sqrt{n}}\left\|G(x)^{-\frac{1}{2}}\nabla g(x)\right\|_2\\ &= \frac{-4r}{\sqrt{n}}\sup_{\|v\|_2=1}\nabla g(x)^\top G(x)^{-\frac{1}{2}}v\\ &= \frac{-4r}{\sqrt{n}}\sup_{\|v\|_2=1}\mathrm{Tr}\left(G(x)^{-1}\mathcal{D}G(x)\left[G(x)^{-\frac{1}{2}}v\right]\right)\\ &\geq -4r\sup_{\|v\|_2=1}\left\|G(x)^{-\frac{1}{2}}\mathcal{D}G(x)\left[G(x)^{-\frac{1}{2}}v\right]G(x)^{-\frac{1}{2}}\right\|_F\\ &\geq -8r\sup_{\|v\|_2=1}\left\|G(x)^{-\frac{1}{2}}v\right\|_{G(x)} = -8r,\end{aligned} \tag{53}$$

where the last inequality follows from SSC of $G$. Now we continue to bound term III. For convenience, we let $h := z - x$ and we have

$$\begin{aligned}\mathcal{D}^2 g(x^*)[h,h] &= \mathrm{Tr}\left(G(x^*)^{-1}\mathcal{D}^2 G(x^*)[h,h]\right)-\mathrm{Tr}\left(G(x^*)^{-1}\mathcal{D}G(x^*)[h]G(x^*)^{-1}\mathcal{D}G(x^*)[h]\right)\\ &\overset{(i)}{\geq} -\|h\|_{G(x^*)}^2-\left\|G(x^*)^{-\frac{1}{2}}\mathcal{D}G(x^*)G(x^*)^{-\frac{1}{2}}\right\|_F^2\\ &\overset{(ii)}{\geq} -\|h\|_{G(x^*)}^2-4\|h\|_{G(x^*)}^2 \overset{(iii)}{\geq} -5(1-\|h\|_{G(x)})^{-2}\|h\|_{G(x)}^2,\end{aligned}$$

where inequality $(i)$ follows from LTSC of $G$, inequality $(ii)$ follows from ASC of $G$, and inequality $(iii)$ follows from self-concordant properties (see Proposition E.3). Using Proposition E.4 with $t = 4$, then we have $\mathbb{P}_z \left( \|h\|_{G(x)} > 4r \right) \le 0.01$, and this implies with probability $\ge 0.99$ we have

$$\text{III} = \mathcal{D}^2 g(x^*)[h, h] \overset{(i)}{\ge} - \cdot 5 \cdot 4 \cdot 16r^2 = -320r^2 \tag{54}$$

where inequality $(i)$ holds due to we assumed $r \le \frac{1}{2}$ and we insert $\|h\|_{G(x)} > 4r$. Now insert the bound of $\text{I}, \text{II}, \text{III}$ in Eq. (52),(53) and (54), together with the fact that we set $r \le 10^{-4}$, then we have

$$\mathbb{P}_{z\sim\mathcal{P}_x} \left\{ \log \frac{p_z(x)}{p_x(z)} \ge -0.01 - 4r - 80r^2 \ge -0.01 \right\} \ge 0.97. \tag{55}$$

Now we continue to bound the term of stationary distribution

$$\begin{aligned}
\frac{e^{-f(z)}}{e^{-f(x)}} &\overset{(i)}{\ge} \exp\left( \nabla f(z)^\top (x - z) \right) \\
&= \exp\left[ (\nabla f(z) - \nabla f(x))^\top (x - z) + \nabla f(x)^\top (x - z) \right] \\
&\overset{(ii)}{\ge} \exp\left( -\beta \|z - x\|_2^2 \right) \exp\left[ \nabla f(x)^\top (x - z) \right] \\
&\overset{(iii)}{\ge} \exp\left( - \|z - x\|_{G(x)}^2 \right) \exp\left[ \nabla f(x)^\top (x - z) \right]
\end{aligned}$$

where the inequality $(i)$ follows from the convexity of $f$, the inequality $(ii)$ follows from the $\beta$-smoothness of $f$, and the inequality $(iii)$ follows from our assumption that $G(x) \succeq \beta I$. In order to control the term $\nabla f(x)^\top (z - x)$, we notice that $\nabla f(x)^\top (z - x)$ is a one dimensional Gaussian variable centered at 0. By the symmetric property of Gaussian distributions, we have

$$\mathbb{P}_{z\sim\mathcal{P}_x} \left\{ \exp\left[ \nabla f(x)^\top (x - z) \right] \ge 1 \right\} \ge 0.5 \tag{56}$$

Now combining Eq. (51), (55) and (56), we have

$$\mathbb{P}_{z\sim\mathcal{P}_x} \left[ \mathbf{1}_K(z) \frac{e^{-f(z)} p_z(x)}{e^{-f(x)} p_x(z)} \ge \exp\left( - \|z - x\|_{G(x)}^2 - 0.01 \right) \right] \ge 0.5 - 0.04 \tag{57}$$

Using the fact that of Gaussian concentration

$$\mathbb{P} \left\{ \|z - x\|_{G(x)}^2 \ge 16r^2 \right\} = \mathbb{P}_{\xi\sim\mathcal{N}(0,I_n)} \left\{ \|\xi\|_2 \ge 4n \right\} \le 2\exp(-8) \le 0.01$$

This translate Eq. (57) into:

$$\mathbb{P}_{z\sim\mathcal{P}_x} \left[ \mathbf{1}_K(z) \frac{e^{-f(z)} p_z(x)}{e^{-f(x)} p_x(z)} \ge \exp\left( -16r^2 - 0.01 \right) \right] \ge 0.46 - 0.01 \tag{58}$$

As a result of Markov inequality (Eq. (50)), we have

$$\|\mathcal{T}_x - \mathcal{P}_x\|_{TV} \le 1 - 0.45 \cdot \exp(-16r^2 - 0.01) \le 0.6,$$

and it holds for all $r \le \min\left\{ r_{\text{ASC}}, 10^{-4} \right\}$. ∎

**Lemma E.6 (close coupling)** *Let $G$ be a SSC, ASC matrix function defined on $K$. Fix any $\gamma \in (0,1)$, there exists $r_\gamma > 0$ such that for all $r \leq r_\gamma$, $x, y \in K$ and*

$$\max\left\{\|y-x\|_{G(x)}, \|y-x\|_{G(y)}\right\} \leq \frac{r}{\sqrt{n}}, \tag{59}$$

*we have the following bound on the TV-distance between transition distributions:*

$$\|\mathcal{T}_y - \mathcal{T}_x\|_{TV} \leq \gamma + \|\mathcal{T}_x - \mathcal{P}_x\|_{TV} + \|\mathcal{P}_x - \mathcal{P}_y\|_{TV} \tag{60}$$

**Proof** This lemma extends the close-coupling technique in Andrieu et al. (2024) to our asymmetric proposals. For convenience, we use $p_x(\cdot)$ to denote the probability density function of $\mathcal{P}_x$. In other words, we define $p_x(z) := \mathcal{N}(z|x, \frac{r^2}{n}G(x)^{-1})$.

Without loss of generality, we assume $e^{-f(x)} \geq e^{-f(y)}$.

Consider the maximal coupling of $(\mathcal{P}_x, \mathcal{P}_y)$, there exists a pair of random vectors $(V_x, V_y)$ such that $V_x \sim \mathcal{P}_x$ and $V_y \sim \mathcal{P}_y$ and $\mathbb{P}\{V_x = V_y\} = 1 - \|\mathcal{P}_x - \mathcal{P}_y\|_{TV}$. Then we draw $U \sim \text{unif}(0,1)$ independent of $(V_x, V_y)$, and we define random vectors $X', Y'$ according to the following rule:

$$\begin{aligned}
X' &:= V_x \mathbb{I}_{\{U \leq \alpha(x, V_x)\}} + x\mathbb{I}_{\{U > \alpha(x, V_x)\}},\\
Y' &:= V_y \mathbb{I}_{\{U \leq \alpha(y, V_y)\}} + y\mathbb{I}_{\{U > \alpha(y, V_y)\}},
\end{aligned}$$

where $\alpha(x,z)$ denotes the acceptance rate $\alpha(x,z) := \min\left\{1, \frac{e^{-f(z)}p_z(x)}{e^{-f(x)}p_x(z)}\mathbf{1}_K(z)\right\}$. For convenience of discussion, we define two events $\mathcal{E}_1$ and $\mathcal{E}_2$,

$$\mathcal{E}_1 := \left\{V_x = V_y, X' = V_x, Y' \neq V_y\right\}, \quad \mathcal{E}_2 := \{V_x = V_y, \alpha(y, V_y) < \alpha(x, V_x)\}.$$

It is straightforward to verify that $X' \sim \mathcal{T}_x$ and $Y' \sim \mathcal{T}_y$, so we have

$$\begin{aligned}
\mathbb{P}\left\{X' = Y'\right\} &\geq \mathbb{P}\left\{V_x = V_y, X' = V_x, Y' = V_y\right\}\\
&= \mathbb{P}\left\{V_x = V_y, X' = V_x\right\} - \mathbb{P}(\mathcal{E}_1)\\
&\geq \mathbb{P}\{V_x = V_y\} + \mathbb{P}\left\{V_x = X'\right\} - 1 - \mathbb{P}(\mathcal{E}_1),
\end{aligned}$$

where the last inequality follows from the fact that $1 \geq \mathbb{P}(B_1 \cup B_2) = \mathbb{P}(B_1) + \mathbb{P}(B_2) - \mathbb{P}(B_1 \cap B_2)$ for any event $B_1, B_2$. Then we insert the coupling equation $\mathbb{P}\{V_x = V_y\} = 1 - \|\mathcal{P}_x - \mathcal{P}_y\|_{TV}$, and the easy-to-verify fact $\mathbb{P}\{V_x = X'\} = 1 - \|\mathcal{P}_x - \mathcal{T}_x\|_{TV}$, we have

$$\begin{aligned}
1 - \mathbb{P}\left\{X' = Y'\right\} &\leq \|\mathcal{P}_x - \mathcal{P}_y\|_{TV} + \|\mathcal{P}_x - \mathcal{T}_x\|_{TV} + \mathbb{P}(\mathcal{E}_1)\\
\|\mathcal{T}_x - \mathcal{T}_y\|_{TV} &\leq \|\mathcal{P}_x - \mathcal{P}_y\|_{TV} + \|\mathcal{P}_x - \mathcal{T}_x\|_{TV} + \mathbb{P}(\mathcal{E}_1),
\end{aligned}$$

where the last inequality follows by the coupling inequality $\|\mathcal{T}_x - \mathcal{T}_y\|_{TV} \leq \mathbb{P}\{X' \neq Y'\}$ when $X' \sim \mathcal{T}_x$ and $Y' \sim \mathcal{T}_y$. In order to prove Eq. (60), we only need to control $\mathbb{P}(\mathcal{E}_1)$ to be smaller than $\gamma$,

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_1) &\overset{(i)}{\leq} \mathbb{P}\{V_x = V_y, \alpha(y, V_y) < U \leq \alpha(x, V_x)\}\\
&= \mathbb{E}_{(V_x, V_y, U)}\left[\mathbb{I}_{\{V_x = V_y\}} \cdot \mathbb{I}_{\{\alpha(y, V_y) < U \leq \alpha(x, V_x)\}}\right]\\
&\overset{(ii)}{=} \mathbb{E}_{(V_x, V_y)}\left\{\mathbb{I}_{\{V_x = V_y, \alpha(y, V_y) < \alpha(x, V_x)\}} \times [\alpha(x, V_x) - \alpha(y, V_y)]\right\}
\end{aligned}$$

where inequality $(i)$ holds since $Y' \neq V_y \Rightarrow U > \alpha(y, V_y)$, and $X' = V_x \Rightarrow U \leq \alpha(x, V_x)$ except on a set with probability 0 (the proposal satisfies $V_x = x$), because our proposal has no point mass. Equality $(ii)$ follows from our assumption that $U$ is independent of $(V_x, V_y)$. Then we continue to control the difference in acceptance rates at $x$ and $y$. For convenience, we denote $\Lambda(x, z)$ to be the following ratio,

$$\Lambda(x, z) := \frac{e^{-f(z)} p_z(x)}{e^{-f(x)} p_x(z)}.$$

Then we have

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_1) &\leq \mathbb{E}_{(V_x,V_y)}\left[\mathbb{I}_{\mathcal{E}_2} \cdot (\alpha(x, V_x) - \alpha(y, V_y))\right] \\
&\overset{(i)}{=} \mathbb{E}_{(V_x,V_y)}\left[\mathbb{I}_{\mathcal{E}_2} \cdot (\min\{1, \Lambda(x, V_x)\mathbf{1}_K(V_x)\} - \Lambda(y, V_y)\mathbf{1}_K(V_y))\right] \\
&\overset{(ii)}{\leq} \mathbb{E}_{(V_x,V_y)}\left[\mathbb{I}_{\mathcal{E}_2} \min\{1, \Lambda(x, V_x)\mathbf{1}_K(V_x) - \Lambda(y, V_y)\mathbf{1}_K(V_y)\}\right] \\
&\leq \underbrace{\mathbb{E}_{(V_x,V_y)}\left[\mathbb{I}_{\mathcal{E}_2 \cap \{V_x, V_y \in K\}} \min\{1, \Lambda(x, V_x) - \Lambda(y, V_y)\}\right]}_{\text{I}} + \underbrace{\mathbb{P}\{V_x \notin K\} + \mathbb{P}\{V_y \notin K\}}_{\text{II}},
\end{aligned}$$

where equality $(i)$ follows since in the event $\mathcal{E}_2$ we have $\alpha(y, V_y) < \alpha(x, V_x) \leq 1$. Inequality $(ii)$ holds since for all $a, b \geq 0$ we have $\min\{1, b\} - a \leq \min(1, b - a)$. We proceed by controlling term I and term II separately. Term II can be controlled easily by Gaussian concentration as in Eq. (51) in Lemma E.5:

$$\text{II} \leq 2\mathbb{P}_{\xi \sim \mathcal{N}(0, I_n)}\left\{\xi^\top \xi > \frac{n}{r^2}\right\} \leq 4\exp\left(-\frac{1}{2r^2}\right) \overset{(i)}{\leq} \frac{\gamma}{2},$$

where we only need to set $r \leq \left(2\log(\frac{8}{\gamma})\right)^{-1/2}$ to ensure inequality $(i)$ holds. Now we proceed to control term I:

$$\begin{aligned}
\text{I} &\overset{(i)}{\leq} \mathbb{E}_{(V_x,V_y)}\left[\mathbb{I}_{\mathcal{E}_2 \cap \{V_x, V_y \in K\}} \min\left\{1, \frac{e^{-f(V_y)}}{e^{-f(y)}}\left(\frac{p_{V_x}(x)}{p_x(V_x)} - \frac{p_{V_y}(y)}{p_y(V_y)}\right)\right\}\right] \\
&\leq \mathbb{E}_{(V_x,V_y)}\left[\mathbb{I}_{\mathcal{E}_2 \cap \{V_x, V_y \in K\}} \min\left\{1, \frac{e^{-f(V_y)} p_{V_y}(y)}{e^{-f(y)} p_y(V_y)}\left|\frac{p_{V_x}(x) p_y(V_y)}{p_x(V_x) p_{V_y}(y)} - 1\right|\right\}\right] \\
&\overset{(ii)}{\leq} \mathbb{E}_{(V_x,V_y)}\left[\mathbb{I}_{\mathcal{E}_2 \cap \{V_x, V_y \in K\}} \min\left\{1, \left|\frac{p_{V_x}(x) p_y(V_y)}{p_x(V_x) p_{V_y}(y)} - 1\right|\right\}\right]
\end{aligned} \tag{61}$$

where inequality $(i)$ holds since $V_x = V_y$ in the event $\mathcal{E}_2$ and we assumed $e^{-f(x)} \geq e^{-f(y)}$. Inequality $(ii)$ holds because on the event $\mathcal{E}_2 \cap \{V_x, V_y \in K\}$ we have

$$\Lambda(y, V_y) = \Lambda(y, V_y)\mathbf{1}_K(y) = \alpha(y, V_y) < 1.$$

For convenience, we let the notation $\Gamma(x, z)$ denote the following expression:

$$\Gamma(x, z) = -\frac{n}{2r^2}(z - x)^\top (G(z) - G(x))(z - x). \tag{62}$$

Now we continue the inequality in Eq. (61):

$$
\begin{aligned}
\mathrm{I} &\leq \mathbb{E}_{(V_x,V_y)}\left[\mathbb{I}_{\mathcal{E}_2\cap\{V_x,V_y\in K\}}\min\left\{\left|\frac{p_{V_x}(x)p_y(V_y)}{p_x(V_x)p_{V_y}(y)}-1\right|,1\right\}\right]\\
&\overset{(i)}{=} \mathbb{E}_{(V_x,V_y)}\left[\mathbb{I}_{\mathcal{E}_2\cap\{V_x,V_y\in K\}}\min\left\{\left|\sqrt{\frac{\det G(y)}{\det G(x)}}\frac{\exp\left[\Gamma(x,V_x)\right]}{\exp\left[\Gamma(y,V_y)\right]}-1\right|,1\right\}\right]\\
&\leq \mathbb{E}_{(V_x,V_y)}\min\left\{\left|\sqrt{\frac{\det G(y)}{\det G(x)}}\frac{\exp\left[\Gamma(x,V_x)\right]}{\exp\left[\Gamma(y,V_y)\right]}-1\right|,1\right\}\\
&\overset{(ii)}{\leq} \mathbb{P}\left\{\left|\sqrt{\frac{\det G(y)}{\det G(x)}}\frac{\exp\left[\Gamma(x,V_x)\right]}{\exp\left[\Gamma(y,V_y)\right]}-1\right|\geq e^{3c}-1\right\}+(e^{3c}-1)\\
&\overset{(iii)}{\leq} \mathbb{P}\left\{\left|\Gamma(y,V_y)\right|\geq c\right\}+\mathbb{P}\left\{\left|\Gamma(x,V_x)\right|\geq c\right\}+(e^{3c}-1)
\end{aligned}
$$

where inequality $(i)$ holds since we insert the proposal densities and we notice $V_x = V_y$ in the event $\mathcal{E}_2$. Inequality $(ii)$ holds for any constant $c \geq 0$, and we would determine $c$ later. Since we have $\sqrt{\det(G(y)G(x)^{-1})} \in [e^{-4r}, e^{4r}]$ due to SSC property in Proposition E.4 and the closeness of $x$ and $y$ in Eq. (59), and we can always set $r \leq c/4$ so that inequality $(iii)$ holds.

Now we set $c := \min\left\{\frac{1}{3}\log(1+\frac{\gamma}{6}), \frac{\gamma}{6}\right\}$, using ASC property in Definition A.3, then there exists $r_c > 0$ such that for any step-size $r \leq r_c$, we have

$$
\mathrm{I} \leq c + c + e^{3c} - 1 \leq \frac{\gamma}{6} + \frac{\gamma}{6} + \frac{\gamma}{6} = \frac{\gamma}{2} \tag{63}
$$

The result is proved if we combine term I and II for all $r < \min\left\{\frac{c}{4}, r_c, \left(2\log(\frac{8}{\gamma})\right)^{-1/2}\right\}$. ■

**Lemma E.7 (Proposal Overlap)** *Assume $r \leq \frac{1}{16}$ to be the step-size, and $x, y \in K$ such that $\|y-x\|_{G(x)} \leq \frac{r}{10\sqrt{n}}$. Recall that we use $\mathcal{P}_x$, $\mathcal{P}_y$ to denote the Gaussian proposal distributions at $x, y \in K$. In other words, $\mathcal{P}_x := \mathcal{N}(x, \frac{r^2}{n}G(x)^{-1})$, $\mathcal{P}_y := \mathcal{N}(x, \frac{r^2}{n}G(y)^{-1})$, then we have the following bound:*

$$
\|\mathcal{P}_x - \mathcal{P}_y\|_{TV} \leq \sqrt{\frac{1}{100} + 24r}
$$

**Proof** By Pinsker's inequality $\|\mathcal{P}_x - \mathcal{P}_y\|_{TV} \leq \sqrt{2\mathrm{KL}(\mathcal{P}_y||\mathcal{P}_x)}$. The KL-divergence between two Gaussian distributions $P = \mathcal{N}(\mu_1, \Sigma_1)$ and $Q = \mathcal{N}(\mu_2, \Sigma_2)$ can be computed analytically:

$$
\mathrm{KL}(Q||P) = \frac{1}{2}\left\{\underbrace{\mathrm{Tr}(\Sigma_1^{-1}\Sigma_2) - n}_{\mathrm{I}} + \underbrace{\log\det\left(\Sigma_1\Sigma_2^{-1}\right)}_{\mathrm{II}} + \underbrace{(\mu_1-\mu_2)^\top\Sigma_1^{-1}(\mu_1-\mu_2)}_{\mathrm{III}}\right\}. \tag{64}
$$

We plug in $P := \mathcal{N}(x, \frac{r^2}{n}G(x)^{-1})$ and $Q := \mathcal{N}(y, \frac{r^2}{n}G(y)^{-1})$, then bound the terms I, II, and III separately. We first bound the determinant ratio I:

$$
\mathrm{I} = \log\det\left(G(y)G(x)^{-1}\right) \overset{(i)}{\leq} 8\sqrt{n}\,\|y-x\|_{G(x)} \leq 8r,
$$

where inequality $(i)$ follows from the properties of SSC in Proposition E.1. Then we bound the trace term II, we denote the eigenvalues of $G(x)^{-\frac{1}{2}}G(y)G(x)^{-\frac{1}{2}}$ to be $\lambda_1, \ldots, \lambda_n$, so according to Proposition E.1 we have

$$\sum_{i=1}^{n} |\lambda_i - 1| \leq \sqrt{n} \left\| G(x)^{-\frac{1}{2}}G(y)G(x)^{-\frac{1}{2}} - I \right\|_F \overset{(i)}{\leq} \sqrt{n} \cdot \frac{2r/\sqrt{n}}{\left(1-\frac{1}{2}\right)^2} = 8r, \tag{65}$$

where inequality $(i)$ holds because we assumed $r \leq \frac{1}{2}$. Since we further assumed $r < \frac{1}{16}$, thus Eq. (65) further implies that $\lambda_i \geq \frac{1}{2}$ for all index $i \in [n]$. Thus we have

$$\mathrm{II} = -n + \mathrm{Tr}\left(G(x)G(y)^{-1}\right) = \sum_{i=1}^{n}\left(\frac{1}{\lambda_i} - 1\right) \leq \sum_{i=1}^{n}\left|\frac{\lambda_i - 1}{\lambda_i}\right| \leq 2\sum_{i=1}^{n}|\lambda_i - 1| \leq 16r.$$

Finally, term III can be bounded as

$$\mathrm{III} = \frac{n}{r^2}(y-x)^\top G(x)(y-x) = \frac{n}{r^2}\|y-x\|_{G(x)}^2 \leq \frac{n}{r^2}\cdot\frac{r^2}{100n} = \frac{1}{100}.$$

■

With the preparation of Lemma E.5, E.6 and E.7, we now give the proof of Lemma B.2.

**Proof** [proof of Lemma B.2] We first notice that $\|y-x\|_{G(y)} \leq \frac{\|y-x\|_{G(x)}}{1-\|y-x\|_{G(x)}}$ by Proposition E.3, and since we can set $r \leq \frac{1}{2}$, we have

$$\max\left\{\|y-x\|_{G(x)}, \|y-x\|_{G(y)}\right\} \leq 2\|y-x\|_{G(x)} \leq \frac{r}{\sqrt{n}},$$

so we can use Lemma E.5 and Lemma E.6 with $\gamma = 0.01$, then there exists $r_0, r_\gamma > 0$ such that for all $r < \min\left\{10^{-4}, r_0, r_\gamma\right\}$ we have

$$\begin{aligned}
\|\mathcal{T}_x - \mathcal{T}_y\|_{TV} &\leq 0.01 + \|\mathcal{T}_x - \mathcal{P}_x\|_{TV} + \|\mathcal{P}_x - \mathcal{P}_y\|_{TV} \\
&\leq 0.01 + 0.6 + \|\mathcal{P}_x - \mathcal{P}_y\|_{TV} \\
&\leq 0.61 + \sqrt{\frac{1}{100} + 24r} \leq \frac{4}{5}.
\end{aligned}$$

■

## Appendix F. Properties of Soft-Threshold Metric

In this section, we briefly verify the SSC, LTSC and ASC of soft-threshold metric as defined in Definition 1.3, and proving the correctness of Lemma D.1.

For the SSC, it is well-known that the unregularized logarithmic metric $H(x) = A_x^\top A_x$ is SSC when $H(x)$ is invertible. Adding a regularization term $\lambda I_n$ is only helping us:

$$\left\| G(x)^{-\frac{1}{2}}\mathcal{D}G(x)[h]G(x)^{-\frac{1}{2}} \right\|_F$$

is becoming smaller since $G(x) \succeq H(x)$ and $\mathcal{D}H(x)[h] = \mathcal{D}G(x)[h]$. The rigorous version of this instinct is shown in Lemma F.1.

Our proof used the fact that the unregularized logarithmic metric $H(x) = A_x^\top A_x$ is strongly self-concordant if $H(x)$ is invertible Laddha et al. (2020), and also the fact that $H(x)$ is SLTSC when $H(x)$ is invertible Kook and Vempala (2024). This is summarized in Lemma F.2. However, in our case, we do not restrict $H(x)$ to be invertible since we allow $K$ to be unbounded and $m < n$. In order to prove $G(x) = H(x) + \lambda I$ is SSC and LTSC, we would use a limit argument: we add artificial constraints so that $H(x) := A_x A_x^\top$ is invertible so we can apply Lemma F.2 and the additivity of LTSC as in Proposition D.6, thus $G(x) = H(x) + \lambda I$ is also SSC and LTSC, then we take the limit and remove all artificial constraints.

**Lemma F.1** *If $A, B, C \in \mathbb{R}^{n\times n}$ are a symmetric matrices, and if $0 \preceq A \preceq C$, then we have*

$$\left\| A^{\frac{1}{2}} B A^{\frac{1}{2}} \right\|_F \leq \left\| C^{\frac{1}{2}} B C^{\frac{1}{2}} \right\|_F .$$

**Proof** Note that $\mathrm{Tr}\left\{ P^{\frac{1}{2}} BPBP^{\frac{1}{2}} \right\} = \left\| P^{\frac{1}{2}} BP^{\frac{1}{2}} \right\|_F^2$ for any symmetric PSD matrix $P$. So we only need to prove $\mathrm{Tr}\left\{ A^{\frac{1}{2}} BABA^{\frac{1}{2}} \right\} \leq \mathrm{Tr}\left\{ C^{\frac{1}{2}} BCBC^{\frac{1}{2}} \right\}$. This is easy to verify:

$$\mathrm{Tr}\left\{ A^{\frac{1}{2}} BABA^{\frac{1}{2}} \right\} \leq \mathrm{Tr}\left\{ A^{\frac{1}{2}} BCBA^{\frac{1}{2}} \right\} = \mathrm{Tr}\left\{ C^{\frac{1}{2}} BABC^{\frac{1}{2}} \right\} \leq \mathrm{Tr}\left\{ C^{\frac{1}{2}} BCBC^{\frac{1}{2}} \right\}.$$

■

**Lemma F.2 (Lemma 4.1 Laddha et al. (2020), Lemma E.1 Kook and Vempala (2024))** *Let $K = \{x | Ax > b\}$ be a convex polytope in $\mathbb{R}^n$ and the logarithmic metric $H(x) := A_x^\top A_x$ is invertible for all $x \in K$. Then we have $H$ is SSC and SLTSC on $K$.*

The remaining task in this section is to verify soft-threshold metric $G$ defined in Definition 1.3 is ASC. This is proved by concentration of Gaussian polynomials, which also appears in Sachdeva and Vishnoi (2016), and the intuition is that adding a regularization term $\lambda I$ only makes the Gaussian concentration tighter. Lemma F.3 is a general bound for Gaussian polynomials, and Lemma F.4 is upper bounding the expectation of specific Gaussian polynomials that appears in our proof. The condition $\sum_{i=1}^m b_i b_i^\top = I$ in the original Lemma F.4 is changed to $\sum_{i=1}^m b_i b_i^\top \preceq I$ due to our regularization, and the proof also changes. To be rigorous, we list the proof of Lemma F.4 here.

**Lemma F.3 (Theorem 6.7 from Janson (1997))** *Let $P$ be a degree $q$ polynomial over $\mathbb{R}^n$, and $\xi \sim \mathcal{N}(0, I_n)$. Then for any $t \geq (2e)^{\frac{q}{2}}$ we have*

$$\mathbb{P}\left[ |P(\xi)| \geq t \left( \mathbb{E}P(\xi)^2 \right)^{1/2} \right] \leq \exp\left( -\frac{q}{2e} t^{2/q} \right) \tag{66}$$

**Lemma F.4 (Adapted from Fact 10 in Sachdeva and Vishnoi (2016))** *Suppose $\xi \sim \mathcal{N}(0, I_n)$ and $b_i \in \mathbb{R}^n$ are vectors for $i \in [m]$ such that $\sum_{i=1}^m b_i b_i^\top \preceq I_n$, then we have the following bounds:*

$$\mathbb{E}\left\{ \left[ \sum_{i=1}^m (b_i^\top \xi)^3 \right]^2 \right\} \leq 15n, \quad \mathbb{E}\left\{ \left[ \sum_{i=1}^m (b_i^\top \xi)^4 \right]^2 \right\} \leq 105n^2,$$

**Proof** Same as in **Fact 10** in Sachdeva and Vishnoi (2016), we get

$$\mathbb{E}\left\{\left[\sum_{i=1}^{m}(b_i^\top \xi)^3\right]^2\right\} = 9\sum_{i,j=1}^{m}(b_i^\top b_i)(b_j^\top b_j)(b_i^\top b_j) + 6\sum_{i,j=1}^{m}(b_i^\top b_j)^3,$$

Following the notations in Sachdeva and Vishnoi (2016), we set $B$ to be the $m \times n$ matrix with its $i$-th row being $b_i^\top$, and $w \in \mathbb{R}^m$ be such that $w_i = b_i^\top b_i$. The first term is simplified to:

$$\sum_{i,j=1}^{m}(b_i^\top b_i)(b_j^\top b_j)(b_i^\top b_j) = w^\top BB^\top w,$$

since we assumed $\sum_{i\in[m]} b_i b_i^\top \preceq I_n$, thus $B^\top B \preceq I_n$. Since all non-zero eigenvalues of $BB^\top$ are the same with $B^\top B$, thus $BB^\top \preceq I_m$. As a result, $w^\top BB^\top w \leq w^\top w$. The remaining arguments in Sachdeva and Vishnoi (2016) go on smoothly, which we omit here. ■

With all the preparations, we now give the proof of Lemma D.1.

**Proof** [Proof of Lemma D.1]

We define the two following local metrics for any $\gamma > 0$. Fix $x \in K$, we define:

$$\begin{aligned} H^{(\gamma)}(x) &:= \sum_{i=1}^{m}\frac{a_i a_i^\top}{(a_i^\top x - b_i)^2} + \sum_{j=1}^{n}\frac{e_j e_j^\top}{(e_j^\top x - \gamma)^2}, \\ G^{(\gamma)}(x) &:= H^{(\gamma)}(x) + \lambda I_n, \end{aligned}$$

where $e_j$ is the unit vector in the $j$-th direction for $j \in [n]$. We add the term $\sum_j \frac{e_j e_j^\top}{(e_j^\top x - \gamma)^2}$ to ensure that $H^{(\gamma)}(x)$ is invertible, so we can use the strong self-concordance of unregularized logarithmic metric. It is also clear that fix $x \in K$, $\lim_{\gamma\to\infty} H^{(\gamma)}(x) = H(x)$ and $\lim_{\gamma\to\infty} G^{(\gamma)}(x) = G(x)$. So we have

$$\begin{aligned} \left\|G^{(\gamma)}(x)^{-\frac{1}{2}}\mathcal{D}G^{(\gamma)}(x)[h]G^{(\gamma)}(x)^{-\frac{1}{2}}\right\|_F &= \left\|G^{(\gamma)}(x)^{-\frac{1}{2}}\mathcal{D}H^{(\gamma)}(x)[h]G^{(\gamma)}(x)^{-\frac{1}{2}}\right\|_F \\ &\overset{(i)}{\leq} \left\|H^{(\gamma)}(x)^{-\frac{1}{2}}\mathcal{D}H^{(\gamma)}(x)[h]H^{(\gamma)}(x)^{-\frac{1}{2}}\right\|_F \\ &\overset{(ii)}{\leq} 2\left\|h\right\|_{H^{(\gamma)}(x)} \leq 2\left\|h\right\|_{G^{(\gamma)}(x)}, \end{aligned}$$

where inequality $(i)$ holds due to $G^{(\gamma)}(x) \succeq H^{(\gamma)}(x)$ and Lemma F.1, and inequality $(ii)$ holds because $H^{(\gamma)}(x)$ is a unregularized logarithmic metric, so is strongly self-concordant. It is clear that for any $x \in K$, $G^{(\gamma)}(x) \to G(x)$ as $\gamma \to \infty$, so we take limits on both sides, and we have

$$\left\|G(x)^{-\frac{1}{2}}\mathcal{D}G(x)[h]G(x)^{-\frac{1}{2}}\right\|_F \leq 2\left\|h\right\|_{G(x)},$$

so we just proved SSC of $G$. Using the same technique, we can prove $G$ is LTSC. We know from Lemma F.2 that $H^{(\gamma)}$ is SLTSC for every $\gamma > 0$, then according to Proposition D.6, $G^{(\gamma)} := H^{(\gamma)} + \lambda I$ is LTSC since $\lambda I$ is clearly SLTSC. So fix any $x \in K$ and $h \in \mathbb{R}^n$ we have

$$\mathrm{Tr}\left\{G^{(\gamma)}(x)\mathcal{D}^2 G^{(\gamma)}(x)[h,h]\right\} \geq -\left\|h\right\|^2_{G^{(\gamma)}(x)},$$

and take the limit $\gamma \to 0$ on both sides, we conclude that $G$ is LTSC.

The proof follows similar steps in Sachdeva and Vishnoi (2016), except that we define the soft-threshold metric in Definition 1.3, so we allow the polytope $K$ to be unbounded, thus $A$ is not necessarily full-rank and $m$ could be smaller than $n$.

We set $\hat{a}_i = \frac{G(x)^{-\frac{1}{2}} a_i}{a_i^\top x - b_i}$, and noticing that $z = x + \frac{r}{\sqrt{n}} G(x)^{-\frac{1}{2}}\xi$, we define term I to be:

$$\begin{aligned}
\mathrm{I} &:= \frac{n}{r^2}\left(\|z-x\|^2_{G(z)} - \|z-x\|^2_{G(x)}\right)\\
&= \xi^\top G(x)^{-\frac{1}{2}}\left[G(z) - G(x)\right] G(x)^{-\frac{1}{2}}\xi\\
&= \xi^\top G(x)^{-\frac{1}{2}}\left\{\sum_{i=1}^m \frac{a_i a_i^\top}{(a_i^\top x - b_i)^2}\left[\frac{1}{(1+\frac{r}{\sqrt{n}}\hat{a}_i^\top\xi)^2} - 1\right]\right\} G(x)^{-\frac{1}{2}}\xi\\
&= \sum_{i=1}^m (\hat{a}_i^\top\xi)^2\left[\frac{1}{(1+\frac{r}{\sqrt{n}}\hat{a}_i^\top\xi)^2} - 1\right]\\
&= \frac{r^2}{n}\sum_{i=1}^m (\hat{a}_i^\top\xi)^4\left[\frac{2}{1+\frac{r}{\sqrt{n}}\hat{a}_i{}^\top\xi} + \frac{1}{(1+\frac{r}{\sqrt{n}}\hat{a}_i{}^\top\xi)^2}\right] - 2\frac{r}{\sqrt{n}}\sum_{i=1}^m(\hat{a}_i^\top\xi)^3.
\end{aligned}$$

We notice that $\sum_{i=1}^m \hat{a}_i\hat{a}_i^\top = G(x)^{-\frac{1}{2}}H(x)G(x)^{-\frac{1}{2}} \preceq I_n$, thus we can apply Lemma F.4. Let $P_1(\xi) = \sum_{i=1}^m(\hat{a}_i^\top\xi)^3$, and $P_2(\xi) = \sum_{i=1}^m(\hat{a}_i^\top\xi)^4$, then we define the following event:

$$\mathcal{E}_0 := \left\{\xi \middle| \, |P_1(\xi)| \geq t\sqrt{15n}, \quad |P_2(\xi)| \geq t\sqrt{105n^2}\right\}.$$

Then we can apply the concentration inequality of Gaussian polynomials Lemma F.3, for any $t > (2e)^2$, we have

$$\mathbb{P}_\xi(\mathcal{E}_0) \leq \exp\left(-\frac{3}{2e}t^{\frac{2}{3}}\right) + \exp\left(-\frac{4}{2e}t^{\frac{1}{2}}\right) \leq \epsilon, \tag{67}$$

where the inequality follows since we set $t = \max\left\{(2e)^2, \left(\frac{e}{2}\log(\frac{2}{\epsilon})\right)^2, \left(\frac{2e}{3}\log\left(\frac{2}{\epsilon}\right)\right)^{\frac{3}{2}}\right\}$. Moreover, for any $\xi \in \mathcal{E}_0$, we have for any integers $i \in [m]$:

$$\left|\frac{r}{\sqrt{n}}\hat{a}_i^\top\xi\right| \leq \frac{r}{\sqrt{n}}\left[\sum_{j=1}^m(\hat{a}_j^\top\xi)^4\right]^{1/4} \leq \frac{r}{\sqrt{n}}\left(t\sqrt{105n^2}\right)^{\frac{1}{4}} \leq rt^{\frac{1}{4}}\cdot(105)^{\frac{1}{8}} < \frac{1}{2}$$

The last inequality holds since we can set $r \leq \frac{1}{2t^{\frac{1}{4}}(105)^{\frac{1}{8}}}$. In conclusion, with probability greater than $1-\epsilon$, we have

$$\begin{aligned}
|\mathrm{I}| &\leq \frac{r^2}{n}\sum_{i=1}^m(\hat{a}_i^\top\xi)^4\left|\frac{2}{1+\frac{r}{\sqrt{n}}\hat{a}_i{}^\top\xi} + \frac{1}{(1+\frac{r}{\sqrt{n}}\hat{a}_i{}^\top\xi)^2}\right| + 2\frac{r}{\sqrt{n}}\left|\sum_{i=1}^m(\hat{a}_i^\top\xi)^3\right|\\
&\leq \frac{r^2}{n}\sum_{i=1}^m 8\cdot(\hat{a}_i^\top\xi)^4 + 2\frac{r}{\sqrt{n}}\left|\sum_{i=1}^m(\hat{a}_i^\top\xi)^3\right|\\
&\leq 8\frac{r^2}{n}t\sqrt{105n^2} + 2\frac{r}{\sqrt{n}}t\sqrt{15n} \leq 8\sqrt{105}tr^2 + 2\sqrt{15}tr \leq 2\epsilon,
\end{aligned}$$

where the last inequality holds as long as we set $r \le \min\left\{1, \frac{1\epsilon}{4\sqrt{105}+\sqrt{15}} \cdot \frac{1}{t}\right\}$. ■

## Appendix G. Warm Initialization & Per-step Complexity

In this section, we talk about the computational complexity for each step of Markov transition, and we also design a feasible warm start.

### G.1. Algebraic Complexity of Each Iteration

For per-step complexity, we are mainly interested in the algebraic complexity of each step. In other words, we assume we can do exact addition, subtraction, multiplication, division over $\mathbb{R}$. Since we need to compute the decomposition $G(x) = Q^\top Q$, we also assume we can compute the exact square root $\sqrt{x}$ for all $x \in \mathbb{R}^+$. In our model, each of the five arithmetic operations $\{+, -, \times, \div, \sqrt{\phantom{x}}\}$ has a unit cost $O(1)$.

We first list a simple lemma, arguing that drawing from a uniform ellipsoid $E(x, G(x), r)$ can be reduced to drawing from the unit ball and computing the decomposition $G(x) = Q^\top Q$.

**Lemma G.1** *Assume $\xi \sim \mathcal{N}(0, I_n)$ is drawn from the standard Gaussian distribution, fix any invertible matrix $Q \in \mathbb{R}^{n\times n}$ satisfying $G(x) := Q^\top Q$, the new random vector $Z$ defined by $Z := x + \frac{r}{\sqrt{n}} Q^{-1}\xi$ satisfies $Z \sim \mathcal{N}(x, \frac{r^2}{n} G(x)^{-1})$.*

For each iteration in Algorithm 1, given current state $x$, using the result in Lemma G.1, we need to do the two following steps:

1. draw $\xi \sim \mathcal{N}(0, I_n)$, compute $z = x + \frac{r}{\sqrt{n}} Q^{-1}\xi$, where $Q \in \mathbb{R}^{n\times n}$ is any invertible matrix satisfying $G(x) = Q^\top Q$.
2. given $z$, compute $\det G(z)$.

It costs $O(n)$ to draw $\xi$ since each component can be drawn i.i.d from one-dimensional standard Gaussian distribution. It is worth mentioning that when analyzing the mixing times of our Markov Chains, we just set $Q := G(x)^{\frac{1}{2}}$, this is a legitimate assignment due to the uniqueness of the square root of $G(x)$. i.e., for any PSD & symmetric matrix $C$, there exists a unique PSD & symmetric matrix $B$ such that $B^2 = C$. However, using only $\{+, -, \times, \div, \sqrt{\phantom{x}}\}$ over (nonnegative) real numbers, we may not compute $G(x)^{\frac{1}{2}}$ exactly. Because this involves computing the eigenvalues of $G(x)$ and the corresponding eigenvectors, and further needs us to exactly solve a polynomial equation of order $n$ over $\mathbb{R}$, which can not be done using these basic arithmetic operations.

We may write $G(x) := cA_x^\top W_x A_x + \lambda I$, where $c > 0$ is some scalar irrelevant to $x$ and $W_x \in \mathbb{R}^{m\times m}$ is a diagonal matrix changing with $x$. For the soft-threshold metric $G$ as defined in Definition 1.3, $c \equiv 1$ and $W_x \equiv I_m$. For the regularized Lewis metric $G$ in Definition 1.4, $c := c_1\sqrt{n}(\log m)^{c_2}$ and $W_x$ is the ridge-Lewis weights as defined in Eq. (3). For now we assume that $w_x$ is known exactly for each $x$, thus $c$ and $W_x$ are known exactly (Later we would discuss a high-precision solver of $w_x$).

We first discuss how to compute $G(x)$ efficiently. We can compute $S_x := \mathrm{Diag}(Ax - b)$ in $O(mn)$ arithmetic operations, then we can compute $A_x := S_x^{-1} A$ in $O(mn)$ arithmetic operations since $S_x$ is diagonal. We then attempt to compute $A_x^\top A_x$ using fast matrix multiplication, if $m \le n$, $A_x^\top A_x$ can be computed in $O(n^\omega)$ operations by filling $A_x$ into a $n \times n$ matrix with 0-entries. Otherwise $A_x^\top A_x$ can be computed in $O(mn^{\omega-1})$ by partitioning $A_x$ into $\lfloor \frac{m}{n} \rfloor$ square matrices. Now we get $G(x)$, and we came across $O(\max\{m,n\} n^{\omega-1})$ arithmetic operations.

Given $G(x)$ as an $n \times n$ matrix, we need to compute an invertible matrix $Q$ such that $G(x) = Q^\top Q$. Since $G(x)$ is symmetric, we first compute an invertible matrix $V \in \mathbb{R}^{n\times n}$ such that $\Lambda = VG(x)V^\top$ for some diagonal matrix $\Lambda$, and this can be done in $O(n^\omega)$ (see Chapter 16.8 about **orthogonal basis transform** in Bürgisser et al. (1997)). Then $Q := \sqrt{\Lambda}\left(V^\top\right)^{-1}$ is the desired invertible matrix, where $Q$ is invertible since $G(x)$ is positive definite, thus $\Lambda \succ 0$. Finally, $Q$ can be computed in $O(n^\omega)$, because computing the square root $\sqrt{\Lambda}$ takes $O(n)$ square-root operations, and computing the inverse of $V^\top \in \mathbb{R}^{n\times n}$ takes $O(n^\omega)$ operations (see Chapter 16.4 about **matrix inversion** in Bürgisser et al. (1997)).

Finally, given $G(x) \in \mathbb{R}^{n\times n}$, we can compute its determinant $\det G(x)$ in $O(n^\omega)$ operations (see Chapter 16.4 about **determinant** in Bürgisser et al. (1997)). Combining all these arithmetic operations together, given that $W_x$ and $E_x$ is known exactly, which is the case for soft-threshold metric $G$, the per-step arithmetic cost is $O\left(\max\{m,n\} n^{\omega-1}\right)$.

For the regularized Lewis metric, Fazel et al. (2022) proposed a quasi-Newton algorithm to compute an $\epsilon$-approximation of the Lewis weights $w_x$ in Eq.(3) using **polylog** $\left(\frac{1}{\epsilon}\right)$ steps of gradient descent, where each descent involves computing leverage scores that costs $O(mn^{\omega-1})$. So the per-step arithmetic cost for regularized Dikin walk using regularized Lewis metric is $\widetilde{O}(\max\{m,n\} n^{\omega-1})$ if we ignore logarithmic factors.

## G.2. Uniform Ball as Warm Initialization

In this section, we discuss how to compute a ball $\mathbb{B}(x_0, r_0)$ such that its uniform distribution is a suitable warm start for the truncated distribution defined in Eq. (1). In addition to the assumptions in Eq. (1), we further assume that $K$ is bounded in a ball of radius $\widetilde{R}$, and $K$ contains a ball of radius $\widetilde{r}$. We list the warmness bound as Lemma G.2.

**Lemma G.2** *Let $\Pi$ be a distribution on $\mathbb{R}^n$ with density $\pi(x) \propto \mathbf{1}_K(x)e^{-f(x)}$, where $K$ is a polytope with $m$ linear constraints, and $f$ is $\alpha$-convex and $\beta$-smooth with condition number $\kappa = \beta/\alpha$, as in Eq. (1).*

*If we further assume that there exist two balls such that $\mathbb{B}(x_1, \widetilde{r}) \subseteq K \subseteq \mathbb{B}(x_2, \widetilde{R})$, then there exists a ball $\mathbb{B}(x_0, r_0)$ such that its uniform distribution is a $M$-warm with respect to $\Pi$, where $M$ satisfies*

$$\log M \le 1 + n \log \frac{3\widetilde{R}}{\widetilde{r}} + n \cdot \max\left\{\frac{1}{2}\log\left(\beta\widetilde{R}^2\right), \log\left(2\beta\widetilde{R}\left\|x^\dagger - x^\star\right\|_2\right)\right\}.$$

*The radius $r_0$ and center $x_0$ of the ball can be computed by*

$$r_0 := \frac{r_1 \widetilde{r}}{\|x_1 - x^\dagger\|_2 + \widetilde{r}}, \quad x_0 := x^\dagger + \frac{r_1}{\widetilde{r} + \|x_1 - x^\dagger\|_2}(x_1 - x^\dagger),$$

*where $x^\dagger := \arg\min_K f(x)$, $x^\star := \arg\min_{\mathbb{R}^n} f(x)$, and $r_1 := \min\left\{\frac{1}{\sqrt{\beta}}, \frac{1}{2\beta\left\|x^\dagger - x^\star\right\|_2}\right\}$.*

**Proof** For convenience, we use $B$ to denote $\mathbb{B}(x_0, r_0)$ in this section. Our idea is to make $B$ close to the mode $x^\dagger = \arg\min_K f(x)$ within the polytope. We first determine a radius $r_1$, such that inside $\mathbb{B}(x^\dagger, r_1)$, the function $e^{-f}$ shrinks no more than a constant factor $C \geq 1$. In other words, for any $x \in \mathbb{B}(x^\dagger, r_1)$, we require:

$$\exp\left[-f(x)\right] \geq \frac{1}{C} \exp\left[-f(x^\dagger)\right],$$

for ease of computation, we choose $C$ to be $e$ here. We then require $B$ to be contained in $\mathbb{B}(x^\dagger, r_1) \cap K$. Under this requirement, the $M$-warmness of the uniform distribution $\mu_0$ over $B := \mathbb{B}(x_0, r_0)$ can be bounded by:

$$\begin{aligned}
\frac{\mu_0(U)}{\Pi(U)} \leq \frac{\mu_0(U)}{\Pi(U \cap B)} &= \frac{\mathrm{vol}(U \cap B)}{\mathrm{vol}(B)} \cdot \frac{\int_K e^{-f(z)} dz}{\int_{K \cap U \cap B} e^{-f(z)} dz} \\
&\overset{(i)}{=} \frac{\mathrm{vol}(U \cap B)}{\mathrm{vol}(B)} \cdot \frac{\int_K e^{-f(z)} dz}{\int_{U \cap B} e^{-f(z)} dz} \\
&\overset{(ii)}{\leq} \frac{\mathrm{vol}(U \cap B)}{\mathrm{vol}(B)} \cdot \frac{e \,\mathrm{vol}(K)}{\int_{U \cap B} e^{-f(z)} dz} \frac{\int_{U \cap B} e^{-f(z)} dz}{\mathrm{vol}(U \cap B)} \\
&\leq \frac{e \,\mathrm{vol}(K)}{\mathrm{vol}(B)} \leq \frac{e \widetilde{R}^n}{r_0^n},
\end{aligned} \tag{68}$$

where equality $(i)$ holds since we assumed $B \subseteq K$ and inequality $(ii)$ results from our assumption $e^{-f(x)} \geq \frac{1}{e} \cdot e^{-f(x^\dagger)}$ for $x \in B \subseteq \mathbb{B}(x^\dagger, r_1)$, thus the relationship also holds for the average over $U \cap B$:

$$\frac{\int_{U \cap B} e^{-f(z) dz}}{\mathrm{vol}(U \cap B)} \geq \frac{1}{e} \cdot e^{-f(x^\dagger)} \geq \frac{1}{e} \frac{\int_K e^{-f(z)} dz}{\mathrm{vol}(K)}.$$

We now try to determine $r_1$ so that inside the ball $\mathbb{B}(x^\dagger, r_1)$, the function $e^{-f}$ shrinks less than a factor of $e$, and this translates to $f$ increases less than 1. Due to the $\beta$-smoothness of the function $f$, the change in $f$ can be controlled by:

$$f(y) - f(x^\dagger) \leq \left\|\nabla f(x^\dagger)\right\|_2 \left\|y - x^\dagger\right\|_2 + \frac{\beta}{2} \left\|y - x^\dagger\right\|_2^2, \tag{69}$$

thus we further control the gradient at $x^\dagger$ by mean value theorem, for some $t \in (0, 1)$, we have

$$\left\|\nabla f(x^\dagger)\right\|_2 \overset{(i)}{=} \left\|\nabla f(x^\dagger) - \nabla f(x^\star)\right\|_2 \leq \left\|\nabla^2 f(x^\star + t(x^\dagger - x^\star))\right\|_2 \left\|x^\star - x^\dagger\right\|_2 \leq \beta \left\|x^\star - x^\dagger\right\|_2,$$

where equality $(i)$ follows from $\nabla f = 0$ at the global mode $x^\star$. As a result, we can set $r_1$ to be:

$$r_1 := \min\left\{\sqrt{\frac{1}{\beta}}, \frac{1}{2\beta \left\|x^\dagger - x^\star\right\|_2}\right\}, \tag{70}$$

insert $r_1$ into Eq. (69), we easily check for any $y \in \mathbb{B}(x^\dagger, r_1)$, we have

$$f(y) - f(x^\dagger) \leq \beta \left\|x^\star - x^\dagger\right\|_2 r_1 + \frac{\beta}{2} r_1^2 \leq 1,$$

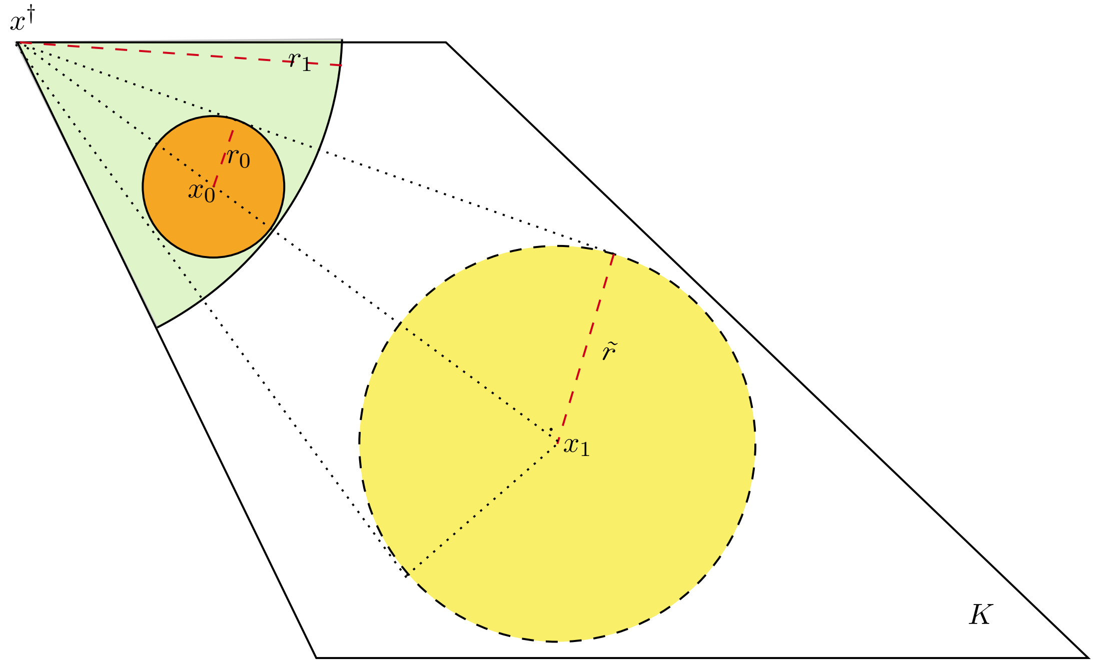


Figure 2: An example of warm-start $\mathbb{B}(x_0, r_0)$ for $K \subseteq \mathbb{R}^2$: Here the mode within the polytope $x^\dagger := \arg\min_K f(x)$ coincides the upper-left vertex of $K$. We need to ensure $\mathbb{B}(x_0, r_0) \subseteq \mathbb{B}(x^\dagger, r_1)$ and $\mathbb{B}(x_0, r_0) \subseteq K$, where the first condition reduces to $\left\|x_0 - x^\dagger\right\|_2 + r_0 \le r_1$, and the later is ensured by $\mathbb{B}(x_0, r_0)$ being inside the convex hull of $\{x^\dagger\} \cup \mathbb{B}(x_1, \widetilde{r})$, and $r_0$ can be easily computed by similarity of cones.

It is worth noting when the mode within the polytope $K$ coincides with the global mode, we have $\left\|x^\dagger - x^\star\right\|_2 = 0$, thus we have $\frac{1}{2\beta\left\|x^\dagger - x^\star\right\|_2} = \infty$ and $r_1 = \sqrt{\frac{1}{\beta}}$. $f(y) - f(x^\dagger) \le 1$ for $y \in \mathbb{B}(x^\dagger, r_1)$ still holds because $\nabla f(x^\dagger) = 0$ in Eq. (69).

After getting the ball $\mathbb{B}(x^\dagger, r_1)$, we require our initial distribution $B = \mathbb{B}(x_0, r_0)$ to be inside $\mathbb{B}(x^\dagger, r_1)$ so that the function $f$ only increases a constant. Moreover, we also want to make sure $\mathbb{B}(x_0, r_0)$ is inside the polytope. This can be ensured by our assumption that $K$ contains a ball of radius $\widetilde{r}$.

These two conditions can be ensured by the following two requirements by simple geometry relations (also see Figure 2 for illustration):

$$\left\|x_0 - x^\dagger\right\|_2 + r_0 \le r_1 \text{ , and } \frac{r_0}{\|x^\dagger - x_0\|_2} = \frac{\widetilde{r}}{\|x_1 - x^\dagger\|_2}.$$

where the first inequality ensures that $\mathbb{B}(x_0, r_0) \subseteq \mathbb{B}(x^\dagger, r_1)$, and the second inequality ensures that $\mathbb{B}(x_0, r_0)$ is included in the convex hull of $x^\dagger \cup \mathbb{B}(x_1, \widetilde{r})$, thus $\mathbb{B}(x_0, r_0) \subseteq K$. From these two inequalities, it is easy to get the following largest $r_0$ and the corresponding center $x_0$:

$$r_0 := \frac{r_1 \widetilde{r}}{\|x_1 - x^\dagger\|_2 + \widetilde{r}}, \quad x_0 := x^\dagger + \frac{r_1}{\widetilde{r} + \|x_1 - x^\dagger\|_2}(x_1 - x^\dagger).$$

The $M$-warmness of our initial distribution (over $\mathbb{B}(x_0, r_0)$) can be bounded by inserting $r_1$ into Eq. (68):

$$M \leq \frac{e\widetilde{R}^n}{r_0^n} = \frac{e\widetilde{R}^n \left(\left\|x_1 - x^\dagger\right\|_2 + \widetilde{r}\right)^n}{r_1^n \widetilde{r}^n} \overset{(i)}{\leq} \frac{e\widetilde{R}^n \left(3\widetilde{R}\right)^n}{r_1^n \widetilde{r}^n}$$
$$= e\left(\frac{3\widetilde{R}}{\widetilde{r}}\right)^n \max\left\{\left(\beta\widetilde{R}^2\right)^{n/2}, \left(2\beta\left\|x^\dagger - x^\star\right\|_2 \widetilde{R}\right)^n\right\},$$

where inequality $(i)$ holds since we $x_1, x^\dagger \in K$ and $\mathrm{diam}(K) \leq 2\widetilde{R}$. Since our sampling algorithm has high accuracy, the mixing time depends on $\log M$:

$$\log M = 1 + n\log\frac{3\widetilde{R}}{\widetilde{r}} + n\cdot\max\left\{\frac{1}{2}\log\left(\beta\widetilde{R}^2\right), \log\left(2\beta\widetilde{R}\left\|x^\dagger - x^\star\right\|_2\right)\right\}.$$

■

## Appendix H. Bounding $\bar{\nu}$-symmetric metrics by cross-ratio distance

In this section, we provide the proof of Proposition E.1 for the sake of rigor and completeness.
**Proof** [Proof of Proposition D.4]

The proof is similar to steps in Laddha et al. (2020), but we notice that the assumption $\|p - x\|_2 \leq \|y - q\|_2$ in Laddha et al. (2020) may not hold, and when the other side hold, we can only derive the conclusion of local metric at $y$ instead of at $x$. Moreover, since we are dealing with possibly unbounded polytopes, we also consider the case when $p$ or $q$ is at infinity.

Throughout the proof, we can assume that both $\|x - y\|_{H(x)}$ and $\|x - y\|_{H(y)}$ are positive. Otherwise, we have at least one of them to be 0, then the RHS of Eq. (36) is 0 so the inequality becomes trivial.

Assume we have the chord $[p, q]$ induced by the line $\overline{xy}$ in the order $p, x, y, q$.

- First, we consider the case when both $p, q$ are bounded. When $\|p - x\|_2 \leq \|y - q\|_2$, then we have $\|p - x\|_2 \leq \|x - q\|_2$, thus $x + x - p \in K$, so we have $p \in K \cap (2x - K)$. According to weak $\bar{\nu}-$symmetry of $H$, we have

  $$\|p - x\|_{H(x)} \leq \sqrt{\bar{\nu}}.$$

  So we have

  $$d_K(x, y) = \frac{\|x - y\|_2 \|p - q\|_2}{\|p - x\|_2 \|q - y\|_2} \geq \frac{\|x - y\|_2}{\|p - x\|_2}$$
  $$= \frac{\|x - y\|_{H(x)}}{\|p - x\|_{H(x)}} \geq \frac{\|x - y\|_{H(x)}}{\sqrt{\bar{\nu}}}.$$

  Next we consider the other case $\|y - q\|_2 < \|p - x\|_2$, then following the same derivation, we have

  $$d_K(x, y) = \frac{\|x - y\|_2 \|p - q\|_2}{\|p - x\|_2 \|q - y\|_2} \geq \frac{\|x - y\|_2}{\|q - y\|_2}$$
  $$= \frac{\|x - y\|_{H(y)}}{\|q - y\|_{H(y)}} \geq \frac{\|x - y\|_{H(y)}}{\sqrt{\bar{\nu}}}.$$

Combining the two results, and we come to the conclusion that in any cases, we have

$$d_K(x,y) \geq \frac{1}{\sqrt{\bar{\nu}}} \min\left\{\|x-y\|_{H(x)}, \|x-y\|_{H(y)}\right\}.$$

- Second, we consider the case when one and only one of $\{p,q\}$ is at infinity. Without loss of generality, we assume $q=\infty$, then we deduce that $x+t(y-x) \in K$ for all $t \geq 0$. Since $x-p$ and $y-x$ are in the same direction, thus $x+(x-p) \in K$, so we have $p \in K \cap (2x-K)$. According to the definition of $\bar{\nu}-$symmetry, $\|p-x\|_{H(x)} \leq \sqrt{\bar{\nu}}$, then we have

$$\begin{aligned} d_K(x,y) &= \frac{\|x-y\|_2}{\|p-x\|_2} = \frac{\|x-y\|_{H(x)}}{\|p-x\|_{H(x)}} \geq \frac{\|x-y\|_{H(x)}}{\sqrt{\bar{\nu}}} \\ &\geq \frac{1}{\sqrt{\bar{\nu}}} \min\left\{\|x-y\|_{H(x)}, \|x-y\|_{H(y)}\right\}. \end{aligned}$$

- Finally, we consider the case that both $p,q$ are at infinity, then $d_K(x,y)=0$ by the definition of cross-ratio distance (Definition A.2). In order to prove Eq. (36), it is adequate to prove $\|x-y\|_{H(x)} = 0$.

  Since both $p,q$ are at infinity, it implies that for all $t \in \mathbb{R}$, $tx+(1-t)y \in K$. By inserting $t := 2-l$ for $l \in \mathbb{R}$, it is easy to check that $tx+(1-t)y \in K \cap (2x-K)$ for $t \in \mathbb{R}$. Then due to our assumption of $\bar{\nu}$-symmetry of $H$, for all $t \in \mathbb{R}$ we have

$$|t-1| \cdot \|y-x\|_{H(x)} = \|tx+(1-t)y-x\|_{H(x)} \leq \sqrt{\bar{\nu}}.$$

  Due to the arbitrariness of $t$, we let $|t-1| \to +\infty$, since $\bar{\nu}$ is finite, so we have $\|y-x\|_{H(x)} = 0$.

■

## Appendix I. Isoperimetry for Weakly Logconcave Measures

In this section, we complete the omitted pieces needed for extending the new isoperimetry to weakly logconcave measures in Section C.2. In Appendix I.1, we show why we can assume the distribution is isotropic by an affine transformation. In Appendix I.2, we show that the combined metric of Euclidean and Hilbert metric actually induces Euclidean topology (Lemma C.6).

### I.1. Reduction to Isotropic Distributions

Let $\mu_\pi, \Sigma_\pi$ denote the mean and the covariance matrix of the log-concave probability distribution $\Pi$ defined in Equation (1). Now suppose the isoperimetric inequality in Lemma B.4 holds for isotropic log-concave distributions ($\Sigma_\pi = I_n$), we show that this implies Lemma B.4 for general covariance matrix $\Sigma_\pi$.

Since we assume $K$ is open, thus $\Sigma_\pi$ is an invertible matrix. We define the following bijective affine map $\mathcal{A}: \mathbb{R}^n \to \mathbb{R}^n$ by $\mathcal{A}(x) := \Sigma_\pi^{-\frac{1}{2}}(x-\mu_\pi)$. Now we let $\Pi^{\mathcal{A}}$ denotes induced measure of $\Pi$ under the mapping $\mathcal{A}$. In other words, for any Borel set $B$ in $K$, we define

$$\Pi^{\mathcal{A}}(B) := \Pi(\mathcal{A}^{-1}(B)). \tag{71}$$

It is easy to verify that the induced measure is logconcave with covariance matrix $I_n$. So for any measurable decomposition $K = S_1 \sqcup S_2 \sqcup S_3$, use the isoperimetric inequality for isotropic and logconcave distributions, we have

$$\Pi^{\mathcal{A}}\left(\mathcal{A}(S_3)\right) \geq c_n \cdot \mathfrak{d}_{\mathcal{A}}(\mathcal{A}(S_1), \mathcal{A}(S_2)) \min\left\{\Pi^{\mathcal{A}}(\mathcal{A}(S_1)), \Pi^{\mathcal{A}}(\mathcal{A}(S_2))\right\} \tag{72}$$

where $c_n$ denotes the constant $c_n := (6 \max\{1, \psi_n\})^{-1}$, and $\mathfrak{d}_{\mathcal{A}}$ is the mixed metric defined by:

$$\mathfrak{d}_{\mathcal{A}}(x, y) := \max\left\{(\log 2)\left\|y - x\right\|_2, d^{\mathcal{H}}_{\mathcal{A}(K)}(x, y)\right\}.$$

Thus Eq. (72) translates to

$$\begin{aligned}
\Pi(S_3) &\geq \inf_{(x,y)\in S_1\times S_2} c_n \cdot \mathfrak{d}_{\mathcal{A}}\left(\mathcal{A}x, \mathcal{A}y\right) \min\left\{\Pi(S_1), \Pi(S_2)\right\} \\
&\overset{(i)}{=} \inf_{(x,y)\in S_1\times S_2} c_n \cdot \max\left\{(\log 2)\left\|y - x\right\|_{\Sigma_\pi^{-1}}, d^{\mathcal{H}}_{\mathcal{A}(K)}(\mathcal{A}x, \mathcal{A}y)\right\} \min\left\{\Pi(S_1), \Pi(S_2)\right\} \\
&\overset{(ii)}{=} \inf_{(x,y)\in S_1\times S_2} c_n \cdot \max\left\{(\log 2)\left\|y - x\right\|_{\Sigma_\pi^{-1}}, d^{\mathcal{H}}_{K}(x, y)\right\} \min\left\{\Pi(S_1), \Pi(S_2)\right\} \\
&\overset{(iii)}{\geq} \inf_{(x,y)\in S_1\times S_2} c_n \cdot \max\left\{(\log 2)\frac{\left\|y - x\right\|_2}{\sqrt{\eta}}, d^{\mathcal{H}}_{K}(x, y)\right\} \min\left\{\Pi(S_1), \Pi(S_2)\right\},
\end{aligned} \tag{73}$$

where inequality $(i)$ holds since we have $\left\|\mathcal{A}x - \mathcal{A}y\right\|_2 = \left\|y - x\right\|_{\Sigma_\pi^{-1}}$ by definition of $\mathcal{A}$, and inequality $(ii)$ holds due to the affine invariance of Hilbert metric, and inequality $(iii)$ holds since $\Sigma_\pi \preceq \eta I_n$.

### I.2. The Mixed Metric induces Euclidean Topology

**Proof** [Proof of Lemma C.6]

Fix any $x, y \in K$, it is clear that $\mathfrak{d}(x, y) \geq 0$. If $\mathfrak{d}(x, y) = 0$, then $\left\|y - x\right\|_2 = 0$, so we have $x = y$. It is also clear that $\mathfrak{d}(x, y) = \mathfrak{d}(y, x)$ by definition. The remaining condition to ensure $\mathfrak{d}$ to be a metric is the triangle inequality. Fix $x, y, z \in K$, we first prove the triangle inequality for Hilbert metric $d^{\mathcal{H}}_K$, which is well-known for bounded open set $K$, see for example De La Harpe (1993). For unbounded $K$, let $\mathbb{B}_r$ denotes the ball in $\mathbb{R}^n$ centered around 0 with radius $r$, then we can prove it by a simple limit argument:

$$\begin{aligned}
d^{\mathcal{H}}_K(x, y) = \log\left(1 + d_K(x, y)\right) &\overset{(i)}{=} \lim_{r\to+\infty} \log\left(1 + d_{K\cap\mathbb{B}_r}(x, y)\right) \\
&= \lim_{r\to+\infty} d^{\mathcal{H}}_{K\cap\mathbb{B}_r}(x, y) \overset{(ii)}{\leq} \lim_{r\to+\infty}\left[d^{\mathcal{H}}_{K\cap\mathbb{B}_r}(x, z) + d^{\mathcal{H}}_{K\cap\mathbb{B}_r}(z, y)\right] \\
&= d^{\mathcal{H}}_K(x, z) + d^{\mathcal{H}}_K(z, y)
\end{aligned}$$

where equality $(i)$ holds due to our definition of cross-ratio distance for unbounded convex sets (see Definition A.2), inequality $(ii)$ holds by applying the triangle inequality for bounded open sets. Thus the triangle inequality for $\mathfrak{d}$ can be proved:

$$\begin{aligned}
\mathfrak{d}(x, y) &\leq \max\left\{\left\|x - z\right\|_2 + \left\|z - y\right\|_2, d^{\mathcal{H}}_K(x, z) + d^{\mathcal{H}}_K(y, z)\right\} \\
&\leq \max\left\{\left\|x - z\right\|_2, d^{\mathcal{H}}_K(x, z)\right\} + \max\left\{\left\|y - z\right\|_2, d^{\mathcal{H}}_K(y, z)\right\} \\
&= \mathfrak{d}(x, z) + \mathfrak{d}(z, y).
\end{aligned}$$

To prove that $\tilde{d}$ induces the same topology as the Euclidean distance $\|\cdot\|_2$, we notice that the identity map $x \mapsto x$ from $(K, \mathfrak{d})$ to $(K, \|\cdot\|_2)$ is clearly continuous, and we proceed to prove its inverse is also continuous.

Fix $x \in K$, then there exists $r_x > 0$ such that $\mathbb{B}(x, r_x) \subseteq K$ by the openness of $K$. Without loss of generality, we can always assume $y$ such that $\|y - x\|_2 < r_x$ Assume $\overline{xy}$ intersects $\partial K$ in the order $p, x, y, q$. Then we have $\|p - x\|_2 \geq r_x$ and $\|y - q\|_2 \geq r_x - \|y - x\|_2$. So we have

$$\mathfrak{d}(x, y) \leq \|y - x\|_2 + \log\left[1 + \frac{\|y - x\|_2}{r_x} + \frac{\|y - x\|_2}{r_x - \|y - x\|_2} + \frac{\|y - x\|_2^2}{r_x(r_x - \|y - x\|_2)}\right]. \tag{74}$$

Treat $r_x > 0$ as a fixed constant, then it is clear that when $\|y - x\|_2 \to 0$, the RHS of Eq. (74) converges to $0$. As a result, we proved the identity map is continuous in both direction, and we established homeomorphism between $(K, \mathfrak{d})$ and $(K, \|\cdot\|_2)$. ■