# Sampling and Optimization on Convex Sets in Riemannian Manifolds of Non-Negative Curvature

**Navin Goyal**                                           NAVINGO@MICROSOFT.COM
*Microsoft Research India*

**Abhishek Shetty**                                ASHETTY1995@GMAIL.COM
*Microsoft Research India*

**Editors:** Alina Beygelzimer and Daniel Hsu

## Abstract

The Euclidean space notion of convex sets (and functions) generalizes to Riemannian manifolds in a natural sense and is called geodesic convexity. Extensively studied computational problems such as convex optimization and sampling in convex sets also have meaningful counterparts in the manifold setting. Geodesically convex optimization is a well-studied problem with ongoing research and considerable recent interest in machine learning and theoretical computer science. In this paper, we study sampling and convex optimization problems over manifolds of non-negative curvature proving polynomial running time in the dimension and other relevant parameters. Our algorithms assume a warm start. We first present a random walk based sampling algorithm and then combine it with simulated annealing for solving convex optimization problems. To our knowledge, these are the first algorithms in the general setting of positively curved manifolds with provable polynomial guarantees under reasonable assumptions, and the first study of the connection between sampling and optimization in this setting.

**Keywords:** Riemannian Manifolds, Geodesic Convex Optimization, Simulated Annealing.

## 1. Introduction

### 1.1. Motivation

Convex geometry in the Euclidean space is a well-developed area of mathematics with connections to many fields. Purely from a geometric perspective, convex sets form a rich class of well-behaved sets with many fascinating properties. Convex optimization, which is a mainstay in the theory of optimization with far reaching applications, provides further motivation for the extensive study of convexity. The notion of convex relaxation has led to advances in our understanding of approximation algorithms. In algorithmic convex geometry, one is interested in constructing algorithms to answer natural questions about convex sets and functions. One important question in the area is sampling uniformly from convex sets and more generally from log-concave distributions. The problem of designing efficient sampling algorithms has led to connections to probability and geometry, raising interesting questions about isoperimetric inequalities and the analysis of Markov operators; see e.g. Vempala (2005) and Vempala (2010) for a survey of the area.

In the Euclidean space, there is an intimate connection between sampling, volume computation (more generally integration) and optimization. Though polynomial time deterministic volume computation for convex bodies is known to have exponential approximation ratio in the worst case, using efficient sampling, one can construct a fully polynomial time randomized approximation scheme for

volume computation Dyer et al. (1991); a long line of follow up work has further improved the time complexity. Similarly, the ability to sample efficiently can be used to design efficient optimization algorithms Bertsimas and Vempala (2004); Kalai and Vempala (2006).

A natural generalization of convexity is the notion of *geodesic convexity* on Riemannian manifolds (see e.g., Udriște (1994)). In addition to being motivated from a mathematical perspective, this notion enjoys applications in theoretical computer science and machine learning. Even in convex geometry in the Euclidean space, giving a Riemannian structure to the convex set has been instructive in understanding the behavior of algorithms such as the interior point methods (e.g. see Nesterov and Todd (2002); Lee and Vempala (2017, 2018)) and is in a sense, a natural way to approach the problem. That said, the theory of algorithmic convex geometry on Riemannian manifolds doesn't seem as well developed as its Euclidean counterpart and it is a natural direction to explore both for its intrinsic interest as well as for applications. In this paper, we ask the following question and take some modest steps towards answering it:

*Is there a theory of algorithmic convex geometry on manifolds? In particular, to what extent do the relations between sampling and optimization that hold in the Euclidean case carry over to manifolds?*

In the Euclidean case, most sampling algorithms are based on geometric random walks. Examples of these walks include the ball walk, where in each step one samples uniformly from the ball at the current point, rejecting the step if it lands outside the convex set and the Hit-and-Run walk, which in each step samples uniformly from a random chord through the current point. These algorithms are generally analyzed by relating their conductance to isoperimetric properties of the underlying set. A technique from convex geometry that has been instrumental in Euclidean convex geometry is that of localization. Localization is a technique used to prove dimension free inequalities by reducing the high dimensional problem to a one-dimensional problem by decomposing the space. The technique has been used to prove isoperimetric and related inequalities for convex bodies and more generally for log-concave measures and has been useful for the analysis of sampling algorithms. Until recently, the existence of a Riemannian analogue of the technique was unclear due to the lack of symmetries in the manifold setting. Klartag (2017) showed that localization technique does indeed extend to the setting of Riemannian manifolds with Ricci curvature lower bounds by exploiting connections to the theory of optimal transport. The connections between optimal transport on manifolds and isoperimetric inequalities had been considered earlier in the context of proving Prékopa–Leindler type inequalities on manifolds with Ricci curvature lower bounds Cordero-Erausquin et al. (2006). See Section C for an extended discussion.

A class of algorithms that relates sampling and optimization in the Euclidean space is the so called cutting plane methods. As an example, consider the center of gravity method which using sampling estimates the center of gravity and then uses a gradient oracle to restrict to a halfspace that contains the minimum. Using the Grünbaum inequality, one can see that each of the two parts of the convex set has a constant fraction of the volume of the convex body and thus each iteration reduces the volume of the set of interest by a constant fraction. It is natural to ask if this technique can be extended to manifolds. Rusciano (2018) proves a version of the Grünbaum inequality in the case of Hadamard manifolds, showing that there exist points such that any halfspaces through these points have large fraction of the volume of the set. Using this they show an oracle complexity result for optimization, where each oracle call returns a gradient at a center point. It remains unclear whether

these methods can be made efficient, due in part to the fact that it is unclear how the required center points can be found efficiently.

Another family of efficient convex optimization algorithms in the Euclidean space is the interior point methods. These algorithms construct barrier functions for the convex sets of interest and use these functions to reduce a constrained optimization problem to an unconstrained optimization problem. As noted earlier, this can be seen as optimizing on a Riemannian manifold with the Hessian structure induced by the barrier function. This connection has been fruitful in constructing optimization algorithms and has also provided motivation for considering optimization on Riemannian manifolds. For optimization on convex sets in Riemannian manifold, one could ask if interior point methods have a natural analog. To our knowledge, there doesn't exist any general construction of this nature. Recently, Abernethy and Hazan (2016) show that the interior point methods and simulated annealing techniques are related in the Euclidean setting: for the universal barrier on the convex body the central path and the heat path of the simulated annealing algorithm coincide suggesting a randomized algorithm for using the barrier using sampling. This connection further motivates studying simulated annealing on manifolds with the hope that this will shed further light on interior point methods in the manifold setting.

On Riemannian manifolds, there exists a natural notion of uniform sampling analogous to the notion of the Lebesgue measure in the Euclidean space. This is given by the Riemannian volume form (Definition 19). With this in mind, it makes sense to ask for random walks that have the uniform distribution as the stationary distribution. It is instructive to note that, in general, the natural ball walk on a compact manifold does not have as its stationary distribution the uniform distribution on the manifolds. By the natural ball walk, we mean the walk where at each step we sample uniformly (i.e. from the measure induced on the ball from the natural Riemannian volume form) from the (geodesic) ball of a fixed radius and move to the sampled point. The issue is that, on manifolds, the volume of balls varies with the center and the stationary distribution of the walk is proportional to this volume. In order to get the right distribution as the stationary point, one can use a Metropolis filter but that requires the knowledge of the volume at each point in order to execute the walk. In the setting of manifolds, it is even unclear how to sample uniformly from the ball at each point as required by the ball walk.

In this paper, we consider manifolds with positive sectional curvature. Note that the positivity of the sectional curvature is a stronger constraint than the positivity of the Ricci curvature, since the former implies the latter. Examples of these include the sphere in $n$ dimensions, the set of orthogonal matrices and Lie groups with bi-invariant metrics. Manifolds of positive curvature have been extensively studied in differential geometry. Their geometry and topology is well understood with characterization theorems such as the Cheeger–Gromoll splitting theorem, Bonnet–Myers theorem and the soul theorem. Their geometry can also be understood using comparison theorems which allow us to relate geometric quantities (such as Jacobi fields) about these manifolds to the corresponding quantities in the Euclidean space (which has zero curvature). The notion of convexity on manifolds has also received attention. For example, Yau (1974) proves the non-existence of convex functions on compact manifolds, which requires us to restrict attention to convex subsets of manifolds of our interest. Optimization on these groups has been considered in literature for various applications such as independent component analysis; for example, see Plumbley (2004).

Sampling and optimization on geodesically convex sets has several applications. For example, recently the geodesically convex optimization on the positive semidefinite matrices has been applied to capacity and operator scaling problems. Closely related is the application of geodesically convex

optimization to the Brascamp–Lieb constants Sra et al. (2018). There are many applications of geodesically convex optimization in machine learning and other areas; for example see Zhang and Sra (2016) for a discussion of the applications. We note however that many of these applications involve manifolds with nonpositive curvature and hence do not fall under our setting.

## 1.2. Previous Work

Literature on sampling algorithms is rich and diverse. Developments in the field have led to several deep results in probability and computer science. For sampling from convex sets, many walks such as the grid walk, the ball walk and the Hit-and-Run walk have been analyzed. Dyer et al. (1991) provide an approximation scheme for the volume of a convex body using sampling, which further motivated the study of sampling on a convex body. Lovász and Simonovits (1990) and Lovász and Simonovits (1993) study mixing time of walks such as the ball walk using localization and isoperimetry and these results were further improved by Kannan et al. (1995). Lovász (1999) and Lovász and Vempala (2006) analyze the hit-and-run walk and show that it has better mixing properties compared to the ball walk. Lovász and Vempala (2007) analyze the sampling from log-concave distributions using the ball and the hit-and-run walk. Lee and Vempala (2017), Lee and Vempala (2018), Mangoubi and Smith (2017), Mangoubi and Vishnoi (2018) and Lee et al. (2018) provide further improvements on sampling uniformly from convex sets and from logconcave densities.

There has been recent interest in sampling on manifolds. Diaconis et al. (2013) surveys methods for sampling on manifolds with emphasis on statistical applications. Lebeau and Michel (2010) analyzes the spectrum of the kernel of the natural ball walk and of the Metropolis adjusted walk on manifolds. They show that the convergence of both the walks to the stationary distributions in terms of the eigenvalues of the Laplace–Beltrami operator on the manifold. As noted earlier, in the setting of manifolds, the natural ball walk does not converge to the uniform distribution.

Mangoubi and Smith (2018) shows that the geodesic walk on manifolds with positive sectional curvature mixes rapidly. Indeed, they show that the walk mixes in time independent of the dimension of the manifold. The metric they consider for comparison with the stationary distribution is the Wasserstein distance, which is the natural choice for their analysis. In their analysis, they use the fact that on manifolds of positive curvature, geodesics that are initially "parallel" tend to move towards each other, a fact captured by the Rauch comparison theorem. As an example of this consider the sphere and two lines of longitude starting at the equator. As we move up towards the north pole, the distance between the end point of the arc reduces till the lines finally intersect at the north pole. This contraction property can easily be converted to statement about average distances, which implies mixing in the Wasserstein metric. They use the sampling algorithm thus obtained to sample from the surface of a convex body using a walk whose mixing time does not depend on the ambient dimension.

Lee and Vempala (2017) consider the problem of sampling from a Riemannian manifold whose metric is specified by the Hessian of a convex function, with the aim to provide faster sampling times for sampling uniformly from a polytope. The motivation for considering the Hessian structure on a convex set is motivated from the theory of interior point methods where the Riemannian structure induced by the barrier function acts as a natural lens through which to view the optimization technique. Lee and Vempala (2017) show that a walk similar to the geodesic walk that we consider mixes rapidly on Hessian manifolds. They use this to obtain a faster algorithm for sam-

pling from a polytope. They state their isoperimetry theorems for Hessian manifolds stated in terms of parameters which are bounded for the particular choice of convex function.

Another perspective on the geodesic walk is as Hamiltonian dynamics on the manifolds with constant potential. This leads to connection to Hamiltonian Monte Carlo, which is a well-studied technique in Markov chain Monte Carlo first considered in Duane et al. (1987). The Hamiltonian dynamics was original considered in physics as an alternate description of classical mechanics. The idea of Hamiltonian Monte Carlo is to sample paths that are solutions of the Hamiltonian equations. In Hamiltonian Monte Carlo, the Hamiltonian dynamics is used to produce a Markov chain whose stationary distribution is the required distribution. For a survey of the area, see Neal (2011) and Betancourt (2017). The Riemannian version of the Hamiltonian Monte Carlo was first considered in Girolami and Calderhead (2011). In a followup work to Lee and Vempala (2017), Lee and Vempala (2018) show convergence guarantees for the walk towards the stationary distribution which is set to be a Gibbs distribution on a manifold. Here too they show, isoperimetric theorems for Hessian manifolds specified by convex functions which has a convex second derivative. They then apply to adapt the Gaussian cooling algorithm on the Hessian manifold induced by the logarithmic barrier to compute volume on polytopes.

Though optimization of geodesically convex functions on manifolds is a problem that has been considered for a long time, it remains an active area of research, e.g. see Udriṣte (1994), Rapcsák (1997) and Absil et al. (2008). It has received attention recently in data science, machine learning and theoretical computer science. For example, recent work on operator scaling and the Brascamp Lieb inequality considers geodesically convex optimization on the manifold of positive semidefinite matrices; e.g. Allen-Zhu et al. (2018), Sra et al. (2018) and Vishnoi (2018). Optimization techniques based on adapting descent algorithms such as gradient descent, accelerated gradient descent and Newton's method have been analyzed in the manifold setting. For an overview of descent based algorithms, see Zhang and Sra (2016), Zhang and Sra (2018), Zhang and Zhang (2018) and references therein. Though analyses of descent algorithms on manifolds have been considered for some time, for example see Udriṣte (1994), most results concerned asymptotic convergence. Zhang and Sra (2016) consider unconstrained optimization on Hadamard manifolds, which are complete, simply connected manifolds with non-positive sectional curvature, and show bounds on the convergence rates of various first order algorithms such as gradient descent and stochastic gradient descent. Allen-Zhu et al. (2018) consider a particular geodesically convex optimization problem on the positive semidefinite cone (seen as a manifold) arising from operator scaling and show exponential rate of convergence of a box-constrained Newton's method using specific properties of the objective function. Zhang and Sra (2018) analyze accelerated methods on convex sets on manifolds but assume that successive iterations remain within the convex set. Zhang et al. (2018) analyze an accelerated stochastic gradient based algorithm on Riemannian manifolds. One advantage that these gradient based methods enjoy is that their rates are independent of the dimension. But when doing constrained optimization on geodesically convex sets they require "projection" to the convex set whose complexity is not clear. To our knowledge, techniques for optimization based on sampling have not been analyzed in the setting of Riemannian manifolds.

### 1.3. Our Contributions

In this paper, we consider the problem of sampling on geodesically convex sets on manifolds. We provide an algorithm that samples from the Riemannian volume form on geodesically convex sets on

manifolds with non-negative lower bounds on the curvature. This corresponds to uniform sampling. More generally, our algorithm can sample from log-concave distributions on geodesically convex sets. We show how to use the sampling algorithm to optimize over geodesically convex sets on manifolds, adapting the simulated annealing algorithm to this setting.

Our sampling algorithm has access to the manifold via an oracle for the exponential map and to the convex set via a membership oracle. We remark that the manifold need not be given extrinsically i.e. by an embedding in an ambient Euclidean space, and our algorithm can work with intrinsic view of manifolds. For sampling on the whole manifold, Mangoubi and Smith (2018) show results that are similar in spirit to our results. In fact, they show that when we are interested in the whole manifold, the geodesic walk mixes in time independent of the dimension, but with the additional requirement that the (sectional) curvature must be lower bounded away from zero. Our results use different techniques, i.e. conductance, to bound the mixing time and require just non-negativity of the curvature. When the whole manifold is considered, our techniques give mixing assuming only bounds on the Ricci curvature. Another point to be noted is that Mangoubi and Smith (2018) show mixing with respect to the Wasserstein metric which on compact metric spaces is weaker than the the total variation metric. But, using standard coupling techniques and bounding the total variation distance between transition kernels of nearby points (as in Theorem 38), bounds on the total variation mixing can be achieved.

The sampling results of Lee and Vempala (2017) and Lee and Vempala (2018) are for Hessian manifolds obtained from barrier functions on convex sets in the Euclidean space, while we work on convex subsets of manifolds with positive curvature. They too use conductance to bound the mixing time. Localization is used to show their conductance results but their results are for Hessian manifolds with conditions on the convex function whose Hessian generates the metric.

To our knowledge, the current paper is the first to explore the connection between sampling and optimization on manifolds. Our algorithm is a zeroth order optimization algorithm requiring access only to an evaluation oracle to the function. Due to this fact, we only assume that the function is convex and has bounded Lipschitz constant (which can be achieved by assuming bounded first order derivatives) while first order and second order methods or zeroth order methods that estimate the gradient require bounds on higher derivatives. We assume regularity conditions on the convex set by assuming that the convex set contains a ball of radius $r$ and has small diameter. To our knowledge, no previous algorithm was known for geodesically-convex optimization in a general settings such as ours. The algorithm of Allen-Zhu et al. (2018) works on the PSD manifold with specific objective function.

We make extensive use of the localization techniques from Klartag (2017). Firstly, we show that the isoperimetric theorem from Kannan et al. (1995) has a natural analogue for manifolds with non-negative curvature and use the localization to show facts about the isoperimetric profile of convex sets in these manifolds. Secondly, we use localization to bound the expectation of the function value when sampled from density proportional to its exponential. This fact is essential for simulated annealing algorithm to return an approximate minimizer of the function. The techniques used to bound the expectation in the Euclidean case do not seem to carry over immediately to the manifold case. We use the localization technique in order to reduce the manifold case to the Euclidean case. Similarly, we use these techniques to relate the distributions for nearby temperatures in the schedule, another essential ingredient for the analysis of the simulated annealing algorithm. Note that Klartag (2017) works with Ricci curvature and thus applies to our setting with assumptions only on the Ricci curvature. However, to assert the convexity of the set of interior points far away from the boundary

of a convex set, we need to use Cheeger and Gromoll (1972) which relies on the positivity of the sectional curvature. The proof of the theorem uses the Rauch comparison theorem for Jacobi fields, comparing Jacobi fields of manifolds with model manifolds with constant curvature. This is indeed the only place where we require sectional curvature lower bounds (instead of the Ricci curvature), and thus if we are working with the whole manifold Ricci curvature suffices for the analysis.

### 1.4. Organization of the Paper

In Section 2 and Appendix A, we provide the preliminary definitions and theorem required for the rest of the paper. In Section 3, we present the algorithms for sampling and optimization. In Section 4, we state the main results regarding the guarantees of the sampling and optimization algorithms. In Section 5, we provide a sketch of the proofs, going over the key ideas informally. The formal proofs of the theorems have been deferred to the appendix.

## 2. Preliminaries

We defer the discussion of required defintions from Riemannian geometry to Appendix A. Here, we begin by extending the notion of convexity to the setting of manifolds. Since the notion of geodesics is the natural manifold generalization of the Euclidean notion of straight lines, a natural requirement for a set on manifolds to be convex would be to require them to contain the geodesics between every two points in the set. A technicality that arises with this definition is that in general there could be multiple geodesics between two points. As an example one could consider the sphere, where there are two geodesics joining any two points. This leads to several notions of geodesic convexity depending on different requirements on the geodesics. We note the main notion here and have an extended discussion in the appendix.

**Definition 1 (Geodesic Convex Set, see Chavel (2006))** *Let $(M, g)$ be a complete Riemannian manifold. A subset $A \subseteq M$ is said to be strongly geodesically convex if for each pair $p, q \in A$, there exists a geodesic $\gamma \subseteq A$ that is the unique length minimizer in $M$ connecting $p$ and $q$ and $\gamma$ is the unique geodesic in $A$ joining $p$ and $q$.*

As an example of strongly convex sets on manifolds, consider spherical caps contained in a hemisphere on the sphere in $n$ dimensions. In general it can be shown that on manifolds, geodesic balls of small enough radius are strongly convex, as could be expected from intuition. We will work with the notion of strongly convex sets. We do not consider the notion of totally convex sets (see Vishnoi (2018)) here, since in the setting of positively curved manifolds, this class is not sufficiently rich. For example, spherical caps are not totally convex. It can be seen that the notion of strongly convex sets on the Euclidean space corresponds to the usual notion of convex sets, since straight line segments are the unique minimizing geodesics in the Euclidean space. This should not be confused with other notions of strong convexity of sets and functions in the Euclidean space. In the sequel, we shall use the term convex set to mean strongly convex set.

**Definition 2 (Geodesically Convex Function)** *Let $(M, g)$ be a Riemannian manifold and let $K \subseteq M$ be a strongly convex subset. Then, a function $f : K \to \mathbb{R}$ is said to be geodesically convex if for every pair $p, q \in K$, and for any geodesic $\gamma : [0, 1] \to K$ joining $p$ and $q$, we have that $f \circ \gamma$ is convex, that is, for all $t \in [0, 1]$,*

$$f\left(\gamma\left(t\right)\right) \leq \left(1 - t\right) f\left(p\right) + t f\left(q\right).$$

7

## 3. Algorithms

In this section, we first describe an algorithm to sample from a convex set on a manifold, given access to the membership oracle for the convex set $K$. We consider a Markov chain whose stationary distribution is the uniform distribution on the convex set. The algorithm in each step samples from the ball uniformly on the tangent space at the current point and moves along the induced geodesic on the manifold for a length specified by the step length. If the geodesic takes us out of the convex set, then we reject the step and resample from the tangent space.

---
**Algorithm 1:** Geodesic Walk on a Manifold $M$

---
**Input:** Dimension $n$, Convex set $K$, Starting Point $X_0 \in K$, Step Size $\delta$, Number of Steps $N$.

**for** $\tau \leq N$ **do**
    Pick $u_{\tau+1} \leftarrow N(0, I)$ in $T_{X_\tau} M$.
    **if** $\exp_{X_\tau}(\delta u_{\tau+1}) \in K$ **then**
        Set $X_{\tau+1} \leftarrow \exp_{X_\tau}(\delta u_{\tau+1})$.
    **else**
        Set $X_{\tau+1} \leftarrow X_\tau$.
    **end**
**end**

**Output:** Point $X_N$ approximately uniformly sampled from $K$.

---

We work with Riemannian manifolds that are complete as metric spaces. The Hopf–Rinow theorem asserts that for such manifolds the exponential map is defined on the whole tangent space. Thus all the oracle calls in Algorithm 1 are well-defined.

Variations of the walk above have been considered in Lee and Vempala (2017), Lee and Vempala (2018) and Mangoubi and Smith (2018). Hamiltonian walk with the Hamiltonian given by $H(x, v) = \frac{1}{2} \langle v, v \rangle_{g^{-1}(x)}$ leads to a variation of the geodesic walk because of the well known fact that the geodesic flow is the Hamiltonian flow for the above Hamiltonian. Here $g$ refers to the Riemannian metric as defined in Definition 13. Note that executing the geodesic walk needs oracle access to the exponential map and a membership oracle to the convex set.

Towards the goal of optimizing convex functions, we adapt the simulated annealing algorithm from the Euclidean setting to the Riemannian setting. Given a function $f$ and a "temperature" $T$, define the probability density $\pi_{f,T} \sim e^{-\frac{f}{T}}$. Intuitively, the function puts more weight on points where the function attains small values and sampling from the distribution is likely to output points near the minima of the function $f$ for low enough temperature $T$. The issue is that sampling from the distribution for a low enough temperature is a priori as hard as solving the initial optimization problem. The way to get around this issue is to set a temperature schedule in which one progressively "cools" the distribution such that the sample from the previous temperature acts to make it easier to sample from the next temperature. Once we reach a low enough temperature, the sample we attain will be close to the optimum point with high probability.

Since we need to sample from a distribution proportional to the $e^{-\frac{f}{T}}$, the natural idea would be to use the Metropolis filter with respect to the original uniform sampling algorithm. This leads to the following algorithm for sampling from the required distribution.

---

**Algorithm 2:** Adapted Geodesic Walk on a Manifold $M$ and function $f$.

---

**Input:** Dimension $n$, Convex set $K$, Convex function $f : K \to \mathbb{R}$, Starting Point $X_0 \in K$,
      Step Size $\eta$, Number of Steps $N$.

**for** $\tau \leq N$ **do**
    Pick $u_{\tau+1} \leftarrow N(0, I)$ in $T_{X_\tau} M$.
    **if** $y = \exp_{X_\tau}(\eta\, u_{\tau+1}) \in K$ **then**
        With probability $\min\left(1, e^{-f(y)+f(X_\tau)}\right)$, set $X_{\tau+1} \leftarrow y$.
        With the remaining probability, set $X_{\tau+1} \leftarrow X_\tau$.
    **else**
        Set $X_{\tau+1} \leftarrow X_\tau$.
    **end**
**end**

**Output:** Point $X_N \in K$ sampled approximately proportional to $e^{-f}$.

---

Given the algorithm for sampling at fixed temperature, we adapt the simulated annealing algorithm for the case of positively curved manifolds. The sequence of temperatures that the optimization takes is called the temperature schedule or the cooling schedule.

---

**Algorithm 3:** Simulated Annealing on Manifold $M$.

---

**Input:** Dimension $n$, Convex set $K$, Convex function $f : K \to \mathbb{R}$, Starting Point $X_0 \in K$,
      Number of Iterations $N$, Temperature Schedule $T_i$.

**for** $\tau \leq N$ **do**
    Sample $X_\tau$ according to distribution $\pi_{f,T_\tau}$ using Algorithm 2 for $\pi_{f,T_\tau}$ with starting point
    $X_{\tau-1}$.
**end**

**Output:** Point $X_N \in K$ that approximately minimizes $f$.

---

The main thing to specify in the design of the algorithm above is the temperature schedule, that is, the sequence of temperatures $T_i$ from which we sample. The aim is to set the schedule such that each temperature is close enough to such that the distributions are similar, while still maintaining a small number of temperature updates. Following Kalai and Vempala (2006), we show that

$$T_{i+1} = \left(1 - \frac{1}{\sqrt{n}}\right) T_i$$

provides us with the required guarantees. We set the initial temperature $T_0$ such that the uniform distribution on convex set is close in the $L_2$ norm to the initial distribution. Note that setting $T_0 = \max_x f(x)$ satisfies this requirement. Since the function is $L$-Lipschitz, we can bound this by $T_0 \leq \min_x f(x) + LD$ where $D$ is the diameter of the convex set.

## 4. Main Results

In this section, we discuss the main results of the paper. The proofs of the theorems discussed here have been deferred to the appendix, while the ideas for the proofs of the main lemmas are discussed

in [Section 5](#). The first result is regarding sampling from a convex set in manifolds with positive curvature.

**Theorem 3** *[Mixing Time Bound] Let $(M, g)$ be an $n$-dimensional Riemannian manifold and $K \subset M$ be a strongly convex subset satisfying the following conditions:*

- *$M$ has non-negative sectional curvature i.e. $Sec_M \geq 0$.*

- *The Riemannian curvature tensor is upper bounded i.e. $\max \|R_M\|_F \leq R$.*

- *$K$ contains a ball of radius $r$.*

- *$K$ has diameter $D$.*

*Let the starting distribution be an $H$-warm start (see [Definition 26](#)). Then, in*

$$
t = O\left( \frac{H^2 D^2 n^3 (R+1)}{r^2 \epsilon^2} \log\left( \frac{H}{\epsilon} \right) \right)
$$

*steps, the distribution of $X_t$, the output of [Algorithm 1](#) is $\epsilon$-close to the uniform distribution on convex set in total variation.*

From the definition, we get that $|Sec_M| \leq Rn^{-1}$ implies $\|R_M\|_F \leq R$. So, the above theorem can be stated in terms of sectional curvature as well. See [Appendix F](#) for the proof of the above theorem. We next state the theorem regarding the optimization of convex functions on convex sets.

**Theorem 4** *[Simulated Annealing] Let $(M, g)$ be a manifold with non-negative Ricci curvature. Let $K \subseteq M$ be a strongly convex set satisfying the requirements in [Theorem 3](#). Let $f : K \to \mathbb{R}$ be a geodesically convex function with Lipschitz constant $L$. Then, starting from a uniform sample from $K$, [Algorithm 3](#) runs for*

$$
O\left( \frac{D^2 n^{7.5} (R+1) L^2}{r^2 \epsilon^2 \delta^6} \log\left( \frac{n}{\delta} \log\left( \frac{T_0(n+1)}{\epsilon\delta} \right) \right) \log^5\left( \frac{T_0 n}{\epsilon\delta} \right) \right)
$$

*steps and with probability $1 - \delta$ outputs a point $x^*$ such that*

$$
f\left(x^*\right) - \min_x f(x) \leq \epsilon.
$$

The requirement to start with the uniform distribution is made for simplicity and can work with a warm start. [Appendix G](#) contains a proof of the above theorem. Note that the dependence on the probability of error can be reduced from polynomial to logarithmic using standard error reduction by running independent trials of the algorithm and outputting the minimum value across the trials.

## 5. Proof Sketch

The dominant strategy for showing that geometric Markov chains mix fast in convex bodies in the Euclidean space goes via the notion of conductance of Markov chain (Definition 25); see e.g. Vempala (2005). Intuitively, the conductance of a Markov chain is a measure of how well-connected the Markov chain is: small conductance implies that there are sets where the walk can get stuck and takes a long time to get out, somewhat like the barbell; large conductance implies that such bottlenecks do not exist. We will adapt this strategy to the manifold setting and show that the conductance of our walk is large, implying that the walk mixes fast. Lower bounding the conductance, as in the case of the Euclidean space, reduces to showing isoperimetric inequalities and bounding the overlap of one-step distributions of the chain. The key tool here is the localization lemma Lovász and Simonovits (1990); Kannan et al. (1995). It is useful for proving many geometric inequalities and in particular isoperimetric properties of convex sets in the Euclidean space. Recently, the classical localization lemma which applies to sets in $\mathbb{R}^n$ was extended to Riemannian manifolds Klartag (2017) to prove isoperimetry theorems on manifolds.

**Theorem 5 (Informal statement of Theorem 35)**  *Given a partition of a strongly geodesically convex set in a Riemannian manifold (with non-negative Ricci curvature) into three subsets $K_1$, $K_2$ and $K_3$ such that $K_1$ and $K_3$ are well separated, then the volume of $K_2$ is large compared to the volume of the minimum of the two separated sets.*

The geodesic walk on the convex set is geometric in nature and thus its conductance is tightly linked with the isoperimetry of the base space. To prove the rapid mixing of the geodesic random walk, we first show:

**Theorem 6 (Informal statement of Theorem 38)**  *For two nearby points on the convex set, the one-step distributions of the geodesic walk from these points are close in total variation distance.*

Next, for any subset $S_1$ of the strongly geodesically convex set $K$ (and its complement in $K$, namely $S_2$), we consider a subset $S_1'$ from which the geodesic walk is unlikely to escape. Using Theorem 6, the aim is to show that $S_1'$ and the corresponding set $S_2'$ in the complement are well separated, thus reducing the conductance to the isoperimetry claim of Theorem 5. One issue with this approach is that there could be a point in the set with very low probability of leaving that point. To deal with this, we restrict to the subset of the convex set which only consists of points with high probability of taking a step away from the current point. Using a theorem of Cheeger and Gromoll it can be shown that this subset is convex. Moreover, we show with another use of localization that it has volume comparable to that of the original set.

**Theorem 7 (Informal statement of Theorem 45)**  *The set of points with high local conductance is a convex subset of the original convex set and large fraction of the size of the original set.*

Using the above, we can show that the conductance of the walk is high, thus showing that the above Markov chain mixes rapidly. Running the Markov chain for sufficiently many steps, gives us a distribution that is close in total variation distance to the stationary distribution, which in this case is the uniform distribution as required.

**Theorem 8 (Informal statement of Theorem 3)** *Given a warm start, Algorithm 1 outputs a point that is approximately sampled from the uniform distribution on a geodesically convex set in time polynomial in the dimension.*

Towards showing the reduction from sampling to optimization on manifolds, we adapt the algorithm from Kalai and Vempala (2006) for the Euclidean setting. We first show that we can sample from the required distributions for each temperature by adapting the algorithm to the case of sampling from the required log-concave distributions.

**Theorem 9 (Informal statement of Theorem 56)** *Given a starting distribution that is close in the $L_2$ distance, Algorithm 2 outputs a point that is approximately sampled from a log-concave distribution in polynomial time.*

One then needs to pick a temperature schedule satisfying two seemingly opposite requirements. Firstly, we require that the schedule makes large enough updates so as to not require too many samples in order to reach the required low temperature. Secondly, we require that the temperatures are close enough that the two distributions we get for these temperatures are close by in $L_2$ distance. We pick a temperature schedule that multiplicatively updates the temperature in each iteration i.e. we pick $T_{i+1} = \left(1 - n^{-0.5}\right) T_i$. This ensures that the temperatures are rapidly decreasing. But, it remains to be shown that the temperatures are indeed close by in $L_2$ distance. The Euclidean version of this theorem is usually shown by constructing an auxiliary function whose log-concavity given the bound. The log-concavity is established using the fact that the marginals of log-concave measure is log-concave. The above strategy does not generalize to the manifold setting. We use localization directly to show the required inequality by reducing to the Euclidean case.

**Theorem 10 (Informal statement of Theorem 62)** *For two adjacent temperatures in the schedule, the two distributions are close in $L_2$ distance.*

Once we can sample from the distribution for each temperature, we need to then show that for low temperatures we get points near the optimal with good probability. The equivalent result in the Euclidean case is shown by reducing general convex functions to linear functions and for linear functions reducing to the case where the convex set is a cone. Again, this approach doesn't seem to generalize to the setting of manifolds. So, we use the localization technique using a more general version stated with respect to a guiding function. The distance from the minimum is used as the guiding function, leading to needles that pass through the minimum of the function, and then we can reduce to the inequality shown in Kalai and Vempala (2006).

**Theorem 11 (Informal statement of Theorem 57)** *For small enough temperature, the expected value of the function sampled from the distribution is close to the minimum value of the function.*

Given the above theorems, we can show the guarantees of the algorithm for optimizing convex functions on strongly convex sets on manifolds.

**Theorem 12 (Informal statement of Theorem 4)** *Given a geodesically convex function $f$ on a geodesically convex set, Algorithm 3 finds a point that is $\epsilon$-close to the optimal point running in time polynomial in the dimension and $\epsilon^{-1}$.*

## References

Jacob Abernethy and Elad Hazan. Faster convex optimization: Simulated annealing with an efficient universal barrier. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, pages 2520–2528. JMLR.org, 2016. URL http://dl.acm.org/citation.cfm?id=3045390.3045656.

P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, Princeton, NJ, 2008. ISBN 978-0-691-13298-3. doi: 10.1515/9781400830244. URL https://doi.org/10.1515/9781400830244. With a foreword by Paul Van Dooren.

Zeyuan Allen-Zhu, Ankit Garg, Yuanzhi Li, Rafael Oliveira, and Avi Wigderson. Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 172–181. ACM, 2018.

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds. *Ann. Probab.*, 43(1):339–404, 2015. ISSN 0091-1798. doi: 10.1214/14-AOP907. URL https://doi.org/10.1214/14-AOP907.

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014. ISBN 978-3-319-00226-2; 978-3-319-00227-9. doi: 10.1007/978-3-319-00227-9. URL https://doi.org/10.1007/978-3-319-00227-9.

Alexandre Belloni, Tengyuan Liang, Hariharan Narayanan, and Alexander Rakhlin. Escaping the local minima via simulated annealing: Optimization of approximately convex functions. In *Conference on Learning Theory*, pages 240–265, 2015.

Dimitris Bertsimas and Santosh Vempala. Solving convex programs by random walks. *Journal of the ACM (JACM)*, 51(4):540–556, 2004.

Michael Betancourt. A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*, 2017.

Isaac Chavel. *Riemannian geometry*, volume 98 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2006. ISBN 978-0-521-61954-7; 0-521-61954-8. doi: 10.1017/CBO9780511616822. URL https://doi.org/10.1017/CBO9780511616822. A modern introduction.

Jeff Cheeger and Detlef Gromoll. On the structure of complete manifolds of nonnegative curvature. *Ann. of Math. (2)*, 96:413–443, 1972. ISSN 0003-486X. doi: 10.2307/1970819. URL https://doi.org/10.2307/1970819.

Dario Cordero-Erausquin, Robert J. McCann, and Michael Schmuckenschläger. Prékopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport. *Ann. Fac. Sci. Toulouse Math. (6)*, 15(4):613–635, 2006. ISSN 0240-2963. URL http://afst.cedram.org/item?id=AFST_2006_6_15_4_613_0.

Persi Diaconis, Susan Holmes, and Mehrdad Shahshahani. Sampling from a manifold. In *Advances in modern statistical theory and applications: a Festschrift in honor of Morris L. Eaton*, volume 10 of *Inst. Math. Stat. (IMS) Collect.*, pages 102–125. Inst. Math. Statist., Beachwood, OH, 2013.

Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid Monte Carlo. *Physics letters B*, 195(2):216–222, 1987.

Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17, January 1991. ISSN 0004-5411. doi: 10.1145/102782.102783. URL http://doi.acm.org/10.1145/102782.102783.

Matthieu Fradelizi and Olivier Guédon. The extreme points of subsets of $s$-concave probabilities and a geometric localization theorem. *Discrete Comput. Geom.*, 31(2):327–335, 2004. ISSN 0179-5376. doi: 10.1007/s00454-003-2868-y. URL https://doi.org/10.1007/s00454-003-2868-y.

Mark Girolami and Ben Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(2):123–214, 2011. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2010.00765.x. URL https://doi.org/10.1111/j.1467-9868.2010.00765.x. With discussion and a reply by the authors.

Adam Tauman Kalai and Santosh Vempala. Simulated annealing for convex optimization. *Mathematics of Operations Research*, 31(2):253–266, 2006.

Ravi Kannan, László Lovász, and Miklós Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13(3-4):541–559, 1995.

Bo'az Klartag. Needle decompositions in Riemannian geometry. *Mem. Amer. Math. Soc.*, 249 (1180):v + 77, 2017. ISSN 0065-9266. doi: 10.1090/memo/1180. URL https://doi.org/10.1090/memo/1180.

Gilles Lebeau and Laurent Michel. Semi-classical analysis of a random walk on a manifold. *Ann. Probab.*, 38(1):277–315, 2010. ISSN 0091-1798. doi: 10.1214/09-AOP483. URL https://doi.org/10.1214/09-AOP483.

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

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*, pages 1115–1121. ACM, 2018.

Yin Tat Lee, Zhao Song, and Santosh S Vempala. Algorithmic theory of ODEs and sampling from well-conditioned logconcave densities. *arXiv preprint arXiv:1812.06243*, 2018.

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

László Lovász and Miklós Simonovits. The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. In *31st Annual Symposium on Foundations of Computer Science, Vol. I, II (St. Louis, MO, 1990)*, pages 346–354. IEEE Comput. Soc. Press, Los Alamitos, CA, 1990. doi: 10.1109/FSCS.1990.89553. URL https://doi.org/10.1109/FSCS.1990.89553.

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. Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *J. Comput. System Sci.*, 72(2):392–417, 2006. ISSN 0022-0000. doi: 10.1016/j.jcss.2005.08.004. URL https://doi.org/10.1016/j.jcss.2005.08.004.

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

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

Oren Mangoubi and Aaron Smith. Rapid mixing of hamiltonian monte carlo on strongly log-concave distributions. *arXiv preprint arXiv:1708.07114*, 2017.

Oren Mangoubi and Aaron Smith. Rapid mixing of geodesic walks on manifolds with positive curvature. *Ann. Appl. Probab.*, 28(4):2501–2543, 08 2018. doi: 10.1214/17-AAP1365. URL https://doi.org/10.1214/17-AAP1365.

Oren Mangoubi and Nisheeth Vishnoi. Dimensionally tight bounds for second-order hamiltonian monte carlo. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 6030–6040. Curran Associates, Inc., 2018. URL http://papers.nips.cc/paper/7842-dimensionally-tight-bounds-for-second-order-hamiltonian-monte-carlo.pdf.

Emanuel Milman. Isoperimetric bounds on convex manifolds. *Concentration, functional inequalities and isoperimetry*, 545:195–208, 2011.

Radford M. Neal. MCMC using Hamiltonian dynamics. In *Handbook of Markov chain Monte Carlo*, Chapman & Hall/CRC Handb. Mod. Stat. Methods, pages 113–162. CRC Press, Boca Raton, FL, 2011.

Yu. E. Nesterov and M. J. Todd. On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Found. Comput. Math.*, 2(4):333–361, 2002. ISSN 1615-3375. doi: 10.1007/s102080010032. URL https://doi.org/10.1007/s102080010032.

Liviu I Nicolaescu. *Lectures on the Geometry of Manifolds*. World Scientific, 2007.

Mark D. Plumbley. Lie group methods for optimization with orthogonality constraints. In Carlos G. Puntonet and Alberto Prieto, editors, *Independent Component Analysis and Blind Signal Separation*, pages 1245–1252, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg. ISBN 978-3-540-30110-3.

Tamás Rapcsák. *Smooth nonlinear optimization in* $\mathbf{R}^n$, volume 19 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 1997. ISBN 0-7923-4680-7. doi: 10.1007/978-1-4615-6357-0. URL https://doi.org/10.1007/978-1-4615-6357-0.

Alexander Rusciano. A Riemannian corollary of Helly's theorem. *arXiv preprint arXiv:1804.10738*, 2018.

Suvrit Sra, Nisheeth K. Vishnoi, and Ozan Yildiz. On Geodesically Convex Formulations for the Brascamp-Lieb Constant. In Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, volume 116 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 25:1–25:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-95977-085-9. doi: 10.4230/LIPIcs.APPROX-RANDOM.2018.25. URL http://drops.dagstuhl.de/opus/volltexte/2018/9429.

Constantin Udrişte. *Convex functions and optimization methods on Riemannian manifolds*, volume 297 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994. ISBN 0-7923-3002-1. doi: 10.1007/978-94-015-8390-9. URL https://doi.org/10.1007/978-94-015-8390-9.

Santosh Vempala. Geometric random walks: a survey. *Combinatorial and computational geometry*, 52(573-612):2, 2005.

Santosh S Vempala. Recent progress and open problems in algorithmic convex geometry. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 8. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2010.

Nisheeth K Vishnoi. Geodesic Convex Optimization: Differentiation on Manifolds, Geodesics, and Convexity. *arXiv preprint arXiv:1806.06373*, 2018.

Shing Tung Yau. Non-existence of continuous convex functions on certain Riemannian manifolds. *Math. Ann.*, 207:269–270, 1974. ISSN 0025-5831. doi: 10.1007/BF01351342. URL https://doi.org/10.1007/BF01351342.

Hongyi Zhang and Suvrit Sra. First-order methods for geodesically convex optimization. In *Conference on Learning Theory*, pages 1617–1638, 2016.

Hongyi Zhang and Suvrit Sra. Towards Riemannian Accelerated Gradient Methods. *arXiv preprint arXiv:1806.02812*, 2018.

Jingzhao Zhang, Hongyi Zhang, and Suvrit Sra. R-SPIDER: A Fast Riemannian Stochastic Optimization Algorithm with Curvature Independent Rate. *arXiv preprint arXiv:1811.04194*, 2018.

Junyu Zhang and Shuzhong Zhang. A cubic regularized Newton's method over Riemannian manifolds. *arXiv preprint arXiv:1805.05565*, 2018.

## Appendix A. Definitions

In this appendix, we note the required preliminary definitions from Riemannian geometry. For a more detailed account of Riemannian geometry see e.g. Chavel (2006) or Nicolaescu (2007).

**Definition 13 (Riemannian Manifold)** *Let $M$ be a smooth manifold and $TM$ be its associated tangent space. A Riemannian metric $g$ is a symmetric positive definite $(0, 2)$ tensor field (a smoothly varying inner product at each point) on $M$. The pair $(M, g)$ is then said to be a Riemannian manifold.*

We will assume in the rest of the paper that the manifold is connected. The Riemannian metric naturally provides a way to measure lengths of paths on the manifold. Given the metric $g_x : T_x M \times T_x M \to \mathbb{R}$, we can define a norm $|v|_x = \sqrt{g_x(v, v)}$. Given a smooth path $\gamma : [a, b] \to M$, define its length to be

$$|\gamma| = \int_a^b |\gamma'(t)|_{\gamma(t)} \, dt.$$

We can then use this notion to define a metric space structure on the manifold. Given any two points $x$ and $y$, define

$$d(x, y) = \inf_{\gamma(0)=x, \gamma(1)=y} |\gamma|.$$

With this metric space structure, we can define the notion of metric balls around points in the manifold. Denote by $B(x, a)$ the geodesic ball around $x$ of radius $a$.

In order to talk about variations as we move along a manifold, we need a way of talking about variations across different tangent spaces and talk about variations of vector fields relative to one another. This is captured by the notion of an affine connection. Affine connections on manifolds can be quite arbitrary and need not behave well the additional Riemannian structure. Given a Riemannian structure, we can define a canonical connection which respects the Riemannian structure. This is captured by the following definition.

**Definition 14 (Levi-Civita Connection)** *Let $(M, g)$ be a Riemannian manifold. An affine connection is said to be the Levi-Civita connection if it is torsion-free i.e.*

$$\nabla_X Y - \nabla_Y X = [X, Y]$$

*for every pair of vector fields $X, Y$ on $M$ and preserves the metric i.e*

$$\nabla g = 0.$$

*Here $[X, Y]$ denotes the Lie bracket between two vector fields.*

In order to extend the notion of convexity, we would like to generalize the notion of straight lines to the Riemannian case. We can do this in two ways. One is to note that in the Euclidean case, the lines are curves that minimize the path length. The second is to note that a particle traveling along a line at constant speed does not experience any acceleration. It can be shown that these two notions are indeed equivalent and points to the fact these definitions are indeed natural extensions of lines in the Euclidean space. For a more detailed treatment of these see Nicolaescu (2007), Chavel (2006) or Vishnoi (2018).

**Definition 15 (Geodesic)** *Let $M$ be a smooth manifold with an affine connection (see Theorem 14) $\nabla$. A smooth curve $\gamma : I \to M$ is a said to be a geodesic if*

$$\nabla_{\dot\gamma}\dot\gamma = 0$$

*where $\dot\gamma = \frac{d\gamma(t)}{dt}$.*

We next define notions of curvature on manifolds.

**Definition 16 (Riemannian Curvature Tensor)** *Let $(M, g)$ be a Riemannian manifold with the associated Levi-Civita connection $\nabla$. Define the Riemannian curvature tensor as*

$$R_M(u, v)w = \nabla_u\nabla_v w - \nabla_v\nabla_u w - \nabla_{[u,v]}w.$$

*Here $u, v$ are vector fields on the manifold.*

**Definition 17 (Ricci Curvature)** *Let $(M, g)$ be a Riemannian manifold with the associated Riemannian curvature tensor $R$. Then, define the Ricci curvature tensor $Ric_M$ to be the trace of the linear map*

$$X \to R(X, Y)Z.$$

*Here $X, Y, Z$ are vector fields on the manifold.*

**Definition 18 (Sectional Curvature)** *Let $(M, g)$ be a Riemannian manifold with the associated Riemannian curvature tensor $R$. Then, define the sectional curvature $Sec_M$ as*

$$Sec_M(u, v) = \frac{\langle R(u, v)v, u\rangle}{\langle u, u\rangle\langle v, v\rangle - \langle u, v\rangle^2}.$$

*where $u, v$ are linearly independent vectors on the tangent space at a point on the manifold. Note that the above definition depends just on the span of the vectors and not on the vectors themselves.*

**Definition 19 (Riemannian Volume)** *Let $(M, g)$ be an orientable Riemannian manifold. Then, then there exists a natural volume form on the manifold which is given in local coordinates as*

$$d\mathrm{Vol} = \sqrt{\det(g)}dx^1 \wedge \cdots \wedge dx^n.$$

Below we note the various definitions of geodesically convex sets. These notions vary based on the requirements on the geodesics in the set.

**Definition 20 (Geodesic Convex Set, see Chavel (2006))** *Let $(M, g)$ be a complete Riemannian manifold. A subset $A \subseteq M$ is said to be*

- *Weakly geodesically convex if for each pair $p, q \in A$, there exists a geodesic $\gamma \subseteq A$ that is the unique length minimizer in $A$ connecting $p$ and $q$.*

- *Geodesically Convex if for each pair $p, q \in A$, there exists a geodesic $\gamma \subseteq A$ that is the unique length minimizer in $M$ connecting $p$ and $q$.*

- *Strongly geodesically convex if for each pair $p, q \in A$, there exists a geodesic $\gamma \subseteq A$ that is the unique length minimizer in $M$ connecting $p$ and $q$ and $\gamma$ is the unique geodesic in $A$ joining $p$ and $q$.*

As in the case of the Euclidean space, we can characterize geodesically convex functions in terms of first and second order conditions. In fact, the characterizations below follow from the Euclidean versions by restricting to the geodesic of interest.

**Theorem 21 (see Udrişte (1994))** *Let $(M, g)$ be a Riemannian manifold and let $K \subset M$ be a strongly convex set. Let $f : K \to \mathbb{R}$ be a function.*

- *If $f$ is differentiable, then for the geodesic $\gamma$ joining points $p$ and $q$, we have*

$$f(p) + \dot{\gamma}(f)(p) \leq f(q)$$

*where $\dot{\gamma}(f)$ denotes the derivative of $f$ along $\gamma$.*

- *If $f$ is twice differentiable, then for the Hessian defined as*

$$Hess_f(X, Y) = \langle \nabla_X \nabla f, Y \rangle,$$

*for two vector fields $X$ and $Y$, we have $Hess_f$ is positive definite.*

In order to sample from a distribution, we use the general technique of Markov Chain Monte Carlo on continuous state space. A Markov chain is a stochastic process where the distribution of the next step depends only on the current state and not the history of the process. Formally,

**Definition 22 (Markov Chain)** *Let $(\Omega, \mathcal{F})$ be a measurable space. Let $Q_0$ be a measure on $(\Omega, \mathcal{F})$. A time homogeneous Markov chain is specified by a probability measure $P_w$ for each $w \in \Omega$. Define the chain by the sequence of measures $Q_i$, given by*

$$Q_{i+1}(A) = \int P_u(A) \, dQ_i(u)$$

The probability measure $P_x$ is said to be the one step of the above chain from the point $x$. By slight abuse of notation, we use $P_x(y)$ for $P_x(\{y\})$, whenever $y$ is a point on the manifold.

An important notion in the study of Markov chains is that of a stationary distribution. A stationary distribution of a Markov chain is a distribution on the state space such that the distribution after one step, starting with this distribution, remains the same. Formally,

**Definition 23 (Stationary Distribution)** *Let $P_x$ be the transition operator at point $x$ of a Markov chain on the state space $\Omega$. A measure $\pi$ on $\Omega$ is said to be stationary if for every measurable subset $A$, we have*

$$\int_{\Omega} P_x(A) \, d\pi(x) = \pi(A).$$

It can be shown under reasonable assumptions that the distribution of a Markov chain converges to its stationary distribution. This fact forms the basis for the use of Markov chains in sampling algorithms. The idea is to set up a chain such that the desired distribution is stationary and run the chain until the distribution is close to the stationary distribution. Below we note a property that implies stationarity and will also be helpful property when transforming a Markov chain with a certain stationary distribution to a Markov chain with a different stationary distribution.

**Definition 24** *A Markov chain is said to be time-reversible with respect to a measure $\pi$ if for any two measurable sets $A, B$*

$$\int_B P_u(A)\, d\pi(u) = \int_A P_u(B)\, d\pi(u).$$

From the above definition, it is clear that if a Markov chain is time-reversible with respect a certain measure then that measure is a stationary distribution for the chain.

One of the primary methods used to bound the mixing time is to bound the conductance of the walk. Intuitively, an impediment to fast mixing is the existence of large sets from which walk has a low chance of leaving, forming a bottleneck. The notion of conductance makes this notion rigorous. It can be shown that the conductance is tightly linked to the spectral gap of a Markov chain, which also controls the rate of mixing. Cheeger's inequality relates these notions, and is widely used in spectral graph theory. Cheeger's inequality was in fact originally proven in the setting of differential geometry, relating the Cheeger isoperimetric constant of a manifold to the eigenvalues of the Laplacian on the manifold.

**Definition 25 (Conductance of a Random Walk)** *Let $P_x$ be the transition operator at point $x$ of a Markov chain on the state space $\Omega$. Define the ergodic flow with respect to a distribution $\pi$ of a set $A$ to be*

$$\Phi(A) = \int_{x \in A} P_x(\Omega \backslash A)\, d\pi(x).$$

*Then, the conductance of the walk is defined to be*

$$\phi = \min_{\pi(A) \leq 1/2} \frac{\Phi(A)}{\pi(A)}.$$

*Similarly, for $0 \leq s \leq 0.5$, we can define the $s$-conductance to be*

$$\phi_s = \min_{s < \pi(A) \leq 1/2} \frac{\Phi(A)}{\pi(A) - s}.$$

*For each point $x \in \Omega$, define the local conductance $\ell(x)$ to be the probability of moving away from the point $x$ i.e.*

$$\ell(x) = 1 - P_x(x).$$

In order to measure the notion of a good start for the random walk, we use the following standard notions.

**Definition 26 (Warm start and $L_2$ distance, see Vempala (2005))** *Let $\mu_1$ and $\mu_2$ be two measures with common support.*

- *We say that $\mu_0$ is $H$-warm with respect to $\mu_1$ if*

$$\sup_A \frac{\mu_0(A)}{\mu_1(A)} \leq H.$$

- *We define the $L_2$ distance of $\mu_0$ with respect to $\mu_1$ as*

$$\|\mu_0/\mu_1\| = \int \left(\frac{d\mu_0}{d\mu_1}\right)^2 d\mu_1 = \int \left(\frac{d\mu_0}{d\mu_1}\right) d\mu_0.$$

### A.1. Oracles

For the problem to be well-posed and useful, we need to specify how the algorithm accesses the manifold. In our setting, we also assume that our algorithm has access to an oracle that takes as input a point on the manifold and real vector $v$ and outputs $\exp_x(v)$. Here, we are associating the tangent space $T_x$ at the point $x$ with $\mathbb{R}^n$. In our algorithms we will need access only to uniformly sampled vectors, which are invariant under orthogonal transformations, in the tangent space, hence the choice of the basis of the tangent space does not affect the computation. The oracle for the exponential map is common practice in the literature on optimization on manifolds, especially ones using first order methods of optimization, for example see Zhang and Sra (2018). Mangoubi and Smith (2018), Lee and Vempala (2017) and Lee and Vempala (2018) also present their algorithm in the presence similar oracles. For the particular manifolds of interest they give methods to construct the required oracle. The computation of the exponential map amounts to solving systems of ordinary differential equations, given by the characterizing equations of geodesics. See Lee and Vempala (2017) for a discussion on the method of construction of the oracle. For natural matrix manifolds the exponential map is well-understood and computationally easier to handle; for example, see Absil et al. (2008).

## Appendix B. Stationary Distribution of the Geodesic Walk

In this section, we show that the stationary distribution of the geodesic random walks on a convex set on a manifold is the uniform distribution on the convex set. This fact follows from using the well known Liouville Theorem for geodesic dynamics (see Chavel (2006)), and was observed in Mangoubi and Smith (2018) for the case of the unconstrained walk on a manifold of positive curvature.

The geodesic walk on the manifold is also a reversible Markov chain. This claim was proved in the more general case of Hamiltonian Monte Carlo in Lee and Vempala (2018), but we present a proof for completeness. The proof as in the case of the stationarity uses the Liouville theorem.

**Theorem 27** *The geodesic walk is a reversible Markov chain.*

**Proof** We show this in the case of unrestricted (i.e., without a metropolis filter to restrict the walk to a convex subset) geodesic walk in compact manifolds; the reversibility of the restricted version then follows from the analysis of the Metropolis filter, for example see Lemma 1.9 in Lovász and Simonovits (1993). In order to show the reversibility, it suffices to show that the transition operator of the walk is self-adjoint. Again, we restrict to the case of fixed step length. The general case where the step lengths are picked independently in each step is similar. Thus consider

$$\int_{x \in M} \int_{u \in S_x} f\left(\exp_x(\delta u)\right) g(x) \, d\mu_x(u) d\mathrm{Vol}\,(x)$$

where $\mu_x$ is the uniform distribution on the unit sphere $S_x$ in the tangent space $T_x$. Let us consider the above integral as an integral over the Liouville measure $\mathcal{L}$.

$$\int f\left(\exp_x(\delta u)\right) g(x) \, d\mathcal{L}\,(x, u).$$

Now note that the joint distribution of $\left((x, u), \left(\exp_x(\delta u), \bar{u}\right)\right)$ is the same as $\left(\left(\exp_y(-v), \bar{v}\right), (y, -v)\right)$, where $y$ is drawn from the uniform measure on the manifold and $v$ is drawn from the unit sphere on the tangent space $T_y$. Here $\bar{u}$ denotes the derivative of the geodesic at the point $\exp_x(\delta u)$. Making the above change of variables, we get the required theorem. To see this note that $\left(\exp_y(-v), -\bar{v}\right)$ is obtained by applying the geodesic flow to $(y, -v)$. From Liouville theorem and the symmetry of the uniform distribution on the unit sphere, we get that this has the same distribution as $(x, u)$. The required claim now follows by change of variables. ∎

With the reversibility given above, we see that the stationary distribution of the geodesic walk is the uniform measure as required. We note this in the following theorem.

**Theorem 28** *The stationary distribution of the geodesic walk on a geodesically convex set $K$ is the uniform distribution on $K$.*

## Appendix C. Needle Decomposition on Riemannian Manifolds

We first recall the localization lemma for the Euclidean space mentioned earlier.

**Theorem 29 (Localization Lemma, Lovász and Simonovits (1990); Kannan et al. (1995))** *Let $g$ and $h$ be real-valued lower semi-continuous measurable function on $\mathbb{R}^n$ satisfying*

$$\int_{\mathbb{R}^n} g(x)dx > 0$$

*and*

$$\int_{\mathbb{R}^n} h(x)dx > 0.$$

*Then, there exists a linear function $\omega : [0, 1] \to \mathbb{R}_+$ and two points $a, b \in \mathbb{R}^n$ such that*

$$\int_0^1 \omega^{n-1}(t)g\left(at + (1 - t)b\right) dt > 0$$

*and*

$$\int_0^1 \omega^{n-1}(t)h\left(at + (1 - t)b\right) dt > 0.$$

The proof of the above theorem follows the strategy of repeated bisection. We find a halfspace for which both the given inequalities are satisfied. Taking this to the limit, we are left with one dimensional interval such that proving the inequality on the original convex set reduces to proving the inequality on the interval with respect to the measure we get from the limiting process. See Klartag (2017) for an overview of the development of the localization technique.

In the Riemannian setting, it is not immediately clear what the right analogue of this bisection argument is. In a recent work Klartag (2017), a Riemannian analogue of the localization lemma was

proven by invoking connections to optimal transport. This localization lemma produces a disintegration of measure for the volume form on the manifold into measures $\mu_I$ supported on geodesic sections such that the measures $\mu_I$ are pushforwards of measures on the real line satisfying the curvature dimension condition on the interval of support.

We next define the notion of curvature-dimension conditions on a weighted manifold, which shall be central in the development of the localization technique over Riemannian manifolds; see Klartag (2017) for more details.

**Definition 30 (Curvature-Dimension Condition)** *Let $(M, d, \mu)$ be a $n$-dimensional weighted Riemannian manifold and let $\rho : M \to \mathbb{R}$ be a smooth function such that the weight measure has density $e^{-\rho}$ with respect to the Riemannian volume measure. Let the generalized Ricci curvature with parameter $N \in (-\infty, 1) \cup [n, +\infty]$ be defined through the equation*

$$
Ric_{\mu,N}(v, v) = \begin{cases} Ric_M(v, v) + Hess_\rho(v, v) - \frac{(\partial_v \rho)^2}{N-n} & N \neq \infty, n \\ Ric_M(v, v) + Hess_\rho(v, v) & N = \infty \\ Ric_M(v, v) & N = n \end{cases}
$$

*for every $x \in M$ and tangent vector $v \in T_x M$. For $\kappa \in \mathbb{R}$, we say that the weighted manifold satisfies the curvature-dimension condition $CD(\kappa, N)$ if for every $x \in M$ and tangent vector $v \in T_x M$, we have*

$$
Ric_{\mu,N}(v, v) \geq \kappa \cdot g(v, v).
$$

The condition defined above comes up naturally in the study of diffusion operators. The notion is used to prove hypercontractivity inequalities and logarithmic Sobolev inequalities. For more information about the connections to diffusion operators, see Bakry et al. (2014). This condition has also been considered in optimal transport in terms of displacement interpolation in the space of probability measures. For discussion on this condition and the relationship to the above curvature dimension condition, see Ambrosio et al. (2015).

When $n$ is positive it can be interpreted as a generalized upper bound on the dimension for weighted manifold and $\kappa$ can be interpreted as a generalized lower bound on the curvature of the weighted manifold. Manifolds in this paper always have non-negative sectional curvature and satisfy $\kappa \geq 0$. We make extensive use of localization for manifolds satisfying $CD(0, N)$ condition with $N = n$ or $\infty$; for example in Theorems 35, 45, 54.

**Definition 31 (Needle, Klartag (2017))** *Let $(M, d, \mu)$ be a weighted Riemannian manifold satisfying the $CD(\kappa, N)$ condition for some $\kappa \in \mathbb{R}$ and $N \in (-\infty, 1) \cup [n, +\infty]$. Let $\eta$ be a measure on $M$. We say that $\eta$ is a $CD(\kappa, N)$ needle if there exists an open interval $A \subset \mathbb{R}$, a smooth function $\Psi : A \to \mathbb{R}$ and a minimizing geodesic $\gamma : A \to M$ such that*

1. *$\eta$ is the pushforward of the density on $A$ that is proportional to $e^{-\Psi}$ under $\gamma$ i.e. for any measurable subset $B$, $\eta(B) = \theta\left(\gamma^{-1}B\right)$, where $\theta \propto e^{-\Psi}$.*

2. *For every $x \in A$, we have*

$$
\Psi''(x) \geq \kappa + \frac{\left(\Psi'(x)\right)^2}{N-1}.
$$

*For the case of $N = \infty$, we intepret the second term on the right hand side to be zero.*

*If the above inequality is an equality, then the measure is said to be a $CD(\kappa, N)$ affine needle.*

Note that the needle condition is equivalent to the $CD(\kappa, N)$ condition on the preimage interval. As we shall see later, the geodesics play the role of the lines that appear in the Euclidean case.

**Theorem 32 (Needle Decomposition, Klartag (2017))** *Let $n \geq 2$, $\kappa \in \mathbb{R}$ and $N \in (-\infty, 1) \cup [n, \infty]$. Let $(M, d, \mu)$ be a weighted complete Riemannian manifold satisfying the $CD(\kappa, N)$ condition. Consider an integrable function $f : M \to \mathbb{R}$ satisfying $\int_M f d\mu = 0$ and $\int_M |f(x)| d(x, x_0) d\mu < \infty$ for some $x_0$. Then, there exists a partition $\Omega$ of $M$ and a measure $\nu$ on $\Omega$ and a family of measure $\{\mu_i\}_{i \in \Omega}$ such that*

1. *For any measurable set $A$,*
$$\mu(A) = \int_\Omega \mu_i(A) d\nu(i).$$
   *This can be seen as a disintegration of measure.*

2. *For almost every $i \in \Omega$, $i$ is an image of a minimizing geodesic and $\mu_i$ is supported on $i$. Furthermore, for almost every $i$, either $i$ is a singleton or $\mu_i$ is a $CD(\kappa, N)$ needle.*

3. *For almost every $i$,*
$$\int_i f d\mu_i = 0.$$

A corollary of the above theorem is the Riemannian analogue of the four function theorem from Kannan et al. (1995). This version of the theorem is easier to use in our setting and will be the application of the needle decomposition to be used in our setting.

**Corollary 33 (Four Function Theorem, Klartag (2017))** *Let $n \geq 2$, $\kappa \in \mathbb{R}$ and $N \in (-\infty, 1) \cup [n, \infty]$. Let $(M, d, \mu)$ be a $n$-dimensional weighted complete Riemannian manifold satisfying the $CD(\kappa, N)$ condition. For $i = 1, 2, 3, 4$, consider integrable functions $f_i : M \to \mathbb{R}^{\geq 0}$ satisfying $\int_M |f_i(x)| d(x, x_0) d\mu < \infty$ for some $x_0$. Assume that there are constants $\alpha, \beta$ such that $f_1^\alpha f_2^\beta \leq f_3^\alpha f_4^\beta$ almost everywhere. If for every $CD(\kappa, N)$ needle $\eta$ (for which the functions are integrable), the following is satisfied*

$$\left(\int f_1 d\eta\right)^\alpha \left(\int f_2 d\eta\right)^\beta \leq \left(\int f_3 d\eta\right)^\alpha \left(\int f_4 d\eta\right)^\beta,$$

*then,*

$$\left(\int f_1 d\mu\right)^\alpha \left(\int f_2 d\mu\right)^\beta \leq \left(\int f_3 d\mu\right)^\alpha \left(\int f_4 d\mu\right)^\beta.$$

In the case of $N = \infty$ or $\kappa = 0$, we can further simplify the above theorem by decomposing the one-dimensional needles into simpler "affine" needles. This reduces showing integral inequalities on manifolds to showing inequalities on the real line involving a small number of real parameters.

**Lemma 34 (Reduction to $N = \infty$ or $\kappa = 0$ affine case, Klartag (2017))** *Let $\mu$ be a $CD(\kappa, N + 1)$ needle on $\mathbb{R}$ for $\kappa = 0$ or $N = \infty$. Consider an integrable, continuous function $f : \mathbb{R} \to \mathbb{R}$ satisfying $\int_M f d\mu = 0$. Then, there exists a partition $\Omega$ of $\mathbb{R}$ and a measure $\nu$ on $\Omega$ and a family of measure $\{\mu_i\}_{i \in \Omega}$ such that*

1. *For any measurable set A,*

$$\mu(A) = \int_\Omega \mu_i(A)\, d\nu(i).$$

2. *For almost every $i \in \Omega$, either $i$ is a singleton or $\mu_i$ is a $CD(\kappa, N+1)$ affine needle.*

3. *For almost every $i$,*

$$\int_i f d\mu_i = 0.$$

The four function theorem was shown in Kannan et al. (1995) for the Euclidean space; for that case we can reduce from general needles to the case of exponential needles. It is conjectured in Klartag (2017) that the above theorem should generalize to the $\kappa \geq 0$ case for positive $N$.

## Appendix D. Isoperimetric Inequality on Manifolds with Positive Curvature Lower Bounds

To show that the random walk mixes fast, we need to show that the conductance of the walk is high. A standard way of showing conductance bounds is through isoperimetry. Bounds on the isoperimetry in convex sets in the Euclidean case were studied by Kannan et al. (1995) and there has been a succession of works in the area. The main idea is that since the walks we are interested in are geometric with transition kernel related to the volumes in the convex body, one expects the conductance of the Markov chain induced by the walk to be related to the isoperimetry profile of the set.

In the following theorem, we show a version of the isoperimetry theorem from Kannan et al. (1995) in the Riemannian setting. Bounds for the Cheeger constant of convex domains of manifolds with Ricci curvature lower bounds were shown in Milman (2011). But in order to apply the isoperimetric theorem to the geometric random walk of interest, we need to get a robust version of the above, similar to the statement in Kannan et al. (1995). We use the localization theorem from Klartag (2017) to get the required statement, similar to the proof of the isoperimetric theorem given in the same paper. In this case, we can reduce the required inequalities largely to the Euclidean log-concave case, since the $CD(\kappa, N)$ condition in this case leads to the log-concavity of the one-dimensional measures. Using the localization technique, we can reduce the required inequalities largely to the Euclidean log-concave case, since the $CD(\kappa, N)$ condition in this case leads to the log-concavity of the one-dimensional measures.

**Theorem 35** *Let $M$ be a manifold with non-negative Ricci curvature, i.e.*

$$Ric_M \geq 0.$$

*Let $K \subset M$ be a strongly convex subset. Then for any pairwise disjoint subsets $K_1, K_2, K_3 \subseteq K$ such that $K = \cup_i K_i$ and $d(K_1, K_3) \geq \epsilon$, we have*

$$\frac{m}{\epsilon \log 2} \text{Vol}(K)\text{Vol}(K_2) \geq \text{Vol}(K_1)\text{Vol}(K_3),$$

*where $m = \frac{1}{\text{Vol}(K)} \int_K d(x,y) d\text{Vol}(y)$ for some $x$.*

**Proof** We will use the localization from Theorem 32 to reduce the problem to an inequality about one-dimensional functions. We will assume that the appropriate sets are closed. For $i = 1, 2, 3$, let $f_i$ be indicators for $K_i$ and let $f_4$ be $(\epsilon \log 2)^{-1} d(x, y)$ for some point $x$ to be picked later. Then, the required equation reduces to

$$\int_M f_1(x)\, d\text{Vol} \int_M f_3(x)\, d\text{Vol} \leq \frac{1}{\epsilon \log 2} \int_M f_2(x) d\text{Vol} \int_M f_4(x)\, d\text{Vol}.$$

To prove this we use Theorem 33. Towards this consider a $CD(0, \infty)$ needle $\eta$ with the associated minimizing geodesic $\gamma$. Note that since $K$ is strongly convex, we can consider $\gamma$ is completely contained in $K$ We need to show the show the following inequality.

$$\int_M f_1(x)\, d\eta \int_M f_3(x)\, d\eta \leq \frac{1}{\epsilon \log 2} \int_M f_2(x) d\eta \int_M f_4(x)\, d\eta.$$

Using the fact that $\eta$ is a push forwards measure, we get

$$\int_{\gamma^{-1}K_1} e^{-\psi(x)} dx \int_{\gamma^{-1}K_3} e^{-\psi(x)} dx \leq \frac{1}{\epsilon \log 2} \int_{\gamma^{-1}K_2} e^{-\psi(x)} dx \int_{\gamma^{-1}K} |x - u|\, e^{-\psi(x)} dx$$

The last integral follows by noting that we can as assume without loss of generality that the point we are interested in lies on the geodesic of interest and using the fact that $\gamma$ is a minimizing geodesic.

Note that since, $\psi$ is a convex function, we have that the $e^{-\psi}$ measure is log-concave. Using the reduction from Kannan et al. (1995), we reduce from the case of arbitrary log concave measures to exponential needles.

**Lemma 36 (Kannan et al. (1995))** *Let $f_1, f_2, f_3, f_4$ be four non-negative continuous functions defined on an interval $[a, b]$ and $\alpha, \beta > 0$. Then, the following are equivalent.*

1. *For every log-concave function $F : \mathbb{R} \to \mathbb{R}$,*

$$\left( \int_a^b F(t) f_1(t) dt \right)^\alpha \left( \int_a^b F(t) f_2(t) dt \right)^\beta \leq \left( \int_a^b F(t) f_3(t) dt \right)^\alpha \left( \int_a^b F(t) f_4(t) dt \right)^\beta;$$

2. *For subinterval $[a', b'] \subseteq [a, b]$, and every real $\delta$,*

$$\left( \int_{a'}^{b'} e^{\delta t} f_1(t) dt \right)^\alpha \left( \int_{a'}^{b'} e^{\delta t} f_2(t) dt \right)^\beta \leq \left( \int_{a'}^{b'} e^{\delta t} f_3(t) dt \right)^\alpha \left( \int_{a'}^{b'} e^{\delta t} f_4(t) dt \right)^\beta.$$

Using Lemma 36, we reduce the required inequality to one about exponential functions.

$$\int_{\gamma^{-1}K_1} e^{\alpha t} dt \int_{\gamma^{-1}K_3} e^{\alpha t} dt \leq \frac{1}{\epsilon \log 2} \int_{\gamma^{-1}K_2} e^{\alpha t} dx \int_{\gamma^{-1}K} |x - u|\, e^{\alpha t} dx$$

Note that $\gamma^{-1}K_i$ is partition of the interval of support $\gamma$ into measurable sets, $J_1$, $J_2$ and $J_3$. The required inequality then follows from the proof of Theorem 5.2 in Kannan et al. (1995). ∎

## Appendix E. Overlap of One-step Distributions

To reduce the conductance of the random walk to the isoperimetric inequality, we need to show that the two points that are close by have single step distributions that are close by in total variation distance. To this end we use the techniques followed by Lee and Vempala (2018). In the statement below, we define the Frobenius norm of the Riemannian curvature tensor as,

$$\|R_M\| = \mathbb{E}_{u,v \sim N\left(0, g^{-1}(x)\right)} \left[ \left\langle R\left(u, \gamma'\right) \gamma', v \right\rangle \right],$$

where $\gamma(t)$ is a geodesic on the manifold. Note that the Frobenius norm implicitly depends on the point on the manifold. Also, recall that for any point $x \in M$, $P_x(x)$ denotes the probability that the geodesic walk, specified by Algorithm 1, starting at a point $x$ stays at the point in one step. The following definition defines a parameter of the manifold upper bounding the curvature.

**Definition 37** *Let $(M, g)$ be a Riemannian manifold. Let $R_M$ denote the Riemannian curvature tensor. Then, define*

$$R = \max_x \|R_M\|_F$$

*where the maximum is taken over all points.*

**Theorem 38 (One Step Overlap)** *Let $(M, g)$ be a complete Riemannian manifold. Let $x, y \in M$ be points on the manifold with $P_x(x), P_y(y) \leq c_2$. Then, for $\delta^2 \leq \frac{1}{100\sqrt{n}R}$, we have*

$$d_{TV}\left(P_x, P_y\right) \leq c_2 + O\left(\frac{1}{\delta}\right) d(x, y) + \frac{1}{25}.$$

**Proof** Note that, since there is a non zero probability of remaining at a given point, the one step distribution is not absolutely continuous with respect to the Riemannian volume. But, if we exclude this point then the measure does indeed become absolutely continuous with respect to the volume. Now, we need to compare the one step distribution of geodesic walk from two different points. To do this, we use the following lemma from Lee and Vempala (2018), noting that the geodesic walk is an instantiation of the Hamiltonian walk with the choice of the Hamiltonian mentioned earlier. Denote by $\tilde{P}_x$ and $\tilde{P}_y$, one step distributions of the geodesic walk without the constraint from the convex set.

**Lemma 39 (see Lemma 25 in Lee and Vempala (2018))** *For $\delta^2 \leq \frac{1}{100\sqrt{n}R}$, we have*

$$d_{TV}\left(\tilde{P}_x, \tilde{P}_y\right) \leq O\left(\frac{1}{\delta}\right) d(x, y) + \frac{1}{25}.$$

This gives us the required result. ■

## Appendix F. Bounds on Conductance of the Geodesic Walk

The following theorem is a version of the general framework of reducing the conductance of a geometric random walk to that of the isoperimetry of the underlying space. We know that nearby points have high overlap of one-step distribution, implying that if two points don't have high overlap, then they must be far apart. Using this we reduce the conductance to isoperimetry of the set given by the Theorem 35.

Since, we are looking at a walk such that at each step there is a probability of step getting rejected, we need to make certain that we are at a point such that local conductance is not too small. This motivates the following definition.

**Definition 40** *Let $K \subseteq M$ be a strongly convex set. Define the set of points with high local conductance with respect to the geodesic walk with step size $\delta$ to be*

$$K_{-\delta} := \{x \in K : P_x(x) = 0\}.$$

The set $K_{-\delta}$ turns out to be nice:

**Lemma 41 (Theorem 1.9 in Cheeger and Gromoll (1972))** *Let $K \subseteq M$ be a strongly convex set on a manifold with nonnegative sectional curvature and let $K_{-\delta}$ be defined as above. Then, $K_{-\delta}$ is a strongly convex subset.*

Next we show that the volume of $K_{-\delta}$ is reasonably large compared to the volume of the original set $K$. To do this we need to use a finer version of the theorem from Klartag (2017). We need the following definition.

**Definition 42 (Strain, Klartag (2017))** *Given a 1-Lipshitz function $u : M \to \mathbb{R}$ and a point $y$, we say that $y$ is a strain point of $u$ if there exist two points $x$ and $z$ with $d(x,z) = d(z,y) + d(y,x)$ such that*

$$u(x) - u(y) = d(x,y) > 0$$

*and*

$$u(z) - u(y) = d(y,z) > 0.$$

*Define the set of strain points on the manifold to be $Strain[u]$.*

**Lemma 43 (Klartag (2017))** *Consider the relation on $Strain[u]$ by*

$$x \sim y \iff |u(x) - u(y)| = d(x,y).$$

*Then, the relation is an equivalence relation on $Strain[u]$ and each equivalence class is an image of a minimizing geodesic.*

Denote by $T^\circ[u]$ the set of equivalence classes. Each element of $T^\circ[u]$ is associated with the image of a minimizing geodesic. Using the above definitions, we state the following more general version of Theorem 32. It can also be shown that the following theorem implies Theorem 32 using the Monge-Kantorovich duality.

**Theorem 44 (Klartag (2017))** *Let $M$ be a manifold satisfying the curvature dimension condition $CD(0, N)$. Let $u : M \to \mathbb{R}$ be a 1-Lipschitz function. Then, there exists a measure $\nu$ on the set $T^{\circ}[u]$ and a family $\{\mu_I\}$ such that*

1. *For any measurable set $A \subseteq M$, the map $I \to \mu_I(A)$ is well defined $\nu$ almost everywhere. For a measurable subset $S \subseteq T^{\circ}[u]$, we have $\pi^{-1}(S) \subseteq Strain[u]$ is a measurable subset of $M$.*

2. *For any measurable set $A \subseteq M$, we have*

$$\mu\left(A \cap Strain[u]\right) = \int_{T^{\circ}[u]} \mu_I(A) \, d\nu(I).$$

3. *For $\nu$-almost every $I \in T^{\circ}[u]$, $\mu_I$ is a $CD(k, N)$-needle supported on $I \subseteq M$. Furthermore, the set $A \subseteq \mathbb{R}$ and the minimizing geodesic $\gamma : A \to M$ may be selected so that $I = \gamma(A)$ and*

$$u\left(\gamma(t)\right) = t.$$

**Theorem 45** *Let $K \subseteq M$ be a strongly convex subset of $M$ containing a ball of radius $r$. Then, for $\epsilon \leq rn^{-1}$, we have*

$$\mathrm{Vol}\left(K - K_{-\epsilon}\right) \leq \frac{en\epsilon}{r}\mathrm{Vol}(K).$$

**Proof** We prove this using localization. Using $\chi$ for the indicator function, we can rewrite our desired inequality as

$$\int_M \chi_{K-K_{-\epsilon}}(x) \, d\mathrm{Vol}(x) \leq \frac{en\epsilon}{r} \int_M \chi_K(x) \, d\mathrm{Vol}(x).$$

Let $x^*$ be a point such that there is a ball of radius $r$ centered around $x^*$ contained within $K$. We use Theorem 44 with respect to the function $u(x) = d(x, x^*)$. Note that the strain sets of $u$ are geodesics through $x^*$ and since $K$ is strongly convex and the manifold is geodesically complete, we can extend these geodesics to the boundary. Using the localization for $CD(0, n)$ manifolds, where $n$ is the dimension of the manifold, we decompose the measure into needles. Thus, the required inequality reduces to showing

$$\int_I \chi_{K-K_{-\epsilon}}(x) d\eta_I \leq \frac{en\epsilon}{r} \int_I \chi_K(x) d\eta_I.$$

for every $CD(0, n)$ needle $\eta_I$. In the special setting of positive curvature, we need not consider general needles but can restrict ourselves to affine needles:

**Lemma 46 (Reduction to $\kappa = 0$ affine case, Klartag (2017), Fradelizi and Guédon (2004))** *Let $\mu$ be a $CD(0, N)$ needle on $\mathbb{R}$. Consider an integrable, continuous function $f : \mathbb{R} \to \mathbb{R}$ satisfying $\int_M f d\mu = 0$. Then, there exists a partition $\Omega$ of $\mathbb{R}$ and a measure $\nu$ on $\Omega$ and a family of measures $\{\mu_i\}_{i \in \Omega}$ such that*

1. *For any measurable set $A$,*

$$\mu(A) = \int_\Omega \mu_i(A) \, d\nu(i).$$

2. *For almost every $i \in \Omega$, either $i$ is a singleton or $\mu_i$ is a $CD(0, N)$ affine needle.*

3. *For almost every $i$,*

$$\int_i f d\mu_i = 0.$$

From the above lemma, we can take the needle to be an affine needle. The condition $CD\left(0, n\right)$ gives us that $\eta_j$ is the pushforward of $e^{-\phi}$, where

$$\phi'' = \frac{(\phi')^2}{n-1}.$$

It is easy to check that the solutions to this equations are of the form

$$\phi\left(x\right) = a_1 - (N-1)\log\left(a_2 + \frac{x}{n-1}\right)$$

for some constants $a_1$ and $a_2$.

Hence the required needles have density

$$e^{-a_1}\left(a_2 + \frac{x}{n-1}\right)^{n-1}.$$

We can now reduce to the following claim about one-dimensional integrals.

**Lemma 47** *Let $[a, b] \subset \mathbb{R}$ be a interval. Then, for every affine function $c_1 x + c_2$ such that $c_1 x + c_2$ is positive on $[a, b]$ and $\epsilon \leq (b-a)n^{-1}$, we have*

$$\int_a^b (c_1 x + c_2)^{n-1} dx \geq \frac{b-a}{\epsilon n e}\int_b^{b+\epsilon} (c_1 x + c_2)^{n-1} dx$$

**Proof** Using the required change of variables $y = c_1 x + c_2$, we reduce to the following case

$$\int_{c_1 a + c_2}^{c_1 b + c_2} y^{n-1} dx \geq \frac{b-a}{\epsilon n}\int_{c_1 b + c_2}^{c_1 b + c_2 + c_1 \epsilon} y^{n-1}dx,$$

where all the limits of integration are non-negative. Setting $c_1 a + c_2 = a'$, $c_1 b + c_2 = b'$ and $\epsilon' = c_1 \epsilon$, we get

$$\int_{a'}^{b'} y^{n-1} dx \geq \frac{b-a}{\epsilon n}\int_{b'}^{b'+\epsilon'} y^{n-1}dx.$$

Expanding the integrals and taking the ratio, we get the following ratio to upper bound.

$$\frac{(b'+\epsilon')^n - b'^n}{b'^n - a'^n} = \frac{b'^n\left(1 + \frac{\epsilon'}{b'}\right)^n - 1}{b'^n - a'^n}$$

$$= \frac{\left(1 + \frac{\epsilon'}{b'}\right)^n - 1}{1 - \frac{a'^n}{b'^n}}.$$

Note that the last expression is polynomial in $\epsilon'$ with zero constant term.

$$\frac{\left(b' + \epsilon'\right)^n - b'^n}{b'^n - a'^n} = \epsilon' \sum_{i=1}^{n} q_i \epsilon'^{i-1}$$

The second factor on the right hand side is monotone in $\epsilon'$. Setting $\epsilon' = \left(b' - a'\right) n^{-1}$

$$\frac{\left(b' + \epsilon'\right)^n - b'^n}{b'^n - a'^n} \leq \epsilon' \sum_{i=1}^{n} q_i \left(\frac{b' - a'}{n}\right)^{i-1}.$$

Dividing and multiplying by $\left(b' - a'\right) n^{-1}$

$$\frac{\left(b' + \epsilon'\right)^n - b'^n}{b'^n - a'^n} \leq \frac{\epsilon' n}{b' - a'} \sum_{i=1}^{n} q_i \left(\frac{b' - a'}{n}\right)^{i}.$$

Substituting this back into the earlier expression gives

$$\frac{\left(b' + \epsilon'\right)^n - b'^n}{b'^n - a'^n} \leq \frac{\epsilon' n}{b' - a'} \frac{\left(1 + \frac{b'-a'}{nb'}\right)^n - 1}{1 - \frac{a'^n}{b'^n}}$$

Setting $\alpha = b'a'^{-1}$

$$\frac{\left(b' + \epsilon'\right)^n - b'^n}{b'^n - a'^n} \leq \frac{\epsilon' n}{b' - a'} \frac{\left(1 + \frac{1-\alpha}{n}\right)^n - 1}{1 - \alpha^n}.$$

Factorizing the numerator and denominator, we get

$$\frac{\left(b' + \epsilon'\right)^n - b'^n}{b'^n - a'^n} \leq \frac{\epsilon'}{b' - a'} \frac{\sum_{i=0}^{n-1}\left(1 + \frac{1-\alpha}{n}\right)^i}{\sum_{i=0}^{n-1} \alpha^i}.$$

Note that the denominator is greater than one.

$$\frac{\left(b' + \epsilon'\right)^n - b'^n}{b'^n - a'^n} \leq \frac{\epsilon'}{b' - a'} \sum_{i=0}^{n-1}\left(1 + \frac{1-\alpha}{n}\right)^i.$$

Note that $1 - \alpha \leq 1$,

$$\frac{\left(b' + \epsilon'\right)^n - b'^n}{b'^n - a'^n} \leq \frac{\epsilon'}{b' - a'} \sum_{i=0}^{n-1}\left(1 + \frac{1}{n}\right)^i$$

$$\leq \frac{\epsilon'}{b' - a'} \sum_{i=0}^{n-1}\left(1 + \frac{1}{n}\right)^n$$

$$\leq \frac{\epsilon'}{b' - a'} \sum_{i=0}^{n-1} e$$

$$\leq \frac{\epsilon' n e}{b' - a'},$$

completing the proof of Lemma 47. ■

This completes the proof of Theorem 45. ■

We next show the bounds on the conductance of our random walk for manifolds of positive curvature.

**Theorem 48 (Conductance in Positive Curvature)** *Let $(M, g)$ be a Riemannian manifold and $K \subseteq M$ be a strongly convex set satisfying the following conditions.*

- *$M$ has positive sectional curvature i.e. $Sec_M \geq 0$.*

- *The Riemannian curvature tensor is upper bounded i.e. $\max \|R\|_F \leq R$.*

- *$K$ contains a ball of radius $r$.*

*Then, for any $0 \leq s \leq 0.5$, the geodesic walk with step length $\delta^2 \leq \frac{1}{100\sqrt{n}R}$ and $\delta \leq \frac{sr}{4n\sqrt{n}}$ has $s$-conductance*

$$\phi_s \geq \Omega\left(\frac{\delta}{m}\right),$$

*where $m = \frac{1}{\mathrm{Vol}(K)} \int_K d(x^*, y) d\mathrm{Vol}(y)$ for some point $x^*$.*

**Proof** By Theorem 45 , $\mathrm{Vol}\left(K_{-\delta\sqrt{n}}\right) \geq \left(1 - \frac{(e-1)\delta n\sqrt{n}}{r}\right) \mathrm{Vol}(K)$. We are considering step length $\delta\sqrt{n}$ because the length of the Gaussian random variable in $n$ dimensions is $\sqrt{n}$ with high probability. Consider an arbitrary measurable subset $S_1$ and its complement $S_2$. Bounding the conductance of the random walk boils down to bounding the probability of transitioning from $S_1$ to $S_2$. In order to do this, we shall look at points in the sets that have a low chance of crossing over to the other side. Since we are bounding the $s$-conductance, we consider $\mathrm{Vol}(S_1), \mathrm{Vol}(S_2) \geq s\mathrm{Vol}(K)$. Since the walks we consider are geometrical in nature, this corresponds to partitioning the set into well separated subsets. With that in mind, define

$$S_1' = \left\{x \in S_1 \cap K_{-\delta} : P_x(S_2) \leq c_1\right\}$$
$$S_2' = \left\{x \in S_2 \cap K_{-\delta} : P_x(S_1) \leq c_1\right\}.$$

We can assume without loss of generality that $\mathrm{Vol}\left(S_1'\right) \geq 0.5\mathrm{Vol}(S_1)$ and $\mathrm{Vol}(S_2') \geq 0.5\mathrm{Vol}(S_2)$. If it were not, it is easy to see that the conductance of these sets is high. Consider any points $x \in S_1'$ and $y \in S_2'$. Then,

$$d_{TV}\left(P_x, P_y\right) \geq 1 - P_x(S_2) - P_y(S_1)$$
$$\geq 1 - 2c_1.$$

Since $x, y \in K_{-\delta}$, the local conductance $P_x(x), P_y(y) = 0$, applying Theorem 38, we get

$$O\left(\delta^{-1}\right) d(x, y) \geq 1 - 2c_1 - 0.04.$$

Since this is true for an arbitrary pair of points $x, y$, we get the same bound for $d(S_1, S_2)$. Denote by $S_3' = K_{-\delta} \cap \left(S_1' \cup S_2'\right)^c$. We now apply Theorem 35 to the partition of $K_{-\delta}$ given by the sets $S_1'$, $S_2'$ and $S_3'$.

$$\text{Vol}\left(S_3'\right) \geq \Omega\left(\delta\right) \frac{1}{m\text{Vol}\left(K_{-\delta}\right)} \text{Vol}\left(S_2'\right) \text{Vol}\left(S_1'\right)$$

This gives us,

$$
\begin{aligned}
\int_{S_1} P_x\left(S_2\right) dx &\geq \frac{c_1}{2} \frac{\text{Vol}(S_3')}{\text{Vol}(K)} \\
&\geq \Omega\left(\delta\right) \frac{\text{Vol}\left(S_2'\right) \text{Vol}\left(S_1'\right)}{m\text{Vol}(K)\text{Vol}\left(K_{-\delta}\right)} \\
&\geq \Omega\left(\delta\right) \frac{\min\{\text{Vol}\left(S_2'\right), \text{Vol}\left(S_1'\right)\}}{m\text{Vol}(K)}.
\end{aligned}
$$

Here, $m = \int_K d(x^*, x)d\text{Vol}(x)$ for some point $x^*$. Note that $m \leq D$ where $D$ is the diameter of the set. This show the required bound on the conductance that we require. ∎

The theorem below shows that Markov chains with high conductance have small mixing times as required. Before that we state a lemma.

**Lemma 49 ($s$-Conductance Implies Fast Mixing , see Vempala (2005))** *Let $\mu_t$ be the distribution of the random walk after $t$ steps starting at $X_0$, distributed as $\mu_0$. Let $H_s = \sup\left\{\left|\mu_0\left(A\right) - \mu(A)\right| : \mu\left(A\right) \leq s\right\}$ where $\mu$ is the stationary distribution of the random walk. Then,*

$$d_{TV}\left(\mu_t, \mu\right) \leq H_s + \frac{H_s}{s}\left(1 - \frac{\phi_s^2}{2}\right)^t.$$

**Theorem 3** *[Mixing Time Bound] Let $(M, g)$ be an $n$-dimensional Riemannian manifold and $K \subset M$ be a strongly convex subset satisfying the following conditions:*

- *$M$ has non-negative sectional curvature i.e. $Sec_M \geq 0$.*

- *The Riemannian curvature tensor is upper bounded i.e. $\max\|R_M\|_F \leq R$.*

- *$K$ contains a ball of radius $r$.*

- *$K$ has diameter $D$.*

*Let the starting distribution be an $H$-warm start (see Definition 26). Then, in*

$$t = O\left(\frac{H^2 D^2 n^3\left(R + 1\right)}{r^2 \epsilon^2} \log\left(\frac{H}{\epsilon}\right)\right)$$

*steps, the distribution of $X_t$, the output of Algorithm 1 is $\epsilon$-close to the uniform distribution on convex set in total variation.*

**Proof** First note that $H_s \leq Hs$. By Theorem 48, we have that for $\delta^2 = O\left(\frac{1}{\sqrt{n}R}\right)$ and $\delta \leq \frac{sr}{4n\sqrt{n}}$ ,

we have $\phi^2 = \Omega\left(\frac{s^2 r^2}{m^2 n^3 \max\{1,R\}}\right)$ and thus from Theorem 49, we have for each $s = \epsilon_1 H^{-1}$,

$$d_{TV}\left(\mu_t, \mu\right) \leq \epsilon_1 + H\left(1 - \Omega\left(\frac{s^2 r^2}{m^2 n^3 \max\{1, R\}}\right)\right)^t .$$

Picking $\epsilon_1 = \epsilon/2$ and $t = O\left(\frac{H^2 m^2 n^3 \max\{1,R\}}{r^2 \epsilon^2} \log\left(H/\epsilon\right)\right)$, we get the desired result. ∎

If we assume a strictly positive lower bound on the Ricci curvature, then by the Bonnet–Myers Theorem we get an upper bound on the diameter of the manifold and thus on the diameter of the convex set.

**Fact 50 (Bonnet–Myers Theorem, see Chavel (2006))** *Let $M$ be a complete Riemannian manifold, of dimension $n \geq 2$, such that there exists a constant $\kappa > 0$ for which*

$$Ric_M \geq (n-1)\kappa.$$

*Then, $M$ is compact with diameter lesser than $\pi/\sqrt{\kappa}$.*

This gives us a diameter-independent bound on mixing time. This can also be achieved by considering diameter independent isoperimetric inequalities under the strict positivity assumption.

With slightly different assumption on the convex subset, we can show bounds on the conductance rather than the $s$-conductance. The proof of the theorem is the same as that of the theorem above but since each point has high conductance, we do not need to restrict to the large sets.

**Theorem 51 (Conductance in Positive Curvature with assumptions on local conductance)** *Let $(M, g)$ be a Riemannian manifold and $K \subseteq M$ be a strongly convex set satisfying the following conditions.*

- *$M$ has non-negative sectional curvature i.e. $Ric_M \geq 0$.*

- *The Riemannian curvature tensor is upper bounded i.e. $\max\|R\|_F \leq R$.*

- *$K$ has diameter $D$.*

- *Each point in $K$ has local conductance greater than $c_2$.*

*Then, the geodesic walk with step length $\delta^2 \leq \frac{1}{100\sqrt{n}R}$ has*

$$\phi \geq \Omega\left(\frac{\delta\left(1 - c_2 - 0.04\right)^2}{m}\right).$$

*where $m = \frac{1}{\mathrm{Vol}(K)} \int_K d(x^*, x) d\mathrm{Vol}(x)$ for some point $x^*$.*

**Proof** The proof is similar to that of Theorem 48. In order to bound the conductance, we shall look at points in the sets that have a low chance of crossing over to the other side. Since the walks we consider are geometrical in nature, this corresponds to partitioning the set into well separated subsets. With that in mind, define

$$S_1' = \left\{ x \in S_1 : P_x\left(S_2\right) \leq c_1 \right\},$$
$$S_2' = \left\{ x \in S_2 : P_x\left(S_1\right) \leq c_1 \right\}.$$

We can assume without loss of generality that $\text{Vol}\left(S_1'\right) \geq 0.5\text{Vol}(S_1)$ and $\text{Vol}(S_2') \geq 0.5\text{Vol}(S_2)$. If it were not, it is easy to see that the conductance of these sets is high. Consider any points $x \in S_1'$ and $y \in S_2'$. Then,

$$
\begin{aligned}
d_{TV}\left(P_x, P_y\right) &\geq 1 - P_x\left(S_2\right) - P_y\left(S_1\right) \\
&\geq 1 - 2c_1.
\end{aligned}
$$

Since, from the assumption, the local conductance $P_x(x)$, $P_y(y) \leq c_2$, applying Theorem 38, we get

$$O\left(\delta^{-1}\right) d\left(x, y\right) \geq 1 - c_2 - 2c_1 - 0.04.$$

Since this is true for an arbitrary pair of points $x, y$, we get the same bound for $d(S_1, S_2)$. Let $S_3' = K \cap \left(S_1' \cup S_2'\right)^c$. We now apply Theorem 35 to the partition of $K$ given by the sets $S_1'$, $S_2'$ and $S_3'$ with $c_1 = \left(1 - c_2 - 0.04\right)/4$:

$$\text{Vol}\left(S_3'\right) \geq \Omega\left(\delta\right) \frac{1}{m\text{Vol}\left(K\right)} \text{Vol}\left(S_2'\right) \text{Vol}\left(S_1'\right).$$

This gives us,

$$
\begin{aligned}
\int_{S_1} P_x\left(S_2\right) dx &\geq \frac{c_1}{2} \frac{\text{Vol}(S_3')}{\text{Vol}(K)} \\
&\geq \Omega\left(\left(1 - c_2 - 0.04\right)^2 \delta\right) \frac{\text{Vol}\left(S_2'\right) \text{Vol}\left(S_1'\right)}{m\text{Vol}\left(K\right)^2} \\
&\geq \Omega\left(\left(1 - c_2 - 0.04\right)^2 \delta\right) \frac{\min\{\text{Vol}\left(S_2'\right), \text{Vol}\left(S_1'\right)\}}{m\text{Vol}(K)}.
\end{aligned}
$$

Note that $m \leq D$ where $D$ is the diameter of the set. This show the required bound on the conductance that we require. ∎

**Lemma 52 (Conductance Implies Fast Mixing , see Vempala (2005))** *Let $\mu_t$ be the distribution of the random walk after $t$ steps starting at $X_0$, distributed as $\mu_0$, with $H = \left\|\mu_0/\mu\right\|$. Then, for each $\epsilon > 0$,*

$$d_{TV}\left(\mu_t, \mu\right) \leq \epsilon + \sqrt{\frac{H}{\epsilon}} \left(1 - \frac{\phi^2}{2}\right)^t.$$

Putting together the previous theorem and lemma we get:

**Theorem 53 (Mixing Time Bound with assumptions on local conductance)**  *Let $(M, g)$ be a manifold satisfying the conditions in [Theorem 51](#) and let $K \subset M$ be a strongly convex subset. Let the starting distribution be an $H$-warm start, then for $\delta^2 \leq \frac{1}{100\sqrt{n}R_1}$ and*

$$t = O\left(\frac{m^2}{\delta^2 \left(1 - c_2 - 0.04\right)^4} \log\left(\frac{H}{\epsilon}\right)\right)$$

*steps, the distribution $\mu_t$ is $\epsilon$ close to the uniform distribution on convex set in total variation.*

## Appendix G.  Reduction from Sampling to Optimization for Positive Curvature

In this section, we focus on the algorithms for solving convex programs on convex sets on manifolds. That is, given access to a membership oracle to a strongly convex set $K \subseteq M$, an evaluation oracle for a convex function $f : K \to \mathbb{R}$ and an error parameter $\epsilon$, the algorithm needs to output a point $y \in K$ such that

$$f(y) - \min_x f(x) \leq \epsilon.$$

Towards this goal, we adapt the simulated annealing algorithm from the Euclidean setting to the Riemannian setting. Given a function $f$ and a "temperature" $T$, define the density

$$\pi_{f,T} \sim e^{-\frac{f}{T}}.$$

Intuitively, the function puts more weight on points where the function attains small values and sampling from the distribution is likely to output points near the minima of the function $f$ for low enough temperature $T$. The issue is that sampling from the distribution for a low enough temperature is a priori as hard as solving the initial optimization problem. The way to get around this issue to set a temperature schedule in which one progressively "cools" the distribution so that the sample from the previous temperature acts to make it easier to sample from the next temperature. Once we reach a low enough temperature, the sample we attain will be close to the optimum point with high probability.

Since we need to sample from a distribution proportional to the $e^{-\frac{f}{T}}$, the natural idea would be to use the Metropolis filter with respect to the original uniform sampling algorithm. This natu-

rally leads to the following algorithm for sampling from the required distribution (this is same as Algorithm 2 reproduced here for convenience):

---

**Algorithm 4:** Adapted Geodesic Walk on a Manifold $M$ and function $f$.

---

**Input:** Dimension $n$, Convex set $K$, Convex function $f : K \to \mathbb{R}$, Starting Point $X_0 \in K$,
      Step Size $\eta$, Number of Steps $N$.

**for** $\tau \leq N$ **do**

    Pick $u_{\tau+1} \leftarrow N(0, I)$ in $T_{X_i} M$ where the covariance is with respect to the metric.

    **if** $y = \exp_{X_\tau}(\eta \, u_{\tau+1}) \in K$ **then**

        With probability $\min\left(1, e^{-f(y)+f(X_\tau)}\right)$, set $X_{\tau+1} \leftarrow y$.

        With the remaining probability, set $X_{\tau+1} \leftarrow X_\tau$.

    **else**

        Set $X_{\tau+1} \leftarrow X_\tau$.

    **end**

**end**

**Output:** Point $X_N \in K$ sampled approximately proportional to $e^{-f}$.

---

We analyze the above algorithm over manifolds with non-negative Ricci curvature. This algorithm serves as a step in simulated annealing. The sequence of temperatures $T_0, T_1, \ldots$ used by simulated annealing is called the temperature schedule or the cooling schedule.

---

**Algorithm 5:** Simulated Annealing on Manifold $M$.

---

**Input:** Dimension $n$, Convex set $K$, Convex function $f : K \to \mathbb{R}$, Starting Point $X_0 \in K$,
      Number of Iterations $N$, Temperature Schedule $T_i$.

**for** $\tau \leq N$ **do**

    Sample $X_\tau$ according to distribution $\pi_{f,T_i}$ using algorithm 4 for $\pi_{f,T_i}$ with starting point
    $X_{\tau-1}$.

**end**

**Output:** Point $X_N \in K$ that approximately minimizes $f$.

---

The main thing to specify in the design of the algorithm above is the temperature schedule, that is, the sequence of temperatures $T_i$ from which we sample. We need to set the schedule so that the distributions $\pi_{f,T_i}$ and $\pi_{f,T_{i+1}}$ are close for each $i$, yet keeping the length of the schedule small. Following Kalai and Vempala (2006), we use

$$T_{i+1} = \left(1 - \frac{1}{\sqrt{n}}\right) T_i$$

which provides us with the required guarantees.

Note that since we are interested in optimizing convex functions, we need to sample from the analogues of log-concave distributions. For showing mixing of the random walk, we follow the same technique as in the uniform case. That is, we show that the walk has high conductance by reducing the conductance of the walk to isoperimetry of the set with respect to the required measure. Towards proving isoperimetry, we adapt Theorem 35 to the above setting again using Theorem 32, but now with respect to a weighted Riemannian manifold. Since the measure of interest has density which is an exponential of a convex function, the question reduces to the positive curvature case.

**Theorem 54** *Let $M$ be a manifold with positive curvature and let $K$ be a strongly convex subset. Let $g : K \to \mathbb{R}$ be a geodesically convex function and consider the measure with density proportional to*

$$\pi_g \sim e^{-g},$$

*with respect to the Riemannian volume form. Then for any disjoint $K_1, K_2, K_3$ such that $K = \cup_i K_i$ and $d(K_1, K_3) \geq \epsilon$,*

$$\frac{m}{c\epsilon} \pi_g(K) \pi_g(K_2) \geq \pi_g(K_1) \pi_g(K_3),$$

*where $m = \int_K d(x, y) d\pi_g(y)$ for some $x$.*

**Proof** We apply Theorem 32 with $d\mu = e^{-g} d\mathrm{Vol}$. The generalized Ricci curvature then becomes

$$Ric_{\mu,N} = Ric_M + Hess_g.$$

Since $g$ is convex, we have $Hess_g(v, v) \geq 0$ for all $v \in TM$. Thus,

$$Ric_{\mu,N} \geq 0$$

as in the setting of Theorem 35. ∎

Next, we extend the one step overlap of the distribution to the current case. We assume smoothness on the function in the following lemma. We remark that the technique from Lovász and Vempala (2007) to smoothen the function by taking local averages is likely to work obviating the need for smoothness assumption though we cannot prove it.

**Theorem 55** *Let $g : M \to \mathbb{R}$ be convex and $L$-Lipschitz, and let $P_x^g$ be the one step distribution induced by the walk in Algorithm 4 starting at point $x \in M$. Let $x, y \in M$ be points such that $P_x(x), P_y(y) \leq c_2$ for some constant $c_2 \leq 1$, then for $\delta^2 = O\left(\frac{1}{\sqrt{n}(1+R)L^2}\right)$, we have*

$$d_{TV}\left(P_x^g, P_y^g\right) \leq c_2 + 0.01 + O\left(\frac{1}{\delta}\right) d(x, y) + \frac{1}{25}$$

**Proof** As before, we account for the probability of rejection in a step. The rejection probability because of the convex set is given as before by

$$P_x(x) \leq c_2.$$

We need to account for the probability of rejection because of the additional Metropolis filter given by the function. The probability of rejection is upper bounded by

$$1 - e^{-\left|g(y) - g(y')\right|}$$

where $y'$ is proposed next point. Note that since $g$ is Lipschitz, we have

$$\left|g\left(y'\right) - g\left(y\right)\right| \leq Ld\left(y, y'\right).$$

Thus, the probability of rejection is upper bounded by

$$1 - e^{-Ld\left(y, y'\right)}.$$

38

From the definition of the walk, we can take $d\left(y, y'\right) \leq O\left(\delta\sqrt{n}\right)$ (the probability that larger vectors are drawn can be absorbed into the total variation distance of the one step distribution without the filter). So the rejection probability is bounded above by

$$1 - e^{-L\delta\sqrt{n}}.$$

We get this to be a constant by setting $\delta^2 = O\left(\frac{1}{n(1+R)L^2}\right)$ assuming $L > 1$. Having bounded the probability of rejection, we can use Theorem 38 to bound the total variation distance. ∎

With the above theorems, we get the following bound for the mixing time.

**Theorem 56 (Mixing Time for Sampling from Gibbs distributions)**  *Let $M$ be a manifold and $K \subseteq M$ be a strongly geodesically convex subset. Let $g : K \to \mathbb{R}$ be a geodesically convex function on $K$ with Lipschitz constant $L$. Let $\pi_{g,T}(x) \sim e^{-\frac{g}{T}}$ and let $m_T = \int_K d(x, x^*)d\pi_{g,T}(x)$ for some $x^* \in K$. Let $r$ be the radius of the largest ball contained within $K$. Let $\mu_t$ be the distribution of the adapted geodesic walk for $\pi_{g,T}$ after $t$ steps, starting from a distribution that is $H$-warm for $\pi_{g,T}$. Then, for*

$$k = O\left(\frac{H^2 m^2 n^3 \left(1 + R\right) L^2}{r^2 T^2 \epsilon^2} \log\left(H/\epsilon\right)\right)$$

*steps, $\mu_k$ is $\epsilon$-close to $\pi_{g,T}$ in total variation.*

Next, we show that sampling from the distribution that we get does indeed give us a point close to the required optimum. That is, for a small enough temperature $T$, the expected value of the function under the distribution is small. Using Markov's inequality, one can then bound the probability that a sample produces a value much larger than the optimal value. In the Euclidean case, this inequality was shown by Kalai and Vempala (2006) for the case of linear functions and was explicitly shown for general convex functions in Belloni et al. (2015).

**Theorem 57**  *Let $K \subseteq M$ be a strongly convex subset of manifold $M$ with non-negative Ricci curvature. Let $g : K \to \mathbb{R}$ be a convex function with minimum value zero. Let $X$ be sampled from $\pi_{g,T}$. Then,*

$$\mathbb{E}_{\pi_{g,T}}\left(g(X)\right) \leq T\left(n + 1\right) + \min_{x \in K} f\left(x\right).$$

**Proof** In order to prove the above theorem, we use a localization technique similar to the one used in Theorem 45. We note the above theorem in terms of integrals in the next theorem, to make it amenable to localization.

**Theorem 58**  *Let $K \subseteq M$ be a convex subset of the manifold $M$ satisfying the curvature dimension condition $CD(0, N)$. For any convex function $g : K \to \mathbb{R}$ with $\min g(x) = 0$, we have*

$$\int_K e^{-g}g \leq (N + 1) \int_K e^{-g}.$$

**Proof** For the sake of contradiction, assume that the inequality is false. Since $g$ is a convex function, it has a unique minimizer $x^*$ in $K$. By assumption, we have $g(x^*) = 0$. Consider the 1-Lipschitz function $u$ defined by

$$u(x) = d\left(x^*, x\right).$$

Note that the strain sets of $u$ are geodesic segments through the minimizing point $x^*$. Applying Theorem 44 with respect to $u$, we get a decomposition of measure $\eta_I$. Note that the support of $\eta_I$ is a geodesic passing through $x^*$. From the decomposition, we get that for some needle $\eta_j$,

$$\int_j \left( e^g g - (N+1) e^g \right) d\eta_j \geq 0.$$

Note that $\eta_j$ is a $CD(0, N)$ needle. We now note the reduction to the affine case.

**Lemma 59 (Reduction to $\kappa = 0$ affine case, Klartag (2017), Fradelizi and Guédon (2004))** *Let $\mu$ be a $CD(0, N)$ needle on $\mathbb{R}$. Consider an integrable, continuous function $f : \mathbb{R} \to \mathbb{R}$ satisfying $\int_M f d\mu = 0$. Then, there exists a partition $\Omega$ of $\mathbb{R}$ and a measure $\nu$ on $\Omega$ and a family of measure $\{\mu_i\}_{i \in \Omega}$ such that*

1. *For any measurable set $A$,*

$$\mu(A) = \int_\Omega \mu_i(A) \, d\nu(i).$$

2. *For almost every $i \in \Omega$, either $i$ is a singleton or $\mu_i$ is a $CD(0, N)$ affine needle.*

3. *For almost every $i$,*

$$\int_i f d\mu_i = 0$$

From the above lemma, we can take the needle to be an affine needle. The condition for $CD(0, N)$, gives us that $\eta_j$ is the pushforward of $e^{-\phi}$, where

$$\phi'' = \frac{(\phi')^2}{N-1}.$$

It is easy to check that the solutions to this equations are of the form

$$\phi(x) = a_1 - (N-1) \log \left( a_2 + \frac{x}{N-1} \right)$$

for some constants $a_1$ and $a_2$. Thus, we have

$$e^{-\phi(x)} = e^{-a_1} \left( a_2 + \frac{x}{N-1} \right)^{N-1}.$$

Thus, we get

$$\int_J \left( e^g g - (N+1) e^g \right) d\eta_j = \int_{\gamma^{-1} j} e^{-a_1} \left( a_2 + \frac{x}{N-1} \right)^{N-1} \left( e^{g(\gamma(t))} g(\gamma(t)) - (N+1) e^{g(\gamma(t))} \right) dt$$

With appropriate change of variables, we get the integral to be

$$\int_a^b t^{N-1} \left( e^{h(t)} h(t) - (N+1) e^{h(t)} \right) dt.$$

Here $h$ is the convex function gotten from $f$ after the affine change of variables. Note that this is an integral over the real numbers, which can be dealt with using elementary tools. We capture this with the lemma below.

**Lemma 60** *Let $h : [a, b] \to \mathbb{R}$ be a convex function with minimum value zero. Then,*

$$\int_a^b e^{-h(z)}h(z)z^{n-1}dz \leq (n+1)\int_a^b e^{-h(z)}z^{n-1}dz.$$

**Proof** Consider the truncated convex cone $K_1$ in $\mathbb{R}^n$ with radii $a$ and $b$. Let $x_1$ denote the coordinate along the axis of the cone, and let the centres of the circles defining the cone be on the $x_1$ axis at $x_1 = a$ and $x_1 = b$ respectively. Consider the convex function $c(x) = h(x_1)$. Note that

$$\int_{K_1} e^{-c(x)}c(x)dx = \int_a^b e^{-h(z)}h(z)z^{n-1}dz.$$

and

$$\int_{K_1} e^{-c(x)}dx = \int_a^b e^{-h(z)}z^{n-1}dz.$$

The required result now follows from the results in Kalai and Vempala (2006) and Belloni et al. (2015). ∎

From Theorem 60, we get a contradiction. Thus, we have the required bound. ∎

Now, we use the above theorem with $N = n$, the dimension of the manifold to get the desired result. ∎

Given that we have shown that we can sample from a fixed temperature, we need to show that sampling from a distribution gives a warm start to the next distribution. We do this by adapting the analysis of log-concave distributions to the manifold case using the localization lemma from Klartag (2017).

**Lemma 61 (adaptation to manifolds of Kalai and Vempala (2006), Lovász and Vempala (2006))** *Let $K$ be a strongly convex subset of $M$ and $f : K \to \mathbb{R}$ be a convex function. For $a \geq 0$, let*

$$Z(a) = \int_K e^{-af(x)}dx.$$

*Then,*

$$Z(a)Z(b) \leq Z\left(\frac{a+b}{2}\right)^2 \left(\frac{(a+b)^2}{4ab}\right)^n.$$

**Proof** We prove this inequality by using the four function theorem. Set $f_1 = e^{-af}$, $f_2 = e^{-bf}$ and $f_3 = e^{-\frac{a+b}{2}f}$ and $f_4 = \left(\frac{(a+b)^2}{4ab}\right)^n e^{-\frac{a+b}{2}f}$. Note that $f_1 f_2 \leq f_3 f_4$ from the AM-GM inequality. Then, from Theorem 59, we have that the above theorem is true if the corresponding inequality is true for $CD(0, N)$ affine needles. From the above argument, we see that this reduces to showing

$$\int_c^d e^{-ah(x)}x^{n-1}dx \int_c^d e^{-bh(x)}x^{n-1}dx \leq \left(\frac{(a+b)^2}{4ab}\right)^n \left(\int_a^b e^{-\frac{a+b}{2}h(x)}dx\right)^2.$$

Here $h$ is the convex function obtained from $f$ after the affine change of variables. The last inequality follows from the inequality in $\mathbb{R}^n$. ∎

**Theorem 62** *Let $T_i$ and $T_{i+1}$ be adjacent temperatures in the algorithm. Then,*

$$\left\| \pi_{g,T_i}/\pi_{g,T_{i+1}} \right\| \leq 5.$$

**Proof** Denote by $\beta_i$ the inverse of $T_i$. From the definition of the norm, we get

$$\left\| \pi_{g,T_i}/\pi_{g,T_{i+1}} \right\| = \frac{Z(\beta_{i+1})}{Z(\beta_i)^2} \int e^{-g(x)\beta_i} e^{\beta_{i+1}g(x)} e^{-g(x)\beta_i} dx$$
$$= \frac{Z(\beta_{i+1})Z(2\beta_i - \beta_{i+1})}{Z(\beta_i)^2}.$$

We use the fact that $a^n Z(a)$ is log-concave to get,

$$\left\| \pi_{g,T_i}/\pi_{g,T_{i+1}} \right\| \leq \left( \frac{\beta_i^2}{(2\beta_i - \beta_{i+1})\beta_{i+1}} \right)^n.$$

By the choice of the temperature schedule, we get,

$$\left\| \pi_{g,T_i}/\pi_{g,T_{i+1}} \right\| \leq \left( \frac{\left(1 - 1/\sqrt{n}\right)^2}{2\left(1 - 1/\sqrt{n}\right) - 1} \right)^n$$
$$\leq \left( 1 + \frac{1}{n - 2\sqrt{n}} \right)^n$$
$$\leq e^{n/(n-2\sqrt{n})}$$
$$\leq 5.$$

$\blacksquare$

Given the above theorem, we write down the running time bound for the simulated annealing algorithm.

**Theorem 4** *[Simulated Annealing] Let $(M, g)$ be a manifold with non-negative Ricci curvature. Let $K \subseteq M$ be a strongly convex set satisfying the requirements in [Theorem 3](#). Let $f : K \to \mathbb{R}$ be a geodesically convex function with Lipschitz constant $L$. Then, starting from a uniform sample from $K$, [Algorithm 3](#) runs for*

$$O\left( \frac{D^2 n^{7.5}(R+1)L^2}{r^2\epsilon^2\delta^6} \log\left( \frac{n}{\delta} \log\left( \frac{T_0(n+1)}{\epsilon\delta} \right) \right) \log^5\left( \frac{T_0 n}{\epsilon\delta} \right) \right)$$

*steps and with probability $1 - \delta$ outputs a point $x^*$ such that*

$$f\left(x^*\right) - \min_x f(x) \leq \epsilon.$$

**Proof** We start with a temperature such that the uniform distribution $\nu$ on the convex set satisfies $\left\| \nu/\mu_0 \right\| \leq c_5$ for some constant $c_5$. From the choice of temperature schedule, we get that after

$I = \sqrt{n} \log \left( \frac{T_0(n+1)}{\epsilon\delta} \right)$ rounds of cooling, we are at the temperature $\frac{\epsilon\delta}{(n+1)}$. Then, from Theorem 57, we get

$$\mathbb{E}\left[ f\left( z \right) \right] \leq \min_x f(x) + \epsilon\delta.$$

Using Markov inequality, we have that with probability $1 - \delta$, we are within $\epsilon$ of the minimum value. In each phase $i$, we sample such that the distribution we sample from is at most $\frac{\delta}{100I}$ in total variation. From Theorem 56, this requires

$$L_i = O\left( \frac{H^2 I^2 m_{T_i}^2 n^3 \left(1 + R\right) L^2}{r^2 T_i^2 \delta^2} \log\left( \frac{HI}{\delta} \right) \right)$$

steps starting from a $H$-warm start. From Theorem 62, we can take each adjacent distribution have $\left\| \pi_{g,T_i}/\pi_{g,T_{i+1}} \right\| \leq 5$. Using a standard trick (see for example, Corollary 3.5 in Vempala (2005)), we can assume that this is a $2000I/\delta$-warm start with probability $1 - \delta/2000I$ and with probability $\delta/2000I$, it is sampled from an arbitrary distribution. This gives us

$$L_i = O\left( \frac{I^4 m_{T_i}^2 n^3 \left(1 + R\right) L^2}{r^2 T_i^2 \delta^4} \log\left( \frac{I}{\delta} \right) \right)$$

Note that $m_{T_i} \leq D$, where $D$ is the diameter of the set $K$ and $T_i \geq \frac{\epsilon\delta}{2n}$, which gives us

$$L_i = O\left( \frac{D^2 n^7 \left(1 + R\right) L^2}{r^2 \epsilon^2 \delta^6} \log\left( \frac{n}{\delta} \log\left( \frac{T_0 \left(n+1\right)}{\epsilon\delta} \right) \right) \log^4\left( \frac{T_0 n}{\epsilon\delta} \right) \right)$$

as required. ∎