

RAPID MIXING OF GEODESIC WALKS ON MANIFOLDS WITH POSITIVE CURVATURE

BY OREN MANGOUBI¹ AND AARON SMITH²

École Polytechnique Fédérale de Lausanne (EPFL) and University of Ottawa

We introduce a Markov chain for sampling from the uniform distribution on a Riemannian manifold \mathcal{M} , which we call the *geodesic walk*. We prove that the mixing time of this walk on any manifold with positive sectional curvature $C_X(u, v)$ bounded both above and below by $0 < m_2 \leq C_X(u, v) \leq M_2 < \infty$ is $\mathcal{O}^*(\frac{M_2}{m_2})$. In particular, this bound on the mixing time does not depend explicitly on the dimension of the manifold. In the special case that \mathcal{M} is the boundary of a convex body, we give an explicit and computationally tractable algorithm for approximating the exact geodesic walk. As a consequence, we obtain an algorithm for sampling uniformly from the surface of a convex body that has running time bounded solely in terms of the curvature of the body.

1. Introduction. Sampling from manifolds has applications to areas such as statistics [6, 11], computer graphics [32], optimization [12] and systems biology [36]. For a simple example, manifolds with nonnegative curvature appear in statistical applications as the level sets of log-concave, or more generally quasiconcave, distributions (see Remark 1). Many natural distributions in statistics are quasiconcave or log-concave, including the “ridge regression” posterior associated with Gaussian priors for logistic regression [14, 15]. Existing computer science literature shows that the problem of sampling from manifolds with nonnegative curvature is tightly linked to the problems of sampling from and computing integrals of quasiconcave distributions [4, 25, 26, 29]. The recent paper [11] has a broader survey on this sampling problem in various fields.

In this paper, we study a simple Markov chain, which we call the *geodesic walk*, that can sample from the uniform distribution of a general Riemannian manifold \mathcal{M} . Our main result is that this walk mixes quickly when \mathcal{M} has positive bounded curvature. We also give results concerning an efficient implementation of this walk in the important special case that \mathcal{M} is the boundary of a convex set.

Received March 2017; revised October 2017.

¹Supported by a Canadian Statistical Sciences Institute (CANSSI) Postdoctoral Fellowship, and by an NSERC Discovery grant.

²Supported by an NSERC Discovery grant.

MSC2010 subject classifications. Primary 60J05; secondary 68W20, 60J20, 53D25.

Key words and phrases. Markov chain Monte Carlo (MCMC), Hamiltonian Monte Carlo (HMC), manifolds, convex bodies, positive curvature, momentum.

REMARK 1. We give a technical caveat: a level set of a quasiconcave distribution need not be smooth, and in particular may not have well-defined sectional curvature. However, there is a standard extension of the notion of sectional curvature to this situation, called the Alexandrov curvature. It turns out that the Alexandrov curvature of the level set of a quasiconcave distribution is always nonnegative [2]. It is in this sense that we say that the compact level sets of quasiconcave distributions have nonnegative curvature. We now restrict our attention to situations in which the sectional curvature is well defined, until a brief discussion in Section 9.

1.1. *Main results.* The geodesic walk, defined precisely in Algorithm 1, is a Markov chain $\{X_i\}_{i \in \mathbb{N}}$ that is well defined on any Riemannian manifold \mathcal{M} . It evolves by selecting, at each time step i , a random tangent vector U_i in the tangent space \mathcal{T}_{X_i} of \mathcal{M} at X_i , and then following the associated geodesic for some period of time. The geodesic walk is a natural Markov chain on a manifold, and it is somewhat similar to the well-studied “ball walk” on a manifold (see, e.g., [22]). It is well known (see Example 7 of [31]) that the ball walk mixes rapidly on manifolds with positive curvature. The first main result of this paper, Theorem 7.1, is that the geodesic walk also mixes rapidly on a manifold with sectional curvature bounded above and below by two constants $0 < m_2 \leq C_x(u, v) \leq M_2 < \infty$.

Although the ball walk is easy to define and mixes rapidly, it is generally not practical to implement a version of the ball walk with the desired stationary distribution on a computer, and so it does not give rise to practical sampling algorithms (see Example 3C of [11], titled “*How not to sample*”). In contrast, we show that the geodesic walk automatically has the correct stationary distribution. Our second main result (Theorems 4.1 and 6.2) is that, unlike the ball walk, the geodesic walk has the uniform measure on the manifold as its unique stationary distribution.

The geodesic walk is also much easier to implement than the ball walk. Our third main result, Theorem 8.1, shows that any reasonably good algorithm for approximating geodesics can be used to simulate an “approximate” geodesic walk that mixes quickly and has near-uniform stationary measure. We also show that Algorithm 2, which is straightforward to implement, gives an efficient “approximate” geodesic walk in this sense.

We believe our $\mathcal{O}^*\left(\frac{M_2}{m_2}\right)$ bound on the mixing time of the geodesic walk is the first dimensionless mixing time bound for an implementable Markov chain with uniform stationary distribution on bounded-positive curvature manifolds in general, and convex body boundaries in particular. One reason our bound does not depend on the dimension is that, in contrast to previously-proposed algorithms, the geodesic walk can take long steps whose length is independent of the dimension. For comparison, the best existing bound that we know of is given for the “billiards walk” studied in [12], which is shown to have mixing time of $\mathcal{O}^*(d^2 \cdot \frac{M_2}{m_2})$ (see Remark 2). That walk cannot mix quickly in high dimensions, even if \mathcal{M} is the surface of a sphere, because the typical step size is $\mathcal{O}\left(\frac{1}{\sqrt{d}}\right)$.

In both statistics and computer science, one of the most important and well-studied special cases occurs when \mathcal{M} is the boundary $\partial\mathcal{K}$ of a convex body $\mathcal{K} \subset \mathbb{R}^{d+1}$. There are many Markov chains for sampling from the uniform distribution on the boundary $\partial\mathcal{K}$ of a set under various assumptions (see, e.g., [3, 9, 12]), all of which involve a walk that moves inside of \mathcal{K} and then reflects off of $\partial\mathcal{K}$. In Section 8, we explain how our geodesic walk can be well approximated by such a “reflecting walk,” if the reflecting walk is allowed to have long-term momentum. Our final main result, Theorem 8.2, shows that this approximate version of the geodesic walk gives us an algorithm for sampling uniformly from the boundary $\partial\mathcal{K}$ of a convex set \mathcal{K} with arbitrarily small error $\varepsilon > 0$ in roughly $\mathcal{O}^*((\frac{m_2}{m_2})^3 \varepsilon^{-1})$ reflections, provided that \mathcal{M} has inner and outer radii of curvature bounded below and above by $\frac{1}{\sqrt{m_2}}$ and $\frac{1}{\sqrt{m_2}}$, respectively. We believe the $\mathcal{O}^*((\frac{m_2}{m_2})^3)$ bound for the implementation of the geodesic walk is often very pessimistic and can be greatly improved in some important special cases (see Remark 7 in Section 8.2). Furthermore, in situations where one has access to better geodesic approximations, such as when symplectic integrators or exact geodesic integrators are available, our results imply computational costs that are much closer to the mixing time $\mathcal{O}^*(\frac{m_2}{m_2})$ of the true geodesic walk.

REMARK 2. Note that the bound in [12] is better for chains with a bound on the diameter of \mathcal{K} that is much smaller than $\frac{2}{\sqrt{m_2}}$, while our bound is better in higher dimensions and when the bound on the inner radius of curvature of \mathcal{K} is much smaller than $\frac{1}{\sqrt{m_2}}$. Also, a practical implementation of the geodesic walk requires additional computational costs that depend on the method used to approximate geodesics.

1.2. *Proof techniques.* The main component of our proof is a coupling argument, where we couple the initial velocity by parallel transport (Section 5). We then use comparison theorems from differential geometry [1, 7, 35] to show that our assumptions of positive curvature bounds imply that the distance between the two chains contracts in expectation over each step in the Markov chain (Section 6). This contraction estimate immediately implies a bound on the Wasserstein mixing time and other relevant quantities [31] (Section 7).

REMARK 3 (Short and long steps). The geodesic walk has an important tuning parameter: the amount of time T to follow the geodesic. If T is very small, then we can couple two copies $\{X_i, Y_i\}_{i \in \mathbb{N}}$ via parallel transport of the momentum vector U_i used at each step (see Definition 1 for a precise definition) to obtain the following contraction estimate (see Proposition 6 of [31]):

$$\mathbb{E}[d(X_{i+1}, Y_{i+1}) | X_i, Y_i] \leq \left(1 - \frac{m_2}{8} T^2 + O(T^3 + T^2 d(X_i, Y_i))\right) d(X_i, Y_i).$$

In other words, for very small step size T , a lower bound on the sectional curvature implies that the geodesic walk contracts. This bound does not (directly) imply useful bounds for the geodesic walk with T large, since the error term in this inequality may become very large.

After studying the mixing properties of the geodesic walk, we study a possible implementation. Using comparison bounds for Markov chains, we show that one can approximate geodesic trajectories with arbitrary accuracy in a dimension-independent number of steps provided one has access to an appropriate oracle, allowing one to generate samples from a stationary distribution that is arbitrarily close to uniform (in Wasserstein transportation distance) in a dimension-independent number of oracle calls. We also show how to construct such an oracle in the special case where \mathcal{M} is the boundary of a convex body (Section 8). In the same special case, we explain how to use pre- and post-processing steps to replace a general manifold \mathcal{M} with a manifold \mathcal{M}' with sectional curvature bounded by $\mathcal{O}(d^2)$ (Section 9), allowing us to sample efficiently even from “pointy” manifolds that have very high sectional curvature at some locations.

We mention that [24] obtains a bound on the mixing time of a related (but different) geodesic walk. Although our Markov chains are somewhat similar, our settings are quite different: their bounds require \mathcal{M} to be the boundary of a convex polytope, and are best when the polytope has few faces; our results require \mathcal{M} to have curvature that is bounded away from both 0 and infinity. We also note that the mixing time bound for the geodesic walk in [24] is not dimension-independent. A final difference is that the authors of [24] state that their proof follows the conductance-based approach that is popular for analyzing geometric random walks [38], while our proof uses a probabilistic coupling argument as introduced in [13]. See recent work such as [30, 31] for more on the relationship between couplings of Markov chains and curvature, and also work such as [20, 33] for some consequences of the existence of such couplings, including various concentration inequalities.

1.3. *List of sections.* The rest of the paper is arranged as follows:

- In Section 2, we go over the assumptions we make about \mathcal{M} and Riemannian geometry preliminaries.
- In Section 3, we define the geodesic walk on the manifold \mathcal{M} .
- In Section 4, we prove that the stationary distribution of the geodesic walk is uniform on \mathcal{M} .
- In Section 5, we define a coupling of two copies of the geodesic walk.
- In Section 6, we prove the contraction bound described above.
- In Section 7, we use this contraction bound to prove a bound on the mixing time.
- In Section 8, we show that one can computationally approximate geodesic trajectories to sample the uniform distribution on \mathcal{M} to arbitrary accuracy in a

dimension-independent number of steps if one has access to an appropriate oracle (Section 8.1). We construct such an oracle explicitly in the special case where \mathcal{M} is the boundary of a convex body (Section 8.2).

- In Section 9, we discuss some future research directions left open by this paper: In Section 9.1, we give pre- and post-processing steps that allow us to assume the target manifold \mathcal{M} satisfies $\mathfrak{M}_2 = \mathcal{O}(d^2)$. This is useful when one wishes to sample from a convex body \mathcal{K} whose boundary does not have bounded curvature, such as the case that \mathcal{K} is a convex polytope. In Section 9.3, we discuss connections between this paper and recent work [27] on the mixing times of the popular Hamiltonian Monte Carlo algorithm.

2. Assumptions and Riemannian geometry preliminaries. In this section, we recall results from Riemannian geometry and introduce definitions and assumptions that we will use in the rest of the paper.

- *Assumptions about our manifold \mathcal{M} :* Throughout this paper, we will assume that:

1. (\mathcal{M}, g) is a closed, connected, second-order differentiable Riemannian manifold, with associated inner product $g \equiv g_x$ on the tangent space $\mathcal{T}_x \equiv \mathcal{T}_x \mathcal{M}$ at $x \in \mathcal{M}$.

2. (\mathcal{M}, g) has bounded positive sectional curvature. That is, there exist constants $0 < m_2 \leq \mathfrak{M}_2 < \infty$ so that, for all $x \in \mathcal{M}$ and $u, v \in \mathcal{T}_x$,

$$(2.1) \quad m_2 \leq C_x(u, v) \leq \mathfrak{M}_2,$$

where $C_x(u, v)$ is the sectional curvature of \mathcal{M} at x in the directions $\frac{u}{\|u\|}$ and $\frac{v}{\|v\|}$.

- *Levi–Civita connection:* The fundamental theorem of Riemannian geometry guarantees the existence and uniqueness of a torsion-free affine connection ∇ on (\mathcal{M}, g) that induces an isometry of tangent spaces via parallel transport (see Theorem 6.8 of [17]); this connection is called the *Levi–Civita connection*.
- *Paths and metrics:* Recall that the length of a segment of a smooth path $h : \mathbb{R}^+ \mapsto \mathcal{M}$ is given by

$$(2.2) \quad \text{length}(h[a, b]) = \int_a^b \sqrt{g_{h(t)}(h'(t), h'(t))} dt.$$

This gives rise to the metric

$$(2.3) \quad \text{dist}(x, y) = \inf_{h:h(0)=x, h(1)=y} \text{length}(h[0, 1])$$

on \mathcal{M} .

- *Curvatures:* Recall that the *Riemann curvature tensor* R at a point $x \in \mathcal{M}$ sends a pair of vectors $u, v \in \mathcal{T}_x$ to a map $R(u, v) : \mathcal{T}_x \mapsto \mathcal{T}_x$ of the tangent space at x to itself, which transforms an element $w \in \mathcal{T}_x$ via the formula

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

where ∇ is the Levi-Civita connection described above and $[\cdot, \cdot]$ is the Lie bracket. This formula allows us to define the *sectional curvature* $R' : \mathcal{T}_x^2 \mapsto \mathbb{R}$ by

$$(2.4) \quad R'(u, v) = \frac{\langle R(u, v)v, u \rangle}{\langle u, u \rangle \langle v, v \rangle - \langle u, v \rangle^2},$$

where $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_g$ is the usual inner product on \mathcal{T}_x associated with a Riemannian manifold (\mathcal{M}, g) .

- *Uniform measures:* We recall two measures, one on the manifold \mathcal{M} and the other on small subsets of the tangent spaces \mathcal{T}_x , that we will call the *uniform measures* on their associated spaces. First, recall that any Riemannian manifold (\mathcal{M}, g) has an associated volume function λ . If $\lambda(\mathcal{M}) < \infty$, we denote by $\text{Unif}(\mathcal{M})$ the probability measure given by

$$(2.5) \quad \text{Unif}(\mathcal{M})[A] = \frac{\lambda(A)}{\lambda(\mathcal{M})}$$

for $A \subset \mathcal{M}$ measurable. We refer to this as the *uniform measure on \mathcal{M}* .

Recall that any choice of the basis \mathcal{B} of the tangent space \mathcal{T}_x of a d -dimensional manifold \mathcal{M} at a point $x \in \mathcal{M}$ gives a natural isometry $\zeta_x = \zeta_{x,\mathcal{B}} : \mathcal{T}_x \mapsto \mathbb{R}^d$. Furthermore, the pull-back of the Lebesgue measure from the unit sphere $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\} \subset \mathbb{R}^d$ to the set $\mathbb{S}(\mathcal{T}_x) = \zeta_x^{-1}(\mathbb{S})$ given by ζ_x does not depend on the choice of basis \mathcal{B} . We denote by $\text{Unif}(\mathbb{S}(\mathcal{T}_x))$ or μ this unique pull-back of the Lebesgue measure by ζ_x , and refer to this measure as the *uniform measure on the unit sphere in \mathcal{T}_x* .

Throughout the paper, if $S_1 \subset S_2$ and we have defined a uniform measure on S_2 , we define the uniform measure on S_1 by

$$\text{Unif}(S_1)[A] = \frac{\text{Unif}(S_2)[A]}{\text{Unif}(S_2)[S_1]}.$$

- *Geodesics:* Define the phase space

$$(2.6) \quad \mathcal{M}^\circ := \{(x, v) : x \in \mathcal{M}, v \in \mathbb{S}(\mathcal{T}_x)\}.$$

Associated to every element $(x, v) \in \mathcal{M}^\circ$, there is a special path $\gamma_{(x,v)} : \mathbb{R}^+ \rightarrow \mathcal{M}$ with $\gamma_{(x,v)}(0) = x$ and $\gamma'_{(x,v)}(0) = v$ that is called the *geodesic*. Roughly speaking, this is the path obtained by starting at point x and traveling with velocity v along the manifold. We will use the following properties of the geodesic [34]:

1. For all $x \in \mathcal{M}$ and $v \in \mathcal{T}_x$, there exists $\varepsilon = \varepsilon(x, v) > 0$ so that

$$\text{length}(\gamma_{(x,v)}[0, t]) = \text{dist}(\gamma_{(x,v)}(0), \gamma_{(x,v)}(t))$$

for all $0 < t < \varepsilon$ (see Section 5 and Exercise 5.9.34 of [34]).

2. Associated to every smooth map $\tau : \mathbb{R}^+ \mapsto \mathcal{M}^\circ$, there is an object called the *Jacobi field*

$$(2.7) \quad J_\tau(t) := \frac{\partial \gamma_\tau(t)}{\partial \tau}$$

that satisfies the Jacobi second-order ordinary differential equation:

$$(2.8) \quad \frac{D^2}{dt^2} J_\tau(t) + R(J_\tau(t), \gamma'_\tau(t))\gamma'_\tau(t) = 0,$$

where R denotes the Riemann curvature tensor of \mathcal{M} and D is the covariant derivative with respect to the Levi–Civita connection (see Section 3.2.4 of [34]).

- *Diameter*: We define $D := \text{Diam}(\mathcal{M})$ to be the diameter of \mathcal{M} . Since the sectional curvature of \mathcal{M} is bounded below by m_2 , the Myers–Bonnet theorem [28] implies that

$$(2.9) \quad D \leq \frac{\pi}{\sqrt{m_2}}.$$

- *Existence of a unique parallel transport*: Associated with any connection on a manifold and any finite-length path $h : [a, b] \mapsto \mathcal{M}$ there exists a map from $\mathcal{T}_{h(a)}$ to $\mathcal{T}_{h(b)}$, called the *parallel transport map*.

We now define the only parallel transport maps that we will use in this paper. The Hopf–Rinow theorem [19] implies that between any pair of points $x, y \in \mathcal{M}$ there exists at least one unit-speed geodesic $\omega(t) \equiv \omega(t; x, y)$, $t \in [0, \text{dist}(x, y)]$ with $\omega(0; x, y) = x$ and $\omega(\text{dist}(x, y); x, y) = y$ with $\omega(t; x, y) = \omega(\text{dist}(x, y) - t; y, x)$. We choose arbitrarily, via the axiom of choice, a particular family $\{\omega(\cdot; x, y)\}_{(x,y) \in \mathcal{M} \times \mathcal{M}}$ of such geodesic paths. Finally, we define $\phi(t; \cdot; x, y) : \mathcal{T}_x \mapsto \mathcal{T}_{\omega(t;x,y)}$ to be the parallel transport map associated with the Levi–Civita connection and the path $\omega(\cdot; x, y)$. We also write $\phi(\cdot; x, y) \equiv \phi(\text{dist}(x, y); \cdot; x, y)$ for shorthand.

Finally, fix an arbitrary point $x \in \mathcal{M}$ and an arbitrary basis $\{b_i\}$ of \mathcal{T}_x . Throughout the paper, we denote by $\{\zeta_y\}_{y \in \mathcal{M}}$ the maps $\zeta_{y, \{\phi(b_i; x, y)\}} : \mathcal{T}_y \mapsto \mathbb{R}^d$ associated with this basis and its parallel transports by ϕ .

- *Wasserstein distance and mixing time*: For any distribution η , we write $X \sim \eta$ when the random variable X has distribution η . For two distributions η, ν on a common measure space (Ω, \mathcal{A}) , define $\Xi(\eta, \nu)$ to be the collection of all distributions ξ on $(\Omega^2, \mathcal{A} \times \mathcal{A})$ with marginal distributions $\xi(\cdot, \Omega) = \eta(\cdot)$, $\xi(\Omega, \cdot) = \nu(\cdot)$. For any $k \in \mathbb{N}$, the Wasserstein- k transportation distance $W_d^{(k)}$

between two measures η and ν on a common metric measure space (Ω, d) is given by

$$W_d^{(k)}(\eta, \nu) = \left(\inf_{(X,Y):(X,Y) \in \Xi(\eta,\nu)} \mathbb{E}[(d(X, Y))^k] \right)^{\frac{1}{k}}.$$

Consider a transition kernel K on a metric measure space (Ω, d) , with unique stationary distribution ξ . We define the Wasserstein-k mixing profile $t_{\text{mix}}^{(k)} : [0, 1] \mapsto \mathbb{N}$ of K to be

$$t_{\text{mix}}^{(k)}(\epsilon) = \inf \left\{ t \geq 0 : \sup_{x \in \Omega} W_d^{(k)}(K^t(x, \cdot), \xi(\cdot)) < \epsilon \right\}.$$

- *Big-O notation:* In this paper, we use the “big-O” notation. Specifically, for any two functions $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$, we write “ $f(z) = \mathcal{O}(g(z))$ ” or “ $f(z) \leq \mathcal{O}(g(z))$ ” if there exists a constant $0 < C < \infty$ and some $Z > 0$ such that

$$f(z) \leq C \cdot g(z)$$

for all $z > Z$. Furthermore, write “ $f(z) = \mathcal{O}^*(g(z))$ ” or “ $f(z) \leq \mathcal{O}^*(g(z))$ ” if there exist constants $0 < c', C' < \infty$ and some $Z' > 0$ such that

$$f(z) \leq C' \cdot g(z) \log(z)^{c'}$$

for all $z > Z'$, that is, the “big- \mathcal{O}^* ” notation suppresses the logarithmic terms that we would otherwise need to take into account when using the “big- \mathcal{O} ” notation.

- *Angles:* For two vectors $v_1, v_2 \in \mathbb{R}^d$, we denote by $\angle(v_1, v_2)$ the angle between the vectors. For any hyperplanes $\mathbf{n}_1^\perp, \mathbf{n}_2^\perp$ with normal vectors \mathbf{n}_1 and \mathbf{n}_2 , respectively, we define

$$\angle(\mathbf{n}_1^\perp, \mathbf{n}_2^\perp) := \angle(\mathbf{n}_1, \mathbf{n}_2).$$

For a plane P of any dimension and any vector v , the angle $\angle(v, P)$ between v and P is defined as the angle between v and its projection $\text{proj}_P v$ onto P :

$$\angle(v, P) \equiv \angle(P, v) := \angle(v, \text{proj}_P v).$$

For any hyperplane \mathbf{n}^\perp with normal vector \mathbf{n} , we define the angle between \mathbf{n}^\perp and a plane P of any dimension as

$$\angle(\mathbf{n}^\perp, P) \equiv \angle(P, \mathbf{n}^\perp) := \frac{\pi}{2} - \angle(\mathbf{n}, P).$$

Finally, we note that since $\angle(v, \text{proj}_P v) = \min_{w \in P} \angle(v, w)$, we have that

$$(2.10) \quad \angle(v, \text{proj}_P v) \leq \angle(v, u)$$

for every vector $u \in P$.

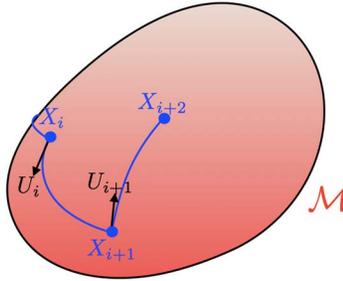


FIG. 1. The geodesic walk X_1, X_2, \dots (blue). Given the current point X_i , the next point is generated by independently sampling a uniform random velocity U_i from the unit sphere on the tangent space at X_i , and running a geodesic trajectory with initial conditions (X_i, U_i) for a fixed time T .

3. The geodesic walk. The geodesic walk $\{X_i\}_{i \in \mathbb{N}}$ on a manifold (Figure 1), with a fixed geodesic step size T is defined precisely in Algorithm 1:

That is, the point X_{i+1} is generated from X_i by running a geodesic trajectory with initial conditions (X_i, U_i) for a fixed time T .

We define the transition kernel \mathbb{K} of the geodesic walk Markov chain $\{X_i\}_{i \in \mathbb{N}}$ by

$$(3.1) \quad \mathbb{K}(Q, R) := \mathbb{P}[X_{i+1} \in R | X_i \sim \text{Unif}(Q)]$$

for all measurable subsets $Q, R \subset \mathcal{M}$.

For the geodesic walk $x = X_1, X_2, \dots$, we define $\Pi_i(\cdot) = \mathbb{K}^i(x, \cdot)$ to be the distribution of X_i , and $\Pi = \lim_{i \rightarrow \infty} \Pi_i$ to be the stationary distribution of the geodesic walk when it exists. In Section 4, we will prove that the uniform measure on \mathcal{M} is a stationary distribution of the geodesic walk (later, we show that this is the only stationary distribution).

We note that, in every step of Algorithm 1, a random variable U_i was constructed. It is straightforward to see that the sequence $\{(X_i, U_i)\}_{i \in \mathbb{N}}$ is a Markov chain on \mathcal{M}° , which we will call the *phase-space* Markov chain.

REMARK 4. Although we write down a fixed integration time T , the algorithm still has the correct stationary distribution if i.i.d. random integration times

Algorithm 1 Geodesic walk

parameters: Integration time T , manifold (\mathcal{M}, g) , number of steps N .

input: $X_1 \in \mathcal{M}$.

output: First N steps $\{X_i\}_{i=1}^N$ of the geodesic walk Markov chain on \mathcal{M} .

- 1: **for** $i = 1, 2, \dots, N - 1$ **do**
 - 2: Sample $U_i \in \mathcal{T}_{X_i}$ from the uniform distribution $\text{Unif}(\mathbb{S}(\mathcal{T}_{X_i}))$ on the unit sphere in \mathcal{T}_{X_i} .
 - 3: Set $X_{i+1} = \gamma_{(X_i, U_i)}(T)$.
 - 4: **end for**
-

T_1, T_2, \dots are sampled at each step (the proof of Theorem 4.1 works essentially as written). In practice, it may be difficult to choose a specific integration time, and we suspect that tricks from the literature on Hamiltonian Monte Carlo can be adapted to this algorithm (see, e.g., [18]).

In referring to the geodesic walk as an “algorithm,” we will assume that we have an oracle for computing geodesic trajectories γ with perfect accuracy; we refer to this oracle as an “idealized geodesic integrator.” We drop this assumption in Section 8, where we discuss a computational implementation of the geodesic walk.

4. The stationary distribution of the geodesic walk. The aim of this section is to prove that the transition kernel \mathbb{K} defined in equation (3.1) has uniform stationary distribution. To do so, we will use the fact that the Liouville measure is invariant under geodesic flow [5, 21, 40].

Recall the definition of the phase space \mathcal{M}° in equation (2.6). The Liouville measure \mathcal{L} is defined to be the measure on \mathcal{M}° with density given by the product of the volume form $d\lambda(x)$ on the manifold and the volume form $d\mu(v)$ on the unit sphere in the tangent space of the manifold at the point x (i.e., the unit-speed velocities):

$$(4.1) \quad \mathcal{L}(\Omega) := \int_{\Omega} d\mu(v) d\lambda(x).$$

Consider a geodesic trajectory on \mathcal{M} with initial position x and initial velocity v . We define $\gamma_{(x,v)}(t)$ and $\varphi_{(x,v)}(t) := \gamma'_{(x,v)}(t) \in \mathcal{T}_{\gamma_{(x,v)}(t)}$ to be, respectively, the position and velocity of this trajectory at time $t \in \mathbb{R}^+$. Finally, we define

$$\psi_{(x,v)}(t) := (\gamma_{(x,v)}(t), \varphi_{(x,v)}(t))$$

to be the location of our geodesic trajectory in the phase space at time $t \in \mathbb{R}^+$.

For notational convenience, we define the map

$$(4.2) \quad \Psi_t(A) := \bigcup_{(x,v) \in A} \psi_{(x,v)}(t)$$

for all $A \subset \mathcal{M}^\circ$ and $t > 0$.

We now prove that the volume measure on any closed manifold \mathcal{M} is a stationary measure of the geodesic walk (Theorem 4.1); this holds even if \mathcal{M} does not have positive curvature. In Theorem 6.2 of Section 6, we prove that the stationary distribution is unique under the additional assumption that \mathcal{M} has bounded positive curvature.

THEOREM 4.1. *The uniform distribution on \mathcal{M} is a stationary distribution of the geodesic walk Markov chain defined in Algorithm 1.*

PROOF. We begin by defining two transition kernels \mathbb{K}_1° and \mathbb{K}_2° on \mathcal{M}° . We define the Markov chain $\{(X_i^{(1)}, v_i^{(1)})\}_{i \in \mathbb{N}}$ on \mathcal{M}° by the deterministic update rule

$$(X_{i+1}^{(1)}, v_{i+1}^{(1)}) = \psi_{(X_i^{(1)}, v_i^{(1)})}(T)$$

and define \mathbb{K}_1° to be the associated transition kernel. We then define the Markov chain $\{(X_i^{(2)}, v_i^{(2)})\}_{i \in \mathbb{N}}$ on \mathcal{M}° by the update rule

$$\begin{aligned} X_{i+1}^{(2)} &= X_i^{(2)}, \\ v_{i+1}^{(2)} &\sim \text{Unif}(\mathbb{S}(\mathcal{T}_{X_{i+1}^{(2)}})) \end{aligned}$$

and define \mathbb{K}_2° to be the associated transition kernel. We observe that \mathbb{K}_1° is deterministic, that \mathbb{K}_2° only ever updates its second coordinate, and finally that

$$(4.3) \quad \mathbb{K}^\circ = \mathbb{K}_2^\circ \mathbb{K}_1^\circ.$$

Since the Liouville measure is invariant under geodesic flow [5, 21, 40], we have

$$(4.4) \quad \mathcal{L}(A) = \mathcal{L}(\Psi_T(A))$$

for every measurable $A \subset \mathcal{M}^\circ$. This implies that the Liouville measure \mathcal{L} is invariant under \mathbb{K}_1° :

$$(4.5) \quad \mathcal{L} \mathbb{K}_1^\circ = \mathcal{L}.$$

It is clear by inspection that

$$(4.6) \quad \mathcal{L} \mathbb{K}_2^\circ = \mathcal{L}.$$

Thus, by equations (4.3), (4.5) and (4.6),

$$(4.7) \quad \mathcal{L} \mathbb{K}^\circ = \mathcal{L}.$$

Let $\{(X_i, U_i)\}_{i \geq 0}$ be a Markov chain started at $(X_1, U_1) \sim \mathcal{L}$ and evolving according to \mathbb{K}° . By the definition of \mathbb{K}° , the marginal process $\{X_i\}_{i \in \mathbb{N}}$ is a Markov chain evolving according to the transition kernel \mathbb{K} . Thus, the marginal distribution of \mathcal{L} on its first coordinate \mathcal{M} must be a stationary distribution for \mathbb{K} . But this marginal distribution is exactly $\text{Unif}(\mathcal{M})$, completing the proof. \square

5. Coupling the geodesic walk. We define a coupling of two copies of the geodesic walk.

DEFINITION 1 (Coupling of geodesic walk). Fix $x, y \in \mathcal{M}$ and $0 < T < \infty$. We define a pair of stochastic processes $\{(X_i, Y_i)\}_{i \in \mathbb{N}}$ with $X_1 = x, Y_1 = y$ as follows:

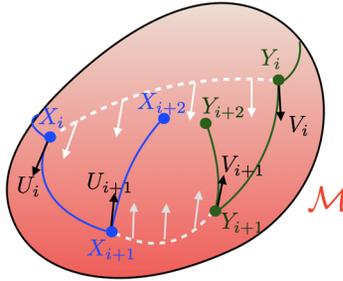


FIG. 2. Coupling of the two geodesic walks X_1, X_2, \dots (blue) and Y_1, Y_2, \dots (green). The coupling is achieved by setting the initial trajectory velocity V_i at Y_i (black) to be the parallel transport of the initial trajectory velocity U_i at X_i along a minimum geodesic from X_i to Y_i (gray/white dashed line). Since \mathcal{M} has positive curvature, the distance between the Markov chains contracts rapidly at each step.

Let $X_1 = x$ and $Y_1 = y$. For $i \in \mathbb{N}$, inductively sample

$$U_i \sim \text{Unif}(\mathbb{S}(\mathcal{T}_{X_i}))$$

and set

$$(5.1) \quad V_i = \phi(U_i; X_i, Y_i),$$

$$(5.2) \quad X_{i+1} = \gamma_{(X_i, U_i)}(T),$$

$$(5.3) \quad Y_{i+1} = \gamma_{(Y_i, V_i)}(T).$$

This coupling is illustrated in Figure 2.

This stochastic process is a valid coupling of two copies of the geodesic walk.

THEOREM 5.1. *Let $\{(X_i, Y_i)\}_{i \in \mathbb{N}}$ be as in Definition 1. Then the marginal processes $\{X_i\}_{i \in \mathbb{N}}$ and $\{Y_i\}_{i \in \mathbb{N}}$ are each Markov chains with transition kernel \mathbb{K} .*

PROOF. It is straightforward to check that $\{X_i\}_{i \in \mathbb{N}}$ is exactly the Markov chain defined in Algorithm 1 (even the notation is the same). Thus, it remains only to check that $\{Y_i\}_{i \in \mathbb{N}}$ has the correct distribution.

To see that $\{Y_i\}_{i \in \mathbb{N}}$ has the correct distribution, looking at Algorithm 1 it is enough to check that (conditional on $\{Y_j\}_{j \leq i}$), V_i has uniform distribution on $\mathbb{S}(\mathcal{T}_{Y_i})$.

By the fundamental theorem of Riemannian geometry, parallel transport using the Levi–Civita connection is an affine transformation. Hence, it preserves the angles between any two vectors u and u' parallel transported to $v = \phi(u; X_i, Y_i)$ and $v' = \phi(u'; X_i, Y_i)$, respectively:

$$(5.4) \quad \angle(u, u') = \angle(v, v').$$

Therefore, denoting the conditional density of U_i conditioned on X_i by $f_{U_i}(u|X_i = x)$, and the conditional density of V_i conditioned on Y_i by $f_{V_i}(v|Y_i = y)$, we have

$$(5.5) \quad f_{V_i}(v|Y_i = y) = f_{U_i}(u|X_i = x) = \frac{1}{\text{Vol}(\mathbb{S}^{d-1})}$$

whenever u is a vector on the unit sphere in the tangent plane of x , and $v = \phi(u; x, y)$ is its parallel transport to Y_i . Since the Levi–Civita connection also preserves magnitude, the fact that u is on the unit sphere in the tangent plane of x implies that v is on the unit sphere in the tangent plane of y .

Since U_i is uniformly distributed on the unit sphere in \mathcal{T}_{X_i} , this implies that V_i is uniformly distributed on the unit sphere in \mathcal{T}_{Y_i} . Therefore, the transition kernel of the Markov chain $\{Y_i\}_{i \in \mathbb{N}}$ must be \mathbb{K} as well. \square

6. Contraction of coupled geodesics. In this section, we prove a contraction bound on the coupled geodesic walks X_1, X_2, \dots and Y_1, Y_2, \dots using the following extension of the Rauch comparison theorem from differential geometry.

LEMMA 1. *Let $x, y \in \mathcal{M}$, let $v_x \in \mathcal{T}_x$ and let $v_y = \phi(v_x; x, y)$. Define $v_x^\perp := \text{proj}_{(\omega'(0;x,y))^\perp} v_x$. Then for all $0 \leq T \leq \frac{\pi}{2\sqrt{\mathfrak{M}_2}}$, we have*

$$(6.1) \quad \text{dist}(\gamma_{(x,v_x)}(T), \gamma_{(y,v_y)}(T)) \leq \text{dist}(x, y) \times \sqrt{\|v_x^\perp\|^2 \cos^2(\sqrt{\mathfrak{m}_2}T) + (1 - \|v_x^\perp\|^2)}.$$

REMARK 5. A lower bound for the geodesic distance with $\|v_x^\perp\| = 1$ is proved in [17]. The authors mention that it is possible to prove an upper bound using a similar construction, although they do not do so explicitly. For this reason, we prove Lemma 1 explicitly here, following a similar presentation in Proposition 7.8 of [37].

PROOF. Note that $v_y = \phi(v_x; x, y)$ is continuous in v_x , and that $\gamma_{(x,v_x)}(T)$ and $\gamma_{(y,v_y)}(T)$ are continuous in v_x and v_y , respectively (see, for instance, Proposition 5.9 in [23]). Therefore, if we can show that Inequality (6.1) holds whenever $\|v_x^\perp\| > 0$, it will also hold whenever $\|v_x^\perp\| = 0$ by continuity in v_x . So without loss of generality, we may assume that $\|v_x^\perp\| > 0$.

Consider the family of geodesics $\mathcal{F} = \{\gamma_{(\omega(\tau;x,y),\phi(\tau;v_x;x,y))}(t) : \tau \in [0, \text{dist}(x, y)]\}$. Let $J_\tau(t)$ be the Jacobi field associated with this family. Since J_τ is a Jacobi field along the geodesic $\gamma_{(\omega(\tau;x,y),\phi(\tau;v_x;x,y))}(t)$, it is possible to decompose $J_\tau(t)$ into two fields $J_\tau(t) = J_\tau^\perp(t) + J_\tau^\parallel(t)$ so that $J_\tau^\perp(t)$ is perpendicular to $\gamma'_{(\omega(\tau;x,y),\phi(\tau;v_x;x,y))}(t)$ and $J_\tau^\parallel(t)$ is parallel to $\gamma'_{(\omega(\tau;x,y),\phi(\tau;v_x;x,y))}(t)$ for all $t \geq 0$ (see, for instance, Corollary B.14 in [8] and the proof of Proposition 7.8 in [37]).

Since \mathcal{M} has curvature bounded above by $\mathfrak{M}_2 > 0$, the Rauch comparison theorem (comparing \mathcal{M} to a sphere of radius $\frac{1}{\sqrt{\mathfrak{M}_2}}$) [16, 35] gives

$$\|J_\tau(t)^\perp\| \geq \cos(\sqrt{\mathfrak{M}_2}t) \times \|\text{proj}_{(\omega'(\tau;x,y))^\perp} \phi(\tau; v_x; x, y)\|$$

for all

$$t \leq \inf\{s > 0 : \cos(\sqrt{\mathfrak{M}_2}s) = 0\} = \frac{\pi}{2\sqrt{\mathfrak{M}_2}}$$

But $\|\text{proj}_{(\omega'(\tau;x,y))^\perp} \phi(\tau; v_x; x, y)\| = \|\text{proj}_{(\omega'(0;x,y))^\perp} v_x\| = \|v_x^\perp\|$, so

$$\|J_\tau(t)^\perp\| \geq \|v_x^\perp\| \cos(\sqrt{\mathfrak{M}_2}t) \quad \forall t \leq \frac{\pi}{2\sqrt{\mathfrak{M}_2}}.$$

Thus,

$$(6.2) \quad \|J_\tau^\perp(t)\| > 0$$

for all $0 \leq t < T$. Therefore, since $\|J_\tau^\perp(t)\|$ has no zeros on $[0, T)$, applying the Rauch comparison theorem a second time gives (this time comparing \mathcal{M} to a sphere of radius $\frac{1}{\sqrt{\mathfrak{m}_2}}$)

$$(6.3) \quad \|J_\tau^\perp(t)\| \leq \|v_x^\perp\| \cos(\sqrt{\mathfrak{m}_2}t)$$

for all $0 \leq t \leq T$.

Define

$$(6.4) \quad F(\tau) := \text{length}(\{\gamma_{(\omega(s;x,y), \phi(s;v_x;x,y))}(T) : s \in [0, \tau]\})$$

for $0 < \tau < \text{dist}(x, y)$. Then

$$(6.5) \quad \frac{dF}{d\tau} = \|J_\tau(T)\|.$$

Therefore, by the fundamental theorem of calculus,

$$\begin{aligned} & \text{dist}(\gamma_{(x,v_x)}(T), \gamma_{(y,v_y)}(T)) \\ &= F(\text{dist}(x, y)) \\ &= \int_0^{\text{dist}(x,y)} \frac{dF}{d\tau} d\tau \\ &\stackrel{\text{Eq. (6.5)}}{=} \int_0^{\text{dist}(x,y)} \|J_\tau(T)\| d\tau \\ &= \int_0^{\text{dist}(x,y)} \sqrt{\|J_\tau(T)^\perp\|^2 + \|J_\tau(T)\|^2} d\tau \\ &\stackrel{\text{Eq. (6.3)}}{\leq} \int_0^{\text{dist}(x,y)} \sqrt{\|v_x^\perp\|^2 \cos^2(\sqrt{\mathfrak{m}_2}T) + (1 - \|v_x^\perp\|^2)} d\tau \\ &= \text{dist}(x, y) \times \sqrt{\|v_x^\perp\|^2 \cos^2(\sqrt{\mathfrak{m}_2}T) + (1 - \|v_x^\perp\|^2)}. \end{aligned} \quad \square$$

THEOREM 6.1. Fix $0 \leq T \leq \frac{\pi}{2\sqrt{m_2}}$ and let $\{(X_i, Y_i)\}_{i \geq 0}$ be coupled as in Definition 1. Then

$$\begin{aligned} \mathbb{E}[\text{dist}(X_{i+1}, Y_{i+1})|X_i, Y_i] &\leq \sqrt{\mathbb{E}[\text{dist}(X_{i+1}, Y_{i+1})^2|X_i, Y_i]} \\ &\leq \left(1 - \frac{m_2}{8}T^2\right) \times \text{dist}(X_i, Y_i). \end{aligned}$$

PROOF. Define $U_i^\perp := \text{proj}_{\omega^\perp(0; X_i, Y_i)} U_i$. Since U_i is sampled from the uniform distribution $\text{Unif}(\mathbb{S}(\mathcal{T}_{X_i}))$ on the sphere of dimension at least 2, for any $0 \leq \alpha \leq 1$ we have

$$(6.6) \quad \mathbb{E}[\alpha^2 \|U_i^\perp\|^2 + (1 - \|U_i^\perp\|^2)] = \alpha^2 \mathbb{E}[\|U_i^\perp\|^2] + (1 - \mathbb{E}[\|U_i^\perp\|^2])$$

$$(6.7) \quad \leq \frac{1}{2}\alpha^2 + \frac{1}{2}$$

By Lemma 1,

$$(6.8) \quad \mathbb{E}[\text{dist}(X_{i+1}, Y_{i+1})^2|X_i, Y_i]$$

$$(6.9) \quad \stackrel{\text{Lemma 1}}{\leq} \text{dist}(X_i, Y_i)^2 \times \mathbb{E}[\|U_i^\perp\|^2 \cos^2(\sqrt{m_2}T) + (1 - \|U_i^\perp\|^2)]$$

$$(6.10) \quad \stackrel{\text{Eq. 6.6}}{\leq} \text{dist}(X_i, Y_i)^2 \times \frac{1}{2}(\cos^2(\sqrt{m_2}T) + 1).$$

By Jensen’s inequality, we have

$$\begin{aligned} \mathbb{E}[\text{dist}(X_{i+1}, Y_{i+1})|X_i, Y_i] &\stackrel{\text{Jensen}}{\leq} \sqrt{\mathbb{E}[\text{dist}(X_{i+1}, Y_{i+1})^2|X_i, Y_i]} \\ &\stackrel{\text{Eq. (6.8)}}{\leq} \text{dist}(X_i, Y_i) \frac{1}{\sqrt{2}} \sqrt{\cos^2(\sqrt{m_2}T) + 1} \\ &= \text{dist}(X_i, Y_i) \sqrt{1 - \frac{1}{2} \sin^2(\sqrt{m_2}T)} \\ &\leq \text{dist}(X_i, Y_i) \sqrt{1 - \frac{m_2}{4}T^2} \\ &\leq \left(1 - \frac{m_2}{8}T^2\right) \times \text{dist}(X_i, Y_i) \end{aligned}$$

where the third inequality holds since $0 \leq T \leq \frac{\pi}{2\sqrt{m_2}}$. \square

THEOREM 6.2. The uniform distribution on \mathcal{M} is the unique stationary distribution of the geodesic walk Markov chain defined in Algorithm 1 for any parameter $0 < T \leq \frac{\pi}{2\sqrt{m_2}}$.

PROOF. By Theorem 4.1, the uniform distribution Π on \mathcal{M} is a stationary distribution of \mathbb{K} . Assume that there is a second stationary distribution $\Pi' \neq \Pi$. Since \mathcal{M} has finite diameter, this implies $0 < W_d^{(1)}(\Pi, \Pi') < \infty$. By Theorem 6.1, however,

$$W_d^{(1)}(\Pi, \Pi') = W_d^{(1)}(\Pi\mathbb{K}, \Pi'\mathbb{K}) \leq \left(1 - \frac{m_2}{8}T^2\right) \times W_d^{(1)}(\Pi, \Pi') < W_d^{(1)}(\Pi, \Pi'),$$

since $0 < W_d^{(1)}(\Pi, \Pi') < \infty$ and $0 < T \leq \frac{\pi}{2\sqrt{m_2}}$. This is a contradiction, so no such distribution Π' exists. \square

7. Bounding the mixing time. In this section, we bound the mixing time (in Wasserstein transportation distance) using the approach of [31].

THEOREM 7.1. *For $k \in \{1, 2\}$, the Wasserstein- k mixing profile $t_{\text{mix}}^{(k)}$ of the Markov chain with transition kernel \mathbb{K} defined in Equation (3.1) of Section 3 and parameter $0 < T \leq \frac{\pi}{2\sqrt{m_2}}$ satisfies*

$$(7.1) \quad t_{\text{mix}}^{(k)}(\epsilon) \leq \left\lceil \frac{\log(\epsilon D^{-1})}{\log(1 - \frac{m_2}{8}T^2)} \right\rceil = \mathcal{O}\left(\frac{1}{m_2 T^2} \cdot \log(D\epsilon^{-1})\right).$$

In particular, for $T = \frac{\pi}{2\sqrt{m_2}}$, the mixing time is bounded by

$$(7.2) \quad t_{\text{mix}}^{(k)}(\epsilon) \leq \left\lceil \frac{\log(\epsilon D^{-1})}{\log(1 - \frac{\pi^2 m_2}{32 m_2^2})} \right\rceil = \mathcal{O}\left(\frac{m_2}{m_2} \cdot \log(D\epsilon^{-1})\right).$$

PROOF. Fix $0 \leq T \leq \frac{\pi}{2\sqrt{m_2}}$. Let $\{(X_i, Y_i)\}_{i \in \mathbb{N}}$ be two copies of the Markov chain with kernel \mathbb{K} , coupled as in Definition 1. Recall that, by Theorem 5.1, this is a valid coupling. By Theorem 6.1, this chain satisfies the following contraction inequality for $k \in \{1, 2\}$:

$$(7.3) \quad (\mathbb{E}[\text{dist}(X_{i+1}, Y_{i+1})^k | X_i, Y_i])^{\frac{1}{k}} \leq \text{dist}(X_i, Y_i) \times \left(1 - \frac{m_2}{8}T^2\right) \quad \forall i \geq 0.$$

Applying this contraction inequality repeatedly, we find

$$(7.4) \quad \begin{aligned} (\mathbb{E}[\text{dist}(X_j, Y_j)^k])^{\frac{1}{k}} &\leq \text{dist}(X_0, Y_0) \times \left(1 - \frac{m_2}{8}T^2\right)^j \\ &\leq D \times \left(1 - \frac{m_2}{8}T^2\right)^j \geq 0. \end{aligned}$$

This bound immediately implies that, for all $0 < \epsilon < 1$,

$$(7.5) \quad t_{\text{mix}}^{(k)}(\epsilon) \leq \left\lceil \frac{\log(\epsilon D^{-1})}{\log(1 - \frac{m_2}{8}T^2)} \right\rceil.$$

Algorithm 2 Numerical approximation of geodesic trajectory

parameters: $\theta > 0, T > 0$.

input: Manifold (\mathcal{M}, g) with sectional curvature bounded above by \mathfrak{M}_2 .

input: Initial point $x_0 \in \mathcal{M}$, initial velocity $v_0 \in \mathcal{T}_{x_0}\mathcal{M}$.

input: Oracle \star returning $\psi_{\theta; (x,v)}^\star$ and $\Delta_{\theta; (x,v)}^\star$ for any $(x, v) \in \mathcal{M}^\circ$.

output: $\gamma_{\theta; (x_0, v_0)}^\dagger(T)$.

1: Set $i = 0, \Delta_0 = 0$.

2: Set $x_0^\dagger = x_0$ and $v_0^\dagger = v_0$.

3: **while** $\sum_{j=0}^i \Delta_j \leq T$ **do**

4: Call the oracle \star to compute $(x_{i+1}^\dagger, v_{i+1}^\dagger) = \psi_{\theta; (x_i^\dagger, v_i^\dagger)}^\star$ and $\Delta_{i+1} = \Delta_{\theta; (x_i^\dagger, v_i^\dagger)}^\star$.

5: Set $i = i + 1$.

6: **end while**

7: Set $\delta = T - \sum_{j=1}^i \Delta_j$.

8: Using the bisection method, iteratively call \star to find a value $\hat{\theta}$ for which

$$\delta - \frac{\theta^k}{\sqrt{\mathfrak{M}_2}} \leq \Delta_{\hat{\theta}; (x_i^\dagger, v_i^\dagger)}^\star \leq \delta. \text{ Set } \Delta_{i+1} := \Delta_{\hat{\theta}; (x_i^\dagger, v_i^\dagger)}^\star.$$

9: Call the oracle \star to compute $x_{i+1}^\dagger = \gamma_{\hat{\theta}; (x_i^\dagger, v_i^\dagger)}^\star$.

10: Set $\gamma_{\theta; (x_0, v_0)}^\dagger(T) = x_{i+1}^\dagger$.

In particular, for $T = \frac{\pi}{2\sqrt{\mathfrak{M}_2}}$, we get:

$$(7.6) \quad t_{\text{mix}}^{(k)}(\epsilon) \leq \left\lceil \frac{\log(\epsilon D^{-1})}{\log(1 - \frac{\pi^2 \mathfrak{m}_2}{32 \mathfrak{M}_2})} \right\rceil. \quad \square$$

8. Approximating geodesics. In this section, we describe and analyze an approximation to the geodesic walk, showing that it can be used to approximately sample from the uniform distribution on \mathcal{M} to arbitrary accuracy using a dimension-independent number of computations.

We begin in Section 8.1 by assuming that we have access to an oracle \star , described at the beginning of Section 8.1, that can approximate short geodesic paths with small error. We use this oracle to construct an ‘‘approximate’’ version of the geodesic walk, given in Algorithm 2. Finally, we prove that, for any fixed error rate $\epsilon > 0$, this approximate geodesic walk can give samples that are within ϵ of the uniform distribution on \mathcal{M} in Wasserstein distance using a dimension-independent number of oracle calls. In Section 8.2, we construct \star explicitly for the special case where \mathcal{M} is the boundary of a convex body, using only the basic convex body oracles used in [12] (Figure 3).

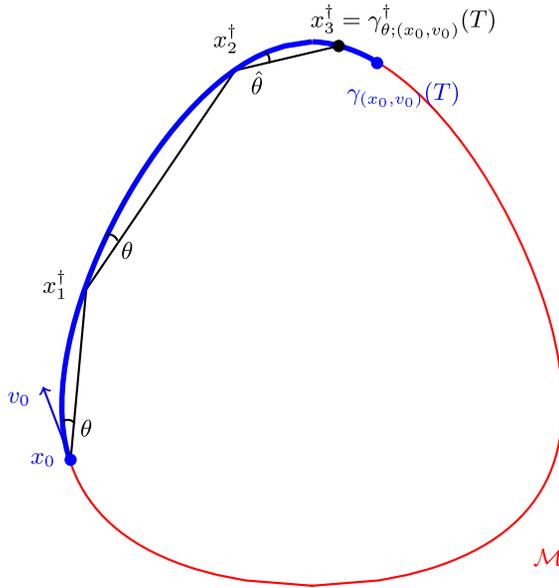


FIG. 3. Algorithm for approximating geodesics. In this example, \mathcal{M} is the boundary of a convex body, so one can generate each step in the approximation as a line ℓ_i leaving the current approximation point x_i^\dagger on the boundary at a fixed angle θ . The next approximation point x_{i+1}^\dagger is the next point when ℓ_i intersects the boundary. Lemmas 2 and 6 show that the final point $x_{i_{\max}^\dagger+1}^\dagger = \gamma_{\theta; (x_0, v_0)}^\dagger(T)$ (here $i_{\max} = 2$) approximates the final point $\gamma_{(x_0, v_0)}(T)$ of the true geodesic trajectory to arbitrary accuracy ε in $i_{\max}(\varepsilon)$ steps, where $i_{\max}(\varepsilon)$ is independent of the dimension, if θ is appropriately chosen.

8.1. Approximating geodesics on general positive-curvature manifolds. In Algorithm 2, we compute recursively an approximation $\gamma_{\theta; (x, v)}^\dagger(T)$ of $\gamma_{(x, v)}(T)$ by repeatedly calling an oracle $\star : \mathcal{M}^\circ \times \mathbb{R}^+ \rightarrow \mathcal{M}^\circ \times \mathbb{R}^+$. We use the following notation for the oracle \star .

DEFINITION 2. An oracle $\star : \mathcal{M}^\circ \times \mathbb{R}^+ \rightarrow \mathcal{M}^\circ \times \mathbb{R}^+$ is a function that has inputs $(x, v) \in \mathcal{M}^\circ$, and $\theta > 0$, and outputs $\psi_{\theta; (x, v)}^\star = (\gamma_{\theta; (x, v)}^\star, \varphi_{\theta; (x, v)}^\star) \in \mathcal{M}^\circ$ and $\Delta_{\theta; (x, v)}^\star \in \mathbb{R}^+$.

The oracle tries to approximate the geodesic trajectory $\psi_{(x, v)}(t)$ at $t = \Delta_{\theta; (x, v)}^\star$, where the step size $\Delta_{\theta; (x, v)}^\star$ is determined by the oracle. The parameter $\theta > 0$ allows the user to adjust the accuracy of the oracle’s approximation.

For every oracle \star and every $\theta > 0$, we define

$$(8.1) \quad \Delta_{\theta; \min}^\star := \inf_{(x, v) \in \mathcal{M}^\circ} \Delta_{\theta; (x, v)}^\star.$$

We make the following assumptions about the oracle \star .

ASSUMPTION 1. There exist multivariate polynomials P and Q with the property that, at every $(x, v) \in \mathcal{M}^\circ$ and every $0 \leq \theta < \frac{1}{70} \frac{m_2}{\mathfrak{M}_2}$, we have

$$(8.2) \quad \frac{1}{\sqrt{\mathfrak{M}_2}} \cdot P\left(\theta, \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{m_2}}, \frac{\sqrt{m_2}}{\sqrt{\mathfrak{M}_2}}\right) \leq \Delta_{\theta;(x,v)}^\star \leq \frac{1}{\sqrt{\mathfrak{M}_2}} \cdot Q\left(\theta, \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{m_2}}, \frac{\sqrt{m_2}}{\sqrt{\mathfrak{M}_2}}\right).$$

Also, for any fixed $0 < a, b < \infty$, $P(\theta, a, b)$ and $Q(\theta, a, b)$ are strictly increasing in θ over the interval $0 < \theta < \frac{1}{70} \frac{m_2}{\mathfrak{M}_2}$, and $P(0, a, b) = Q(0, a, b) = 0$.

We also assume the following.

ASSUMPTION 2. There exist constants $\alpha > 0$, $\beta > 0$, and $k \in \mathbb{N}$ such that

$$(8.3) \quad \text{dist}(\gamma_{(x,v)}(\Delta_{\theta;(x,v)}^\star), \gamma_{\theta;(x,v)}^\star) \leq \alpha \cdot \theta^k \cdot \Delta_{\theta;(x,v)}^\star$$

and also

$$(8.4) \quad \|\varphi_{\theta;(x,v)}^\star - \overline{\varphi_{(x,v)}(\Delta_{\theta;(x,v)}^\star)}\| \leq \beta \cdot \sqrt{\mathfrak{M}_2} \cdot \theta^k \cdot \Delta_{\theta;(x,v)}^\star$$

for every $(x, v) \in \mathcal{M}^\circ$ and $0 \leq \theta < \frac{1}{70} \frac{m_2}{\mathfrak{M}_2}$, where $\overline{\varphi_{(x,v)}(\Delta_{\theta;(x,v)}^\star)} := \phi(\varphi_{(x,v)}(\Delta_{\theta;(x,v)}^\star); \gamma_{(x,v)}(\Delta_{\theta;(x,v)}^\star), \gamma_{\theta;(x,v)}^\star)$.

For the rest of Section 8.1, we will assume that the oracle \star satisfies Assumptions 1 and 2.

Using the notation of Algorithm 1, the geodesic walk Markov chain is defined recursively by

$$(8.5) \quad X_{i+1} := \gamma_{(X_i, U_i)}(T) \quad \forall i = 0, 1, 2, \dots,$$

where U_i is uniformly distributed on $\mathbb{S}(\mathcal{T}_{X_i})$. For fixed $\theta > 0$, we can use Algorithm 2 to recursively generate an approximate geodesic walk $X_1^\theta, X_2^\theta, \dots$ coupled to the geodesic walk X_1, X_2, \dots , using the following recursion:

$$(8.6) \quad \begin{aligned} X_0^\theta &= X_0, \\ X_{i+1}^\theta &= \gamma_{\theta;(X_i^\theta, U_i')}^\dagger(T) \quad \forall i = 0, 1, 2, \dots, \end{aligned}$$

where $U_i' = \phi(U_i; X_i, X_i^\theta)$. For $i \in \mathbb{N}$, we let Π_i^θ be the distribution of X_i^θ .

The following lemma shows that our approximate geodesic chain remains close to the actual geodesic chain.

LEMMA 2. For every $(x_0, v_0) \in \mathcal{M}^\circ$, every $0 \leq \theta < \frac{1}{70} \frac{m_2}{\mathfrak{M}_2}$, and every $0 \leq T \leq \frac{\pi}{2\sqrt{\mathfrak{M}_2}}$, Algorithm 2 makes at most $\lceil \frac{T}{\Delta_{\theta;\min}^\star} \rceil + \mathcal{O}(\log(\frac{\sqrt{\mathfrak{M}_2}}{\sqrt{m_2}} \times \frac{\sqrt{\mathfrak{M}_2}}{\theta^k}))$ calls of the oracle \star and returns a point $\gamma_{\theta;(x_0, v_0)}^\dagger(T)$ such that

$$(8.7) \quad \text{dist}(\gamma_{(x_0, v_0)}(T), \gamma_{\theta;(x_0, v_0)}^\dagger(T)) \leq \left[1 + \frac{\pi}{2}\alpha + \left(\frac{\pi}{2}\right)^2 \beta\right] \cdot \frac{1}{\sqrt{\mathfrak{M}_2}} \cdot \theta^k.$$

PROOF. *Counting the number of \star oracle calls:* Define $i_{\max} := \max\{i \in \mathbb{N} : \sum_{j=1}^{i+1} \Delta_j \leq T\}$. The number of \star oracle calls required for Step 4 in Algorithm 2 is at most $\lceil \frac{T}{\Delta_{\theta; \min}^{\star}} \rceil$, since the largest possible value of i such that $\sum_{j=1}^{i+1} \Delta_j \leq T$ is $\lfloor \frac{T}{\Delta_{\theta; \min}^{\star}} \rfloor$. Assumption 1 implies that the number of \star oracle calls the bisection method makes in Step 8 of Algorithm 2 to find a value $\hat{\theta}$ for which $\Delta_{i+1} := \Delta_{\hat{\theta}; (x_i^{\dagger}, v_i^{\dagger})}^{\star}$ satisfies $\delta - \frac{\theta^k}{\sqrt{\mathfrak{M}_2}} \leq \Delta_{i+1} \leq \delta$ is at most $\mathcal{O}(\log(\frac{\sqrt{\mathfrak{M}_2}}{\sqrt{m_2}} \times \frac{\sqrt{\mathfrak{M}_2}}{\theta^k}))$. Step 9 makes exactly one call of the oracle \star .

Bounding the error: Fix $i \in \mathbb{N}$ and let $x_i^{\dagger}, v_i^{\dagger}, \Delta_i, \theta$, and $\hat{\theta}$ be as in Algorithm 2. Define $(x_i, v_i) := \psi_{(x_{i-1}^{\dagger}, v_{i-1}^{\dagger})}(\Delta_i)$, and $\bar{v}_i := \phi(v_i; x_i, x_i^{\dagger})$. By inequality (8.4), we have $\|v_i^{\dagger} - \bar{v}_i\| \leq \beta \cdot \sqrt{\mathfrak{M}_2} \cdot \theta^k \cdot \Delta_i$. Therefore, comparing to a triangle in a sphere of radius $\frac{1}{\sqrt{\mathfrak{M}_2}}$ by the Toponogov triangle comparison theorem (see Chapter 11 of [34]), we have

$$(8.8) \quad \text{dist}(\gamma_{(x_i^{\dagger}, \bar{v}_i)}(t), \gamma_{(x_i^{\dagger}, v_i^{\dagger})}(t)) \leq t \cdot \|v_i^{\dagger} - \bar{v}_i\| \leq t \cdot \beta \cdot \sqrt{\mathfrak{M}_2} \cdot \theta^k \cdot \Delta_i$$

for all $0 \leq t \leq \frac{\pi}{2\sqrt{\mathfrak{M}_2}}$.

By inequality (8.3), we have $\text{dist}(x_i, x_i^{\dagger}) \leq \alpha \cdot \theta^k \cdot \Delta_i$. Therefore, by Lemma 1,

$$(8.9) \quad \text{dist}(\gamma_{(x_i, v_i)}(t), \gamma_{(x_i^{\dagger}, \bar{v}_i)}(t)) \leq \text{dist}(x_i, x_i^{\dagger}) \leq \alpha \cdot \theta^k \cdot \Delta_i$$

for all $0 \leq t \leq \frac{\pi}{2\sqrt{\mathfrak{M}_2}}$.

Define the “remaining time after the i ’th step” to be $T_i := \sum_{j=i+1}^{i_{\max}+1} \Delta_j$. Then

$$\begin{aligned} & \text{dist}(\gamma_{(x_0, v_0)}(T), \gamma_{\hat{\theta}; (x_0, v_0)}(T)) \\ & \leq \text{dist}(\gamma_{(x_0, v_0)}(T), \gamma_{(x_0, v_0)}(T_0)) \\ & \quad + \text{dist}(\gamma_{(x_0, v_0)}(T_0), x_{i_{\max}}^{\dagger}) \\ & \leq T - T_0 + \text{dist}(\gamma_{(x_0, v_0)}(T_0), x_{i_{\max}}^{\dagger}) \\ & \leq T - T_0 + \sum_{i=1}^{i_{\max}+1} (\text{dist}(\gamma_{(x_i, v_i)}(T_i), \gamma_{(x_i^{\dagger}, \bar{v}_i)}(T_i)) \\ & \quad + \text{dist}(\gamma_{(x_i^{\dagger}, \bar{v}_i)}(T_i), \gamma_{(x_i^{\dagger}, v_i^{\dagger})}(T_i))) \\ & \leq \frac{\theta^k}{\sqrt{\mathfrak{M}_2}} + \sum_{i=1}^{i_{\max}+1} (\alpha \theta^k \cdot \Delta_i + T_i \cdot \beta \sqrt{\mathfrak{M}_2} \cdot \theta^k \Delta_i) \\ & \leq \frac{\theta^k}{\sqrt{\mathfrak{M}_2}} + [\alpha \cdot \theta^k + T_0 \cdot \beta \cdot \sqrt{\mathfrak{M}_2} \cdot \theta^k] \sum_{i=0}^{i_{\max}+1} \Delta_i \end{aligned}$$

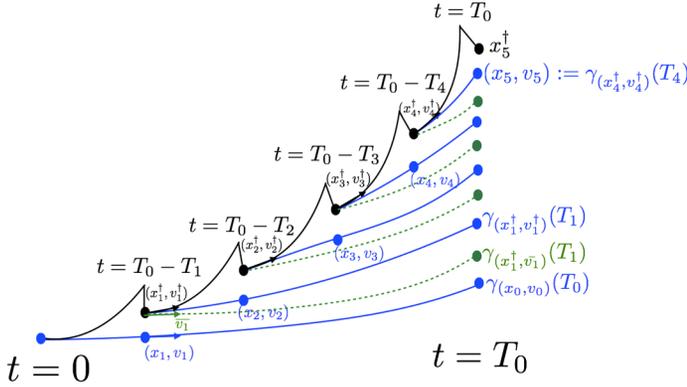


FIG. 4. This is an illustration of the proof of Lemma 2. Steps taken by the oracle \star [which is illustrated here as going along a short curved path and then projecting its position $x_i^\dagger := \gamma_{(x_{i-1}, v_{i-1}^\dagger)}^\star$ (black dot) and velocity $v_i^\dagger := \gamma_{(x_{i-1}, v_{i-1}^\dagger)}^\star$ (black arrow) back onto the phase space] are in black. The true geodesic paths are blue curves or green dashed curves. Only the geodesic path $\gamma_{(x_0, v_0)}(t)$ on the bottom belongs to the true geodesic walk. We imagine the other geodesic paths to help us bound the error. The distance between the blue dot x_i and black dot x_i^\dagger at each time $t = T_0 - T_i$, and the angle between the velocities \bar{v}_i and v_i^\dagger , where v_i is the velocity at the blue dot and $\bar{v}_i := \phi(v_i, x_i, x_i^\dagger)$, are bounded because of our assumptions on the accuracy of the oracle \star . The distance from any green dot at $t = T_0$ to the blue dot directly below it is bounded using Lemma 1. The distance from that same green dot to the blue dot directly above it is bounded using the Toponogov triangle comparison theorem.

$$\begin{aligned} &\leq \frac{\theta^k}{\sqrt{\mathfrak{M}_2}} + [\alpha \cdot \theta^k + T_0 \cdot \beta \cdot \sqrt{\mathfrak{M}_2} \cdot \theta^k] \cdot T_0 \\ &\leq \frac{\theta^k}{\sqrt{\mathfrak{M}_2}} + \left[\alpha + \frac{\pi}{2\sqrt{\mathfrak{M}_2}} \cdot \beta \cdot \sqrt{\mathfrak{M}_2} \right] \cdot \frac{\pi}{2\sqrt{\mathfrak{M}_2}} \cdot \theta^k, \\ &= \left[1 + \frac{\pi}{2}\alpha + \left(\frac{\pi}{2}\right)^2 \beta \right] \cdot \frac{1}{\sqrt{\mathfrak{M}_2}} \cdot \theta^k. \end{aligned}$$

The second inequality holds because $\gamma_{(x_0, v_0)}(t)$ has unit velocity for every $t \in \mathbb{R}$. The third inequality is the triangle inequality (see Figure 4). The bound on $T - T_0$ in the fourth inequality is a consequence of Step 8 of Algorithm 2, while the bounds on the other two terms follow from equations (8.8) and (8.9). The last inequality uses our assumption that $T_0 \leq T \leq \frac{\pi}{2\sqrt{\mathfrak{M}_2}}$. \square

We show that the random walk described by Algorithm 2 can be used to sample approximately uniformly from \mathcal{M} . We first need a simple generic bound.

LEMMA 3. Let K be a transition kernel on metric space (Ω, d) with unique stationary measure μ . Assume that there exists some contraction coefficient $\kappa > 0$

so that K satisfies

$$(8.10) \quad W_d^{(1)}(K(x, \cdot), K(y, \cdot)) \leq (1 - \kappa)d(x, y)$$

for all $x, y \in \Omega$. Let Q be a transition kernel on (Ω, d) with stationary measure ν . Assume that there exists some $\delta \geq 0$ so that

$$(8.11) \quad \sup_{x \in \Omega} W_d^{(1)}(K(x, \cdot), Q(x, \cdot)) < \delta.$$

For fixed $x \in \Omega$ and $Y \sim \mu$, define the eccentricity $E(x) = \mathbb{E}[d(x, Y)]$ and assume that $E(x) < \infty$ for all $x \in \Omega$. Then Q satisfies

$$(8.12) \quad W_d^{(1)}(Q^t(x, \cdot), \mu) \leq (1 - \kappa)^t E(x) + \frac{\delta}{\kappa}$$

for all $x \in \Omega$ and $t \in \mathbb{N}$. Under the further assumption that $\sup_x E(x) < \infty$, we also have

$$(8.13) \quad W_d^{(1)}(\mu, \nu) \leq \frac{\delta}{\kappa}.$$

PROOF. Fix a tolerance $\eta > 0$ and starting points $x, y \in \Omega$. We begin by constructing a coupling of $Q(x, \cdot)$ and $K(y, \cdot)$.

By inequality (8.10), it is possible to couple two random variables $X' \sim K(x, \cdot)$ and $Y' \sim K(y, \cdot)$ so that

$$(8.14) \quad \mathbb{E}[d(X', Y')] \leq (1 - \kappa)d(x, y) + \eta.$$

Furthermore, by inequality (8.11) and the standard gluing lemma (see Chapter 1 of [39]), it is possible to couple X', Y' to the random variable $X \sim Q(x, \cdot)$ so that

$$(8.15) \quad \mathbb{E}[d(X, X')] \leq \delta + \eta.$$

Combining inequalities (8.14) and (8.15), it is possible to couple $X \sim Q(x, \cdot)$ and $Y' \sim K(y, \cdot)$ so that

$$(8.16) \quad \begin{aligned} \mathbb{E}[d(X, Y')] &\leq \mathbb{E}[d(X, X')] + \mathbb{E}[d(X', Y')] \\ &\leq \delta + (1 - \kappa)d(x, y) + 2\eta. \end{aligned}$$

We denote by M the kernel on Ω^2 given by this coupling of $Q(x, \cdot)$ and $K(y, \cdot)$.

Let $x \in \Omega$, and let $Y \sim \mu$. Let $\{(X_t, Y_t)\}_{t \geq 0}$ be a Markov chain evolving according to the kernel M with initial conditions $(X_0, Y_0) = (x, Y)$. By inequality (8.16), we have for all $t \in \mathbb{N}$ that

$$\begin{aligned} \mathbb{E}[d(X_t, Y_t)] &\leq \mathbb{E}[(1 - \kappa)\mathbb{E}[d(X_{t-1}, Y_{t-1})] + \delta + 2\eta] \\ &\leq (1 - \kappa)^2 \mathbb{E}[d(X_{t-2}, Y_{t-2})] + (\delta + 2\eta)(1 + (1 - \kappa)) \\ &\leq \dots \\ &\leq (1 - \kappa)^t \mathbb{E}[d(X_0, Y_0)] + \frac{\delta + 2\eta}{\kappa}. \end{aligned}$$

Since this holds for arbitrary $\eta > 0$ and $Y_t \sim \mu$ for all $t \in \mathbb{N}$, we have

$$(8.17) \quad W_d^{(1)}(Q^t(x, \cdot), \mu) \leq (1 - \kappa)^t E(x) + \frac{\delta}{\kappa}.$$

This completes the proof of inequality (8.12). Inequality (8.13) follows immediately from letting t go to infinity. \square

We apply this bound.

LEMMA 4. *Let $\epsilon > 0$. Set $T = \frac{\pi}{2\sqrt{\mathfrak{M}_2}}$ and set*

$$(8.18) \quad \theta(\epsilon) = \min\left(\left[\epsilon \cdot \sqrt{\mathfrak{M}_2} \cdot \frac{\frac{\pi^2}{32} \frac{m_2}{\mathfrak{M}_2}}{2(1 + \frac{\pi}{2}\alpha + (\frac{\pi}{2})^2\beta)}\right]^{\frac{1}{k}}, \frac{1}{70} \frac{m_2}{\mathfrak{M}_2}\right),$$

$$(8.19) \quad I(\epsilon) = \left\lceil \frac{\log\left(\frac{\epsilon\sqrt{m_2}}{2\pi}\right)}{\log\left(1 - \frac{\pi^2}{32} \frac{m_2}{\mathfrak{M}_2}\right)} \right\rceil.$$

Then the distribution $\Pi_i^{\theta(\epsilon)}$ of the i 'th step of the chain described in Equation (8.6) satisfies

$$(8.20) \quad W_d^{(1)}(\Pi, \Pi_i^{\theta(\epsilon)}) \leq \epsilon$$

for all $i > I(\epsilon)$.

PROOF. We apply Lemma 3, with $K = \mathbb{K}$ and Q the kernel of the Markov chain in equation (8.6) with parameters $\theta = \theta(\epsilon)$ and $T = \frac{\pi}{2\sqrt{\mathfrak{M}_2}}$. We keep the notation of Lemma 3 and check that the assumptions are satisfied.

By Theorem 6.1, inequality (8.10) is satisfied with

$$\kappa = \frac{\pi^2}{32} \frac{m_2}{\mathfrak{M}_2}.$$

By Lemma 2, inequality (8.11) is satisfied with

$$\delta = \frac{1 + \frac{\pi}{2}\alpha + (\frac{\pi}{2})^2\beta}{\sqrt{\mathfrak{M}_2}} \cdot \theta^k.$$

Finally, by inequality (2.9), the diameter of \mathcal{M} satisfies

$$D \leq \frac{\pi}{\sqrt{m_2}}.$$

Applying Lemma 3 with these values of κ , δ and D completes the proof of this theorem. \square

This immediately gives the following dimension-free bound on the number of oracle calls required to obtain a sample from the uniform distribution on \mathcal{M} with error less than some fixed $\epsilon > 0$.

THEOREM 8.1. Fix $\epsilon > 0$ and let $\theta(\epsilon), I(\epsilon)$ be as in Equation (8.18). Also let $X_{I(\epsilon)+1}^{\theta(\epsilon)}$ be the $(I(\epsilon) + 1)$ 'st point generated by Equation (8.6) with parameters $\theta = \theta(\epsilon), T = \frac{\pi}{2\sqrt{\mathfrak{M}_2}}$.

Then the distribution $\Pi_{I(\epsilon)+1}^{\theta(\epsilon)}$ of $X_{I(\epsilon)+1}^{\theta(\epsilon)}$ satisfies

$$(8.21) \quad W_d^{(1)}(\Pi_{I(\epsilon)+1}^{\theta(\epsilon)}, \Pi) \leq \epsilon,$$

and the number $N(\epsilon)$ of \star oracle calls used to generate $X_{I(\epsilon)+1}^{\theta(\epsilon)}$ satisfies

$$(8.22) \quad N(\epsilon) \leq \frac{\log\left(\frac{\epsilon\sqrt{\mathfrak{m}_2}}{2\pi}\right)}{\log\left(1 - \frac{\pi^2}{32} \frac{\mathfrak{m}_2}{\mathfrak{M}_2}\right)} \left(\left\lceil \frac{\pi}{2\sqrt{\mathfrak{M}_2}\Delta_{\theta(\epsilon); \min}^*} \right\rceil + \mathcal{O}\left(\log\left(\frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \times \frac{\sqrt{\mathfrak{M}_2}}{\theta^k(\epsilon)}\right)\right) \right)$$

$$(8.23) \quad = \mathcal{O}^*\left(\frac{\mathfrak{M}_2}{\mathfrak{m}_2} \times \frac{1}{\sqrt{\mathfrak{M}_2}\Delta_{\theta(\epsilon); \min}^*}\right).$$

PROOF. Inequality (8.21) follows immediately from Lemma 4.

By Lemma 2,

$$\begin{aligned} N(\epsilon) &\leq I(\epsilon) \times \left(\left\lceil \frac{T}{\Delta_{\theta(\epsilon); \min}^*} \right\rceil + \mathcal{O}\left(\log\left(\frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \times \frac{\sqrt{\mathfrak{M}_2}}{\theta^k(\epsilon)}\right)\right) \right) \\ &\leq \frac{\log\left(\frac{\epsilon\sqrt{\mathfrak{m}_2}}{2\pi}\right)}{\log\left(1 - \frac{\pi^2}{32} \frac{\mathfrak{m}_2}{\mathfrak{M}_2}\right)} \times \left(\left\lceil \frac{\pi}{2\sqrt{\mathfrak{M}_2}\Delta_{\theta(\epsilon); \min}^*} \right\rceil + \mathcal{O}\left(\log\left(\frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \times \frac{\sqrt{\mathfrak{M}_2}}{\theta^k(\epsilon)}\right)\right) \right). \end{aligned}$$

This completes the proof of the theorem. \square

8.2. Approximating geodesics on convex body boundaries. In this section, we show how to construct an oracle that satisfies the requirements of Lemma 2 in the important special case when $\mathcal{M} = \partial\mathcal{K}$ is the boundary of a convex body \mathcal{K} . To build our oracle \star (Algorithm 3), we use only the same basic inclusion-exclusion oracle and oracle for the normal line to $\partial\mathcal{K}$ used in the stochastic billiards algorithm of [12]. In particular, one can construct an intersection oracle that returns the intersection of any line with $\partial\mathcal{K}$ using the bisection method and the inclusion-exclusion oracle, which we will use to build \star .

In the remainder of the paper, for any line segment $\ell_{a,b} \subset \mathbb{R}^{d+1}$ with endpoints $a, b \in \mathbb{R}^{d+1}$, we denote its Euclidean length by $\text{length}(\ell_{a,b}) := \|b - a\|$. For every $y \in \mathcal{M} \equiv \partial\mathcal{K}$, define $\mathbf{n}(y)$ to be the unit normal vector to \mathcal{M} at y pointing into \mathcal{K} , and denote by “ $D_u w(y)$ ” the directional derivative in the direction u of a vector-valued function $w : \mathcal{M} \mapsto \mathbb{R}^{d+1}$. We will denote by $\mathbf{n}^\perp(\cdot)$ the unique codimension-1 subspace orthogonal to $\mathbf{n}(\cdot)$.

Algorithm 3 An oracle \star in the special case where \mathcal{M} is the boundary of a convex body \mathcal{K}

input: Intersection oracle for the convex body \mathcal{K} .

input: Oracle for the normal vector of $\mathcal{M} = \partial\mathcal{K}$.

input: Initial point $x \in \mathcal{M}$, initial velocity $v \in \mathcal{T}_x\mathcal{M}$, parameter $\theta > 0$.

output: $\Delta_{\theta;(x,v)}^{\star}, \psi_{\theta;(x,v)}^{\star} := (\gamma_{\theta;(x,v)}^{\star}, \varphi_{\theta;(x,v)}^{\star})$.

- 1: Generate the unique unit vector u pointing into \mathcal{K} at x at an angle θ to the tangent plane $\mathcal{T}_x\mathcal{M}$, whose projection \tilde{u} onto $\mathcal{T}_x\mathcal{M}$ satisfies $\angle(\tilde{u}, v) = 0$ (using the normal vector oracle).
 - 2: Let $\ell := \{su : s \in \mathbb{R}\}$ be the line passing through x in the direction u , and let P be the 2-plane spanned by v and u .
 - 3: Solve for the other intersection point x_1^\dagger in $\ell \cap \mathcal{M}$ by calling the intersection oracle.
 - 4: Set \tilde{v}_1 to be the projection of u onto the 1-dimensional subspace $Q := \mathbf{n}^\perp(x_1^\dagger) \cap P$, and set $v_1^\dagger = \frac{\tilde{v}_1}{\|\tilde{v}_1\|}$.
 - 5: Set $(\gamma_{\theta;(x,v)}^{\star}, \varphi_{\theta;(x,v)}^{\star}) = (x_1^\dagger, v_1^\dagger)$.
 - 6: Set $\Delta_{\theta;(x,v)}^{\star} = \|x_1^\dagger - x\|$.
-

In this section, we will make a slightly strengthened assumption about the curvature. Toward this end, we define the inner radius of curvature at a point $x \in \partial\mathcal{K}$ to be the radius of the largest sphere in \mathbb{R}^{d+1} that is tangent to $\partial\mathcal{K}$ at x and contained in \mathcal{K} . Similarly, we define the outer radius of curvature to be radius of the smallest sphere that is tangent to $\partial\mathcal{K}$ at x that contains \mathcal{K} .

ASSUMPTION 3. $\mathcal{M} = \partial\mathcal{K}$ has inner radius of curvature uniformly bounded below by $\frac{1}{\sqrt{\mathfrak{M}_2}}$ and outer radius of curvature uniformly bounded above by $\frac{1}{\sqrt{\mathfrak{m}_2}}$.

In particular, the above assumptions on the inner and outer radii of curvature imply our previous assumptions that the sectional curvature is bounded above and below by \mathfrak{M}_2 and \mathfrak{m}_2 , respectively, although the converse is not true.

The following lemma (Lemma 5) relates the geodesic distance on a convex body to the Euclidean distance in the ambient space. We will use Lemma 5 to prove the main result of this section (Lemma 6).

Before we state Lemma 5, we define $\text{dist}_{\mathcal{N}}(x, y)$ to be the geodesic distance between two points $x, y \in \mathcal{N}$ in a manifold \mathcal{N} . For every set $S = \{x, y\}$ consisting of two distinct points $x, y \in \mathcal{N}$, we also define $\text{dist}_{\mathcal{N}}(S) := \text{dist}_{\mathcal{N}}(x, y)$.

LEMMA 5. Let $q^{(1)}$ and $q^{(2)}$ be two points on the boundary $\partial\mathcal{K}$ of a convex body \mathcal{K} , with inner and outer radii of curvature bounded below and above by $\frac{1}{\sqrt{\mathfrak{M}_2}}$

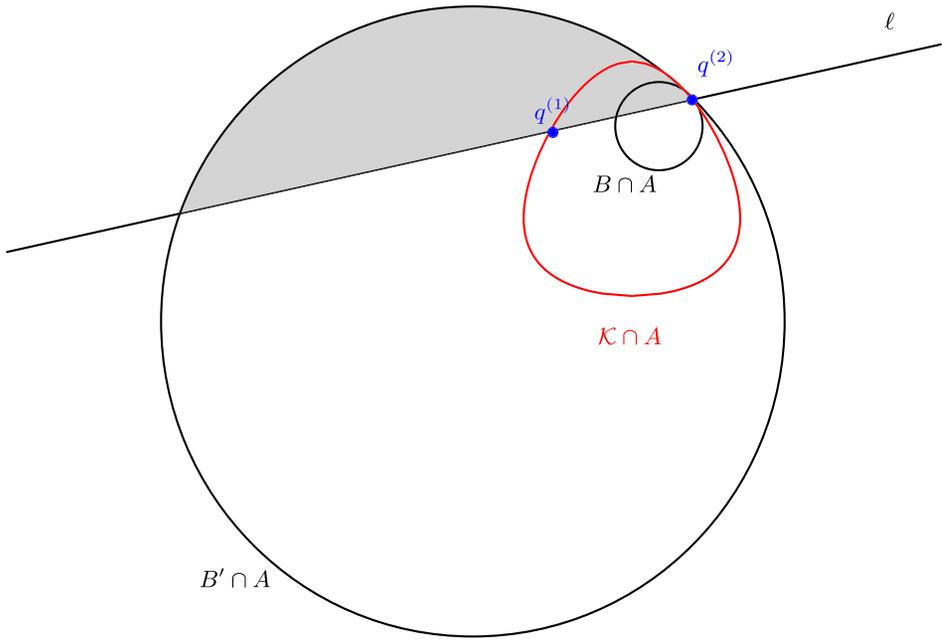


FIG. 5. An illustration of the convex geometry in the proof of Lemma 5.

and $\frac{1}{\sqrt{m_2}}$, respectively. Then

$$(8.24) \quad \text{dist}_{\partial\mathcal{K}}(q^{(1)}, q^{(2)}) \leq 2\pi \frac{\sqrt{M_2}}{\sqrt{m_2}} \|q^{(2)} - q^{(1)}\|.$$

PROOF. Let ℓ be the line passing through both $q^{(1)}$ and $q^{(2)}$. Let B and B' be balls tangent to $\partial\mathcal{K}$ at $q^{(2)}$ of radius $\frac{1}{\sqrt{M_2}}$ and $\frac{1}{\sqrt{m_2}}$, respectively. By assumption,

$$(8.25) \quad B \subseteq \mathcal{K} \subseteq B'.$$

Since B , \mathcal{K} and B' are all convex, we therefore have that (see Figure 5),

$$(8.26) \quad \begin{aligned} \text{dist}_{\partial\mathcal{K}}(\ell \cap \partial\mathcal{K}) &\leq \text{dist}_{\partial\mathcal{K} \cap A}(\ell \cap \partial\mathcal{K}) \\ &\leq \text{dist}_{\partial B'}(\ell \cap \partial B') + \text{length}(\ell \cap B') \\ &\leq 2 \text{dist}_{\partial B'}(\ell \cap \partial B'), \end{aligned}$$

where A is the 2-plane containing the great circle (i.e., the spherical geodesic) in B' that connects the two points in $\ell \cap \partial B'$.

Since the ratio $\frac{\text{dist}_{\partial B'}(\ell \cap \partial B')}{\text{length}(\ell \cap B')}$ is maximized when $\ell \cap B'$ is the diameter of B' , it must be true that

$$(8.27) \quad \text{dist}_{\partial B'}(\ell \cap \partial B') \leq \pi \text{length}(\ell \cap B').$$

Assuming without loss of generality that $q^{(2)}$ is the origin, we note

$$(8.28) \quad B' = \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} B,$$

so that

$$(8.29) \quad \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} (\ell \cap B) = \ell \cap \left(\frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} B \right) = \ell \cap B'.$$

Hence,

$$(8.30) \quad \begin{aligned} & \text{length}(\ell \cap B') \\ &= \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \text{length}(\ell \cap B) \leq \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \text{length}(\ell \cap \partial K) \\ &= \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \|q^{(2)} - q^{(1)}\|. \end{aligned}$$

Therefore, combining equations (8.26), (8.27) and (8.30), we get:

$$\begin{aligned} \text{dist}_{\partial K}(q^{(1)}, q^{(2)}) &= \text{dist}_{\partial K}(\ell \cap \partial K) \leq 2 \text{dist}_{\partial B'}(\ell \cap \partial B') \\ &\leq 2\pi \text{length}(\ell \cap B') \leq 2\pi \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \|q^{(2)} - q^{(1)}\|. \quad \square \end{aligned}$$

LEMMA 6. *Let \mathcal{K} be a convex body and $\mathcal{M} = \partial\mathcal{K}$ be its boundary. Set $\alpha = 5\pi \frac{\mathfrak{M}_2}{\mathfrak{m}_2}$ and $\beta = 10 \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}}$. Fix $(x, v) \in \mathcal{M}^\circ$. For $0 < \theta < \frac{1}{70} \frac{\mathfrak{m}_2}{\mathfrak{M}_2}$ that satisfies $\beta \cdot \sqrt{\mathfrak{M}_2} \cdot \theta \cdot \Delta_{\theta; (x, v)}^* < 1$, let $\psi_{\theta; (x, v)}^* := (\gamma_{\theta; (x, v)}^*, \varphi_{\theta; (x, v)}^*)$ and $\Delta_{\theta; (x, v)}^*$ be generated by Algorithm 3. Then*

$$(8.31) \quad \text{dist}(\gamma_{(x, v)}(\Delta_{\theta; (x, v)}^*), \gamma_{\theta; (x, v)}^*) \leq \alpha \cdot \theta \cdot \Delta_{\theta; (x, v)}^*$$

and

$$(8.32) \quad \|\varphi_{\theta; (x, v)}^* - \overline{\varphi_{(x, v)}(\Delta_{\theta; (x, v)}^*)}\| \leq \beta \cdot \sqrt{\mathfrak{M}_2} \cdot \theta \cdot \Delta_{\theta; (x, v)}^*.$$

Moreover, the function $\Delta_{\theta; (x, v)}^*$ satisfies Assumption 1, with

$$(8.33) \quad \frac{\theta}{2\sqrt{\mathfrak{M}_2}} \leq \Delta_{\theta; (x, v)}^* \leq \frac{\theta}{\sqrt{\mathfrak{m}_2}}$$

for all $(x, v) \in \mathcal{M}^\circ$.

REMARK 6. The most difficult part of this proof is checking the bound 8.32 on the error in the velocity approximation $\varphi_{\theta; (x, v)}^*$, which requires carefully relating the intrinsic geometry of $\partial\mathcal{K}$ to its embedding in Euclidean

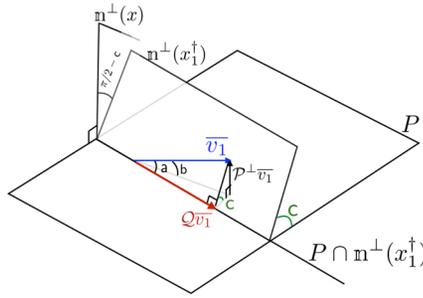


FIG. 6. An illustration of the planes and vectors used to bound the error in the velocity in the proof of Lemma 6.

space. We first show that the component of the error *orthogonal* to a certain 2-plane P is $\mathcal{O}(\theta^2)$, and thus negligible. The component of the error *contained* in P is dealt with separately via a geometric argument, illustrated in Figure 6. The chosen 2-plane P is the plane spanned by the vector v and the normal vector to \mathcal{K} at x . These two vectors represent the initial velocity and acceleration of the geodesic path, and so roughly speaking the plane P captures the zeroth- and first-order terms in our approximation of the velocity.

PROOF. We begin by proving inequality (8.33). Let $u, P, Q, \ell, \Delta_{\theta;(x,v)}^*$, and $\psi_{\theta;(x,v)}^*$ be as in Algorithm 3. Recall that Algorithm 3 sets $\Delta_{\theta;(x,v)}^* = \text{length}(\ell \cap \mathcal{K})$. Since the tangent sphere to $\partial\mathcal{K}$ at x of radius $\frac{1}{\sqrt{\mathfrak{M}_2}}$ is contained in \mathcal{K} , and the tangent sphere to $\partial\mathcal{K}$ at x of radius $\frac{1}{\sqrt{m_2}}$ contains \mathcal{K} (Figure 5), and $0 < \theta < \frac{1}{70}$, we have

$$(8.34) \quad \frac{\theta}{2\sqrt{\mathfrak{M}_2}} \leq \frac{1}{\sqrt{\mathfrak{M}_2}} \sin(\theta) \leq \Delta_{\theta;(x,v)}^* \leq \frac{1}{\sqrt{m_2}} \sin(\theta) \leq \frac{\theta}{\sqrt{m_2}}.$$

Here, we used the fact that

$$(8.35) \quad \frac{1}{2}z \leq \sin(z) \quad \text{for } 0 \leq z \leq \frac{\pi}{2},$$

a fact we will use again later in this proof. Hence, the function $\Delta_{\theta;(x,v)}^*$ satisfies Assumption 1. This completes the proof of inequality (8.33).

Next, we prove inequality (8.31). Define $(x_1, v_1) := \psi_{(x,v)}(\Delta_{\theta;(x,v)}^*)$, and $(x_1^\dagger, v_1^\dagger) := \psi_{\theta;(x,v)}^*$. By Assumption 3, the principal curvatures of \mathcal{M} are bounded above by $\sqrt{\mathfrak{M}_2}$, so $\|\gamma''_{(x,v)}(t)\| \leq \sqrt{\mathfrak{M}_2}$ for all $x, v \in \mathcal{M}^\circ$ and $t \in \mathbb{R}$. Therefore,

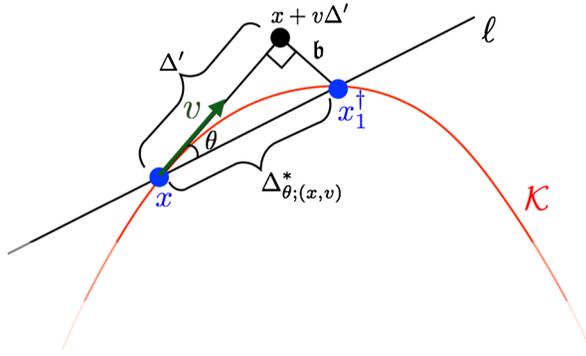


FIG. 7. Bounding the error in the position in the proof of Lemma 6.

since $\gamma_{(x,v)}(t) = x + vt + \int_0^t \int_0^\tau \gamma''_{(x,v)}(s) \, ds \, d\tau$, it must be true that

$$\begin{aligned}
 & \|x + vt - \gamma_{(x,v)}(t)\| \\
 (8.36) \quad &= \left\| x + vt - \left(x + vt + \int_0^t \int_0^\tau \gamma''_{(x,v)}(s) \, ds \, d\tau \right) \right\| \\
 &= \left\| \int_0^t \int_0^\tau \gamma''_{(x,v)}(s) \, ds \, d\tau \right\| \leq \int_0^t \int_0^\tau \|\gamma''_{(x,v)}(s)\| \, ds \, d\tau \\
 &\leq \int_0^t \int_0^\tau \sqrt{\mathfrak{M}_2} \, ds \, d\tau = \frac{1}{2} \sqrt{\mathfrak{M}_2} \cdot t^2
 \end{aligned}$$

for every $t \geq 0$.

Let Δ' be the unique number such that $\langle x + v\Delta' - x_1^\dagger, v \rangle = 0$, and let $\mathfrak{b} := \|(x + v\Delta') - x_1^\dagger\|$ (Figure 7). Then the line segment connecting $(x + v\Delta')$ and x_1^\dagger is orthogonal to v . Then we have

$$\begin{aligned}
 (8.37) \quad & \mathfrak{b} = \sin(\theta) \times \Delta_{\theta;(x,v)}^*, \\
 & \Delta' = \cos(\theta) \times \Delta_{\theta;(x,v)}^*.
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 (8.38) \quad & \|x_1 - x_1^\dagger\| \leq \|x_1 - (x + v\Delta_{\theta;(x,v)}^*)\| \\
 & \quad + \|(x + v\Delta_{\theta;(x,v)}^*) - (x + v\Delta')\| + \|(x + v\Delta') - x_1^\dagger\| \\
 &= \|x_1 - (x + v\Delta_{\theta;(x,v)}^*)\| + \|(\Delta_{\theta;(x,v)}^* - \Delta')v\| + \mathfrak{b} \\
 & \stackrel{\text{Eq. (8.37)}}{=} \|\gamma_{(x,v)}(\Delta_{\theta;(x,v)}^*) - (x + v\Delta_{\theta;(x,v)}^*)\| \\
 & \quad + |\Delta_{\theta;(x,v)}^* - \cos(\theta)\Delta_{\theta;(x,v)}^*| \times \|v\| + \sin(\theta)\Delta_{\theta;(x,v)}^*
 \end{aligned}$$

$$\begin{aligned}
 &\stackrel{\text{Eq. (8.36)}}{\leq} \frac{1}{2}\sqrt{\mathfrak{M}_2} \cdot (\Delta_{\theta; (x,v)}^*)^2 \\
 &\quad + \Delta_{\theta; (x,v)}^* \times (1 - \cos(\theta)) \times 1 + \sin(\theta) \times \Delta_{\theta; (x,v)}^* \\
 &= \frac{1}{2}\sqrt{\mathfrak{M}_2} \cdot (\Delta_{\theta; (x,v)}^*)^2 + \Delta_{\theta; (x,v)}^* \times [1 - \cos(\theta) + \sin(\theta)] \\
 &\leq \frac{1}{2}\sqrt{\mathfrak{M}_2} \cdot (\Delta_{\theta; (x,v)}^*)^2 + \Delta_{\theta; (x,v)}^* \times 2\theta
 \end{aligned}$$

since $1 - \cos(\theta) + \sin(\theta) \leq 2\theta$ for $\theta \geq 0$.

Hence,

$$\begin{aligned}
 &\text{dist}(\gamma_{(x,v)}(\Delta_{\theta; (x,v)}^*), \gamma_{\theta; (x,v)}^*) \\
 &= \text{dist}(x_1, x_1^\dagger) \stackrel{\text{Lemma 5}}{\leq} 2\pi \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \|x_1 - x_1^\dagger\| \\
 &\stackrel{\text{Eq. (8.38)}}{\leq} 2\pi \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \left[\frac{1}{2}\sqrt{\mathfrak{M}_2} \times (\Delta_{\theta; (x,v)}^*)^2 + \Delta_{\theta; (x,v)}^* \times 2\theta \right] \\
 &\stackrel{\text{Eq. (8.34)}}{\leq} 2\pi \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \left[\frac{1}{2}\sqrt{\mathfrak{M}_2} \times \Delta_{\theta; (x,v)}^* \times \frac{\theta}{\sqrt{\mathfrak{m}_2}} + \Delta_{\theta; (x,v)}^* \times 2\theta \right] \\
 &\leq 2\pi \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \left[\frac{1}{2} \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \times \Delta_{\theta; (x,v)}^* \times \theta + 2 \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \times \Delta_{\theta; (x,v)}^* \times \theta \right] \\
 &= 5\pi \left(\frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \right)^2 \times \Delta_{\theta; (x,v)}^* \times \theta = \alpha \times \Delta_{\theta; (x,v)}^* \times \theta,
 \end{aligned}$$

where the last inequality holds since $\frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \geq 1$. This completes the proof of inequality (8.31).

Finally, we prove inequality (8.32). For $(x, v) \in \mathcal{M}^\circ$, define $H_x(v) := \|\gamma''_{(x,v)}(0)\|$. By Assumption 3,

$$(8.39) \quad \sqrt{\mathfrak{m}_2} \leq H_x(v) \leq \sqrt{\mathfrak{M}_2}$$

and

$$(8.40) \quad \|D_u \mathbf{n}(y)\| \leq \sqrt{\mathfrak{M}_2}$$

at every $(x, v), (y, u) \in \mathcal{M}^\circ$.

Recall that P is the plane spanned by the vectors $\mathbf{n}(x)$ and u [this 2-plane is also the 2-plane of curvature associated with the curve $\gamma_{(x,v)}(t)$ at $t = 0$], and let P^\perp be its orthogonal complement. Denote by \mathcal{P} and \mathcal{P}^\perp the operators on \mathbb{R}^{d+1} that project a vector onto P and P^\perp , respectively.

Define $(x(t), v(t)) := \psi_{(x,v)}(t)$. Since $\mathcal{P}^\perp v(0) = 0$ and $\mathcal{P}^\perp \mathbf{n}(\gamma_{(x,v)}(0)) = \mathcal{P}^\perp \mathbf{n}(x) = 0$, we have

$$\begin{aligned}
 & \|\mathcal{P}^\perp v(t)\| \\
 &= \left\| \int_0^t \mathcal{P}^\perp H_{x(\tau)}(v(\tau)) \mathbf{n}(\gamma_{(x,v)}(\tau)) \, d\tau \right\| \\
 (8.41) \quad & \stackrel{\text{Eq. (8.39)}}{\leq} \int_0^t \|\sqrt{\mathfrak{M}_2} \mathcal{P}^\perp \mathbf{n}(\gamma_{(x,v)}(\tau))\| \, d\tau \\
 &= \sqrt{\mathfrak{M}_2} \int_0^t \left\| \int_0^\tau D_{v(s)} \mathcal{P}^\perp \mathbf{n}(\gamma_{(x,v)}(s)) \, ds + \mathcal{P}^\perp \mathbf{n}(\gamma_{(x,v)}(0)) \right\| \, d\tau \\
 &\leq \sqrt{\mathfrak{M}_2} \int_0^t \int_0^\tau \|D_{v(s)} \mathcal{P}^\perp \mathbf{n}(\gamma_{(x,v)}(s))\| \, ds \, d\tau \\
 & \stackrel{\text{Eq. (8.40)}}{\leq} \sqrt{\mathfrak{M}_2} \int_0^t \int_0^\tau \sqrt{\mathfrak{M}_2} \, ds \, d\tau = \frac{(\sqrt{\mathfrak{M}_2})^2}{2} t^2.
 \end{aligned}$$

Therefore, equation (8.41) gives

$$(8.42) \quad \|\mathcal{P}^\perp v_1\| = \|\mathcal{P}^\perp v(\Delta_{\theta; (x,v)}^\star)\| \leq \frac{(\sqrt{\mathfrak{M}_2})^2}{2} (\Delta_{\theta; (x,v)}^\star)^2.$$

Now define $\bar{v}_1(t) := \phi(t; v_1; x_1, x_1^\dagger)$ to be the parallel transport of v_1 from x_1 to x_1^\dagger along the distance-minimizing unit-speed geodesic $\omega(t) \equiv \omega(t; x_1, x_1^\dagger)$. Let $t^\sharp := \text{dist}(x_1, x_1^\dagger)$ and let $\bar{v}_1 := \bar{v}_1(t^\sharp)$. Then

$$\frac{d}{dt} \overline{v_1(t)} = \left\langle \overline{v_1(t)}, \frac{d}{dt} \mathbf{n}(\omega(t)) \right\rangle \mathbf{n}(\omega(t)).$$

Therefore, by a similar calculation to that in equation (8.41),

$$\begin{aligned}
 & \|\mathcal{P}^\perp v_1 - \mathcal{P}^\perp \bar{v}_1\| \\
 &= \left\| \int_0^{t^\sharp} \mathcal{P}^\perp \frac{d}{d\tau} \overline{v_1(\tau)} \, d\tau \right\| \\
 &= \int_0^{t^\sharp} \left\| \mathcal{P}^\perp \left\langle \overline{v_1(\tau)}, \frac{d}{d\tau} \mathbf{n}(\omega(\tau)) \right\rangle \mathbf{n}(\omega(\tau)) \right\| \, d\tau \\
 &\leq \int_0^{t^\sharp} \left\| \frac{d}{d\tau} \mathbf{n}(\omega(\tau)) \right\| \times \|\mathcal{P}^\perp \mathbf{n}(\omega(\tau))\| \, d\tau \\
 & \stackrel{\text{Eq. (8.40)}}{\leq} \int_0^{t^\sharp} \sqrt{\mathfrak{M}_2} \|\mathcal{P}^\perp \mathbf{n}(\omega(\tau))\| \, d\tau \\
 (8.43) \quad &= \sqrt{\mathfrak{M}_2} \int_0^{t^\sharp} \left\| \int_0^\tau \frac{d}{ds} \mathcal{P}^\perp \mathbf{n}(\omega(s)) \, ds + \mathcal{P}^\perp \mathbf{n}(\omega(0)) \right\| \, d\tau
 \end{aligned}$$

$$\begin{aligned}
 &\leq \sqrt{\mathfrak{M}_2} \int_0^{t^\sharp} \left\| \int_0^\tau \frac{d}{ds} \mathcal{P}^\perp \mathbf{n}(\omega(s)) ds \right\| + \|\mathcal{P}^\perp \mathbf{n}(x_1)\| d\tau \\
 &= \sqrt{\mathfrak{M}_2} \int_0^{t^\sharp} \left\| \int_0^\tau \frac{d}{ds} \mathcal{P}^\perp \mathbf{n}(s) ds \right\| \\
 &\quad + \left\| \int_0^{\Delta_{\theta; (x,v)}^\star} D_{v(s)} \mathcal{P}^\perp \mathbf{n}(\gamma_{(x,v)}(s)) ds + \mathcal{P}^\perp \mathbf{n}(\gamma_{(x,v)}(0)) \right\| d\tau \\
 &\stackrel{\text{Eq. (8.40)}}{\leq} \sqrt{\mathfrak{M}_2} \int_0^{t^\sharp} \int_0^\tau \sqrt{\mathfrak{M}_2} ds + \int_0^{\Delta_{\theta; (x,v)}^\star} \sqrt{\mathfrak{M}_2} ds d\tau \\
 &= (\sqrt{\mathfrak{M}_2})^2 \left[\frac{1}{2} (t^\sharp)^2 + t^\sharp \times \Delta_{\theta; (x,v)}^\star \right].
 \end{aligned}$$

Since $0 < \theta < \frac{1}{70} \frac{m_2}{\mathfrak{M}_2}$ by assumption, inequality (8.31) implies that

$$(8.44) \quad t^\sharp \leq \Delta_{\theta; (x,v)}^\star.$$

Therefore, by equation (8.43),

$$\begin{aligned}
 (8.45) \quad \|\mathcal{P}^\perp v_1 - \mathcal{P}^\perp \bar{v}_1\| &\leq \frac{3}{2} (\sqrt{\mathfrak{M}_2})^2 (\Delta_{\theta; (x,v)}^\star)^2 \\
 &\leq 2 (\sqrt{\mathfrak{M}_2})^2 (\Delta_{\theta; (x,v)}^\star)^2.
 \end{aligned}$$

Hence, equations (8.42) and (8.45) imply that

$$\begin{aligned}
 (8.46) \quad \|\mathcal{P}^\perp \bar{v}_1\| &\leq \|\mathcal{P}^\perp v_1\| + \|\mathcal{P}^\perp v_1 - \mathcal{P}^\perp \bar{v}_1\| \\
 &\leq \frac{(\sqrt{\mathfrak{M}_2})^2}{2} \cdot (\Delta_{\theta; (x,v)}^\star)^2 + 2 (\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta; (x,v)}^\star)^2 \\
 &= \frac{5}{2} (\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta; (x,v)}^\star)^2.
 \end{aligned}$$

Let $\mathfrak{b} := \angle(\bar{v}_1, P)$. Since \mathfrak{b} is the angle between a vector and a plane, $0 \leq \mathfrak{b} \leq \frac{\pi}{2}$, so $\sin(\mathfrak{b}) \geq \frac{1}{2} \mathfrak{b}$ by equation (8.35). Thus,

$$(8.47) \quad \frac{1}{2} \mathfrak{b} \leq \sin(\mathfrak{b}) = \frac{\|\mathcal{P}^\perp \bar{v}_1\|}{\|\bar{v}_1\|} = \frac{\|\mathcal{P}^\perp \bar{v}_1\|}{1} \stackrel{\text{Eq. (8.46)}}{\leq} \frac{5}{2} (\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta; (x,v)}^\star)^2,$$

and so

$$(8.48) \quad \mathfrak{b} \leq 5 (\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta; (x,v)}^\star)^2.$$

Since $\theta < \frac{1}{70} \frac{m_2}{\mathfrak{M}_2}$ by assumption, we have

$$(8.49) \quad \sqrt{\mathfrak{M}_2} \Delta_{\theta; (x,v)}^\star \stackrel{\text{Eq. (8.33)}}{\leq} \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{m_2}} \theta < \frac{1}{70} \frac{\sqrt{m_2}}{\sqrt{\mathfrak{M}_2}} \leq \frac{1}{70}.$$

Since the radius of curvature of $\partial\mathcal{K}$ is uniformly bounded below by $\frac{1}{\sqrt{\mathfrak{M}_2}}$, this implies

$$\begin{aligned}
 \angle(\mathbf{n}^\perp(x), \mathbf{n}^\perp(x_1^\dagger)) &= \angle(\mathbf{n}(x), \mathbf{n}(x_1^\dagger)) \leq \sqrt{\mathfrak{M}_2} \text{dist}(x, x_1^\dagger) \\
 (8.50) \qquad &\stackrel{\text{Lemma 5}}{\leq} \sqrt{\mathfrak{M}_2} \times 2\pi \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \|x_1^\dagger - x\| \\
 &= 2\pi\sqrt{\mathfrak{M}_2} \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \Delta_{\theta; (x, v)}^*.
 \end{aligned}$$

Also, since $\mathbf{n}(x)$ is contained in the plane P , we have that

$$(8.51) \qquad \angle(P, \mathbf{n}^\perp(x)) = \frac{\pi}{2}.$$

Now,

$$\begin{aligned}
 \angle(\text{proj}_P \mathbf{n}(x), \text{proj}_P \mathbf{n}(x_1^\dagger)) &= \angle(\mathbf{n}(x), \text{proj}_P \mathbf{n}(x_1^\dagger)) \\
 (8.52) \qquad &\leq \angle(\mathbf{n}(x), \mathbf{n}(x_1^\dagger)) + \angle(\mathbf{n}(x_1^\dagger), \text{proj}_P \mathbf{n}(x_1^\dagger)) \\
 &\leq \angle(\mathbf{n}(x), \mathbf{n}(x_1^\dagger)) + \angle(\mathbf{n}(x_1^\dagger), \mathbf{n}(x)) = 2\angle(\mathbf{n}(x), \mathbf{n}(x_1^\dagger)),
 \end{aligned}$$

where the first equality is true since $\mathbf{n}(x) \in P$ implies that $\text{proj}_P \mathbf{n}(x) = \mathbf{n}(x)$, the first inequality is the triangle inequality for angles between vectors, and the second inequality is true by inequality (2.10) and the fact that $\mathbf{n}(x) \in P$.

Define $\mathbf{c} := \angle(P, \mathbf{n}^\perp(x_1^\dagger))$. Equation (8.52) implies that

$$\begin{aligned}
 \angle(\text{proj}_P \mathbf{n}(x_1^\dagger), \mathbf{n}(x_1^\dagger)) &\leq \angle(\text{proj}_P \mathbf{n}(x_1^\dagger), \text{proj}_P \mathbf{n}(x)) \\
 (8.53) \qquad &+ \angle(\text{proj}_P \mathbf{n}(x), \mathbf{n}(x)) + \angle(\mathbf{n}(x), \mathbf{n}(x_1^\dagger)) \\
 &\stackrel{\text{Eq. (8.52)}}{\leq} 2\angle(\mathbf{n}(x), \mathbf{n}(x_1^\dagger)) + \angle(\text{proj}_P \mathbf{n}(x), \mathbf{n}(x)) + \angle(\mathbf{n}(x), \mathbf{n}(x_1^\dagger)) \\
 &= 3\angle(\mathbf{n}(x), \mathbf{n}(x_1^\dagger)) + \angle(\text{proj}_P \mathbf{n}(x), \mathbf{n}(x)) = 3\angle(\mathbf{n}(x), \mathbf{n}(x_1^\dagger)) + 0,
 \end{aligned}$$

so

$$\begin{aligned}
 \mathbf{c} &= \angle(P, \mathbf{n}^\perp(x_1^\dagger)) = \frac{\pi}{2} - \angle(\text{proj}_P \mathbf{n}(x_1^\dagger), \mathbf{n}(x_1^\dagger)) \\
 (8.54) \qquad &\stackrel{\text{Eq. (8.53)}}{\geq} \frac{\pi}{2} - 3\angle(\mathbf{n}(x), \mathbf{n}(x_1^\dagger)) \\
 &\stackrel{\text{Eq. (8.50)}}{\geq} \frac{\pi}{2} - 3 \times 2\pi\sqrt{\mathfrak{M}_2} \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \Delta_{\theta; (x, v)}^* \stackrel{\text{Eq. (8.49)}}{>} \frac{\pi}{4}.
 \end{aligned}$$

Therefore, $\mathbf{n}^\perp(x_1^\dagger)$ does not contain the 2-plane P , so $\mathbf{n}^\perp(x_1^\dagger) \cap P$ must be a 1-dimensional subspace. Define \mathcal{Q} and \mathcal{Q}^\perp to be the projection operators onto the

one-dimensional subspace $Q := \mathbf{n}^\perp(x_1^\dagger) \cap P$ and its orthogonal complement, respectively. We will now show that $\angle(Q\bar{v}_1, v_1^\dagger) = 0$.

$$\begin{aligned}
 \|v(t) - v(0)\| &= \left\| \int_0^t H_{x(\tau)}(v(\tau)) \mathbf{n}(\gamma_{(x,v)}(\tau)) \, d\tau \right\| \\
 (8.55) \quad &\stackrel{\text{Eq. (8.39)}}{\leq} \int_0^t \|\sqrt{\mathfrak{M}_2} \mathbf{n}(\gamma_{(x,v)}(\tau))\| \, d\tau \\
 &= \int_0^t \|\sqrt{\mathfrak{M}_2}\| \times 1 \, d\tau = \sqrt{\mathfrak{M}_2} t.
 \end{aligned}$$

Also,

$$\begin{aligned}
 \|v_1 - \bar{v}_1\| &= \left\| \int_0^{t^\sharp} \frac{d}{d\tau} \overline{v_1(\tau)} \, d\tau \right\| \\
 (8.56) \quad &= \int_0^{t^\sharp} \left\| \left\langle \overline{v_1(\tau)}, \frac{d}{d\tau} \mathbf{n}(\omega(\tau)) \right\rangle \mathbf{n}(\omega(\tau)) \right\| \, d\tau \\
 &\stackrel{\text{Eq. (8.40)}}{\leq} \int_0^{t^\sharp} \sqrt{\mathfrak{M}_2} \|\mathbf{n}(\omega(\tau))\| \, d\tau \\
 &= \int_0^{t^\sharp} \sqrt{\mathfrak{M}_2} \times 1 \, d\tau = \sqrt{\mathfrak{M}_2} t^\sharp \stackrel{\text{Eq. (8.44)}}{\leq} \sqrt{\mathfrak{M}_2} \Delta_{\theta; (x,v)}^*.
 \end{aligned}$$

Hence,

$$\begin{aligned}
 \|v - \bar{v}_1\| &\leq \|v - v_1\| + \|v_1 - \bar{v}_1\| \\
 (8.57) \quad &= \|v(0) - v(\Delta_{\theta; (x,v)}^*)\| + \|v_1 - \bar{v}_1\| \\
 &\stackrel{\text{Eq. (8.55), (8.56)}}{\leq} \sqrt{\mathfrak{M}_2} \Delta_{\theta; (x,v)}^* + \sqrt{\mathfrak{M}_2} \Delta_{\theta; (x,v)}^* \\
 &\stackrel{\text{Eq. (8.49)}}{\leq} \frac{3}{70}.
 \end{aligned}$$

But $\|\bar{v}_1\| = \|v\| = 1$, so \bar{v}_1, v and $v - \bar{v}_1$ form an isosceles triangle whose height bisects the angle $\angle(v, \bar{v}_1)$. Hence,

$$(8.58) \quad \sin\left(\frac{1}{2}\angle(v, \bar{v}_1)\right) = \frac{1}{2} \|v - \bar{v}_1\| \stackrel{\text{Eq. (8.57)}}{\leq} \frac{1}{2} \times \frac{3}{70}.$$

Since $\|v - \bar{v}_1\| \leq \frac{3}{70} < 1 = \|\bar{v}_1\| = \|v\| = 1$, $\angle(v, \bar{v}_1)$ must be the smallest angle of this isosceles triangle, so $\angle(v, \bar{v}_1) \leq \frac{\pi}{3}$. Therefore, since $0 \leq \angle(v, \bar{v}_1) \leq \frac{\pi}{3}$,

$$\frac{1}{2}\angle(v, \bar{v}_1) \stackrel{\text{Eq. (8.35)}}{\leq} 2 \sin\left(\frac{1}{2}\angle(v, \bar{v}_1)\right) \stackrel{\text{Eq. (8.58)}}{\leq} \frac{3}{70}.$$

Hence,

$$(8.59) \quad \angle(v, \bar{v}_1) \leq \frac{6}{70}.$$

Now, any unit normal vector of the line $P \cap \mathbf{n}^\perp(\cdot)$ that is also contained in the plane P is equal (up to a scalar factor) to $\text{proj}_P \mathbf{n}(\cdot)$. Therefore,

$$\begin{aligned}
 \angle(P \cap \mathbf{n}^\perp(x), P \cap \mathbf{n}^\perp(x_1^\dagger)) &= \angle(\text{proj}_P \mathbf{n}(x), \text{proj}_P \mathbf{n}(x_1^\dagger)) \\
 (8.60) \qquad \qquad \qquad &\stackrel{\text{Eq. (8.52)}}{\leq} 2\angle(\mathbf{n}(x), \mathbf{n}(x_1^\dagger)) \\
 &= 2\angle(\mathbf{n}^\perp(x), \mathbf{n}^\perp(x_1^\dagger)).
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \angle(u, P \cap \mathbf{n}^\perp(x_1^\dagger)) &\leq \angle(u, P \cap \mathbf{n}^\perp(x)) + \angle(P \cap \mathbf{n}^\perp(x), P \cap \mathbf{n}^\perp(x_1^\dagger)) \\
 (8.61) \qquad \qquad \qquad &\stackrel{\text{Eq. (8.60)}}{\leq} \angle(u, P \cap \mathbf{n}^\perp(x)) + 2\angle(\mathbf{n}^\perp(x), \mathbf{n}^\perp(x_1^\dagger)) \\
 &= \theta + 2\angle(\mathbf{n}^\perp(x), \mathbf{n}^\perp(x_1^\dagger)) \\
 &\stackrel{\text{Eq. (8.50)}}{\leq} \theta + 2 \times 2\pi\sqrt{\mathfrak{M}_2} \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \Delta_{\theta;(x,v)}^\star \\
 &\stackrel{\text{Eq. (8.49)}}{\leq} \theta + 2 \times \frac{2\pi}{70}.
 \end{aligned}$$

But v_1^\dagger is the (normalized) projection of u onto $P \cap \mathbf{n}^\perp(x_1^\dagger)$, so $\angle(u, v_1^\dagger) = \angle(u, P \cap \mathbf{n}^\perp(x_1^\dagger))$. Hence,

$$\begin{aligned}
 \angle(v, v_1^\dagger) &\leq \angle(v, u) + \angle(u, v_1^\dagger) = \theta + \angle(u, P \cap \mathbf{n}^\perp(x_1^\dagger)) \\
 (8.62) \qquad \qquad \qquad &\stackrel{\text{Eq. (8.61)}}{\leq} \theta + \theta + 2 \times \frac{2\pi}{70} \leq \frac{3}{10},
 \end{aligned}$$

since $0 < \theta < \frac{1}{70} \frac{\mathfrak{m}_2}{\mathfrak{M}_2} \leq \frac{1}{70} \times 1$ by assumption. Therefore,

$$(8.63) \quad \angle(\bar{v}_1, v_1^\dagger) \leq \angle(\bar{v}_1, v) + \angle(v, v_1^\dagger) \stackrel{\text{Eqs. (8.59), (8.62)}}{\leq} \frac{6}{70} + \frac{3}{10} < \frac{\pi}{5}.$$

Now, observe that by equation (8.44),

$$\text{dist}(x, x_1^\dagger) \leq \text{dist}(x, x_1) + \text{dist}(x_1, x_1^\dagger) \leq \Delta_{\theta;(x,v)}^\star + t^\sharp \stackrel{\text{Eq. (8.44)}}{\leq} 2\Delta_{\theta;(x,v)}^\star.$$

Define $\mathbf{a} := \angle(\bar{v}_1, \mathbf{n}^\perp(x_1^\dagger) \cap P)$. Since $\sin(\mathbf{a}) \times \sin(\mathbf{c}) = \sin(\mathbf{b})$ (Figure 6), we have

$$(8.64) \quad \sin(\mathbf{a}) = \frac{\sin(\mathbf{b})}{\sin(\mathbf{c})} \stackrel{\text{Eq. (8.54)}}{\leq} \frac{\sin(\mathbf{b})}{\sin(\frac{\pi}{4})} \stackrel{\text{Eq. (8.47)}}{\leq} \sqrt{2} \times 5(\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta;(x,v)}^\star)^2.$$

Since α is the angle between a vector and a subspace, $0 \leq \alpha \leq \frac{\pi}{2}$, and so

$$\frac{\|Q\bar{v}_1\|}{\|\bar{v}_1\|} = \cos(\alpha) \geq 1 - \sin^2(\alpha) \stackrel{\text{Eq. (8.64)}}{\geq} 1 - (5\sqrt{2}(\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta;(x,v)}^*)^2)^2.$$

Since $\|\bar{v}_1\| = 1$, we get that

$$(8.65) \quad \|Q\bar{v}_1\| \geq 1 - (5\sqrt{2}(\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta;(x,v)}^*)^2)^2.$$

But $v_1^\dagger \in P \cap \mathbf{n}^\perp(x_1^\dagger)$ and $Q\bar{v}_1 \in P \cap \mathbf{n}^\perp(x_1^\dagger)$, so both v_1^\dagger and $Q\bar{v}_1$ lie in the same 1-dimensional subspace, $P \cap \mathbf{n}^\perp(x_1^\dagger)$. Since we have by equation (8.63) that $\angle(\bar{v}_1, v_1^\dagger) < \frac{\pi}{5}$ and both v_1^\dagger and $Q\bar{v}_1$ lie in the same 1-dimensional subspace, it must be true that

$$(8.66) \quad \angle(Q\bar{v}_1, v_1^\dagger) = 0.$$

Therefore,

$$(8.67) \quad \begin{aligned} & \|v_1^\dagger - Q\bar{v}_1\| \\ & \stackrel{\text{Eq. (8.66)}}{=} \|v_1^\dagger\| - \|Q\bar{v}_1\| \\ & \stackrel{\text{Eq. (8.65)}}{\leq} 1 - (1 - (5\sqrt{2}(\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta;(x,v)}^*)^2)^2) \\ & = (5\sqrt{2}(\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta;(x,v)}^*)^2)^2. \end{aligned}$$

Also, by equation (8.64) we have

$$(8.68) \quad \begin{aligned} \|\bar{v}_1 - Q\bar{v}_1\| &= \|\bar{v}_1 - Q^\perp \bar{v}_1\| = \|\bar{v}_1\| \times \sin(\alpha) \\ &= 1 \times \sin(\alpha) \leq 5\sqrt{2}(\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta;(x,v)}^*)^2. \end{aligned}$$

Therefore, by equations (8.67) and (8.68),

$$(8.69) \quad \begin{aligned} \|v_1^\dagger - \bar{v}_1\| &\leq \|v_1^\dagger - Q\bar{v}_1\| + \|Q\bar{v}_1 - \bar{v}_1\| \\ &\leq (5\sqrt{2}(\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta;(x,v)}^*)^2)^2 + 5\sqrt{2}(\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta;(x,v)}^*)^2 \\ &\leq 10(\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta;(x,v)}^*)^2, \end{aligned}$$

where the last inequality holds since $(\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta;(x,v)}^*)^2 < \frac{1}{70^2}$ by equation (8.49). Hence,

$$\begin{aligned} & \|\varphi_{\theta;(x,v)}^* - \overline{\varphi_{(x,v)}(\Delta_{\theta;(x,v)}^*)}\| \\ & = \|v_1^\dagger - \bar{v}_1\| \stackrel{\text{Eq. (8.69)}}{\leq} 10(\sqrt{\mathfrak{M}_2})^2 \cdot (\Delta_{\theta;(x,v)}^*)^2 \\ & \stackrel{\text{Eq. (8.34)}}{\leq} 10(\sqrt{\mathfrak{M}_2})^2 \cdot \frac{\theta}{\sqrt{\mathfrak{m}_2}} \cdot \Delta_{\theta;(x,v)}^* = \beta \cdot \sqrt{\mathfrak{M}_2} \cdot \theta \cdot \Delta_{\theta;(x,v)}^*. \quad \square \end{aligned}$$

We conclude with a bound on the total computational effort that is required to sample approximately from the uniform distribution on $\partial\mathcal{K}$ with error less than any fixed $\varepsilon > 0$.

THEOREM 8.2. *Fix $\varepsilon > 0$, let α, β be defined as in Lemma 6, and let $I(\varepsilon), \theta(\varepsilon)$ be defined as in equation (8.18). Let $\{X_i^{\theta(\varepsilon)}\}_{i \in \mathbb{N}}$ be the Markov chain defined in equation (8.6), with the oracle \star provided by Algorithm 3. Then the distribution $\Pi_{I(\varepsilon)+1}^{\theta(\varepsilon)}$ of $X_{I(\varepsilon)+1}^{\theta(\varepsilon)}$ satisfies*

$$(8.70) \quad W_d^{(1)}(\Pi_{I(\varepsilon)+1}^{\theta(\varepsilon)}, \Pi) \leq \varepsilon,$$

and the intersection oracle and normal vector oracle in Algorithm 3 are called no more than

$$(8.71) \quad N(\varepsilon) = \mathcal{O}\left(\left(\frac{\mathfrak{M}_2}{\mathfrak{m}_2}\right)^3 \times \frac{\log\left(\frac{2\pi}{\varepsilon\sqrt{\mathfrak{m}_2}}\right)}{\varepsilon \cdot \sqrt{\mathfrak{M}_2}}\right)$$

times.

PROOF. Inequality 8.70 is an application of Lemma 4 (since Assumption 2 is satisfied by Lemma 6). To prove Inequality 8.71, by Theorem 8.1,

$$\begin{aligned} N(\varepsilon) &\leq \frac{\log\left(\frac{\varepsilon\sqrt{\mathfrak{m}_2}}{2\pi}\right)}{\log\left(1 - \frac{\pi^2}{32} \frac{\mathfrak{m}_2}{\mathfrak{M}_2}\right)} \times \left[\left\lceil \frac{2\pi\sqrt{\mathfrak{M}_2}}{2\sqrt{\mathfrak{M}_2}\theta(\varepsilon)} \right\rceil + \mathcal{O}\left(\log\left(\frac{\mathfrak{M}_2}{\mathfrak{m}_2} \times \frac{\sqrt{\mathfrak{M}_2}}{\theta(\varepsilon)}\right)\right) \right] \\ &= \mathcal{O}\left(\frac{\log\left(\frac{\varepsilon\sqrt{\mathfrak{m}_2}}{2\pi}\right)}{\log\left(1 - \frac{\pi^2}{32} \frac{\mathfrak{m}_2}{\mathfrak{M}_2}\right)} \left[\frac{1}{\varepsilon \cdot \sqrt{\mathfrak{M}_2} \cdot \frac{\pi^2}{32} \frac{\mathfrak{m}_2}{\mathfrak{M}_2}} \left[1 + \frac{\pi}{2}\alpha + \left(\frac{\pi}{2}\right)^2 \beta \right] \right] \right) \\ &= \mathcal{O}\left(\frac{\log\left(\frac{\varepsilon\sqrt{\mathfrak{m}_2}}{2\pi}\right)}{\log\left(1 - \frac{\pi^2}{32} \frac{\mathfrak{m}_2}{\mathfrak{M}_2}\right)} \right. \\ &\quad \times \left. \left[\frac{1}{\varepsilon \cdot \sqrt{\mathfrak{M}_2} \cdot \frac{\pi^2}{32} \frac{\mathfrak{m}_2}{\mathfrak{M}_2}} \times \left[1 + \frac{\pi}{2} \cdot 5\pi \left(\frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}}\right)^2 + \left(\frac{\pi}{2}\right)^2 \cdot 10 \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}} \right] \right] \right) \\ &= \mathcal{O}\left(\frac{\log\left(\frac{\varepsilon\sqrt{\mathfrak{m}_2}}{2\pi}\right)}{\log\left(1 - \frac{\pi^2}{32} \frac{\mathfrak{m}_2}{\mathfrak{M}_2}\right)} \times \frac{1 + \frac{\pi}{2} \cdot 5\pi \left(\frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}}\right)^2 + \left(\frac{\pi}{2}\right)^2 \cdot 10 \frac{\sqrt{\mathfrak{M}_2}}{\sqrt{\mathfrak{m}_2}}}{\varepsilon \cdot \sqrt{\mathfrak{M}_2} \cdot \frac{\pi^2}{32} \frac{\mathfrak{m}_2}{\mathfrak{M}_2}}\right) \\ &= \mathcal{O}\left(\left(\frac{\mathfrak{M}_2}{\mathfrak{m}_2}\right)^3 \times \frac{\log\left(\frac{2\pi}{\varepsilon\sqrt{\mathfrak{m}_2}}\right)}{\varepsilon \cdot \sqrt{\mathfrak{M}_2}}\right), \end{aligned}$$

where the first inequality is due to inequality (8.22) and Lemma 6, and the remaining lines are obtained by the definitions of $\theta(\varepsilon)$, α and β . \square

REMARK 7. If the smoothness of the curvature of \mathcal{M} is uniformly bounded below (i.e., if the derivative of the curvature is uniformly bounded above), we can replace the ratio $\frac{\mathfrak{M}_2}{m_2}$ appearing in Lemmas 5 and 6 by a constant \mathfrak{M}_3 that depends only on a lower bound on the smoothness of the curvature, not on the curvature itself, so that we get $\alpha = 2\pi\mathfrak{M}_3$ and $\beta = \sqrt{\mathfrak{M}_3}$. The number of intersection and normal line oracle calls is then $\mathcal{O}^*((\frac{\mathfrak{M}_2}{m_2})^2 \cdot \mathfrak{M}_3 \cdot (\varepsilon\sqrt{\mathfrak{M}_2})^{-1})$, which is quadratic in the curvature ratio $\frac{\mathfrak{M}_2}{m_2}$.

9. Discussion and future work. We give informal discussions of three natural questions left open by this paper.

9.1. *Sampling from convex polytopes with the geodesic walk.* In this paper, we presented the geodesic walk, gave bounds on its mixing properties for general manifolds \mathcal{M} with bounded positive curvature, and analyzed a simple efficient implementation of the walk in the special case that $\mathcal{M} = \partial\mathcal{K}$ is the boundary of a convex set \mathcal{K} . However, it is clear that our bounds can be very poor for some natural manifolds, even in this special case. In this section, we discuss a natural open question: how strong is the assumption that $\partial\mathcal{K}$ has bounded positive curvature, and how can this assumption be weakened in practice?

We first note that the assumption of *nonnegative* curvature is very weak. Recall that the *Alexandrov curvature* is a generalization of the usual notion of sectional curvature to more general metric spaces, and that the boundary of a convex body in \mathbb{R}^d always has *nonnegative* Alexandrov curvature (see, e.g., [2] for a definition of Alexandrov curvature and survey of relevant results). However, the assumption that the curvature is *bounded* is much stronger: it is straightforward to check that the boundary of any polytope will have Alexandrov curvature that is not bounded away from either 0 or $+\infty$.

Fortunately, it is possible to obtain efficient samples from the boundaries of convex bodies that do not satisfy our assumptions by using appropriate pre- and post-processing steps. We now give an algorithm that allows us to replace the task of sampling from the boundary of a convex body $\partial\mathcal{K}$ with the task of sampling from a related “rounded” body $\partial\mathcal{K}'$ whose sectional curvature is bounded from above by $\mathcal{O}(d^2)$. The construction is based on essentially the same intuition as the pre- and post-processing described in Section 5 of [38] for sampling from the interior of a convex set, though the required analysis is slightly more delicate. We first require some simple definitions.

DEFINITION 3 (ε -thickening and projection). For fixed $\varepsilon > 0$ and convex body $\mathcal{K} \subset \mathbb{R}^d$, we define $\mathcal{K}_\varepsilon = \{x \in \mathbb{R}^d : \inf_{y \in \mathcal{K}} \|x - y\| \leq \varepsilon\}$ and $\mathcal{M}_\varepsilon = \partial\mathcal{K}_\varepsilon$.

We define the projection-like map $f_\varepsilon : \mathcal{M}_\varepsilon \mapsto \mathcal{M} \equiv \partial\mathcal{K}$ by

$$f_\varepsilon(x) = \arg \min_{y \in \mathcal{M}} \|x - y\|.$$

Algorithm 4 Simple rejection sampling algorithm

parameters: Convex set \mathcal{K} and parameters $\varepsilon > 0, N \in \mathbb{N}$.

output: Samples X_1, \dots, X_N from the uniform measure on $\mathcal{M} \equiv \partial\mathcal{K}$.

- 1: **for** $i = 1, 2, \dots, N$ **do**
 - 2: Sample $Z_i \sim \text{Unif}(\mathcal{M}_\varepsilon)$.
 - 3: Propose $Y_i = f_\varepsilon(Z_i)$.
 - 4: Accept the sample Y_i and set $X_i = Y_i$ with probability $H_\varepsilon(Z_i)$. Otherwise, reject the sample and go back to Step 2.
 - 5: **end for**
-

We define $H_\varepsilon(x)$ to be the magnitude of the determinant of the Hessian of f_ε at x , when it exists. We note the following.

LEMMA 7. *With notation as above,*

$$0 \leq H_\varepsilon(x) \leq 1$$

for all $x \in \mathcal{M}_\varepsilon$.

PROOF. This is an immediate consequence of the fact that f_ε is a contraction; see [42] for a proof of this fact. \square

As a consequence of this lemma, for all $\varepsilon > 0$ the following is a valid rejection-sampling algorithm for sampling from \mathcal{M} if you know how to sample from \mathcal{M}_ε .

The rejection rate of this algorithm is given by

$$(9.1) \quad \mathbb{P}[\text{reject}] = \frac{\text{Area}(\mathcal{M})}{\text{Area}(\mathcal{M}_\varepsilon)}.$$

By Crofton’s formula, for $\varepsilon = \frac{1}{20d}$ we have

$$(9.2) \quad \text{Area}(\mathcal{M}_\varepsilon) \leq \text{Area}(\mathcal{M}) + \frac{1}{2}.$$

Assuming without loss of generality that $\text{Area}(\mathcal{M}) \geq 1$ after appropriate rescaling and choosing $\varepsilon = \frac{1}{20d}$, we have by formulas (9.1) and (9.2) that the rejection probability of Algorithm 4 is uniformly bounded away from 1:

$$(9.3) \quad \mathbb{P}[\text{reject}] \leq \frac{2}{3}.$$

Thus, in order to be able to sample from the uniform distribution on \mathcal{M} , it is enough to be able to sample from the uniform distribution on \mathcal{M}_ε as long as $\varepsilon < \frac{1}{20d}$. We have gained something important in this replacement: while \mathcal{M} may have unbounded sectional curvature, the sectional curvature of \mathcal{M}_ε is at most ε^{-2} . To see this last fact, consider a point $x \in \mathcal{M}_\varepsilon$ and let $y \in \mathcal{M}$ satisfy $\|x - y\| = \varepsilon$

(such a nearest point exists by the definition of \mathcal{M}_ε). Then the ball $B_\varepsilon(y)$ of radius ε centered at y is entirely contained in the interior of \mathcal{M}_ε , and x is on the boundary of both $B_\varepsilon(y)$ and \mathcal{M}_ε . Therefore, the radius of curvature of \mathcal{M}_ε at x must be at least ε (recalling that \mathcal{M}_ε has nonnegative curvature because \mathcal{M} is the boundary of a convex body), and hence the sectional curvature at x must be at most ε^{-2} . Since this applies for every point $x \in \mathcal{M}_\varepsilon$, we conclude that the sectional curvature of \mathcal{M}_ε is at most ε^{-2} .

Inequality (9.3) allows us to conclude that, using Algorithm 4, we can always transform the problem of sampling uniformly from a manifold \mathcal{M} with the problem of sampling uniformly from a slightly perturbed manifold \mathcal{M}_ε with bounded sectional curvature $\sup_{x \in \mathcal{M}_\varepsilon} \sup_{u, v \in \mathcal{T}_x \mathcal{M}_\varepsilon} C_x(u, v) = \mathcal{O}(d^2)$.

Further discussion of pre- and post-processing is out of the scope of this paper. We leave two related questions for future work:

1. Is there another perturbation of the manifold that guarantees an even smaller upper bound on the sectional curvature?
2. Is there another perturbation of the manifold that guarantees a useful *lower* bound on the sectional curvature?

9.2. Sectional or Ricci curvature. Readers familiar with previous work on mixing for walks on manifolds with “positive curvature,” such as [31], may be surprised that our main conditions are stated in terms of the *sectional* curvature rather than the *Ricci* curvature. Roughly speaking, a manifold has positive sectional curvature if it is positively curved “in all directions,” while it has positive Ricci curvature under the weaker condition that it is positively curved “on average.” It is natural to ask: are the main results of the paper true if we replace our bound on the sectional curvature with the weaker analogous assumption on the Ricci curvature?

If we consider Algorithm 1 for very small integration time T , the answer is yes. Indeed, it is straightforward to check that, in the limit as T goes to 0, Algorithm 1 has the same contraction bound as the ball walk studied in Example 4 of [31]. However, choosing T to be very small will generally result in a very inefficient algorithm; it would be much more useful if we could obtain a similar result for T large.

For longer integration times, we do not know the answer to the question in any generality, but mention that some positive results are possible using existing (but more advanced) results from differential geometry. For example, we could find an “averaged” version of the contraction result in Theorem 6.1 if we could replace references to Rauch’s theorem to an analogue that assumed only a positive Ricci curvature, rather than the stronger assumption of positive sectional curvature. There has been substantial research on finding such analogues, under the name of “comparison geometry” (see the survey [41]). We do not summarize these results here, except to note that there exist some positive results that would apply to our algorithm, including those in [10].

9.3. *Faster mixing time bounds for Hamiltonian Monte Carlo.* Hamiltonian Monte Carlo (HMC), a Markov chain algorithm that is in many ways analogous to the geodesic walk, is popular in applications such as statistics and machine learning where one wishes to sample from a target distribution $\pi : \mathbb{R}^d \rightarrow \mathbb{R}^+$. The HMC Markov chain proceeds on \mathbb{R}^d in a similar way as the geodesic walk on a manifold \mathcal{M} , except that the geodesic trajectories are replaced by the trajectory of a particle moving according to the laws of classical mechanics in a potential well with potential $U = -\log(\pi)$. When π is a high-dimensional product measure, most HMC trajectories are closely approximated by a geodesic on one of the level sets of π , due to the concentration of measure phenomenon. In a similar vein, the paper [37] uses concentration of measure to approximate HMC with geodesic trajectories on a related manifold, called the Jacobi manifold.

Using a different approach, the paper [27] obtained the best current mixing time bound $\mathcal{O}^*\left(\left(\frac{M_2}{m_2}\right)^2\right)$ for HMC on strongly log-concave distributions π , when the eigenvalues of the Hessian matrix of π are uniformly bounded above and below by positive constants M_2 and m_2 , respectively. As each level set S of π has sectional curvature ratio $\frac{\mathfrak{M}_2(S)}{\mathfrak{m}_2(S)} \leq \frac{M_2}{m_2}$, where $\mathfrak{M}_2(S)$ and $\mathfrak{m}_2(S)$ are, respectively, the maximum and minimum values of the sectional curvature of S , the mixing time bound $\mathcal{O}^*\left(\left(\frac{M_2}{m_2}\right)^2\right)$ obtained in [27] for HMC on the distribution π is slower than the mixing time bound of $\mathcal{O}^*\left(\frac{\mathfrak{M}_2}{\mathfrak{m}_2}\right)$ we obtained in this paper for the geodesic walk on any of the level sets S of π .

Since HMC trajectories closely approximate geodesic paths, we conjecture that one should be able to obtain a stronger mixing time bound of $\mathcal{O}^*\left(\frac{M_2}{m_2}\right)$ for HMC that matches the mixing time bound we obtained in this paper for the geodesic walk on any of the level sets S of π . We leave as an open problem whether the methods used in this paper to bound the mixing time of the geodesic walk can be extended to obtain the conjectured mixing time bound of $\mathcal{O}^*\left(\frac{M_2}{m_2}\right)$ for HMC.

Acknowledgments. We are grateful to Yin Tat Lee and A. B. Dieker for helpful discussions. We also thank the anonymous reviewer for helpful comments and suggestions.

REFERENCES

- [1] ABRESCH, U. and MEYER, W. T. (1997). Injectivity radius estimates and sphere theorems. In *Comparison Geometry* (Berkeley, CA, 1993–94). *Math. Sci. Res. Inst. Publ.* **30** 1–47. Cambridge Univ. Press, Cambridge. [MR1452866](#)
- [2] ALEXANDROV, A. D. (2005). *Convex Polyhedra*. Springer, Berlin. [MR2127379](#)
- [3] BÉLISLE, C. J. P., ROMEIJN, H. E. and SMITH, R. L. (1993). Hit-and-run algorithms for generating multivariate distributions. *Math. Oper. Res.* **18** 255–266. [MR1250117](#)
- [4] BELKIN, M., NARAYANAN, H. and NIYOGI, P. (2006). Heat flow and a faster algorithm to compute the surface area of a convex body. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on* 47–56. IEEE, New York.
- [5] BONAHOON, F. and SÖZEN, Y. (2003). Variation of the Liouville measure of a hyperbolic surface. *Ergodic Theory Dynam. Systems* **23** 729–758.

- [6] BORNN, L., SHEPHARD, N. and SOLGI, R. (2015). Moment conditions and Bayesian non-parametrics. Preprint. Available at [arXiv:1507.08645](https://arxiv.org/abs/1507.08645).
- [7] CHEEGER, J. and EBIN, D. G. (2008). *Comparison Theorems in Riemannian Geometry* **365**. Am. Math. Soc., Providence.
- [8] CHOW, B. and KNOPF, D. (2004). *The Ricci Flow: An Introduction* **110**. American Mathematical Society, Providence.
- [9] COMETS, F., POPOV, S., SCHÜTZ, G. M. and VACHKOVSKAIA, M. (2009). Billiards in a general domain with random reflections. *Arch. Ration. Mech. Anal.* **191** 497–537.
- [10] DAI, X. and WEI, G. (1995). A comparison-estimate of Toponogov type for Ricci curvature. *Math. Ann.* **303** 297–306.
- [11] DIACONIS, P., HOLMES, S. and SHAHSHAHANI, M. (2013). Sampling from a manifold. In *Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton. Inst. Math. Stat. (IMS) Collect.* **10** 102–125. IMS, Beachwood, OH. [MR0298716](https://arxiv.org/abs/1308.4072)
- [12] DIEKER, A. B. and VEMPALA, S. S. (2015). Stochastic billiards for sampling from the boundary of a convex set. *Math. Oper. Res.* **40** 888–901.
- [13] DOBRUŠIN, R. L. (1970). Definition of a system of random variables by means of conditional distributions. *Teor. Veroyatnost. i Primenen.* **15** 469–497. [MR0298716](https://arxiv.org/abs/1708.07114)
- [14] DURMUS, A. and MOULINES, E. (2016). Sampling from strongly log-concave distributions with the unadjusted Langevin algorithm. Preprint. Available at [arXiv:1605.01559](https://arxiv.org/abs/1605.01559).
- [15] DURMUS, A. and MOULINES, E. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.* **27** 1551–1587. [MR3678479](https://arxiv.org/abs/1708.07114)
- [16] ESCHENBURG, J.-H. (1994). *Comparison Theorems in Riemannian Geometry*. Dipartimento di matematica, Univ. delli studi di Trento.
- [17] GUDMUNDSSON, S. (2008). *An Introduction to Riemannian Geometry*. Lund Univ.
- [18] HOFFMAN, M. D. and GELMAN, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623.
- [19] HOPF, H. and RINOW, W. (1931). Über den Begriff der vollständigen differentialgeometrischen Fläche. *Comment. Math. Helv.* **3** 209–225.
- [20] JOULIN, A. and OLLIVIER, Y. (2010). Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.* **38** 2418–2442. [MR2683634](https://arxiv.org/abs/1002.3610)
- [21] KLINGENBERG, W. P. (1995). *Riemannian Geometry*, 2nd ed. de Gruyter, Berlin.
- [22] LEBEAU, G. and MICHEL, L. (2010). Semi-classical analysis of a random walk on a manifold. *Ann. Probab.* **38** 277–315. [MR2599200](https://arxiv.org/abs/1002.3610)
- [23] LEE, J. M. (2006). *Riemannian Manifolds: An Introduction to Curvature* **176**. Springer, Berlin.
- [24] LEE, Y. T. and VEMPALA, S. S. (2017). Geodesic walks in polytopes. In *STOC’17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing* 927–940. ACM, New York. [MR3678240](https://arxiv.org/abs/1708.07114)
- [25] LOVÁSZ, L. and VEMPALA, S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures Algorithms* **30** 307–358.
- [26] MANGOUBI, O. Sampling the interior of a convex body from its boundary. *In preparation*.
- [27] MANGOUBI, O. and SMITH, A. (2017). Rapid mixing of hamiltonian Monte Carlo on strongly log-concave distributions. Preprint. Available at [arXiv:1708.07114](https://arxiv.org/abs/1708.07114).
- [28] MYERS, S. B. (1941). Riemannian manifolds with positive mean curvature. *Duke Math. J.* **8** 401–404. [MR0004518](https://arxiv.org/abs/1708.07114)
- [29] NARAYANAN, H. and NIYOGI, P. (2008). Sampling hypersurfaces through diffusion. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques* 535–548. Springer, Berlin.
- [30] OLLIVIER, Y. A visual introduction to Riemannian curvatures and some discrete generalizations. *Analysis and Geometry of Metric Measure Spaces: Lecture Notes of the 50th Séminaire de Mathématiques Supérieures (SMS), Montréal* 197–219.

- [31] OLLIVIER, Y. (2009). Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256** 810–864.
- [32] ÖZTIRELI, A. C., ALEXA, M. and GROSS, M. (2010). Spectral sampling of manifolds. *ACM Transactions on Graphics (TOG)* **29** 168.
- [33] PAULIN, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20**.
- [34] PETERSEN, P. (2006). *Riemannian Geometry*, 2nd ed. *Graduate Texts in Mathematics* **171**. Springer, New York. [MR2243772](#)
- [35] RAUCH, H. E. (1951). A contribution to differential geometry in the large. *Ann. of Math. (2)* **54** 38–55. [MR0042765](#)
- [36] RHEE, C-H., ZHOU, E. and QIU, P. (2014). An iterative algorithm for sampling from manifolds. In *Proceedings of the 2014 Winter Simulation Conference* 574–585. IEEE Press, New York.
- [37] SEILER, C., RUBINSTEIN-SALZEDO, S. and HOLMES, S. (2014). Positive curvature and Hamiltonian Monte Carlo. In *Advances in Neural Information Processing Systems* 586–594.
- [38] VEMPALA, S. (2005). Geometric random walks: A survey. *Combinatorial and Computational Geometry* **52** 2.
- [39] VILLANI, C. (2008). *Optimal Transport: Old and New* **338**. Springer, Berlin.
- [40] WALCZAK, P. G. (1994). Existence of smooth invariant measures for geodesic flows of foliations of Riemannian manifolds. *Proc. Amer. Math. Soc.* **120** 903–906.
- [41] ZHU, S. (1997). The comparison geometry of Ricci curvature. In *Comparison Geometry (Berkeley, CA, 1993–94)*. *Math. Sci. Res. Inst. Publ.* **30** 221–262. Cambridge Univ. Press, Cambridge. [MR1452876](#)
- [42] <http://math.stackexchange.com/users/8508/robert-israel>. Surface area of a convex set less than that of its enclosing sphere? Mathematics stack exchange. Available at <http://math.stackexchange.com/q/58556> (version: 2011-08-19).

ÉCOLE POLYTECHNIQUE FÉDÉRALE
DE LAUSANNE (EPFL)
IC IINFCOM THL3, STATION 14
1015 LAUSANNE
SWITZERLAND
E-MAIL: omangoubi@gmail.com

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF OTTAWA
585 KING EDWARD AVENUE
OTTAWA ON K1N 7N5
CANADA
E-MAIL: smith.aaron.matthew@gmail.com