Geodesic Slice Sampling on the Sphere

Michael Habeck[1]   MICHAEL.HABECK@UNI-JENA.DE

Mareike Hasenpflug[2]   MAREIKE.HASENPFLUG@UNI-PASSAU.DE

Shantanu Kodgirwar[1]   SHANTANU.KODGIRWAR@UNI-JENA.DE

Daniel Rudolf[2]   DANIEL.RUDOLF@UNI-PASSAU.DE

[1]*Friedrich Schiller University Jena, 07743 Jena, Germany*
[2]*Universität Passau, Innstraße 33, 94032 Passau, Germany*

Editor: Anthony Lee

Abstract

Probability measures on the sphere form an important class of statistical models and are used, for example, in modeling directional data or shapes. Due to their widespread use, but also as an algorithmic building block, efficient sampling of distributions on the sphere is highly desirable. We propose a shrinkage based and an idealized geodesic slice sampling Markov chain, designed to generate approximate samples from distributions on the sphere. In particular, the shrinkage-based version of the algorithm can be implemented such that it runs efficiently and has no tuning parameters. We verify reversibility and prove that under weak regularity conditions geodesic slice sampling is uniformly ergodic. Numerical experiments show that the proposed slice samplers achieve excellent mixing on challenging targets including distributions arising in rigid-registration problems and mixtures of von Mises-Fisher distributions. In these settings our approach outperforms standard samplers such as random-walk Metropolis-Hastings and Hamiltonian Monte Carlo.

Keywords: Markov Chain Monte Carlo, slice sampling, spherical distributions

1. Introduction

In recent years, with the advent of sampling methods based on Markov chains, Bayesian inference with posterior distributions on manifolds attracted considerable attention. In particular, various Markov chain algorithms for approximate sampling on different manifolds have been developed (see Byrne and Girolami, 2013; Lan et al., 2014; Goyal and Shetty, 2019; Lie et al., 2023; Beskos and Kamatani, 2022). Here we focus on the prototypical case of the sphere embedded in $\mathbb{R}^d$ with its inherent geometrical features as the underlying manifold. Following the slice sampling paradigm, we introduce and analyze an efficient way of approximate sampling of distributions on the sphere.

Let $\mathbb{S}^{d-1}$ be the $(d-1)$-dimensional Euclidean unit sphere in $\mathbb{R}^d$ and let σ_{d-1} denote the volume measure on $\mathbb{S}^{d-1}$. For $p\colon \mathbb{S}^{d-1} \to (0,\infty)$ satisfying

$$Z := \int_{\mathbb{S}^{d-1}} p(x)\sigma_{d-1}(\mathrm{d}x) \in (0,\infty), \tag{1}$$

we are interested in a target distribution π of the form

$$\pi(\mathrm{d}x) = \frac{1}{Z}\, p(x)\sigma_{d-1}(\mathrm{d}x), \tag{2}$$

such that p/Z is the probability density function of π relative to σ_{d-1}. In a Bayesian setting, π can be considered a posterior distribution determined by likelihood p and prior measure σ_{d-1}. Throughout the paper, we assume that p can be evaluated, which is a common minimal requirement in the Markov chain Monte Carlo (MCMC) literature. Our motivation for considering MCMC on the sphere is twofold:

1. *Posteriors naturally defined on the sphere require efficient sampling:* To provide some examples, sampling distributions on the sphere plays an important role in directional statistics and shape analysis (see Mardia and Jupp, 2000), and Bayesian inverse problems on $\mathbb{S}^2$ occur, e.g., in astrophysics or geophysics (see Marignier et al., 2023). Moreover, Bayesian density estimation (Holbrook et al., 2020) requires sampling of a posterior on an infinite-dimensional sphere, which usually is approximated by truncating the dimension, ending up with $\mathbb{S}^{d-1}$ for large d. Also the problem of 3D rigid registration can be reduced to sampling from a distribution on $\mathbb{S}^3$ (see Section 4.1).

2. *Spherical sampling can be used as tool within transforming Markov chains:* Lan et al. (2014) introduced a Hamiltonian Monte Carlo (HMC) scheme for sampling from spherical distributions that leads, after suitable transformations, to an efficient exploration of distributions in $\mathbb{R}^d$ constrained by inequalities. Moreover, recently Yang et al. (2024) demonstrated that MCMC algorithms on $\mathbb{S}^{d-1}$ can target distributions in $\mathbb{R}^d$ by mapping them to the sphere by means of the stereographic projection and then sampling from the transformed target on the sphere. This is further developed by Bell et al. (2025) with an adaptive MCMC strategy. The authors report that the "stereographic scheme" performs well in the presence of heavy-tailed distributions, as shown by both theoretical and empirical evidence. Bell et al. (2025) also indicate that the combination with slice sampling on the sphere, proposed here, is particularly effective and outperforms competing methods.

The advantage of the slice sampling paradigm is that it enables an automatic, location-dependent traversal of the space via density based regions, without requiring any laborious manual tuning of parameters to ensure efficient exploration of the state space. This usually results in an algorithm that is less sensitive to certain degenerations of the target distribution such as multimodality or anisotropy, see also Neal (2003); Murray et al. (2010). A transition mechanism for realizing a corresponding Markov chain works as follows: Given current state $x \in \mathbb{S}^{d-1}$, a superlevel set $L(t)$ of p is randomly determined by choosing a level $t \in (0, p(x))$. Then, the next Markov chain instance is specified by (mimicking) sampling of the normalized reference measure σ_{d-1} restricted to the level set. The latter step requires some care regarding its algorithmic design. The main contributions of our paper are:

- We introduce an ideal geodesic slice sampler with corresponding Markov kernel H for approximately simulating π. The kernel H is implemented by moving along a randomly chosen great circle through the current state. The resulting sampling on the great circle (understood as geodesic) intersected with the level set can be algorithmically realized by using an univariate acceptance/rejection approach.

- To increase the computational efficiency we propose a modification of H by employing a shrinkage procedure (cf. Murray et al., 2010; Neal, 2003) and call the resulting Markov kernel $\widetilde{H}$.

- We show that if p is lower semicontinuous, i.e., the strict superlevel sets of p are open, then both, H and $\widetilde{H}$ are well-defined and reversible with respect to (w.r.t.) the target distribution π.

- We prove uniform ergodicity for lower semicontinuous p that satisfy a boundedness condition, i.e., we provide total variation distance estimates of the nth marginal of a Markov chain with transition kernel H or $\widetilde{H}$ to π, see Theorem 6 and Theorem 16.

- We test our algorithms on challenging targets such as a posterior distribution arising in rigid registration and mixtures of von Mises-Fisher distributions. We observe that, in a variety of settings, our slice samplers outperform standard samplers such as random walk Metropolis Hastings (RWMH) and Hamiltonian Monte Carlo (HMC) on spherical distributions.

Let us comment on how our contributions fit into the literature. Lie et al. (2023) present two MCMC algorithms to sample distributions on the sphere using push forward kernels of a preconditioned Crank-Nicolson algorithm and of elliptical slice sampling, respectively. The crucial difference to our work is that Lie et al. (2023) assume that the target density is defined relative to an angular Gaussian distribution. Instead, here we consider target densities relative to the volume measure of the sphere. Observe that with increasing dimension both reference measures become "increasingly singular" to each other. Furthermore, Marignier et al. (2023) propose a proximal MCMC method to sample from posterior distributions of inverse imaging problems on $\mathbb{S}^2$. An infinite-dimensional setting of Bayesian density estimation is treated by Holbrook et al. (2020) by an HMC algorithm on a sphere. Other MCMC approaches have been developed on more general manifolds that usually cover $\mathbb{S}^{d-1}$ as a special case. This includes the HMC algorithm for manifolds embedded in the Euclidean space introduced by Byrne and Girolami (2013) which uses the geodesic flow, and the HMC algorithm for manifolds that are the fibre of a smooth map proposed by Brubaker et al. (2012). Furthermore, Goyal and Shetty (2019) investigate a Metropolis-Hastings-like geodesic walk on manifolds with non-negative curvature. Similarly, the Metropolis-Hastings algorithm of Zappa et al. (2018) based on projections from the tangent spaces can be applied to $\mathbb{S}^{d-1}$. Here it is also worth to mention that several geodesic walks exist that target the uniform distribution on subsets of the sphere. We refer to Section 2.3 for more details. Finally, note that recently there has been some theoretical progress in the investigation of convergence properties of slice sampling (see for example Natarovskii et al., 2021a,b; Łatuszyński and Rudolf, 2024; Hasenpflug et al., 2024) that very much influenced the presentation and proof arguments of our theoretical results.

The outline of our paper is as follows: We start by introducing notation and details regarding the geometry of the sphere. In Section 3, we formulate the ideal and shrinkage slice sampler in terms of the transition kernel and transition mechanism. Here we also prove the reversibility and uniform ergodicity statements. Next we illustrate the applicability of our approach in different scenarios by numerical experiments. We conclude with a discussion.

2. Preliminaries

In this section we provide some general notation, give a brief introduction to geometrical features of the Euclidean unit sphere and explain how these can be leveraged to "walk the sphere".

2.1 Setting and Notation

Let λ be the Lebesgue measure on $\mathbb{R}$, and let $\|\cdot\|$ be the Euclidean norm induced by the standard inner product $x^T y$, where $x, y \in \mathbb{R}^d$ for $d \in \mathbb{N}$. Moreover, define $\mathrm{Id}_d \in \mathbb{R}^{d \times d}$ to be the d-dimensional identity matrix. Throughout this paper we assume $d \geq 3$ and consider the Euclidean unit sphere

$$\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d \mid \|x\| = 1\}$$

equipped with its Borel-σ-algebra $\mathcal{B}(\mathbb{S}^{d-1})$. By σ_{d-1} we denote the canonical volume measure on $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}))$ that serves as reference measure for distributions, cf. (2). To assess the difference of distributions ν, μ on $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}))$ we use the total variation distance

$$d_{tv}(\mu, \nu) := \sup_{A \in \mathcal{B}(\mathbb{S}^{d-1})} |\mu(A) - \nu(A)|.$$

On a generic Borel space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ define the restriction of a measure ν to a set $A \in \mathcal{B}(\mathcal{X})$ by $\nu|_A$, that is $\nu|_A(B) = \nu(A \cap B)$ for all $B \in \mathcal{B}(\mathcal{X})$. Whenever appropriate, for a set $A \in \mathcal{B}(\mathcal{X})$ we denote by $\mathcal{U}_A$ the uniform distribution on A. For example, for $\mathcal{X} = \mathbb{R}$ with $A \in \mathcal{B}(\mathbb{R})$ the probability measure $\mathcal{U}_A$ refers to $\frac{1}{\lambda(A)} \lambda|_A$ if $\lambda(A) \in (0, \infty)$ and for $\mathcal{X} = \mathbb{S}^{d-1}$ the distribution $\mathcal{U}_{\mathbb{S}^{d-1}}$ refers to $\frac{1}{\omega_{d-1}} \sigma_{d-1}$, where $\omega_{d-1} := \sigma_{d-1}(\mathbb{S}^{d-1})$. Sometimes it is more convenient to work with random variables. Therefore, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a sufficiently rich probability space on which all random variables occurring in this paper are defined. If a random variable X has distribution ν, we write $X \sim \nu$.

2.2 The Geometry of the Sphere

To approximately sample from distributions on the sphere we rely on its (special) geometry. The key objects in this context are great circles and "equators". More formally regarding the latter, for given $x \in \mathbb{S}^{d-1}$ we call

$$\mathbb{S}_x^{d-2} := \{v \in \mathbb{S}^{d-1} \mid v^T x = 0\}$$

the great subsphere with pole x. This $(d-2)$-dimensional subsphere is the intersection of $\mathbb{S}^{d-1}$ and the $(d-1)$-dimensional hyperplane perpendicular to x, and in this sense can be thought of as a generalization of the equator. We also equip $\mathbb{S}_x^{d-2}$ with the corresponding volume measure, which we denote as μ_x. Since $\mathbb{S}_x^{d-2}$ is essentially "a tilted $\mathbb{S}^{d-2}$", we have $\mu_x(\mathbb{S}_x^{d-2}) = \omega_{d-2}$ for all $x \in \mathbb{S}^{d-1}$, cf. Appendix A. Accordingly $\frac{1}{\omega_{d-2}} \mu_x$ coincides with $\mathcal{U}_{\mathbb{S}_x^{d-2}}$. In our transition mechanism, sampling w.r.t. the latter distribution frequently appears, such that we provide a procedure for performing it in Appendix A.

The great circles of the sphere are the "largest" 1-dimensional subspheres spanning $\mathbb{S}^{d-1}$. Rigorously, for every pair (x, v), where $x \in \mathbb{S}^{d-1}$ and $v \in \mathbb{S}_x^{d-2}$, we define the great circle

$$\gamma_{(x,v)} : \mathbb{R} \to \mathbb{S}^{d-1}, \qquad \theta \mapsto \cos(\theta)x + \sin(\theta)v.$$

4

Intuitively, we can also think of $\gamma_{(x,v)}$ as the curve obtained by "moving from the point x in the direction of v with a constant velocity". This interpretation originates from Riemannian geometry and following its terminology we may use the terms great circle and geodesic interchangeably for the object $\gamma_{(x,v)}$. Note that all great circles passing through a point $x \in \mathbb{S}^{d-1}$ are of the from $\gamma_{(x,v)}$ for some $v \in \mathbb{S}^{d-2}_x$, i.e., the great circles through $x \in \mathbb{S}^{d-1}$ are parametrized by $\mathbb{S}^{d-2}_x$.

Observe that, due to the periodicity of sine and cosine, all great circles are 2π-periodic. This is exploited when we incorporate great circles into our slice sampling approach.

We continue by illuminating an interaction between the geometric and the measure theoretic structure of the sphere with the help of the map

$$T_\theta : \mathcal{S}\mathbb{S}^{d-1} \to \mathcal{S}\mathbb{S}^{d-1}, \qquad (x,v) \mapsto \big(\cos(\theta)x + \sin(\theta)v, \ \sin(\theta)x - \cos(\theta)v \big)$$

for $\theta \in \mathbb{R}$, where for brevity we define $\mathcal{S}\mathbb{S}^{d-1} := \bigcup_{x \in \mathbb{S}^{d-1}} \big(\{x\} \times \mathbb{S}^{d-2}_x \big)$. If we interpret for a pair $(x,v) \in \mathcal{S}\mathbb{S}^{d-1}$ the first component as position on the sphere and the second component as "direction of view", then applying the map T_θ to (x,v) can be thought of as "following the great circle $\gamma_{(x,v)}$ from x in direction v for a length of θ and then doing a U-turn back on the spot". This intuition is underpinned by the following lemma, proven in Appendix B, which essentially tells us that following the great circle corresponding to the pair $T_\theta(x,v)$ is the same as following the original great circle $\gamma_{(x,v)}$ in reverse direction with an offset of θ.

Lemma 1 *Let $x \in \mathbb{S}^{d-1}$, $v \in \mathbb{S}^{d-2}_x$. For all $\theta, r \in \mathbb{R}$ we have $\gamma_{(x,v)}(\theta - r) = \gamma_{T_\theta(x,v)}(r)$. In particular, this implies $\gamma_{T_\theta(x,v)}(\theta) = x$.*

Moreover, central to our proof techniques is the observation that the measure on $\mathcal{S}\mathbb{S}^{d-1}$ that "zips up" the volume measures μ_x on the single fibers of $\mathcal{S}\mathbb{S}^{d-1}$ through the volume measure σ_{d-1} on $\mathbb{S}^{d-1}$ is invariant under T_θ.

Lemma 2 *Let $F : \mathcal{S}\mathbb{S}^{d-1} \to \mathbb{R}$ be a function such that the left-hand side of the equation below exists. Then, for all $\theta \in \mathbb{R}$ we have*

$$\int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-2}_x} F(x,v)\mu_x(\mathrm{d}v)\sigma_{d-1}(\mathrm{d}x) = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-2}_x} F\left(T_\theta(x,v)\right) \mu_x(\mathrm{d}v)\sigma_{d-1}(\mathrm{d}x). \qquad (3)$$

For the convenience of the reader we provide the proof of the former lemma in Appendix B. For identity (3) in a setting of more general manifolds we refer to (Goyal and Shetty, 2019, Proof of Theorem 27).

2.3 Geodesic Walk on the Sphere

Based on great circles we may "walk the sphere" starting at some state $x \in \mathbb{S}^{d-1}$ by iteratively performing the following:

1. Choose a great circle γ through the given point x randomly.

2. Choose the next point on the great circle γ randomly.

For the first step in the outlined transition we can use the fact that great circles through $x \in \mathbb{S}^{d-1}$ are parametrized by $\mathbb{S}^{d-2}_x$. Namely, we realize "a random great circle" as $\gamma_{(x,v)}$,

where v is a sample from $\mathcal{U}_{\mathbb{S}_x^{d-2}}$, see Algorithm 5 in Appendix A for the simulation of this probability measure. The distribution of the randomly chosen point on the great circle in step 2 significantly influences the stationary distribution of the random walk on the sphere. In this section, we consider it to be independent of x, which causes the stationary distribution to be the uniform one. In Section 3, we explain how the second step can be modified so as to obtain a Markov chain with a desired stationary probability measure.

Using a distribution τ on $[0, 2\pi)$ to sample on the great circle, we obtain the following Markov kernel

$$K_\tau(x, A) := \int_{\mathbb{S}_x^{d-2}} \int_{[0,2\pi)} \mathbb{1}_A \left(\gamma_{(x,v)}(\theta) \right) \tau(\mathrm{d}\theta) \, \mathcal{U}_{\mathbb{S}_x^{d-2}}(\mathrm{d}v)$$

with $x \in \mathbb{S}^{d-1}, A \in \mathcal{B}(\mathbb{S}^{d-1})$. The corresponding transition mechanism is described in Algorithm 1, where τ can be interpreted as the distribution of the step-size.

Algorithm 1 Geodesic random walk on the sphere.

---

input: current state $x \in \mathbb{S}^{d-1}$
output: next state x'

  1: Draw $V \sim \mathcal{U}_{\mathbb{S}_x^{d-2}}$, call the result v.   {Perform Algorithm 5 with input x.}
  2: Draw $\Theta \sim \tau$, call the result θ.
  3: Set $x' = \cos(\theta)x + \sin(\theta)v$.

By Lemma 2 and Lemma 1 we have for any $A, B \in \mathcal{B}(\mathbb{S}^{d-1})$ that

$$\int_B K_\tau(x, A) \, \sigma_{d-1}(\mathrm{d}x) = \int_{[0,2\pi)} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}_x^{d-2}} \mathbb{1}_B(x) \mathbb{1}_A\big(\gamma_{(x,v)}(\theta)\big) \, \frac{\mu_x(\mathrm{d}v)}{\omega_{d-2}} \, \sigma_{d-1}(\mathrm{d}x) \, \tau(\mathrm{d}\theta)$$

$$= \int_{[0,2\pi)} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}_x^{d-2}} \mathbb{1}_B\big(\gamma_{(x,v)}(\theta)\big) \mathbb{1}_A(x) \, \frac{\mu_x(\mathrm{d}v)}{\omega_{d-2}} \, \sigma_{d-1}(\mathrm{d}x) \, \tau(\mathrm{d}\theta) = \int_A K_\tau(x, B) \, \sigma_{d-1}(\mathrm{d}x),$$

which yields that K_τ is reversible w.r.t. $\mathcal{U}_{\mathbb{S}^{d-1}}$, see also Lee and Vempala (2018) or Goyal and Shetty (2019, Theorem 27).

Note that, depending on how τ is chosen, K_τ coincides with different random walks that have been discussed in the literature. We add some remarks related to this point:

Related work. If the step-size is chosen to be constant, that is, $\tau = \delta_\varepsilon$ for δ_ε being the Dirac measure on $\mathbb{R}$ at $\varepsilon \in (0, 2\pi)$, then K_{δ_ε} coincides with the geodesic walk from Mangoubi and Smith (2018) on the sphere. They provide dimension independent mixing time results for this walk on manifolds with bounded positive sectional curvature.

For targeting the uniform distribution on the sphere, the geodesic random walk introduced by Goyal and Shetty (2019, Algorithm 1, for their $K = \mathbb{S}^{d-1}$) corresponds to K_τ with τ being the distribution of εR, where R is a chi-distributed random variable with $d-1$ degrees of freedom and fixed $\varepsilon > 0$. For manifolds with non-negative sectional curvature and bounded Riemannian curvature tensor, Goyal and Shetty (2019) provide mixing time results if their unfiltered walk is modified to target the uniform distribution of a strongly geodesically convex subset of the ambient manifold.

Moreover, the kernel K_{δ_ε} is related to the retraction-based random walk introduced by Schwarz et al. (2025), where the geodesics are replaced by retractions, i.e., second order

approximations of the geodesics. Schwarz et al. (2025) show that for $\varepsilon \to 0$ this algorithm can be used to sample paths of the Brownian motion on a general manifold. Note, however, that this causes the stationary distribution of the resulting Markov chain to deviate from the uniform distribution.

General transitions according to K_τ can also be interpreted as the repeated action of random rotations. The d-dimensional rotation matrix

$$G(y, z, \theta) = \operatorname{Id}_d + (\cos(\theta) - 1)(yy^T + zz^T) + \sin(\theta)(yz^T - zy^T) \tag{4}$$

is the Givens rotation acting in the plane spanned by two orthogonal directions $y, z \in \mathbb{S}^{d-1}$ where θ is the rotation angle (see Givens, 1958). In this view, a transition of K_τ to the next state X' is achieved by drawing an axis $V \sim \mathcal{U}_{\mathbb{S}_x^{d-2}}$ and an angle $\Theta \sim \tau$ to form a random Givens rotation that is applied to the current state x, i.e., $X' = G(V, x, \Theta)\, x$.

A suitable reformulation of the latter can be viewed as a generalization of Kac's random walk on the sphere (see e.g., Kac, 1956; Pillai and Smith, 2017). To perform a transition of Kac's walk, draw $1 \leq I < J \leq d$ and $\Theta \sim \mathcal{U}_{(0,2\pi)}$ randomly to generate the next state X' by rotating the current state x, i.e., $X' = G(e_I, e_J, \Theta)\, x$, where $\{e_i\}_{i=1}^d$ is the standard basis of $\mathbb{R}^d$. Kac's walk also approximately simulates the uniform distribution, but chooses the plane of rotation from a discrete set, whereas in the generalized version the plane of rotation changes continuously and always contains the current state x. Pillai and Smith (2017) show optimal mixing time results for Kac's walk.

3. Geodesic Slice Sampling in $\mathbb{S}^{d-1}$

Following the slice sampling paradigm, we construct Markov chains for approximate sampling of π that rely on exploring the sphere along suitable 1-dimensional objects, the great circles. For this purpose let

$$L(t) := \{x \in \mathbb{S}^{d-1} \mid p(x) > t\}, \qquad t \in (0, \infty),$$

be the superlevel set of p. Recall that p specifies π, cf. (2).

The rough idea is to choose a great circle $\gamma_{(x,v)}$ and a level t randomly and then, from the intersection of the corresponding superlevel set and the great circle, draw the next state suitably. For $x \in \mathbb{S}^{d-1}$, $v \in \mathbb{S}_x^{d-2}$ and $t \in (0, \infty)$ we require

$$L(x, v, t) := \{\theta \in [0, 2\pi) \mid p(\gamma_{(x,v)}(\theta)) > t\},$$

called geodesic level set, which contains all points on the great circle $\gamma_{(x,v)}$ with the function value of p being greater than t. Naturally, this implies that level sets and geodesic level sets are linked via the identity

$$\mathbb{1}_{L(t)}\big(\gamma_{(x,v)}(\theta)\big) = \mathbb{1}_{L(x,v,t)}(\theta). \tag{5}$$

Exploiting lower semicontinuity of p we are able to identify a regime, where the geodesic level sets have strictly positive Lebesgue measure.

Lemma 3 *For p being lower semicontinuous, any $x \in \mathbb{S}^{d-1}$, $t \in (0, p(x))$ and $v \in \mathbb{S}_x^{d-2}$ we have $\lambda(L(x, v, t)) > 0$. Moreover, the essential supremum norm $\|p\|_\infty$ of p w.r.t. σ_{d-1} coincides with $\sup_{x \in \mathbb{S}^{d-1}} p(x)$.*

Proof By the lower semicontinuity $p^{-1}((t, \infty)) = L(t)$ is open, and by the fact that $x \in L(t)$, it is also non-empty. Moreover, since $\gamma_{(x,v)}$ is continuous, we have that $\gamma_{(x,v)}^{-1}(L(t))$ is again open and non-empty. Together with the 2π-periodicity of $\gamma_{(x,v)}$, this implies that $L(x, v, t) = \gamma_{(x,v)}^{-1}(L(t)) \cap [0, 2\pi)$ has non-empty interior, such that $\lambda(L(x, v, t)) > 0$. The final assertion follows by observing that $L(\|p\|_\infty)$ is open (by the lower semicontinuity) and by the definition of $\|p\|_\infty$ satisfies $\sigma_{d-1}(L(\|p\|_\infty)) = 0$. Consequently, we have $L(\|p\|_\infty) = \emptyset$, such that $\|p\|_\infty = \sup_{x \in \mathbb{S}^{d-1}} p(x)$. ∎

In the following we always assume that p is lower semicontinuous, which implies by the previous lemma the well-definedness of the slice sampling schemes that we introduce now.

3.1 Ideal Geodesic Slice Sampling

We start by presenting an acceptance/rejection sampling based version of the geodesic slice sampler. For $x \in \mathbb{S}^{d-1}$, $t \in (0, p(x))$ and $A \in \mathcal{B}(\mathbb{S}^{d-1})$ let

$$H_t(x, A) := \int_{\mathbb{S}_x^{d-2}} \int_{L(x,v,t)} \mathbb{1}_A(\gamma_{(x,v)}(\theta)) \frac{\mathrm{d}\theta}{\lambda(L(x, v, t))} \frac{\mu_x(\mathrm{d}v)}{\omega_{d-2}}$$

$$= \int_{\mathbb{S}_x^{d-2}} \int_{L(x,v,t)} \mathbb{1}_A(\gamma_{(x,v)}(\theta)) \, \mathcal{U}_{L(x,v,t)}(\mathrm{d}\theta) \, \mathcal{U}_{\mathbb{S}_x^{d-2}}(\mathrm{d}v).$$

The kernel $H_t(x, \cdot)$ can be simulated by first choosing a random great circle $\gamma_{(x,v)}$, and then, by sampling a point from the uniform distribution on $L(x, v, t)$, the intersection of the great circle and the level set $L(t)$. The *ideal geodesic slice sampler* is given by the algorithm that implements a transition corresponding to the Markov kernel

$$H(x, A) := \frac{1}{p(x)} \int_0^{p(x)} H_t(x, A) \, \mathrm{d}t = \int_0^{p(x)} H_t(x, A) \, \mathcal{U}_{(0,p(x))}(\mathrm{d}t), \qquad (6)$$

where $x \in \mathbb{S}^{d-1}, A \in \mathcal{B}(\mathbb{S}^{d-1})$. In words, a level $t \in (0, p(x))$ is chosen uniformly distributed before $H_t(x, \cdot)$ is performed. A single transition of H is described in Algorithm 2 and a graphical illustration on $\mathbb{S}^2$ is provided in Figure 1.

Algorithm 2 Ideal geodesic slice sampler.

---

input: current state $x \in \mathbb{S}^{d-1}$
output: next state x'
 1: Draw $V \sim \mathcal{U}_{\mathbb{S}_x^{d-2}}$, call the result v.   {Perform Algorithm 5 with input x.}
 2: Draw $T \sim \mathcal{U}_{(0,p(x))}$, call the result t.
 3: **repeat**
 4:     Draw $\Theta \sim \mathcal{U}_{(0,2\pi)}$, call the result θ.
 5:     Set $x' = \cos(\theta)x + \sin(\theta)v$.
 6: **until** $p(x') > t$.

Remark 4 *The geodesic slice sampler can be viewed as an instance of the procedure described in Section 2.3 where the second step uses a distribution on the great circle that depends on the initial point x and the chosen great circle γ.*
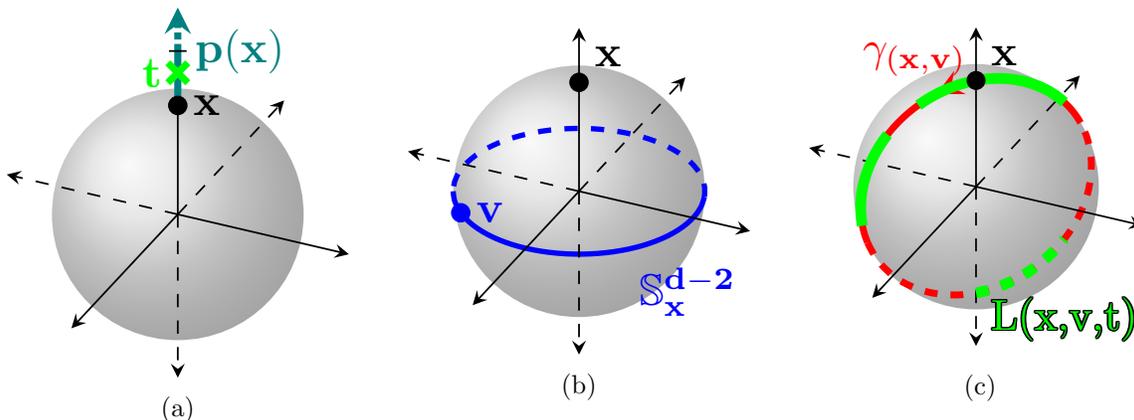
Figure 1: Transition mechanism of the ideal geodesic slice sampler. (a) Sample a random level t at the current point x. (b) Sample a random point v on $\mathbb{S}_x^{d-2}$. (c) Sample uniformly from the geodesic level set $L(x, v, t)$.

We establish the correctness of the ideal geodesic slice sampler by arguing for reversibility w.r.t. π and by proving a quantitative convergence guarantee. The proof of the following result is provided in Appendix C.

Proposition 5 *For π defined as in (2) with lower semicontinuous p, the kernel H is reversible w.r.t. π.*

Now we present our main result for the ideal geodesic slice sampler: Imposing a boundedness assumption on p, we provide an explicit convergence rate for the convergence to the stationary distribution in the total variation distance.

Theorem 6 *For p being lower semicontinuous with $\|p\|_\infty < \infty$ we have*

$$\sup_{x \in \mathbb{S}^{d-1}} d_{tv}\left(H^n(x, \cdot), \pi\right) \leq \left(1 - \frac{\sup_{t>0}\left[t \cdot \mathcal{U}_{\mathbb{S}^{d-1}}(L(t))\right]}{\sqrt{2\pi}\sqrt{d-1}\,\|p\|_\infty}\right)^n, \qquad \forall n \in \mathbb{N}, \qquad (7)$$

where $\sup_{t>0}\left[t \cdot \mathcal{U}_{\mathbb{S}^{d-1}}(L(t))\right] \in (0, \|p\|_\infty]$.

Let us comment on Theorem 6.

Remark 7 *The right-hand side of (7) depends on p through*

$$r_p := \sup_{t>0}\left[t \cdot \mathcal{U}_{\mathbb{S}^{d-1}}(L(t))\right] / \|p\|_\infty .$$

Observe that for lower semicontinuous p the size of $r_p \in (0, 1]$ solely depends on the decreasing level set function $t \mapsto \mathcal{U}_{\mathbb{S}^{d-1}}(L(t))$. This function plays an even stronger role in controlling the convergence of ideal slice sampling. It is known to be the only relevant quantity for geometric ergodicity and spectral gap estimates, cf. Roberts and Rosenthal (1999); Natarovskii et al. (2021b).

Remark 8 *The statement of the theorem contains the fact that for $\|p\|_\infty < \infty$ we have $r_p > 0$, such that the right-hand side of (7) converges exponentially to zero. We provide an example where r_p is computed explicitly. For $\delta \in (0,1)$, $S \subseteq \mathbb{S}^{d-1}$ open and $\mathcal{U}_{\mathbb{S}^{d-1}}(S) = \gamma \in (0,1]$ consider $p(x) = \mathbb{1}_S(x) + \delta \cdot \mathbb{1}_{\mathbb{S}^{d-1}\setminus S}(x)$. Note that for the corresponding level sets we have $\mathcal{U}_{\mathbb{S}^{d-1}}(L(t)) = \mathbb{1}_{(0,\delta)}(t) + \gamma \cdot \mathbb{1}_{[\delta,1)}(t)$ which implies $r_p = \max\{\gamma, \delta\}$.*

Remark 9 *We can interpret the ideal geodesic slice sampler as hit-and-run on the sphere. The hit-and-run algorithm has been introduced for approximate sampling of distributions on $\mathbb{R}^d$ (see Bélisle et al., 1993). A transition of the algorithm that starts in $x \in \mathbb{R}^d$ is performed in the following way: A direction v is chosen uniformly at random in the unit ball centered around x to construct a straight line ℓ through x in direction v. Then the next point is chosen w.r.t. the distribution of interest restricted to the line ℓ. Since on the sphere straight lines correspond to great circles, ideal geodesic slice sampling can be considered as hit-and-run where simulating the target distribution restricted to the great circle is performed by slice sampling. Investigations on the performance of the hit-and-run algorithm in $\mathbb{R}^d$ w.r.t. the dimension attracted a lot of attention, cf. Lovász and Vempala (2007). Regarding this aspect note that for $\varepsilon \in (0,1)$ we require $n \geq \sqrt{2(d-1)\pi}\, r_p^{-1} \log(\varepsilon^{-1})$ for achieving $\sup_{x \in \mathbb{S}^{d-1}} d_{tv}(H^n(x,\cdot),\pi) < \varepsilon$, which illuminates a moderate dependence on d.*

In the proof of Theorem 6 we apply Meyn and Tweedie (2009, Theorem 16.2.4). For the convenience of the reader we provide a reformulation of the relevant parts of this result.

Lemma 10 (Meyn and Tweedie, 2009, Theorem 16.2.4) *Let $P : \mathbb{S}^{d-1} \times \mathcal{B}(\mathbb{S}^{d-1}) \to [0,1]$ be a Markov kernel with stationary probability measure η. Assume there exists a non-zero measure ν on $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}))$ such that $P(x,A) \geq \nu(A)$ for all $x \in \mathbb{S}^{d-1}$ and all $A \in \mathcal{B}(\mathbb{S}^{d-1})$. Then*

$$\sup_{x \in \mathbb{S}^{d-1}} d_{tv}(P^n(x,\cdot),\eta) \leq (1 - \nu(\mathbb{S}^{d-1}))^n, \qquad \forall n \in \mathbb{N}.$$

We add one more auxiliary result which is proven in Appendix D.

Lemma 11 *For $x \in \mathbb{S}^{d-1}$ and $A \in \mathcal{B}(\mathbb{S}^{d-1})$ we have*

$$\int_0^\pi \int_{\mathbb{S}_x^{d-2}} \mathbb{1}_A(\gamma_{(x,v)}(\theta))\, \mathcal{U}_{\mathbb{S}_x^{d-2}}(\mathrm{d}v)\, \mathrm{d}\theta \geq \frac{\sqrt{2\pi}}{\sqrt{d-1}}\, \mathcal{U}_{\mathbb{S}^{d-1}}(A).$$

Now we turn to proof of the theorem:
Proof *(Proof of Theorem 6.)* Let $x \in \mathbb{S}^{d-1}$ and $s \in (0, \|p\|_\infty)$. Observe that for all $v \in \mathbb{S}_x^{d-2}$ and $\theta \in [0,2\pi)$ with $\gamma_{(x,v)}(\theta) \in L(s)$ holds

$$\frac{1}{p(x)} \int_0^{p(x)} \mathbb{1}_{[0,p(\gamma_{(x,v)}(\theta)))}(t)\, \mathrm{d}t = \min\left\{1, \frac{p(\gamma_{(x,v)}(\theta))}{p(x)}\right\} \geq \frac{s}{\|p\|_\infty}. \tag{8}$$

Observe that $L(x, v, t) \subseteq [0, 2\pi)$ such that $\lambda(L(x, v, t)) \leq 2\pi$ for all $v \in \mathbb{S}_x^{d-2}$ and all $t \in (0, \infty)$. Then, using Lemma 11 this implies

$$
\begin{aligned}
H(x, A) &= \frac{1}{p(x)} \int_0^{p(x)} \int_{\mathbb{S}_x^{d-2}} \frac{1}{\lambda(L(x, v, t))} \int_{L(x,v,t)} \mathbb{1}_A\big(\gamma_{(x,v)}(\theta)\big) \, d\theta \, \mathcal{U}_{\mathbb{S}_x^{d-2}}(dv) \, dt \\
&\geq \frac{1}{p(x)} \int_0^{p(x)} \int_{\mathbb{S}_x^{d-2}} \frac{1}{\lambda(L(x, v, t))} \int_{L(x,v,t)} \mathbb{1}_{A \cap L(s)}\big(\gamma_{(x,v)}(\theta)\big) \, d\theta \, \mathcal{U}_{\mathbb{S}_x^{d-2}}(dv) \, dt \\
&\geq \frac{1}{2\pi} \int_{\mathbb{S}_x^{d-2}} \int_0^{2\pi} \mathbb{1}_{A \cap L(s)}\big(\gamma_{(x,v)}(\theta)\big) \frac{1}{p(x)} \int_0^{p(x)} \mathbb{1}_{[0, p(\gamma_{(x,v)}(\theta)))}(t) \, dt \, d\theta \, \mathcal{U}_{\mathbb{S}_x^{d-2}}(dv) \\
&\underset{(8)}{\geq} \frac{s}{2\pi \|p\|_\infty} \int_0^\pi \int_{\mathbb{S}_x^{d-2}} \mathbb{1}_{A \cap L(s)}\big(\gamma_{(x,v)}(\theta)\big) \, \mathcal{U}_{\mathbb{S}_x^{d-2}}(dv) \, d\theta \geq \frac{s \, \mathcal{U}_{\mathbb{S}^{d-1}}(A \cap L(s))}{\sqrt{2\pi}\sqrt{d-1} \, \|p\|_\infty},
\end{aligned}
$$

for all $A \in \mathcal{B}(\mathbb{S}^{d-1})$. By Lemma 10 this implies

$$
\sup_{x \in \mathbb{S}^{d-1}} d_{tv}\left(H^n(x, \cdot), \pi\right) \leq \left(1 - \frac{s \, \mathcal{U}_{\mathbb{S}^{d-1}}(L(s))}{\sqrt{2\pi}\,\sqrt{d-1}\,\|p\|_\infty}\right)^n
$$

for any $s \in (0, \|p\|_\infty)$. For $s \geq \|p\|_\infty$ the statement holds trivially. Therefore taking the infimum over $s > 0$ on the right hand-side gives (7). By the lower semicontinuity of p we have that $L(s)$ is open and non-empty, such that $\mathcal{U}_{\mathbb{S}^{d-1}}(L(s)) \in (0, 1]$ and consequently $0 < s \, \mathcal{U}_{\mathbb{S}^{d-1}}(L(s)) < \|p\|_\infty$ for all $s \in (0, \|p\|_\infty)$, which provides the final claim. ∎

3.2 Geodesic Shrinkage Slice Sampling

Now we modify the ideal geodesic slice sampler by replacing the acceptance/rejection step of Algorithm 2 by a shrinkage procedure that has been also used in elliptical slice sampling (see Murray et al., 2010) and originates in Neal's bracketing procedure on the interval $[0, 2\pi)$ (see Neal, 2003). We pick candidates from subsets of $[0, 2\pi)$ that shrink until an accepted sample is generated. Intuitively, this strategy reduces the number of rejections per iteration, because candidates will be drawn from nested intervals that contain a neighborhood of the current state. Note that this approach can still be viewed as a realization of the procedure discussed in Section 2.3 in the sense of Remark 4.

In Algorithm 3 we provide the shrinkage procedure as subroutine that is called as $shrink(x, v, t)$ with input $x \in \mathbb{S}^{d-1}$, $v \in \mathbb{S}_x^{d-2}$ and level $t \in (0, p(x))$. Roughly, it generates a point from a specified geodesic level set $L(x, v, t)$ by drawing points uniformly from a segment of (one winding) of the great circle $\gamma_{(x,v)}$ until we hit $L(x, v, t)$, while the rejected points are used to successively shrink the segment of the great circle. We work with this procedure as a black box by encapsulating it in the distribution of its output. Considering the subroutine of Algorithm 3 as random variable yields the following definition.

Definition 12 *For lower semi-continuous p, for $x \in \mathbb{S}^{d-1}$, $v \in \mathbb{S}_{d-2}^x$ and $t \in (0, p(x))$ we set*

$$
Q_{x,v,t}(A) := \mathbb{P}\big((shrink(x, v, t) \mod 2\pi) \in A\big), \qquad \forall A \in \mathcal{B}([0, 2\pi)),
$$

where $shrink(x, v, t)$ is determined by Algorithm 3.

11

Algorithm 3 Shrinkage procedure, called as shrink(x, v, t).

---

input: current state $x \in \mathbb{S}^{d-1}$, direction $v \in \mathbb{S}^{d-2}_x$, level $t \in (0, p(x))$
output: parameter $\theta \in L(x, v, t)$
  1: Draw $\Theta \sim \mathcal{U}_{(0, 2\pi)}$, call the result θ.
  2: Set $\theta_{\min} = \theta - 2\pi$ and set $\theta_{\max} = \theta$.
  3: **while** $p\left(\cos(\theta)x + \sin(\theta)v\right) \leq t$ **do**
  4:     **if** $\theta < 0$ **then**
  5:         Set $\theta_{\min} = \theta$.
  6:     **else**
  7:         Set $\theta_{\max} = \theta$.
  8:     **end if**
  9:     Draw $\Xi \sim \mathcal{U}_{(\theta_{\min}, \theta_{\max})}$, call the result θ.
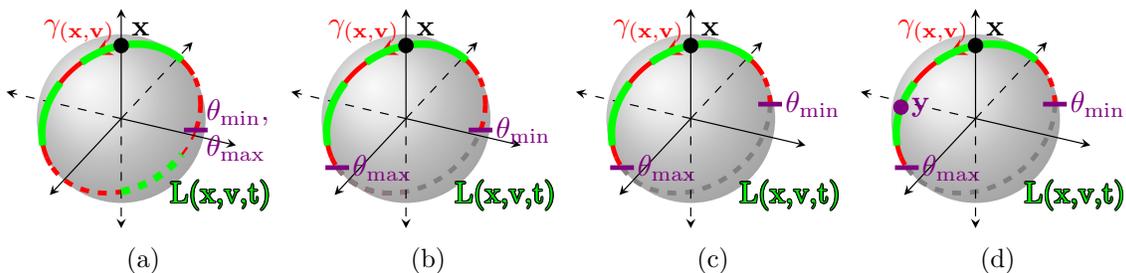10: **end while**

Figure 2: Shrinkage procedure of the geodesic shrinkage slice sampler. (*a*) First proposal is rejected. (*b*) Second proposal is rejected and becomes new right bound of the proposal interval. (*c*) Third proposal is rejected and becomes new left bound of the proposal interval. (*d*) Fourth proposal lies in geodesic level set $L(x, v, t)$ and is accepted.

Remark 13 *The distribution $Q_{x,v,t}$ coincides with the kernel of the shrinkage procedure defined by Hasenpflug et al. (2024, Algorithm 2.2 with $\theta_{in} = 0$, $S = L(x, v, t)$). For details we refer to Appendix E. The lower semicontinuity of p ensures that $Q_{x,v,t}$ is well-defined, cf. Hasenpflug et al. (2024). Intuitively, the openness of the superlevel sets guarantees that x lies in a subinterval of the image of the geodesic level set $\gamma_{(x.v)}\left(L(x, v, t)\right)$ for all $v \in \mathbb{S}^{d-2}_x$ and all $t \in (0, p(x))$. Hence, while shrinking the sampling region on $\gamma_{(x,v)}([0, 2\pi))$, the probability to hit the level set for a given level $t \in (0, p(x))$ is strictly positive.*

A single transition of the *geodesic shrinkage slice sampler* is presented in Algorithm 4 and for an illustration of the shrinkage scheme on a great circle on $\mathbb{S}^2$ we refer to Figure 2. To define the corresponding kernel $\widetilde{H}$, we need the following auxiliary level set kernels $\widetilde{H}_t$. Those correspond to drawing a random great circle through the starting point x, and then running the shrinkage procedure to generate a point from the intersection of the random great circle and the level set $L(t)$. That is, for $x \in \mathbb{S}^{d-1}$, $t \in (0, p(x))$ and $A \in \mathcal{B}(\mathbb{S}^{d-1})$ we

---

Algorithm 4 Geodesic shrinkage slice sampler.

---

input: current state $x \in \mathbb{S}^{d-1}$
output: next state x'

  1: Draw $V \sim \mathcal{U}_{\mathbb{S}_x^{d-2}}$, call the result v.   {Perform Algorithm 5 with input x.}
  2: Draw $T \sim \mathcal{U}_{(0,p(x))}$, call the result t.
  3: Draw $\Theta \sim Q_{x,v,t}$, call the result θ.   {Perform Algorithm 3 with input x, v, t.}
  4: Set $x' = \cos(\theta)x + \sin(\theta)v$.

---

set

$$\widetilde{H}_t(x, A) = \int_{\mathbb{S}_x^{d-2}} \int_{[0,2\pi)} \mathbb{1}_A\big(\gamma_{(x,v)}(\theta)\big)\, Q_{x,v,t}(\mathrm{d}\theta)\, \mathcal{U}_{\mathbb{S}_x^{d-2}}(\mathrm{d}v)$$

and let

$$\widetilde{H}(x, A) = \frac{1}{p(x)} \int_0^{p(x)} \widetilde{H}_t(x, A)\, \mathrm{d}t = \int_0^{p(x)} \widetilde{H}_t(x, A)\, \mathcal{U}_{(0,p(x))}(\mathrm{d}t).$$

In words, the kernel $\widetilde{H}$ can be described as first sampling a random level t below the current value of the unnormalized density p and then running the auxiliary kernel $\widetilde{H}_t$ associated to that random level.

Remark 14 *Observe that due to the 2π-periodicity of the great circle on the sphere, we have*

$$\cos\big(shrink(x, v, t) \mod 2\pi\big)x + \sin\big(shrink(x, v, t) \mod 2\pi\big)v$$
$$= \cos\big(shrink(x, v, t)\big)x + \sin\big(shrink(x, v, t)\big)v$$

for all $x \in \mathbb{S}^{d-1}$, $v \in \mathbb{S}_{d-2}^x$ and $t \in (0, p(x))$. Therefore, we may introduce this modulo operation when transitioning from the algorithmic formulation to a Markov kernel.

In Appendix E we prove that $\widetilde{H}$ is reversible with respect to our target distribution, which implies that π is the stationary distribution of this kernel. We obtain the following statement.

Proposition 15 *For π defined as in (2) with lower semicontinuous p, the kernel $\widetilde{H}$ is reversible w.r.t. π.*

For the geodesic shrinkage slice sampler we also provide a statement about convergence to the target distribution.

Theorem 16 *For p being lower semicontinuous with $\|p\|_\infty < \infty$ we have*

$$\sup_{x \in \mathbb{S}^{d-1}} d_{tv}\left(\widetilde{H}^n(x, \cdot), \pi\right) \leq \left(1 - \frac{\sup_{t>0}\left[t \cdot \mathcal{U}_{\mathbb{S}^{d-1}}(L(t))\right]}{\sqrt{2\pi}\sqrt{d-1}\,\|p\|_\infty}\right)^n, \qquad \forall n \in \mathbb{N}, \qquad (9)$$

where $\sup_{t>0}\left[t \cdot \mathcal{U}_{\mathbb{S}^{d-1}}(L(t))\right] \in (0, \|p\|_\infty]$.

13

We follow the same strategy of proof as in Theorem 6, i.e., we show that the whole state space is a small set for $\widetilde{H}$ and then apply Lemma 10. To this end, the crucial idea is to estimate the distribution $Q_{x,v,t}$ by its restriction to the event that the loop in Algorithm 3 terminates after the first iteration.

Proof *(Proof of Theorem 16.)* For arbitrary $x \in \mathbb{S}^{d-1}$, $s \in (0, \|p\|_\infty)$, for any $v \in \mathbb{S}_x^{d-2}$ and $\theta \in [0, 2\pi)$ with $\gamma_{(x,v)}(\theta) \in L(s)$ we have (cf. Equation 8) that

$$\frac{1}{p(x)} \int_0^{p(x)} \mathbb{1}_{\left[0, p\left(\gamma_{(x,v)}(\theta)\right)\right)}(t) \, \mathrm{d}t \geq \frac{s}{\|p\|_\infty}. \tag{10}$$

Moreover, for all $t \in (0, p(x))$ we obtain with $\Xi \sim \mathcal{U}_{(0,2\pi)}$ that

$$Q_{x,v,t}(B) \geq \mathbb{P}\big(\Xi \in B \cap L(x,v,t)\big) = \frac{1}{2\pi}\lambda\big(B \cap L(x,v,t)\big), \quad B \in \mathcal{B}([0,2\pi)). \tag{11}$$

Then

$$\widetilde{H}(x,A) \underset{(11)}{\geq} \frac{1}{p(x)} \int_0^{p(x)} \int_{\mathbb{S}_x^{d-2}} \frac{1}{2\pi} \int_{L(x,v,t)} \mathbb{1}_A\big(\gamma_{(x,v)}(\theta)\big) \, \mathrm{d}\theta \, \mathcal{U}_{\mathbb{S}_x^{d-2}}(\mathrm{d}v) \, \mathrm{d}t$$

$$\geq \frac{1}{2\pi} \int_{\mathbb{S}_x^{d-2}} \int_{[0,2\pi)} \mathbb{1}_{A \cap L(s)}\big(\gamma_{(x,v)}(\theta)\big) \frac{1}{p(x)} \int_0^{p(x)} \mathbb{1}_{\left[0, p\left(\gamma_{(x,v)}(\theta)\right)\right)}(t) \mathrm{d}t \, \mathrm{d}\theta \, \mathcal{U}_{\mathbb{S}_x^{d-2}}(\mathrm{d}v)$$

$$\underset{(10)}{\geq} \frac{s}{2\pi \|p\|_\infty} \int_{\mathbb{S}_x^{d-2}} \int_{[0,2\pi)} \mathbb{1}_{A \cap L(s)}\big(\gamma_{(x,v)}(\theta)\big) \, \mathrm{d}\theta \, \mathcal{U}_{\mathbb{S}_x^{d-2}}(\mathrm{d}v) \geq \frac{s \cdot \mathcal{U}_{\mathbb{S}^{d-1}}(A \cap L(s))}{\sqrt{2\pi}\sqrt{d-1}\,\|p\|_\infty},$$

where the last inequality follows by Lemma 11. Therefore, by Lemma 10 and by taking the infimum over $s > 0$ on the right hand-side we get

$$\sup_{x \in \mathbb{S}^{d-1}} d_{tv}\left(\widetilde{H}^n(x,\cdot),\pi\right) \leq \left(1 - \frac{\sup_{t>0}\left[t \cdot \mathcal{U}_{\mathbb{S}^{d-1}}(L(t))\right]}{\sqrt{2\pi}\sqrt{d-1}\,\|p\|_\infty}\right)^n, \quad \forall n \in \mathbb{N}.$$

Finally, note that $\sup_{t>0}\left[t \cdot \mathcal{U}_{\mathbb{S}^{d-1}}(L(t))\right] \in (0, \|p\|_\infty]$ follows already from Theorem 6. ∎

We comment on the result.

Remark 17 *Note that the right hand-side of* (9) *coincides with the the right hand-side of* (7) *stated in Theorem 6, that addressed the uniform ergodicity statement for ideal slice sampling. This is due to the proof technique, since in the small set estimate, we lower bound both kernels with the same expression appearing in Lemma 11. Intuitively, it is clear that* $\sup_{x \in \mathbb{S}^{d-1}} d_{tv}(H^n(x,\cdot),\pi)$ *is smaller than* $\sup_{x \in \mathbb{S}^{d-1}} d_{tv}(\widetilde{H}^n(x,\cdot),\pi)$*, since the shrinkage procedure just adaptively imitates the acceptance/rejection step of the ideal one to gain computational efficiency. Exactly this gain in efficiency leads in applications to a potentially better accuracy to cost ratio, although the performance per Markov chain transition may be worse. We also point out that Remark 7 and Remark 8 apply to Theorem 16 as well.*

14

4. Numerical Illustrations

We apply geodesic slice sampling on the sphere for approximate sampling on various distributions on the sphere and compare it to tailor-made and general purpose MCMC algorithms including random-walk Metropolis-Hastings (RWMH) and Hamiltonian Monte Carlo (HMC). The RWMH algorithm on the sphere is very similar to the standard RWMH, for details refer to Appendix F. Lan et al. (2014) have developed a modified version of HMC for sampling spherical distributions. The details of the HMC algorithm can be found in Appendix G. Both RWMH and HMC have a step-size parameter that is automatically tuned for each target, see Appendix H. For a prespecified sampler we denote by $(x_n)_{n \in \mathbb{N}} \subset \mathbb{S}^{d-1}$ a realization of the corresponding Markov chain.

4.1 Rigid Registration of Biomolecular Structures

First, we discuss an application of the geodesic slice sampler on the sphere to a common inference problem in structural bioinformatics, computer vision, and robotics. Two 3D objects are represented by point clouds, $Q = \{\boldsymbol{q}_i\}_{i=1}^{I}$ and $P = \{\boldsymbol{p}_j\}_{j=1}^{J}$, where $\boldsymbol{q}_i, \boldsymbol{p}_j \in \mathbb{R}^3$. The task is to find a 3D rotation matrix $\boldsymbol{R} \in SO(3)$ that best superimposes both point clouds. This is a rigid registration or pose estimation problem. In general, the point clouds have different size, and the correspondence between points in Q and P is unknown.

We use a probabilistic model inspired by the coherent point drift (CPD) method (Myronenko and Song, 2010) to address this task. According to the model, a point in Q is either generated from a Gaussian mixture model whose components have equal weights and are centered on the rotated source P, or it is an outlier that has no corresponding point in P. Outliers are generated from a uniform distribution $\mathcal{U}_B$ over the bounding box B of the target point cloud (i.e. $B \subset \mathbb{R}^3$ is the smallest rectangle such that $Q \subset B$). We denote the probability density function w.r.t. the Lebesgue measure of this distribution as p_o. The probability that a point is an outlier is $\omega \in [0, 1]$. The probabilistic model, i.e., the conditional density again w.r.t. the Lebesgue measure, for a single point in Q is

$$p(\boldsymbol{q}_i \mid P, \boldsymbol{R}, \omega, \sigma) = \omega p_o(\boldsymbol{q}_i) + \frac{1 - \omega}{J (2\pi\sigma^2)^{3/2}} \sum_{j=1}^{J} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{q}_i - \boldsymbol{R}\boldsymbol{p}_j\|^2\right). \quad (12)$$

We can represent rotation matrices via unit quaternions (Horn, 1987) and thereby map the rigid registration problem to a distribution over $\mathbb{S}^3$. This mapping is achieved by representing rotation matrices with unit quaternions:

$$\boldsymbol{R}(x) = \begin{pmatrix} 1 - 2(x_3^2 + x_4^2) & 2(x_2x_3 - x_1x_4) & 2(x_2x_4 + x_1x_3) \\ 2(x_2x_3 + x_1x_4) & 1 - 2(x_2^2 + x_4^2) & 2(x_3x_4 - x_1x_2) \\ 2(x_2x_4 - x_1x_3) & 2(x_3x_4 + x_1x_2) & 1 - 2(x_2^2 + x_3^2) \end{pmatrix} \quad (13)$$

where $x = (x_1, x_2, x_3, x_4)^T \in \mathbb{S}^3$. Therefore, probabilistic pose estimation boils down to generating samples from a distribution with density p over $\mathbb{S}^3$, where $x \in \mathbb{S}^3$ encodes a 3D rotation matrix. The density p w.r.t the volume measure on $\mathbb{S}^3$ of the overall posterior is (up to normalization) the product of the densities in (12), that is,

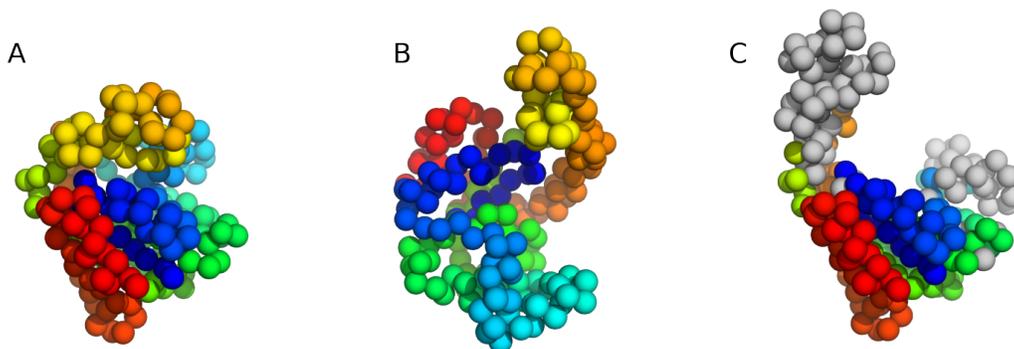$$p(x) \propto \prod_{i=1}^{I} p(\boldsymbol{q}_i \mid P, \boldsymbol{R}(x), \omega, \sigma). \quad (14)$$

15

Figure 3: Rigid registration of point clouds. Each point cloud consists of 214 carbon-α atom positions. Points are colored according to their sequence position: N-terminal amino acids are shown in blue, C-terminal points are red. (A) Target point cloud showing AK in closed configuration. (B) Source point cloud showing AK in open configuration and in an orientation that differs from the orientation of the target. Note that target and source points with identical color correspond to each other. However, due to the non-rigid conformational change not all source points will generate target points. (C) Source rotated optimally so as to generate the target with high probability according to the CPD model (12). Note that the points in the source that do not generate a target point are shown in gray. These are located in the moveable domains of the AK structure (for example the orange domain that is closed in the target, but open in the source).

In the following, we assume that the fraction of outliers ω and the standard deviation σ are known such that the only unknown parameter is the unit quaternion x encoding a rotation of the source relative to the target. We consider an example from structural bioinformatics and test MCMC samplers to tackle this rigid registration task. Adenylate kinase (AK) undergoes a large conformational change upon ligand binding. This internal structural change cannot be modeled as a global rigid transformation. Only the core domain, comprising a 60% subset of all points, can be related to each other via a rigid transformation. In our tests, the target and source point clouds show the closed and open conformation of AK (the protein data bank (PDB) codes 1AKE and 4AKE; see Fig. 3). We fix $\sigma = 1\text{Å}$ and $\omega = 0.4$ (motivated by the fact that the core domain contains roughly 60% of all points); the bounding box volume of 1AKE is $5.73 \times 10^4 \text{Å}^3$. To guide the visual inspection, we use the *chainbow* coloring from blue to red offered by the Pymol software. Note that the posterior (14) is invariant under reordering of the points.

The posterior density (14) has many local peaks as is clear from Figure 4A showing a slice through p along a random great circle passing through the posterior mode. Rigid registration by sampling from posterior (14) is challenging, because standard MCMC samplers tend to get stuck in subordinate peaks and miss the global maximum. To gain an overview of the posterior, we systematically evaluate p over a regular tessellation of $\mathbb{S}^3$ based on the 600-cell (Straub et al., 2017). The 600-cell is the 4D analog of the regular icosahedron and can be

used to approximate $\mathbb{S}^3$ at varying degrees of resolution. A single cell is a 3D tetrahedron whose corners lie in $\mathbb{S}^3$. By splitting each tetrahedron into eight equally-sized tetrahedra we obtain a tessellation at a higher resolution. Starting from the 600-cell encoding 300 unique rotations, we repeat the splitting process four times to obtain 1279264 unit quaternions that tessellate $\mathbb{S}^3$ in a regular fashion. Evaluation of p at these quaternions offers an accurate discretization of the posterior that allows us to study its shape (see Figure 4B). The global maximum of $\log p$ on the quaternions of the 600-cell tessellation is -2192.89. Although, an overwhelming fraction of quaternions achieves only very small posterior probabilities, almost all probability mass is concentrated in the maximum posterior peak: the fractional volume of $\mathbb{S}^3$ covered by quaternions $x \in \mathbb{S}^3$ with $\log p(x) > -2300$ is close to zero (1.27×10^{-4} based on the 600-cell tessellation), but the cumulative probability is close to one ($1 - 4 \times 10^{-44}$). In hindsight, due to the high concentration of the target distribution one could argue that the ability to identify the mode, effectively serving as an optimizer, is of primary importance, while posterior-based uncertainty quantification is less relevant.

To investigate the robustness of probabilistic rigid registration with the geodesic slice samplers, RWMH and HMC, we launched 200 MCMC chains starting from random initial orientations that are generated by sampling unit quaternions uniformly from $\mathbb{S}^3$. We assess the performance of the samplers by counting how many among all 200 runs reached a log posterior probability greater than -2300 and consider these simulations a "success", because it located the dominant posterior mode rather than getting stuck in a subordinate peak. As shown in Figure 4C, the fraction of successful slice sampling runs increases with the number of MCMC iterations. As expected, the rejection strategy converges faster to the posterior mode than the shrinkage strategy, because it explores the sample space more exhaustively. The success rate of the other two MCMC samplers, RWMH and HMC, remains constantly low reaching only up to 7% after 2000 iterations. This demonstrates that geodesic slice sampling can deal with multi-modal posteriors, whereas RWMH and HMC tend to get stuck independent of the number of MCMC iterations. Both geodesic slice samplers reach a success rate of greater than 50% for about 50 iterations. The shrinkage strategy achieves 100% success rate after 1500 iterations, whereas the rejection strategy is at 100% success rate already after 200 iterations. In contrast, the performance of RWMH and HMC is quite poor with a success rates ranging between 3–7%. A similarly poor performance is also observed for a custom-made Gibbs sampler (not shown).

Rigid registration is often just one of many sampling tasks within the context of a more complex probabilistic model. Therefore, it is desirable that highly probable rotations are sampled after few MCMC iterations within a short computation time. This is not achieved by the rejection strategy (see Fig. 4D), because the dominant posterior mode is rather narrow resulting in many rejections. The shrinkage strategy, although requiring more steps to successfully detect the posterior peak with high probability, clearly outperforms any of the other samplers when taking also the computation time into consideration.

4.2 Mixture of von Mises-Fisher Distributions

To study the performance on a multi-modal target, we test the slice samplers as well as RWMH and HMC also on a K-component mixture model of von Mises-Fisher (vMF) distributions in d dimensions. The vMF distribution is defined by the unnormalized density
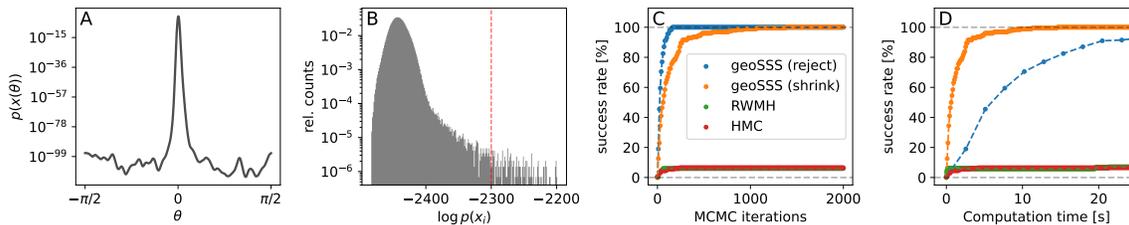
Figure 4: Probabilistic rigid registration of the open and closed structure of adenylate kinase based on posterior p given in (14). (A) Slice through p along a great circle $x(\theta) = \cos\theta\,x_1 + \sin\theta\,x_2$ where $x_1 \in \mathbb{S}^3$ is the global optimum and $x_2 \in \mathbb{S}^2_{x_1}$. (B) Marginal distribution of log posterior probabilities $\log p(x_i)$ based on a discretization of $\mathbb{S}^3$ using $\sim 1.28 \times 10^6$ regularly placed unit quaternions $x_i \in \mathbb{S}^3$. (C) Success rate against number of Markov chain iterations for all four samplers. (D) Success rate against computation time.

$$p_{\mathrm{vMF}}(x; \mu, \kappa) = \exp(\kappa\mu^T x), \quad x \in \mathbb{S}^{d-1}, \tag{15}$$

where $\kappa > 0$ is the concentration parameter and $\mu \in \mathbb{S}^{d-1}$. For a given $K \in \mathbb{N}$ our target distribution is a mixture of vMF distributions where each component has the same weight $1/K$ and a shared concentration parameter κ, i.e., the corresponding unnormalized density takes the form (note that the normalizers of each component are identical due to the shared concentration parameter)

$$p_{\mathrm{mix}}(x) = \sum_{k=1}^{K} p_{\mathrm{vMF}}(x; \mu_k, \kappa), \quad x \in \mathbb{S}^{d-1}, \tag{16}$$

where every $\mu_k \in \mathbb{S}^{d-1}$ with $k \in \{1, \ldots, K\}$ is sampled w.r.t. the uniform distribution on the unit sphere and then fixed.

For this target distribution we also consider a Metropolis algorithm with mixture proposal based on a random walk and independent sampling, which is referred to as mixture-MH. The random walk part allows local exploration, whereas the independence sampling, regarding the uniform distribution on the sphere, encourages global moves. In this scheme the mixture hyperparameter $\alpha \in [0, 1]$ appears: With probability α a random walk and with probability $1 - \alpha$ an independent sampling proposal is used. The resulting Metropolis kernel can also be interpreted as a mixture of RWMH and independent Metropolis. For implementation details we refer to Algorithm 8 under Appendix F.

We evaluate both geoSSS variants against RWMH, HMC and the mixture-MH (for suitable $\alpha \in [0, 1]$) on a 10-dimensional mixture of vMF distributions with $K = 5$ vMF-mixture components on the unit sphere $\mathbb{S}^9$. We set $\kappa = 100$ and draw 10^6 samples. To gain some insight into the performance of the samplers on the multi-modal target, we first look at the marginal distributions (see Fig. 5). We estimate marginal histograms from the samples and compare them with the marginal distributions of the vMF mixture that we computed analytically. Both variants of geoSSS represent the underlying mixture significantly better
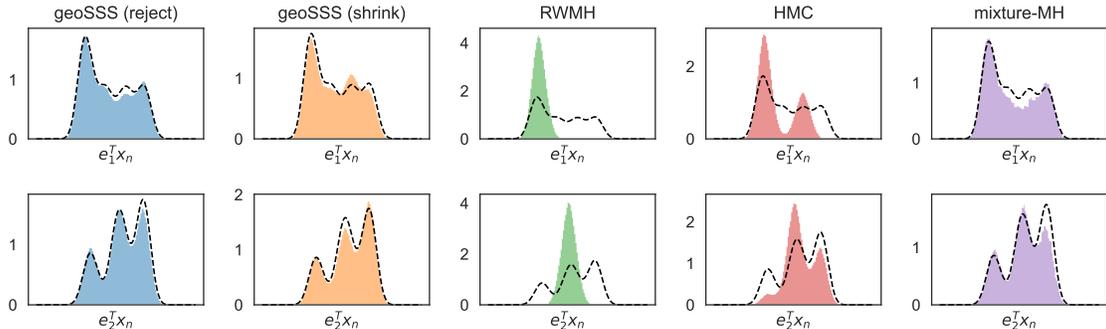
18

Figure 5: Marginal distributions (shown in colors) for the mixture of vMF distributions ($d = 10, \kappa = 100, K = 5$) where columns corresponds to MCMC methods and rows correspond to the dimension of the marginal (only the first two dimensions are shown). Here, $\{e_i\}_{i=1}^{10}$ denotes the standard basis in $\mathbb{R}^{10}$. The black dashed line indicates the true marginals computed analytically for the mixture of vMF distribution.

compared to HMC and especially RWMH which seems to explore only a single component of the vMF mixture. The mixture-MH is almost on-par with our GeoSSS (shrink) variant. However, this represents the best result achieved by sweeping through α values[1] with 10^6 MCMC steps for each case and therefore represents an additional computational burden of tuning this hyperparameter along with the step-size of RWMH for every target density.

Next we study quality measures to obtain more quantitative insights into the sampling performance. RWMH is the fastest to decorrelate, which is reflected by the rapid decay of the autocorrelation function (ACF) (see Fig. 6A). However, the reason for the fast decorrelation is that RWMH fails to explore the entire distribution, but is stuck in a single component. The geoSSS variants decorrelate much faster as compared to mixture-MH and HMC. Note that the slow, linear decay of the ACF for HMC results from the fact that it spends a large fraction of the simulation time in a single mode and only occasionally escapes to another mode (also HMC only visits three out of all five modes in the course of the simulation).

As a measure for the quality of approximate sampling we consider the frequency with which the samplers visit the K modes of the target mixture model (16). Ideally, these frequencies should be constant, because each component of the mixture has the same weight and concentration parameter. We contrast the empirical frequency q_k with which the k-th mode is visited by a sampler with the uniform distribution by using the Kullback-Leibler

---

[1] For the $d = 10, \kappa = 100, K = 5$ vMF mixture target, the best mixing probability $\alpha = 0.2$ corresponds to the lowest KL divergence value (see Fig. 6B). For the systematic evaluation, we varied $\alpha \in [0, 1]$ with a grid spacing of 0.1.
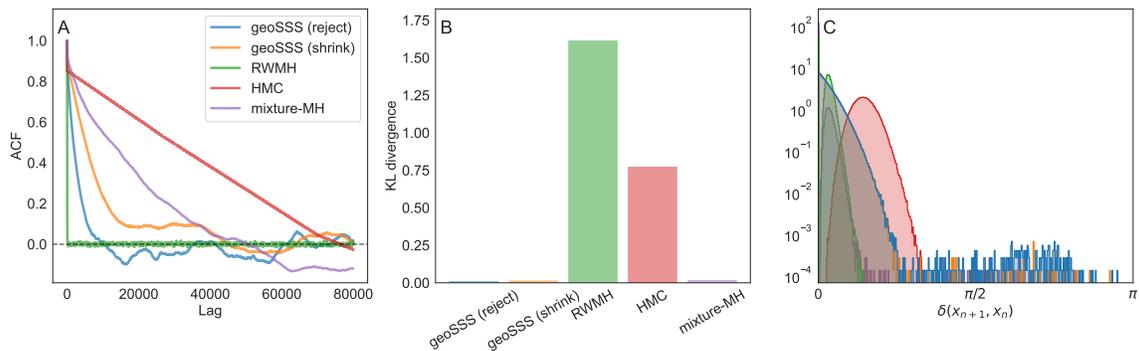
Figure 6: (A) ACF plotted for the first dimension from the 10-dimensional mixture of vMF distributions with $K = 5$ and $\kappa = 100$. (B) Kullback-Leibler divergence between the empirical and ideal frequency with which each mode is visited by the four MCMC samplers. (C) Geodesic distance (log-scale) between successive approximate samples.

(KL) divergence

$$\mathrm{KL}(q \mid \ell) := \sum_{k=1}^{K} q_k \log(q_k/\ell_k)$$

where $\ell = (\ell_1, \ldots, \ell_K)$ with $\ell_k = 1/K$ for $k \in \{1, \ldots, K\}$ is the ideal distribution and $q = (q_1, \ldots, q_K) \in [0,1]^K$. To estimate q_k, we assign each sample x_n to the component under which x_n is most likely, i.e.,

$$k_n = \underset{k \in \{1, \ldots, K\}}{\operatorname{argmax}} \ p_{\mathrm{vMF}}(x_n; \mu_k, \kappa).$$

Based on the allocations k_n, we can count how often each mode is visited by the Markov chain. A good MCMC sampler should produce $q_k \approx 1/K$ resulting in $\mathrm{KL}(q \mid \ell) \approx 0$. The more the KL divergence differs from the minimum value of zero, the greater is the mismatch between the ideal and empirical frequency of mode visits. As seen from Fig. 6B, geoSSS (reject) achieves the lowest KL divergence whereas geoSSS (shrink) and mixture-MH perform comparably in achieving low KL divergence values. However, RWMH and HMC produce large KL divergences, indicating a complete failure to detect some of the components and, as a result, a poorer representation of the target distribution compared to geoSSS. Indeed, RWMH is stuck in a single mode in the course of the simulation. HMC fares better, but also misses two out of five modes.

Another quality measure is the geodesic or great circle distance between successive samples given as

$$\delta(x_{n+1}, x_n) := \arccos(x_{n+1}^T x_n) \,.$$

An efficient MCMC algorithm should explore the sphere rapidly by making large leaps from one sample to the next. The great circle distance between successive samples (see Fig.
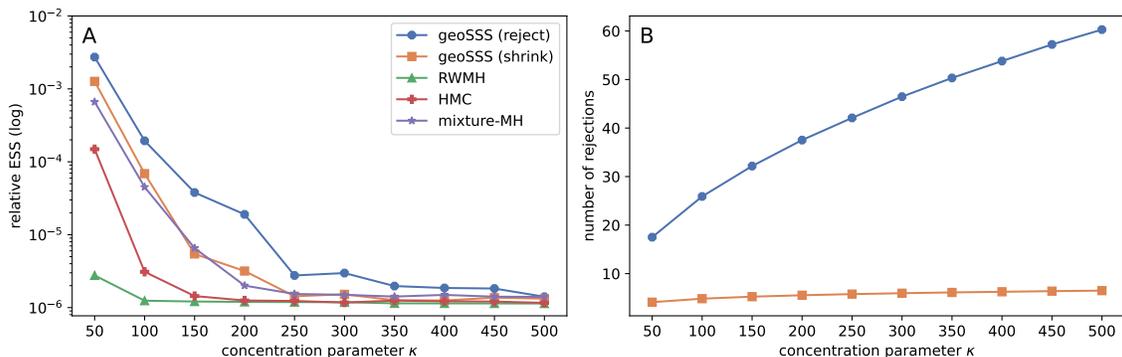
20

Figure 7: Comparing the MCMC samplers for the mixture of vMF distribution ($d = 10$, $K = 5$) by varying κ from 50 to 500. (A) Considering the first dimension, estimating the relative effective sample size (ESS) for 10 chains. (B) Number of rejections per MCMC step.

6C) shows that the slice samplers and mixture-MH explore the sphere more efficiently than HMC and RWMH.

We further evaluate the performance of the samplers by fixing the components on the sphere and increasing the concentration parameter κ from 50 to 500, therefore making the distributions "spikier". We estimate the effective sample size (ESS) for each method by considering 10 chains and 10^6 MCMC steps from the first dimension, resulting in ESS values per method for each κ. Figure 7A shows the estimated relative[2] ESS for each of the four samplers as a function of the concentration parameter κ. Overall, we observe that sampling the mixture model becomes more challenging as the components become more concentrated with increasing κ for all methods. Notably, the geoSSS (reject) variant achieves the best relative ESS. We can again observe the trend that geoSSS (shrink) and mixture-MH (with the mixture hyperparameter fine-tuned for every κ corresponding to the highest relative ESS) are comparable as well as that both outperform RWMH and HMC, particularly for lower κ values. At large κ values, all samplers perform similarly poorly and fail to produce a reliable approximation of the target. It is worth noting that although the rejection-based geoSSS yields the highest ESS, it is at the cost of generating a significantly larger number of rejected samples in comparison to the shrinkage-based geoSSS. This is demonstrated in Fig. 7B, where it is observed that the rejection-based geoSSS generates an increasingly larger number of rejected samples for higher values of κ. For instance, for $\kappa = 50$, approximately 17 rejections happen per MCMC step, while for $\kappa = 500$, around 60 rejections occur. On the other hand, the number of rejections for the shrinkage-based geoSSS increases at a much slower rate with κ. It can be seen here that for $\kappa = 50$, we have approximately 4 rejections per MCMC step, and for $\kappa = 500$, approximately just 6 rejections per MCMC step occur. The number of rejections directly determines the number of log probability

---

[2]Here and elsewhere the relative ESS refers to the estimated ESS divided by the total number of MCMC steps.

evaluations and thereby the computational costs of the slice samplers. RWMH and HMC, on the other hand, have a fixed computational budget per MCMC step. In case of RWMH, one step requires a single log probability evaluation, whereas an HMC step involves one log probability evaluation as well as multiple gradient evaluations, one for each leapfrog step. In our settings, we use 10 leapfrog integration steps. In summary, RWMH has the lowest computational costs, but also the smallest ESS. The computational costs of shrinkage-based geoSSS is smaller than the costs of HMC, whilst it also performs better than HMC in terms of the other evaluation criteria.

4.3 Curved Distribution on the Sphere

For a target density where the 'mass' accumulates along narrow connected regions, HMC is expected to show superior performance. To study such a sampling problem, we define a spherical distribution along a curve. The curve is created by picking 10 consecutive points on the $(d-1)$-sphere and to ensure a smooth path, we connect every two successive points $x, y \in \mathbb{S}^{d-1}$ using spherical linear interpolation (slerp), i.e., we apply

$$t \mapsto \frac{\sin(\theta\,(1-t))x + \sin(\theta\,t)y}{\sin(\theta)},$$

as map from $[0,1]$ to $\mathbb{S}^{d-1}$, where $\theta = \arccos(x^T y)$, cf. (Hanson, 1995). If $\mu(t)$ with $t \in [0,1]$ is the curve obtained by concatenating and rescaling the slerps between the 9 pairs of successive points, then we define the spherical distribution with unnormalized density

$$p(x) = \exp\left(\kappa \max_{t \in [0,1]} x^T \mu(t)\right), \qquad x \in \mathbb{S}^{d-1}, \tag{17}$$

which we also call curved von Mises-Fisher distribution (curved vMF). The maximum of $x^T\mu(t)$ for a single slerp can be computed in closed form. The total maximum over the entire curve is then the maximum over the maxima of slerps connecting two successive points. The probability distribution corresponding to (17) concentrates probability mass around the curve $\mu(t)$. As in the case for the standard vMF distribution, the parameter $\kappa > 0$ controls the concentration.

For our initial sampling tests for the curved vMF target, we consider $\mathbb{S}^2$ and choose $\kappa = 500$. Again, we run all four MCMC samplers. Rejection-based geoSSS and HMC[3] take roughly the same amount of time, with HMC taking slightly longer. However, both are almost twice as slow as shrinkage-based geoSSS. RWMH is by far the fastest method (refer to Appendix I for run times).

Figure 8 shows the target density and samples of the four MCMC algorithms generated in the last 5000 iterations. The underlying target density shown here is evaluated by uniformly distributing 3×10^4 points on the 2-sphere using the method of Saff and Kuijlaars (1997). Visual inspection shows that the slice samplers and also HMC explore the target much more rapidly than RWMH, which is also reflected in the high autocorrelation (see Fig. 9A). To verify this observation, we measure the discrepancy between the discretized target and

---

[3]More details on the implementation of the required gradient can be found in the Python library `https://github.com/microscopic-image-analysis/geosss` and in the following note `geosss/scripts/gradient-curve.pdf`

Figure 8: First panel from the left shows the curved vMF target density on the 2-sphere with concentration parameter $\kappa = 500$. Remaining panels show the last 5000 samples from the four different MCMC samplers targeting this density.
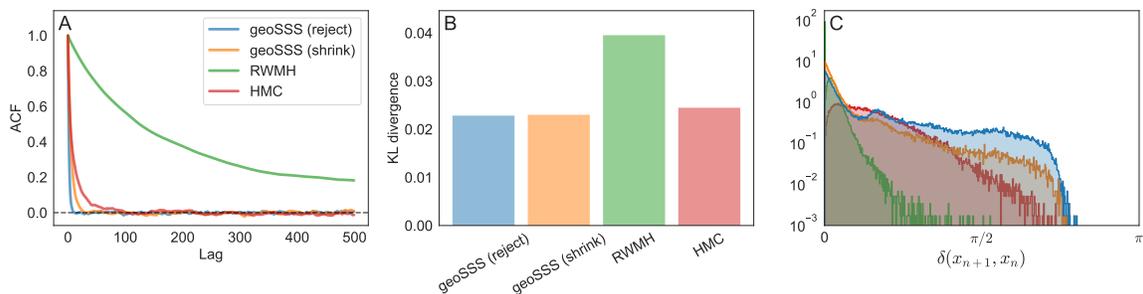

Figure 9: (A) ACF of the first dimension plotted for the curved vMF target on 2-sphere based on 10^5 samples. (B) KL divergence between a discretized version of the curved vMF target on 2-sphere and a histogram compiled from the MCMC samples. (C) Geodesic distance (log-scale) between successive approximate samples for the curved von Mises-Fisher distribution (17) on 2-sphere.

the histogram generated by the sampling procedure using KL divergence (see Fig. 9B). Moreover, we monitor the distance between successive samples (see Fig. 9C). Across all three evaluation criteria we consider, RWMH is clearly disfavored. On the other hand, geoSSS and HMC show a similar performance on this low-dimensional target.

However, this picture changes as the dimensionality of the target increases from $d = 3$ to $d = 5$. We compute the relative ESS for the first dimension by varying the concentration parameter. By using 10 chains, fixing the dimension $d = 5$ and drawing 10^6 samples per chain, we observe (see Fig. 10A), that HMC consistently outperforms the remaining samplers across all values of κ. The trend of decreasing ESS with increasing κ aligns with results from the previous section on the mixture of vMF distributions. HMC is expected to perform exceptionally well on such targets, even if the dimensionality is increased further. To assess the performance of the remaining samplers under similar conditions, we repeat the ESS calculations, however in this case, we fix the concentration parameter with $\kappa = 500$ and vary the dimensions from 3 to 24. As observed in Fig. 10B, while the ESS values of
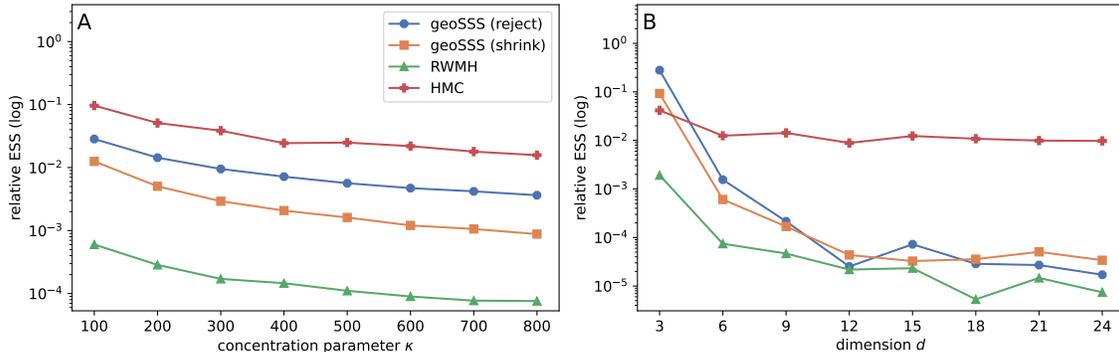
Figure 10: Comparing the relative ESS values. (A) Fixing the dimension $d = 5$ and varying the concentration paramater. (B) Fixing $\kappa = 500$ and varying the dimensions.

the slice samplers are better than RWMH, all three exhibit rapid decline with increasing dimension. As expected, in contrast, HMC demonstrates superior performance, maintaining fairly consistent ESS values across all dimensions.

5. Summary

We introduce two slice sampling based MCMC-methods on the sphere, the ideal geodesic and geodesic shrinkage slice samplers, that use movements on great circles. For the Markov kernels of both samplers we are able to establish reversibility and explicit convergence statements under mild assumptions. In numerical experiments, we see that in particular for moderately concentrated target distributions with a moderate number of dimensions our slice samplers perform well and can compete with RWMH and HMC on the sphere, while having the additional advantage of being free of tuning parameters. For multimodal target distributions we observe that the ideal geodesic slice sampler and the geodesic shrinkage slice sampler even outperform RWMH and HMC. However, we observe that the performance of our methods deteriorates for increasing concentration and increasing dimension of the target distribution. This dependence on concentration and dimension already appears in Theorem 6 and Theorem 16. Regarding the comparison between the two geodesic slice samplers, the ideal version seems to outperform the shrinkage based sampler in terms of Markov chain transitions. However, this comes with a (significantly) greater computational cost per transition.

Acknowledgments

Code Availability

We have implemented our geoSSS algorithms, including the spherical RWMH and HMC methods (detailed in Appendices F and G), in our `geosss` package available at `https://github.com/microscopic-image-analysis/geosss`. You can find the scripts used to generate the numerical illustrations in the `scripts` directory, and the protein example data in the `data` directory. Also the precomputed results are archived on Science Data Bank `https://doi.org/10.57760/sciencedb.30181`.

Appendix A. Simulation of $\mathcal{U}_{\mathbb{S}_x^{d-2}}$

Implementing the slice sampling algorithms proposed in this paper involves sampling from $\mathcal{U}_{\mathbb{S}_x^{d-2}} = \frac{1}{\omega_{d-2}} \mu_x$ for $x \in \mathbb{S}^{d-1}$. A scheme for performing this is provided in Algorithm 5 and justified as follows. It exploits that $\mathbb{S}_x^{d-2}$ can be interpreted "as tilted $\mathbb{S}^{d-2}$, such that it lies in the $(d-1)$-dimensional hyperplane orthogonal to x". We formalize the arguments and require further objects. For $x \in \mathbb{S}^{d-1}$ let

$$E_x := \{ y \in \mathbb{R}^d \mid y^T x = 0 \}$$

be the $(d-1)$-dimensional hyperplane with normal vector x. Since this is a $(d-1)$-dimensional linear subspace of $\mathbb{R}^d$, we can choose an orthonormal basis $e_1^{(x)}, \ldots, e_{d-1}^{(x)} \in E_x$ of E_x. The isometric mapping

$$\varphi_x : \mathbb{R}^{d-1} \to E_x, \qquad (y_1, \ldots, y_{d-1}) \mapsto \sum_{i=1}^{d-1} y_i e_i^{(x)} \tag{18}$$

describes the rotation of $\mathbb{R}^{d-1}$ into E_x. Therefore it maps the unit sphere in $\mathbb{R}^{d-1}$ to the unit sphere in E_x, i.e., $\varphi_x(\mathbb{S}^{d-2}) = \mathbb{S}_x^{d-2}$. Isometries preserve volume, such that

$$\mu_x(A) = \sigma_{d-2}\left(\varphi_x^{-1}(A) \right), \qquad \forall\, A \in \mathcal{B}(\mathbb{S}_x^{d-2}), \tag{19}$$

provides the crucial relation between the volume measure on $\mathbb{S}^{d-2}$ and the volume measure μ_x on the tilted version $\mathbb{S}_x^{d-2}$, (see also Munkres, 1991, Exercise 25.4). In particular, $\mu_x(\mathbb{S}_x^{d-2}) = \omega_{d-2}$.

We prove the validity of Algorithm 5 based upon this relationship between σ_{d-2} and μ_x. For $x \in \mathbb{S}^{d-1}$ we establish an expression for the mapping $y \mapsto y - (x^T y)\, x$ in terms of φ_x that describes the projection of $\mathbb{R}^d$ onto E_x. Then, we deduce that Algorithm 5 realizes a transformed normalized standard normally distributed random variable under the isometry φ_x. By (19) this implies the desired result formalized as follows.

Lemma 18 *Let Y be a random variable distributed according to the d-dimensional standard normal distribution $\mathcal{N}(0, \mathrm{Id}_d)$. Then, for any $x \in \mathbb{S}^{d-1}$ we have $\frac{Y - x^T Y\, x}{\|Y - x^T Y\, x\|} \sim \mathcal{U}_{\mathbb{S}_x^{d-2}}$.*

Proof Let $U \in \mathbb{R}^{d \times (d-1)}$ be the matrix with columns $e_1^{(x)}, \ldots, e_{d-1}^{(x)}$ and let $\widetilde{U} \in \mathbb{R}^{d \times d}$ be the matrix with columns $e_1^{(x)}, \ldots, e_{d-1}^{(x)}, x$, where $e_1^{(x)}, \ldots, e_{d-1}^{(x)}$ are defined as above. Since $e_1^{(x)}, \ldots, e_{d-1}^{(x)}, x$ is an orthonormal basis of $\mathbb{R}^d$, we may write $y \in \mathbb{R}^d$ as $y = \sum_{i=1}^{d-1} \alpha_i e_i^{(x)} + \alpha_d x$ where $\alpha := (\alpha_1, \ldots, \alpha_d)^T = \widetilde{U}^T y$. Therefore, we have

$$y - (x^T y)\, x = \sum_{i=1}^{d-1} \alpha_i e_i^{(x)} = \varphi_x\left((\alpha_1, \ldots, \alpha_{d-1})^T \right) = \varphi_x(U^T y).$$

Algorithm 5 Sampling from $\mathcal{U}_{\mathbb{S}_x^{d-2}}$.

---

input point $x \in \mathbb{S}^{d-1}$
output sample v from $\mathcal{U}_{\mathbb{S}_x^{d-2}}$
  1: Draw $Y \sim \mathcal{N}(0, \mathrm{Id}_d)$, call the result y.
  2: Set $z = y - (x^T y)\, x$ and $v = z / \|z\|$.

Since $Y \sim \mathcal{N}(0, \mathrm{Id}_d)$ and $\widetilde{U}$ is orthogonal, also $\widetilde{U}^T Y \sim \mathcal{N}(0, \mathrm{Id}_d)$. Observe that the distribution of $U^T Y$ is a marginal distribution of the distribution of $\widetilde{U}^T Y$. Hence $U^T Y \sim \mathcal{N}(0, \mathrm{Id}_{d-1})$. This implies $\frac{U^T Y}{\|U^T Y\|} \sim \mathcal{U}_{\mathbb{S}^{d-2}} = \frac{1}{\omega_{d-2}} \sigma_{d-2}$. Using (19) and that φ_x is a linear isometry, we obtain for $A \in \mathcal{B}(\mathbb{S}_x^{d-2})$ that

$$
\mathbb{P}\left(\frac{Y - x^T Y\, x}{\|Y - x^T Y\, x\|} \in A \right) = \mathbb{P}\left(\varphi_x\left(\frac{U^T Y}{\|U^T Y\|} \right) \in A \right) = \mathbb{P}\left(\frac{U^T Y}{\|U^T Y\|} \in \varphi_x^{-1}(A) \right)
$$

$$
= \frac{1}{\omega_{d-2}} \sigma_{d-2}\left(\varphi_x^{-1}(A) \right) = \frac{1}{\omega_{d-2}} \mu_x(A) = \mathcal{U}_{\mathbb{S}_x^{d-2}}(A). \qquad \blacksquare
$$

Appendix B. Proof of Lemma 1 and Lemma 2

We start with the proof of Lemma 1 that follows by an elementary calculation.

Proof *(Proof of Lemma 1.)* For $x \in \mathbb{S}^{d-1}$, $v \in \mathbb{S}_x^{d-2}$ and $\theta \in \mathbb{R}$, using classic trigonometric identities of sine and cosine, we obtain for all $r \in \mathbb{R}$ that

$$
\gamma_{T_\theta(x,v)}(r) = \cos(r)\big(\cos(\theta)x + \sin(\theta)v \big) + \sin(r)\big(\sin(\theta)x - \cos(\theta)v \big)
$$

$$
= \cos(\theta - r)x + \sin(\theta - r)v = \gamma_{(x,v)}(\theta - r).
$$

This implies $\gamma_{T_\theta(x,v)}(\theta) = \gamma_{(x,v)}(0) = \cos(0)x + \sin(0)v = x$. $\qquad \blacksquare$

We turn to the proof of Lemma 2. For convenience we introduce

$$
\mathcal{L}(A) := \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}_x^{d-2}} \mathbb{1}_A(x, v) \mu_x(\mathrm{d}v) \sigma_{d-1}(\mathrm{d}x), \qquad A \in \mathcal{B}(\mathcal{S}\mathbb{S}^{d-1}), \tag{20}
$$

for the measure on $\mathcal{S}\mathbb{S}^{d-1} = \bigcup_{x \in \mathbb{S}^{d-1}} \left(\{x\} \times \mathbb{S}_x^{d-2} \right)$ that "sews" up the volume measure on the fibers of $\mathcal{S}\mathbb{S}^{d-1}$ by the volume measures on $\mathbb{S}^{d-1}$, and call $\mathcal{L}$ the *Liouville measure*. We also use the following map.

Definition 19 *Let $\theta \in \mathbb{R}$. The function*

$$
\psi_\theta : \mathcal{S}\mathbb{S}^{d-1} \to \mathcal{S}\mathbb{S}^{d-1}, \qquad (x, v) \mapsto \big(\cos(\theta)x + \sin(\theta)v, -\sin(\theta)x + \cos(\theta)v \big)
$$

is called the geodesic flow on the sphere.

To prove Lemma 2 we exploit that we can write T_θ as a composition of the geodesic flow and a "sign flip" in the second component. Then Lemma 2 follows by invariance properties of the Liouville measure. Note that when naming $\mathcal{L}$ and ψ_θ we adhere to the terminology of Riemannian geometry.

Proof *(Proof of Lemma 2.)* Let $\theta \in \mathbb{R}$, and note that the Liouville measure is invariant w.r.t. the geodesic flow, that is,

$$
\mathcal{L}\left(\psi_\theta^{-1}(A) \right) = \mathcal{L}(A), \quad \forall A \in \mathcal{B}(\mathcal{S}\mathbb{S}^{d-1}),
$$

(see e.g., Chavel, 1984, Section V.2). Moreover, define

$$
\iota : \mathcal{S}\mathbb{S}^{d-1} \to \mathcal{S}\mathbb{S}^{d-1}, \qquad (x, v) \mapsto (x, -v).
$$

27

Observe that the Liouville measure is invariant under ι (see e.g., Paternain, 1999, Lemma 1.34). We can express the map T_θ as $T_\theta = \iota \circ \psi_\theta$. Therefore the invariance of the Liouville measure under ψ_θ and ι yields

$$\int_{\mathcal{S}\mathbb{S}^{d-1}} F\left(T_\theta(y)\right) \mathcal{L}(\mathrm{d}y) = \int_{\mathcal{S}\mathbb{S}^{d-1}} (F \circ \iota \circ \psi_\theta)(y)\, \mathcal{L}(\mathrm{d}y) = \int_{\mathcal{S}\mathbb{S}^{d-1}} (F \circ \iota)(y)\, \mathcal{L}(\mathrm{d}y)$$

$$= \int_{\mathcal{S}\mathbb{S}^{d-1}} F(y)\, \mathcal{L}(\mathrm{d}y). \qquad\blacksquare$$

Appendix C. Proof of Proposition 5

A useful tool for showing reversibility of Markov kernels exhibiting the same structure as the ideal geodesic slice sampling kernel is Lemma 1 by Łatuszyński and Rudolf (2024). It applies to subsets of $\mathbb{R}^d$ in its original formulation, but can be extended to arbitrary σ-finite measure spaces. For the convenience of the reader we adapt the relevant parts of the aforementioned lemma to our setting.

Lemma 20 *Let*

$$P(x, A) \coloneqq \frac{1}{p(x)} \int_0^{p(x)} P_t(x, A \cap L(t))\, \mathrm{d}t, \qquad x \in \mathbb{S}^{d-1}, A \in \mathcal{B}(\mathbb{S}^{d-1}),$$

be a Markov kernel where $P_t : L(t) \times \mathcal{B}(L(t)) \to [0,1]$ for $t \in (0, \|p\|_\infty)$ are themselves Markov kernels. If P_t is reversible with respect to $\mathcal{U}_{L(t)}$ for all $t \in (0, \|p\|_\infty)$, then P is reversible with respect to π.

Remark 21 *By the fact that p is lower semicontinuous for any $t \in (0, \|p\|_\infty)$ we have that $L(t)$ is open and non-empty. Therefore, $\sigma_{d-1}(L(t)) \in (0, \infty)$ such that $\mathcal{U}_{L(t)}$ is well defined.*

We add an auxiliary result w.r.t. the volume of the geodesic level sets under the map T_θ.

Lemma 22 *Let $x \in \mathbb{S}^{d-1}$, $v \in \mathbb{S}_x^{d-2}$ and $t \in (0, \|p\|_\infty)$. Then, for all $\theta \in \mathbb{R}$ we have $\lambda\big(L(T_\theta(x,v),t)\big) = \lambda\big(L(x,v,t)\big)$.*

Proof Let $x \in \mathbb{S}^{d-1}$, $v \in \mathbb{S}_x^{d-2}$ and $t \in (0, \|p\|_\infty)$. For $\theta \in \mathbb{R}$ set

$$w_\theta \colon [0, 2\pi) \to [0, 2\pi), \quad r \mapsto (\theta - r) \cdot \mathbb{1}_{(-\infty, \theta]}(r) + (\theta - r + 2\pi) \cdot \mathbb{1}_{(\theta, \infty)}(r).$$

Due to the 2π-periodicity of sine and cosine, we have $L(T_\theta(x,v), t) = w_\theta^{-1}\big(L(x,v,t)\big)$. We obtain $\lambda\left(L(T_\theta(x,v),t)\right) = \lambda\left(w_\theta^{-1}\big(L(x,v,t)\big)\right) = \lambda\left(L(x,v,t)\right)$, since the Lebesgue measure λ is invariant under w_θ. $\qquad\blacksquare$

Proof *(Proof of Proposition 5.)* For $t \in (0, \|p\|_\infty)$ and $A, B \in \mathcal{B}(\mathbb{S}^{d-1})$ Lemma 2 implies

$$\omega_{d-2} \int_{B \cap L(t)} H_t(x, A)\, \sigma_{d-1}(\mathrm{d}x)$$

$$= \int_{B \cap L(t)} \int_{\mathbb{S}_x^{d-2}} \frac{1}{\lambda(L(x,v,t))} \int_{L(x,v,t)} \mathbb{1}_A\big(\gamma_{(x,v)}(\theta)\big)\, \mathrm{d}\theta\, \mu_x(\mathrm{d}v)\, \sigma_{d-1}(\mathrm{d}x)$$

$$= \int_{[0,2\pi)} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}_x^{d-2}} \frac{\mathbb{1}_{L(T_\theta(x,v),t)}(\theta)\, \mathbb{1}_{B \cap L(t)}(\gamma_{(x,v)}(\theta))\, \mathbb{1}_A\big(\gamma_{T_\theta(x,v)}(\theta)\big)}{\lambda(L(T_\theta(x,v),t))}\, \mu_x(\mathrm{d}v)\, \sigma_{d-1}(\mathrm{d}x)\, \mathrm{d}\theta.$$

Then (5), Lemma 1 and Lemma 22 yield

$$\omega_{d-2} \int_{B \cap L(t)} H_t(x, A)\, \sigma_{d-1}(\mathrm{d}x)$$

$$= \int_{[0,2\pi)} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}_x^{d-2}} \frac{1}{\lambda(L(x, v, t))} \mathbb{1}_{L(t)}(x) \mathbb{1}_B(\gamma_{(x,v)}(\theta)) \mathbb{1}_{L(x,v,t)}(\theta) \mathbb{1}_A(x) \mu_x(\mathrm{d}v) \sigma_{d-1}(\mathrm{d}x)\, \mathrm{d}\theta$$

$$= \omega_{d-2} \int_{A \cap L(t)} H_t(x, B)\, \sigma_{d-1}(\mathrm{d}x).$$

Hence, H_t is reversible w.r.t. $\mathcal{U}_{L(t)} = \frac{1}{\sigma_{d-1}(L(t))} \sigma_{d-1}|_{L(t)}$. Lemma 20 then implies that H is reversible with respect to π. ∎

Appendix D. Proof of Lemma 11

In this section we prove the integral estimate of Lemma 11. A major part of deriving it consists of handling the measure resulting from exploring the sphere along the great circle passing through a fixed $x \in \mathbb{S}^{d-1}$. At first sight, one could think that this measure is σ_{d-1}. Keeping in mind though that, interpreting x as the "north pole", the great circles lie much "denser" at the poles than at the equator, it becomes rather clear that σ_{d-1} can only be a lower estimate for the measure obtained by exploring $\mathbb{S}^{d-1}$ via the great circles through x.

To prove the desired statement, we make use of an expression for σ_{d-1} in terms of polar coordinates (see Schilling, 2017, Corollary 16.19). Define the polar coordinate transformations

$$f_{d-1} : (0, \pi)^{d-2} \times (-\pi, \pi) \to \mathbb{R}^d \setminus \{(x_1, \ldots, x_d)^T \in \mathbb{R}^d \mid x_d = 0, x_{d-1} \leq 0\},$$

$$(\theta_1, \ldots, \theta_{d-1}) \mapsto \begin{pmatrix} \cos(\theta_1) \\ \sin(\theta_1)\cos(\theta_2) \\ \prod_{i=1}^{2} \sin(\theta_i)\cos(\theta_3) \\ \vdots \\ \prod_{i=1}^{d-2} \sin(\theta_i)\cos(\theta_{d-1}) \\ \prod_{i=1}^{d-1} \sin(\theta_i) \end{pmatrix},$$

and the absolute value of the determinant of their Jacobians

$$J_{d-1}(\theta_1, \ldots, \theta_{d-1}) = \sin^{d-2}(\theta_1)\sin^{d-3}(\theta_2) \cdot \ldots \cdot \sin(\theta_{d-2}).$$

Then for all $A \in \mathcal{B}(\mathbb{S}^{d-1})$ we have

$$\sigma_{d-1}(A) = \int_0^\pi \cdots \int_0^\pi \int_{-\pi}^\pi \mathbb{1}_A\big(f_{d-1}(\theta_1, \ldots, \theta_{d-1})\big) J_{d-1}(\theta_1, \ldots, \theta_{d-1})\, \mathrm{d}\theta_{d-1} \ldots \mathrm{d}\theta_1. \quad (21)$$

Proof *(Proof of Lemma 11.)* Recall that $\mu_x = \varphi_x(\sigma_{d-2})$ where φ_x is defined in (18). This allows us to shift from $\mathbb{S}_x^{d-2}$ to $\mathbb{S}^{d-2}$, where we may use the previous explicit expression in

polar coordinates. That is, applying (21) to σ_{d-2} we get

$$\int_0^\pi \int_{\mathbb{S}_x^{d-2}} \mathbb{1}_A\big(\gamma_{(x,v)}(\theta)\big) \mu_x(\mathrm{d}v) \, \mathrm{d}\theta = \int_0^\pi \int_{\mathbb{S}^{d-2}} \mathbb{1}_A\big(\gamma_{(x,\varphi_x(y))}(\theta)\big) \sigma_{d-2}(\mathrm{d}y) \, \mathrm{d}\theta$$

$$= \int_0^\pi \int_0^\pi \cdots \int_0^\pi \int_{-\pi}^\pi \mathbb{1}_A\left(\gamma_{(x,\varphi_x(f_{d-2}(\theta_1,\ldots,\theta_{d-2})))}(\theta)\right) J_{d-2}(\theta_1,\ldots,\theta_{d-2}) \mathrm{d}\theta_{d-2}\ldots \mathrm{d}\theta_1 \, \mathrm{d}\theta.$$

Now, we can use the outer integral (corresponding to travelling along the great circles) to add a dimension in the explicit polar coordinate representation of the volume measure on the sphere at the cost of introducing a correction term. To this end, extend the map φ_x to

$$\widetilde{\varphi}_x : \mathbb{S}^{d-1} \to \mathbb{S}^{d-1}, \qquad (y_1,\ldots,y_d) \mapsto y_1 x + \sum_{i=1}^{d-1} y_{i+1} e_i^{(x)}.$$

Observe that this reparametrization of $\mathbb{S}^{d-1}$ is compatible with the geodesic structure of the sphere, because it respects the basis of $\mathbb{S}_x^{d-2}$ chosen by the map φ_x. More precisely,

$$\gamma_{(x,\varphi_x(f_{d-2}(\theta_1,\ldots,\theta_{d-2})))}(\theta) = \widetilde{\varphi}_x\big(f_{d-1}(\theta,\theta_1,\ldots,\theta_{d-2})\big).$$

Furthermore, defining

$$g : \mathbb{R}^d \setminus \big\{(y_1,\ldots,y_d) \in \mathbb{R}^d \mid |y_1| > 1\big\} \to \mathbb{R}_+, \qquad (y_1,\ldots,y_d) \mapsto \frac{1}{\sin^{d-2}(\arccos(y_1))},$$

we have

$$J_{d-2}(\theta_1,\ldots,\theta_{d-2}) = g\big(f_{d-1}(\theta,\theta_1,\ldots,\theta_{d-2})\big) J_{d-1}(\theta,\theta_1,\ldots,\theta_{d-2}).$$

Therefore

$$\int_0^\pi \int_{\mathbb{S}_x^{d-2}} \mathbb{1}_A\big(\gamma_{(x,v)}(\theta)\big) \mu_x(\mathrm{d}v) \, \mathrm{d}\theta$$

$$= \int_0^\pi \int_0^\pi \cdots \int_0^\pi \int_{-\pi}^\pi \mathbb{1}_A\left(\widetilde{\varphi}_x\big(f_{d-1}(\theta,\theta_1,\ldots,\theta_{d-2})\big)\right) g\big(f_{d-1}(\theta,\theta_1,\ldots,\theta_{d-2})\big)$$

$$\cdot J_{d-1}(\theta,\theta_1,\ldots,\theta_{d-2}) \, \mathrm{d}\theta_{d-2}\ldots \mathrm{d}\theta_1 \, \mathrm{d}\theta.$$

Applying (21) for σ_{d-1} we obtain

$$\int_0^\pi \int_{\mathbb{S}_x^{d-2}} \mathbb{1}_A\big(\gamma_{(x,v)}(\theta)\big) \mu_x(\mathrm{d}v) \, \mathrm{d}\theta = \int_{\mathbb{S}^{d-1}} \mathbb{1}_A\big(\widetilde{\varphi}_x(y)\big) g(y) \, \sigma_{d-1}(\mathrm{d}y).$$

Observe that $0 \le \sin^{d-2} \circ \arccos \le 1$, such that $g \ge 1$. Moreover, σ_{d-1} is invariant under the orthogonal map $\widetilde{\varphi}_x$. Hence,

$$\int_0^\pi \int_{\mathbb{S}_x^{d-2}} \mathbb{1}_A\big(\gamma_{(x,v)}(\theta)\big) \mu_x(\mathrm{d}v) \, \mathrm{d}\theta \ge \int_{\mathbb{S}^{d-1}} \mathbb{1}_A\big(\widetilde{\varphi}_x(y)\big) \sigma_{d-1}(\mathrm{d}y) = \sigma_{d-1}(A).$$

Finally, take into account that $\mathcal{U}_{\mathbb{S}^{d-1}} = \frac{1}{\omega_{d-1}} \sigma_{d-1}$, $\mathcal{U}_{\mathbb{S}_x^{d-2}} = \frac{1}{\omega_{d-2}} \mu_x$ and $\frac{\omega_{d-1}}{\omega_{d-2}} \ge \frac{\sqrt{2\pi}}{\sqrt{d-1}}$, cf. (Mathé and Novak, 2007, Lemma 6) as well as the formulas for $\omega_{d-1}, \omega_{d-2}$ in terms of the Gamma function. ∎

Appendix E. Proof of Proposition 15

Hasenpflug et al. (2024) provide a formal description, with a reversibility result, of the shrinkage scheme that we use in the proof of Proposition 15. For convenience of the reader we restate it and state the corresponding reversibility result.

To this end we call a set $S \in \mathcal{B}([0, 2\pi))$ *open on the circle* if for all $\theta \in S$ there exists $\varepsilon > 0$ such that $\{a \mod 2\pi \mid |a - \theta| < \varepsilon, a \in \mathbb{R}\} \subseteq S$. Moreover, for $a, b \in [0, 2\pi)$ define generalized intervals

$$
I(a,b) := \begin{cases} [0,b) \cup [a, 2\pi), & a > b \\ [a,b), & a < b \\ [0, 2\pi), & a = b, \end{cases} \qquad \bar{I}(a,b) := \begin{cases} [0,b) \cup [a, 2\pi), & a > b \\ [a,b), & a < b \\ \emptyset, & a = b, \end{cases}
$$

that appear in Algorithm 6, which is called as $\overline{\mathrm{shrink}}(\theta, S)$ for $S \in \mathcal{B}([0, 2\pi))$ being open on the circle and $\theta \in S$.

Algorithm 6 Algorithm 2.2 from Hasenpflug et al. (2024) with input $S \in \mathcal{B}([0, 2\pi))$ and $\theta \in S$, called by $\overline{\mathrm{shrink}}(\theta, S)$.

---

input: current state $\theta \in S$
output: step-size α

1: Set $i := 1$ and draw $\Lambda_i \sim \mathcal{U}_{(0, 2\pi)}$, call the result a_i.
2: Set $a_i^{\min} := a_i$ and $a_i^{\max} := a_i$.
3: **while** $a_i \notin S$ **do**
4:    **if** $a_i \in \bar{I}(a_i^{\min}, \theta)$ **then**
5:       Set $a_{i+1}^{\min} := a_i$ and $a_{i+1}^{\max} := a_i^{\max}$.
6:    **else**
7:       Set $a_{i+1}^{\min} = a_i^{\min}$ and $a_{i+1}^{\max} = a_i$.
8:    **end if**
9:    Draw $\Lambda_{i+1} \sim \mathcal{U}_{I(a_{i+1}^{\min}, a_{i+1}^{\max})}$, call the result a_{i+1}.
10:   Set $i := i + 1$.
11: **end while**
12: Return $\alpha := a_i$.

For a given set $S \in \mathcal{B}([0, 2\pi))$ that is open on the circle the transition kernel defined by Algorithm 6 is denoted by

$$
\bar{Q}_S : S \times \mathcal{B}(S) \to [0, 1], \tag{22}
$$

that is,

$$
\bar{Q}_S(\theta, A) = \mathbb{P}\big(\overline{\mathrm{shrink}}(\theta, S) \in A\big), \qquad \forall \theta \in [0, 2\pi), A \in \mathcal{B}([0, 2\pi)).
$$

Similarly as in Hasenpflug et al. (2024) observe that for $S = L(x, v, t)$ with $x \in \mathbb{S}^{d-1}$, $v \in \mathbb{S}_x^{d-2}$ and $t \in (0, p(x))$, holds

$$
\begin{aligned}
Q_{x,v,t}(\cdot) &= \mathbb{P}\big(\mathrm{shrink}(x, v, t) \in \cdot \mod 2\pi\big) = \mathbb{P}\big(\overline{\mathrm{shrink}}(0, L(x, v, t)) \in \cdot\big) \\
&= \bar{Q}_{L(x,v,t)}(0, \cdot),
\end{aligned} \tag{23}
$$

where the distribution $Q_{x,v,t}$ is provided in Definition 12. We state two useful properties of the kernel of the shrinkage procedure $\bar{Q}_S$ that are proven by Hasenpflug et al. (2024).

Lemma 23 (Hasenpflug et al., 2024, Theorem 2.10) *Let $S \in \mathcal{B}([0, 2\pi))$ be non-empty and open on the circle. Then, $\bar{Q}_S$ defined in (22) is reversible w.r.t. $\mathcal{U}_S$.*

Lemma 24 (Hasenpflug et al., 2024, Lemma 2.12) *Let $S \in \mathcal{B}([0, 2\pi))$ be non-empty and open on the circle, $\theta \in [0, 2\pi)$, and define $g_\theta : [0, 2\pi) \to [0, 2\pi)$ as $a \mapsto (\theta - a) \mod 2\pi$. Then*

$$\bar{Q}_{g_\theta^{-1}(S)} \left(g_\theta^{-1}(\alpha), g_\theta^{-1}(A) \right) = \bar{Q}_S(\alpha, A), \qquad \forall \alpha \in S, A \in \mathcal{B}(S).$$

Intuitively, the map g_θ from the previous lemma corresponds to following a line in reverse direction with an offset of θ. The lemma tells us if we apply this motion to all inputs of the shrinkage procedure simultaneously, its effects cancel. We formulate the consequences for our setting.

By Lemma 1 and the 2π-periodicity of sine and cosine, we have

$$\gamma_{(x,y)} \left(g_\theta(a) \right) = \gamma_{T_\theta(x,v)}(a), \qquad \forall \, x \in \mathbb{S}^{d-1}, v \in \mathbb{S}_x^{d-2}, \theta, a \in [0, 2\pi).$$

This implies for all $x \in \mathbb{S}^{d-1}, v \in \mathbb{S}_x^{d-2}, \theta \in [0, 2\pi), t \in (0, p(x))$ and $B \in \mathcal{B}(\mathbb{S}^{d-1})$ that

$$g_\theta^{-1} \left(L(x, v, t) \right) = L(T_\theta(x, v), t) \qquad \text{and} \qquad g_\theta^{-1} \left(\gamma_{(x,v)}^{-1}(B) \right) = \gamma_{T_\theta(x,v)}^{-1}(B).$$

Hence, by applying Lemma 24 for $A = \gamma_{(x,v)}^{-1}(B)$, $\alpha = \theta$ and $S = L(x, v, t)$, we obtain

$$\bar{Q}_{L(T_\theta(x,v),t)} \left(0, \gamma_{T_\theta(x,v)}^{-1}(B) \right) = \bar{Q}_{g_\theta^{-1}\left(L(x,v,t) \right)} \left(0, g_\theta^{-1} \left(\gamma_{(x,v)}^{-1}(B) \right) \right) = \bar{Q}_{L(x,v,t)}(\theta, \gamma_{(x,v)}^{-1}(B)). \tag{24}$$

To prove Proposition 15 we show reversibility of $\widetilde{H}_t$ w.r.t. $\mathcal{U}_{L(t)}$ for all $t \in (0, \|p\|)_\infty$ and then conclude the assertion by Lemma 20. To achieve this, we use the invariance of the Liouville measure under T_θ (which corresponds to "a forward move with a U-turn"). This boils down to starting the shrinkage procedure at a random point on the geodesic level set.

Proof *(Proof of Proposition 15.)* Let $t \in (0, \|p\|_\infty)$ and $A, B \in \mathcal{B}(\mathbb{S}^{d-1})$. Due to the lower semicontinuity of p, the level set $L(t)$ is open. Thus, $L(x, v, t)$ is non-empty and open on the circle for all $x \in L(t)$ with $v \in \mathbb{S}_x^{d-2}$. Moreover, Lemma 3 implies that $\lambda(L(x, v, t)) > 0$ for all $x \in L(t)$ and $v \in \mathbb{S}_x^{d-2}$. Hence, we have, exploiting (23), that

$$\omega_{d-2} \int_{B \cap L(t)} \widetilde{H}_t(x, A) \, \sigma_{d-1}(\mathrm{d}x) = \int_{B \cap L(t)} \int_{\mathbb{S}_x^{d-2}} \bar{Q}_{L(x,v,t)}(0, \gamma_{(x,v)}^{-1}(A)) \, \mu_x(\mathrm{d}v) \, \sigma_{d-1}(\mathrm{d}x)$$

$$= \int_{[0,2\pi)} \int_{B \cap L(t)} \int_{\mathbb{S}_x^{d-2}} \frac{\mathbb{1}_{L(x,v,t)}(\theta)}{\lambda(L(x, v, t))} \bar{Q}_{L(x,v,t)}(0, \gamma_{(x,v)}^{-1}(A)) \mu_x(\mathrm{d}v) \, \sigma_{d-1}(\mathrm{d}x) \, \mathrm{d}\theta.$$

Then, Lemma 2 yields

$$\omega_{d-2} \int_{B \cap L(t)} \widetilde{H}_t(x, A) \, \sigma_{d-1}(\mathrm{d}x) = \int_{[0,2\pi)} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}_x^{d-2}} \mathbb{1}_{B \cap L(t)} \left(\gamma_{(x,v)}(\theta) \right) \frac{\mathbb{1}_{L(T_\theta(x,v),t)}(\theta)}{\lambda(L(T_\theta(x, v), t))}$$

$$\times \bar{Q}_{L(T_\theta(x,v),t)}(0, \gamma_{T_\theta(x,v)}^{-1}(A)) \, \mu_x(\mathrm{d}v) \, \sigma_{d-1}(\mathrm{d}x) \, \mathrm{d}\theta.$$

Using (24), (5), Lemma 1 and Lemma 22, we obtain

$$
\omega_{d-2} \int_{B \cap L(t)} \widetilde{H}_t(x, A) \, \sigma_{d-1}(\mathrm{d}x)
$$

$$
= \int_{[0,2\pi)} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}_x^{d-2}} \mathbb{1}_B(\gamma_{(x,v)}(\theta)) \frac{\mathbb{1}_{L(x,v,t)}(\theta) \mathbb{1}_{L(t)}(x)}{\lambda(L(x,v,t))} \bar{Q}_{L(x,v,t)}(\theta, \gamma_{(x,v)}^{-1}(A)) \mu_x(\mathrm{d}v) \sigma_{d-1}(\mathrm{d}x) \mathrm{d}\theta
$$

$$
= \int_{L(t)} \int_{\mathbb{S}_x^{d-2}} \frac{1}{\lambda(L(x,v,t))} \int_{\gamma_{(x,v)}^{-1}(B) \cap L(x,v,t)} \bar{Q}_{L(x,v,t)}(\theta, \gamma_{(x,v)}^{-1}(A)) \mathrm{d}\theta \, \mu_x(\mathrm{d}v) \, \sigma_{d-1}(\mathrm{d}x).
$$

As $L(x, v, t)$ is open on the circle and non-empty for all $x \in L(t)$ and $v \in \mathbb{S}_x^{d-2}$, we may apply Lemma 23 and get

$$
\omega_{d-2} \int_{B \cap L(t)} \widetilde{H}_t(x, A) \, \sigma_{d-1}(\mathrm{d}x)
$$

$$
= \int_{L(t)} \int_{\mathbb{S}_x^{d-2}} \frac{1}{\lambda(L(x,v,t))} \int_{\gamma_{(x,v)}^{-1}(A) \cap L(x,v,t)} \bar{Q}_{L(x,v,t)}(\theta, \gamma_{(x,v)}^{-1}(B)) \mathrm{d}\theta \, \mu_x(\mathrm{d}v) \, \sigma_{d-1}(\mathrm{d}x).
$$

Performing the same arguments in reversed order, we get

$$
\omega_{d-2} \int_{B \cap L(t)} \widetilde{H}_t(x, A) \, \sigma_{d-1}(\mathrm{d}x) = \omega_{d-2} \int_{A \cap L(t)} \widetilde{H}_t(x, B) \, \sigma_{d-1}(\mathrm{d}x).
$$

Thus, by Lemma 20, we obtain reversibility of $\widetilde{H}$ with respect to π. ∎

Appendix F. Random-walk Metropolis-Hastings on the Sphere

The random-walk Metropolis-Hastings (RWMH) algorithm uses an isotropic Gaussian proposal kernel. As suggested by Lie et al. (2023), for current state $x \in \mathbb{S}^{d-1}$ we choose an auxiliary point $\sqrt{r}x$ in the ambient space $\mathbb{R}^d$ by generating a radius $\sqrt{r}$, with r being a realization of $R \sim \text{Gamma}(d/2, 1/2)$. Then, given $\sqrt{r}x$ we sample a realization y from $Y \sim \text{Normal}\left(\sqrt{r}x, \varepsilon^2 \text{Id}_d\right)$, where ε is the step-size of the random walk. Since this y does not yet lie on the sphere, we radially project and propose $y/\|y\|$, which finally is accepted or rejected using the usual acceptance ratio. In Algorithm 7 we provide the corresponding pseudocode. Note that we also introduce mixture-MH, which can be interpreted as a variant that combines RWMH with independent Metropolis. For details we refer to Algorithm 8, which is specifically employed for testing the mixture of vMF distributions.

Appendix G. Hamiltonian Monte Carlo on the Sphere

Since RWMH suffers from diffusive behavior, we also test spherical Hamiltonian Monte Carlo (HMC) suggested by Lan et al. (2014) as an alternative MCMC approach. In spherical HMC, the sample space is first augmented by *momenta* or *velocities* $v \in \mathbb{R}^d$ that are in the the tangent space E_x to $\mathbb{S}^{d-1}$ at the current sample x. The momenta follow a standard Normal distribution, and the Markov chain is generated in the space $\{(x, v) \mid x \in \mathbb{S}^{d-1}, v \in E_x\}$. Diffusive behavior is suppressed by the special type of proposal step that solves Hamilton's equations of motion by using a leapfrog integrator (Neal, 2011). During leapfrog integration, the gradient of the log probability $\nabla \log p(x)$ guides the Markov chain, thereby reaching

---

Algorithm 7 Reprojected RWMH on $\mathbb{S}^{d-1}$ for step-size $\varepsilon > 0$.

---

input: current state $x \in \mathbb{S}^{d-1}$
output: next state x'

  1: Draw $R \sim \mathrm{Gamma}\,(d/2, 1/2)$, call the result r.
  2: Draw $Y \sim \mathrm{Normal}\,\big(\sqrt{r}\,x, \varepsilon^2 \mathrm{Id}_d\big)$, call the result y.
  3: Set $z = y/\|y\|$.
  4: Draw $U \sim \mathcal{U}_{(0,1)}$, call the result u.
  5: **if** $u \leq \min\{1, p(z)/p(x)\}$ **then**
  6:    $x' = z$
  7: **else**
  8:    $x' = x$
  9: **end if**

Algorithm 8 Mixture of RWMH and Independence Sampler on $\mathbb{S}^{d-1}$ with mixture hyper-parameter $\alpha \in [0, 1]$ and step-size $\varepsilon > 0$.

---

input: current state $x \in \mathbb{S}^{d-1}$
output: next state x'

  1: Draw $U_1 \sim \mathcal{U}_{(0,1)}$, call the result u_1.
  2: **if** $u_1 < \alpha$ **then**
  3:    // Use random walk proposal
  4:    Draw $R \sim \mathrm{Gamma}\,(d/2, 1/2)$, call the result r.
  5:    Draw $Y \sim \mathrm{Normal}\,\big(\sqrt{r}\,x, \varepsilon^2 \mathrm{Id}_d\big)$, call the result y.
  6:    Set $z = y/\|y\|$.
  7: **else**
  8:    // Use independent proposal
  9:    Draw $Y \sim \mathrm{Normal}\,(0, \mathrm{Id}_d)$, call the result y.
10:    Set $z = y/\|y\|$.
11: **end if**
12: Draw $U_2 \sim \mathcal{U}_{(0,1)}$, call the result u_2.
13: **if** $u_2 \leq \min\{1, p(z)/p(x)\}$ **then**
14:    $x' = z$
15: **else**
16:    $x' = x$
17: **end if**

---

nearby modes in much shorter time than RWMH. The spherical HMC algorithm is detailed in Algorithm 9. In contrast to standard HMC, spherical HMC involves a rotation of the positions and momenta during leapfrog integration where the rotation matrix is a Givens rotation $G(v/\|v\|, x, \theta)$ in the plane spanned by the momentum v and the position x (see lines 8–10 in Algorithm 9 and Equation 4). In addition to the step-size parameter $\varepsilon > 0$, we also need to choose the number of leapfrog steps $T \in \mathbb{N}$. In our experiments, we always set $T = 10$.

Appendix H. Step-size Tuning for RWMH and HMC

Both RWMH and spherical HMC involve a step-size parameter ε. Because a good choice of ε depends on the particular shape of our target distribution, we first find ε automatically during a burn-in phase. During burn-in, we increase ε by a factor 1.02, if the proposal (based on the current value of ε) is accepted. We decrease ε by a factor 0.98 if the proposed state is rejected. After a burn-in phase, the value of ε is kept fixed.

Algorithm 9 Hamiltonian Monte Carlo on $\mathbb{S}^{d-1}$ for step-size $\varepsilon > 0$ and number of integration steps T.

---

input: current state $x \in \mathbb{S}^{d-1}$
output: next state x'

1: Draw $V \sim \text{Normal}(0, \text{Id}_d)$, call the result v.
2: Set $v_1 = (\text{Id}_d - xx^T)\,v$, $x_1 = x$ and $t = 1$.
3: **while** $t \leq T$ **do**
4:    Set $v_{t+1/2} = v_t + \frac{\varepsilon}{2}\left(\text{Id}_d - x_t x_t^T\right) \nabla \log p(x_t)$.
5:    Set $R = G(v_{t+1/2}/\|v_{t+1/2}\|, x_t, \varepsilon\,\|v_{t+1/2}\|)$.
6:    Set $x_{t+1} = R\,x_t$.
7:    Set $v_{t+1} = R\,v_{t+1/2} + \frac{\varepsilon}{2}\left(\text{Id}_d - x_{t+1} x_{t+1}^T\right) \nabla \log p(x_{t+1})$.
8:    Set $t := t + 1$.
9: **end while**
10: Draw $U \sim \mathcal{U}_{(0,1)}$, call the result u.
11: **if** $u < \min\left\{1, \exp\left(\|v_1\|^2/2 - \|v_{T+1}\|^2/2\right)\,p(x_{T+1})/p(x)\right\}$ **then**
12:    Set $x' = x_{T+1}$.
13: **else**
14:    Set $x' = x$.
15: **end if**

Appendix I. Run Times of Numerical Illustrations

We summarize the run times in seconds (s) for all the numerical experiments tested on Intel i7-10510U CPU in Table A1. In addition to the tests reported in Section 4, we provide a Jupyter notebook for testing the geodesic slice samplers on Bingham targets in various dimensions.[4] We also report run times for these targets. For brevity we name these experiments "Bingham-d10" for the Bingham distribution with $d = 10$, $\kappa_d = 30$ and $N = 10^5$; "Bingham-d50" for the Bingham distribution with $d = 50$, $\kappa_d = 300$ and $N = 10^5$; "vMF-d10" for the mixture of vMF distribution with $d = 10$, $\kappa = 100$, $K = 5$ and $N = 10^6$; "curved-vMF-d3" for the curve vMF distribution with $d = 3$, $\kappa = 500$ and $N = 10^6$; "curved-vMF-d5" for curve vMF with $d = 5$, $\kappa = 800$ and $N = 10^6$.

---

[4] `github.com/microscopic-image-analysis/geosss/scripts/Bingham.ipynb`

Experiments	geoSSS (reject)	geoSSS (shrink)	RWMH	HMC
Bingham-d10	13.9 s	8.5 s	3.5 s	28.4 s
Bingham-d50	36.5 s	13.1 s	3.5 s	32.0 s
vMF-d10[*]	2601.6 s	812.9 s	274.4 s	1898.9 s
curve-vMF-d3	17106.9 s	5583.2 s	1841.0 s	12292.9 s
curve-vMF-d5	42483.7 s	5509.0 s	1355.7 s	8659.0 s

[*] The extensively tuned mixture-MH sampler employed only for this test took 577.0 s.

Table A1: Comparison of run times for each MCMC method across different numerical experiments.

References

Claude J. P. Bélisle, H. Edwin Romeijn, and Robert L. Smith. Hit-and-run algorithms for generating multivariate distributions. *Mathematics of Operations Research*, 18(2): 255–266, 1993.

Cameron Bell, Krzystof Łatuszyński, and Gareth O. Roberts. Adaptive stereographic MCMC. arXiv:2408.11780, 2025.

Alexandros Beskos and Kengo Kamatani. MCMC algorithms for posteriors on matrix spaces. *Journal of Computational and Graphical Statistics*, 31(3):721–738, 2022.

Marcus Brubaker, Mathieu Salzmann, and Raquel Urtasun. A family of MCMC methods on implicitly defined manifolds. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics*, volume 22, pages 161–172, 2012.

Simon Byrne and Mark Girolami. Geodesic Monte Carlo on embedded manifolds. *Scandinavian Journal of Statistics*, 40(4):825–845, 2013.

Isaac Chavel. *Eigenvalues in Riemannian geometry*. Academic Press, Orlando, 1984.

Wallace Givens. Computation of plain unitary rotations transforming a general matrix to triangular form. *Journal of the Society for Industrial and Applied Mathematics*, 6(1): 26–50, 1958.

Navin Goyal and Abhishek Shetty. Sampling and optimization on convex sets in Riemannian manifolds of non-negative curvature. In *Proceedings of the 32nd Conference on Learning Theory*, volume 99, pages 1519–1561, 2019.

Andrew J. Hanson. Rotations for N-dimensional graphics. In Alan W. Paeth, editor, *Graphics Gems V*, pages 55–64. Academic Press, Boston, 1995.

Mareike Hasenpflug, Viacheslav Telezhnikov, and Daniel Rudolf. Reversibility of elliptical slice sampling revisited. *Bernoulli*, 31(2):1377–1401, 2024.

Andrew Holbrook, Shiwei Lan, Jeffrey Streets, and Babak Shahbaba. Nonparametric Fisher geometry with application to density estimation. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, volume 124, pages 101–110, 2020.

Berthold K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 4(4):629–642, 1987.

Marc Kac. Foundations of kinetic theory. In *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, volume 3, pages 171–197. 1956.

Shiwei Lan, Bo Zhou, and Babak Shahbaba. Spherical Hamiltonian Monte Carlo for constrained target distributions. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32, pages 629–637, 2014.

Krzysztof Łatuszyński and Daniel Rudolf. Convergence of hybrid slice sampling via spectral gap. *Advances in Applied Probability*, 56(4):1440–1466, 2024.

Yin Tat Lee and Santosh S. Vempala. Convergence rate of Riemannian Hamiltonian Monte Carlo and faster polytope volume computation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, page 1115–1121, 2018.

Han Cheng Lie, Daniel Rudolf, Björn Sprungk, and T. J. Sullivan. Dimension-independent Markov chain Monte Carlo on the sphere. *Scandinavian Journal of Statistics*, 5(4):1818–1858, 2023.

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

Oren Mangoubi and Aaron Smith. Rapid mixing of geodesic walks on manifolds with positive curvature. *The Annals of Applied Probability*, 28(4):2501–2543, 2018.

Kanti V. Mardia and Peter E. Jupp. *Directional statistics*. John Wiley & Sons, Ltd., Chichester, 2000.

Augustin Marignier, Jason D. McEwen, Ana M. G. Ferreira, and Thomas D. Kitching. Posterior sampling for inverse imaging problems on the sphere in seismology and cosmology. *RAS Techniques and Instruments*, 2(1):20–32, 2023.

Peter Mathé and Erich Novak. Simple Monte Carlo and the Metropolis algorithm. *Journal of Complexity*, 23(4):673–696, 2007.

Sean P. Meyn and Richard L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, 2nd edition, 2009.

James R. Munkres. *Analysis on manifolds*. Addison-Wesley Publishing Company, Redwood City, 1991.

Iain Murray, Ryan Adams, and David MacKay. Elliptical slice sampling. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, volume 9, pages 541–548, 2010.

Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2262–2275, 2010.

Viacheslav Natarovskii, Daniel Rudolf, and Björn Sprungk. Geometric convergence of elliptical slice sampling. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 7969–7978, 2021a.

Viacheslav Natarovskii, Daniel Rudolf, and Björn Sprungk. Quantitative spectral gap estimate and Wasserstein contraction of simple slice sampling. *The Annals of Applied Probability*, 31(2):806 – 825, 2021b.

Radford M. Neal. Slice sampling. *The Annals of Statistics*, 31(3):705–767, 2003.

Radford M. Neal. MCMC using Hamiltonian dynamics. In Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng, editors, *Handbook of Markov chain Monte Carlo*, pages 113–162. CRC Press, Boca Raton, 2011.

Gabriel P. Paternain. *Geodesic flows.* Birkhäuser, Boston, 1999.

Natesh S. Pillai and Aaron Smith. Kac's walk on n-sphere mixes in $n \log n$ steps. *The Annals of Applied Probability*, 27(1):631–650, 2017.

Gareth O. Roberts and Jeffrey S. Rosenthal. Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(3):643–660, 1999.

Edward B. Saff and Amo B. J. Kuijlaars. Distributing many points on a sphere. *The Mathematical Intelligencer*, 19:5–11, 1997.

René L. Schilling. *Measures, integrals and martingales.* Cambridge University Press, Cambridge, 2nd edition, 2017.

Simon Schwarz, Michael Herrmann, Anja Sturm, and Max Wardetzky. Efficient random walks on Riemannian manifolds. *Foundations of Computational Mathematics*, 25:145–161, 2025.

Julian Straub, Trevor Campbell, Jonathan P. How, and John W. Fisher III. Efficient global point cloud alignment using Bayesian nonparametric mixtures. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2403–2412, 2017.

Jun Yang, Krzysztof Łatuszyński, and Gareth O. Roberts. Stereographic Markov chain Monte Carlo. *The Annals of Statistics*, 52(6), 2024.

Emilio Zappa, Miranda Holmes-Cerfon, and Jonathan Goodman. Monte Carlo on manifolds: Sampling densities and integrating functions. *Communications on Pure and Applied Mathematics*, 71(12):2609–2647, 2018.