Received: 25 February 2022 | Revised: 23 March 2023 | Accepted: 6 April 2023

DOI: 10.1111/sjos.12653

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Dimension-independent Markov chain Monte Carlo on the sphere

Han Cheng Lie[1] | **Daniel Rudolf**[2] | **Björn Sprungk**[3] | **T. J. Sullivan**[4,5]

[1]Institut für Mathematik, Universität Potsdam, Potsdam, Germany

[2]Fakultät für Informatik und Mathematik, Universität Passau, Passau, Germany

[3]Fakultät für Informatik und Mathematik, Technische Universität Bergakademie Freiberg, Freiberg, Germany

[4]Mathematics Institute and School of Engineering, University of Warwick, Coventry, UK

[5]Alan Turing Institute, London, UK

Correspondence
Han Cheng Lie, Institut für Mathematik, Universität Potsdam, Campus Golm, Karl-Liebknecht-Str. 24-25, Potsdam 14476, Germany.
Email: hanlie@uni-potsdam.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Numbers: 318763901, 389483880, 390685689, 415980428, 432680300; Excellence Initiative; Felix Bernstein Institute for Mathematical Statistics in the Biosciences; Freie Universität Berlin; Zuse Institut Berlin

Abstract

We consider Bayesian analysis on high-dimensional spheres with angular central Gaussian priors. These priors model antipodally symmetric directional data, are easily defined in Hilbert spaces and occur, for instance, in Bayesian density estimation and binary level set inversion. In this paper we derive efficient Markov chain Monte Carlo methods for approximate sampling of posteriors with respect to these priors. Our approaches rely on lifting the sampling problem to the ambient Hilbert space and exploit existing dimension-independent samplers in linear spaces. By a push-forward Markov kernel construction we then obtain Markov chains on the sphere which inherit reversibility and spectral gap properties from samplers in linear spaces. Moreover, our proposed algorithms show dimension-independent efficiency in numerical experiments.

KEYWORDS

dimension independence, directional statistics, high-dimensional manifolds, level set inversion, Markov chain Monte Carlo

1 | INTRODUCTION

The Markov chain Monte Carlo (MCMC) method is a standard tool for computational probability and recent years have seen increasing interest in *dimension-independent* MCMC schemes, that is,

those whose statistical efficiency and mixing rates do not degenerate to zero as the dimension of the sample space tends to infinity. We mention here the preconditioned Crank–Nicolson (pCN) scheme of Cotter et al. (2013)—see also Neal (1999, equation (15)) and Beskos et al. (2008)—and the elliptical slice sampler (ESS) of Murray et al. (2010), both of which rely on a Gaussian reference or prior measure. Recently, the pCN scheme has been combined with geometric MCMC methods (Beskos et al., 2017; Rudolf & Sprungk, 2018) and extended to classes of non-Gaussian priors (Chen et al., 2018), and for the ESS geometric ergodicity was shown by Natarovskii et al. (2021a).

In this work, we study whether these dimension-independent sampling schemes could also be modified for Bayesian analysis on high-dimensional manifolds. As a starting point we focus on Bayesian inference on high-dimensional spheres (Watson, 1983). We then consider the case of the unit sphere in a general Hilbert space. We formulate some results in more generality, for example, by replacing the ambient Hilbert space and the unit sphere with a pair of topological spaces that are related by a measurable mapping. This allows us to extend some of our results to manifolds that are more general than the unit sphere. Our choice of the sphere is further motivated by particular inverse problems on function spaces such as level set inversion—more precisely, *binary classification*—where one is essentially interested only in recovering the pointwise *sign* of a function $u : D \to \mathbb{R}$ on some domain D. Thus, u and αu for $\alpha > 0$ yield equivalent classifications and, hence, it is natural to consider the inverse problem just on some unit sphere of functions.

Previous works on MCMC methods on manifolds—such as those of Brubaker et al. (2012), Byrne and Girolami (2013), Diaconis et al. (2013), Mangoubi and Smith (2018), and Zappa et al. (2018)—derive algorithms which are based on the Hausdorff or surface measure as reference measure. However, despite their use of geometric structure, the performance of such methods typically still degrades as the dimension of the sample space increases to infinity—one reason being the degeneration of the target density with respect to the Hausdorff measure.

1.1 | Contribution

In this paper, we aim to construct dimension-independent MCMC methods in order to sample efficiently from target measures on high-dimensional spheres. We identified the angular central Gaussian (ACG) distribution as a suitable reference measure for this purpose. The ACG models antipodally symmetric directional data and is an alternative to the Bingham distribution (Tyler, 1987). ACG distributions and their mixtures have been applied in finite-dimensional directional-statistical problems such as geomagnetism (Tyler, 1987, section 8), imaging in neuroscience (Tabelow et al., 2012), and materials science (Franke et al., 2016, section 4). ACG distributions have been generalized to the projected normal distribution, for which the initial Gaussian distribution may have nonzero mean (Wang & Gelfand, 2013).

The ACG distribution is defined as the radial projection onto the sphere of a centered Gaussian measure on the ambient Hilbert space and thus yields a well-defined reference measure even in infinite-dimensional Hilbert spaces. Moreover, the ACG distribution can be applied in an acceptance-rejection method for sampling from several families of distributions on spheres and similar manifolds (Kent et al., 2018). Thus, we anticipate that our proposed methods could also be exploited for dimension-independent MCMC for posteriors with other priors, for example, Bingham, Fisher–Bingham, or von Mises–Fisher priors. However, we leave this question for future research, and focus on posteriors given with respect to the ACG prior in this paper.

The particular structure of the ACG prior allows us to lift the sampling problem to the ambient Hilbert space. Thus, we can exploit existing dimension-independent MCMC algorithms on linear spaces, for example, the pCN algorithm mentioned earlier. In order to obtain Markov chains on

the sphere, we use *push-forward Markov kernels* as introduced by Rudolf and Sprungk (2022). This approach then yields specific MCMC algorithms that first draw from a suitable distribution a point on the ray defined by the current position on the sphere, and then take a step using a dimension-independent transition kernel. The resulting state is finally "reprojected" to the sphere.

In summary, our contributions are as follows:

(1) We propose two easily implementable MCMC algorithms that generate reversible Markov chains on high-dimensional spheres, where the chains have as their invariant distribution a given posterior with respect to an ACG prior;
(2) We prove uniform ergodicity of the suggested Markov chains in finite-dimensional settings;
(3) We provide theoretical and numerical evidence for dimension-independent statistical efficiency of the proposed algorithms.

Moreover, our numerical experiments show that some other existing MCMC methods for sampling on manifolds—see Section 1.3 for an overview—exhibit decreasing statistical efficiency as the state space dimension increases. Thus, we provide a first contribution to dimension-independent MCMC on manifolds, and thereby demonstrate the feasibility of efficient Bayesian analysis on high-dimensional spheres.

1.2 | Outline

The remainder of this paper is structured as follows. Section 1.3 overviews some related work in this area and Section 1.4 sets out some basic notation. In Section 2 we recall two basic MCMC algorithms that are valid in infinite-dimensional Hilbert spaces. Basic definitions and properties related to MCMC, in particular the Metropolis–Hastings (MH) and slice sampling paradigms, are provided in Appendix B for completeness. In Section 3 we make our main theoretical contributions by developing a general framework for obtaining dimension-independent MCMC methods on manifolds. In particular, we derive and analyse two sampling methods on the sphere. These methods are subjected to numerical tests, in the context of Bayesian binary classification and density estimation, in Section 4. Some closing remarks are given in Section 5. In the appendix we further recall some key facts about Gaussian and ACG measures (Appendix A), describe related existing MCMC algorithms on the sphere (Appendix C) and provide technical auxiliary results (Appendix D).

1.3 | Overview of related work

Classical references that treat statistical inference on the sphere include those of Watson (1983), who focuses exclusively on spheres in finite-dimensional Euclidean spaces, and Mardia and Jupp (2000), who focus on circular data but also treat spheres, Stiefel and Grassmann manifolds, and general manifolds. Special manifolds such as the Stiefel and Grassmann manifolds have also been studied by Chikuse (2003). A recent treatment that focuses on modern developments in directional statistics is given by Ley and Verdebout (2017). Srivastava and Jermyn (2009) consider the infinite-dimensional unit sphere of diffeomorphisms in the context of computer vision, and then apply a Bayesian method for shape identification. However, none of the cited works treat MCMC sampling methods or Bayesian inference on high-dimensional manifolds.

Regarding sampling on embedded manifolds, Hamiltonian Monte Carlo methods are considered by Brubaker et al. (2012) and Byrne and Girolami (2013), for instance, and Diaconis

et al. (2013) propose a Gibbs sampler. Moreover, Mangoubi and Smith (2018) study the so-called "geodesic walk" algorithm and establish Wasserstein contraction under the assumption that the manifold has bounded, positive curvature. The geodesic walk algorithm of Mangoubi and Smith (2018) chooses a random element uniformly from the unit sphere in the tangent space and moves a fixed time or step size along the corresponding geodesic. This could be used as a proposal in an MH algorithm. Similarly, Zappa et al. (2018) developed an MH algorithm on manifolds where the proposed point is generated by a normally distributed tangential move into the ambient Euclidean space which is then suitably projected back to the manifold. We will compare our algorithms particularly to that of Zappa et al. (2018) and the geodesic walk algorithm of Mangoubi and Smith (2018). For the specific problem of designing MCMC samplers on the sphere, Lan et al. (2014) considered Hamiltonian Monte Carlo for distributions that undergo several transformations in order to be defined on the unit sphere. Their approach has been used by Holbrook et al. (2020) to perform Bayesian nonparametric density estimation based on the Bingham distribution as the prior.

We also mention the work of Yang et al. (2022), which considers high-dimensional MCMC methods for sampling from heavy-tailed distributions. Their work uses stereographic projection to the sphere to prove desirable mixing properties for the resulting MCMC samplers as the dimension increases. Two important differences between their work and our work are that they focus on sampling from heavy-tailed distributions on Euclidean spaces, while we consider sampling from the sphere in general Hilbert spaces and focus on the ACG prior.

1.4 | Preliminaries and notation

Throughout, $(\Omega, \mathcal{A}, \mathbb{P})$ will be a fixed probability space, which we assume to be rich enough to serve as a common domain of definition for all random variables under consideration.

Given a topological space $\mathbb{X}$, $\mathcal{P}(\mathbb{X})$ denotes the space of probability measures on the Borel σ-algebra $\mathcal{B}(\mathbb{X})$ of $\mathbb{X}$. Given another topological space $\mathbb{Y}$, $T_\sharp\mu \in \mathcal{P}(\mathbb{Y})$ denotes the *push-forward* or *image measure* of $\mu \in \mathcal{P}(\mathbb{X})$ under a measurable map $T : \mathbb{X} \to \mathbb{Y}$, that is,

$$(T_\sharp\mu)(E) := \mu(T^{-1}(E)) \equiv \mu(\{x \in \mathbb{X} | T(x) \in E\}) \quad \text{for each } E \in \mathcal{B}(\mathbb{Y}). \tag{1}$$

The range of a map T is denoted $\mathrm{ran}(T)$. Throughout this paper, we use 'measurability' to refer to Borel measurability of a mapping between topological spaces or Borel measurability of a subset.

The absolute continuity of one measure $\mu \in \mathcal{P}(\mathbb{X})$ with respect to another measure ν will be denoted by $\mu \ll \nu$.

We denote the s-dimensional Hausdorff measure by $\mathcal{H}^s$. If E is an s-dimensional measurable set, then $\mathcal{H}_E$ denotes the restriction of $\mathcal{H}^s$ to E.

We denote the uniform distribution on a bounded subset $G \subset \mathbb{R}^d$ by $\mathrm{U}[G]$ and the normal distribution with mean element m and covariance operator C by $\mathrm{N}(m, C)$. For the convenience of the reader we provide a short overview of Gaussian measures on Hilbert spaces in Appendix A.1.

2 | MCMC IN HILBERT SPACES

We consider the case in which $\mathbb{X}$ is a separable Hilbert space $\mathbb{H}$ and the target or posterior distribution $\nu \in \mathcal{P}(\mathbb{H})$ is determined by a density with respect to a mean-zero Gaussian reference or

Algorithm 1. pCN-MH algorithm on $\mathbb{H}$

Given: prior $\nu_0 = \mathrm{N}(0, C)$ and target ν as in (2)
Initial: step size $s \in (0, 1]$ and state $x_0 \in \mathbb{H}$
for $k \in \mathbb{N}_0$ **do**
  Draw a sample w_k of $\mathrm{N}(0, C)$ and set $y_{k+1} := \sqrt{1 - s^2} x_k + s w_k$
  Compute $\alpha := \min\{1, \exp(\Phi(x_k) - \Phi(y_{k+1}))\}$
  Draw a sample u of $\mathrm{U}[0, 1]$
  if $u \leqslant \alpha$ **then**
    Set $x_{k+1} = y_{k+1}$
  else
    Set $x_{k+1} = x_k$
  end if
end for

prior measure $\nu_0 = \mathrm{N}(0, C)$ with covariance operator C via

$$\frac{\mathrm{d}\nu}{\mathrm{d}\nu_0}(x) \propto \exp(-\Phi(x)), \quad \nu_0\text{-a.e. } x \in \mathbb{H}, \tag{2}$$

with measurable $\Phi : \mathbb{H} \to \mathbb{R}$ satisfying

$$\int_{\mathbb{H}} \exp(-\Phi(x))\ \nu_0(\mathrm{d}x) < \infty.$$

In this setting we state two popular approaches for generating ν-reversible Markov chains $(X_k)_{k \in \mathbb{N}}$. The first is the pCN-MH algorithm (Algorithm 1; Cotter et al., 2013; Neal, 1999).

Here a possible new state y_{k+1} of the Markov chain given the current state $X_k = x_k$ is drawn according to the pCN-MH proposal kernel $Q(x_k, \cdot)$ where

$$Q(x, \mathrm{d}y) := \mathrm{N}\left(\sqrt{1 - s^2} x, s^2 C\right),$$

with $s \in (0, 1]$ denoting a step size parameter. The state y_{k+1} is accepted as the new state X_{k+1} only with probability $\alpha(x_k, y_{k+1})$, where the acceptance probability function α is given by

$$\alpha(x, y) := \min\{1, \exp(\Phi(x) - \Phi(y))\};$$

otherwise, the Markov chain remains at $X_{k+1} := x_k$. Algorithm 1 describes how to realise a Markov chain with pCN-MH transition kernel.

Next, we consider the ESS algorithm suggested by Murray et al. (2010). In this reference it is stated in a finite-dimensional setting, but the ESS algorithm can be lifted also to infinite-dimensional settings. Given $X_k = x_k$ we first choose a slice $\mathbb{H}_t := \{x \in \mathbb{H} : \exp(-\Phi(x)) \geqslant t\}$ at random by drawing t according to $t \sim \mathrm{U}[0, \exp(-\Phi(x_k))]$. We then sample a new state $X_{k+1} = y$ where $y \in \mathbb{H}_t$ according to the restriction of $\nu_0 = \mathrm{N}(0, C)$ to $\mathbb{H}_t$. In order to achieve the second step in an approximate way, the ESS employs a certain transition mechanism using randomly drawn ellipses in $\mathbb{H}$ and a shrinkage procedure. We state this transition in Algorithm 3, which we call *shrink-ellipse* (x, t). Thus, the ESS sampler is a hybrid slice sampler. Its algorithmic realisation is described in Algorithm 2.

Algorithm 2. ESS algorithm on $\mathbb{H}$

Given: prior $\nu_0 = \mathrm{N}(0, C)$ and target ν as in (2)
Initial: state $x_0 \in \mathbb{H}$
for $k \in \mathbb{N}_0$ **do**
  Draw sample $t \sim \mathrm{U}[0, \exp(-\Phi(x_k))]$
  Set $x_{k+1} = \text{shrink-ellipse}(x_k, t)$ (see Algorithm 3)
end for

Algorithm 3. Elliptical shrinkage (shrink-ellipse(x, t))

Input: state $x \in \mathbb{H}$ and level $t \in (0, \infty)$
Output: state y in level set $\mathbb{H}_t$
Draw a sample $w \sim \mathrm{N}(0, C)$
Draw a sample $\theta \sim \mathrm{U}[0, 2\pi]$
Set $\theta_{\min} = \theta - 2\pi$ and $\theta_{\max} = \theta$
while $\exp(-\Phi(y)) < t$ **do**
  if $\theta < 0$ **then**
    Set $\theta_{\min} = \theta$
  else
    Set $\theta_{\max} = \theta$
  end if
  Draw a sample $\theta \sim \mathrm{U}[\theta_{\min}, \theta_{\max}]$
  Set $y = \cos(\theta)x + \sin(\theta)w$
end while

It can be shown that the transition kernel of the ESS sampler has ν as its invariant distribution; see Murray et al. (2010) and Hasenpflug et al. (2023, theorem 3.2) for further details. For a more comprehensive introduction to MCMC and the MH and slice sampling approaches, see Appendix B.

Remark 1. As noted by Murray et al. (2010), both the ESS algorithm and the pCN algorithm draw proposal states from ellipses that are accepted or rejected. In the pCN algorithm, the random proposal X' satisfies $X' = \sqrt{1 - s^2}x + sW$, where $W \sim \mathrm{N}(0, C)$. For a fixed realization w of W and for varying $s \in (0, 1)$, the set $\{\sqrt{1 - s^2}x + sw | s \in (0, 1)\}$ is half of the ellipse passing through x and w centred at the origin, since $\sqrt{1 - s^2}x + sw = \cos(\theta)x + \sin(\theta)w$ for $\theta = \arcsin(s)$. In the elliptical slice sampling algorithm, a full ellipse instead of a half ellipse is used, thus providing a larger set of potential proposal states. Moreover, one never remains at the current state. Intuitively, using a larger set of potential proposal states might lead to faster convergence, as measured by the number of Markov chain steps.

3 | MCMC ON THE SPHERE

In this section, we construct and analyse MCMC algorithms for approximate sampling from a probability distribution μ on a high-dimensional unit sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ where μ admits a density

with respect to an ACG reference or prior measure μ_0. The ACG measure is given as follows. Consider the unit sphere $\mathbb{S} := \{x \in \mathbb{H} | \, \|x\| = 1\}$ of a separable and possibly infinite-dimensional Hilbert space $\mathbb{H}$ as well as a centred Gaussian measure $\mathrm{N}(0, C)$ on $\mathbb{H}$. Furthermore, let $\Pi^{\mathbb{S}} : \mathbb{H} \to \mathbb{S}$ denote the radial projection to the sphere

$$\Pi^{\mathbb{S}} : \mathbb{H} \to \mathbb{S}, \quad \Pi^{\mathbb{S}}(x) := \begin{cases} \frac{x}{\|x\|}, & \text{if } x \neq 0, \\ \bar{z}, & \text{if } x = 0, \end{cases} \tag{3}$$

with a fixed but arbitrary $\bar{z} \in \mathbb{S}$. Then we call the probability measure

$$\mu_0 := \Pi^{\mathbb{S}}_{\sharp} \mathrm{N}(0, C),$$

the *angular central Gaussian measure* with parameter C and denote it by $\mu_0 = \mathrm{ACG}(C)$. In the case where $\mathbb{H} = \mathbb{R}^d$ with the usual Euclidean norm and $C \in \mathbb{R}^{d \times d}$ being symmetric and positive definite, one can show that the density $\rho : \mathbb{S}^{d-1} \to [0, \infty)$ of $\mu_0 = \mathrm{ACG}(C)$ with respect to the $(d-1)$-dimensional Hausdorff measure on the sphere is

$$\rho(\bar{x}) = \frac{\Gamma(d/2)}{2\pi^{d/2}\sqrt{\det C}} \, \|\bar{x}\|_C^{-d};$$

see Appendix A.2 for details, including the definition (A4) of $\|\bar{x}\|_C$. We shall write bars over symbols to distinguish elements of $\mathbb{S}$ from elements of $\mathbb{H}$. Thus, $x \in \mathbb{H}$, while $\bar{x} \in \mathbb{S}$.

Consider a given target or posterior measure $\mu \in \mathcal{P}(\mathbb{S})$ which is absolutely continuous with respect to an ACG reference or prior measure $\mu_0 := \mathrm{ACG}(C)$, that is,

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(\bar{x}) \propto \exp(-\overline{\Phi}(\bar{x})), \quad \bar{x} \in \mathbb{S}, \tag{4}$$

where $\overline{\Phi} : \mathbb{S} \to \mathbb{R}$ denotes a measurable function that satisfies

$$\int_{\mathbb{S}} \exp(-\overline{\Phi}(\bar{x})) \, \mu_0(\mathrm{d}\bar{x}) < \infty.$$

The ACG prior allows us to define an equivalent sampling problem in the ambient Hilbert space.

Lifting to ambient Hilbert space

Define the measurable function $\Phi : \mathbb{H} \to \mathbb{R}$ by

$$\Phi(x) := \overline{\Phi}(\Pi^{\mathbb{S}}(x)), \qquad x \in \mathbb{H}, \tag{5}$$

where $\Pi^{\mathbb{S}} : \mathbb{H} \to \mathbb{S}$ is the radial projection to the sphere from (3), and define a target measure $\nu \in \mathcal{P}(\mathbb{H})$ via

$$\frac{\mathrm{d}\nu}{\mathrm{d}\nu_0}(x) \propto \exp(-\Phi(x)), \quad \nu_0\text{-a.e. } x \in \mathbb{H}, \tag{6}$$

where $\nu_0 = \mathrm{N}(0, C)$. Using $\mu_0 = \mathrm{ACG}(C) = \Pi^{\mathbb{S}}_{\sharp}\nu_0$ and using the construction of Φ, we obtain $\mu = \Pi^{\mathbb{S}}_{\sharp}\nu$. We show this result in a slightly more general form, that is, for an arbitrary measurable map $T : \mathbb{X} \to \mathbb{Y}$ between two arbitrary topological spaces $\mathbb{X}$ and $\mathbb{Y}$. In particular, one can apply Proposition 1 to more general manifolds in Hilbert spaces, provided that these manifolds can be expressed as the images of a measurable mapping.

Proposition 1. *Let $\nu_0 \in \mathcal{P}(\mathbb{X})$ and $T : \mathbb{X} \to \mathbb{Y}$ be measurable. Let $\overline{\Phi} : \mathbb{Y} \to \mathbb{R}$ be a measurable function that satisfies $Z = \int_{\mathbb{Y}} \exp(-\overline{\Phi}(T(x)))\ \nu_0(dx) < \infty$. Define ν by*

$$\frac{d\nu}{d\nu_0}(x) = \frac{1}{Z}\exp(-\overline{\Phi}(T(x))), \qquad \nu_0\text{-a.e. } x \in \mathbb{X}.$$

Then

$$\frac{dT_\sharp\nu}{dT_\sharp\nu_0}(\bar{x}) = \frac{1}{Z}\exp(-\overline{\Phi}(\bar{x})), \qquad T_\sharp\nu_0\text{-a.e. } \bar{x} \in \mathbb{Y}.$$

Proof. Let $A \in \mathcal{B}(\mathbb{Y})$. We shall show that

$$(T_\sharp\nu)(A) = \frac{1}{Z}\int_A \exp(-\overline{\Phi}(\bar{x}))\ T_\sharp\nu_0(\mathrm{d}\bar{x}). \tag{7}$$

To this end, let $X \sim \nu_0$ and $\overline{X} := T(X)$, i.e., $\overline{X} \sim T_\sharp\nu_0$, be random variables on the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ that we fixed in Section 1.4. Then

$$\begin{aligned}
(T_\sharp\nu)(A) &= \frac{1}{Z}\int_{T^{-1}(A)} \exp(-\overline{\Phi}(T(x)))\ \nu_0(\mathrm{d}x) \\
&= \frac{1}{Z}\int_{X^{-1}(T^{-1}(A))} \exp(-\overline{\Phi}(T(X(\omega))))\ \mathbb{P}(\mathrm{d}\omega) \quad \text{since } \nu_0 = \mathbb{P}\circ X^{-1} \\
&= \frac{1}{Z}\int_{\overline{X}^{-1}(A)} \exp(-\overline{\Phi}(\overline{X}(\omega)))\ \mathbb{P}(\mathrm{d}\omega) \quad \text{since } \overline{X} := T(X) \\
&= \frac{1}{Z}\int_A \exp(-\overline{\Phi}(\bar{x}))\ T_\sharp\nu_0(\mathrm{d}\bar{x}) \quad \text{since } \overline{X} \sim T_\sharp\nu_0,
\end{aligned}$$

which establishes (7) and completes the proof. ∎

The idea of sampling the push-forward $\mu = T_\sharp\nu$ of a measure ν defined on the ambient Hilbert space $\mathbb{H}$ is crucial for the construction of the following algorithms. In particular, we shall exploit suitable transition kernels for sampling from $\nu \in \mathcal{P}(\mathbb{H})$ in order to construct Markov chains on $\mathbb{S}$ with invariant distribution $T_\sharp\nu$, where $T = \Pi^{\mathbb{S}}$. To this end, we use the framework of push-forward transition kernels, which we describe in Section 3.3. Before describing this framework, we first discuss two simpler approaches for generating μ-invariant Markov chains on the sphere $\mathbb{S}$ based on transition kernels K in the ambient space $\mathbb{H}$, and explain why they are unsuitable.

3.1 | Naïve approaches and their shortcomings

Given a Markov chain $(X_n)_{n\in\mathbb{N}}$ with ν-reversible transition kernel K, one can also consider $(T(X_n))_{n\in\mathbb{N}}$ as a sequence of random variables on $\mathbb{Y}$. In our prototypical setting where $\mathbb{Y} = \mathbb{S}$, the

stochastic process $(T(X_n))_{n\in\mathbb{N}}$ is simply the projection of the Markov chain $(X_n)_{n\in\mathbb{N}}$ onto $\mathbb{S}$ via $T = \Pi^{\mathbb{S}}$. Hence, one can think of this as a simple projection approach. If the law $\mathbb{P}\circ(X_n)^{-1}$ of X_n converges to ν in the total variation norm as $n \to \infty$, then the law $\mathbb{P}\circ(T(X_n))^{-1}$ of $T(X_n)$ will also converge to $T_\sharp\nu$ in the total variation norm, since

$$\left\|\mathbb{P}\circ(T(X_n))^{-1} - T_\sharp\nu\right\|_{\mathrm{TV}} \leqslant \left\|\mathbb{P}\circ(X_n)^{-1} - \nu\right\|_{\mathrm{TV}},$$

due to

$$\left\|T_\sharp\rho - T_\sharp\nu\right\|_{\mathrm{TV}} = \sup_{A\in\mathcal{B}(\mathbb{S})} |\rho(T^{-1}(A)) - \nu(T^{-1}(A))| = \sup_{A\in\sigma(T)} |\rho(A) - \nu(A)| \leqslant \|\rho - \nu\|_{\mathrm{TV}}$$

with $\mathbb{P}\circ(T(X_n))^{-1} = T_\sharp\rho$ and $\mathbb{P}\circ(X_n)^{-1} = \rho$, where $\sigma(T)$ denotes the Borel σ-algebra generated by T. However, the sequence $(T(X_n))_{n\in\mathbb{N}}$ fails, in general, to be a Markov chain (Glover & Mitro, 1990). In particular, we provide an explicit counterexample in the case of $T = \Pi^{\mathbb{S}}$ in Appendix D. More generally, Rosenblatt (1966, theorem 3) considers general Markov processes $(X_i)_{i\in I}$, which may be discrete or continuous in time or space, and gives sufficient and necessary conditions on a measurable mapping T such that $(T(X_i))_{i\in I}$ is again a Markov process.

Another related approach can be constructed as follows. One could simply define the transition kernel

$$\overline{K}(\bar{x}, A) := K(\bar{x}, T^{-1}(A)) \qquad \forall \bar{x} \in \mathbb{Y}, A \in \mathcal{B}(\mathbb{Y}). \tag{8}$$

We shall refer to the transition kernel $\overline{K}$ in (8) as the "naïve reprojection kernel." We call $\overline{K}$ "naïve" because it does not perform averaging with respect to the regular conditional distribution $\nu_{|T}(\bar{x}, \cdot)$ of $X \sim \nu$ given $T(X) = \bar{x}$. In (11) below, we describe a kernel—the so-called 'push-forward transition kernel'—that does perform this averaging. In the setting where the topological spaces $\mathbb{X}$ and $\mathbb{Y}$ satisfy $\mathbb{Y} \subset \mathbb{X}$, one realizes $\bar{y}$ with respect to $\overline{K}(\bar{x}, \cdot)$ by first choosing y according to $K(\bar{x}, \cdot)$ and then setting $\bar{y} := T(y)$, as illustrated in Figure 1. That is, one first transitions from $\bar{x} \in \mathbb{Y}$ to a state y in the ambient space $\mathbb{X}$, and then "reprojects" this state y into $\bar{y} \in \mathbb{Y}$ using the mapping T.

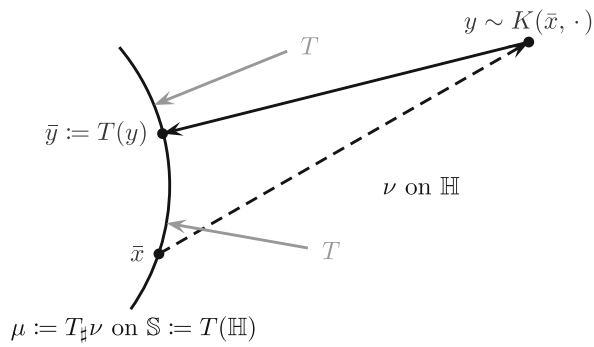

FIGURE 1 Illustration of the steps for drawing states using the naïve reprojection kernel $\overline{K}$ in (8) for $\mathbb{X} = \mathbb{H}$, $\mathbb{Y} = \mathbb{S}$, and $T = \Pi^{\mathbb{S}}$ in (3). Starting from $\bar{x}$, an intermediate state $y \in \mathbb{H}$ is drawn from the ν-reversible transition kernel $K(\bar{x}, \cdot)$ on $\mathbb{H}$. The next state drawn from the naïve reprojection kernel $\overline{K}(\bar{x}, \cdot)$ is then $\bar{y} := T(y) = \Pi^{\mathbb{S}}(y)$. Solid arrows indicate deterministic maps, whereas dashed arrows indicate randomized maps, that is, draws from transition kernels.

Unlike the projection approach described earlier, this method yields a Markov chain. However, as numerical experiments show, the naïve reprojection kernel $\overline{K}$ does not have μ as its stationary distribution, even if K is ν-invariant or ν-reversible. To see why, recall that $\mu = T_\sharp \nu$. Hence, $\overline{K}$ is μ-invariant if and only if

$$\int_{\mathbb{Y}} \overline{K}(\bar{x}, A)\ \mu(\mathrm{d}\bar{x}) = \int_{\mathbb{Y}} K(\bar{x}, T^{-1}(A))\ \mu(\mathrm{d}\bar{x}) = \int_{\mathbb{X}} K(T(x), T^{-1}(A))\ \nu(\mathrm{d}x) = \nu(T^{-1}(A)),$$

for all $A \in \mathcal{B}(\mathbb{Y})$. If K is ν-invariant, then by definition $\nu(T^{-1}(A)) = \int_{\mathbb{X}} K(x, T^{-1}(A))\ \nu(\mathrm{d}x)$. This yields the following necessary and sufficient condition for the μ-invariance of $\overline{K}$:

$$\int_{\mathbb{X}} K(x, T^{-1}(A))\ \nu(\mathrm{d}x) = \int_{\mathbb{X}} K(T(x), T^{-1}(A))\ \nu(\mathrm{d}x), \qquad \forall A \in \mathcal{B}(\mathbb{Y}).$$

Based upon numerical experiments, we argue that this condition is not necessarily satisfied in the setting where $\mathbb{X} = \mathbb{H}$, $\mathbb{Y} = \mathbb{S}$, and $T = \Pi^{\mathbb{S}}$. Let $\mathbb{H} = \mathbb{R}^3$, $\nu = \mathrm{N}(0, C)$ with covariance matrix $C \in \mathbb{R}^{3\times 3}$, and consider the ν-reversible pCN proposal kernel

$$K(x) = \mathrm{N}\left(\sqrt{1-s^2}x, s^2 C\right), \quad \text{with} \quad s = 0.7 \ \text{ and } C = \begin{pmatrix} 1.25 & 0.33 & -1.62 \\ 0.33 & 0.42 & -0.09 \\ -1.62 & -0.09 & 2.85 \end{pmatrix}.$$

We now estimate and compare the probability density function of the marginals of $\mu = \Pi^{\mathbb{S}}_\sharp \nu$ and $\mu\overline{K}$ by kernel density estimation based on 10^6 independent samples of μ and $\mu\overline{K}$, respectively. Each sample was generated according to the following procedure: (1) Draw a sample x from ν and set $\bar{x} := \Pi^{\mathbb{S}}(x)$, so that $\bar{x}$ is a sample draw from μ; (2) draw another sample w from ν and set $y := \sqrt{1-s^2}\bar{x} + sw$, so that $\bar{y} = \Pi^{\mathbb{S}}(y)$ is a sample draw from $\mu\overline{K}$. The results are displayed in Figure 2. The important observation is that the marginals of μ (dashed yellow line) and $\mu\overline{K}$ (dotted blue line) differ. Hence, $\overline{K}$ is not μ-invariant in this case. Note that the marginals of μ (dashed yellow line) coincide with the marginals of $\mu(\Pi^{\mathbb{S}}_\sharp K)$ (solid red line). Here, $\Pi^{\mathbb{S}}_\sharp K$ is the reprojection kernel $T_\sharp K$ that performs averaging with respect to the regular conditional distribution of $X \sim \nu$ given $T(X) = \bar{x}$, for $T = \Pi^{\mathbb{S}}$. In Section 3.3 below, we define $T_\sharp K$ in (11) for general measurable mappings T between Polish spaces and show, among other things, that it is μ-reversible for $\mu = T_\sharp \nu$ and ν-reversible K.

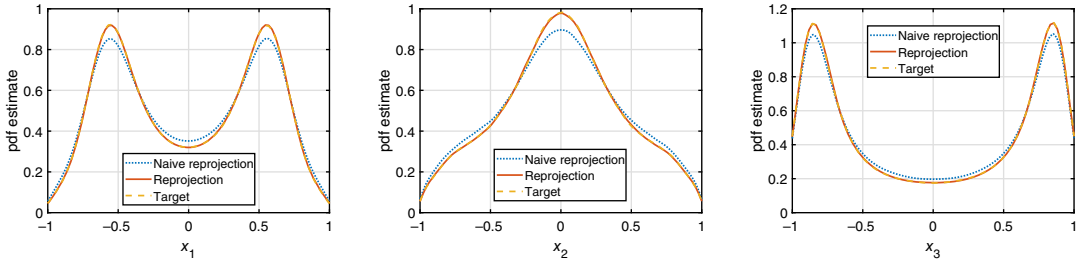

FIGURE 2 Comparison of marginals of $\mu = \Pi^{\mathbb{S}}_\sharp \nu$ (target), $\mu\overline{K}$ (naïve reprojection) with $\overline{K}$ as in (8) with preconditioned Crank–Nicolson proposal kernel K and $T = \Pi^{\mathbb{S}}$, and $\mu(\Pi^{\mathbb{S}}_\sharp K)$ (reprojection) with $\Pi^{\mathbb{S}}_\sharp K$ as in (11) for the same K and T.

3.2 | The reprojection method

We present now a simple method for defining a μ-reversible Markov chain $(\overline{X}_n)_{n\in\mathbb{N}}$ on $\mathbb{S}$, by using a ν-reversible transition kernel K on $\mathbb{H}$. The method employs the concept of *push-forward transition kernels* which we explain for the general setting of Polish spaces $\mathbb{X}$, $\mathbb{Y}$ connected via a measurable mapping $T : \mathbb{X} \to \mathbb{Y}$. For more details of this approach we refer to (Rudolf & Sprungk, 2022). For the particular algorithms that we consider later, we focus on the specific case of $\mathbb{Y} = \mathbb{S}$, $\mathbb{X} = \mathbb{H}$, and T the radial projection map $T = \Pi^{\mathbb{S}}$ given in (3).

Given a ν-invariant transition kernel K on $\mathbb{X}$ and a measurable map $T : \mathbb{X} \to \mathbb{Y}$, we define the *push-forward transition kernel* $T_\sharp K$ on $\mathbb{Y}$ as follows:

$$T_\sharp K(\overline{x}, A) := \mathbb{E}\left[K(X, T^{-1}(A))|T(X) = \overline{x}\right], \qquad X \sim \nu, \tag{9}$$

where $\overline{x} \in \mathbb{Y}$ and $A \in \mathcal{B}(\mathbb{Y})$. If T is bijective, then

$$T_\sharp K(\overline{x}, A) = K(T^{-1}(\overline{x}), T^{-1}(A)).$$

In the following, we also use the shorter notation $\overline{K} = T_\sharp K$. Below, we summarise some important properties of push-forward transition kernels that are inherited from the original transition kernel.

Lemma 1 (Rudolf & Sprungk, 2022). *Let $\mathbb{X}$ and $\mathbb{Y}$ be Polish spaces, $T : \mathbb{X} \to \mathbb{Y}$ be a measurable mapping, K be a ν-invariant transition kernel on $\mathbb{X}$, and $\mu := T_\sharp \nu$.*

(a) If K is reversible with respect to ν, then $T_\sharp K$ is reversible with respect to μ.
(b) If K has an L^2_ν-spectral gap, then $T_\sharp K$ has an L^2_μ-spectral gap and

$$\mathrm{gap}_\mu(T_\sharp K) \geqslant \mathrm{gap}_\nu(K).$$

(c) If K is an MH kernel with proposal kernel $Q : \mathbb{X} \times \mathcal{B}(\mathbb{X}) \to [0,1]$ and acceptance probability $\alpha : \mathbb{X} \times \mathbb{X} \to [0,1]$ such that

$$\alpha(x, y) = \overline{\alpha}(T(x), T(y)) \qquad \forall x, y \in \mathbb{X},$$

for a measurable $\overline{\alpha} : \mathbb{Y} \times \mathbb{Y} \to [0,1]$, then $T_\sharp K$ is an MH kernel with acceptance probability $\overline{\alpha}$ and proposal kernel $\overline{Q}$ given by

$$\overline{Q}(\overline{x}, A) := \int_{\mathbb{X}} Q(x, T^{-1}(A))\ \nu_{|T}(\overline{x}, dx), \qquad \forall \overline{x} \in \mathbb{Y},\ A \in \mathcal{B}(\mathbb{Y}),$$

where $\nu_{|T}(\overline{x}, \cdot)$ denotes the regular conditional distribution of $X \sim \nu$ given $T(X) = \overline{x}$.

See Appendix B.1 for the definition of the spectral gap of a transition kernel K.

The last item in the above lemma also shows that one can simulate push-forward transition kernels by exploiting the regular conditional distribution $\nu_{|T} : \mathbb{Y} \times \mathcal{B}(\mathbb{X}) \to [0,1]$ of $X \sim \nu$ given $T(X) = \overline{x}$. We recall that $\nu_{|T}$ possesses the properties of a transition kernel and satisfies

$$\nu_{|T}(T(X), A) = \mathbb{P}\left(X \in A \,\middle|\, T(X)\right) \qquad \mathbb{P}\text{-almost surely}, \tag{10}$$

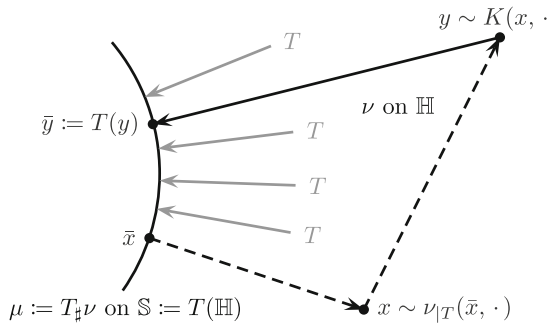

FIGURE 3 Illustration of the steps in the reprojection method from Section 3.3. The reprojection method defines a μ-reversible transition kernel on $\mathbb{S}$ in terms of a ν-reversible transition kernel K on the ambient space $\mathbb{H}$, where $\mathbb{S} := T(\mathbb{H})$, $\mu := T_\sharp \nu$, $\nu_{|T}(\bar{x}, \cdot)$ is the regular conditional distribution of $X \sim \nu$ given $T(X) = \bar{x}$, and $T = \Pi^{\mathbb{S}}$. Solid arrows indicate deterministic maps, whereas dashed arrows indicate randomized maps, that is, draws from transition kernels.

for any $A \in \mathcal{B}(\mathbb{X})$. Given this regular conditional distribution, the disintegration theorem yields the representation

$$T_\sharp K(\bar{x}, A) = \int_{\mathbb{X}} K(x, T^{-1}(A))\ \nu_{|T}(\bar{x}, \mathrm{d}x), \tag{11}$$

for general transition kernels K. Thus, the push-forward transition kernel $T_\sharp K$ can be realised by the following mechanism.

Transition Mechanism 1. Given the current state $\bar{x} \in \mathbb{Y}$ one obtains the next state $\bar{y} \in \mathbb{Y}$ as follows:

(1) Draw $X \sim \nu_{|T}(\bar{x}, \cdot)$ and call the realisation $x \in \mathbb{X}$;
(2) Draw $Y \sim K(x, \cdot)$, call the realisation $y \in \mathbb{X}$ and return $\bar{y} := T(y) \in \mathbb{Y}$.

We now consider the specific case of $\mathbb{X} = \mathbb{H}$ being a Hilbert space, $\mathbb{Y} = \mathbb{S}$ its unit sphere and $T = \Pi^{\mathbb{S}}$ being the radial projection defined in (3). In order to obtain a μ-reversible Markov chain on $\mathbb{S}$, we can consider the push-forward transition kernels $\overline{K} = T_\sharp K$ of ν-reversible transition kernels K on the ambient Hilbert space $\mathbb{H}$—such as the pCN-MH kernel or the ESS kernel—provided that we can also simulate the regular conditional distribution $\nu_{|T}$ for the lifted target ν in (6). The resulting algorithm is illustrated in Figure 3. In particular, by going randomly from $\bar{x}$ to x in the ambient space, performing a transition from x to y by using K, and then by "reprojecting" deterministically from y to $\bar{y} = \Pi^{\mathbb{S}}(y)$, we end up on the sphere $\mathbb{S}$. Since this is performed at each iteration of the Markov chain, we name this the *reprojection method*.

Next, we derive the regular conditional distribution $\nu_{|\Pi^{\mathbb{S}}}$ for Gaussian measures $\nu = \mathrm{N}(0, C)$ on $\mathbb{H} = \mathbb{R}^d$, and state the resulting reprojected pCN-MH algorithm as well as the reprojected ESS algorithm.

Simulating the conditional distribution $\nu_{|\Pi^{\mathbb{S}}}$

We first prove a proposition about the regular conditional distributions $\nu_{|T}$ in the more general setting of Polish spaces $\mathbb{X}$, $\mathbb{Y}$. We then apply this proposition to derive an explicit description for

$\nu_{|\Pi^{\mathbb{S}}}$. One can modify this procedure for manifolds in Hilbert spaces that are more general than the unit sphere $\mathbb{S}$, for example, manifolds that can be described by a measurable mapping $T : \mathbb{H} \to \mathbb{H}$, by replacing $\Pi^{\mathbb{S}}$ with T.

Proposition 2. *Let $\mathbb{X}$ and $\mathbb{Y}$ be Polish spaces equipped with a measurable mapping $T : \mathbb{X} \to \mathbb{Y}$ and $\nu_0 \in \mathcal{P}(\mathbb{X})$. Let $\overline{\Phi} : \mathbb{Y} \to \mathbb{R}$ be measurable with $Z := \int_{\mathbb{Y}} \exp(-\overline{\Phi}(T(x)))\ \nu_0(dx) < \infty$, define $\Phi : \mathbb{X} \to \mathbb{R}$ by $\Phi(x) := \overline{\Phi}(T(x))$ for $x \in \mathbb{X}$, and let $\nu \in \mathcal{P}(\mathbb{X})$ be given by*

$$\frac{d\nu}{d\nu_0}(x) = \frac{1}{Z}\exp(-\Phi(x)), \quad \nu_0\text{-a.e. } x \in \mathbb{X}.$$

Furthermore, let $\nu_{0|T}$ be the regular conditional distribution of $X_0 \sim \nu_0$ given $T(X_0)$, and let $\nu_{|T}$ be the regular conditional distribution of $X \sim \nu$ given $T(X)$. Then

$$\nu_{|T}(T(x), \cdot) = \nu_{0|T}(T(x), \cdot),$$

for ν_0-a.e. $x \in \mathbb{X}$.

The hypotheses of the proposition differ from the hypotheses of Proposition 1 only in the additional assumption that $\mathbb{X}$ and $\mathbb{Y}$ are Polish spaces. This assumption ensures that we may apply the disintegration theorem to obtain regular conditional distributions. We can weaken the assumption by requiring that $\mathbb{X}$ and $\mathbb{Y}$ be merely Radon spaces.

Proof of Proposition 2. The regular conditional distribution $\nu_{0|T} : \mathbb{Y} \times \mathcal{B}(\mathbb{X}) \to [0, 1]$ of X_0 given $T(X_0)$ is defined as a Markov kernel satisfying for every $A \in \mathcal{B}(\mathbb{X})$ and $B \in \mathcal{B}(\mathbb{Y})$,

$$\mathbb{P}(X_0 \in A,\ T(X_0) \in B) = \int_B \nu_{0|T}(y, A)\ T_\sharp\nu_0(\mathrm{d}y) = \int_{T^{-1}(B)} \nu_{0|T}(T(x), A)\ \nu_0(\mathrm{d}x).$$

Moreover, given the regular conditional distribution $\nu_{0|T}$ we can express the conditional expectation of $g(X_0)$ given $T(X_0)$ for any measurable $g : \mathbb{X} \to \mathbb{R}$ as

$$\mathbb{E}\left[g(X_0)|T(X_0)\right] = \int_{\mathbb{X}} g(x)\ \nu_{0|T}(T(X_0), \mathrm{d}x) \qquad \mathbb{P}\text{-almost surely.} \tag{12}$$

We ask now for the regular conditional distribution $\nu_{|T}$ of $X \sim \nu$ given $T(X)$. Analogously, this is a Markov kernel $\nu_{|T} : \mathbb{Y} \times \mathcal{B}(\mathbb{X}) \to [0, 1]$ satisfying

$$\mathbb{P}(X \in A,\ T(X) \in B) = \int_B \nu_{|T}(y, A)\ T_\sharp\nu(\mathrm{d}y) = \int_{T^{-1}(B)} \nu_{|T}(T(x), A)\ \nu(\mathrm{d}x),$$

for every $A \in \mathcal{B}(\mathbb{X})$ and $B \in \mathcal{B}(\mathbb{Y})$. Since $\mathbb{P}(X \in A,\ T(X) \in B) = \nu(A \cap T^{-1}(B))$, the statement follows if

$$\nu(A \cap T^{-1}(B)) = \int_{T^{-1}(B)} \nu_{0|T}(T(x), A)\ \nu(\mathrm{d}x) \qquad \forall A \in \mathcal{B}(\mathbb{X})\ \forall B \in \mathcal{B}(\mathbb{Y}), \tag{13}$$

because $\nu_{0|T}$ is then a valid regular conditional distribution of $X \sim \nu$ given $T(X)$. For $A \in \mathcal{B}(\mathbb{X})$ and $B \in \mathcal{B}(\mathbb{Y})$,

$$\nu(A \cap T^{-1}(B)) = \int_{\mathbb{X}} \mathbb{1}_A(x)\ \mathbb{1}_B(T(x))\ \frac{1}{Z}\mathrm{e}^{-\Phi(x)}\ \nu_0(\mathrm{d}x) = \mathbb{E}\left[\frac{1}{Z}\mathrm{e}^{-\Phi(X_0)}\ \mathbb{1}_A(X_0)\ \mathbb{1}_B(T(X_0))\right],$$

where $X_0 \sim \nu_0$. Using the law of total expectation and the hypothesis that $\Phi(x) = \overline{\Phi}(T(x))$ for every $x \in \mathbb{X}$, we obtain

$$\begin{aligned}\nu(A \cap T^{-1}(B)) &= \mathbb{E}\left[\mathbb{E}\left[\frac{1}{Z}\mathrm{e}^{-\Phi(X_0)}\ \mathbb{1}_A(X_0)\ \mathbb{1}_B(T(X_0))\ \middle|\ T(X_0)\right]\right] \\ &= \mathbb{E}\left[\frac{1}{Z}\mathrm{e}^{-\Phi(X_0)}\ \mathbb{1}_B(T(X_0))\ \mathbb{E}\left[\mathbb{1}_A(X_0)|T(X_0)\right]\right].\end{aligned}$$

Applying now (12) to $g(x) = \mathbb{1}_A(x)$ yields $\mathbb{E}\left[\mathbb{1}_A(X_0)|T(X_0)\right] = \nu_{0|T}(T(X_0), A)$ and, thus,

$$\begin{aligned}\nu(A \cap T^{-1}(B)) &= \mathbb{E}\left[\frac{1}{Z}\mathrm{e}^{-\Phi(X_0)}\ \mathbb{1}_B(T(X_0))\ \nu_{0|T}(T(X_0), A)\right] \\ &= \int_{T^{-1}(B)} \frac{1}{Z}\mathrm{e}^{-\Phi(x)}\ \nu_{0|T}(T(x), A)\ \nu_0(\mathrm{d}x) \\ &= \int_{T^{-1}(B)} \nu_{0|T}(T(X_0), A)\ \nu(\mathrm{d}x),\end{aligned}$$

which shows (13). ∎

We now turn to the setting of a finite dimensional sphere, where $\mathbb{X} = \mathbb{R}^d$, $\mathbb{Y} = \mathbb{S}^{d-1}$, and $T = \Pi^{\mathbb{S}}$. In order to implement the reprojection method, it suffices by Proposition 2 to simulate the regular conditional distribution of $X_0 \sim \mathrm{N}(0, C)$ given $\Pi^{\mathbb{S}}(X_0) = \overline{x}$. In the following result $\mathrm{Gam}(a, b)$ denotes the Gamma distribution with shape parameter $a > 0$ and inverse scale parameter $b > 0$.

Proposition 3. *Let $\nu_0 = N(0, C)$ be given on $\mathbb{R}^d$ and let $\nu_{0|\Pi^{\mathbb{S}}}(\overline{x}, \cdot)$ denote the conditional distribution of $X_0 \sim \nu_0$ given $\Pi^{\mathbb{S}}(X_0) = \overline{x}$. Then, for a nonnegative real-valued random variable R satisfying $R^2 \sim \mathrm{Gam}\left(\frac{d}{2}, \frac{1}{2}\overline{x}^\top C^{-1}\overline{x}\right)$,*

$$R\overline{x} \sim \nu_{0|\Pi^{\mathbb{S}}}(\overline{x}, \cdot).$$

Proof. We can write $X_0 \sim \nu = \mathrm{N}(0, C)$ as $X_0 = R\overline{X}_0$, $\overline{X}_0 := \Pi^{\mathbb{S}}(X_0)$, $R := \|X_0\|$. Thus, the condition of $\Pi^{\mathbb{S}}(X_0) = \overline{x}$ yields the following conditional density of R:

$$f_{R|\Pi^{\mathbb{S}}(X_0)=\overline{x}}(r) \ \propto\ r^{d-1}\exp\left(-\frac{1}{2}(\overline{x}^\top C^{-1}\overline{x})r^2\right) = r^{d-1}\left(\exp\left(-\frac{1}{2}\overline{x}^\top C^{-1}\overline{x}\right)\right)^{r^2}.$$

By the change of variables $r \mapsto r^2 =: r_2$ we obtain the following probability density for R^2 given $\Pi^{\mathbb{S}}(X_0) = \overline{x}$:

$$f_{R^2|\Pi^{\mathbb{S}}(X_0)=\overline{x}}(r_2) \ \propto\ \frac{r_2^{(d-1)/2}}{2r_2^{1/2}}\left(\mathrm{e}^{-\frac{1}{2}\overline{x}^\top C^{-1}\overline{x}}\right)^{r_2} \ \propto\ r_2^{d/2-1}\mathrm{e}^{-\left(\frac{1}{2}\overline{x}^\top C^{-1}\overline{x}\right)r_2}.$$

Thus, R^2 conditioned on $\Pi^{\mathbb{S}}(X_0) = \overline{x}$ is distributed according to $\mathrm{Gam}(\frac{d}{2}, \frac{1}{2}\overline{x}^\top C^{-1}\overline{x})$, as desired. ∎

Algorithm 4. Reprojected pCN-MH algorithm on $\mathbb{S}^{d-1}$

Given: ACG prior $\mu_0 = \mathrm{ACG}(C)$ and target μ as in (4)
Initial: step size $s \in (0, 1]$ and state $\bar{x}_0 \in \mathbb{S}^{d-1}$
for $k \in \mathbb{N}_0$ **do**
  Draw a sample r_k^2 of $\mathrm{Gam}\left(d/2, \frac{1}{2}\bar{x}_k^\top C^{-1}\bar{x}_k\right)$ and set $x_k = r_k\bar{x}_k$
  Draw a sample w_k of $\mathrm{N}(0, C)$ and set $y_{k+1} := \sqrt{1-s^2}x_k + sw_k$
  Set $\bar{y}_{k+1} := y_{k+1}/\|y_{k+1}\|$
  Compute $a := \min\{1, \exp\left(\bar{\Phi}(\bar{x}_k) - \bar{\Phi}(\bar{y}_{k+1})\right)\}$
  Draw a sample u of $\mathrm{U}[0, 1]$
  if $u \leqslant a$ **then**
    Set $\bar{x}_{k+1} = \bar{y}_{k+1}$
  else
    Set $\bar{x}_{k+1} = \bar{x}_k$
  end if
end for

Algorithm 5. Reprojected ESS algorithm on $\mathbb{S}^{d-1}$

Given: ACG prior $\mu_0 = \mathrm{ACG}(C)$ and target μ as in (4)
Initial: state $\bar{x}_0 \in \mathbb{S}^{d-1}$
for $k \in \mathbb{N}_0$ **do**
  Draw a sample $t \sim \mathrm{U}[0, \exp(-\bar{\Phi}(\bar{x}_k))]$
  Draw a sample r_k^2 of $\mathrm{Gam}\left(d/2, \frac{1}{2}\bar{x}_k^\top C^{-1}\bar{x}_k\right)$ and set $x_k = r_k\bar{x}_k$
  Set $x_{k+1} = \text{shrink-ellipse}(x_k, t)$ (see Algorithm 3)
  Set $\bar{x}_{k+1} = x_{k+1}/\|x_{k+1}\|$
end for

Resulting algorithms

We now provide two explicit algorithms for approximate sampling of target measures μ on $\mathbb{S}^{d-1}$ as given in (4). Algorithms 4 and 5 result from applying the push-forward transition kernel approach to the pCN-MH algorithm and the ESS algorithm on $\mathbb{H}$, respectively. The pCN-MH algorithm and the ESS algorithm on $\mathbb{H}$ were stated in Algorithms 1 and 2. According to Lemma 1(c), Algorithm 4 yields an MH algorithm on $\mathbb{S}^{d-1}$. Its acceptance probability is simply $\overline{\alpha}(\bar{x}, \bar{y}) = \min\{1, \exp\left(\overline{\Phi}(\bar{y}) - \overline{\Phi}(\bar{x})\right)\}$, for $\bar{x}, \bar{y} \in \mathbb{S}^{d-1}$, and its proposal kernel $\overline{Q} : \mathbb{S}^{d-1} \times \mathcal{B}(\mathbb{S}^{d-1}) \to [0, 1]$ admits a proposal density $\overline{q} : \mathbb{S}^{d-1} \times \mathbb{S}^{d-1} \to (0, \infty)$ with respect to the Hausdorff measure on $\mathbb{S}^{d-1}$ given by

$$\overline{q}(\bar{x}, \bar{y}) = \int_0^\infty \int_0^\infty q(r\bar{x}, r'\bar{y})\, f_{\bar{x}}(r)\, (r')^{d-1}\, \mathrm{d}r\, \mathrm{d}r' > 0, \qquad \bar{x}, \bar{y} \in \mathbb{S}^{d-1}, \tag{14}$$

where q denotes the proposal density of the pCN proposal kernel $Q(x, \cdot) = \mathrm{N}(\sqrt{1-s^2}x, s^2C)$, $x \in \mathbb{R}^d$, and $f_{\bar{x}}(r)$ denotes the conditional density of $R = \|X\|$ for $X \sim \mathrm{N}(0, C)$ given $\Pi^{\mathbb{S}}(X) = \bar{x}$.

According to the proof of Proposition 3, the density $f_{\bar{x}}(r)$ takes the form

$$f_{\bar{x}}(r) := \frac{1}{c_{\bar{x}}} r^{d-1} \exp\left(-\frac{r^2}{2}\bar{x}^\top C^{-1}\bar{x}\right), \qquad c_{\bar{x}} := \int_0^\infty r^{d-1} \exp\left(-\frac{r^2}{2}\bar{x}^\top C^{-1}\bar{x}\right) \mathrm{d}r. \tag{15}$$

Remark 2. Due to the generality of the pushforward Markov kernel approach, it is also possible to combine the reprojection methodology that we proposed with other common MH algorithms, such as the dimension-independent Hamiltonian Monte Carlo (HMC) algorithm of Beskos et al. (2011). However, an extensive investigation of a reprojected HMC algorithm that shows advantages and disadvantages, for example, in comparison to the work of Lan et al. (2014), is beyond the scope of this paper.

3.3 | Uniform and geometric ergodicity

We investigate the exponential convergence behaviour of the transition kernels that correspond to the Markov chains that are realised either by Algorithm 4 or 5. Since the underlying state space $\mathbb{S}^{d-1}$ is compact, we aim for uniform ergodicity. The associated transition kernel $\overline{K} : \mathbb{S}^{d-1} \times \mathcal{B}(\mathbb{S}^{d-1}) \to [0,1]$ is said to be *uniformly ergodic*, if there are $\kappa \in [0,1)$ and $c < \infty$ such that

$$\left\| \overline{K}^n(\bar{x}) - \mu \right\|_{\mathrm{TV}} \leqslant c\, \kappa^n \qquad \forall \bar{x} \in \mathbb{S}^{d-1}. \tag{16}$$

It is well known (Meyn & Tweedie, 2009, theorem 16.0.2) that uniform ergodicity of a Markov chain is equivalent to the *smallness* of the whole state space. A set $B \in \mathcal{B}(\mathbb{S}^{d-1})$ is called *small* with respect to a transition kernel $\overline{K}$ if there exists some $m \in \mathbb{N}$ and a nonzero measure ϕ on $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}))$ such that

$$\overline{K}^m(\bar{x}, A) \geqslant \phi(A) \qquad \forall A \in \mathcal{B}(\mathbb{S}^{d-1}),\ \bar{x} \in B. \tag{17}$$

In particular, if (17) holds for $B = \mathbb{S}^{d-1}$, then Meyn and Tweedie (2009, theorem 16.2.4) yield that

$$\left\| \overline{K}^n(\bar{x}) - \mu \right\|_{\mathrm{TV}} \leqslant (1 - \phi(\mathbb{S}^{d-1}))^{n/m-1}. \tag{18}$$

By exploiting the particular structure (11) of push-forward transition kernels, we obtain the following result.

Theorem 1. *Let $\overline{\Phi} : \mathbb{S}^{d-1} \to \mathbb{R}$ be uniformly bounded. Then the transition kernels corresponding to the Markov chains realised by Algorithms 4 and 5 are uniformly ergodic.*

Proof. The idea of the proof is to show that, in both cases, the state space is small.

We first consider the reprojected pCN-MH kernel $\overline{K}$. The boundedness of $\overline{\Phi}$, that is, $\underline{c} \leqslant \overline{\Phi}(\bar{x}) \leqslant \bar{c}$, yields the following lower bound on the acceptance probability:

$$\bar{\alpha}(\bar{x}, \bar{y}) = \min\left\{1, \exp\left(\overline{\Phi}(\bar{y}) - \overline{\Phi}(\bar{x})\right)\right\} \geqslant \exp(\underline{c} - \bar{c}) > 0.$$

Hence, by the corresponding MH form of the reprojected pCN-MH kernel $\overline{K}$ stated in Lemma 1, for any $A \in \mathcal{B}(\mathbb{S}^{d-1})$ and any $\bar{x} \in \mathbb{S}^{d-1}$,

$$\overline{K}(\bar{x}, A) \geqslant \int_A \bar{\alpha}(\bar{x}, \bar{y})\, \overline{Q}(\bar{x}, \mathrm{d}\bar{y}) \geqslant \exp(\underline{c} - \bar{c})\, \overline{Q}(\bar{x}, A).$$

Recall that $\overline{Q}$ possesses the density $\overline{q}$ given in (14). Note that $\overline{x} \mapsto \overline{x}^\top C^{-1}\overline{x}$ is bounded on $\mathbb{S}^{d-1}$, such that there exists for every $r > 0$ a lower bound $\underline{f}(r) > 0$ satisfying $f_{\overline{x}}(r) \geqslant \underline{f}(r) > 0$ for every $\overline{x} \in \mathbb{S}^{d-1}$. Moreover, the density $q(\cdot, y)$ of the pCN proposal kernel $Q(x, \cdot) = \mathrm{N}(\sqrt{1-s^2}x, s^2 C)$ is continuous. Therefore, $q(x, y)$ is uniformly bounded away from zero for any $x, y \in \mathbb{R}^d$ with $\|x\|, \|y\| \leqslant 1$. Hence, there exists some $\epsilon > 0$ such that, for every $\overline{x}, \overline{y} \in \mathbb{S}^{d-1}$, $\overline{q}(\overline{x}, \overline{y})$ in (14) satisfies $\overline{q}(\overline{x}, \overline{y}) \geqslant \epsilon$. This implies that $\overline{Q}(\overline{x}, A) \geqslant \epsilon \mathcal{H}_{\mathbb{S}^{d-1}}(A)$, for the Hausdorff measure $\mathcal{H}_{\mathbb{S}^{d-1}}$ on $\mathbb{S}^{d-1}$. Thus, $\mathbb{S}^{d-1}$ is small with respect to $\overline{K}$ with $\phi(A) := \epsilon \exp(\underline{c} - \overline{c})\mathcal{H}_{\mathbb{S}^{d-1}}(A)$.

Next, we consider the reprojected ESS kernel. Since Φ on $\mathbb{H} = \mathbb{R}^d$ is constructed from $\overline{\Phi}$ on $\mathbb{S}$ by (5), the boundedness of $\overline{\Phi}$ implies the boundedness of Φ. Hence, any compact set $B \subset \mathbb{R}^d$ is small with respect to the ESS transition kernel K. The measure with respect to which the smallness property holds is $\phi = \epsilon_B \lambda_B$, where $\epsilon_B > 0$ denotes a constant and λ_B is the Lebesgue measure restricted to a compact set B with positive d-dimensional Lebesgue measure; see Natarovskii et al. (2021a, lemma 3.4). We now use this fact in order to show the smallness of $\mathbb{S}^{d-1}$ with respect to the reprojected ESS transition kernel $\overline{K}$ for the measure $\phi = \epsilon \mathcal{H}_{\mathbb{S}^{d-1}}$ for appropriately chosen $\epsilon > 0$, see below. To this end, we apply the representation (11) with $\mathbb{X} = \mathbb{R}^d$, $\mathbb{Y} = \mathbb{S}^{d-1}$ and $T = \Pi^{\mathbb{S}}$:

$$\overline{K}(\overline{x}, A) = \int_0^\infty K(r\overline{x}, T^{-1}(A)) f_{\overline{x}}(r)\, \mathrm{d}r,$$

with $f_{\overline{x}}$ as in (15). Again, since $\overline{x} \mapsto \overline{x}^\top C^{-1}\overline{x}$ is bounded on $\mathbb{S}^{d-1}$, there exists for every $r > 0$ a lower bound $\underline{f}(r) > 0$ such that, for every $\overline{x} \in \mathbb{S}^{d-1}$, $f_{\overline{x}}(r) \geqslant \underline{f}(r) > 0$. Now fix $B := B_1(0) = \{x \in \mathbb{R}^d | \, \|x\| \leqslant 1\}$ and note that B is small with respect to K. Thus

$$\overline{K}(\overline{x}, A) \geqslant \int_0^1 K(r\overline{x}, T^{-1}(A)) \underline{f}(r)\, \mathrm{d}r \geqslant \epsilon_B \lambda_B(T^{-1}(A)) \int_0^1 \underline{f}(r)\, \mathrm{d}r = \epsilon \mathcal{H}_{\mathbb{S}^{d-1}}(A)$$

where $\epsilon := \epsilon_B \int_0^1 \underline{f}(r)\, \mathrm{d}r \int_0^1 u^{d-1}\, \mathrm{d}u$, since $\lambda_B(T^{-1}(A)) = \int_A \int_0^1 u^{d-1} \mathrm{d}u\, \mathcal{H}_{\mathbb{S}^{d-1}}(\mathrm{d}\overline{x})$. ∎

The boundedness assumption on $\overline{\Phi}$ in Theorem 1 is rather mild. It is satisfied if $\overline{\Phi} : \mathbb{S}^{d-1} \to \mathbb{R}$ is continuous. For example, in the Bayesian level set inversion and Bayesian density estimation problems considered in Section 4, the corresponding $\overline{\Phi}$ is bounded.

Theorem 1 yields uniform ergodicity in finite dimension. In the last paragraph of the proof of Theorem 1, we considered the measure $\phi := \epsilon \exp(\underline{c} - \overline{c})\mathcal{H}_{\mathbb{S}^{d-1}}$ for the reprojected pCN-MH kernel $\overline{K}$ and the measure $\phi := \epsilon \mathcal{H}_{\mathbb{S}^{d-1}}$ for the reprojected ESS kernel. Supposing that the prefactors ϵ and $\epsilon \exp(\underline{c} - \overline{c})$ do not grow in d and substituting these choices of ϕ in (18) we observe that the corresponding $\kappa = (1 - \phi(\mathbb{S}^{d-1}))^{1/m}$ in (16), with $c = (1 - \phi(\mathbb{S}^{d-1}))^{-1}$, converges exponentially quickly to 1 as $d \to \infty$. This is because the $(d-1)$-dimensional Hausdorff measure of $\mathbb{S}^{d-1}$ is given by $\frac{2\pi^{d/2}}{\Gamma(d/2)}$, where $\Gamma(\cdot)$ is the Gamma function, and because of the asymptotic behavior of $\Gamma(\cdot)$.

In the subsequent section, we present a dimension-independent convergence behavior, but in the context of geometric ergodicity as in (B5), and not in the context of uniform ergodicity as in (16).

Dimension-independent geometric ergodicity

In order to study the geometric ergodicity of Markov chains generated by the reprojected pCN-MH and reprojected ESS algorithms, we can exploit Lemma 1. This lemma states that the spectral gaps of the reprojected transition kernels $\overline{K}$ of Algorithms 4 and 5 are at least as large as the spectral gaps of the transition kernels K of Algorithms 1 and 2, respectively. In order to describe a dimension-independent spectral gap, we introduce the following notation: Given $\mu_0 = \mathrm{ACG}(C)$ with nondegenerate, trace-class covariance operator $C : \mathbb{H} \to \mathbb{H}$ on an infinite-dimensional separable Hilbert space $\mathbb{H}$, let $\{e_j | j \in \mathbb{N}\}$ be a complete orthonormal system in $\mathbb{H}$ consisting of the eigenvectors of C. We now construct finite-dimensional approximations to the infinite-dimensional setting as follows: For $d \in \mathbb{N}$, let $\mu_0^{(d)} = \mathrm{ACG}(C_d)$ denote the ACG measure on $\mathbb{S}^{d-1}$ resulting from the marginal of $\mathrm{N}(0, C)$ on $\mathrm{span}\{e_1, \ldots, e_d\}$ and consider the target measure $\mu^{(d)}$ on $\mathbb{S}^{d-1}$ given by

$$\frac{\mathrm{d}\mu^{(d)}}{\mathrm{d}\mu_0^{(d)}}(\bar{x}) \propto \exp(-\overline{\Phi}(\bar{x})), \quad \bar{x} \in \mathbb{S}^{d-1}. \tag{19}$$

In order to apply $\overline{\Phi}$ to $\bar{x} \in \mathbb{S}^{d-1}$, we view $\mathbb{S}^{d-1}$ as the "equatorial" subsphere $\{\bar{x}_1 e_1 + \ldots + \bar{x}_d e_d \in \mathbb{H} : \bar{x} = (\bar{x}_1, \ldots, \bar{x}_d) \in \mathbb{S}^{d-1}\}$ of $\mathbb{S} \subset \mathbb{H}$. Let

$$\frac{\mathrm{d}\nu^{(d)}}{\mathrm{d}\nu_0^{(d)}}(x) \propto \exp(-\Phi(x)), \quad x \in \mathbb{R}^d, \tag{20}$$

where $\nu_0^{(d)} = \mathrm{N}(0, C_d)$ and $\Phi(x) := \overline{\Phi}(\Pi^{\mathbb{S}}(x))$ for $x \in \mathbb{H}$, as in (5). In order to apply Φ to $x \in \mathbb{R}^d$, we view $\mathbb{R}^d$ as the subspace $\{x_1 e_1 + \ldots + x_d e_d \in \mathbb{H} : x = (x_1, \ldots, x_d) \in \mathbb{R}^d\}$ of $\mathbb{H}$.

For a reminder of the definition of $\mathrm{gap}_\mu(K)$ for a given measure μ and transition kernel K we refer to (B4). By Lemma 1 we obtain the following result.

Proposition 4. *Let $\mu^{(d)}$ and $\nu^{(d)}$ be as in (19) and (20), respectively. Let $K^{(d)}$ denote the pCN-MH transition kernel targeting $\nu^{(d)}$ using a step size $s \in (0, 1]$ in the proposal. If there exists a $\beta > 0$ such that*

$$\inf_{d \in \mathbb{N}} \mathrm{gap}_{\nu^{(d)}}\left(K^{(d)}\right) \geqslant \beta, \tag{21}$$

then the reprojected pCN-MH transition kernel $\overline{K}^{(d)} := (\Pi^{\mathbb{S}})_\sharp K^{(d)}$ targeting $\mu^{(d)}$ on $\mathbb{S}^{d-1}$ satisfies

$$\inf_{d \in \mathbb{N}} \mathrm{gap}_{\mu^{(d)}}\left(\overline{K}^{(d)}\right) \geqslant \beta. \tag{22}$$

The same statement holds for the reprojected ESS transition kernel $\overline{K}^{(d)} := (\Pi^{\mathbb{S}})_\sharp K^{(d)}$.

Dimension independence of the spectral gap of the ESS transition kernel has been demonstrated in the literature by numerical experiments (Natarovskii et al., 2021a). However, to the best of our knowledge, no theoretical proof is available. Therefore, we focus on the pCN-MH algorithm, for which (21) was shown by Hairer et al. (2014) under certain assumptions on Φ. For convenience, we summarise their result:

Theorem 2. *Let $K^{(d)}$ denote the pCN-MH transition kernel targeting $\nu^{(d)}$ using the step size $s \in (0,1]$ in the proposal. Suppose the following conditions hold*:

(a) There exist some $R > 0$ and $\underline{\alpha} \in \mathbb{R}$ such that, for all $x \in \mathbb{H}$ with $\|x\| > R$,

$$\Phi(y) < \Phi(x) - \_\alpha \qquad \text{for all } \ y \in \mathbb{H} \text{ with } \left\| y - \sqrt{1-s^2}x \right\| \leqslant \frac{1}{2}\left(1 - \sqrt{1-s^2}\right)\|x\|.$$

(b) The function $e^{-\Phi}$ is integrable with respect to $\nu_0 = N(0, C)$.

(c) For every $\gamma > 0$ there exists some $C_\gamma < \infty$ such that

$$|\Phi(x) - \Phi(y)| \leqslant C_\gamma e^{\gamma r} \qquad \text{for all } \ x, y \in \mathbb{H} \text{ with } \ \|x\|, \|y\| \leqslant r.$$

Then there exists a $\beta > 0$ such that (21) holds.

The conditions of Theorem 2 are satisfied for constant Φ. Thus, the pCN-MH transition kernel exhibits a dimension-independent spectral gap when targeting the prior $\nu = \nu_0 = \mathrm{N}(0, C)$. By means of results by Vollmer (2015) and Rudolf and Sprungk (2018), this can then be lifted to bounded perturbations of the prior measure, such as μ as in (5) for bounded $\overline{\Phi}$. This yields our final result.

Theorem 3. *Let $\overline{\Phi} : \mathbb{S}^{d-1} \to \mathbb{R}$ be uniformly bounded. Then the transition kernel corresponding to the Markov chain realized by Algorithm 4 has a dimension-independent spectral gap in the sense of (22).*

Proof. Let $K^{(d)}$ denote the pCN-MH transition kernel in $\mathbb{R}^d$ for an arbitrary step size $s \in (0,1]$ and dimension $d \in \mathbb{N}$ with target $\nu^{(d)}$, and let $K_0^{(d)}$ denote the pCN-MH transition kernel that targets the prior $\nu_0^{(d)}$. Note that $K_0^{(d)}$ coincides with the corresponding proposal kernel. By Theorem 2 we know that there exists a $\beta > 0$ such that

$$\inf_{d \in \mathbb{N}} \mathrm{gap}_{\nu_0^{(d)}}\left(K_0^{(d)}\right) \geqslant \beta.$$

By Rudolf and Sprungk (2018, theorem 11) the transition operator $K^{(d)}$ is positive, and hence,

$$\mathrm{gap}_{\nu_0^{(d)}}\left(K_0^{(d)}\right) = 1 - \Lambda_{\nu_0^{(d)}}\left(K_0^{(d)}\right), \qquad \mathrm{gap}_{\nu^{(d)}}\left(K^{(d)}\right) = 1 - \Lambda_{\nu^{(d)}}\left(K^{(d)}\right),$$

where $\Lambda_\mu(K)$ denotes the supremum of the spectrum of the restriction of a μ-invariant transition operator K to $L^2_{0,\mu}$, where $L^2_{0,\mu} := \{f \in L^2(\mu) | \int f(u)\mu(\mathrm{d}u) = 0\}$, and $L^2(\mu) := \{f : \mathbb{H} \to \mathbb{R} | \ \|f\|_{L^2_\mu} := \int |f|^2 \mu(\mathrm{d}u) < \infty\}$. In the reversible case, $\Lambda_\mu(K) = \sup_{f \in L^2_{0,\mu}} \langle Kf, f\rangle_{L^2_\mu} / \|f\|_{L^2_\mu}$. Now, if $\overline{\Phi}$ is bounded, then so is Φ. A comparison result (Vollmer, 2015, theorem 3.3) states that

$$1 - \Lambda_{\nu^{(d)}}\left(K^{(d)}\right) \geqslant \exp\left(4\inf\overline{\Phi} - 4\sup\overline{\Phi}\right)\left(1 - \Lambda_{\nu_0^{(d)}}\left(K_0^{(d)}\right)\right).$$

Applying this comparison result yields

$$\inf_{d \in \mathbb{N}} \mathrm{gap}_{\nu^{(d)}}\left(K^{(d)}\right) \geqslant \exp\left(4\inf\overline{\Phi} - 4\sup\overline{\Phi}\right)\beta > 0,$$

which by Proposition 4 yields the statement. ∎

4 | NUMERICAL ILLUSTRATIONS

We now demonstrate the dimension-independent performance of the reprojected pCN-MH algorithm and the reprojected ESS algorithm on two applications. We state these algorithms in Algorithm 4 and Algorithm 5, respectively. Section 4.1, rooted in inverse problems, treats an application to Bayesian binary classification or level set inversion. Section 4.4 has a more statistical flavour and considers an application to Bayesian density estimation. Readers whose main interest is in the second application may proceed directly to Section 4.4 but may wish to recall the definition (30) of the root mean square jump distance with respect to the Riemannian metric on the sphere. We will use these applications to illustrate the dimension-dependent performance of the geodesic random walk-MH algorithm of Mangoubi and Smith (2018) and the MH algorithm of Zappa et al. (2018), which we state in Algorithm 6 and Algorithm 7, respectively.

4.1 | Bayesian binary level set inversion

For the convenience of the reader, we give a self-contained description of (Bayesian) binary level set inversion, following the presentation of Iglesias et al. (2016). Readers who are interested in technical details concerning the random fields perspective of level set inversion may consult appendix 2 of that paper. For simplicity, we focus on the single-phase Darcy flow model or groundwater flow problem on a bounded domain $D \subset \mathbb{R}^k$, $k = 1, 2, 3$, with closure $\overline{D}$ and boundary ∂D. This problem is described by the elliptic partial differential equation (PDE)

$$-\nabla \cdot (e^u \nabla p) = f \quad \text{in} \quad D, \tag{23a}$$

$$p = \kappa \quad \text{on} \quad \partial D, \tag{23b}$$

where p denotes a fluid pressure field, f describes sources and sinks, and u is the log-permeability parameter. Let $\mathcal{U} := L^\infty(D)$ and $\mathcal{V} := \{\tilde{p} \in H^1(D) | \tilde{p} = \kappa \text{ on } \partial D\}$, where $H^1(D)$ is the Sobolev space of functions in $L^2(D)$ whose first-order weak derivatives have finite $L^2(D)$ norm. If $u \in \mathcal{U}$, then $e^u \in L^\infty(D)$, and given a suitable boundary condition κ and source term $f \in L^2(D)$, a unique weak solution $p \in \mathcal{V}$ of (23) exists. Denote the solution map that maps the log-permeability u to the corresponding solution p by $\Upsilon : \mathcal{U} \to \mathcal{V}$. Then Υ is locally Lipschitz continuous, see for example, Bonito et al. (2017).

We assume now that the domain D is divided into two disjoint regions $D_0, D_1 \subset D$, that is, $\overline{D} = D_0 \cup D_1$. The subdomains D_i describe the location of different materials, for example, background and abnormal material, with different constant log-permeabilities $u_0, u_1 \in \mathbb{R}$. In particular, in binary level set inversion we assume that u takes the form

$$u(t) = u_0 \mathbb{1}_{D_0}(t) + u_1 \mathbb{1}_{D_1}(t), \qquad t \in D, \tag{24}$$

where $\mathbb{1}_{D_i}$ denotes the indicator function of D_i and where the values u_0, u_1 are known a priori. The goal is then to infer the location of D_1 and, hence, $D_0 = \overline{D} \setminus D_1$ based on noisy observations of p, that is,

$$y = O \circ \Upsilon(u) + \eta, \tag{25}$$

where $O : \mathcal{V} \to \mathbb{R}^J$ denotes for some $J \in \mathbb{N}$ a bounded linear observation operator and η describes observational noise. We assume that $\eta \sim \mathrm{N}(0, \Sigma)$ with known covariance $\Sigma \in \mathbb{R}^{J\times J}$.

In the level set approach we then introduce a so-called *level set function* $g \in \mathbb{S}(L^2(D)) \cap C(\overline{D})$ as well as the *level set map* $\mathcal{L} : C(\overline{D}) \to \mathcal{U}$ defined by

$$g \mapsto \mathcal{L}(g) := u_0 \mathbb{1}_{D_0(g)} + u_1 \mathbb{1}_{D_1(g)}, \qquad D_1(g) := \{t \in D | g(t) \geqslant 0\}, \tag{26}$$

and $D_0(g) = \{t \in D | g(t) < 0\}$, respectively. We can formulate the level set inverse problem as the problem of inferring the data-generating level set function g, given the model that the observations $y = O \circ \Upsilon \circ \mathcal{L}(g) + \eta$ are generated by a unique $g \in \mathbb{S}(L^2(D)) \cap C(\overline{D})$. We argue that the unit sphere in the function space $L^2(D)$ or $C(\overline{D})$ is an advantageous setting for the level set inverse problem: if we considered the problem in the ambient space $g \in C(\overline{D})$, then g becomes nonidentifiable, because $\mathcal{L}(\alpha g) = \mathcal{L}(g)$ for any $\alpha > 0$. For computational convenience, below we choose to work with the unit sphere in the Hilbert space $L^2(D)$ instead of the unit sphere in the Banach space $C(\overline{D}) \subset L^2(D)$.

For the Bayesian approach to level set inversion, we use a prior for the level set function g in the form of a series expansion

$$g = g(\overline{X}) = \sum_{i=1}^{\infty} \overline{X}_i \phi_i, \tag{27}$$

where the $\phi_i \in C(\overline{D})$ are given and form an orthonormal system in $L^2(D)$, and the $\overline{X}_i$ are random coefficients such that $\overline{X} = (\overline{X}_i)_{i\in\mathbb{N}} \in \mathbb{S}(\ell^2)$ almost surely; see for example Iglesias et al. (2016) and Dunlop et al. (2017). Then we have $g(\overline{X}) \in \mathbb{S}(L^2(D))$ almost surely. A common choice for the system $\{\phi_i | i \in \mathbb{N}\}$ are the eigenfunctions of a covariance operator $C : L^2(D) \to L^2(D)$ given by $C\phi(t) = \int_D c(s,t)\, \phi(s)\, \mathrm{d}s$ and a continuous covariance function $c \in C(\overline{D} \times \overline{D})$. The latter then yields $\phi_i \in C(\overline{D})$ for every $i \in \mathbb{N}$, and by Mercer's theorem we have $g(\overline{X}) \in C(\overline{D})$ almost surely. In summary, we obtain a reformulation of the original level set inverse problem, as the problem of inferring the sequence $\overline{x} \in \mathbb{S}(\ell^2)$ that corresponds to the data-generating level set function $g(\overline{x})$, given the noisy data $y = O \circ \Upsilon \circ \mathcal{L}(g(\overline{x})) + \eta$. For the prior μ_0 on $\mathbb{S}(\ell^2)$ in (4) we consider the ACG measure $\mu_0 = \mathrm{ACG}(\Lambda)$ where $\Lambda = \mathrm{diag}(\lambda_i : i \in \mathbb{N})$ involves the eigenvalues $(\lambda_i)_{i\in\mathbb{N}}$ of the covariance operator C, that is, $C\phi_i = \lambda_i \phi_i$. The associated distribution of $g(\overline{X})$, $\overline{X} \sim \mathrm{ACG}(\Lambda)$, on $\mathbb{S}(L^2(D))$ is then $\mathrm{ACG}(C)$.

Remark 3 (Boundedness of $\overline{\Phi}$). Given that $\eta \sim \mathrm{N}(0, \Sigma)$, the negative log-likelihood $\overline{\Phi}$ for Bayesian level set inversion on $\mathbb{S}$ takes the form

$$\overline{\Phi}(\overline{x}) := \frac{1}{2} \left\| \Sigma^{-1/2} \left(y - O \circ \Upsilon \circ \mathcal{L}(g(\overline{x})) \right) \right\|^2.$$

Since the range of $\mathcal{L}$ is bounded in $L^\infty(D)$, i.e. $\|\mathcal{L}(g)\|_{L^\infty(D)} \leqslant \max_i |u_i|$, and since Υ and O are locally Lipschitz continuous and bounded respectively, we obtain that also $\overline{\Phi}$ is bounded on $\mathbb{S}(\ell^2)$. Thus, the setting of Bayesian level set inversion satisfies the assumptions of Theorem 1 and Theorem 3. In particular, for finite-dimensional approximations of the Bayesian level set inversion problem that are obtained by truncating the expansion (27) after d terms, Algorithms 4, 5, and 7 yield uniformly ergodic Markov chains on $\mathbb{S}^{d-1}$ targeting the corresponding posterior $\mu^{(d)}$. Here, the corresponding posterior $\mu^{(d)}$ on $\mathbb{S}^{d-1}$ may be obtained according to (19) or (20). To the

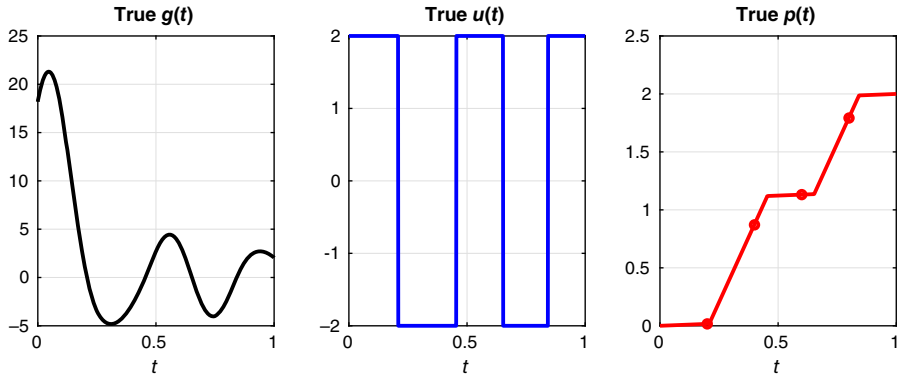

FIGURE 4 True $g^\dagger$, $u^\dagger$, $p^\dagger$, and true observations $o^\dagger = (p^\dagger(0.2),\dots,p^\dagger(0.8))^\top$.

best of our knowledge, uniform ergodicity on $\mathbb{S}^{d-1}$ of the Markov chain generated by Algorithm 6 has not been established.

Problem setting

We consider the elliptic problem (23) on $D = [0,1]$ with Dirichlet boundary conditions $p(0) = 0$ and $p(1) = 2$. For u we assume a form as in (24) based on a decomposition of D into two subregions D_0, D_1 with corresponding values $u_0 = -2$ and $u_1 = 2$.

For the level set function g such that $u = \mathcal{L}(g)$ we assume a series expansion (27) based on the eigensystem of the covariance function

$$c(s,t) = \left(1 + \sqrt{3}\frac{|t-s|}{0.1}\right)\exp\left(-\sqrt{3}\frac{|t-s|}{0.1}\right), \tag{28}$$

that is, a Whittle–Matérn covariance with variance $\sigma^2 = 1$, correlation length $\rho = 0.1$ and smoothness $\nu = 1.5$. For computational reasons we truncate the representation (27) after d terms and then infer the d coefficients $\bar{x}_1,\dots,\bar{x}_d$ given noisy observations of $p(0.2)$, $p(0.4)$, $p(0.6)$, and $p(0.8)$.

The assumed noise model is $\eta \sim \mathrm{N}(0,\Sigma)$ where $\Sigma = \mathrm{diag}(\sigma_1^2,\dots,\sigma_4^2)$ and $\sigma_i^2 = p^\dagger(0.2i)/10$ where $p^\dagger$ denotes the "true" solution resulting from the "true" coefficient vector $x^\dagger = (1,2,3,4,5,1,1,1,0,\dots,0)$ in the Karhunen–Loève expansion (27) of $g^\dagger$. For an illustration of $g^\dagger$, $u^\dagger$ and $p^\dagger$, see Figure 4.

Given $\eta \sim \mathrm{N}(0,\Sigma)$ the negative log-likelihood for observed data $y \in \mathbb{R}^4$ is then

$$\Phi(x) := \frac{1}{2}\sum_{j=1}^{4}\sigma_j^{-2}\left|y_j - F_j(x)\right|^2, \qquad x \in \mathbb{R}^d,$$

where F_j denotes the forward mapping $x = (x_i)_{i=1}^d \mapsto g \mapsto u \mapsto p \mapsto p(0.2j)$. As described in the previous subsection, $F_j(\alpha x) = F_j(x)$ for every $x \in \mathbb{R}^d$ and $\alpha > 0$. Thus, Φ is invariant under the radial projection map $\Pi^{\mathbb{S}}$, that is, $\Phi \circ \Pi^{\mathbb{S}} = \Phi$, and we can consider Bayesian level set inversion on the sphere $\mathbb{S}^{d-1}$ with corresponding prior

$$\mu_0 = \mathrm{ACG}(\Lambda_d), \qquad \Lambda_d = \mathrm{diag}(\lambda_1,\dots,\lambda_d).$$

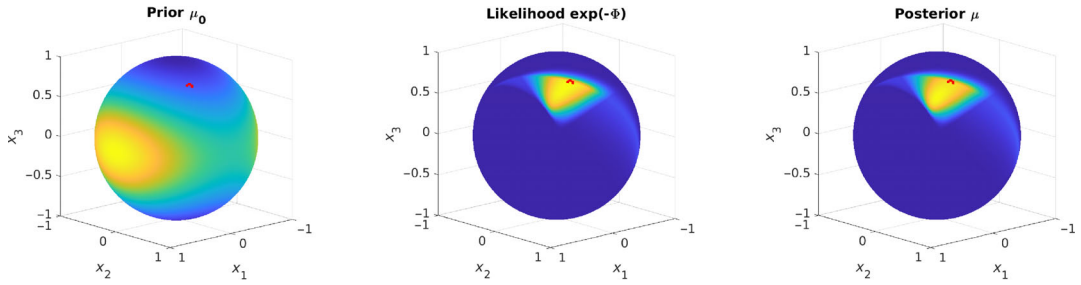

FIGURE 5 Prior, likelihood and posterior for dimension $d = 3$; the red spot indicates the (projected) truth $\overline{x}^{\dagger}$.

The posterior μ then takes the form (4) with $\overline{\Phi}(\overline{x}) = \Phi(\overline{x})$. Figure 5 illustrates the prior, the likelihood and the resulting posterior for dimension $d = 3$.

Remark 4. The eigenpairs (λ_j, ϕ_j) of the covariance operator C associated to c in (28) are computed numerically via a discretization of D and C using a grid size of length $\delta t = 10^{-3}$. The elliptic problem (23) is solved numerically using the same discretization. Note that the solution p of (23) on $D = [0, 1]$ is given by $\Upsilon(u)(t) = p(t) = 2S_t(e^{-u})/S_1(e^{-u})$ with $S_t(f) = \int_0^t f(s)ds$. We evaluate S_t using the trapezoidal rule on the given grid.

MCMC on the sphere

We now apply the four MCMC algorithms described in Section 3 in order to sample approximately from the posterior μ in various dimensions d. In particular, we aim to compute the posterior expectation of the following quantity of interest:

$$f(\overline{x}) = \left(\int_D \exp(-u(t, \overline{x}))\, \mathrm{d}t \right)^{-1}, \tag{29}$$

where $u(\cdot, \overline{x}) = \mathcal{L}(g(\overline{x}))$. We may interpret $f(\overline{x})$ as the effective homogenized permeability field over the one-dimensional domain D; see for example Alexanderian (2015, section 2).

First, we show in Figure 6 the thinned realizations $\overline{x}_{100k}$, $k = 1, \ldots, 100$ of the Markov chains generated by the reprojected pCN-MH algorithm, the geodesic random walk-MH algorithm based on Mangoubi and Smith (2018), and the MH algorithm of Zappa et al. (2018) for $d = 3$, subsampled every 100 steps.

All three MH algorithms were tuned to an average acceptance rate of roughly 23%. We ran the algorithms for another 10^6 iterations after a burn-in period of $5 \cdot 10^4$ iterations. All three runs yielded similar estimates for the posterior expectation of f. We provide the corresponding estimate plus/minus the half-length of a 95% confidence interval based on asymptotic variance estimates via the empirical autocorrelation functions of $(f(\overline{x}_k))_k$:

$$\begin{aligned} \text{reprojected pCN-MH:} &\quad 0.420 \pm 1.299 \cdot 10^{-3} \\ \text{geodesic random walk-MH:} &\quad 0.419 \pm 1.271 \cdot 10^{-3} \\ \text{MH by Zappa et al.:} &\quad 0.420 \pm 1.478 \cdot 10^{-3}. \end{aligned}$$

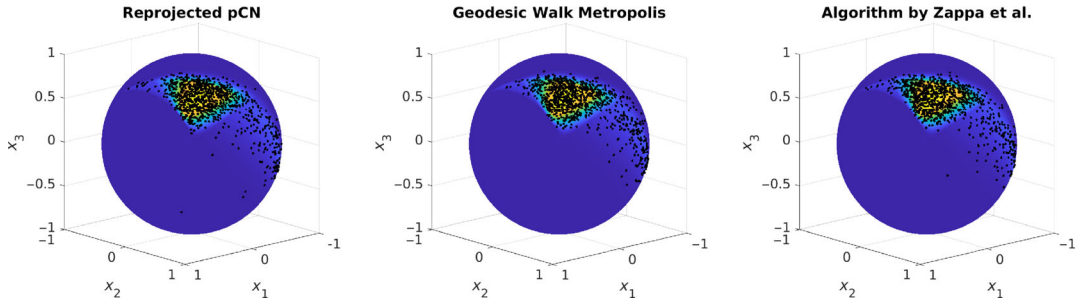

FIGURE 6 Thinned realizations $\bar{x}_{100k}$, $k = 1,\ldots,100$ of the Markov chains generated by Algorithms 4, 6, and 7 (left, centre, and right, respectively) for $d = 3$, subsampled every 100 steps.

All three estimates exhibit similar accuracies. Recall the root mean squared jump distance with respect to the Riemannian metric on the sphere given by

$$\text{RMSJD} := \sqrt{\frac{1}{n-1}\sum_{k=1}^{n-1} d_{\mathbb{S}}^2(\bar{x}_k, \bar{x}_{k+1})}. \tag{30}$$

In dimension $d = 3$, the three MH algorithms yielded similar estimates for the root mean squared jump distance:

reprojected pCN-MH:	0.202
geodesic random walk-MH:	0.234
MH by Zappa et al.:	0.185.

Next, we tested all four algorithms, including now the reprojected ESS algorithm, for increasing dimensions $d = 10, 20, 40, 80, 160, 320, 640$. In particular, we display the following quantities in Figure 7:

(i) the estimated posterior expectation of f given by the arithmetic mean of $(f(\bar{x}_k))_k$;
(ii) the estimated integrated autocorrelation time of the (approximately stationary) time series $(f(\bar{x}_k))_k$ as a measure for the MCMC error for computing the posterior expectation of f;
(iii) the root mean squared jump distance as a measure of how well the Markov chain explores the sphere.

An important observation from Figure 7 is that the two reprojected MCMC methods show dimension-independent efficiency in terms of integrated autocorrelation time and root mean squared jump distance. In contrast, the two MH algorithms relying on the surface measure as reference measure—namely, the geodesic random walk-MH algorithm based on Mangoubi and Smith (2018) and the MH algorithm of Zappa et al. (2018)—shows a clear decrease in efficiency as the dimension d of the state space increases. In particular, these methods lead to less accurate estimates of the posterior mean; see the left plot in Figure 7. While it seems that the ESS algorithm yields a higher efficiency, the higher efficiency comes at an increased cost: on average, the ESS algorithm required ≈ 3.8 tries until it hit the level set. Thus, the computational cost of the ESS algorithm was roughly four times higher than the computational cost of the pCN algorithm.

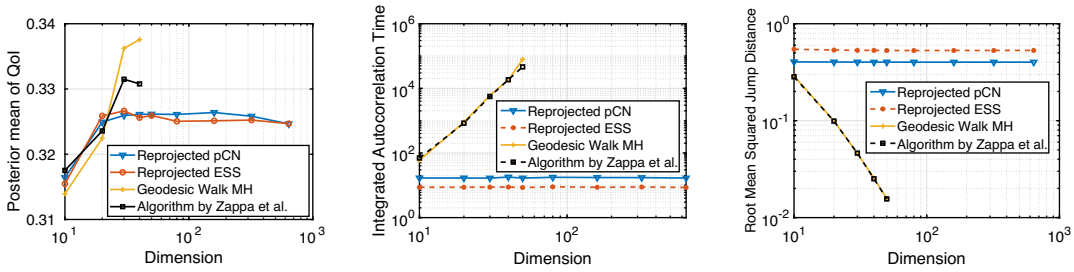

FIGURE 7 Estimated posterior mean (left), integrated autocorrelation time (middle) for f as in (29) and root mean squared jump distance (right) for Algorithms 4, 5, 6, and 7 as applied to the Bayesian level set inverse problem.

4.2 | Bayesian density estimation

We now consider a second application of Bayesian inference on high-dimensional spheres. We follow the approach of Holbrook et al. (2020) to nonparametric Bayesian density estimation: Given data $y_1,\dots,y_n \in D$, for D a bounded smooth domain in $\mathbb{R}^k$, we infer the Lebesgue probability density function p of the data, where p belongs to

$$\mathrm{P} := \left\{ p : D \to [0,\infty) \,\middle|\, \int_D p(y)\, \mathrm{d}y = 1 \right\}.$$

The set P is the unit simplex in the Banach space $L^1(D)$. Instead of inferring p directly, we instead infer the *square root* $g = \sqrt{p}$ of p, where g belongs to

$$\mathrm{Q} := \left\{ g : D \to \mathbb{R} \,\middle|\, \int_D g^2(y)\, \mathrm{d}y = 1 \right\}.$$

The set Q coincides with the unit sphere of the function space $L^2(D)$.

Data

We use the British coal mine disaster dataset that was studied in Holbrook et al. (2020). This dataset consists of the dates of 191 disasters recorded between March 1851 and March 1962 which can be found in Hand et al. (1994, data set 204). We aim to estimate the underlying probability density between the years 1850 and 1965. However, for computations and simplicity, we scale the data to lie within $D = [0,1]$ by a suitable affine transformation. Thus, we would like to infer a square root density $g \in \mathbb{S} = \mathrm{Q}$, where $\mathbb{S}$ denotes the unit sphere of $L^2([0,1])$.

Prior and posterior

For numerical discretization and constructing a prior model for g, we expand g with respect to a suitable orthonormal system $\{\phi_i | i \in \mathbb{N}\}$ in $L^2(D)$, $D = [0,1]$. Here, we choose the same system

as in Holbrook et al. (2020). which is based on a mean-zero Gaussian process model for g with a Whittle–Matérn covariance (cf. (28)):

$$g = g(\overline{X}) = \sum_{i=1}^{\infty} \overline{X}_i \, \phi_i, \qquad \phi_1 \equiv 1, \phi_i(y) = \sqrt{2}\cos(\pi \, (i-1) \, y), i > 1, \tag{31}$$

where $\overline{X} = (\overline{X}_i)_{i\in\mathbb{N}}$ are suitable random coefficients almost surely belonging to the unit sphere in ℓ^2, such that the resulting $g = g(\overline{X})$ belongs to the unit sphere in $L^2([0,1])$. Here, we assume as prior for $\overline{X}$ that $\overline{X} \sim \mathrm{ACG}(\Lambda)$ where $\Lambda = \mathrm{diag}(\lambda_i : i \in \mathbb{N})$ with $\lambda_i = \sigma^2(\kappa + \pi^2(i-1)^2)^{-r}$, $i \geqslant 1$. We choose $\sigma = 0.5$, $\kappa = 0.1$ and $r = 1$ for our experiments. We chose these parameter values because of their similarity to the values used by Holbrook et al. (2020). The resulting prior for g is then $\mathrm{ACG}(C)$ where $C : L^2(D) \to L^2(D)$ denotes the covariance operator $Cv(y) = \int_D c(y,y')v(y') \, \mathrm{d}y'$ using the corresponding covariance function

$$c(y,y') = \sum_{i=0}^{\infty} \lambda_i \, \phi_i(y) \, \phi_i(y').$$

Given $g \in \mathrm{Q}$, the likelihood $L(y;g)$ for observing the data $y = (y_1,\ldots,y_n) \in D^n$ is

$$L(y;g) = \prod_{i=1}^{n} g^2(y_i) = \exp\left(\sum_{j=1}^{n} \log(g^2(y_j))\right) = \exp\left(2\sum_{j=1}^{n} \log(|g(y_j)|)\right).$$

Thus, using the series representation in (31), we obtain as likelihood for y given coefficients $\overline{x}$ in the unit sphere of ℓ^2,

$$L(y;\overline{x}) = \exp\left(-\overline{\Phi}(\overline{x})\right), \qquad \overline{\Phi}(\overline{x}) := -2\sum_{j=1}^{n} \log\left(\left|\sum_{i=0}^{\infty} \overline{x}_i \, \phi_i(y_j)\right|\right).$$

Note that $\overline{\Phi} : \ell^2 \to \mathbb{R}$ is bounded. Given the data y, the resulting posterior for the coefficients $\overline{X}$ follows the form (4) with $\mu_0 = \mathrm{ACG}(\Lambda)$, and the assumptions of Theorem 1 are satisfied.

A quantity of interest for this problem is the posterior expectation for the probability mass of p between 0.435 and 0.574. This quantity of interest is the probability of a coal mine disaster between the years 1900 and 1916. It can be written as

$$f(\overline{x}) := \int_{0.435}^{0.574} g^2(y,\overline{x}) \, \mathrm{d}y = \left(\sum_{i,k=1}^{\infty} w_{i,k}\overline{x}_i \, \overline{x}_k\right)^2, \tag{32}$$

where $w_{i,k} := \int_{0.435}^{0.574} \phi_i(y) \, \phi_k(y) \, \mathrm{d}y$, i.e., f is a quadratic function of $\overline{x}$.

MCMC simulations

We truncated the expansion in (31) after d terms for $d = 10$, 20, 30, 40, 50, 100, 200, 400, and 800 and sampled approximately from the resulting truncated posterior $\mu^{(d)}$ on $\mathbb{S}^{d-1}$ for the coefficients $\overline{x}^{(d)} = (\overline{x}_1,\ldots,\overline{x}_d)$ using the prior $\mu_0^{(d)} = \mathrm{ACG}(\Lambda_d)$ with $\Lambda_d := \mathrm{diag}(\lambda_1,\ldots,\lambda_d)$. To this end, we applied Algorithms 4, 5, 6, and 7 and used them to compute the expectation of the quantity

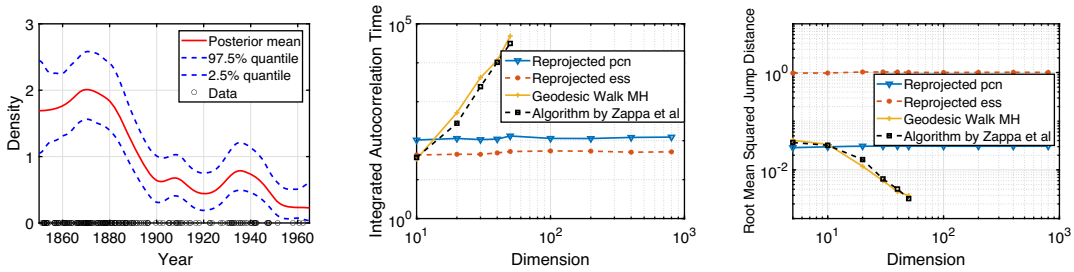

FIGURE 8 Estimated posterior mean and pointwise posterior quantiles for the density p based on truncation of (31) to the first $d = 30$ summands (left), estimated integrated autocorrelation time for quantity of interest f as in (32) (middle), and root mean squared jump distance (right) for Algorithms 4, 5, 6, and 7 as applied to the Bayesian density estimation problem.

f in (32) with respect to the truncated posterior $\mu^{(d)}$. After a burn-in period of 10^5 iterations, we ran the algorithms for 10^6 iterations and compared their efficiency. We quantified their efficiency in terms of the estimated integrated autocorrelation time for f and the root mean squared jump distance. We display the results in Figure 8. As in Figure 7, the results exhibit dimension-independent efficiency of the two reprojected MCMC methods, whereas the geodesic random walk-MH algorithm and the MH algorithm of Zappa et al. (2018) show a clear and drastic deterioration of efficiency as d increases.

5 | CLOSING REMARKS

In this paper, we proposed efficient MCMC algorithms for sampling target measures on high-dimensional spheres that are absolutely continuous with respect to an ACG prior. Our algorithms exploit the structure of the ACG prior by lifting the sampling problem on the sphere to a sampling problem in the ambient Hilbert space. This allows us to apply existing MCMC algorithms on linear spaces—such as the pCN-MH algorithm—for which there are theoretical results concerning dimension-independent efficiency.

Using the technique of push-forward Markov kernels, we then obtained transition kernels on the sphere that inherit many properties of the transition kernels in the ambient Hilbert space, for example, reversibility with respect to the desired target measure. Under fairly mild conditions, we showed the uniform ergodicity of Markov chains generated by our algorithms, and provided theoretical arguments for the dimension independence of their spectral gaps.

Using binary classification and Bayesian density estimation as test problems, we compared the performance of our methods to that of MH algorithms based on the geodesic random walk proposal of Mangoubi and Smith (2018) and the approach of Zappa et al. (2018). Our results illustrated the robustness of our algorithms as the dimension of the state space increased. In comparison, the statistical efficiency of the two other existing algorithms decreased as the dimension of the state space increased.

Based on our work, several interesting questions for future research remain. A theoretical analysis of the dimension independence of the (reprojected) ESS transition kernel remains an open issue. Here Markov chain comparison techniques—as have been developed by, for example, Peskun (1973), Andrieu and Vihola (2016), and Rudolf and Sprungk (2018)—may be useful for establishing the inheritance of a spectral gap from the pCN-MH transition kernel to the ESS

transition kernel. Additionally, it seems promising to modify our algorithms so that they can be applied to sample from target measures with respect to other common priors such as Bingham distributions, by using the acceptance-rejection approach of Kent et al. (2018), for example.

ACKNOWLEDGMENTS

The authors wish to thank Stephan Huckemann and Michael Habeck for helpful comments and, in particular, Andre Wibisono for the inspiring discussion which led to the proof of Theorem 3. In addition, the authors thank the associate editor and the two referees for their careful reading of the manuscript and helpful feedback. Open Access funding enabled and organized by Projekt DEAL.

ORCID

Han Cheng Lie https://orcid.org/0000-0002-6905-9903
Daniel Rudolf https://orcid.org/0000-0002-4492-5203
Björn Sprungk https://orcid.org/0000-0001-8359-6863
T. J. Sullivan https://orcid.org/0000-0003-4081-8301

REFERENCES

Alexanderian, A. (2015). Expository paper: A primer on homogenization of elliptic PDEs with stationary and ergodic random coefficient functions. *The Rocky Mountain Journal of Mathematics*, *45*, 703–735.

Andrieu, C., & Vihola, M. (2016). Establishing some order amongst exact approximations of MCMCs. *Annals of Applied Probability*, *26*, 2661–2696.

Asmussen, S., & Glynn, P. W. (2011). A new proof of convergence of MCMC via the ergodic theorem. *Statistics & Probability Letters*, *81*, 1482–1485.

Beskos, A., Girolami, M., Lan, S., Farrell, P. E., & Stuart, A. M. (2017). Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics*, *335*, 327–351.

Beskos, A., Pinski, F. J., Sanz-Serna, J. M., & Stuart, A. M. (2011). Hybrid Monte Carlo on Hilbert spaces. *Stochastic Processes and their Applications*, *121*, 2201–2230.

Beskos, A., Roberts, G., Stuart, A. M., & Voss, J. (2008). MCMC methods for diffusion bridges. *Stochastics and Dynamics*, *8*, 319–350.

Bogachev, V. I. (1998). *Gaussian measures*. American Mathematical Society.

Bonito, A., Cohen, A., DeVore, R., Petrova, G., & Welper, G. (2017). Diffusion coefficients estimation for elliptic partial differential equations. *SIAM Journal on Mathematical Analysis*, *49*, 1570–1592.

Brubaker, M., Salzmann, M., & Urtasun, R. (2012). *A family of MCMC methods on implicitly defined manifolds*. In N. D. Lawrence & M. Girolami (Eds.), *Proceedings of the fifteenth international conference on artificial intelligence and statistics. Proceedings of machine learning research* (Vol. *22*, pp. 161–172). PMLR.

Byrne, S., & Girolami, M. (2013). Geodesic Monte Carlo on embedded manifolds. *Scandinavian Journal of Statistics*, *40*, 825–845.

Chen, V., Dunlop, M. M., Papaspiliopoulos, O., & Stuart, A. M. (2018). Dimension-robust MCMC in Bayesian inverse problems. *arXiv:1803.03344*.

Chikuse, Y. (2003). *Statistics on special manifolds*. Springer.

Cotter, S. L., Roberts, G. O., Stuart, A. M., & White, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statistical Science*, *28*, 424–446.

Diaconis, P., Holmes, S., & Shahshahani, M. (2013). *Sampling from a manifold*. In *Advances in modern statistical theory and applications. A festschrift in honor of Morris L. Eaton* (pp. 102–125). Institute of Mathematical Statistics.

Douc, R., Moulines, E., Priouret, P., & Soulier, P. (2018). *Markov chains*. Springer.

Dunlop, M. M., Iglesias, M. A., & Stuart, A. M. (2017). Hierarchical Bayesian level set inversion. *Statistics and Computing*, *27*, 1555–1584.

Fan, J., Jiang, B., & Sun, Q. (2021). Hoeffding's inequality for general Markov chains and its applications to statistical learning. *Journal of Machine Learning Research*, *22*, 1–35 https://www.jmlr.org/papers/volume22/19-479/19-479.pdf

Franke, J., Redenbach, C., & Zhang, N. (2016). On a mixture model for directional data on the sphere. *Scandinavian Journal of Statistics*, *43*, 139–155.

Glover, J., & Mitro, J. (1990). Symmetries and functions of Markov processes. *Annals of Probability*, *18*, 655–668.

Goyal, N., & Shetty, A. (2019). *Sampling and optimization on convex sets in Riemannian manifolds of non-negative curvature*. In A. Beygelzimer & D. Hsu (Eds.), *Proceedings of the thirty-second conference on learning theory*. *Proceedings of machine learning research* (Vol. *99*, pp. 1519–1561). PMLR.

Habeck, M., Hasenpflug, M., Kodgirwar, S., & Rudolf, D. (2023). Geodesic slice sampling on the sphere. *arXiv:2301.08056*.

Hairer, M., Stuart, A. M., & Vollmer, S. J. (2014). Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *Annals of Applied Probability*, *24*, 2455–2490.

Hand, D. J., Daly, F., McConway, K., Lunn, D., & Ostrowski, E. (1994). *A handbook of small data sets*. Chapman & Hall.

Hasenpflug, M., Natarovskii, V., & Rudolf, D. (2023). Reversibility of elliptical slice sampling revisited. *arXiv:2301.02426*.

Holbrook, A., Lan, S., Streets, J., & Shahbaba, B. (2020). *Nonparametric fisher geometry with application to density estimation*. In J. Peters & D. Sontag (Eds.), *Proceedings of the 36th conference on uncertainty in artificial intelligence*. *Proceedings of machine learning research* (Vol. *124*, pp. 101–110). PMLR.

Iglesias, M. A., Lu, Y., & Stuart, A. M. (2016). A Bayesian level set method for geometric inverse problems. *Interfaces and Free Boundaries*, *18*, 181–217.

Kechris, A. S. (1995). *Classical descriptive set theory*. Springer-Verlag.

Kent, J. T., Ganeiber, A. M., & Mardia, K. V. (2018). A new unified approach for the simulation of a wide class of directional distributions. *Journal of Computational and Graphical Statistics*, *27*, 291–301.

Lan, S., Zhou, B., & Shahbaba, B. (2014). *Spherical Hamiltonian Monte Carlo for constrained target distributions*. In E. P. Xing & T. Jebara (Eds.), *Proceedings of the 31st international conference on machine learning*. *Proceedings of machine learning research* (Vol. *32*, pp. 629–637). PMLR.

Łatuszyński, K., Miasojedow, B., & Niemiro, W. (2013). Nonasymptotic bounds on the estimation error of MCMC algorithms. *Bernoulli*, *19*, 2033–2066.

Łatuszyński, K., & Niemiro, W. (2011). Rigorous confidence bounds for MCMC under a geometric drift condition. *Journal of Complexity*, *27*, 23–38.

Łatuszyński, K., & Rudolf, D. (2014). Convergence of hybrid slice sampling via spectral gap. *arXiv:1409.2709*.

Ley, C., & Verdebout, T. (2017). *Modern directional statistics*. CRC Press.

Li, Y., & Walker, S. G. (2023). A latent slice sampling algorithm. *Computational Statistics and Data Analysis*, *179*(107652), 15.

Mangoubi, O., & Smith, A. (2018). Rapid mixing of geodesic walks on manifolds with positive curvature. *Annals of Applied Probability*, *28*, 2501–2543.

Mardia, K. V., & Jupp, P. E. (2000). *Directional statistics*. John Wiley & Sons, Ltd.

Meyn, S., & Tweedie, R. L. (2009). *Markov chains and stochastic stability*. Cambridge University Press, second edn.

Murray, I., Adams, R., & MacKay, D. (2010). *Elliptical slice sampling*. In Y. W. Teh & M. Titterington (Eds.), *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. *Proceedings of machine learning research* (Vol. *9*, pp. 541–548). PMLR.

Natarovskii, V., Rudolf, D., & Sprungk, B. (2021a). *Geometric convergence of elliptical slice sampling*. In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th international conference on machine learning*. *Proceedings of machine learning research* (Vol. *139*, pp. 7969–7978). PMLR.

Natarovskii, V., Rudolf, D., & Sprungk, B. (2021b). Quantitative spectral gap estimate and Wasserstein contraction of simple slice sampling. *Annals of Applied Probability*, *31*, 806–825.

Neal, R. M. (1999). *Regression and classification using Gaussian process priors. (with discussion)*. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics 6. Proceedings of the sixth Valencia international meeting, 1998* (pp. 475–501). Clarendon Press.

Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, *31*, 705–767.

Novak, E., & Rudolf, D. (2014). *Computation of expectations by Markov chain Monte Carlo methods*. In S. Dahlke, W. Dahmen, M. Griebel, W. Hackbusch, K. Ritter, R. Schneider, C. Schwab, & H. Yserentant (Eds.), *Extraction*

of quantifiable information from complex systems. Lecture notes in computational science and engineering (pp. 397–411). Springer.
Paulin, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electronic Journal of Probability*, *20*, 1–32.
Peskun, P. H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika*, *60*, 607–612.
Rosenblatt, M. (1966). Functions of Markov processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *5*, 232–243.
Rudolf, D. (2012). Explicit error bounds for Markov chain Monte Carlo. *Dissertationes Mathematicae*, *485*, 1–93.
Rudolf, D., & Schweizer, N. (2015). Error bounds of MCMC for functions with unbounded stationary variance. *Statistics & Probability Letters*, *99*, 6–12.
Rudolf, D., & Sprungk, B. (2018). On a generalization of the preconditioned Crank–Nicolson Metropolis algorithm. *Foundations of Computational Mathematics*, *18*, 309–343.
Rudolf, D., & Sprungk, B. (2022). Robust random walk-like Metropolis–Hastings algorithms for concentrating posteriors. *arXiv:2202.12127*.
Srivastava, A., & Jermyn, I. H. (2009). Looking for shapes in two-dimensional cluttered point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*, 1616–1629.
Tabelow, K., Voss, H. U., & Polzehl, J. (2012). Modeling the orientation distribution function by mixtures of angular central Gaussian distributions. *Journal of Neuroscience Methods*, *203*, 200–211.
Tierney, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Annals of Applied Probability*, *8*, 1–9.
Tyler, D. E. (1987). Statistical analysis for the angular central Gaussian distribution on the sphere. *Biometrika*, *74*, 579–589.
Vollmer, S. J. (2015). Dimension-independent MCMC sampling for inverse problems with non-Gaussian priors. *SIAM/ASA Journal on Uncertainty Quantification*, *3*, 535–561.
Wang, F., & Gelfand, A. E. (2013). Directional data analysis under the general projected normal distribution. *Statistical Methodology*, *10*, 113–127.
Watson, G. S. (1983). *Statistics on spheres*. In *University of Arkansas Lecture notes in the mathematical sciences* (Vol. *6*). John Wiley & Sons, Inc..
Yang, J., Latuszynski, K., & Roberts, G. (2022). Stereographic Markov chain Monte Carlo. *arXiv.2205.12112*.
Zappa, E., Holmes-Cerfon, M., & Goodman, J. (2018). Monte Carlo on manifolds: Sampling densities and integrating functions. *Communications on Pure and Applied Analysis*, *71*, 2609–2647.

How to cite this article: Lie, H. C., Rudolf, D., Sprungk, B., & Sullivan, T. J. (2023). Dimension-independent Markov chain Monte Carlo on the sphere. *Scandinavian Journal of Statistics*, 1–41. https://doi.org/10.1111/sjos.12653

APPENDIX A. GAUSSIAN MEASURES AND ANGULAR CENTRAL GAUSSIAN MEASURES

A.1 Gaussian measures on Hilbert spaces

We first briefly recall some basic notions related to measures, especially Gaussian measures, on Hilbert spaces. A standard reference in this area is the book of Bogachev (1998).

In the following, we consider a separable Hilbert space $\mathbb{H}$ with norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$. Suppose $\mu \in \mathcal{P}(\mathbb{H})$ with finite second moment $\int_{\mathbb{H}} \|x\|^2 \, \mu(\mathrm{d}x)$ is given. Then the *mean element* $a \in \mathbb{H}$ and *covariance operator* $C : \mathbb{H} \to \mathbb{H}$ of μ are determined by

$$0 = \int_{\mathbb{X}} \langle v, x - a\rangle \, \mu(\mathrm{d}x) \qquad \text{for all } \; v \in \mathbb{H}, \tag{A1}$$

$$\langle Cu, v\rangle = \int_{\mathbb{X}} \langle u, x - a\rangle\langle v, x - a\rangle \, \mu(\mathrm{d}x) \qquad \text{for all } \; u, v \in \mathbb{H}. \tag{A2}$$

When $a = 0$, the measure μ is said to be *centred*.

In this paper, we focus on *Gaussian measures* $\mu = \mathrm{N}(a, C)$ on $\mathbb{H}$. Such measures are equivalently determined by the property that every one-dimensional linear image is Gaussian on $\mathbb{R}$, that is, identifying $\ell \in \mathbb{H}$ with the continuous linear functional $\mathbb{H} \to \mathbb{R}, x \mapsto \langle \ell, x \rangle$,

$$\mu = \mathrm{N}(a, C) \in \mathcal{P}(\mathbb{H}) \Leftrightarrow \forall \ell \in \mathbb{H}, \ell_\sharp \mu = \mathrm{N}(\langle \ell, a \rangle, \langle C\ell, \ell \rangle) \in \mathcal{P}(\mathbb{R});$$

or by the form of the characteristic function, that is,

$$\mu = \mathrm{N}(a, C) \in \mathcal{P}(\mathbb{H}) \Leftrightarrow \forall v \in \mathbb{H}, \int_{\mathbb{X}} \exp(i\langle v, x \rangle)\ \mu(\mathrm{d}x) = \exp\left(i\langle v, a \rangle - \frac{1}{2}\langle Cv, v \rangle\right).$$

The covariance operator C of a probability measure on $\mathbb{H}$ is always self-adjoint and positive semi-definite, and so its eigenvalues are all real and nonnegative. Furthermore, the assumption of finite second moment implies that these eigenvalues are summable, that is, C is a trace-class operator.

Given any self-adjoint and positive-definite operator $C : \mathbb{H} \to \mathbb{H}$, we define for $x, y \in \mathrm{ran}(C^{1/2})$ the weighted inner product and norm

$$\langle x, y \rangle_C := \langle C^{-1/2}x, C^{-1/2}y \rangle, \tag{A3}$$

$$\|x\|_C := \sqrt{\langle x, x \rangle_C}, \tag{A4}$$

If C is the covariance operator of a Gaussian measure $\mu = \mathrm{N}(a, C)$, then $\mathrm{ran}(C^{1/2})$ is the *Cameron–Martin space* of μ, and (A3) and (A4) are often called the *precision* inner product and norm, respectively.

The topological support of $\mathrm{N}(0, C)$—that is, the smallest closed set of full measure—is the closure in $\mathbb{H}$ of $\mathrm{ran}(C)$ (Bogachev, 1998, theorem 3.6.1). A sufficient condition for the density of $\mathrm{ran}(C)$ in $\mathbb{H}$ is the strict positivity of C, that is, that it has no null eigenvalue. The containments $\mathrm{ran}(C) \subseteq \mathrm{ran}(C^{1/2}) \subseteq \mathbb{H}$ always hold, and are strict when the Cameron–Martin space $\mathrm{ran}(C^{1/2})$ has infinite dimension.

A.2 Angular central Gaussian measures on spheres in Hilbert spaces

We consider the unit sphere $\mathbb{S} := \{x \in \mathbb{H} | \ \|x\| = 1\}$ in a separable Hilbert space $\mathbb{H}$. For $\mathbb{H} = \mathbb{R}^d$, we emphasise the dimension and denote the unit sphere by $\mathbb{S}^{d-1}$. The sphere $\mathbb{S}$ is equipped with the following metric $d_{\mathbb{S}} : \mathbb{S} \times \mathbb{S} \to [0, \pi]$:

$$d_{\mathbb{S}}(\bar{x}, \bar{y}) := \arccos\big(\langle \bar{x}, \bar{y} \rangle\big) = 2\arcsin\Big(\frac{1}{2}\|\bar{x} - \bar{y}\|\Big) \qquad \bar{x}, \bar{y} \in \mathbb{S}, \tag{A5}$$

where the second identity follows from elementary trigonometry. The definition of $d_{\mathbb{S}}$ by the arcsine function also extends to the unit sphere $\mathbb{S}$ in general Banach spaces. Moreover, it yields the Lipschitz equivalence—and hence the topological equivalence—of $d_{\mathbb{S}}$ and the metric on $\mathbb{S}$ induced by the norm $\|\cdot\|$ of the ambient space $\mathbb{H}$:

$$\|\bar{x} - \bar{y}\| \leqslant d_{\mathbb{S}}(\bar{x}, \bar{y}) \leqslant \frac{\pi}{2}\|\bar{x} - \bar{y}\| \qquad \forall \bar{x}, \bar{y} \in \mathbb{S}, \tag{A6}$$

since $g(r) = 2\arcsin\left(\frac{1}{2}r\right)/r$ varies between 1 and $\pi/2$ for $r \in [0, 2]$. Thus, the metric $d_{\mathbb{S}}$ is topologically equivalent to the norm of the ambient space on $\mathbb{S}$. We record this result in Lemma 2

below. According to Kechris (1995, proposition 3.3(ii)), a closed subset of a Polish space is always itself Polish in the relative (subspace) topology. Thus, $(\mathbb{S}, d_{\mathbb{S}})$ is a Polish space and $\mathcal{B}(\mathbb{S}) = \{B \cap \mathbb{S} | B \in \mathcal{B}(\mathbb{H})\}$.

Lemma 2. *The topology on $\mathbb{S}$ generated by $d_{\mathbb{S}}$ coincides with the relative topology on $\mathbb{S}$ generated by the norm $\|\cdot\|$ of $\mathbb{H}$.*

Proof. This is a consequence of the Lipschitz equivalence of the metrics (A6). ∎

Fix an arbitrary $\bar{z} \in \mathbb{S}$. Recall the radial projection to the sphere defined in (3),

$$\Pi^{\mathbb{S}} : \mathbb{H} \to \mathbb{S}, \quad \Pi^{\mathbb{S}}(x) := \begin{cases} \frac{x}{\|x\|}, & \text{if } x \neq 0, \\ \bar{z}, & \text{if } x = 0. \end{cases}$$

The choice of $\bar{z}$ is not important in what follows, but we fix $\bar{z}$ in order to ensure that $\Pi^{\mathbb{S}}$ is a measurable mapping from $\mathbb{H}$ into $\mathbb{S}$.

Recall that the ACG measure on $\mathbb{S}$ associated to a centred Gaussian measure $\mathrm{N}(0, C)$ on $\mathbb{H}$ is given by

$$\mu := \Pi^{\mathbb{S}}_{\sharp} \mathrm{N}(0, C),$$

and denoted by $\mu = \mathrm{ACG}(C)$. In other words, under $\mathrm{ACG}(C)$, each $E \in \mathcal{B}(\mathbb{S})$ is assigned the Gaussian $\mathrm{N}(0, C)$-measure of the cone $\{\alpha u \in \mathbb{H} | \alpha > 0, u \in E\}$. Note that $\mathrm{ACG}(C) = \mathrm{ACG}(\lambda C)$ for $\lambda \neq 0$ and so it can sometimes be useful to fix a normalization for C, for example, by taking its leading eigenvalue to be unity.

Corollary 1. *Let $C_E := \{\alpha e \in \mathbb{H} | \alpha > 0, e \in E\}$ denote the cone spanned by a subset $E \subseteq \mathbb{S}$. The set E is open in $(\mathbb{S}, d_{\mathbb{S}})$ if and only if C_E is open in $(\mathbb{H}, \|\cdot\|)$.*

Proof. Suppose that C_E is open in $\mathbb{H}$. Then $E = C_E \cap \mathbb{S}$ is, by Lemma 2, open in $(\mathbb{S}, d_{\mathbb{S}})$.

Let $\Pi^{\mathbb{S}}|_{\mathbb{H}\setminus\{0\}}$ be the restriction of the radial projection $\Pi^{\mathbb{S}}$ defined in (3) to $\mathbb{H} \setminus \{0\}$. For E in $(\mathbb{S}, d_{\mathbb{S}})$, $(\Pi^{\mathbb{S}}|_{\mathbb{H}\setminus\{0\}})^{-1}(E) = C_E$. Equip $\mathbb{H} \setminus \{0\}$ with the subspace topology. Then $\Pi^{\mathbb{S}}|_{\mathbb{H}\setminus\{0\}}$ is continuous, and if E is open, then C_E is open in $(\mathbb{H} \setminus \{0\}, \|\cdot\|)$, and hence in $(\mathbb{H}, \|\cdot\|)$. ∎

Whenever $\mathrm{N}(0, C)$ has support equal to $\mathbb{H}$, Corollary 1 immediately implies that the induced $\mathrm{ACG}(C)$ measure has support equal to $\mathbb{S}$; thus, in view of Appendix A.1, $\mathrm{ACG}(C)$ is a strictly positive measure on $\mathbb{S}$ whenever C is positive definite.

In the case where $\mathbb{H} = \mathbb{R}^d$ with the usual Euclidean norm and $C \in \mathbb{R}^{d\times d}$ is symmetric and positive definite, $\mu = \mathrm{ACG}(C)$ admits a density $\rho : \mathbb{S}^{d-1} \to [0, \infty)$ with respect to the $(d-1)$-dimensional Hausdorff measure on the sphere. This density is given by

$$\begin{aligned} \rho(\bar{x}) &= \int_0^\infty \frac{\exp\left(-\frac{1}{2}(r\bar{x}) \cdot C^{-1}(r\bar{x})\right)}{\sqrt{\det(2\pi C)}} r^{d-1} \, \mathrm{d}r \\ &= \frac{\Gamma(d/2)}{2\left(\frac{1}{2}\|\bar{x}\|_C^2\right)^{d/2}\sqrt{\det(2\pi C)}} \end{aligned}$$

$$= \frac{\Gamma(d/2)}{2\pi^{d/2}\sqrt{\det C}} \|\bar{x}\|_C^{-d}.$$

In the second equation, we used the fact that, for $a > 0$ and $n \in \mathbb{N}$, $\int_0^\infty a^{-r^2} r^{n-1}\, \mathrm{d}r = \frac{\Gamma(n/2)}{2(\log a)^{n/2}}$ (Tyler, 1987, equation (1)).

Remark 5. The above definitions and properties can be readily adapted to the image of a centred Gaussian measure on a projective space, that is, the quotient of $\mathbb{H}$ by the equivalence relation $u \sim v \Leftrightarrow u = \lambda v$ for some $\lambda \neq 0$, that is, $\mathbb{S}$ with antipodal points identified. Note that, while ACG(C) on $\mathbb{S}$ is always a multimodal distribution with at least two modes, its image on the projective space can be unimodal, which may be desirable in applications.

APPENDIX B. MARKOV CHAIN MONTE CARLO

Markov chain methods provide a standard tool for approximate sampling of complicated distributions, such as posterior distributions appearing in the Bayesian analysis of data. We recall here notions related to MCMC on a general state space $\mathbb{X}$ equipped with a target probability measure $\mu \in \mathcal{P}(\mathbb{X})$, the MH and slice sampling paradigms, and make some particular points about MCMC on infinite-dimensional Hilbert spaces.

B.1 General notions

By a *Markov kernel* on a topological space $\mathbb{X}$ we mean a function $K : \mathbb{X} \times \mathcal{B}(\mathbb{X}) \to [0, 1]$ such that $K(x, \cdot) \in \mathcal{P}(\mathbb{X})$ for each $x \in \mathbb{X}$, and $K(\cdot, A)$ is a measurable function for each $A \in \mathcal{B}(X)$. A sequence of random variables $(X_n)_{n\in\mathbb{N}}$, mapping from $(\Omega, \mathcal{A}, \mathbb{P})$ to $\mathbb{X}$, is a (*time-homogeneous*) *Markov chain* with *transition kernel* K on $\mathbb{X}$ if

$$\mathbb{P}(X_{n+1} \in A | X_1, \ldots, X_n) = \mathbb{P}(X_{n+1} \in A | X_n) = K(X_n, A), \quad n \in \mathbb{N},\ A \in \mathcal{B}(\mathbb{X}),$$

where K is a Markov kernel. In this paper, we shall distinguish between Markov kernels and transition kernels. A transition kernel is a Markov kernel associated to a time-homogeneous Markov chain. We abuse notation and also use K to denote the transition operator induced by the kernel K; the transition operator acts on functions $f : \mathbb{X} \to \mathbb{R}$ via

$$(Kf)(x) := \int_{\mathbb{X}} f(y)\, K(x, \mathrm{d}y) = \mathbb{E}\left[f(X_{n+1}) | X_n = x\right] \quad \text{for } x \in \mathbb{X}. \tag{B1}$$

The idea of MCMC is to construct a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition kernel K such that the distribution of X_n converges to μ as $n \to \infty$. Ideally, this convergence will be "fast" and the correlation between successive random variables X_n and X_{n+1} will also be weak.

A necessary condition for a transition kernel K to have μ as a limiting distribution is that μ is an *invariant distribution* of K, that is,

$$\mu(A) = \mu K(A) := \int_{\mathbb{X}} K(x, A)\, \mu(\mathrm{d}x), \quad \text{for all } A \in \mathcal{B}(\mathbb{X}). \tag{B2}$$

Reversibility (or *detailed balance*) of a transition kernel K on $\mathbb{X}$ with respect to μ refers to the property that

$$K(x, \mathrm{d}y)\mu(\mathrm{d}x) = K(y, \mathrm{d}x)\mu(\mathrm{d}y), \tag{B3}$$

that is, that the measure $K(x, \mathrm{d}y)\mu(\mathrm{d}x)$ on $\mathbb{X} \times \mathbb{X}$ is symmetric. If (B3) holds we say that K is μ-reversible. Reversibility of a transition kernel K with respect to μ implies that K has μ as an invariant distribution, although the converse is generally false.

If μ is an invariant distribution of K, and if some nonrestrictive regularity conditions such as φ-irreducibility and Harris recurrence hold, then a strong law of large numbers holds, that is,

$$\lim_{n\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(X_n) = \int_{\mathbb{X}} f(x)\ \mu(\mathrm{d}x) \quad \mathbb{P}\text{-a.s.},$$

for any μ-integrable $f : \mathbb{X} \to \mathbb{R}$ (Asmussen & Glynn, 2011; Meyn & Tweedie, 2009). This shows that the "MCMC-time average" $\frac{1}{N}\sum_{n=1}^{N} f(X_n)$ can be used to approximate the mean of f with respect to the distribution of interest μ. For more quantitative statements the spectral gap of a Markov chain (or its transition kernel) is a crucial quantity. For a transition kernel K that is reversible with respect to μ (inducing the transition operator K via (B1)) we define

$$\mathrm{gap}_\mu(K) := 1 - \|K\|_{L^2_{0,\mu} \to L^2_{0,\mu}}, \tag{B4}$$

where we recall that $L^2_{0,\mu} := \{f \in L^2(\mu) | \int f(u)\ \mu(\mathrm{d}u) = 0\}$ and where $\|K\|_{L^2_{0,\mu} \to L^2_{0,\mu}}$ denotes the norm of the operator K, which we view as an element of the space of bounded linear operators from $L^2_{0,\mu}$ to itself. An L^2_μ-spectral gap, that is, $\mathrm{gap}_\mu(K) > 0$, leads to desirable properties of a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition kernel K. For instance, it implies a central limit theorem, see for example Douc et al. (2018, section 22.5) and the relevant references therein. In particular, an explicit lower bound on $\mathrm{gap}_\mu(K)$ leads to an estimate of the total variation distance and a mean squared error bound of the time average $\frac{1}{N}\sum_{n=1}^{N} f(X_n)$. More precisely, it is well known, for example, by virtue of Novak and Rudolf (2014, lemma 2 and lemma 3), that

$$\|\xi K^n - \mu\|_{\mathrm{TV}} \leqslant (1 - \mathrm{gap}_\mu(K))^n \left\| \frac{\mathrm{d}\xi}{\mathrm{d}\mu} - 1 \right\|_{L^2_\mu}, \tag{B5}$$

where $\|\xi K^n - \mu\|_{\mathrm{TV}} := \sup_{A\in\mathcal{B}(\mathbb{X})} |\xi K^n(A) - \mu(A)|$ denotes the total variation distance between $\xi K^n = \mathbb{P}_{X_{n+1}}$ and μ, $\xi := \mathbb{P}_{X_1}$, and $\|f\|^p_{L^p_\mu} := \int_{\mathbb{X}} |f|^p\ \mathrm{d}\mu$ for $f : \mathbb{X} \to \mathbb{R}$. Furthermore, Rudolf (2012) shows that, for every $p > 2$, there exists an explicit constant c_p such that, for $f \in L^p(\mu)$ with $\|f\|^p_{L^p_\mu} \leqslant 1$,

$$\mathbb{E}\left[\left|\frac{1}{N}\sum_{n=1}^{N} f(X_n) - \int_{\mathbb{X}} f \mathrm{d}\mu\right|^2\right] \leqslant \frac{2}{N \cdot \mathrm{gap}_\mu(K)} + \frac{c_p \left\|\frac{\mathrm{d}\xi}{\mathrm{d}\mu} - 1\right\|_\infty}{N^2 \cdot \mathrm{gap}_\mu(K)^2}.$$

There are also other non-asymptotic bounds that consider different convergence assumptions on the Markov chain and the underlying error criterion. For details we refer to (Fan et al., 2021; Łatuszyński et al., 2013; Łatuszyński & Niemiro, 2011; Paulin, 2015; Rudolf & Schweizer, 2015) and the references therein.

In the following we briefly discuss two popular methods for the construction of a transition kernel K that is reversible with respect to μ.

B.2 MH algorithms

Probably the most popular method for constructing a transition kernel that is reversible with respect to μ is given by the MH algorithm. Given a Markov kernel Q on $\mathbb{X}$ which serves as a proposal mechanism, that is, a "proposal kernel," and given a function $\alpha : \mathbb{X} \times \mathbb{X} \to [0,1]$ which provides acceptance probabilities and depends on μ and Q, a transition from a state x to a state y in the MH algorithm proceeds as follows.

Transition Mechanism 2. Let Q be a proposal kernel and $\alpha : \mathbb{X} \times \mathbb{X} \to [0,1]$ be an acceptance probability function. Given the current state $x \in \mathbb{X}$, one obtains the next state $y \in \mathbb{X}$ as follows:

(1) Draw $X' \sim Q(x, \cdot)$ and $U \sim \mathrm{U}[(0,1)]$ independently and denote the realisations by x' and u respectively;
(2) If $u < \alpha(x, x')$, return $y := x'$, otherwise return $y := x$.

The algorithm above can be rewritten as a transition kernel:

$$K(x, \mathrm{d}y) = \alpha(x,y)Q(x, \mathrm{d}y) + r(x)\delta_x(\mathrm{d}y), \qquad r(x) := 1 - \int_{\mathbb{X}} \alpha(x,y)\, Q(x, \mathrm{d}y), \tag{B6}$$

where δ_x denotes the Dirac measure on $\mathbb{X}$ at $x \in \mathbb{X}$ and the function r is called the "rejection probability." It remains to specify α. Let $\sigma^+, \sigma^- \in \mathcal{P}(\mathbb{X} \times \mathbb{X})$ with

$$\begin{aligned} \sigma^+(\mathrm{d}x, \mathrm{d}x') &:= Q(x, \mathrm{d}x')\mu(\mathrm{d}x), \\ \sigma^-(\mathrm{d}x, \mathrm{d}x') &:= \sigma^+(\mathrm{d}x', \mathrm{d}x) = Q(x', \mathrm{d}x)\mu(\mathrm{d}x'), \end{aligned} \tag{B7}$$

and set

$$\alpha(x, x') := \min\left\{1, \frac{\mathrm{d}\sigma^-}{\mathrm{d}\sigma^+}(x, x')\right\}.$$

Then the transition kernel K defined in (B6) is reversible with respect to μ (Tierney, 1998). Of course, the Radon–Nikodym derivative $\frac{\mathrm{d}\sigma^-}{\mathrm{d}\sigma^+}(x, x')$ does not necessarily exist. In the finite-dimensional Euclidean setting where $\mathbb{X} = \mathbb{R}^d$, the derivative $\frac{\mathrm{d}\sigma^-}{\mathrm{d}\sigma^+}$ is often just the ratio of Lebesgue densities. However, in infinite-dimensional spaces the absolute continuity $\sigma^- \ll \sigma^+$ requires the choice of a suitable proposal kernel Q.

B.3 Slice sampling

Suppose that a σ-finite reference measure μ_0 on $\mathbb{X}$ is given that satisfies $\mu \ll \mu_0$. Additionally, we assume that the probability density function $\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}$ satisfies

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(x) \propto \exp(-\Phi(x)), \quad \mu_0\text{-a.e. } x \in \mathbb{X},$$

for a measurable function $\Phi : \mathbb{X} \to \mathbb{R}$ such that $\exp(-\Phi)$ is integrable with respect to μ_0. For some $s > 0$, define

$$\mathbb{X}_s := \{x \in \mathbb{X} | \exp(-\Phi(x)) \geqslant s\}, \tag{B8}$$

to be the super-level set of $\exp(-\Phi)$ to level s. Let $\|\exp(-\Phi)\|_\infty := \mu_0\text{-esssup}_{x\in\mathbb{X}} |\exp(-\Phi(x))|$ and note that $\mu_0(\mathbb{X}_s) \in (0, \infty)$ for $s \in (0, \|\exp(-\Phi)\|_\infty)$. Here,

$$\mu_0\text{-esssup}_{x\in\mathbb{X}} f(x)$$

denotes the essential supremum with respect to μ_0 of $f : \mathbb{X} \to \mathbb{R}$. Define the probability measure $\mu_{0,s} \in \mathcal{P}(\mathbb{X})$ by

$$\mu_{0,s}(A) := \frac{\mu_0(A \cap \mathbb{X}_s)}{\mu_0(\mathbb{X}_s)}, \quad A \in \mathcal{B}(\mathbb{X}),$$

that is, $\mu_{0,s}$ is the normalised restriction of μ_0 to $\mathbb{X}_s$. In the idealised slice sampling algorithm, a transition from a state x to a state y proceeds as follows.

Transition Mechanism 3 (Idealised slice sampling). Given the current state $x \in \mathbb{X}$ one obtains the next state $y \in \mathbb{X}$ as follows:

(1) Draw $S \sim \mathrm{U}[(0, \exp(-\Phi(x)))]$ and let s be the realization.
(2) Draw $Y \sim \mu_{0,s}$ and let the state y be the realization of Y.

The corresponding transition kernel takes the form

$$K(x, \mathrm{d}y) = \frac{1}{\exp(-\Phi(x))} \int_0^{\exp(-\Phi(x))} \mu_{0,s}(\mathrm{d}y)\, \mathrm{d}s,$$

and it can be readily shown that K is reversible with respect to μ.

Remark 6. In the case that $\mathbb{X} = \mathbb{R}^d$ and μ_0 is the Lebesgue measure, $\mu_{0,s}$ coincides with the uniform distribution on $\mathbb{X}_s$. In that setting the corresponding method is known as simple (uniform) slice sampling and recent results (Natarovskii et al., 2021b, theorem 3.10) concerning the spectral gap indicate under which conditions robust convergence behavior with respect to the dimension is present.

However, the main issue with the idealised slice sampling algorithm is that the second step in Transition Mechanism 3 may be difficult to implement, because the implicit assumption of being able to draw samples from $\mu_{0,s}$ for an arbitrary level s may be very restrictive. Whenever one is not able to sample from a distribution exactly, one can try to use a suitable Markov chain step.

Namely, one substitutes the second step of Transition Mechanism 3 by performing a transition, depending on x and s, which (at least) has stationary distribution $\mu_{0,s}$. Such approaches are known as *hybrid slice sampling* strategies; see for example, Łatuszyński and Rudolf (2014) for some theory and comparison results. There are several different methods in the literature that are feasible in finite-dimensional settings, and can—from an algorithmic perspective—be lifted to infinite-dimensional scenarios; see for example, (Li & Walker, 2023; Murray et al., 2010; Neal, 2003).

APPENDIX C. TWO RANDOM WALK-LIKE MCMC ALGORITHMS ON MANIFOLDS

We describe now two MH algorithms from the literature. These algorithms use random walk-like proposals that are defined on general, finite-dimensional manifolds $\mathbb{M}$. We compare the algorithms described in Section 3.3 with these algorithms on numerical examples in Section 4. In the following paragraphs we first describe the corresponding MH algorithm for a general manifold $\mathbb{M}$ and then provide the particular algorithm for the case of sampling on a sphere $\mathbb{M} = \mathbb{S}^{d-1}$. Throughout, we assume that $\mathbb{M}$ admits a Hausdorff measure $\mathcal{H}_{\mathbb{M}}$ and that the target probability measure μ on $\mathbb{M}$ is given by an unnormalized density ρ with respect to $\mathcal{H}_{\mathbb{M}}$, that is,

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mathcal{H}_{\mathbb{M}}}(\bar{x}) \propto \rho(\bar{x}), \qquad \bar{x} \in \mathbb{M}. \tag{C1}$$

When $\mathbb{M} = \mathbb{S}^{d-1}$ and μ is a posterior measure with respect to an ACG prior $\mu_0 = \mathrm{ACG}(C)$ as in (4),

$$\rho(\bar{x}) = \frac{\exp(-\overline{\Phi}(\bar{x}))}{\|\bar{x}\|_C^d}.$$

C.1 Geodesic random walk-MH algorithm

Assume that $\mathcal{H}_{\mathbb{M}}(\mathbb{M}) < \infty$. Then the uniform measure $\mathrm{U}(\mathbb{M})$ on $\mathbb{M}$ is defined via $\mathrm{U}(\mathbb{M})(A) := \mathcal{H}_{\mathbb{M}}(A)/\mathcal{H}_{\mathbb{M}}(\mathbb{M})$ for $A \in \mathcal{B}(\mathbb{M})$. The geodesic random walk as described in Mangoubi and Smith (2018) yields a Markov chain $(\overline{X}_k)_{k\in\mathbb{N}}$ on $\mathbb{M}$ with $\mathrm{U}(\mathbb{M})$ as its limit distribution. For any $\bar{x} \in \mathbb{M}$, denote the tangent space at $\bar{x}$ by $\mathcal{T}_{\bar{x}}\mathbb{M}$, and for any vector $v \in \mathcal{T}_{\bar{x}}\mathbb{M}$, denote by $\gamma_{\bar{x},v}$ the unique geodesic $\gamma_{\bar{x},v} : [0,\infty) \to \mathbb{M}$ on $\mathbb{M}$ that satisfies $\gamma_{\bar{x},v}(0) = \bar{x}$ and $\gamma'_{\bar{x},v}(0) = v$, where γ' denotes the first derivative.

Next, we present how a transition from a state $\bar{x}$ to a state $\bar{y}$ proceeds in the geodesic random walk algorithm.

> **Transition Mechanism 4 (Geodesic random walk).** Given the current state $\bar{x} \in \mathbb{M}$ one obtains the next state $\bar{y} \in \mathbb{M}$ for fixed $t > 0$ as follows:
>
> (1) Draw from the uniform distribution on the unit sphere in $\mathcal{T}_{\bar{x}}\mathbb{M}$ and call the result v;
> (2) Set $\bar{y} = \gamma_{\bar{x},v}(t)$.

Hence, in order to sample approximately from μ as in (C1), we can employ the geodesic random walk kernel as a proposal kernel in an MH algorithm. The resulting acceptance probability α involves ratios of the unnormalized density ρ in (C1). In Goyal and Shetty (2019, theorem 27)—and, in a more general setting, Habeck et al. (2023, section 2.5)—it is shown that the corresponding geodesic random walk transition kernel is reversible with respect to $\mathrm{U}(\mathbb{M})$. Therefore,

Algorithm 6. Geodesic random walk-MH on $\mathbb{S}^{d-1}$ (Mangoubi & Smith, 2018)

Given: time $t \in (0, \frac{\pi}{2}]$ and initial state $\bar{x}_0 \in \mathbb{S}^{d-1}$
for $k \in \mathbb{N}_0$ **do**
  Compute ONB matrix $U_{\bar{x}_k} \in \mathbb{R}^{d\times(d-1)}$ of $\bar{x}_k^{\perp}$ according to Remark 7
  Draw a sample w_k from $\mathrm{U}(\mathbb{S}^{d-2})$ and set $v_k := U_{\bar{x}_k} w_k$
  Set $\bar{y}_{k+1} := \cos(t)\bar{x}_k + \sin(t)v_k$
  Compute $a := \min\{1, \rho(\bar{y}_{k+1})/\rho(\bar{x}_k)\}$
  Draw a sample u from $\mathrm{U}([0,1])$
  if $u \leqslant a$ **then**
    Set $\bar{x}_{k+1} = \bar{y}_{k+1}$
  else
    Set $\bar{x}_{k+1} = \bar{x}_k$
  end if
end for

the resulting algorithm for realizing the Metropolized geodesic random walk on $\mathbb{M} = \mathbb{S}^{d-1}$ is given as in Algorithm 6.

Remark 7. In order to sample from the uniform distribution on the unit sphere in $\mathcal{T}_{\bar{x}}\mathbb{M}$ we can use the tools outlined by Zappa et al. (2018) for the case of constrained m-dimensional manifolds embedded into $\mathbb{R}^d$, $d > m$, via

$$\mathbb{M} = \left\{ x \in \mathbb{R}^d \,\middle|\, q_i(x) = 0 \quad \forall i = 1,\dots,L \right\}, \qquad L \geqslant d - m, \tag{C2}$$

where $q_i : \mathbb{R} \to \mathbb{R}$ are smooth functions. In particular, an orthonormal basis (ONB) of $\mathcal{T}_{\bar{x}}\mathbb{M}$ can be obtained by a QR decomposition of the Jacobian of $Q(x) := (q_1(x),\dots,q_L(x))$ evaluated at $x = \bar{x}$. Let us denote such a basis by $u_1,\dots,u_m \in \mathbb{R}^d$ and store the vectors in the columns of $U_{\bar{x}} = [u_1|\dots|u_m] \in \mathbb{R}^{d\times m}$. Drawing a sample w from $\mathrm{U}(\mathbb{S}^{m-1})$ and setting $v = U_{\bar{x}}w$ yields a sample according to the uniform distribution on the sphere in $\mathcal{T}_{\bar{x}}\mathbb{M}$. If $\mathbb{M} = \mathbb{S}^{d-1}$, then there is only one constraint function, that is, $q_1(x) = \|x\|^2 - 1$. Hence, the tangent space $\mathcal{T}_{\bar{x}}\mathbb{M}$ coincides with the orthogonal complement $\bar{x}^{\perp}$ of $\mathrm{span}\{\bar{x}\}$.

Remark 8 (Ergodicity). In Mangoubi and Smith (2018, theorem 7.1) it is shown that the geodesic random walk, that is, the proposal kernel of Algorithm 6, possesses a mixing time with respect to a Wasserstein distance that depends only on the positive curvature of the manifold $\mathbb{M}$, given a suitably small integration time t. Thus, in case of the sphere $\mathbb{M} = \mathbb{S}^{d-1}$, the mixing time is independent of the dimension d for any $t \in (0, \frac{\pi}{2}]$. However, this statement solely holds for the proposal chain. As we see in Section 4, the MH algorithm based on the geodesic random walk proposal shows deteriorating efficiency as $d \to \infty$. Besides that, also the uniform ergodicity of Algorithm 6, or the geodesic random walk itself, is not obvious and left open for future research.

C.2 MH algorithm by Zappa et al. (2018).

The MH algorithm presented in Zappa et al. (2018) shows some similarities with our method, in the sense that it first proposes a new state in the ambient Euclidean space which is then projected to the manifold $\mathbb{M}$. The manifold $\mathbb{M}$ is assumed to be of the form (C2). Given a current state $\bar{x} \in \mathbb{M}$, we first draw a tangent vector $v \in \mathcal{T}_{\bar{x}}\mathbb{M}$, but this time with respect to an isotropic multivariate Gaussian measure $\mathrm{N}(0, s^2 I_{d-m})$. Here, we can employ again the technique explained in Remark 7, by using an ONB matrix $U_{\bar{x}} \in \mathbb{R}^{d\times(d-m)}$ of $\mathcal{T}_{\bar{x}}\mathbb{M}$, drawing w from $\mathrm{N}(0, s^2 I_{d-m})$, and defining the resulting sample $v = U_{\bar{x}} w$. We then consider $x := \bar{x} + v \in \mathbb{R}^d$, and project x to some $\bar{y} \in \mathbb{M}$ by

$$\bar{y} := x + w \quad \text{with suitable} \quad w \in (\mathcal{T}_{\bar{x}}\mathbb{M})^{\perp}.$$

In general, computing w and $\bar{y}$ requires solving a nonlinear system; see Zappa et al. (2018, equation (2.5)). For $\mathbb{M} = \mathbb{S}^{d-1}$ the situation is rather easy: Since $(\mathcal{T}_{\bar{x}}\mathbb{M})^{\perp} = \operatorname{span}\{\bar{x}\}$, $w = a\bar{x}$ for some $a \in \mathbb{R}$ satisfying $\|\bar{y}\| = \|(1+a)\bar{x} + v\| = 1$. Since $\bar{x} \perp v$, if $\|v\|^2 \leqslant 1$, then

$$\bar{y} = \sqrt{1 - \|v\|^2}\,\bar{x} + v.$$

If $\|v\|^2 > 1$, then x cannot be projected back to the sphere along $\operatorname{span}\{\bar{x}\}$. In this case, x is rejected and the Markov chain remains at its current state, that is, in the kth iteration it is realized as $\bar{x}_{k+1} = \bar{x}_k$. In case of a successful proposal $\bar{y}$ we still require a Metropolis step, where the correct acceptance probability for $\bar{y}$ also requires the ingredients of the reverse move from $\bar{y}$ to $\bar{x}$. That is, we require $\tilde{v} \in \mathcal{T}_{\bar{y}}\mathbb{M}$ such that

$$\bar{x} = \bar{y} + \tilde{v} + \tilde{w}, \qquad \tilde{w} \in (\mathcal{T}_{\bar{y}}\mathbb{M})^{\perp}.$$

The acceptance probability in the MH algorithm targeting μ as in (C1) is then given by

$$\alpha(\bar{x}, \bar{y}) = \min\left\{1, \frac{\rho(\bar{y})\; p(\bar{y}, \tilde{v})}{\rho(\bar{x})\; p(\bar{x}, v)}\right\},$$

where $p(\bar{y}, \tilde{v}) \propto \exp\left(-\left\|U_{\bar{y}}^{\top}\tilde{v}\right\|^2 / 2s^2\right)$ denotes the proposal density for the tangential moves. Since $U_{\bar{y}}$ is orthonormal, $\left\|U_{\bar{y}}^{\top}\tilde{v}\right\| = \|\tilde{v}\|$. Moreover, for $\mathbb{M} = \mathbb{S}^{d-1}$, the vector $\tilde{v} \in \mathcal{T}_{\bar{y}}\mathbb{M}$ for going from $\bar{y} = \sqrt{1 - \|v\|^2}\,\bar{x} + v$ back to $\bar{x}$ can be computed easily by projecting $\bar{x} - \bar{y}$ onto $\mathcal{T}_{\bar{y}}\mathbb{M} = (\bar{y})^{\perp}$, which yields

$$\tilde{v} = \bar{x} - \bar{y} - \langle \bar{x} - \bar{y}, \bar{y}\rangle \bar{y} = \|v\|^2 \bar{x} - \sqrt{1 - \|v\|^2}\, v.$$

In particular, since $\bar{x} \perp v$, we obtain that $\|\tilde{v}\| = \|v\|$. Hence, for $\mathbb{M} = \mathbb{S}^{d-1}$ the acceptance probability is just $\alpha(\bar{x}, \bar{y}) = \min\{1, \rho(\bar{y})/\rho(\bar{x})\}$. We summarize the resulting MH algorithm in Algorithm 7.

Remark 9 (Ergodicity). As stated in Zappa et al. (2018, section 2.1), their proposed MH algorithm yields uniform ergodicity for compact $\mathbb{M}$ and continuous ρ. This follows by standard arguments for MH algorithms with continuous proposal densities bounded away from zero on compact state spaces, see, for example, Douc et al. (2018,

Algorithm 7. MH algorithm on $\mathbb{S}^{d-1}$ (Zappa et al., 2018)

Given: step size $s > 0$ and initial state $\bar{x}_0 \in \mathbb{S}^{d-1}$
for $k \in \mathbb{N}_0$ **do**
  Compute ONB matrix $U_{\bar{x}_k} \in \mathbb{R}^{d\times(d-1)}$ of $\bar{x}_k^{\perp}$ according to Remark 7
  Draw a sample w_k of $\mathrm{N}(0, s^2 I_{d-1})$ and set $v_k := U_{\bar{x}_k} w_k$
  if $\|v\| > 1$ **then**
    Set $\bar{x}_{k+1} = \bar{x}_k$
  else
    Set $\bar{y}_{k+1} := \sqrt{1 - \|v\|^2}\bar{x}_k + v_k$
    Compute $a := \min\{1, \rho(\bar{y}_{k+1})/\rho(\bar{x}_k)\}$
    Draw a sample u of $\mathrm{U}[0, 1]$
    if $u \leqslant a$ **then**
      Set $\bar{x}_{k+1} = \bar{y}_{k+1}$
    else
      Set $\bar{x}_{k+1} = \bar{x}_k$
    end if
  end if
end for

example 15.3.2). In particular, Algorithm 7 yields uniformly ergodic Markov chains under the conditions of Theorem 1.

APPENDIX D. COUNTEREXAMPLE FOR THE MARKOVIANITY OF $(\Pi^{\mathbb{S}}(X_n))_{n\in\mathbb{N}}$

Let $\mathbb{H} = \mathbb{R}^d$, $d > 1$, and $\Pi^{\mathbb{S}}$ be the radial projection onto the unit sphere in $\mathbb{H}$ as in (3), where we have fixed $\bar{z} = e_d = (0,\dots,0,1)^\top$. Moreover, let $(X_n)_{n\in\mathbb{N}}$ denote a Markov chain on $\mathbb{H}$ with transition kernel K.

If $(\Pi^{\mathbb{S}}(X_n))_{n\in\mathbb{N}}$ were again a Markov chain, then we would have for the "upper half" of the sphere $H_d := \{x = (x_1,\dots,x_d)^\top \in \mathbb{S} | x_d \geqslant 0\}$ that

$$\mathbb{P}(\Pi^{\mathbb{S}}(X_2) \in H_d \mid \Pi^{\mathbb{S}}(X_1) = e_d) = \mathbb{P}(\Pi^{\mathbb{S}}(X_2) \in H_d \mid \Pi^{\mathbb{S}}(X_1) = e_d, \Pi^{\mathbb{S}}(X_0) = e_d),$$

or, $\mathbb{P}(\Pi^{\mathbb{S}}(X_{n+2}) \in H_d \mid \Pi^{\mathbb{S}}(X_{n+1}) = e_d) = \mathbb{P}(\Pi^{\mathbb{S}}(X_{n+2}) \in H_d \mid \Pi^{\mathbb{S}}(X_{n+1}) = e_d, \Pi^{\mathbb{S}}(X_n) = e_d)$ for any $n \in \mathbb{N}$. By definition of $\Pi^{\mathbb{S}}$, this is equivalent to

$$\frac{\mathbb{P}(X_{2,d} > 0, X_{1,d} > 0)}{\mathbb{P}(X_{1,d} > 0)} = \frac{\mathbb{P}(X_{2,d} > 0, X_{1,d} > 0, X_{0,d} > 0)}{\mathbb{P}(X_{1,d} > 0, X_{0,d} > 0)},$$

where $X_n = (X_{n,1},\dots,X_{n,d})^\top$ denotes the state vector of the Markov chain. Consider now a Markov chain $(X_n)_{n\in\mathbb{N}}$ with initial distribution $X_0 \sim \mathrm{N}(0, I)$ and Gaussian random walk transition kernel $K(x,\cdot) = \mathrm{N}(x, s^2 I)$ for a fixed step size $s > 0$. Then each component of the states is a Markov chain

independent of the other components. Hence,

$$\frac{\mathbb{P}(X_{2,d} > 0, X_{1,d} > 0)}{\mathbb{P}(X_{1,d} > 0)} = \frac{\mathbb{P}(W_0 + sW_1 + sW_2 > 0, W_0 + sW_1 > 0)}{\mathbb{P}(W_0 + sW_1 > 0)},$$

for $W_0, W_1, W_2 \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$ and, analogously,

$$\frac{\mathbb{P}(X_{2,d} > 0, X_{1,d}, X_{0,d} > 0)}{\mathbb{P}(X_{1,d} > 0, X_{0,d} > 0)} = \frac{\mathbb{P}(W_0 + sW_1 + sW_2 > 0, W_0 + sW_1 > 0, W_0 > 0)}{\mathbb{P}(W_0 + sW_1 > 0, W_0 > 0)}.$$

The expressions for these probabilities are difficult to evaluate exactly. Therefore, we perform a Monte Carlo integration using 10^8 samples and obtain the following estimates for $s = 1$:

$$\frac{\mathbb{P}(W_0 + sW_1 + sW_2 > 0, W_0 + sW_1 > 0)}{\mathbb{P}(W_0 + sW_1 > 0)} \approx 0.8041,$$

$$\frac{\mathbb{P}(W_0 + sW_1 + sW_2 > 0, W_0 + sW_1 > 0, W_0 > 0)}{\mathbb{P}(W_0 + sW_1 > 0, W_0 > 0)} \approx 0.8333.$$

We now show that the above observation is not restricted to the random walk transition kernel. Consider a stationary Markov chain generated by the pCN proposal kernel $Q(x, \cdot) = \mathrm{N}(\sqrt{1-s^2}x, s^2 I)$ and initial distribution $X_0 \sim \mathrm{N}(0, I)$. We again obtain independent scalar Markov chains in each component:

$$\frac{\mathbb{P}(X_{2,d} > 0, X_{1,d} > 0)}{\mathbb{P}(X_{1,d} > 0)} = \frac{\mathbb{P}((1-s^2)W_0 + s\sqrt{1-s^2}W_1 + sW_2 > 0, \sqrt{1-s^2}W_0 + sW_1 > 0)}{\mathbb{P}(\sqrt{1-s^2}W_0 + sW_1 > 0)},$$

as well as

$$\frac{\mathbb{P}(X_{2,d} > 0, X_{1,d}, X_{0,d} > 0)}{\mathbb{P}(X_{1,d} > 0, X_{0,d} > 0)} = \frac{\mathbb{P}((1-s^2)W_0 + s\sqrt{1-s^2}W_1 + sW_2 > 0, \sqrt{1-s^2}W_0 + sW_1 > 0, W_0 > 0)}{\mathbb{P}(\sqrt{1-s^2}W_0 + sW_1 > 0, W_0 > 0)},$$

where $W_0, W_1, W_2 \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$. Monte Carlo integration with the stationary pCN-generated Markov chain, $s = 0.5$, and 10^8 samples yields

$$\frac{\mathbb{P}(X_{2,d} > 0, X_{1,d} > 0)}{\mathbb{P}(X_{1,d} > 0)} \approx 0.8333, \qquad \frac{\mathbb{P}(X_{2,d} > 0, X_{1,d}, X_{0,d} > 0)}{\mathbb{P}(X_{1,d} > 0, X_{0,d} > 0)} \approx 0.8620.$$

Thus, these Markov chains serve as counterexamples to the claim that $(\Pi^{\mathbb{S}}(X_n))_{n\in\mathbb{N}}$ is again a Markov chain.