STEREOGRAPHIC MARKOV CHAIN MONTE CARLO

BY JUN YANG[1,a], KRZYSZTOF ŁATUSZYŃSKI[2,b] AND GARETH O. ROBERTS[2,c]

[1]*Department of Mathematical Sciences, University of Copenhagen,* [a]*jy@math.ku.dk*

[2]*Department of Statistics, University of Warwick,* [b]*k.g.latuszynski@warwick.ac.uk,* [c]*gareth.o.roberts@warwick.ac.uk*

High-dimensional distributions, especially those with heavy tails, are notoriously difficult for off-the-shelf MCMC samplers: the combination of unbounded state spaces, diminishing gradient information, and local moves results in empirically observed "stickiness" and poor theoretical mixing properties—lack of geometric ergodicity. In this paper, we introduce a new class of MCMC samplers that map the original high-dimensional problem in Euclidean space onto a sphere and remedy these notorious mixing problems. In particular, we develop random-walk Metropolis type algorithms as well as versions of the Bouncy Particle Sampler that are uniformly ergodic for a large class of light and heavy-tailed distributions and also empirically exhibit rapid convergence in high dimensions. In the best scenario, the proposed samplers can enjoy the "blessings of dimensionality" that the convergence is faster in higher dimensions.

1. Introduction. Bayesian analysis relies heavily on Markov chain Monte Carlo (MCMC) methods to explore complex posterior distributions. In most typical settings, such distributions have support contained within a subset S of $\mathbb{R}^d$ for some $d > 0$, and then it is natural to construct appropriate MCMC algorithms directly on S. In practice, this is how the vast majority of algorithms are constructed, although there are intrinsic problems with this approach. For instance, it is now well established (Mengersen and Tweedie (1996)) that the popular vanilla MCMC workhorse, the random-walk Metropolis (RWM) algorithm fails to be uniformly ergodic for any target density π when S is unbounded. All existing generic MCMC methods are built upon local proposal mechanisms and are similarly afflicted. Lack of uniform ergodicity results in sensitivity of the algorithm's convergence to its starting value, and potentially long burn-in periods.

Moreover, these convergence problems are exacerbated for target distributions with heavy (heavier than exponential) tails. For instance, in such cases RWM and the Metropolis-adjusted Langevin algorithm (MALA) fail to be even geometrically ergodic (i.e., they converge at a rate slower than any geometric rate) (Roberts and Tweedie (1996a), Jarner and Hansen (2000), Roberts and Tweedie (1996b)). In practice, this manifests itself on the algorithm trajectories by the presence of infrequent excursions of heavy-tailed duration into the target distribution tails. This can lead to further theoretical and practical problems, for example, the absence of a central limit theorem (CLT) for all L^2 functions, for example, Jarner and Roberts (2007), and instability of Monte Carlo estimators with large and difficult-to-quantify mean square errors which are highly sensitive to initial values.

Important modern innovations in MCMC algorithms have come from Piecewise Deterministic Markov Processes (PDMPs) (Bernard, Krauth and Wilson (2009), Bouchard-Côté, Vollmer and Doucet (2018), Bierkens, Fearnhead and Roberts (2019)) which offer for the first time generic recipes for the construction of nonreversible MCMC and often yield substantial gains in algorithm efficiency as a result. Although not completely necessary, it turns

out to be convenient and natural to construct PDMPs as continuous time algorithms. While the practical use of PDMPs for posterior exploration is still in its infancy, these methods offer substantial promise. However, PDMPs are still localised algorithms and as such also suffer from the lack of uniform and/or geometric ergodicity on unfounded state spaces and/or heavy-tailed targets (Vasdekis and Roberts (2022, 2023), Andrieu, Dobson and Wang (2021), Andrieu et al. (2021)).

In contrast, when the target density tails are exponential or lighter, RWM, MALA, and related algorithms are generally geometrically ergodic under weak regularity conditions (Roberts and Tweedie (1996a), Jarner and Hansen (2000)). Some transformation strategies for achieving this are described in Sherlock, Fearnhead and Roberts (2010), Johnson and Geyer (2012), with closely related strategies being proposed in Kamatani (2018), although these approaches are unlikely to regain uniform ergodicity. When S is bounded, then MCMC algorithms are generally easily shown to be uniformly ergodic (see, e.g., Mengersen and Tweedie (1996) for RWM). This naturally suggests that transformations designed to compactify S might facilitate the construction of more robust families of MCMC algorithms. However, finding generic solutions to the construction of such transformations which lead to well-behaved densities on the transformed space is challenging.

It is, however, far easier to construct general transformations to transform $\mathbb{R}^d$ to a *pre*-compact space. The most celebrated of such transformations is *stereographic projection* which was known to the ancient Egyptians; see Coxeter (1961) for a modern description. It maps $\mathbb{R}^d \to \mathbb{S}^d \setminus N$ (where N denotes the *North Pole*), that is, the d-sphere excluding its North Pole. This paper will explore the use of stereographically projected algorithms. One iteration of such an algorithm takes the current state $\mathbf{x} \in \mathbb{R}^d$, transforms to $\mathbb{S}^d \setminus N$ through inverse stereographic projection, carrying out an appropriately constructed MCMC step on $\mathbb{S}^d \setminus N$ before returning to $\mathbb{R}^d$ by stereographic projection. The contributions of our work are as follows.

1. We present the Stereographic Projection Sampler (SPS), an efficient and practical implementation of the above programme for RWM. It employs a simple reprojection step to ensure the Markov chain remains on $\mathbb{S}^d \setminus N$. We provide a dimension and scale dependent recipe for choosing the radius of S. See Section 2.1.

2. We prove that for continuous positive densities, π on $\mathbb{R}^d$ with tails no heavier than those of a d-dimensional multivariate Student's t distribution with d degrees of freedom, SPS is uniformly ergodic. The tail conditions on π are in fact necessary for geometric ergodicity. See Theorem 2.1 and Remark 2.1.

3. We give a high-dimensional analysis of SPS for the stylised family of product i.i.d. form targets, by maximizing the expected squared jumping distance (ESJD) as well as by showing for a variant of SPS that each component converges (as $d \to \infty$) to a Langevin diffusion. This affords a direct and uniformly favorable comparison with the Euclidean RWM algorithm. See Section 5.

4. We also introduce the Stereographic Bouncy Particle Sampler (SBPS) which replaces the random walk move in SPS with a PDMP algorithm that follows great circle trajectories interspersed with abrupt direction changes. See Section 2.2.

5. We prove the uniform ergodicity of SBPS under weaker conditions than those for SPS, specifically requiring tails to be lighter than those of a d-dimensional multivariate Student's t distribution with $d - 1/2$ degrees of freedom. See Theorem 2.2 and Corollary 2.1.

6. For isotropic targets (i.e., spherically symmetric targets), both light- and heavy-tailed, we (informally) demonstrate that the proposed SPS and SBPS can enjoy a *blessings of dimensionality* effect which implies that SPS and SBPS converge arbitrarily faster in higher dimensions than their Euclidean analogues. See Section 3 and Section 6.

7. We introduce generalisations of stereographic projection which are suitable for elliptical targets. See Section 4. The framework we established in this paper opens opportunities for developing other mappings from $\mathbb{R}^d$ to $\mathbb{S}^d$ for other classes of targets. See Section 7.

In very recent work, Lie et al. (2023) presents a related methodology for exploring distributions defined directly on a manifold, providing supporting theory showing that under suitable conditions their method's spectral gap is dimension-independent. However, the focus of their work is very different. Their work uses a different reprojection scheme based on Cotter et al. (2013) and does not consider distributions defined directly on $\mathbb{R}^d$. Moreover, they consider densities that are absolutely continuous with respect to a Gaussian measure in the infinite-dimensional limit. This is arguably a very restrictive class. So their dimension-independent results are not really comparable with our results. An alternative reprojection scheme is introduced in Zappa, Holmes-Cerfon and Goodman (2018). Compared to that method, the reprojection scheme used in this paper has the advantage that reprojection from a fixed point on the tangent space is always possible and is a much more natural approach for the hypersphere. The Zappa, Holmes-Cerfon and Goodman (2018) method does have advantages for use in more general manifolds, but this is not of any use to us here.

There are strong theoretical reasons for wanting to construct the algorithm dynamics directly on a manifold of positive curvature such as a hypersphere (Mangoubi and Smith (2018), Ollivier (2009), Mijatović, Mramor and Uribe Bravo (2018)). However, quite naturally, the existing literature has concentrated primarily on the case of Brownian motion which clearly has the uniform invariant distribution on S, or on the geodesic walk designed specifically to target uniform distributions on manifolds. The uniform distribution on S maps via stereographic projection to the Student's t distribution with d degrees of freedom on $\mathbb{R}^d$. Therefore, we can immediately lift theory from existing theory to (informally) show that SPS on such target densities has a dimension-free convergence time. In our paper, we go further. Rapid convergence extends to a large family of spherically symmetric target densities (essentially excluding only very heavy-tailed distributions).

However the proposed algorithms are not only applicable for stylized classes of target distributions. To showcase the practical value of stereographic samplers in more realistic statistical contexts, we shall demonstrate the utility of our methodology on a Bayesian analysis of a Cauchy regression model. Full specification of the model, its parameter values and other details will be given in Section 6.1.

EXAMPLE 1.1. Consider $Y_i \sim \text{Cauchy}(\alpha + \beta^T X_i, \gamma)$, where $\{X_i, Y_i\}_{i=1}^n$ are the design matrices and responses, respectively, $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^{d-2}$, $\gamma \in \mathbb{R}^+$ are the parameters. Assume a flat prior for α and β, a Gamma(a, b) prior for γ (i.e., $\pi_0(\gamma) \propto \gamma^{a-1} \exp(-b\gamma)$), the posterior can be written by

$$(1) \qquad \pi(\alpha, \beta, \gamma) \propto \frac{\gamma^{a-1-n} \exp(-b\gamma)}{\prod_{i=1}^n \{1 + [\frac{Y_i - (\alpha + \beta^T X_i)}{\gamma}]^2\}}.$$

For sampling the posterior, we compare SPS and RWM in Figure 1, where we plot the trace-plots for α, β_1 (the first coordinate of β), and $\log(\gamma)$, as well as (α, β_1) for both algorithms. From the figure, one can see clearly that: (1) RWM which is not geometric ergodic completely failed; (2) the proposed SPS which is uniformly ergodic converged extremely fast.

Our paper is structured as follows. Section 2 introduces both SPS and SBPS algorithms and presents formal ergodicity and uniform ergodicity results for both methods. Sections 3 to 5 provide additional results for SPS and its generalisations. In Section 3, a detailed analysis of SPS for isotropic densities is provided, while in Section 4, a generalised version of the

Bayesian Cauchy Regression (d=11, n=15, accept rates SPS=0.57, RWM=0.69)
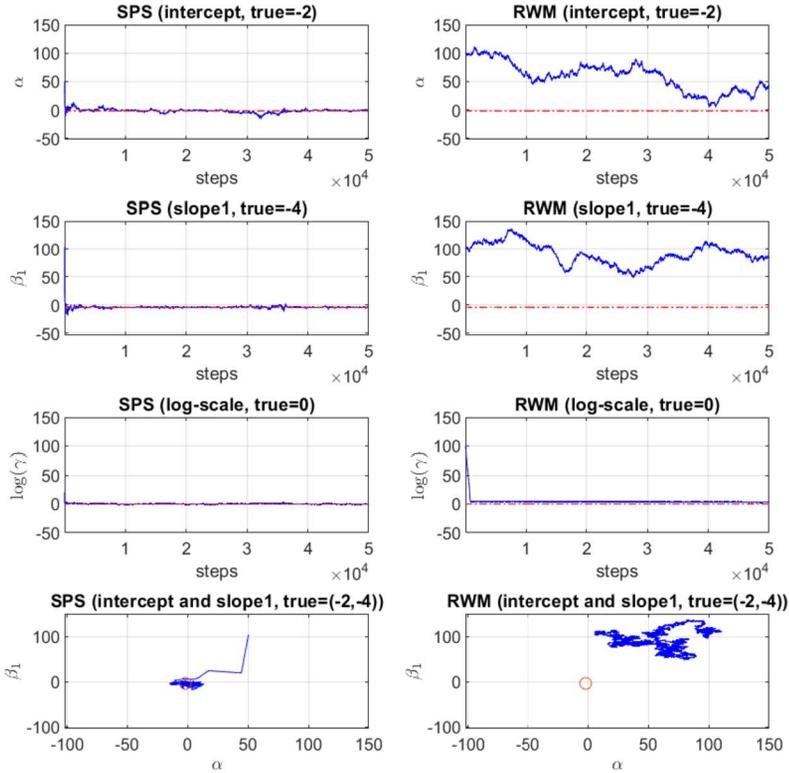
FIG. 1. *Bayesian Cauchy Regression. Red dotted lines and circles represent the parameters of the true sampling distribution.*

algorithm is introduced. Section 4 also gives robustness results to departures from isotropy. In Section 5, we provide a high-dimensional analysis of SPS for the stylised family of product i.i.d. target densities and give positive comparisons to the Euclidean RWM algorithm. Numerical studies on both SPS and SBPS are provided in Section 6 to illustrate our theory and we conclude with the discussions on SPS and SBPS in Section 7. The Supplementary Material (Yang, Łatuszyński and Roberts (2024)) contains all the technical proofs and some additional simulations.

2. Stereographic Markov chain Monte Carlo.

2.1. *Stereographic Projection Sampler* (*SPS*). We begin by giving a brief description of stereographic projection and then describe in detail two novel MCMC samplers that exploit the properties of stereographic projection.

Let $\mathbb{S}^d$ denote the *unit sphere* in $\mathbb{R}^{d+1}$ centered at the origin. A stereographic projection describes a bijection from $\mathbb{S}^d \setminus \{(0, \ldots 0, 1)\}$ to $\mathbb{R}^d$. Within this paper, we shall restrict attention to projections indexed by a single parameter $R \in \mathbb{R}^+$ and described by the mapping

$$x = \mathrm{SP}(z) := \left(R \frac{z_1}{1 - z_{d+1}}, \ldots, R \frac{z_d}{1 - z_{d+1}} \right)^T,$$

with Jacobian determinant at $x \in \mathbb{R}^d$ satisfying

$$(2) \qquad J_{\mathrm{SP}}(x) \propto \left(R^2 + \|x\|^2 \right)^d,$$
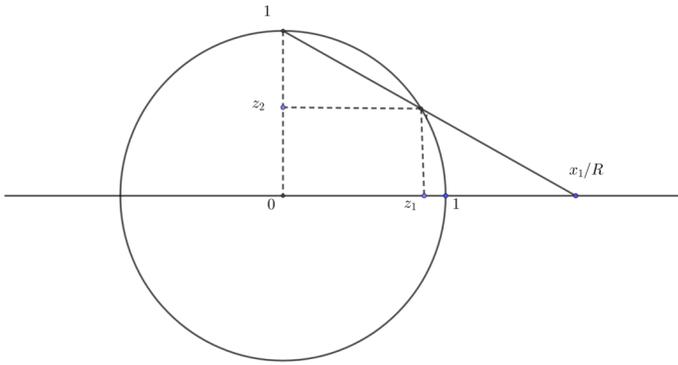
FIG. 2. *Illustration of Stereographic Projection* $\mathrm{SP} : \mathbb{S} \to \mathbb{R}$.

and inverse $\mathrm{SP}^{-1} : \mathbb{R}^d \to \mathbb{S}^d \setminus \{(0, \ldots 0, 1)\}$ given by

$$(3) \qquad z_i = \frac{2R x_i}{\|x\|^2 + R^2} \quad \forall 1 \le i \le d, \qquad z_{d+1} = \frac{\|x\|^2 - R^2}{\|x\|^2 + R^2}.$$

See Figure 2 for a geometric illustration of this stereographic projection (in the case $d = 1$ for simplicity) and the Supplementary Material (Yang, Łatuszyński and Roberts (2024), S8.13) for the proof of the Jacobian determinant equation (2).

Now suppose we wish to sample from a target density $\pi(x)$ where $x \in \mathbb{R}^d$. Our aim is to take advantage of our stereographic bijection to construct an MCMC sampler directly on $\mathbb{S}^d$, projecting the output back onto $\mathbb{R}^d$. We denote the transformed target as $\pi_S(z)$ then for $x = \mathrm{SP}(z)$ we have

$$(4) \qquad \pi_S(z) \propto \pi(x)(R^2 + \|x\|^2)^d.$$

First, we just consider a random-walk Metropolis algorithm with a step size h on the unit sphere. Note that we describe our algorithms on the unit sphere. The radius R will only be taken into account when projecting to $\mathbb{R}^d$. Figure 3 illustrates how the algorithm moves are constructed on the unit sphere, and pseudo-code for the algorithm is given in Algorithm 1. (Note that we could very easily have constructed more general Metropolis–Hastings algorithms.)

The symmetry of the reprojection mechanism in the proposal of the algorithm means that SPS is indeed a valid MCMC algorithm for π. More formally, we have the following result.

PROPOSITION 2.1. *If $\pi(x)$ is positive and continuous in $\mathbb{R}^d$, then SPS gives rise to an ergodic Markov chain on $\mathbb{R}^d$ with invariant distribution π.*
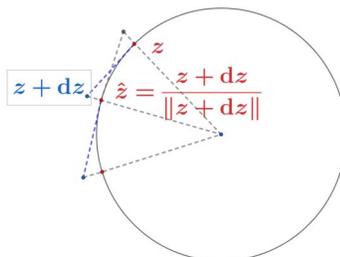
FIG. 3. *Illustration of SPS proposing $\hat{z} = \mathrm{SP}^{-1}(\hat{X})$ from $z = \mathrm{SP}^{-1}(x)$. By symmetricity, the proposal distribution is the same as proposing z from $\hat{z}$.*

Algorithm 1: Stereographic Projection Sampler (SPS)

- Let the current state be $X^d(t) = x$;
- Compute the proposal $\hat{X}$:
  - Let $z := \mathrm{SP}^{-1}(x)$;
  - Sample independently $\mathrm{d}\tilde{z} \sim \mathcal{N}(0, h^2 I_{d+1})$;
  - Let $\mathrm{d}z := \mathrm{d}\tilde{z} - \frac{(z^T \cdot \mathrm{d}\tilde{z})z}{\|z\|^2}$ and $\hat{z} := \frac{z+\mathrm{d}z}{\|z+\mathrm{d}z\|}$;
  - The proposal $\hat{X} := \mathrm{SP}(\hat{z})$.
- $X^d(t+1) = \hat{X}$ with probability $1 \wedge \frac{\pi(\hat{X})(R^2+\|\hat{X}\|^2)^d}{\pi(x)(R^2+\|x\|^2)^d}$; otherwise $X^d(t+1) = x$.

PROOF. See the Supplementary Material (Yang, Łatuszyński and Roberts (2024), S8.1). □

We shall address the chain's uniform ergodicity properties. Recall that a Markov chain X on state space E with transition kernel P is *uniformly ergodic* if $\forall \epsilon > 0$, there exists, $N \in \mathbb{N}$ such that $\|P^N(x, \cdot) - \pi\|_{\mathrm{TV}} \le \epsilon$, $\forall x \in E$, where π denotes the chain's (unique) invariant distribution and $\|\cdot\|_{\mathrm{TV}}$ represents total variation distance. Intuitively, for uniformly ergodic chains, we cannot have "arbitrarily bad" starting values.

It is long recognised that random-walk Metropolis (RWM) algorithms on bounded spaces are usually uniformly ergodic, while the same algorithms on unbounded spaces are never uniformly ergodic (Mengersen and Tweedie (1996)). Therefore, given the compactness of $\mathbb{S}^d$, it is reasonable to hope that SPS might be uniformly ergodic under mild regularity conditions on π. Since the actual state space of SPS is in fact $\mathbb{S}^d \setminus N$ which is not compact, this question is more complicated than in the Euclidean state space case as we need to consider the properties of the transformed density near N. However, our first main result confirms that we do get uniform ergodicity if and only if the transformed density on the sphere is bounded at N.

THEOREM 2.1. *If $\pi(x)$ is positive and continuous in $\mathbb{R}^d$, then SPS is uniformly ergodic if and only if*

$$(5) \qquad \sup_{x \in \mathbb{R}^d} \pi(x)\big(R^2 + \|x\|^2\big)^d < \infty.$$

PROOF. See the Supplementary Material (Yang, Łatuszyński and Roberts (2024), S8.2). □

EXAMPLE 2.1. If the target $\pi(x)$ where $x \in \mathbb{R}^d$ is multivariate Student's t distribution with degrees of freedom no smaller than d, then the SPS algorithm is *uniformly ergodic*.

REMARK 2.1. We make the following remarks:

1. The condition equation (5) is necessary: if $\sup_{x \in \mathbb{R}^d} \pi(x)(R^2 + \|x\|^2)^d = \infty$, then the chain is not even geometrically ergodic (Roberts and Tweedie (1996a), Proposition 5.1);

2. The traditional RWM algorithm is *not* uniformly ergodic if the support of π is $\mathbb{R}^d$ (Mengersen and Tweedie (1996), Theorem 3.1) and *not* geometrically ergodic for any heavy-tailed target distribution (Jarner and Hansen (2000), Corollary 3.4);

3. The condition that $\pi(x)$ is positive and continuous in $\mathbb{R}^d$ in both Proposition 2.1 and Theorem 2.1 can be relaxed. We used it here just for the simplicity of the proof.

Algorithm 2: Stereographic Bouncy Particle Sampler (SBPS)

- Initialize $z^{(0)} \in \mathbb{S}^d$ and $v^{(0)}$ such that $v^{(0)} \cdot z^{(0)} = 0$ and $\|v^{(0)}\| = 1$.
- Simulate BPS on unit sphere: for $i = 1, 2, \ldots$
  - Simulate bounce time τ_{bounce} of a Poisson process of intensity

$$\chi(t) = \lambda\big(\sin(t)v^{(i-1)} + \cos(t)z^{(i-1)}, \cos(t)v^{(i-1)} - \sin(t)z^{(i-1)}\big),$$

  where

$$\lambda(z, v) := \max\big\{0, \big[-v \cdot \nabla_z \log \pi_S(z)\big]\big\}.$$

  - Simulate refreshment time $\tau_{\text{refresh}} \sim \text{Exponential}(\lambda_{\text{refresh}})$.
  - Let $\tau_i = \min\{\tau_{\text{bounce}}, \tau_{\text{refresh}}\}$ and

$$z^{(i)} = \sin(\tau_i)v^{(i-1)} + \cos(\tau_i)z^{(i-1)}.$$

  - If $\tau_i = \tau_{\text{refresh}}$, sample new $v^{(i)}$ independently

$$v^{(i)} \sim \text{Uniform}\big\{v : z^{(i)} \cdot v = 0, \|v\| = 1\big\}$$

  - If $\tau_i = \tau_{\text{bounce}}$, compute

$$v^{(i)} = v_{\text{temp}} - 2\bigg[\frac{v_{\text{temp}} \cdot \tilde{\nabla}_z \log \pi_S(z^{(i)})}{\tilde{\nabla}_z \log \pi_S(z^{(i)}) \cdot \tilde{\nabla}_z \log \pi_S(z^{(i)})}\bigg]\tilde{\nabla}_z \log \pi_S(z^{(i)}),$$

  where

$$v_{\text{temp}} = \cos(\tau_i)v^{(i-1)} - \sin(\tau_i)z^{(i-1)}$$

$$\tilde{\nabla}_z \log \pi_S(z^{(i)}) = \nabla_z \log \pi_S(z^{(i)}) - \big[z^{(i)} \cdot \nabla_z \log \pi_S(z^{(i)})\big]z^{(i)}.$$

  - If $\sum_{j=1}^{i} \tau_j \geq T$ (where T is some constant time), exit.
- Return $x = \text{SP}(z)$ where z denotes BPS on unit sphere.

2.2. *Stereographic Bouncy Particle Sampler* (*SBPS*). Many recent innovations in MCMC algorithm construction have focused on nonreversible methods, most particularly those described by *piecewise deterministic Markov processes* (PDMPs) (see, e.g., Bouchard-Côté, Vollmer and Doucet (2018), Bierkens, Fearnhead and Roberts (2019) and Davis (1984) for theoretical background). PDMPs are continuous-time processes that have stochastic jumps at event times of a point process, but where the state evolves deterministically between the event times. In this subsection, we shall demonstrate that we can readily incorporate these methods within our projective framework. We shall concentrate on a version of the *Bouncy Particle Sampler* (BPS) (Bouchard-Côté, Vollmer and Doucet (2018)) as this adapts naturally to our context. The algorithm (Algorithm 2) is described as follows.

PDMPs utilise an auxiliary random variable v, which in the case of the Stereographic Bouncy Particle Sampler (SBPS) has stationary distribution uniformly distributed on $\mathbb{S}^d$ and independently of x. One feature of PDMP algorithms such as SBPS is the option to include *refresh* moves that contribute to the intensity of their constituting point process (according to some possibly x-dependent hazard rate) and which independently refresh v by sampling it from its invariant distribution (uniform on $\mathbb{S}^d$). In the description below, we restrict ourselves to the case where this refresh rate is constant.

Again, we demonstrate that SBPS is indeed a valid MCMC algorithm, at least under certain conditions on π and the refresh rate in the algorithm. Note that the conditions on π and

λ_{refresh} in Proposition 2.2 are not required for invariance, only for ensuring ϕ-irreducibility of the algorithm.

PROPOSITION 2.2. *Suppose that* $\lambda_{\text{refresh}} > 0$ *and* $\pi > 0$ *for all* $x \in \mathbb{R}^d$. *Then SBPS gives rise to an ergodic Markov chain on* $\mathbb{R}^d$ *with invariant distribution for* (x, v) *with joint density on* $\mathbb{R}^d \times \mathbb{S}^d$. *The marginal density on* $\mathbb{R}^d$ *is proportional to* $\pi(x)$.

PROOF. The proof of this result is routine (but tedious), following the lines of existing results for PDMPs in the literature. We give a sketch proof in Yang, Łatuszyński and Roberts (2024), S8.3. $\square$

We shall prove a uniform ergodicity result analogous to that of Theorem 2.1. For a Markov process X with state space E, transition semi-group P, and invariant distribution π, it is uniformly ergodic if $\forall \epsilon > 0$ there exists T such that $\|P^T(x, \cdot) - \pi\|_{\text{TV}} \le \epsilon$, $\forall x \in E$.

THEOREM 2.2. *If* $\pi(x)$ *is positive in* $\mathbb{R}^d$ *with continuous first derivative in all components, then SBPS is* uniformly ergodic *if*

$$\limsup_{\{x : \|x\| \to \infty\}} \sum_{i=1}^{d} \left(\frac{\partial \log \pi(x)}{\partial x_i} x_i \right) + 2d < \frac{1}{2}.$$

PROOF. See the Supplementary Material (Yang, Łatuszyński and Roberts (2024), S8.4). $\square$

COROLLARY 2.1. *If the target* $\pi(x)$ *where* $x \in \mathbb{R}^d$ *is multivariate Student's t distribution with degrees of freedom larger than* $d - \frac{1}{2}$, *then the SBPS algorithm is* uniformly ergodic.

PROOF. See the Supplementary Material (Yang, Łatuszyński and Roberts (2024), S8.5). $\square$

REMARK 2.2. We make the following remarks:

1. This is the first known PDMP algorithm that is uniformly ergodic for a large family of target distributions including heavy-tailed targets. For comparison, the Euclidean BPS is only known to be geometrically ergodic under certain restrictive conditions on the target (Deligiannidis, Bouchard-Côté and Doucet (2019), Durmus, Guillin and Monmarché (2020)) and is known not to be geometrically ergodic for any heavy-tailed target distribution (Vasdekis and Roberts (2023)).

2. We conjecture that the best possible condition for Theorem 2.2 is

$$\limsup_{\{x : \|x\| \to \infty\}} \sum_{i=1}^{d} \left(\frac{\partial \log \pi(x)}{\partial x_i} x_i \right) + 2d < 1.$$

We explain the reason for our conjecture in Yang, Łatuszyński and Roberts (2024), S8.6. Therefore, we conjecture that Corollary 2.1 is only loose by $\frac{1}{2}$ degree. That is, if the target $\pi(x)$ where $x \in \mathbb{R}^d$ is multivariate Student's *t* distribution with degree of freedom larger than $d - 1$, then the SBPS algorithm is conjectured to be *uniformly ergodic*.

To finish this section, we present some typical sample paths obtained by implementing SBPS.

EXAMPLE 2.2. In Figure 4, we show the proposed SBPS without and with refreshment for standard Gaussian target in two dimensions. Note that the SBPS without refreshment is not irreducible in this case. The same issue exists for the Euclidean BPS for standard Gaussian targets. This issue can be fixed by adding the refreshment.
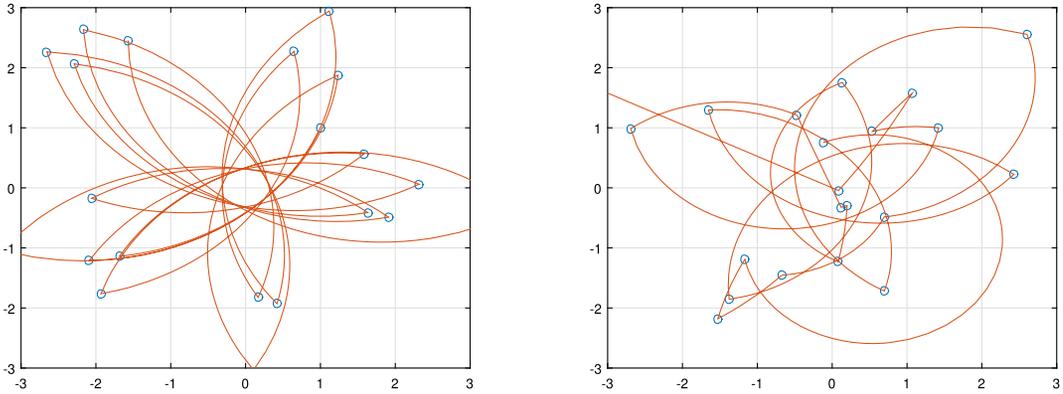
FIG. 4. *SBPS without (left) and with (right) refreshment for target distribution* $\mathcal{N}(0, I_2)$. *Note that the SBPS chain is not irreducible without refreshment.*

3. SPS: Isotropic targets.

In this section, we consider *isotropic targets*, which is also called *spherical symmetric targets*.[1] That is, $\pi(x)$ is only a function of $\|x\|$. Recall the mapping and equation (3) and the transformed target $\pi_S(z)$ in equation (4). We can see that, for any isotropic target, the transformed target $\pi_S(z)$ is only a function of the "latitude" z_{d+1}. In this sense, isotropic targets are the "best" targets for stereographically projected algorithms.

We shall assume that all second moments exist, and without loss of further generality, we suppose $\pi(x)$ satisfies that

$$\mathbb{E}_{X \sim \pi}\big[\|X\|^2\big] = d.$$

Under this assumption, the "optimal" radius should be chosen as $R = \sqrt{d}$, since it maps the concentration region of π to the neighborhood of the "equator" of the sphere. Throughout this section, we study this best scenario. We will study the robustness to R and the optimal scaling of SPS for any given R in Section 5 for another family of targets.

Under the above assumptions, it suffices to study the path of the absolute value of the "latitude" z_{d+1} of SPS. Informally, the "stationary phase" of SPS is the period in which $z_{d+1} = \mathcal{O}(d^{-1/2})$ and the "transient phase" is the period in which $|z_{d+1}|$ is larger than $\mathcal{O}(d^{-1/2})$. Somehow surprisingly, by analyzing the proposed "latitude" $\hat{z}_{d+1}$, we can (informally) show that the SPS enjoys the *blessings of dimensionality*: the number of iterations for the "latitude" of SPS to decrease to $\mathcal{O}(d^{-1/2})$ *decreases* with the dimension d. This implies (informally) that the "transient phase" of SPS is $\mathcal{O}(1)$ and it decreases with dimension.

3.1. *Analysis of the proposal distribution.*

We assume the chain starts from either the North Pole or the South Pole. By the assumptions, the chain is in the "stationary phase" once it is around the "equator".

In the following, we give useful approximations for the proposed "latitude" $\hat{z}_{d+1}$ in both the transient phase and stationary phase. This can be used to analyze the behavior of isotropic targets. See Yang, Łatuszyński and Roberts (2024, S9.1, for some simulations using the result from Lemma 3.1.

---

[1]Note that in some literature, isotropic distribution could be used to denote a distribution with zero mean and identity covariance matrix, which is different from our definition in this paper.

LEMMA 3.1. *Let z_i be the current ith coordinate and $\hat{z}_i$ be the ith coordinate of the proposal, where $i = 1, \ldots, d + 1$. Then we have the following expression*:

$$(6) \qquad \hat{z}_i = \frac{1}{\sqrt{1 + h^2(U^2 + U_\perp^2)}} \left(z_i + \sqrt{1 - z_i^2} hU \right),$$

where $U \sim \mathcal{N}(0, 1)$ and $U_\perp^2 \sim \chi_{d-1}^2$ which is independent with U. Furthermore, if $h = \mathcal{O}(d^{-1/2})$, then we have the following coordinatewise approximation:

$$(7) \qquad \hat{z}_i = \frac{1}{\sqrt{1 + h^2 U_\perp^2}} \left[\left(1 - \frac{1}{2} h^2 U^2 \right) z_i - \sqrt{1 - z_i^2} hU \right] + \mathcal{O}_\mathbb{P}(h^3 + dh^4 z_i).$$

As special cases of equation (7), z_{d+1} *is the current "latitude" and $\hat{z}_{d+1}$ be the proposed "latitude". Then, in the transient phase, if $z_{d+1}^2 = 1 - o(h^2)$, we have*

$$(8) \qquad \hat{z}_{d+1} = \frac{1}{\sqrt{1 + h^2(d - 1)}} \left(1 - \frac{1}{2} h^2 U^2 \right) z_{d+1} + \mathcal{O}_\mathbb{P}(d^{-1/2}),$$

and in the stationary phase, if $z_{d+1} = \mathcal{O}(d^{-1/2})$, we have

$$(9) \qquad \hat{z}_{d+1} = \frac{1}{\sqrt{1 + h^2(d - 1)}} (z_{d+1} - hU) + \mathcal{O}_\mathbb{P}(d^{-1}).$$

PROOF. See the Supplementary Material (Yang, Łatuszyński and Roberts (2024), S8.7). □

The above lemma informally suggests that in the "transient phase", the proposed "latitude" is almost deterministic, whereas in the "stationary phase", the proposed "latitude" approximately follows an autoregressive process. For example, if we take h to be constant and $i = d + 1$, then equation (6) shows that the proposal "latitude" $\hat{z}_{d+1}$ decreases to $\mathcal{O}(d^{-1/2})$ faster when dimension is larger since the concentration of $U_\perp^2/d$. If the proposals will be accepted by a probability bounded away from 0, the "transient phase" of SPS only involve $\mathcal{O}(1)$ iterations and the number even decreases with the dimension. As an example, in the next subsection, we show a class of isotropic targets such that the SPS enjoys the *blessings of dimensionality*.

3.2. *Examples of isotropic targets.* We denote the multivariate Student's t distributions by $\pi_\nu(x)$ where ν is the degree of freedom. We denote the standard multivariate Gaussian as the limit $\pi_\infty(x)$. That is,

$$\pi_\nu(x) \propto \left(1 + \frac{1}{\nu} \|x\|^2 \right)^{-(\nu+d)/2}, \qquad \pi_\infty(x) \propto \exp \left(-\frac{1}{2} \|x\|^2 \right).$$

Then the logarithm of the likelihood ratio can be written as a function of z_{d+1} and $\hat{z}_{d+1}$:

$$\log \frac{\pi_\nu(\hat{X})}{\pi_\nu(x)} + d \log \frac{R^2 + \|\hat{X}\|^2}{R^2 + \|x\|^2} = d \left(g_{\nu/d}(z_{d+1}) - g_{\nu/d}(\hat{z}_{d+1}) \right),$$

where $g_k(z) := \frac{k+1}{2} \log(k + \frac{1+z}{1-z}) + \log(1 - z)$. If $k = \nu/d \to \infty$, we have $g_k(z)$ converges to $g_\infty(z) := \frac{1}{1-z} - \frac{1}{2} + \log(1 - z)$ up to a constant. See Yang, Łatuszyński and Roberts (2024), S9.2, for the plots of the function of $g_k(z)$ for different values of k.

EXAMPLE 3.1. (Multivariate Student's t distribution with DoF $\nu = d$) It can be easily verified that $g_1(z) = \log(2)$. Therefore, any proposal will be accepted since the acceptance rate is always 1 whatever h is.

Algorithm 3: Generalized Stereographic Projection Sampler (GSPS)

- Let the current state be $X^d(t) = x$;
- Compute the proposal $\hat{X}$:
  - Let $z := \text{GSP}^{-1}(x)$;
  - Sample independently $d\tilde{z} \sim \mathcal{N}(0, h^2 I_{d+1})$;
  - Let $dz := d\tilde{z} - \frac{(z^T \cdot d\tilde{z})z}{\|z\|^2}$ and $\hat{z} := \frac{z+dz}{\|z+dz\|}$;
  - The proposal $\hat{X} := \text{GSP}(\hat{z})$.
- $X^d(t+1) = \hat{X}$ with probability $1 \wedge \frac{\pi(\hat{X})(R^2+\|\hat{X}\|_{\Lambda,Q}^2)^d}{\pi(x)(R^2+\|x\|_{\Lambda,Q}^2)^d}$; otherwise $X^d(t+1) = x$.

EXAMPLE 3.2. (Multivariate Student's t distribution with DoF $\nu > d$, including Gaussian) From Lemma 3.1, one can see informally: (i) in the "transient phase", as the proposed "latitude" is almost deterministic which has higher target density, the acceptance probability is almost 1 starting from either the North Pole or the South Pole; (ii) In the "stationary phase", as the proposed "latitude" approximately follows an autoregressive process, the acceptance rate is well-approximated by a positive constant. This suggests that as long as the target has a "lighter tail" than multivariate Student's t with DoF d, the "transient phase" of SPS takes at most $\mathcal{O}(1)$ steps. For comparison, for the standard multivariate Gaussian target, the "transient phase" of the Euclidean RWM takes $\mathcal{O}(d)$ steps (Christensen, Roberts and Rosenthal (2005)).

REMARK 3.1. Our theoretical results don't cover the cases of SPS for targets with heavier tails, such as multivariate Student's t with DoF $\nu < d$. In this case, the SPS cannot start from the North Pole since the first proposal of the SPS will be rejected with probability 1. One might consider to start from the South Pole. However, the SPS could get stuck at the South Pole if the proposal variance is large (even if the origin is the mode of the target density). See Yang, Łatuszyński and Roberts (2024, S8.12, for comments. In practice, we suggest choosing the initial state of the SPS as a random state uniformly sampled on the sphere.

4. SPS: Extension to elliptical targets.

4.1. *Extensions of stereographic projection.* Same as the previous section, we denote the state of the Markov chain by $x = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$. Suppose $z = (z_1, \ldots, z_{d+1})^T$ is the coordinates of a *unit sphere in* $\mathbb{R}^{d+1}$ (that is, $\|z\| = 1$).

Now, we map $z \in \mathbb{S}^d$ to $x \in \mathbb{R}^d$ by the generalized stereographic projection (GSP)

$$x = \text{GSP}(z) := Q\left(R\sqrt{\lambda_1}\frac{z_1}{1 - z_{d+1}}, \ldots, R\sqrt{\lambda_d}\frac{z_d}{1 - z_{d+1}}\right)^T,$$

where R is the radius parameter, $Q^T = Q^{-1} \in \mathbb{R}^{d \times d}$ is a rotation matrix, $\{\lambda_1, \ldots, \lambda_d\}$ are nonnegative constants. That is, the GSP is obtained by "stretching" via $\{\lambda_i\}$ and "rotating" via Q from the SP.

Defining the norm

$$\|x\|_{\Lambda,Q}^2 := x^T Q \Lambda^{-1} Q^T x,$$

where $\Lambda = \text{Diag}\{\lambda_1, \ldots, \lambda_d\}$, we have the following Generalized Stereographic Projection Sampler (GSPS, Algorithm 3) with parameters R, h, Λ, Q.

Note that even though GSPS can be formulated alternatively as using "ellipsoid" instead of "sphere", this is not our formulation. We still use the unit sphere to compute $\hat{z}$ for GSPS and only replace the mapping SP by GSP. That is, we only "stretch" and "rotate" the proposal $\hat{z}$ via GSP when projecting to $\mathbb{R}^d$.

The following example shows that we can extend isotropic targets to elliptical targets using the GSPS.

EXAMPLE 4.1. Suppose $\pi(x)$ is multivariate Student's t with covariance matrix $\frac{\nu}{\nu-2}\Sigma = \frac{\nu}{\nu-2}Q\Lambda Q^T$ where $\Lambda = \text{Diag}\{\lambda_1, \ldots, \lambda_d\}$:

$$\pi(x) \propto \left(1 + \frac{1}{\nu}x^T\Sigma^{-1}x\right)^{-(\nu+d)/2}.$$

Then, for the GSPS with the corresponding Q and Λ, and $R^2 = \sum_i \lambda_i$, the acceptance rate is always 1 for any h.

The GSPS naturally suggests that one can estimate $\Sigma = Q\Lambda Q^T$ under the adaptive MCMC framework, which results in adaptive GSPS. Indeed, if both Λ and Q are known, then one can normalize GSPS and reduce it to SPS, which is GSPS with $\Sigma = I_d$.

As the robustness to estimations of the mean and the covariance matrix of the target is the key to the success of adaptive GSPS, we study the robustness of Gaussian targets in the next section.

4.2. *Robustness for Gaussian targets.* We consider multivariate Gaussian targets with mean vector μ and covariance matrix Σ:

$$\pi_{\mu,\Sigma}(x) \propto \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right),$$

where

$$\Sigma = \text{Diag}(\lambda_1, \ldots, \lambda_d), \qquad \mu = (\mu_1, \ldots, \mu_d)^T.$$

We are interested in the robustness when $\mu \neq 0$ and $\Sigma \neq I_d$.

For comparison with the traditional RWM in $\mathbb{R}^d$ on the orders of stepsize, we recall that the optimal scaling theory gives an optimal stepsize of $\mathcal{O}(d^{-1/2})$ for RWM (Roberts, Gelman and Gilks (1997)). Our stepsize h is defined on the unit sphere. When projecting to $\mathbb{R}^d$, we multiply by $R = \mathcal{O}(d^{1/2})$. Therefore, the optimal stepsize of RWM roughly corresponds to $h = \mathcal{O}(d^{-1})$ in our setting. In this subsection, we study how large the stepsize h can be so the expected acceptance probability of SPS goes to 1. This roughly corresponds to a stepsize of $\mathcal{O}(hd^{1/2})$ in $\mathbb{R}^d$. Our results show that the order of the stepsize of SPS when projected back to $\mathbb{R}^d$ is always no smaller than the order of the optimal stepsize of RWM.

THEOREM 4.1. *Assume there exist constants $C < \infty$ and $c > 0$ such that $c \leq \lambda_i \leq C$ and $|\mu_i| \leq C$ for all $i = 1, \ldots, d$. Furthermore, assume $R = d^{1/2}$ and*

$$(10) \qquad \left|\sum_{i=1}^d \mu_i^2 - \sum_{i=1}^d (1-\lambda_i)\right| = \mathcal{O}(d^\alpha),$$

where $\alpha \leq 1$. Then, under stationarity that $X \sim \pi$, the expected acceptance probability converges to 1 as $d \to \infty$

$$\mathbb{E}_{X \sim \pi_{\mu,\Sigma}}\mathbb{E}_{\hat{X}|X}\left[1 \wedge \frac{\pi_{\mu,\Sigma}(\hat{X})(R^2 + \|\hat{X}\|^2)^d}{\pi_{\mu,\Sigma}(X)(R^2 + \|X\|^2)^d}\right] \to 1,$$

for all h such that

(11)
$$h = o\left(\frac{d^{-1}}{\sqrt{\max\{\frac{1}{d}\sum_i |1-\lambda_i|, \frac{1}{d}\sum_i \mu_i^2\}}} \wedge d^{-(\frac{1}{2}\vee\alpha)}\right).$$

PROOF.   See the Supplementary Material (Yang, Łatuszyński and Roberts (2024), S8.8).
□

Note that for the standard Gaussian target, we know that the acceptance rate goes to 1 as $d \to \infty$ for any h. Recall that for traditional RWM in dimension k, the optimal stepsize is $\mathcal{O}(k^{-1/2})$ in $\mathbb{R}^k$ under the optimal scaling framework (Roberts, Gelman and Gilks (1997)), which corresponds to $h = \mathcal{O}(k^{-1/2}d^{-1/2})$ in the SPS setting. Therefore, Theorem 4.1 suggests that the "effective dimension" is determined by $\{\mu_i\}$ and $\{\lambda_i\}$: as long as $\{\lambda_i\}$ and $\{\mu_i\}$ are uniformly bounded, the "effective dimension" of SPS is never larger than d, which is the "effective dimension" of the traditional RWM. Furthermore, we can compare the "effective dimension" for SPS with d using Theorem 4.1 under different settings.

EXAMPLE 4.2.   Suppose $\mu_i = 0$ for all $i = 1, \ldots, d$. Moreover, $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 2$ and $\lambda_{k+1} = \cdots = \lambda_d = 1$, then as long as k is a fixed number, $\frac{1}{d}\sum_i |1-\lambda_i| = \mathcal{O}(kd^{-1})$. By Theorem 4.1, when $h = o(k^{-1/2}d^{-1/2})$ then the acceptance rate goes to 1 as $d \to \infty$. This suggests that the "effective dimension" for SPS is no more than k.

EXAMPLE 4.3.   Suppose $\lambda_i = 0$ for all $i = 1, \ldots, d$. Furthermore, $\mu_1 = \mu_2 = \cdots = 1$ and $\mu_{k+1} = \cdots = \mu_d = 0$ where k is a fixed number. Then we have $\frac{1}{d}\sum_i \mu_i^2 = \mathcal{O}(kd^{-1})$. By Theorem 4.1, the acceptance rate goes to 1 when $h = o(k^{-1/2}d^{-1/2})$, which implies the "effective dimension" for SPS is no more than k.

One can consider Theorem 4.1 in the context of adaptive MCMC for Gaussian targets, in which $\{\mu_i\}$ and $\{1-\lambda_i\}$ represent the "estimation errors" of the coordinate means and eigenvalues of the covariance matrix of the target. One can choose the "radius" R of the stereographic projection properly to satisfy equation (10) with $\alpha = \frac{1}{2}$. This is to properly scale R so that the "latitude" on the unit sphere of the SPS is $\mathcal{O}_{\mathbb{P}}(d^{-1/2})$. Then, according to Theorem 4.1, the "effective dimension" of SPS is smaller than d if $\frac{1}{d}\sum_i |1-\lambda_i| = o(1)$ and $\frac{1}{d}\sum_i \mu_i^2 = o(1)$. However, the above results on "effective dimension" are based on Theorem 4.1, which only hold for Gaussian targets. For more results on robustness and optimal acceptance rate for adaptive MCMC, we will study another family of targets in Section 5.

Finally, we show two examples that the result of Theorem 4.1 is tight. First, consider the special case that $\mu_i = \mu > 0$ and $\lambda_i = \sigma^2$ for all i. We assume $\mu^2 = 1 - \sigma^2$ so that equation (10) holds for any $\alpha \leq 1/2$. Then equation (11) suggests that the acceptance probability goes to zero if $h = o(d^{-1}/\mu)$. On the other hand, the target in this special case is a product i.i.d. target with marginal distribution $\mathcal{N}(\mu, 1-\mu^2)$. By our optimal scaling results in Section 5, we know that the acceptance probability does not go to zero if $h = \mathcal{O}(d^{-1}/\mu)$ (see Lemma 5.1 and Corollary 5.2). Therefore, equation (11) is tight. Second, consider the special case that $\mu_i = \mu = 0$ and $\lambda_i = \sigma^2 \neq 1$ for all i. Then equation (10) holds for $\alpha = 1$. In this case, equation (11) suggests that the acceptance probability goes to zero if $h = o(d^{-\alpha}) = o(d^{-1})$. On the other hand, the target in this special case is a product i.i.d. target with marginal $\mathcal{N}(0, \sigma^2)$ where $\sigma^2 \neq 1$. If we properly rescale R and σ^2, it reduces to the case of Lemma 5.1 where the target is the standard Gaussian but $R \neq d^{1/2}$. Therefore, by Lemma 5.1, the acceptance probability doesn't converge to zero if $h = \mathcal{O}(d^{-1})$. Therefore, equation (11) is again tight.

5. SPS in high-dimensional problems. In this section, we shall give a case study of the behavior of SPS for high-dimensional target densities. We shall consider various limits of SPS as $d \to \infty$ though to have tractability of this limit we need to consider a very specialised class of target densities. To this end, analogous to Roberts, Gelman and Gilks (1997) we assume the target $\pi(x)$ has a product i.i.d. form.

5.1. *Assumptions on π.* We assume the target $\pi(x)$ has a product i.i.d. form:

$$\pi(x) = \prod_{i=1}^{d} f(x_i). \tag{12}$$

Without loss of generality, we assume f is normalized such that

$$\mathbb{E}_f(X^2) = \int x^2 f(x)\,\mathrm{d}x = 1, \qquad \mathbb{E}_f(X^6) < \infty. \tag{13}$$

We further assume f'/f is Lipschitz continuous, $\lim_{x \to \pm\infty} x f'(x) = 0$, and

$$\mathbb{E}_f\left[\left(\frac{f'(X)}{f(X)}\right)^8\right] < \infty, \qquad \mathbb{E}_f\left[\left(\frac{f''(X)}{f(X)}\right)^4\right] < \infty, \qquad \mathbb{E}_f\left[\left(\frac{Xf'(X)}{f(X)}\right)^4\right] < \infty. \tag{14}$$

REMARK 5.1. Under the assumption $\mathbb{E}_f(X^2) = 1$, by Cauchy–Schwarz inequality

$$\mathbb{E}_f\left[((\log f)')^2\right] \geq 1,$$

where the equality is achieved by standard Gaussian (or truncated standard Gaussian). For univariate Student's t distribution with any DoF $= \nu > 2$, rescaling by a factor $\sqrt{\frac{\nu-2}{\nu}}$, we can obtain a target density f_ν with $\mathbb{E}_{f_\nu}(X^2) = 1$ and

$$C_\nu := \mathbb{E}_{f_\nu}\left[((\log f_\nu)')^2\right] = \left(\frac{\nu}{\nu-2}\right)\left(\frac{\nu+1}{\nu}\right)\left(\frac{\nu+4}{\nu+3}\right)\sqrt{\frac{\nu+4}{\nu}} > 1,$$

where $\nu \to \infty$ recovers the case for the standard Gaussian target. See Table 1 for values of C_ν for different ν. One can see that C_ν is very close to 1 for even a medium size of ν.

In this section, we consider f has full support in $\mathbb{R}$. Then the product i.i.d. target π is not isotropic unless it is the standard Gaussian.

5.2. *Acceptance probability.* To make progress, we need a detailed understanding of the high-dimensional behavior of the acceptance probability of SPS. It turns out that to optimally apply SPS, we need to scale R to be $\mathcal{O}(d^{1/2})$. Not doing this will concentrate mass at either the North or South polls in an ultimately degenerate way. We shall thus assume R to be scaled in this way in what follows. Therefore, we consider $R = \sqrt{\lambda d}$ for a fixed constant λ. To simplify the final result, we replace the step size h by another parameter ℓ via

$$h = \frac{1}{\sqrt{d-1}}\left[\frac{1}{(1 - \frac{\ell^2}{2d}\frac{4\lambda}{(1+\lambda)^2})^2} - 1\right]^{1/2} \tag{15}$$

TABLE 1
Examples of C_ν and $C_\nu/(C_\nu - 1)$ for different ν in Remark 5.1 and Remark 5.4

	$\nu = 3$	$\nu = 5$	$\nu = 10$	$\nu = 20$	$\nu = 50$	$\nu = 100$
C_ν	7.1285	3.0187	1.7521	1.3336	1.1250	1.0612
$C_\nu/(C_\nu - 1)$	1.1632	1.4954	2.3297	3.9977	8.9990	17.3328

which implies $\frac{1}{\sqrt{1+h^2(d-1)}} = 1 - \frac{\ell^2}{2d}\frac{4\lambda}{(1+\lambda)^2}$. Note that ℓ is simply a reparameterization of h.

When ℓ is a fixed constant, h is scaled as $\mathcal{O}(d^{-1})$ since $h \approx \frac{\ell}{d}\sqrt{4\lambda/(1+\lambda)^2}$.

LEMMA 5.1. *Under the assumptions on π in Section 5.1, suppose the current state $X \sim \pi$, and the parameter of the algorithm $R = \sqrt{\lambda d}$, where $\lambda > 0$ is a fixed constant. We reparameterize h by ℓ according to equation* (15). *Then, if either $\lambda \neq 1$ or f is not the standard Gaussian density, there exists a sequence of sets $\{F_d\}$ such that $\pi(F_d) \to 1$ and*

$$\sup_{X \in F_d} \mathbb{E}_{\hat{X}|X}\left[\left|1 \wedge \frac{\pi(\hat{X})(R^2 + \|\hat{X}\|^2)^d}{\pi(X)(R^2 + \|X\|^2)^d} - 1 \wedge \exp(W_{\hat{X}|X})\right|\right] = o(d^{-1/4}\log(d)),$$

where $W_{\hat{X}|X} \sim \mathcal{N}(\mu, \sigma^2)$ and

$$\mu = \frac{\ell^2}{2}\left\{\frac{4\lambda}{(1+\lambda)^2} - \mathbb{E}_f[((\log f)')^2]\right\}, \qquad \sigma^2 = \ell^2\left\{\mathbb{E}_f[((\log f)')^2] - \frac{4\lambda}{(1+\lambda)^2}\right\}.$$

PROOF.   See the Supplementary Material (Yang, Łatuszyński and Roberts (2024), S8.9). □

REMARK 5.2.   Lemma 5.1 requires either $\lambda \neq 1$ or π is not the standard multivariate Gaussian. If $\lambda = 1$ and π is the standard multivariate Gaussian, the case reduces to isotropic targets discussed in Section 3.2, so the Gaussian approximation in Lemma 5.1 doesn't hold.

5.3. *Optimisation and robustness of SPS.*   Note that Lemma 5.1 suggests that the expected acceptance probability in the stationary phase

$$\mathbb{E}_{X \sim \pi}\mathbb{E}_{\hat{X}|X}\left[1 \wedge \frac{\pi(\hat{X})(R^2 + \|\hat{X}\|^2)^d}{\pi(X)(R^2 + \|X\|^2)^d}\right] \to 2\Phi\left(-\frac{\sigma}{2}\right).$$

Furthermore, as $\mathbb{E}[\|\hat{X} - X\|^2] \to \ell^2$, we obtain the commonly used approximation of Expected Squared Jumping Distance (ESJD):

$$\mathbb{E}[\|\hat{X} - X\|^2] \cdot \mathbb{E}\left[1 \wedge \frac{\pi(\hat{X})(R^2 + \|\hat{X}\|^2)^d}{\pi(X)(R^2 + \|X\|^2)^d}\right]$$

(16)

$$\to 2\ell^2 \cdot \Phi\left(-\frac{\ell}{2}\sqrt{\mathbb{E}_f[((\log f)')^2] - \frac{4\lambda}{(1+\lambda)^2}}\right).$$

In this subsection, we shall explicitly consider the optimisation of SPS, and take a close look at its relative performance in comparison to standard Euclidean RWM. All the results in this section to date have used the convenient restriction that $\mathbb{E}_f(X^2) = 1$. However, at this point we shall need to generalise this notion. Therefore, consider density f to have mean m and variance s^2.

We consider SPS with $R = \sqrt{\lambda(s^2 + m^2)d}$. and stepsize $h = \frac{1}{\sqrt{d-1}}[\frac{1}{(1-\frac{\ell^2}{2d}\frac{4\lambda}{(1+\lambda)^2})^2} - 1]^{1/2}$.

We shall denote this algorithm's approximate limiting ESJD by $E(\ell, m, s, \lambda)$ and its corresponding acceptance probability: $A(\ell, m, s, \lambda)$. The following result is a direct consequence of equation (16) applied to a scaled density (dividing by $\sqrt{s^2 + m^2}$):

$$E(\ell, m, s, \lambda) = \sqrt{s^2 + m^2} \cdot 2\ell^2\Phi\left(\frac{-\ell}{2}\sqrt{\tilde{I}(s, m, \lambda)}\right),$$

$$A(\ell, m, s, \lambda) = 2\Phi\left(\frac{-\ell}{2}\sqrt{\tilde{I}(s, m, \lambda)}\right),$$

where $\tilde{I}(s, m, \lambda) = (s^2 + m^2)I - \frac{4\lambda}{(1+\lambda)^2}$ and $I = \mathbb{E}_f((\log f)'(X)^2)$.

We note that we can readily recover the the corresponding quantities for RWM: $E(\ell, m, s, \infty)$ and $A(\ell, m, s, \infty)$. In particular, to consider robustness, we shall consider the relative performance ratio

$$\tilde{R}(m, s, \lambda) := \frac{\sup_\ell E(\ell, m, s, \lambda)}{\sup_\ell E(\ell, m, s, \infty)}.$$

To get an expression for $\tilde{R}$, we shall follow the standard approach of Roberts, Gelman and Gilks (1997) of expressing the efficiency in terms of acceptance rate. To that end, we define

$$\tilde{E}(A, m, s, \lambda) := E\big(\ell(A), m, s, \lambda\big),$$

where ℓ is chosen to be the unique solution to $A(\ell(A), m, s, \lambda) = A$. A simple calculation yields the following.

COROLLARY 5.1. *We have* $\tilde{E}(A, m, s, \lambda) = A \cdot \Phi^{-1}(\frac{A}{2}) \cdot \frac{4(s^2 + m^2)}{\tilde{I}(s, m, \lambda)}$ *and* $\tilde{R}(m, s, \lambda) = \frac{(s^2 + m^2)I}{\tilde{I}(s, m, \lambda)} = \frac{1}{1 - \alpha \cdot \beta \cdot \gamma}$, *where* $\alpha = \frac{4\lambda}{(1 + \lambda)^2}$, $\beta = \frac{s^2}{s^2 + m^2}$, $\gamma = \frac{1}{s^2 I}$. *Since all of* α, β, γ *are in* $[0, 1]$, *In this situation, SPS is never less efficient than Euclidean RWM.*

REMARK 5.3. Primarily, Corollary 5.1 is a robustness result. However, it also highlights the (only) ways in which SPS can fail to outperform Euclidean RWM. The three constants α, β, γ characterise different sensitivities of the algorithm.

α measures the penalty due to misspecification of the sphere radius. $4\lambda/(1 + \lambda)^2$ is optimised at $\lambda = 1$ when there is no penalty for misspecification of the target distribution dispersion.

β describes the penalty due to mislocation of the hyper-sphere. It is seen that the optimal choice is to locate the sphere at the mean of the target distribution.

γ is a distribution-specific penalty. It is straightforward by to check by functional calculus that $\gamma \leq 1$ with equality achieved only in the case where the target density is Gaussian. Thus γ is characterising proximity to Gaussianity.

From this result, it can be seen that we can only achieve super-efficiency (convergence complexity more rapid than $\mathcal{O}(d)$) when all three parameters, α, β and γ are close to (and converging to) unity.

5.4. *Maximizing ESJD.* In this subsection, we consider $\lambda = 1$ only (i.e., $R = \sqrt{d}$) and prove that the approximate limiting ESJD in equation (16) is indeed the limiting maximum ESJD. We will demonstrate (again analogously to Roberts, Gelman and Gilks (1997)) that its limit is optimised by targeting an acceptance probability of 0.234.

DEFINITION 5.1. Expected Squared Jumping Distance (ESJD):

$$\text{ESJD} := \mathbb{E}_{X \sim \pi} \mathbb{E}_{\hat{X} \mid X} \left[\|\hat{X} - X\|^2 \left(1 \wedge \frac{\pi(\hat{X})(R^2 + \|\hat{X}\|^2)^d}{\pi(X)(R^2 + \|X\|^2)^d}\right) \right].$$

THEOREM 5.1. *Under the assumptions on the target in Section 5.1, suppose f is not the standard Gaussian density, the SPS chain is in the stationary phase, and the radius parameter is chosen as $R = \sqrt{d}$ and reparameterize h by ℓ according to equation (15) with $\lambda = 1$. Then, as $d \to \infty$, we have* $\text{ESJD} \to 2\ell^2 \cdot \Phi(-\frac{\ell}{2}\sqrt{\mathbb{E}_f[((\log f)')^2]} - 1)$.

PROOF. See the Supplementary Material (Yang, Łatuszyński and Roberts (2024), S8.10). □

Algorithm 4: Revised Stereographic Projection Sampler (RSPS)

- Let the current state be $X^d(t) = x$;
- Compute the proposal $\hat{X}$:
  - Let $z := \mathrm{SP}^{-1}(x)$;
  - Sample independently $d\tilde{z}', d\tilde{z}'' \sim \mathcal{N}(0, h^2 I_{d+1})$;
  - Let $dz' := d\tilde{z} - \frac{(z^T \cdot d\tilde{z}')z}{\|z\|^2}$ and $dz'' := d\tilde{z} - \frac{(z^T \cdot d\tilde{z}'')z}{\|z\|^2}$;
  - Let $\hat{z}' := \frac{z+dz'}{\|z+dz'\|}$ and $\hat{z}'' := \frac{z+dz''}{\|z+dz''\|}$;
  - Two independent proposals $\hat{X}' := \mathrm{SP}(\hat{z}')$ and $\hat{X}'' := \mathrm{SP}(\hat{z}'')$;
  - The proposal $\hat{X} := (\hat{X}'_1, \hat{X}''_{2:d})$.
- $X^d(t+1) = \hat{X}$ with probability $1 \wedge \frac{\pi(\hat{X})(R^2+\|\hat{X}\|^2)^d}{\pi(x)(R^2+\|x\|^2)^d}$; otherwise $X^d(t+1) = x$.

COROLLARY 5.2. *The maximum ESJD is approximately* $\frac{1.3}{\mathbb{E}_f[((\log f)')^2]-1}$, *which is achieved when the acceptance rate is about* 0.234. *The optimal* $\hat{\ell} \approx \frac{2.38}{\sqrt{\mathbb{E}_f[((\log f)')^2]-1}}$ *and the optimal* $\hat{h} = \frac{1}{\sqrt{d-1}}[(1 - \frac{\hat{\ell}^2}{2d})^{-2} - 1]^{1/2} \approx \frac{\hat{\ell}}{\sqrt{d(d-1)}} \approx \frac{2.38}{\sqrt{d(d-1)}} \frac{1}{\sqrt{\mathbb{E}_f[((\log f)')^2]-1}}$.

REMARK 5.4. Compared with the maximum ESJD of RWM, the maximum ESJD of SPS is $\frac{\mathbb{E}_f[((\log f)')^2]}{\mathbb{E}_f[((\log f)')^2]-1}$ times larger. For example, for the class of distributions defined in Remark 5.1 indexed by ν, we can compute $C_\nu/(C_\nu - 1)$. See Table 1 for examples of $C_\nu/(C_\nu - 1)$ for different ν. One can see that the maximum ESJD of SPS can be much larger than the maximum ESJD of RWM even for a medium size of ν.

5.5. *Diffusion limit.*   Continuing our analogy to the Euclidean RWM case, we shall provide a diffusion limit result, giving a more explicit description of SPS for high-dimensional situations. However, it is difficult to obtain a diffusion limit directly for SPS. Instead, we shall slightly change the original SPS algorithm in a way that is asymptotically negligible (as d gets large) but which greatly facilitates our limiting diffusion approach. The revised algorithm, Algorithm 4, is called RSPS.

Note that the difference between SPS and RSPS is that in RSPS two independent proposals $\hat{X}'$ and $\hat{X}''$ are first computed. Then the final proposal is composed by the first coordinate of $\hat{X}'$ and other coordinates of $\hat{X}''$, that is, $\hat{X} := (\hat{X}'_1, \hat{X}''_{2:d})$. This design guarantees $\hat{X}_1$ is independent with $\hat{X}_{2:d}$ conditional on the current state, which is the technical condition needed to prove the following result on a diffusion limit.

THEOREM 5.2.   *Under the assumptions on π in Section* 5.1, *suppose f is not the standard Gaussian density and the RSPS chain $\{X^d(t)\}$ starts from the stationarity, that is, $X^d(0) \sim \pi$, and the radius parameter is chosen as $R = \sqrt{d}$. Writing $X^d(t) = (X^d_1(t), \ldots, X^d_d(t))$, we let $U^d(t) := X^d_1(\lfloor dt \rfloor)$ be the sequence of the first coordinates of $\{X^d(t)\}$ sped-up by a factor of d. Then, as $d \to \infty$, we have $U^d \Rightarrow U$, where $\Rightarrow$ denotes weak convergence in Skorokhod topology, and U satisfies the following Langevin SDE:*

$$dU(t) = (s(\ell))^{1/2} dB(t) + s(\ell)\frac{f'(U(t))}{2f(U(t))} dt,$$

where $s(\ell) := 2\ell^2 \Phi(-\ell \frac{\sqrt{\mathbb{E}_f[((\log f)')^2]-1}}{2})$ is the speed measure for the diffusion process, and $\Phi(\cdot)$ being the standard Gaussian cumulative density function.

PROOF. See the Supplementary Material (Yang, Łatuszyński and Roberts (2024), S8.11).
□

COROLLARY 5.3. *The optimal acceptance rate for RSPS is* 0.234 *and the maximum speed of the diffusion limit is* $s(\hat{\ell}) \approx \frac{1.3}{\mathbb{E}_f[((\log f)')^2]-1}$ *where* $\hat{\ell} \approx \frac{2.38}{\sqrt{\mathbb{E}_f[((\log f)')^2]-1}}$. *The optimal* $\hat{h} = \frac{1}{\sqrt{d-1}}[(1-\frac{\hat{\ell}^2}{2d})^{-2}-1]^{1/2} \approx \frac{\hat{\ell}}{\sqrt{d(d-1)}} \approx \frac{2.38}{\sqrt{d(d-1)}}\frac{1}{\sqrt{\mathbb{E}_f[((\log f)')^2]-1}}$.

REMARK 5.5. Compared with the maximum speed of the diffusion limit of RWM (Roberts, Gelman and Gilks (1997)), the maximum speed of the diffusion limit of RSPS is $\frac{\mathbb{E}_f[((\log f)')^2]}{\mathbb{E}_f[((\log f)')^2]-1}$ times larger.

REMARK 5.6. A reasonable conjecture is that the same diffusion limit holds for the original SPS algorithm. In order to establish the same diffusion limit, if we follow the same arguments as in the proof of Theorem 5.1, it is required to show the acceptance rate term becomes "asymptotically independent" with the first coordinate as a rate of $\mathcal{O}(d^{-1/2})$. However, our current technical arguments in Theorem 5.1 can achieve a rate of $\mathcal{O}(d^{-1/8})$ which is not enough for establishing the weak convergence to a diffusion limit. Therefore, we only prove the diffusion limit for RSPS in this paper and leave the proof for SPS as an open problem.

6. Simulations. In this section, we study the proposed SPS and SBPS through numerical examples. In most of the examples, we consider $d = 100$ dimensions and two choices of target distributions, the heavy-tailed multivariate Student's t target with d degree of freedom and the standard Gaussian target. By default, we choose $R = \sqrt{d}$ for SPS and SBPS. We refer to Yang, Łatuszyński and Roberts (2024), S9, for additional simulations such as different choices of R.

6.1. *SPS*: *Bayesian Cauchy regression.* Here we return to Example 1.1. In Figure 1, we choose $a = b = 0.1$, $d = 11$, $n = 15$, and $R = \sqrt{d}$. Random design $X_i \sim \mathcal{N}(0, I) \in \mathbb{R}^d$ and responses $\{Y_i\}_{i=1}^n$ are generated by $Y_i = \alpha_0 + \beta_0^T X_i + \epsilon_i$ where $\alpha_0 = -2$, $\beta_0 = (-4, -3, -2, -1, 0, 1, 2, 3, 4)^T$, $\{\epsilon_i\}$ are i.i.d. zero-mean Cauchy distribution with scale $\gamma_0 = 1$. We take a logarithm transformation for γ and implement both SPS and RWM in $\mathbb{R}^d$. RWM starts from $(100, 100, \ldots, 100)$ and SPS starts from the North Pole. From the figure, one can see clearly that: (1) RWM which is not geometric ergodic completely failed; (2) the proposed SPS which is uniformly ergodic converged extremely fast.

6.2. *SPS*: *ESJD per dimension.* In this example, we study the ESJD for SPS and its robustness to the choice of the radius R. We first tune the proposal stepsize h of SPS to get different acceptance rates. Then we plot ESJD per dimension for varying acceptance rates as the efficiency curve of SPS. Figure 5 shows eight efficiency curves for different choices of R. The target distribution is multivariate Student's t distribution with $d = 100$ degrees of freedom. In this setting, $R = \sqrt{d}$ is optimal. The four subplots in the first column are for $R < \sqrt{d}$ and the second column contains four cases of $R > \sqrt{d}$.

Although we do not plot the ESJD for RWM, the maximum ESJD per dimension for RWM is known to have an order of $\mathcal{O}(d^{-1})$. In all cases in Figure 5, SPS has a much larger ESJD than RWM. For the two subplots in the first row of Figure 5, since R is closer to $\sqrt{d}$, the acceptance rate cannot be lower than 0.5 whatever the proposal variance is. For all the other efficiency curves, an interesting observation is that the maximum ESJDs are always achieved when the acceptance rate is around 0.234. This suggests the optimal acceptance rate 0.234 is quite robust and not limited to product i.i.d. targets, which is similar to the case of optimal acceptance rate 0.234 for RWM (Yang, Roberts and Rosenthal (2020)).
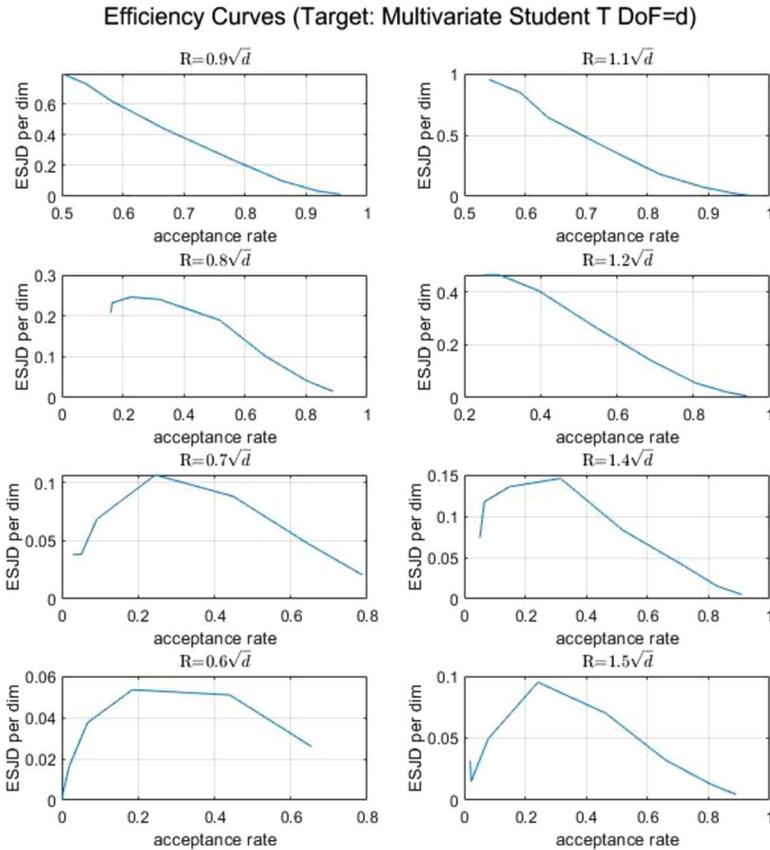
FIG. 5. *Efficiency Curves (ESJD per dimension) as functions of the acceptance rate of SPS for different choices of R. Target distribution is multivariate Student's t distribution with DoF = d = 100. When R is close to $\sqrt{d}$ (such as $R = 0.9\sqrt{d}$ and $R = 1.1\sqrt{d}$), the acceptance rate is always larger than 0.234.*

6.3. *SBPS: ESS per switch.* In this example, we study the efficiency curves of SBPS and BPS in terms of ESS per Switch versus the refresh rate. The first subplot of Figure 6 contains the proportion of refreshments in all the N events for varying refresh rates. It is clear that as the refresh rate increases, the proportion of refreshments increases. In the other three subplots of Figure 6, we plot the logarithm of ESS per Switch as a function of the refresh rate for three cases, the 1st coordinate, the negative log-density, and the squared 1st coordinate, respectively. For each efficiency curve, $N = 1000$ events are simulated. We use random initial states for our SBPS. For comparison, BPS starts from stationarity to avoid slow mixing. As SBPS and BPS are continuous-time processes, each unit time period is discretized into 5 samples. The target is standard Gaussian with $d = 100$. We refer to Yang, Łatuszyński and Roberts (2024), S9, for additional efficiency curves for heavy-tailed targets.

According to Figure 6, the ESS per Switch of SBPS is much larger than the ESS per Switch of BPS for all cases (actually the gap becomes larger in higher dimensions). For both the 1st coordinate and the negative log-target density, the ESS per Switch of SBPS can be larger than 1 if the refresh rate is relatively low. For BPS, however, even starting from stationarity, the ESS per Switch is always much smaller than 1.

7. Discussion. We have explored the use of stereographically projected algorithms and developed SPS and SBPS that are uniformly ergodic for a large class of light and heavy-tailed targets and also exhibit fast convergence in high-dimensions. The framework established

FIG. 6. *Efficiency curves (ESS per Switch) for SBPS and BPS for varying refresh rate. N = 1000 events are simulated, Random initial values for SBPS and BPS starting from stationarity. Each unit time is discretized to 5 samples. Target distribution: standard Gaussian with d = 100. The first subplot is the proportion of refreshment events in all N events. The other three subplots are ESS for the 1st coordinate, the negative log density, and the squared 1st coordinate.*

opens new opportunities for developing MCMC algorithms for sampling high-dimensional and/or heavy-tailed distributions. We finish the paper with some future directions.

- *Adaptive MCMC*: the proposed GSPS algorithm fits the adaptive MCMC framework very well (Roberts and Rosenthal (2009)). The tuning parameters of GSPS such as the covariance matrix, the location as well as the radius of the sphere can be tuned adaptively. Empirical study of the sensitivity of GSPS to the tuning parameters and extending GSPS for sampling multimodal distributions (Pompe, Holmes and Łatuszyński (2020)) are important future directions.
- *Quantitative bounds*: we know SPS is uniform ergodicity for a large class of targets and dimension-free for isotropic targets. However, we do not establish quantitative bounds for the mixing time. The quantitative bound and its dimension dependence (e.g., Yang and Rosenthal (2023)) would be an interesting direction for future work.
- *Scaling limit for SBPS*: we know SBPS is dimension-free for isotropic targets. For product i.i.d. targets, we establish optimal scaling results for SPS but not for SBPS. Recently, scaling limits for the traditional BPS have been developed in different settings (Bierkens, Kamatani and Roberts (2022), Deligiannidis et al. (2021)). It would be interesting to obtain such scaling limits for SBPS for nonisotropic targets to compare with BPS directly.
- *Stereographic MALA, HMC, and others*: another future direction is to develop new MCMC samplers based on other popular MCMC algorithms, such as (Riemann) MALA and Hamiltonian Monte Carlo (HMC) (Girolami and Calderhead (2011)) and other PDMPs, or based on other mappings from $\mathbb{R}^d$ to $\mathbb{S}^d$ than the (generalized) stereographic projection.

SUPPLEMENTARY MATERIAL

Supplement to "Stereographic Markov chain Monte Carlo" (DOI: 10.1214/24-AOS2426SUPP; .pdf). The supplement contains technical proofs and some additional simulations.

REFERENCES

ANDRIEU, C., DOBSON, P. and WANG, A. Q. (2021). Subgeometric hypocoercivity for piecewise-deterministic Markov process Monte Carlo methods. *Electron. J. Probab.* **26** 1–26. MR4269208 https://doi.org/10.1214/21-EJP643

ANDRIEU, C., DURMUS, A., NÜSKEN, N. and ROUSSEL, J. (2021). Hypocoercivity of piecewise deterministic Markov process-Monte Carlo. *Ann. Appl. Probab.* **31** 2478–2517. MR4332703 https://doi.org/10.1214/20-aap1653

BERNARD, E. P., KRAUTH, W. and WILSON, D. B. (2009). Event-chain Monte Carlo algorithms for hard-sphere systems. *Phys. Rev. E* **80** 056704.

BIERKENS, J., FEARNHEAD, P. and ROBERTS, G. (2019). The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* **47** 1288–1320. MR3911113 https://doi.org/10.1214/18-AOS1715

BIERKENS, J., KAMATANI, K. and ROBERTS, G. O. (2022). High-dimensional scaling limits of piecewise deterministic sampling algorithms. *Ann. Appl. Probab.* **32** 3361–3407. MR4497848 https://doi.org/10.1214/21-aap1762

BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2018). The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113** 855–867. MR3832232 https://doi.org/10.1080/01621459.2017.1294075

CHRISTENSEN, O. F., ROBERTS, G. O. and ROSENTHAL, J. S. (2005). Scaling limits for the transient phase of local Metropolis–Hastings algorithms. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 253–268. MR2137324 https://doi.org/10.1111/j.1467-9868.2005.00500.x

COTTER, S. L., ROBERTS, G. O., STUART, A. M. and WHITE, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statist. Sci.* **28** 424–446. MR3135540 https://doi.org/10.1214/13-STS421

COXETER, H. S. M. (1961). *Introduction to Geometry*. Wiley, New York. MR0123930

DAVIS, M. H. A. (1984). Piecewise-deterministic Markov processes: A general class of nondiffusion stochastic models. *J. Roy. Statist. Soc. Ser. B* **46** 353–388. MR0790622

DELIGIANNIDIS, G., BOUCHARD-CÔTÉ, A. and DOUCET, A. (2019). Exponential ergodicity of the bouncy particle sampler. *Ann. Statist.* **47** 1268–1287. MR3911112 https://doi.org/10.1214/18-AOS1714

DELIGIANNIDIS, G., PAULIN, D., BOUCHARD-CÔTÉ, A. and DOUCET, A. (2021). Randomized Hamiltonian Monte Carlo as scaling limit of the bouncy particle sampler and dimension-free convergence rates. *Ann. Appl. Probab.* **31** 2612–2662. MR4350970 https://doi.org/10.1214/20-aap1659

DURMUS, A., GUILLIN, A. and MONMARCHÉ, P. (2020). Geometric ergodicity of the bouncy particle sampler. *Ann. Appl. Probab.* **30** 2069–2098. MR4149523 https://doi.org/10.1214/19-AAP1552

GIROLAMI, M. and CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 123–214. MR2814492 https://doi.org/10.1111/j.1467-9868.2010.00765.x

JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030 https://doi.org/10.1016/S0304-4149(99)00082-4

JARNER, S. F. and ROBERTS, G. O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Stat.* **34** 781–815. MR2396939 https://doi.org/10.1111/j.1467-9469.2007.00557.x

JOHNSON, L. T. and GEYER, C. J. (2012). Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *Ann. Statist.* **40** 3050–3076. MR3097969 https://doi.org/10.1214/12-AOS1048

KAMATANI, K. (2018). Efficient strategy for the Markov chain Monte Carlo in high-dimension with heavy-tailed target probability distribution. *Bernoulli* **24** 3711–3750. MR3788187 https://doi.org/10.3150/17-BEJ976

LIE, H. C., RUDOLF, D., SPRUNGK, B. and SULLIVAN, T. J. (2023). Dimension-independent Markov chain Monte Carlo on the sphere. *Scand. J. Stat.* **50** 1818–1858. MR4677377 https://doi.org/10.1111/sjos.12653

MANGOUBI, O. and SMITH, A. (2018). Rapid mixing of geodesic walks on manifolds with positive curvature. *Ann. Appl. Probab.* **28** 2501–2543. MR3843835 https://doi.org/10.1214/17-AAP1365

MENGERSEN, K. L. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882 https://doi.org/10.1214/aos/1033066201

MIJATOVIĆ, A., MRAMOR, V. and URIBE BRAVO, G. (2018). Projections of spherical Brownian motion. *Electron. Commun. Probab.* **23** 1–12. MR3852266 https://doi.org/10.1214/18-ECP162

OLLIVIER, Y. (2009). Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256** 810–864. MR2484937 https://doi.org/10.1016/j.jfa.2008.11.001

POMPE, E., HOLMES, C. and ŁATUSZYŃSKI, K. (2020). A framework for adaptive MCMC targeting multimodal distributions. *Ann. Statist.* **48** 2930–2952. MR4152629 https://doi.org/10.1214/19-AOS1916

ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. MR1428751 https://doi.org/10.1214/aoap/1034625254

ROBERTS, G. O. and ROSENTHAL, J. S. (2009). Examples of adaptive MCMC. *J. Comput. Graph. Statist.* **18** 349–367. MR2749836 https://doi.org/10.1198/jcgs.2009.06134

ROBERTS, G. O. and TWEEDIE, R. L. (1996a). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158 https://doi.org/10.1093/biomet/83.1.95

ROBERTS, G. O. and TWEEDIE, R. L. (1996b). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273 https://doi.org/10.2307/3318418

SHERLOCK, C., FEARNHEAD, P. and ROBERTS, G. O. (2010). The random walk Metropolis: Linking theory and practice through a case study. *Statist. Sci.* **25** 172–190. MR2789988 https://doi.org/10.1214/10-STS327

VASDEKIS, G. and ROBERTS, G. O. (2022). A note on the polynomial ergodicity of the one-dimensional zig-zag process. *J. Appl. Probab.* **59** 895–903. MR4480086 https://doi.org/10.1017/jpr.2021.97

VASDEKIS, G. and ROBERTS, G. O. (2023). Speed up zig-zag. *Ann. Appl. Probab.* **33** 4693–4746. MR4674062 https://doi.org/10.1214/23-aap1930

YANG, J., ŁATUSZYŃSKI, K. and ROBERTS, G. O. (2024). Supplement to "Stereographic Markov chain Monte Carlo." https://doi.org/10.1214/24-AOS2426SUPP

YANG, J., ROBERTS, G. O. and ROSENTHAL, J. S. (2020). Optimal scaling of random-walk Metropolis algorithms on general target distributions. *Stochastic Process. Appl.* **130** 6094–6132. MR4140028 https://doi.org/10.1016/j.spa.2020.05.004

YANG, J. and ROSENTHAL, J. S. (2023). Complexity results for MCMC derived from quantitative bounds. *Ann. Appl. Probab.* **33** 1459–1500. MR4564431 https://doi.org/10.1214/22-aap1846

ZAPPA, E., HOLMES-CERFON, M. and GOODMAN, J. (2018). Monte Carlo on manifolds: Sampling densities and integrating functions. *Comm. Pure Appl. Math.* **71** 2609–2647. MR3869037 https://doi.org/10.1002/cpa.21783