Sticky Manifold Piecewise Deterministic Monte Carlo for Bayesian Model Averaging

Hirofumi Shiba [1,2], Luke Hardcastle[3] and Kengo Kamatani [1,2]

[1]Institute of Statistical Mathematics, Tokyo, Japan
[2]Graduate University for Advanced Studies (SOKENDAI), Tokyo, Japan
[3]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

November 16, 2026

Abstract

Assigning positive prior probability to lower-dimensional submodels enriches the range of uncertainty that Bayesian models can represent. The resulting posterior distributions combine continuous parameter uncertainty with discrete model uncertainty. Efficient movement between these submodels, however, can be difficult to design. We propose the sticky manifold construction as a principled, modular framework for combining piecewise deterministic Markov processes (PDMPs) across different submodels. The integration is rejection-free, requires no additional likelihood evaluations, and retains the non-reversible and ballistic nature of the component PDMP dynamics. We prove that the resulting process leaves the intended mixed-dimensional target invariant. We also develop a manifold PDMP algorithm based on coordinate charts to provide the within-submodel dynamics and a blow-up procedure that resolves singular sets into codimension-one sticky boundaries. This geometric construction makes the remaining design choices explicit; component dynamics and unsticking kernels can be chosen freely, subject to easily verifiable velocity compatibility conditions. Finally, we demonstrate the sticky manifold algorithms in Bayesian model averaging examples, propagating uncertainty over pure versus mixed states in quantum state tomography, rank in Bayesian principal component analysis, and parametric families in survival extrapolation.

1 Introduction

Statistical models are commonly defined such that the number of model parameters is unknown. In the Bayesian paradigm, with a prior on all unknown quantities, Markov

MSC2020 subject classifications: 5C05 (Primary); 60J25, 62F15 (Secondary)
Keywords and phrases: Trans-dimensional MCMC, mixed-dimensional distributions, Rejection-free sampling, geometric MCMC, Rao–Blackwellization, Manifold sampling

Chain Monte Carlo (MCMC) samplers are then required to sample across several submodels. This is most commonly achieved using the celebrated reversible jump MCMC algorithm (Green, 1995, 2003). Example settings include finite mixture models (Richardson and Green, 1997), variable selection (Dellaportas et al., 2002), autoregressive models (Vermaak et al., 2004), Bayesian neural networks (Andrieu et al., 2001), and matrix factorisation (Lopes and West, 2004; Frühwirth-Schnatter et al., 2025), with application areas in finance, biology and epidemiology. For a recent comprehensive review see Craiu et al. 2026, Chapter 7.

Efficiency of reversible jump methods primarily depends on the design of efficient across-model jumps, a non-trivial task in general settings (Ehlers and Brooks, 2008). In the case of nested submodels (e.g., finite mixture models and variable selection) a useful perspective is to recast the problem as sampling from nested subspaces. This motivates the widely used split-merge proposals of Richardson and Green (1997), the design of local across-model moves around centring points where likelihoods cancel (Brooks et al., 2003), and the same, gain, and lose proposals of Holmes-Cerfon (2020) that can be constructed without explicit manifold charts. Recent works have combined this perspective with Piecewise Deterministic Markov Processes (PDMPs) in the context of variable selection, designing samplers that traverse a sequence of axis-aligned hyperplanes (Chevallier et al., 2023; Bierkens et al., 2023b) using tractable continuous-time dynamics. In this paper, we extend this approach to design PDMP algorithms that traverse a general sequence of nested submanifolds and models embedded in a common encompassing model, generalising this approach to a wider class of problems beyond the case of variable selection.

Piecewise Deterministic Markov Processes are a class of Markov processes defined by deterministic dynamics interrupted by jumps at random event times (Davis, 1984). When interest is in sampling from a target measure π on the state space $\mathbb{R}^d$, recent works introduce PDMPs defined on the augmented state space $(x_t, v_t) \in \mathbb{R}^{2d}$, such that the marginal stationary distribution of x_t is π. Example processes include the bouncy particle sampler (Bouchard-Côté et al., 2018), the zig-zag sampler (Bierkens et al., 2019), forward event-chain Monte Carlo (Michel et al., 2020), the boomerang sampler (Bierkens et al., 2020), and randomised Hamiltonian Monte Carlo (Bou-Rabee and Sanz-Serna, 2017; Kleppe, 2022). These samplers have attracted considerable attention in the MCMC literature in recent years, as velocity persistence allows for non-reversible, ballistic exploration of the state space, improving sampling efficiency (Bierkens and Roberts, 2017; Andrieu and Livingstone, 2021). Further, they exhibit attractive properties such as robustness to length-scale differences (Bierkens et al., 2025; Power and Vasdekis, 2025; Lu and Luo, 2026) and scalability to large datasets (Bierkens et al., 2019, 2020; Agrawal et al., 2025), in comparison to other gradient-informed MCMC schemes.

Recent works have shown that these processes can be designed to sample from target measures consisting of mixtures of atomic and continuous components (Chevallier et al., 2023; Bierkens et al., 2023b,a). These targets are common in Bayesian variable selection with spike and slab priors where the prior is constructed coordinate-wise as a mixture of a Dirac mass at zero and a continuous density (Mitchell and Beauchamp, 1988). PDMPs with tractable deterministic dynamics can target these measures by deterministically sticking to the spike when the process intersects it, and then unsticking after a random time (Chevallier et al., 2023; Bierkens et al., 2023b). We refer to these dynamics throughout as sticky PDMPs following the terminology of Bierkens et al. (2023b). Under this con-

struction trans-dimensional moves involve no computation of the likelihood and require no additional tuning parameters, unlike most reversible jump MCMC algorithms. Similar considerations have motivated the design of PDMP methods for piecewise continuous densities and state spaces with boundaries (Bierkens et al., 2023a; Chevallier et al., 2024; Bierkens et al., 2018), with several applications of PDMP samplers in the statistical literature focusing on target distributions defined on non-trivial spaces (Koskela, 2022; Hardcastle et al., 2025a,b; van den Bergh and Marsman, 2026; Arnstad et al., 2026).

1.1 Our contributions

In its simplest form, we consider target measures of the form

$$\pi(dx) \propto \exp(-U(x))\Big(\omega\pi_{\mathbb{R}^d}(x)\,dx + (1-\omega)\pi_{\mathcal{M}}(x)\mathcal{H}^{d-1}_{\mathcal{M}}(dx)\Big), \tag{1}$$

where $\pi_{\mathbb{R}^d}$ is a density with respect to the d-dimensional Lebesgue measure (denoted dx) and $\pi_{\mathcal{M}}$ is a density on the embedded Riemannian manifold $\mathcal{M}$ with respect to the normalised Hausdorff measure (denoted $\mathcal{H}^{d-1}_{\mathcal{M}}$).

Our primary contribution is to show that the following process, that we term the sticky manifold piecewise deterministic Markov process (SM-PDMP) and illustrate in Figure 1, is π-invariant:

(1) In the ambient space, the process is given by a standard Euclidean PDMP with $\exp(-U(x))\pi_{\mathbb{R}^d}(x)\,dx$ as the marginal invariant measure for x.

(2) When the ambient process intersects $\mathcal{M}$, it deterministically sticks to the manifold, $\mathcal{M}$.

(3) On $\mathcal{M}$, the process is given by a transformation of a standard PDMP sampler on the embedded manifold, with $\exp(-U(x))\pi_{\mathcal{M}}(x)\,\mathcal{H}^{d-1}_{\mathcal{M}}$ as the marginal invariant measure for x; see Algorithm 1 in Section 2.

(4) The process leaves the manifold with an unsticking rate and kernel that preserves the velocity of the process relative to $\mathcal{M}$ before the sticking event, allowing the process to move from $\mathcal{M}$ to the ambient space. This rate is given by

$$\Lambda(x) = \|v^{\perp}_{\sigma}\|\frac{\omega}{1-\omega}\frac{\pi_{\mathbb{R}^d}(x)}{\pi_{\mathcal{M}}(x)}, \tag{2}$$

where $v^{\perp}_{\sigma}$ is the component of the velocity that is orthogonal to the manifold at the hitting time σ.

The above construction admits the original sticky PDMP algorithm (Bierkens et al., 2023b) as a special case and, in particular, it retains the attractive property that the unsticking rate (2) does not depend on the negative log-likelihood U and can be computed only through priors. Our full algorithm (Algorithm 2) further allows stochastic transformation upon sticking and unsticking and applies beyond the target (1) to the cases with a nested sequence of manifolds, in which the sampler can iteratively stick to the adjacent lower dimensional submanifolds, and cases with multiple manifolds. We also propose the blow-up construction as a key to this generalisation, while also using it to turn a singular subset into sticky boundaries. This geometric treatment also enables the clear separation of the geometric construction of state spaces (Sections 2.3 and 3.2) and invariance proofs

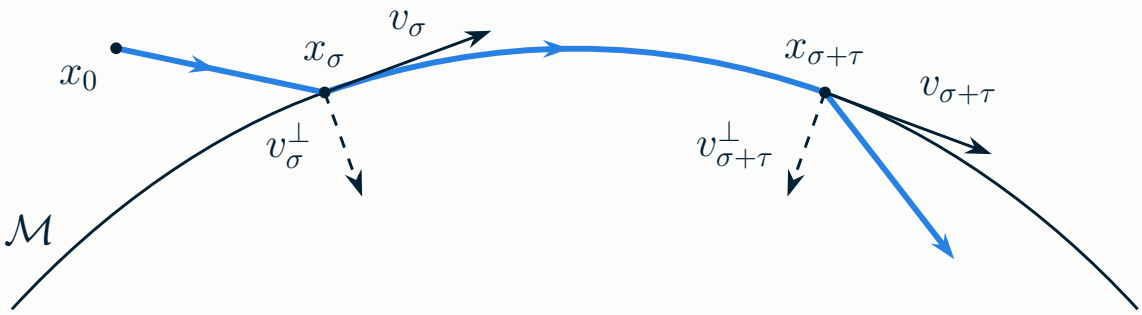


Figure 1: An illustration of the sticky manifold piecewise deterministic Markov process, whose trajectory is in blue. When it intersects $\mathcal{M}$, the dynamics are confined within $\mathcal{M}$, with the normal velocity component frozen and parallel-transported along $\mathcal{M}$, until the first arrival time τ from a Poisson process with rate Λ (2) elapses. See Section 3 for details.

(Sections 2.4 and 3.3) from the remaining algorithmic choices of specific dynamics and unsticking kernels (Sections 2.2 and 3.1). Our invariance results centre on Theorem 3.5, whose proof rests on a flow-wise Gauss–Green formula that provides a common basis for further generalisations. Extending the extended-generator approach of Chevallier et al. (2023, 2024), the formula applies to non-linear flows, test functions that are only absolutely continuous along the flow, and general curved boundaries.

Before introducing our sticky manifold framework in Section 3, we separately treat its subroutine that does manifold sampling within $\mathcal{M}$ in Section 2. This method extends existing PDMPs to target measures defined on a Riemannian manifold $\mathcal{M}$ based on their coordinate charts. When a suitable chart is available, our method provides a largely plug-and-play route to manifold sampling, allowing existing PDMP methodology to be reused rather than redesigned for each manifold.

1.2 A motivating example

Target measures in the form of (1) commonly arise in Bayesian model averaging. To motivate this, we consider the simple example of fitting a Gaussian mixture model with a prior on the number of components,

$$f(y) = \sum_{k=1}^{K} \frac{1}{K} h(y \mid \mu_k, \sigma^2),$$

where $h(y \mid \mu, \sigma^2)$ denotes a Gaussian density function with mean μ and variance σ^2. For simplicity of exposition, we assume that components are equally weighted with shared variance σ^2, and that the prior on K is restricted to $K \in \{1, 2\}$. This then translates to the implied prior on μ

$$\pi(d\mu) = \omega \pi_{\mathbb{R}^2}(\mu)\, d\mu + (1-\omega)\pi_{\mathcal{M}}(\mu)\, \mathcal{H}^1_{\mathcal{M}}(d\mu), \quad \mathcal{M} = \{\mu \in \mathbb{R}^2 \mid \mu_1 = \mu_2\}.$$

When the density $\pi_{\mathcal{M}}$ is the normalised restriction of $\pi_{\mathbb{R}^2}$ on $\mathcal{M}$, the unsticking rate (2) becomes constant

$$\Lambda = \frac{\omega}{1-\omega} \|v_\sigma^\perp\| Z, \qquad Z := \int_{\mathcal{M}} \pi_{\mathbb{R}^2}(x)\, \mathcal{H}^1(dx).$$

Employing the forward event-chain Monte Carlo dynamics (Michel et al., 2020) in both the ambient space and the submanifold, our method produces a single continuous non-reversible trajectory traversing the two components, as shown in Figure 2 (a). For comparison, panel (c) shows the standard reversible-jump MCMC with a random walk proposal, and (b) a simple PDMP analogue where the same trans-dimensional jumps are triggered at a constant Poisson rate $\lambda > 0$ within the PDMP dynamics; see also Hardcastle et al. (2025a). As we demonstrate with a variety of examples in Section 4, the main strength of our construction is that the same principle extends beyond this linear example, regardless of the geometry of the submanifold or the particular choice of within-model PDMP dynamics.

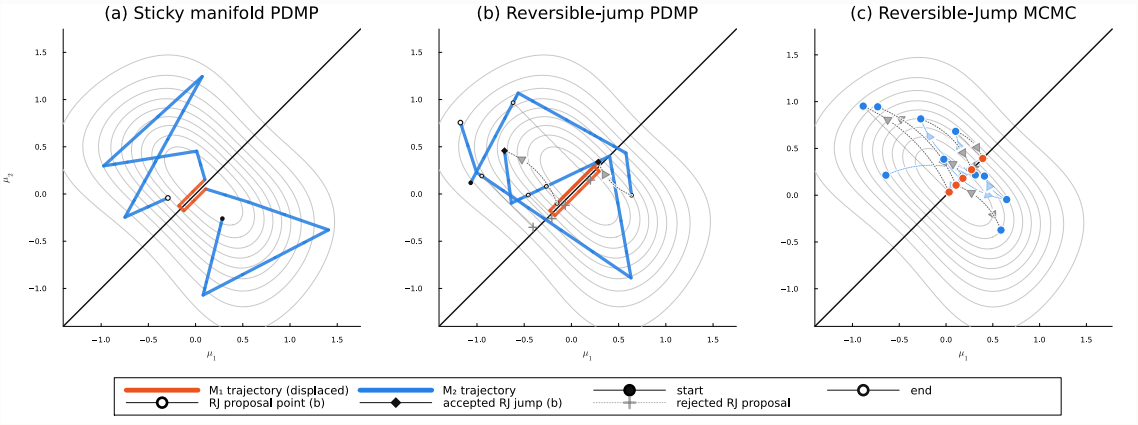


Figure 2: Representative trajectories on the common posterior geometry: (a) sticky manifold forward event-chain Monte Carlo, (b) reversible-jump forward event-chain Monte Carlo in the sense of Hardcastle et al. (2025a), and (c) reversible-jump random walk Metropolis–Hastings. In panel (b), the trajectory crosses the manifold without sticking, whereas in panel (a), it sticks every time it intersects the manifold. The dotted black arrows in panel (b) and (c) indicate the reversible-jump moves, while the dotted blue arrows in panel (c) indicate the random-walk moves. In (b) and (c), there are some rejected reversible-jump proposals depicted by the plus mark (+), while the sticking-and-unsticking mechanism is rejection-free. Manifold trajectories are displaced slightly in the normal direction for visibility.

2 The coordinate-based manifold PDMP

In this Section we introduce a mechanism that adapts standard PDMP samplers with piecewise linear dynamics to sample from target measures defined on Riemannian manifolds. Existing constructions either retain Hamiltonian dynamics, as in the randomised Riemannian manifold HMC (Whalley et al., 2024), or are targeted at hyperspheres, such as the spherical bouncy particle sampler (Yang et al., 2024) and the event-chain Monte Carlo for continuous-spin models (Michel et al., 2015; Nishikawa et al., 2015). In contrast, our approach based on coordinate charts can be applied to any PDMP sampler with tractable deterministic dynamics when the manifold admits a global chart.

2.1 Euclidean PDMP algorithms

We briefly recall the Euclidean PDMP dynamics used as a building block. A piecewise linear sampler uses the Euclidean geodesic flow $\phi_t(x, v) = (x + tv, v)$ between events. The bouncy particle sampler (Bouchard-Côté et al., 2018), the zig-zag sampler (Bierkens et al., 2019), and forward event-chain Monte Carlo (Michel et al., 2020) belong to this class, while the boomerang sampler (Bierkens et al., 2020) and randomised Hamiltonian Monte Carlo (Bou-Rabee and Sanz-Serna, 2017) do not. Stochastic events occur with state-dependent rate $\lambda(x, v)$, and the state is updated according to a transition kernel Q, in most instances only updating the velocity components. Different choices of λ and Q lead to different dynamics and algorithms.

A key computational challenge is simulating the event times $\{\tau\}_{i=1}^{n}$. Theoretically, given τ_{i-1}, the next event time is given by

$$\tau_i := \tau_{i-1} + \arg\min_{s \geq 0} \left\{ \int_0^s \lambda \circ \phi_t(x_{\tau_{i-1}}, v_{\tau_{i-1}})\, dt = E \right\}, \tag{3}$$

for $E \sim \mathrm{Exp}(1)$. When a closed-form solution is unavailable, however, this equation must be solved numerically (Pagani et al., 2024; Chevallier et al., 2026). Alternatively, when there is an upper bound of λ from which event times can be simulated exactly, one obtains exact samples of τ_i by accepting or rejecting the candidate event time. This latter approach is called Poisson thinning (Lewis and Shedler, 1979; Lemaire et al., 2018), and some methods aim to automatically generate the upper bound (Corbella et al., 2022; Sutton and Fearnhead, 2023; Andral and Kamatani, 2024). See Fearnhead et al. 2025, Chapter 5 for an accessible overview.

2.2 The manifold PDMP algorithm

Given a target probability measure $\pi^g(x)\mathcal{H}_g^p(dx)$ on the Riemannian manifold $(\mathcal{M}, g)$ that admits a global chart $\varphi : \mathcal{M} \to \mathbb{R}^p$, we consider the following two-step procedure

(1) Sample from the pushforward density

$$\pi^\varphi(x) := \pi^g(\xi) J_\varphi^g(\xi)^{-1}, \qquad \xi = \varphi^{-1}(x), \tag{4}$$

where J_φ^g is the Jacobian of φ with respect to the Riemannian metric g of $\mathcal{M}$.

(2) Transform the output via φ^{-1} to recover samples on $\mathcal{M}$.

Since the weighted target π^φ is now defined on a Euclidean region $\varphi(\mathcal{M})$, one can use standard PDMP algorithms, such as those of Bierkens et al. (2023a, 2018), to obtain samples $X_t \sim \pi^\varphi$. Proposition 2.2 below then ensures that the transformed samples $\varphi^{-1}(X_t)$ obey the original position law π^g.

Algorithm 1 Manifold PDMP with a global chart φ.

1: Initialise $(\xi_0, \eta_0) \in T(\mathcal{M})$
2: Initialise $X_0 \leftarrow \varphi(\xi_0)$ and $V_0 \leftarrow D\varphi(\xi_0)\eta_0$
3: Run a standard PDMP on $\varphi(\mathcal{M}) \subset \mathbb{R}^p$ targeting π^φ (4)
4: Recover the sample of π^g by $\xi_t \leftarrow \varphi^{-1}(X_t)$

When we use this as a subroutine within the sticky manifold PDMP algorithm (Algorithm 2), the metric g is set to the induced metric from the embedding $\mathcal{M} \subset \mathbb{R}^d$ so that π^g corresponds to $\pi_{\mathcal{M}}$ in (1). The Jacobian $J^g_\varphi(\xi)^{-1} = J^g_{\varphi^{-1}}(\xi)$ accounts for the artificial acceleration caused by the transformation φ^{-1} at $\xi \in \mathcal{M}$. In our numerical examples of survival analysis (Section 4.5), we use Algorithm 1 to sample from four parametric subfamilies, while to sample from the unit sphere $\mathbb{S}^2$ (Sections 4.2 and 4.3) or from the Stiefel manifold $O(20)$ (Section 4.4), we simulate the geodesic flow explicitly.

Remark 2.1 (Extension to general manifolds). In this work, we mainly consider manifolds with a global chart or explicit geodesic flows. We however suggest that a global chart may not be necessary by working with an atlas, $\{\varphi_i\}_{i=1}^n$, of $\mathcal{M}$ and dividing $\mathcal{M}$ into a disjoint sum of U_i's. The above process can then transition between local charts using the transition map $\varphi_j \circ \varphi_i^{-1}$ and its derivative to transform the position and velocity, respectively. More generally, one may include the chart index as a discrete component of the state and introduce jumps between coordinate patches, with their rates and transition kernels chosen to preserve the global target distribution. This leads to a setting structurally similar to the reversible-jump PDMP framework of Chevallier et al. (2023).□

2.3 Geometric foundations: pullback construction of the state space

For the manifold PDMPs considered here, the natural state space is the tangent bundle, generalising the position-velocity product space in the Euclidean case. Accounting for this bundle structure is instrumental in developing a consistent mathematical formulation and subsequently proving invariance. The state spaces of both manifold PDMPs and sticky manifold PDMPs are constructed geometrically as pullback bundles, independently of the specific PDMP dynamics employed. We therefore review this construction before establishing the invariance of the manifold PDMP (Algorithm 1) in Proposition 2.2 and the sticky manifold PDMP (Algorithm 2) in Theorem 3.5. Here, we introduce only the geometric concepts needed for these constructions; readers are referred to Lang (1995) for a thorough treatment.

A p-dimensional *manifold* $\mathcal{M}$ is a space that looks locally like Euclidean space $\mathbb{R}^p$ through prescribed transformations called *local coordinate charts* $\{\varphi_i\}$. We say $\mathcal{M}$ is *smooth* when all transition maps $\varphi_i \circ \varphi_j^{-1}$ are smooth in the usual sense wherever they are defined. In this work, we focus on *embedded* manifolds $\mathcal{M} \subset \mathbb{R}^d$. The difference $d - p$ is called the *codimension*. For any smooth manifold $\mathcal{M}$, such an embedding exists for sufficiently large d.

At a point $\xi \in \mathcal{M}$, the *tangent space* $T_\xi(\mathcal{M})$ consists of all *tangent vectors* η that can be realised by smooth curves through ξ; η belongs to $T_\xi(\mathcal{M})$ if there exists a smooth curve $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ such that $\gamma(0) = \xi$ and $\dot\gamma(0) = \eta$. The space $T_\xi(\mathcal{M})$ forms a p-dimensional vector space, and the collection of all tangent spaces forms the *tangent bundle*, $T(\mathcal{M}) := \sqcup_{\xi \in \mathcal{M}} T_\xi(\mathcal{M}) = \{(\xi, \eta) \mid \eta \in T_\xi(\mathcal{M})\}$. For any bundle E or a subset of it, we denote its *fibre* at x by $E_x := \{v \mid (x, v) \in E\}$. Tangent vectors represent the natural generalisation of velocity in Euclidean space to Riemannian manifolds, and the state spaces for manifold PDMPs and sticky manifold PDMPs will all be formulated as bundles.

In this work, two canonical constructions are repeatedly used: *pullback* and *lifting*. *Pullback* and its dual, *pushforward*, are universal constructions for many mathematical objects. The probability measure π_* satisfying $\pi_*(A) = \pi(\phi^{-1}(A))$ is called the *pushforward*

of π along a mapping $\phi : \mathcal{M} \to \mathcal{N}$. When ϕ is invertible, π is the *pullback* of π_*. A tangent bundle can also be pulled back along ϕ by defining

$$\phi^*(T(\mathcal{N})) := \{(\xi, \eta) \mid \xi \in \mathcal{M}, \eta \in T_{\phi(\xi)}(\mathcal{N})\}.$$

When ϕ is the global chart, $\phi^*(T(\mathcal{N}))$ is the state space of manifold PDMPs. We define a pullback of a PDMP in Section 2.4. A *lift* of a probability measure π on $\mathcal{M}$ to its tangent bundle is any distribution defined on $T(\mathcal{M})$ with position marginal π. When π is the target distribution of a PDMP, the choice of lift corresponds to the choice of velocity distribution. In Section 3.2, we discuss a lifting of manifold PDMP dynamics.

2.4 Pulling back PDMPs through coordinate charts

In this subsection, we consider the pullback of PDMP (X_t, V_t) through a coordinate chart $\varphi : \mathcal{M} \to \mathcal{U} \subset \mathbb{R}^p$, defined by

$$\varphi^* Z_t := (\varphi^{-1}(X_t), D\varphi^{-1}(X_t)V_t), \tag{5}$$

where the position process is pulled back by φ, while the velocity is pulled back by the derivative $D\varphi$. When π is an invariant measure of X_t, this pullback PDMP dynamics leaves the pullback distribution $\pi^* := (\varphi^{-1})_*\pi$ invariant, and π^* is described through the Jacobian. We defer the proof to Appendix A.

Proposition 2.2 (Invariant distribution of pullback PDMPs). Let $(\mathcal{M}, g)$ be a p-dimensional Riemannian manifold, $\mathcal{U} \subset \mathbb{R}^p$ an open set, and $\varphi : \mathcal{M} \to \mathcal{U}$ be a diffeomorphism. Consider a PDMP (X_t, V_t) on $\mathcal{U} \times \mathbb{R}^p$ with an invariant distribution $\pi(x)\,dx\nu_x(dv)$. Then, the pullback PDMP process (5) on $T(\mathcal{M})$ has an invariant distribution

$$\pi(\varphi(\xi))J^g_\varphi(\xi)\mathcal{H}^p_g(d\xi)\nu^\varphi_\xi(d\eta), \qquad (\xi, \eta) \in T(\mathcal{M}) \tag{6}$$

where J^g_φ is the Jacobian of φ computed with the Riemannian metric g, $\mathcal{H}^p_g$ is the Hausdorff measure on $(\mathcal{M}, g)$, and $\nu^\varphi_\xi := (D\varphi(\xi)^{-1})_*\nu_{\varphi(\xi)}$ is the pullback distribution. □

Remark 2.3 (The choice of g). The Jacobian factor in (6) disappears when g is the pullback of the Euclidean metric g_{Euclid} through φ, i.e., $g = \varphi^* g_{\text{Euclid}}$, or, in other words, when φ is isometric; see Section 2.5 for an example. However, the target distribution (1) of sticky manifold PDMPs needs to be specified with respect to the induced metric $g = i^* g_{\text{Euclid}}$ through the embedding $i : \mathcal{M} \to \mathbb{R}^d$, which differs from $\varphi^* g_{\text{Euclid}}$ in general. □

Theorem A.2 shows that the pulled-back PDMP follows the geodesics of the induced metric $\varphi^* g_{\text{Euclid}}$. The choice of parametrisation is therefore crucial and may result in both improvement and worsening of the geometry of the transformed target, as well as the resulting computational complexity of simulating the continuous time process.

2.5 Illustration: two charts of the unit hypersphere

We illustrate how different coordinate charts give rise to different weighted targets, using the one-dimensional unit hypersphere $\mathbb{S}^1$ as an example. It has two natural coordinate maps, $\varphi_1 : \mathbb{S}^1 \setminus \{(1,0)\} \to (0, 2\pi)$ and $\varphi_2 : \mathbb{S}^1 \setminus \{(0,1)\} \to \mathbb{R}$, given by

$$\varphi_1(x, y) = \begin{cases} \arccos x & y \geq 0, \\ 2\pi - \arccos x & y < 0. \end{cases} \qquad \varphi_2(x, y) = \frac{x}{1 - y},$$

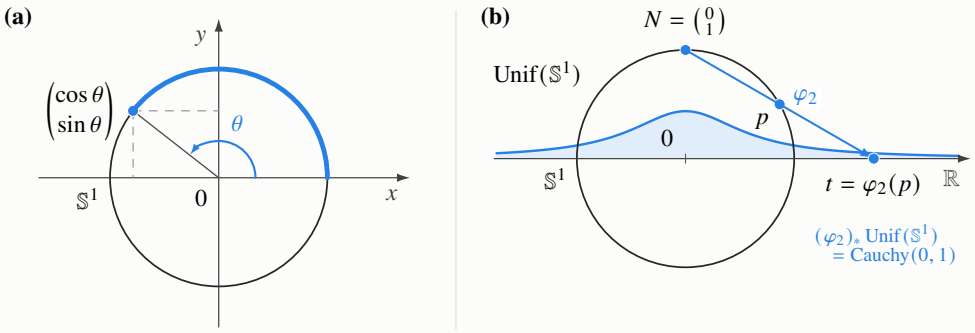


Figure 3: Two parametrisations for the one-dimensional unit hypersphere. (a) represents the arc length parametrisation φ_1. In this case, the Jacobian is constant, $J^g_{\varphi_1} = 1$ and φ_1 is a Riemannian isometry. (b) represents the stereographic projection, pushing-forward the uniform distribution to the standard Cauchy distribution.

respectively (Figure 3). The Jacobians with respect to the metric g on $\mathbb{S}^1$ induced by the ambient Euclidean metric are

$$J^g_{\varphi_1^{-1}}(x) = 1, \qquad J^g_{\varphi_2^{-1}}(x) = \frac{2}{1+x^2},$$

respectively. Thus, to sample from the uniform measure on $\mathbb{S}^1$, i.e., $\pi^*(x) = (2\pi)^{-1}$, the required weighted proxy targets are

$$\pi_1(x) = \frac{1}{2\pi} \quad (x \in (0, 2\pi)), \qquad \pi_2(x) = \frac{1}{\pi}\frac{1}{1+x^2} \quad (x \in \mathbb{R}),$$

calculated via Equation (4) for φ_1 and φ_2 respectively. The simplicity of π_1 derives from the constant Jacobian 1, that means φ_1 is a Riemannian isometry. In this case, the weight is absorbed by the reference measure $\mathcal{H}^p_g$ in Equation (6). By contrast, the weighted target π_2 is nontrivial, being the standard Cauchy distribution. This correspondence is utilised to efficiently sample from heavy-tailed distributions on $\mathbb{R}^d$; see Yang et al. (2024); Bell et al. (2025).

Remark 2.4 (Higher-dimensional unit spheres). We consider sampling on the two-dimensional unit sphere in Sections 4.2 and 4.3. This sphere $\mathbb{S}^2$ or any higher-dimensional sphere differs from $\mathbb{S}^1$ in that no isometric coordinate chart φ is available due to the non-trivial sectional curvature. Thus, isometric charts like φ_1 are not available in higher dimensions, while the stereographic projection φ_2 can be naturally generalised. We can, however, simulate the geodesic flow directly as constrained dynamics through the canonical embedding $\mathbb{S}^2 \to \mathbb{R}^3$ (e.g., Yang et al., 2024),

$$\begin{cases} x_t = \cos(\|v_0\|t)x_0 + \sin(\|v_0\|t)v_0/\|v_0\| \\ v_t = -\|v_0\|\sin(\|v_0\|t)x_0 + \cos(\|v_0\|t)v_0. \end{cases}$$

□

3 The sticky manifold PDMP

We now introduce the sticky manifold PDMP sampler. Throughout this Section we focus on a single embedded Riemannian manifold $\mathcal{M} \subset \mathbb{R}^d$ of codimension one that is closed in $\mathbb{R}^d$. We also assume that $\mathcal{M}$ is *orientable*: a global unit normal vector field $n : \mathcal{M} \to \mathbb{R}^d$ exists and can be evaluated at each point $x \in \mathcal{M}$. We define the orthogonal decomposition of the velocity as

$$v^\perp := \langle v, n(x)\rangle n(x), \qquad v^{\nparallel} := v - v^\perp,$$

at any point $x \in \mathcal{M}$. We do not impose strong assumptions on the choice of PDMP sampler for sampling in the ambient space or on the submanifold, allowing any standard PDMP sampler mentioned in Section 1 and manifold PDMP sampler in Section 2 to be incorporated as subroutines.

3.1 Algorithmic description

We present the algorithm in a slightly more general form than the one outlined in the Introduction (Section 1.1), allowing stochastic velocity transformations upon sticking and unsticking. These transformations are specified by the kernels Q_{stick} and Q_{unstick}, respectively. The construction in the Introduction corresponds to taking these kernels to be Delta measures associated with deterministic velocity transformations. More generally, we allow the sticking kernel Q_{stick} to transform the tangential velocity component to meet the requirements of the chosen manifold dynamics while leaving the normal component unchanged. During the manifold dynamics, this normal component is parallel-transported along $\mathcal{M}$. The unsticking kernel Q_{unstick} uses the current tangential velocity and the stored normal component to generate the velocity with which the ambient dynamics resume. The same unsticking rate (2) continues to be valid when both kernels preserve the norm of the normal component and satisfy the velocity compatibility condition (8). This condition requires Q_{stick} to map the incoming ambient velocity distribution to the prescribed joint distribution of the manifold velocity and the stored normal component, and Q_{unstick} vice versa. Given a PDMP sampler with position marginal $e^{-U(x)}\pi_{\mathbb{R}^d}(x)\,dx$ and a manifold PDMP sampler with $e^{-U(x)}\pi_{\mathcal{M}}(x)\mathcal{H}^{d-1}_{\mathcal{M}}(dx)$, the sticky manifold PDMP algorithm can be summarised as follows:

Algorithm 2 Sticky manifold PDMP.

1: Initialise $z_0 = (x_0, v_0) \in \mathbb{R}^d \times \mathcal{V}$
2: **while** $t < T$ **do**
3: Run a standard PDMP until $\sigma = \inf\{s > t \mid x_s \in \mathcal{M}\}$
4: Sample a new velocity $v_\sigma \sim Q_{\text{stick}}((x_{\sigma-}, v_{\sigma-}), -)$
5: Start a manifold PDMP in $\mathcal{M}$ from $(x_\sigma, v_\sigma^{\nparallel})$ until τ, the first Poisson arrival with rate Λ
6: Parallel-transport the normal component by setting $v^\perp_{\tau-} \leftarrow \langle v_\sigma, n(x_\sigma)\rangle n(x_{\tau-})$
7: Recover the ambient velocity by sampling $v_\tau \sim Q_{\text{unstick}}((x_{\tau-}, v_{\tau-}), -)$
8: **end while**

In the remainder of this subsection, we list examples of kernels $Q_{\text{stick}}, Q_{\text{unstick}}$ in a few important situations.

Example 3.1 (SM-PDMP with the identity sticking and side-flipping unsticking kernels). When both ambient and manifold velocities are standard Gaussian, we can choose the sticking kernel to be the identity kernel $Q_{\text{stick}}((x,v),-) = \delta_{(x,v)}$ and the unsticking kernel to be a side-flipping kernel $Q_{\text{unstick}}((x^{\pm},v),-) = \delta_{(x^{\mp},v)}$, where x^+ and x^- represent the same point of $\mathcal{M}$ on the two boundary faces. The side flip changes only the boundary-face label, leaving the physical position and velocity unchanged. Relative to the new face, the retained normal velocity points outwards, allowing the ambient dynamics to resume in the blown-up state space (Section 3.2). This choice reduces Algorithm 2 to the one outlined in Section 1.1. These kernels satisfy condition (8), since the ambient standard Gaussian velocity decomposes into independent standard Gaussian tangential and normal components, and side flipping preserves the ambient velocity distribution. □

Example 3.2 (SM-PDMP with unit velocities). When the invariant velocity distributions are the uniform measures on $\mathcal{V} = \mathbb{S}^{d-1}$ and $\mathbb{S}^{d-2}$, a modest transformation satisfying (8) is a deterministic sticking map $\psi_x : \mathbb{S}^{d-1} \to \mathbb{S}^{d-2} \times (-1,1)$ defined by

$$\psi_x(v) = \left(\frac{\not{v}}{\|\not{v}\|}, \langle n(x), v \rangle \right),$$

where $\not{v}$ depends on the position x. Upon unsticking, the inverse transformation is applied:

$$\psi_x^{-1}(\not{u}, u^{\perp}) = \sqrt{1-(u^{\perp})^2}\, \not{u} + u^{\perp} n(x).$$

This sticking transformation ψ_x pushes forward $\text{Unif}(\mathbb{S}^{d-1})$ to the direct product of $\text{Unif}(\mathbb{S}^{d-2})$ and the distribution of $2B-1$ for $B \sim \text{Beta}((d-1)/2, (d-1)/2)$. This verifies condition (9), and hence the velocity compatibility condition (8). Alternatively, we may resample the normal component from a flux-weighted distribution upon unsticking. This approach is used with the tangential part of ψ_x in Chevallier et al. (2023, Section 4.2), who consider coordinate-aligned linear subspaces. When combined with the Rao–Blackwellised unsticking rate (Corollary 3.8), this approach avoids storing the normal component but discards the memory of the incoming normal speed, potentially reducing persistence across model transitions. □

Example 3.3 (SM-Zigzag with velocity refreshment). For the zig-zag sampler, velocity refreshment upon sticking becomes crucial, even for non-linear submanifolds and tilted lines (Sections 1.2 and C.2). Since the tangent spaces of these submanifolds are no longer aligned with the coordinate axes, the new velocity needs to be transformed upon sticking, unlike the coordinate-aligned case of Chevallier et al. (2023, Section 4.1). A straightforward way to ensure velocity compatibility (8) is to completely refresh the tangential velocity by sampling a new manifold velocity from $\text{Unif}(\{\pm 1\})$ upon sticking, while resuming the ambient dynamics upon unsticking with the stored normal velocity. This specification is used in the numerical examples of Section C.2. □

More generally, partial stochastic refreshments can be incorporated as long as the more general flux balance condition (10) is met, being related to the discussion in Bierkens et al. (2023a, Section 2.7). Previous works suggest that this may lead to faster mixing when designed effectively (Michel et al., 2020; Wu and Robert, 2020; Tran and Kleppe, 2025). Note that Gaussian velocity distributions do not guarantee that the velocity of the manifold PDMP induced by its coordinate chart is an isotropic Gaussian distribution (see

Section 2.5). We discuss the velocity compatibility condition under such a setting further in Section 3.5.

3.2 Blow-up construction of the state space

In this technical section we mathematically define a non-reversible Markov process that connects the PDMP dynamics in the ambient space with the dynamics on $\mathcal{M}$. We establish Π-invariance of the resulting process in Section 3.3. We construct the state space, E, as a pullback along the blow-down map, followed by a lifting of manifold dynamics. In E, states on the manifold $\mathcal{M}$ correspond to *frozen boundaries* $\mathfrak{F}$, whose state is represented by $(x, \vec{v}, v^\perp)$ with $(x, \vec{v}) \in T(\mathcal{M})$ and $v^\perp$ perpendicular to $\mathcal{M}$. Importantly, the state of the manifold PDMP $(x, \vec{v}) \in T(\mathcal{M})$ is lifted by adding the ambient normal velocity $v^\perp$, that is parallel-transported along the submanifold while the process develops within $\mathfrak{F}$. This construction enables us to: *i*) Distinguish the two sides of the manifold, denoted by $\mathcal{M}^+$ and $\mathcal{M}^-$, thereby deciding from which side the process stuck to $\mathcal{M}$ and from which side it should leave; *ii*) Provide the normal velocity when the process restarts in the ambient space. This augmented state $(x, \vec{v}, v^\perp) \in \mathfrak{F}$ provides the complete information needed to define a Markov process representing the sticky manifold PDMP algorithm.

The radial blow-up of $\mathcal{M}$ in $\mathbb{R}^d$ gives an enlarged space $[\mathbb{R}^d; \mathcal{M}]$ equipped with a smooth surjection $\beta : [\mathbb{R}^d; \mathcal{M}] \to \mathbb{R}^d$, called the *blow-down* map. As illustrated in Figure 4, the blown-up space $[\mathbb{R}^d; \mathcal{M}]$ has a boundary that consists of two copies of the original manifold $\beta^{-1}(\mathcal{M}) = \mathcal{M}^+ \sqcup \mathcal{M}^-$. Apart from this boundary, the interior is identical to $\mathbb{R}^d \setminus \mathcal{M}$, and the blow-down map β is an identity map on it. Note that this construction generalises the approach in Bierkens et al. (2023b) for the coordinate-aligned linear space $\{y = 0\}$, where two copies of the origin 0 are introduced as 0^- and 0^+. The precise geometric construction of $[\mathbb{R}^d; \mathcal{M}]$ and β, and the unit vector field ∂_r pointing outwards follows the standard radial real blow-up procedure used in differential geometry; see Appendix B.1 for full details. It replaces the submanifold with its spherical normal bundle, thereby retaining the direction of approach. The state space of the sticky manifold PDMP is then defined as the pullback bundle $E := \beta^*(T(\mathbb{R}^d))$ along the blow-down map β. The frozen boundary is defined by the misalignment of the velocity v with the vector field ∂_r

$$\mathfrak{F} := \{(\xi, v) \in E \mid \xi \in \mathcal{M}^\pm, \langle v, \partial_r(\xi)\rangle < 0\}.$$

In words, a state is confined in $\mathcal{M}^\pm$ when the normal velocity points inwards, the opposite direction to ∂_r. Here $\mathfrak{F}$ represents the stuck states over $\mathcal{M}$, retaining both the tangential velocity $\vec{v}$ and the ambient normal velocity $v^\perp$. For a point $x = \beta(\xi) \in \mathcal{M}$ in the manifold, we write $\xi = x^+$ or $\xi = x^-$ depending on which face the position belongs to in the blown-up space. We similarly define the ambient boundary as the subsets of $\Gamma := \partial T(\mathbb{R}^d \setminus \mathcal{M})$ comprising states that satisfy $\langle v, \partial_r(x)\rangle < 0$ and $\langle v, \partial_r(x)\rangle > 0$ by Γ_{in} and Γ_{out}, respectively. Although Γ_{in} coincides with $\mathfrak{F}$ as a set, we distinguish the two for clarity. The sticking kernel Q_{stick} is defined from Γ^{in} to $\mathfrak{F}$ and Q_{unstick} from $\mathfrak{F}$ to Γ^{out}.

Lastly, we need to lift the manifold dynamics on $T(\mathcal{M})$ to the frozen boundary $\mathfrak{F}$. We choose the *horizontal lift* for this purpose: for a state $(\xi, \vec{v}, v^\perp) \in \mathfrak{F}$, the normal velocity $v^\perp$ is parallel-transported in the normal bundle along $\mathcal{M}$, while $(\xi, \vec{v})$ evolves according to the manifold dynamics. Given two PDMP generators $\mathcal{L}_{\mathbb{R}^d}$ and $\mathcal{L}_{\mathcal{M}}$ in $\mathbb{R}^d$ and $\mathcal{M}$ respectively, and denoting the horizontal lift by $\mathcal{L}^{\text{lift}}_{\mathcal{M}}$, the generator of the sticky manifold PDMP

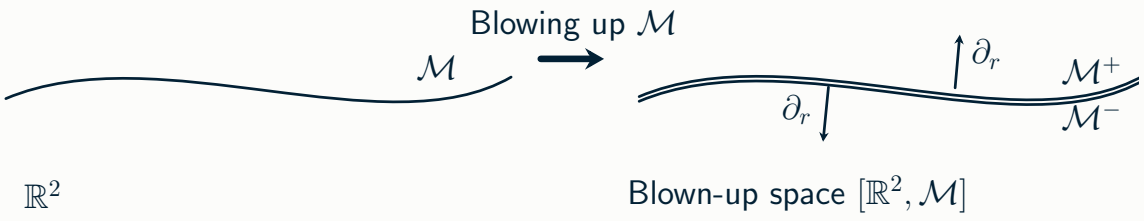


Figure 4: Radial real blow-up of a smooth embedded submanifold $\mathcal{M} \subset \mathbb{R}^2$. The blow-up replaces $\mathcal{M}$ by two boundary faces, $\mathcal{M}^+$ and $\mathcal{M}^-$, representing the two normal directions. There exists a natural vector field ∂_r induced from our construction, representing the outward normal direction on each face.

on the state space E is then defined by

$$\mathcal{L}f = \begin{cases} \mathcal{L}_{\mathbb{R}^d} f, & \text{on } E \setminus \mathfrak{F}, \\ \mathcal{L}^{\text{lift}}_{\mathcal{M}} f + \Lambda(Q_{\text{unstick}} f - f), & \text{on } \mathfrak{F}. \end{cases} \tag{7}$$

The state-dependent unsticking rate Λ determines when the process leaves $\mathcal{M}$ via the kernel Q_{unstick}. Note that there is no jump term in (7) that corresponds to the sticking move; the sticking move is encoded in the boundary condition $Q_{\text{stick}} f = \lim_{t \searrow 0} f \circ \phi_{-t}$, where ϕ_t is the ambient flow, satisfied by the test functions f in the generator domain $\mathcal{D}(\mathcal{L})$; see Davis (1984, Theorem 5.5) and Davis (1993, Theorem 26.14). See Appendix B.2 for mathematical details and the well-definedness result.

3.3 Invariance

Lifting the position target π in (1) to the state space E constructed in Section 3.2, we establish the invariance of the possibly unnormalised finite measure $\Pi = \omega \Pi_{E \setminus \mathfrak{F}} + (1 - \omega)\Pi_{\mathfrak{F}}$, where

$$\Pi_{E \setminus \mathfrak{F}}(dxdv) = \pi_{\mathbb{R}^d}(x)\, dx\, \nu_{E \setminus \mathfrak{F}, x}(dv), \quad (x, v) \in E \setminus \mathfrak{F},$$
$$\Pi_{\mathfrak{F}}(dxdv) = \pi_{\mathcal{M}}(x) \mathcal{H}^{d-1}_{\mathcal{M}}(dx) \nu_{\mathfrak{F}, x}(dv), \quad (x, v) \in \mathfrak{F}.$$

Here, $\mathcal{H}^{d-1}_{\mathcal{M}}$ is the Hausdorff measure induced from the ambient metric or, equivalently, the Riemannian volume measure, and $\nu_{E \setminus \mathfrak{F}}$ and $\nu_{\mathfrak{F}}$ are probability kernels representing conditional velocity distributions. Throughout this subsection, we absorb the common factor e^{-U} in (1) into the component densities and, with a slight abuse of notation, continue to denote the resulting possibly unnormalised densities by $\pi_{\mathbb{R}^d}$ and $\pi_{\mathcal{M}}$ as the result only depends on the ratio between them. All proofs are deferred to Appendix B.4.

Assumption 3.4 . We assume that the ambient PDMP and the lifted manifold PDMP leave $\Pi_{E \setminus \mathfrak{F}}$ and $\Pi_{\mathfrak{F}}$ invariant, respectively, and that the embedded Riemannian manifold $\mathcal{M} \subset \mathbb{R}^d$ is orientable, $d - 1$ dimensional, and closed in $\mathbb{R}^d$. Additionally, we work under the technical conditions ensuring well-definedness, regularity of $\Pi_{E \setminus \mathfrak{F}}$ and $\Pi_{\mathfrak{F}}$, and integrability of the flow and the overall jump intensity, outlined in Assumption B.2. □

Define the restriction of the ambient velocity distribution on the inward and outward boundaries $\Gamma_{\text{in}}, \Gamma_{\text{out}}$ by

$$\nu_x^{\text{in/out}}(B) := \sum_{\xi \in \beta^{-1}(x)} \nu_{E \setminus \mathfrak{F}, \xi}(B_\xi),$$

respectively, for a Borel measurable set B in $\Gamma_{\text{in/out},x} := \{(\xi, v) \in \Gamma_{\text{in/out}} \mid \beta(\xi) = x\}$. We use the standard notation $\mu Q(d\zeta) := \int Q(z, d\zeta)\mu(dz)$ for a measure μ and a probability kernel Q.

Theorem 3.5 (Invariance of sticky manifold PDMP). Under Assumption 3.4, suppose that the two probability kernels $Q_{\text{stick}}, Q_{\text{unstick}}$ preserve the position $\beta(x)$ and the norm of the normal velocity component $\|v^\perp\|$. Additionally, assume that

$$\nu_x^{\text{in}} Q_{\text{stick},x} = \nu_{\mathfrak{F},x}, \qquad \nu_{\mathfrak{F},x} Q_{\text{unstick},x} = \nu_x^{\text{out}}, \tag{8}$$

for $\mathcal{H}^{d-1}_{\mathcal{M}}$-almost all $x \in \mathcal{M}$. Then, the sticky manifold PDMP with unsticking rate (2) leaves Π invariant. □

Remark 3.6 (Velocity pushforward condition for deterministic kernels). Suppose further that the kernels are deterministic and given by

$$Q_{\text{stick},x}(v, -) = \delta_{\psi_x(v)}, \qquad Q_{\text{unstick},x}(\eta, -) = \delta_{R_x \circ \psi_x^{-1}(\eta)},$$

where $\psi_x : \Gamma_{\text{in},x} \to \mathfrak{F}_x$ is a smooth diffeomorphism and $R_x : \Gamma_{\text{in},x} \to \Gamma_{\text{out},x}$ is a side-flipping map that takes $(y, v) \in \Gamma_{\text{in},x}$ to the state $(y', v) \in \Gamma_{\text{out},x}$ with the same velocity v but opposite position of the blown-up front face $y' \in \beta^{-1}(x) \setminus \{y\}$. Condition (8) then reads

$$(\psi_x)_* \nu_x^{\text{in}} = \nu_{\mathfrak{F},x}, \qquad (R_x \circ \psi_x^{-1})_* \nu_{\mathfrak{F},x} = \nu_x^{\text{out}}. \tag{9}$$

When $(R_x)_* \nu_x^{\text{in}} = \nu_x^{\text{out}}$, these conditions reduce to $(\psi_x)_* \nu_x^{\text{in}} = \nu_{\mathfrak{F},x}$. □

Remark 3.7 (General rate function beyond the velocity compatibility condition). The proof of Theorem 3.5 reveals a more general picture. In full generality, the unsticking rate involves velocity distributions as well, while the velocity compatibility (8) is a sufficient condition for the velocity factor to disappear from the rate function. Define the boundary measures

$$J_{\text{in/out}}(dz) := 1_{\Gamma_{\text{in/out}}}(z) \|v^\perp\| \pi_{\mathbb{R}^d}(x) \mathcal{H}^{d-1}_{\mathcal{M}}(dx) \nu_{E \setminus \mathfrak{F}}(x, dv).$$

$J_{\text{in/out}}$ contains additional factors of the normal velocity norm $\|v^\perp\|$, ambient position density $\pi_{\mathbb{R}^d}$, and volume measure $\mathcal{H}^{d-1}_{\mathcal{M}}$. For $J_{\text{in/out}}$, we assume the following balance condition:

$$J_{\text{in}} Q_{\text{stick}} Q_{\text{unstick}} = J_{\text{out}}, \qquad \omega J_{\text{in}} Q_{\text{stick}} = (1 - \omega) \Lambda \Pi_{\mathfrak{F}}. \tag{10}$$

This is a generalisation of (8). Indeed, the assumptions of Theorem 3.5 imply

$$(\|v^\perp\| \nu_x^{\text{in}}) Q_{\text{stick},x} = \|v^\perp\| \nu_{\mathfrak{F},x}, \qquad (\|v^\perp\| \nu_{\mathfrak{F},x}) Q_{\text{unstick},x} = \|v^\perp\| \nu_x^{\text{out}}.$$

This can be interpreted as a fibre-wise balance condition that, in turn, implies (10) after applying the coarea formula. Under (10) and absolute continuity $J_{\text{in}} Q_{\text{stick}} \ll \Pi_{\mathfrak{F}}$, the rate given by the Radon–Nikodym derivative

$$\Lambda(\zeta) = \frac{\omega}{1 - \omega} \frac{d(J_{\text{in}} Q_{\text{stick}})}{d\Pi_{\mathfrak{F}}}(\zeta), \qquad \Pi_{\mathfrak{F}}\text{-a.e.} \tag{11}$$

makes the sticky manifold PDMP Π-invariant. □

3.4 Rao–Blackwellisation and extensions

We establish a two-stage Rao–Blackwellisation within our sticky manifold framework that preserves the target position marginal. First, we average over the retained normal velocity in the rate, obtaining a rate analogous to that in reversible-jump BPS (Chevallier et al., 2023, Proposition 2). When this averaged rate is constant, we further replace exponential sticking times by their expectations, extending the construction of Chin and Nishimura (2025) for the sticky zig-zag sampler. Together, these operations remove both sources of randomness in the sticking time. Our numerical experiments demonstrate improvements in estimation accuracy and qualitatively different behaviours in the model-index autocorrelation function (Section 4).

Corollary 3.8 (Rao–Blackwellized unsticking rate). Assume the conditions of Theorem 3.5 with the standard Gaussian ambient velocity, and that the rate (2) can be written as $\Lambda = c\|v^\perp\|$ for a measurable function $c(x) > 0$. Then, replacing this rate with the velocity-averaged one $\Lambda_{\mathrm{RB}} = c\sqrt{2/\pi}$ still leaves the position marginal of Π invariant, but now with reweighted velocity distribution $\nu^{\mathrm{RB}}_{\mathfrak{F},x}(dv) = \sqrt{\pi/2}\|v^\perp\|\nu_{\mathfrak{F},x}(dv)$. □

Corollary 3.9 (Rao–Blackwellized sticking time). Under the assumptions of Corollary 3.8, suppose further that c is constant. Replace the exponential unsticking time with the deterministic residence time $T = \Lambda_{\mathrm{RB}}^{-1} = c^{-1}\sqrt{\pi/2}$. The resulting process has an invariant distribution whose position marginal distribution is identical to that of the original case. □

Later, we also consider the case where the ambient space is a region $U \subset \mathbb{R}^d$ and $\mathcal{M} = \partial U$ is the boundary of U; see examples in Section 4.3 and 4.4. In this case, the unsticking rate has to be altered accordingly.

Corollary 3.10 (One-sided case). Under Assumption 3.4, assume that the ambient dynamics are confined within an open region $U \subset \mathbb{R}^d$ with boundary $\mathcal{M} = \partial U$, and that the ambient velocity distribution is everywhere spherically symmetric. Further suppose that Q_{stick} and Q_{unstick} preserve the position x together with the norm of the normal velocity component $\|v^\perp\|$ and satisfy the compatibility condition

$$\nu_x^{\mathrm{in}} Q_{\mathrm{stick},x} = \frac{1}{2}\nu_{\mathfrak{F},x}, \qquad \frac{1}{2}\nu_{\mathfrak{F},x} Q_{\mathrm{unstick},x} = \nu_x^{\mathrm{out}}.$$

Then, half of the rate (2)

$$\Lambda(x) = \frac{\|v_\sigma^\perp\|}{2}\frac{\omega}{1-\omega}\frac{\pi_{\mathbb{R}^d}(x)}{\pi_{\mathcal{M}}(x)}$$

ensures the correct position marginal of the invariant distribution. □

Remark 3.11 (Nested strata). The preceding construction extends directly to a finite nested sequence of submanifolds $\mathcal{M}_0 \subset \mathcal{M}_1 \subset \cdots \subset \mathcal{M}_n$ by a repeated application of the Gauss–Green argument (Lemma B.5), provided that each adjacent inclusion is of codimension one, transitions occur only between adjacent strata, and the corresponding assumptions and balance conditions hold at every interface. Note that a minor modification arises in the mixture weights, with the pair $\omega, 1-\omega$ replaced by a sequence of weights $\omega_0, \omega_1, \cdots, \omega_n$ that sum to one. □

Remark 3.12 (Submanifolds overlapping over a null set). In the survival extrapolation example (Section 4.5), we consider three codimension-one manifolds $\mathcal{M}_{\mathrm{W}}, \mathcal{M}_{\Gamma}, \mathcal{M}_{\mathrm{A}}$ that

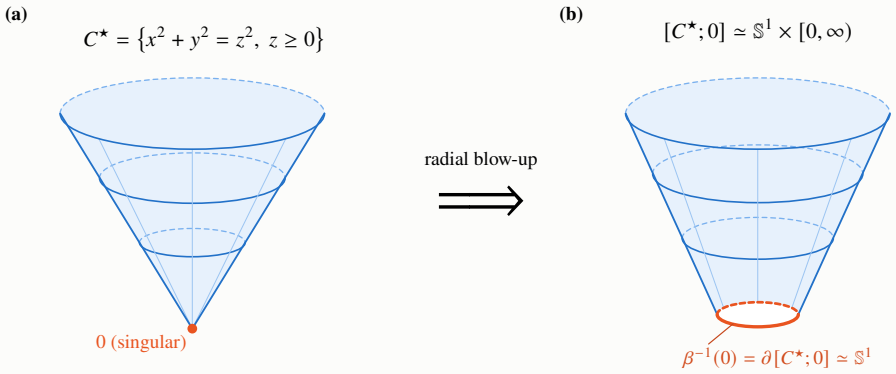


Figure 5: Blow-up of a half cone, or equivalently the space of all symmetric semi-definite rank-one matrices. (a) Singular cone $C^\star$ has its apex as a point of singularity. (b) Represents the blown-up manifold with its front face given by $\beta^{-1}(0) \simeq \mathbb{S}^1$.

intersect at $\{\lambda = \sigma = 1\}$. Due to this overlap, the union $\mathcal{M}_\mathrm{W} \cup \mathcal{M}_\Gamma \cup \mathcal{M}_\mathrm{A}$ is not an embedded manifold. We nevertheless regard the three submodels as distinct labelled components within the disjoint union $\widetilde{M} := \cup_{i \in \{\mathrm{W},\Gamma,\mathrm{A}\}} \{i\} \times \mathcal{M}_i$; a trajectory evolving in one of the components is not allowed to transition directly to the other components. Although $\widetilde{M}$ is no longer an embedded submanifold, our theorem also extends to this case by applying the invariance argument component-wise, given that the ambient trajectory hits two or more hypersurfaces simultaneously with probability zero. □

3.5 Illustration: two charts of the quadratic cone and blow-up

The punctured cone, defined by

$$C = \left\{(x, y, z)^\top \in \mathbb{R}^3 \,\middle|\, x^2 + y^2 = z^2, z > 0\right\}, \tag{12}$$

is perhaps the simplest example that naturally produces a singularity. For C, we consider two coordinate charts: the projection coordinate chart and the polar coordinate chart. The *projection coordinate* $\varphi : C \to \mathbb{R}^2 \setminus \{0\}$, defined by $\varphi(x, y, z) = (x, y)$, has constant Jacobian, $J^g_\varphi = 2^{-1/2}$, and as such does not affect the weighted target (4). This simplifies the algorithm when the sole aim is sampling on C. A different perspective arises, however, when the aim is to also sample from the apex 0 of C.

Example 3.13 (Cone apex as the null matrix). In Section 4.4, we deal with the space of the $d \times d$ positive semi-definite symmetric matrices $\mathrm{Sym}_d(\mathbb{R})_+$. When $d = 2$, it is a solid cone with three strata $\mathcal{M}_i = \{\mathrm{rank}(G) = i\}$ of rank $i = 0, 1, 2$ matrices. In this context, C corresponds to the stratum of rank 1 positive symmetric matrices $\mathcal{M}_1$ and its apex to the null matrix singleton $\mathcal{M}_0 = \{0\}$ through the isometry $\Phi : \mathbb{R}^3 \to \mathrm{Sym}_2(\mathbb{R})$ given by

$$\Phi(x, y, z) = \frac{1}{\sqrt{2}} \begin{pmatrix} z + x & y \\ y & z - x \end{pmatrix}.$$

Since the two strata $\mathcal{M}_1, \mathcal{M}_0$ differ in dimension by two, the sampler defined by the projection coordinate chart is unable to sample from the apex. In contrast, the polar coordinate allows the apex to be hit with positive probability. □

Polar coordinates (r, θ) on C are defined by the inverse of a parametrisation β

$$\beta : (0, \infty) \times \mathbb{S}^1 \ni (r, \theta) \mapsto (r \cos\theta, r \sin\theta, r) \in C.$$

Note that the cone with apex $C^\star := C \cup \{0\} = \{x^2 + y^2 = z^2, z \geq 0\}$ is not a smooth manifold, although its punctured version C is a two-dimensional smooth manifold on which $\beta^{-1} : C \to (0, \infty) \times \mathbb{S}^1$ is defined and smooth. Blowing up $C^\star$ results in a manifold with boundary $[C^\star; 0] := [0, \infty) \times \mathbb{S}^1$, on which the polar parametrisation β naturally extends. This extension $\overline{\beta} : [C^\star; 0] \to C^\star$ is the blow-down map (Figure 5). Importantly, the point $0 \in C^\star$ is now replaced with a codimension one front face $\{0\} \times \mathbb{S}^1 = \partial[C^\star; 0]$. This allows a manifold PDMP sampler on $[C^\star; 0]$ with the polar coordinate (r, θ) to stick to this boundary with a non-zero hitting probability.

Remark 3.14 (Jacobian of the blow-up coordinates). The Jacobian of the chart is $J^g_\beta(r, \theta) = \sqrt{2}r$ for the induced metric g of C; the weighted target (4) is thus $\pi(r, \theta) = \pi^*(\beta(r, \theta)) \cdot \sqrt{2}r$. Therefore, for naïve PDMP dynamics on the blown-up space to reach $r = 0$ with positive probability, π^* must diverge sufficiently rapidly as $r \searrow 0$ to compensate for the vanishing Jacobian. □

Example 3.15 (Higher-dimensional rank cones). We can generalise the above construction to higher dimensions using lifting. First, the polar coordinate chart generalises to a spectral parametrisation over $\mathrm{Sym}_d(\mathbb{R})_+$. Write

$$\mathcal{M}_k := \{G \in \mathrm{Sym}_d(\mathbb{R})_+ \mid \mathrm{rank}(G) = k\}, \qquad k = 0, \cdots, d.$$

The dimension of $\mathcal{M}_k$ is $k(2d - k + 1)/2$, so that the dimensions of successive rank strata differ by $d - k + 1$. We define the spectral lift by

$$\widetilde{\mathcal{M}}_k = C_k \times \mathrm{St}(d, d), \qquad \beta_k(r, Q) = Q\,\mathrm{diag}(r_1, \dots, r_k, 0, \dots, 0)Q^\top, \qquad C_k = \{r_1 > \cdots > r_k > 0\},$$

with C_0 a singleton and $\beta_0(Q) = 0$. The columns of $Q \in \mathrm{St}(d, d) = O(d)$ form an orthonormal eigenbasis of $\beta_k(r, Q)$. The map β_k is a smooth fibre bundle over the subset of $\mathcal{M}_k$ with distinct positive eigenvalues, with fibre $\{\pm 1\}^k \times O(d - k)$. Extending r_k to zero produces the sticking boundary

$$\{r_1 > \cdots > r_{k-1} > 0,\ r_k = 0\} \times O(d) \simeq \widetilde{\mathcal{M}}_{k-1}.$$

Thus, every adjacent rank transition becomes codimension-one in the lifted space. □

4 Numerical examples

4.1 Overview

In this Section, we first validate our methodology on two toy examples where $\mathcal{M}$ corresponds to linear subspaces of $\mathbb{R}^d$ (Section C.2) and to hyperspheres (Section 4.2). The former illustrates how our rate (2) accounts for changing incoming flux as the slope of the line varies; the latter demonstrates the effectiveness of Rao–Blackwellisation while highlighting a limitation of our method: infrequent manifold visits may slow down between-model mixing. We then apply the sampling approach to three examples of applied statistical interest arising in quantum state tomography (Section 4.3), Bayesian principal component analysis (Section 4.4), and survival analysis (Section 4.5). The quantum state tomography example highlights the computational value of model transitions that require

no additional likelihood evaluations. Bayesian principal component analysis provides a higher-dimensional setting for joint inference on rank and covariance. Finally, the survival analysis example demonstrates how a single construction accommodates multiple parametric families without the need to develop several bespoke reversible-jump proposals.

In these examples, we mainly use forward event-chain Monte Carlo (FECMC) as the default PDMP dynamics, given its favourable performance (Michel et al., 2020; Shiba and Kamatani, 2026), although our sticky manifold construction is agnostic to the choice of within-model dynamics. Our main comparator is the celebrated reversible jump algorithm (Green, 1995), defining between-model proposals that are bespoke to each application. For within-model sampling, we consider two alternatives, Hamiltonian Monte Carlo proposals (RJ-HMC) and standard FECMC dynamics (RJ-FECMC, Section 1.2), on the same parametrisation. In the former case, the algorithm alternates between within and across model proposals, while in the latter reversible jump proposals are triggered with a time-homogeneous Poisson rate $\lambda > 0$, both tuned in pilot runs to yield a comparable number of model transitions to the sticky manifold algorithm (SM-FECMC). RJ-FECMC serves as a bridge between SM-FECMC and RJ-HMC that helps distinguish the effects of within-model dynamics from those of the model-transition mechanism. To yield competitive performance, we tuned the integration step T and step size ϵ of the HMC dynamics based on the effective sample size per CPU time, while we did not allow any tuning or preconditioning for FECMC dynamics in SM-FECMC and RJ-FECMC. For a given sampler, its statistical efficiency is measured through the root mean squared error (RMSE) over 100 independent runs of the same budget. The ground truth is either computed by numerical quadrature or approximated by a long run, whose error or Monte Carlo standard error is verified to be sufficiently smaller than the RMSE. Further details on implementation are in Appendix C.

4.2 Varying hypersphere radii

In this section we sample from an equal mixture distribution of the d-dimensional standard Gaussian $\pi_{\mathbb{R}^d}$ and its restriction $\pi_{\mathcal{M}}$ to the hypersphere $\mathcal{M}_R := \{\|x\| = R\} \subset \mathbb{R}^d$. Using Gaussian velocity (Example 3.1), the rate is $\Lambda_R = \|v^\perp\|\chi_d(R)$, where χ_d is the chi density function with d degrees of freedom. Figure 6 (b) and (c) show the sample paths of $\|x\|$ for $d = 10$, and $R = \sqrt{d}$ and 1, respectively, in consideration that the mass concentrates around $\|x\| = \sqrt{d}$ as d becomes large. This highlights that, when $\mathcal{M}$ is close to the high-density region of the target, sticking times are short to account for shorter excursions away from $\mathcal{M}$, and vice versa. Thus, when the density ratio $\pi_{\mathbb{R}^d}/\pi_{\mathcal{M}}$ is small and manifold visits are rare, the mixing of the marginal model-index process can be limited by the longer sojourn, even if the mixture weights of the target distribution are equal $\omega = 1 - \omega = 1/2$. This deterioration is analogous to slow mixing in the reversible-jump methods with poorly tuned proposal distributions.

Table 1 compares four variants of the sticky manifold PDMP sampler for $R = \sqrt{d}, 1$ across varying dimensions. In particular these variants are *i*) The standard sampler; *ii*) The standard sampler where in addition $v^\perp$ is refreshed at refreshment events; *iii*) The standard sampler with the averaged unsticking rate (Corollary 3.8); *iv*) The previous sampler with deterministic sticking time (Corollary 3.9). For each sampler, RMSE changes

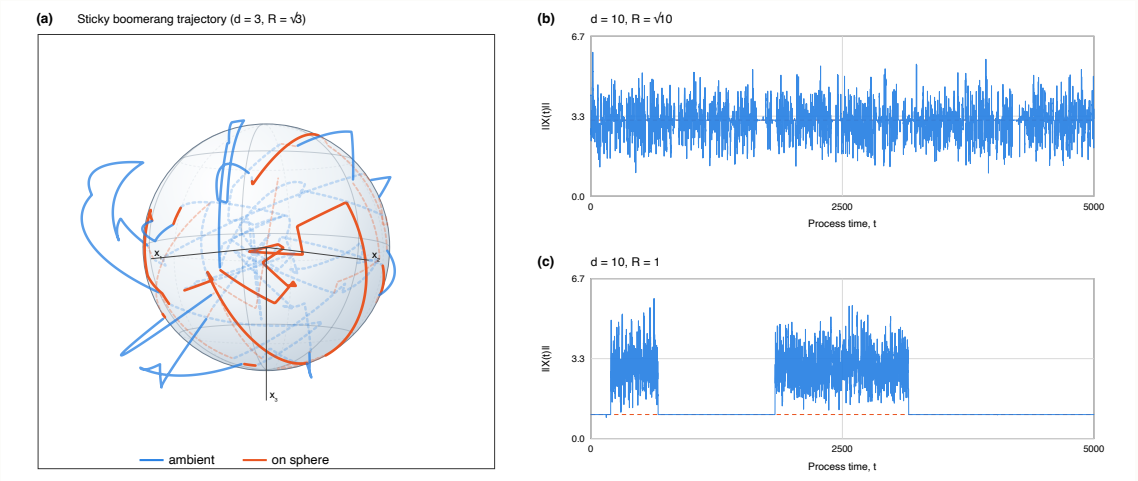


Figure 6: (a) Trajectory of the sticky boomerang sampler for $d = 3$ and $R = \sqrt{d}$; blue segments evolve in the ambient space while orange segments on the sphere along its geodesic flow (see Remark 2.4), with back-facing portions shown faint and dashed. (b), (c) Radius traces for $d = 10$ with (b) $R = \sqrt{d}$ and (c) $R = 1$. Both targets assign mass $1/2$ to the sphere. The smaller sphere is reached rarely and produces longer residence episodes.

only moderately as d increases when $R = \sqrt{d}$. When $R = 1$, however, RMSE increases rapidly with no model transitions for $d = 20, 50$. In both cases the Rao–Blackwellisation procedures improve sampler efficiency, most notably for the fixed unsticking time.

Table 1: RMSE for estimating the ambient mixture weight $\omega = 1/2$, mean number of trans-dimensional jumps, and the empirical variance of sticking times. Each condition uses 1000 independent stationary-started runs of time horizon $T = 10^4$.

	FECMC			+ normal refreshment			+ averaged rate			+ fixed unsticking time		
d	RMSE	jumps	Sticking time variance	RMSE	jumps	Sticking time variance	RMSE	jumps	Sticking time variance	RMSE	jumps	Sticking time variance
$R = \sqrt{d}$												
3	0.0242	4263.6	37.83	0.0096	4260.4	7.45	0.0089	4258.0	5.51	**0.0067**	4262.6	0.00
5	0.0222	4356.7	38.98	0.0098	4357.5	7.13	0.0093	4351.4	5.27	**0.0078**	4356.8	0.00
10	0.0212	4439.7	39.19	0.0110	4429.2	6.94	0.0103	4425.6	5.10	**0.0089**	4427.0	0.00
20	0.0227	4457.0	36.78	0.0126	4470.1	6.81	0.0117	4465.3	5.02	**0.0104**	4463.1	0.00
50	0.0247	4486.8	52.61	0.0150	4482.2	6.77	0.0140	4491.1	4.96	**0.0129**	4488.7	0.00
$R = 1$												
3	0.0279	3854.1	60.04	0.0124	3860.0	8.84	0.0113	3863.5	6.71	**0.0102**	3866.1	0.00
5	0.0401	1284.5	367.90	0.0252	1283.7	67.79	0.0242	1287.8	59.91	**0.0221**	1287.2	0.00
10	0.2349	12.4	6.40×10^5	0.2156	12.5	5.05×10^5	0.1954	12.5	5.24×10^5	**0.1743**	12.7	0.00
20	0.5000	0.0	—	0.5000	0.0	—	0.5000	0.0	—	0.5000	0.0	—
50	0.5000	0.0	—	0.5000	0.0	—	0.5000	0.0	—	0.5000	0.0	—

4.3 The Bloch sphere in quantum state tomography

Quantum state tomography estimates a quantum state from measurements on identically prepared systems (Blume-Kohout, 2010). Experiments often aim to prepare a pure quantum state, but noise can produce a mixed state. The state of a single qubit is represented

by a Bloch vector $r = (r_x, r_y, r_z) \in B^3 := \{\|r\| \leq 1\}$, with pure states on the boundary $\mathbb{S}^2 = \partial B^3$ and mixed states in the interior. To quantify uncertainty between the two models, we therefore assign positive prior mass to both the sphere and the interior, yielding a mixed-dimensional prior for Bayesian model averaging (Ferrie, 2014). Specifically, we use an equal mixture of uniform distributions on B^3 and $\mathbb{S}^2$.

A measurement in the direction $a \in \mathbb{S}^2$ yields $Y_a \in \{\pm 1\}$ with likelihood

$$\mathrm{P}[Y_a = y \mid r] = \frac{1 + y a^\top r}{2}, \qquad y \in \{-1, +1\}.$$

Synthetic data are generated from $r_\alpha = \alpha(1,1,1)^\top/\sqrt{3}$, with $\alpha \in \{1, .99\}$ and $N = 100$ independent measurements in uniformly random directions. We use Rao–Blackwellized sticky manifold FECMC (RB-SM-FECMC) with fixed sticking times (Corollary 3.9), and the unsticking rate is halved to account for one-sidedness (see Corollary 3.10). For RJ-HMC, we use $T = 3$ and $\epsilon \in \{0.08, 0.09\}$, guided by pilot searches maximising the ESS per CPU second, and geodesic Monte Carlo (Byrne and Girolami, 2013) on $\mathbb{S}^2$. The reversible-jump proposal is tailored to the spherical geometry (Brooks et al., 2003): it proposes a radius $R \in (0,1)$ when entering the ball and radially projects onto $\mathbb{S}^2$ in the reverse direction. Its frequency is calibrated separately for each (α, ϵ) to match RB-SM-FECMC's model-transition count per CPU second. Each algorithm is initialised from an exact posterior sample from a rejection sampler and run for one second 100 times. We report RMSEs of four quantities: the posterior pure state probability $p_{\mathbb{S}^2} := \mathrm{P}[\|r\| = 1 \mid y]$, two conditional posterior means $\mu_{\mathbb{S}^2} := \mathrm{E}[r \mid \|r\| = 1, y]$, $\mu_{B^3} := \mathrm{E}[r \mid \|r\| < 1, y]$, and the posterior mean $\mathrm{E}[r \mid y]$.

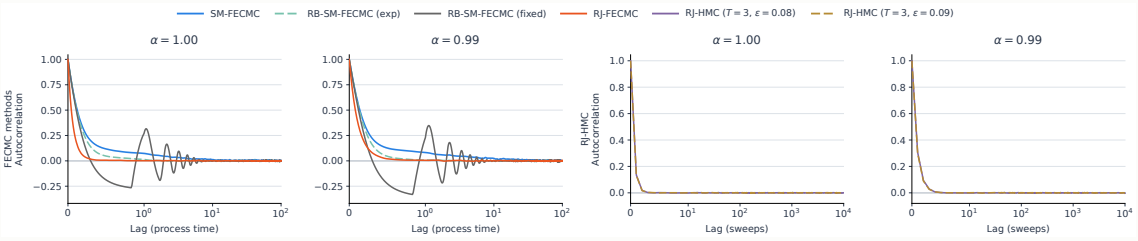


Figure 7: Model-indicator autocorrelation. Only RB-SM-FECMC with fixed sticking times (gray) plunges into the negative region, followed by damped oscillations. RB-SM-FECMC (exp) represents the version with the Rao–Blackwellised rate (Corollary 3.8), which leads to exponential sticking times.

At a budget of one CPU second, RB-SM-FECMC achieves smaller RMSEs than RJ-FECMC for all reported quantities across the two scenarios (Table 2, panel (a)). It also estimates the posterior pure-state probability $p_{\mathbb{S}^2}$ more accurately than RJ-HMC, despite approximately matched model-transition counts. RJ-HMC, however, gives more accurate estimates of the conditional posterior means $\mu_{\mathbb{S}^2}$ and μ_{B^3}, as well as the model-averaged posterior mean $\mathrm{E}[r \mid y]$. The good mixing of the model indicator process under RB-SM-FECMC is reflected in the autocorrelation function (Figure 7), in which RB-SM-FECMC exhibits negative regions and damped oscillations, whereas these features are absent from the other algorithms. These negative correlations may contribute to the reduced RMSE through antithetic effects.

Table 2: RMSE from 100 paired runs under two budgets: (a) 1 process CPU second and (b) 10^6 potential or gradient evaluations. Both panels use the same data and initial states. N_U and $N_{\nabla U}$ count evaluations of the potential U and its gradient, respectively, including directional derivatives in the latter. Bold values indicate the smallest RMSE.

α	Sampler	RMSE				Transitions	N_U	$N_{\nabla U}$
		$p_{\mathbb{S}^2}$	$\mu_{\mathbb{S}^2}$	μ_{B^3}	$\mathrm{E}[r \mid y]$			
(a) 1 process CPU second								
1.00	RB-SM-FECMC	$\mathbf{4.91}\times10^{-4}$	3.56×10^{-4}	8.34×10^{-4}	3.30×10^{-4}	1.69×10^{5}	0	5.80×10^{5}
	RJ-FECMC	9.90×10^{-4}	6.21×10^{-4}	1.12×10^{-3}	6.28×10^{-4}	1.69×10^{5}	7.55×10^{5}	1.71×10^{5}
	RJ-HMC ($\epsilon = 0.08$)	6.40×10^{-4}	2.60×10^{-4}	$\mathbf{5.77}\times10^{-4}$	2.46×10^{-4}	1.71×10^{5}	1.36×10^{6}	2.26×10^{6}
	RJ-HMC ($\epsilon = 0.09$)	6.25×10^{-4}	$\mathbf{2.36}\times10^{-4}$	6.54×10^{-4}	$\mathbf{2.25}\times10^{-4}$	1.71×10^{5}	1.36×10^{6}	2.25×10^{6}
0.99	RB-SM-FECMC	$\mathbf{4.87}\times10^{-4}$	4.33×10^{-4}	6.84×10^{-4}	3.48×10^{-4}	1.61×10^{5}	0	6.18×10^{5}
	RJ-FECMC	1.59×10^{-3}	5.26×10^{-4}	9.58×10^{-4}	5.37×10^{-4}	1.59×10^{5}	5.99×10^{5}	2.77×10^{5}
	RJ-HMC ($\epsilon = 0.08$)	7.91×10^{-4}	2.38×10^{-4}	$\mathbf{5.13}\times10^{-4}$	2.39×10^{-4}	1.62×10^{5}	1.32×10^{6}	2.38×10^{6}
	RJ-HMC ($\epsilon = 0.09$)	8.32×10^{-4}	$\mathbf{2.22}\times10^{-4}$	5.22×10^{-4}	$\mathbf{2.36}\times10^{-4}$	1.64×10^{5}	1.32×10^{6}	2.38×10^{6}
(b) 10^6 combined potential/first-derivative evaluations								
1.00	RB-SM-FECMC	$\mathbf{3.61}\times10^{-4}$	$\mathbf{2.92}\times10^{-4}$	$\mathbf{6.05}\times10^{-4}$	$\mathbf{2.58}\times10^{-4}$	2.92×10^{5}	0	1.00×10^{6}
	RJ-FECMC	9.16×10^{-4}	5.82×10^{-4}	1.07×10^{-3}	5.95×10^{-4}	1.83×10^{5}	8.16×10^{5}	1.84×10^{5}
	RJ-HMC ($\epsilon = 0.08$)	1.15×10^{-3}	4.98×10^{-4}	1.20×10^{-3}	4.64×10^{-4}	4.73×10^{4}	3.77×10^{5}	6.23×10^{5}
	RJ-HMC ($\epsilon = 0.09$)	1.11×10^{-3}	4.40×10^{-4}	1.16×10^{-3}	4.02×10^{-4}	4.74×10^{4}	3.76×10^{5}	6.24×10^{5}
0.99	RB-SM-FECMC	$\mathbf{4.10}\times10^{-4}$	$\mathbf{3.17}\times10^{-4}$	$\mathbf{5.44}\times10^{-4}$	$\mathbf{2.58}\times10^{-4}$	2.61×10^{5}	0	1.00×10^{6}
	RJ-FECMC	1.48×10^{-3}	4.93×10^{-4}	8.99×10^{-4}	5.10×10^{-4}	1.81×10^{5}	6.84×10^{5}	3.16×10^{5}
	RJ-HMC ($\epsilon = 0.08$)	1.41×10^{-3}	5.34×10^{-4}	1.06×10^{-3}	5.01×10^{-4}	4.39×10^{4}	3.56×10^{5}	6.44×10^{5}
	RJ-HMC ($\epsilon = 0.09$)	1.43×10^{-3}	4.76×10^{-4}	9.98×10^{-4}	4.31×10^{-4}	4.44×10^{4}	3.56×10^{5}	6.44×10^{5}

Computing the unsticking rate (2) does not involve likelihood evaluations and therefore RB-SM-FECMC is the only sampler compared in panel (a) of Table 2 that requires no evaluations of U during trajectory simulation, as recorded by $N_U = 0$. To distinguish efficiency per target evaluation from the remaining simulation costs in this three-dimensional example, panel (b) of Table 2 reports a complementary result at a fixed budget of $N_U + N_{\nabla U} = 10^6$. Each evaluation of U or its full gradient ∇U counts as one epoch, including all evaluations required for event-time simulation. Under this cost measure, RB-SM-FECMC achieves the smallest observed RMSE for every reported quantity in all scenarios.

4.4 Blown-up rank strata in Bayesian principal component analysis

In this example, we apply the blow-up construction in Section 3.5 to Bayesian principal component analysis (Bishop, 1998; Zhang et al., 2004) with unknown signal rank. Specifically, for $y_i \in \mathbb{R}^{20}$ and known σ^2,

$$y_i|x_i \sim N_{20}(x_i, \sigma^2 I_{20}), \quad x_i \sim N_{20}(0, G), \quad i = 1, \cdots, n.$$

We generate $n = 80$ observations with $\sigma = 0.5$ in six scenarios, where G has eigenvalues A: $(2, 1.5, 1, 0.5, 0, \dots, 0)$; B: $(1, 1, 1, 0.1, \dots, 0.1)$; C: $(1, 1, 1, 0.1, 0.1, 0.1, 0, \dots, 0)$; D: $(1, 1, 1, 1, 1, 0, \dots, 0)$;

E: $(1, 0.95, \ldots, 0.05)$; O: $(0, \ldots, 0)$. We consider a mixed-dimensional prior on G with equal weights on each rank stratum $\mathcal{M}_i$, in which $G = U_k A_k U_k^\top$ where $A_k \sim W_k(k+1, (k+1)^{-1} I_k)$ is Wishart-distributed and U_k is uniformly distributed on the Stiefel manifold $\mathrm{St}(k, d)$. The full-rank stratum $\mathcal{M}_{20}$ has dimension 210. We report RMSEs for three quantities: the posterior rank-probability vector $(p_0, \ldots, p_{20})$, where $p_k = \mathrm{P}[\mathrm{rank}(G) = k \mid y]$, the posterior mean covariance $\mathrm{E}[G \mid y]$, and the posterior mean signal (or denoised signal) $\mathrm{E}[x \mid y]$. SM-FECMC uses Rao–Blackwellised rates (Corollary 3.8). HMC dynamics combine reflective dynamics on the constrained Euclidean regions C_k with geodesic Monte Carlo (Byrne and Girolami, 2013) on the Stiefel manifold $O(20)$, and use integration step T and step size ϵ that are guided by pilot searches maximising the ESS per CPU time, while we did not allow any tuning or preconditioning for FECMC dynamics. All three algorithms are run on the same spectral lift coordinates (Example 3.15) for 3 CPU seconds over 100 times with shared initial states from a separate long run.

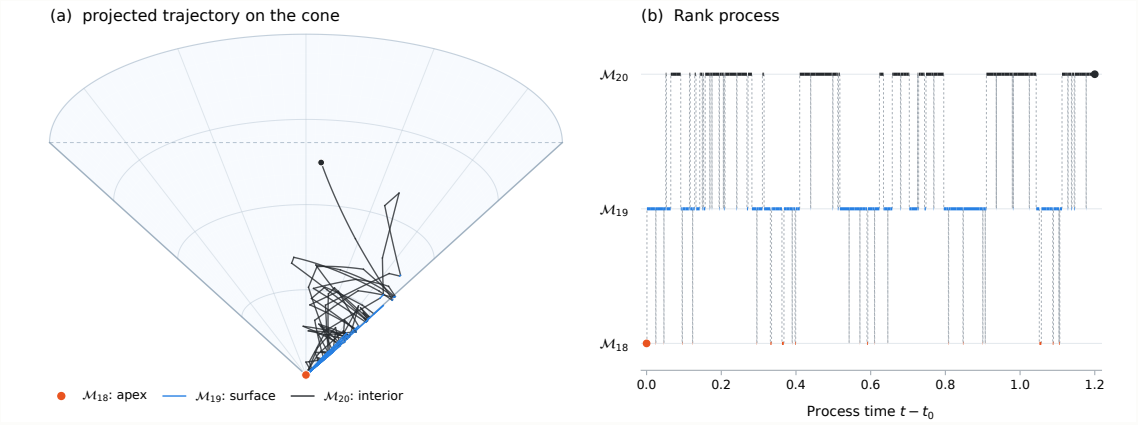


Figure 8: An illustrative SM-FECMC posterior trajectory in scenario E. (a) The trajectory of the Schur-complement $S_t := D_t - B_t^\top A_t^{-1} B_t \in \mathrm{Sym}_2(\mathbb{R})_+$ via the conical parametrisation (Example 3.13), where $U^{-1} G_t U = (A_t \; B_t; B_t^\top \; D_t)$ for a $U \in O(20)$. (b) The corresponding rank process over a short time interval.

Table 3 shows that either SM-FECMC or RJ-FECMC performs best for the posterior mean of G across all scenarios, while RJ-HMC with $(T, \epsilon) = (3, 0.05)$ performs best in scenario C for the posterior mean of x. Interestingly, across scenarios A–E, the RMSE ratio of RJ-HMC against either SM-FECMC or RJ-FECMC tends to increase as the posterior modal rank $k^\star$ increases (Figure 9). Moreover, the average acceptance probability of HMC proposals with the best-performing tuning also decreases accordingly (Appendix C.5.7). This may reflect the increasing number of order constraints on the eigenvalues, near whose boundaries the characteristic length scales of the posterior, and hence the acceptance rates, differ significantly (Appendix C.5.1), requiring small step sizes that may be suboptimal for other parts of the posterior and for dynamics on $O(20)$. In fact, the best-performing RJ-HMC in scenarios B and E uses $T = 10$ integration steps with a smaller step size ϵ, while in the other scenarios it uses $T = 3$. In contrast, the PDMP dynamics adapt to such anisotropy and are relatively robust to multiscale structure without further tuning or preconditioning, consistent with previous results (Bierkens et al., 2025; Lu and Luo, 2026).

Table 3: RMSEs for the six scenarios from 100 paired replicates with 3 CPU seconds. $k^\star$ denotes the rank with the maximum posterior probability $\hat{p}_{k^\star}$, which is computed from the reference run.

Scenario	Sampler	Rank-vector RMSE	Mean-G RMSE	Denoising RMSE	Mean rank transitions
A $(2, 1.5, 1, 0.5, 0, \dots, 0)$ $k^\star = 6$, $\hat{p}_{k^\star} = 0.459$	SM-FECMC	1.12×10^{-2}	$\mathbf{1.46 \times 10^{-2}}$	$\mathbf{4.95 \times 10^{-3}}$	3.96×10^{4}
	RJ-FECMC	1.15×10^{-2}	1.60×10^{-2}	5.25×10^{-3}	3.62×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.04$)	1.05×10^{-2}	2.31×10^{-2}	6.26×10^{-3}	3.96×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.05$)	$\mathbf{8.95 \times 10^{-3}}$	2.11×10^{-2}	5.81×10^{-3}	4.02×10^{4}
B $(1, 1, 1, 0.1, \dots, 0.1)$ $k^\star = 9$, $\hat{p}_{k^\star} = 0.450$	SM-FECMC	$\mathbf{1.27 \times 10^{-2}}$	$\mathbf{1.74 \times 10^{-2}}$	$\mathbf{9.63 \times 10^{-3}}$	2.85×10^{4}
	RJ-FECMC	1.32×10^{-2}	1.79×10^{-2}	9.77×10^{-3}	2.97×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.03$)	3.03×10^{-2}	1.05×10^{-1}	6.44×10^{-2}	3.01×10^{4}
	RJ-HMC ($T = 10$, $\epsilon = 0.007$)	2.09×10^{-2}	5.86×10^{-2}	2.70×10^{-2}	2.86×10^{4}
C $(1, 1, 1, 0.1, 0.1, 0.1, 0, \dots, 0)$ $k^\star = 4$, $\hat{p}_{k^\star} = 0.484$	SM-FECMC	1.34×10^{-2}	$\mathbf{1.09 \times 10^{-2}}$	4.92×10^{-3}	3.52×10^{4}
	RJ-FECMC	1.35×10^{-2}	1.13×10^{-2}	5.40×10^{-3}	3.82×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.04$)	9.03×10^{-3}	1.77×10^{-2}	5.33×10^{-3}	3.73×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.05$)	$\mathbf{8.56 \times 10^{-3}}$	1.50×10^{-2}	$\mathbf{4.80 \times 10^{-3}}$	3.71×10^{4}
D $(1, 1, 1, 1, 1, 0, \dots, 0)$ $k^\star = 7$, $\hat{p}_{k^\star} = 0.462$	SM-FECMC	1.07×10^{-2}	$\mathbf{1.79 \times 10^{-2}}$	$\mathbf{7.64 \times 10^{-3}}$	5.85×10^{4}
	RJ-FECMC	1.23×10^{-2}	2.05×10^{-2}	8.36×10^{-3}	5.64×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.04$)	$\mathbf{9.34 \times 10^{-3}}$	3.10×10^{-2}	9.53×10^{-3}	5.91×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.05$)	1.07×10^{-2}	2.96×10^{-2}	9.80×10^{-3}	5.91×10^{4}
E $(1, 0.95, 0.9, \dots, 0.05)$ $k^\star = 20$, $\hat{p}_{k^\star} = 0.606$	SM-FECMC	$\mathbf{2.13 \times 10^{-2}}$	$\mathbf{5.15 \times 10^{-2}}$	$\mathbf{1.42 \times 10^{-2}}$	6.45×10^{4}
	RJ-FECMC	2.94×10^{-2}	6.15×10^{-2}	1.70×10^{-2}	6.91×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.03$)	1.65×10^{-1}	1.05×10^{0}	3.10×10^{-1}	7.28×10^{4}
	RJ-HMC ($T = 10$, $\epsilon = 0.007$)	2.41×10^{-2}	1.59×10^{-1}	4.50×10^{-2}	6.71×10^{4}
O $(0, \dots, 0)$ $k^\star = 0$, $\hat{p}_{k^\star} = 0.816$	SM-FECMC	8.30×10^{-3}	5.40×10^{-4}	8.39×10^{-4}	3.76×10^{4}
	RJ-FECMC	6.53×10^{-3}	$\mathbf{4.61 \times 10^{-4}}$	$\mathbf{7.13 \times 10^{-4}}$	3.58×10^{4}
	RJ-HMC ($T = 2$, $\epsilon = 0.1$)	$\mathbf{5.82 \times 10^{-3}}$	5.84×10^{-4}	8.82×10^{-4}	3.92×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.05$)	9.75×10^{-3}	8.28×10^{-4}	1.26×10^{-3}	3.77×10^{4}

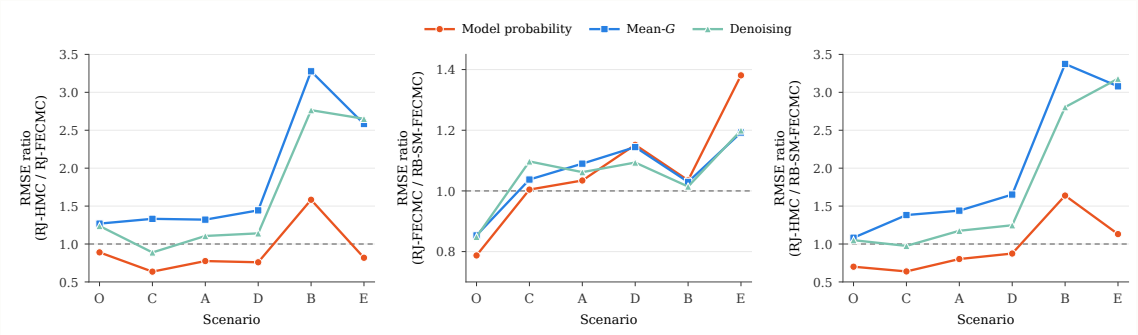


Figure 9: Relative RMSEs across the six PPCA scenarios. Left: RJ-HMC relative to RJ-FECMC; middle: RJ-FECMC relative to SM-FECMC; right: RJ-HMC relative to SM-FECMC. For each scenario and quantity, the RJ-HMC ratio uses the smallest observed RMSE among all configurations.

This robustness to scale differences provides a further reason to use PDMP dynamics for trans-dimensional sampling, in which posterior geometry may vary across strata,

whereas RJ-HMC may benefit from stratum-wise tuning or preconditioning. For example, in scenario C, the average acceptance probability of RJ-HMC with $(T, \epsilon) = (3, 0.05)$ ranges from 69.9% in $\mathcal{M}_3$ to 7.4% in $\mathcal{M}_6$ (Table 12). Additionally, RJ-HMC is sensitive to step-size tuning; Figure 10 shows that no single setting is uniformly competitive with SM-FECMC and RJ-FECMC, whereas the baseline SM-FECMC dynamics required no tuning.

Comparison with RJ-FECMC highlights a computational advantage of the rejection-free nature of the sticky manifold framework. In addition to the cost of evaluating acceptance probabilities, rejected reversible-jump proposals may incur computational overhead on interrupted event-time simulation. Additional results suggest that SM-FECMC achieves longer trajectories under the same CPU budget by avoiding these costs (Appendix C.5.7).

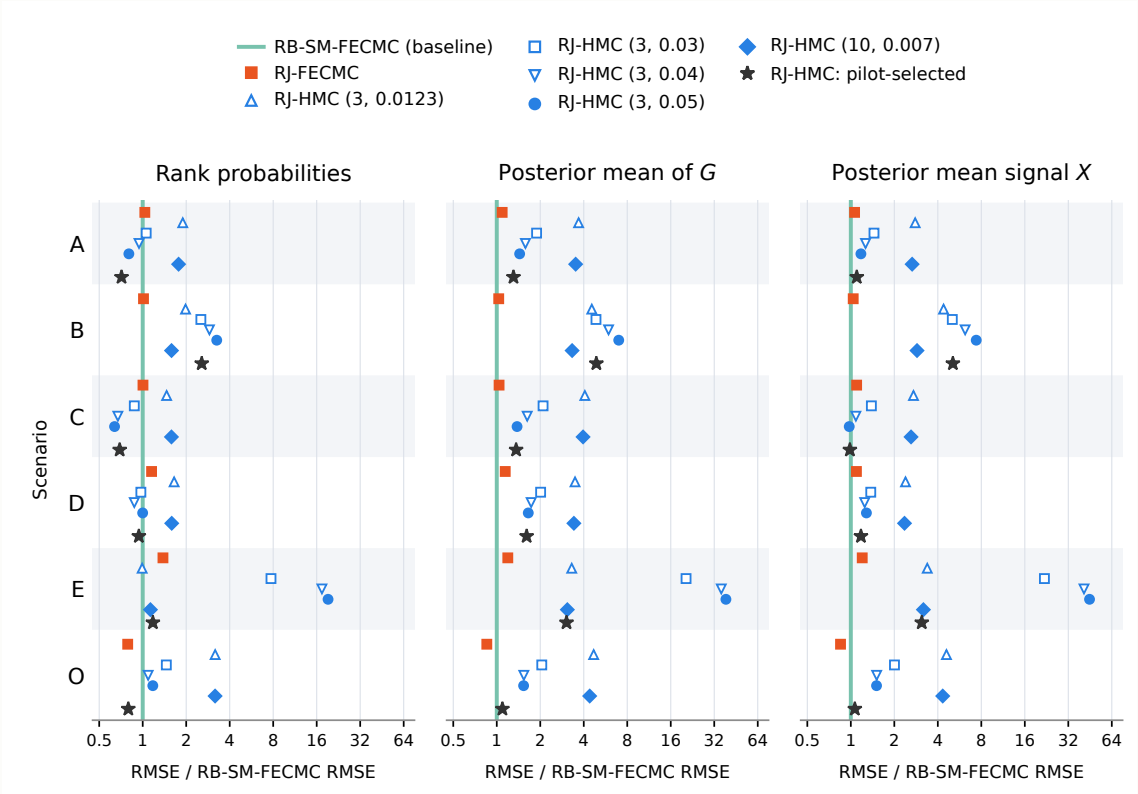


Figure 10: Sensitivity of RJ-HMC to hyperparameters across six scenarios. Blue symbols use four common configurations, while stars use scenario-specific settings selected by pilot runs based on ESS-per-CPU, which took about 2 hours for each scenario on average; see Appendix C.5.5 for details.

4.5 Model-averaged survival extrapolation

In the following we consider an example based on the application of survival models to time-to-event data arising from clinical trials. Concretely, we consider time-to-event data $y \in \mathbb{R}_+^N$ administratively censored at y_+ and where the estimand of interest is mean survival over a lifetime time-horizon $(0, y_\infty)$, requiring survival curves to be extrapolated over the interval (y_+, y_∞). This task commonly arises in the cost-effectiveness evaluation

of novel medical interventions in publicly funded healthcare systems, and precludes the use of standard non-parametric methods (e.g. Kaplan–Meier estimates) as they lack a mechanism to extrapolate beyond final observation times. See Baio (2020); Hardcastle et al. (2025b) for more general introductions to this problem setting.

A common solution is to assume that y are generated by a parametric survival distribution e.g. Weibull or log-Normal (Latimer, 2011), however extrapolations are then sensitive to the choice of survival model. A solution proposed in the literature is the use of Bayesian model averaging to both account for this uncertainty and to incorporate prior information into extrapolations through priors on the choice of underlying model (Negrín et al., 2017; Gallacher et al., 2021; Hardcastle et al., 2025a).

We propose a novel computational approach to this problem using the sticky manifold PDMP sampler and the generalised gamma family of distributions (Stacy, 1962), a three-parameter distribution that contains the log-normal, Weibull, gamma, and ammag distributions as special cases. Using the parametrisation (μ, σ, λ) of Prentice (1974) these cases correspond to the submanifolds $\mathcal{M}_{\text{LN}} = \{\lambda = 0\}$, $\mathcal{M}_{\text{W}} := \{\lambda = 1\}$, $\mathcal{M}_{\Gamma} = \{\lambda = \sigma\}$, and $\mathcal{M}_{\text{A}} := \{\lambda = 1/\sigma\}$ (Figure 11). We place prior mass on these submanifolds, and then sample from the resulting posterior using the sticky manifold PDMP (Remark 3.12), avoiding the design of multiple reversible jump steps, and facilitating model averaging over the parametric sub-families.

We consider the following prior over $\theta = (\alpha, \beta, \eta, \lambda)$ with $\eta := \log \sigma$:

$$\pi(d\theta) = \omega_{\text{GG}} \pi_{\text{GG}}(\theta)\, d\theta + \sum_{i \in \{\text{LN,W,}\Gamma\text{,A}\}} \omega_i \pi_i(\theta) \mathcal{H}^3_{\mathcal{M}_i}(d\theta), \qquad \sum_{i \in \{\text{GG,LN,W,}\Gamma\text{,A}\}} \omega_i = 1,$$

where $\pi_{\text{GG}}(\theta)\, d\theta$ assigns independent Gaussian priors to the elements of θ and $\pi_i \mathcal{H}^3_{\mathcal{M}_i}$ is its restriction onto the constrained manifold. In particular, the prior weights ω_i can be chosen to reflect prior information about the choice of submodel based on previous analyses.

We apply the model to data for 810 leukemia patients from the CLL-8 trial that informed NICE technology appraisal TA174 (Williams et al., 2017) evaluating the efficacy of the novel immunotherapy Rituximab. Final observation times occur at $y_+ = 4$, and we evaluate mean survival up to $y_\infty = 30$. We consider two versions of the model, the first with control and treatment arms related through an accelerated failure time assumption, and the second where independent models are fitted for control and treatment arms. This latter approach is common in health technology assessment as it avoids strong assumptions on the relationship between treatment arms. We implement the model with uniform prior weights, and further consider a second hypothetical scenario where previous analyses indicate that the log-Normal model is the most appropriate model for the control arm, with $\omega_{\text{LN}} = 1/2$. This acts as a prior for control and treatment groups under the accelerated failure time assumption, but is only applied to the model for the control arm when the groups are modelled independently.

Running the SM-FECMC algorithm with $T = 10^5$ results in the trajectories shown in panel (a) of Figure 11, with the sampler placing non-negligible posterior probability on all four parametric sub-families. Example hazard and survival functions for the control group when treatment arms are modelled separately with an informative prior are shown in panel (b) and (c) of Figure 11, with 95% credible intervals under the model averaged posterior and the log-normal posterior shown in grey and blue respectively. Note

the larger uncertainty intervals in the extrapolation period under the averaged model, accounting for the epistemic uncertainty that selecting a single parametric model does not address. Full posterior model probabilities are included in Appendix C. For uniform ω_i the ammag model is the most favoured model with 55%, 54% and 46% posterior probabilities. All other models have $< 20\%$ posterior probability with the generalised-gamma receiving the least posterior mass in all cases. With the informative prior, the log-normal model goes from 11% posterior probability to 31% posterior probability under the accelerated failure time model, and 12% to 34% posterior probability in the independent control arm model. The impact on mean survival is shown in Table 4. Note that model averaging increases the uncertainty in estimates of mean survival when compared to selecting a single parametric model. Further, the impact of the informative prior has a small impact under the accelerated failure time assumption, but reduces the expected difference in mean survival when the arms are modelled separately by about a year. Given the current NICE threshold of £35,000 per quality adjusted life year, this difference would potentially be large enough to alter the final funding decision for the treatment.

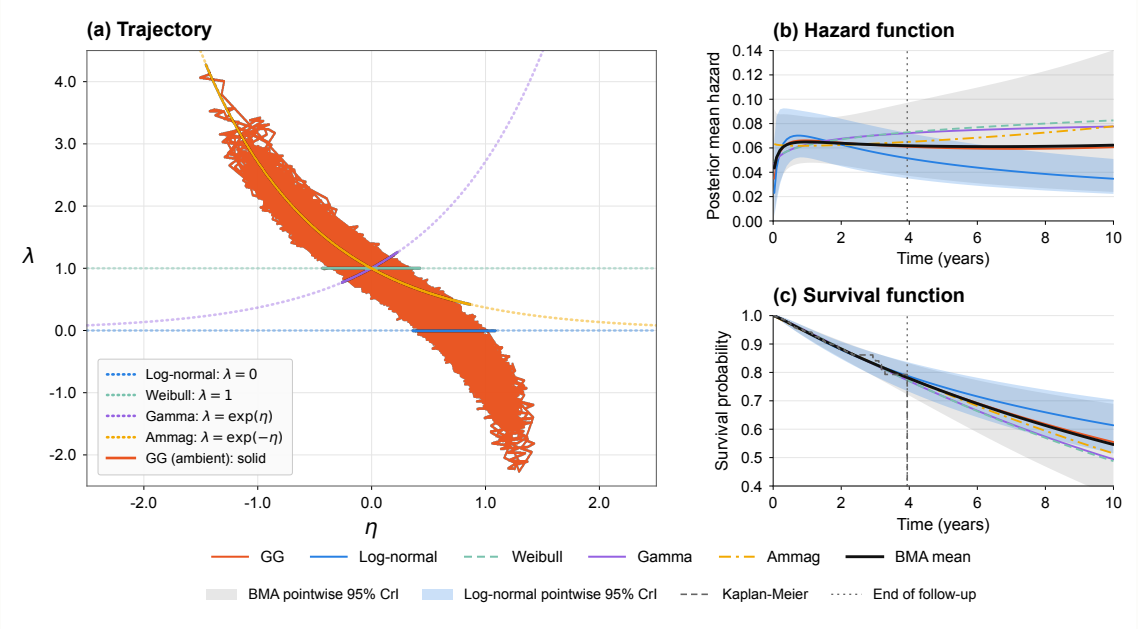


Figure 11: (a) Full SM-FECMC trajectory for the joint AFT model with uniform prior model probabilities on the (η, λ) plane with $\eta = \log \sigma$. (b,c) Posterior mean hazard and survival functions for the control arm fitted separately under the informative model prior.

5 Discussion

We have developed the trans-dimensional moves based on sticking and unsticking that are rejection-free, require no additional likelihood evaluation, and apply to both linear and non-linear submanifolds under the same principle. In numerical experiments, we demonstrated regimes in which the sticky manifold framework leads to better estimates with Rao–Blackwellisation than the reversible-jump counterparts, a high-dimensional ex-

Table 4: Estimates of 30-year restricted mean survival time (years). Entries are posterior means (95% credible intervals).

				Model averaged	
Model	Gen-Gamma	Log-normal	Ammag	Uniform prior	Informative prior
Treatment arms modelled together					
Control	15.63 (8.82, 20.70)	18.16 (15.74, 20.48)	13.08 (8.49, 18.02)	13.85 (8.81, 19.48)	14.82 (9.02, 19.96)
Treatment	17.89 (11.56, 22.05)	19.85 (17.47, 22.13)	16.04 (11.33, 20.41)	16.59 (11.65, 21.22)	17.33 (11.93, 21.63)
Difference	2.26 (−0.56, 5.73)	1.70 (−0.55, 3.89)	2.97 (−0.64, 6.72)	2.74 (−0.61, 6.44)	2.51 (−0.58, 6.24)
Treatment arms modelled separately					
Control	14.81 (7.86, 20.53)	17.19 (14.17, 20.08)	12.74 (8.35, 17.95)	13.35 (8.45, 19.14)	14.33 (8.70, 19.55)
Treatment	18.25 (11.31, 22.98)	20.10 (17.17, 22.73)	16.55 (11.37, 21.34)	16.98 (11.53, 21.90)	16.98 (11.53, 21.90)
Difference	3.45 (−5.71, 12.34)	2.92 (−1.20, 6.93)	3.80 (−3.59, 10.67)	3.63 (−4.42, 11.01)	2.65 (−5.28, 10.58)

ample that highlights the potential of PDMP dynamics in varying posterior geometry, and an application where multiple non-linear submodel constraints can be handled without bespoke jump proposals. A broader goal is to give practitioners greater freedom to choose statistically-motivated parametrisations, rather than computationally-motivated ones. Realising this goal more generally will require practical strategies for choosing coordinates and dynamics and for simulating boundary hitting and event times.

The blow-up procedure (Section 3.5) is a geometric solution to high codimensions, as illustrated in the Bayesian PCA example (Section 4.4). However, it may limit the range of priors that can be used with our sticky manifold construction, for regular priors on the associated radial blow-up coordinate translate into densities on the original space with a diverging Jacobian factor near the sticking boundaries (see Remarks 3.14 and C.1). If, instead, a prior does not diverge sufficiently rapidly in the original space, a naïve application of PDMP dynamics on the blow-up coordinates will have zero probability of hitting the front face, producing reducible dynamics. Additional procedures are therefore required to accommodate such priors within the blow-up procedure.

In our numerical experiments, RJ-FECMC was used primarily as a bridge to disentangle the effects of within-model PDMP dynamics from those of the model-transition mechanism, rather than to identify an optimal sampler. In practice, reversible-jump and sticky-manifold transitions could complement each other rather than serve as competing components in trans-dimensional PDMP sampling. Sticky-manifold transitions require no additional likelihood evaluations, while reversible-jump moves may facilitate between-model exploration, particularly when boundary hits are infrequent. Further improvements may come from adaptive tuning and preconditioning of the within-model dynamics. Such methods have been developed for both PDMP-based samplers (Bertazzi and Bierkens, 2022; Chevallier and Raab, 2025; Chevallier et al., 2026) and HMC (Hoffman and Gelman, 2014; Bou-Rabee et al., 2026).

There are several immediate extensions to the example in Section 4.5, for example, using this approach for distributional selection in more complex survival models (e.g. Hardcastle et al., 2025a) or combining the model and sampling architecture with standard sticky PDMP dynamics to perform simultaneous distributional and variable selection.

Acknowledgements

The authors thank Jong Hee Park, Yuki Shiraito, Alexandros Beskos, Samuel Livingstone, and Gianluca Baio for insightful discussions. They thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme *Monte Carlo sampling: beyond the diffusive regime*, where initial discussions on this work took place. They are also grateful to the Institute of Statistical Mathematics for its hospitality during LH's research visits from 17 to 31 May 2025 and from 24 October to 7 November 2026, which were instrumental in the development of this work. HS is supported by JST BOOST, Japanese Grant Number JPMJBS2412. LH was supported by EPSRC grant EP/W523835/1 while a PhD student at University College London, and was further supported by Medical Research Council Career Development Award UKRI332. KK was supported by JST CREST, Grant Number JPMJCR2115.

References

S. Agrawal, J. Bierkens, and G. O. Roberts. Large Sample Scaling Analysis of the Zig-Zag Algorithm for Bayesian Inference. *arXiv:2411.14983*, 2025.

C. Andral and K. Kamatani. Automated Techniques for Efficient Sampling of Piecewise-Deterministic Markov Processes, 2024.

C. Andrieu and S. Livingstone. Peskun-Tierney Ordering for Markovian Monte Carlo: Beyond the Reversible Scenario. *The Annals of Statistics*, 49(4):1958–1981, 2021.

C. Andrieu, N. d. Freitas, and A. Doucet. Robust Full Bayesian Learning for Radial Basis Networks. *Neural Computation*, 13(10):2359–2407, 10 2001. ISSN 0899-7667.

A. Arnstad, G. O. Storvik, F. Pagani, and L. Rønneberg. Sparse Bayesian Neural Networks via Sticky Piecewise Deterministic Markov Processes, 2026.

F. Baccelli and P. Brémaud. *Elements of Queueing Theory: Palm Martingale Calculus and Stochastic Recurrences*, volume 26 of *Stochastic Modelling and Applied Probability*. Springer Berlin, Heidelberg, 2 edition, 2003.

G. Baio. survHE: Survival Analysis for Health Economic Evaluation and Cost-Effectiveness Modeling. *Journal of Statistical Software*, 95(14):1–47, 2020.

C. Bell, K. Łatuszyński, and G. O. Roberts. Adaptive Stereographic MCMC, 2025.

A. Bertazzi and J. Bierkens. Adaptive Schemes for Piecewise Deterministic Monte Carlo Algorithms. *Bernoulli*, 28(4):2404–2430, 2022.

J. Bierkens and G. Roberts. A Piecewise Deterministic Scaling Limit of Lifted Metropolis-Hastings in the Curie-Weiss Model. *The Annals of Applied Probability*, 27(2):846 – 882, 2017.

J. Bierkens, A. Bouchard-Côté, A. Doucet, A. B. Duncan, P. Fearnhead, T. Lienart, G. Roberts, and S. J. Vollmer. Piecewise Deterministic Markov Processes for Scalable Monte Carlo on Restricted Domains. *Statistics & Probability Letters*, 136:148–154, 2018. ISSN 0167-7152. The role of Statistics in the era of big data.

J. Bierkens, P. Fearnhead, and G. Roberts. The Zig-Zag Process and Super-Efficient Sampling for Bayesian Analysis of Big Data. *The Annals of Statistics*, 47(3):1288–1320, 2019.

J. Bierkens, S. Grazzi, K. Kamatani, and G. Roberts. The Boomerang Sampler. In Hal Daumé III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 908–918. PMLR, 13–18 Jul 2020.

J. Bierkens, S. Grazzi, G. Roberts, and M. Schauer. Methods and Applications of PDMP Samplers with Boundary Conditions. *arXiv:2303.08023*, 2023a.

J. Bierkens, S. Grazzi, F. van der Meulen, and M. Schauer. Sticky PDMP Samplers for Sparse and Local Inference Problems. *Statistics and Computing*, 33(1):8, 2023b.

J. Bierkens, K. Kamatani, and G. O. Roberts. Scaling of Piecewise Deterministic Monte Carlo for Anisotropic Targets. *Bernoulli*, 31(3):2323 – 2350, 2025.

C. Bishop. Bayesian PCA. In M. Kearns, S. Solla, and D. Cohn, editors, *Advances in Neural Information Processing Systems*, volume 11. MIT Press, 1998.

R. Blume-Kohout. Optimal, Reliable Estimation of Quantum States. *New Journal of Physics*, 12(4):043034, apr 2010.

V. I. Bogachev. *Measure Theory*. Springer Berlin, Heidelberg, 2007.

N. Bou-Rabee and J. M. Sanz-Serna. Randomized Hamiltonian Monte Carlo. *The Annals of Applied Probability*, 27(4):2159 – 2194, 2017.

N. Bou-Rabee, B. Carpenter, T. S. Kleppe, and S. Liu. The Within-Orbit Adaptive Leapfrog No-U-Turn Sampler. *Journal of Machine Learning Research*, 27(113):1–64, 2026.

A. Bouchard-Côté, S. J. Vollmer, and A. Doucet. The Bouncy Particle Sampler: A Nonreversible Rejection-Free Markov Chain Monte Carlo Method. *Journal of the American Statistical Association*, 113(522):855–867, 2018.

S. P. Brooks, P. Giudici, and G. O. Roberts. Efficient Construction of Reversible Jump Markov Chain Monte Carlo Proposal Distributions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(1):3–39, 02 2003. ISSN 1369-7412.

S. Byrne and M. Girolami. Geodesic Monte Carlo on Embedded Manifolds. *Scandinavian Journal of Statistics*, 40(4):825–845, 2013.

A. Chevallier and E. Raab. Covariance-Adaptive Bouncy Particle Samplers via Split Lagrangian Dynamics, 2025.

A. Chevallier, P. Fearnhead, and M. Sutton. Reversible Jump PDMP Samplers for Variable Selection. *Journal of the American Statistical Association*, 118(544):2915–2927, 2023.

A. Chevallier, S. Power, A. Q. Wang, and P. Fearnhead. PDMP Monte Carlo Methods for Piecewise Smooth Densities. *Advances in Applied Probability*, 56(4):1153–1194, 2024.

A. Chevallier, S. Power, and M. Sutton. Towards Practical PDMP Sampling: Metropolis Adjustments, Locally Adaptive Step-sizes, and NUTS-Based Time Lengths, 2026.

A. Chin and A. Nishimura. Smoothing Out Sticking Points: Sampling from Discrete-Continuous Mixtures with Dynamical Monte Carlo by Mapping Discrete Mass into a Latent Universe. *arXiv:2511.07340*, 2025.

A. Corbella, S. E. F. Spencer, and G. O. Roberts. Automatic Zig-Zag Sampling in Practice. *Statistics and Computing*, 32(6):107, 2022.

R. V. Craiu, D. Vats, G. L. Jones, S. Brooks, A. Gelman, and X.-L. Meng, editors. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2026.

M. H. A. Davis. Piecewise-Deterministic Markov Processes: A General Class of Non-Diffusion Stochastic Models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3):353–376, 1984.

M. H. A. Davis. *Markov Models and Optimization*, volume 49 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 1993.

P. Dellaportas, J. J. Forster, and I. Ntzoufras. On Bayesian Model and Variable Selection Using MCMC. *Statistics and Computing*, 12(1):27–36, 2002.

P. Diaconis, S. Holmes, and M. Shahshahani. Sampling from a Manifold. In *Advances in modern statistical theory and applications: a Festschrift in honor of Morris L. Eaton*, volume 10, pages 102–126. Institute of Mathematical Statistics, 2013.

A. Durmus, A. Guillin, and P. Monmarché. Piecewise Deterministic Markov Processes and their Invariant Measures. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 57(3):1442 – 1475, 2021.

R. S. Ehlers and S. P. Brooks. Adaptive Proposal Construction for Reversible Jump MCMC. *Scandinavian Journal of Statistics*, 35(4):677–690, 2008.

P. Fearnhead, C. Nemeth, C. J. Oates, and C. Sherlock. *Scalable Monte Carlo for Bayesian Learning*. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2025.

H. Federer. *Geometric Measure Theory*. Classics in Mathematics. Springer Berlin, Heidelberg, 1996.

C. Ferrie. Quantum Model Averaging. *New Journal of Physics*, 16(9):093035, sep 2014.

S. Frühwirth-Schnatter, D. Hosszejni, and H. F. Lopes. Sparse Bayesian Factor Analysis When the Number of Factors Is Unknown (with Discussion). *Bayesian Analysis*, 20(1): 213–344, 2025.

D. Gallacher, P. Kimani, and N. Stallard. Extrapolating Parametric Survival Models in Health Technology Assessment: A Simulation Study. *Medical Decision Making*, 41(1): 37–50, 2021. PMID: 33283635.

P. J. Green. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, 82(4):711–732, 1995.

P. J. Green. Trans-dimensional Markov Chain Monte Carlo. In P. J. Green, N. L. Hjort, and S. Richardson, editors, *Highly Structured Stochastic Systems*. Oxford University Press, 05 2003. ISBN 9780198510550.

D. Grieser. *Basics of the b-Calculus*, pages 30–84. Birkhäuser Basel, Basel, 2001. ISBN 978-3-0348-8253-8.

L. Hardcastle, S. Livingstone, and G. Baio. Averaging Polyhazard Models using Piecewise Deterministic Monte Carlo with Applications to Data with Long-Term Survivors. *The Annals of Applied Statistics*, 19(4):3179 – 3202, 2025a.

L. Hardcastle, S. Livingstone, and G. Baio. Diffusion Piecewise Exponential Models for Survival Extrapolation using Piecewise Deterministic Monte Carlo. *arXiv:2505.05932*, 2025b.

M. D. Hoffman and A. Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(47):1593–1623, 2014.

M. Holmes-Cerfon. Simulating Sticky Particles: A Monte Carlo Method to Sample a Stratification. *The Journal of Chemical Physics*, 153(16):164112, 10 2020. ISSN 0021-9606.

T. S. Kleppe. Connecting the Dots: Numerical Randomized Hamiltonian Monte Carlo with State-Dependent Event Rates. *Journal of Computational and Graphical Statistics*, 31 (4):1238–1253, 2022.

J. Koskela. Zig-Zag Sampling for Discrete Structures and Nonreversible Phylogenetic MCMC. *Journal of Computational and Graphical Statistics*, 31(3):684–694, 2022.

S. Lang. *Differential and Riemannian Manifolds*, volume 160 of *Graduate Texts in Mathematics*. Springer New York, 3 edition, 1995.

N. Latimer. NICE DSU Technical Support Document 14: Survival Analysis for Economic Evaluations alongside Clinical Trials–Extrapolation with Patient-Level Data. *Report by the Decision Support Unit*, 2011.

V. Lemaire, M. Thieullen, and N. Thomas. Exact Simulation of the Jump Times of a Class of Piecewise Deterministic Markov Processes. *Journal of Scientific Computing*, 75(3): 1776–1807, 2018.

P. A. W. Lewis and G. S. Shedler. Simulation of Nonhomogeneous Poisson Processes by Thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.

H. F. Lopes and M. West. Bayesian Model Assessment in Factor Analysis. *Statistica Sinica*, 14:41–67, 2004.

J. Lu and Y. Luo. Entropy Hypocoercivity and Query Complexity for the Bouncy Particle Sampler, 2026.

R. B. Melrose. Real Blow Ups. Lecture Note, 2008.

M. Michel, J. Mayer, and W. Krauth. Event-Chain Monte Carlo for Classical Continuous Spin Models. *Europhysics Letters*, 112(2):20003, nov 2015.

M. Michel, A. Durmus, and S. Sénécal. Forward Event-Chain Monte Carlo: Fast Sampling by Randomness Control in Irreversible Markov Chains. *Journal of Computational and Graphical Statistics*, 29(4):689–702, 2020.

T. J. Mitchell and J. J. Beauchamp. Bayesian Variable Selection in Linear Regression. *Journal of the American Statistical Association*, 83(404):1023–1032, 1988.

R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 1982.

R. M. Neal. *Handbook of Markov Chain Monte Carlo*, chapter MCMC Using Hamiltonian Dynamics, pages 113–162. Chapman and Hall/CRC, 2011.

M. A. Negrín, J. Nam, and A. H. Briggs. Bayesian Solutions for Handling Uncertainty in Survival Extrapolation. *Medical Decision Making*, 37(4):367–376, 2017. PMID: 27281336.

Y. Nishikawa, M. Michel, W. Krauth, and K. Hukushima. Event-Chain Algorithm for the Heisenberg Model: Evidence for $z \simeq 1$ Dynamic Scaling. *Phys. Rev. E*, 92:063306, Dec 2015.

F. Pagani, A. Chevallier, S. Power, T. House, and S. Cotter. NuZZ: Numerical Zig-Zag for General Models. *Statistics and Computing*, 34(1):61, 2024.

A. A. Pourzanjani, R. M. Jiang, B. Mitchell, P. J. Atzberger, and L. R. Petzold. Bayesian Inference over the Stiefel Manifold via the Givens Representation. *Bayesian Analysis*, 16 (2):639 – 666, 2021.

S. Power and G. Vasdekis. Some Aspects of Robustness in Modern Markov Chain Monte Carlo. *arXiv:2511.21563*, 2025.

R. L. Prentice. A Log Gamma Model and Its Maximum Likelihood Estimation. *Biometrika*, 61(3):539–544, 12 1974. ISSN 0006-3444.

S. Richardson and P. J. Green. On Bayesian Analysis of Mixtures with an Unknown Number of Components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(4):731–792, 1997.

H. Shiba and K. Kamatani. Diffusive Scaling Limits of Forward Event-Chain Monte Carlo: Provably Efficient Exploration with Partial Refreshment. *arXiv:2602.17087*, 2026.

E. W. Stacy. A Generalization of the Gamma Distribution. *The Annals of Mathematical Statistics*, 33(3):1187 – 1192, 1962.

M. Sutton and P. Fearnhead. Concave-Convex PDMP-based Sampling. *Journal of Computational and Graphical Statistics*, 32(4):1425–1435, 2023.

J. H. Tran and T. S. Kleppe. Numerical Generalized Randomized Hamiltonian Monte Carlo for Piecewise Smooth Target Densities, 2025.

D. van den Bergh and M. Marsman. Accelerating Bayesian Variable Selection using Piecewise Deterministic Markov Processes, 2026.

A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner. Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2):667 – 718, 2021.

J. Vermaak, C. Andrieu, A. Doucet, and S. J. Godsill. Reversible Jump Markov Chain Monte Carlo Strategies for Bayesian Model Selection in Autoregressive Processes. *Journal of Time Series Analysis*, 25(6):785–809, 2004.

P. A. Whalley, D. Paulin, and B. Leimkuhler. Randomized time Riemannian Manifold Hamiltonian Monte Carlo. *Statistics and Computing*, 34(1):48, 2024.

C. Williams, J. D. Lewsey, D. F. Mackay, and A. H. Briggs. Estimation of Survival Probabilities for Use in Cost-effectiveness Analyses: A Comparison of a Multi-state Modeling Survival Analysis Approach with Partitioned Survival and Markov Decision-Analytic Modeling. *Medical Decision Making*, 37(4):427–439, 2017. PMID: 27698003.

C. Wu and C. P. Robert. Coordinate Sampler: A Non-Reversible Gibbs-Like MCMC Sampler. *Statistics and Computing*, 30(3):721–730, 2020.

J. Yang, K. Łatuszyński, and G. O. Roberts. Stereographic Markov Chain Monte Carlo. *The Annals of Statistics*, 52(6):2692–2713, 2024.

Z. Zhang, K. L. Chan, J. T. Kwok, and D.-Y. Yeung. Bayesian Inference on Principal Component Analysis Using Reversible Jump Markov Chain Monte Carlo. In *Proceedings of the 19th National Conference on Artificial Intelligence*, AAAI'04, page 372–377. AAAI Press, 2004. ISBN 0262511835.

A Manifold PDMP

A.1 The Hausdorff measure and area formula

For a Riemannian manifold $\mathcal{M}$, the volume measure induced from the Riemannian volume form (see Lang, 1995, Chapter X) agrees (on all Borel measurable sets) with the Hausdorff measure induced from the Riemannian distance, with the normalisation convention of Federer (1996, 2.10.2). This fact is proven in Federer, 1996, 3.2.46. Throughout this paper, we refer to this measure as the Hausdorff measure, and we provide a definition following Federer, 1996, Section 2.10 and Bogachev, 2007, Section 3.10 (iii).

Definition A.1 (Hausdorff measure). Let $(\mathcal{M}, g)$ be a q-dimensional Riemannian manifold, dist_g the Riemannian distance, and α_q be the volume of the q-dimensional unit ball in $\mathbb{R}^q$. Define a function $\zeta : 2^{\mathcal{M}} \to [0, \infty]$ by

$$\zeta^q(S) = \frac{\alpha_q}{2^q}(\mathrm{diam}_g\, S)^q, \qquad \emptyset \neq S \subset \mathcal{M},$$

where $\mathrm{diam}_g\, S := \sup \mathrm{dist}_g(S \times S)$. We define $\zeta^q(\emptyset) = 0$. This ζ^q defines an outer measure in the sense of Carathéodory (see Bogachev, 2007, Definition 1.11.1), i.e., a monotone and σ-subadditive set function $\mathcal{H}^q$ by

$$\mathcal{H}^q(A) := \lim_{\delta \searrow 0} \mathcal{H}^q_\delta(A) \qquad A \subset \mathcal{M},$$

$$\mathcal{H}^q_\delta(A) = \inf\left\{\sum_{i=1}^{\infty} \zeta(S_i) \,\middle|\, A \subset \bigcup_{i=1}^{\infty} S_i, \operatorname{diam}_g(S_i) \le \delta\right\}.$$

The resulting complete measure $\mathcal{H}^q$, constructed from Carathéodory's extension theorem (Bogachev, 2007, Theorem 1.11.4), is, in fact, a restriction of $\mathcal{H}^q$ onto a complete σ-algebra and will be called *Hausdorff measure* on the Riemannian manifold $(\mathcal{M}, g)$. We sometimes write $\mathcal{H}^q$ as $\mathcal{H}^q_g$ to clarify the Riemannian distance on which it depends. □

The Hausdorff measure $\mathcal{H}^q$ defined above is the same as the Lebesgue measure when $\mathcal{M} = \mathbb{R}^q$, including the domain of the definition.

For subsequent reference, we give the area formula, which is an extension of the change of measures formula (Federer, 1996; Diaconis et al., 2013). For a Riemannian manifold $(\mathcal{M}, g)$,

$$\int_A g(f(x)) J_m f(x) \lambda^m(\mathrm{d}x) = \int_{\mathcal{M}} g(y) |\{x \in A : f(x) = y\}| \mathcal{H}^m_g(\mathrm{d}y),$$

where $f : \mathbb{R}^m \to \mathcal{M}$ is Lipschitz continuous, $A \subset \mathbb{R}^m$, and $|\cdot|$ denotes the cardinality of a set.

A.2 Proof of Proposition 2.2

Proof. Defining the mapping $T_\varphi : T(\mathcal{M}) \to \mathcal{U} \times \mathbb{R}^p$ as $T_\varphi(\xi, \eta) := (\varphi(\xi), D\varphi(\xi)\eta)$, the above pullback construction can be expressed as

$$T_\varphi^{-1}(X_t, V_t) = (\varphi^{-1}(X_t), D\varphi^{-1}(X_t)V_t).$$

Since T_φ is again a diffeomorphism, actually a bundle isomorphism, we see that $T_\varphi^{-1}(X_t, V_t)$ is again a PDMP, whose transition semigroup (P^φ_t) satisfies the intertwining relationship

$$(P^\varphi_t F) = P_t(F \circ T_\varphi^{-1}) \circ T_\varphi,$$

where (P_t) is the semigroup of the original PDMP (X_t, V_t). Using this conjugacy and the area formula (Federer, 1996, Theorem 3.2.5, 3.2.46), we conclude that

$$\mu^\varphi(d\xi d\eta) := \pi(\varphi(\xi)) J^g_\varphi(\xi) \mathcal{H}^p_g(d\xi) \nu^\varphi_\xi(d\eta), \qquad (\xi, \eta) \in T(\mathcal{M})$$

is invariant with respect to P^φ_t, where $\nu^\varphi_\xi := (D\varphi(\xi)^{-1})_* \nu_{\varphi(\xi)}$ is the pullback probability measure. The calculation goes as follows

$$\begin{aligned}
\int_{T(\mathcal{M})} P^\varphi_t F(\xi, \eta) \mu^\varphi(d\xi d\eta) &= \int_{T(\mathcal{M})} P_t(F \circ T_\varphi^{-1})(T_\varphi(\xi, \eta)) \mu^\varphi(d\xi d\eta) \\
&= \int_{\mathcal{U} \times \mathbb{R}^p} P_t(F \circ T_\varphi^{-1})(x, v) \pi(x)\, dx \nu_x(dv) \\
&= \int_{\mathcal{U} \times \mathbb{R}^p} F(T_\varphi^{-1}(x, v)) \pi(x)\, dx \nu_x(dv) = \int_{T(\mathcal{M})} F(\xi, \eta) \mu^\varphi(d\xi d\eta),
\end{aligned}$$

where we used the invariance of (X_t, V_t) in the second-to-last equality. ■

A.3 Dynamics of manifold PDMPs

Theorem A.2 (Pullback of PDMP generators). Let $(\mathcal{M}, g)$ be a p-dimensional Riemannian manifold, $\mathcal{U} \subset \mathbb{R}^p$ an open set, and $\varphi : \mathcal{M} \to \mathcal{U}$ be a diffeomorphism. Assume that the PDMP (X_t, V_t) on $\mathcal{U} \times \mathbb{R}^p$ has a strong generator of the form

$$\mathcal{L}f(z) = \langle v, \nabla_x f(z) \rangle + \lambda(z)(Qf - f)(z), \qquad z = (x, v),$$

for a function $f : \mathcal{U} \times \mathbb{R}^p \to \mathbb{R}$ in the domain. Then, the generator of the induced PDMP process $Z_t^\varphi := (\varphi^{-1}(X_t), D\varphi^{-1}(X_t)V_t)$ on $T(\mathcal{M})$ is given by, for a test function $F : T(\mathcal{M}) \to \mathbb{R}$,

$$\mathcal{L}^\varphi F(\zeta) = \mathcal{L}_D^\varphi F(\zeta) + \lambda^\varphi(\zeta)(Q^\varphi F - F)(\zeta), \qquad \zeta = (\xi, \eta), \tag{13}$$
$$\mathcal{L}_D^\varphi = v^i \frac{d}{dx^i} - \Gamma^i_{jk} v^j v^k \frac{d}{dv^i},$$

where $\mathcal{L}_D^\varphi$ is the metric spray of $g_\varphi := \varphi^* g_E$, the pullback metric of the Euclidean metric g_E. Additionally, if $\mathcal{L}$ admits $C_c^1(\mathcal{U} \times \mathbb{R}^p)$ as a core, then $C_c^1(T(\mathcal{M}))$ is also a core of $\mathcal{L}^\varphi$. □

Proof. We discuss locally by fixing a chart for $\mathcal{M}$. Denote the generator of $Z = (X, V)$ on $\mathcal{U} \times \mathbb{R}^d$ by

$$\mathcal{L}f(z) = \langle v, \nabla_x f(z) \rangle + \lambda(z)(Qf - f)(z), \qquad z = (x, v),$$

for a function $f : \mathcal{U} \times \mathbb{R}^d \to \mathbb{R}$ in the domain. Then, defining the bundle isomorphism as $T_\varphi(\xi, \eta) := (\varphi(\xi), D\varphi(\xi)\eta)$, the process $T_\varphi^{-1}(X, V)$ on $T(\mathcal{M})$ is again a PDMP. Its generator can be written as

$$\mathcal{L}^\varphi F(\zeta) = \langle \eta, \nabla_\xi F(\zeta) \rangle - \langle \Gamma_\varphi(\zeta), \nabla_\eta F(\zeta) \rangle + \lambda^\varphi(\zeta)(Q^\varphi F - F)(\zeta), \qquad \zeta = (\xi, \eta),\ F \in C_c^1(T(\mathcal{M})), \tag{14}$$

where $\lambda^\varphi := \lambda \circ T_\varphi$, $Q^\varphi F := Q(F \circ T_\varphi^{-1}) \circ T_\varphi$, and

$$\Gamma_\varphi(\zeta) := -D^2\varphi^{-1}(x)[v, v], \qquad (x, v) = T_\varphi(\xi, \eta)$$

is the Christoffel acceleration term that pushes the PDMP flow along the geodesics of the metric induced through φ, i.e., $g_\varphi := \varphi^* g_{\mathbb{R}^d}$. This construction works for a general manifold $\mathcal{M}$ with the deterministic part of $\mathcal{L}^\varphi$ replaced with the metric spray of g_φ. It is straightforward to check that the diagram commutes

$$\mathcal{L}f \circ T_\varphi = \mathcal{L}^\varphi(f \circ T_\varphi)$$

for any $f : \mathcal{U} \times \mathbb{R}^d \to \mathbb{R}$ in the domain of $\mathcal{L}$. This indicates that the precomposition map $T_\varphi^*(f) = f \circ T_\varphi$ defines an isometry between $C_c^1(\mathcal{U} \times \mathbb{R}^d)$ and $C_c^1(T(\mathcal{M}))$ with respect to the graph norm. Therefore, $C_c^1(T(\mathcal{M}))$ is a core of $\mathcal{L}^\varphi$, given that $C_c^1(\mathcal{U} \times \mathbb{R}^d)$ is a core of $\mathcal{L}$, the set of all smooth functions with compact supports (Durmus et al., 2021, Corollary 24). ■

B Sticky manifold PDMP

B.1 Blow-up construction and tangent-normal splitting

For an embedded submanifold $\mathcal{M} \hookrightarrow \mathbb{R}^d$ with codimension at least one, merely taking the boundary results in $\partial(\mathbb{R}^d \setminus \mathcal{M}) = \overline{\mathcal{M}}$. To retain the information as to which side the process approached the submanifold $\mathcal{M}$, we consider *blowing-up* $\mathcal{M}$. For a general treatment of blow-ups, especially as a technique of singularity resolution in geometry, see Melrose (2008, Section 1.5) or Grieser (2001, Section 2.3). Here, we define the blow-up of $\mathcal{M}$ in $\mathbb{R}^d$ as a set

$$[\mathbb{R}^d; \mathcal{M}] := (\mathbb{R}^d \setminus \mathcal{M}) \sqcup S(N_{\mathcal{M}}(\mathbb{R}^d)),$$

where $S(N_{\mathcal{M}}(\mathbb{R}^d)) = (N_{\mathcal{M}}(\mathbb{R}^d) \setminus 0_{\mathcal{M}})/\mathbb{R}_{>0}$ is the unit sphere subbundle of the normal bundle $N_{\mathcal{M}}(\mathbb{R}^d)$. We equip this set with a surjection $\beta : [\mathbb{R}^d; \mathcal{M}] \to \mathbb{R}^d$ such that

$$\beta(y) = y \quad (y \in \mathbb{R}^d \setminus \mathcal{M}), \qquad \beta(x, s) = x \quad ((x, s) \in S(N_{\mathcal{M}}(\mathbb{R}^d))),$$

called a *blow-down map*. We define a manifold structure on $[\mathbb{R}^d; \mathcal{M}]$ so that β is smooth and $[\mathbb{R}^d; \mathcal{M}]$ has boundary $\beta^{-1}(\mathcal{M})$, that is also called *front face*. To do so, we use a restriction of the exponential map $\exp|_{N_{\mathcal{M}}(\mathbb{R}^d)} : N_{\mathcal{M}}(\mathbb{R}^d) \to \mathbb{R}^d$, that is diffeomorphic on an open neighbourhood U of the zero section $0_{\mathcal{M}} = \{(x, 0)\}_{x \in \mathcal{M}} \subset N_{\mathcal{M}}(\mathbb{R}^d)$ (Lang, 1995, Theorem IV.5.3). A key observation is that this neighbourhood, after excluding the zero section $U \setminus 0_{\mathcal{M}}$, admits a polar coordinate expression in the fibre direction $0 \neq v \in N_{\mathcal{M}}(\mathbb{R}^d)$, i.e.,

$$(x, v) = (x, rs), \qquad x \in \mathcal{M},\ r > 0,\ s \in S_x(N_{\mathcal{M}}(\mathbb{R}^d)), \tag{15}$$

where $S_x(N_{\mathcal{M}}(\mathbb{R}^d))$ is the unit sphere in the fibre of x in $N_{\mathcal{M}}(\mathbb{R}^d)$. The angular part s represents the direction of approach to $x \in 0_{\mathcal{M}}$, which corresponds to $\mathcal{M}$ via β, and the radial part $r > 0$ indicates the distance to $0_{\mathcal{M}} \simeq \mathcal{M}$. Together with x, the triplet (x, r, s) induces local coordinates on $U \setminus 0_{\mathcal{M}}$, an open set in $N_{\mathcal{M}}(\mathbb{R}^d)$. By extending this local coordinate to $r = 0$, we obtain local coordinates of the blown up manifold $[\mathbb{R}^d; \mathcal{M}]$ and the blow-down map $\beta(x, r, s) = \exp(x, rs)$, which is smooth since it is a smooth extension of the restricted exponential map $\exp|_{N_{\mathcal{M}}(\mathbb{R}^d)}$.

As a result, each point in the front face $(x, s) \in \beta^{-1}(\mathcal{M})$ is distinguished by the direction of approach s to $x \in \mathcal{M}$. When the submanifold $\mathcal{M}$ is coorientable and of codimension 1, the normal bundle is one dimensional and trivial: $N_{\mathcal{M}}(\mathbb{R}^d) \simeq \mathcal{M} \times \mathbb{R}$. In this case, fixing a coorientation, $\beta^{-1}(\mathcal{M}) \simeq \mathcal{M} \times \{\pm 1\}$ and we see that the boundaries of $[\mathbb{R}^d; \mathcal{M}]$ consist of two copies of $\mathcal{M}$.

The state space E of the sticky manifold PDMP is the pullback of the tangent bundle $T(\mathbb{R}^d)$ along the blow-down map β, denoted by

$$E := \beta^*(T(\mathbb{R}^d)) = \{(x, v) \mid x \in [\mathbb{R}^d; \mathcal{M}], v \in T_{\beta(x)}(\mathbb{R}^d)\}.$$

Since the blown-up manifold $[\mathbb{R}^d; \mathcal{M}]$ can be decomposed into the interior $\mathbb{R}^d \setminus \mathcal{M}$ and the boundary, which we denote by $\mathcal{M}_+ \sqcup \mathcal{M}_-$ assuming $\mathcal{M}$ is of codimension one and coorientable, we have a decomposition, after an appropriate identification,

$$E = T(\mathbb{R}^d \setminus \mathcal{M}) \sqcup T_{\beta(\mathcal{M}_+)}(\mathbb{R}^d) \sqcup T_{\beta(\mathcal{M}_-)}(\mathbb{R}^d),$$

where $T_{\mathcal{M}}(\mathbb{R}^d) = \{(x, v) \in T(\mathbb{R}^d) \mid x \in \mathcal{M}\}$ is a restriction of $T(\mathbb{R}^d)$. This vector bundle admits a canonical splitting induced by the Euclidean structure (see also Lang, 1995, Section IV.3),

$$T_{\mathcal{M}}(\mathbb{R}^d) \simeq T(\mathcal{M}) \oplus N_{\mathcal{M}}(\mathbb{R}^d),$$

where $\oplus$ is the direct sum of vector bundles (Lang, 1995, Section III.4). In other words, on a boundary point $x \in \mathcal{M}_+$, the velocity $v \in \beta^*(T(\mathbb{R}^d))_x = T_{\beta(x)}(\mathbb{R}^d)$ can be decomposed into the tangent component $v_{\mathcal{M}_+}$ and the normal component $v^{\perp}_{\mathcal{M}_+}$.

Next we define the frozen boundary $\mathfrak{F}$ in E. The normal component $v^{\perp}_{\mathcal{M}_+}$ is *pointing outwards from* $\mathcal{M}$ when it is positively aligned with ∂_r, derived from the local coordinate

(15), i.e. $\langle v^{\perp}_{\mathcal{M}_+}, \partial_r \rangle > 0$. The opposite direction is said to be *pointing inwards*. The states in the subset

$$\mathfrak{F}^{\pm} := \{(x, v) \in \beta^*(T(\mathbb{R}^d)) \mid x \in \mathcal{M}_{\pm}, \langle v^{\perp}_{\mathcal{M}_\pm}, \partial_r \rangle < 0\}$$

are considered to be 'stuck' to $\mathcal{M}$; $\mathfrak{F} := \mathfrak{F}^+ \cup \mathfrak{F}^-$ is called the 'frozen boundary' (see also Bierkens et al., 2023b, Section 2.1). The pair $(x, v_{\mathcal{M}_+})$ is naturally identified with a point in $T(\mathcal{M})$, describing dynamics confined in $\mathcal{M}$. The remaining normal component $v^{\perp}_{\mathcal{M}_+}$ is considered to be an auxiliary variable that retains the old velocity. The base variable $x \in \mathcal{M}_{\pm}$ retains the information of which side the process entered the submanifold $\mathcal{M}$. At a thawing event, $x \in \mathcal{M}_+$ is mapped to the corresponding point in the opposite boundary component $\mathcal{M}_-$, possibly with an appropriate modification of velocity, so that the normal component $v^{\perp}_{\mathcal{M}_+}$ now points outwards. The process will then leave $\mathcal{M}$, ensuring the whole procedure will yield a well-defined Markov process.

The above construction generalises naturally to the cases that the codimension of $\mathcal{M}$ is larger than one. In such cases, a subtle difference arises in that the gauge dimension of $s \in S(N_{\mathcal{M}}(\mathbb{R}^d))$ becomes non-trivial. Motion along these fibres is invisible under the blow-down map, although it may influence the direction of subsequent departures from the front face. As in our probabilistic PCA example (Section 4.4), we specify this gauge dynamics as $\dot{\theta} \sim N(0, 1)$ in the rank-one stratum.

B.2 Lifting manifold dynamics to the frozen boundary

To align it with the blown-up state space E, we need to horizontally lift the manifold PDMP generator along $\overline{\beta} : \mathfrak{F} \to T(\mathcal{M})$, which is defined by

$$\overline{\beta}(x, v) = (x, P_x v),$$

where $P_x : T_x([\mathbb{R}^d; \mathcal{M}]) \to T_x(\mathcal{M})$ denotes the orthogonal projection. Although other ways of lifting are possible, we do not consider them in this work.

We can alternatively describe this locally. Let the deterministic part of the generator on $\mathcal{M}$ be given by

$$\mathcal{L}_D = \eta^i \frac{}{\xi^i} - \Gamma^i_{jk}(\xi)\eta^j\eta^k \frac{}{\eta^i},$$

for local coordinates $(\xi, \eta) \in T(\mathcal{M})$, and Γ being the Christoffel symbols of the connection induced by, say, a chosen coordinate chart φ. Then, the lifted generator is given by the same form

$$\mathcal{L}^{\text{lift}}_D = (\vec{v})^i \frac{}{x^i} - \Gamma^i_{jk}(x)(\vec{v})^j(\vec{v})^k \frac{}{(\vec{v})^i},$$

but now for a global parametrisation $(x, \vec{v}, a) \in \mathfrak{F}$ induced by the decomposition

$$v = \vec{v} + v^{\perp} \in T_x([\mathbb{R}^d; \mathcal{M}]) = T_x(\mathcal{M}) \oplus N_x(\mathcal{M}),$$

with $a = \|v^{\perp}\|$ when $(x, v) \in \mathfrak{F}^+$ and $a = -\|v^{\perp}\|$ when $(x, v) \in \mathfrak{F}^-$. In particular, it follows that $\mathcal{L}^{\text{lift}}_D(\|v^{\perp}\|) = 0$. In other words, the normal component $v^{\perp}$ is parallel-transported along $\mathcal{M}^{\pm}$, with the norm $\|v^{\perp}\|$ preserved.

Equivalently, $\mathcal{L}^{\text{lift}}_{\mathcal{M}}$ can be characterised by the following properties, which will be used in the subsequent proof. First, the coordinate chart $(x, \vec{v}, v^{\perp})$ on $\mathfrak{F}$ has an inverse mapping $\Psi : T(\mathcal{M}) \times (\mathbb{R} \setminus \{0\}) \to \mathfrak{F}$ defined by

$$\Psi(x, v, a) = \begin{cases} (x, v - a\partial_r) \in \mathfrak{F}^+ & a > 0, \\ (x, v + a\partial_r) \in \mathfrak{F}^- & a < 0. \end{cases} \tag{16}$$

which is a diffeomorphism. This can be understood as a kind of trivialisation of the 'tangled' tangent bundle $\mathfrak{F}$. Then,

$$(\mathcal{L}^{\text{lift}}_{\mathcal{M}} f) \circ \Psi(x, v, a) = \mathcal{L}_{\mathcal{M}}(f \circ \Psi(\cdot, \cdot, a))(x, v). \tag{17}$$

The lifted dynamics preserve the invariance property. We give a proof in the Gaussian velocity case.

Proposition B.1 (Invariance of the lifted manifold dynamics). Assume invariant manifold dynamics before lifting. Under Assumption B.2 and Gaussian velocity $\nu_{E\setminus\mathfrak{F}}(x, -) = \nu_{\mathfrak{F}}(x, -) = N_d(0, I_d)$,

$$\int_{\mathfrak{F}} \mathcal{L}^{\text{lift}}_{\mathcal{M}} f \, d\Pi_{\mathfrak{F}} = 0,$$

for every $f \in \mathcal{D}(\mathcal{L}^{\text{lift}}_{\mathcal{M}})$. □

Proof. The general strategy is that $\mathfrak{F}$ has a natural foliation with respect to $a \in \mathbb{R} \setminus \{0\}$ and we employ Fubini's theorem after trivialising the foliation induced by the conserved value: the normal velocity. This is done by $\Psi_a(x, v) := \Psi(x, v, a)$ defined in (16). Recall that the diffeomorphism $\Psi : T(\mathcal{M}) \times (\mathbb{R} \setminus \{0\}) \to \mathfrak{F}$ satisfies (17). The fact that $\nu = \gamma^d$ is the d-dimensional standard Gaussian distribution simplifies the proof since an orthogonal transformation of a standard Gaussian is again a standard Gaussian and any one-dimensional projection is again standard Gaussian independent of the rest, rendering a product disintegration

$$\Psi_x^* \gamma^d = \gamma^{d-1} \otimes \gamma^1$$

on each fibre $T_x([\mathbb{R}^d; \mathcal{M}])$ identical over all $x \in \mathcal{M}^{\pm}$. Therefore, using Fubini's theorem after a change of variables along Ψ,

$$\begin{aligned}
\int_{\mathfrak{F}} \mathcal{L}^{\text{lift}}_{\mathcal{M}} f \, d\Pi_{\mathfrak{F}} &= \int_{\mathbb{R}} \int_{T(\mathcal{M})} (\mathcal{L}^{\text{lift}}_{\mathcal{M}} f) \circ \Psi(x, \vec{v}, a) \, \pi_{\mathcal{M}}(x) \mathcal{H}^{d-1}_{\mathcal{M}}(dx) \gamma^{d-1}(d\vec{v}) \gamma^1(da) \\
&= \int_{\mathbb{R}} \left(\int_{T(\mathcal{M})} \mathcal{L}_{\mathcal{M}}(f \circ \Psi(\cdot, \cdot, a))(x, \vec{v}) \, \pi_{\mathcal{M}}(x) \mathcal{H}^{d-1}_{\mathcal{M}}(dx) \gamma^{d-1}(d\vec{v}) \right) \gamma^1(da),
\end{aligned}$$

where we used the intertwining relation (17) in the last equation. The term inside the parentheses equals zero since $f \circ \Psi(\cdot, \cdot, a) \in \mathcal{D}(\mathcal{L}_{\mathcal{M}})$ for almost all $a \in \mathbb{R}$. ■

The proof extends to the general case in which there exists a general disintegration $\Psi^* \Pi_{\mathfrak{F}}(dx d\vec{v} da) = \Pi^a_{\mathfrak{F}}(dx d\vec{v}) \nu^{\perp}_x(da)$ and $\Pi^a_{\mathfrak{F}}$ is invariant for almost all a.

B.3 Assumptions for the theorems

Assumption B.2 . For ambient PDMP dynamics in $\mathbb{R}^d$ and manifold PDMP dynamics in an orientable closed submanifold $\mathcal{M} \subset \mathbb{R}^d$ of codimension one, we assume the following conditions:

[A1] Both dynamics are velocity-jump PDMPs that have deterministic flows generated by a smooth second-order vector field $\mathfrak{X}_{(x,v)} = (v, F(x, v))$ (Lang, 1995, Section IV.3), satisfying the standard condition (24.8) of Davis (1993).

[A2] The overall number of jumps is finite: $\mathrm{E}_z[N_t] < \infty$ for all $z \in E$ and $t > 0$.

[A3] The unsticking rate Λ is locally integrable along the flow.

[A4] Almost surely, the velocity is not tangent to $\mathcal{M}$ whenever the position lies within $\mathcal{M}$, including the initial state.

[A5] The ambient velocity distribution $\nu_{E\setminus\mathfrak{F}}(x, dv)$ has density, denoted by $\nu_x(v)$, with respect to a reference measure $\rho_x(dv)$ on the fibre $\mathcal{V}_x(\mathbb{R}^d)$ of a smooth fibre bundle $\mathcal{V}(\mathbb{R}^d) \subset T(\mathbb{R}^d)$, whose total space is an embedded submanifold without boundary. The deterministic vector field $\mathcal{X}$ is tangent to $\mathcal{V}(\mathbb{R}^d)$. Further, the joint density $m(z) := \pi_{\mathbb{R}^d}(x)\nu_x(v)$ is of class $C^1(\mathcal{V}(\mathbb{R}^d))$, and the joint measure

$$\Omega(dx, dv) := dx\rho_x(dv)$$

defines a smooth positive density on $\mathcal{V}(\mathbb{R}^d)$ (Lang, 1995, Section XI.4, p.304).

[A6] For $\mathcal{H}^{d-1}_{\mathcal{M}}$-almost all $\xi \in \mathcal{M}$, $\pi_{\mathcal{M}}(\xi) = 0$ implies $\pi_{\mathbb{R}^d}(\xi) = 0$.

[A7] The velocity and the acceleration component F of the vector field $\mathfrak{X}$ have Π-integrable norm:

$$\int_E \Big(\|v\| + \|F(z)\| \Big) \Pi(dz) < \infty.$$

[A8] The overall jump intensity is Π-integrable:

$$\int_E \lambda(z)\Pi(dz) < \infty.$$

□

The last two integrability assumptions, [A7] and [A8], are necessary in a functional analytic argument of Lemma B.8, while they are not used in the main local arguments that focus on test functions with compact support. The regularity assumption [A5] is required to perform integration by parts in Lemma B.7. The Gaussian case corresponds to $\mathcal{V}_x(\mathbb{R}^d) = T_x(\mathbb{R}^d)$, while the two primary non-Gaussian examples of the fibre subbundle $\mathcal{V}(\mathbb{R}^d)$ are the unit sphere bundle $\mathcal{V}_x(\mathbb{R}^d) = \mathbb{S}^{d-1}$ and the discrete bundle $\mathcal{V}_x(\mathbb{R}^d) = \{\pm 1\}^d$. [A6] is required to obtain the unsticking rate function Λ. The first four assumptions, [A1]-[A4], provide the well-definedness of the sticky manifold PDMP.

Proposition B.3 (Well-definedness of sticky manifold PDMP). Assume [A1], [A2], [A3], and [A4] in Assumption B.2. Then the process (X_t, V_t) defined by the sticky manifold PDMP algorithm (Algorithm 2) is a well-defined PDMP and has (7) as its extended generator. Additionally, the domain $\mathcal{D}(\mathcal{L})$ is composed of measurable functions $f : E \to \mathbb{R}$ that satisfy the following conditions

(1) Along the deterministic flow ϕ_t, $t \mapsto f(\phi_t(z))$ is absolutely continuous on each bounded time interval for any $z \in E$.

(2) the cumulative jump increment is locally integrable, i.e.,

$$\mathrm{E}_z\left[\sum_{t=0}^{\sigma_n} |\Delta f(X_t, V_t)|\right] < \infty,$$

for an increasing sequence of stopping times $\{\sigma_n\}$.

(3) Boundary condition:

$$f_{\mathrm{in}}(z) = Q_{\mathrm{stick}} f(z) = \int_{\mathfrak{F}} Q_{\mathrm{stick}}(z, d\zeta) f(\zeta), \qquad z \in \Gamma_{\mathrm{in}},$$

where $f_{\mathrm{in}}(z) := \lim_{t \nearrow 0} f \circ \phi_t(z)$.

□

Proof. By considering the state space E as the direct sum of two components $E_1 := E \setminus \mathfrak{F}, E_2 := \mathfrak{F}$, we can apply Theorem 26.14 of Davis (1993) after checking Conditions 1 to 4 in (24.8) of Davis (1993). We check these four conditions in turn:

(1) The regularity of the vector field $\mathfrak{X}$ on each component, together with the non-explosion of its flow, is assumed in [A1].

(2) On $\mathfrak{F}$, the intensity of the manifold PDMP jump $\lambda_{\mathfrak{F}}$ is assumed to be locally integrable along the flow in [A1]. Together with [A3], we obtain the locally integrable rate of the sticky manifold PDMP on $\mathfrak{F}$.

(3) Our construction of sticking and unsticking mechanisms, denoted here also by $Q_{\text{stick}}, Q_{\text{unstick}}$, ensures $Q_{\text{stick}}(z, \{z\}) = Q_{\text{unstick}}(z, \{z\}) = 0$.

(4) The overall number of jumps, including forced jumps, is finite given [A2]. This is because a sticking event only takes place after an unsticking event.

By Theorem 26.14 of Davis (1993), we conclude that $\mathcal{D}(\mathcal{L})$ is as such. ■

Corollary B.4 (measure-separating property of the extended generator domain). Denote the subset of $\mathcal{D}(\mathcal{L})$ of bounded functions f satisfying $\mathcal{L}f \in L^1(\Pi)$ by $\mathcal{D}_0$. Then, under the same assumptions of Proposition B.3 in addition to [A7] and [A8], $\mathcal{D}_0$ separates probability measures on E: for two probability measures μ, ν, the following condition implies $\mu = \nu$

$$\int_E f d\mu = \int_E f d\nu, \qquad f \in \mathcal{D}_0. \tag{18}$$

□

Proof. We show that (18) implies that the same equality holds for an arbitrary $g \in C_c^1(E)$. This will conclude the measure separating property for $\mathcal{D}_0$ since $C_c^1(E)$ is measure separating (Bogachev, 2007, Example 2.12.10). This will be concluded from the bounded convergence theorem by taking a bounded sequence $\{f_n\} \subset \mathcal{D}_0$ that converges to g pointwise. f_n assumes the same value as g on $\mathfrak{F}$, while it converges to the value of $g_{\text{in}} := Q_{\text{stick}}g$ at $\mathfrak{F}$ from within $E \setminus \mathfrak{F}$. To construct f_n, we first define a flow-box local coordinate along $\mathfrak{X}$ on $E \setminus \mathfrak{F}$. For a point $z \in E \setminus \mathfrak{F}$, define

$$\tau(z) := \inf\{t > 0 \mid \phi_t(z) \in \Gamma_{\text{in}}\},$$

and $b(z) := \phi_{\tau(z)}(z)$ when $\tau(z) < \infty$ where ϕ_t is the flow of $\mathfrak{X}$. Then, using a standard smooth cut-off function $0 \le \chi \le 1$ with $\chi(0) = 1$ and $\operatorname{supp}\chi \subset [0, 1]$, define f_n as

$$f_n(z) = (1 - \chi(n\tau(z)))g(z) + \chi(n\tau(z))g_{\text{in}}(b(z))$$

when $z \in E \setminus \mathfrak{F}$ satisfies $\tau(z) < \infty$, and $f_n(z) = g(z)$ otherwise. We can check $f_n \in \mathcal{D}_0$ as follows. First, $f_n \in \mathcal{D}(\mathcal{L})$ since $(f_n)_{\text{in}} = g_{\text{in}} = Q_{\text{stick}}f$ and f_n is absolutely continuous along the flow:

$$\mathfrak{X}f_n(z) = n\chi'(n\tau(z))g(z) + (1 - \chi(n\tau(z)))\mathfrak{X}g - n\chi'(n\tau(z))g_{\text{in}}(b(z))$$

on $z \in E \setminus \mathfrak{F}$ such that $\tau(z) < \infty$. Second, $\mathcal{L}f_n$ is integrable since

$$\int_E \|\mathcal{L}f_n\| \, d\Pi \le \int_E \|\mathfrak{X}g\| \, d\Pi + 2n\|\chi'\|_\infty \|g\|_\infty \Pi(E \setminus \mathfrak{F}) + 2\|g\|_\infty \int_E \lambda \, d\Pi < \infty,$$

given $\|\mathfrak{X}_z\|$ is Π-integrable from [A7] and λ is Π-integrable from [A8]. Finally, $\|f_n\|_\infty \le \|g\|_\infty$ implies that the sequence $\{f_n\}$ is uniformly bounded. ■

B.4 Proof of Theorem 3.5

We will see that the essential condition is the equation of measures (21), from which the unsticking rate is obtained under the absolute continuity condition [A6]. Recall the definition of the inward and outward weighted boundary measure flux:

$$J_{\text{in/out}}(dz) = 1_{\Gamma_{\text{in/out}}}(z)|\langle v, \partial_r(x)\rangle|\pi_{\mathbb{R}^d}(x)\mathcal{H}^{d-1}_{\mathcal{M}}(dx)\nu_{E\setminus\mathfrak{F}}(x,dv).$$

Proof. Denote the subset of $\mathcal{D}(\mathcal{L})$ of bounded functions f satisfying $\mathcal{L}f \in L^1(\Pi)$ by $\mathcal{D}_0$. We prove $\int_E \mathcal{L}f d\Pi = 0$ for all $f \in \mathcal{D}_0$. This implies invariance when $\mathcal{D}_0$ separates probability measures on E (Davis, 1993, Proposition 34.7), which is indeed the case in our context (Corollary B.4). This approach, establishing $\int_E \mathcal{L}f d\Pi = 0$ for f in a measure-separating domain that is larger than the strong generator domain, is also taken in Davis (1993, Theorem 34.31) and Chevallier et al. (2024, Section 3). The latter work is closely related to our proof; we here extend it by providing a flow-wise Gauss–Green formula (Lemma B.5). This can be of independent interest since it accommodates a wide range of PDMP samplers with non-linear flows, such as the boomerang sampler and manifold PDMPs discussed in Section 2.

For every $f \in \mathcal{D}_0$, using the generator expression (7), we have a decomposition

$$\int_E \mathcal{L}f d\Pi = \omega \int_{E\setminus\mathfrak{F}} \mathcal{L}_{\mathbb{R}^d} f d\Pi_{E\setminus\mathfrak{F}} + (1-\omega)\int_{\mathfrak{F}} \mathcal{L}^{\text{lift}}_{\mathcal{M}} f d\Pi_{\mathfrak{F}} + (1-\omega)\int_{\mathfrak{F}} \Lambda(Q_{\text{unstick}}f - f)\, d\Pi_{\mathfrak{F}}. \quad (19)$$

The second term equals zero by the assumed invariance of the lifted manifold dynamics. To match the first term with the third, the Gauss–Green divergence theorem (Lang, 1995, Theorem XIII.3.2) would imply

$$\int_{E\setminus\mathfrak{F}} \mathcal{L}_{\mathbb{R}^d} f d\Pi_{E\setminus\mathfrak{F}} = \int_{\Gamma_{\text{in}}} f_{\text{in}}\, dJ_{\text{in}} - \int_{\Gamma_{\text{out}}} f dJ_{\text{out}} \quad (20)$$

where $f_{\text{in}}(z) := \lim_{t\searrow 0} f(\phi_{-t}(z))$ on $z \in \Gamma_{\text{in}}$ for the ambient flow ϕ. However, the standard result only covers the case $f|_{E\setminus\mathfrak{F}} \in C^1_c(T(\mathbb{R}^d))$. We therefore apply Lemma B.5 to extend the result to an arbitrary $f \in \mathcal{D}_0$ with compact support, and an approximation argument to the general case $f \in \mathcal{D}_0$ (Lemma B.8). Using the boundary condition $f_{\text{in}} = Q_{\text{stick}}f$, (19) now reads

$$\int_E \mathcal{L}f d\Pi = \omega \int_{\mathfrak{F}} Q_{\text{stick}} f dJ_{\text{in}} - \omega \int_{\Gamma_{\text{out}}} f dJ_{\text{out}} + (1-\omega)\int_{\mathfrak{F}} \Lambda(Q_{\text{unstick}}f - f)\, d\Pi_{\mathfrak{F}}.$$

The assumptions on the kernels ensure the following equalities:

$$\omega J_{\text{in}} Q_{\text{stick}} = (1-\omega)\Lambda\Pi_{\mathfrak{F}}, \qquad (1-\omega)(\Lambda\Pi_{\mathfrak{F}})Q_{\text{unstick}} = \omega J_{\text{out}}. \quad (21)$$

This balance condition (21) concludes (19) equals zero, and rate Λ is the Radon–Nikodym derivative

$$\Lambda(\zeta) = \frac{\omega}{1-\omega}\frac{d(J_{\text{in}}Q_{\text{stick}})}{d\Pi_{\mathfrak{F}}}(\zeta) = |\langle \eta, \partial_r(\xi)\rangle|\frac{\omega}{1-\omega}\frac{\pi_{\mathbb{R}^d}(\xi)}{\pi_{\mathcal{M}}(\xi)}. \quad (22)$$

■

B.5 Lemmas for the theorems

Behind the proof, there are two key techniques. The first is disintegration over foliations, similar to the approach in Chevallier et al. 2024, Theorem 2, where the authors prove an integration by parts formula along the linear foliation $x + \mathbb{R}v$ parametrised by $v \in \mathbb{R}$. This approach naturally aligns with the martingale approach of Davis (1984, 1993), as the domain of the extended generator is described by conditions on the foliation induced by the deterministic flow; see Davis 1984, Theorem 5.5 and Davis 1993, Theorem 26.14. Accordingly, the divergence theorem is replaced by the fundamental theorem of calculus along each one-dimensional leaf. The second key technique is related to the blow-up construction of the state space: the trivialisation of the bundle $\mathfrak{F}$ through Ψ given in (16). This connects the two foliations, one in the ambient space and the other in the manifold.

The following lemma is the core of the proof of Theorem 3.5. To compute the first term in (19), we could use the standard Gauss–Green divergence theorem and the Leibniz rule in $\mathbb{R}^d \setminus \mathcal{M}$ to derive the result if f were smooth. However, this time f is only assumed to be absolutely continuous along the flow, or we have to establish the Feller property and identify functional-analytic arguments about a core. Therefore, we consider foliating the state space along the flow ϕ_t of the vector field $\mathfrak{X}$ corresponding to $\mathcal{L}_D$ and use the fundamental theorem of calculus in each one-dimensional leaf. This approach is similar to Chevallier et al. 2024, Theorem 4, but we first construct a tubular neighbourhood and coordinate chart induced by $\mathfrak{X}$ to straighten the flow, a similar idea to the straightening-out theorem, also known as the flow-box theorem (Lang, 1995, Proposition IV.2.13). The resulting proof accommodates PDMPs with a non-linear flow such as the boomerang sampler and randomised Hamiltonian Monte Carlo.

Lemma B.5 (**flowwise Gauss–Green formula on $E \setminus \mathfrak{F}$**). Under Assumption B.2, it holds that

$$\int_{E\setminus\mathfrak{F}} \mathcal{L}_{\mathbb{R}^d} f \, d\Pi_{E\setminus\mathfrak{F}} = \int_{\Gamma_{\mathrm{in}}} f_{\mathrm{in}} \, dJ_{\mathrm{in}} - \int_{\Gamma_{\mathrm{out}}} f \, dJ_{\mathrm{out}}$$

for every bounded compactly supported Borel function that has integrability $\mathfrak{X}f \in L^1(\Pi_{E\setminus\mathfrak{F}})$, absolute continuity along the ambient flow ϕ, and limits $f_{\mathrm{in}}(z) := \lim_{t\searrow 0} f(\phi_{-t}(z))$ on $z \in \Gamma_{\mathrm{in}}$. In particular, it holds for every compactly supported $f \in \mathcal{D}_0$. □

Remark B.6 (**Key differences to existing theorems**). Similar Gauss–Green type formulae are also established in Bierkens et al. (2023a, Proposition 2.7) and Chevallier et al. (2024, Proposition 10). These two results rely on complementary sets of assumptions, in that neither set of assumptions completely subsumes the other. Among several technical conditions, Bierkens et al. (2023a, Proposition 2.7) assumes C_c^1-regularity of a test function f, while allowing a potentially curved flow of the form $\dot{x} = s(x)v$ with speed function s. By contrast, Chevallier et al. (2024, Proposition 10) only imposes a substantially weaker regularity condition, namely absolute continuity along the foliation, but it is confined to linear dynamics. Our theorem may be viewed as a common generalisation of the two, simultaneously accommodating non-linear second-order flows, including boomerang and Hamiltonian dynamics, and dealing directly with non-differentiable test functions in the domain of the extended generator. □

Proof. We mainly consider the standard Gaussian case; the non-Gaussian case can be similarly deduced by replacing $T(\mathbb{R}^d)$ by $\mathcal{V}(\mathbb{R}^d)$, dz by the density $\Omega(dz)$, and div by div_Ω.

We then have $\Pi_{E\setminus\mathfrak{F}}(dz) = m(z)dz$ for $m(z) = \pi_{\mathbb{R}^d}\gamma^d(v)$ with Gaussian density γ^d. From the ambient invariance, the left-hand side reduces to

$$\int_{E\setminus\mathfrak{F}} \mathcal{L}_{\mathbb{R}^d} f(z)m(z)\,dz = \int_{E\setminus\mathfrak{F}} (m(z)\mathcal{L}_D f(z) + f(z)\operatorname{div}(m\mathfrak{X})(z))\,dz, \tag{23}$$

by Lemma B.7. The last integrand is actually the Lie derivative $\mathcal{L}_{\mathfrak{X}}(f\Omega_m)$ for the volume form Ω_m with density m when f is smooth. Therefore we could apply the Gauss–Green divergence theorem to derive the boundary term if f were assumed to be smooth. Now that all terms in the right-hand side are local and do not involve jump terms, we proceed with the calculation locally by foliating $E \setminus \mathfrak{F}$.

There exists a countable and locally finite open covering $(U_i)_{i\in I}$ of the frozen boundary $\mathfrak{F}$, and a corresponding family of open sets $(V_i)_{i\in I}$, open in $\mathbb{R}^{2d}$, with a diffeomorphism

$$\Phi_i : (-\epsilon_i, \epsilon_i) \times U_i \to V_i \overset{\text{open}}{\subset} \mathbb{R}^{2d} \tag{24}$$

defined by $\Phi_i(t,\zeta) = \phi_t(\zeta)$ on a relatively open set U_i in $\mathfrak{F}$, since the flow $\Phi(t,\zeta) = \phi_t(\zeta)$ has a full-rank differential at each point $(0,\zeta)$ satisfying $\zeta \in \mathfrak{F}$, similar to the argument of the flow-box theorem (Lang, 1995, Lang IV.2.13). Note that, by construction, every $(x,v) \in \mathfrak{F}$ satisfies $v^\perp \neq 0$; the set $\{v = 0\}$ can be disregarded since it has Lebesgue measure zero. We first show how to pull back an integral on V_i to $(-\epsilon_i,\epsilon_i)\times U_i$.

We first calculate the Jacobian of Φ_i. On a point $(0,\zeta)$, $\zeta \in U_i$, its derivative is given by

$$D(\Phi_i)_{(0,\zeta)}(s,\alpha) = s\mathfrak{X}(\zeta) + \alpha, \qquad \alpha \in T_\zeta(\mathfrak{F}),$$

where $\alpha \in T_\zeta(\mathfrak{F})$ is identified with a point in $\alpha \in T_\zeta(E) \simeq \mathbb{R}^{2d}$. The subfibre $T_\zeta(\mathfrak{F})$ forms a $2d-1$ dimensional subspace in $T_\zeta(E)$, Therefore, the Jacobian of the linear mapping $D(\Phi_i)_{(0,\zeta)}$ is the absolute value of the component of $\mathfrak{X}$ orthogonal to $T_\zeta(\mathfrak{F})$, i.e., $|\langle\eta, \partial_r(\xi)\rangle|$, since $\mathfrak{X}$ is a second-order vector field. Since $\Phi_i(t,\zeta) = \phi_t(\Phi_i(0,\zeta))$, we obtain the Jacobian at $(t,\zeta) \in (-\epsilon_i,\epsilon_i)\times U_i$ as

$$J_{\Phi_i}(t,\zeta) = |\det D\phi_t(\zeta)||\langle\eta,\partial_r(\xi)\rangle|.$$

Given the Jacobian, we can pull back an integral on V_i through the change of variable formula. For any integrable function $F(\zeta)$, application of the Fubini theorem gives

$$\begin{aligned}\int_{V_i} F\,d\Pi_{E\setminus\mathfrak{F}} &= \int_{-\epsilon_i}^{\epsilon_i}\int_{U_i} F(\phi_t(\zeta))m(\phi_t(\zeta))J_{\Phi_i}(t,\zeta)\,dt d\zeta \\ &= \int_{U_i} |\langle\eta,\partial_r(\xi)\rangle|\,d\zeta \int_{-\epsilon_i}^{\epsilon_i} F(\phi_t(\zeta))m(\phi_t(\zeta))J_{\phi_-(\zeta)}(t)\,dt,\end{aligned}$$

where $J_{\phi_-(\zeta)}(t) := |\det D\phi_t(\zeta)|$. Observe that the family $\{\Phi_i((-\epsilon,\epsilon),\zeta)\}_{\zeta\in\mathfrak{F}}$ provides a one-dimensional foliation of V_i. Substituting $F(\zeta) = \mathcal{L}_D f(\zeta) + f(\zeta)\operatorname{div}(m\mathfrak{X})(\zeta)/m(\zeta)$, we continue the calculation of (23) as

$$= \int_{U_i} |\langle\eta,\partial_r(\xi)\rangle|\,d\zeta \int_{-\epsilon_i}^{\epsilon_i} \Big(m(\phi_t(\zeta))\mathcal{L}_D f(\phi_t(\zeta)) + f(\phi_t(\zeta))\operatorname{div}(m\mathfrak{X})(\phi_t(\zeta))\Big)J_{\phi_-(\zeta)}(t)\,dt \tag{25}$$

$$= \int_{U_i} |\langle\eta,\partial_r(\xi)\rangle|\,d\zeta \int_{-\epsilon_i}^{\epsilon_i} \frac{d}{dt}\Big(f(\phi_t(\zeta))m(\phi_t(\zeta))J_{\phi_-(\zeta)}(t)\Big)\,dt, \tag{26}$$

where we used $\mathcal{L}_D f(\phi_t(\zeta)) = \dfrac{d}{dt} f(\phi_t(\zeta))$ and that

$$\frac{d}{dt}\Big(m(\phi_t(z))|\det D\phi_t(z)|\Big) = \operatorname{div}(m\mathfrak{X})(\phi_t(z))|\det D\phi_t(z)|,$$

which is a ϕ_t pullback of the relationship $\mathcal{L}_{\mathfrak{X}}\Omega_m = \operatorname{div}(m\mathfrak{X})\Omega$, a result that follows from the Cartan formula (Lang, 1995, Proposition V.5.3) where $\Omega = dz$ is the canonical volume form (see also Bogachev, 2007, Theorem 9.10.6). Since $f(\phi_t(\zeta))$ is absolutely continuous only on $(-\epsilon, 0) \cup (0, \epsilon)$, using the fundamental theorem of calculus to the integrand $G(t,\zeta) := f(\phi_t(\zeta))m(\phi_t(\zeta))|\det D\phi_t(\zeta)|$, we have (26) equals

$$\int_{U_i} \Big(G(\epsilon,\zeta) - G(-\epsilon,\zeta) - G(0+,\zeta) + G(0-,\zeta)\Big)|\langle \eta, \partial_r(\xi)\rangle|\, d\zeta. \tag{27}$$

In the last step, we will get rid of these $G(\pm\epsilon,\zeta)$.

Finally, we get back to the global integral (23). We first extend (V_i), the open covering of $\mathfrak{F}$, to an open covering of $\operatorname{supp}(f)$, still denoted by (V_i). Through a partition of unity subordinated to (V_i), we have

$$\begin{aligned}
&\int_{E\setminus\mathfrak{F}} \Big(m(z)\mathcal{L}_D f(z) + f(z)\operatorname{div}(m\mathfrak{X})(z)\Big)\, dz \\
&= \sum_{i\in I} \int_{V_i} \Big(m(z)\mathcal{L}_D(\chi_i f)(z) + \chi_i f(z)\operatorname{div}(m\mathfrak{X})(z)\Big)\, dz \\
&= \sum_{i\in I} \int_{U_i} \Big(\chi_i(\phi_{0-}(\zeta))f(\phi_{0-}(\zeta)) - \chi_i(\phi_{0+}(\zeta))f(\phi_{0+}(\zeta))\Big) m(\zeta)|\langle \eta, \partial_r(\xi)\rangle|\, d\zeta \\
&= \int_{\Gamma_{\text{in}}} f_{\text{in}}(\zeta)\, m(\zeta)|\langle \eta, \partial_r(\xi)\rangle|\, d\zeta - \int_{\Gamma_{\text{out}}} f(\zeta)\, m(\zeta)|\langle \eta, \partial_r(\xi)\rangle|\, d\zeta
\end{aligned}$$

■

Lemma B.7 (Implication of the ambient invariance). Define $m(z) := \pi_{\mathbb{R}^d}(x)\nu_x^d(v)$. Then, under Assumption B.2,

$$\int_{T(\mathbb{R}^d)} \mathcal{L}_I f(z) m(z)\, dz = \int_{T(\mathbb{R}^d)} f(z)\operatorname{div}(m\mathfrak{X})(z)\, dz,$$

for every bounded, compactly supported Borel measurable function f. □

Proof. This can be directly checked for many standard PDMPs, including randomised Hamiltonian Monte Carlo, bouncy particle sampler, Zig-Zag sampler, boomerang sampler, etc. Nonetheless, we provide a proof in this generic setting. Before considering an arbitrary bounded compactly supported Borel function f, we first show the relationship for differentiable $f \in C_c^1(T(\mathbb{R}^d))$. The regularity assumptions of [A1] and [A5] enable us to use the Gauss divergence theorem (Lang, 1995, Theorem XIII.3.2) as follows

$$\int_{T(\mathbb{R}^d)} \mathcal{L}_I f m\, d\Omega = -\int_{T(\mathbb{R}^d)} \mathcal{L}_D f m\, d\Omega = \int_{T(\mathbb{R}^d)} f \operatorname{div}_\Omega(m\mathfrak{X})\, d\Omega,$$

where we used the assumed ambient invariance in the first equality. To extend the result for any bounded measurable function f, we first realise that both the leftmost and rightmost sides define the same linear functional on $C_c^1(T(\mathbb{R}^d))$. This equality first extends to $C_c(T(\mathbb{R}^d))$ by a standard approximation argument under the integrability assumption of [A8]. Then, they have to coincide as a Borel measure by the uniqueness of the Riesz representation (Bogachev, 2007, Theorem 7.11.3). ■

Lemma B.8 (Approximation argument for the flowwise Gauss–Green formula). Assume that, under Assumption B.2,

$$\int_{E\setminus\mathfrak{F}} \mathcal{L}_{\mathbb{R}^d} f \, d\Pi_{E\setminus\mathfrak{F}} = \int_{\Gamma_{\text{in}}} f_{\text{in}} \, dJ_{\text{in}} - \int_{\Gamma_{\text{out}}} f \, dJ_{\text{out}} \tag{28}$$

has been proven for every bounded, compactly supported Borel function f satisfying the condition in Lemma B.5. Further suppose the flux balance condition (10)

$$J_{\text{in}} Q_{\text{stick}} Q_{\text{unstick}} = J_{\text{out}}, \qquad \omega J_{\text{in}} Q_{\text{stick}} = (1-\omega)\Lambda \Pi_{\mathfrak{F}}.$$

Then, the same identity holds for every $f \in \mathcal{D}_0$. □

Proof. Take a sequence of non-negative cutoff functions $\{\rho_n\} \subset C_c^\infty(E)$ such that $\rho_n \nearrow 1$ a.s., for example,

$$\rho_n(x,v) := \rho\left(\frac{\|\beta(x)\|^2 + \|v\|^2}{n^2}\right)$$

for a smooth decreasing function $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\rho([0,1]) = \{1\}$ and $\rho([4,\infty)) = \{0\}$. For any $f \in \mathcal{D}_0$, we can then check that $f_n := \rho_n f$ satisfies the required conditions of Lemma B.5. f_n also has compact support since

$$\operatorname{supp} \rho_n \subset \{(x,v) \in E \mid \|\beta(x)\|^2 + \|v\|^2 \le 4n^2\},$$

where the right-hand side is $\overline{\beta}^{-1}(B_{2n})$ for the lifted bundle morphism $\overline{\beta}$ and closed ball B_{2n} with radius $2n$; this is a compact set because $\mathcal{M} \subset \mathbb{R}^d$ is assumed to be closed. We have, for the decomposition of the ambient generator $\mathcal{L}_{\mathbb{R}^d} = \mathcal{L}_D + \mathcal{L}_J$ and $\mathcal{L}_J f = \lambda(Qf - f)$,

$$\mathcal{L}_{\mathbb{R}^d} f_n = \rho_n \mathcal{L}_{\mathbb{R}^d} f + f\mathcal{L}_D \rho_n + \int_E f(z')(\rho_n(z') - \rho_n(z))\lambda(z) Q(z, dz'). \tag{29}$$

The convergence of the left-hand side in (28) $\int_{E\setminus\mathfrak{F}} \mathcal{L}_{\mathbb{R}^d} f_n \, d\Pi_{E\setminus\mathfrak{F}} \to \int_{E\setminus\mathfrak{F}} \mathcal{L}_{\mathbb{R}^d} f \, d\Pi_{E\setminus\mathfrak{F}}$ can be established term by term:

The first term Since $f \in \mathcal{D}_0$, it satisfies $\mathcal{L}f \in L^1(\Pi)$. Therefore, by the Lebesgue dominated convergence theorem,

$$\int_E \rho_n \mathcal{L} f \, d\Pi \xrightarrow{n\to\infty} \int_E \mathcal{L} f \, d\Pi.$$

The second term Since $\mathfrak{X}$ is a second-order vector field [A7],

$$\begin{aligned} |\mathcal{L}_D \rho_n| &= \frac{2}{n^2}\rho'\left(\frac{\|\beta(x)\|^2 + \|v\|^2}{n^2}\right)\Big(\langle x, v\rangle + \langle v, F(x,v)\rangle\Big) \\ &\le \frac{2}{n^2} C\Big(\|v\| 2n + \|F(x,v)\| 2n\Big) = \frac{C}{n}(\|v\| + \|F(x,v)\|), \end{aligned}$$

where $C := \max_{x\in[0,4]} |\rho'(x)|$. Under [A7], the integral can be evaluated as follows:

$$\begin{aligned} \left|\int_E f\mathcal{L}_D\rho_n \, d\Pi\right| &\le \|f\|_\infty \int_E |\mathcal{L}_D \rho_n| \, d\Pi \\ &\le \|f\|_\infty \int_E \frac{C}{n}(\|v\| + \|F(z)\|) \, \Pi(dz) \xrightarrow{n\to\infty} 0. \end{aligned}$$

The third term The Π-integral of the third term converges to 0 by the Lebesgue dominated convergence theorem under the Π-integrability of the intensity λ [A8].

The finiteness of J_{in} and J_{out} follows from the flux balance condition combined with [A8]:

$$J_{\text{in}}(\Gamma_{\text{in}}) = J_{\text{out}}(\Gamma_{\text{out}}) = \frac{1-\omega}{\omega} \int_{\mathfrak{F}} \Lambda \, d\Pi_{\mathfrak{F}} < \infty.$$

We therefore conclude similarly the convergence of the right-hand side in (28)

$$\int_{\Gamma_{\text{in}}} (f_n)_{\text{in}} \, dJ_{\text{in}} - \int_{\Gamma_{\text{out}}} f_n \, dJ_{\text{out}} \xrightarrow{n\to\infty} \int_{\Gamma_{\text{in}}} f_{\text{in}} \, dJ_{\text{in}} - \int_{\Gamma_{\text{out}}} f \, dJ_{\text{out}}$$

■

B.6 Proofs of corollaries

Proof of Remark 3.7. The assumed balance conditions imply

$$\omega J_{\text{in}} Q_{\text{stick}} = (1-\omega)\Lambda\Pi_{\mathfrak{F}}, \qquad (1-\omega)(\Lambda\Pi_{\mathfrak{F}})Q_{\text{unstick}} = \omega J_{\text{out}}.$$

The proof of Theorem 3.5 uses velocity compatibility only to establish these two identities. The same Gauss–Green and approximation arguments therefore apply and establish invariance of Π. ■

Proof of Corollary 3.8. Define the weighted target on the blown-up front face by $\Pi^{\text{RB}}_{\mathfrak{F}}(dz) = \frac{\|v^{\perp}\|}{\sqrt{2/\pi}}\Pi_{\mathfrak{F}}(dz)$ and on the entire space by $\Pi^{\text{RB}} = \omega\Pi_{E\setminus\mathfrak{F}} + (1-\omega)\Pi^{\text{RB}}_{\mathfrak{F}}$. Since the lifted manifold dynamics preserves $\|v^{\perp}\|$, invariance of $\Pi_{\mathfrak{F}}$ therefore implies invariance of $\Pi^{\text{RB}}_{\mathfrak{F}}$. Moreover, the velocity compatibility condition and the standard Gaussian ambient velocity law give

$$\int \|v^{\perp}\| \, \nu_{F,x}(dv) = \sqrt{\frac{2}{\pi}},$$

so this re-weighting preserves the position marginal. Finally,

$$\Lambda_{\text{RB}}\Pi^{\text{RB}}_{\mathfrak{F}} = c\|v^{\perp}\|\Pi_{\mathfrak{F}} = \Lambda\Pi_{\mathfrak{F}}.$$

Consequently, the original flux identities give

$$\omega J_{\text{in}} Q_{\text{stick}} = (1-\omega)\Lambda_{\text{RB}}\Pi^{\text{RB}}_{\mathfrak{F}}, \qquad (1-\omega)\Lambda_{\text{RB}}\Pi^{\text{RB}}_{\mathfrak{F}} Q_{\text{unstick}} = \omega J_{\text{out}}.$$

The general flux-balance result 3.7 establishes invariance of Π^{RB}. ■

Proof of Corollary 3.9. We augment the frozen boundary with an age variable $a \in [0, T)$:

$$\hat{\mathfrak{F}} = \mathfrak{F} \times [0, T), \qquad \hat{E} = (E \setminus \mathfrak{F}) \sqcup \hat{\mathfrak{F}}.$$

We lift the manifold dynamics to $\hat{\mathfrak{F}}$ by increasing the age at unit speed. Sticking initialises the age at zero, while reaching $a = T$ triggers Q_{unstick}. Define the lifted target by

$$\widehat{\Pi} = \omega\Pi_{E\setminus\mathfrak{F}} + (1-\omega)\Pi^{\text{RB}}_{\mathfrak{F}} \otimes \frac{da}{T}, \qquad 0 \le a < T,$$

whose projection onto E is Π_{RB}. Let $\widehat{\mathcal{L}}$ be the extended generator. Writing f_0 for the value at age zero and f_T for the limit at $a = T$, $f \in \mathcal{D}(\widehat{\mathcal{L}})$ satisfies the boundary conditions

$$f^{\mathrm{in}} = Q_{\mathrm{stick}} f_0, \qquad f_T = Q_{\mathrm{unstick}} f.$$

The class

$$\widehat{D}_0 = \{f \in D(\widehat{\mathcal{L}}) : f \text{ is bounded}, \quad \widehat{\mathcal{L}} f \in L^1(\widehat{\Pi})\}$$

separates probability measures by the same argument as Corollary B.4. The estimates of that proof also apply in the product coordinates (17) with the additional unit age drift since the manifold dynamics are independent of the conserved normal coordinate, whose Gaussian law is now weighted by $\|v^\perp\|$.

For $f \in \widehat{D}_0$, the ambient flow-wise Gauss–Green formula B.5 and the associated approximation argument give

$$\omega \int_{E \setminus \mathfrak{F}} \widehat{\mathcal{L}} f \, d\Pi_{E \setminus \mathfrak{F}} = \omega \int_{\mathfrak{F}} f_0 \, d(J_{\mathrm{in}} Q_{\mathrm{stick}}) - \omega \int_{\Gamma_{\mathrm{out}}} f \, dJ_{\mathrm{out}}.$$

On the augmented frozen boundary, applying Dynkin's formula to the lifted manifold process and integrating it against its invariant measure $\Pi^{\mathrm{RB}}_{\mathfrak{F}}$ over $a \in [0, T]$ yield

$$\int_0^T \int_{\mathfrak{F}} \widehat{\mathcal{L}} f(z, a) \, \Pi^{\mathrm{RB}}_{\mathfrak{F}}(dz) \, da = \int_{\mathfrak{F}} (f_T - f_0) \, d\Pi^{\mathrm{RB}}_{\mathfrak{F}}.$$

Boundedness of f and integrability of $\widehat{\mathcal{L}} f$ justify localization and passage to the terminal value f_T. By Corollary 3.8, the flux identities read

$$\omega J_{\mathrm{in}} Q_{\mathrm{stick}} = \frac{1-\omega}{T} \Pi^{\mathrm{RB}}_{\mathfrak{F}}, \qquad \frac{1-\omega}{T} \Pi^{\mathrm{RB}}_{\mathfrak{F}} Q_{\mathrm{unstick}} = \omega J_{\mathrm{out}}.$$

Adding the two component contributions and using the boundary conditions therefore gives

$$\int_{\widehat{E}} \widehat{\mathcal{L}} f \, d\widehat{\Pi} = 0, \qquad f \in \widehat{D}_0.$$

The same extended-generator criterion as in Theorem 3.5 proves invariance of $\widehat{\Pi}$. ∎

Proof of Corollary 3.10. For every $x \in \mathcal{M}$, let $\mathcal{V}_x^+ := \{v \in \mathcal{V}_x \mid \langle v, \partial_r(x) \rangle < 0\}$ denote the half-fibre of velocities directed inwards (towards $\mathcal{M}$). The measures ν_x^{in} and ν_x^{out} are now the restrictions of the ambient velocity law to the incoming and outgoing half-fibres, respectively, having total mass $1/2$ respectively. Since both kernels preserve the position and $\|v^\perp\|$, the velocity compatibility conditions imply

$$(\|v^\perp\| \nu_x^{\mathrm{in}}) Q_{\mathrm{stick},x} = \frac{1}{2} \|v^\perp\| \nu_{\mathfrak{F},x}, \qquad \left(\frac{1}{2} \|v^\perp\| \nu_{\mathfrak{F},x} \right) Q_{\mathrm{unstick},x} = \|v^\perp\| \nu_x^{\mathrm{out}}.$$

Multiplying by $\pi_{\mathbb{R}^d}(x) \mathcal{H}^{d-1}_{\mathcal{M}}(dx)$ therefore gives

$$(J_{\mathrm{in}} Q_{\mathrm{stick}})(dx, dv) = \frac{1}{2} \|v^\perp\| \pi_{\mathbb{R}^d}(x) \mathcal{H}^{d-1}_{\mathcal{M}}(dx) \nu_{\mathfrak{F},x}(dv), \qquad J_{\mathrm{in}} Q_{\mathrm{stick}} Q_{\mathrm{unstick}} = J_{\mathrm{out}}.$$

Consequently, the rate

$$\Lambda(x, v) = \frac{\omega}{2(1-\omega)} \|v^\perp\| \frac{\pi_{\mathbb{R}^d}(x)}{\pi_{\mathcal{M}}(x)}$$

satisfies

$$\omega J_{\mathrm{in}} Q_{\mathrm{stick}} = (1-\omega) \Lambda \Pi_{\mathfrak{F}}.$$

Thus both flux-balance conditions in Remark 3.7 hold, and the claimed invariance follows. ∎

We have an alternative approach to Corollary 3.10, in which we formally reduce the one-sided case to the two-sided case. We artificially consider the two faces $\mathcal{M}^+$ and $\mathcal{M}^-$, where one of the faces will never be hit directly from U. In this approach, we can retain the rate (2) for unsticking, but after the first sticking event the process is interpreted to restick to the other face of the submanifold. Equivalently, the sampler returns to U at the second arrival of the inhomogeneous Poisson process with rate (2).

C Experimental details

The samplers were implemented in Julia; experiments used Julia 1.13.3. The CPU-budget comparisons in Sections 4.3 and 4.4 used an Apple M5 processor (10 CPU cores, 64-bit ARM) running macOS 26.6.2. Each timed sampling process used one Julia sampling thread and one BLAS thread. The code, experimental settings, and reproduction instructions are available at .

C.1 Gaussian mixture model (Section 1.2)

In Section 1.2, we discussed a two-component Gaussian mixture model that gives rise to the submanifold $\mathcal{M}_{\pi/4}$ using the notation in Section C.2.

C.1.1 Reversible-jump proposals

The reversible-jump proposals are given by the following kernels:

$$P_{\text{split}}\left(\begin{pmatrix}\mu\\ \mu\end{pmatrix}, -\right) = N_2\left(\begin{pmatrix}\mu\\ \mu\end{pmatrix}, \frac{1}{2}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\right),$$

$$P_{\text{merge}}\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, -\right) = \delta_{\left(\frac{\mu_1+\mu_2}{2}, \frac{\mu_1+\mu_2}{2}\right)}.$$

In words, the split kernel proposes a displacement of law $N(0,1)$ in the normal direction to $\mathcal{M}_{\pi/4}$, while the merge kernel deterministically projects the sampler to the foot of the perpendicular. Through a coordinate change $(t,n)^\top = (\frac{\mu_1+\mu_2}{\sqrt{2}}, \frac{\mu_1-\mu_2}{\sqrt{2}})^\top$, the detailed balance condition between the two points $(t,0)$ and (t,n) requires

$$(1-\omega)\pi_{\mathcal{M}}(t)L(t,0)\,dt\,P_{\text{split}}(t,dn)\alpha_{\text{split}}(t,n) = \omega\pi_{\mathbb{R}^2}(t,n)L(t,n)\,dn\,P_{\text{merge}}((t,n),dt)\alpha_{\text{merge}}(t,n)$$

$$\therefore \qquad (1-\omega)L(t,0)\phi(n)\alpha_{\text{split}}(t,n) = \omega\phi(n)L(t,n)\alpha_{\text{merge}}(t,n)$$

to hold, where α_{split} and α_{merge} are acceptance probabilities, $L(t,n)$ the likelihood, and $\phi(x)$ denotes the density of the standard Gaussian distribution. Observe that the proposal density and the prior distribution match and cancel out. A natural choice is then

$$\alpha_{\text{split}}(n) = \frac{\omega}{1-\omega}\frac{L(t,n)}{L(t,0)} \wedge 1, \quad \alpha_{\text{merge}}(n) = \frac{1-\omega}{\omega}\frac{L(t,0)}{L(t,n)} \wedge 1.$$

The resulting RJ-MCMC makes proposals using these kernels after every random walk Metropolis step, which is $N_2(0, I_2/4)$ in the ambient space and $N_1(0, 1/4)$ in the manifold chart μ.

C.1.2 Algorithm design

The SM-FECMC uses the uniform velocity distribution and the identity map on sticking and unsticking just as in Example 3.1. However, the ambient velocity is the uniform distribution on the unit hypersphere $\mathrm{Unif}(\mathbb{S}^1)$. This still satisfies the velocity compatibility condition (8) since the FECMC within $\mathcal{M}_{\pi/4}$ leaves the pushforward of $\mathrm{Unif}(\mathbb{S}^1)$ on $T(\mathcal{M}_{\pi/4})$ invariant. We do not perform any orthogonal switch within the FECMC algorithm since the problem is two dimensional.

The reversible-jump FECMC uses the same reversible-jump proposal kernels and acceptance probability as the reversible-jump random walk Metropolis–Hastings. The velocity is retained after splitting to keep the momentum persistent.

C.1.3 Prior specification and the thawing rate

The trajectories in Figure 2 were generated from the following prior specification:

$$\omega = 0.8, \quad \pi_{\mathbb{R}^2}(\mu) = \frac{1}{2\pi} e^{-\frac{\|\mu\|^2}{2}}, \quad \pi_{\mathcal{M}_{\pi/4}}(\mu_1) = \frac{1}{\sqrt{2\pi}} e^{-\mu_1^2}.$$

Note that for the coordinate chart $\varphi : \mu \mapsto \mu_1$ of $\mathcal{M}_{\pi/4}$, the Hausdorff measure is $\varphi_* \mathcal{H}^1_{\mathcal{M}_{\pi/4}}(d\mu) = \sqrt{2}\, d\mu_1$. Thus, the pushforward of the manifold measure is

$$\varphi_* \pi_{\mathcal{M}_{\pi/4}} \mathcal{H}^1_{\mathcal{M}_{\pi/4}} = \frac{1}{\sqrt{\pi}} e^{-\mu_1^2} d\mu_1,$$

namely $N_1(0, 1/2)$. The rate now reads

$$\Lambda(x) = \|v_\sigma^\perp\| \frac{\omega}{1-\omega} \frac{\pi_{\mathbb{R}^2}(x)}{\pi_{\mathcal{M}_{\pi/4}}(x)} = \|v_\sigma^\perp\| \frac{\omega}{1-\omega} \frac{1}{\sqrt{2\pi}}.$$

C.1.4 Data generation

We generated $n = 10$ data points from

$$Y_i \overset{\text{i.i.d.}}{\sim} \frac{1}{2} N(0.1, 1) + \frac{1}{2} N(-0.1, 1).$$

The likelihood is thus

$$L = \prod_{i=1}^{10} \left(\tfrac{1}{2} \phi(y_i - \mu_1) + \tfrac{1}{2} \phi(y_i - \mu_2) \right).$$

C.2 Tilting linear subspaces

C.2.1 Main results

Consider a one-dimensional linear subspace in $\mathbb{R}^2$ with angle θ, $\mathcal{M}_\theta := \{x_2 = x_1 \tan\theta\}$, and a target

$$\pi_\theta(dx) = \omega \gamma^2(x)\, dx + (1-\omega) \gamma^1(x) \mathcal{H}^1_{\mathcal{M}_\theta}(dx),$$

where γ^d denotes the d-dimensional standard Gaussian distribution. The unsticking rate (2) is $\Lambda_\theta(x) = \dfrac{\|v_\sigma^\perp\|}{\sqrt{2\pi}} \dfrac{\omega}{1-\omega}$ (Example 3.3). When $\theta = 0$, this coincides with the coordinate-aligned sticky zig-zag rate (Bierkens et al., 2023b, Equation (1.3)). To validate our theory,

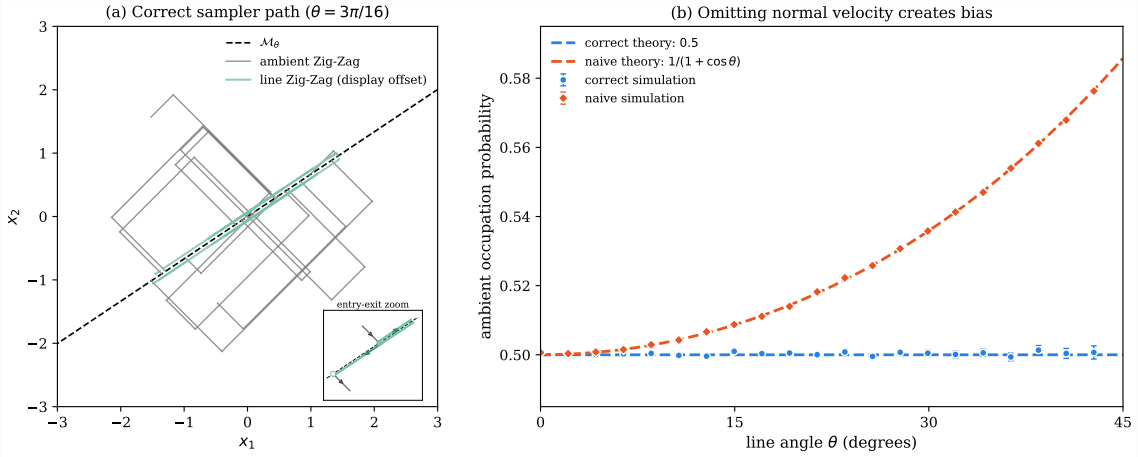


Figure 12: (a) A trajectory of the sticky zig-zag sampler with $\theta = 3\pi/16$. The dynamics on $\mathcal{M}$ are displaced slightly in the normal direction for visibility. The inset represents one sticking event and one unsticking event. (b) Mean estimates of the ambient mass $\omega = 1/2$ over 100 independent runs, with error bars of ± 1.96 Monte Carlo standard errors computed over the replicates. The 21 simulated angles are equally spaced in $[0, 0.95\pi/4]$.

we compare our algorithm with a naïve sampler that retains the rate of Bierkens et al. (2023b, Equation (1.3)) at every angle $\theta \in [0, \pi/4)$, omitting the velocity factor $\|v_\sigma^\perp\|$. We set $\omega = 1/2$ and employ the zig-zag dynamics in both the ambient space and the manifold. Figure 12 (b) shows that the two samplers coincide at $\theta = 0$, while the bias in the naïve sampler increases as θ approaches $\pi/4$. The sticky manifold rate leads to unbiased estimates, illustrating the validity and necessity of the correction factor $\|v^\perp\|$ implied by our theory.

C.2.2 Algorithm design

The SM-Zigzag algorithm follows Example 3.3. Each run starts in the ambient space, with independent position and velocity drawn from γ^2 and $\mathrm{Unif}(\{\pm 1\}^2)$, respectively. After a burn-in of 5000 time units, we estimate ω by the fraction of continuous time spent in the ambient space over a further 50,000 time units. Event times are simulated analytically (Bouchard-Côté et al., 2018, Example 1).

C.2.3 Remarks

Analytic expression of the bias

There are only four velocity patterns $v \in \{\pm 1\}^2$ in the two-dimensional zig-zag sampler. Only two patterns of them produce manifold hitting, whose norms are $\cos\theta + \sin\theta$ and $\cos\theta - \sin\theta$. As the ambient velocity distribution assigns them with equal probability, its average is $\mathrm{E}[|V^\perp|] = \cos\theta$. The correct rate (2) is thus $\cos\theta$ times the naïve rate on average. For the naïve sampler, the detailed balance condition on the marginal model indicator process is

$$p_\theta \int_{\partial(E\setminus\mathfrak{F})} \gamma^2(x)\|v^\perp\| \, \mathcal{H}^1_{\mathcal{M}_\theta}(x)\nu_{E\setminus\mathfrak{F}}(dv) = (1-p_\theta)\int_{\mathfrak{F}} \Lambda(x)\gamma^1(x)\, \mathcal{H}^1_{\mathcal{M}_\theta}\nu_{\mathfrak{F}}(dv)$$

for the state $(p_\theta, 1-p_\theta)$. This suggests the invariant distribution is

$$p_\theta = \frac{\omega}{\omega + (1-\omega)\cos\theta}.$$

What happens at $\theta = \pi/4$

This edge case is qualitatively different in that the velocity compatibility condition (8) no longer holds in its current form, because two of the velocity patterns, $(1,1)$ and $(-1,-1)$, are tangent to the line and therefore $\nu_x^{\text{in}}(\Gamma_{\text{in},x}) = 1/2$. Based on the general result (Remark 3.7) and

$$\nu_x^{\text{in}} Q_{\text{stick},x} = \frac{1}{2}\nu_{\mathfrak{F},x},$$

the rate should be

$$\Lambda = \frac{1}{2}\|v_\sigma^\perp\|\frac{\omega}{1-\omega}\frac{\pi_{\mathbb{R}^d}}{\pi_{\mathcal{M}}}.$$

This is essentially because, half of the velocity patterns, $(\pm 1, \pm 1)$, do not produce boundary hitting in $\theta = \pi/4$, which reduces the amount of the incoming probability flow J_{in}. The rate therefore has to be half to account for this with prolonged manifold sticking time.

C.3 Concentric sphere example (Section 4.2)

C.3.1 Algorithm design

For all four SM-PDMP algorithms, the within-manifold dynamics are identical: geodesic flow with global velocity refreshment at rate $\lambda = 1$, which can be seen as the geodesic bouncy particle sampler applied to the uniform distribution (Yang et al., 2024; Bell et al., 2025). The three ambient dynamics also perform global refreshment at rate 1. At a sphere hit, write $n = x/R$ and decompose the incoming velocity as $v = u + an$, retaining the tangent velocity u and signed normal component a. On the sphere, only the tangential velocity is refreshed, while FECMC + normal refreshment also redraws the ambient normal velocity at each sphere refresh. The tabulated runs select each initial stratum with probability $1/2$ and draw positions and auxiliary velocities from its conditional stationary distribution.

C.3.2 Additional results on Rao–Blackwellisation

Without normal refreshment or Rao–Blackwellisation, the normal speed $\|v^\perp\|$ at sphere entry has the flux-weighted density

$$f(\|v^\perp\|) = \|v^\perp\| \exp\left(-\frac{\|v^\perp\|^2}{2}\right).$$

under stationarity (Baccelli and Brémaud, 2003, Remark 1.9.4). The sphere sticking time satisfies $\tau \,|\, \|v^\perp\| \sim \text{Exp}(\chi_d(R)\|v^\perp\|)$, thereby giving $\text{Var}[\tau] = \infty$. This variance explosion is not confined to this particular setting but rather stems from the fact that the sticking time distribution is a scale mixture of the exponential distribution. Normal refreshment

makes this variance finite while preserving the mean sticking time and stationary transition intensity:

$$\mathrm{E}[\tau] = \frac{1}{\chi_d(R)}\sqrt{\frac{\pi}{2}}, \qquad \mathrm{E}[J_T] = T\chi_d(R)\sqrt{\frac{2}{\pi}},$$

where J_T counts ambient-sphere transitions in both directions during a stationary run of duration T. Thus, normal refreshment changes higher order moments while preserving the mean transition count.

Figure 13 shows the sticking-time distributions for standard SM-FECMC and SM-FECMC with normal refreshment. We observe that the heavy tails of long manifold sojourns disappear in the normally refreshed versions. Based on our experience, we repeatedly observe situations where small $\|v^{\perp}\|$ causes extremely long manifold stays that exhaust the given computational budget. This phenomenon has typically been observed when we look into very biased runs, causing an inflated RMSE. Shorter manifold stays are also shrunk towards the mean, while this effect is more subtle and less evident in the plots. Extremely short stays are retained because no normal velocity refreshment takes place during them.

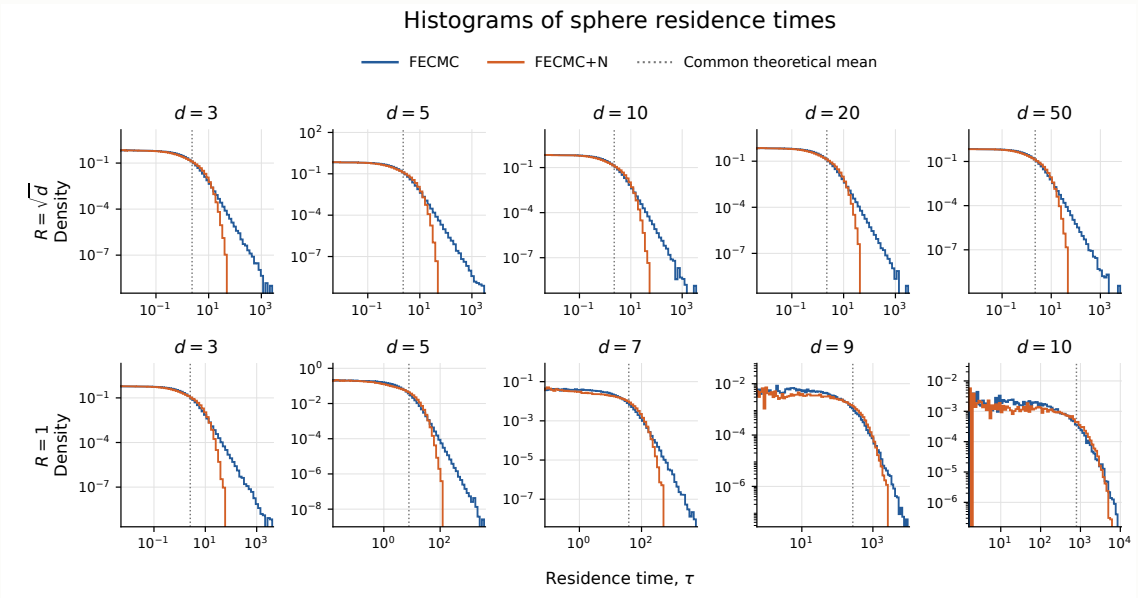


Figure 13: Histograms of sphere residence times for FECMC (blue) and FECMC+N (orange). Top: $R = \sqrt{d}$ with $d = 3, 5, 10, 20, 50$; bottom: $R = 1$ with $d = 3, 5, 7, 9, 10$. Each condition uses 1000 independent stationary-start runs of duration $T = 10^4$. Both axes are logarithmic. Dotted vertical lines indicate the theoretical mean, common to both samplers.

A notable property in the autocorrelation function is observed for the Rao–Blackwellised version with fixed sticking times (Figure 14).

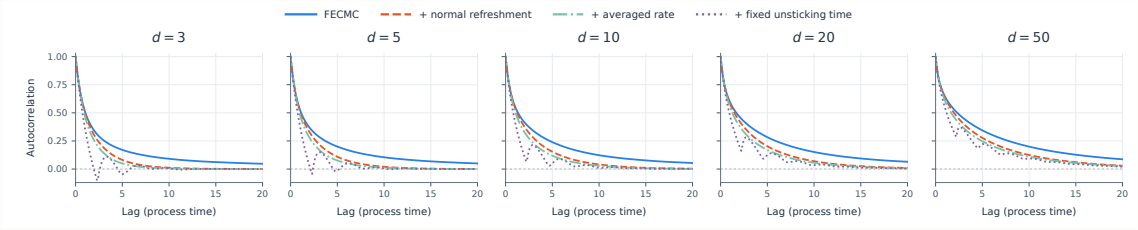


Figure 14: Autocorrelation of the model-indicator process for the four FECMC variants as in Table 1, with $R = \sqrt{d}$. Curves show averages over 1000 stationary-start runs of duration $T = 10^4$.

C.4 Quantum state tomography example (Section 4.3)

C.4.1 Algorithm design

FECMC uses straight flows in the ball and geodesic flows on the sphere, without global velocity refreshment nor orthogonal switch. The radial RJ proposal uses $R \sim \text{Beta}(6, 1)$. Within RJ-FECMC, upon unsticking from $s \in \mathbb{S}^2$, the sampler reconstructs $v = u + \epsilon|v^\perp|s$, with a uniformly chosen sign $\epsilon \in \{-1, +1\}$. RJ-HMC uses the identity mass matrix, specular boundary reflections within the ball (Neal, 2011, Section 5.5.1.5), and geodesic HMC on the sphere (Byrne and Girolami, 2013). Thinning uses locally constructed analytic bounds. The CPU time comparison includes only trajectory simulation and recording, while excluding observable integration. FECMC estimates use continuous-time averages evaluated by analytic integration along trajectories. Vector RMSEs use the usual Euclidean norm.

C.4.2 Tuning of HMC

For HMC, we conducted pilot runs with identity mass matrices. For each scenario and each integration step $T \in \{1, 2, 3, 4, 5, 10\}$, we selected a common step size ϵ used across B^3 and $\mathbb{S}^2$ by an adaptive grid search. The objective was the minimum of the ball and sphere ESSs of the three Bloch coordinates per CPU time. The initial grid comprised 27 values from 0.001 to 0.8. The validation point estimates were highest at $T = 3$ in all scenarios (Table 5), with selected step sizes ranging from 0.079 to 0.091.

Table 5: HMC tuning for each scenario and integration-step count T. S is the smallest ESS per CPU second over three coordinates on either $\mathbb{S}^2$ or B^3, displayed in units of 10^5.

	$\alpha = 1.00$		$\alpha = 0.99$	
T	$\hat{\epsilon}$	$S/10^5$	$\hat{\epsilon}$	$S/10^5$
1	0.18250	3.97	0.19750	4.91
2	0.12500	5.57	0.13000	8.20
3	0.07875	5.64	0.08625	8.89
4	0.05875	5.45	0.06625	8.38
5	0.05000	5.01	0.05625	7.54
10	0.02500	3.71	0.02750	5.92

We here note that HMC is sensitive to this step-size tuning and it may be difficult to find the optimal step size with available quantities such as ESS.

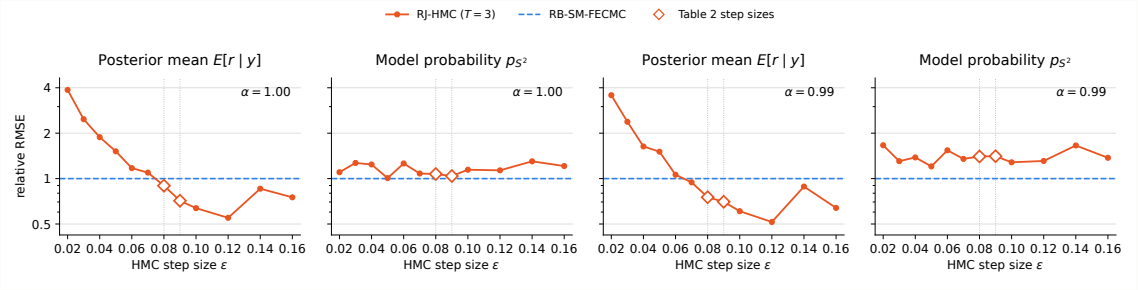


Figure 15: Step-size sensitivity of RJ-HMC with $T = 3$ integration steps. $\epsilon = 0.08, 0.09$ used in Table 2 were suggested from pilot runs based on ESS per CPU time, which took approximately one hour per each setting; see Appendix C.4.2 for details. $\epsilon = 0.12$ that yields the smallest RMSE in the plot was not selected in the pilot runs.

C.4.3 Prior specification and the thawing rate

Since the uniform ball and sphere densities are $3/(4\pi)$ and $1/(4\pi)$, the one-sided unsticking rate is

$$\frac{1}{2}\|v^{\perp}\|\frac{3/4\pi}{1/4\pi} = \|v^{\perp}\|\frac{3}{2},$$

where the factor $1/2$ arises from the one-sidedness of the boundary (Corollary 3.10). The Rao–Blackwellised sticking time is therefore $\sqrt{2\pi}/3$.

C.4.4 Initialisation and RMSE computation

Initial states are generated by conditional posterior rejection sampling with model probabilities computed by numerical quadrature, which also supplies the reference values for RMSE. This uses rejection sampling with global quadratic likelihood envelopes. Reference values are computed by numerical quadrature in spherical coordinates. Numerical stability is assessed by increasing the quadrature orders; the resulting changes are negligible relative to the reported Monte Carlo RMSEs.

C.5 Bayesian PCA example (Section 4.4)

C.5.1 Prior specification and length-scale differences

We use the spectral lift of Example 3.15 with $d = 20$. Relative to $dr\,\mathcal{H}(dQ)$, where $\mathcal{H}$ is the Hausdorff measure derived from the Frobenius metric (or equivalently normalized Haar measure) on $O(20)$, the prior on the rank-k stratum has density

$$f_k(r) = Z_k^{-1}\exp\left(-\frac{k+1}{2}\sum_{i=1}^{k} r_i\right)\prod_{1\le i<j\le k}(r_i - r_j), \qquad Z_k = \frac{2^k\prod_{j=0}^{k-1} j!}{(k+1)^{k(k+1)/2}}, \tag{30}$$

with $f_0 = Z_0 = 1$ (Muirhead, 1982, Corollary 3.2.19 p.107). The Vandermonde determinant factor introduces divergence in the gradient and the Hessian of the potential

$U_k := -\log f_k$ along the order constraints of C_k. This makes the boundaries from the order constraints inaccessible for the PDMP and the exact Hamiltonian dynamics.

Remark C.1 . (Pourzanjani et al., 2021, Section 5.3) consider a related parametrisation to the spectral lift, while defining a uniform (improper) prior on the restricted (unbounded) region C_k. This yields a factor of inverse Vandermonde determinant in the original matrix space, producing divergences of the density near regions with overlapping eigenvalues. □

The left panel in Figure 16 shows how the typical shortest length scale $\ell_{\min} := 1/\sqrt{\lambda_{\max}(\nabla^2 U)}$ of the posterior potential U differs within the main stratum of each scenario, while the right panel similarly shows that of the condition number $\kappa_2 := \lambda_{\max}(\nabla^2 U)/\lambda_{\min}(\nabla^2 U)$. From the eight reference RB-SM-FECMC chains (Section C.5.6), the medians of the nearest and farthest 10% of the discretised samples to the order constraint boundaries are plotted. As a result, the average acceptance probability of HMC proposals on typical points varies within a stratum (Table 6). Each entry in Table 6 averages estimates at eight fixed positions, randomly selected from each chain in the nearest or farthest area. At each position, rejection probability is estimated by averaging it over 128 Gaussian momentum draws.

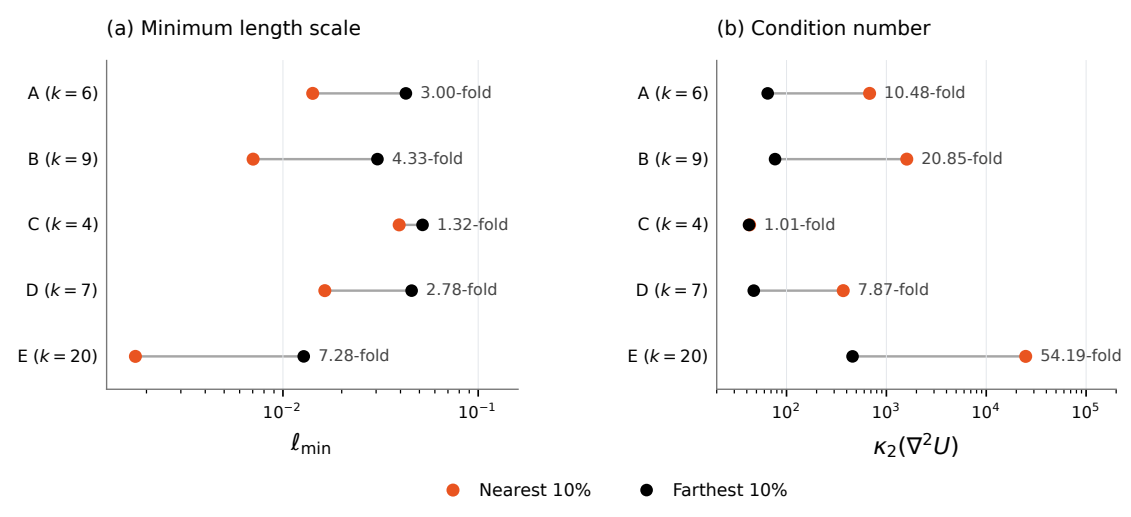


Figure 16: Variation of the shortest length scale (left) and the condition number (right) within the main rank stratum for each scenario. The two end-points are the medians computed from the nearest and farthest 10% of the output of the reference runs.

Such multiscale structure limits the potential of HMC. Stable step sizes are governed by the most constrained direction, which may be suboptimal for other parts of the posterior and for dynamics on $O(20)$, especially when the condition number is large. This is a mechanism commonly reported for HMC (Neal, 2011, Section 5.4.2.2).

C.5.2 Thawing rates

We use Gaussian velocity law $(v_r, (\sqrt{2}\Omega_{ij})_{i<j}) \sim N(0, I_{k+190})$ for $\Omega \in T_I(O(20))$, i.e., $\Omega^\top = -\Omega$, such that the geodesic flow is described as $\dot{r} = v_r$ and $\dot{Q} = Q\Omega$ and the velocity

Table 6: Estimated within-stratum HMC rejection probabilities at the posterior modal rank. All rows use identity mass, three leapfrog steps and $\epsilon \simeq 0.01234$.

Scenario	Rank	Near boundary	Far from boundary
A	6	29.5%	3.5%
B	9	36.0%	10.6%
C	4	2.9%	1.2%
D	7	17.0%	5.0%
E	20	99.0%	77.3%

compatibility condition is met at every interface.

At $r_{k+1} = 0$, Theorem 3.5, Corollaries 3.8 and 3.10 together give the Rao–Blackwellised rate

$$\Lambda_{k\to k+1}(r) = \frac{1}{2}\frac{f_{k+1}(r,0)}{f_k(r)}\sqrt{\frac{2}{\pi}} = \frac{1}{\sqrt{2\pi}}\frac{Z_k}{Z_{k+1}}\exp\left(-\frac{1}{2}\sum_{i=1}^{k} r_i\right)\prod_{i=1}^{k} r_i, \qquad k = 0, 1, \cdots 19.$$

C.5.3 Invariance proof

Theorem 3.5 assumes an Euclidean ambient space, which precludes the spectral lift parametrisation. However, the invariance argument using the extended Gauss–Green formula (Lemma B.5) in the proof also extends to product ambient spaces $N = U \times K$, where $U \subset \mathbb{R}^d$ is open and K is a compact Riemannian manifold without boundary, with sticking interfaces of the form $\mathcal{M} \times K$. Imposing the intrinsic counterparts of the previous assumptions and require any additional boundary terms to vanish, the flowwise Gauss–Green argument applies in local coordinates with respect to the Riemannian volume density, and a partition of unity yields the same identity. Under the product metric, the unit normal is of the form $(n_{\mathcal{M}}, 0)$, so the normal flux factor is again $|\langle v_r, n_{\mathcal{M}}\rangle|$.

C.5.4 Algorithm design and reversible-jump proposals

Both PDMPs use FECMC dynamics on the spectral lift. No global refreshment is used, while orthogonal switch is performed on every reflection event. Thinning uses locally constructed analytic bounds. RB-SM-FECMC discards the incoming radial velocity upon sticking and re-samples the outgoing velocity from the Rayleigh density $ve^{-v^2/2}$ $(v > 0)$ upon thawing. RJ-FECMC and RJ-HMC reflect at $r_k = 0$. RJ-HMC alternates a fixed-rank HMC update with the identity mass matrix on the spectral lift with the same birth–death kernel as RJ-FECMC. The orthogonal frame evolves along geodesics on $O(20)$ within FECMC and HMC (Byrne and Girolami, 2013).

Both reversible-jump algorithms propose births of new eigenvalues from densities

$$q_k(s \mid r) = \frac{a_k e^{-a_k s}}{1 - e^{-a_k r_k}}, \qquad 0 < s < r_k, \qquad a_k = \frac{k+2}{2},$$

with $r_0 = \infty$. Death proposals delete the smallest eigenvalue. RJ-FECMC uses Poisson clocks with rates λ_k, same for each direction. Writing the likelihood by $L(G) = p(Y \mid G)$, the birth acceptance probabilities are $\min(1, A_k)$, where

$$A_k = \frac{f_{k+1}(r,s)\, L(\beta_{k+1}((r,s),Q))}{f_k(r)\, L(\beta_k(r,Q))\, q_k(s \mid r)}.$$

The reverse death uses $\min\{1, A_k^{-1}\}$.

C.5.5 Tuning of HMC

For each scenario, pilot experiments compared $T \in \{2, 3, 5, 10, 20\}$ and searched for ϵ using a logarithmic grid followed by local refinement based on effective sample size per CPU second. Each candidate was evaluated using four chains of the complete RJ-HMC sampler (Table 7). Guided by these comparisons, we chose four representative configurations $(T, \epsilon) \in \{(3, 0.03), (3, 0.04), (3, 0.05), (10, 0.007)\}$ for the main experiments.

Table 7: Summary of the RJ-HMC pilot experiments, reporting the harmonic mean of ESSs over all components of the three quantities per process CPU second. Bold values indicate the largest validation point estimate within each scenario.

Scenario	Quantity	$T = 2$	$T = 3$	$T = 5$	$T = 10$	$T = 20$
A	ϵ	0.0538	0.0494	0.0381	0.0247	0.0123
	ESS/CPU sec	719.7	**922.6**	705.4	469.5	292.2
B	ϵ	0.0349	0.0320	0.0247	0.0160	0.00951
	ESS/CPU sec	173.2	**345.6**	342.3	323.2	217.5
C	ϵ	0.0538	0.0587	0.0381	0.0349	0.0269
	ESS/CPU sec	883.4	**994.0**	799.1	503.0	275.2
D	ϵ	0.0453	0.0415	0.0293	0.0147	0.00951
	ESS/CPU sec	536.3	**596.3**	503.2	360.9	232.8
E	ϵ	0.0135	0.0123	0.00951	0.00673	0.00476
	ESS/CPU sec	36.1	43.9	46.2	**50.7**	49.3
O	ϵ	0.0987	0.0698	0.0587	0.0494	0.0320
	ESS/CPU sec	**1318.0**	1148.5	916.8	545.3	292.3

C.5.6 Data generation, initialisation and RMSE computation

We generate each dataset with $n = 80$ and $\sigma = 0.5$:

$$X = Q_\star D_\star^{1/2} Z, \qquad Y = X + \sigma E,$$

where $D_\star$ the diagonal matrix with the specified signal eigenvalues, $Q_\star$ is uniform (Haar-distributed) eigendirections on $O(20)$, and Z, E have independent standard Gaussian entries. Each dataset is shared by all methods and replicates. The eigenvalue scenarios are selected to cover different patterns of posterior rank uncertainty (Figure 17).

For scenarios A–E, both initial positions and reference values were obtained using independent long RB-SM-FECMC runs from dispersed starts. Pilot runs are used to determine burn-in time $T_{\text{burn-in}}$ that satisfies (1) rank-normalised split $\widehat{R} < 1.01$ (Vehtari et al., 2021), (2) pooled effective sample sizes were at least 1,000, (3) time averages between first and second halves of $[T_{\text{burn-in}}, 2T_{\text{burn-in}}]$ match with a prescribed tolerance, and (4) each rank with estimated probability between 0.001 and 0.999 has at least 20 visits across chains. We then ran 100 independent preparation chains per scenario and, after confirming the same diagnostics, selected one position uniformly from $[T_{\text{burn-in}}, 2T_{\text{burn-in}}]$. These positions were shared across all samplers.

Reference values were computed from eight separate RB-SM-FECMC chains, up to both (i) $T_{\text{burn-in}}$ has passed and (ii) the criteria (1)-(4) have been met. Batch-means

Table 8: Total measured sampling CPU time for searching and validating all five integration step counts $T \in \{2, 3, 5, 10, 20\}$ in each scenario.

Scenario	Selected (T, ϵ)	Sampling CPU time (min)
A	(3, 0.04935)	84
B	(3, 0.032)	74
C	(3, 0.05869)	88
D	(3, 0.04150)	80
E	(10, 0.006727)	238
O	(2, 0.09870)	102

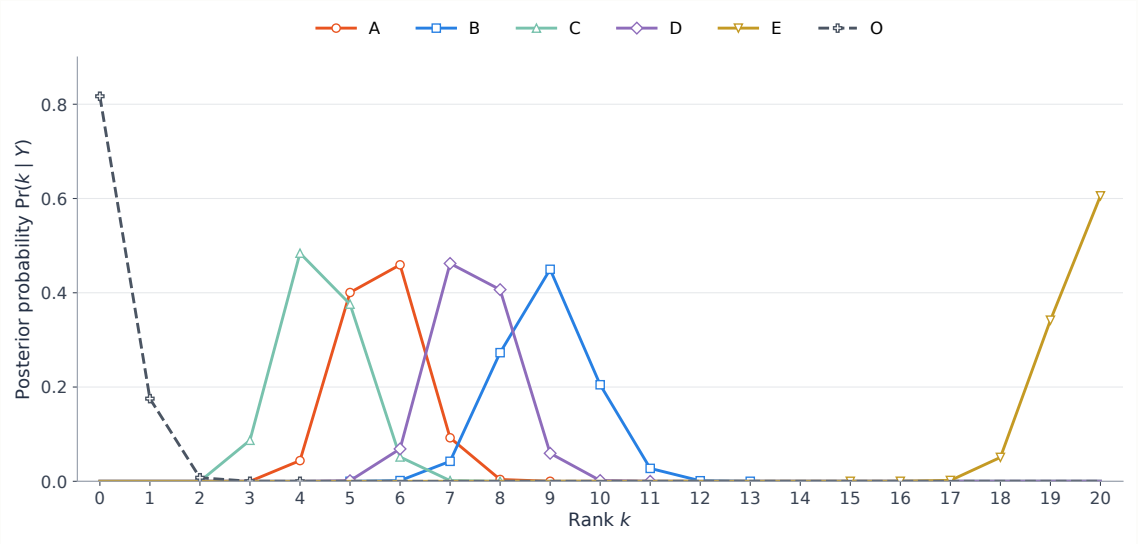


Figure 17: Estimated posterior rank probabilities $\widehat{p_k}$ $(k = 0, \ldots, 20)$ from the long reference runs for the six scenarios A–E and O.

estimators are then used to estimate Monte Carlo standard errors. These reference simulations are prolonged until the MCSEs are less than 10% of the smallest RMSE across methods for every scenario and estimand (Table 9). RMSEs are the square roots of the averages of the following squared errors over 100 replicates:

$$\|\hat{p} - p_{\text{ref}}\|_2^2, \qquad \|\widehat{G} - G_{\text{ref}}\|_F^2, \qquad \frac{1}{n}\|(\hat{B} - B_{\text{ref}})Y\|_F^2,$$

where $p = (p_0, \ldots, p_{20})$ and $B = \mathrm{E}[G(G + \sigma^2 I)^{-1} \mid Y]$.

Table 9: Precision of the independent RB-SM-FECMC reference computation. Each MCSE is followed by its percentage of the smallest RMSE among the six methods for that scenario and observable.

Scenario	Horizon H	Burn-in b	Rank-vector MCSE (%)	Mean-G MCSE (%)	Denoising MCSE (%)	$\widehat{R}_{\max}$
A	65536	4096	7.51×10^{-4} (8.39%)	1.09×10^{-3} (7.47%)	3.26×10^{-4} (6.60%)	1.00090
B	43008	4096	1.08×10^{-3} (8.44%)	1.39×10^{-3} (8.00%)	7.55×10^{-4} (7.84%)	1.00046
C	135168	8192	7.40×10^{-4} (8.64%)	5.94×10^{-4} (5.47%)	2.81×10^{-4} (5.86%)	1.00020
D	51200	4096	7.14×10^{-4} (7.64%)	1.24×10^{-3} (6.94%)	5.29×10^{-4} (6.92%)	1.00077
E	24576	2048	2.10×10^{-3} (9.87%)	3.24×10^{-3} (6.29%)	8.93×10^{-4} (6.31%)	1.00086
O	2150400	4096	4.15×10^{-4} (7.14%)	2.67×10^{-5} (5.79%)	4.13×10^{-5} (5.80%)	1.00000

C.5.7 Additional results

Here we restate the full results (Table 10), with Table 11 and Figure 18 showing RMSE ratios for the posterior mean of the true signal x. The modal posterior rank k^* increases in the order of scenario O, C, A, D, B, and E.

Table 10: Bayesian PCA comparison using 100 paired replicates per scenario and a three-second process-CPU stopping threshold. RJ-FECMC and RJ-HMC are tuned to match the model-transition count of RB-SM-FECMC using independent pilot and validation runs.

Scenario	Sampler	Rank-vector RMSE	Mean-G RMSE	Denoising RMSE	Mean rank transitions
A $(2, 1.5, 1, 0.5, 0, \dots, 0)$ $k^\star = 6$, $\hat{p}_{k^\star} = 0.459$	RB-SM-FECMC	1.12×10^{-2}	$\mathbf{1.46 \times 10^{-2}}$	$\mathbf{4.95 \times 10^{-3}}$	3.96×10^{4}
	RJ-FECMC	1.15×10^{-2}	1.60×10^{-2}	5.25×10^{-3}	3.62×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.03$)	1.18×10^{-2}	2.76×10^{-2}	7.16×10^{-3}	3.85×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.04$)	1.05×10^{-2}	2.31×10^{-2}	6.26×10^{-3}	3.96×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.05$)	$\mathbf{8.95 \times 10^{-3}}$	2.11×10^{-2}	5.81×10^{-3}	4.02×10^{4}
	RJ-HMC ($T = 10$, $\epsilon = 0.007$)	1.98×10^{-2}	5.15×10^{-2}	1.31×10^{-2}	3.85×10^{4}
B $(1, 1, 1, 0.1, \dots, 0.1)$ $k^\star = 9$, $\hat{p}_{k^\star} = 0.450$	RB-SM-FECMC	$\mathbf{1.27 \times 10^{-2}}$	$\mathbf{1.74 \times 10^{-2}}$	$\mathbf{9.63 \times 10^{-3}}$	2.85×10^{4}
	RJ-FECMC	1.32×10^{-2}	1.79×10^{-2}	9.77×10^{-3}	2.97×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.03$)	3.03×10^{-2}	1.05×10^{-1}	6.44×10^{-2}	3.01×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.04$)	4.13×10^{-2}	1.23×10^{-1}	7.44×10^{-2}	3.22×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.05$)	3.90×10^{-2}	1.24×10^{-1}	7.52×10^{-2}	3.23×10^{4}
	RJ-HMC ($T = 10$, $\epsilon = 0.007$)	2.09×10^{-2}	5.86×10^{-2}	2.70×10^{-2}	2.86×10^{4}
C $(1, 1, 1, 0.1, 0.1, 0.1, 0, \dots, 0)$ $k^\star = 4$, $\hat{p}_{k^\star} = 0.484$	RB-SM-FECMC	1.34×10^{-2}	$\mathbf{1.09 \times 10^{-2}}$	4.92×10^{-3}	3.52×10^{4}
	RJ-FECMC	1.35×10^{-2}	1.13×10^{-2}	5.40×10^{-3}	3.82×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.03$)	1.17×10^{-2}	2.27×10^{-2}	6.82×10^{-3}	3.59×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.04$)	9.03×10^{-3}	1.77×10^{-2}	5.33×10^{-3}	3.73×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.05$)	$\mathbf{8.56 \times 10^{-3}}$	1.50×10^{-2}	$\mathbf{4.80 \times 10^{-3}}$	3.71×10^{4}
	RJ-HMC ($T = 10$, $\epsilon = 0.007$)	2.12×10^{-2}	4.30×10^{-2}	1.28×10^{-2}	3.61×10^{4}
D $(1, 1, 1, 1, 1, 0, \dots, 0)$ $k^\star = 7$, $\hat{p}_{k^\star} = 0.462$	RB-SM-FECMC	1.07×10^{-2}	$\mathbf{1.79 \times 10^{-2}}$	$\mathbf{7.64 \times 10^{-3}}$	5.85×10^{4}
	RJ-FECMC	1.23×10^{-2}	2.05×10^{-2}	8.36×10^{-3}	5.64×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.03$)	1.04×10^{-2}	3.60×10^{-2}	1.05×10^{-2}	5.81×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.04$)	$\mathbf{9.34 \times 10^{-3}}$	3.10×10^{-2}	9.53×10^{-3}	5.91×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.05$)	1.07×10^{-2}	2.96×10^{-2}	9.80×10^{-3}	5.91×10^{4}
	RJ-HMC ($T = 10$, $\epsilon = 0.007$)	1.70×10^{-2}	6.12×10^{-2}	1.80×10^{-2}	5.54×10^{4}
E $(1, 0.95, 0.9, \dots, 0.05)$ $k^\star = 20$, $\hat{p}_{k^\star} = 0.606$	RB-SM-FECMC	$\mathbf{2.13 \times 10^{-2}}$	$\mathbf{5.15 \times 10^{-2}}$	$\mathbf{1.42 \times 10^{-2}}$	6.45×10^{4}
	RJ-FECMC	2.94×10^{-2}	6.15×10^{-2}	1.70×10^{-2}	6.91×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.03$)	1.65×10^{-1}	1.05×10^{0}	3.10×10^{-1}	7.28×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.04$)	3.71×10^{-1}	1.85×10^{0}	5.82×10^{-1}	7.52×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.05$)	4.09×10^{-1}	1.99×10^{0}	6.35×10^{-1}	7.32×10^{4}
	RJ-HMC ($T = 10$, $\epsilon = 0.007$)	2.41×10^{-2}	1.59×10^{-1}	4.50×10^{-2}	6.71×10^{4}
O $(0, \dots, 0)$ $k^\star = 0$, $\hat{p}_{k^\star} = 0.816$	RB-SM-FECMC	8.30×10^{-3}	5.40×10^{-4}	8.39×10^{-4}	3.76×10^{4}
	RJ-FECMC	6.53×10^{-3}	$\mathbf{4.61 \times 10^{-4}}$	$\mathbf{7.13 \times 10^{-4}}$	3.58×10^{4}
	RJ-HMC ($T = 2$, $\epsilon = 0.1$)	$\mathbf{5.82 \times 10^{-3}}$	5.84×10^{-4}	8.82×10^{-4}	3.92×10^{4}
	RJ-HMC ($T = 3$, $\epsilon = 0.05$)	9.75×10^{-3}	8.28×10^{-4}	1.26×10^{-3}	3.77×10^{4}

Table 11: HMC acceptance rates and relative RMSEs.

Scenario	Posterior modal rank $k^\star$	T	ϵ	acceptance rate of HMC	RMSE ratio for G	RMSE ratio for X	acceptance rate of RJ
O	0	2	0.1	0.907	**1.083**	**1.050**	0.069
		3	0.05	0.974	1.534	1.506	0.069
C	4	3	0.03	0.761	2.092	1.388	0.242
		3	0.04	0.625	1.625	1.084	0.242
		3	0.05	0.499	**1.382**	**0.975**	0.242
		10	0.007	0.958	3.956	2.611	0.242
A	6	3	0.03	0.654	1.887	1.447	0.256
		3	0.04	0.504	1.579	1.264	0.256
		3	0.05	0.407	**1.440**	**1.175**	0.256
		10	0.007	0.931	3.516	2.657	0.256
D	7	3	0.03	0.477	2.012	1.374	0.255
		3	0.04	0.317	1.730	**1.247**	0.255
		3	0.05	0.206	**1.652**	1.282	0.255
		10	0.007	0.853	3.418	2.353	0.255
B	9	3	0.03	0.298	6.073	6.692	0.289
		3	0.04	0.138	7.095	7.724	0.288
		3	0.05	0.054	7.161	7.804	0.290
		10	0.007	0.829	**3.374**	**2.804**	0.290
E	20	3	0.03	3.40×10^{-4}	20.420	21.868	0.157
		3	0.04	6.40×10^{-6}	35.793	41.091	0.170
		3	0.05	3.81×10^{-8}	38.542	44.884	0.163
		10	0.007	0.183	**3.079**	**3.179**	0.151

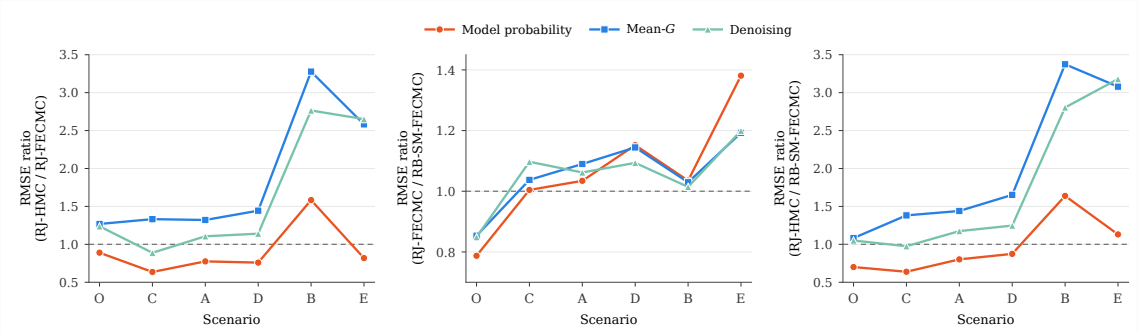


Figure 18: Relative RMSEs across the six PPCA scenarios. Left: RJ-HMC relative to RJ-FECMC; middle: RJ-FECMC relative to RB-SM-FECMC; right: RJ-HMC relative to RB-SM-FECMC. For each scenario and quantity, the RJ-HMC ratio uses the smallest observed RMSE among all configurations.

RJ-HMC shows varying acceptance probabilities across different rank strata (Table 12). This is due to the differences in length scale, which is reflected in the increased number of rejections from order constraints violation. The Vandermonde determinant in the prior (30) serves as the potential barrier to the order constraints, so exact Hamiltonian flows would not reach these boundaries.

Comparison between SM-FECMC and RJ-FECMC reveals an advantage of the sticky manifold transition. Additional computational efforts spent on RJ proposals are reflected in reduced trajectory lengths in process time; see Table 13. The rejection-free property

Table 12: Acceptance probabilities of the RJ-HMC with $(T, \epsilon) = (3, 0.05)$ in scenarios C and D.

Scenario	Rank k	Posterior rank probability (%)	HMC acceptance rate (%)	Rejection due to ordering violations (%)
C	3	8.7	69.9	3.8
	4	48.4	62.1	3.8
	5	37.6	35.6	28.1
	6	5.2	7.4	80.8
D	7	46.2	33.0	28.3
	8	40.7	4.1	86.6

Table 13: Relative trajectory lengths based on process time and model-probability RMSEs of RJ-FECMC compared with RB-SM-FECMC. The RJ-HMC column reports the smallest RMSE ratio among the RJ-HMC configurations

			Model probability RMSE ratio		Mean-G RMSE ratio		Denoising RMSE ratio	
Scenario	Mean trajectory length (%)	RJ-FECMC RJ acceptance probability	RJ-FECMC	RJ-HMC	RJ-FECMC	RJ-HMC	RJ-FECMC	RJ-HMC
A	85.5	0.2560	1.034	0.802	1.090	1.440	1.062	1.175
B	97.5	0.2892	1.035	1.639	1.030	3.374	1.014	2.804
C	89.9	0.2415	1.004	0.638	1.037	1.382	1.097	0.975
D	78.2	0.2550	1.152	0.874	1.144	1.652	1.094	1.247
E	73.4	0.2161	1.381	1.131	1.192	3.079	1.199	3.179
O	96.6	0.1174	0.787	0.701	0.853	1.083	0.849	1.050

works particularly favourably in scenarios D and E, while the two between-model transition mechanisms are comparable in scenarios A, B, and C, where RMSEs for model probability estimation are comparable between SM-FECMC and RJ-FECMC. Things are more nuanced between SM-FECMC and RJ-HMC because within-stratum dynamics are also different.

C.6 Survival extrapolation example (Section 4.5)

C.6.1 Algorithm design

The ambient dynamics are FECMC (Michel et al., 2020) with Gaussian velocity. No orthogonal switch is performed, while global refreshment is performed with rate 0.5. Within each submanifold, the arc length parametrisation is used within manifold PDMP (Algorithm 1). As this parametrisation is isometric, we continue to use standard Gaussian velocity on it, as in Example 3.1. Event times are computed by thinning with analytic bounds.

C.6.2 The model

The accelerated failure time model is specified with a treatment-dependent location parameter as

$$\mu_i = \alpha + \beta z_i$$

with respect to the treatment indicator z_i, where $z_i = 0$ denotes the control group and $z_i = 1$ the treatment group for patient i. The separate model fits $\theta_g = (\mu_g, \eta_g, \lambda_g)$ to both groups $g \in \{0, 1\}$ respectively. Restricted mean survival time at time τ is defined by

$$\mathrm{RMST}(\tau) := \int_0^\tau S(t)\, dt,$$

$S(t)$ is the survival function.

C.6.3 Prior specification and the thawing rates

In the ambient generalised gamma component, we assign mutually independent priors

$$\alpha \sim N(\log 2.5, 1.5^2), \quad \beta \sim N(0, 1), \quad \eta := \log \sigma \sim N(0, 0.75^2), \quad \lambda \sim N(0, 1).$$

For each submodel $i \in \{\mathrm{LN}, \mathrm{W}, \Gamma, \mathrm{A}\}$, we define the priors as pull-backs through the associated inclusion map $i : \mathcal{M}_i \to \mathbb{R}^4$. In other words, they share the same density up to a normalising factor, with respect to the Hausdorff measure $\mathcal{H}^3_{\mathcal{M}_i}$ such that

$$d\mathcal{H}^3_{\mathcal{M}_i} = \sqrt{1 + (r_i'(\eta))^2}\, d\alpha d\beta d\eta,$$

with $\lambda = r_i(\eta)$ being the constraints for the submodel i. The Rao–Blackwellised unsticking time (Corollary 3.9) is therefore

$$\tau_i = \sqrt{\frac{\pi}{2}} \frac{\omega_i}{\omega_{\mathrm{GG}}} \frac{1}{Z_i},$$

where

$$Z_i := \int_{\mathbb{R}^3} \phi(\alpha; \log 2.5, 1.5^2)\phi(\beta; 0, 1)\phi(\eta; 0, 0.75^2)\phi(r_i(\eta); 0, 1)\, \mathcal{H}^3_{\mathcal{M}_i}(d\alpha d\beta d\eta).$$

C.6.4 Data, initialisation, and output analysis

We analysed all 810 records in the `ta174` dataset as distributed with the R package `survHE` (Baio, 2020). We use the observed overall survival times t_i in years, death indicators δ_i, and treatment indicators z_i for all 810 patients. The data comprise 93 deaths and 717 right-censored observations, with a maximum follow-up of four years. The likelihood is

$$L(\theta) = \prod_{i=1}^{810} f(t_i \mid z_i, \theta)^{\delta_i} S(t_i \mid z_i, \theta)^{1-\delta_i}.$$

The sampler is initialised in the GG component at the posterior mode obtained by Nelder–Mead optimisation. The initial velocity is randomly drawn from $N(0, I_4)$, and the first 200 units of process time are discarded as burn-in. The Monte Carlo standard errors are reported in Tables 14 and 15.

Table 14: Posterior model probabilities with Monte Carlo standard errors for each survival model.

Model prior	Gen-Gamma	Log-normal	Weibull	Gamma	Ammag
Treatment arms modelled together					
Uniform prior	8.91 (0.136)	10.54 (0.242)	15.07 (0.304)	10.39 (0.200)	55.09 (0.464)
Informative prior	6.66 (0.088)	30.93 (0.719)	11.75 (0.284)	8.16 (0.165)	42.49 (0.651)
Treatment arms modelled separately					
Control					
Uniform prior	10.50 (0.125)	11.68 (0.233)	18.64 (0.277)	13.17 (0.185)	46.00 (0.418)
Informative prior	7.89 (0.094)	34.29 (0.784)	13.71 (0.309)	10.12 (0.179)	34.00 (0.637)
Treatment					
Uniform prior	9.21 (0.111)	9.53 (0.213)	16.17 (0.327)	10.97 (0.212)	54.12 (0.485)
Informative prior	7.07 (0.070)	28.62 (0.611)	12.87 (0.320)	8.68 (0.214)	42.76 (0.520)

Table 15: Monte Carlo standard errors of estimates of the 30-year restricted mean survival time difference.

	Model averaged				
Model prior	Posterior mean	MCSE	Relative MCSE (%)	MCSE range across batch counts	Half-run difference
Treatment arms modelled together					
Uniform prior	2.7422	0.00468	0.17	0.00329–0.00478	−0.00679
Informative prior	2.5094	0.00956	0.38	0.00894–0.00999	−0.01537
Treatment arms modelled separately					
Uniform prior	3.6319	0.02126	0.59	0.01917–0.02141	−0.01335
Informative prior	2.6499	0.04274	1.61	0.03750–0.04625	+0.08710

C.6.5 Additional results

Prior and posterior model probabilities are shown in Table 16. The results in Section 4.5 for the independent models for treatment arms are based on the informative prior for

the control arm and uniform prior for the treatment arm. We also include the posterior probabilities for when the treatment arm also uses the informative prior.

Table 16: Prior and posterior model probabilities for the joint AFT model and the treatment-only and control-only models.

	Joint model		Treatment only		Control only	
Model	Prior	Posterior	Prior	Posterior	Prior	Posterior
Uniform prior						
GG	0.200	0.0891	0.200	0.0921	0.200	0.1050
Log-normal	0.200	0.1054	0.200	0.0953	0.200	0.1168
Weibull	0.200	0.1507	0.200	0.1617	0.200	0.1864
Gamma	0.200	0.1039	0.200	0.1097	0.200	0.1317
Ammag	0.200	0.5509	0.200	0.5412	0.200	0.4600
Informative prior						
GG	0.125	0.0666	0.125	0.0707	0.125	0.0789
Log-normal	0.500	0.3093	0.500	0.2862	0.500	0.3429
Weibull	0.125	0.1175	0.125	0.1287	0.125	0.1371
Gamma	0.125	0.0816	0.125	0.0868	0.125	0.1012
Ammag	0.125	0.4249	0.125	0.4276	0.125	0.3400

Tha full table of Table 17 is as follows:

Table 17: Estimates of 30-year restricted mean survival time (years). Entries are posterior means (95% credible intervals).

						Model averaged	
Model	Gen-Gamma	Log-normal	Weibull	Gamma	Ammag	Uniform prior	Informative prior
Treatment arms modelled together							
Control	15.63 (8.82, 20.70)	18.16 (15.74, 20.48)	13.10 (9.53, 16.95)	13.14 (10.08, 16.55)	13.08 (8.49, 18.02)	13.85 (8.81, 19.48)	14.82 (9.02, 19.96)
Treatment	17.89 (11.56, 22.05)	19.85 (17.47, 22.13)	15.96 (12.04, 19.82)	15.99 (12.49, 19.61)	16.04 (11.33, 20.41)	16.59 (11.65, 21.22)	17.33 (11.93, 21.63)
Difference	2.26 (−0.56, 5.73)	1.70 (−0.55, 3.89)	2.86 (−0.58, 6.32)	2.85 (−0.58, 6.28)	2.97 (−0.64, 6.72)	2.74 (−0.61, 6.44)	2.51 (−0.58, 6.24)
Treatment arms modelled separately							
Control	14.81 (7.86, 20.53)	17.19 (14.17, 20.08)	12.25 (8.30, 16.78)	12.44 (9.02, 16.41)	12.74 (8.35, 17.95)	13.35 (8.45, 19.14)	14.33 (8.70, 19.55)
Treatment	18.25 (11.31, 22.98)	20.10 (17.17, 22.73)	16.33 (11.37, 20.75)	16.28 (12.08, 20.42)	16.55 (11.37, 21.34)	16.98 (11.53, 21.90)	16.98 (11.53, 21.90)
Difference	3.45 (−5.71, 12.34)	2.92 (−1.20, 6.93)	4.08 (−2.56, 10.21)	3.84 (−1.89, 9.31)	3.80 (−3.59, 10.67)	3.63 (−4.42, 11.01)	2.65 (−5.28, 10.58)