Check for updates

Adv. Appl. Probab. **56,** 1153–1194 (2024)
*doi:*10.1017/apr.2023.61

PDMP MONTE CARLO METHODS FOR PIECEWISE SMOOTH DENSITIES

AUGUSTIN CHEVALLIER,* *Université de Strasbourg*
SAM POWER, *University of Bristol*
ANDI Q. WANG, *University of Warwick*
PAUL FEARNHEAD, *Lancaster University*

Abstract

There has been substantial interest in developing Markov chain Monte Carlo algorithms based on piecewise deterministic Markov processes. However, existing algorithms can only be used if the target distribution of interest is differentiable everywhere. The key to adapting these algorithms so that they can sample from densities with discontinuities is to define appropriate dynamics for the process when it hits a discontinuity. We present a simple condition for the transition of the process at a discontinuity which can be used to extend any existing sampler for smooth densities, and give specific choices for this transition which work with popular algorithms such as the bouncy particle sampler, the coordinate sampler, and the zigzag process. Our theoretical results extend and make rigorous arguments that have been presented previously, for instance constructing samplers for continuous densities restricted to a bounded domain, and we present a version of the zigzag process that can work in such a scenario. Our novel approach to deriving the invariant distribution of a piecewise deterministic Markov process with boundaries may be of independent interest.

Keywords: Bayesian inference; bouncy particle sampler; Hamiltonian Monte Carlo; Markov chain Monte Carlo; non-reversible samplers; zigzag process

2020 Mathematics Subject Classification: Primary 62-08
Secondary 65C05; 60J25

1. Introduction

In recent years there has been substantial interest in using continuous-time *piecewise deterministic Markov processes* (PDMPs) as the basis for Markov chain Monte Carlo (MCMC) algorithms. These ideas started in the statistical physics literature [2, 22, 28] and have led to a number of new MCMC algorithms, such as the bouncy particle sampler (BPS) [8], the zigzag (ZZ) process [5], and the coordinate sampler (CS) [33], amongst many others. See [18] and [31] for an introduction to the area. One potential benefit which is associated with these samplers are that they are non-reversible, and it is known that non-reversible samplers can mix faster than their reversible counterparts [3, 9, 13].

Informally, a PDMP process evolves according to a deterministic flow—defined via an ordinary differential equation—for a random amount of time, before exhibiting an instantaneous

Received 16 December 2021; accepted 11 December 2023.
* Email address: a.chevallier@unistra.fr

© The Author(s), 2024. Published by Cambridge University Press on behalf of Applied Probability Trust.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

transition, and then following a (possible different) deterministic flow for another random amount of time, and so on.

Initial PDMP samplers were defined to sample target distributions which were continuously differentiable (C^1) on $\mathbb{R}^d$, but there is interest in extending them to more general target distributions. To date, this has been achieved for sampling from distributions defined on the union of spaces of different dimensions [7, 10] and for sampling from distributions on restricted domains [4], binary distributions [26], and phylogenetic trees [21]. Here we rigorously consider a further extension to sampling from target distributions on $\mathbb{R}^d$ which are piecewise C^1. That is, they can be defined by partitioning $\mathbb{R}^d$ into a countable number of regions, with the target density being C^1 on each region. We call such densities *piecewise smooth*. Such target distributions arise in a range of statistical problems, such as latent threshold models [24], binary classification [25], and changepoint models [29]. The importance of this extension of PDMP samplers is also indicated by the usefulness of extensions of Hamiltonian Monte Carlo (HMC) to similar problems [1, 14, 27, 34].

The challenge with extending PDMP samplers to piecewise smooth densities is the need to specify the appropriate dynamics when the sampler hits a discontinuity in the density. Essentially, we need to specify the dynamics so that the PDMP has our target distribution μ as its invariant distribution. Current samplers are justified based on considering the infinitesimal generator, denoted by A, of the PDMP. Informally, the generator is an operator that acts on functions and describes how the expectation of that function of the state of the PDMP changes over time. The idea is that if we average the generator applied to a function and this is zero for a large set of functions $f \in \mathcal{F}$,

$$\int Af \, \mathrm{d}\mu = 0, \tag{1}$$

then the distribution μ must be the invariant distribution. Whilst intuitively this makes sense, many papers use this intuitive reasoning without giving a formal proof that the distribution they average over is in fact the invariant distribution; see for instance [18, 31]. One important exception is the work of [16], where precise conditions and functional-analytic proofs are given which establish that smooth, compactly supported functions form a *core* for the generator. This then rigorously establishes μ-invariance of the PDMP. In this work, we actually take a different approach from [16]: we make use of the particular nature of the PDMPs under consideration and give conditions under which (1) directly holds for the whole domain $\mathcal{D}(A)$, then show that a large class of functions $\mathcal{F} \subset \mathcal{D}(A)$ separate measures, without attempting to show that $\mathcal{F}$ is a core. This avoids the need for highly technical proofs as in [16].

Furthermore, once we introduce discontinuities, this complicates the definition of the generator. The impact of these discontinuities is realized in terms of the *domain* of the generator, and this necessitates the use of additional arguments which take account of the impact of the discontinuity when applying arguments based on integration by parts.

More specifically, we can see the challenge of dealing with discontinuities and some of the contributions of this paper by considering the previous work of [4], who consider designing PDMP samplers when the target distribution is only compactly supported on $\mathbb{R}^d$—a special case of our present work. It should be noted that there is a sign error in the condition given by Equation (4) of [4], and furthermore their proof was not complete—see Appendix E. In this work, we will provide a full proof.

The paper is structured as follows. In the next section we give a brief introduction to PDMPs and some common PDMP samplers. Then in Section 3 we give general conditions for

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

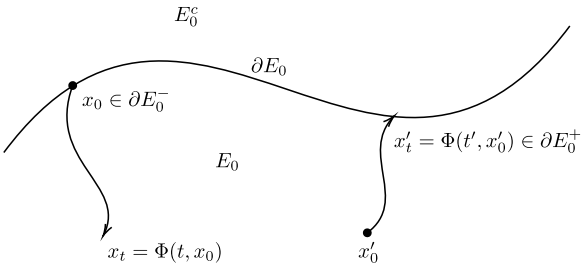


FIGURE 1. Exit boundary $\left(\partial E_0^+\right)$ and entrance boundary $\left(\partial E_0^+\right)$: x_0 is in the entrance boundary ∂E_0^-, while x_t' is in the exit boundary ∂E_0^+. The arrows represent the flow $\Phi(\cdot, \cdot)$.

checking the invariant distribution of a PDMP. The result in this section formalizes the informal argument used by previous authors. We then develop these results for the specific cases where the PDMP has active boundaries—for example, due to a compact support, or when we wish to sample from a mixture of densities defined on spaces of differing dimensions. The results in this section can be used to address some shortcomings of [4] for sampling on a bounded domain, and have been used to justify the reversible jump PDMP algorithm of [10]. In Section 5 we use our results to provide a simple sufficient condition on the dynamics for a PDMP to admit a specific piecewise smooth density as its invariant density. Various proofs and technical assumptions for these results are deferred to the appendices. We then show how to construct dynamics at the discontinuities for a range of common PDMP samplers, and empirically compare these samplers on some toy examples—with a view to gaining intuition as to their relative merits, particularly when we wish to sample from high-dimensional piecewise smooth densities. The paper ends with a discussion.

2. PDMP basic properties

2.1. General PDMP construction

For this work, we require a general construction of *piecewise deterministic Markov processes* (PDMPs) in spaces featuring boundaries. We will follow the construction of Davis in [11, p. 57], and largely make use of the notation therein.

Let K be a countable set, and for $k \in K$, let E_k^0 be an open subset of $\mathbb{R}^{d_k}$. Let E^0 be their disjoint union:

$$E^0 := \left\{(k, z) : k \in K, z \in E_k^0\right\}.$$

For any $k \in K$, we have a Lipschitz vector field on E_k^0 that induces a flow $\Phi_k(t, z)$.

In this setting, trajectories may reach the boundaries of the state. Hence we define the entrance and exit boundaries using the flow, $\partial^- E_k^0$ and $\partial^+ E_k^0$ respectively (see Figure 1):

$$\partial^{\pm} E_k^0 = \left\{z \in \partial E_k^0 | z = \Phi_k(\pm t, \xi) \text{ for some } \xi \in E_k^0 \text{ and } t > 0\right\}.$$

Note that it may be possible for a point to be both an entrance and an exit boundary.

We then have $\partial_1 E_k^0 := \partial^- E_k^0 \backslash \partial^+ E_k^0$ as in [11, p. 57], and also

$$E_k := E_k^0 \cup \partial_1 E_k^0.$$

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

Finally, the full state space is the disjoint union,

$$E := \bigcup_k (\{k\} \times E_k).$$

The active boundary (that is, the exit boundary) is then defined as

$$\Gamma := \bigcup_k \left(\{k\} \times \partial^+ E_k^0\right).$$

These are points on the boundary that the deterministic flow can hit.

We will denote the state of a PDMP on E at time t by $Z_t \in E$. A detailed construction of the PDMP is provided by Davis [11, p. 59], but we provide here a summary of the quantities that define a PDMP (Z_t):

(i) An event rate $\lambda(z)$, with $z \in E$. An event occurs in $[t, t+h]$ with probability $\lambda(Z_t)h + o(h)$.

(ii) A jump kernel defined for $z \in E \cup \Gamma$: $Q(\cdot|z)$ with $Q(\cdot|z)$ a probability measure on E. At each event time T_i, the state will change according to the jump kernel: $Z_{T_i} \sim Q(\cdot|Z_{T_i-})$.

(iii) The deterministic flow Φ which determines the behaviour of Z_t between jumps.

(iv) For any trajectory Z_t such that

$$\lim_{t \uparrow t_0} Z_t = Z_{t_0-} \in \Gamma,$$

the state will change according to the jump kernel: $Z_{t_0} \sim Q(\cdot|Z_{t_0-})$.

Remark 1. The trajectory never enters Γ, which is not in the domain.

2.2. PDMP samplers

In the case of most PDMP samplers, the state space is constructed by augmenting the space of interest with an auxiliary velocity space $\mathcal{V}_k$: $E_k^0 = U_k \times \mathcal{V}_k$. The deterministic flow is then typically given by free transport, i.e. $\Phi_k(t, x, v) = (x + tv, v)$, though other examples exist [6, 30, 31]. The use of such simple dynamics allows for the exact simulation of the process dynamics, without resorting to numerical discretization.

For the purposes of this work, it will be useful to introduce three of the more popular classes of PDMP sampler, which we will then be able to refer back to as running examples. Each of these processes works on a velocity-augmented state space and follows free-transport dynamics, and so they differ primarily in (i) the set of velocities which they use, and (ii) the nature of the jumps in the process. We describe the dynamics for each process if we wish to sample from a density $\pi(x)$ on $\mathbb{R}^d$:

1. The *bouncy particle sampler* [8] uses a spherically symmetric velocity space, given by either $\mathbb{R}^d$ equipped with the standard Gaussian measure, or the unit sphere equipped with the uniform surface measure. 'Bounce' events occur at rate $\lambda(x, v) = \langle v, -\nabla \log \pi(x) \rangle_+$, and at such events, the velocity deterministically jumps to $v' = \left(I - 2\frac{(\nabla \log \pi(x))(\nabla \log \pi(x))^\top}{(\nabla \log \pi(x))^\top (\nabla \log \pi(x))}\right) v$, i.e. a specular reflection against the level set of $\log \pi$ at x.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

2. The *zigzag process* [5] uses $\{\pm 1\}^d$ as its velocity space, equipped with the uniform measure. There are now d different types of bounce events, corresponding to the coordinates of the velocity vector. Bounces of type i occur at rate $\lambda_i(x, v) = (-v_i \partial_i \log \pi(x))_+$, and at such events, v_i is deterministically replaced by $-v_i$.
3. The *coordinate sampler* [20, 33] uses $\{\pm e_i\}_{i=1}^d$ as its velocity space, equipped with the uniform measure, where e_i is the ith coordinate vector. Bounce events again happen at rate $\lambda(x, v) = \langle v, -\nabla \log \pi(x)\rangle_+$. At such events, the velocity is resampled from the full velocity space, with probability proportional to $\lambda(x, -v')$.

When $d = 1$, all of these processes are identical. Additionally, all three processes can be supplemented with 'refreshment' events, which occur at a rate independent of v, and modify the velocity in a way which leaves its law invariant. This can include full resampling, autoregressive resampling in the case of spherical velocities, coordinate-wise resampling in the case of velocity laws with independent coordinates, and other variations.

From the above definitions, it is easy to see that the event rates only make sense when π is sufficiently smooth, and that in the presence of discontinuities, complications in defining the process will arise. Furthermore, it is not a priori clear which processes will work best in the presence of such discontinuities.

2.3. Review: extended generator and semigroup

We collect some basic definitions and facts which will be crucial for our later results.

Let $\mathcal{B}(E)$ denote the set of bounded measurable functions $E \to \mathbb{R}$. For any $f \in \mathcal{B}(E)$, we recall the definition of the semigroup P_t associated to the process Z_t:

$$P_t f(z) = \mathbb{E}_z[f(Z_t)], \qquad z \in E,$$

where $\mathbb{E}_z$ is the expectation with respect to $\mathbb{P}_z$, with $\mathbb{P}_z$ the probability such that $\mathbb{P}_z[Z_0 = z] = 1$.

Proposition 1. *The semigroup P_t is a contraction for the sup norm:*

$$\|P_t f\|_\infty \le \|f\|_\infty,$$

for all $f \in \mathcal{B}(E)$ and $t \ge 0$.

Proof. See [11, p. 28]. □

The semigroup is said to be *strongly continuous for* $f \in \mathcal{B}(E)$ if $\lim_{t \downarrow 0} \|P_t f - f\|_\infty = 0$. Let $\mathcal{B}_0$ be the set of functions for which P_t is strongly continuous:

$$\mathcal{B}_0 := \left\{ f \in \mathcal{B}(E) : \lim_{t \downarrow 0} \|P_t f - f\|_\infty = 0 \right\}.$$

Lemma 1. *We have that $\mathcal{B}_0 \subset \mathcal{B}(E)$ is a Banach space with sup norm $\|\cdot\|_\infty$, and P_t maps $\mathcal{B}_0 \to \mathcal{B}_0$ for any $t \ge 0$.*

Proof. See [11, p. 29]. □

Let us write $(A, \mathcal{D}(A))$ for the infinitesimal generator (also referred to as the *strong generator*) of the semigroup (P_t). By definition, for all $f \in \mathcal{D}(A)$,

$$Af = \lim_{t \to 0} \frac{1}{t}(P_t f - f),$$

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

with this limit being taken in $\|\cdot\|_\infty$, with

$$\mathcal{D}(A) = \left\{ f \in \mathcal{B}_0 : \left\| \frac{1}{t}(P_t f - f) - g \right\|_\infty \to 0, \text{ for some } g \in \mathcal{B}(E) \right\}. \tag{2}$$

Since g in (2) is a limit of functions in a Banach space, if such a g exists, it must be unique, and Af is well-defined.

Lemma 2. *Let $f \in \mathcal{D}(A)$. Then $Af \in \mathcal{B}_0$. In other words, $A : \mathcal{D}(A) \to \mathcal{B}_0$.*

Proof. This is immediate since P_t maps $\mathcal{B}_0 \to \mathcal{B}_0$. □

We now define the *extended generator* $(\mathfrak{A}, \mathcal{D}(\mathfrak{A}))$: $\mathcal{D}(\mathfrak{A})$ is the set of (potentially unbounded) measurable functions $f : E \to \mathbb{R}$ such that there exists a measurable function $h : E \to \mathbb{R}$ with $t \mapsto h(Z_t)$ $\mathbb{P}_z$-integrable almost surely for each initial point z, and such that the process

$$C_t^f := f(Z_t) - f(Z_0) - \int_0^t h(Z_s)\,\mathrm{d}s, \quad t \geq 0, \tag{3}$$

is a local martingale; see [11, (14.16)]. For $f \in \mathcal{D}(\mathfrak{A})$, $\mathfrak{A}f = h$, for h as in (3).

Proposition 2. *The extended generator is an extension of the infinitesimal generator:*

1. *$\mathcal{D}(A) \subset \mathcal{D}(\mathfrak{A})$.*
2. *$Af = \mathfrak{A}f$ for any $f \in \mathcal{D}(A)$.*

Proof. See [11, p. 32]. □

To simplify notation, we will define the action of our probability kernel, Q, on a function, f, as

$$(Qf)(z) = \int_E f(y) Q(\mathrm{d}y|z).$$

We will assume throughout this work that the *standard conditions* of Davis [11, (24.8)] hold. Under this assumption on PDMPs, $(\mathfrak{A}, \mathcal{D}(\mathfrak{A}))$ are fully characterized in [11, (26.14)]. In particular, the set $\mathcal{D}(\mathfrak{A})$ is entirely known, and for all $f \in \mathcal{D}(\mathfrak{A})$,

$$\mathfrak{A}f(z) = \Xi f(z) + \lambda(z)\{(Qf)(z) - f(z)\}, \tag{4}$$

where Ξ is the differential operator associated to the deterministic flow of the PDMP. The PDMP samplers we are interested in (see Section 2.2) have flow corresponding to free transport, with corresponding Ξ operator for continuously differentiable f,

$$\Xi f(x, v) = v \cdot \nabla_x f(x, v).$$

Remark 2. To reiterate, while the domain of the strong generator $\mathcal{D}(A)$ is not known, the domain $\mathcal{D}(\mathfrak{A})$ is known and $\mathcal{D}(A) \subset \mathcal{D}(\mathfrak{A})$.

3. A general framework for the invariant measure of a PDMP

A challenge in the piecewise smooth setting is that the usual approach to constructing and working with PDMPs does not work without changing the topology. In particular, existing results concerning the invariant measure of such processes require the process to be Feller. For PDMPs with boundaries, this is in fact not the case in general [11, (27.5)].

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

3.1. Strong continuity of the semigroup

First, we give a general result that is not tied to our specific context and is valid for any PDMP that follows Davis's construction, [11, Section 24, (24.8)].

Let $\mathcal{F}$ be the space of C^1 functions contained in $\mathcal{D}(\mathfrak{A})$ with compact support.

Proposition 3. *Assume that the deterministic flow and the jump rates are bounded on any compact set, and that Qf has compact support whenever f has compact support. Then $\mathcal{F} \subset \mathcal{B}_0$. In other words, the semigroup P_t of the process is strongly continuous on $\mathcal{F}$:*

$$P_t f \to f \text{ for all } f \in \mathcal{F},$$

in $\|\cdot\|_\infty$, as $t \downarrow 0$.

Proof. Let $f \in \mathcal{F}$. Since $f \in \mathcal{D}(\mathfrak{A})$,

$$C_t^f := f(Z_t) - f(Z_0) - \int_0^t \mathfrak{A}f(Z_s)\,\mathrm{d}s, \quad t \geq 0,$$

is a local martingale. Furthermore, by examining the expression for $\mathfrak{A}f$, one sees that it can be rewritten as

$$\mathfrak{A}f(z) = \Xi f(z) + \lambda(z)Qf(z) - \lambda(z)f(z)$$

from (4). Since f and Qf have compact support and are bounded, using the assumptions on Ξ and λ, we deduce that $\mathfrak{A}f$ is bounded.

Since f and $\mathfrak{A}f$ are bounded, C_t^f is bounded for any fixed t, which implies that it is a martingale. More precisely, consider the stopped process $C_{t\wedge T}^f$ for any $T > 0$. This is a uniformly bounded local martingale, and is hence a true martingale.

We have $C_0^f = 0$; hence, for any starting point z and $t > 0$,

$$\mathbb{E}_z[C_t^f] = 0.$$

Hence

$$P_t f(z) - f(z) = \int_0^t P_s \mathfrak{A}f(z)\,\mathrm{d}s,$$

where we used Fubini's theorem to swap the integral and the expectation. Since P_t is a contraction for the sup norm, we see that

$$\begin{aligned}\|P_t f - f\|_\infty &\leq \int_0^t \|P_s \mathfrak{A}f\|_\infty\,\mathrm{d}s \\ &\leq \int_0^t \|\mathfrak{A}f\|_\infty\,\mathrm{d}s \\ &\leq t\|\mathfrak{A}f\|_\infty.\end{aligned}$$

We thus conclude that $P_t f - f \to 0$ as $t \downarrow 0$. □

Remark 3. The set $\mathcal{F}$ does not capture every function of $\mathcal{B}_0$, nor is it invariant under P_t. We will not attempt to prove that $\mathcal{F}$ is a core of the infinitesimal generator.

Recall that a set of functions $\mathcal{F}_0 \subset \mathcal{B}(E)$ separates measures if for any probability measures μ_1, μ_2 on E, $\int f\,\mathrm{d}\mu_1 = \int f\,\mathrm{d}\mu_2$ for all $f \in \mathcal{F}_0$ implies that $\mu_1 = \mu_2$. In order to study the invariant measure through the semigroup and its effect on functions, it is important to consider sets

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

of functions which separate measures. Therefore, we will now show that $\mathcal{F}$ separates measures on E^0.

Proposition 4. *Assume that the jump kernel Q is such that for any $z \in \Gamma$, the measure $Q(\cdot\,|z)$ is supported on the boundary $\cup_k \partial^- E_k^0$. Then $\mathcal{F}$ separates measures on E^0.*

Proof. Consider the set of C^1 functions $f : E \to \mathbb{R}$ which are compactly supported on each open set E_k^0. The collection of such functions separates measures on E^0.

We will show that such functions belong to $\mathcal{D}(\mathfrak{A})$, and hence to $\mathcal{F}$, by using the explicit characterization of $\mathcal{D}(\mathfrak{A})$ from [11, Theorem 26.14].

Firstly, if f is C^1 with compact support, Conditions 1 and 3 of [11, Theorem 26.14] are automatically satisfied.

It remains to check the boundary condition:

$$f(z) = \int_E f(y) Q(\mathrm{d}y|z) = Qf(z), \quad z \in \Gamma. \tag{5}$$

However, since we are considering only functions f which are compactly supported on each E_k^0, it holds that $f(z) = 0$ for any $z \in \Gamma$ on the boundary. Recalling that $Q(\cdot\,|z)$ is supported only on the boundary, we have that $Qf(z) = 0$ also for any $z \in \Gamma$. It hence follows that the boundary condition is satisfied. □

Remark 4. The assumption of Proposition 4 is restrictive and does not cover, for example, the case of variable selection described in [10]. However, it is possible to weaken the assumption to cover the variable selection case. A sketch is available in Appendix A.

3.2. Invariant measure

We now turn to giving conditions for the invariant measure of our PDMP. The following lemma will be important within the proof, as it allows us to ignore contributions from the boundary when calculating expectations over the path of the PDMP.

Lemma 3. *For all $z \in E$, the process starting from z spends a negligible amount of time on the boundary: for any $t > 0$,*

$$\int_0^t 1_{Z_s \in E \setminus E^0} \, \mathrm{d}s = 0,$$

$\mathbb{P}_z$-almost surely.

Proof. In Davis's construction, the number of events, including jumps at the boundary, is countable for every trajectory. Hence, for every trajectory of the process, the set of times for which $Z_t \in E \setminus E^0$ is countable and therefore negligible. □

Theorem 1. *Let μ be a measure on E. Assume the following conditions hold:*

1. *The vector field of the deterministic flow and the jump rates are bounded on any compact set.*
2. *Qf has compact support whenever f has compact support.*
3. *For any $z \in \Gamma$, the measure $Q(\cdot|z)$ is supported on the boundary $\cup_k \partial^- E_k^0$.*
4. *We have $\mu(E \setminus E^0) = 0$.*

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

5. *For all* $f \in \mathcal{D}(A)$, $\int_E Af \, \mathrm{d}\mu = 0$.

Then μ *is invariant.*

Proof. Using this fact and Proposition 1.5 from Ethier and Kurtz [17] (or [11, (14.10)]), we note that for any $f \in \mathcal{B}_0$ and $t > 0$, we have that $\int_0^t P_s f \, \mathrm{d}s \in \mathcal{D}(A)$, the domain of the strong generator. We also have that

$$\int_E P_t f \, \mathrm{d}\mu = \int_E \left[f + A \int_0^t P_s f \, \mathrm{d}s \right] \mathrm{d}\mu = \int_E f \, \mathrm{d}\mu, \tag{6}$$

where we have used our assumption that $\int Ag \, \mathrm{d}\mu = 0$ for any $g \in \mathcal{D}(A)$ and taken $g = \int_0^t P_s f \, \mathrm{d}s$.

Let $f_B(z) := f(z)1_{z \in E \setminus E^0}$ and $f_0(z) := f(z)1_{z \in E^0}$ be a decomposition of f with $f = f_0 + f_B$. Let $1_B(z) = 1_{z \in E \setminus E^0}$ be the indicator function of the boundary $E \setminus E^0$.

From Lemma 3, $\int_0^t P_s 1_B(z) \, \mathrm{d}s = 0$ for all $z \in E$. Hence we have that $\int_E \left[\int_0^t P_s 1_B(z) \, \mathrm{d}s \right] \mathrm{d}\mu(z) = 0$, and by using Fubini's theorem, that

$$\int_0^t \int_E P_s \, 1_B \, \mathrm{d}\mu \, \mathrm{d}s = 0.$$

By the non-negativity of $P_t 1_B$, there exists a null set $\mathcal{N} \subset \mathbb{R}^+$ such that for all $t \in \mathbb{R}^+ \setminus \mathcal{N}$,

$$\int_E P_t \, 1_B \, \mathrm{d}\mu = 0.$$

For all z, $f_B(z) \leq \|f\|_\infty 1_B(z)$, so for all $t \notin \mathcal{N}$, $\int_E P_t f_B \, \mathrm{d}\mu = 0$. Hence $\int_E P_t f \, \mathrm{d}\mu = \int_E P_t f_0 \, \mathrm{d}\mu$ for all $t \notin \mathcal{N}$. Since μ is supported on E^0, $\int_E f \, \mathrm{d}\mu = \int_E f_0 \, \mathrm{d}\mu$ and we deduce using (6) that for all $t \notin \mathcal{N}$,

$$\int_E P_t f_0 \, \mathrm{d}\mu = \int_E f_0 \, \mathrm{d}\mu. \tag{7}$$

Let μ_t be the law Z_t with $Z_0 \sim \mu$. Let μ_t^0 and μ_t^B be the measures defined by $\mu_t^0(A) = \mu_t(A \cap E^0)$ and $\mu_t^B(A) = \mu_t(A \cap (E \setminus E^0))$. Using (7), for all $t \notin \mathcal{N}$,

$$\int_{E^0} f_0 \, \mathrm{d}\mu_t^0 = \int_E P_t f_0 \, \mathrm{d}\mu = \int_E f_0 \, \mathrm{d}\mu = \int_{E_0} f_0 \, \mathrm{d}\mu^0.$$

Since $\mathcal{F}$ separates measures on E_0 by Proposition 4, $\mu_t^0 = \mu^0$ for all $t \notin \mathcal{N}$. Furthermore, $\mu(E) = \mu_t(E)$ and $\mu(E) = \mu^0(E)$; hence $\mu_t^0(E) = \mu_t(E)$ and $\mu_t^B(E) = 0$. Thus $\mu_t = \mu_t^0 = \mu^0 = \mu$ for all $t \notin \mathcal{N}$.

Let $t_1, t_2 \notin \mathcal{N}$. Then $\mu_{t_1} = \mu_{t_2} = \mu$, and for all functions g which are measurable and bounded, it holds that

$$\int_E P_{t_1+t_2} g \, \mathrm{d}\mu = \int_E P_{t_1}(P_{t_2} g) \, \mathrm{d}\mu = \int_E P_{t_2} g \, \mathrm{d}\mu_{t_1} = \int_E P_{t_2} g \, \mathrm{d}\mu = \int_E g \, \mathrm{d}\mu.$$

Hence $\mu_{t_1+t_2} = \mu$. To conclude, since $\mathcal{N}$ is a null set, for all $t > 0$, there exist $t_1, t_2 \notin \mathcal{N}$ such that $t = t_1 + t_2$, and therefore $\mu_t = \mu$. □

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

4. PDMP samplers with active boundaries

Let U_k be an open set of $\mathbb{R}^{d_k}$ for all $k \in K$, and let $\mathcal{V}^k \subset \mathbb{R}^{d_k}$ be the velocity space associated to the PDMP sampler on U_k. We consider the state space defined following the construction of Davis described in Section 2.1: first, set

$$E_k^0 = U_k \times \mathbb{R}^{d_k}.$$

Remark 5. We do not take $E_k^0 = U_k \times \mathcal{V}^k$ because E_k^0 must be open and $\mathcal{V}^k$, the set of velocities of the PDMP sampler, might not be.

Let π be a measure on the disjoint union $\cup_k U_k$ with a density π_k on each U_k, where π_k can be extended continuously to the closure $\bar{U}_k$. Let p_k be the marginal velocity probability distribution on $\mathbb{R}^{d_k}$ with support on $\mathcal{V}^k$, and let $\mu(k, x, v) = \pi_k(x)p_k(v)$ be a density on $\cup_{k\in K} U_k \times \mathcal{V}^k$ with respect to a suitable dominating measure (which will have discrete components when $\mathcal{V}^k$ is a finite set), defining a measure on E.

The core result of this section relies on integration by parts, and as such requires extra assumptions on the sets U_k. For clarity of exposition, we give here an intuitive version of the required assumptions; a detailed version can be found in Assumptions 3 and 4 in the appendix.

Assumption 1. *Assumptions 3 and 4 can be informally described as follows:*

(i) *U_k has no interior discontinuities on a $(d_k - 1)$-dimensional subset.*

(ii) *The boundary ∂U_k can be decomposed into a finite union of smooth parts, on each of which the normal is well-defined.*

(iii) *The set of corner points, on which the normals of the boundary are not defined, is negligible in an appropriate sense.*

(iv) *For each $x \in U_k$, $v \in \mathcal{V}^k$, there is a finite number of intersections between each line $x + \mathbb{R}v$ and ∂U_k.*

(v) *For each $v \in \mathcal{V}^k$, the projection of the points on the boundary, which are tangent to the velocity, onto $H_v = \text{span}(v)^\perp$ is negligible in an appropriate sense.*

Let N_k be the subset of points x on the boundary ∂U_k where the normal $n(x)$ is properly defined (see (10) of the appendix for a precise statement).

Assumption 2.

(i) *We have $\int |\lambda(z)| \, \mathrm{d}\mu < \infty$.*

(ii) *For all $k \in K$, π_k is C^1 in $\bar{U}_k$.*

(iii) *For any $k \in K$, and any $v \in \mathcal{V}^k$, $\nabla \pi_k \cdot v$ is in $L_1(\text{Leb})$.*

Theorem 2. *If Assumptions 2, 3, and 4 hold, then for all $f \in \mathcal{D}(A)$,*

$$\int_E Af \, \mathrm{d}\mu = -\int_{E^0} f(k, x, v)\nabla_x \mu \cdot v \, \mathrm{d}x \, \mathrm{d}v + \int_{E^0} \lambda(z)[Qf(z) - f(z)] \, \mathrm{d}\mu$$
$$+ \sum_{k\in K} \int_{x\in\partial U_k \cap N_k, v\in\mathcal{V}^k} f(k, x, v)\pi_k(x)\langle n(x), v\rangle \, \mathrm{d}\sigma(x) \, \mathrm{d}p(v),$$

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

where $f(k, x, v)$ is defined as $f(k, x, v) := \lim_{t \downarrow 0} f(k, x - tv, v)$, for $x \in \partial U_k \cap N_k$, $v \in \mathbb{R}^{d_k}$ such that $\langle n(x), v \rangle > 0$, and σ is the Lebesgue measure induced on the surface ∂U_k.

Proof (sketch). The outline of the proof is that we first use the definition of the generator acting on a function Af and then rearrange the resulting integral using integration by parts. If our PDMP has no boundary, this will give just the first two terms on the right-hand side [18, 31]. The effect of the boundaries is to introduce the additional terms when performing integration by parts. For full details of the proof, see Section B.2. □

Corollary 1. *If Assumptions 2, 3, and 4 hold, then for all $f \in \mathcal{D}(A)$,*

$$\int_E Af \, \mathrm{d}\mu = -\int_{E^0} f(k, x, v) \nabla_x \mu \cdot v \, \mathrm{d}x \, \mathrm{d}v + \int_{E^0} \lambda(z)[Qf(z) - f(z)] \, \mathrm{d}\mu$$
$$+ \sum_{k \in K} \int_{\partial^- E_k} [f(k, x, v) - Qf(k, x, -v)] \pi_k(x) \langle n(x), v \rangle \, \mathrm{d}\sigma(x) \, \mathrm{d}p(v). \tag{8}$$

Proof. Plugging the boundary condition (5) that must be satisfied by any $f \in \mathcal{D}(A)$ into the result of Theorem 2 yields the result. □

To show that a PDMP has the correct invariant distribution, we need to show that $\int_E Af \, \mathrm{d}\mu = 0$. It is difficult to give simple criteria for this in full generality, so in what follows we will focus on samplers where the position is unchanged at the boundary. This is consistent with the dynamics of current PDMP samplers where trajectories are continuous, and events of the PDMP only change the velocity. A natural approach for constructing other sampling behaviour at the boundary is to work with samplers for which $f(k, x, v) \nabla_x \mu \cdot v \, \mathrm{d}x \, \mathrm{d}v + \int_{E^0} \lambda(z)[Qf(z) - f(z)] \, \mathrm{d}\mu = 0$. In this case, we only need to verify that the dynamics at the boundary are such that the remaining term in (8) simplifies to 0: see [19] for a condition for a class of moves on the boundary for which this term vanishes. In more generality, such as for the variable selection sampler of [10], we can also introduce additional events into the sampler to compensate for the boundary term.

5. PDMP samplers for piecewise continuous densities

Let π be a density on $\mathbb{R}^d$, and let $\{U_k : k \in K\}$ be a finite collection of disjoint open subsets of $\mathbb{R}^d$, such that $\overline{\cup_k U_k} = \mathbb{R}^d$, which satisfy our technical Assumptions 3 and 4. We assume that π is C^1 on each $\bar{U}_k$. We are now in the same setting as in the previous section.

Let $\partial U = \bigcup_{k \in K} (\partial U_k \cap N_k)$ be the union of the boundaries, i.e. the set of discontinuities of π. We consider now only points on the boundary where exactly two sets U_{k_1}, U_{k_2} intersect, and where the respective normals are well-defined; the set of points where more sets intersect or the normal is ill-defined form a null set by assumption and thus have no impact on the resulting invariant distribution.

In the following we will restrict ourselves to transition kernels on the boundary that keep the location, x, unchanged and only update the velocity, v.

For each such x in ∂U, there exist $k_1(x)$ and $k_2(x)$ such that $x \in \bar{U}_{k_1}$ and $x \in \bar{U}_{k_2}$. We will define the ordering of the labels so that $\pi_{k_1}(x) < \pi_{k_2}(x)$. Let $n(x)$ be the outer normal for U_{k_1}. Thus this is the normal that points to the region $k_2(x)$, which is the region that has higher density, under π, at x.

Let $\mathcal{V}_x^+ = \{v \in \mathcal{V} | \langle v, n(x) \rangle > 0\}$ and $\mathcal{V}_x^- = \{v \in \mathcal{V} | \langle v, n(x) \rangle < 0\}$. Thus $\mathcal{V}_x^+$ is the set of velocities that would move the position x into $k_2(x)$, thereby increasing the density under π, and $\mathcal{V}_x^-$ is the set of velocities that move the position into $k_1(x)$.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

For $x \in \partial U$, let l_x be the following (unnormalized) density on $\mathcal{V}$:

$$l_x(v) = \begin{cases} |\langle n(x), v\rangle| p(v) \pi_{k_2(x)}(x) & \forall v \in \mathcal{V}_x^+, \\ |\langle n(x), v\rangle| p(v) \pi_{k_1(x)}(x) & \forall v \in \mathcal{V}_x^-. \end{cases}$$

This is just proportional to the density $p(v)$ weighted by the size of the velocity in the direction of the normal $n(x)$ and weighted by the density at x in the region, either $k_1(x)$ or $k_2(x)$, that the velocity is moving toward.

Consider the transition $(x, v) \to (x', v')$ obtained by first flipping the velocity and then applying the Markov kernel Q. Since we assume that Q only changes the velocity, there exists a Markov kernel Q'_x such that this transition can be described as $\delta_x(\mathrm{d}x') Q'_x(\mathrm{d}v'|v)$. One can equally define this kernel by $Q'_x(A|v) = Q(\{x\} \times A|x, -v)$ for any measurable A.

Theorem 3. *Assume for all $v \in \mathcal{V}_x$ that $p(v) = p(-v)$, and that the transition kernel Q only changes the velocity, and define the family of kernels Q'_x for $x \in \partial U$ as above. Furthermore, assume that*

$$-\int_{E^0} f(k, x, v) \nabla_x \mu \cdot v \, \mathrm{d}x \, \mathrm{d}v + \int_{E^0} \lambda(z)[Qf(z) - f(z)] \, \mathrm{d}\mu = 0,$$

and that

$$\forall x \in \partial U, \, l_x \text{ is an invariant density of } Q'_x. \tag{9}$$

Then $\int_E Af \, \mathrm{d}\mu = 0$ for all $f \in \mathcal{D}(A)$, and μ is the invariant distribution of the process.

Proof. Starting from Theorem 2, for all $f \in \mathcal{D}(A)$,

$$\begin{aligned} \int_E Af \, \mathrm{d}\mu = &- \int_{E^0} f(k, x, v) \nabla_x \mu \cdot v \, \mathrm{d}x \, \mathrm{d}v + \int_{E^0} \lambda(z)[Qf(z) - f(z)] \, \mathrm{d}\mu \\ &+ \sum_{k \in K} \int_{x \in \partial U_k \cap N_k, v \in \mathcal{V}} f(k, x, v) \pi_k(x) \langle n(x), v\rangle \, \mathrm{d}\sigma(x) \, \mathrm{d}p(v). \end{aligned}$$

By assumption, $-\int_{E^0} f(k, x, v) \nabla_x \mu \cdot v \, \mathrm{d}x \, \mathrm{d}v + \int_{E^0} \lambda(z)[Qf(z) - f(z)] \, \mathrm{d}\mu = 0$, and so we simplify the integral of Af to

$$\int_E Af \, \mathrm{d}\mu = \sum_{k \in K} \int_{x \in \partial U_k \cap N_k, v \in \mathcal{V}} f(k, x, v) \pi_k(x) \langle n(x), v\rangle \, \mathrm{d}\sigma(x) \, \mathrm{d}p(v).$$

To simplify notation, in the rest of the proof we write k_1 for $k_1(x)$ and k_2 for $k_2(x)$. We rewrite the previous equation as follows:

$$\begin{aligned} \int_E Af \, \mathrm{d}\mu = &\int_{x \in \partial U} \int_{v \in \mathcal{V}_x^+} \left[f(k_2, x, v) \pi_{k_2}(x) - f(k_1, x, v) \pi_{k_1}(x) \right] |\langle n(x), v\rangle| \, \mathrm{d}p(v) \, \mathrm{d}\sigma(x) \\ &- \int_{x \in \partial U} \int_{v \in \mathcal{V}_x^-} \left[f(k_2, x, v) \pi_{k_2}(x) - f(k_1, x, v) \pi_{k_1}(x) \right] |\langle n(x), v\rangle| \, \mathrm{d}p(v) \, \mathrm{d}\sigma(x). \end{aligned}$$

Using the fact that if $v \in \mathcal{V}_x^-$ then $-v \in \mathcal{V}_x^+$, we can rewrite the right-hand side as

$$\begin{aligned} \int_{x \in \partial U} \Bigg(&\int_{v \in \mathcal{V}_x^+} \left[f(k_2, x, v) \pi_{k_2}(x) - f(k_1, x, v) \pi_{k_1}(x) \right] |\langle n(x), v\rangle| \, \mathrm{d}p(v) \\ &- \int_{v \in \mathcal{V}_x^+} \left[f(k_2, x, -v) \pi_{k_2}(x) - f(k_1, x, -v) \pi_{k_1}(x) \right] |\langle n(x), v\rangle| \, \mathrm{d}p(-v) \Bigg) \, \mathrm{d}\sigma(x). \end{aligned}$$

A sufficient condition for $\int_E Af \, \mathrm{d}\mu = 0$ is that for all x the integral over v in the brackets is 0.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

Any $f \in \mathcal{D}(A)$ satisfies the boundary condition (5) on Γ. For $v \in \mathcal{V}_x^+$, we have $(k_1, x, v) \in \Gamma$ and $(k_2, x, -v) \in \Gamma$; hence,

$$f(k_1, x, v) = \int f(z')Q(\mathrm{d}z'|(k_1, x, v)),$$

$$f(k_2, x, -v) = \int f(z')Q(\mathrm{d}z'|(k_2, x, -v)).$$

Thus our sufficient condition for $\int_E Af \, \mathrm{d}\mu = 0$ becomes

$$\int_{\mathcal{V}_x^+} [f(k_2, x, v)\pi_{k_2}(x) - Qf(k_1, x, v)\pi_{k_1}(x)]|\langle n(x), v\rangle| \, \mathrm{d}p(v)$$

$$= \int_{\mathcal{V}_x^+} [Qf(k_2, x, -v)\pi_{k_2}(x) - f(k_1, x, -v)\pi_{k_1}(x)]|\langle n(x), v\rangle| \, \mathrm{d}p(-v).$$

Using again the fact that if $v \in \mathcal{V}_x^+$ then $-v \in \mathcal{V}_x^-$, this condition can be rewritten as follows:

$$\int_{\mathcal{V}_x^+} f(k_2, x, v)\pi_{k_2}(x)|\langle n(x), v\rangle| \, \mathrm{d}p(v) + \int_{\mathcal{V}_x^-} f(k_1, x, v)\pi_{k_1}(x)|\langle n(x), v\rangle| \, \mathrm{d}p(v)$$

$$= \int_{\mathcal{V}_x^+} Qf(k_2, x, -v)\pi_{k_2}(x)|\langle n(x), v\rangle| \, \mathrm{d}p(-v)$$

$$+ \int_{\mathcal{V}_x^-} Qf(k_1, x, -v)\pi_{k_1}(x)|\langle n(x), v\rangle| \, \mathrm{d}p(-v).$$

We can then write this in terms of l_x and Q'_x by introducing a function $f'(x, v)$ that is defined as

$$f'(x, v) = f(k_1, x, v) \text{ if } v \in \mathcal{V}_x^-, \text{ and } f'(x, v) = f(k_2, x, v) \text{ if } v \in \mathcal{V}_x^+.$$

Then, using $p(v) = p(-v)$ and the definitions of l_x and Q'_x, our sufficient condition becomes

$$\int_{\mathcal{V}_x} f'(x, v) \, \mathrm{d}l_x(v) = \int_{\mathcal{V}_x} \left(\int_{\mathcal{V}_x} f'(x, v')Q'_x(\mathrm{d}v'|v) \right) \mathrm{d}l_x(v).$$

This is true if l_x is an invariant density of Q'_x. □

6. Boundary kernels for usual PDMP samplers

We give here possible Markov kernels for the bouncy particle sampler, the zigzag sampler, and the coordinate sampler. Since the condition of Theorem 3 only depends on the velocity distribution, any two processes that share the same velocity distribution can use the same boundary Markov kernels. We present two approaches to constructing appropriate kernels on the boundary.

6.1. Sampling l using Metropolis–Hastings

Recall from Theorem 3 that a valid kernel for the velocity when we hit a boundary can be constructed as follows: first, construct a transition kernel Q'_x which leaves l_x invariant; then, if

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

the current velocity is v, simulate a new velocity from $Q'_x(\cdot|-v)$. That is, we can simulate a new velocity by (i) flipping the velocity and (ii) applying Q'_x.

The simplest choice of Q'_x is just the identity map, i.e. a kernel that keeps the velocity. However, this would correspond to a transition kernel on the boundary which simply flips the velocity, thus forcing the PDMP to retrace its steps. Where possible, we can improve on this by choosing Q'_x to be a kernel which samples from l_x, though it should be noted that this choice may be difficult to implement.

When $\mathcal{V}$ is bounded, an alternative is to define Q'_x as a Metropolis–Hastings kernel [15] targeting l_x, with proposals from a uniform sampler of $\mathcal{V}$. The algorithm starting from v then proceeds as follows:

1. Sample v' uniformly in $\mathcal{V}$.
2. Accept $v^* = v'$ with probability $\alpha = \frac{l_x(v')}{l_x(v)}$; otherwise set $v^* = v$.

Of course, it is also possible to iterate the Metropolis–Hastings kernel several times to get a good sample from l_x at a reasonable cost.

6.2. Limiting behaviours

A natural strategy for constructing the transition kernel for the velocity at a boundary is to consider the limiting behaviour of the sampler for a family of continuous densities which tend to a piecewise discontinuous density in an appropriate limit. We will do this for a density π with one discontinuity on a hyperplane with normal n: $\pi(x) = c_0 1_{\langle x,n\rangle<0} + c_1 1_{\langle x,n\rangle\geq 0}$ with $c_0 < c_1$. In the following we will assume $c_0 > 0$, but the extension of the arguments to the case $c_0 = 0$ is straightforward.

We can approximate π by a continuous density π_k such that $\nabla \log(\pi_k)$ is piecewise constant:

$$\pi_k(x) = c_0 1_{\langle x,n\rangle\in(\infty,-C/k]} + c_1 \exp\{k\langle x, n\rangle\} 1_{\langle x,n\rangle\in(-C/k,0]} + c_1 1_{\langle x,n\rangle\in(0,\infty)},$$

where $C = \log(c_1/c_0)$. As $k \to \infty$ we can see that π_k converges to π. In the following we will call the region where $\langle x, n\rangle \in [-C/k, 0]$ the boundary region of π_k, as this is approximating the boundary defined by the discontinuity in π.

The advantage of using the densities π_k is that the resulting behaviour of standard PDMP samplers is tractable, and, as we will see, the distribution of the change in velocity from entering to exiting the boundary region of π_k will not depend on k. The choice of k does affect the change in position of the PDMP as it moves through the boundary region. The effect of increasing k is just to reduce the time spent passing through the boundary region. In the limit as $k \to \infty$ this becomes instantaneous, and the PDMP's position will be unchanged.

We consider this limiting behaviour for the bouncy particle sampler, the coordinate sampler, and the zigzag process. Whilst we derive the transition distribution for the velocity in each case from this limiting behaviour, we will demonstrate that each distribution is valid for the corresponding sampler directly by showing that it satisfies our condition (9). The proofs of the propositions in this section are deferred to Appendix D.

6.2.1. *Limiting behaviour of the bouncy particle sampler.* Consider the bouncy particle sampler dynamics for sampling from π_k for a trajectory that enters the boundary region, and ignore any refresh events. If the state of the bouncy particle sampler is (x, v), then dynamics are such that events occur at a rate $\max\{0, -\langle v, \nabla \log \pi_k(x)\rangle\}$, and at an event the velocity is reflected in $\nabla \log \pi_k(x)$. Whilst in the boundary region, $\nabla \log \pi_k(x) = kn$.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

For any v such that $\langle v, n\rangle > 0$, it is clear that $\lambda(x, v) = 0$ for all x. Hence, the trajectory through the boundary region will be a straight line.

Let v be such that $\langle v, n\rangle < 0$. Without loss of generality assume that the trajectory enters the boundary region at $t = 0$ with $\langle x_0, n\rangle = 0$. If no jumps occur, the trajectory will exit the boundary region at some time t_e, where $\langle x_{t_e}, n\rangle = -C/k$, which implies $t_e = -C/(k\langle v, n\rangle)$. For such a trajectory, the Poisson rate whilst passing through the boundary region is $\lambda = -k\langle v, n\rangle$. Remembering that $C = \log(c_1/c_0)$, we have that the probability of a trajectory passing through the region in a straight line is

$$\exp\{-\lambda t_e\} = \exp\left\{-\log\left(\frac{c_1}{c_0}\right)\right\} = \frac{c_0}{c_1},$$

which does not depend on k.

Finally, the probability of an event that changes the velocity in the boundary region is thus $1 - c_0/c_1$. If there is an event, then the velocity reflects in the normal and becomes $v' = v - 2\langle v, n\rangle$. As $\langle v', n\rangle > 0$, no further events will occur whilst passing through the boundary region.

Hence the probability transition kernel assigns probabilities

$$Q(x, v'|x, v) = \begin{cases} 1, & v' = v \text{ and } \langle v, n\rangle > 0, \\ c_0/c_1, & v' = v \text{ and } \langle v, n\rangle < 0, \\ 1 - c_0/c_1, & v' = v - 2\langle n, v\rangle n \text{ and } \langle v, n\rangle < 0. \end{cases}$$

If we translate this into the corresponding transition kernel at a general discontinuity, for a trajectory that hits the boundary defined by the continuity at a point with unit normal $n = n(x)$, then

$$Q_{\mathrm{BPS}}(x, v'|x, v) = \begin{cases} 1, & v' = v \text{ and } v \in \mathcal{V}_x^+, \\ \pi_{k_1(x)}(x)/\pi_{k_2(x)}(x), & v' = v \text{ and } v \in \mathcal{V}_x^-, \\ 1 - \pi_{k_1(x)}(x)/\pi_{k_2(x)}(x), & v' = v - 2\langle n, v\rangle n \text{ and } v \in \mathcal{V}_x^-. \end{cases}$$

That is, if the trajectory is moving to the region of lower probability density, then it passes through the discontinuity with a probability proportional to the ratio in the probability densities. Otherwise, it reflects off the surface of discontinuity.

Proposition 5. *The transition kernel of the velocity, Q'_x, derived from Q_{BPS} satisfies* (9).

This result holds for either implementation of the bouncy particle sampler, i.e. where the distribution of the velocity is uniform on the sphere, or is an isotropic multivariate Gaussian. Examination of the proof of the proposition shows that we only require that $p(v)$ is spherically symmetric.

6.2.2. *Limiting behaviour of the coordinate sampler.* In the coordinate sampler, the velocity is always in the direction of one of the coordinates of x, and we will denote the set of $2d$ possible velocities by $\mathcal{V}$. The dynamics of the coordinate sampler are similar to those of the bouncy particle sampler except that the transition kernel at an event is different. At an event, the probability of the new velocity being $v' \in \mathcal{V}$ is proportional to $\max\{0, \langle v', \nabla \log \pi_k(x)\rangle\}$.

The calculations for the transition kernel of the coordinate sampler for a trajectory that enters the boundary region of π_k is similar to that for the bouncy particle sampler, except that, if there is an event, the distribution of the new velocity changes to that for the coordinate sampler.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

The resulting probability transition kernel, expressed for a general discontinuity, is

$$Q_{\mathrm{CS}}(x, v'|x, v) = \begin{cases} 1, & v' = v \text{ and } v \in \mathcal{V}_x^+ \\ \pi_{k_1(x)}(x)/\pi_{k_2(x)}(x), & v' = v \text{ and } v \in \mathcal{V}_x^-, \\ (1 - \pi_{k_1(x)}(x)/\pi_{k_2(x)}(x))\frac{\langle v', n\rangle}{K}, & v' \in \mathcal{V}_x^+ \text{ and } v \in \mathcal{V}_x^-, \end{cases}$$

where $K = \sum_{v \in \mathcal{V}_x^+} \langle v, n\rangle$ is a normalizing constant for the distribution of the new velocity if it changes.

That is, a trajectory moving to the higher-probability region is unaffected by the discontinuity. A trajectory moving to a lower-probability region either passes through the discontinuity or bounces. The bounce direction is chosen at random from $v' \in \mathcal{V}_x^+$ with probability equal to the component of v' in the direction of the normal at the discontinuity, n.

Proposition 6. *The transition kernel of the velocity, Q'_x, derived from Q_{CS} satisfies* (9).

6.2.3. *Limiting behaviour of the zigzag sampler.* For the zigzag process the velocities are of the form $\{\pm 1\}^d$. Given the positions, events occur independently for each component. That is, if $v_i \in \{\pm 1\}$ is the component of the velocity in the ith coordinate axis, then this velocity will flip, i.e. change sign, at a rate $\max\{0, -v_i \partial \log \pi_k(x)/\partial x_i\}$. For the boundary region of $\pi_k(x)$ we have

$$\frac{\partial \log \pi_k(x)}{\partial x_i} = k n_i,$$

where n_i is the ith component of the normal n.

If we consider the dynamics of zigzag once it enters the boundary region of $\pi_k(x)$, then each velocity component with $v_i n_i < 0$ will potentially flip. Whilst travelling through the boundary region, the rate at which such a v_i will flip will be $-v_i n_i k$. If a component flips, then the new velocity satisfies $v_i' = -v_i$, and hence $v_i' n_i > 0$ and the rate of any future flip while in the boundary region will be 0. Thus, each component will flip at most once whilst the trajectory passes through the boundary region.

The resulting dynamics are somewhat complicated, but can easily be simulated using the following algorithm:

(a) For $i = 1, \ldots, d$ simulate τ_i as independent realizations of an exponential random variable with rate $k \max\{-n_i v_i, 0\}$. If $n_i v_i \geq 0$ then set $\tau_i = \infty$.

(b) Calculate the time t^* at which we leave the boundary as the smallest value $t > 0$ for which

$$\sum_{i=1}^{n} v_i n_i (\tau_i - |t - \tau_i|) = \pm C/k,$$

or

$$\sum_{i=1}^{n} v_i n_i (\tau_i - |t - \tau_i|) = 0.$$

(c) The new velocity has $v_i' = v_i$ if $\tau_i > t^*$ and $v_i' = -v_i$ otherwise.

The key idea is that whilst the trajectory remains within the boundary region, each velocity component flips independently with its own rate. Step (a) then simulates the time at which each

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

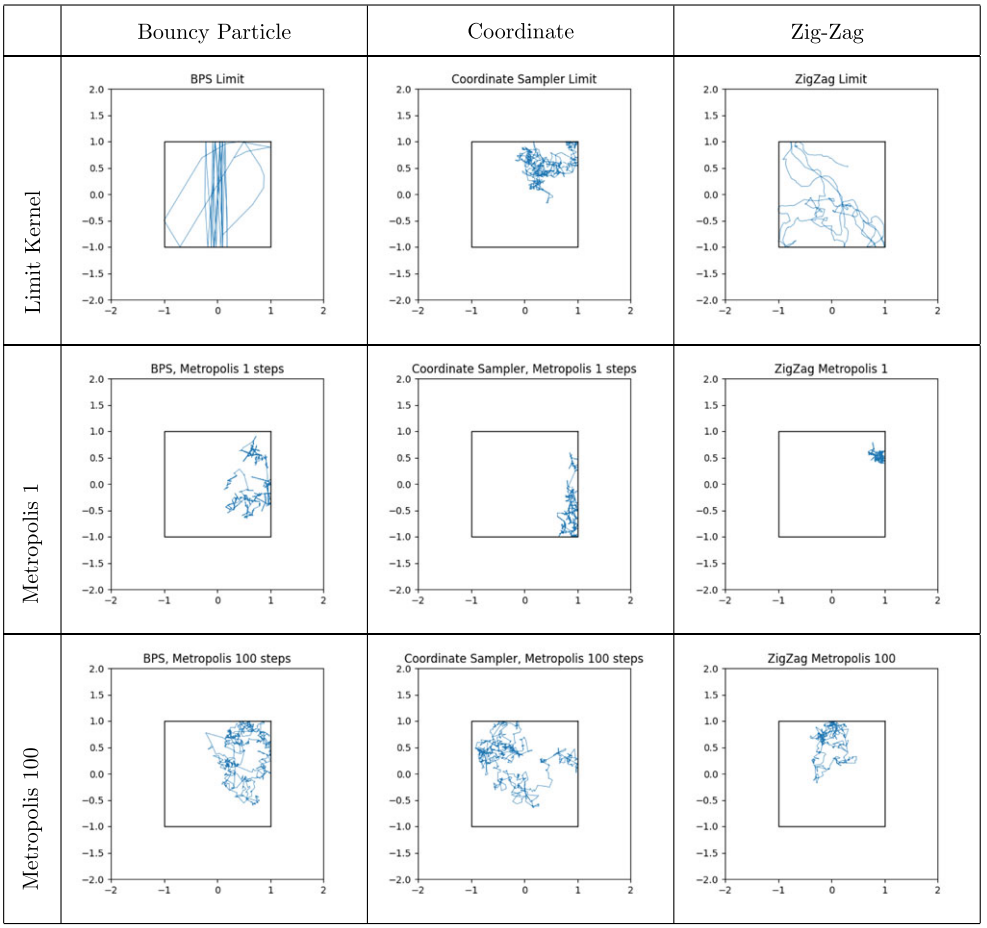


FIGURE 2. Example trajectories for the bouncy particle sampler (left), coordinate sampler (middle), and zigzag process (right) for simulating from a 100-dimensional Gaussian distribution restricted to a cube for different transitions on the boundary. For the zigzag process and coordinate sampler, the basis is rotated by a random matrix R. We show the dynamics for the first two coordinates only. The different transitions correspond to the limiting behaviour in Sections 6.2.1–6.2.3 (top); using a single Metropolis–Hastings step to sample from l_x (middle); and using 100 Metropolis–Hastings steps to sample from l_x (bottom).

component of the velocity would switch. Step (b) calculates, for the event times simulated in 1, the time t^* at which the trajectory will leave the boundary. There are two possibilities, the first corresponding to passing through the boundary, the second to bouncing back to the region where we started. For the first of these possibilities we have two options, corresponding to C/k and $-C/k$, to allow for the two possible directions with which we can enter the boundary region. Then in Step (c) we calculate the velocity once the trajectory exits the boundary region—using the fact that a velocity component will have flipped if and only if $\tau_i < t^*$.

It is simple to show that the distribution of the new velocity, v', simulated in Step (c) is independent of k, as the value of k only introduces a scale factor into the definition of the event times τ_i and the exit time t^*. Thus for a general discontinuity, we define the probability

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

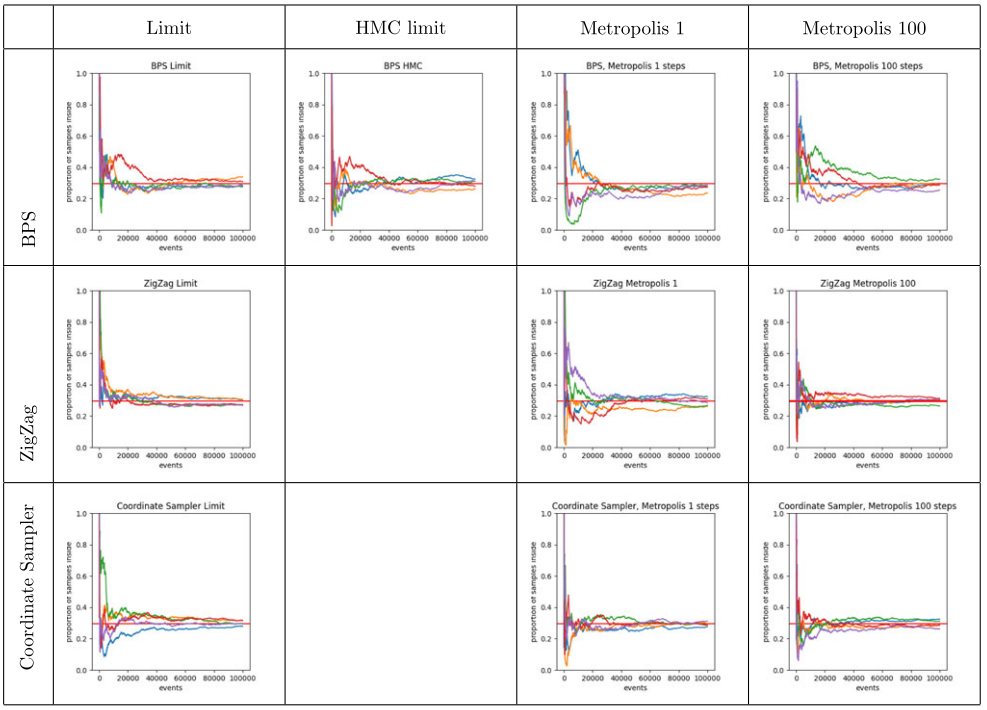


FIGURE 3. MCMC estimates over time of $\mathbb{E}_\pi[1_{[-1,1]^d}(X)]$, the expectation of the indicator function of $[-1,1]^d$ under π, for different samplers, repeated 5 times for each sampler. Here $d=20$ and the expectation is computed with respect to the d-dimensional density (proportional to) $\pi(x) = \alpha_{\text{in}} e^{-\frac{\|x\|^2}{2\sigma_{\text{in}}^2}} 1_{R^{-1}x\in[-1,1]^d} + \alpha_{\text{out}} e^{-\frac{\|x\|^2}{2\sigma_{\text{out}}^2}} 1_{R^{-1}x\notin[-1,1]^d}$, with $(\alpha_{\text{in}}, \alpha_{\text{out}}, \sigma_{\text{in}}, \sigma_{\text{out}}) = (1, 1, 2, 0.8)$, where R is a random rotation matrix. The red line corresponds to the true value of the expectation. For the bouncy particle sampler, a refresh rate of 5 is used. Each sampler ran for 100000 events.

transition kernel, Q_{ZZ}, as corresponding to the above algorithm with $k=1$, with $n=n(x)$ and $C=\log(\pi_{k_2(x)}(x)/\pi_{k_1(x)}(x))$, the log of the ratio in probability density at x for the two regions.

Proposition 7. *The transition kernel of the velocity Q'_x induced by Q_{ZZ} satisfies* (9).

6.3. Bouncy particle sampler with reflection and refraction

The bouncy particle sampler can be implemented using a Gaussian velocity distribution, similarly to Hamiltonian Monte Carlo. In that case, it is possible to use the boundary kernel developed for Hamiltonian Monte Carlo which reflects and refracts on the boundary [1]. This kernel is fully deterministic and is as follows. As before, let k_1 and k_2 denote the regions on either side of the discontinuity, with the density being higher in the region k_2, and let n be the normal that points to k_2. Let $\Delta U = \log(\pi_{k_2}(x)) - \log(\pi_{k_1}(x)) > 0$ be the (positive) difference in log density between the two sides of the boundary. If $\langle v, n\rangle < 0$ and $\langle v, n\rangle^2 > 2\Delta U$, then the process crosses the boundary and the outgoing velocity is $v' = v + \left(\sqrt{\langle v, n\rangle^2 - 2\Delta U} - \langle v, n\rangle\right)n$ (refraction). Otherwise, if $\langle v, n\rangle < 0$ and $\langle v, n\rangle^2 < 2\Delta U$, the

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

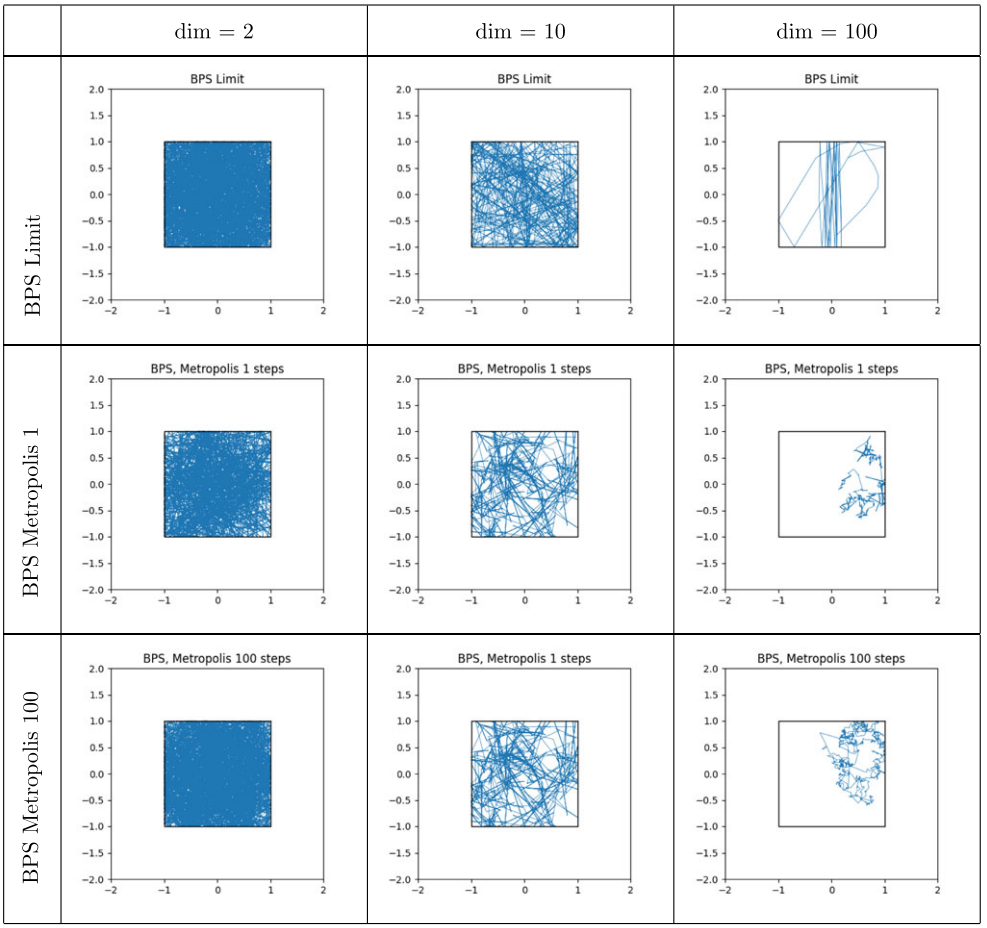


FIGURE 4. Example trajectories for the bouncy particle sampler for simulating from a d-dimensional Gaussian distribution restricted to a cube, for $d = 2$, 10, 100, and for different transitions on the boundary. For $d = 10$, 100 we show the dynamics for the first two coordinates only. The different transitions correspond to the limiting behaviour, Q_{BPS}, of Section 6.2.1 (top); using a single Metropolis–Hastings step to sample from l_x (middle); and using 100 Metropolis–Hastings steps to sample from l_x (bottom).

boundary is not crossed and the output velocity is $v' = v - 2\langle v, n\rangle n$ (reflection). If $\langle v, n\rangle > 0$ then, as we are moving to the region with higher density, we always refract, and the new velocity is $v' = v + \left(\sqrt{\langle v, n\rangle^2 + 2\Delta U} - \langle v, n\rangle\right)n$.

The idea of the refraction is that only the velocity in the direction of the normal, n, is changed. If the PDMP is moving to a region with higher density, then the velocity in the direction of n is increased; if it is moving fast enough into a region with lower density, it is reduced.

Let Q_{RR} be the probability kernel defined by the reflection and refraction process. The following result gives the validity of this kernel.

Proposition 8. *The transition kernel of the velocity, Q'_x induced by Q_{RR} satisfies* (9).

We give examples of using this boundary kernel in Figure 3 and Appendix F.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

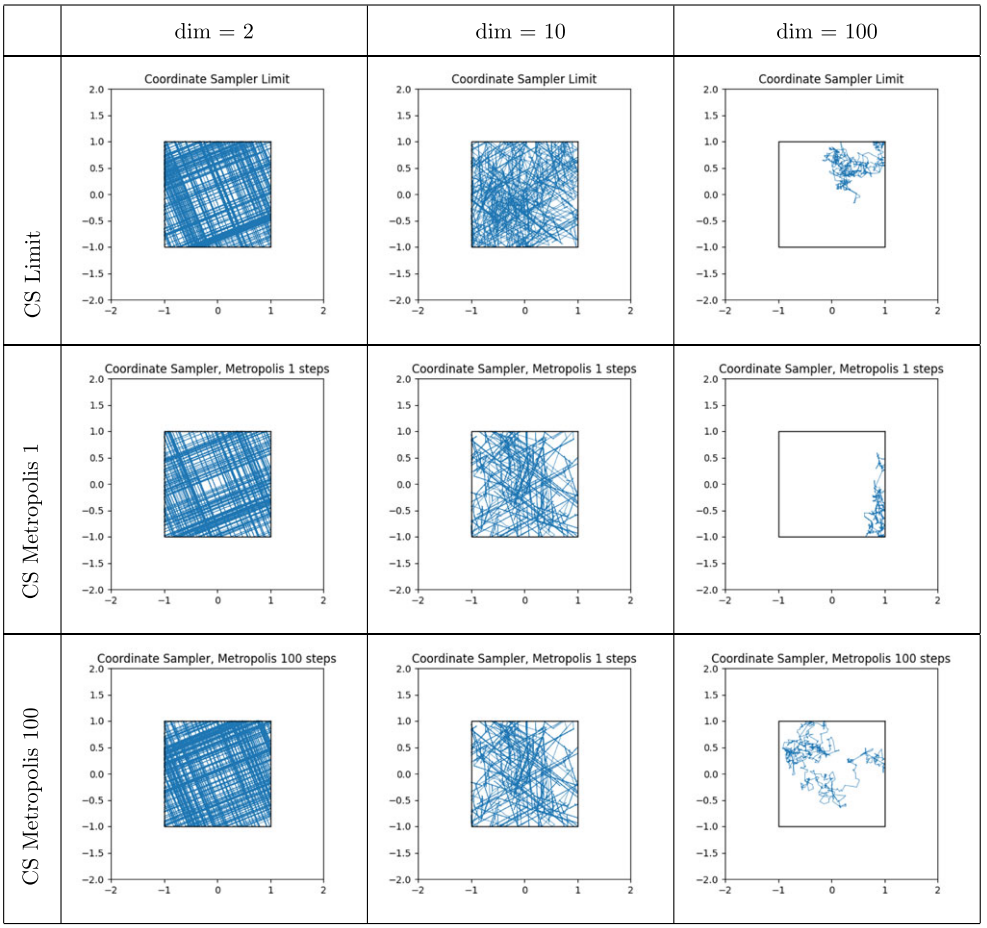


FIGURE 5. Example trajectories for the coordinate sampler for simulating from a d-dimensional Gaussian distribution restricted to a cube, for $d = 2$, 10, 100, and for different transitions on the boundary. For $d = 10$, 100 we show the dynamics for the first two coordinates only. The different transitions correspond to the limiting behaviour, Q_{CS}, of Section 6.2.2 (top); using a single Metropolis–Hastings step to sample from l_x (middle); and using 100 Metropolis–Hastings steps to sample from l_x (bottom).

7. Comparison of samplers

We now present some simulation results that aim to illustrate the theory, and show how different samplers and different choices of kernel at the discontinuities behave. We do this by considering a simple model for which it is easy to see the boundary behaviour, with a target density

$$\pi(x) = \alpha_{\mathrm{in}} e^{-\frac{\|x\|^2}{2\sigma_{\mathrm{in}}}} 1_{x\in[-1,1]^d} + \alpha_{\mathrm{out}} e^{-\frac{\|x\|^2}{2\sigma_{\mathrm{out}}}} 1_{x\notin[-1,1]^d},$$

which is Gaussian inside and outside the hypercube $[-1, 1]^d$, with a discontinuous boundary on the hypercube.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

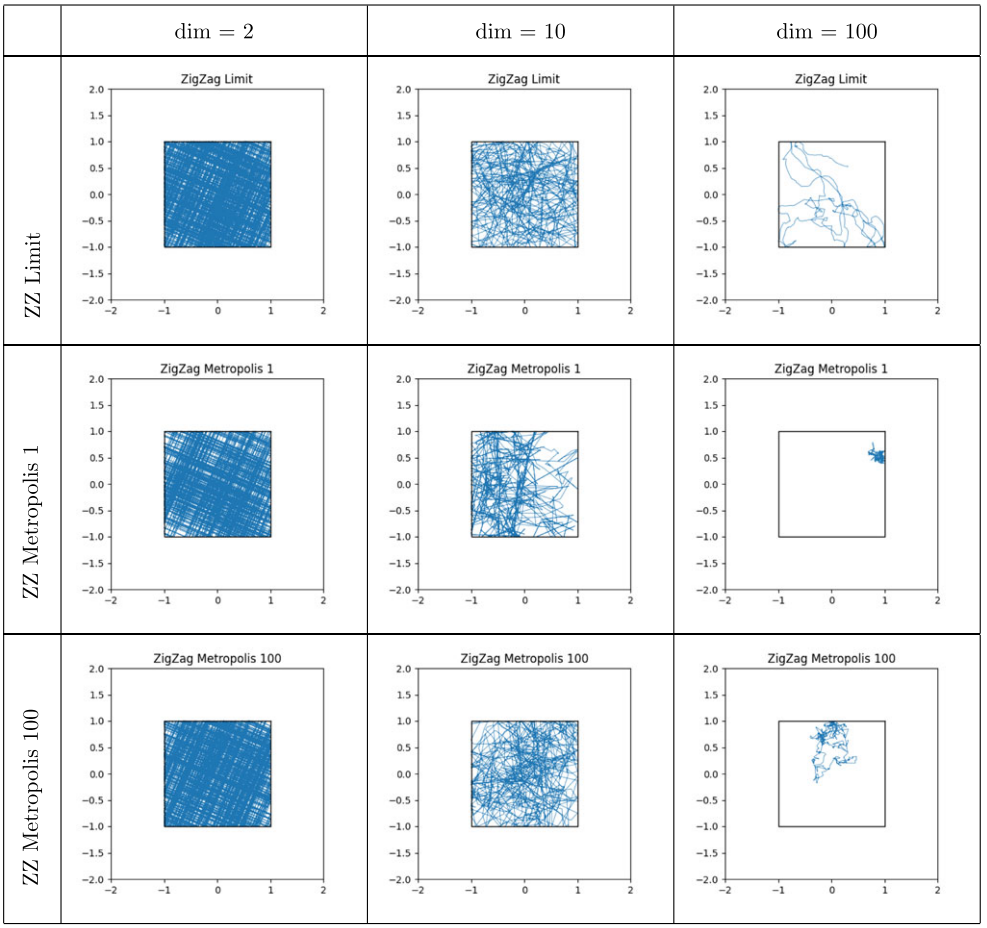


FIGURE 6. Example trajectories for the zigzag sampler for simulating from a d-dimensional Gaussian distribution restricted to a cube, for $d = 2$, 10, 100, and for different transitions on the boundary. For $d = 10$, 100 we show the dynamics for the first two coordinates only. The different transitions correspond to the limiting behaviour, Q_{ZZ}, of Section 6.2.3 (top); using a single Metropolis–Hastings step to sample from l_x (middle); and using 100 Metropolis–Hastings steps to sample from l_x (bottom).

For algorithms such as the zigzag process and the coordinate sampler, the choice of basis is extremely important. In particular, we expect the zigzag process to perform very well for product measures if the velocity basis is properly chosen. Hence we use a rotated basis where we generate a random rotation matrix R and rotate the canonical basis by R. (For results for zigzag with the canonical basis, see Appendix F.) The random matrix R is obtained by computing the polar decomposition of a random matrix composed of independent and identically distributed standard normal random variables.

Since the goal of the experiments is to highlight the boundary behaviour, and not a general comparison between the bouncy particle sampler, the zigzag process, and the coordinate sampler, we only perform basic tuning of these algorithms, in particular with respect to the refresh rate which is necessary for the bouncy particle sampler to be ergodic (without

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

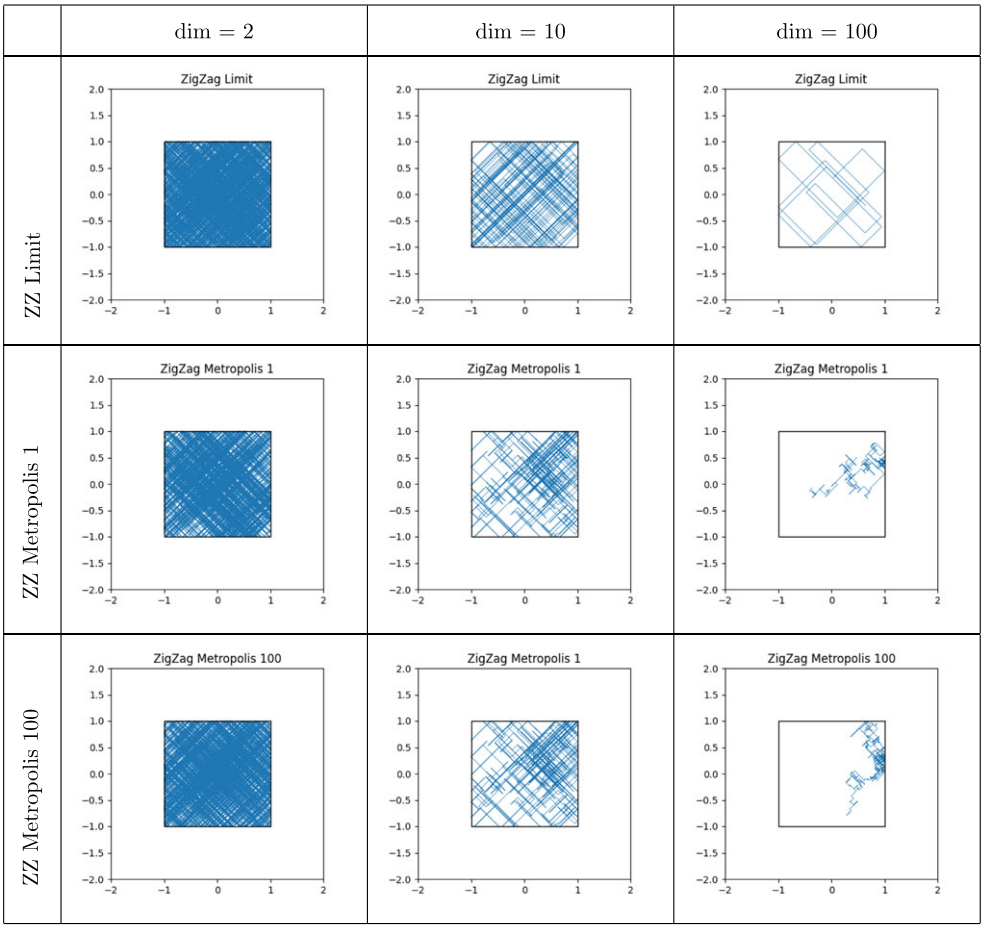


FIGURE 7. Example trajectories for the zigzag sampler for simulating from a d-dimensional Gaussian distribution restricted to a cube, for $d = 2$, 10, 100; and for different transitions on the boundary—using the canonical basis. For $d = 10$, 100 we show the dynamics for the first two coordinates only. The different transitions correspond to the limiting behaviour, Q_{ZZ} of Section 6.2.3 (top); using a single Metropolis–Hastings step to sample from l_x (middle); and using 100 Metropolis–Hastings steps to sample from l_x (bottom).

boundaries). For each sampler we consider a range of transition kernels at the discontinuity. These are the Metropolis–Hastings kernel of Section 6.1; using 1 or 100 iterations of the Metropolis–Hastings kernel; and using the kernel derived from the limiting behaviour in Sections 6.2.1–6.2.3. We have implemented all methods in dimensions $d = 2$, 10, and 100, though we only present results for $d = 100$ here in Figure 2, with the full results shown in Appendix F.

An example of the resulting trajectories for $d = 100$, for the case of a Gaussian restricted to the cube, i.e. $\alpha_{\text{out}} = 0$, can be found in Figure 2. Additional trajectory examples can be found in the appendix for other dimensions in Figures 4–7. There are a number of obvious qualitative conclusions that can be drawn. First, using a single Metropolis–Hastings kernel leads to poor

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

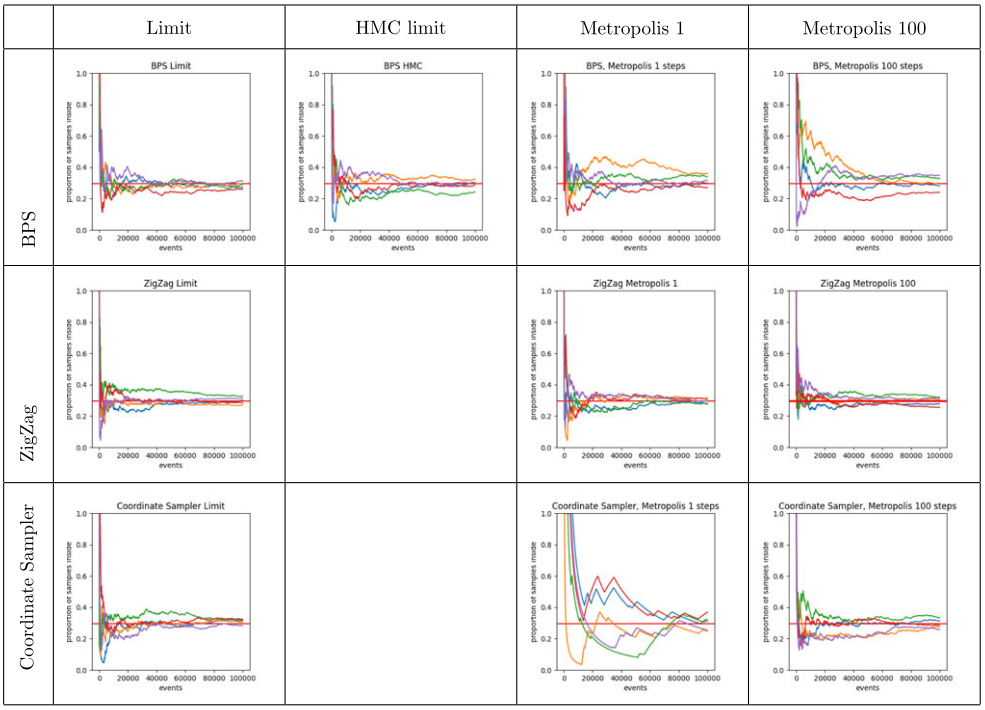


FIGURE 8. MCMC estimates over time of $\mathbb{E}_\pi[1_{[-1,1]^d}(X)]$, the expectation of the indicator function of $[-1, 1]^d$ under π, for different samplers, repeated 5 times for each sampler. Here $d = 20$ and the expectation is computed with respect to the d-dimensional density (proportional to) $\pi(x) = \alpha_{\text{in}} e^{-\frac{\|x\|^2}{2\sigma_{\text{in}}^2}} 1_{x\in[-1,1]^d} + \alpha_{\text{out}} e^{-\frac{\|x\|^2}{2\sigma_{\text{out}}^2}} 1_{x\notin[-1,1]^d}$, with $(\alpha_{\text{in}}, \alpha_{\text{out}}, \sigma_{\text{in}}, \sigma_{\text{out}}) = (1, 1, 2, 0.8)$. The red line corresponds to the true value. For the bouncy particle sampler, a refresh rate of 5 is used. Each sampler ran for 100000 events.

exploration for these examples—the trajectories often double back on themselves when they hit the boundary, and the trajectories for all three algorithms explore only a small part of the sample space. Increasing the number of Metropolis–Hastings kernels qualitatively improves exploration noticeably, but does introduce diffusive-like behaviour. For the bouncy particle sampler and the zigzag process, the kernel derived from the limiting behaviour allows for smaller changes in the velocity at the boundary. We see this as the trajectories look qualitatively different from the Metropolis kernels, with the diffusive behaviour being suppressed. Overall the bouncy particle sampler with the limiting kernel appears to mix best—though this may in part be because this sampler is known to mix well for low-dimensional summaries of the target, but less well for global properties [12].

When the density π is a product, i.e. the basis used in the zigzag process is not rotated by a random matrix R, the zigzag process does not display diffusive behaviour, as can be seen in Figure 7 in Appendix F.

Finally, while the case $\alpha_{\text{out}} = 0$ allows a clear visualization of the trajectories, we provide simulations for the general case $\alpha_{\text{out}} \neq 0$ in Figure 3, and in Figures 8–10 of Appendix F. Figures 9 and 10 provide example trajectories in dimension 20, while we verify in

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

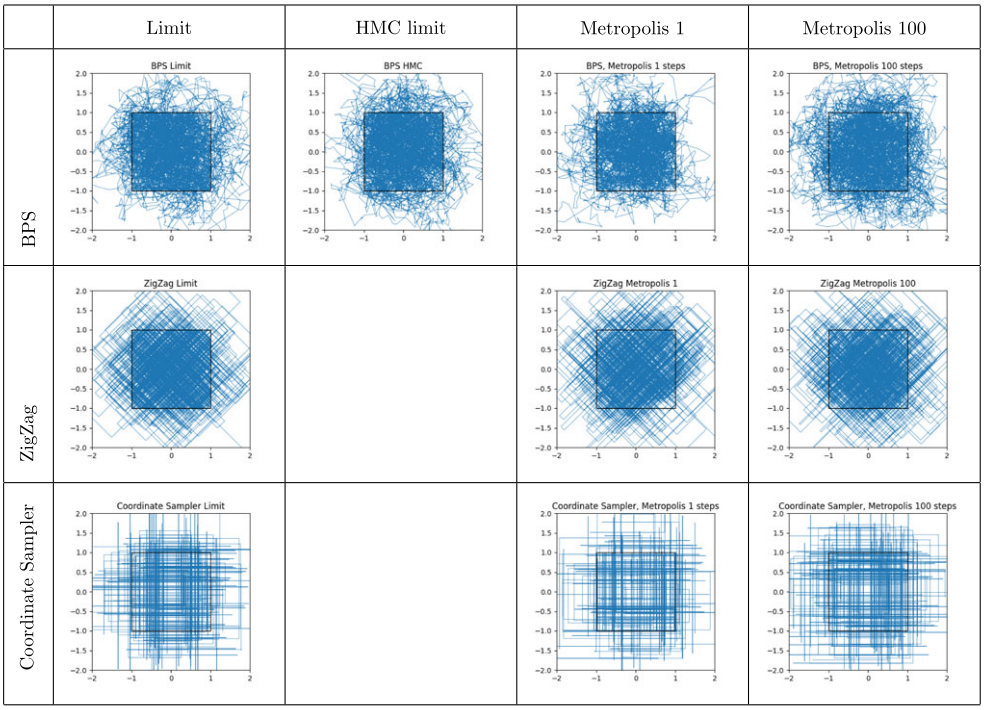


FIGURE 9. Example trajectories for each sampler using the 20-dimensional density (proportional to) $\pi(x) = \alpha_{\text{in}} e^{-\frac{\|x\|^2}{2\sigma_{\text{in}}^2}} 1_{x\in[-1,1]^d} + \alpha_{\text{out}} e^{-\frac{\|x\|^2}{2\sigma_{\text{out}}^2}} 1_{x\notin[-1,1]^d}$, with $(\alpha_{\text{in}}, \alpha_{\text{out}}, \sigma_{\text{in}}, \sigma_{\text{out}}) = (1, 1, 2, 0.8)$. For the bouncy particle sampler, a refresh rate of 5 is used. Each sampler ran for 10000 events.

Figures 3 and 8 that the integral with respect to a test function converges to the true value to show the validity of our boundary kernels, using a random orthogonal basis and the canonical basis, respectively. It is clear the variability is worse for Metropolis compared to limiting behaviours. In dimension 20 with a rotated basis, the zigzag process with the limiting behaviour seems better overall, but the simulation of the boundary kernel is significantly more difficult than for the other processes. To make a fair comparison between kernels, we would need to rescale the time axis, taking into account the complexity of each kernel.

8. Discussion

This paper focuses on PDMP-based MCMC samplers to sample densities which are only piecewise smooth. In particular, we presented a general framework for showing invariance of a given target, and then specialize to the case of the common PDMP samplers, namely the bouncy particle sampler, coordinate sampler, and zigzag sampler, when the target is piecewise smooth. Our general framework avoids the general functional-analytic approach of establishing that a given set of functions is a core [16, 17]. Rather, we make use of specific properties of the PDMP processes which we are interested in.

Since we do not use the functional-analytic and operator-theoretic framework of [16], we lose some of their powerful results, such as the existence of a tractable core for the generator

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

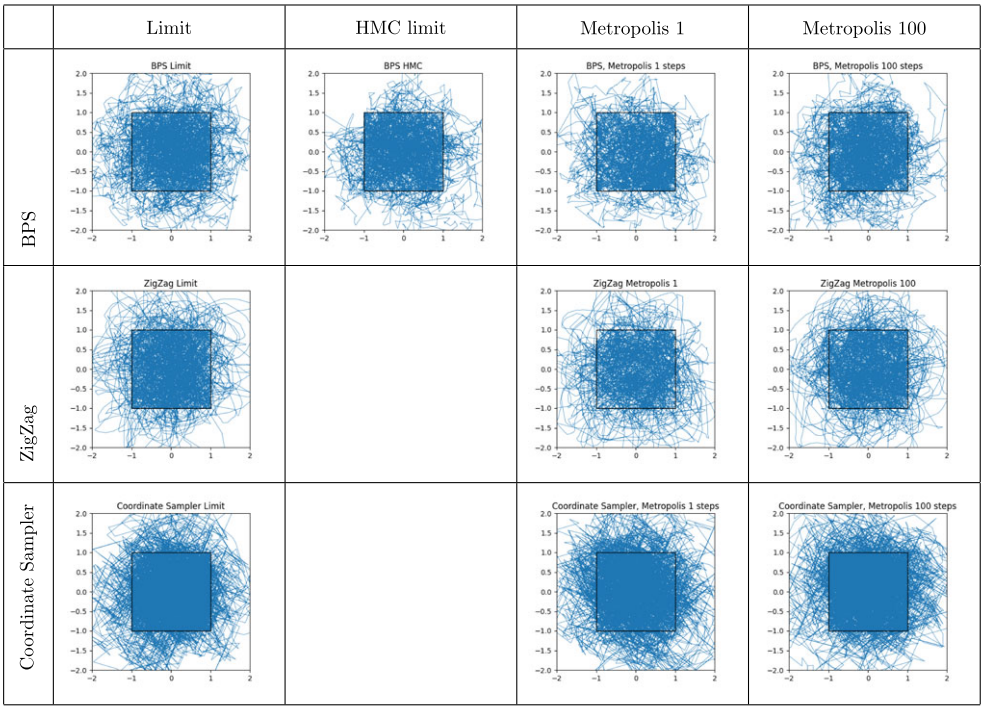


FIGURE 10. Example trajectories for each sampler using the 20-dimensional density (proportional to) $\pi(x) = \alpha_{\text{in}} e^{-\frac{\|x\|^2}{2\sigma_{\text{in}}^2}} 1_{R^{-1}x \in [-1,1]^d} + \alpha_{\text{out}} e^{-\frac{\|x\|^2}{2\sigma_{\text{out}}^2}} 1_{R^{-1}x \notin [-1,1]^d}$, with $(\alpha_{\text{in}}, \alpha_{\text{out}}, \sigma_{\text{in}}, \sigma_{\text{out}}) = (1, 1, 2, 0.8)$ and R a random rotation matrix. For the bouncy particle sampler, a refresh rate of 5 is used. Each sampler ran for 10000 events.

and useful perturbation results [16, Section 10]. However, our conditions are arguably more transparent and our general approach is considerably less technically demanding.

When the target π possesses discontinuities, we find that PDMP-based samplers display a surprisingly rich set of behaviours at the boundary, as evidenced by our empirical results, which demonstrate that the choice of jump kernel at the boundary is crucial. We see that the *limiting kernels* compare favourably to Metropolis–Hastings-based jump kernels.

Ideas from [23] for designing the distribution of the change in velocity at a jump event for standard PDMP samplers may possibly be adapted to our boundary kernels to improve their mixing properties, as there is an equivalence between the densities l_x that we introduce in Theorem 3 and the densities that are left-invariant at jumps of a PDMP (see [23]).

There remain several avenues for future exploration. We believe it is possible to weaken Assumptions 1 and 2. For example Assumption 1(ii) could be relaxed to allow for a countable union of smooth parts; it should be possible to remove Assumption 1(v) using Sard's theorem; and Assumption 2(ii) could be relaxed to the following: for all $x \in U_k$, for all $v \in \mathcal{V}^k$, $t \to \pi_k(x + tv)$ is absolutely continuous. Our chosen set of Assumptions 3 and 4 are sufficient to allow an application of integration by parts, but a simpler and more transparent set of sufficient assumptions would also be desirable. Finally, we conjecture that *non-local moves* into PDMP samplers, for example based on [32] or otherwise, might also be useful in boosting convergence in the presence of significant discontinuities.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

Appendix A. Weakened assumptions for Proposition 4

As mentioned in Remark 4, the assumption of Proposition 4 can be weakened to the following: for any compact K_1, there exists a compact K_2 such that $K_1 \subset K_2$ and for any $z \in K_2^c \cap \Gamma$, $Q(K_2, z) = 0$. In other words, it is impossible to jump from the boundary outside of K_2 to the inside of K_2.

For the sake of brevity, we only give an outline of the proof:

1. Let g_0 be a function with compact support. The boundary condition (5) is not satisfied.
2. For any $\alpha > 0$, build a sequence g_{i+1}^α such that g_{i+1}^α is C^1, $g_{i+1}^\alpha = f$ on the domain of f and $g_{i+1}^\alpha = Q g_i^\alpha$ on Γ, and $g_{i+1}^\alpha(z) = g_i^\alpha(z)$ if $d(z, \Gamma) > \alpha$.
3. Show that this sequence converges to g_α, and that g_α is C^1 with compact support and satisfies the boundary condition (5), and hence is in $\mathcal{F}$. Furthermore, $g_\alpha \to f$ pointwise when $\alpha \to 0$.
4. Use dominated convergence to conclude.

This covers the case of variable selection.

Appendix B. Theorem 2

B.1 Precise assumptions

Assumption 3. *For each $k \in K$ we assume the following:*

(i) We have $\dim_H ((\overline{U_k})^o \setminus U_k) \leq d_k - 2$, *where* $\dim_H$ *is the Hausdorff dimension.*

(ii) There is a finite collection $\{W_1^k, \ldots, W_l^k\}$ *of disjoint open sets, and a second collection of disjoint open sets,* $\{\Omega_1^k, \ldots, \Omega_l^k\}$, *such that each* $W_i^k, \Omega_i^k \subset \mathbb{R}^{d_k - 1}$ *with* $W_i^k \subset \overline{W_i^k} \subset \Omega_i^k$ *for each* $i = 1, \ldots l$.

(iii) The boundaries satisfy $\dim \partial W_i^k \leq d_k - 2$.

(iv) We have C^1 (injective) embeddings $\phi_i^k : \Omega_i^k \to \mathbb{R}^{d_k}$, *and also have continuous normals* $n_i : \Omega_i^k \to S^{d_k - 1}$.

(v) Set $M_i := \phi(\overline{W_i^k})$, *for each* $i = 1, \ldots, l$. *Then we have* $\partial U_k = M_1 \cup \cdots \cup M_l$.

(vi) The intersections satisfy $\dim M_i \cap M_j \leq d - 2$ *for any* $i \neq j$.

Let

$$N_k = \{x \in \partial U_k | \exists! i \text{ such that } \exists u \in W_i \text{ such that } \phi(u) = x\} \tag{10}$$

be the set of points of ∂U_k for which the normal $n(x) = n(u)$ is well-defined. Since $\dim_H ((\overline{U_k})^o \setminus U_k) \leq d_k - 2$, for all points for which the normal exists, the boundary separates U_k and $\mathbb{R}^{d_k} \setminus U_k$, and does not correspond to an 'internal' boundary of U_k that is removed. By convention, we assume that $n(x)$ is the outer normal.

Assumption 4. *We make the following assumptions: for all $v \in \mathcal{V}$, there is a refinement* $\{W_1^v, ..., W_m^v\}$ *and* $\{\Omega_1^v, ..., \Omega_m^v\}$ *of the boundary decomposition such that the following hold:*

(i) This new decomposition satisfies the previous assumptions.

(ii) For all $0 < i \leq m$, *there exists* $j \in \{1, \ldots, l\}$ *such that* $\phi(W_i^v) \subset \phi(W_j)$.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

(iii) *For all $x \in W_i^v$, $y \in W_j$ such that $\phi(x) = \phi(y)$, we have $n_i(x) = n_i(y)$.*

(iv) *For each $0 < i \le m$, $\dim_H p_v(M_i^v) \le n-2$, where $M_i^v = \{\phi_i(x) : x \in \overline{W}_i^v$ and $\langle v, n_i(x)\rangle = 0\}$ and p_v is the orthogonal projection on $H_v = \operatorname{span}(v)^\perp$.*

(v) *For all $0 < i \le m$ and all $x \in \mathbb{R}^n$, the sets $M_i^{v+} \cap (x + \mathbb{R}v)$ and $M_i^{v-} \cap (x + \mathbb{R}v)$ have at most one element, where $M_i^{v+} = \{x = \phi(y) \in M_i : \langle n_i(y), v\rangle > 0\}$ and $M_i^{v-} = \{x = \phi(y) \in M_i : \langle n_i(y), v\rangle < 0\}$.*

B.2. Proof of Theorem 2

We abuse notation and write $\partial_v f$ for the derivative at $t = 0$ of $f(x + tv, v)$ with respect to t for a fixed v, which corresponds to the term Ξf.

B.2.1. *Integrability.* We give two integrability lemmas that will be useful for the following proof.

Lemma 4. *Let $f \in \mathcal{D}(A)$. We know that*

(i) *f is bounded,*

(ii) *Af is bounded, and*

(iii) *$f \in \mathcal{D}(\mathfrak{A})$ and $Af = \mathfrak{A}f$.*

Proof. Items (i) and (ii) are immediate since $\mathcal{D}(A) \subset \mathcal{B}_0 \subset \mathcal{B}(E)$, and $A : \mathcal{D}(A) \to \mathcal{B}_0 \subset \mathcal{B}(E)$.

Item (iii) follows from the fact that for $f \in \mathcal{D}(A)$ there is the Dynkin formula, which exactly implies C_t^f is a true martingale ([11, (14.13)]), hence also a local martingale. □

Lemma 5. *We have $\lambda(k, x, v)(Qf(k, x, v) - f(k, x, v)) \in L_1(\mu)$ and $\partial_v f(k, x, v) \in L_1(\mu)$.*

Proof. Since Af is bounded, $Af \in L_1(\mu)$. Since f is bounded and with (i) of Assumption 2, we have $\lambda(k, x, v)(Qf(k, x, v) - f(k, x, v)) \in L_1(\mu)$. Hence $\partial_v f(k, x, v) \in L_1(\mu)$. □

This means that we can treat each term independently.

B.2.2. *Integration of the infinitesimal generator over E.*

Proposition 9. *Let $f \in \mathcal{D}(A)$ and let $k \in K$. For all $v \in \mathcal{V}^k$ we have*

$$\int_{U_k} \pi_k(x)\partial_v f(k, x, v)\,\mathrm{d}x = -\int_{U_k} f(k, x, v)\partial_v \pi_k(x)\,\mathrm{d}x\,\mathrm{d}v + \int_{\partial U_k \cap N_k} \pi_k(x) f(k, x, v)\,|\langle n(x), v\rangle|\,\mathrm{d}\sigma(x),$$

where σ is the Lebesgue measure of the boundary (seen as a Riemannian manifold).

Proof. We would like to use an integration-by-parts result on the integral in question, which is precisely detailed in Appendix C.

So now let $v \in \mathcal{V}^k$. Assumptions 3 and 4 imply that U_k satisfies Assumption 6 of Appendix C. Furthermore, $f \in \mathcal{D}(A) \subset \mathcal{D}(\mathfrak{A})$; thus for all $x \in U_k$, $t \to f(k, x + tv, v)$ is absolutely continuous. Finally, Lemma 5 implies that $\partial_v f(k, x, v) \in L_1(\mu)$; hence we can use Proposition 10 of Appendix C on U_k with the function $z \mapsto f(z)\pi(z)$. In this context, for $x \in \partial U_k$, if $\langle n(x), v\rangle > 0$,

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

$$\pi_k(x^-)f(k, x^-, v) = \lim_{t\uparrow 0} \pi_k(x+tv)f(k, x+tv, v) = \pi_k(x)f(k, x, v),$$

$$\pi_k\big(x^+\big)f\big(k, x^+, v\big) = 0;$$

otherwise, if $\langle n(x), v\rangle > 0$,

$$\pi_k(x^-)f(k, x^-, v) = 0,$$

$$\pi_k(x^+)f(k, x^+, v) = \lim_{t\downarrow 0} \pi_k(x+tv)f(k, x+tv, v) = \pi_k(x)f(k, x, v).$$

This yields

$$\int_{U_k} \pi_k(x)\partial_v f(k, x, v)\,\mathrm{d}x = -\int_{U_k} f(k, x, v)\partial_v \pi_k(x)\,\mathrm{d}x\,\mathrm{d}v + \int_{\partial U_k \setminus C_v} \pi_k(x)f(k, x, v)\,\langle n(x), v\rangle\,\mathrm{d}\sigma(x),$$

where we have removed the absolute value around $\langle n(x), v\rangle$ to account for the sign difference of $\pi_k(x^-)f(k, x^-, v) - \pi_k(x^+)f(k, x^+, v)$. Furthermore, $\partial U_k \setminus C_v \subset \partial U_k \cap N_k$, and these two sets differ by a set of zero measure. Hence

$$\int_{U_k} \partial_v \pi_k(x)f(k, x, v)\,\mathrm{d}x = -\int_{U_k} f(k, x, v)\partial_v \pi_k(x)\,\mathrm{d}x\,\mathrm{d}v + \int_{\partial U_k \cap N_k} \pi_k(x)f(k, x, v)\,\langle n(x), v\rangle\,\mathrm{d}\sigma(x),$$

which concludes the proof. □

Lemma 6. *We have*

$$\int_{E_k} \partial_v f(k, x, v)\mu(k, x, v)\,\mathrm{d}x\,\mathrm{d}v = -\int_{U_k\times\mathcal{V}^k} f(k, x, v)\nabla_x \mu \cdot v\,\mathrm{d}x\,\mathrm{d}v + \int_{(\partial U_k\cap N_k)\times\mathcal{V}^k} f(k, x, v)\mu(k, x, v)\langle v, n(x)\rangle\,\mathrm{d}\sigma(x)\,\mathrm{d}v.$$

Proof. Since $\partial_v f$ is in $L_1(\mu)$, $\partial_v f(k, x, v)\mu(k, x, v)$ is integrable. Hence by Fubini's theorem,

$$\int_{U_k\times\mathcal{V}^k} \partial_v f(k, x, v)\mu(k, x, v)\,\mathrm{d}x\,\mathrm{d}v = \int_{\mathcal{V}^k}\int_{U_k} \partial_v f(k, x, v)\mu(k, x, v)\,\mathrm{d}x\,\mathrm{d}v.$$

Using Proposition 9,

$$\int_{U_k\times\mathcal{V}^k} \partial_v f(k, x, v)\mu(k, x, v)\,\mathrm{d}x\,\mathrm{d}v = -\int_{\mathcal{V}^k}\int_{U_k} f(k, x, v)(\nabla\pi_k(x)\cdot v)p_k(v)\,\mathrm{d}x\,\mathrm{d}v + \int_{\mathcal{V}^k}\int_{\partial U_k\cap N_k} f(k, x, v)\mu(k, x, v)\langle v, n(x)\rangle\,\mathrm{d}\sigma(x)\,\mathrm{d}v.$$

Using (iii) from Assumption 2, $f(k, x, v)(\nabla\pi_k(x)\cdot v)p_k(v)$ is integrable, and $f(k, x, v)\mu(k, x, v)$ is bounded. Hence we can use Fubini a second time on both terms to get the result. □

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

Appendix C. Theorem: integration by parts

Assumption 5. (Informal geometrical assumption.) *Let U be an open set in $\mathbb{R}^n$ such that the following hold:*

- $\bar{U} = \mathbb{R}^n$*;*
- *the boundary ∂U can be decomposed as a finite union of smooth closed sub-manifolds with piecewise C^1 boundaries in $\mathbb{R}^n$;*
- *for any $x, v \in \mathbb{R}^n$, the intersection $\partial U \cup \{x + \mathbb{R}v\}$ is finite (not taking into account the points where v is tangent to ∂U);*
- $\dim_H (N_v) \leq n - 2$*, where N_v is the subset of ∂U where the normal is ill-defined; and*
- $\dim_H p_v(M^v) \leq n - 2$*, where $M^v = \{x \in \partial U$ such that $\langle v, n(x)\rangle = 0\}$ and p_v is the orthogonal projection on $H_v = v^\perp$,*

with $\dim_H$ being the Hausdorff dimension.

These assumptions are made precise in Assumption 6 of the next section.

Proposition 10. *Let U be an open set of $\mathbb{R}^n$ satisfying Assumption 6, and let ∂U be its boundary. Let f and g be measurable functions from U to $\mathbb{R}$ such that the following hold:*

(i) f is bounded;

(ii) for any sequence $(y_n) \subset U$ with $\|y_n\| \to \infty$, $\lim_{n\to\infty} g(y_n) = 0$;

(iii) for each $x, v \in \mathbb{R}^n$, the functions $t \mapsto f(x + tv)$ and $t \mapsto g(x + tv)$ are absolutely continuous on $U \cap (x + \mathbb{R}v)$ and $\partial_t f$, $\partial_t g \in L^1(U)$.

Fix $v \in \mathbb{R}^n$. Then, using the convention

$$f(x^+) = \lim_{t \downarrow 0} f(x + tv) \text{ and } f(x^-) = \lim_{t \uparrow 0} f(x + tv)$$

and

$$g(x^+) = \lim_{t \downarrow 0} g(x + tv) \text{ and } g(x^-) = \lim_{t \uparrow 0} g(x + tv),$$

we have

$$\int_U \partial_v f(x)\, g(x)\, \mathrm{d}x = \int_{\partial U \setminus N_v} (g(x^-) f(x^-) - f(x^+) g(x^+))\, |\langle n(x), v\rangle|\, \mathrm{d}\sigma(x) - \int_U f(x)\, \partial_v g(x)\, \mathrm{d}x,$$

where the second term is integrated with respect to the Lebesgue measure of the boundary (seen as a Riemannian manifold) and N_v is the set of points where the normal is ill-defined.

Proof. The proof is a corollary of the next section. □

C.1. Integration over open domain

Assumption 6. *Let U be an open set in $\mathbb{R}^n$ such that for each $v \in S^{n-1}$ (the unit sphere in $\mathbb{R}^n$), the following hold:*

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

1. *There exist* $W_1 \subset \overline{W}_1 \subset \Omega_1, \dots, W_k \subset \overline{W}_k \subset \Omega_k$, $2k$ *open sets in* $\mathbb{R}^{n-1}$, *with* $\dim_H \partial W_i \leq n-2$ *(these open sets may depend on* v *because of Item 6 of this assumption).*
2. *There exist* $\phi_i : \Omega_i \to \mathbb{R}^n$, $i = 1, \dots k$, *which are* C^1 *one-to-one maps such that the differential* $D\phi_i(x)$ *is one-to-one for all* $x \in \Omega_i$. *This implies that there is a continuous normal* $n_i : \Omega_i \to S^{n-1}$.
3. *We have* $\partial U = M_1 \cup \dots \cup M_k$ *where the sets* $M_i = \phi_i(\overline{W}_i)$ *are closed.*
4. *We have* $\dim_H M_i \cap M_j \leq n-2$ *for all* $i \neq j$.
5. *Let* $W_i^0 = \{x \in W_i : v \cdot n_i(x) = 0\}$). *For each* i, $\dim_H p_v(M_i^0) \leq n-2$, *where* $M_i^0 = \phi_i(W_i^0)$ *and* p_v *is the orthogonal projection on* $H = H_v = v^\perp$.
6. *Let* $W_i^+ = \{x \in W_i \in M_i : n_i(x) \cdot v > 0\}$ *and* $W_i^- = \{x \in W_i : n_i(x) \cdot v < 0\}$. *For all* i *and all* $y \in \mathbb{R}^n$, *the sets* $M_i^+ \cap (y + \mathbb{R}v)$ *and* $M_i^- \cap (y + \mathbb{R}v)$ *have at most one element, where* $M_i^+ = \phi_i(W_i^+)$ *and* $M_i^- = \phi_i(W_i^-)$.

Assumption 7. *Let* $f : U \to \mathbb{R}$ *be a measurable function such that the following hold:*

(i) *For each* y, $v \in \mathbb{R}^n$, *the function* $t \mapsto f(y + tv)$ *is absolutely continuous on every bounded interval* I *such that* $y + Iv \subset U$.

(ii) *We have* $\lim_{\|y\| \to \infty} f(y) = 0$ *(that is, for any sequence* $(y_n) \subset U$ *with* $\|y_n\| \to \infty$*).*

(iii) *If* U *is not bounded, then for each* $v \in \mathbb{R}^n$, $\partial_v f \in L^1(U)$.

We extend f to $\mathbb{R}^n \setminus \partial U$ with $f(y) = 0$ for every $y \notin \overline{U}$, so that we can suppose that $U = \mathbb{R}^n \setminus \partial U$.

Theorem 4. *Let* U *be an open set of* $\mathbb{R}^n$ *satisfying Assumption 6, for some fixed* $v \in \mathbb{R}^n$ *with* $\|v\| = 1$. *Let*

$$\mathcal{N}_v = \left(\bigcup_{i=1}^{k} (\phi_i(\partial W_i) \cup M_i^0) \right) \bigcup \left(\bigcup_{1 \leq i < j \leq k} M_i \cap M_j) \right)$$

be the set on which normals are ill-defined. Then for any f *satisfying Assumption 7, the following hold:*

1. *We have* $\dim_H \mathcal{N}_v \leq 2$ *and* $\dim_H p_v(\mathcal{N}_v) \leq n-2$.
2. *For each* $y \in \partial U \setminus \mathcal{N}_v$, *the limits*

$$\lim_{t \downarrow 0} f(y + tv) = f(y^+) \text{ and } \lim_{t \uparrow 0} f(y + tv) = f(y^-)$$

exist.

3. *The normal* $n(y)$ *is well-defined at each point* $y \in \partial U \setminus \mathcal{N}_v$, *and*

$$\int_U \partial_v f(y)\, \mathrm{d}y = \int_{\partial U \setminus \mathcal{N}_v} (f(y^-) - f(y^+))\, |n(y) \cdot v|\, \mathrm{d}\sigma(y),$$

where σ *is the Lebesgue measure on* ∂U.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

We can use the theorem with a product $f = g\pi$ where

- $g : U \to \mathbb{R}$ is measurable, bounded, and absolutely continuous on each set $U \cap (y + \mathbb{R}v)$, $y \in \mathbb{R}^n$, and $\partial_t g \in L^1(U)$; and
- $\pi : U \to \mathbb{R}$ is in $C^1(U) \cap L^1(U)$ and is bounded with bounded derivatives, $\lim_{\|y\|\to\infty} \pi(y) = 0$, and the derivative $\partial_v \pi \in L^1(U)$.

C.2. Proof of Theorem 4

For the first point, let

$$\mathcal{N}_v = \left(\bigcup_{i=1}^{k} (\phi_i(\partial W_i) \cup M_i^0) \right) \bigcup \left(\bigcup_{1\le i<j\le k} M_i \cap M_j) \right),$$

$$\mathcal{N} = p_v^{-1}(p_v(\mathcal{N}_v)) \text{ and } V = H \setminus p_v(\mathcal{N}_v).$$

By Assumptions 6.1, 6.4, and 6.5, $\dim_H(p_v(\mathcal{N}_v)) \le n-2$; therefore $p_v(\mathcal{N}_v)$ has zero H-Lebesgue measure.

For the second point, the fact that $f(y+)$ and $f(y-)$ exist is a direct consequence of Assumption 7(i).

Finally we consider the third point. For all $z \in H$, define

$$E(z) = \{t \in \mathbb{R} : z + tv \in \partial U\}.$$

By Assumption 6.6, the set $E(z)$ has $2k$ elements at most for all $z \in V$. Set

$$U' = U \setminus \mathcal{N}.$$

Since $\dim_H (p_v(\mathcal{N}_v)) \le n-2$, it follows that $\dim_H \mathcal{N} \le n-1$. Therefore,

$$\int_U \partial_v f(y)\, \mathrm{d}y = \int_{U'} \partial_v f(y)\, \mathrm{d}y.$$

By Fubini's theorem,

$$\int_{U'} \partial_v f(y)\, \mathrm{d}y = \int_V \left(\int_{\mathbb{R}\setminus E(z)} \partial_t f(z+tv)\, \mathrm{d}t \right) \mathrm{d}z.$$

By Assumption 7, for almost all $z \in V$ and for each connected component (a, b) of $\mathbb{R} \setminus E(z)$,

$$\int_{(a,b)} \partial_v f(z+tv)\, \mathrm{d}t = f((z+tb)^-) - f((z+ta)^+).$$

Taking into account that $\lim_{t\to\pm\infty} f(z+tv) = 0$, we obtain

$$\int_U \partial_t f(y)\, \mathrm{d}y = \int_V \sum_{t\in E(z)} (f((z+tv)^-) - f((z+tv)^+))\, \mathrm{d}z.$$

Now we want to see that the latter integral is equal to

$$\int_{\partial U\setminus\mathcal{N}_v} (f(y^-) - f(y^+))\, |n(y)\cdot v|\, \mathrm{d}\sigma(y).$$

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

Set

$$I = \{1, \dots, k\} \times \{+, -\},$$
$$J(z) = \{(i, s) \in I : \exists t \in E(z),\, z + tv \in M_i^s\} \text{ for } z \in H,$$
$$V_J = \{z \in V : J(z) = J\} \text{ for } J \subset I, \text{ and}$$
$$V_i^s = p_v(M_i^s) \text{ for } (i, s) \in I.$$

By Assumption 6.6, for each $(i, s) \in I$, the map $p_v \circ \phi_i : W_i^s \to V_i^s$ is a bijection, so that we can define the map $F_i^s : V_i^s \to M_i^s$ by $F_i^s(z) = \phi_i((p_v \circ \phi_i)^{-1}(z))$. Using the definition of the set $V = H \setminus \mathcal{N}$, we see that for each $z \in V$ and each $t \in E(z)$, there exists $(i, s) \in I$ unique such that $z + tv = F_i^s(z) \in M_i^s$. Furthermore, for each $(i, s) \in I$, $V_i^s \cap V = \cup_{J \ni (i,s)} V_J$. It follows that

$$\begin{aligned}
\int_U \partial_t f(y)\, \mathrm{d}y &= \int_V \sum_{t \in E(z)} (f((z + tv)^-) - f((z + tv)^+))\, \mathrm{d}z \\
&= \sum_{J \subset I} \int_{V_J} \sum_{t \in E(z)} (f((z + tv)^-) - f((z + tv)^+))\, \mathrm{d}z \\
&= \sum_{J \subset I} \int_{V_J} \sum_{(i,s) \in J} (f(F_i^s(z)^-) - f(F_i^s(z)^+))\, \mathrm{d}z \\
&= \sum_{(i,s) \in I} \sum_{J \ni (i,s)} \int_{V_J} (f(F_i^s(z)^-) - f(F_i^s(z)^+))\, \mathrm{d}z \\
&= \sum_{(i,s) \in I} \int_{V_i^s \cap V} (f(F_i^s(z)^-) - f(F_i^s(z)^+))\, \mathrm{d}z.
\end{aligned}$$

Since the differential of each ϕ_i is always one-to-one and since $D\phi_i(x)(u).u$ is never orthogonal to v for $x \in W_i^s$ and $u \neq 0$, the local inverse function theorem implies that the maps F_i^s are C^1. Furthermore, the image of F_i^s is M_i^s and the normal to M_i^s at $y = F_i^s(z)$ is $n(y) = \pm n_i((p_v \circ \phi)^{-1}(z))$. Therefore,

$$\int_{V_i^s \cap V} (f(F_i^s(z)^-) - f(F_i^s(z)^+))\, \mathrm{d}z = \int_{M_i^s \setminus \mathcal{N}_v} (f(y^-) - f(y^+))|n(y)|\, \mathrm{d}\sigma(y).$$

Finally, since $\partial U \setminus \mathcal{N}_v = \cup_{(i,s) \in I}(M_i^s \setminus \mathcal{N}_v)$,

$$\sum_{(i,s) \in I} \int_{V_i^s \cap V} (f(F_i^s(z)^-) - f(F_i^s(z)^+))\, \mathrm{d}z = \int_{\partial U \setminus \mathcal{N}_v} (f(y^-) - f(y^+))|n(y)|\, \mathrm{d}\sigma(y).$$

Appendix D. Validity of transition kernels derived from limiting behaviour

D.1. Bouncy particle sampler: proof of Proposition 5

For the specified Q_{BPS}, we first derive the form of the associated probability kernel on velocities, Q_x', remembering that Q_x' is obtained by flipping the velocity and then applying the

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

transition defined by Q_{BPS}. This gives

$$Q'_x(v'|v) = \begin{cases} 1, & v' = -v \text{ and } v \in \mathcal{V}_x^-, \\ \pi_{k_1(x)}(x)/\pi_{k_2(x)}(x), & v' = -v \text{ and } v \in \mathcal{V}_x^+, \\ 1 - \pi_{k_1(x)}(x)/\pi_{k_2(x)}(x), & v' = -v + 2\langle n, v\rangle n \text{ and } v \in \mathcal{V}_x^+. \end{cases}$$

The transition kernel Q'_x allows for two possible transitions, either $v' = -v$ or $v' = v - 2\langle n, v\rangle n$. These transitions have the following properties:

(i) In the first case, if $v \in \mathcal{V}_x^+$ then $v' \in \mathcal{V}_x^-$, and vice versa. In the second case, if $v \in \mathcal{V}_x^+$ then $v' \in \mathcal{V}_x^+$.

(ii) For either transition, $\langle v', v'\rangle = \langle v, v\rangle$, and $|\langle n, v'\rangle| = |\langle n, v\rangle|$. Furthermore, by the spherical symmetry of $p(v)$ for the bouncy particle sampler, the first of these means that $p(v) = p(v')$.

We need to show that $l_x(dv') = \int Q'_x(dv'|v) l_x(v)\, \mathrm{d}v$, where

$$l_x(v) = \begin{cases} |\langle n, v\rangle| p(v) \pi_{k_2(x)}(x) & \forall v \in \mathcal{V}_x^+, \\ |\langle n, v\rangle| p(v) \pi_{k_1(x)}(x) & \forall v \in \mathcal{V}_x^-. \end{cases}$$

We will show this holds first for $v' \in \mathcal{V}_x^+$ and then for $v' \in \mathcal{V}_x^-$.

If $v' \in \mathcal{V}_x^+$ then there are two possible transitions, from $v = -v' \in \mathcal{V}_x^-$ and from $v \in \mathcal{V}_x^+$ where $v = -v' + 2\langle v', n\rangle n$. Let $v^* = -v' + 2\langle v', n\rangle n$. Since the Jacobian of both transformations is 1,

$$\begin{aligned} \int Q'_x(dv'|v) l_x(v)\, \mathrm{d}v &= 1 \cdot \left(|\langle n, -v'\rangle| p(-v') \pi_{k_1(x)}(x)\right) \mathrm{d}v' \\ &\quad + \left(1 - \frac{\pi_{k_1(x)}(x)}{\pi_{k_2(x)}(x)}\right) \left(|\langle n, v^*\rangle| p(v^*) \pi_{k_2(x)}(x)\right) \mathrm{d}v', \\ &= |\langle n, v'\rangle| p(v') \left(\pi_{k_1(x)}(x) + \left(1 - \frac{\pi_{k_1(x)}(x)}{\pi_{k_2(x)}(x)}\right) \pi_{k_2(x)}(x)\right) \mathrm{d}v', \end{aligned}$$

where the last equality comes from applying Property (ii) of the transition. The last expression simplifies to $l_x(\mathrm{d}v')$ as required.

For $v' \in \mathcal{V}_x^-$ we have only one transition and thus

$$\int Q'_x(dv'|v) l_x(v) dv = \left(\frac{\pi_{k_1(x)}(x)}{\pi_{k_2(x)}(x)}\right) \left(|\langle n, -v'\rangle| p(-v') \pi_{k_2(x)}(x)\right) \mathrm{d}v',$$

which, again using Property (ii), is $l_x(v')$.

D.2. Coordinate sampler: proof of Proposition 6

We follow a similar argument to that of the previous section. First we write down the form of Q'_x derived from Q_{CS}:

$$Q'_x(v'|v) = \begin{cases} 1, & v' = -v \text{ and } v \in \mathcal{V}_x^-, \\ \pi_{k_1(x)}(x)/\pi_{k_2(x)}(x), & v' = -v \text{ and } v \in \mathcal{V}_x^+, \\ (1 - \pi_{k_1(x)}(x)/\pi_{k_2(x)}(x)) \frac{\langle v', n\rangle}{K}, & v' \in \mathcal{V}_x^+ \text{ and } v \in \mathcal{V}_x^+, \end{cases}$$

where $K = \sum_{v \in \mathcal{V}_x^+} |\langle n, v\rangle|$.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

For the coordinate sampler, $p(v) = 1/(2d)$ for each of the possible values for v. Now we need to show $l_x(v') = \sum Q'_x(v'|v) l_x(v)$. We will consider the cases $v' \in \mathcal{V}_x^+$ and $v' \in \mathcal{V}_x^-$ separately. For the latter case, the argument is the same as for the bouncy particle sampler. Thus we just present the case for $v' \in \mathcal{V}_x^+$.

We have

$$\begin{aligned}
\sum Q'_x(v'|v) l_x(v) &= 1 \cdot \left(|\langle n, -v' \rangle| p(-v') \pi_{k_1(x)}(x) \right) \\
&\quad + \langle v', n \rangle \left(1 - \frac{\pi_{k_1(x)}(x)}{\pi_{k_2(x)}(x)} \right) \sum_{v \in \mathcal{V}_x^+} \frac{|\langle n, v \rangle|}{K} p(v) \pi_{k_2(x)}(x) \\
&= p(v') \langle v', n \rangle \left(\pi_{k_1(x)}(x) + \left(1 - \frac{\pi_{k_1(x)}(x)}{\pi_{k_2(x)}(x)} \right) \pi_{k_2(x)}(x) \sum_{v \in \mathcal{V}_x^+} \frac{|\langle n, v \rangle|}{K} \right) \\
&= p(v') \langle v', n \rangle \left(\pi_{k_1(x)}(x) + \left(1 - \frac{\pi_{k_1(x)}(x)}{\pi_{k_2(x)}(x)} \right) \pi_{k_2(x)}(x) \right).
\end{aligned}$$

The second equality comes from the fact that $p(v)$ is constant for all $v \in \mathcal{V}$. The third equality comes from the definition of K.

D.3. Zigzag sampler: proof of Proposition 7

As discussed, the kernel for the velocity does not depend on k. Without loss of generality, we can set $k = 1$ for implementing the algorithm that defines Q_{ZZ}. We need to show that Q'_x keeps l_x invariant. We will prove this by showing that the following stronger detailed balance condition holds:

$$l_x(v) Q'_x(v'|v) = l_x(v') Q'_x(v|v'), \quad \forall v, v' \in \mathcal{V}.$$

As $Q'_x(v'|v) = Q_{ZZ}(v'| - v)$, writing the detailed balance condition for pairs $-v$ and v', we have that it suffices to show

$$l_x(-v) Q_{ZZ}(v'|v) = l_x(v') Q_{ZZ}(-v| - v'), \quad \forall v, v' \in \mathcal{V}.$$

By a slight abuse of notation, let $k(v) = k_1(x)$ if $v \in \mathcal{V}_x^-$ and $k(v) = k_2(x)$ if $v \in \mathcal{V}_x^+$. Then we can write $l_x(v) = |\langle n, v \rangle| p(v) \pi_{k(v)}(x)$. Thus, using the fact that $p(v)$ defines a uniform distribution on $\mathcal{V}$, we have that the detailed balance condition simplifies to

$$|\langle n, v \rangle| \pi_{k(-v)}(x) Q_{ZZ}(v'|v) = |\langle n, v' \rangle| \pi_{k(v')}(x) Q_{ZZ}(-v| - v'), \quad \forall v, v' \in \mathcal{V}.$$

This can be viewed as matching the probability that we have a velocity v and transition to v' with one where we flip the velocities, starting at $-v'$ and transitioning to $-v$.

We show that the detailed balance condition holds separately for different combinations of whether $v \in \mathcal{V}_x^+$ or $v \in \mathcal{V}_x^-$ and whether $v' \in \mathcal{V}_x^+$ or $v' \in \mathcal{V}_x^-$.

First assume $v \in \mathcal{V}_x^+$ and $v' \in \mathcal{V}_x^-$. This corresponds to a trajectory that is moving from the lower- to the higher-density region, but that reflects off the boundary and stays in the lower-density region. It is straightforward to see that the events that change the velocity only increase $\langle n, v \rangle$, the speed at which the trajectory moves through the boundary region to the higher-density region. Thus a transition from $\mathcal{V}_x^+$ to $\mathcal{V}_x^-$ is impossible, and $Q_{ZZ}(v'|v) = 0$. Similarly,

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

$-v' \in \mathcal{V}_x^+$ and $-v \in \mathcal{V}_x^-$ so $Q_{\mathrm{ZZ}}(-v|-v') = 0$. Hence the detailed balance conditions trivially hold in this case.

Next assume $v \in \mathcal{V}_x^-$ and $v' \in \mathcal{V}_x^+$. This corresponds to a trajectory that is moving from the higher- to the lower-density region, but that reflects off the boundary and stays in the higher-density region. In this case $k(-v) = k(v')$ and thus the detailed balance condition becomes

$$|\langle n, v\rangle| Q_{\mathrm{ZZ}}(v'|v) = |\langle n, v'\rangle| Q_{\mathrm{ZZ}}(-v|-v'), \quad \forall v \in \mathcal{V}_x^-,\ v' \in \mathcal{V}_x^+. \tag{11}$$

To prove that the detailed balance condition holds, we will first obtain an expression for $Q_{\mathrm{ZZ}}(v'|v)$, and then introduce a coupling between a transition for v to v' and one from $-v'$ to $-v$ to link it to a similar expression for $Q_{\mathrm{ZZ}}(-v|-v')$.

The randomness in the algorithm that defines Q_{ZZ} only comes through the randomness of the event times simulated in Step (a) of Section 6.2.3. Remember that τ_i is the time at which component i of the velocity will switch, if the trajectory is still within the boundary region. Each τ_i is (conditionally) independent of the others, and has an exponential distribution with rate $\max\{0, -n_i v_i\}$, where n_i is the component of the ith coordinate of the unit normal n. If $n_i v_i \geq 0$, then $\tau_i = \infty$.

It is helpful to introduce three sets of components:

- Let $\mathcal{S}_1$ be the set of components i such that $v'_i = v_i$ and $n_i v_i < 0$.
- Let $\mathcal{S}_2$ be the set of components i such that $v'_i = -v_i$.
- Let $\mathcal{S}_3$ be the set of components i such that $v'_i = v_i$ and $n_i v_i \geq 0$.

So $\mathcal{S}_1$ is the set of components of the velocity v that are moving the particle towards the low-density region, and are unchanged by the transition to v'; $\mathcal{S}_2$ is the set of components that flip during the transition from v to v'; and $\mathcal{S}_3$ is the set of components of the velocity v that are moving the particle towards the high-density region, and are unchanged by the transition to v'.

Only components i of the velocity for which $n_i v_i < 0$ can change during the transition from v to v'. This means that if there exists $i \in \mathcal{S}_2$ such that $n_i v_i \geq 0$, then the transition from v to v' is impossible. By the same argument, the transition from $-v'$ to $-v$ is impossible. Thus in this case $Q_{\mathrm{ZZ}}(v'|v) = Q_{\mathrm{ZZ}}(-v|-v') = 0$ and detailed balance trivially holds. So in the following we will assume that $n_i v_i < 0$ for $i \in \mathcal{S}_2$.

By a similar argument we have that the set $\mathcal{S}_1$ is the set of indices of the velocity that could have changed during the transition from v to v', but did not, whereas $\mathcal{S}_3$ is the set of indices of the velocity that could never have changed during the transition.

To ease notation, let $m = |\mathcal{S}_2|$, the number of indices in the set $\mathcal{S}_2$, and note that $m \geq 1$ as $v \neq v'$. Without loss of generality we can relabel the coordinates so that $\mathcal{S}_2 = \{1, \ldots, m\}$, and we will use $\tau_{1:m}$ to denote the vector of event times for the coordinates in $\mathcal{S}_2$.

We now introduce a function of time, t, that depends on $\tau_{1:n}$. This is

$$h(t; \tau_{1:n}) = \sum_{i=1}^{m} n_i v_i (t - 2\max\{0, t - \tau_i\}) + \sum_{i \in \mathcal{S}_1} n_i v_i t + \sum_{i \in \mathcal{S}_3} n_i v_i t.$$

This can be viewed as the net distance travelled by the trajectory up to time t in the direction of the normal n, given that only velocity coordinates in $\mathcal{S}_2$ can change, and these change at times $\tau_{1:m}$. This function is important as it determines when the trajectory leaves the boundary region, and determines the termination of the simulation algorithm in Step (b). As $v \in \mathcal{V}_x^-$ and $v' \in \mathcal{V}_x^+$, and the changes in velocity in the direction of n are monotone as we flip components,

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

we have that $h(t;\tau_{1:n})$ is strictly decreasing at $t=0$, is strictly increasing for large enough t, and is unimodal. As $h(0;\tau_{1:m})=0$, this means that there is a unique $t^*(\tau_{1:m})>0$ such that $h(t^*(\tau_{1:m});\tau_{1:m})=0$. This is the exit time from the boundary region calculated in Step (b) of the algorithm.

We can now define the set $\mathcal{T}$ of values of $\tau_{1:m}$ that are consistent with a transition from v to v'. The conditions are that all components of the velocity must flip before t^*, and that the trajectory must not pass through the boundary region—see the other stopping criteria in Step (b) of the algorithm. This gives us that

$$\mathcal{T}=\left\{\tau_{1:m}:\tau_i\le t^*(\tau_{1:m}),\ i=1,\dots,m;\ \min_{0<t<t^*(\tau_{1:m})}h(t;\tau_{1:n})>-C\right\}.$$

The probability of a transition from v to v' is thus the probability that $\tau_{1:m}\in\mathcal{T}$ times the probability that $\tau_i>t^*(\tau_{1:m})$ for $i\in\mathcal{S}_1$. As each τ_i, $i\in\mathcal{S}_1$ or $i\in\mathcal{S}_2$, has an independent exponential distribution with rate $-n_iv_i$,

$$Q_{\mathrm{ZZ}}(v'|v)=\int_{\mathcal{T}}\left(\prod_{i\in\mathcal{S}_1}\exp\{n_iv_it^*(\tau_{1:m})\}\right)\left(\prod_{i=1}^m(-n_iv_i)\exp\{n_iv_i\tau_i\}\,\mathrm{d}\tau_{1:m}\right).$$

Now consider the reverse transition, from $-v'$ to $-v$. Under our existing definitions of $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$, we have that $\mathcal{S}_2$ is still the set of indices that the flip for the transition from $-v'$ to $-v$, but now $\mathcal{S}_1$ is the set of components of the velocity that could never have flipped, while $\mathcal{S}_3$ is the set of components that could have flipped but did not.

We can define the same quantities for the reverse transition from $-v'$ to $-v$. We will use tildes to denote quantities that relate to this transition. So $\tilde\tau_{1:m}$ will be the vector of flip times for components in $i\in\mathcal{S}_2$. We have

$$\tilde h(t;\tilde\tau_{1:n})=\sum_{i=1}^m n_iv_i(t-2\max\{0,t-\tilde\tau_i\})-\sum_{i\in\mathcal{S}_1}n_iv_it-\sum_{i\in\mathcal{S}_3}n_iv_it,$$

using the fact that $-v'_i=v_i$ for $i\in\mathcal{S}_2$ and $-v'_i=-v_i$ otherwise. By the same argument as above, there is a unique $\tilde t^*(\tilde\tau_{1:m})>0$ such that $\tilde h(\tilde t^*(\tilde\tau_{1:m});\tilde\tau_{1:m})=0$. The set of possible values of $\tilde\tau_{1:m}$ that are consistent with the transition from $-v'$ to v is

$$\tilde{\mathcal{T}}=\left\{\tilde\tau_{1:m}:\tilde\tau_i\le\tilde t^*(\tilde\tau_{1:m}),\ i=1,\dots,m;\ \min_{0<t<\tilde t^*(\tilde\tau_{1:m})}\tilde h(t;\tilde\tau_{1:n})>-C\right\}.$$

Finally we can write down the transition probability as before, remembering that the rate of flipping for components $i\in\mathcal{S}_2$ is $-n_iv_i$ as before, but for $i\in\mathcal{S}_3$ it is n_iv_i. Thus

$$Q_{\mathrm{ZZ}}(-v|-v')=\left(\prod_{i\in\mathcal{S}_3}\exp\{-n_iv_i\tilde t^*(\tilde\tau_{1:m})\}\right)\left(\int_{\tilde{\mathcal{T}}}\prod_{i=1}^m(-n_iv_i)\exp\{n_iv_i\tilde\tau_i\}\,\mathrm{d}\tilde\tau_{1:m}\right).\tag{12}$$

To relate the two transition probabilities, we introduce a coupling between $\tau_{1:m}$ and $\tilde\tau_{1:m}$, so $\tilde\tau_{1:m}=g(\tau_{1:m})$, where

$$\tilde\tau_i=g(\tau_{1:m})_i=t^*(\tau_{1:m})-\tau_i.$$

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

This coupling is a natural one. If we consider the path through the boundary region given by $\tau_{1:m}$ that transitions from v to v', we can reverse that path to get a path that transitions from $-v'$ to $-v$. For the forward path a flip of component i at time τ_i occurs at a time $t^*(\tau_{1:m}) - \tau_i$ prior to the end of the path. Thus for the reverse path the flip would occur at time $t^*(\tau_{1:m}) - \tau_i$.

It is straightforward to show that if $\tilde{\tau}_{1:m} = g(\tau_{1:m})$ then $h(t; \tau_{1:m}) = \tilde{h}(t^*(\tau_{1:m}) - t; \tilde{\tau}_{1:m})$. This result is intuitive; it is saying the distance of the forward trajectory within the boundary region at time t is equal to the distance of the backward trajectory within the boundary region at time $t^*(\tau_{1:m}) - t$. This immediately implies that $t^*(\tau_{1:m}) = \tilde{t}^*(\tilde{\tau}_{1:m})$: the exit times for the forward and backward trajectories are the same. Furthermore, if we consider the second constraint on $\tau_{1:m}$ in the definition of $\mathcal{T}$, then we have

$$\min_{0<t<t^*(\tau_{1:m})} h(t; \tau_{1:n}) = \min_{0<t<\tilde{t}^*(\tilde{\tau}_{1:m})} \tilde{h}(t; \tilde{\tau}_{1:n}),$$

for $\tilde{\tau}_{1:m} = g(\tau_{1:m})$. Combining this with the fact that $\tau_i \leq t^*(\tau_{1:m})$, we have $\tilde{\tau}_{1:m} \leq \tilde{t}^*(\tilde{\tau}_{1:m})$. We have that the function g maps $\tau_{1:m} \in \mathcal{T}$ to $\tilde{\tau}_{1:m} \in \tilde{\mathcal{T}}$. Furthermore, the function g is invertible, and by similar arguments we have that g^{-1} maps $\tilde{\tau}_{1:m} \in \tilde{\mathcal{T}}$ to $\tau_{1:m} \in \mathcal{T}$. Hence g is a bijection from $\mathcal{T}$ to $\tilde{\mathcal{T}}$.

The function g defines a linear map between $\tau_{1:m}$ and $\tilde{\tau}_{1:m}$. For $\tau_{1:m} \in \mathcal{T}$ we have that, by definition of $t^*(\tau_{1:m})$,

$$\sum_{i=1}^{m} n_i v_i (2\tau_i - t^*(\tau_{1:m})) + \sum_{i=m+1}^{d} n_i v_i t^*(\tau_{1:m}) = 0. \tag{13}$$

This gives that

$$t^*(\tau_{1:m}) = \sum_{i=1}^{m} \left(\frac{-2n_i v_i}{K} \right) \tau_i, \text{ where } K = -\sum_{i=1}^{m} n_i v_i + \sum_{i=m+1}^{d} n_i v_i.$$

Furthermore, using that v' is equal to v except that v_i is flipped for $i = 1, \ldots, m$, we have $K = \langle v', n \rangle$.

Let $b_{1:m}$ be the $1 \times m$ vector whose ith entry is $b_i = 2v_i n_i / K$. If we let 1_m denote the $1 \times m$ vector of ones, and I_m the $m \times m$ identity matrix, then we have

$$\tilde{\tau}_{1:m} = g(\tau_{1:m}) = (b_{1:m} 1_{1:m}^\top - I_m)\tau_{1:m} = A\tau_{1:m},$$

where the $m \times m$ matrix $A = (b_{1:m} 1_{1:m}^\top - I_m)$. In the following argument we will make the change of variables $\tilde{\tau}_{1:m} = g(\tau_{1:m}) = A\tau_{1:m}$, and we will need the determinant of the Jacobian of this transformation. Using the matrix determinant lemma, this is given by

$$|\det(A)| = |(1 - 1_{1:m}^\top b_{1:m})||\det(-I_m)| = \left| 1 - \sum_{i=1}^{m} b_i \right|.$$

This simplifies to

$$\left| 1 - \sum_{i=1}^{m} b_i \right| = \left| 1 + \sum_{i=1}^{m} \frac{2v_i n_i}{\langle v', n \rangle} \right| = \left| \frac{\langle v', n \rangle + \sum_{i=1}^{m} 2v_i n_i}{\langle v', n \rangle} \right| = \left| \frac{\langle v, n \rangle}{\langle v', n \rangle} \right|.$$

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

So now, taking the definition of $Q_{ZZ}(v'|v)$ and applying the change of variables $\tilde{\tau}_{1:m} = g(\tau_{1:m})$, we get

$$\begin{aligned}
Q_{ZZ}(v'|v) &= \int_{\mathcal{T}} \prod_{i\in\mathcal{S}_1} \exp\{n_i v_i t^*(\tau_{1:m})\} \prod_{i=1}^{m} ((-n_i v_i)\exp\{n_i v_i \tau_i\})\ \mathrm{d}\tau_{1:m} \\
&= \int_{\tilde{\mathcal{T}}} \exp\left\{\sum_{i\in\mathcal{S}_1} n_i v_i \tilde{t}^*(\tilde{\tau}_{1:m})\right\} \left|\frac{\langle v, n\rangle}{\langle v', n\rangle}\right| \prod_{i=1}^{m} \left((-n_i v_i)\exp\{n_i v_i(\tilde{t}^*(\tilde{\tau}_{1:m}) - \tilde{\tau}_i)\}\right)\ \mathrm{d}\tilde{\tau}_{1:m} \\
&= \left|\frac{\langle v, n\rangle}{\langle v', n\rangle}\right| \int_{\tilde{\mathcal{T}}} \left(\prod_{i=1}^{m}(-n_i v_i)\right) \exp\left\{\sum_{i\in\mathcal{S}_1} n_i v_i \tilde{t}^*(\tilde{\tau}_{1:m}) + \sum_{i=1}^{m} n_i v_i(\tilde{t}^*(\tilde{\tau}_{1:m}) - \tilde{\tau}_i)\right\}\ \mathrm{d}\tilde{\tau}_{1:m} \\
&= \left|\frac{\langle v, n\rangle}{\langle v', n\rangle}\right| \int_{\tilde{\mathcal{T}}} \left(\prod_{i=1}^{m}(-n_i v_i)\right) \exp\left\{\sum_{i\in\mathcal{S}_3} -n_i v_i \tilde{t}^*(\tilde{\tau}_{1:m}) + \sum_{i=1}^{m} n_i v_i \tilde{\tau}_i\right\}\ \mathrm{d}\tilde{\tau}_{1:m},
\end{aligned}$$

where the final equality comes from the definition of $\tilde{t}^*(\tilde{\tau}_{1:m}) = t^*(\tau_{1:m})$, using (13) after substituting in $\tau_i = \tilde{t}^*(\tilde{\tau}_{1:m}) - \tilde{\tau}_i$.

By comparing the final expression with (12), we get that

$$Q_{ZZ}(v'|v) = \left|\frac{\langle v, n\rangle}{\langle v', n\rangle}\right| Q_{ZZ}(-v|v'),$$

which satisfies (11) as required.

The final combination involves $v, v' \in \mathcal{V}_x^+$ and $-v', -v \in \mathcal{V}_x^-$, or vice versa. The detailed balance condition in this case becomes

$$\pi_{k_1(x)}(x)|\langle n, v\rangle|Q_{ZZ}(v'|v) = \pi_{k_2(x)}(x)|\langle n, v'\rangle|Q_{ZZ}(-v|-v'), \quad \forall v, v' \in \mathcal{V}_x^-.$$

We can show this using an argument similar to the one above, with the same coupling of paths from v to v' with paths from v' to v. The main differences are as follows. First, the definition of $\mathcal{T}$ is simplified to

$$\left\{\tau_{1:m} : \tau_i \le t^*(\tau_{1:m}),\ i = 1, \ldots, m\right\},$$

as, by monotonicity of the changes in velocity, we do not need to check whether the other exit condition in Step (b) holds. Second, the definition of $t^*(\tau_{1:m})$ changes, with it being the value of t for which $h(t; \tau_{1:m}) = C$. For $\tau_{1:m} \in \mathcal{T}$, this becomes

$$\sum_{i=1}^{m} n_i v_i (2\tau_i - t^*(\tau_{1:m})) + \sum_{i=m+1}^{d} n_i v_i t^*(\tau_{1:m}) = C$$

because of the different exit condition in Step (b). We have similar changes to the definitions of $\tilde{\mathcal{T}}$ and $\tilde{t}^*(\tilde{\tau}_{1:m})$.

However, we can define $Q_{ZZ}(v'|v)$ and $Q_{ZZ}(-v|-v')$ in a similar way. Furthermore we can use the same linear transformation g, which is still a bijection between $\mathcal{T}$ and $\tilde{\mathcal{T}}$. Whilst the definition t^* has changed, this only introduces an additive constant into the linear transformation defined by g, and thus the Jacobian of the transformation is unchanged. Following the

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

argument above, we thus get to the same expression for $Q_{\mathrm{ZZ}}(v'|v)$ after making the change of variables:

$$Q_{\mathrm{ZZ}}(v'|v) = \left|\frac{\langle v, n\rangle}{\langle v', n\rangle}\right| \int_{\tilde{\mathcal{T}}} \left(\prod_{i=1}^{m} (-n_i v_i)\right)$$

$$\times \exp\left\{\sum_{i\in\mathcal{S}_1} n_i v_i \tilde{t}^*(\tilde{\tau}_{1:m}) + \sum_{i=1}^{m} n_i v_i (\tilde{t}^*(\tilde{\tau}_{1:m}) - \tilde{\tau}_i)\right\} \mathrm{d}\tilde{\tau}_{1:m}.$$

Now substituting in our new definition of $\tilde{t}^*(\tilde{\tau}_{1:m}) = t^*(\tau_{1:m})$ we get

$$Q_{\mathrm{ZZ}}(v'|v) = \left|\frac{\langle v, n\rangle}{\langle v', n\rangle}\right| Q_{\mathrm{ZZ}}(-v|v') \exp\{C\},$$

where the additional factor of $\exp\{C\}$ is due to the different definition of t^*. As $C = \log(\pi_{k_2(x)}(x)/\pi_{k_1(x)}(x))$ this rearranges to

$$\pi_{k_1(x)}(x)|\langle v', n\rangle| Q_{\mathrm{ZZ}}(v'|v) = \pi_{k_2(x)}(x)|\langle v, n\rangle| Q_{\mathrm{ZZ}}(-v|v'),$$

as required.

D.4. Bouncy particle sampler with reflect and refract: proof of Proposition 8

We first write down the form of $Q_x{}'$ derived from Q_{RR}. To do this it is helpful to define functions that define the new velocity at a reflection or refraction. Remember that ΔU is the change in log density as we move from the region with lower density, k_1, to the region with higher density, k_2, and is strictly positive. We will define the reflection of a velocity v by $L(v) = v - 2\langle v, n\rangle n$. For refractions we need to distinguish between transitions from k_1 to k_2 and vice versa. The refraction for a transition from k_1 to k_2 is $R_1(v) = \sqrt{\langle v, n\rangle^2 + 2\Delta U} n + v - \langle v, n\rangle n$, and for k_2 to k_1 it is $R_2(v) = \sqrt{\langle v, n\rangle^2 - 2\Delta U} n + v - \langle v, n\rangle n$, with R_2 being defined only for v such that $\langle v, n\rangle^2 > 2\Delta U$.

Using these functions, we can write the transition kernel $Q_x{}'$ as

$$Q_x{}'(v'|v) = \begin{cases} 1 & \text{for } v \in \mathcal{V}_x^- \text{ and } v' = R_1(-v), \\ 1 & \text{for } v \in \mathcal{V}_x^+,\ \langle v, n\rangle^2 > 2\Delta U, \text{ and } v' = R_2(-v), \\ 1 & \text{for } v \in \mathcal{V}_x^+,\ \langle v, n\rangle^2 \le 2\Delta U, \text{ and } v' = L(-v), \\ 0 & \text{otherwise.} \end{cases}$$

We need to show that $l_x(dv') = \int Q_x'(dv'|v) l_x(v) dv$ with

$$l_x(v) = \begin{cases} |\langle n, v\rangle| p(v) \pi_{k_2(x)}(x) & \forall v \in \mathcal{V}_x^+, \\ |\langle n, v\rangle| p(v) \pi_{k_1(x)}(x) & \forall v \in \mathcal{V}_x^-. \end{cases}$$

Since, for each v', there is only a finite number of v such that $Q_x'(v'|v) > 0$, the former is equivalent to $l_x(v') = \sum Q_x'(v'|v) l_x(v) J_{v,v'}$, where $J_{v,v'}$ is the Jacobian of the deterministic transformation from v to v'.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

If $v' \in \mathcal{V}_x^+$ and $\langle v', n\rangle^2 \leq 2\Delta U$, then the only v for which $Q'_x(v'|v)$ is non-zero is $v = L(-v')$. The result trivially holds as this transition has unit Jacobian, $\|v\| = \|v'\|$, so $p(v) = p(v')$; $|\langle n, v\rangle| = |\langle n, v'\rangle|$; $v' \in \mathcal{V}_x^+$ and $Q'_x(v'|v) = 1$.

If $v' \in \mathcal{V}_x^+$ and $\langle v, n\rangle^2 > 2\Delta U$, then the only v for which $Q'_x(v'|v)$ is non-zero is for v such that $v' = R_1(v)$, which is equivalent to $v = R_2(-v') \in \mathcal{V}_x^-$. So as $p(v) = C_1 \exp\{-\frac{1}{2}\|v\|^2\}$ for some constant C_1,

$$\begin{aligned}\sum Q'_x(v'|v) l_x(v) &= l_x(R_2(-v')) \\ &= |\langle n, R_2(-v')\rangle| \cdot p(R_2(-v')) \cdot \pi_{k_1}(x) \\ &= |\langle n, R_2(-v')\rangle| \left(C_1 \exp\left\{-\frac{1}{2}\|R_2(-v')\|^2\right\}\right) \pi_{k_1}(x). \end{aligned} \tag{14}$$

Now

$$\begin{aligned}\|R_2(-v')\|^2 &= \left\|\sqrt{\langle v', n\rangle^2 - 2\Delta U} n - v' + \langle v', n\rangle n\right\|^2 = \langle v', n\rangle^2 - 2\Delta U + \|v' - \langle v', n\rangle n\|^2 \\ &= \|v'\|^2 - 2\Delta U,\end{aligned}$$

where the third equality comes from n being orthogonal to $v' - \langle v', n\rangle n$. This gives

$$\left(C_1 \exp\left\{-\frac{1}{2}(\|v'\|^2 - 2\Delta U)\right\}\right) \pi_{k_1}(x) = p(v') \exp\{\Delta U\} \pi_{k_1}(x).$$

The definition of ΔU gives that $\exp\{\Delta U\}\pi_{k_1}(x) = \pi_{k_2}(x)$.

Now $v' = \sqrt{\langle v, n\rangle^2 + 2\Delta U} n - v + \langle v, n\rangle n$. Up to a sign change, this only changes the velocity in the direction of n. Thus the Jacobian of the transformation is given by

$$J_{v,v'}^{-1} = \left|\frac{d\langle v', n\rangle}{d\langle v, n\rangle}\right| = \left|\frac{d}{d\langle v, n\rangle}\sqrt{\langle v, n\rangle^2 + 2\Delta U}\right| = \left|\frac{\langle v, n\rangle}{\sqrt{\langle v, n\rangle^2 + 2\Delta U}}\right| = \frac{|\langle v, n\rangle|}{|\langle v', n\rangle|}.$$

Thus, substituting in these expressions, we have

$$\begin{aligned}\sum Q'_x(v'|v) l_x(v) J_{v,v'} &= |\langle n, R_2(-v')\rangle| p(v') \pi_{k_2}(x) \frac{|\langle v', n\rangle|}{|\langle v, n\rangle|} \\ &= |\langle n, v\rangle| p(v') \pi_{k_2}(x) \frac{|\langle v', n\rangle|}{|\langle v, n\rangle|} \\ &= |\langle n, v'\rangle| p(v') \pi_{k_2}(x) \\ &= l_x(v'),\end{aligned}$$

as required.

An equivalent calculation holds for $v' \in \mathcal{V}_x^-$, which is omitted for brevity.

Appendix E. Comments on reference [4]

Equation (4) of [4] should read $Q_b(x, u, \mathrm{d}v)\rho(\mathrm{d}u) = Q_b(x, -v, -\mathrm{d}u)\rho(-\mathrm{d}v)$, $x \in \partial O$, for u in the exit boundary and v in the entrance boundary. The sketch of the proof given in the

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

appendix of [4] also requires a correction. The condition (S2) on the domain of the generator has to be modified: it only holds for v in the exit boundary, and not for every v. This mistake, combined with the sign error in Equation (4), is what allows [4] to reach the conclusion that $\int \mathcal{L}f \, \mathrm{d}\pi \, \mathrm{d}\rho = 0$, missing the fact that the boundary term for the exit velocities cancels out the one for the entrance velocities. A full proof is provided in this manuscript, with proper care for the entrance/exit velocities.

Appendix F. Additional simulation results

For $\alpha_{\text{out}} = 0$, Figures 4–6 show trajectories for the bouncy particle sampler, the coordinate sampler, and the zigzag process for dimensions $d = 2, 10, 100$ for the sampling from a Gaussian restricted to a cube. Figure 7 shows trajectories for the zigzag process if we use the canonical basis—in this case the distributions of all coordinates are independent, and the zigzag process benefits from this by being able to run independent dynamics for each coordinates.

For $\alpha_{\text{out}} > 0$, Figures 8 and 9 show trajectories and Monte Carlo estimates along trajectories in dimension 20, in the case where the canonical basis is not rotated. In this case, the zigzag sampler clearly performs the best. Figure 10 shows trajectories when the basis is rotated.

Acknowledgements

This work was supported by EPSRC grants EP/R018561/1 and EP/R034710/1. A. C. did most of this work while at the University of Lancaster, and A. Q. W. did most of this work while at the University of Bristol.

Funding information

There are no funding bodies to thank in relation to the creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] Afshar, H. M. and Domke, J. (2015). Reflection, refraction, and Hamiltonian Monte Carlo. In *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, Neural Information Processing Systems Foundation, San Diego, CA, pp. 3007–3015.

[2] Bernard, E. P., Krauth, W. and Wilson, D. B. (2009). Event-chain Monte Carlo algorithms for hard-sphere systems. *Phys. Rev. E* **80**, article no. 056704.

[3] Bierkens, J. (2016). Non-reversible Metropolis–Hastings. *Statist. Comput.* **26**, 1213–1228.

[4] Bierkens, J. *et al.* (2018). Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains. *Statist. Prob. Lett.* **136**, 148–154.

[5] Bierkens, J., Fearnhead, P. and Roberts, G. (2019). The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* **47**, 1288–1320.

[6] Bierkens, J., Grazzi, S., Kamatani, K. and Roberts, G. (2020). The boomerang sampler. In *Proc. 37th International Conference on Machine Learning* (Proceedings of Machine Learning Research **119**), Machine Learning Research Press, pp. 908–918.

[7] Bierkens, J., Grazzi, S., van der Meulen, F. and Schauer, M. (2021). *Sticky PDMP samplers for sparse and local inference problems*. Preprint. Available at https://arxiv.org/abs/2103.08478.

[8] Bouchard-Côté, A., Vollmer, S. J. and Doucet, A. (2018). The bouncy particle sampler: a nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113**, 855–867.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press

[9] CHEN, F., LOVÁSZ, L. AND PAK, I. (1999). Lifting Markov chains to speed up mixing. In *STOC '99: Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, Association for Computing Machinery, New York, pp. 275–281.
[10] CHEVALLIER, A., FEARNHEAD, P. AND SUTTON, M. (2023). Reversible jump PDMP samplers for variable selection. *J. Amer. Statist. Assoc.* **118**, 2915–2927.
[11] DAVIS, M. H. A. (1993). *Markov Models and Optimization*. Routledge, New York.
[12] DELIGIANNIDIS, G., PAULIN, D., BOUCHARD-CÔTÉ, A. AND DOUCET, A. (2021). Randomized Hamiltonian Monte Carlo as scaling limit of the bouncy particle sampler and dimension-free convergence rates. *Ann. Appl. Prob.* **31**, 2612–2662.
[13] DIACONIS, P., HOLMES, S. AND NEAL, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Prob.* **10**, 726–752.
[14] DINH, V., BILGE, A., ZHANG, C. AND MATSEN, F. A., IV (2017). Probabilistic path Hamiltonian Monte Carlo. In *Proc. 34th International Conference on Machine Learning* (Proceedings of Machine Learning **70**), Machine Learning Research Press, pp. 1009–1018.
[15] DUNSON, D. B. AND JOHNDROW, J. (2020). The Hastings algorithm at fifty. *Biometrika* **107**, 1–23.
[16] DURMUS, A., GUILLIN, A. AND MONMARCHÉ, P. (2021). Piecewise deterministic Markov processes and their invariant measures. *Ann. Inst. H. Poincaré Prob. Statist.* **57**, 1442–1475.
[17] ETHIER, S. N. AND KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. John Wiley, New York.
[18] FEARNHEAD, P., BIERKENS, J., POLLOCK, M. AND ROBERTS, G. O. (2018). Piecewise deterministic Markov processes for continuous-time Monte Carlo. *Statist. Sci.* **33**, 386–412.
[19] GRAZZI, S. (2023). Design and applications of Monte Carlo methods based on piecewise deterministic Markov processes. Doctoral Thesis, Delft University of Technology.
[20] HARLAND, J., MICHEL, M., KAMPMANN, T. A. AND KIERFELD, J. (2017). Event-chain Monte Carlo algorithms for three- and many-particle interactions. *Europhys. Lett.)* **117**, article no. 30001.
[21] KOSKELA, J. (2022). Zig-zag sampling for discrete structures and nonreversible phylogenetic MCMC. *J. Comput. Graph. Statist.* **31**, 684–694.
[22] KRAUTH, W. (2021). Event-chain Monte Carlo: foundations, applications, and prospects. *Frontiers Phys.* **9**, article no. 663457.
[23] MICHEL, M., DURMUS, A. AND SÉNÉCAL, S. (2020). Forward event-chain Monte Carlo: fast sampling by randomness control in irreversible Markov chains. *J. Comput. Graph. Statist.* **29**, 689–702.
[24] NAKAJIMA, J. AND WEST, M. (2013). Bayesian analysis of latent threshold dynamic models. *J. Business Econom. Statist.* **31**, 151–164.
[25] NISHIMURA, A., DUNSON, D. B. AND LU, J. (2020). Discontinuous Hamiltonian Monte Carlo for discrete parameters and discontinuous likelihoods. *Biometrika* **107**, 365–380.
[26] PAKMAN, A. (2017). Binary bouncy particle sampler. Preprint. Available at https://arxiv.org/abs/1711.00922v1.
[27] PAKMAN, A. AND PANINSKI, L. (2014). Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *J. Comput. Graph. Statist.* **23**, 518–542.
[28] PETERS, E. A. J. F. AND DE WITH, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E* **85**, article no. 026703.
[29] RAFTERY, A. E. AND AKMAN, V. (1986). Bayesian analysis of a Poisson process with a change-point. *Biometrika* **73**, 85–89.
[30] TERENIN, A. AND THORNGREN, D. (2018). A piecewise deterministic Markov process via (r, θ) swaps in hyperspherical coordinates. Preprint. Available at http://arxiv.org/abs/1807.00420.
[31] VANETTI, P., BOUCHARD-CÔTÉ, A., DELIGIANNIDIS, G. AND DOUCET, A. (2017). Piecewise-deterministic Markov chain Monte Carlo. Preprint. Available at http://arxiv.org/abs/1707.05296.
[32] WANG, A. Q., POLLOCK, M., ROBERTS, G. O. AND STEINSALTZ, D. (2021). Regeneration-enriched Markov processes with application to Monte Carlo. *Ann. Appl. Prob.* **31**, 703–735.
[33] WU, C. AND ROBERT, C. P. (2020). Coordinate sampler: a non-reversible Gibbs-like MCMC sampler. *Statist. Comput.* **30**, 721–730.
[34] ZHOU, G. (2020). Mixed Hamiltonian Monte Carlo for mixed discrete and continuous variables. In *NIPS '20: Proceedings of the 34th Conference on Neural Information Processing Systems*, Neural Information Processing Systems Foundation, San Diego, CA, pp. 17094–17104.

https://doi.org/10.1017/apr.2023.61 Published online by Cambridge University Press