Scaling of piecewise deterministic Monte Carlo for anisotropic targets

JORIS BIERKENS[1,a] , KENGO KAMATANI[2,b] and GARETH O. ROBERTS[3,c]

[1]*Delft Institute of Applied Mathematics, Delft, Netherlands,* [a]*joris.bierkens@tudelft.nl*
[2]*Institute of Statistical Mathematics, Tokyo, Japan,* [b]*kamatani@ism.ac.jp*
[3]*Department of Statistics, University of Warwick, Coventry, United Kingdom,* [c]*Gareth.O.Roberts@warwick.ac.uk*

Piecewise deterministic Markov processes (PDMPs) are a type of continuous-time Markov process that combine deterministic flows with jumps. Recently, PDMPs have garnered attention within the Monte Carlo community as a potential alternative to traditional Markov chain Monte Carlo (MCMC) methods. The Zig-Zag sampler and the Bouncy Particle Sampler are commonly used examples of the PDMP methodology which have also yielded impressive theoretical properties, but little is known about their robustness to extreme dependence or anisotropy of the target density. It turns out that PDMPs may suffer from poor mixing due to anisotropy and this paper investigates this effect in detail in the stylised but important Gaussian case. To this end, we employ a multi-scale analysis framework in this paper. Our results show that when the Gaussian target distribution has two scales, of order 1 and ϵ, the computational cost of the Bouncy Particle Sampler is of order ϵ^{-1}, and the computational cost of the Zig-Zag sampler is ϵ^{-2}. In comparison, the cost of the traditional MCMC methods such as RWM is of order ϵ^{-2}, at least when the dimensionality of the small component is more than 1. Therefore, there is a robustness advantage to using PDMPs in this context.

Keywords: Markov process; Monte Carlo methods; multi-scale analysis

1. Introduction

Piecewise deterministic Markov processes (PDMPs) are continuous-time non-reversible Markov processes driven by deterministic flows and jumps. Recently, PDMPs have attracted the attention of the Monte Carlo community as a possible alternative to traditional Markov chain Monte Carlo methods. The Zig-Zag sampler Bierkens, Fearnhead and Roberts (2019) and the Bouncy Particle Sampler Bouchard-Côté, Vollmer and Doucet (2018), Peters and de With (2012) are commonly used examples of the PDMP methodology to approximate target distributions defined in $\mathbb{R}^d$. Both Markov samplers operate in subsets of $\mathbb{R}^d \times \mathbb{R}^d$, with states represented as (x, v). When these processes achieve ergodicity, the empirical averages of the Markov processes in the x-space converge to the corresponding expectations under the original target distribution, by the law of large numbers.

These samplers have yielded impressive theoretical properties, for example in high-dimensional contexts Andrieu et al. (2021), Bierkens, Kamatani and Roberts (2022), Deligiannidis et al. (2021), in terms of asymptotic variance reduction Bierkens and Duncan (2017), and in the context of MCMC with large data sets Bierkens, Fearnhead and Roberts (2019).

It is well-known that all MCMCs suffer from deterioration of mixing properties in the context of anisotropy of the target distribution. For example the mixing of random walk Metropolis, Metropolis-Adjusted Langevin and Hamiltonian schemes can all become arbitrarily bad in any given problem for increasingly inappropriately scaled proposal distributions (see for example the studies in Beskos et al. (2013, 2018), Graham, Thiery and Beskos (2022), Roberts, Gelman and Gilks (1997), Roberts and Rosenthal (1998, 2001)). Meanwhile the Gibbs sampler is known to perform poorly in the context of highly correlated target distributions, see for example Roberts and Sahu (1997), Zanella and Roberts

(2021). In principle these problems can be overcome using preconditioning (a priori reparameterisations of the state space), often carried out using adaptive MCMC (Andrieu and Thoms, 2008, Roberts and Rosenthal, 2009). However, effective preconditioning can be very difficult in higher-dimensional problems. Therefore, it is important to understand the extent of this deterioration of mixing in the face of anisotropy or dependence.

PDMPs can also suffer from poor mixing due to anisotropy and this paper will study this effect in detail in the stylised but important Gaussian case. In principle, continuous time processes are well-suited to this problem because they do not require a rejection scheme, unlike traditional MCMC methods which suffer from a scaling issue. However, this does not guarantee that the processes will be effective for this problem. Even if the process mixes well, it may require a large number of jumps, resulting in high computational cost. To understand this quantitatively, it is important to determine the influence that anisotropies in the target distribution have on the speed convergence to equilibrium.

However, for continuous-time processes a further consideration relevant to the computational efficiency of an algorithm involves some measure of the computational effort expended per unit stochastic process time. It is difficult to be precise about how to measure this in general. Here we will use a commonly adopted surrogate: the number of jumps per unit time. Effectively this assumes that we have access to an efficient implementation scheme for our PDMP via an appropriate Poisson thinning scheme, which is not always available in practice (see Section 2.3 for the detail).

This quantification of the computational cost can be achieved by the scaling analysis which first appeared in Roberts, Gelman and Gilks (1997) in the literature under mathematically formal argument. In this paper, we follow Beskos et al. (2018) that studied the random walk Metropolis algorithm for anisotropic target distributions. To avoid technical difficulties, they considered a simple target distribution with both small and large-scale components, where these two components are conditionally independent. The scale of the noise of the algorithm must be appropriately chosen in advance, otherwise, the algorithm will be frozen. On the other hand, the piecewise deterministic Markov processes naturally fit the scale without artificial parameter tuning. However, from a theoretical point of view, this makes the analysis complicated since the process changes its dynamics depending on the target distribution. Therefore, we study a simple Gaussian example and identify the factors that change the dynamics for a better understanding of the processes for anisotropic target distributions.

The main contributions of this paper are thus the following:

1. We give a theoretical and rigorous study of the robustness of Zig-Zag and BPS algorithms in the case of increasingly extreme anisotropy (where one component is scaled by a factor ϵ which converges to 0 in the limit). Specifically, we derive weak convergence results that characterise the limiting dynamics of the slow component in the extreme anisotropy limit.
2. For the Zig-Zag, we see that, unless the slow component is either (a) perfectly aligned with one of the possible velocity directions, or (b) independent of the fast component, the limiting behaviour is a reversible Ornstein–Uhlenbeck process that mixes in time $O(\epsilon^{-1})$. Considering that the number of switches per unit time scales as $O(\epsilon^{-1})$ (by Theorem 3.3), the total complexity grows as $O(\epsilon^{-2})$.
3. For the BPS in Theorem 3.4 we show that its limiting behaviour is a deterministic dynamical system for which we can explicitly write down its generating ODE. Mixing occurs in $O(1)$ time, which when combined with a cost of $O(\epsilon^{-1})$ switches per unit time, yields an overall complexity of $O(\epsilon^{-1})$.

The paper is organised as follows. In Section 2 we explore the problem in a 2-dimensional setting where we can easily illustrate and motivate our results. Here we describe the limiting processes using the homogenisation and averaging techniques of Pavliotis and Stuart (2008). Section 3 then describes our main results in detail and signposts the proofs. Section 4 then gives the detailed proofs for the Zig-Zag and BPS respectively, supported by technical material from the supplementary material (Bierkens,

Kamatani and Roberts, 2025). More specifically, in Section 4, we illustrate the properties of the leading-order generator. Additionally, in the supplementary material, we demonstrate weak convergence for the Zig-Zag sampler through homogenisation, and for the Bouncy Particle Sampler through averaging methods. A final discussion is given in Section 5.

2. Exhibit of main results in the two-dimensional case

To better understand our result, we first consider the two-dimensional case, which is easier to interpret than the general result. Consider the situation of a centered two-dimensional Gaussian target distribution with one large and one small component. Indeed, let $\Sigma^\epsilon = U^\top (\Lambda^\epsilon)^2 U$, where

$$U = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{and} \quad \Lambda^\epsilon = \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \end{pmatrix}. \tag{1}$$

We will show that the Markov processes of the Zig-Zag sampler and the Bouncy Particle Sampler converges to limit processes as $\epsilon \to 0$ if we choose a correct scaling in space and time.

To illustrate the impact of ϵ, we will use a reparametrised variable for our scaling analysis:

$$y = (\Lambda^\epsilon)^{-1} U x. \tag{2}$$

By this variable scaling, the invariant distribution of the Markov semigroup corresponding to $\xi = (y, v)$ is $\mu = \mathcal{N}_d(0, I_d) \otimes \mathcal{U}(\{-1, +1\}^d)$ for the Zig-Zag sampler, and $\mu = \mathcal{N}_d(0, I_d) \otimes \mathcal{N}_d(0, I_d)$ for the Bouncy Particle Sampler.

2.1. The Zig-Zag sampler in the two dimensional case

The Markov generator of the two-dimensional Zig-Zag sampler with reparametrisation is

$$\begin{aligned}
\mathcal{L}^\epsilon f(y, v) = {} & (v_1 \cos\theta - v_2 \sin\theta) \partial_{y_1} f(y, v) + \epsilon^{-1} (v_1 \sin\theta + v_2 \cos\theta) \partial_{y_2} f(y, v) \\
& + \left(v_1 (y_1 \cos\theta + \epsilon^{-1} y_2 \sin\theta) \right)_+ (\mathcal{F}_1 - \mathrm{id}) f(y, v) \\
& + \left(v_2 (-y_1 \sin\theta + \epsilon^{-1} y_2 \cos\theta) \right)_+ (\mathcal{F}_2 - \mathrm{id}) f(y, v),
\end{aligned} \tag{3}$$

where $\mathcal{F}_i f(y, v) = f(y, F_i(v))$ and F_i represents the operation that flips the sign of the i-th component. Figure 1 displays the trajectories of the Zig-Zag sampler in a two-dimensional scenario, with $\theta = 0, \pi/6$ and $\pi/4$ and $\epsilon = 0.01$.

Our first observation from Figure 1 is that the process of (y_2, v_1, v_2) exhibits good mixing, whereas the process of y_1 does not. This result is expected, as the terms in the expression for $\mathcal{L}^\epsilon$ that correspond to the dynamics of (y_2, v_1, v_2) are proportional to ϵ^{-1}, whereas the dynamics of y_1 are of order $O(1)$. When ϵ is sufficiently small, the dynamics of (y_2, v_1, v_2) are described by the operator $\mathcal{L}_0$, which is derived from $\mathcal{L}^\epsilon$ by retaining only terms proportional to ϵ^{-1}:

$$\begin{aligned}
\mathcal{L}_0 f(y, v) = {} & (v_1 \sin\theta + v_2 \cos\theta) \, \partial_{y_2} f(y, v) \\
& + (v_1 y_2 \sin\theta)_+ (\mathcal{F}_1 - \mathrm{id}) f(y, v) + (v_2 y_2 \cos\theta)_+ (\mathcal{F}_2 - \mathrm{id}) f(y, v).
\end{aligned}$$

Our second observation is that the Zig-Zag process corresponding to y_1 generally demonstrates diffusive behaviour, except in some specific situations, which correspond to the figures on the left and
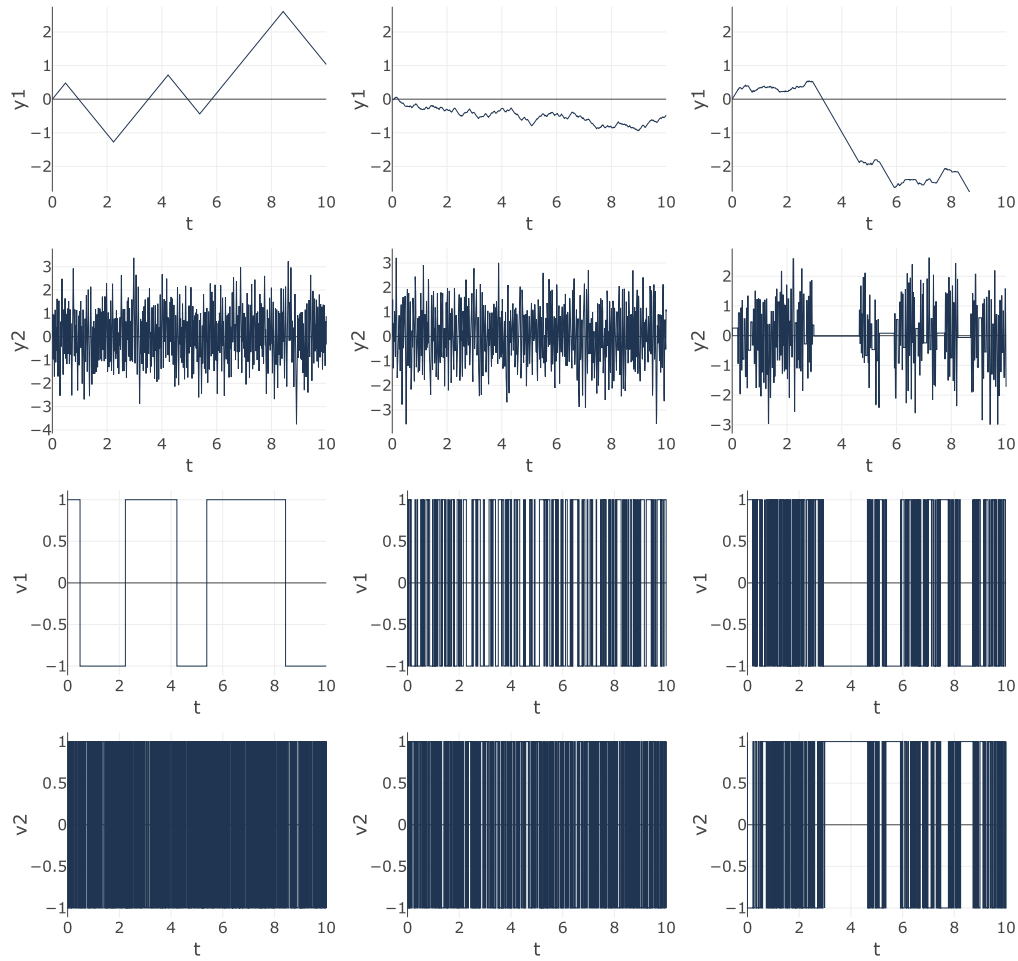
Figure 1. Paths of (y, v) for $\theta = 0$ (left), $\theta = \pi/6$ (centre), and $\theta = \pi/4$ (right), and y_1, y_2 (1st and 2nd rows) and v_1, v_2 (3rd and 4th rows) of the Zig-Zag sampler where $\epsilon = 0.01$.

right in Figure 1. The behaviour in these exceptional scenarios will be discussed in Section 2.1.1. For the time being, we make the following assumption.

Assumption 2.1. $\theta \in [0, 2\pi)$ but $\theta \neq n\pi/4$ ($n = 0, \ldots, 7$).

To clarify the asymptotic behavior, the solution of the Poisson equation plays a crucial role, as it enables the separation of the fast dynamics associated with the y_2 process and the slow dynamics associated with the y_1 process. Let $\chi(y_2, v)$ denote the solution to the Poisson equation

$$\mathcal{L}_0 \chi(y_2, v) = -(v_1 \cos \theta - v_2 \sin \theta). \tag{4}$$
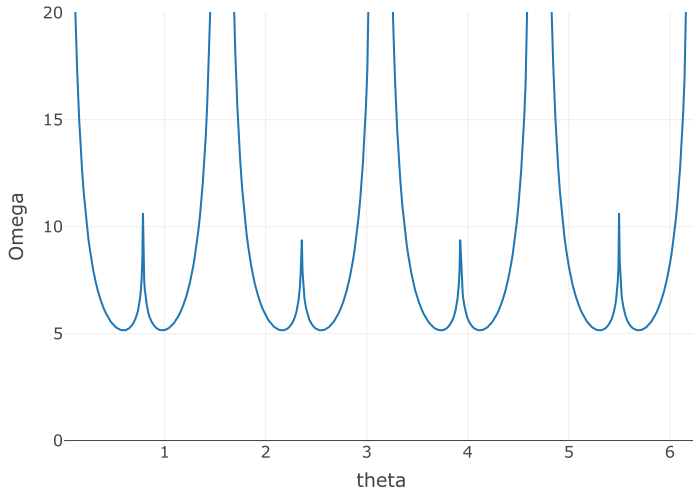
Figure 2. The value of the diffusion coefficient $\Omega = \Omega(\theta)$ for $\theta \in [0, 2\pi)$ for the Zig-Zag sampler. The function takes $+\infty$ when $\theta = n\pi/4, n = 0, 1, \dots, 7$.

Let $\Omega = -2\mu(\chi \, \mathcal{L}_0 \chi)$ which can be shown to admit the expression

$$
\Omega = \begin{cases}
\dfrac{8}{\sqrt{\pi}} \dfrac{\operatorname{arctanh} \sqrt{|\tan\theta|}}{(1 + |\sin 2\theta|)\sqrt{|\sin 2\theta|}} + \dfrac{\sqrt{2\pi}}{|\sin\theta|(1 + |\sin 2\theta|)} & \text{if } |\sin\theta| < |\cos\theta| \\[4ex]
\dfrac{8}{\sqrt{\pi}} \dfrac{\operatorname{arctanh} \sqrt{1/|\tan\theta|}}{(1 + |\sin 2\theta|)\sqrt{|\sin 2\theta|}} + \dfrac{\sqrt{2\pi}}{|\cos\theta|(1 + |\sin 2\theta|)} & \text{if } |\sin\theta| > |\cos\theta|
\end{cases}
\tag{5}
$$

for $\theta \neq n\pi/4$ ($n = 0, \dots, 7$). See Figure 2. The expression of χ, along with the derivation of Ω, can be found in Section C.

Let $y_1^\epsilon(t)$ be the y_1-coordinate of the Zig-Zag sampler process. The following is the formal statement of our second observation. Let $\mathbb{D}[0, T]$ be the space of càdlàg processes on $[0, T]$ equipped with the Skorokhod topology.

Theorem 2.2. *Under Assumption 2.1, and when the process is stationary, the ϵ^{-1}-time scaled process $y_1^\epsilon(\epsilon^{-1}t)$ converges weakly in $\mathbb{D}[0, T]$ for any $T > 0$ to the Ornstein–Uhlenbeck process*

$$
dX_t = -\frac{\Omega}{2} X_t dt + \Omega^{1/2} dW_t,
$$

where W_t is the one-dimensional standard Wiener process.

The convergence of this process is a consequence of the multi-dimensional result established in Theorem 3.2. Figure 3 shows that the theory agrees well with a numerical experiment.

2.1.1. The fully and diagonally aligned cases of the Zig-Zag sampler

In this section, we will examine the special cases excluded in Assumption 2.1. When θ is $\theta = n\pi/4$ ($n = 0, \dots, 7$), then Ω diverges. These eight cases can be further divided into two categories. When $\theta = n\pi/4$

(a) $\theta \in (0, \pi/4)$ (b) Zoomed in at a sub-interval.
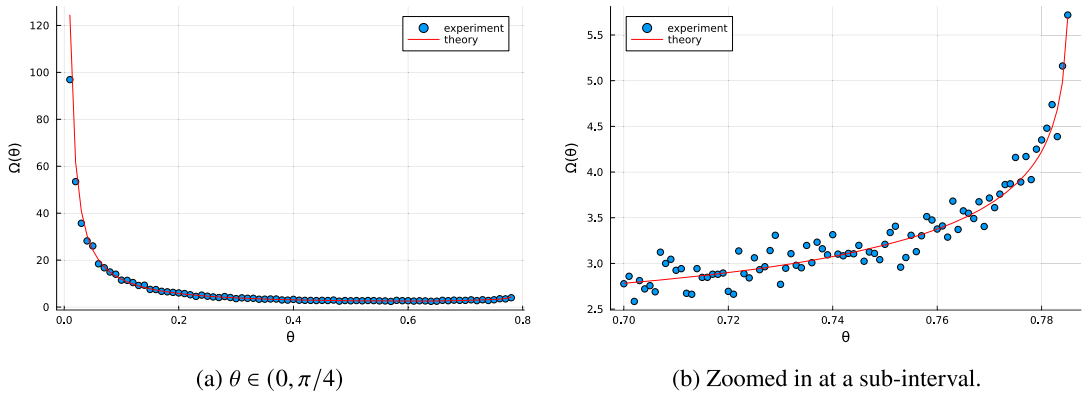
Figure 3. Numerical validation of the theoretical value $\Omega = \Omega(\theta)$ in the range $\theta \in (0, \pi/4)$. This is achieved by statistical estimation of the diffusion coefficient for a grid of values for θ, interpreting the process $y_1^\epsilon(t)$ as a diffusion. This is for $\epsilon = 0.001$ and a trajectory with time horizon $T = 2.0$.

and n is even, either $\cos\theta$ or $\sin\theta = 0$ and as a result, y_1 and y_2 are independent. The path of y_1 in this scenario corresponds to the one-dimensional Zig-Zag sampler and exhibits good mixing behaviour without diffusive behaviour. We refer to these cases as the *'fully aligned' scenario*.

On the other hand, if θ equals $n\pi/4$ and n is an odd number, the process remains diffusive but now allows for large jumps. These cases are referred to as the *'diagonally aligned' scenario*. We give an informal but explicit description of how the trajectories behave in this case, and how this differs in comparison to the other cases.

We will make use of the following simple result.

Lemma 2.3. *Suppose X is a positive random variable with hazard rate $(a + \gamma t)_+$ for some real number a and positive constant γ, i.e.,*

$$\mathbb{P}(X \geq t) = \exp\left(-\int_0^t (a + \gamma s)_+ \, \mathrm{d}s\right), \quad t \geq 0.$$

Then for any positive constant c,

$$\mathbb{P}(X + a\gamma^{-1} \geq c\gamma^{-1/2}) = \begin{cases} \exp\left(-c^2/2 + a_+^2/2\gamma\right), & c \geq a\gamma^{-1/2}, \\ 1, & 0 \leq c < a\gamma^{-1/2}, \end{cases}$$

and

$$\mathbb{E}(X + a\gamma^{-1}) = a_+/\gamma + \gamma^{-1/2}\sqrt{2\pi}\,\Phi(-a_+\gamma^{-1})\,e^{a_+^2/2\gamma}.$$

By (3), we have for a starting position $(y_1(0), y_2(0), v_1, v_2)$ that

$$y_1(t) = y_1(0) + (v_1\cos\theta - v_2\sin\theta)t, \quad y_2(t) = y_2(0) + \epsilon^{-1}(v_1\sin\theta + v_2\cos\theta)t$$

until next jump. By the expression from the jump rates in (3) the jump rates of the first coordinate and the second coordinate are summarised as

$$\lambda_1(t) = (c_1 + \epsilon^{-1}a_1 + (1 - \gamma_1)t + \epsilon^{-2}\gamma_1 t)_+, \quad \lambda_2(t) = (c_2 + \epsilon^{-1}a_2 + (1 - \gamma_2)t + \epsilon^{-2}\gamma_2 t)_+,$$

where

$$\gamma_1 = \sin\theta^2 + v_1 v_2 \sin\theta\cos\theta, \quad \gamma_2 = \cos\theta^2 + v_1 v_2 \sin\theta\cos\theta$$

and

$$a_1 = y_2 v_1 \sin\theta, \quad a_2 = y_2 v_2 \cos\theta,$$

and $c_1, c_2 \in \mathbb{R}$ are constants depending on the initial positions. Note that, if $v_1 \sin\theta \neq -v_2 \cos\theta$, then

$$\gamma_1 + \gamma_2 = (v_1 \sin\theta + v_2 \cos\theta)^2 > 0 \implies \gamma_1 > 0 \text{ or } \gamma_2 > 0.$$

On the other hand, $v_1 \sin\theta = -v_2 \cos\theta$ implies that $\gamma_1 = \gamma_2 = 0$ and one of a_1 and a_2 is positive unless $y_2 = 0$. Note that, $|v_1 \sin\theta| = |v_2 \cos\theta|$ is satisfied if and only if θ is in the diagonally aligned cases. Moreover, in this case, the two events $v_1 \sin\theta = -v_2 \cos\theta$ and $v_1 \sin\theta \neq -v_2 \cos\theta$ occur one after the other and the events are interchanged by jumps.

So by Lemma 2.3, we identify two cases for (v_1, v_2):

1. If $v_1 \sin\theta \neq -v_2 \cos\theta$ then the next jump time is an $O_P(\epsilon)$ random variable;
2. If $v_1 \sin\theta = -v_2 \cos\theta$ then the next jump time is an $O_P(\epsilon |y_2|^{-1})$ random variable.

In fact it is easy to see that from starting values $(y_1(0), y_2(0))$ on $\mathbb{R} \times \mathbb{R}$ and with any initial velocities, the switch random variables corresponding to v_1 and v_2 have explicit hazard rates, many of them in the form of that described in Lemma 2.3 with differing γ values summarised in the following result.

Proposition 2.4. *With the notation introduced above,*

(i) *The next jump distribution is stochastically bounded above by a random variable of the type appearing in Lemma 2.3 with*

$$\gamma = \epsilon^{-2} \max\{\gamma_1, \gamma_2\}$$

*with $a = \max\{c_1, c_2\} + \epsilon^{-1} \max\{a_1, a_2\}$.*

(ii) *If $\theta \neq (2n+1)\pi/4$, then uniformly over the state space, as $\epsilon \to 0$ and the next jump distribution is bounded above by a $O_P(\epsilon)$ random variable.*

(iii) *If $\theta = (2n+1)\pi/4$, then two subsequent jump distributions are of the order of $O(\epsilon |y_2|^{-1})$ and $O_P(\epsilon)$ in turn.*

Thanks to these observations, and also by the simulation results, we conjecture that the process y_1 converges to a diffusion with jumps in this case. Here, the jump occurs when $|y_2|$ is small. However, this case is technically challenging. In a real-world application the narrow posterior distribution will likely not be perfectly aligned with the diagonal, and in an exceptional case of perfect diagonal alignment we expect a better performance compared to our worst case analysis for the non-aligned case.

2.2. The Bouncy Particle Sampler in the two dimensional case

Next, we present the asymptotic behaviour for the Bouncy Particle Sampler (BPS) for $d = 2$. The BPS is rotationally invariant and we may therefore assume $U = I_d$ without loss of generality. The associated generator in terms of the coordinates (x, v) is given by equation (9). With the reparametrisation in (2), the generator can be expressed as

$$\mathcal{L}^\epsilon f(y, v) = v_1 \partial_{y_1} f(y, v) + \epsilon^{-1} v_2 \partial_{y_2} f(y, v) + (v_1 y_1 + \epsilon^{-1} v_2 y_2)_+ (\mathcal{B}^\epsilon - \mathrm{id}) f(y, v),$$
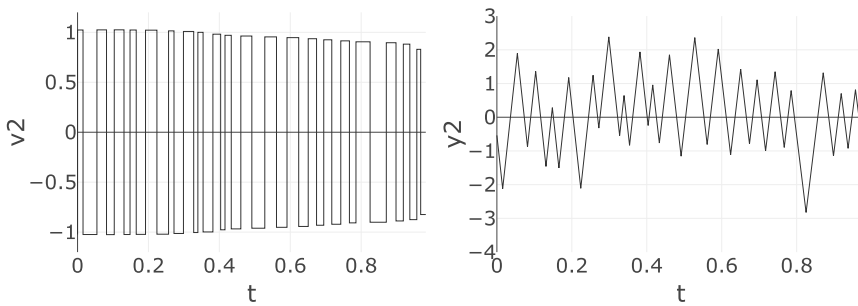
Figure 4. Paths of v_2 (left) and y_2 (right) for the Bouncy Particle Sampler for $\epsilon = 0.01$.

where we omit the refreshment term for simplicity (see Remarks 2.1 and 2.2 below), and where the operator $\mathcal{B}^\epsilon$ is given as the pullback of the reflection

$$(y,v) \quad \mapsto \quad \left(y, \quad v - 2\frac{v_1 y_1 + \epsilon^{-1} v_2 y_2}{y_1^2 + \epsilon^{-2} y_2^2}\left(\begin{matrix} y_1 \\ \epsilon^{-1} y_2 \end{matrix}\right)\right) =: \left(y, v + \left(\begin{matrix} \Delta^\epsilon v_1 \\ \Delta^\epsilon v_2 \end{matrix}\right)\right).$$

Observe that the jump size $\Delta^\epsilon v_1$ of v_1 induced by $\mathcal{B}^\epsilon$ satisfies

$$\epsilon^{-1}\Delta^\epsilon v_1 \longrightarrow_{\epsilon \to 0} \quad -2\frac{v_2\, y_2}{y_2^2}\, y_1. \tag{6}$$

Our first observation is that, as shown in Figure 4, the process of $\text{sgn}(v_2)$ and that of y_2 mixes well but that of (y_1, v_1) does not. This is not surprising given the form of the generator. When ϵ is small enough, the process of (y_2, v_2) is asymptotically dominated by the generator $\mathcal{L}_0$ corresponding to the one-dimensional Zig-Zag sampler

$$\mathcal{L}_0 f(y,v) = v_2 \partial_{y_2} f(y,v) + (v_2 y_2)_+ (\mathcal{B}^0 - \text{id}) f(y,v),$$

where $\mathcal{B}^0$ is given as the pullback of the coordinate reflection

$$(x,v) \mapsto \left(x, \left(\begin{matrix} v_1 \\ -v_2 \end{matrix}\right)\right).$$

Note that the value of $|v_2|$ remains unchanged by the dynamics generated by $\mathcal{L}_0$.

The second observation is that, as in Figure 5, the process (y_1, v_1) (and $|v_2|$) mixes slowly and vaguely follows a periodic orbit. The process of y_1 satisfies $\frac{d}{dt}y_1 = v_1$. Although the process of v_1 is driven by the reflection jump, the jump size becomes so small that it asymptotically follows

$$\frac{d}{dt}v_1 = (v_2 y_2)_+ \left(-2\frac{v_2 y_2}{y_2^2}y_1\right) = -2v_2^2\, 1_{\{v_2 y_2 > 0\}} y_1$$

as $\epsilon \to 0$ by (6). Observe that the value of $|v| = \sqrt{v_1^2 + v_2^2}$ does not change during the dynamics. Thus, for $\kappa^2 := |v(0)|^2$, we have $|v_2| = \sqrt{\kappa^2 - v_1^2}$. Also, the fast mixing component $(y_2, \text{sgn}(v_2))$ is averaged

(a) $\epsilon = 0.1$

(b) $\epsilon = 0.01$

(c) $\epsilon = 10^{-4}$

(d) $\epsilon = 10^{-5}$
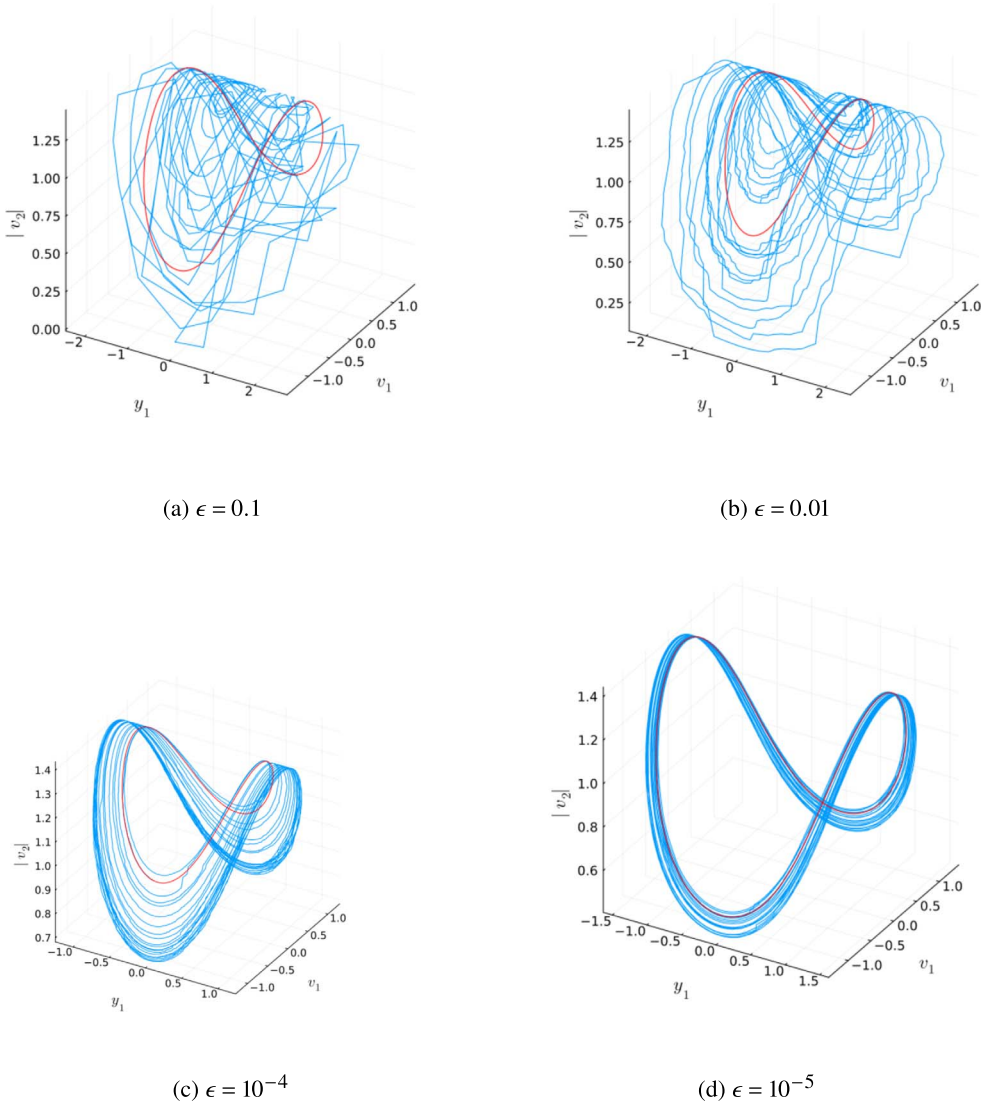
Figure 5. Numerical experiment illustrating how the $(y_1(t), v_1(t), |v_2|(t))$ statistics of a BPS trajectory (in blue) converge to the fluid limit (in red) given by (7).

out by $\mathcal{N}(0,1) \otimes \mathcal{U}(\{-1,+1\})$, and we obtain the dynamics

$$\frac{d}{dt}\begin{pmatrix} y_1 \\ v_1 \end{pmatrix} = \begin{pmatrix} v_1 \\ -(\kappa^2 - v_1^2)y_1 \end{pmatrix}. \tag{7}$$

See the red lines of Figure 5.

Let $(y_1^\epsilon(t), v_1^\epsilon(t))$ be the (y_1, v_1)-coordinate of the Bouncy Particle Sampler process. We have the following convergence result.

Theorem 2.5. *The stationary process* $(y_1^\epsilon(t), v_1^\epsilon(t))$ *converges to the solution of the differential equation (7) in the sense that*

$$\sup_{0 \le t \le T} |y_1^\epsilon(t) - y_1(t)| + |v_1^\epsilon(t) - v_1(t)| \longrightarrow_{\epsilon \to 0} 0$$

in probability for any $T > 0$ *where* $(y_1^\epsilon(0), v_1^\epsilon(0)) = (y_1(0), v_1(0))$, $\kappa^2 = |v^\epsilon(0)|^2$.

Proof. The claim follows from the general result Theorem 3.4 replacing $c(\alpha, \beta)$ by $\mathbb{E}[(v_2 y_2)_+^2 / |y_2|^2] = 1/2$ since the Skorokhod topology is equivalent to the local uniform topology when the limit process is continuous as described in page 124 of Section 12 of Billingsley (1999). □

Remark 2.1. For simplicity we have assumed the refreshment to be equal to zero. The only thing that would change for a fixed refreshment rate would be that the process (y_1, y_2, v_1, v_2) would experience refreshments at $O_P(1)$ random times. The trajectory between refreshments would not be affected by these refreshments, aside from having a new initial velocity vector. For the multi-dimensional case with refreshment jumps, refer to Theorem 3.4.

Remark 2.2. In the absence of refreshment, the limit process retains the value of $y_1^2 + \log(|v(0)|^2 - v_1^2)$, suggesting that the process is non-ergodic. However, with the introduction of refreshment, proving the process's ergodicity becomes straightforward. As this is beyond the scope of this paper, we do not include the details.

2.3. Discussion of the two-dimensional case

We summarise the results of the two-dimensional case in Table 1. Here, we estimate the time until convergence using a factor that accelerates convergence to a limiting process, which exhibits mixing properties. Additionally, we estimate the computational cost per unit time of the processes based on the number of jump events per unit time (see the discussion at the end of the section). The combined effort is then calculated as the product of these two factors.

Both the Zig-Zag Sampler (ZZS) and the Bouncy Particle Sampler (BPS) are capable of simulating anisotropic target distribution. However, the asymptotic behaviour of the two methods is systematically different. The former converges to a diffusion process and the latter to a deterministic periodic orbit. Moreover, the number of jumps to convergence is $O(\epsilon^{-2})$ for the ZZS (except for the non-generic, aligned cases), and $O(\epsilon^{-1})$ for the BPS. Therefore, the BPS is efficient in this scenario, which is in stark contrast to the high-dimensional, factorised scenario studied in Bierkens, Kamatani and Roberts (2022).

The behaviour of the Zig-Zag sampler is divided into three categories: Fully aligned ($\theta = n\pi/4$, n: even), diagonally-aligned ($\theta = n\pi/4$, n: odd), and non-aligned ($\theta \neq n\pi/4$). We are not yet fully able to understand the asymptotic behaviour of the diagonally aligned case. However, we conjecture that it converges to a diffusion with jumps.

For a discussion on a multi-dimensional scenario and a comparison of the computational complexity between PDMPs and the random walk Metropolis algorithm, please refer to Section 5.

Note here that estimating the computational cost is challenging since the implementation methods for PDMPs are not as established as those for Markov chain Monte Carlo methods. We assume the use of the Poisson thinning strategy to simulate jump times. This strategy employs a starting Poisson process with a higher conditional intensity function than that of the process we aim to simulate. We pick the true jump times from jump times of the starting Poisson process by the ratio of the conditional intensity functions.

Method	♯ Events/unit time	Time until convergence	Combined effort
ZZ ($\theta \neq n\pi/4$)	$O(\epsilon^{-1})$	$O(\epsilon^{-1})$	$O(\epsilon^{-2})$
ZZ ($\theta = n\pi/2$)	$O(\epsilon^{-1})$	$O(1)$	$O(\epsilon^{-1})$
BPS	$O(\epsilon^{-1})$	$O(1)$	$O(\epsilon^{-1})$

Table 1. Computational effort of the piecewise deterministic processes. The Zig-Zag sampler in the case of diagonally-aligned is not presented in this table.

Simulating the starting Poisson process is assumed to be straightforward, whereas calculating the ratio of the conditional intensities is computationally demanding due to the need to evaluate the derivative of the log of the negative density of the target distribution. Therefore, it is reasonable to estimate that the computational cost is linearly related to the number of jump events in the starting Poisson process. We assume that the number of jumps in the starting process is estimated by a constant multiple of that in the target process, and this multiplier is consistent across all processes listed in the table. Under this assumption, the computational effort per unit time can be estimated by the number of jump events in the target process.

It is important to note, however, that this is just an assumption and other estimates could also be valid depending on different factors and assumptions, including variability in implementation techniques or computational capabilities.

2.4. Non-Gaussian distributions

Our proof strategy assumes that the target distribution is multivariate Gaussian. In this section we investigate whether our results are seen to hold empirically beyond the Gaussian regime.

In Figure 6 we establish numerically the dependence of the estimated asymptotic variance (with respect to the second moment) as a function of the anisotropy $1/\epsilon$, both for a Gaussian distribution with covariance matrix Σ^ϵ and for a heavy-tailed multivariate Student distribution with parameters $\nu = 4$ (degrees of freedom) and covariance matrix Σ^ϵ.

Here the asymptotic variance of a Markov process (X_t^ϵ) satisfying a CLT with respect to a function f is defined as

$$\sigma_{\infty,\epsilon}^2(f) = \lim_{T \to \infty} \text{Var}\left(\frac{1}{\sqrt{T}} \int_0^T f(X_s^\epsilon)\, ds\right).$$

It may be checked that, if $(X_{t/\epsilon^\beta}^\epsilon) \Rightarrow (X_t^\infty)$ for a limiting process X^∞ as $\epsilon \downarrow 0$, along with certain additional conditions regarding the convergence of probability laws and ergodic properties, then $\sigma_{\infty,\epsilon}^2 \sim \epsilon^{-\beta} \sigma_\infty^2$, where σ_∞^2 is the asymptotic variance of (X_t^∞). Therefore, the estimated asymptotic variance may be used to estimate the asymptotic speed decrease as influenced by ϵ. The slopes of the regression lines plotted in these figures are displayed in Table 2 and give an empirical measure of the appropriate time scaling of the process with varying anisotropy.

Our implementation for the experiment is efficient in the sense that it leads to an approximate 0.4 acceptance probability of proposed switches for all values of ϵ for the Student distribution, illustrating that in this experiment the computational efficiency of the process does not deteriorate with increasing anisotropy, as argued in Subsection 2.3.

(a) $\theta = 0$

(b) $\theta = \pi/6$

(c) $\theta = \pi/4$

(d) $\theta = \pi/3$
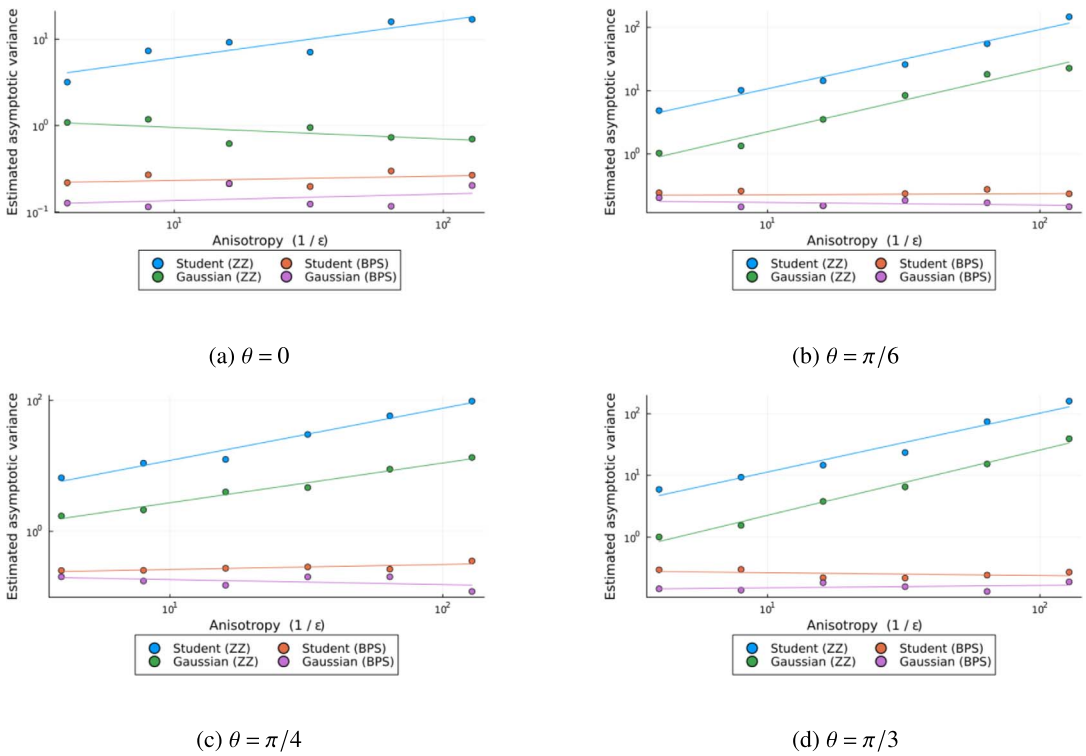
Figure 6. Log-log plot of the dependence of the estimated asymptotic variance on $1/\epsilon$ for a multivariate Gaussian distribution and a multivariate student distribution with $\nu = 4$ degrees of freedom. Plotted are the (geometric) mean over 100 experiments of the asymptotic variance, for simulations with a time horizon $T = 10/\epsilon$. The time horizon is made to vary with ϵ in order to estimate asymptotic variance reliably.

θ	0	$\pi/6$	$\pi/4$	$\pi/3$
Gaussian (ZZ)	-0.13 ± 0.07	1.00 ± 0.07	0.61 ± 0.08	1.06 ± 0.07
Gaussian (BPS)	0.07 ± 0.07	-0.04 ± 0.07	-0.08 ± 0.06	0.04 ± 0.07
Student (ZZ)	0.43 ± 0.07	0.94 ± 0.07	0.80 ± 0.07	0.96 ± 0.07
Student (BPS)	0.05 ± 0.07	0.02 ± 0.06	0.08 ± 0.06	$.05 \pm 0.06$

Table 2. Regression slopes corresponding to Figure 6, indicating the dependence of the logarithm of the asymptotic variance on $\log(1/\epsilon)$. An estimated slope of β indicates scaling of the asymptotic variance by a factor $1/\epsilon^\beta$. The theory of this paper predicts $\beta \approx 1$ for Zig-Zag, and $\beta \approx 0$ for BPS, in the Gaussian case. This table and Figure 6 suggest that these observations may also apply beyond the Gaussian regime.

3. Main results and proof strategy

Let $d \geq 2$ be an integer, and let K and L be a partition of $\{1, \ldots, d\}$ that is,

$$K \cup L = \{1, \ldots, d\}, \quad K \cap L = \varnothing$$

and $0 < \sharp K = k$ and $0 < \sharp L = l = d - k$. The target probability distribution is $\mathcal{N}_d(0, \Sigma^\epsilon)$, where the $d \times d$-symmetric positive definite matrix Σ^ϵ has the eigen decomposition

$$\Sigma^\epsilon = U^\top (\Lambda^\epsilon)^2 U, \ \Lambda^\epsilon = \mathrm{diag}(\Lambda_K, \epsilon \, \Lambda_L),$$

where Λ_K and Λ_L are $k \times k$ and $l \times l$ diagonal matrices with strictly positive diagonal elements, and U is an orthogonal matrix with decomposition

$$U = \begin{pmatrix} U_{KK} & U_{KL} \\ U_{LK} & U_{LL} \end{pmatrix}.$$

In this paper, we will study two kinds of extended generators, the Zig-Zag sampler defined on $\mathbb{R}^d \times \{-1, +1\}^d$ and the Bouncy Particle Sampler defined on $\mathbb{R}^d \times \mathbb{R}^d$:

$$\text{ZZS}: \quad \mathfrak{L}^\epsilon f(x, v) = v^\top \partial_x f(x, v) + \sum_{i=1}^{d} (v_i ((\Sigma^\epsilon)^{-1} x)_i)_+ (\mathcal{F}_i - \mathrm{id}) f(x, v) \tag{8}$$

$$\text{BPS}: \quad \mathfrak{L}^\epsilon f(x, v) = v^\top \partial_x f(x, v) + (v^\top (\Sigma^\epsilon)^{-1} x)_+ \, (\mathfrak{B}^\epsilon - \mathrm{id}) f(x, v) + \rho \, (\mathcal{R} - \mathrm{id}) \, f(x, v), \tag{9}$$

where the reflection operator $\mathfrak{B}^\epsilon$ is a deterministic operation such that

$$(x, v) \quad \mapsto \quad \left(x, \quad v - 2 v^\top (\Sigma^\epsilon)^{-1} x \, \frac{(\Sigma^\epsilon)^{-1} x}{|(\Sigma^\epsilon)^{-1} x|^2} \right).$$

The constant $\rho \geq 0$ is called the refreshment rate, and the refresh operator $\mathcal{R}$ is defined by

$$(\mathcal{R} f)(x, v) = \int_{\mathbb{R}^d} f(x, w) \phi(w) \mathrm{d} w$$

where $\phi(w)$ is the probability density function of the d-dimensional standard normal distribution. The domains of the extended generators are the set of bounded functions on $(x, v) \in \mathbb{R}^d \times \{-1, +1\}^d$ and $(x, v) \in \mathbb{R}^d \times \mathbb{R}^d$ such that each function is differentiable in x. The cores of the extended generators have also been studied such as Durmus, Guillin and Monmarché (2021), Holderrieth (2021).

3.1. Main results for the Zig-Zag sampler

With the reparametrisation (2), the extended generator of the Zig-Zag process $\xi^\epsilon(t) = (y^\epsilon(t), v^\epsilon(t))$ is

$$\mathcal{L}^\epsilon f(y, v) = ((\Lambda^\epsilon)^{-1} U v)^\top \partial_y f(y, v) + \sum_{i=1}^{d} (v_i (U^\top (\Lambda^\epsilon)^{-1} y)_i)_+ \, (\mathcal{F}_i - \mathrm{id}) f(y, v). \tag{10}$$

We decompose the d-dimensional vector y into K and L components, i.e., $y = (y_K, y_L)$. When ϵ is small, except for some special cases, the process of (y_L, v) mixes well but that of y_K does not. We are interested in the latter behaviour since the slowest dynamics reflects the total computational cost of Monte Carlo methods. The limit process of the faster dynamics (y_L, v) is characterised by the extended generator

$$\mathcal{L}_0 = (\Lambda_L^{-1} (U v)_L)^\top \partial_{y_L} + \sum_{i=1}^{d} (v_i (U_{L,\cdot}^\top \Lambda_L^{-1} y_L)_i)_+ \, (\mathcal{F}_i - \mathrm{id}), \tag{11}$$

where, $U_{\mathrm{L},\cdot} = (U_{\mathrm{LK}}, U_{\mathrm{LL}})$. Before discussing the asymptotic result, we describe the special cases mentioned above. In the two-dimensional case described in Section 2.1, $\theta = n\pi/4$ ($n = 0, \ldots, 7$) were excluded in the analysis, which corresponds to the case where the Markov process with generator $\mathcal{L}_0$ is not ergodic as a process of (y_2, v). Ergodicity is related to the $d \times d$-positive semidefinite matrix

$$\Theta_{\mathrm{L}} = U_{\mathrm{L},\cdot}^{\top} \Lambda_{\mathrm{L}}^{-2} U_{\mathrm{L},\cdot}.$$

If $v_i(\Theta_{\mathrm{L}} v)_i > 0$, then the i-th component v_i is flippable with direction v, in the sense that if the process maintains direction v, there is a positive probability of switching the sign of the i-th component of v when t is large enough. Indeed,

$$(v_i(t)(U_{\mathrm{L},\cdot}^{\top} \Lambda_{\mathrm{L}}^{-1} y_{\mathrm{L}}(t))_i) = (v_i(U_{\mathrm{L},\cdot}^{\top} \Lambda_{\mathrm{L}}^{-1} y_{\mathrm{L}}(0))_i + t \, v_i(\Theta_{\mathrm{L}} v)_i) > 0$$

when t is large enough, where $v(t) = (v_1(t), \ldots, v_d(t)) = v$, and $y_{\mathrm{L}}(t) = y_{\mathrm{L}}(0) + t \, \Lambda_{\mathrm{L}}^{-1}(Uv)_{\mathrm{L}}$. We assume the following condition for flippability, which ensures the ergodicity of the process.

Assumption 3.1. $v^{\top}\Theta_{\mathrm{L}} v > 0$ for any $v \in \{-1, 0, +1\}^d \backslash \mathbf{0}$.

As described above, the process $(y_{\mathrm{L}}^{\epsilon}(t), v^{\epsilon}(t))$ converges weakly to a Markov process with an extended generator $\mathcal{L}_0$. However, in the same time scale, the process $y_{\mathrm{K}}^{\epsilon}(t)$ becomes degenerate. To address this, we must scale time appropriately. Specifically, we estimate $O(\epsilon^{-1})$ time for mixing. This is because the time for the process $y_{\mathrm{K}}^{\epsilon}$ to converge is accelerated by a factor of n, and the limiting Ornstein–Uhlenbeck process exhibits strong mixing properties. The proof will be presented at the end of Section A.

Theorem 3.2. *As $\epsilon \to 0$, under Assumption 3.1, and if $\xi^{\epsilon}(0)$ follows the stationary distribution $\mathcal{N}(0, I_d) \otimes \mathcal{U}(\{-1, +1\}^d)$, then $y_{\mathrm{K}}^{\epsilon}(\epsilon^{-1}t)$ converges to an Ornstein–Uhlenbeck process defined in (18) in the Skorokhod topology.*

Under the same conditions as in Theorem 3.2, we examine the quantity N_T^{ϵ} which represents the number of jumps in the process $(y^{\epsilon}(t), v^{\epsilon}(t))$ over the time interval $[0, T]$. The number of jumps, N_T^{ϵ}, is an indicator of the computational cost per unit time. Let $\Sigma_{i,i}^{\epsilon}$ and $(\Theta_{\mathrm{L}})_{i,i}$ represent the i-th diagonal entries of the matrices Σ^{ϵ} and Θ_{L}, respectively. The proof is postponed to Section A.

Based on the following theorem, we can conclude that $\mathbb{E}[N_T^{\epsilon}] = O(\epsilon^{-1}T)$. This implies that $O(\epsilon^{-2})$ jumps are required to achieve mixing, as we make precise in the following result.

Theorem 3.3. *Let $\theta_{i,i}^{\epsilon}$ be the (i,i)-th element of $(\Sigma^{\epsilon})^{-1}$ for $i = 1, \ldots, d$. Under the same conditions as Theorem 3.2, the expected number of jumps of the process $(y^{\epsilon}(t), v^{\epsilon}(t))$ in the time interval $[0, T]$ is*

$$\mathbb{E}[N_T^{\epsilon}] = \frac{T}{2\sqrt{2\pi}} \sum_{i=1}^{d} \sqrt{\theta_{i,i}^{\epsilon}}.$$

In particular,

$$\lim_{\epsilon \to 0} \epsilon \, \mathbb{E}[N_T^{\epsilon}] = \frac{T}{2\sqrt{2\pi}} \sum_{i=1}^{d} \sqrt{(\Theta_{\mathrm{L}})_{i,i}}.$$

3.2. Main results for the Bouncy Particle Sampler

With the reparametrisation (2), the extended generator of the Bouncy Particle Sampler process $\xi^\epsilon(t) = (y^\epsilon(t), v^\epsilon(t))$ is

$$\mathcal{L}^\epsilon = ((\Lambda^\epsilon)^{-1}v)^\top \partial_y + (v^\top (\Lambda^\epsilon)^{-1} y)_+ \ (\mathcal{B}^\epsilon - \mathrm{id}) + \rho \ (\mathcal{R} - \mathrm{id}), \tag{12}$$

where the reflection operator is defined by $\mathcal{B}^\epsilon f(y,v) = f(y, B^\epsilon(y)v)$ and

$$B^\epsilon(y)v = v - 2v^\top (\Lambda^\epsilon)^{-1} y \ \frac{(\Lambda^\epsilon)^{-1} y}{|(\Lambda^\epsilon)^{-1} y|^2}. \tag{13}$$

Unlike Zig-Zag sampler case, without loss of generality, we can assume $\Sigma^\epsilon = (\Lambda^\epsilon)^2$ since the process is rotationally invariant. Let $y^\epsilon(t) = (y_\mathrm{K}^\epsilon(t), y_\mathrm{L}^\epsilon(t))$ and $v^\epsilon(t) = (v_\mathrm{K}^\epsilon(t), v_\mathrm{L}^\epsilon(t))$. In two dimensional case in Section 2.2, (y_2, v_2) mixes faster and (y_1, v_1) mixes slower. Similarly, in the general case, $(y_\mathrm{L}, v_\mathrm{L})$ mixes faster except some statistics described below, and $(y_\mathrm{K}, v_\mathrm{K})$ mixes slower.

The limit behaviour depends on the value of Λ_L. Using a different matrix would result in a different limit process, although the rate would remain unchanged. For the sake of simplicity, we assume Λ_L is the identity matrix. In this case, $\alpha = |v_\mathrm{L}|^2$, and $\beta = |v_\mathrm{L}|^2 |y_\mathrm{L}|^2 - (v_\mathrm{L}^\top y_\mathrm{L})^2$ are fixed constants, and $(y_\mathrm{L}, v_\mathrm{L})$ is always on a two dimensional hyperplane. This hyperplane will be refreshed at the refreshment jump time. In the asymptotic analysis, we need to study not only $(y_\mathrm{K}, v_\mathrm{K})$ but also (α, β) together with the effect of refreshment. The limit process is a little complicated, and so we will describe it Section B. The proof of the following theorem is also in the section. Let $(\alpha^\epsilon(t), \beta^\epsilon(t))$ be the corresponding process of the statistics (α, β).

Theorem 3.4. *As $\epsilon \to 0$, if $\Lambda_\mathrm{L} = I_\mathrm{L}$ and if $\xi^\epsilon(0)$ follows the stationary distribution $\mathcal{N}(0, I_d) \otimes \mathcal{N}(0, I_d)$, the Markov process $(y_\mathrm{K}^\epsilon(t), v_\mathrm{K}^\epsilon(t), \alpha^\epsilon(t), \beta^\epsilon(t))$ converges to an ordinary differential equation with jumps which will be defined in (19) in the supplementary material (Bierkens, Kamatani and Roberts, 2025) in the Skorokhod topology.*

Let N_T^ϵ be the number of jumps of the process $(y^\epsilon(t), v^\epsilon(t))$ within the time interval of $[0, T]$. Based on the following theorem, we can conclude that $\mathbb{E}[N_T^\epsilon] = O(\epsilon^{-1} T)$. This implies that $O(\epsilon^{-1})$ jumps are required to achieve mixing. The proof of the following theorem is in Section B.

Theorem 3.5. *Under the same conditions as those for Theorem 3.4, the expected number of jumps of $(y^\epsilon(t), v^\epsilon(t))$ in the time interval $[0, T]$ is*

$$\mathbb{E}[N_T^\epsilon] \le \frac{T}{2}(\mathrm{tr}((\Lambda^\epsilon)^{-1}) + 2\rho).$$

Also,

$$\lim_{\epsilon \to 0} \epsilon \ \mathbb{E}[N_T^\epsilon] = \frac{T}{2\sqrt{\pi}} \ \frac{\Gamma((l+1)/2)}{\Gamma(l/2)}.$$

3.3. Proof strategy

Our approach is to employ a multiscale analysis Pavliotis and Stuart (2008) to determine the limiting behaviours of both the Zig-Zag sampler and the Bouncy Particle Sampler. Multiscale analysis serves as

a comprehensive framework to simplify problems characterised by multiple scales. This framework is divided into two schemes: averaging and homogenisation. Averaging is often described as a first-order expansion and is seen as a result of the law of large numbers. Conversely, homogenisation is referred to as a second-order expansion and can emerge as a result of the central limit theorem.

More specifically, we examine the backward Kolmogorov equation for the Zig-Zag sampler and the Bouncy Particle Sampler. We introduce a perturbation of the corresponding extended generator through a scale factor ϵ. Our objective is to expand this equation and identify the limit. By identifying the limit, then we will show convergences of the corresponding processes using martingale theory.

The hightest order term in the generator expansions describes the dynamics of the fast components. Heuristically, these average out to a stationary distribution, conditional on the slow components that may be considered fixed. The asymptotic dynamics of the slow components may then be computed by averaging the lower order term in the generator with respect to the obtained stationary distribution of the fast components. This provides a heuristic description of the *averaging approach*.

In some cases the speed of the slow components averages out to have zero in the chosen time scale. In such cases an extra step of the expansion might be necessary, requiring us to look at an even longer, diffusive time scale. This is the situation where the *homogenisation approach* is useful.

The averaging approach has been employed in the Markov chain Monte Carlo literature by works such as Sherlock et al. (2015), Beskos et al. (2018) and Sherlock and Thiery (2022). However the application of the homogenisation approach seems to represent a novel contribution to this field. We define $u(x, v, t) = \mathbb{E}[\phi(\xi^\epsilon(t))|\xi^\epsilon(0) = (x, v)]$ for a real-valued function ϕ where $\xi^\epsilon(t)$ represents either the Zig-Zag process or the bouncy particle process. Under certain regularity conditions, the function u satisfies the backward Kolmogorov equation which can be expressed as

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \mathcal{L}^\epsilon u,$$

where $\mathcal{L}^\epsilon$ is the extended generator. The scaling limit of the extended generator $\mathcal{L}^\epsilon$ can be analysed by using the expansion of this differential equation in powers of a small parameter ϵ.

To understand the scaling limit of the Zig-Zag sampler, we employ the *homogenisation* approach. Specifically, we solve the differential equation

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \epsilon^{-1} \, \mathcal{L}^\epsilon u,$$

where the factor ϵ on the left-hand side represents time scaling. The extended generator $\mathcal{L}^\epsilon$ is then expanded into $\mathcal{L}^\epsilon = \epsilon^{-1}\mathcal{L}_0 + \mathcal{L}_1 + O(\epsilon)$ and the function u is expanded into

$$u = u_0 + \epsilon u_1 + \epsilon^2 u_2 + O(\epsilon^3).$$

This leads to a system of equations

$$\begin{cases} 0 & = & \mathcal{L}_0 u_0 \\ 0 & = & \mathcal{L}_0 u_1 + \mathcal{L}_1 u_0 \\ \dfrac{\mathrm{d}u_0}{\mathrm{d}t} & = & \mathcal{L}_0 u_2 + \mathcal{L}_1 u_1 \end{cases} \tag{14}$$

at orders $O(\epsilon^{-2}), O(\epsilon^{-1})$ and $O(1)$ of the backward Kolmogorov equation. The system can be used to derive a second-order differential equation for a function u_0, given by

$$\frac{\mathrm{d}u_0}{\mathrm{d}t} = f(u_0),$$

which characterise the limit of the process $\xi^\epsilon(t)$. The expression of f will be described later.

The Bouncy Particle Sampler can be analysed using the *averaging* approach of Pavliotis and Stuart (2008). This technique involves a different expansion of the Kolmogorov backward equation leading to a set of equations given by

$$
\begin{cases}
0 & = \quad \mathcal{L}_0 u_0 \\
\dfrac{\mathrm{d}u_0}{\mathrm{d}t} & = \quad \mathcal{L}_0 u_1 + \mathcal{L}_1 u_0.
\end{cases}
$$

Solving this set of equations leads to a first-order differential equation that characterises the scaling limit of the Bouncy Particle Sampler.

4. Properties of the leading-order generators

We will structure our convergence proofs based on the methodology presented in Section 3.3. Essentially, each proof, whether for Zig-Zag or Bouncy Particle Sampler, is divided into two parts. In the first part, we illustrate the properties of the leading-order generator. In the second part, we prove weak convergence via homogenization for the Zig-Zag sampler and through averaging for the Bouncy Particle Sampler. While the proof for the second part is lengthy, it's straightforward. Therefore, we've included all proofs related to this second part in the supplementary material (Bierkens, Kamatani and Roberts, 2025). Meanwhile, in the following sections of the main manuscript, we will focus exclusively on the first part.

4.1. Properties of the leading-order generator for the Zig-Zag sampler

The reparametrised extended generator of the Zig-Zag sampler (2) has a formal expansion

$$
\mathcal{L}^\epsilon = \epsilon^{-1} \mathcal{L}_0 + \mathcal{L}_1 + O(\epsilon),
$$

where

$$
\begin{aligned}
\mathcal{L}_0 &= (\Lambda_{\mathrm{L}}^{-1}(Uv)_{\mathrm{L}})^\top \partial_{y_{\mathrm{L}}} + \sum_{i=1}^{d} (v_i (U_{\mathrm{L},.}^\top \Lambda_{\mathrm{L}}^{-1} y_{\mathrm{L}})_i)_+ (\mathcal{F}_i - \mathrm{id}) \\
\mathcal{L}_1 &= (\Lambda_{\mathrm{K}}^{-1}(Uv)_{\mathrm{K}})^\top \partial_{y_{\mathrm{K}}} + \sum_{i=1}^{d} (v_i (U_{\mathrm{K},.}^\top \Lambda_{\mathrm{K}}^{-1} y_{\mathrm{K}})_i) \, 1(v_i (U_{\mathrm{L},.}^\top \Lambda_{\mathrm{L}}^{-1} y_{\mathrm{L}})_i \geq 0) \, (\mathcal{F}_i - \mathrm{id}).
\end{aligned}
\tag{15}
$$

For the expression of $\mathcal{L}_1$, we applied the expression $(a + \epsilon^{-1}b)_+ - \epsilon^{-1}b_+ - a \cdot 1(b > 0) = 0$ under the condition $|a| \leq \epsilon^{-1}|b|$. For detailed information, please refer to the supplementary material (Bierkens, Kamatani and Roberts, 2025).

We consider the operator $\mathcal{L}_0$ as acting on real-valued function of (y_{L}, v), where y_{K} is fixed. We establish the ergodicity of $\mathcal{L}_0$ by following the methodology outlined in Bierkens, Roberts and Zitt (2019). However, we need to make some modifications to the arguments in Bierkens, Roberts and Zitt (2019) since the velocity in $\mathcal{L}_0$ is not aligned with the coordinates.

If E is a topological space, a Markov process is called a T-process if there exists a probability measure a on $[0, \infty)$, a lower semicontinuous, non-trivial semi Markov kernel T such that $P_a(x, A) \geq T(x, A)$ ($x \in E, A \in \mathcal{E}$) where $P_a(x, A) = \int P_t(x, A) a(\mathrm{d}t)$. Let

$$
\mu = \mathcal{N}_{\mathrm{L}}(0, I_{\mathrm{L}}) \otimes \mathcal{U}(\{-1, +1\}^d).
$$

Lemma 4.1. *Under Assumption 3.1, the Markov process corresponding to the extended generator $\mathcal{L}_0$ is a T-process, ergodic and the invariant measure is μ.*

Proof. The core of the generator is $C_c^\infty(\mathbb{R}^l \times \mathbb{R}^d)$ as proven in Corollary 6.2 of Holderrieth (2021). Consequently, μ is the invariant measure of the semigroup by Proposition 4.9.2 of Ethier and Kurtz (1986). The rest of the proof is essentially the same as that of Section 2.2 and Section 3 of Bierkens, Roberts and Zitt (2019). Section 2.2 shows that for a multivariate normal target distribution, any state z is reachable from any other state z' in the sense that there is a piecewise deterministic path of $\mathcal{L}_0$ from z' to z such that the corresponding jump intensity is strictly positive at each jump point. In our case, $e_1, \ldots, e_d$ in Corollary 2 of the paper are replaced by $\Lambda_L^{-1}(Ue_i)_L$ $(i = 1, \ldots, d)$, where $e_i \in \mathbb{R}^d$ is a vector which is 1 at the i-th component and all other components are 0. Therefore, the number of velocity vectors is larger than the dimension of y_L, and the velocity vectors are angled. Nonetheless, we show that still, the argument in Section 2.2 is applicable to this case.

We apply Lemmas 1-3 of their paper to this case. By Assumption 3.1, Lemmas 1 and 3 of their paper can be directly applicable and so we focus on Lemma 2. The lemma states that for any $y, y' \in \mathbb{R}^d$, the state $(y', -v)$ is reachable from (y, v). Observe that

$$\Lambda_L^{-1}(Ue_i)_L = \Lambda_L^{-1}U_{LL}(e_i)_L \ (i \in L).$$

By applying a linear transformation $x \mapsto U_{LL}^{-1}\Lambda_L x$, the argument in Lemma 2 of Bierkens, Roberts and Zitt (2019) can be directly applied to the irreducibility proof of $\mathcal{L}_0$ if we omit $\Lambda_L^{-1}(Ue_i)_L$ $(i \in K)$. The proof will be carried out if we can handle the omitted component. On the other hand, thanks to Assumption 3.1, the components in K are asymptotically flippable in the sense that any coordinate can be switched after some time. Therefore, the argument in Lemma 2 can be applied to the current case after switching all K components. From this, the assertion follows.   □

We prove exponential ergodicity of $\mathcal{L}_0$ by showing exponential drift inequality. A slight modification of the drift function in Lemma 11 of Bierkens, Roberts and Zitt (2019) can be applied in this case. In the following, we apply the linear operators $\mathcal{L}_0$ and $\mathcal{L}_1$ to vector valued functions coordinatewise.

In the supplementary material, we show that the first order term u_1 of the perturbation solution to the backward Kolmogorov equation is constituted by a linear combination of $\chi_0(y_L, v)$. Here, $\chi_0(y_L, v)$ is itself the solution to the equation $\mathcal{L}_0\chi_0 = v$. The next proposition show the existence of χ_0.

Proposition 4.2. *The extended generator $\mathcal{L}_0$ is exponentially ergodic as an operator of the variable (y_L, v). In particular, there is a solution $\chi_0 : \mathbb{R}^l \times \mathbb{R}^d \to \mathbb{R}^d$ of the Poisson equation $\mathcal{L}_0\chi_0(y_L, v) = v$, and it is unique up to a constant. Moreover, for any $c > 0$ there is a constant c_0 such that $|\chi_0| \le c_0 \exp(c|y_L|^2)$. In particular, $\chi_0(y_L, v)$ has an arbitrary order of moments.*

Proof. As in Lemma 11 of Bierkens, Roberts and Zitt (2019), the exponential drift inequality can be determined by the drift function

$$V(y_L, v) = \exp\left(\alpha \, |y_L|^2/2 + \sum_{i=1}^d \kappa(v_i(U_{L,\cdot}^\top \Lambda_L^{-1} y_L)_i)\right),$$

where $0 < \alpha < 1$ and $\kappa(s) = \text{sgn}(s)\log(1 + |s|)$. By Theorem 3.2 of Glynn and Meyn (1996), since v is bounded, there is a solution to the Poisson equation for $\mathcal{L}_0\chi_0 = v$ and the solution satisfies $|\chi_0| \le c_0(V + 1)$ for some constant $c_0 > 0$. Thus the first claim follows since $\kappa(s) \le |s| \le 1 + |s|^2$. Also, since we can take $\alpha \le 1/m$ for any $m \in \mathbb{N}$, the solution has the m-th order of moment.   □

Lemma 4.3. *The solution $\chi_0(y_L, v)$ of the Poisson equation $\mathcal{L}_0 \chi_0 = v$ is in the domain of the extended generator $\mathcal{L}^\epsilon$, as a function of (y, v).*

Proof. The solution χ_0 is included in the domain of the extended generator $\mathcal{L}_0$ by construction. The domain of $\mathcal{L}^\epsilon$ is characterised by conditions (i-iii) of Theorem 5.5 of Davis (1984). The map $t \mapsto \chi_0(y_L + t \Lambda_L^{-1}(Uv)_L, v)$ is absolutely continuous, so it satisfies condition (i). Moreover, $|\chi_0| \le c_0 \exp(c|y_L|^2)$ for constants $c, c_0 > 0$, which means that the local integrability condition for $(\mathcal{F}_i - \text{id})\chi_0$ in (iii) of Theorem 5.5 is also satisfied for $i = 1, \ldots, d$. The generator $\mathcal{L}^\epsilon$ is no boundary, so it satisfies condition (ii). By Theorem 5.5, the function χ_0 is in the domain of the extended generator $\mathcal{L}^\epsilon$. $\qquad\square$

In the supplementary material, we identify the limit equation satisfied by u_0 which is expressed through expectations of χ_0. For this identification, we use a linear form of χ_0 denoted as $\chi : \mathbb{R}^l \times \mathbb{R}^d \to \mathbb{R}^k$:

$$\chi(y_L, v) = -(\Lambda_K^{-1}(U \chi_0)_K). \tag{16}$$

This is the solution of the Poisson equation

$$\mathcal{L}_0 \chi(y_L, v) = -\mathcal{L}_1 \psi(y, v) = -(\Lambda_K^{-1}(Uv)_K). \tag{17}$$

The limit equation of u_0 is a second order differential equation of y_K corresponding to the Ornstein–Uhlenbeck process

$$dX(t) = -\frac{1}{2}\Upsilon X(t)dt + \Omega^{1/2} \, dW(t), \tag{18}$$

where $W(t)$ is the k-dimensional standard Wiener process, and Ω and Υ are $k \times k$ matrices defined by

$$\Omega = -\mu((\mathcal{L}_0\chi)\,\chi^\top + \chi\,(\mathcal{L}_0\chi)^\top), \quad \Upsilon = -2\mu\left(\chi\,(\mathcal{L}_0\chi)^\top\right) \implies \Omega = \frac{\Upsilon + \Upsilon^\top}{2}. \tag{19}$$

Lemma 4.4. *The $k \times k$-matrix Ω is a symmetric positive definite matrix.*

Proof. Observe that the matrix can be expressed as

$$\Omega = -\mu\left(((\mathcal{L}_0 + \mathcal{L}_0^*)\chi)\,\chi^\top\right)$$

using the adjoint operator $\mathcal{L}_0^*$. Here, the adjoint operator is

$$\mathcal{L}_0^* = -(\Lambda_L^{-1}(Uv)_L)^\top \partial_{y_L} + \sum_{i=1}^{d}(-v_i(U_{L,.}^\top \Lambda_L^{-1} y_L)_i)_+ \, (\mathcal{F}_i - \text{id}).$$

By the expression of the adjoint operator, we have

$$\mathcal{L}_0 + \mathcal{L}_0^* = \sum_{i=1}^{d} a_i \, (\mathcal{F}_i - \text{id}), \quad a_i(y_L) = |(U_{L,.}^\top \Lambda_L^{-1} y_L)_i|.$$

Since $\mathcal{F}_i \mathcal{F}_i = \text{id}$, we have

$$-2 \sum_{v \in \{-1,+1\}^d} ((\mathcal{F}_i - \text{id})\chi)\,\chi^\top = \sum_{v \in \{-1,+1\}^d} ((\mathcal{F}_i - \text{id})\chi)((\mathcal{F}_i - \text{id})\chi)^\top \ge 0.$$

Therefore,

$$\Omega = -\sum_{i=1}^{d} \mu\left(a_i\left((\mathcal{F}_i - \mathrm{id})\chi\right)\chi^{\top}\right) \geq 0$$

Thus Ω is positive semidefinite. We show that it is indeed positive definite. If it is not positive definite, then there is a non-zero vector $u \in \mathbb{R}^k$ such that

$$u^{\top}\Omega u = \sum_{i=1}^{d} \mu(a_i \, |u^{\top}(\mathcal{F}_i - \mathrm{id})\chi|^2)/2 = 0.$$

On the other hand, $a_i(y_{\mathrm{L}}) = |(U_{\mathrm{L},\cdot}^{\top}\Lambda_{\mathrm{L}}^{-1}y_{\mathrm{L}})_i| \neq 0$ almost surely since the equality occurs when y is on the $d-1$ dimensional hyperplane, which is a null set for the Lebesgue measure. Therefore, the above equation implies that $u^{\top}(\mathcal{F}_i - \mathrm{id})\chi = (\mathcal{F}_i - \mathrm{id})(u^{\top}\chi) = 0$ almost surely for $i = 1,\dots d$. This implies that $u^{\top}\chi$ is in the null set of $\mathcal{F}_i - \mathrm{id}$ for any i. Therefore, $u^{\top}\chi$ does not depend on v. Thus

$$-u^{\top}(\Lambda_{\mathrm{K}}^{-1}(Uv)_{\mathrm{K}}) = \mathcal{L}_0(u^{\top}\chi) = u^{\top}\partial_{y_{\mathrm{L}}}\chi(y_{\mathrm{L}})\,(\Lambda_{\mathrm{L}}^{-1}(Uv)_{\mathrm{L}}) \quad v \in \{-1,+1\}^d,$$

where $\partial_{y_{\mathrm{L}}}\chi = (\partial_{y_j}\chi_i)_{i\in\mathrm{K},j\in\mathrm{L}}$, is a $k \times l$ matrix. Since the above equation is linear in v, the equation holds for any $v \in \mathbb{R}^d$. However, the left-hand side depends only on $(Uv)_{\mathrm{K}}$ and the right-hand side depends only on $(Uv)_{\mathrm{L}}$, so the equality holds only if both ends are 0. However, this is impossible since the left side is strictly positive if $u = (Uv)_{\mathrm{K}} \neq 0$. Thus Ω is positive definite. $\square$

Lemma 4.5. $\mu((\mathcal{L}_1(\chi \circ \psi))(y_{\mathrm{K}},\cdot)) = -\Upsilon\, y_{\mathrm{K}}/2$, and $\mu(m) = \Omega$ for

$$m(y_{\mathrm{L}},v) = \sum_{i=1}^{d}(v_i(U_{\mathrm{LL}}^{\top}\Lambda_{\mathrm{L}}^{-1}y_{\mathrm{L}})_i)_+\,((\mathcal{F}_i - \mathrm{id})\chi(y_{\mathrm{L}},v))^{\otimes 2}. \tag{20}$$

Proof. By construction, we have

$$\begin{cases} a_i & := v_i(U_{\mathrm{K},\cdot}^{\top}\Lambda_{\mathrm{K}}^{-1}y_{\mathrm{K}})_i, \\ b_i & := v_i(U_{\mathrm{L},\cdot}^{\top}\Lambda_{\mathrm{L}}^{-1}y_{\mathrm{L}})_i \end{cases} \implies \mathcal{L}_1(\chi \circ \psi)(y,v) = \sum_{i=1}^{d} a_i\, 1(b_i \geq 0)(\mathcal{F}_i - \mathrm{id})\chi(y_{\mathrm{L}},v).$$

Observe that

$$\mathcal{F}_i a_i = -a_i,\ \mathcal{F}_i b_i = -b_i \implies (\mathcal{F}_i - \mathrm{id})(a_i 1(b_i \geq 0)) = -a_i - a_i 1(b_i = 0). \tag{21}$$

From this fact, since $(\mathcal{F}_i - \mathrm{id})$ is μ-reversible, $\mu((\mathcal{L}_1(\chi \circ \psi))(y_{\mathrm{K}},\cdot))$ is

$$\sum_{i=1}^{d} \mu(a_i\, 1(b_i \geq 0)\,(\mathcal{F}_i - \mathrm{id})\chi) = \sum_{i=1}^{d} \mu(\chi\,(\mathcal{F}_i - \mathrm{id})a_i 1(b_i \geq 0))$$

$$= -\sum_{i=1}^{d} \mu(\chi\, a_i) = -\frac{\Upsilon}{2}\, y_{\mathrm{K}}.$$

Thus the first claim follows.

For the second claim, let

$$
\begin{cases}
\alpha_i & := v_i (U_{\mathrm{L},\cdot}^\top \Lambda_{\mathrm{L}}^{-1} y_{\mathrm{L}})_i, \\
\beta & := (Uv)_{\mathrm{L}}
\end{cases}
\implies \quad \mathcal{L}_0 = \beta^\top \partial_{y_{\mathrm{L}}} + \sum_{i=1}^{d} (\alpha_i)_+ (\mathcal{F}_i - \mathrm{id}).
$$

Let χ_i be the i-th component of $\chi(y_{\mathrm{L}}, v) \in \mathbb{R}^d$. Observe that there are two identities

$$
(\mathcal{F}_i - \mathrm{id})(\chi_k)\chi_l + (\mathcal{F}_i - \mathrm{id})(\chi_l)\chi_k - (\mathcal{F}_i - \mathrm{id})(\chi_k \chi_l) = -(\mathcal{F}_i - \mathrm{id})(\chi_k)(\mathcal{F}_i - \mathrm{id})(\chi_l),
$$

$$
(\partial_{y_{\mathrm{L}}} \chi_k)\chi_l \quad + (\partial_{y_{\mathrm{L}}} \chi_l)\chi_k \quad - \partial_{y_{\mathrm{L}}}(\chi_k \chi_l) \quad = 0,
$$

where we used the fact that $\mathcal{F}_i(fg) = \mathcal{F}_i(f)\mathcal{F}_i(g)$. From this fact, the (k,l)-th component of $m(y_{\mathrm{L}}, v)$ is,

$$
m_{kl}(y_{\mathrm{L}}, v) = \sum_{i=1}^{d} (\alpha_i)_+ (\mathcal{F}_i - \mathrm{id})(\chi_k)(\mathcal{F}_i - \mathrm{id})(\chi_l)
$$

$$
= -\mathcal{L}_0(\chi_l)\chi_k - \mathcal{L}_0(\chi_k)\chi_l + \mathcal{L}_0(\chi_l \chi_k).
$$

Since $\mathcal{L}_0$ is μ-invariant, the expectation of the third term in the right-hand side is 0. Thus we have

$$
\mu(m) = - \int (\mathcal{L}_0(\chi)\chi^\top + \chi \mathcal{L}_0(\chi)^\top)\mu(\mathrm{d}y_{\mathrm{L}}, \mathrm{d}v) = \Omega.
$$

$\square$

4.2. Properties of the leading-order generator of the Bouncy Particle Sampler

The generator of the Bouncy Particle Sampler (12) has a formal expansion

$$
\mathcal{L}^\epsilon = \epsilon^{-1} \mathcal{L}_0 + \mathcal{L}_1 + o(\epsilon), \tag{22}
$$

where the leading term is

$$
\mathcal{L}_0 = (\Lambda_{\mathrm{L}}^{-1} v_{\mathrm{L}})^\top \partial_{y_{\mathrm{L}}} + (v_{\mathrm{L}}^\top \Lambda_{\mathrm{L}}^{-1} y_{\mathrm{L}})_+ (\mathcal{B}^0 - \mathrm{id}) \tag{23}
$$

with a reflection operator $\mathcal{B}^0$ defined by $\mathcal{B}^0 f(x,v) = f(x, B^0(y)v)$ where

$$
B^0(y)v = \begin{pmatrix} v_{\mathrm{K}} \\ v_{\mathrm{L}} - 2v_{\mathrm{L}}^\top \Lambda_{\mathrm{L}}^{-1} y_{\mathrm{L}} \dfrac{\Lambda_{\mathrm{L}}^{-1} y_{\mathrm{L}}}{|\Lambda_{\mathrm{L}}^{-1} y_{\mathrm{L}}|^2} \end{pmatrix}.
$$

The operator $\mathcal{L}_0$ serves as an extended generator for the Markov process associated with the Bouncy Particle Sampler, in the absence of refreshment jumps (as outlined in Theorem 5.5 of Davis (1984)). The operator $\mathcal{L}_1$ will be discussed subsequently. In order to conduct multivariate analysis, it is necessary to identify the null space of $\mathcal{L}_0$. The null space is contingent upon Λ_{L} and necessitates a distinct analysis for varying null spaces. To simplify the analysis, we shall only consider the simplest case.

Assumption 4.6. $\Lambda_{\mathrm{L}} = I_{\mathrm{L}}$.
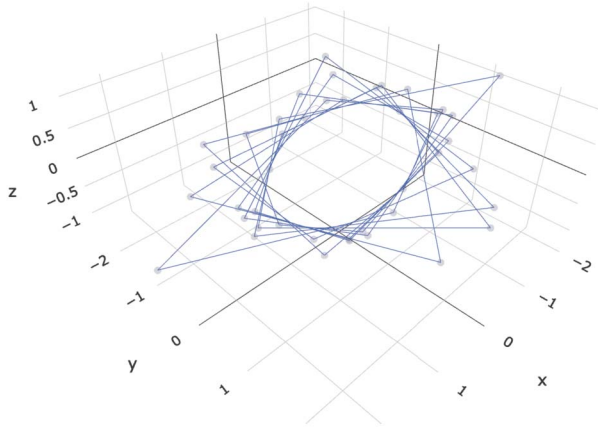
Figure 7. Typical path of the process $y_L(t)$.

In this scenario, the Markov process $(y_L(t), v_L(t))$ with extended generator $\mathcal{L}_0$ is by design constrained to the two-dimensional hyperplane spanned by $y_L(0)$ and $v_L(0)$, as depicted in Figure 7. By specifying two linearly independent vectors as the basis for this hyperplane, the dynamics of $y_L(t)$ and $v_L(t)$ are determined by at most four scalar parameters. Moreover,

$$
\begin{cases}
\alpha & := |v_L|^2 \\
\beta & := |v_L|^2 |y_L|^2 - (v_L^\top y_L)^2
\end{cases}
\tag{24}
$$

and

$$
r := \sqrt{\beta/\alpha}
$$

do not change through the dynamics. Let $f : \mathbb{R} \times \mathbb{S} \to \mathbb{R}$ such that $f(z, e)$ is differentiable with respect to z. Consider the operator

$$
Hf(z, e) = \alpha^{1/2} \left[\partial_z f(z, e) + z_+ (f(-z, b(z, e)) - f(z, e)) \right].
\tag{25}
$$

In (27), we will provide the expression for the vector $b(z, e)$, which can be uniquely determined based on the fixed orientation and the variables z and e. Let $(z(t), e(t))$ be the transformation of the process $(y_L(t), v_L(t))$ by

$$
z = |v_L|^{-1} v_L^\top y_L, \quad e = r^{-1} (y_L - z v_L / |v_L|).
\tag{26}
$$

Observe that e is a unit vector. Note that if $z(t)$ is continuous at t, then $z(t)' = |v_L(t)|$, and if not, $z(t) = -z(t-)$. We show that the process has the extended generator H.

Proposition 4.7. *Under Assumption 4.6, given the vectors $y_L(0)$ and $v_L(0)$, which are assumed to be linearly independent, we define Π as the oriented hyperplane spanned by these vectors. Then, we can assert that $(z(t), e(t))$ forms an ergodic Markov process with extended generator H, and its invariant distribution is $\mathcal{N}(0, 1) \otimes \mathcal{U}(\mathbb{S})$.*
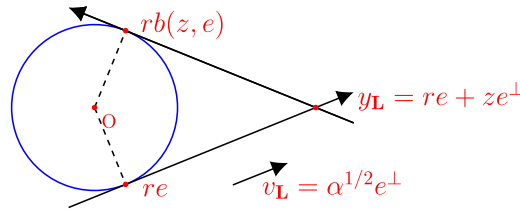
Figure 8. Trajectory of $y_L(t)$

Proof. Observe that linearly independence implies $\alpha > 0$. Let $0 = \tau_0 < \tau_1 < \cdots$ be the jump times of the Markov process $(y_L(t), v_L(t))$. For each $n = 0, 1, \ldots$, $v_L(t) = v_L(\tau_n)$ and

$$y_L(t) = y_L(\tau_n) + v_L(\tau_n)\,(t - \tau_n)$$

forms a line segment in each time interval $t \in [\tau_n, \tau_{n+1})$. By the constraint equation (24),

$$r^2 = |y_L|^2 - (v_L^\top y_L)^2 / |v_L|^2 \quad \Longrightarrow \quad |y_L| \geq r$$

and equality holds if and only if $z = y_L^\top v_L = 0$. Therefore, the line segment is tangent to the circle of radius r. Also, since

$$z(t) = z(\tau_n) + (t - \tau_n)\alpha^{1/2}, \quad t \in [\tau_n, \tau_{n+1})$$

and $z(\tau_n) = -z(\tau_n -) < 0$, the tangency point $z(t) = 0$ is included in each line segment. From this observation, for each line segment, we can define the orientation, clockwise or counterclockwise, by observing the direction at the tangency point.

We claim that the orientation of all line segments are consistent. Otherwise, there exists a set of line segments with differing orientations, such as $(y_L(t) : \tau_n \leq t < \tau_{n+1})$ and $(y_L(t) : \tau_{n+1} \leq t < \tau_{n+2})$. At $y(\tau_{n+1})$, there are two tangent lines to the two-dimensional circle of radius r. If the two line segments are on different tangent lines, then the orientation must be the same, given that the path $(y_L(t) : \tau_n \leq t < \tau_{n+2})$ is connected. Thus, if orientations are different, the two line segments are on the same tangent line, and the orientations of the two line segments must have opposite signs, i.e. $v_L(\tau_{n+1}-) = -v_L(\tau_{n+1})$. This can only occur when $v_L(t)/|v_L(t)| = y_L(t)/|y_L(t)|$ at $t = \tau_{n+1}-$. However, by turning back the time it implies linearly dependence of $v_L(0)$ and $y_L(0)$ which contradicts the initial assumption.

Observe that

$$e^\perp := \alpha^{-1/2} v_L \quad \Longrightarrow \quad z = y_L^\top e^\perp \text{ and } e = r^{-1}(y_L - z\, e^\perp) \quad \Longrightarrow \quad e \perp e^\perp.$$

The unit vector $e^\perp$ is solely determined by the vector e together with the orientation of the segments. By construction, $e \perp v_L$ and $r\, e = y_L - z v_L / |v_L|$ is on the line segment. Thus $r\, e$ is the point of tangency. See Figure 8. For given Π, α, β and the orientation of the process, there is a map $\varphi(z, e) = (y_L, v_L)$ from $\mathbb{R} \times \mathbb{S}$ to $\mathbb{R}^l \times \mathbb{R}^l$ defined by

$$v_L = \alpha^{1/2} e^\perp, \quad y_L = r e + z e^\perp.$$

If there is a reflection jump event at $t > 0$, then $e^\perp$ will jump to

$$(e^\perp)^* = e^\perp - 2 \frac{v_L^\top y_L}{|v_L|} \frac{y_L}{|y_L|^2} = \frac{-2rz}{r^2 + z^2} e + \frac{r^2 - z^2}{r^2 + z^2} e^\perp,$$

where we used $|y_\mathrm{L}|^2 = r^2 + z^2$. Also, z jumps to $-z$, and e jumps to

$$b(z,e) = r^{-1}(y_\mathrm{L} - (-z)\,(e^\perp)^*) = \frac{r^2 - z^2}{r^2 + z^2}e + \frac{2rz}{r^2 + z^2}e^\perp. \tag{27}$$

Since $(y_\mathrm{L}(t), v_\mathrm{L}(t))$ is a Markov process, $(z(t), e(t))$ is also a Markov process. The Markov process $(z(t), e(t))$ is ergodic if there is an invariant probability measure by Proposition 4.8 below and Theorem 1 of Kulik and Scheutzow (2015). We show that the invariant probability measure is $\mathcal{N}(0,1) \otimes \mathcal{U}(\mathbb{S})$. To see this, observe that the extended generator of the Markov process $(z(t), e(t))$ is

$$\widetilde{H}f(z,e) = \mathcal{L}_0(f \circ \varphi^{-1})(\varphi(z,e))$$

since for $Y_t := (z(t), e(t))$, $X_t = \varphi(X_t) = (y_\mathrm{L}(t), v_\mathrm{L}(t))$ and $g = f \circ \varphi^{-1}$, the process

$$M_T := f(Y_T) - f(Y_0) - \int_0^T (\widetilde{H}f)(Y_t)\mathrm{d}t = g(X_T) - g(X_0) - \int_0^T (\mathcal{L}_0 g)(X_t)\mathrm{d}t$$

is a local martingale when g is the domain of the extended generator of $\mathcal{L}_0$. We see that the extended generator $\widetilde{H}$ coincides H in (25). Observe that we have the expression

$$(z,e) = \varphi^{-1}(y_\mathrm{L}, v_\mathrm{L}) = (\alpha^{-1/2}\,y_\mathrm{L}^\top v_\mathrm{L}, r^{-1}(y_\mathrm{L} - (\alpha^{-1/2}y_\mathrm{L}^\top v_\mathrm{L})\,(\alpha^{-1/2}v_\mathrm{L}))).$$

By differentiating both sides of the equation, we obtain

$$v_\mathrm{L}^\top \partial_{y_\mathrm{L}} z = \alpha^{1/2}, \qquad v_\mathrm{L}^\top \partial_{y_\mathrm{L}} e = 0.$$

Together with the fact that the (z,e) jumps to $(-z, b(z,e))$ at the reflection jump, the expression (25) follows. By integration by parts formula, we have

$$\int_{-\infty}^{\infty} \{\partial_z f(z,e) + z_+(f(-z,e) - f(z,e))\}\,\phi(z)\mathrm{d}z = \int_{-\infty}^{\infty} \{\partial_z f(z,e) - z f(z,e)\}\,\phi(z)\mathrm{d}z = 0$$

for any absolutely continuous function $f(\cdot, e)$ such that $\int |\partial_z f(z,e)|\phi(z)\mathrm{d}z < \infty$ for each $e \in \mathbb{S}$. On the other hand,

$$\int_{\mathbb{S}} f(-z, b(z,e))\mathrm{d}e = \int_{\mathbb{S}} f(-z,e)\mathrm{d}e.$$

Therefore, the expectation of Hf with respect to $\mathcal{N}(0,1) \otimes \mathcal{U}(\mathbb{S})$ is always 0. Hence the claim follows from Theorem 21 and Corollary 22 of Durmus, Guillin and Monmarché (2021). □

Proposition 4.8. *Let P_t be the Markov kernel of $(z(t), e(t))$. Then for $(z,e), (z^*, e^*) \in \mathbb{R} \times \mathbb{S}$, there exists $T > 0$ such that*

$$P_t((z,e), \cdot) \not\perp P_t((z^*, e^*), \cdot) \tag{28}$$

for all $t \geq T$.

The proof of the proposition is provided in the supplementary material Bierkens, Kamatani and Roberts (2025).

Corollary 4.9. *Let* $y_{\mathrm{L}}(0), v_{\mathrm{L}}(0) \in \mathbb{R}^{\mathrm{L}}$ *be fixed constants such that the two vectors are linearly independent. Then*

$$\mathbb{E}\left[\frac{(v_{\mathrm{L}}(t)^\top y_{\mathrm{L}}(t))_+^2}{|y_{\mathrm{L}}(t)|^2}\right] = \mathbb{E}\left[\frac{\alpha\, z(t)_+^2}{\beta/\alpha + z(t)^2}\right] \longrightarrow_{t\to\infty} c(\alpha,\beta)$$

where, for $r = (\beta/\alpha)^{1/2}$,

$$c(\alpha,\beta) = \alpha - \sqrt{2\pi}\,\alpha r e^{r^2/2}\Phi(-r) = \alpha\left[1 - r\frac{\Phi(-r)}{\phi(r)}\right], \tag{29}$$

where Φ *is the cumulative distribution function of the normal distribution.*

Proof. The equation (24) yields the first equation. Since the invariant measure of $z(t)$ is the standard normal distribution, by the law of large numbers and by Proposition 4.7, the expectation is

$$\int_{\mathbb{R}} \frac{\alpha x^2}{r^2 + x^2}\phi(x)\mathrm{d}x = \alpha - \alpha r^2 \int_{\mathbb{R}} \frac{1}{r^2 + x^2}\phi(x)\mathrm{d}x.$$

Since the characteristic function of the probability density function $\phi(x)$ of the standard normal distribution is $\sqrt{2\pi}\phi(x)$, we have

$$\begin{aligned}
\int_{\mathbb{R}} \frac{1}{r^2 + x^2}\phi(x)\mathrm{d}x &= (2\pi)^{-1/2}\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{1}{r^2 + x^2}e^{\mathrm{i}xu}\phi(u)\mathrm{d}x\mathrm{d}u \\
&= (2\pi)^{-1/2}r^{-1}\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{1}{1 + x^2}e^{\mathrm{i}xur}\phi(u)\mathrm{d}x\mathrm{d}u \\
&= (\pi/2)^{1/2}r^{-1}\int_{\mathbb{R}} e^{-|u|r}\phi(u)\mathrm{d}u
\end{aligned}$$

where we used the fact that the characteristic function of the Cauchy distribution $\pi^{-1}(1 + x^2)^{-1}\mathrm{d}x$ is $\exp(-|x|)$. This expectation is

$$2\,(\pi/2)^{1/2}r^{-1}\int_0^\infty e^{-ur}\phi(u)\mathrm{d}u = 2\,(\pi/2)^{1/2}r^{-1}\,\Phi(-r)e^{r^2/2}.$$

This proves the claim. □

5. Discussion

In this study, we aimed to determine the scaling limit of the Zig-Zag sampler and the Bouncy Particle Sampler for a special class of anisotropic target distributions. Here, we would like to compare PDMPs with MCMC algorithms, such as the random walk Metropolis algorithm. As described in Beskos et al. (2018), the computational effort for the random walk Metropolis algorithm is $O(\epsilon^{-1})$ when the dimension of the smaller component l equals 1, based on the Markov jump process limit. However, in general, the computational complexity is $O(\epsilon^{-2})$ as discussed in Beskos et al. (2018). Therefore, even after selecting the optimal scaling for the random walk Metropolis algorithm, the BPS has a better complexity of $O(\epsilon^{-1})$, and the ZZS has the same complexity of $O(\epsilon^{-2})$ when $l \neq 1$, in terms of orders of magnitude of ϵ^{-1}. We estimate that the cost per unit time for PDMPs is on the order of ϵ^{-1} by Theorems

3.3 and 3.5. Thus, our theoretical results demonstrate that the Bouncy Particle Sampler exhibits superior convergence rate compared to random walk Metropolis chains for anisotropic target distributions, supporting the use of piecewise deterministic Markov processes for anisotropic target distributions.

As we discussed in Subsection 2.3, it is important to emphasise that our complexity analysis is strongly related to the implementational cost of either of these algorithms, and this cost can vary substantially between different applications. Also target distributions with significant conditional independence structure can be less expensive to explore using Zig-Zag than the Bouncy Particle Sampler, see for example Bierkens, Fearnhead and Roberts (2019), Chevallier, Fearnhead and Sutton (2023), and this could offset the theoretical advantages that BPS has in these contexts.

Our findings also highlight the importance of pre-conditioning of the Zig-Zag sampler, which is a coordinate-dependent method. One approach to improving the performance of the Zig-Zag sampler is to transpose the state space so that the components are roughly uncorrelated. It is helpful to set the scale and direction such that the number of jumps in each direction are roughly equal. However, the effective direction and scale of the state space can be position-dependent. In these cases, the Bouncy Particle Sampler may have an advantage due to its relative simplicity.

The asymptotic behaviour of these processes is sensitive to the target distribution. In particular, we find that the Bouncy Particle Sampler can become degenerate and trapped in a low-dimensional subspace in the limit if there are no refreshment jumps. The specific low-dimensional subspace is determined by the structure of the target distribution. Also, we only studied the simplest scenario for the covariance matrix. More general scenario such as more than two scales were out of scope for this paper.

Multi-scale analysis provides an efficient framework for the scaling limit analysis of both Markov chains and Markov processes. By using this approach, we are able to effectively separate the contributions of the two scales of the Markov processes using the solution of the Poisson equation. This also simplifies the structure of the proof. We believe that this framework could be useful for the scaling limit analysis of many other Markov chains and processes.

Finally, recent work by Andrieu et al. (2021) provides L^2-exponential convergence for general target distributions, identifying the exponential rate even in high-dimensional scenarios. Their results can also be applied to anisotropic target distributions and provide an upper bound for the convergence rates. However, it is worth noting that it is currently unclear how to achieve the convergence rate described in this paper as an upper bound for their study.

Funding

Supplementary Material

Supplement to "Scaling of piecewise deterministic Monte Carlo for anisotropic targets" (DOI: 10.3150/24-BEJ1807SUPP; .pdf). The supplementary material provides proofs for the results presented in Sections 3.1, 3.2, and for Proposition 4.8. Additionally, it includes the proof of the results for the two-dimensional Zig-Zag sampler analysis described in Section 2.1.

References

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Stat. Comput.* **18** 343–373. MR2461882 https://doi.org/10.1007/s11222-008-9110-y

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

Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J.-M. and Stuart, A. (2013). Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli* **19** 1501–1534. MR3129023 https://doi.org/10.3150/12-BEJ414

Beskos, A., Roberts, G., Thiery, A. and Pillai, N. (2018). Asymptotic analysis of the random walk Metropolis algorithm on ridged densities. *Ann. Appl. Probab.* **28** 2966–3001. MR3847978 https://doi.org/10.1214/18-AAP1380

Bierkens, J. and Duncan, A. (2017). Limit theorems for the zig-zag process. *Adv. in Appl. Probab.* **49** 791–825. MR3694318 https://doi.org/10.1017/apr.2017.22

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

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

Bierkens, J., Kamatani, K. and Roberts, G.O. (2025). Supplement to "Scaling of piecewise deterministic Monte Carlo for anisotropic targets." https://doi.org/10.3150/24-BEJ1807SUPP

Bierkens, J., Roberts, G.O. and Zitt, P.-A. (2019). Ergodicity of the zigzag process. *Ann. Appl. Probab.* **29** 2266–2301. MR3983339 https://doi.org/10.1214/18-AAP1453

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. New York: Wiley. A Wiley-Interscience Publication. MR1700749 https://doi.org/10.1002/9780470316962

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

Chevallier, A., Fearnhead, P. and Sutton, M. (2023). Reversible jump PDMP samplers for variable selection. *J. Amer. Statist. Assoc.* **118** 2915–2927. MR4681630 https://doi.org/10.1080/01621459.2022.2099402

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

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

Durmus, A., Guillin, A. and Monmarché, P. (2021). Piecewise deterministic Markov processes and their invariant measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 1442–1475. MR4291455 https://doi.org/10.1214/20-aihp1125

Ethier, S.N. and Kurtz, T.G. (1986). *Markov Processes: Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR0838085 https://doi.org/10.1002/9780470316658

Glynn, P.W. and Meyn, S.P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931. MR1404536 https://doi.org/10.1214/aop/1039639370

Graham, M.M., Thiery, A.H. and Beskos, A. (2022). Manifold Markov chain Monte Carlo methods for Bayesian inference in diffusion models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1229–1256. MR4494159 https://doi.org/10.1111/rssb.12497

Holderrieth, P. (2021). Cores for piecewise-deterministic Markov processes used in Markov chain Monte Carlo. *Electron. Commun. Probab.* **26** Paper No. 57, 12. MR4346861 https://doi.org/10.1214/21-ecp430

Kulik, A. and Scheutzow, M. (2015). A coupling approach to Doob's theorem. *Atti Accad. Naz. Lincei, Rend. Lincei, Mat. Appl.* **26** 83–92. MR3345324 https://doi.org/10.4171/RLM/694

Pavliotis, G.A. and Stuart, A.M. (2008). *Multiscale Methods: Averaging and Homogenization. Texts in Applied Mathematics* **53**. New York: Springer. MR2382139

Peters, E.A.J.F. and de With, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E* **85** 026703. https://doi.org/10.1103/PhysRevE.85.026703

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

Roberts, G.O. and Rosenthal, J.S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 255–268. MR1625691 https://doi.org/10.1111/1467-9868.00123

Roberts, G.O. and Rosenthal, J.S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statist. Sci.* **16** 351–367. MR1888450 https://doi.org/10.1214/ss/1015346320

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

Roberts, G.O. and Sahu, S.K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B* **59** 291–317. MR1440584 https://doi.org/10.1111/1467-9868.00070

Sherlock, C. and Thiery, A.H. (2022). A discrete bouncy particle sampler. *Biometrika* **109** 335–349. MR4430961 https://doi.org/10.1093/biomet/asab013

Sherlock, C., Thiery, A.H., Roberts, G.O. and Rosenthal, J.S. (2015). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.* **43** 238–275. MR3285606 https://doi.org/10.1214/14-AOS1278

Zanella, G. and Roberts, G. (2021). Multilevel linear models, Gibbs samplers and multigrid decompositions (with discussion). *Bayesian Anal.* **16** 1308–1390. MR4381136 https://doi.org/10.1214/20-BA1242