



Polynomial Convergence Rates of Piecewise Deterministic Markov Processes

Gareth O. Roberts¹ · Jeffrey S. Rosenthal²

Received: 24 August 2022 / Revised: 27 December 2022 / Accepted: 3 January 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

We consider piecewise-deterministic Markov processes such as the Bouncy Particle sampler, on target densities with polynomial tails. Using direct drift condition methods, we provide bounds on the polynomial order of the processes' convergence rate to stationary, on both one-dimensional and high-dimensional state spaces, in both total variation distance and f -norm.

Keywords PDMP · MCMC · Bouncy Particle sampler · Zig Zag algorithm · Convergence rate · Polynomial convergence · Infinitesimal generator · Drift condition

Mathematics Subject Classification (2020) Primary 60J25 · Secondary 60J22

1 Introduction

Markov chain Monte Carlo (MCMC) algorithms have become an indispensable part of statistical computation; see e.g. Brooks et al. (2011) and the many references therein. Piecewise-deterministic Markov processes (PDMP), such as the Bouncy Particle sampler (Bouchard-Côté et al. 2018) and the Zig-Zag algorithm (Bierkens et al. 2019), have emerged as a non-reversible alternative to traditional Metropolis-based MCMC. They are of great theoretical interest and also some practical relevance; see e.g. Bierkens (2021) and the references therein. An important question about PDMP is their rate of convergence, i.e. how quickly they converge to their target stationary distribution. For sufficiently lightly-tailed targets, geometric ergodicity has been established under certain conditions (Deligiannidis et al. 2019). However, if the target distribution has tails which are heavier than exponential, then geometric ergodicity does not apply.

✉ Jeffrey S. Rosenthal
jeff@math.toronto.edu

Gareth O. Roberts
Gareth.O.Roberts@warwick.ac.uk

¹ Department of Statistics, University of Warwick, Coventry CV4 7AL, UK

² Department of Statistics, University of Toronto, Toronto, Ontario M5S 3G3, Canada

In this paper, we instead focus on *polynomial* convergence rates of certain PDMP. That topic was previously approached using the concept of hypocoercivity in Andrieu et al. (2021a, b), but here we proceed using direct drift condition methods. We specifically consider the Bouncy Particle sampler (Bouchard-Côté et al. 2018), for a given target density π in \mathbf{R}^d . This PDMP has, at each time, a location $x \in \mathbf{R}^d$ and a velocity $v \in \mathbf{R}^d$ with $|v| = 1$. It proceeds primarily by deterministically moving x through \mathbf{R}^d at the fixed velocity v . It also reflects v along π 's contour lines at hazard rate $\lambda(x, v) = [-v(\log \pi)'(x)]^+$. In addition, it refreshes at some specified hazard rate (which could depend on the current position x), at which point it replaces the velocity v by an independent draw from the uniform distribution Ψ on the unit sphere in \mathbf{R}^d . This process is known (Bouchard-Côté et al. 2018) to be irreducible with stationary density π , and to converge to π exponentially quickly for sufficiently light-tailed target densities π .

This paper examines the polynomial convergence rate of this PDMP to target densities π which are heavy-tailed. We first consider one-dimensional heavy-tailed targets (for which polynomial convergence rates of the Zig-Zag Process was also considered in Vasdekis and Roberts 2021). For targets with tails comparable to a t -distribution with r degrees of freedom, we derive sharp bounds on polynomial convergence (Theorem 4). In particular, we prove that the polynomial convergence order in total variation distance is precisely r , in the sense that $\lim_{t \rightarrow \infty} t^a \|P^t(x, \cdot) - \pi(\cdot)\|_{TV}$ equals 0 for $a < r$ and infinity for $a > r$. We also prove convergence in the $V^{(1-\alpha)p}$ -norm (see Section 3) at polynomial order approaching $(1-p)r$, for any $p \in [0, 1)$. We then consider high-dimensional PDMP, and compute their infinitesimal generator applied to an appropriate drift function (Theorem 5). We specialise this generator computation to target densities with polynomial tails proportional to $(1 + |x|^2)^{-(r+d)/2}$ (Corollary 7), and use this to derive specific bounds on their polynomial convergence rate (Theorem 8) in both total variation distance and f -norm. Our theorem shows that for $r > (2\pi - 1)d$, the process converges in total variation distance at polynomial order approaching $(r + d)\sqrt{2\pi/d} - 1$.

This paper is organised as follows. In Section 2, we present some computer simulations to illustrate the convergence of PDMP to stationarity. In Section 3, we review general polynomial convergence rate bounds for continuous-time processes as in Fort and Roberts (2005), and present some corollaries adapting those results to our needs. In Section 4, we consider one-dimensional PDMP, and prove an exact characterisation (Theorem 4) of the polynomial convergence order in that case. In Section 5, we prove a general result (Theorem 5) which derives the infinitesimal generator of PDMP acting on certain choices of drift function, which we then apply to target densities with polynomial tails (Corollary 7). In Section 6, we apply these results to derive specific polynomial rate bounds for high-dimensional PDMP (Theorem 8). Finally, in Section 7, we present an auxiliary computation about expected values with respect to the refresh distribution Ψ , which is used in the proof of Theorem 8.

2 Computer Simulations of the PDMP

We begin by performing some computer simulations. Suppose the state space \mathcal{X} is the one-dimensional real line \mathbf{R} , with C^1 target density $\pi(x)$. In this case, the Piecewise Deterministic Markov Process (PDMP) enlarges the state space to $\mathcal{X} \times \{-1, 1\}$, and expands π to $\pi(x, v) = \frac{1}{2}\pi(x)$ for $v \in \{-1, 1\}$. It then proceeds by moving with fixed constant velocity v , except reflecting from v to $-v$ with hazard rate $\lambda(x, v) = [-v(\log \pi)'(x)]^+$.

(We omit refreshes, i.e. take the refresh rate to be zero, since refreshes are not required in one dimension.)

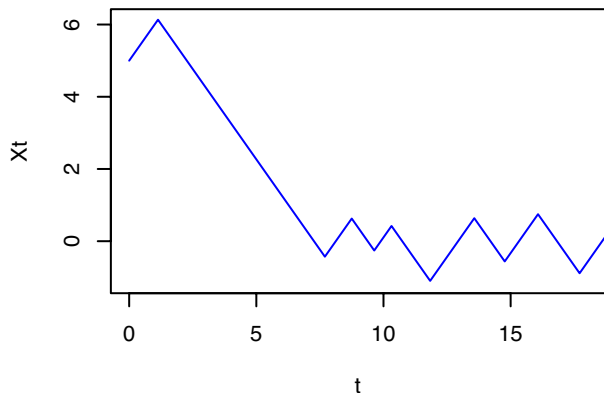
Simulating this process requires that we identify the reflection times, which arise in continuous time according to the hazard rate $\lambda(x, v)$. This could be approximated by advancing time in small discrete increments, but the errors in such approximations are difficult to control. Instead, we proceed as follows. Suppose the process is currently at some state (X_t, V_t) at time $t \geq 0$, and we wish to simulate the next T units of time. We first find some value M for which we must have $\lambda(X_s, V_s) \leq M$ for all $t \leq s \leq t + T$.

Then, we simulate a Poisson process with constant rate M for the next T time units. We then use ‘‘Poisson thinning’’ to proceed through those times points in order, with acceptance probability $\lambda(x, v)/M$, until the first one is accepted and hence the next reflection time is identified. At that point, we discard the remaining Poisson time points, and continue the simulation anew from the identified reflection time. In this way, the reflection times are simulated accurately, without any discretisation error. (The R script that we used for our simulations is available for inspection at: probability.ca/Rpoly.)

We first simulate this process where $\pi(x) = (1 + x^2)^{-3}$ (so, π is essentially a t -distribution with parameter $r = 5$). In this case, $\lambda(x, v) := [-v(\log \pi)'(x)]^+ = \frac{(1+r)(xv)^+}{1+x^2}$, which is maximised at $M := \lambda(1, 1) = (1 + r)/2$. (So, in this case $\lambda(x, v)$ has a constant upper bound M , but in general M might depend on X_t and V_t and T .) A typical run of this process is shown in Fig. 1, starting with $X_0 = 5$ and $V_0 = +1$. We see that the process moves at constant velocity ± 1 , with reflections at appropriate random times to preserve stationarity.

To illustrate the convergence of this process X_t to its stationary distribution π , we consider (inspired by total variation distance, see next section) the expected values of functionals $g : \mathcal{X} \rightarrow \mathbf{R}$, specifically the difference between the expected value $\mathbf{E}[g(X_t)]$ at time t of our process, compared to the stationary expected value $\pi(g) := \mathbf{E}_\pi[g(X)]$. By repeating the simulation a large number of times, we obtain a mean value and 95% confidence interval for $\mathbf{E}[g(X_t)]$ at various times t , for three different functionals $g_1(x) = x$, $g_2(x) = \mathbf{1}_{x>0}$, and $g_3(x) = x^2$. In each case, we compare $\mathbf{E}[g(X_t)]$ to the corresponding stationary expectation $\pi(g)$ (equal to 0 and 1/2 and 1/3, respectively), at various times t . The results are shown in Fig. 2. The mean value of each of the three functionals when running the process (blue) converges quickly to its stationary value (red). This provides confirmation that our PDMP process is indeed converging to the correct distribution. But how quickly?

Fig. 1 A typical piecewise-deterministic Markov process (PDMP) run for the Student’s t -distribution π , starting with $X_0 = 5$ and $V_0 = +1$, with reflections at random times to make π be stationary



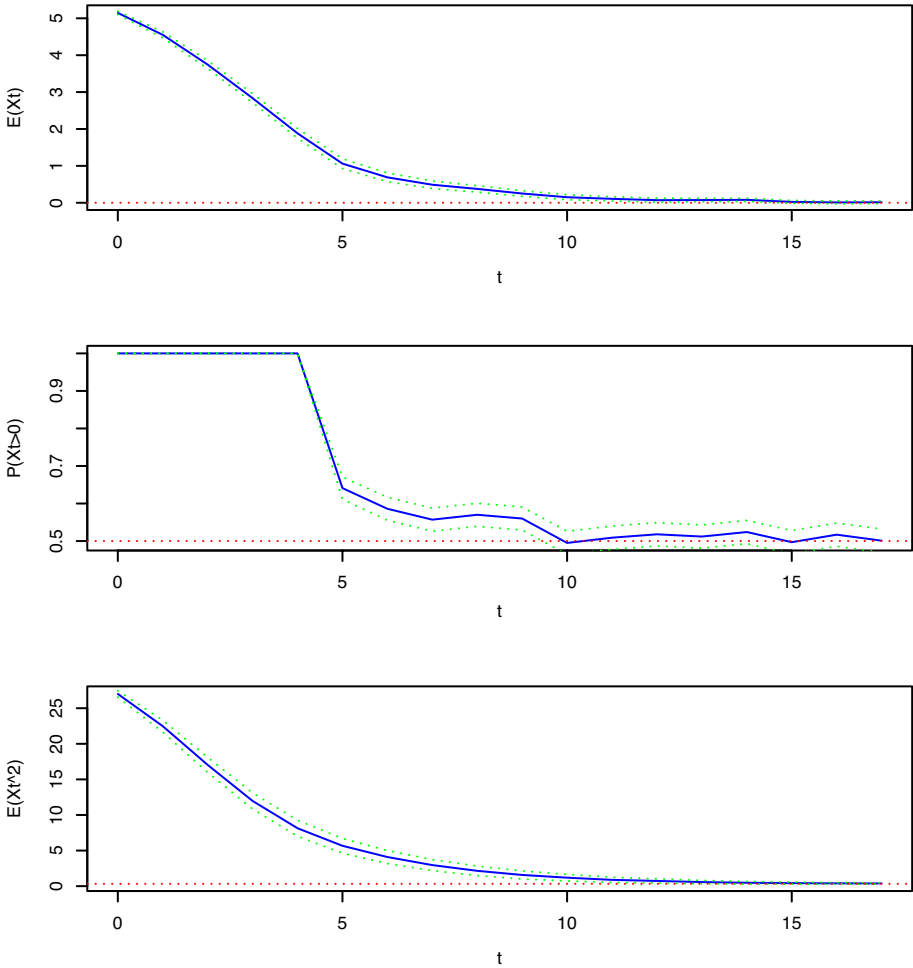
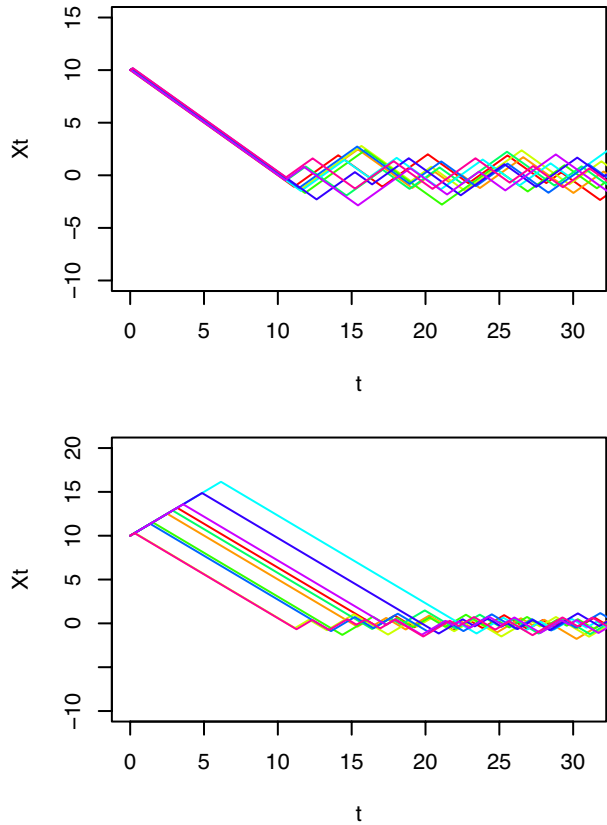


Fig. 2 The mean value (blue) and 95% confidence interval (green dotted) for $E[g(X_t)]$ at various times t , for the three different functionals $g_1(x) = x$ (top), $g_2(x) = \mathbf{1}_{x>0}$ (middle), and $g_3(x) = x^2$ (bottom), compared to the corresponding stationary expectation (red), when running a PDMP for the Student’s t -distribution

In fact, the convergence rate of this process is heavily affected by the tail behaviour of the target distribution $\pi(x)$. To illustrate this, consider a second example where the target $\bar{\pi}(x) = e^{-x^2/2}$ corresponds to a standard Gaussian distribution. Then the tails of $\pi(x)$ are much heavier than those of $\bar{\pi}(x)$. Its hazard rate is equal to $\bar{\lambda}(x, \nu) := [-\nu (\log \bar{\pi})'(x)]^+ = (x\nu)^+$ which grows much faster than $\lambda(x, \nu)$ above. Here $M := \sup_{t \leq s \leq t+T} \lambda(X_s, V_s) = |X_t| + T$, which depends on X_t . In particular, the process for $\bar{\pi}$ will return to the origin much more quickly and consistently than for π , leading to much faster convergence. This is illustrated in Fig. 3, which shows ten runs of the process for $\bar{\pi}$ (top) and for π (bottom) when started with $X_0 = 10$ and $V_0 = +1$, with the π processes much more variable in their return times. The difference also arises in Fig. 4, which shows ten runs for each target, but this time started with $X_0 = 1000$ and

Fig. 3 Ten PDMP runs for the Gaussian target $\bar{\pi}$ (top) and for the Student's t -distribution target π (bottom), started with $X_0 = 10$ and $V_0 = +1$; the π (bottom) processes are much more variable



$V_0 = +1$, i.e. much farther out in the tails, with the π processes even more variable in their return times.

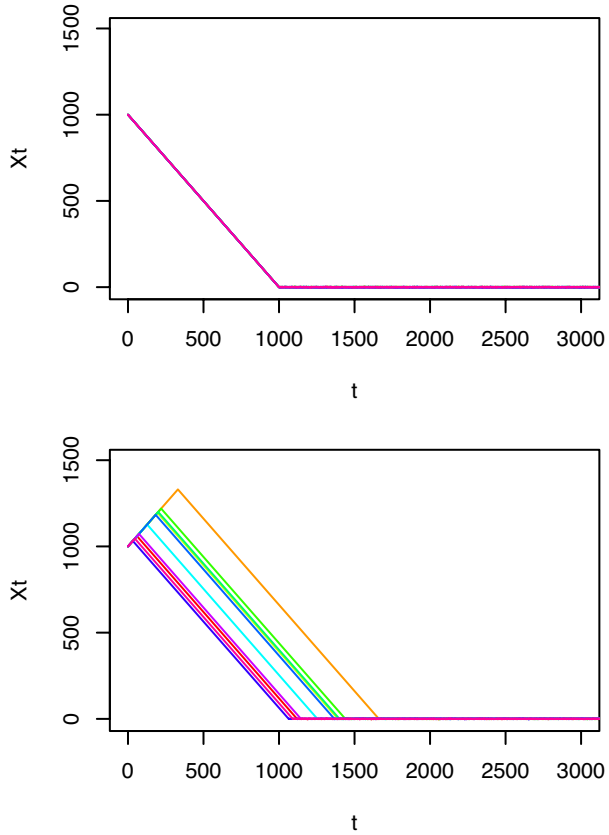
Due to the heavy polynomial tails of the Student's t -distribution $\pi(x)$, the convergence to π cannot be exponentially quick, i.e. “geometrically” ergodic. But it might still be polynomially ergodic. To investigate that question, we next to turn our attention to the theory of polynomial convergence rates of Markov processes.

3 Polynomial Convergence Rates of Markov Processes

Quantitative convergence rates of discrete-time geometrically ergodic Markov chains have a long history, see e.g. Rosenthal (2002) and the many references therein. More recently, focus has turned to polynomial ergodicity, e.g. Fort and Moulines (2000) and Jarner and Roberts (2002). Most of these results are in discrete time, but Fort and Roberts (2005) yields the following continuous-time polynomial convergence bound. To state it, recall that if μ and ν are two probability distributions on \mathcal{X} , and $f : \mathcal{X} \rightarrow (0, \infty)$, then the f -norm distance between μ and ν is defined as

$$\|\mu(\cdot) - \nu(\cdot)\|_f := \sup_{\substack{g : \mathcal{X} \rightarrow \mathbf{R} \\ |g| \leq f}} |\mathbf{E}_\mu(g) - \mathbf{E}_\nu(g)|,$$

Fig. 4 Ten PDMP runs for the Gaussian target $\bar{\pi}$ (top) and for the Student's t -distribution target π (bottom), started with $X_0 = 1000$ and $V_0 = +1$; the π (bottom) processes are again more variable



and the total variation distance between μ and ν is defined as

$$\|\mu(\cdot) - \nu(\cdot)\|_{TV} := \sup_{g: \mathcal{X} \rightarrow \mathbf{R}, |g| \leq 1} |\mathbf{E}_\mu(g) - \mathbf{E}_\nu(g)|.$$

Let $P^t(x, A) = \mathbf{P}[X_t \in A \mid X_0 = x]$ be the time- t transition probabilities. Also, recall that a continuous-time Markov process $\{X_t\}$ has an *infinitesimal generator* \mathcal{A} which acts on appropriate functions $f : \mathcal{X} \rightarrow \mathbf{R}$ by

$$\mathcal{A}f(x) := \lim_{\delta \searrow 0} \frac{\mathbf{E}[f(X_\delta) \mid X_0 = x] - f(x)}{\delta}$$

(for background about generators see e.g. Ethier and Kurtz 1986). Then we have:

Proposition 1 *Suppose a continuous-time Markov process on state space $\mathcal{X} \subseteq \mathbf{R}^d$ has stationary distribution π , and infinitesimal generator \mathcal{A} , and there is $\alpha \in (0, 1)$ and $c > 0$ and $b_0 < \infty$ and a closed petite set $C \subseteq \mathcal{X}$ and a drift function $V \geq 1$ with $\sup_{x \in C} V(x) < \infty$ such that $\mathcal{A}V(x) \leq -c(V(x))^{1-\alpha} + b_0 \mathbf{1}_C(x)$ for all $x \in \mathcal{X}$. Then for any $p \in [0, 1)$ and $x \in \mathcal{X}$,*

$$\lim_{t \rightarrow \infty} t^{(1-p)(1-\alpha)/\alpha} \|P^t(x, \cdot) - \pi(\cdot)\|_{V^{(1-\alpha)p}} = 0,$$

i.e. the process converges to stationary in the $V^{(1-\alpha)p}$ -norm at polynomial order $(1-p)(1-\alpha)/\alpha$. In particular, setting $p = 0$,

$$\lim_{t \rightarrow \infty} t^{(1-\alpha)/\alpha} \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} = 0,$$

i.e. $\|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \leq O(t^{-(1-\alpha)/\alpha})$, i.e. the process converges to stationarity in total variation distance at polynomial order $(1-\alpha)/\alpha$.

Proof This result follows from Corollary 6 of Fort and Roberts (2005) upon setting their $\eta = 1$ and $c_\eta = c$ and $b = 0$, and using that $t \leq 1 + t$. (Note that the “ b ” in their convergence equation is different from the “ b ” in their drift equation (8), which we here refer to as “ b_0 ”.) ■

To continue, recall that a function $V : \mathbf{R}^d \rightarrow \mathbf{R}$ is *norm-like* if $\lim_{|x| \rightarrow \infty} V(x) = \infty$. Also, a subset $C \subseteq \mathcal{X}$ is *small* if there are $t_0 > 0$ and a non-zero σ -finite measure ν on \mathcal{X} such that $P^{t_0}(x, A) \geq \nu(A)$ for all measurable $A \subseteq \mathcal{X}$, or *petite* if that condition is replaced by $\int_0^\infty P^t(x, A) dt \geq \nu(A)$. Call a process *compact-small* if all compact sets are small; this holds for many processes (see e.g. Meyn and Tweedie 1993), and in particular it holds for Bouncy Particle samplers with refresh rates which are bounded below on sets which are compact in the extended state space (Deligiannidis et al. 2019, Lemma 2), or which are in just one dimension (cf. Bierkens et al. 2019), so it holds for all of our applications here. In terms of these various definitions, we have:

Corollary 2 *Suppose a continuous-time compact-small Markov process on state space $\mathcal{X} \subseteq \mathbf{R}^d$ has stationary distribution π , and infinitesimal generator \mathcal{A} , and there is $\alpha \in (0, 1)$ and $c, c_0 > 0$ and $\Delta < \infty$ and a continuous norm-like drift function $V \geq c_0 > 0$ such that $\mathcal{A}V$ is bounded on compact sets and $\mathcal{A}V(x) \leq -c(V(x))^{1-\alpha}$ for all $x \in \mathcal{X}$ with $V(x) \geq \Delta$. Then, again, for any $p \in [0, 1)$ and $x \in \mathcal{X}$,*

$$\lim_{t \rightarrow \infty} t^{(1-p)(1-\alpha)/\alpha} \|P^t(x, \cdot) - \pi(\cdot)\|_{V^{(1-\alpha)p}} = 0.$$

Proof First of all, by replacing V by V/c_0 and c by c/c_0^α if necessary, we can assume that $c_0 = 1$. Then, let $C = \{x \in \mathcal{X} : V(x) \leq \Delta\}$. This C is closed by continuity of V , and is bounded since V is norm-like, so C is compact. Hence, by the compact-small property, C is small, and hence also petite. Then $b_0 := \sup_{x \in C} \mathcal{A}V(x) < \infty$ since $\mathcal{A}V$ is bounded on compact sets. This result now follows from Proposition 1, by noting that if $\mathcal{A}V(x) \leq -c(V(x))^{1-\alpha}$ when $V(x) \geq \Delta$ then $\mathcal{A}V(x) \leq -c(V(x))^{1-\alpha} + b_0 \mathbf{1}_C(x)$ for all $x \in \mathcal{X}$. ■

Corollary 3 *Suppose a continuous-time compact-small Markov process on state space $\mathcal{X} \subseteq \mathbf{R}^d$ has stationary distribution π , and infinitesimal generator \mathcal{A} , and there is $\beta > 1$ and $c_0, c_1 > 0$ and $\delta > 0$ and a drift function $V(x) \geq \max(c_0, c_1 |x|^\beta)$ such that*

$$\mathcal{A}V(x) \leq -\delta |x|^{\beta-1} [1 + o(|x|)], \quad |x| \rightarrow \infty.$$

Then for any $p \in [0, 1)$ and $x \in \mathcal{X}$,

$$\lim_{t \rightarrow \infty} t^{(1-p)(\beta-1)} \|P^t(x, \cdot) - \pi(\cdot)\|_{V^{(1-\alpha)p}} = 0,$$

and in particular

$$\lim_{t \rightarrow \infty} t^{\beta-1} \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} = 0.$$

Proof Since $V(x) \geq c_1 |x|^\beta$, it follows that $|x| \leq [V(x)/c_1]^{1/\beta}$, so for all large $|x|$,

$$\mathcal{AV} \leq -\delta |x|^{\beta-1} [1 + o(|x|)] \leq -\frac{\delta}{2} |x|^{\beta-1} \leq -\frac{\delta}{2} \left([V(x)/c_1]^{1/\beta}\right)^{\beta-1} = -c V(x)^{1-(1/\beta)}$$

where $c = \frac{\delta}{2}(c_1)^{1-(1/\beta)}$. Hence, we can apply Corollary 2 with $\alpha = 1/\beta \in (0, 1)$. The result then follows since $(1 - \alpha)/\alpha = (1 - \frac{1}{\beta}) / (1/\beta) = \beta - 1$. ■

Remark Although we focus here on the polynomial order of the convergence rates, using the above general polynomial bound results, it is also possible to use a similar approach to obtain actual quantitative (computable) bounds on the distance to stationarity of PDMP, similar in spirit to Rosenthal (2002) and the references therein; by using the related results of Fort and Moulines (2003).

4 Convergence Rate in One Dimension

Suppose again that the state space \mathcal{X} is the one-dimensional real line \mathbf{R} , with C^1 target density $\pi(x)$. Again consider the algorithm which enlarges the state space to $\mathcal{X} \times \{-1, 1\}$, and expands π to $\pi(x, v) = \frac{1}{2}\pi(x)$ for $v \in \{-1, 1\}$, and moves with fixed constant velocity v , except reflecting from v to $-v$ with hazard rate $\lambda(x, v) = [-v(\log \pi)'(x)]^+$, and with zero refresh rate.

We know that if π has heavy tails, then this process cannot converge exponentially quickly. However, it might still converge polynomially quickly. Polynomial convergence rates for the related Zig-Zag process on one-dimensional heavy-tailed targets have been studied in Vasdekis and Roberts (2021). In this section, we present a result which gives precise polynomial convergence rates for the Bouncy Particle sampler, including a generalisation to f -norm convergence.

Consider now the specific example where $\pi(x) = (1 + x^2)^{-(1+r)/2}$ for some fixed constant $r \geq 1$, at least when $|x| \geq \Delta \geq 1$ (so, π is essentially a Student's t -distribution). Then we have:

Theorem 4 *The above one-dimensional PDMP converges to stationarity in total variation distance at polynomial rate equal to r . More precisely, for any $x \in \mathcal{X}$,*

$$\lim_{t \rightarrow \infty} t^a \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} = \begin{cases} 0, & a < r \\ \infty, & a > r \end{cases}$$

Furthermore, for any $p \in [0, 1)$, for appropriate drift function V as defined in the proof, the process converges to stationarity in the $V^{(1-\alpha)p}$ -norm at polynomial order approaching $(1-p)r$, i.e. for any $a < r$,

$$\lim_{t \rightarrow \infty} t^{(1-p)a} \|P^t(x, \cdot) - \pi(\cdot)\|_{V^{(1-\alpha)p}} = 0.$$

Proof First, for the assumed form of π , we have for $|x| \geq \Delta$ that

$$\lambda(x, v) := [-v(\log \pi)'(x)]^+ = \frac{(1+r)(xv)^+}{1+x^2} = \begin{cases} \frac{(1+r)xv}{1+x^2}, & xv > 0 \\ 0, & xv < 0 \end{cases}$$

It follows that $\lambda(x, v) \geq (1+r)/(1+x)$ for $v = +1$ and $x \geq \Delta \geq 1$. Next, for some $\beta > 1$ and $K > 1$ to be determined later, let

$$V(x, v) = \begin{cases} K(1+|x|)^\beta, & xv > 0 \\ (1+|x|)^\beta, & xv < 0 \end{cases}$$

Next, note that this process has infinitesimal generator \mathcal{A} which acts (cf. Ethier and Kurtz 1986; Deligiannidis et al. 2019) on appropriate C^1 functions $f : \mathcal{X} \times \{-1, 1\} \rightarrow \mathbf{R}$ by:

$$\mathcal{A}f(x, v) = v \frac{\partial f}{\partial x} + \lambda(x, v) [f(x, -v) - f(x, v)]. \tag{1}$$

Hence, for $v = +1$ and $x \geq \Delta$,

$$\begin{aligned} \mathcal{A}V(x, v) &= \beta K(1+x)^{\beta-1} - \lambda(x, v)(K-1)(1+x)^\beta \\ &\leq -(1+x)^{\beta-1} [(K-1)(r+1) - \beta K]. \end{aligned}$$

Suppose it holds that

$$1 < \beta < 1+r, \quad \text{and} \quad K > (1+r)/(1+r-\beta). \tag{2}$$

Then $(K-1)(r+1) - \beta K > 0$, so $\mathcal{A}V(x, v) < 0$ for $v = +1$ and $x \geq \Delta$. Meanwhile, for $x \geq \Delta$ and $v = -1$ we have $\lambda(x, v) = 0$, so we compute from (1) that

$$\mathcal{A}V(x, v) = -\beta(1+x)^{\beta-1}.$$

Combining these two calculations, it follows that if $K^* = \min[\beta, (K-1)(r+1) - \beta K]$, then assuming (2), we have for $x \geq \Delta$ and either $v = +1$ or $v = -1$ that

$$\mathcal{A}V(x, v) \leq -K^*(1+x)^{\beta-1} = -K^*(V(x))^{(\beta-1)/\beta} = -K^*(V(x))^{1-(1/\beta)}.$$

By symmetry, this condition also holds for $x \leq -\Delta$, i.e. it holds whenever $|x| \geq \Delta$. This shows that the assumptions of Corollary 2 hold with $\alpha = 1/\beta$, so $(1-\alpha)/\alpha = \beta - 1$. Hence, that corollary gives that

$$\lim_{t \rightarrow \infty} t^{(1-p)(\beta-1)} \|P^t(x, \cdot) - \pi(\cdot)\|_{V^{(1-\alpha)p}} = 0,$$

and in particular with $p = 0$,

$$\lim_{t \rightarrow \infty} t^{\beta-1} \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} = 0.$$

It remains to ensure that (2) holds. But (2) can be satisfied for any $\beta < 1+r$ by using a sufficiently large K . It follows that the polynomial order $\beta - 1$ can be made $\geq r - \epsilon$ for any $\epsilon > 0$, i.e. we can take $\beta - 1 = a$ for any $a < r$, which gives the claimed upper bounds.

Finally, for the lower bound, note that since the process never moves faster than speed 1, we must have $P^t((x, \pm 1), (t, \infty)) = 0$ for $x \leq 0$, and similarly that $P^t((x, \pm 1), (-\infty, -t)) = 0$ for $x \geq 0$. Hence, for any $x \in \mathbf{R}$, by symmetry,

$$\|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \geq \frac{1}{2} \pi((t, \infty)),$$

which to first order as $t \rightarrow \infty$ is

$$\int_t^\infty (1+x^2)^{-(1+r)/2} dx \approx \int_t^\infty (x^2)^{-(1+r)/2} dx = \int_t^\infty x^{-(1+r)} dx = t^{-r}/(1+r) = \Omega(t^{-r}).$$

This completes the proof. ■

5 Multi-Dimensional Generator Bounds

We now turn to PDMP on $\mathcal{X} = \mathbf{R}^d$. At each time, the process has position x and velocity v with $|v| = 1$. The process primarily moves at fixed constant velocity v . It also reflects along π 's contour lines at the hazard rate

$$\lambda(x, v) = \left(-(\nabla \log \pi) \cdot v \right)^+.$$

And it refreshes, by drawing a new v independently from the uniform distribution Ψ on the unit sphere in \mathbf{R}^d , with refresh rate which we take to be $s/|x|$ for some choice of $s > 0$ to be determined later, which does not depend on x (but might still depend on d). (This choice of $|x|^{-1}$ refresh rate decay helps avoid diffusive behaviour for large $|x|$, and makes the process self-similar in the sense that multiplying it by a constant preserves the trajectories just at a slower speed, and also balances the influence of refreshes with those of the continuous dynamics and reflections as we shall see, thus facilitating our calculations and analysis.)

To proceed, consider a drift function of the form

$$V(x, v) = W(C_{x,v}) 1 + |x|^\beta,$$

for some $\beta > 1$, where

$$C_{x,v} = x \cdot v / |x|$$

is the cosine of the angle between x and v , and $W(C) \geq 1$ is a function which will be chosen later. We assume that W has right-hand first derivatives (at least), denoted $W'(C)$. Let $E := \mathbf{E}_\Psi[W(C_{x,U})]$ be the expected value of $W(C_{x,U})$ where $U \sim \Psi$. Extending (1) to multiple dimensions, and including the refreshes at rate $s/|x|$, this process has infinitesimal generator \mathcal{A} which acts on appropriate C^1 functions $f : \mathcal{X} \times \{-1, 1\} \rightarrow \mathbf{R}$ by

$$\mathcal{A}f(x, v) = v \cdot \nabla_x f(x) + \lambda(x, v) [f(x, -v) - f(x, v)] + \frac{s(x)}{|x|} [E - f(x, v)]$$

(for background see e.g. Ethier and Kurtz 1986; Deligiannidis et al. 2019). Then we have:

Theorem 5 *The above PDMP has infinitesimal generator satisfying*

$$AV(x, v) = |x|^{\beta-1} B(x, v) [1 + O(|x|^{-\beta})], \quad \text{as } |x| \rightarrow \infty,$$

where

$$\begin{aligned} B(x, v) = & \left[W(C_{x,v}) \beta C_{x,v} + W'(C_{x,v})(1 - C_{x,v}^2) \right] \\ & + \left[\lambda(x, v) |x| [W(-C_{x,v}) - W(C_{x,v})] \right] \\ & + \left[s(E - W(C_{x,v})) \right]. \end{aligned}$$

The proof of Theorem 5 requires a simple gradient lemma:

Lemma 6 *For any $a \in \mathbf{R}$, $\nabla_x(|x|^a) = a|x|^{a-2}x$.*

Proof If $h(x) = |x|^2 = x^2$, then $\nabla_x h(x) = 2x$. Hence, by the chain rule,

$$\nabla_x(|x|^a) = \nabla_x(h(x)^{a/2}) = (a/2)[h(x)]^{(a/2)-1}(2x) = a|x|^{a-2}x. \quad \blacksquare$$

Proof of Theorem 5 We wish to compute the generator AV . Write this as $\mathcal{A}_1V + \mathcal{A}_2V + \mathcal{A}_3V$, where \mathcal{A}_1 is the contribution from the continuous dynamics, and \mathcal{A}_2 is the contribution from reflections, and \mathcal{A}_3 is the contribution from refreshing.

We begin with \mathcal{A}_1V (the continuous dynamics). We compute that

$$\mathcal{A}_1V(x, v) = \frac{\partial V(x, v)}{\partial t} = \sum_{i=1}^d \frac{\partial V(x, v)}{\partial x_i} \frac{\partial x_i}{\partial t} = \left(\nabla_x V(x, v) \right) \cdot v.$$

Now, since $\nabla_x x \cdot v = v$, Lemma 6 with $a = 1$ gives

$$\nabla_x C_{x,v} = \nabla_x \left(\frac{x \cdot v}{|x|} \right) = \frac{|x|v - x \cdot v|x|^{-1}x}{|x|^2} = \frac{v}{|x|} - \frac{x \cdot v x}{|x|^3} = \frac{v}{|x|} - \frac{C_{x,v}x}{|x|^2}.$$

Hence,

$$\left(\nabla_x C_{x,v} \right) \cdot v = \frac{v \cdot v}{|x|} - \frac{C_{x,v}x \cdot v}{|x|^2}.$$

And, Lemma 6 with $a = \beta$ gives

$$\nabla_x |x|^\beta = \beta|x|^{\beta-2}x.$$

Hence, by the product rule for derivatives,

$$\begin{aligned} \nabla_x V(x, v) = & W(C_{x,v}) \nabla_x (|x|^\beta) + (1 + |x|^\beta) W'(C_{x,v}) \nabla_x C_{x,v} \\ = & W(C_{x,v}) \beta|x|^{\beta-2}x + 1 + |x|^\beta W'(C_{x,v}) \left(\frac{v}{|x|} - \frac{C_{x,v}x}{|x|^2} \right). \end{aligned}$$

Then since $v \cdot v = 1$ and $x \cdot v = |x| C_{x,v}$,

$$\begin{aligned}
\mathcal{A}_1 V(x, v) &= \left(\nabla_x V(x, v) \right) \cdot v \\
&= W(C_{x,v}) \beta |x|^{\beta-2} x \cdot v + 1 + |x|^\beta W'(C_{x,v}) \left(\frac{1}{|x|} - \frac{C_{x,v} x \cdot v}{|x|^2} \right) \\
&= W(C_{x,v}) \beta |x|^{\beta-1} C_{x,v} + 1 + |x|^\beta W'(C_{x,v}) \left(\frac{1}{|x|} - \frac{C_{x,v}^2}{|x|} \right) \\
&= |x|^{\beta-1} \left[W(C_{x,v}) \beta C_{x,v} + W'(C_{x,v})(1 - C_{x,v}^2) \right] [1 + O(|x|^{-\beta})].
\end{aligned}$$

We next consider $\mathcal{A}_2 V$ (reflections). They occur at rate $\lambda(x, v)$, and change v to $-v$, hence $C_{x,v}$ to $-C_{x,v}$, so they change $V(x, v)$ from $W(C_{x,v}) 1 + |x|^\beta$ to $W(-C_{x,v}) 1 + |x|^\beta$, which is a change of $[W(-C_{x,v}) - W(C_{x,v})] 1 + |x|^\beta$. It follows that

$$\begin{aligned}
\mathcal{A}_2 V(x, v) &= \lambda(x, v) [W(-C_{x,v}) - W(C_{x,v})] 1 + |x|^\beta \\
&= \lambda(x, v) |x| [W(-C_{x,v}) - W(C_{x,v})] |x|^{\beta-1} [1 + O(|x|^{-\beta})].
\end{aligned}$$

Finally, we consider $\mathcal{A}_3 V(x, v)$ (refreshing). Refreshes occur at rate $s/|x|$, and replace the current velocity v with a fresh i.i.d. draw from the spherically-symmetric distribution Ψ on $\{z \in \mathbf{R}^d : |z| = 1\}$. This changes $V(x, v)$ from $W(C_{x,v})(1 + |x|^\beta)$ to $W(C_{x,U})(1 + |x|^\beta)$ where $U \sim \Psi$, which is a difference of $[W(C_{x,U}) - W(C_{x,v})](1 + |x|^\beta)$. Hence,

$$\mathcal{A}_3 V(x, v) = \frac{s}{|x|} [E - W(C_{x,v})] 1 + |x|^\beta = s [E - W(C_{x,v})] |x|^{\beta-1} [1 + O(|x|^{-\beta})]$$

where $E := \mathbf{E}_\Psi[W(C_{x,U})]$.

Putting this all together, the claim follows since $\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3$. ■

We now assume that π has polynomial tails as in a Student's t -distribution, i.e. that

$$\pi(x) \propto (1 + |x|^2)^{-(r+d)/2} \tag{3}$$

at least for $|x| \geq \Delta$. Theorem 5 then gives:

Corollary 7 *If π is given by 3, then the above PDMP has infinitesimal generator satisfying*

$$\mathcal{A}V(x, v) = |x|^{\beta-1} M(C_{x,v}) \left[1 + O(|x|^{-\beta}) \right], \quad \text{as } |x| \rightarrow \infty$$

where

$$\begin{aligned}
M(C) &= \left[W(C) \beta C + W'(C)(1 - C^2) \right] \\
&\quad + \left[(r + d) C^+ [W(-C) - W(C)] \right] \\
&\quad + \left[s(E - W(C)) \right].
\end{aligned}$$

Proof Here for $|x| > \Delta$ we have $\log \pi(x) = -((r + d)/2) \log(1 + |x|^2)$, so

$$\nabla \log \pi(x) = -\frac{r+d}{2} \frac{2x}{1+|x|^2} = -(r+d) \frac{x}{1+|x|^2},$$

and

$$\lambda(x, v) = (r + d) \frac{x \cdot v^+}{1 + |x|^2} = (r + d) C_{x,v}^+ |x|^{-1} [1 + O(|x|^{-2})].$$

The result then follows from Theorem 5. ■

6 Multi-Dimensional Convergence Rates

In this section, we prove the following bound on the polynomial convergence rate of high-dimensional PDMP:

Theorem 8 *For the above PDMP, for all sufficiently large $d \in \mathbf{N}$, with π as in (3) with $r > (2\pi - 1)d$, we have for any $a < (r + d)\sqrt{2\pi/d} - 1$ and any $p \in [0, 1)$ that*

$$\lim_{t \rightarrow \infty} t^{(1-p)a} \|P^t(x, \cdot) - \pi(\cdot)\|_{V^{(1-p)p}} = 0$$

for appropriate choice of refresh parameter s and drift function V as defined in the proof. In particular,

$$\lim_{t \rightarrow \infty} t^a \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} = 0.$$

That is, the process converges to stationarity in total variation distance at polynomial order approaching $(r + d)\sqrt{2\pi/d} - 1$. On the other hand, for any $a > r$,

$$\lim_{t \rightarrow \infty} t^a \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} = \infty.$$

Proof To obtain specific convergence rate bounds, we need to choose the function $W(C)$ in the drift function $V(x, v) = W(C_{x,v}) 1 + |x|^\beta$. After considering many possible choices, including some complicated ones, we eventually settled on the simple piecewise-linear choice

$$W(C) = 1 + m C \mathbf{1}_{C < 0} := \begin{cases} 1, & C \geq 0 \\ 1 + m C, & C < 0 \end{cases} \tag{4}$$

for some $m \in (0, 1)$. For this $W(C)$, let $M(C)$ be as in Corollary 7. Note that $V(x, v) := W(C_{x,v})(1 + |x|^\beta) \geq (1 + m(-1))(1) = 1 - m =: c_0 > 0$. Hence, by Corollary 3, it suffices to find values $s > 0$ and $m > 0$ (perhaps depending on d) such that $\sup_{C \in [-1, 1]} M(C) < 0$ for all sufficiently large d .

To proceed, let $k = r/d$, so $k > 2\pi - 1$. Then $(k + 1)/\sqrt{2\pi} > \sqrt{2\pi}$. Hence, we can find small enough $\epsilon > 0$ that $\xi := (1 - \epsilon)^3(k + 1)/\sqrt{2\pi} > \sqrt{2\pi}$. Then set $s = \xi\sqrt{d}$ (so $s > \sqrt{2\pi d}$), and $m = 1/2$ (so $4m^2 = 1 > 1 - \epsilon$), and $\beta = (1 - \epsilon)^2(k + 1)\sqrt{d/2\pi}$ so $\beta > (1 - \epsilon)^3(k + 1)\sqrt{d/2\pi} = \xi\sqrt{d}$ and also

$$\begin{aligned} \beta^2 &= (1 - \epsilon)^4(k + 1)^2 d / 2\pi \\ &= (1 - \epsilon)\xi(k + 1)d / \sqrt{2\pi} \\ &= (1 - \epsilon)4m^2\xi(k + 1)d / \sqrt{2\pi}. \end{aligned} \tag{5}$$

We now consider separately the cases $C < 0$ and $C \geq 0$.

For $C < 0$, it follows from 4 that $W(C) = 1 + mC$ and $W'(C) = m$ and $C^+ = 0$, so

$$\begin{aligned} M(C) &= \beta C + m\beta C^2 + m - mC^2 + s(E - 1 - mC) \\ &= C^2m(\beta - 1) + C(\beta - ms) + (m - s(1 - E)). \end{aligned}$$

Hence

$$M(0^-) := \lim_{C \nearrow 0} M(C) = m - s(1 - E).$$

Next we use Lemma 9 below, which states that $E = 1 - m\sqrt{1/2\pi d} [1 + O(\frac{1}{d})]$ so

$$1 - E = m\sqrt{1/2\pi d} [1 + O(\frac{1}{d})], \tag{6}$$

and hence

$$\begin{aligned} M(0^-) &= m\left(1 - s\sqrt{1/2\pi d} [1 + O(\frac{1}{d})]\right) \\ &= m\left(1 - \xi\sqrt{1/2\pi} [1 + O(\frac{1}{d})]\right), \end{aligned}$$

which is < 0 for all sufficiently large d since $\xi > \sqrt{2\pi}$. Also,

$$\begin{aligned} M(-1) &= m(\beta - 1) - \beta + ms + (m + s(E - 1)) \\ &= m(\beta - 1) - \beta + ms + m\left(1 - s\sqrt{1/2\pi d} [1 + O(\frac{1}{d})]\right), \end{aligned}$$

so for sufficiently large d (since $s > \sqrt{2\pi d}$ and $-m < 0$),

$$M(-1) < (m - 1)\beta + ms = (m - 1)\beta + m\xi\sqrt{d} < 0$$

since $m = 1/2$ and $\beta > \xi\sqrt{d}$. Furthermore, for $C < 0$, $M''(C) = 2(m\beta - m) = 2m(\beta - 1) > 0$, i.e. M is convex. It follows that

$$\sup_{C \in (-1, 0)} M(C) \leq \sup_{0 \leq \lambda \leq 1} \lambda M(-1) + (1 - \lambda)M(0^-) = \max [M(-1), M(0^-)] < 0.$$

For $C \geq 0$, it follows from (4) that $W(C) = 1$ and $W'(C) = 0$ and $C^+ = C$, and also $W(-C) = 1 - mC$, so

$$M(C) = \beta C - (r + d)mC^2 + s(E - 1).$$

Hence, $M'(C) = \beta - 2(r + d)mC$. So, on $[0, 1]$, the function M is first increasing and then decreasing, with a maximum where $\beta - 2(r + d)mC = 0$ so $C = \beta/2(r + d)m$. Hence, again using (6),

$$\begin{aligned} \sup_{C \in [0, 1]} M(C) &= M\left(\beta/2(r + d)m\right) \\ &= \beta^2 / [2(r + d)m] - (r + d)\beta^2 m / [4(r + d)^2 m^2] + s(E - 1) \\ &= \beta^2 / [4(r + d)m] - m\xi\sqrt{1/2\pi} [1 + O(\frac{1}{d})]. \end{aligned}$$

Then, using the bound (5) and that $(k + 1)d = r + d$, this is

$$\begin{aligned} &\leq (1 - \epsilon)4m^2\xi(k + 1)(d/\sqrt{2\pi})/[4(r + d)m] - m\xi\sqrt{1/2\pi} [1 + O(\frac{1}{d})] \\ &= (1 - \epsilon)m\xi(1/\sqrt{2\pi}) - m\xi\sqrt{1/2\pi} [1 + O(\frac{1}{d})] = [-\epsilon + O(\frac{1}{d})] m\xi/\sqrt{2\pi}, \end{aligned}$$

so it must be < 0 for all sufficiently large d .

The above results show that $\sup_{C \in [-1,1]} M(C) < 0$. The stated convergence in $V^{(1-\alpha)p}$ -norm then follows from Corollary 3. And since this convergence holds for any choice $\beta = (1 - \epsilon)^2(k + 1)\sqrt{d/2\pi}$ for sufficiently small $\epsilon > 0$, it holds for any $\beta < (k + 1)\sqrt{d/2\pi}$. Hence, the stated conclusion holds for any $a = \beta - 1 < (k + 1)\sqrt{d/2\pi} - 1$, as claimed.

For the lower bound, similar to Theorem 4 we have since $|v| = 1$ that

$$\|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \geq \frac{1}{2} \pi(S_t)$$

where $S_t = \{x \in \mathbf{R}^d : |x| \geq t\}$. But for large t , we have using polar coordinates that

$$\begin{aligned} \pi(S_t) &\propto \int_{|x| \geq t} (1 + |x|^2)^{-(r+d)/2} dx \propto \int_{\rho=t}^\infty (1 + \rho^2)^{-(r+d)/2} \rho^{d-1} d\rho \\ &\geq \int_{\rho=t}^\infty (\rho^2)^{-(r+d)/2} \rho^{d-1} d\rho = \int_{\rho=t}^\infty \rho^{-r-1} d\rho = \left. \frac{-\rho^{-r}}{r} \right|_{\rho=t}^{\beta=\infty} = \frac{t^{-r}}{r} \propto t^{-r}, \end{aligned}$$

so $\|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \geq \Omega(t^{-r})$, and hence $\lim_{t \rightarrow \infty} t^a \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} = \infty$ for $a > r$. ■

7 An Expectation Computation

To complete the proof of Theorem 8, we require the following computation:

Lemma 9 For $W(C)$ as in (4), consider the expected value $E := \mathbf{E}_\Psi[W(C_{x,U})]$, where $U \sim \Psi$ where Ψ is the uniform distribution on the unit sphere in \mathbf{R}^d for some $d > 1$, and $x \neq 0$ is any fixed vector in \mathbf{R}^d , and $C_{x,U}$ is the cosine of the angle between x and U . Then

$$E = 1 - m \frac{1/(d - 1)}{\sqrt{\pi} \Gamma(\frac{d-1}{2})/\Gamma(\frac{d}{2})} = 1 - m \sqrt{1/2\pi d} \left[1 + O\left(\frac{1}{d}\right) \right] \text{ as } d \rightarrow \infty.$$

To prove Lemma 9, we first need another lemma giving the $C_{x,U}$ density function:

Lemma 10 Let $U \sim \Psi$ as in Lemma 9. Then the quantity $C_{x,U}$ has density function on $[-1, 1]$ proportional to $f(c) = (1 - c^2)^{(d-3)/2}$.

Proof Let (e_1, \dots, e_d) be an orthonormal basis of \mathbf{R}^d with $e_1 = x/|x|$, and write $Z = (Z_1, \dots, Z_d)$ in this basis where $\{Z_i\}$ are i.i.d. $N(0, 1)$. Then the unit vector $Z/|Z|$ has uniform distribution Ψ , so $C_{x,U}$ has the same distribution as

$$\frac{x}{|x|} \cdot \frac{Z}{|Z|} = (1, 0, \dots, 0) \cdot \frac{(Z_1, Z_2, \dots, Z_d)}{|Z|} = \frac{Z_1}{|Z|}.$$

Therefore, $C_{x,U}^2$ has the same distribution as

$$\frac{Z_1^2}{|Z|^2} = \frac{Z_1^2}{Z_1^2 + (Z_2^2 + \dots + Z_d^2)} = \frac{\chi^2(1)}{\chi^2(1) + \chi^2(d-1)} \sim \text{Beta}\left(\frac{1}{2}, \frac{d-1}{2}\right),$$

using the general property that if $X \sim \chi^2(\alpha)$ and $Y \sim \chi^2(\beta)$ are independent, then $\frac{X}{X+Y} \sim \text{Beta}\left(\frac{\alpha}{2}, \frac{\beta}{2}\right)$. Hence, $C_{x,U}^2$ has density function on $[0, 1]$ proportional to $h(c) = c^{\frac{1}{2}-1}(1-c)^{\frac{d-1}{2}-1} = c^{-1/2}(1-c)^{(d-3)/2}$.

Then, $|C_{x,U}| = \sqrt{C_{x,U}^2} = g(C_{x,U}^2)$ where $g(c) = \sqrt{c}$ and $g^{-1}(c) = c^2$. So, by the change-of-variable formula, $|C_{x,U}|$ has density on $[0, 1]$ proportional to

$$h(g^{-1}(c)) \left| \frac{d}{dc} g^{-1}(c) \right| = h(c^2) \left| \frac{d}{dc} c^2 \right| = c^{-1} (1-c^2)^{(d-3)/2} |2c| \propto (1-c^2)^{(d-3)/2}.$$

Finally, since $C_{x,U}$ is symmetric about 0, the density of $C_{x,U}$ on all of $[-1, 1]$ must also be proportional to $(1-c^2)^{(d-3)/2}$. ■

Proof of Lemma 9 We compute using Lemma 10 that

$$E := \mathbf{E}[W(C_{x,U})] = 1 + m \mathbf{E}[C_{x,U} \mathbf{1}_{C_{x,U} < 0}] = 1 + m \frac{\int_{-1}^0 c(1-c^2)^{(d-3)/2} dc}{\int_{-1}^1 (1-c^2)^{(d-3)/2} dc}.$$

But for $d > 1$, it can be computed that $\int_{-1}^0 c(1-c^2)^{(d-3)/2} = -\frac{1}{d-1}$, and $\int_{-1}^1 (1-c^2)^{(d-3)/2} = \sqrt{\pi} \Gamma(\frac{d-1}{2}) / \Gamma(\frac{d}{2})$. Hence,

$$E = 1 - m \frac{1/(d-1)}{\sqrt{\pi} \Gamma(\frac{d-1}{2}) / \Gamma(\frac{d}{2})}. \quad (7)$$

Next, we use Stirling's Approximation, which says (e.g. Jameson 2015) that for all $x > 0$, we have

$$\sqrt{2\pi} x^{x-1/2} e^{-x} \leq \Gamma(x) \leq \sqrt{2\pi} x^{x-1/2} e^{-x} e^{1/(12x)}.$$

It follows that as $x \rightarrow \infty$,

$$\Gamma(x) = \sqrt{2\pi} x^{x-1/2} e^{-x} \left[1 + O\left(\frac{1}{x}\right) \right].$$

Hence, as $d \rightarrow \infty$,

$$\Gamma\left(\frac{d-1}{2}\right) / \Gamma\left(\frac{d}{2}\right) = \frac{\left(\frac{d-1}{2}\right)^{\frac{d-2}{2}} e^{-\frac{d-1}{2}}}{\left(\frac{d}{2}\right)^{\frac{d-1}{2}} e^{-\frac{d}{2}}} \left[1 + O\left(\frac{1}{d}\right) \right] = \left(\frac{d-1}{d}\right)^{\frac{d-2}{2}} \sqrt{2e/d} \left[1 + O\left(\frac{1}{d}\right) \right].$$

Now, as $x \rightarrow 0$, $e^x = 1 + x + O(x^2)$, i.e. $1 + x = e^x + O(x^2) = e^x [1 + O(x^2)]$. So, as $d \rightarrow \infty$, $\frac{d-1}{d} = 1 - \frac{1}{d} = e^{-1/d} [1 + O(d^{-2})]$, whence

$$\left(\frac{d-1}{d}\right)^d = \left(e^{-1/d} [1 + O(d^{-2})]\right)^d = e^{-1} \left[1 + O\left(\frac{1}{d}\right) \right],$$

and so

$$\left(\frac{d-1}{d}\right)^{\frac{d-2}{2}} = \left[\left(\frac{d-1}{d}\right)^d\right]^{1/2} \left(\frac{d}{d-1}\right) = (e^{-1})^{1/2} \left[1 + O\left(\frac{1}{d}\right)\right].$$

It follows that

$$\Gamma\left(\frac{d-1}{2}\right) / \Gamma\left(\frac{d}{2}\right) = e^{-1/2} \sqrt{2e/d} \left[1 + O\left(\frac{1}{d}\right)\right] = \sqrt{2/d} \left[1 + O\left(\frac{1}{d}\right)\right].$$

Therefore, from (7),

$$E = 1 - m \frac{1/(d-1)}{\sqrt{\pi} \sqrt{2/d}} \left[1 + O\left(\frac{1}{d}\right)\right] = 1 - m \sqrt{1/2\pi d} \left[1 + O\left(\frac{1}{d}\right)\right],$$

as claimed. ■

Author Contribution This project was an equal collaboration between the two authors.

Funding GOR was supported by EPSRC grants Bayes for Health (R018561) and CoSInES (R034710). JSR was supported by NSERC discovery grant RGPIN-2019-04142.

Data Availability This paper introduces no new data nor materials, aside from the R script for the computer simulations which is freely available at: probability.ca/Rpoly

Declarations

Conflict of Interest The authors declare that they have no financial nor non-financial competing interests.

References

- Andrieu C, Dobson P, Wang AQ (2021a) Subgeometric hypocoercivity for piecewise-deterministic Markov process Monte Carlo methods. *Elec J Prob* 26:1–26
- Andrieu C, Durmus A, Nüsken N, Roussel J (2021b) Hypocoercivity of piecewise deterministic Markov process-Monte Carlo. *Ann Appl Prob* 31(5):2478–2517
- Bierkens J (2021) Piecewise deterministic Monte Carlo web resource page. <https://diamweb.ewi.tudelft.nl/~joris/pdmps.html>
- Bierkens J, Fearnhead P, Roberts GO (2019) The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data. *Ann Stat* 47(3):1288–1320
- Bierkens J, Roberts GO, Zitt PA (2019) Ergodicity of the zigzag process. *Ann Appl Prob* 29(4):2266–2301
- Bouchard-Côté A, Vollmer SJ, Doucet A (2018) The bouncy particle sampler: a nonreversible rejection-free Markov chain Monte Carlo method. *J Amer Stat Assoc* 113(522):855–867
- Brooks S, Gelman A, Jones G, Meng X-L (eds) (2011) *Handbook of Markov chain Monte Carlo*. Chapman & Hall, New York
- Deligiannidis G, Bouchard-Côté A, Doucet A (2019) Exponential ergodicity of the Bouncy Particle sampler. *Ann Stat* 47(3):1268–1287
- Ethier SN, Kurtz TG (1986) *Markov processes, characterization and convergence*. Wiley, New York
- Fort G, Moulines E (2000) V-subgeometric ergodicity for a Hastings-Metropolis algorithm. *Stat Prob Lett* 49:401–410
- Fort G, Moulines E (2003) Polynomial ergodicity of Markov transition kernels. *Stoch Proc Appl* 103(1):57–99
- Fort G, Roberts GO (2005) Subgeometric ergodicity of strong Markov processes. *Ann Appl Prob* 15(2):1565–1589
- Jameson GJO (2015) A simple proof of Stirling’s formula for the gamma function. *Math Gazette* 99(544):68–74
- Jarner SF, Roberts GO (2002) Polynomial convergence rates of Markov chains. *Ann Appl Prob* 12:224–247
- Meyn SP, Tweedie RL (1993) *Markov chains and stochastic stability*. Springer-Verlag, London. Available at probability.ca/MT

- Rosenthal JS (2002) Quantitative convergence rates of Markov chains: a simple account. *Elec Comm Prob* 7:123–128
- Vasdekis G, Roberts GO (2021) A note on the polynomial ergodicity of the one-dimensional Zig-Zag process. *J Appl Prob*. <https://arxiv.org/abs/2106.11357v2>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.