# Polynomial Convergence Rates of Piecewise Deterministic Markov Processes

by

Gareth O. Roberts[1] and Jeffrey S. Rosenthal[2]

(Version of: December 16, 2021)

**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.

## 1. Introduction.

Markov chain Monte Carlo (MCMC) algorithms have become an indispensible part of statistical computation; see e.g. [6] and the many references therein. Piecewise-deterministic Markov processes (PDMP), such as the Bouncy Particle sampler [5] and the Zig-Zag algorithm [4], 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. [3] 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 [7]. However, if the target distribution has tails which are heavier than exponential, then geometric ergodicity does not apply.

In this paper, we instead focus on *polynomial* convergence rates of certain PDMP. That topic was previously approached using the concept of hypocoercivity in [1, 2], but here we proceed using direct drift condition methods. We specifically consider the Bouncy Particle sampler [5], for a given target density $\pi$ 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 $\pi$'s contour lines at hazard rate $\lambda(x, v) = \left[-v\,(\log\pi)'(x)\right]^+$. In addition, it refreshes at some specified hazard rate (which

[1]Department of Statistics, University of Warwick, CV4 7AL, Coventry, U.K. Email: `Gareth.O.Roberts@warwick.ac.uk`. Supported in part by EPSRC grants EP/K014463/1 (i-Like) and EP/D002060/1 (CRiSM).

[2]Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Email: `jeff@math.toronto.edu`. Web: `http://probability.ca/jeff/` Supported in part by NSERC of Canada.

could depend on the current position $x$), at which point it replaces the velocity $v$ by an independent draw from the uniform distribution $\Psi$ on the unit sphere in $\mathbf{R}^d$. This process is known [5] to be irreducible with stationary density $\pi$, and to converge to $\pi$ exponentially quickly for sufficiently light-tailed target densities $\pi$.

In this paper, we instead examine the polynomial convergence rate of this PDMP to target densities $\pi$ 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 [14]). For targets with tails like 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 $\lim_{t\to\infty} t^a \, \|P^t(x, \cdot) - \pi(\cdot)\|_{TV}$ equals 0 for $a < r$ and infinity for $a > r$. We also prove convergence in the $V^{(1-\alpha)p}$-norm (see Section 2) at polynomial order approaching $(1-p)r$, for any $p \in [0, 1)$. We then consider high-dimensional PDMP, and compute their infinitesimal generator (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. In particular, we prove 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 review general polynomial convergence rate bounds for continuous-time processes as in [10], and present some corollaries adapting those results to our needs. In Section 3, we consider one-dimensional PDMP, and prove an exact characterisation (Theorem 4) of the polynomial convergence order in that case. In Section 4, we prove a general result (Theorem 5) which gives the infinitesimal generator of PDMP for certain choices of drift function, which we then apply to target densities with polynomial tails (Corollary 7). In Section 5, we apply our generator computations to derive specific polynomial rate bounds for high-dimensional PDMP (Theorem 8). Finally, in Section 6, we present an auxiliary computation about expected values with respect to the refresh distribution $\Psi$, which is used in the proof of Theorem 8.

# 2. Polynomial Convergence Rates.

Quantitative convergence rates of discrete-time geometrically ergodic Markov chains have a long history, see e.g. [13] and the many references therein. More recently, focus has turned to polynomial ergodicity, e.g. [8, 12]. Most of these results are in discrete time, but [10] yields the following continuous-time polynomial convergence bound. To state it, recall that if $\mu$ and $\nu$ are two probability distributions on $\mathcal{X}$, and $f : \mathcal{X} \to (0, \infty)$, then the $f$-norm distance between $\mu$ and $\nu$ is defined as

$$\|\mu(\cdot) - \nu(\cdot)\|_f \;:=\; \sup_{\substack{g:\mathcal{X}\to\mathbf{R} \\ |g|\le f}} \left|\mathbf{E}_\mu(g) - \mathbf{E}_\nu(g)\right|,$$

and the total variation distance between $\mu$ and $\nu$ is defined as

$$\|\mu(\cdot) - \nu(\cdot)\|_{TV} \;:=\; \sup_{\substack{g:\mathcal{X}\to\mathbf{R} \\ |g|\le 1}} \left|\mathbf{E}_\mu(g) - \mathbf{E}_\nu(g)\right|.$$

Then we have:

**Proposition 1.** *If a continuous-time Markov process on state space $\mathcal{X} \subseteq \mathbf{R}^d$ has stationary distribution $\pi$, 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 $V \ge 1$ with $\sup_{x\in C} V(x) < \infty$ such that $\mathcal{A}V(x) \le -c\,(V(x))^{1-\alpha} + b_0\,\mathbf{1}_C(x)$ for all $x \in \mathcal{X}$, then for any $p \in [0,1)$ and $x \in \mathcal{X}$,*

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

*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 $p = 0$,*

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

*i.e. $\|P^t(x,\cdot) - \pi(\cdot)\|_{TV} \;\le\; O(t^{-(1-\alpha)/\alpha})$, 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 [10] upon setting their $\eta = 1$ and $c_\eta = c$ and $b = 0$, and using that $t \le 1 + t$. (Note that the "$b$" in their convergence equation is different from the "$b$" in their drift equation (8), which we here refer to as "$b_0$".) ∎

Recall that a function $V : \mathbf{R}^d \to \mathbf{R}$ is *norm-like* if $\lim_{|x|\to\infty} V(x) = \infty$. Then we have:

**Corollary 2.** *If a continuous-time Markov process on state space $\mathcal{X} \subseteq \mathbf{R}^d$ has stationary distribution $\pi$, 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 $V \ge c_0 > 0$ such that $\mathcal{A}V$ is bounded on compact sets and $\mathcal{A}V(x) \le -c\,(V(x))^{1-\alpha}$ for all $x \in \mathcal{X}$ with $V(x) \ge \Delta$, then again for any $p \in [0,1)$ and $x \in \mathcal{X}$,*

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

**Proof.** First of all, by replacing $V$ by $V/c_0$ and $c$ by $c/c_0^\alpha$ if necessary, we can assume that $c_0 = 1$. Then, let $C = \{x \in \mathcal{X} : V(x) \le \Delta\}$. This $C$ is closed by continuity of $V$, and is bounded since $V$ is norm-like, so $C$ is compact. It then follows from Lemma 2 of [7] that $C$ is small and hence 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 $\mathcal{A}V(x) \le -c\,(V(x))^{1-\alpha}$ when $V(x) \ge \Delta$ then $\mathcal{A}V(x) \le -c\,(V(x))^{1-\alpha} + b_0\,\mathbf{1}_C(x)$ for all $x \in \mathcal{X}$. ∎

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

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

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

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

*and in particular*

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

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

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

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 $(1-\alpha)/\alpha = (1-\frac{1}{\beta})\big/(1/\beta) = \beta - 1$. ∎

**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 (computible) bounds on the distance to stationarity of PDMP, similar in spirit to [13] and the references therein; by using the related results of [9].

# 3. Convergence Rate in One Dimension.

Suppose first that the state space $\mathcal{X}$ is the one-dimensional real line $\mathbf{R}$, with $C^1$ target density $\pi(x)$. Polynomial convergence rates for the related Zig-Zag process on one-dimensional heavy-tailed targets have been studied in [14]. 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.

To proceed, consider the algorithm which exhances the state space to $\mathcal{X}\times\{-1,1\}$, and expands $\pi$ to $\pi(x,v)=\frac{1}{2}\pi(x)$ for $v\in\{-1,1\}$. The PDMP proceeds by moving with fixed constant velocity $v$, except reflecting from $v$ to $-v$ with hazard rate $\lambda(x,v)=\big[-v\,(\log\pi)'(x)\big]^+$. (We omit refreshes, i.e. take the refresh rate to be zero, since refreshes are not required in one dimension.) This process has infinitesimal generator on $C^1$ functions $f:\mathcal{X}\times\{-1,1\}\to\mathbf{R}$ given by

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

Consider now the specific example where $\pi(x)=(1+x^2)^{-(1+r)/2}$ is a $t$-distribution, for a fixed constant $r\geq 1$ (the "degrees of freedom"), at least when $|x|\geq\Delta$.

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

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

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

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

**Proof.** Assume that $|x|\geq\Delta$. We compute that

$$\lambda(x,v) \;=\; \begin{cases} \frac{(1+r)x}{1+x^2}\,, & xv>0 \\ 0\;\;\;\;\;, & xv<0 \end{cases}$$

It follows that $\lambda(x,v)\geq(1+r)/(1+x)$ for $v=+1$ and $x\geq 1$. Next, for some $\beta>0$ and $K>1$ to be determined later, let

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

Then we compute that for $x,v>0$,

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

Suppose it holds that

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

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

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

Hence, if $K^* = \min\big[\beta,\;(K-1)(r+1) - \beta K\big]$, then assuming (1), we have for $x > 0$ and $v \in \{-1, 1\}$ that

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

where $\alpha = 1/\beta$. By symmetry, this condition also holds for $x < 0$, i.e. it holds whenever $|x| \ge \Delta$. Applying Corollary 2 with $\alpha = 1/\beta$, so $(1-\alpha)/\alpha = \beta - 1$, gives

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

and hence with $p = 0$,

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

It remains to ensure that (1) holds. But (1) 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 $\ge r - \epsilon$ 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 $x \le 0$, and similarly that $P^t((x, \pm 1),\;(-\infty, -t)) = 0$ for $x \ge 0$. Hence, for any $x \in \mathbf{R}$, by symmetry,

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

which for large $t$ 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. ∎

# 4. 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 $\pi$'s contour lines at the hazard rate

$$\lambda(x, v) \;=\; \Big( - (\nabla \log \pi) \cdot v \Big)^{+}.$$

And it refreshes, by drawing a new $v$ independently from the uniform distribution $\Psi$ on the unit sphere in $\mathbf{R}^d$, with refresh rate which we take to be $s/|x|$ for some constant $s > 0$ (which might depend on $d$).

To proceed, we consider a drift function of the form

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

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 $U \sim \Psi$. Then we have:

**Theorem 5.** *The above PDMP has infinitesimal generator given as $|x| \to \infty$ by*

$$\mathcal{A}V(x, v) \;=\; |x|^{\beta - 1} \, B(x, v) \;\; \left[1 + O\left(|x|^{-\beta}, |x|^{-2}\right)\right]$$

*where*

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

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,

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

**Proof of Theorem 5.** We wish to compute the generator $\mathcal{A}V$. 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). If $x$ has velocity $v$, then

$$\frac{\partial|x|}{\partial t} \;=\; |v|\,C_{x,v} \;=\; C_{x,v}$$

since our dynamics always has $|v|=1$. Hence,

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

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

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

Hence,

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

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

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

Hence, by the product rule for derivatives,

$$\nabla_xV(x,v) \;=\; W(C_{x,v})\,\nabla_x\left(|x|^\beta\right)+\left(1+|x|^\beta\right)W'(C_{x,v})\,\nabla_xC_{x,v}$$

$$=\; W(C_{x,v})\,\beta|x|^{\beta-2}x+\left(1+|x|^\beta\right)W'(C_{x,v})\left(\frac{v}{|x|}-\frac{C_{x,v}\,x}{|x|^2}\right).$$

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

$$\mathcal{A}_1V(x,v) \;=\; \Big(\nabla_xV(x,v)\Big)\cdot v$$

$$=\; W(C_{x,v})\,\beta|x|^{\beta-2}(x\cdot v)+\left(1+|x|^\beta\right)W'(C_{x,v})\left(\frac{1}{|x|}-\frac{C_{x,v}\,(x\cdot v)}{|x|^2}\right)$$

$$=\; W(C_{x,v})\,\beta|x|^{\beta-1}C_{x,v}+\left(1+|x|^\beta\right)W'(C_{x,v})\left(\frac{1}{|x|}-\frac{{C_{x,v}}^2}{|x|}\right)$$

$$\approx\; |x|^{\beta-1}\Big[W(C_{x,v})\,\beta\,C_{x,v}+W'(C_{x,v})(1-{C_{x,v}}^2)\Big]\,.$$

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

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

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 $\Psi$ on $\{z\in\mathbf{R}^d : |z|=1\}$. This changes $V(x,v)$ from $W(C_{x,v})\big(1+|x|^\beta\big)$ to $W(C_{x,U})\big(1+|x|^\beta\big)$ where $U\sim\psi$, which is a difference of $[W(C_{x,U})-W(C_{x,v})]\big(1+|x|^\beta\big)$. Hence,

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

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 $\pi$ has polynomial tails as in a $t$-distribution, i.e. that

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

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

**Corollary 7.** *If $\pi$ is given by (2), then the above PDMP has infinitesimal generator given as $|x|\to\infty$ by*

$$\mathcal{A}V(x,v) \;=\; |x|^{\beta-1}\,M(C_{x,v})\,\Big[1+O\big(|x|^{-\beta},|x|^{-2}\big)\Big]$$

*where*

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

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

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

and

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

The result then follows from Theorem 5. ∎

# 5. 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, with $\pi$ as in (2) with $r > (2\pi - 1)d$, we have for any $a < (r+d)\sqrt{2\pi/d} - 1$ any $p \in [0,1)$ and all sufficiently large $d \in \mathbf{N}$ that*

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

*and in particular*

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

*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 $a > r$,*

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

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

$$W(C) \;=\; 1 + m\,C\,\mathbf{1}_{C<0} \;:=\; \begin{cases} 1, & C \geq 0 \\ 1 + m\,C, & C < 0 \end{cases} \tag{3}$$

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

$$\beta^2 \;=\; (1-\epsilon)^4(k+1)^2 d/2\pi \;=\; (1-\epsilon)\xi(k+1)d/\sqrt{2\pi} \;<\; 4m^2\xi(k+1)d/\sqrt{2\pi}\,. \tag{4}$$

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

For $C < 0$, it follows from (3) that $W(C) = 1 + m\,C$ and $W'(C) = m$ and $C^+ = 0$, so

$$M(C) \;=\; \beta C + m\beta C^2 + m - mC^2 + s(E - 1 - mC)$$

$$= \; C^2 m(\beta-1) + C(\beta - ms) + (m - s(1-E))\,.$$

Hence

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

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

$$1 - E \;=\; m\sqrt{1/2\pi d}\,\left[1 + O(\frac{1}{d})\right], \tag{5}$$

and hence

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

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

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

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

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

For $C \geq 0$, it follows from (3) that $W(C) = 1$ and $W'(C) = 0$ and $C^+ = C$, and also $W(-C) = 1 - mC$, so

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

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 (5).

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

which is $< 0$ for sufficiently large $d$ by (4).

The above results show that $\sup_{C\in[-1,1]} M(C) < 0$. The stated upper bound then follows from Corollary 3. And since it applies for any choice $\beta = (1-\epsilon)^2(k+1)\sqrt{d/2\pi}$ for sufficiently small $\epsilon > 0$, it applies for any $\beta < (k+1)\sqrt{d/2\pi}$, leading to the conclusion for any $a = \beta - 1 < (k+1)\sqrt{d/2\pi} - 1$.

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

$$\pi(S_t) \ \propto\ \int_{|x|\geq t} (1+|x|^2)^{-(r+d)/2}\,dx \ \propto\ \int_{\rho=t}^{\infty} (1+\rho^2)^{-(r+d)/2}\ \rho^{d-1}\,d\rho$$

$$\geq\ \int_{\rho=t}^{\infty} (\rho^2)^{-(r+d)/2}\ \rho^{d-1}\,d\rho \ =\ \int_{\rho=t}^{\infty} \rho^{-r-1}\,d\rho \ =\ \frac{-\rho^{-r}}{r}\Big|_{\rho=t}^{\rho=\infty} \ =\ \frac{t^{-r}}{r} \ \propto\ t^{-r}\,,$$

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

## 6. An Expectation Computation.

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

**Lemma 9.** *For $W(C)$ as in (3), consider the expected value $E := \mathbf{E}_\Psi[W(C_{x,U})]$, where $U \sim \Psi$ where $\Psi$ 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*

$$E \ =\ 1 - m\frac{1/(d-1)}{\sqrt{\pi}\,\Gamma(\frac{d-1}{2})/\Gamma(\frac{d}{2})} \ =\ 1 - m\sqrt{1/2\pi d}\,\Big[1 + O\Big(\frac{1}{d}\Big)\Big] \quad \text{as } d\to\infty\,.$$

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

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

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

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

using the general property that if $X \sim \chi^2(\alpha)$ and $Y \sim \chi^2(\beta)$ are independent, then $\frac{X}{X+Y} \sim \text{Beta}(\frac{\alpha}{2}, \frac{\beta}{2})$. 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

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

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,

$$E \;=\; 1-m\,\frac{1/(d-1)}{\sqrt{\pi}\;\Gamma(\frac{d-1}{2})\;/\;\Gamma(\frac{d}{2})}. \tag{6}$$

Next, we use Stirling's Approximation, which says (e.g. [11]) that for all $x>0$, we have

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

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

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

Hence, as $d\to\infty$,

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

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

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

and so

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

It follows that

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

Therefore, from (6),

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

as claimed. ∎

# References


[1] C. Andrieu, P. Dobson, and A.Q. Wang (2021), Subgeometric hypocoercivity for piecewise-deterministic Markov process Monte Carlo methods. Elec. J. Prob. **26**, 1–26.

[2] C. Andrieu, A. Durmus, N. Nüsken, and J. Roussel (2021), Hypocoercivity of piecewise deterministic Markov process-Monte Carlo. Ann. Appl. Prob. **31(5)**, 2478–2517.

[3] J. Bierkens (2021), Piecewise Deterministic Monte Carlo web resource page. https://diamweb.ewi.tudelft.nl/∼joris/pdmps.html

[4] J. Bierkens, P. Fearnhead, and G.O. Roberts (2019), The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data. Ann. Stat. **47(3)**, 1288–1320.

[5] A. Bouchard-Côté, S.J. Vollmer, and A. Doucet (2018), The bouncy particle sampler: a nonreversible rejection-free Markov chain Monte Carlo method. J. Amer. Stat. Assoc. **113(522)**, 855–867.

[6] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, eds. (2011), Handbook of Markov chain Monte Carlo. Chapman & Hall, New York.

[7] G. Deligiannidis, A. Bouchard-Côté, and A. Doucet (2019), Exponential ergodicity of the Bouncy Particle sampler. Ann. Stat. **47(3)**, 1268–1287.

[8] G. Fort and E. Moulines (2000), V-subgeometric ergodicity for a Hastings-Metropolis algorithm. Stat. Prob. Lett. **49**, 401–410.

[9] G. Fort and E. Moulines (2003), Polynomial ergodicity of Markov transition kernels. Stoch. Proc. Appl. **103(1)**, 57–99.

[10] G. Fort and G.O. Roberts (2005), Subgeometric ergodicity of strong Markov processes. Ann. Appl. Prob. **15(2)**, 1565–1589.

[11] G.J.O. Jameson (2015), A simple proof of Stirling's formula for the gamma function. Math. Gazette **99(544)**, 68–74.

[12] S.F. Jarner and G.O. Roberts (2002), Polynomial convergence rates of Markov chains. Ann. Appl. Prob. **12**, 224–247.

[13] J.S. Rosenthal (2002), Quantitative convergence rates of Markov chains: A simple account. Elec. Comm. Prob. **7**, 123–128.

[14] G. Vasdekis and G.O. Roberts (2021), A note on the polynomial ergodicity of the one-dimensional Zig-Zag process. https://arxiv.org/abs/2106.11357v2 J. Appl. Prob., to appear.