A NOTE ON THE POLYNOMIAL ERGODICITY OF THE ONE-DIMENSIONAL ZIG-ZAG PROCESS

G. VASDEKIS,* *University of Warwick*

G. O. ROBERTS,** *University of Warwick*

Abstract

We prove polynomial ergodicity for the one-dimensional Zig-Zag process on heavy tailed targets and identify the exact order of polynomial convergence of the process when targeting Student distributions.

Keywords: Piecewise Deterministic Markov Process, Markov Chain Monte Carlo, Polynomial Ergodicity.

2020 Mathematics Subject Classification: Primary 60J25

Secondary 65C05, 60F05.

1. Introduction

The Zig-Zag process is a Piecewise Deterministic Markov Process (PDMP) that was recently used as a new way to construct MCMC algorithms. The one dimensional Zig-Zag appeared in [8] as a scaling limit of the Lifted Metropolis-Hastings (see [21, 12]) applied to the Curie-Weiss model (see [19]). The process was later extended in higher dimensions in [5] and has been proposed as a way to sample from posterior distributions in a Bayesian setting (see also [14, 22]). Since then, its properties have been extensively studied in the literature (see for example [4, 6, 10, 7] etc.).

[9] proves ergodicity and exponential ergodicity of the Zig-Zag process in arbitrary dimension, however a crucial assumption required for exponential ergodicity in that

---

* Postal address: Department of Statistics, University of Warwick, Coventry, CV4 7AL

* Email address: Giorgos.Vasdekis.1@warwick.ac.uk

** Postal address: Department of Statistics, University of Warwick, Coventry, CV4 7AL

** Email address: Gareth.O.Roberts@warwick.ac.uk

work is that the target density has exponential or lighter tails. In [23] the converse result was proven. The Zig-Zag sampler fails to be exponentially ergodic when the target distribution is heavy tailed. On the other hand, it was shown in Theorem 1 of [9] that the process will converge to the invariant measure under very mild assumptions, including the heavy tailed case. Furthermore, [1] used hypocoercivity techniques (see also [2]) to prove polynomial rates of convergence for the Zig-Zag process on heavy tailed targets in arbitrary dimension.

In this note we will focus on the one-dimensional Zig-Zag process and prove polynomial ergodicity in the heavy tailed scenario. The result applies in the special case where the tails of the target decay in the same manner as a Student distribution with ν degrees of freedom. In that case, we prove that the polynomial rate of convergence is arbitrarily close to ν, but not more than ν. This improves upon the result stated in [1] in the special case where $d = 1$, although their work provides convergence results for higher dimensions as well.

The rest of this paper is organised as follows. In Section 2 we recall the definition of the one-dimensional Zig-Zag process and we state the main result concerning its polynomial ergodicity in heavy tailed targets. In Section 3 we provide proof of this result. Finally, in Section 4 we discuss the rates of polynomial convergence of the Zig-Zag process and compare them with the ones of other state of the art Metropolis-Hastings algorithms.

2. Results

We begin with recalling the definition of the one-dimensional Zig-Zag process. Let $E = \mathbb{R} \times \{-1, +1\}$, $U \in C^1(E)$ and $\lambda : E \to \mathbb{R}_{\geq 0}$ with

$$\lambda(x, \theta) = [\theta U'(x)]^+ + \gamma(x), \tag{1}$$

where γ is a non-negative integrable function and we write $a^+ = \max\{a, 0\}$. The one-dimensional Zig-Zag process $(Z_t)_{t \geq 0} = (X_t, \Theta)_{t \geq 0}$ is a continuous time Markov process with state space E which evolves as follows. If the process starts from $(x, \theta) \in E$, then $X_t = x + t\theta$ and $\Theta_t = \theta$ for all $t < T_1$, where T_1 is the first arrival time of a non-homogeneous Poisson process with rate $m_1(s) = \lambda(x + s\theta, \theta)$. Then $X_{T_1} = x + T_1\theta$, $\Theta_{T_1} = -\theta$. Then, $X_t = X_{T_1} + (t - T_1)\Theta_{T_1}$ and $\Theta_t = \Theta_{T_1}$ for all $t \in (T_1, T_1 + T_2)$,

where T_2 is the first arrival time of a Poisson process with intensity $m_2(s) = \lambda(X_{T_1} + s\Theta_{T_1}, \Theta_{T_1})$. The process is then defined inductively up to time T_n for all $n \in \mathbb{R}$.

It is proven in [5] that with this choice of λ, the process has measure μ invariant, where

$$\mu = \pi \bigotimes \frac{1}{2}\delta_{\{-1,+1\}} \tag{2}$$

and

$$\pi(dx) = \frac{1}{Z}\exp\{-U(x)\}dx \tag{3}$$

with $Z = \int_{\mathbb{R}} \exp\{-U(y)\}dy < \infty$.

Since this note concerns the polynomial ergodicity of the one-dimensional Zig-Zag process, we now recall the definition of polynomial ergodicity.

Definition 1. Let $(Z_t)_{t\geq 0}$ be a Markov process with state space E, having invariant probability measure μ and let $k > 0$. We say that the process is polynomially ergodic of order k if there exists a function $M : E \to \mathbb{R}_{>0}$ such that for all $z \in E$ and $t \geq 0$

$$\|\mathbb{P}_z(Z_t \in \cdot) - \mu(\cdot)\|_{TV} \leq \frac{M(z)}{t^k},$$

where $\mathbb{P}_z$ denotes the law of the process starting from z.

We will make the following assumption, typically verified in practice.

Assumption 1. *Assume that there exists an $\nu > 0$ and a compact set $C \subset \mathbb{R}$ such that for all $x \notin C$,*

$$|U'(x)| \geq \frac{1+\nu}{|x|}. \tag{4}$$

Remark 1. This assumption directly implies that there exists a c' such that for all $x \in \mathbb{R}$, $U(x) \geq (1+\nu)\log(|x|) - c'$. This is an assumption made in [9] in order to prove non-evanescence of the Zig-Zag process and it's a natural assumption given that the function $\exp\{-U(x)\}$ must be integrable.

We also need the following assumption for the refresh rate γ.

Assumption 2. *Assume that the refresh rates satisfy*

$$\lim_{|x|\to\infty} \frac{\gamma(x)}{|U'(x)|} = 0. \tag{5}$$

Assumption 2 ensures that the bouncing events will vastly outnumber the refresh events, at least in the tails of the target. Whilst we have not been able to establish the necessity of this assumption for the conclusions of Theorem 1, some control over $\frac{\gamma(x)}{|U'(x)|}$ is definitely needed. Indeed in the regime where $\lim_{|x|\to\infty} \frac{\gamma(x)}{|U'(x)|} = +\infty$ (as would be the case with constant refresh rate and heavy-tailed target) random direction changes would outnumber systematic ones, leading to random walk/diffusive behaviour commonly associated with slow convergence. In fact in this case, the algorithm would resemble a Random Walk Metropolis algorithm, which is known to converge at slower polynomial rate. Therefore, we believe that some control of the refresh rate relative to $|U'|$ should be assumed for guarantees about specific rates of convergence to hold. This can be further supported by simulation studies. In Figure 1 we present the Mean Square Error of estimating a tail probability ($\mathbb{P}(X \geq 5)$ for a standard Cauchy target) using Zig-Zag with differing refresh rates. We consider $\gamma(x) = 0$, $\gamma(x) = |U'(x)|$ and $\gamma(x) = 1$. For each algorithm we generated 1000 independent realisations, all starting from $(-5, +1)$ and all realisations run until time $T = 10^4$. For each time less than T the average square error of the true probability (approximately equal to 0.0628) is reported. It is clear that the smaller refresh rate leads to more rapid convergence.

Our main result is the following.

Theorem 1. (Polynomial ergodicity of Zig-Zag.) *Suppose that U satisfies Assumption 1 and let C and $\nu > 0$ as in (4). Suppose further that the refresh rate satisfies Assumption 2. Then, for any $k < \nu$, there exist constants $B, \delta > 0$ and $\beta \in (0, 1)$ such that if we let*

$$V_{\beta,\delta}(x, \theta) = \exp\{\beta U(x) + \delta \operatorname{sgn}(x)\theta\} \tag{6}$$

then for all $(x, \theta) \in \mathbb{R} \times \{-1, +1\}$,

$$\|\mathbb{P}_{x,\theta}(Z_t \in \cdot) - \mu(\cdot)\|_{TV} \leq \frac{BV_{\beta,\delta}(x, \theta)}{t^{1+k}} + \frac{B}{t^k}, \tag{7}$$

i.e. the process is polynomially ergodic of order k.

Remark 2. By carefully inspecting the proof of Theorem 1, we observe that Assumption 2 can be weakened in the following sense. If ν is as in Assumption 1 and if we fix $k < \nu$, then in order to prove polynomial ergodicity of order k, it suffices that there
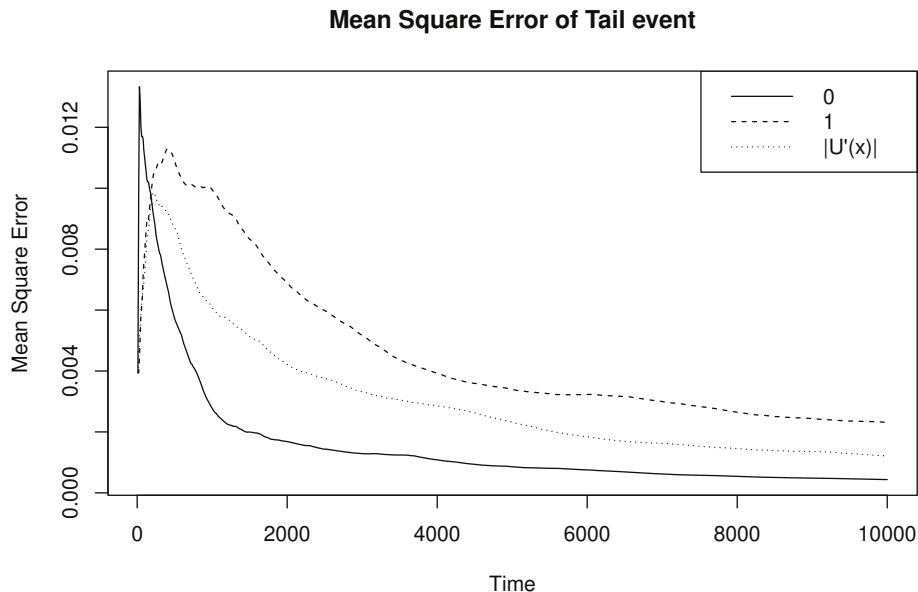
Mean Square Error of Tail event

FIGURE 1: *Mean Square Error over time of three Zig-Zag algorithms with different refresh rates $\gamma(x)$, described in the upper right corner. The algorithms target the Cauchy distribution and are used to estimate the probability mass the Cauchy distribution assigns to the event $[5, +\infty)$. For all three algorithms, 1000 independent realisations were generated and for each time less than 10^4 the average square error between the approximation of the probability and the actual probability (approximately 0.0628) are reported. Evidently, the fastest convergence is achieved by lowering the values of γ.*

exists a small $\eta > 0$ and that we ask that

$$\lim_{|x| \to \infty} \frac{\gamma(x)}{|U'(x)|} \leq M$$

where

$$M = M(k) = \left(\frac{k+1}{\nu + 1} (1+\eta) - \eta \right) \frac{\left(1 - \frac{1+k}{1+\nu} (1+\eta) \right)^{1+\eta}}{1 - \left(1 - \frac{1+k}{1+\nu} (1+\eta) \right)^{1+\eta}}. \tag{8}$$

We observe however that M is not uniform in k. More precisely, assuming that η is small enough so that $M(k) > 0$ for all k, we have $\lim_{k \to \nu} M(k) = 0$. A proof of this remark will be given in Section 3.

An immediate corollary of Theorem 1 is the following characterisation of the order of polynomial convergence of the one-dimensional Zig-Zag on Student distributions. For the following lower bound on the total variation distance from the invariant measure

we use a type of argument similar to the proof of Theorem 2.1 in [23], suggested to us by Professor Anthony Lee in private communication.

Corollary 1. *Let* $\nu > 0$ *and suppose* π *is a Student distribution with* ν *degrees of freedom, i.e.*

$$\pi(x) = \frac{1}{Z} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} \tag{9}$$

and the Zig-Zag process targets μ *as in (2), having refresh rate that satisfies Assumption 2. For all* $k < \nu$, *there exist* $\beta \in (0,1)$ *and* $\delta, B > 0$ *such that for all* $(x, \theta) \in \mathbb{R} \times \{-1, +1\}$,

$$\|\mathbb{P}_{x,\theta}(Z_t \in \cdot) - \mu(\cdot)\|_{TV} \leq \frac{B V_{\beta,\delta}(x)}{t^{1+k}} + \frac{B}{t^k},$$

i.e. for all $k < \nu$, *the process is polynomially ergodic of order* k. *Furthermore, for any* $k > \nu$, *the process is not polynomially ergodic of order* k. *More specifically, there exists a constant* C' *such that for all* $t > 0$ *large enough*

$$\|\mathbb{P}_{0,+1}(Z_t \in \cdot) - \mu(\cdot)\|_{TV} \geq \frac{C'}{t^\nu}.$$

Corollary 1 illustrates that Theorem 1 is close to being tight. At least for Student distributions with ν degrees of freedom, it is established that the process is polynomially ergodic for order $k < \nu$ and it is not for any order $k > \nu$. Whether the process is polynomially ergodic of order ν is still an open question, but probably of limited practical importance.

3. Proofs

Before we prove Theorem 1 we recall the form of the strong generator of the Zig-Zag process.

Definition 2. We define $\mathcal{L}$ to be the operator acting on any $f \in C^1(E)$ so that for all $(x, \theta) \in E$,

$$\mathcal{L}f(x,\theta) = \theta f'(x,\theta) + \left([\theta U'(x)]^+ + \gamma(x)\right)(f(x,-\theta) - f(x,\theta)). \tag{10}$$

It can be proven (see for example [11]) that $\mathcal{L}$ is the restriction on C^1 of the strong generator of $(Z_t)_{t \geq 0}$.

General conditions for polynomial ergodicity can be found in Theorems 3.2 and 3.4 of [13] and Theorem 1.2 of [3] (see also Corollary 6 in [15] for some earlier results on sub-geometric convergence). Here we use a result found and proved in the presented form in the unpublished lecture notes [16].

Theorem 2. (Hairer [16] Theorem 4.1.) *Let* $(X_t)_{t\geq 0}$ *a continuous time Markov process on* X *with strong generator* $\mathcal{L}$. *Suppose that there exists a function* $V : X \to [1, +\infty)$ *and a constant* K *such that for all* $x \in X$

$$\mathcal{L}V(x) \leq K - f(V) \tag{11}$$

for a function $f : [0, +\infty) \to [0, +\infty)$ *strictly concave, increasing, with* $f(0) = 0$, $\lim_{s\to +\infty} f(s) = +\infty$. *Suppose further that all the sub-level sets of* V *are pre-compact and small. Then, the following hold:*

1. *There exists a unique invariant measure* μ *for the process such that* $\int f(V(x))\mu(dx) < \infty$.

2. *Let* $H_f(u) = \int_1^u 1/f(s)ds$, *then there exists a constant* $B > 0$ *such that for every* $x \in X$
$$\|\mathbb{P}_x(X_t \in \cdot) - \mu(\cdot)\|_{TV} \leq \frac{BV(x)}{H_f^{-1}(t)} + \frac{B}{f \circ H_f^{-1}(t)}. \tag{12}$$

Proof of Theorem 1. Suppose that U satisfies Assumption 1 and let $k < \nu$. Select a such that $k = a/(1 - a)$ so that $a < 1 - 1/(1 + \nu)$. For any $\tilde{\beta} \in (0, 1)$ we have that there exists a $c_0 > 0$ such that for all $x \notin C$

$$\left(V_{\tilde{\beta},\delta}(x, \theta)\right)^{1-a} |U'(x)| \geq c_0 \exp\{\tilde{\beta}(1 - a)(1 + \nu)\log|x|\}\frac{1 + \nu}{|x|} = c_0(1 + \nu)|x|^{\tilde{\beta}(1-a)(1+\nu)-1}$$

Since $(1-a)(1+\nu)-1 > 0$, there exists a β close to 1 such that $\beta(1-a)(1+\nu)-1 > 0$, so

$$\lim_{|x|\to\infty} V_\beta^{1-a}(x, \theta)|U'(x)| = +\infty. \tag{13}$$

Now, $V_{\beta,\delta} \in C^1$ and $\lim_{|x|\to\infty} V_{\beta,\delta}(x, \theta) = +\infty$ so all the level sets are compact. Since the process is positive Harris recurrent and some skeleton is irreducible (see [9]) we get from Proposition 6.1 of [20] that the level sets are also small. Since $\lim_{|x|\to\infty} U(x) = +\infty$ it is clear that $V_{\beta,\delta}$ is bounded below away from 0 so by multiplying with an

appropriate constant we can assume that $V_{\beta,\delta}(x,\theta) \geq 1$ for all (x,θ). We calculate

$$\mathcal{L}V_{\beta,\delta}(x,\theta) = V_{\beta,\delta}(x,\theta)\left(\theta\beta U'(x) + \left([\theta U'(x)]^+ + \gamma(x)\right)\left(\exp\{-2\theta\operatorname{sgn}(x)\delta\} - 1\right)\right).$$

Note that due to Assumption 1 and that $U(x) \xrightarrow{|x|\to\infty} +\infty$, we get that there exists a compact set $\tilde{C}$ such that for all $x \notin \tilde{C}$ $\operatorname{sgn}(U'(x)) = \operatorname{sgn}(x)$. Therefore, when $x \notin \tilde{C}$ and $\theta \operatorname{sgn}(x) > 0$,

$$\frac{\mathcal{L}V_{\beta,\delta}(x,\theta)}{V_{\beta,\delta}^a(x,\theta)} \leq V_{\beta,\delta}^{1-a}(x,\theta)|U'(x)|\left[\beta + \left(\frac{\gamma(x)}{|U'(x)|} + 1\right)(\exp\{-2\delta\} - 1)\right]$$

and when $\theta \operatorname{sgn}(x) < 0$

$$\frac{\mathcal{L}V_{\beta,\delta}(x,\theta)}{V_{\beta,\delta}^a(x,\theta)} \leq V_{\beta,\delta}^{1-a}(x,\theta)|U'(x)|\left[-\beta + \frac{\gamma(x)}{|U'(x)|}(\exp\{2\delta\} - 1)\right].$$

Overall we have for $x \notin \tilde{C}$

$$\frac{\mathcal{L}V_{\beta,\delta}(x,\theta)}{V_{\beta,\delta}^a(x,\theta)} \leq V_{\beta,\delta}^{1-a}(x,\theta)|U'(x)|\max\left\{\beta + \left(\frac{\gamma(x)}{|U'(x)|} + 1\right)(\exp\{-2\delta\} - 1), \left[-\beta + \frac{\gamma(x)}{|U'(x)|}(\exp\{2\delta\} - 1)\right]\right\}.$$

$$(14)$$

Recall that $\frac{\gamma(x)}{|U'(x)|} \xrightarrow{|x|\to\infty} 0$. Fix $\delta > -1/2\log(1-\beta)$ and by possibly increasing $\tilde{C}$, we have that there exists a constant $c' > 0$ such that for all $x \notin \tilde{C}$

$$\max\left\{\beta + \left(\frac{\gamma(x)}{|U'(x)|} + 1\right)(\exp\{-2\delta\} - 1), -\beta + \frac{\gamma(x)}{|U'(x)|}(\exp\{2\delta\} - 1)\right\} < -c' < 0.$$

Combining this with (13) and (14) we get that $V_{\beta,\delta}$ satisfies (11) with $f(u) = cu^a$ for $c = c'/2$ and with K being an appropriate constant that bounds the continuous function $\mathcal{L}V_{\beta,\delta} + f(V_{\beta,\delta})$ inside $\tilde{C}$.

Therefore, all the conditions of Theorem 2 are satisfied. Note that $H_f(s) = \int_1^s 1/f(u)du = c^{-1}\int_1^s u^{-a}du = \frac{1}{c(1-a)}(s^{1-a} - 1)$ so

$$H_f^{-1}(t) = (1 + c(1-a)t)^{1/(1-a)}$$

and therefore

$$f \circ H_f^{-1}(t) = c\left(1 + c(1-a)t\right)^{a/(1-a)}$$

Since we picked a such that $k = a/(1-a)$, meaning that $k+1 = 1/(1-a)$, (7) follows.

$\square$

Proof of Remark 2. Fix $k < \nu$ and set a such that $k = a/(1-a)$, therefore $1 - a = 1/(1+k)$. Our goal is to find an appropriate upper bound on $\frac{\gamma(x)}{|U'(x)|}$ that will guarantee that the RHS of (14) is negative for appropriately chosen β and δ. For this it suffices to prove that for some $\beta, \delta > 0$, the maximum appearing in the RHS of (14) is negative, bounded away from zero and that (13) holds. These two conditions will guarantee that the drift condition (11) holds for $V_{\beta,\delta}$ and we can conclude similar to the proof of Theorem 1. Since from the proof of that theorem

$$\left(V_{\tilde{\beta},\delta}(x,\theta)\right)^{1-a} |U'(x)| \geq c_0(1+\nu)\,|x|^{\tilde{\beta}(1-a)(1+\nu)-1}\,,$$

in order to guarantee (13), it suffices to pick $\beta = (1+\eta)(1+k)/(1+\nu)$ for some $\eta > 0$. From the discussion after (14), we can pick $\delta = -\frac{1}{2}(1+\eta)\log(1-\beta) = -\frac{1}{2}(1+\eta)\log\left(1 - (1+\eta)(1+k)/(1+\nu)\right)$. With this choice of β, δ the first part of the maximum of the RHS in (14) will be negative, bounded away from zero. The second part of the maximum writes

$$-\beta + \frac{\gamma(x)}{|U'(x)|}(\exp\{2\delta\}-1) = -(1+\eta)\frac{1+k}{1+\nu} + \frac{\gamma(x)}{|U'(x)|}\left(\left(\frac{1}{(1-(1+\eta)(1+k)/(1+\nu))}\right)^{(1+\eta)} - 1\right).$$

Therefore, solving the inequality $-\beta + \frac{\gamma(x)}{|U'(x)|}(\exp\{2\delta\} - 1) < -\eta$ (which would guarantee that the maximum on the RHS of (14) is negative, bounded away from zero), we get

$$\frac{\gamma(x)}{|U'(x)|} \leq M(k)$$

where $M(k)$ as in (8). $\qquad\square$

Proof of Corollary 1. Let π as in (9). Then for all $\delta' > 0$ there exists a compact set C with $0 \in C$, such that for all $x \notin C$

$$|U'(x)| = \frac{(\nu+1)|x|}{\nu + x^2} \geq \frac{1+\nu-\delta'}{|x|}$$

Therefore for every $\delta' > 0$, the distribution satisfies Assumption 1 where the ν in that assumption is equal to $\nu - \delta'$. From Proposition 1, for all $k < \nu$, there exists a $\beta \in (0,1), \delta > 0$ and $B > 0$ such that for all $(x,\theta) \in \mathbb{R} \times \{-1,+1\}$,

$$\|\mathbb{P}_{x,\theta}(Z_t \in \cdot) - \mu(\cdot)\|_{TV} \leq \frac{BV_{\beta,\delta}(x)}{t^{1+k}} + \frac{B}{t^k}.$$

Now, suppose that the Zig-Zag starts from $x = 0$, $\theta = +1$. There exists a C_0 and $K > 0$ such that for all $|x| \geq K$, $\pi(x) \geq C_0|x|^{-\nu-1}$. Fix a time $t > K$. Let $A_t = \{x : x > t\}$. After less or equal to time t has passed, the Zig-Zag will not have hit A_t. We therefore get for all $t > K$,

$$\|\mathbb{P}_{0,+1}(Z_t \in \cdot) - \mu(\cdot)\|_{TV} \geq |\mathbb{P}_{0,+1}(X_t \in A_t^c) - \pi(A_t^c)| = \pi(A_t^c) = \int_{|x|>t} \pi(x)dx \geq$$

$$\geq 2C_0 \int_t^{+\infty} x^{-\nu-1}dx = \frac{C_0}{\nu}\frac{1}{t^\nu}.$$

$\square$

4. Discussion

It is interesting to compare the Zig-Zag polynomial convergence rates with the ones of the one-dimensional Random Walk Metropolis (RWM) and the Metropolis-adjusted Langevin algorithm (MALA) algorithms. In fact, it is shown in [18], Propositions 4.1 and 4.3 (see also [17]) that when targeting a Student distribution with ν degrees of freedom with any finite variance proposal RWM or with MALA, one gets polynomial order of convergence $(\nu/2)^-$, i.e. for any $\epsilon > 0$, the polynomial rate of convergence is at least $\nu/2 - \epsilon$. It is however proven not to be $\nu/2$. In [17] the authors provide these lower bounds for the convergence rates, while in [18] they provide the upper bound. As proven in this note, the one-dimensional Zig-Zag has polynomial rate of convergence ν^- in the same setting, which is better than RWM or MALA. This phenomenon was also observed in simulations in [4]. We conjecture that the advantage of the Zig-Zag is due its momentum, which, in a one-dimensional, uni-modal setting with zero refresh rate, will force the process to never switch direction before it hits the mode. This diminishes any possible diffusive behaviour of the process at the tails and helps the algorithm converge faster. We should note here that better polynomial rates (and more precisely, arbitrarily better rates) can be achieved for the Random Walk Metropolis (RWM) if one introduces a proposal with heavier tails. However, a natural analogue of this modification is to allow the Zig-Zag to speed up and move faster in areas of lower density. This idea is further discussed in [23] and proven to be able to provide exponentially ergodic algorithms even on heavy tailed targets, which can outperform the simple Zig-Zag in the sense of having better effective sample size per number of

likelihood evaluations.

Acknowledgements

The authors would like to acknowledge Professor Anthony Lee for an indication of the proof of the lower bound of the total variation distance in Corollary 2.5. They would also like to acknowledge Andrea Bertazzi, George Deligiannidis and Krzysztof Latuszyński for helpful discussions. Furthermore, they would like to thank the anonymous referee and the associated editor for the helpful comments that improved this note.

Funding information

Competing interests

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

References

[1] ANDRIEU, C., DOBSON, P. AND WANG, A. Q. (2021). Subgeometric hypocoercivity for piecewise-deterministic Markov process Monte Carlo methods. *Electron. J. Probab.: 1-26*. 10.1214/21-EJP643.

[2] ANDRIEU, C., DURMUS, A., NÜSKEN, N. AND ROUSSEL, J. (2021). Hypocoercivity of piecewise deterministic Markov Process-Monte carlo. *To appear in Annals of Applied Probability*.

[3] BAKRY, D., CATTIAUX, P. AND GUILLIN, A. (2008). Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *Journal of Functional Analysis* **254,** 727 – 759.

[4] BIERKENS, J. AND DUNCAN, A. (2017). Limit theorems for the zig-zag process. *Advances in Applied Probability* **49,** 791–825.

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

[6] BIERKENS, J., KAMATANI, K. AND ROBERTS, G. High-dimensional scaling limits of piecewise deterministic sampling algorithms 2018. Available at https://arxiv.org/abs/1807.11358.

[7] BIERKENS, J., NYQUIST, P. AND C., S. M. Large deviations for the empirical measure of the zig-zag process 2019. To appear in Annals of Applied Probability. Available at https://arxiv.org/abs/1912.06635.

[8] BIERKENS, J. AND ROBERTS, G. O. (2017). A piecewise deterministic scaling limit of lifted metropolis–hastings in the curie–weiss model. *The Annals of Applied Probability* **27,** 846–882.

[9] BIERKENS, J., ROBERTS, G. O. AND ZITT, P. A. (2019). Ergodicity of the zigzag process. *Ann. Appl. Probab.* **29,** 2266–2301.

[10] BIERKENS, J. AND VERDUYN LUNEL, S. M. Spectral analysis of the zigzag process 2019. To appear in Annales de l'Institut Henri Poincaré (B) Probabilitès et Statistiques. Available at https://arxiv.org/abs/1905.01691.

[11] DAVIS, M. (2018). *Markov Models and Optimization.* Chapman and Hall/CRC Monographs on Statistics and Applied Probability. Routledge.

[12] DIACONIS, P., HOLMES, S. AND NEAL, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.* **10,** 726–752.

[13] DOUC, R., FORT, G. AND GUILLIN, A. (2009). Subgeometric rates of convergence of f-ergodic strong markov processes. *Stochastic Processes and their Applications* **119,** 897 – 923.

[14] FEARNHEAD, P., BIERKENS, J., POLLOCK, M. AND ROBERTS, G. O. (2018). Piecewise deterministic Markov processes for continuous-time Monte Carlo. *Statist. Sci.* **33,** 386–412.

[15] FORT, G. AND ROBERTS, G. O. (2005). Subgeometric ergodicity of strong Markov processes. *Ann. Appl. Probab.* **15,** 1565–1589.

[16] HAIRER, M. (2016). Convergence of Markov Processes. *Unpublished Lecture Notes*. Available at http://www.hairer.org/Teaching.html.

[17] JARNER, S. AND ROBERTS, G. O. (2007). Convergence of Heavy-tailed Monte Carlo Markov Chain Algorithms. *Scandinavian Journal of Statistics* **34,** 781–815.

[18] JARNER, S. F. AND TWEEDIE, R. L. (2003). Necessary conditions for geometric and polynomial ergodicity of random-walk-type. *Bernoulli* **9,** 559–578.

[19] LEVIN, D., LUCZAK, M. AND PERES, Y. (2007). Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. *Probability Theory and Related Fields* **146,** 223–265.

[20] MEYN, S. P. AND TWEEDIE, R. L. (1993). Stability of Markovian Processes II: Continuous-Time Processes and Sampled Chains. *Advances in Applied Probability* **25,** 487–517.

[21] TURITSYN, K. S., CHERTKOV, M. AND VUCELJA, M. (2011). Irreversible Monte Carlo algorithms for efficient sampling. *Physica D: Nonlinear Phenomena* **240,** 410 – 414.

[22] VANETTI, P., BOUCHARD-CÔTÉ, A., DELIGIANNIDIS, G. AND DOUCET, A. Piecewise-Deterministic Markov Chain Monte Carlo 2017. Available at https://arxiv.org/abs/1707.05296.

[23] VASDEKIS, G. AND ROBERTS, G. O. Speed Up Zig-Zag 2021. Available at https://arxiv.org/abs/2103.16620.