SPEED UP ZIG-ZAG

BY G. VASDEKIS[1,a] AND G. O. ROBERTS[2,b]

[1]*Department of Statistics, University College London,* [a]*G.Vasdekis@ucl.ac.uk*

[2]*Department of Statistics, University of Warwick,* [b]*Gareth.O.Roberts@warwick.ac.uk*

The Zig-Zag process is a piecewise deterministic Markov process, efficiently used for simulation in an MCMC setting. As we show in this article, it fails to be exponentially ergodic on heavy tailed target distributions. We introduce an extension of the Zig-Zag process by allowing the process to move with a nonconstant speed function *s*, depending on the current state of the process. We call this process Speed Up Zig-Zag (SUZZ). We provide conditions that guarantee stability properties for the SUZZ process, including nonexplosivity, exponential ergodicity in heavy tailed targets and central limit theorem. Interestingly, we find that using speed functions that induce explosive deterministic dynamics may lead to stable algorithms that can even mix faster. We further discuss the choice of an efficient speed function by providing an efficiency criterion for the one-dimensional process and we support our findings with simulation results.

CONTENTS

1. Introduction. Piecewise deterministic Markov processes (PDMP) have recently emerged as a new way to construct MCMC algorithms. Traditional MCMC algorithms employ discrete time Markov chains to generate samples from a target distribution which is invariant for the chain, and subsequently use these samples to numerically estimate intractable expectations of functions of interest. By construction, standard MCMC algorithms like Random Walk Metropolis [40], MALA [2], etc. are time-reversible with respect to their target distribution. However, there is by now substantial evidence that reversible MCMC methods can be significantly outperformed (in terms of mixing times and variances of estimators) by nonreversible ones (see, e.g., [3, 15, 19, 20, 29, 35, 47]). Some PDMPs such as the Bouncy Particle Sampler [13] and the Zig-Zag sampler [6] can be implemented directly and free from numerical error, providing a source of genuinely nonreversible MCMC algorithms.

The one-dimensional Zig-Zag algorithm appeared in [10] as a scaling limit of the lifted Metropolis–Hastings (see [19, 49]) applied to the Curie–Weiss model (see [36]), although a simpler version of the process was introduced in [28] as the telegraph process (see also [25, 26, 32]). The process was later extended in higher dimensions in [6] and has been proposed as a PDMP which can be used as an MCMC algorithm to target posterior distributions (see also [23, 51]). In [6], the authors also introduce some variants of the algorithm that use the technique of subsampling, improving computational efficiency when the target distribution is obtained from a Bayesian analysis involving a large data set. Further literature on the topic includes [4, 5, 7–9, 12].

[11] proves ergodicity and exponential ergodicity of the Zig-Zag process in arbitrary dimension. A crucial assumption required for exponential ergodicity in that work is that the target density has exponential or lighter tails. This paper will demonstrate the converse: the Zig-Zag sampler fails to be exponentially ergodic when the target distribution has tails thicker than any exponential distribution, that is, it is a heavy tailed. In fact, polynomial rates of convergence have been proven in [1] for the process in arbitrary dimension, while [53] proves tight polynomial rates of convergence in the total variation distance, for the one-dimensional process, when the target has tails that decay like a Student distribution.

In order to address the problem of slow mixing on heavy tails, we introduce a variant of the Zig-Zag process, called Speed Up Zig-Zag (SUZZ). The idea behind the process has a similar spirit to the work of [38] and [44]. In our case, instead of only permitting the process to move with unit speed, we allow it to have a positive position-dependant speed. This assists the exploration of the tails and subsequent return to the high density areas of the distribution more rapidly. We note that if the speed function is large enough, the solution to the ODE that governs the behaviour of the SUZZ process may potentially explode in finite time. A large part of the theory in this article focuses on proving that such dynamics are mathematically acceptable in the context of MCMC. Furthermore, for carefully chosen speed functions, these ODEs and the induced SUZZ process can be numerically simulated exactly. Although explosive deterministic dynamics have been mentioned in the past (see, e.g., Example 2.1.3 of [45]), to the best of our knowledge, this is the first use of explosive dynamics within the literature of PDMPs for MCMC.

The rest of this paper is organised as follows. In Section 2 we recall the definition of the Zig-Zag process and we prove its lack of exponential ergodicity on heavy tails. Motivated by this slow convergence result, in Section 3 we define the Speed Up Zig-Zag (SUZZ) process and we establish stability and convergence properties. Theorem 3.1 proves that under certain conditions on the speed function, the process is nonexplosive. Theorem 3.2 proves that the process has the distribution of interest as invariant. Theorem 3.3 proves that the process is exponentially ergodic and Theorem 3.4 that it satisfies a central limit theorem. Theorem 3.5 proves that when the target has light tails, the SUZZ process is exponentially ergodic, essentially under the same conditions as in the original Zig-Zag, while Proposition 3.1 proves exponential ergodicity of the SUZZ process for a family of heavy tailed distributions, with some

specific, practical choices of speed functions. Corollary 3.1 proves exponential ergodicity of the original Zig-Zag on light tailed targets, relaxing the assumptions of [11]. Furthermore, focusing on the one-dimensional SUZZ process, Theorem 3.6 proves that under explosive deterministic dynamics the process is uniformly ergodic and provides weaker assumptions to prove that the process is exponentially ergodic. In Section 4 we focus on the one-dimensional process and discuss how the choice of the speed function can improve algorithmic efficiency. Using Proposition 4.1 we write the asymptotic variance of the one-dimensional process as a function of the speed function, which allows us to introduce a minimisation problem characterising the optimal speed function for one-dimensional SUZZ within an MCMC context. Finally in Section 5 we describe some numerical results, comparing the efficiency of different algorithms on one-dimensional and twenty-dimensional distributions. The Appendices contain the proofs of the main results along with some other useful information (e.g., how to formally construct the process or how to solve the deterministic ODE and construct the deterministic paths of the process).

2. The original Zig-Zag process.

Here we give a brief introduction to the Zig-Zag process (which in this article we will refer to as original Zig-Zag), recalling some basic properties and proving it has a subexponential convergence rate for heavy-tailed distributions. The d-dimensional original Zig-Zag process $(Z_t)_{t \geq 0} = ((X_t, \Theta_t))_{t \geq 0}$ is a PDMP with state space $E = \mathbb{R}^d \times \{\pm 1\}^d$. One can think of the process as a particle moving in $\mathbb{R}^d$ along one of 2^d possible straight lines. When the process is at point $(x, \theta) \in E$ the particle is at point $x \in \mathbb{R}^d$ and moves with constant velocity $\theta \in \{\pm 1\}^d$. This means that the process moves according to the ODE

$$(1) \qquad \begin{cases} \dfrac{dX_t}{dt} = \theta, \\ \dfrac{d\Theta}{dt} = 0. \end{cases}$$

To each of the d coordinates, we let T_i denote the first event of a nonhomogeneous Poisson process of rate $m_i(t) = \lambda_i^{ZZ}(x + \theta t, \theta)$, for $i = 1, \ldots, d$ and for some function $\lambda_i^{ZZ} : E \to [0, +\infty)$. We assume that all m_i are locally integrable. Let $T = \min_{i \in \{1, \ldots, d\}} T_i$ and $j = \arg \min_{i \in \{1, \ldots, d\}}\{T_i\}$. The process moves with velocity θ until time T at which time its velocity changes to $F_j(\theta) \in \{\pm 1\}^d$, where

$$(2) \qquad \begin{cases} F_j(\theta)_j = -\theta_j, \\ F_j(\theta)_i = \theta_i \qquad \text{for } i \neq j, \end{cases}$$

and proceeds to move again with constant velocity $F_j(\theta)$ until it switches again, etc. (see Figure 1 for traceplots of the process).

In [6] the goal was to target a probability measure on E of the form

$$(3) \qquad \mu(dx, d\theta) = \frac{1}{2^d H} \exp\{-U(x)\} \, dx \, d\theta$$

for some $U \in C^1$ with $H = \int_{\mathbb{R}^d} \exp\{-U(x)\} \, dx < \infty$. It is proven that the original Zig-Zag process has μ as invariant distribution when the rate functions are chosen according to

$$(4) \qquad \lambda_i^{ZZ}(x, \theta) = [\theta_i \partial_i U(x)]^+ + \gamma_i(x, \theta_{-i}),$$

where we write ∂_i to denote the operator of the partial derivative on the i coordinate, $a^+ = \max\{a, 0\}$ and γ_i is a nonnegative function that does not depend on the i component of θ. The special case where $\gamma_i(x, \theta_{-i}) = 0$ for all x, i is known as the *canonical Zig-Zag*.

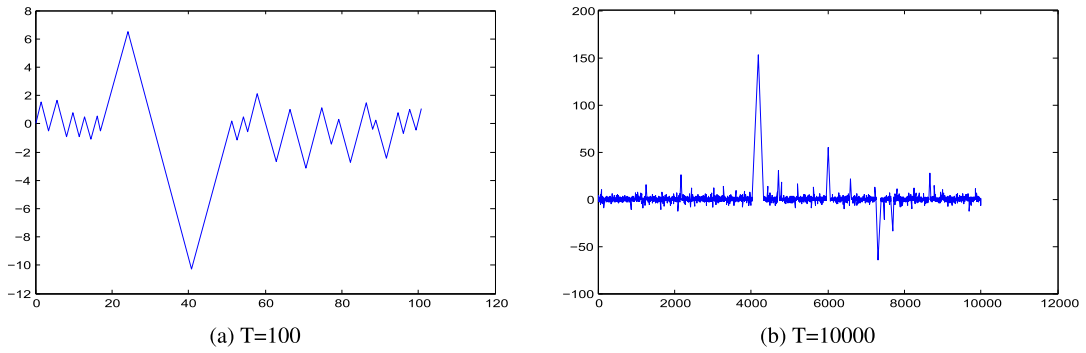(a) T=100   (b) T=10000

FIG. 1.   *Trace plots of original Zig-Zag process, ran until time T, targeting one-dimensional Cauchy distribution.*

REMARK 2.1.   We note that in many MCMC applications the goal is to target a measure

$$(5) \qquad\qquad \pi(dx) = 1/H \exp\{-U(x)\}\, dx$$

in $\mathbb{R}^d$. Technically, the original Zig-Zag process targets a measure μ on $E = \mathbb{R}^d \times \{-1, +1\}^d$, whose marginal distribution on $\mathbb{R}^d$ is π and whose marginal distribution on $\{-1, +1\}^d$ is the uniform. One can then use the projection of the process on $\mathbb{R}^d$ to generate samples from the measure of interest π. Throughout this work we shall denote π the measure of interest in $\mathbb{R}^d$ and μ the measure on E given by (3).

Bierkens Roberts and Zitt [11] demonstrates that assuming that U grows at least linearly in the tails and appropriate smoothness conditions hold, the original Zig-Zag process $(Z_t)_{t\geq0}$ converges to μ exponentially fast, that is, there exist $M : E \to [1, +\infty)$ and $\rho < 1$ such that for any $(x, \theta) \in E$

$$(6) \qquad\qquad \left\| \mathbb{P}_{x,\theta}(Z_t \in \cdot) - \mu(\cdot) \right\|_{\mathrm{TV}} \leq M(x, \theta)\rho^t.$$

If (6) holds, we say that the process is exponentially ergodic.

However, [11] does not cover the case where U grows sublinearly. In this scenario traditional MCMC algorithms based on random walk or Langevin proposals are known to converge at subexponential rates; see, for example, [30, 39, 46]. We will observe similar behaviour for the original Zig-Zag sampler. Figure 2 provides the Q-Q plots of one-dimensional

(a) Student(1) distribution   (b) Student(8) distribution   (c) Student(500) distribution
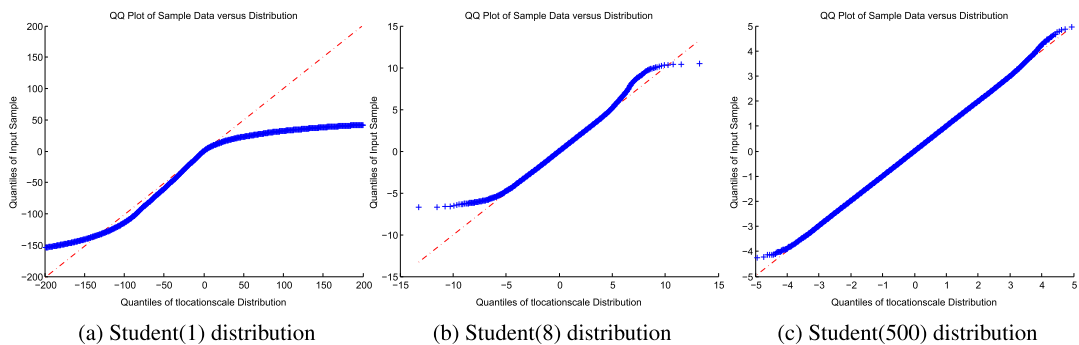
FIG. 2.   *Q-Q plots of canonical Zig-Zag for three Student distributions with increasing degrees of freedom. With Student(ν) we denote the Student distribution with ν degrees of freedom. Each algorithm runs until $N = 10^4$ switches of direction occur. The closer the curve is to the diagonal line, the better the algorithm approximates the distribution. Clearly, the approximation is better when the target has higher degrees of freedom, that is, lighter tails.*

canonical Zig-Zag processes, targeting Student distributions with three different degrees of freedom. Each algorithm runs until $N = 10^4$ switches of direction occur. The figure indicates that the process is less stable when targeting a Student distribution with lower degrees of freedom, which has more mass at the tails. This instability is characterised by infrequent and unstable (heavy-tailed) excursions. In this article we will be mainly dealing with distributions that assign more mass at the tails than any exponential distribution. We give the following definition.

DEFINITION 2.1. We say that a measure π on $\mathbb{R}^d$ is heavy-tailed if for any $a > 0$, if $B(0, R)$ is the ball of radius R, centred at 0, then

$$(7) \qquad \lim_{R \to +\infty} \pi\big(B(0, R)^c\big) \exp\{aR\} = +\infty.$$

The following simple negative result for the original Zig-Zag sampler on heavy-tailed distributions was made known to us in personal correspondence with Professor Anthony Lee.

THEOREM 2.1 (Nonexponential ergodicity). *Suppose that the original Zig-Zag targets a heavy tailed distribution. Then the process is not exponentially ergodic.*

PROOF OF THEOREM 2.1. Suppose that the original Zig-Zag starts from $x = 0$, $\theta \in \{-1, +1\}^d$. For any $t > 0$, let $A_t = \{x : \|x\|_2 > t\}$ be the complement of the ball of radius t and let us fix a time $t > 0$. Note that the process will always move with some velocity $\eta \in \{-1, +1\}^d$, and note that for any such η we have $\|\eta\|_2 = \sqrt{d}$. Since the process moves with constant speed equal to $\sqrt{d}$, the original Zig-Zag will not have hit $A_{\sqrt{d}t}$ by time t. Therefore,

$$\big\|\mathbb{P}_{0,\theta}(X_t \in \cdot) - \pi(\cdot)\big\|_{\mathrm{TV}} \ge \big|\mathbb{P}_{0,\theta}(X_t \in A_{\sqrt{d}t}) - \pi(A_{\sqrt{d}t})\big| = \pi(A_{\sqrt{d}t}) = \pi\big(B(0, \sqrt{d}t)^c\big).$$

So if we were to have exponential ergodicity, we would have that there exists $M > 0$ and $\rho < 1$ such that for all $t > 0$,

$$\pi\big(B(0, t)^c\big) \le \big\|\mathbb{P}_{0,\theta}(X_{\sqrt{d}^{-1}t} \in \cdot) - \pi(\cdot)\big\|_{\mathrm{TV}} \le M\big(\rho^{\sqrt{d}^{-1}}\big)^t,$$

which creates a contradiction as π has heavy tails. $\square$

Note that the same proof can be used for any other algorithm with constant speed function such as the Bouncy Particle Sampler with refreshing velocities taken from the unit sphere. Since these processes move with constant speed, they will not be able to explore the tails of the distribution sufficiently, which will result in a bad estimation of a target distribution that has heavy tails. Constant velocities are used mainly for the simplicity of the deterministic paths they provide. However, there are other types of deterministic paths that we can simulate exactly which do not move with unit speed. This raises the question why not allow the original Zig-Zag process to move with nonconstant velocities. We introduce this algorithm in the following section.

3. Speed Up Zig-Zag.

3.1. *Definition of the algorithm.* In order to address the problem of slow mixing on heavy tails we introduce a variant of the original Zig-Zag process. Instead of allowing the process to move with unit velocity, we allow it to have a positive speed depending on the current position. Since in high dimensions this might create a system of ODEs that is nonimplementable, we only allow the process to move in directions $\{\pm 1\}^d$ as the original Zig-Zag does.

The state space will, again, be $E = \mathbb{R}^d \times \{\pm 1\}^d$. However, when the process is at point $(x, \theta) \in E$, it will move along the path $\{x + \theta t, t \geq 0\}$ with speed function s that depends on the current position. Typically, this speed will increase the further the process is from the mode. After a random time that will depend on a Poisson process, as in the original Zig-Zag, the process will stop, one of the coordinates of θ will switch sign and the process will start moving again in the new direction. This will create excursions that tend to leave the area of high density and visit the tails quite often. At the same time, when the process is at the tails of the distribution, it can speed up and return to the centre fast enough. We shall call this process *Speed Up Zig-Zag (SUZZ)*, with state space $E \cup \{\partial\}$ (where ∂ is a graveyard state that is needed for technical reasons), with C^2 *speed function* $s : \mathbb{R}^d \to (0, \infty)$ and *rate functions* $\lambda_i : E \to [0, +\infty)$ for all $i \in \{1, \ldots, d\}$.

The SUZZ process starting from $(x, \theta) \in E$ evolves as follows. Consider the following ODE system:

(8)
$$\begin{cases} \dfrac{dX_t}{dt} = \theta s(X_t), & X_0 = x, \\ \dfrac{d\Theta_t}{dt} = 0, & \Theta_0 = \theta. \end{cases}$$

The procedure to solve (8) will be given in Appendix B. When the speed function is superlinear, the solution to (8) explodes in finite time $t^*(x, \theta)$. Let (X_t, Θ_t) denote the solution to (8) until time $t^*(x, \theta) \in (0, +\infty]$. For each coordinate $i \in \{1, \ldots, d\}$, we let T_1^i denote the first event of a nonhomogeneous Poisson process of rate $m_i(t) = \lambda_i(X_t, \theta)$, for $i = 1, \ldots, d$. Let $T_1 = \min_{i \in \{1, \ldots, d\}} T_1^i$ and $j = \arg\min_{i \in \{1, \ldots, d\}} \{T_1^i\}$. The SUZZ process is defined until time T_1 to be the solution $(X_t, \Theta_t)_{t \leq T_1}$ to (8). At time T_1 the direction/velocity Θ_{T_1} of the process switches from θ to $F_j(\theta)$, as in (2). In the case where $T_1 > t^*(x, \theta)$, the process is defined as the solution to (8) until $t^*(x, \theta)$ and then it moves to the graveyard state ∂. If $T_1 < t^*(x, \theta)$, the process starts again from the new starting point $(X_{T_1}, F_j[\theta])$ and evolves as before until time T_2 when the velocity switches again. Then the process starts again from the new position etc. This inductively defines the process until time

(9)
$$\xi = \lim_{n \to +\infty} T_n.$$

In the case where $\xi < \infty$, ξ is the first time that the process has had infinitely many switches of direction and the process moves to the graveyard state at time ξ.

A different way to describe the process is through its generator. We will later see that the class of functions with compact support and continuous first derivative, C_c^1, is contained in the domain of the strong generator of the SUZZ process and for any $f \in C_c^1$ the strong generator is given by

$$\mathcal{L}f(x, \theta) := \sum_{i=1}^{d} \theta_i s(x) \partial_i f(x, \theta) + \lambda_i(x, \theta)\big(f(x, F_i(\theta)) - f(x, \theta)\big).$$

The process is therefore defined as a piecewise deterministic process in [18] would be. The difference is that we allow the deterministic dynamics to have a finite explosion time, which Davis in [18] does not. We therefore need to be more careful in the analysis of the process. Let O_m be the ball of radius m centred around the origin 0. We define

(10)
$$\zeta_m = \inf\{t \geq 0 : X_t \notin O_m\}$$

and let

(11)
$$\zeta = \lim_{m \to \infty} \zeta_m.$$

The two random variables ξ and ζ quantify two types of explosion that can occur for the process. The first is that the process could have infinitely many switches in finite time and the second, that the process might diverge to infinity in finite time.

An algorithmic description of the process is given in Section 1 of the Supplementary Material (Vasdekis and Roberts (2023), [54]) and a more formal construction of the process is given in Appendix A.

3.2. *Stability and convergence properties.* In this section we will study the process in more detail and provide some convergence results to the distribution of interest. We will ultimately provide assumptions that ensure exponential ergodicity of the process.

Similar to the original Zig-Zag case, we will assume that we are trying to target the measure μ introduced in (3) using a SUZZ with speed function s. Throughout this article we will assume that the rates are of the form

$$(12) \qquad \lambda_i(x, \theta) = [\theta_i A_i(x)]^+ + \gamma_i(x, \theta_{-i}),$$

where γ_i is a nonnegative, locally bounded, integrable function that does not depend on the ith component of θ and

$$(13) \qquad A_i(x) = s(x)\partial_i U(x) - \partial_i s(x).$$

Note that if we use the constant speed function $s \equiv 1$, we retrieve the original Zig-Zag rates when targeting μ. We will later prove that if the rates satisfy (12) and some extra regularity conditions hold, the SUZZ process leaves the measure μ in (3) invariant.

Before we focus more on whether the SUZZ process targets the right distribution, we first need to consider some explosivity issues the process might have. Even in one dimension, picking a large enough speed function s can lead to deterministic dynamics that explode in finite time. On a first glance, following deterministic dynamics that explode in finite time seems to be nonimplementable, therefore nondesirable. However, as will be proven in Theorem 3.1, frequent direction changes will almost surely rule out trajectories actually reaching ∞. Moreover, since the deterministic dynamics can reach infinity in finite time, when one reverses the time, the dynamics "come down from infinity" in finite time. This means that the time it takes to return to areas of high density may be independent of where the process starts from. This paves the way for the SUZZ algorithm to be uniformly ergodic.

One can allow the deterministic dynamics to be explosive, as long as a switching Poisson process is also introduced, having a very large intensity that will switch the direction of the process before it reaches the explosion time. We will provide conditions, the rates should satisfy, for the process to be a.s. nonexplosive, even if the deterministic dynamics themselves are explosive.

Before that, we need to properly define how can the process explode.

DEFINITION 3.1. Let ζ be as in (11). The process is called nonexplosive if $\zeta = +\infty$ a.s.

We begin with the most essential assumption for the speed function.

ASSUMPTION 3.1 (Speed growth). $\lim_{\|x\| \to \infty} \exp\{-U(x)\}s(x) = 0$.

REMARK 3.1. We are imposing Assumption 3.1 in order to ensure that the process will have switched the deterministic dynamics before they reach the explosion time. To see an example of how things could go wrong, consider a one-dimensional SUZZ with speed function s targeting a distribution that has U as minus log-likelihood. Assume that there exists a x_0 such that for all $x \geq x_0$, $(U(x) - \log s(x))' > 0$, as would typically be the case.

Suppose that the process starts from $(x, +1)$, $x \geq x_0$. The process evolves under the deterministic dynamics given in (8) until the explosion time $t^* = t^*(x, +1)$. Consider the Poisson process with rate $\{m(t) = \lambda(X_t, \theta), t \geq 0\}$. Then

$$\int_0^{t^*} m(t)\,dt = \int_0^{t^*} \lambda(X_t, +1)\,dt$$

$$= \int_x^{+\infty} \lambda(u, +1)\frac{1}{s(u)}\,du$$

$$= \int_x^{+\infty} U'(u) - \frac{s'(u)}{s(u)}\,du$$

$$= \lim_{u \to +\infty} U(u) - \log s(u) - (U(x) - \log s(x)).$$

Therefore, assuming that Assumption 3.1 does not hold and let us say that $s(u) \times \exp\{-U(u)\} \xrightarrow{u \to +\infty} a > 0$ we get that

$$\lim_{u \to +\infty} U(u) - \log(s(u)) = \log(a^{-1}) < \infty,$$

and therefore

$$\mathbb{P}_{x,+1}(\text{no switches until time } t^*(x, +1)) = \exp\left\{ -\int_0^{t^*(x,+1)} m(t)\,dt \right\} > 0.$$

Therefore, if $t^*(x, +1) < \infty$ then the process has a positive probability to explode. The same situation is experienced in higher dimensions assuming that for all coordinates i, $\partial_i(U(x) - \log s(x)) > 0$ for all $x = (x_1, \ldots, x_d)$ for which x_i is positive and very large. Furthermore, as we will see in Proposition D.1, assuming that Assumption 3.1 holds, forces the SUZZ process to a.s. switch direction before it reaches the explosion time. This forces us to adopt Assumption 3.1.

We will, also, make the following assumption.

ASSUMPTION 3.2. Assume that for the refresh rates there exists $\bar{\gamma}$ such that for all $i \in \{1, \ldots, d\}$, $(x, \theta) \in E$, $\gamma_i(x, \theta_{-i}) = \gamma_i(x) \leq \bar{\gamma}$.
Furthermore, assume that there exists $R > 0$ and $A > 0$ so that for all $x \notin B(0, R)$

$$(14) \qquad \sum_{i=1}^d |A_i(x)| > A > \max\{3d\bar{\gamma}, 4d(d-1)\bar{\gamma}\}.$$

REMARK 3.2. When all refresh rates are zero, then $\bar{\gamma} = 0$ and (14) means that the overall switching rate is bounded away from zero, which seems essential in order to gain exponential ergodicity. More generally, the A_i's describe the intention of the algorithm to switch from a direction leading to lower density areas, while the γ_i's describe the intention of the algorithm to switch direction randomly. Large values of γ_i would lead the algorithm to a random walk behaviour (see also [5]) and might decrease the convergence rate. Therefore, (14) could be seen as a quantitative upper bound for the refresh rate.

We also have to assume the following.

ASSUMPTION 3.3. If we iteratively define the functions $h_n : [0, +\infty) \to [0, +\infty)$ such that

$$(15) \qquad h_0(x) = x$$

and for $n \geq 1$

$$(16) \qquad h_n(x) = \log\big(1 + h_{n-1}(x)\big),$$

then there exists an $n \in \mathbb{N}$ such that

$$(17) \qquad \lim_{\|x\| \to \infty} \frac{h_n(s(x)\|\nabla(U(x) - \log s(x))\|_1)}{U(x) - \log s(x)} = 0.$$

REMARK 3.3. Recall that, due to Assumption 3.1, we have that $U(x) - \log s(x) \xrightarrow{|x| \to \infty} +\infty$. Therefore, in practice Assumption 3.3 will almost always hold.

Finally, we make one more assumption.

ASSUMPTION 3.4. For all $j \in \{1, \dots, d\}$, $A_j \in C^1$ and for any $\delta > 0$,

$$(18) \qquad \lim_{\|x\| \to \infty} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{s(x)}{\sum_{k=1}^{d} |A_k(x)|} \frac{|\partial_i A_j(x)|}{(1 + |A_j(x)|)(1 + \log(1 + \delta|A_j(x)|))} = 0.$$

REMARK 3.4. This is a technical assumption used to prove the results of this section and it can be quite difficult to verify in practice for multidimensional targets. We believe, however, that it is not necessary for the results to hold. For example, in Section 3.3 the desired properties for the SUZZ process are directly proved for a family of targets and with speed functions that do not satisfy Assumption 3.4.

We note, however, that Assumption 3.4 generalises one made in [11] to prove exponential ergodicity of the original Zig-Zag. Indeed, when $s(x) = 1$, Assumption 3.4 writes

$$\lim_{\|x\| \to \infty} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{1}{\|\nabla U(x)\|_1} \frac{|\partial_i \partial_j U(x)|}{(1 + |\partial_j U(x)|)(1 + \log(1 + \delta|\partial_j U(x)|))} = 0$$

for all $\delta > 0$. This is weaker than

$$\lim_{\|x\| \to \infty} \frac{\|\operatorname{Hess}(U)(x)\|}{\|\nabla U(x)\|_1} = 0,$$

assumed in [11]. The reader can see Example 5.2.9 of [52] for one-dimensional examples where the target has tails asymptotically similar to those of a Student distribution and it is verified that (18) holds.

Our first main result is the following.

THEOREM 3.1 (Nonexplosion). *Assume that $s \in C^2$ is strictly positive, the rates satisfy* (12) *and Assumptions* 3.1, 3.2, 3.3 *and* 3.4 *hold. Then the process is nonexplosive, meaning that if ζ as in* (11), *then $\zeta = +\infty$ a.s. Furthermore, if ξ as in* (9), *then $\xi = +\infty$ a.s.*

Furthermore, if we pick the switching rates according to (12), then our nonexplosive process leaves the target distribution of interest invariant. For this we need to make the following assumption in the case where the deterministic dynamics of the process are explosive.

ASSUMPTION 3.5.

$$(19) \qquad \lim_{\|x\| \to \infty} \|x\|^{d-1} s(x) \exp\{-U(x)\} = 0.$$

REMARK 3.5. This is a stronger assumption than Assumption 3.1, since it imposes a more strict upper bound on the growth of the speed functions we can use. However, it still allows a lot of flexibility on the growth of s. Consider for example, a d-dimensional Student distribution with ν degrees of freedom, that is, $\pi(x) = \exp\{-U(x)\} \sim 1/|x|^{d+\nu}$, where we write $a(x) \sim b(x)$ to denote that $\lim_{|x| \to \infty} \frac{a(x)}{b(x)} = c$ for a constant $c > 0$. Assumption 3.5 implies that $s(x)/|x|^{1+\nu} \xrightarrow{|x| \to \infty} 0$. Therefore, if $s(x) \sim |x|^{1+k}$, we have to impose the condition that $k < \nu$.

As will be seen in the proof of Theorem 3.2, this assumption is only needed in the case of explosive deterministic dynamics.

We then have the following.

THEOREM 3.2 (Invariant measure). *Assume that the rates satisfy* (12) *and Assumptions* 3.2, 3.3, 3.4 *and* 3.5 *hold. Assume that* $s \in C^2$ *is strictly positive. Then the SUZZ process has the measure* μ *in* (3) *as invariant.*

Crucially, under some further conditions on the speed function s, the SUZZ process is exponentially ergodic even when targeting some heavy tailed distributions.

THEOREM 3.3 (Exponential ergodicity of SUZZ). *Let* $(Z_t)_{t \geq 0} = (X_t, \Theta_t)_{t \geq 0}$ *be a SUZZ process with strictly positive speed function* $s \in C^2$. *Suppose that the rates satisfy* (12) *and Assumptions* 3.2, 3.3, 3.4 *and* 3.5 *hold. Assume further that the function* $U - \log s \in C^3$ *and has a nondegenerate local minimum, that is, there exists an* $x_0 \in \mathbb{R}^d$ *local minimum for* $U - \log s$ *such that the Hessian matrix* $\mathrm{Hess}(U - \log s)(x_0)$ *is strictly positive definite. Finally, assume that* μ *introduced in* (3) *is a probability measure. Then the SUZZ process is exponentially ergodic, that is, there exists some* $M : E \to [1, +\infty)$ *and* $\rho < 1$ *such that for any* $(x, \theta) \in E$,

$$\left\| \mathbb{P}_{x,\theta}(Z_t \in \cdot) - \mu(\cdot) \right\|_{\mathrm{TV}} \leq M(x, \theta)\rho^t. \tag{20}$$

An immediate result due to Theorem 2 of [14] is the following CLT.

THEOREM 3.4 (Central limit theorem). *Suppose that all the assumptions of Theorem* 3.3 *hold. Let* $\{Y_n, n \geq 0\}$ *be any skeleton of the SUZZ process (i.e., for some* $\delta > 0$, $Y_n = Z_{n\delta}$ *for all* $n \in \mathbb{N}$) *and let* $f : E \to \mathbb{R}$ *such that there exists an* $\epsilon > 0$ *with* $\mathbb{E}_\mu[f^{2+\epsilon}] < \infty$. *Then there exists a* $\gamma_f^2 \in [0, \infty)$ *such that*

$$\frac{\sum_{k=1}^n (f(Y_k) - \mu(f))}{\sqrt{n}} \xrightarrow[D]{n \to \infty} Z \tag{21}$$

for some $Z \sim \mathcal{N}(0, \gamma_f^2)$.

Finally, in the case where the target has lighter tails (such that the gradient of the log-likelihood does not decay to zero) we can prove the convergence results for SUZZ under conditions that can be easily verified.

ASSUMPTION 3.6. Assume that $U - \log s \in C^2$ and there exists an $\tilde{M} > 0$ such that the refresh rates $\gamma_i(x)$ of the SUZZ process satisfy $\gamma_i(x) \leq \tilde{M}s(x)$ for all $x \in \mathbb{R}^d$. Assume further that for some $n \in \mathbb{N}$, if h_n as in (16),

$$\lim_{\|x\| \to \infty} \frac{h_n(\|\nabla(U(x) - \log s(x))\|)}{U(x) - \log s(x)} = 0,$$

$$\lim_{\|x\| \to \infty} \frac{\|\mathrm{Hess}(U(x) - \log s(x))\|}{\|\nabla((U(x) - \log s(x))\|} = 0,$$

and that there exists $R > 0$ and $A > 0$ so that for all $x \notin B(0, R)$

$$\|\nabla(U(x) - \log s(x))\|_1 > A > \max\{3d\tilde{M}, 4d(d-1)\tilde{M}\}. \tag{22}$$

THEOREM 3.5. *Let* $(Z_t)_{t\geq 0} = (X_t, \Theta_t)_{t\geq 0}$ *be a SUZZ process with speed function* $s \in C^2$ *bounded away from* 0.

- *Assume that the rates satisfy* (12) *and Assumptions* 3.1 *and* 3.6 *hold. Then the SUZZ process is nonexplosive.*
- *Assume further that either Assumption* 3.5 *holds or the deterministic dynamics are nonexplosive. Then the SUZZ process has the measure* μ *in* (3) *as invariant.*
- *Assume further that the function* $U - \log s \in C^3$ *and has a nondegenerate local minimum, in the sense of Theorem* 3.3. *Finally, assume that* μ *introduced in* (3) *is a probability measure. Then the SUZZ process is exponentially ergodic.*
- *Assuming the assumptions of the previous bullet, let* $\{Y_n, n \geq 0\}$ *be any skeleton of the SUZZ process (i.e., for some* $\delta > 0$, $Y_n = Z_{n\delta}$ *for all* $n \in \mathbb{N}$*) and let* $f : E \to \mathbb{R}$ *such that there exists an* $\epsilon > 0$ *with* $\mathbb{E}_\mu[f^{2+\epsilon}] < \infty$. *Then the CLT result of* (21) *holds.*

The conditions of Theorem 3.5 can be seen as direct generalisations of assumptions made in [11] for the original Zig-Zag. Therefore, Theorem 3.5 guarantees that for reasonable speed functions, the convergence properties of the original Zig-Zag carry over in SUZZ. This allows one to see the speed function as a tuning parameter for the original Zig-Zag, which could potentially increase the efficiency of the algorithm even in cases where the original Zig-Zag works well.

3.3. *Stability and convergence for practical choices of speed functions.* Assumption 3.4 used in Theorems 3.1, 3.2, 3.3 and 3.4 can be difficult or impossible to verify for some practical choices of speed functions. For this reason, in this section we will focus our attention on these particular, practical speed functions and we will establish convergence properties for a class of targets, some of which we will also use in simulations in Section 5.

We will consider two speed functions, namely

$$s(x) = \left(1 + \|x\|_2^2\right)^{\frac{1+k}{2}} \tag{23}$$

for $k = 0$ and $k = 1$. We will refer to the SUZZ algorithms induced by these two functions as SUZZ(0) and SUZZ(1) respectively. Note that SUZZ(0) has nonexplosive deterministic dynamics, while SUZZ(1) has explosive ones, since the speed function grows super-linearly.

We have the following.

PROPOSITION 3.1. *Assume that the target is of the form*

$$\pi(x) = \frac{1}{H} \exp\{-\left(1 + \|x\|_2^2\right)^{a/2}\} \tag{24}$$

for some $a \in (0, 1)$ *or of the form*

$$\pi(x) = \frac{1}{H}\left(1 + \frac{1}{\nu}\|x\|_2^2\right)^{-\frac{\nu+d}{2}} \tag{25}$$

for some ν *satisfying*

$$\nu > \frac{27}{2}d^3 + 2 - d. \tag{26}$$

Assume also that s *is as in* (23) *for* $k = 0$ *or* $k = 1$. *Then the SUZZ process with refresh rates* $\gamma_i \equiv 0$ *is nonexplosive, has* μ *as invariant, is geometrically ergodic and satisfies the CLT as in Theorem* 3.4.

REMARK 3.6. Following the proof of Proposition 3.1 in Appendix H, we can more generally have the conclusion of Proposition 3.1 when the speed function is such that there exists a $K > 0$ and $M_0 > 0$ such that for all $\|x\|_2 \geq K$, $\frac{s(x)}{\|x\|_2} \geq M_0$ and the target is such that for all $\|x\|_2 \geq K$,

$$A_i(x) = c(x)(B \cdot x)_i,$$

where B is a positive definite matrix such that for all $i \in \{1, \ldots, d\}$, $b_{ii} - \sum_{j \neq i} |b_{ij}| \geq m > 0$ and if $M = \max\{\sum_{j=1}^{d} |b_{ij}|, i = 1, \ldots, d\}$, then c satisfies for all $\|x\|_2 \geq K$

$$c(x)\frac{\|x\|_2^2}{s(x)} > \frac{27}{2}\frac{M}{m}d^3.$$

3.4. *Comparison with results on the original Zig-Zag.* In this section we will translate the assumptions and the results of Section 3.2 in the case of the original Zig-Zag process, which arises when we use the constant speed function $s(x) = 1$. In this setting, we will see that all the assumptions made in Section 3.2 are weaker versions of assumptions made in [11]. This will serve as a way to justify our assumptions and at the same time will allow us to prove exponential ergodicity of the original Zig-Zag process under weaker assumptions than the ones of Theorem 2 of [11].

Our first observation is that in the original Zig-Zag case where $s \equiv 1$, Assumption 3.5 is implied by the following growth condition.

ASSUMPTION 3.7. There exists $\epsilon > 0$, $c' \in \mathbb{R}$ such that for all $x \in \mathbb{R}^d$, $U(x) \geq (d + \epsilon)\log(\|x\|) - c'$.

REMARK 3.7. Assumption 3.7 is Growth Condition 2 of [11], assumed in order to prove nonevanescence of the original Zig-Zag process.

Second, we observe that in the setting of the original Zig-Zag, Assumption 3.2 is the following.

ASSUMPTION 3.8. Assume that for the refresh rates there exists $\bar{\gamma}$ such that for all $i \in \{1, \ldots, d\}$, $(x, \theta) \in E$, $\gamma_i(x, \theta_{-i}) = \gamma_i(x) \leq \bar{\gamma}$.

Furthermore, assume that there exists $R > 0$ and $A > 0$ so that for all $x \notin B(0, R)$

$$\|\nabla U(x)\|_1 > A > \max\{3d\bar{\gamma}, 4d(d-1)\bar{\gamma}\}.$$

REMARK 3.8. We observe that this is a weaker version of Growth Condition 3 of [11], necessary for proving exponential ergodicity of the original Zig-Zag process. Instead of asking that $\lim_{\|x\|\to\infty} \|\nabla U(x)\|_1 = +\infty$, we only ask that the limit is bounded below by a constant that may depend on the dimension of the space. Note as well that in the canonical Zig-Zag case, $\bar{\gamma} = 0$, so we only ask for $\|\nabla U\|_1$ to be bounded away from zero at the tails.

Furthermore, in the case of the original Zig-Zag, Assumption 3.3 is the following.

ASSUMPTION 3.9. If $h_0(x) = x$ and for all $n \in \mathbb{N}$, h_n is defined as in (16), then there exists an $n \in \mathbb{N}$ such that

$$(27) \qquad \lim_{\|x\|\to\infty} \frac{h_n(\|\nabla U(x)\|_1)}{U(x)} = 0.$$

REMARK 3.9. We note that Assumption 3.9 is almost always true in any practical setting where $U \in C^1$. Furthermore, it is a relaxed version of Growth Condition 3 of [11].

Finally, in the case of the original Zig-Zag, Assumption 3.4, is equivalent to the following.

ASSUMPTION 3.10. $U \in C^2$ and for all $\delta > 0$

$$(28) \qquad \lim_{\|x\| \to \infty} \frac{1}{\|\nabla U(x)\|_1} \sum_{j=1}^{d} \frac{\sum_{i=1}^{d} |\partial_i \partial_j U(x)|}{(1 + |\partial_j U(x)|)(1 + \log(1 + \delta|\partial_j U(x)|))} = 0.$$

REMARK 3.10. As mentioned in Remark 3.4, Assumption 3.10 is weaker than Growth Condition 3 of [11].

Using these assumptions, we see that an immediate corollary of Theorem 3.3 is the following.

COROLLARY 3.1 (Exponential ergodicity of original Zig-Zag). *Let $(Z_t)_{t \geq 0} = (X_t, \Theta_t)_{t \geq 0}$ be a d-dimensional original Zig-Zag process. Assume that $U \in C^3$, and has a nondegenerate local minimum. Assume further that Assumptions 3.7, 3.8, 3.9 and 3.10 hold. Then the original Zig-Zag process is exponentially ergodic, that is, there exist $M : E \to [1, +\infty)$, and $\rho < 1$ such that for any $(x, \theta) \in E$,*

$$(29) \qquad \left\| \mathbb{P}_{x,\theta}(Z_t \in \cdot) - \mu(\cdot) \right\|_{\mathrm{TV}} \leq M(x, \theta)\rho^t.$$

3.5. *Space transformation and uniform ergodicity.* When we focus on the one-dimensional process, we can prove that it is a space transformation of an original, one-dimensional Zig-Zag process. We have the following.

PROPOSITION 3.2 (One-dimensional SUZZ as space transformation). *Consider a one-dimensional SUZZ process $Z_t = (X_t, \Theta_t)_{t \geq 0}$ with strictly positive speed function $s \in C^2$, targeting a measure μ as in (3). Assume that the rates satisfy (12) and let*

$$(30) \qquad f(x) = \int_0^x \frac{1}{s(u)} \, du$$

and

$$(31) \qquad \pm M^{\pm} = \lim_{x \to \pm\infty} f(x) \in \mathbb{R} \cup \{-\infty, +\infty\}.$$

Then the process $(Y_t, \Theta_t)_{t \geq 0}$, where $Y_t = f(X_t)$, is a one-dimensional original Zig-Zag process, defined on $(-M^-, M^+) \times \{-1, +1\}$. If the SUZZ process is nonexplosive, then $(Y_t, \Theta_t)_{t \geq 0}$ has invariant measure ν where

$$(32) \qquad \nu(dy, d\theta) = \frac{1}{\tilde{H}} \exp\{-\tilde{U}(y)\} \, dy \, d\theta$$

and

$$(33) \qquad \tilde{U}(y) = U(f^{-1}(y)) - \log s(f^{-1}(y)).$$

Using Proposition 3.2, we can prove that the one-dimensional SUZZ process with explosive deterministic dynamics is uniformly ergodic. This means that it is exponentially ergodic and the mixing time can be bounded by a quantity that does not depend on the starting point. This is a consequence of the fact that explosive deterministic dynamics have ∞ as entrance boundary. Our current proof, presented in Appendix I, heavily relies on the fact that the one-dimensional process is a space transformation of an original Zig-Zag.
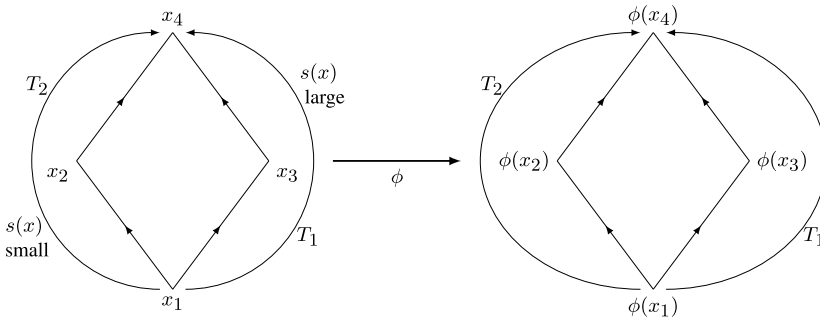
FIG. 3. *Figure explaining why a two-dimensional SUZZ is not in general a space transformation ϕ of an original Zig-Zag process. The figure on the left shows two possible SUZZ paths from x_1 to x_4, passing via either x_2 or x_3 and switching directions there. Assuming that the SUZZ process was a space transformation of an original Zig-Zag, via a function ϕ, the figure on the right shows the two paths of the space transformed original Zig-Zag process, from $\phi(x_1)$ to $\phi(x_4)$, via either $\phi(x_2)$ or $\phi(x_3)$. The times T_1 and T_2 to traverse the two paths from x_1 to x_4 on the left figure, depending on the speed function, do not have to be the same. The same times, T_1, T_2, would also be the times to traverse the two paths from $\phi(x_1)$ to $\phi(x_4)$ on the ϕ-transformed right figure. However, if the ϕ-transformed right figure was an original Zig-Zag, moving with constant unit speed, these two times would had to be the same.*

THEOREM 3.6 (Exponential and uniform ergodicity in one dimension). *Consider a one-dimensional SUZZ process $Z_t = (X_t, \Theta_t)_{t \geq 0}$ with strictly positive speed function $s \in C^2$. Assume that the rates satisfy* (12) *and Assumptions* 3.1 *and* 3.2 *hold. Then the process is nonexplosive, it has μ defined in* (3) *as invariant and is exponentially ergodic. Assume further, that for some $x \in \mathbb{R}$ and for any $\theta = \pm 1$ the deterministic flow of the process $\{\Phi_t(x, \theta), t \geq 0\}$ has a finite explosion time $t^*(x, \theta)$. Then the process is uniformly ergodic, that is, there exists a $M > 0$ and $\rho < 1$ such that for any $(x, \theta) \in E$ and $t \geq 0$,*

$$\|\mathbb{P}_{x,\theta}(Z_t \in \cdot) - \mu(\cdot)\|_{\mathrm{TV}} \leq M\rho^t.$$

We emphasise, however, that the SUZZ algorithm cannot necessarily be written as a space transformation of an original Zig-Zag in dimension higher than 1. In Figure 3 we illustrate the contradiction that may occur if such a space transformation were to exist. We consider a $d = 2$ case, and assume (to reach a contradiction) that there does exist such a transformation ϕ. The left figure represents the movement of a two-dimensional SUZZ process starting from x_1 and ending at x_4. The right figure represents the movement of the ϕ-space transformed process, assumed to be an original Zig-Zag, starting from $\phi(x_1)$ and ending at $\phi(x_4)$. There are two paths from x_1 to x_4, passing through and switching at x_2 or x_3 respectively. If the speed function $s(x)$ takes smaller values on the path via x_2, then the process arrives at x_4 faster via the x_3 path rather than via the x_2 path. The same thing must hold for the ϕ-transformed process, that is, the process arrives to $\phi(x_4)$ faster via $\phi(x_3)$ rather than $\phi(x_2)$. However, the transformed process moves with constant unit speed, as it is an original Zig-Zag process. Furthermore, the two paths from $\phi(x_1)$ to $\phi(x_4)$, passing either via $\phi(x_2)$ or via $\phi(x_3)$ have the same length. Therefore the time it takes for the transformed process to traverse either of the two paths from $\phi(x_1)$ to $\phi(x_4)$ is the same. This gives a contradiction and establishes that the SUZZ process cannot be spaced transformed to an original Zig-Zag process in dimension $d \geq 2$.

As a result of this discussion, we cannot rely on the existence of such a transformation between SUZZ and original Zig-Zag, so results for SUZZ cannot easily be obtained from those for original Zig-Zag by simple transformation arguments. This also means that we do not currently have a way to extend the uniform ergodicity result of Theorem 3.6 to higher

dimensions, since the proof heavily relies on the space transformation property of the one-dimensional SUZZ process. Simulation results, however, seem to suggest that the starting position does not heavily influence the algorithmic performance, and we suspect the uniform ergodicity holds for higher dimensions as well.

4. Choice of speed function. A natural objective is to choose the speed function that generates an algorithm which is as efficient as possible. To get some intuition into how to achieve this, consider the one-dimensional SUZZ process $(X_t, \Theta_t)_{t\geq 0}$ with speed function s and f as given in (30). From Proposition 3.2, $(Y_t, \Theta_t) = (f(X_t), \Theta_t)$ is an original Zig-Zag process, targeting a measure with negative log-density given by $\tilde{U}(y) = U(f^{-1}(y)) - \log s(f^{-1}(y))$, defined on a subset of $\mathbb{R}$. Therefore, instead of using SUZZ, one could equivalently use the original Zig-Zag (Y_t, Θ_t), target the potential $\tilde{U}$ and then use the path of $f^{-1}(Y_t)$ as a way to sample from the measure of interest. This is very similar in spirit to the work of [31]. In summary, in the one-dimensional case, the goal of choosing the most efficient s, boils down to choosing an invertible space transformation f, and analysing an original Zig-Zag algorithm on the transformed potential

$$\tilde{U}(y) = U(f^{-1}(y)) - \log s(f^{-1}(y)) = U(f^{-1}(y)) + \log f'(f^{-1}(y)).$$

A natural candidate suggested by this is the choice $s(x) = H \exp\{U(x)\}$ leading to the space transformation

$$f(x) = \int_0^x \frac{1}{H} \exp\{-U(u)\} du = F(x) - \pi((-\infty, 0]),$$

where F is the CDF of π. Using this SUZZ is equivalent to running an original Zig-Zag on the measure with negative log-density $\tilde{U} \equiv 0$ (i.e., the Lebesgue measure) and then transforming the values back according to the function $f^{-1}(y)$. There are similarities here with inverse CDF sampling. However, this choice of s is precluded by Assumption 3.1 as it leads to an explosive SUZZ, corresponding to the transformed Zig-Zag process eventually hitting the boundary of the transformed space (either $f(+\infty)$ or $f(-\infty)$).

This discussion suggests that we might obtain an efficient method by picking s such that $s(x) \exp\{-U(x)\}$ decays to zero as $|x| \to \infty$ slowly. While the equivalence of the SUZZ to an original Zig-Zag with appropriate transformation is only valid in the one-dimensional case, this strategy for choosing s can be applied quite generally in multidimensional settings.

4.1. *A computational efficiency criterion in the one-dimensional case.* Computational efficiency of the algorithm goes far beyond qualitative convergence results such as exponential ergodicity. The actual cost of implementing MCMC algorithms is controlled by the number of computational operations that need to be performed to obtain a desirable amount of samples from the target distribution. In our setting the computational cost comes largely from evaluating the gradient of the log-likelihood of the target, which is needed in order to sample the direction switches. Therefore, in order to understand the algorithmic efficiency, we must study the number of the gradient log-likelihood evaluations needed to be performed until we get enough samples from the target. This section will try to answer this question for the one-dimensional SUZZ process.

In an ideal setting, using Poisson thinning in a perfect way (see [37]), and for any choice of speed function, the number of gradient log-likelihood evaluations (and therefore the computational cost) would be equal to the number of switches of direction. In practice, the actual number of gradient log-likelihood evaluations depends on the tightness of the bounds used in this Poisson thinning operation and is therefore difficult to use as a consistent metric. Therefore we shall instead use the number of direction switches as a unit for measuring the

implementation cost of the algorithm. Our goal now is to define a quantity that depends on the speed function and provides a way to measure the performance of the algorithm per implementation cost.

For the remainder of this section, we focus on dimension one and closely follow [5].

Let N_T be the expected number of switches until time T, that is, the average implementation cost of the algorithm. Then $N_T = \mathbb{E}[\int_0^T \lambda(X_s, \Theta_s)\,ds]$. Since the process is Harris recurrent, we have a law of large numbers and

$$
N_0 := \lim_{T \to \infty} \frac{N_T}{T} = \int \lambda(x, \theta)\,d\mu(x, \theta)
$$

$$
(34) \qquad\qquad = \frac{1}{2H} \sum_{\theta = \pm 1} \int_{\mathbb{R}} \exp\{-U(x)\}\lambda(x, \theta)\,dx
$$

$$
= \frac{1}{2H} \int_{\mathbb{R}} \exp\{-U(x)\}|s(x)U'(x) - s'(x)|\,dx.
$$

Consider a functional of interest $g \in L^2(\mu)$, whose integral under μ we are trying to approximate. Let $(Z_t)_{t \geq 0}$ be a SUZZ process targeting μ. Assume without loss of generality that $\mu(g) = 0$ and consider the estimator

$$
(35) \qquad\qquad g_T = \frac{1}{T} \int_0^T g(Z_s)\,ds.
$$

If the process satisfies a CLT then there exists an asymptotic variance $\gamma_g^2 \in [0, +\infty)$ such that

$$
(36) \qquad\qquad \lim_{T \to +\infty} T \cdot \mathrm{Var}(g_T) = \gamma_g^2.
$$

A way to measure the efficiency of the algorithm is the effective sample size (ESS) (see [50]) which approximates the number of independent samples the algorithm has generated from the target until time T. It is defined as

$$
(37) \qquad\qquad \mathrm{ESS}(T) = \frac{\mathrm{Var}_\mu(g)}{\mathrm{Var}(g_T)}.
$$

Since the cost of implementing the algorithm is the average number of switches, it seems natural to consider the quantity of ESS per average number of direction switches in order to evaluate the efficiency of the algorithm. Combining (34), (36) and (37) we get

$$
(38) \qquad\qquad \lim_{T \to \infty} \frac{\mathrm{ESS}(T)}{N_T} = \frac{\mathrm{Var}_\mu(g)}{\gamma_g^2 N_0}.
$$

Therefore, in order to choose the optimal s that makes the algorithm the most efficient we need to minimize the quantity $\gamma_g^2 N_0$ over different speed functions. N_0 is written in terms of s in (34). We will now present a proposition that describes the asymptotic variance γ_g^2 in terms of s. Before that, we need to make an assumption. Let

$$
(39) \qquad V(x, \theta) = \exp\left\{aU(x) - a\log s(x) + \frac{1}{2}\delta|A_i(x)|h_n(|A_i(x)|)\right\},
$$

where h_n as in (16), and $a, \delta > 0$ small enough so that if $\mathcal{L}$ is the operator defined for all $f \in C^1(E)$ as

$$
(40) \qquad \mathcal{L}f(x, \theta) = \theta s(x)f'(x, \theta) + \left([\theta U'(x)]^+ + \gamma(x)\right)\left(f(x, -\theta) - f(x, \theta)\right),
$$

then there exist $c, b > 0$ and a compact set C, such that for all $(x, \theta) \in E$,

$$
(41) \qquad\qquad \mathcal{L}V(x, \theta) \leq -cV(x, \theta) + b1_{(x, \theta) \in C}.
$$

The fact that (41) holds for a, δ small enough will be later proved in Appendix D, in the proof of Theorem 3.1. We now assume the following.

ASSUMPTION 4.1. Let $\mathcal{L}$ be the operator in (40). Let V be defined as in (39) for $a, \delta > 0$ small enough such that (41) holds. Assume that there exists a $C > 0$ such that for all $g \in L^1(E)$ satisfying $|g(x, \theta)| \leq V(x, \theta)$ for all $(x, \theta) \in E$, there exists a ϕ such that

$$-\mathcal{L}\phi = g$$

and such that for all $(x, \theta) \in E$

$$\left|\phi(x, \theta)\right| \leq C V(x, \theta).$$

This assumption is a result proven in [27] in the case where $\mathcal{L}$ is the extended generator of a process, in the sense that for any $f \in C^1(E)$ the process

$$M_t = f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_s)\,ds$$

is a martingale. However, since we allow the process to have explosive deterministic dynamics, we can only guarantee that M_t is a local martingale. We note here that in [27] the authors claim that Assumption 4.1 holds in our case as well, that is, when $\mathcal{L}$ only induces a local martingale. However, to the best of our knowledge this is not something proven in the literature. Therefore, we make this assumption here and we present the following result under Assumption 4.1. This result describes the asymptotic variance γ_g^2 in terms of the speed function s.

PROPOSITION 4.1. *Assume that the rates satisfy (12) and Assumptions 3.1, 3.2, 3.3 and 3.4 hold. Let $g : E \to \mathbb{R}$ in the domain of $\mathcal{L}$, with $\mu(g) = 0$ and assume that $|g(x, \theta)| \leq V(x, \theta)$ for all $(x, \theta) \in E$, where V is the function defined in (39) for some $a < 1$, $\delta > 0$ small enough such that (41) holds. Finally, assume that Assumption 4.1 holds. Then, if $Z_t = (X_t, \Theta_t)$ is the one-dimensional SUZZ process with speed function s, starting from the invariant measure μ, we have*

$$\frac{1}{\sqrt{T}} \int_0^T g(Z_s)\,ds \xrightarrow{T \to \infty} N\left(0, \gamma_g^2\right)$$

in distribution where

(42) $$\gamma_g^2 = \frac{1}{2H} \int_{\mathbb{R}} |s(x)U'(x) - s'(x)| \frac{1}{s^2(x)\exp\{-U(x)\}} k^2(x)\,dx$$

and

(43) $$k(x) = \int_x^{+\infty} \left(g(y, +1) + g(y, -1)\right) \exp\{-U(y)\}\,dy.$$

Proposition 4.1 and equations (34) and (38) indicate that for a given function $g : E \to \mathbb{R}$, satisfying the assumptions of Proposition 4.1, we need to pick a speed function s in order to minimize the quantity

(44) $$J[r] := \gamma_g^2 N_0 = \int_{\mathbb{R}} |r'(x)|\,dx \int_{\mathbb{R}} \frac{|r'(x)|}{r^2(x)} k^2(x)\,dx,$$

where

$$r(x) = s(x)\exp\{-U(x)\},$$

and we need to impose the condition

$$\lim_{|x| \to \infty} r(x) = 0,$$

so that Assumption 3.1 holds.

We will call the functional J the *inverse algorithmic efficiency*. Note that J is invariant under constant scaling of function s. This is in accordance to the fact that we do not gain or lose any efficiency by speeding up Zig-Zag with a constant speed, for example, by having velocities of the form $\{\pm 2\}$.

REMARK 4.1. The result of Proposition 4.1 can be generalised in the case where $\mu(g)$ is not necessarily zero. In the general case, the function k in (43) used to define $J[r]$ would be

$$k(x) = \int_x^{+\infty} (g(y, +1) + g(y, -1) - 2\mu(g)) \exp\{-U(y)\} \, dy.$$

In practice, $\mu(g)$ is not a known quantity. Then one can use the asymptotically unbiased estimator g_T in (35) instead of $\mu(g)$ to calculate an approximation of the inverse efficiency J.

Ideally, we would like to pick a speed function such that r minimises (44). Note, however, that minimising (44) is not a well-posed problem. Indeed, let r_0 be a function such that $J[r_0] < \infty$, $r_0(0) = 1$ and $\lim_{|x| \to \infty} r_0(x) = 0$. For any $n \in \mathbb{N}$ let

(45)
$$\begin{cases} r_n(x) = 1 & |x| \leq n, \\ r_n(x) = r_0(x - n) & x > n, \\ r_n(x) = r_0(x + n) & x < -n. \end{cases}$$

Then $J[r_n] \xrightarrow{n \to \infty} 0$. At the same time, the only functions that satisfy $J[r] = 0$ are the constant ones and since we impose the condition that $\lim_{|x| \to \infty} r(x) = 0$, the only function r that satisfies $J[r] = 0$ is the function $r \equiv 0$.

Note, however, that the nth term of the minimising sequence r_n is equal to 1 on $[-n, n]$ and this means that $s(x) = \exp\{U(x)\}$ for $x \in [-n, n]$. Heuristically, and as discussed in the beginning of Section 4, one could expect good performance in the ideal case where $s(x)$ could be set equal to $\exp\{U(x)\}$ for $x \in [-n, n]$ for some large n.

In Table 1 we present some examples, comparing the efficiency of different algorithms. As target distribution we consider a normal with mean zero and variance one, an exponential with parameter one, symmetrically extended to the negative reals (Laplace), a Student distribution with 3 degrees of freedom, and a distribution of the form

$$\pi(x) = \frac{1}{H} \exp\{-(1 + x^2)^{1/4}\},$$

which we will call subexponential, since it has tails heavier than any exponential, but it does not decay polynomially fast. For each of these densities, except for the Student(3), we are

TABLE 1
J values, as introduced in (44), for various SUZZ algorithms targeting various distributions; SUZZ(k) denotes the SUZZ algorithm with speed function of the form $s(x) = (1 + x^2)^{(1+k)/2}$; Smallest value for every column in bold

Algorithmic inverse efficiency in one dimension				
Algorithms	Normal	Exponential	Subexponential(0.5)	Student(3)
Original Zig-Zag	16	80	57,044	34.2457
SUZZ(0)	4.9817	26.3397	**3,536**	7.9736
SUZZ(1)	**4.4259**	**7.1017**	45,948	**2.4708**
SUZZ(2)	14.9568	19.0364	1,315,827	11.7397
SUZZ(3)	45.6342	30.0839	24,623,975,012	–

estimating the expectation of the distribution, that is, we set $g(x, \theta) = x$. For the Student(3) we are estimating the expectation the function $g(x) = \text{sgn}(x) \log(1 + |x|)$. This is to guarantee that the function g verifies the growth assumptions of Proposition 4.1. Regarding the speed function, we use the original Zig-Zag (i.e., $s(x) = 1$), and we also use the speed functions

$$(46) \qquad\qquad\qquad s(x) = (1 + x^2)^{(1+k)/2}$$

for $k = 0, 1, 2, 3$. These algorithms will be denoted by SUZZ(k), where k is the parameter in the exponent of the speed function. Choosing $k = 0$ induces nonexplosive deterministic dynamics, whereas choosing $k > 0$ induces explosive ones. We will verify the assumptions used in Theorem 3.6 for these speed functions is Appendix K. In Table 1 we compare the inverse efficiencies of all the algorithms for all four targets. In order to numerically estimate the integrals arising in the definition of $J[r]$ we use the `integrate` function and the `polyroot` library of R. We should emphasize that since we do not take into account some normalisation constants and since in the case of Student(3) distribution we are estimating a different observable, the comparison in Table 1 should only be made column-wise (i.e., for a given distribution compare different algorithms).

For any target distribution, the algorithms SUZZ(0) and SUZZ(1) provide better results than the original Zig-Zag algorithm. Furthermore, for all targets except for the subexponential, the SUZZ(2) algorithm performs better than the original Zig-Zag. SUZZ(3) does not seem to perform that well and it only has better efficiency than the original Zig-Zag on the exponential target. Note that we do not present an efficiency value for SUZZ(3) on the Student(3) target since this algorithm does not satisfy Assumption 3.1 and will in fact explode in finite time a.s. It is also worth noting that for all the targets, with the exception of the subexponential one, the inverse efficiency function of the algorithms seems to be "quadratic" with respect to k and seems to be minimised when $k = 1$.

Finally, we should note that the notion of inverse algorithmic efficiency is so far restricted to one-dimensional setting. Generalising this to higher dimensions would involve solving the Poisson equation of the multidimensional SUZZ process and is subject to further research.

5. Numerical simulations.

In this section we will present some computational results that aim to highlight the behaviour of SUZZ and compare it with original Zig-Zag. We will present results for one-dimensional and twenty-dimensional targets. As already suggested in Section 4.1, the one-dimensional SUZZ can vastly outperform the original Zig-Zag. However, it will be seen that there are significant advantages in using a speed function in higher dimensions as well.

First, we present numerical results on a one-dimensional Student target with three degrees of freedom, denoted by Student(3), that is, a target with density given by $\pi(x) = \frac{1}{H}(1 + \frac{1}{3}x^2)^{-2}$. We used the Zig-Zag algorithm (ZZ) along with SUZZ(0) and SUZZ(1) algorithms, where SUZZ(k) indicates the SUZZ algorithm with speed function given by (46). We emphasise here that even though the deterministic dynamics of SUZZ(1) explode in finite time, the process will a.s. not explode due to Theorem 3.6. Finally, we also used a Random Walk Metropolis algorithm on a transformed state space, introduced in [31] as a method that is geometrically ergodic even on heavy tailed targets. The proposal distribution is a one-dimensional Normal(0, 1) and the parameters of the space transformation are tuned using the guidance of the discussion in [31]. We will be referring to this algorithm as Transformed Random Walk Metropolis (TRWM). For each of the four algorithms presented, we simulated 25 independent realisations of each process, until $N = 10^4$ switches of direction occurred for the ZZ or SUZZ algorithms. In the case of TRWM we simulated for $N = 10^4$ steps. To construct a sample from the ZZ and the SUZZ algorithms, we used the position of the process every δ time units (δ-skeletons). Here δ is different for each algorithm and it is chosen in the

following way. For each algorithm, we first run an initial run, which created a path of time length $S(N)$. Then we fixed $\delta = \frac{S(N)}{N}$ so that for this run, the size of the skeleton was equal to $N = 10^4$. We used this fixed δ for all other runs of the algorithm, expecting each future skeleton to have a size roughly equal to $N = 10^4$, which was indeed the case. This was done in order to guarantee fairness between the performance evaluation across all algorithms. More precisely, since we use the number of switches ($N = 10^4$) as a unit to measure computational cost, it would make sense for all the algorithms that run for the same number of switches to produce roughly the same number of samples. For TRWM, since we run the algorithm for $N = 10^4$ steps, the sample generated had a size of $N = 10^4$. To analyse the performance of the algorithms we have used the effective sample size (ESS) (see [17]), computed using `coda` from R. The ESSs were calculated after we transformed the sample via the function

$$(47) \qquad f(x) = \text{sgn}(x) \log(1 + |x|),$$

so that we can guarantee that the variance of the ESS is finite. All simulations were performed using MATLAB in a computer with i7-8550U CPU and 1.80 GHz.

We present our results in Table 2. We present average and median ESS across 25 realizations (standard deviation in parenthesis). We also report the median ESS per likelihood evaluation and per minute of implementation time. The best performance is highlighted in bold letters. It is clear that both SUZZ algorithms outperform both the original Zig-Zag and the TRWM, in all criteria based on ESS, (ESS per switches, per likelihood evaluations and per implementation time). It is also interesting that the algorithm with the explosive deterministic dynamics seems to perform the best. This is consistent with Table 1 where the inverse algorithmic efficiency of SUZZ(1) is the smallest of all algorithms targeting the Student(3).

As a second example, we used two SUZZ algorithms and ZZ to target a one-dimensional Cauchy distribution (i.e., $\pi(x) = \frac{1}{H}(1 + x^2)^{-1}$). For the SUZZ algorithms we used speed functions of the form

$$(48) \qquad s(x) = \max\{1, |x|^{1+k}\}$$

for $k = 0$ and $k = 0.5$, denoted by SUZZ(0) and SUZZ(0.5). In Figure 4 we present the Q-Q plots for these three algorithms against the Cauchy target. All the algorithms run for $N = 10^4$ number of switches. It is clear that the SUZZ algorithms far outperform the original Zig-Zag process and the SUZZ algorithm with explosive deterministic dynamics ($k = 0.5$) seems to have the optimal performance.

TABLE 2
SUZZ, ZZ and TRWM algorithms targeting a one-dimensional Student(3) distribution. For SUZZ(k), we use the speed function as in (46). The algorithms ran until $N = 10^4$ switches (or steps for the TRWM) occurred and the average ESS (with standard deviation in a parenthesis) along with the median ESS are presented. The median ESS per average likelihood evaluations and per average minutes of implementation time is also presented. All ESSs are calculated after we transform the sample via the function f as in (47). The best performance is highlighted with bold letters

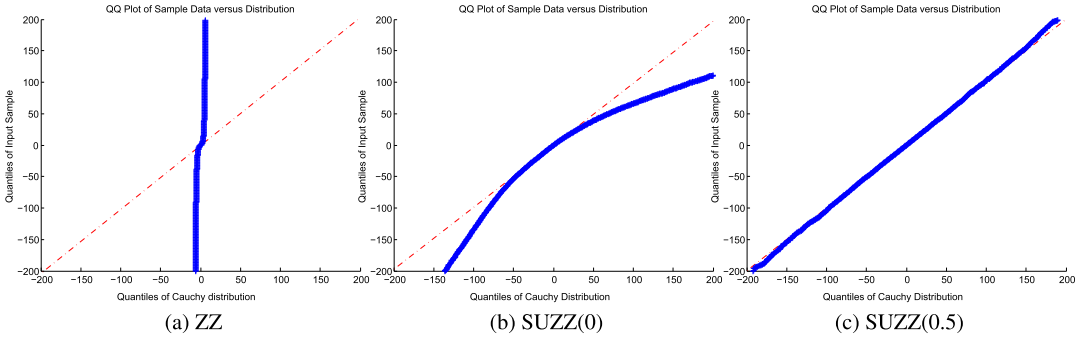|	One-dimensional Student(3), number of switches $N = 10^4$			
Algorithms	ESS(SD)	Median ESS	ESS/Lik.Eval.	ESS/min
ZZ	5272.9 (1274.0)	5675.6	$1.5 \cdot 10^{-4}$	15,765.6
SUZZ(0)	20,755.8 (718.1)	20,779.2	$3.0 \cdot 10^{-2}$	31,483.6
SUZZ(1)	**46,346.2 (3154.6)**	**46,397.8**	**$3.4 \cdot 10^{-2}$**	**154,659.3**
TRWM	29.8 (14.0)	22.8	$0.2 \cdot 10^{-2}$	3257.1

FIG. 4. *Q-Q plots of various one-dimensional SUZZ algorithms targeting a Cauchy distribution. The algorithms have run until $N = 10^4$ switches of direction have occurred. The sample is created using the δ-skeleton of the process for $\delta = 0.1$. As SUZZ(k) we denote the SUZZ algorithm with speed function of the form (48) and k the parameter appearing in the equation.*

Next, we present results on two twenty-dimensional targets. The first target is a twenty-dimensional distribution with density of the form

$$(49) \qquad \pi(x) = \frac{1}{H} \exp\{-(1 + \|x\|_2^2)^{1/4}\},$$

which we will call Subexponential(0.5), since the tails decay like $\exp\{-\|x\|_2^{0.5}\}$, slower than any exponential target but faster than any polynomial.

The second target is a twenty-dimensional Student distribution, with 3 degrees of freedom (denoted by Student(3)), with scale matrix given by B where

$$(50) \qquad \begin{aligned} B(i, j) &= 5, \quad i \neq j, \qquad B(i, i) = 30, \quad i = 1, 2, 3, \\ B(i, i) &= 20, \quad i = 4, 5, \quad \text{and} \quad B(i, i) = 10, \quad i = 6, \dots, 20. \end{aligned}$$

This means that

$$(51) \qquad \pi(x) = \frac{1}{H}\left(1 + \frac{1}{3}x^T B^{-1} x\right)^{-23/2}.$$

The first target is in the setting of Proposition 3.1, therefore we know that SUZZ will be exponentially ergodic. Even though the second target is not in the setting of that proposition and no theoretical guarantees are established for the rate of convergence, we will see that the SUZZ process leads to numerical gains for both targets, compared to the original Zig-Zag and to TRWM algorithm, introduced in the one-dimensional simulations. We conjecture that some of the assumptions made in this document might not be necessary and the class of targets on which SUZZ can work well could be larger.

For both distributions we used four different algorithms to target them and we compare their performances. First of all, we used an original Zig-Zag process (ZZ). We also used two SUZZ processes, SUZZ(0) and SUZZ(1), where SUZZ(k) denotes the SUZZ process with speed function given by (23). Note that SUZZ(0) has nonexplosive deterministic dynamics, while SUZZ(1) has explosive ones. We present a general way to construct the deterministic dynamics for this type of speed functions in Appendix B. Finally, we also used the Transformed Random Walk Metropolis (TRWM) algorithm, described in the one-dimensional simulations. The proposal distribution we used was a twenty-dimensional normal with identity covariance matrix and the parameters of the space transformation were tuned using the guidance of the discussion in [31]. For each of the four algorithms presented, we simulated 25 independent realisations of each process, until $N = 10^6$ switches of direction occurred for the ZZ or SUZZ algorithms. In the case of TRWM we simulated for $N = 10^6$ steps. Having

TABLE 3

SUZZ, ZZ and TRWM algorithms targeting a twenty-dimensional Subexponential(0.5) *distribution, with density given by* (49). *For SUZZ*(k), *we use the speed function as in* (23). *The algorithms ran until* $N = 10^6$ *switches* (*or steps for the TRWM*) *occurred and the average ESS* (*with standard deviation in a parenthesis*) *along with the median ESS is presented. The median ESS per average likelihood evaluations and per average minutes of implementation time is also presented. All ESS concern the first coordinate of the twenty-dimensional process and are calculated after we transform the sample via the function* f *as in* (47). *An estimation of probabilities assigned to various squares of* $\mathbb{R}^{20}$ *by the target distribution is also presented. The square denoted by "Sq a" means that target assigns probability a inside the square. The best performance is highlighted with bold letters*

Algorithms	ESS(SD)	Median ESS	Sq. 0.9	Sq. 0.99	Sq. 0.999	ESS/Lik.Eval.	ESS/min
ZZ	103,661.4 (6347.7)	104,665.5	0.9008	0.9905	0.9991	$0.3 \cdot 10^{-3}$	124.9
SUZZ(0)	**142,663.2** (1511.3)	**142,382.6**	**0.8998**	**0.9899**	**0.9990**	$3.9 \cdot 10^{-3}$	2847.7
SUZZ(1)	134,561.8 (2453.4)	134,140.1	0.8982	0.9897	**0.9990**	$\mathbf{6.3 \cdot 10^{-3}}$	**4471.3**
TRWM	2767.0 (68.5)	2753.5	0.8994	0.9902	0.9992	$2.8 \cdot 10^{-3}$	1966.8

simulated a continuous time path, in order to construct a sample from the ZZ and the SUZZ algorithms we used the same procedure as in the one-dimensional simulations. We used the δ-skeleton of the process, where δ was chosen after an initial run of the algorithm such that the sample size was roughly equal to the number of direction switches ($N = 10^6$). As mentioned in the one-dimensional simulations, setting δ this way guarantees fairness between the performance evaluation across all algorithms. For TRWM, since we run the algorithm for $N = 10^6$ steps, the sample generated had a size of $N = 10^6$. All simulations were performed using MATLAB in a computer with i7-8550U CPU and 1.80 GHz.

We present our results in Tables 3, 4 and 5. In Tables 3 and 4 we report results on the subexponential(0.5) target, while in Table 5 we report results concerning the Student(3) target. We present average and median ESS across 25 realizations (standard deviation in parenthesis) and empirical probabilities of squares centered around 0, containing 0.9, 0.99 and 0.999 of the mass of the target and denoted Sq. 0.9, Sq 0.99 and Sq. 0.999 respectively. These squares were estimated using mvtnorm and adaptMCMC of R. We also report the median ESS per likelihood evaluation and per minute of implementation time. For all algorithms, we consider the ESS of the first coordinate of the process, computed using the routine coda of R, but we note here that we recovered similar results when using the routine mcmcse to estimate the multivariate ESS of the twenty-dimensional algorithms. For the

TABLE 4

SUZZ, ZZ and TRWM algorithms targeting a twenty-dimensional Subexponential(0.5) *distribution, with density given by* (49). *For SUZZ*(k), *we use the speed function as in* (23). *The algorithms ran until* $N = 10^6$ *switches* (*or steps for the TRWM*) *occurred and the average ESS* (*with standard deviation in a parenthesis*) *along with the median ESS are presented. The median ESS per average likelihood evaluations and per average minutes of implementation time are also presented. All ESS concern the first coordinate of the twenty-dimensional process and are calculated without transforming the sample. The best performance is highlighted with bold letters*

Algorithms	ESS(SD)	Median ESS	ESS/Lik.Eval.	ESS/min
ZZ	54,925.7 (3113.5)	55,460.3	$0.1 \cdot 10^{-3}$	66.2
SUZZ(0)	80,123.7 (1067.4)	80,015.9	$2.2 \cdot 10^{-3}$	1600.3
SUZZ(1)	**92,356.6** (1459.1)	**92,214.6**	$\mathbf{4.3 \cdot 10^{-3}}$	**3073.8**
TRWM	2162.5 (56.7)	2162.6	$2.2 \cdot 10^{-3}$	1544.7

TABLE 5

SUZZ, ZZ and TRWM algorithms targeting a twenty-dimensional Student(3) distribution with scale matrix given by (50). For SUZZ(k), we use the speed function as in (23). The algorithms ran until $N = 10^6$ switches (or steps for the TRWM) occurred and the average ESS (with standard deviation in a parenthesis) along with the median ESS are presented. The median ESS per average likelihood evaluations and per average minutes of implementation time are also presented. All ESS concern the first coordinate of the twenty-dimensional process and are calculated after we transform the sample via the function f as in (47). An estimation of probabilities assigned to various squares of $\mathbb{R}^{20}$ by the target distribution is also presented. The square denoted by "Sq a" means that target assigns probability a inside the square. The best performance is highlighted with bold letters

20-dimensional Student(3), number of switches $N = 10^6$, with space transformation.							
Algorithms	ESS(SD)	Median ESS	Sq. 0.9	Sq. 0.99	Sq. 0.999	ESS/Lik.Eval.	ESS/min
ZZ	16,095.0 (717.8)	16,151.2	0.8980	0.9888	**0.9989**	$0.3 \cdot 10^{-3}$	734.1
SUZZ(0)	**25,882.6** (421.6)	**25,943.5**	0.8978	0.9892	0.9986	**$1.4 \cdot 10^{-3}$**	1005.6
SUZZ(1)	23,002.8 (511.0)	23,052.3	0.8994	**0.9898**	0.9989	$1.1 \cdot 10^{-3}$	**1746.4**
TRWM	1153.3 (80.6)	1158.5	**0.8996**	**0.9902**	**0.9991**	$1.2 \cdot 10^{-3}$	827.5

Student(3) distribution (Table 5), the ESSs were calculated after we transformed the sample via the function (47) so that we can guarantee that the variance of the ESS is finite, and the computation of ESS consistent across all 25 realisations of the chains. We did the same for the subexponential(0.5) distribution (Table 3), but for that target we also present the ESS without any transformation of the sample (Table 4), since the variance of the ESS is finite when estimating the expectation of this target.

All four algorithms provided a decent estimation of the probabilities of the squares, which can increase our trust that all algorithms converged to the right distribution. In terms of ESS, we observe that all SUZZ algorithms vastly outperformed the ZZ algorithm in terms of every criterion we used, that is, ESS per number of switches, per number of likelihood evaluations and per implementation minutes. This shows that using a speed function in the context of PDMP algorithms can lead to significant benefits. Furthermore, the SUZZ algorithms can compare favourably to a state of the art algorithm like the TRWM in all three criteria (ESS per switches, ESS per likelihood evaluations and ESS per implementation minutes). For example, the SUZZ(1) algorithm has twice better ESS per implementation time than the TRWM on the Student target. Notably, if the criterion is ESS per number of switches, which gives a theoretical upper bound on the ESS per likelihood evaluations for the SUZZ algorithm, the SUZZ algorithms perform at least 20 times better than the TRWM.

We also note that although in our simulations the TRWM takes a lot less time to be implemented, there does not seem to be enough space to further reduce the implementation time of the TRWM code. On the other hand, the code of SUZZ is quite more complicated and a more qualified programmer could probably reduce the implementation time even further. More specifically, most of the simulation time was spent in finding the maximum of the rate function over a specific time horizon in order to perform Poisson thinning. If one could reduce the time spent in this type of maximisation subroutines, one could significantly reduce the implementation time of SUZZ. Furthermore, while the likelihood evaluations of TRWM are always equal to the number of steps of the algorithm, one could try to further reduce the number of likelihood evaluations of the SUZZ algorithm if one has access to extra information on the structure of the target. One could also use ideas from [48], for example, by adapting the time horizon over which the optimisation of the rate takes place, taking into account the previous switching times. This can be done without losing any theoretical guarantees since any choice of time horizon leads to stochastically identical algorithms. Furthermore, one can use ideas from [16] to further optimise the Poisson thinning procedure and the implementation time, for example, with the use of automatic differentiation schemes.

Finally, we should note that we tested SUZZ algorithms on targets where Assumption 3.1 is not verified and the process will a.s. explode in finite time. Specifically, we targeted a one-dimensional Cauchy $\pi(x) = \frac{1}{Z}(1 + x^2)^{-1}$ with SUZZ(1), that is, speed function $s(x) = 1 + x^2$. Very quickly there were numerical issues, with MATLAB reporting NaN. A diagnostic test we propose for one to check possible explosivity is to construct a large square (e.g., $[-10^8, 10^8]^d$ for a d-dimensional target) and change the process so that whenever it hits the boundary of that square, a switch of direction occurs. Meanwhile one can count the number of times the process hit the boundary of the square. If the proportion of direction switches due to hitting the boundary over the overall number of switches is large, there is a good chance that the algorithm explodes and should not be used.

APPENDIX A: A FORMAL CONSTRUCTION OF THE SUZZ PROCESS

Formally, the SUZZ process is constructed as follows.

Let $(\tilde{E}_n)_{n \in \mathbb{N}}$ be i.i.d. exponential random variables with mean 1 and $(u_n)_{n \in \mathbb{N}}$ be i.i.d. uniform in $[0, 1]$ random variables, independent of the $\tilde{E}_n$'s. Suppose that the process starts from $(x, \theta) \in E$. Let $\{\Phi_t(x, \theta)\}_{t \geq 0}$ be the flow on $\mathbb{R}^d$ that solves the ODE system

$$(52) \qquad \begin{cases} \dfrac{d\Phi_t(x, \theta)}{dt} = \theta s(\Phi_t(x, \theta)) & t \in [0, t^*(x, \theta)), \\ \Phi_0(x, \theta) = x, \end{cases}$$

where $t^*(x, \theta) = \sup\{t \geq 0 : \Phi_t(x, \theta) \in \mathbb{R}^d\}$ is the explosion time of the flow, with the convention that if the flow does not explode this is set $t^*(x, \theta) = +\infty$. As we will see in Appendix B, if $s \in C^1$, the ODE system (52) has a unique solution, and this solution flow moves in a straight line parallel to $\theta \in \{\pm 1\}^d$. We define for all $(y, \eta) \in E$

$$(53) \qquad \lambda(y, \eta) = \sum_{i=1}^{d} \lambda_i(y, \eta).$$

Let

$$\tau_1 = \inf\left\{t \in [0, t^*(x, \theta)) : \int_0^t \lambda(\Phi_u(x, \theta), \theta) \, du \geq \tilde{E}_1\right\},$$

where we will always use the convention that $\inf \varnothing = +\infty$. Let $T_1 = \tau_1$.

If $\tau_1 = \infty$ we set $(X_t, \Theta_t) = (\Phi_t(x, \theta), \theta)$ for all $t < t^*(x, \theta)$ and $(X_t, \Theta_t) = \partial$ for $t \geq t^*(x, \theta)$.

If $\tau_1 < \infty$, we define i_1 to be the a.s. unique $i \in \{1, \ldots, d\}$ that satisfies

$$u_1 \in \left[\frac{\sum_{k=1}^{i-1} \lambda_k(\Phi_{\tau_1}(x, \theta), \theta)}{\lambda(\Phi_{\tau_1}(x, \theta), \theta)}, \frac{\sum_{k=1}^{i} \lambda_k(\Phi_{\tau_1}(x, \theta), \theta)}{\lambda(\Phi_{\tau_1}(x, \theta), \theta)}\right].$$

We set $(X_t, \Theta_t) = (\Phi_t(x, \theta), \theta)$ for all $0 \leq t < T_1$. Then set $(X_{T_1}, \Theta_{T_1}) = (\Phi_{\tau_1}(x, \theta), F_{i_1}(\theta))$, where $F_{i_1}(\theta) \in \{\pm 1\}^d$ as in (2).

We then continue the construction inductively, for any $n \in \mathbb{N}$. If $T_n < \infty$ and assuming that the process is constructed until time T_n, we then consider the flow $\{\Phi_t(X_{T_n}, \Theta_{T_n})\}_{t \geq 0}$, we let

$$(54) \qquad \tau_{n+1} = \inf\left\{t \in [0, t^*(X_{T_n}, \Theta_{T_n})) : \int_0^t \lambda(\Phi_u(X_{T_n}, \Theta_{T_n}), \Theta_{T_n}) \, du \geq \tilde{E}_{n+1}\right\},$$

and set $T_{n+1} = T_n + \tau_{n+1}$.

If $\tau_{n+1} = \infty$, we set $(X_{T_n+t}, \Theta_{T_n+t}) = (\Phi_t(X_{T_n}, \Theta_{T_n}), \Theta_{T_n})$ for all $t < t^*(X_{T_n}, \Theta_{T_n})$ and $(X_{T_n+t}, \Theta_{T_n+t}) = \partial$ for $t \geq t^*(X_{T_n}, \Theta_{T_n})$.

If $\tau_{n+1} < \infty$, we define i_{n+1} to be the a.s. unique $i \in \{1, \ldots, d\}$ that satisfies

$$u_{n+1} \in \left[\frac{\sum_{k=1}^{i-1} \lambda_k(\Phi_{\tau_{n+1}}(X_{T_n}, \Theta_{T_n}), \Theta_{T_n})}{\lambda(\Phi_{\tau_{n+1}}(X_{T_n}, \Theta_{T_n}), \Theta_{T_n})}, \frac{\sum_{k=1}^{i} \lambda_k(\Phi_{\tau_{n+1}}(X_{T_n}, \Theta_{T_n}), \Theta_{T_n})}{\lambda(\Phi_{\tau_{n+1}}(X_{T_n}, \Theta_{T_n}), \Theta_{T_n})} \right].$$

Then set $(X_{T_n+t}, \Theta_{T_n+t}) = (\Phi_t(X_{T_n}, \Theta_{T_n}), \Theta_{T_n})$ for all $0 \le t < T_{n+1} - T_n$. Then set $(X_{T_{n+1}}, \Theta_{T_{n+1}}) = (\Phi_{\tau_{n+1}}(X_{T_n}, \Theta_{T_n}), F_{i_{n+1}}(\Theta_{T_n}))$.

This defines the process until time ξ as in (9). In the case where $\xi < \infty$, ξ is the first time that the process has had infinitely many switches of direction. We set $(X_t, \Theta_t) = \partial$ for all $t \ge \xi$.

APPENDIX B: SOLUTION OF ODE (8)

Here we explain why the ODE system (8) has a unique solution when $s \in C^1$. Note that the solution flow $\{\Phi_t(x, \theta), t \ge 0\}$, representing the solution after time t when starting from (x, θ), must move in a straight line in $\mathbb{R}^d$, parallel to $\theta \in \{\pm 1\}^d$. Therefore, if the process starts from $(x, \theta) = (x_1, \ldots, x_d; \theta_1, \ldots, \theta_d)$, then at time t the position $\Phi_t(x, \theta) = (X^1(t), \ldots, X^d(t))$ satisfies for all i,

$$(55) \qquad X^i(t) = x_i + \theta_1\theta_i(X^1(t) - x_1) = y_i + \theta_1\theta_i X^1(t),$$

where

$$y_i = x_i - \theta_1\theta_i x_1,$$

and where we omit including the dependence of $X^i(t)$ on (x, θ) for notation convenience.

Therefore, as long as we solve $\{X^1(t), t \ge 0\}$, we will have completely identified the solution $\{\Phi_t(x, \theta), t \ge 0\}$. Also, from (8) and (55) we get that $X^1(t)$ satisfies

$$\frac{dX^1(t)}{dt} = \theta_1 s(X^1(t), y_2 + \theta_1\theta_2 X^1(t), \ldots, y_d + \theta_1\theta_d X^1(t)).$$

Separating the variables we get that $X^1(t)$ satisfies

$$\theta_1 t = \int_{x_1}^{X_t^1} \frac{1}{s(u, y_2 + \theta_1\theta_2 u, \ldots, y_d + \theta_1\theta_d u)} \, du,$$

which has a unique solution

$$(56) \qquad X^1(t) = f^{-1}(f(x_1) + \theta_1 t),$$

where

$$f(a) = \int_0^a \frac{1}{s(u, y_2 + \theta_1\theta_2 u, \ldots, y_d + \theta_1\theta_d u)} \, du$$

is invertible since it is continuous and strictly increasing. The solution is defined until the explosion time

$$(57) \qquad t^*(x, \theta) = \theta_1 \int_{x_1}^{\theta_1 \cdot \infty} \frac{1}{s(u, y_2 + \theta_1\theta_2 u, \ldots, y_d + \theta_1\theta_d u)} \, du \in (0, +\infty],$$

where we use the convention $\theta_1 \cdot \infty$ to denote $+\infty$ if $\theta_1 = +1$ and $-\infty$ if $\theta_1 = -1$.

A natural family of speed functions, for which the ODE (8) can be solved analytically is

$$s(x) = (1 + x^T B x)^{\frac{1+n}{2}},$$

where $n = 0, 1, 2, 3, \ldots$ and B a positive definite matrix. For $n = 0$, the ODE solution is nonexplosive for all starting value $(x, \theta) \in \mathbb{R}^d \times \{-1, +1\}^d$, while for $n > 0$, the solution explodes in finite time. Here we will present two specific examples of speed functions that are used throughout the numerical simulations, along with the solution to the ODE they induce. More details can be found in Section 5.8.2 of [52].

- *Nonexplosive deterministic dynamics:* A natural speed function to use is

$$s(x) = (1 + x^T x)^{1/2}.$$

The solution to the ODE (8) starting from $(x, \theta) \in E$, is given by (55), where

$$X^1(t) = \frac{b^2(t) - a}{2b(t)} - \frac{c}{d}, \quad t \in [0, +\infty),$$

and

$$b(t) = \left(Y_0 + \sqrt{Y_0^2 + a}\right) \cdot \exp\{\sqrt{d}\theta_1 t\}, \quad Y_0 = x_1 + \frac{c}{d},$$

$$a = \frac{1 + \|y\|_2^2}{d} - \frac{c^2}{d^2}, \quad c = \theta_1(y \cdot \theta), \quad y_i = x_i - \theta_1 \theta_i x_1.$$

- *Explosive deterministic dynamics:* We use the speed function

$$s(x) = (1 + x^T x).$$

The solution to the ODE (8) starting from $(x, \theta) \in E$, is given by (55), where

$$X^1(t) = -c_1 + c_2 \tan\left(\arctan\left(\frac{x_1 + c_1}{c_2}\right) + \theta_1 c_2 dt\right), \quad t \in [0, t^*(x, \theta)),$$

and

$$c_2 = \sqrt{\frac{1 + y_2^2 + \cdots + y_d^2}{d} - c_1^2}, \quad c_1 = \frac{(y \cdot \theta)}{d}\theta_1, \quad y_i = x_i - \theta_1 \theta_i x_1,$$

and where the explosion time is

$$t^*(x, \theta) = \frac{\frac{\pi}{2} - \theta_1 \arctan\left(\frac{x_1 + c_1}{c_2}\right)}{c_2 d}.$$

APPENDIX C: THE GENERATOR

PROPOSITION C.1. *Let $(Z_t)_{t \geq 0} = (X_t, \Theta_t)_{t \geq 0}$ be a SUZZ process with strictly positive speed function $s \in C^2$. For any function $f \in C^1(E)$ and any $(x, \theta) \in E$,*

(58)
$$\lim_{t \to 0} \frac{\mathbb{E}_{x,\theta}[f(Z_t)] - f(x, \theta)}{t}$$

$$= \mathcal{L}f(x, \theta)$$

$$:= \sum_{i=1}^{d} \theta_i s(x) \partial_i f(x, \theta) + \lambda_i(x, \theta)\big(f\big(x, F_i(\theta)\big) - f(x, \theta)\big).$$

Before we prove Proposition C.1 we need the following technical results.

DEFINITION C.1. Let $m \in \mathbb{N}$. We denote with O_m the ball in $\mathbb{R}^d$, centred around 0, having radius m. Recall that $\zeta_m = \inf\{t \geq 0 : X_t \notin O_m\}$ is the first exit time of the O_m for the SUZZ process and $\zeta = \lim_{m \to \infty} \zeta_m$ is the explosion time of the process. Finally, recall that $T_1, T_2, \ldots$ are the switching times of the process and $\xi = \lim_{n \to \infty} T_n$.

LEMMA C.1. *Assume that the speed function $s \in C^1$ is strictly positive and the rate functions λ_i are locally bounded for all i. Then almost surely $\zeta_m < \xi$ for all $m \in \mathbb{N}$. Therefore, a.s. $\zeta \leq \xi$.*

PROOF OF LEMMA C.1. Let λ be as in (53) and let $m \in \mathbb{N}$. Let $\bar{\lambda}$ be an upper bound of λ on O_m and let $\tilde{E}_1, \tilde{E}_2, \ldots$ be a configuration of the i.i.d. exponential random variables, with expectation 1, used to construct the $(X_t, \Theta_t)_{t \geq 0}$ process, such that $\xi \leq \zeta_m$. We will show that the event of such a configuration has probability zero. In that configuration, for all $t < \xi$, we have $t < \zeta_m$, therefore $X_t \in O_m$. By the definition of the switching times T_k in (54) (and writing $T_0 = 0$) we get

$$\sum_{k=1}^{\infty} \tilde{E}_k = \sum_{k=1}^{\infty} \int_{T_{k-1}}^{T_k} \lambda(X_t, \Theta_t) \, dt \leq \xi \bar{\lambda},$$

and therefore

$$\mathbb{P}(\{\xi < \infty\} \cap \{\xi \leq \zeta_m\}) \leq \mathbb{P}\left(\sum_{k=1}^{\infty} \tilde{E}_k \leq \xi \bar{\lambda} < \infty\right) = 0.$$

Let t_m be the maximum time it takes for a flow that starts from inside O_m and solves (8) to exit O_m. For any n, on the event $\{\tilde{E}_n \geq \bar{\lambda} t_m\}$, if the process has not escaped the ball O_m until the $n - 1$th switch, it does so following the dynamics before the nth switch occurs. Since $\mathbb{P}(\tilde{E}_n \geq \bar{\lambda} t_m) = a > 0$ we have for all n, $\mathbb{P}(\zeta_m > T_n) \leq (1 - a)^n$ and therefore $\mathbb{P}(\zeta_m = +\infty) = 0$.

Overall this gives

$$\mathbb{P}(\xi \leq \zeta_m) = \mathbb{P}(\{\xi \leq \zeta_m\} \cap \{\xi < \infty\}) + \mathbb{P}(\{\xi \leq \zeta_m\} \cap \{\xi = \infty\}) \leq \mathbb{P}(\zeta_m = \infty) = 0,$$

which concludes the proof. $\square$

LEMMA C.2. *If the speed function $s \in C^1$ and the rate functions are locally bounded, then for all $x \in \mathbb{R}^d$ and any neighbourhood U_x of x there exists a time $t > 0$ such that for any $\theta \in \{\pm 1\}^d$ if the SUZZ starts from (x, θ), then $X_s \in U_x$ for all $0 \leq s \leq t$.*

PROOF OF LEMMA C.2. Let $x \in \mathbb{R}^d$ and consider a small neighbourhood U_x of x. Let $\bar{s}$ be an upper bound for s on U_x. Take t small enough so that $t\sqrt{d}\bar{s} < \text{dist}(x, \partial U_x)$, where dist denotes the Euclidean distance between a point and a set. This way, any path starting from x, and following a straight line, with speed function s in each coordinate, for time less than t, will not have exit U_x. Then any path moving in directions $\{\pm 1\}^d$, with speed function $s(x)$ in each component, that switches direction finitely many times, will not have exit U_x. From Lemma C.1, a.s. the original Zig-Zag process will switch direction finitely many times until it exits the bounded set U_x and this proves that the process a.s. stays inside U_x until time t. $\square$

PROOF OF PROPOSITION C.1. Fix a starting point $(x, \theta) \in \mathbb{R}^d$. We know from Lemma C.2 that for some neighbourhood U_x of x and for small t_0, if the process starts from (x, θ), then a.s. $X_s \in U_x$ for all $s \in [0, t_0]$. Therefore, the quantity $\mathbb{E}_{x,\theta}[f(Z_s)]$ is well-defined for all $s \in [0, t_0]$ so the limit makes sense. For the rest of the proof we will always assume that $t \leq t_0$.

Write $S_i(t) = \{$the i coordinate switches before time t and is the first coordinate to switch$\}$, for $i = 1, \ldots, d$ and $S_0(t) = \{$ no coordinate switches until time $t\}$. Note that if the process starts from (x, θ) and if T_i is the first arrival time of the Poisson process with intensity $t \to \lambda_i(\Phi_t(x, \theta), \theta)$, then $\mathbb{P}(T_i \geq t) = \exp\{-\int_0^t \lambda_i(\Phi_u(x, \theta), \theta) \, du\}$ therefore the density of T_i is

$$f_{T_i}(t) = \lambda_i(\Phi_t(x, \theta), \theta) \exp\left\{-\int_0^t \lambda_i(\Phi_u(x, \theta), \theta) \, du\right\}.$$

For $f \in C^1(E)$, conditioning on the coordinate which was the first to switch before t (or whether no switch occurred), we write

$$
\frac{\mathbb{E}_{x,\theta}[f(X_t, \Theta_t)] - f(x,\theta)}{t}
$$

(59)
$$
= \mathbb{P}_{x,\theta}(S_0(t))\frac{f(\Phi_t(x,\theta),\theta) - f(x,\theta)}{t}
$$

$$
+ \sum_{i=1}^{d} \frac{\mathbb{E}_{x,\theta}[(f(X_t, \Theta_t) - f(x,\theta))1_{S_i(t)}]}{t}.
$$

If we write $\lambda(x,\theta) = \sum_{i=1}^{d} \lambda_i(x,\theta)$ then we observe that

$$
\mathbb{P}_{x,\theta}(S_0(t)) = \exp\left\{-\int_0^t \lambda(\Phi_u(x,\theta),\theta)\,du\right\} \xrightarrow{t\to 0} 1
$$

therefore

(60)
$$
\lim_{t\to 0} \mathbb{P}_{x,\theta}(S_0(t))\frac{f(\Phi_t(x,\theta),\theta) - f(x,\theta)}{t} = \sum_{i=1}^{d} \theta_i s(x)\partial_i f(x,\theta).
$$

Furthermore, for all $i \in \{1, \ldots, d\}$, conditioning on the first switch occurring at time $u \le t$ and being of the ith coordinate, we have

$$
\mathbb{E}_{x,\theta}\big[(f(X_t, \Theta_t) - f(x,\theta))1_{S_i(t)}\big]
$$

$$
= \int_0^t \mathbb{E}_{\Phi_u(x,\theta), F_i(\theta)}\big[f(X_{t-u}, \Theta_{t-u}) - f(x,\theta)\big]
$$

$$
\times \mathbb{P}_{x,\theta}(\text{no switches until time } u \text{ for any component } j \ne i)
$$

$$
\cdot f_{T_i}(u)\,du
$$

$$
= \int_0^t \mathbb{E}_{\Phi_u(x,\theta), F_i(\theta)}\big[f(X_{t-u}, \Theta_{t-u}) - f(x,\theta)\big]\lambda_i(\Phi_u(x,\theta),\theta)
$$

(61)
$$
\times \exp\left\{-\int_0^u \lambda(\Phi_{u'}(x,\theta),\theta)\,du'\right\}du
$$

$$
= \int_0^t \mathbb{E}_{\Phi_u(x,\theta), F_i(\theta)}\big[f(X_{t-u}, \Theta_{t-u}) - f(x, F_i(\theta))\big]\lambda_i(\Phi_u(x,\theta),\theta)
$$

$$
\times \exp\left\{-\int_0^u \lambda(\Phi_{u'}(x,\theta),\theta)\,du'\right\}du
$$

$$
+ \int_0^t (f(x, F_i(\theta)) - f(x,\theta))\lambda_i(\Phi_u(x,\theta),\theta)
$$

$$
\times \exp\left\{-\int_0^u \lambda(\Phi_{u'}(x,\theta),\theta)\,du'\right\}du.
$$

Let $\epsilon > 0$. Since $f \in C^0$ we can assume that the neighbourhood U_x is small enough such that if $y \in U_x$, then $|f(y, F_i(\theta)) - f(x, F_i(\theta))| < \frac{\epsilon}{4}$. From Lemma C.2 we know that if the process starts from (x,θ), then for any path that switches direction finitely many times by time t, we have for all $u' \le t$, $X_{u'} \in U_x$. Therefore, for any $u \le t$, if the process starts from $(\Phi_u(x,\theta), F_i(\theta))$, then $X_{t-u} \in U_x$ a.s. Let M be such that for any $y \in U_x$ and $\eta \in \{-1, +1\}^d$, $|f(y,\eta)| \le M$. Then, for any $u \le t$,

$$
\mathbb{E}_{\Phi_u(x,\theta), F_i(\theta)}\big[|f(X_{t-u}, \Theta_{t-u}) - f(x, F_i(\theta))|\big]
$$

$$
\le \mathbb{P}_{\Phi_u(x,\theta), F_i(\theta)}(\text{no switch by time } t - u)|f(\Phi_{t-u}(\Phi_u(x,\theta), F_i(\theta)), F_i(\theta))
$$

$$- f(x, F_i(\theta))|$$

$$+ \mathbb{P}_{\Phi_u(x,\theta), F_i(\theta)}(\text{switch occurs by time } t - u)2M < \frac{\epsilon}{4} + \frac{\epsilon}{4} = \frac{\epsilon}{2},$$

when t is small enough since λ is bounded on U_x. Therefore, since $\lambda_i \in C^0$ for all i and $\Phi_u(x, \theta)$ is continuous over u, we get

(62)
$$\frac{1}{t}\left|\int_0^t \mathbb{E}_{\Phi_u(x,\theta), F_i(\theta)}\big[f(X_{t-u}, \Theta_{t-u}) - f(x, F_i(\theta))\big]\lambda_i(\Phi_u(x, \theta), \theta)\right.$$
$$\times \exp\left\{-\int_0^u \lambda(\Phi_{u'}(x, \theta), \theta) \, du'\right\} du\bigg|$$
$$\leq \epsilon \lambda_i(x, \theta).$$

Furthermore, from the fundamental theorem of calculus,

(63)
$$\frac{1}{t}\int_0^t (f(x, F_i(\theta)) - f(x, \theta))\lambda_i(\Phi_u(x, \theta), \theta) \exp\left\{-\int_0^u \lambda(\Phi_{u'}(x, \theta), \theta) \, du'\right\} du$$
$$\xrightarrow{t\to 0} (f(x, F_i(\theta)) - f(x, \theta))\lambda_i(x, \theta).$$

Combining (61), (62) and (63) we get that

$$\lim_{t\to 0} \frac{\mathbb{E}_{x,\theta}[(f(X_t, \Theta_t) - f(x, \theta))1_{S_i(t)}]}{t} = \lambda_i(x, \theta)\big(f(x, F_i(\theta)) - f(x, \theta)\big)$$

for all $i \in \{1, \ldots, d\}$, which combined with (59) and (60) proves the result. $\square$

APPENDIX D: PROOF OF NONEXPLOSIVITY OF SUZZ (THEOREM 3.1)

Before we prove nonexplosivity of the process, we prove the following useful result, which is of independent interest. The result states that under Assumption 3.1, independently of the starting point, the process cannot follow the deterministic dynamics until the explosion time, but has to switch direction beforehand. Naturally, this is strongly connected with the notion of nonexplosion and justifies the existence of Assumption 3.1.

PROPOSITION D.1. *Assume that $s \in C^2$ is strictly positive, the rates satisfy* (12) *and Assumption* 3.1 *hods. For any starting point (x, θ), let $t^*(x, \theta)$ be the explosion time of the deterministic flow solving the ODE* (52), *and let T_1 be the first switching time of the process. Then $\mathbb{P}_{x,\theta}(T_1 < t^*(x, \theta)) = 1$.*

PROOF OF PROPOSITION D.1. Suppose that the process starts from $(x, \theta) = (x_1, \ldots, x_d; \theta_1, \ldots, \theta_d)$. As proven in Appendix B, for any $t < T_1$ we have $X_t = (X_t^1, \ldots, X_t^d)$ with X_t^i as in (55).

Consider the Poisson process with rate $\{m(t) = \lambda(X_t, \theta) = \sum_{i=1}^d \lambda_i(X_t, \theta), t \geq 0\}$. If y_i as in (55), then using the fact that $a^+ \geq a$ for any $a \in \mathbb{R}$, and using the notation $\theta_1 \cdot \infty$ to denote $+\infty$ if $\theta_1 = +1$ or $-\infty$ when $\theta_1 = -1$, we get

$$\int_0^{t^*(x,\theta)} m(t) \, dt$$
$$= \int_0^{t^*(x,\theta)} \lambda(X_t^1, y_2 + \theta_1\theta_2 X_t^1, \ldots, y_d + \theta_1\theta_d X_t^1) \, dt$$
$$= \int_{x_1}^{\theta_1 \cdot \infty} \lambda(u, y_2 + \theta_1\theta_2 u, \ldots, y_d + \theta_1\theta_d u)$$

$$\times \frac{1}{s(u, y_2 + \theta_1\theta_2 u, \ldots, y_d + \theta_1\theta_d u)}\theta_1\, du$$

$$\geq \int_{x_1}^{\theta_1 \cdot \infty} \sum_{i=1}^{d} \theta_1\theta_i\, \partial_i U(u, y_2 + \theta_1\theta_2 u, \ldots, y_d + \theta_1\theta_d u)$$

$$- \theta_1\theta_i \frac{\partial_i s(u, y_2 + \theta_1\theta_2 u, \ldots, y_d + \theta_1\theta_d u)}{s(u, y_2 + \theta_1\theta_2 u, \ldots, y_d + \theta_1\theta_d u)}\, du$$

$$= \int_{x_1}^{\theta_1 \cdot \infty} \frac{d}{du}[U(u, y_2 + \theta_1\theta_2 u, \ldots, y_d + \theta_1\theta_d u)$$

$$- \log s(u, y_2 + \theta_1\theta_2 u, \ldots, y_d + \theta_1\theta_d u)]\, du$$

$$= \lim_{u \to \theta_1 \cdot \infty} U(u, c_2 + \theta_1\theta_2 u, \ldots, c_d + \theta_1\theta_d u)$$

$$- \log s(u, c_2 + \theta_1\theta_2 u, \ldots, c_d + \theta_1\theta_d u) - C = +\infty$$

by Assumption 3.1. Therefore, $\mathbb{P}_{x,\theta}(T_1 \geq t^*(x, \theta)) = \exp\{-\int_0^{t^*(x,\theta)} m(t)\, dt\} = 0$. ☐

To prove that the process will not explode, we use standard techniques from [42] which depend on the generator of the process. Although we can define the operator $\mathcal{L}$ in (58), we cannot immediately conclude that this is the strong generator of the process. This is because the strong generator is defined through uniform convergence and we only defined $\mathcal{L}$ in (58) as a point-wise limit. This complicates the proof. However, the techniques in [42] only require us to use the generator of the process restricted in a bounded domain, which we introduce now.

DEFINITION D.1. Let O_m be the ball of radius m centered around 0 and let $E_m = O_m \times \{-1, +1\}^d$. Starting from $(x, \theta) \in E$ we define the stopped m-process as the restriction of the SUZZ on O_m, stopped when exiting O_m, that is, $(Z_t^m)_{t \geq 0} = (X_t^m, \Theta_t^m)_{t \geq 0} = (X_{t \wedge \zeta_m}, \Theta_{t \wedge \zeta_m})_{t \geq 0}$.

Since the switching rate of Z^m is bounded as the process is defined on a bounded set and λ_i are locally bounded, we have that for any $T > 0$, if N_T is the number of switching events before time T, then $\mathbb{E}_{x,\theta}[N_T] < \infty$ for any $(x, \theta) \in E_m$. Therefore Z^m is a PDMP that can be seen in the setting of [18] and we have the following as a result of Theorem 5.5 of [18].

PROPOSITION D.2. *Let $\mathcal{L}$ be the operator defined in (58). The extended generator $\mathcal{L}^m$ for Z^m has domain $\mathcal{D}(\mathcal{L}^m) \supset C^1(E)$ and for any function $f \in C^1(E)$ we have*

$$\mathcal{L}^m f(x, \theta) = \mathcal{L}f(x, \theta)1_{x \in O_m}.$$

Let $n \in \mathbb{N}$ such that (17) holds. Let A_i be as in (12). For some $a \in (0, 1)$ and $\delta > 0$, consider the function

$$(64) \qquad V(x, \theta) = \exp\left\{aU(x) - a\log s(x) + \sum_{i=1}^{d} \phi(\theta_i A_i(x))\right\},$$

where

$$(65) \qquad \phi(s) = \frac{1}{2}\operatorname{sgn}(s)h_{n+1}(\delta|s|),$$

and h_{n+1} as in (16). The proof of nonexplosion relies on the following lemma.

LEMMA D.1. *Assume that the rates satisfy* (12) *and Assumptions* 3.1, 3.2, 3.3 *and* 3.4 *hold. Let* $\mathcal{L}$ *be the operator defined in* (58). *Then there exist* $a \in (0, 1)$ *and* $\delta > 0$ *for which* V *introduced in* (64) *is a norm-like function, that is,* $\lim_{\|x\| \to \infty} V(x, \theta) = +\infty$ *and there exists a compact set* C *and* $b, c > 0$ *such that for all* $(x, \theta) \in E$

$$(66) \qquad \mathcal{L}V(x, \theta) \leq -cV(x, \theta) + b\mathbf{1}_{(x,\theta) \in C}.$$

PROOF OF LEMMA D.1. One can verify that $V \in C^1$ therefore $V \in D(\mathcal{L})$. Note that

$$V(x, F_i(\theta)) - V(x, \theta) = V(x, \theta)(\exp\{\phi(-\theta_i A_i(x)) - \phi(\theta_i A_i(x))\} - 1).$$

We then calculate

$$(67) \qquad \frac{\mathcal{L}V(x, \theta)}{V(x, \theta)} = \sum_{i=1}^{d} \Bigg\{ \theta_i a A_i(x) + \sum_{j=1}^{d} \theta_i \theta_j s(x) \partial_i A_j(x) \phi'(\theta_j A_j(x)) $$
$$+ [(\theta_i A_i(x))^+ + \gamma_i(x)](\exp\{\phi(-\theta_i A_i(x)) - \phi(\theta_i A_i(x))\} - 1) \Bigg\}.$$

Note that

$$\exp\{\phi(-u) - \phi(u)\} = (1 + h_n(\delta|u|))^{-\operatorname{sgn}(u)}$$

and

$$\phi'(u) = \frac{\delta}{2} \frac{1}{(1 + \delta|u|)} \prod_{k=1}^{n} \frac{1}{1 + h_k(\delta|u|)} \leq \frac{\delta}{2} \frac{1}{(1 + \delta|u|)(1 + \log(1 + \delta|u|))}.$$

Consider the ith component of the sum in the RHS of (67) and the following cases.

Case 1: $\theta_i A_i(x) \geq 0$. Then the ith component of the sum in the RHS of (67) can be written as

$$a|\theta_i A_i(x)| + \sum_{j=1}^{d} \theta_i \theta_j s(x) \partial_i A_j(x) \phi'(\theta_j A_j(x))$$

$$(68) \qquad + (|\theta_i A_i(x)| + \gamma_i(x))\left(\frac{1}{1 + h_n(\delta|A_i(x)|)} - 1\right)$$

$$\leq |A_i(x)|\left[a - 1 + \frac{1}{1 + h_n(\delta|A_i(x)|)}\right] + \sum_{j=1}^{d} \theta_i \theta_j s(x) \partial_i A_j(x) \phi'(\theta_j A_j(x)),$$

where we used that $\gamma_i(x) \geq 0$.

Case 2: $\theta_i A_i(x) < 0$. Then the ith component of the sum in the RHS of (67) can be written as

$$a\theta_i A_i(x) + \sum_{j=1}^{d} \theta_i \theta_j s(x) \partial_i A_j(x) \phi'(\theta_j A_j(x))$$

$$+ [(\theta_i A_i(x))^+ + \gamma_i(x)] \underbrace{\log(1 + \cdots \log(1 + \delta|A_i(x)|) \cdots)}_{n}$$

$$\leq -a|A_i(x)| + \bar{\gamma} \underbrace{\log(1 + \cdots \log(1 + \delta|A_i(x)|) \cdots)}_{n}$$

$$+ \sum_{j=1}^{d} \theta_i \theta_j s(x) \partial_i A_j(x) \phi'(\theta_j A_j(x))$$

$$\leq |A_i(x)|(-a + \bar{\gamma}\delta) + \sum_{j=1}^{d} \theta_i \theta_j s(x) \partial_i A_j(x) \phi'(\theta_j A_j(x)),$$

where we used that $\log(1 + x) \leq x$. Combining the two different cases, we get overall that,

$$
\begin{aligned}
(69) \quad \frac{\mathcal{L}V(x,\theta)}{V(x,\theta)} &\leq \sum_{i=1}^{d} |A_i(x)| \max\left\{ a - 1 + \frac{1}{1 + h_n(\delta|A_i(x)|)}, -a + \delta\bar{\gamma} \right\} \\
&\quad + \sum_{i=1}^{d} \sum_{j=1}^{d} \theta_i \theta_j s(x) \partial_i A_j(x) \phi'(\theta_j A_j(x)).
\end{aligned}
$$

Let us set $\epsilon = \bar{\gamma} > 0$, if $\bar{\gamma} > 0$ or $0 < \epsilon < A/(2d^2)$, if $\bar{\gamma} = 0$. Since we have assumed that $A > 3d\bar{\gamma}$ in Assumption 3.2, we get that $\frac{A}{d} > \bar{\gamma} + 2\epsilon$. We will choose $\delta > 0$ small enough to be specified later and given δ, we set $a = \delta\bar{\gamma} + \delta\epsilon$. Then the second part of the maximum of (69) is equal to $-\delta\epsilon < 0$.

Consider the function

$$f(z) = \max\left\{ -\delta\epsilon, \delta\epsilon + \delta\bar{\gamma} - 1 + \frac{1}{1 + h_n(\delta z)} \right\},$$

so that the first term of the RHS of (69) is equal to

$$\sum_{i=1}^{d} |A_i(x)| f(|A_i(x)|).$$

Our goal will be to show that $\sum_{i=1}^{d} |A_i(x)| f(|A_i(x)|) < 0$ for $\|x\|$ large enough. One can verify that

$$f(A) < 0 \quad \Longleftrightarrow \quad A \geq P(\delta) = \frac{1}{\delta} h_n^{-1}\left(\frac{\delta\epsilon + \delta\bar{\gamma}}{1 - \delta\epsilon - \delta\bar{\gamma}} \right)$$

and

$$f(A) = -\delta\epsilon \quad \Longleftrightarrow \quad A \geq M(\delta) = \frac{1}{\delta} h_n^{-1}\left(\frac{2\delta\epsilon + \delta\bar{\gamma}}{1 - 2\delta\epsilon - \delta\bar{\gamma}} \right),$$

where h_n as in (16). Now, from L'Hôpital's rule,

$$\lim_{\delta \to 0} M(\delta) = \bar{\gamma} + 2\epsilon, \qquad \lim_{\delta \to 0} P(\delta) = \bar{\gamma} + \epsilon$$

so if we choose δ small enough, we have $M(\delta) < A/d$. Suppose $k = \arg\max\{|A_i(x)| : i = 1, \ldots, d\}$ so that

$$|A_k(x)| \geq \frac{\sum_{i=1}^{d} |A_i(x)|}{d} > \frac{A}{d} > M(\delta).$$

Therefore

$$(70) \quad |A_k(x)| f(|A_k(x)|) \leq -|A_k(x)|\delta\epsilon \leq -\frac{\sum_{i=1}^{d} |A_i(x)|}{d}\delta\epsilon.$$

For any other coordinate i, the contribution to the sum $\sum_{i=1}^{d} |A_i(x)| f(|A_i(x)|)$ will be positive if and only if $|A_i(x)| \leq P(\delta)$. Then, using that $\frac{1}{1 + h_n(\delta z)} \leq 1$, we can bound

$$f(z) \leq \delta\epsilon + \delta\bar{\gamma}$$

for $z \geq 0$, and therefore

$$(71) \qquad \sum_{i \neq k}^{d} |A_i(x)| f(|A_i(x)|) \leq (d-1) P(\delta)(\delta \epsilon + \delta \bar{\gamma}).$$

Recall that when $\bar{\gamma} > 0$, we have picked $\epsilon = \bar{\gamma}$, so due to (14) we get

$$(72) \qquad A > d(d-1) \frac{(\bar{\gamma} + \epsilon)^2}{\epsilon}.$$

On the other hand, if $\bar{\gamma} = 0$ we have picked $\epsilon < A/(2d^2)$ so (72) holds in this case as well. Therefore

$$\lim_{\delta \to 0} \frac{1}{\sum_{i=1}^{d} |A_i(x)|} (d-1)(\bar{\gamma} + \epsilon) P(\delta) - \frac{\epsilon}{d} = \frac{1}{\sum_{i=1}^{d} |A_i(x)|} (d-1)(\bar{\gamma} + \epsilon)^2 - \frac{\epsilon}{d} < 0.$$

Combining this with (70) and (71) we get

$$\sum_{i=1}^{d} |A_i(x)| f(|A_i(x)|) = |A_k(x)| f(A_k(x)) + \sum_{i \neq k} |A_i(x)| f(|A_i(x)|)$$

$$(73) \qquad \leq \delta \sum_{i=1}^{d} |A_i(x)| \left[-\frac{\epsilon}{d} + \frac{1}{\sum_{i=1}^{d} |A_i(x)|} (d-1)(\bar{\gamma} + \epsilon) P(\delta) \right]$$

$$\leq -c \sum_{i=1}^{d} |A_i(x)|$$

for some $c > 0$, assuming δ is small enough.

To finish the proof of the drift condition (66), let us consider the last term of the RHS in (69). Here, due to (18) and assuming that $x \notin C$ for some compact set large enough, we can write

$$\sum_{j=1}^{d} \theta_i \theta_j s(x) \partial_i A_j(x) \phi'(\theta_j A_j(x))$$

$$\leq \sum_{i,j=1}^{d} \frac{\delta}{2} \frac{s(x) |\partial_i A_j(x)|}{(1 + \delta |A_j(x)|)(1 + \log(1 + \delta |A_j(x)|))}$$

$$(74) \qquad = \left(\sum_{k=1}^{d} |A_k(x)| \right) \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{s(x) |\partial_i A_j(x)|}{\sum_{k=1}^{d} |A_k(x)|}$$

$$\times \frac{1}{(\delta^{-1} + |A_j(x)|)(1 + \log(1 + \delta |A_j(x)|))}$$

$$\leq \sum_{i=1}^{d} |A_i(x)| \frac{c}{2}.$$

Then, combining (69), (73) and (74) we get for $x \notin C$,

$$\frac{\mathcal{L}V(x, \theta)}{V(x, \theta)} \leq -\frac{c}{2} \sum_{i=1}^{d} |A_i(x)| \leq -\frac{c}{2} A,$$

and this proves (66) since V and $\mathcal{L}V$ are bounded on C.

Finally, we need to prove that $\lim_{\|x\|\to\infty} V(x,\theta) = +\infty$. For this, we can find $C, C' > 0$ such that for all $\|x\|$ large enough,

$$
\begin{aligned}
V(x,\theta) &\geq\quad C \exp\left\{ aU(x) - a\log s(x) - \sum_{i=1}^{d} \frac{1}{2} h_{n+1}(\delta|A_i(x)|) \right\} \\
&\geq\quad C \exp\{aU(x) - a\log s(x)\} \prod_{i=1}^{d}(1 + h_n(\delta|A_i(x)|))^{-1/2} \\
&\geq\quad C \exp\{aU(x) - a\log s(x)\}(1 + h_n(s(x)\|\nabla(U(x) - \log s(x))\|_1))^{-d/2} \\
&\geq\quad C'((U(x) - \log s(x))(1 + h_n(s(x)\|\nabla(U(x) - \log s(x))\|_1))^{-1})^{d/2} \\
&\xrightarrow{\|x\|\to\infty} +\infty,
\end{aligned}
$$

due to (17) and Assumption 3.1, and where we used that $\delta \leq 1$ in the third inequality. This completes the proof. □

PROOF OF THEOREM 3.1.   Under the assumptions of Theorem 3.1, we get from Lemma D.1 and Proposition D.2 that there exists a norm-like function V and constants $c, b > 0$ such that $\mathcal{L}^m V(x,\theta) \leq cV(x,\theta) + b$ for all $m \in \mathbb{N}$. The assumptions of Theorem 2.1 in [42] are satisfied and this proves that the process is nonexplosive, that is, if ζ as in (11) then $\zeta = +\infty$ a.s. Finally, from Lemma C.1 if ξ as in (9) then $\xi = +\infty$ a.s.   □

APPENDIX E: PROOF OF THEOREM 3.2 (INVARIANT MEASURE)

For this section, we first recall the definition of the strong generator of the process.

DEFINITION E.1.   Let $(P^t)_{t\geq 0}$ be the transition semigroup of the process. We define $\mathcal{D}(\mathcal{A})$ to be the set of all the Borel functions f such that the limit

$$
\lim_{t\to 0} \frac{P^t f - f}{t}
$$

exists in the uniform norm (over $(x,\theta) \in E$). We define the strong generator as the operator $\mathcal{A}$, acting on any $f \in \mathcal{D}(\mathcal{A})$ as

$$
\mathcal{A}f(x,\theta) = \lim_{t\to 0} \frac{P^t f(x,\theta) - f(x,\theta)}{t}.
$$

We begin by formally proving that a large class of functions belong to the domain of the strong generator of the SUZZ process and for these functions the strong generator is given by the operator $\mathcal{L}$ introduced in (58).

LEMMA E.1.   *Let us assume that the rates satisfy* (12), *that Assumptions* 3.2, 3.3, 3.4 *and* 3.5 *hold and that* $(Z_t)_{t\geq 0}$ *is a SUZZ process with speed function* s. *If* $f \in C_c^1(E)$ *then* f *is in the domain of the strong generator* $\mathcal{A}$ *of* Z *and* $\mathcal{A}f = \mathcal{L}f$, *where* $\mathcal{L}$ *is the operator defined in* (58). *This means that*

$$
\left| \frac{\mathbb{E}_{x,\theta}[f(X_t, \Theta_t)] - f(x,\theta)}{t} - \mathcal{L}f(x,\theta) \right| \xrightarrow{t\to 0} 0
$$

uniformly in $(x,\theta) \in E$.

PROOF OF LEMMA E.1. Let K be a compact set that contains the support of f and let

$$K' = \{(x, \theta) : \text{there exists a } y, \text{ with } \|x - y\| < \epsilon, (y, \theta) \in K\} = K + \epsilon B(0, 1)$$

for some $\epsilon > 0$. Let $\bar{s}$ be an upper bound on the speed function s on K'. Then for all $t < t_1 = \epsilon/(\sqrt{d}\bar{s})$, if the process starts from any $(x, \theta) \notin K'$, then the process will not have hit K until time t, and since the support of f is contained in K, $P^t f(x, \theta) = \mathbb{E}_{x,\theta}[f(X_t)] = 0$, for all $(x, \theta) \notin K'$. Note also that for all $(x, \theta) \notin K'$, $\mathcal{L}f(x, \theta) = 0$ and $f(x, \theta) = 0$.

Now, let us focus on $(x, \theta) \in K'$. Pick

$$K'' = K' + \epsilon B(0, 1)$$

and let $\bar{\bar{s}}$ be an upper bound of s on K''. Then for all $t \leq t_2 = \epsilon/(\sqrt{d}\bar{\bar{s}}) < t_1$, the process starting from K' will not have exited K'' by time t and if we cover K'' by some O_m for some large m, then a.s. $Z(t) = Z^m(t)$ for all $t \leq t_2$ as long as we start from somewhere in K'. Then, for any $(x, \theta) \in K'$ and any $t < t_2$

$$\frac{\mathbb{E}_{x,\theta}[f(Z_t)] - f(x, \theta)}{t} - \mathcal{L}f(x, \theta) = \frac{\mathbb{E}_{x,\theta}[f(Z_t^m)] - f(x, \theta)}{t} - \mathcal{L}^m f(x, \theta),$$

so overall for all $t < t_2$

$$\sup_{(x,\theta) \in E} \left| \frac{\mathbb{E}_{x,\theta}[f(Z_t)] - f(x, \theta)}{t} - \mathcal{L}f(x, \theta) \right|$$

$$\leq \sup_{(x,\theta) \in K'} \left| \frac{\mathbb{E}_{x,\theta}[f(Z_t^m)] - f(x, \theta)}{t} - \mathcal{L}^m f(x, \theta) \right| \xrightarrow{t \to 0} 0,$$

where the convergence can be seen to hold using the proof of Proposition 15b of [21]. □

We also have the following.

LEMMA E.2. *Assume that the assumptions of Lemma* E.1 *hold. If* $f \in C_c^1$, *then for all* $t_0 > 0$, $P^{t_0} f$ *is differentiable along the deterministic flow of the SUZZ process, that is, for all* $(x, \theta) \in E$ *there exists a function* $DP^{t_0} f : E \to \mathbb{R}$ *such that for all* $(x, \theta) \in E$,

$$\lim_{t \to 0} \frac{P^{t_0} f(\Phi_t(x, \theta), \theta) - P^{t_0} f(x, \theta)}{t} = DP^{t_0} f(x, \theta).$$

PROOF OF LEMMA E.2. Fix $t_0 > 0$ and let's write $g = P^{t_0} f$ for notational convenience. Due to Lemma E.1, f is in the domain of the strong generator of the SUZZ process, therefore, using standard results (see, e.g., [22]), g is also in the domain of the strong generator of the SUZZ process. From the proof of Proposition C.1, rewriting (59) we get

$$\mathbb{P}_{x,\theta}(S_0(t)) \frac{g(\Phi_t(x, \theta), \theta) - g(x, \theta)}{t}$$

$$= \frac{\mathbb{E}_{x,\theta}[g(X_t, \Theta_t)] - g(x, \theta)}{t} - \sum_{i=1}^{d} \frac{\mathbb{E}_{x,\theta}[(g(X_t, \Theta_t) - g(x, \theta))1_{S_i(t)}]}{t}.$$

The first term of the RHS is finite since g is in the domain of the strong generator of SUZZ and the second term can be seen to be finite using the same argument as in the proof of Proposition C.1. Since $\lim_{t \to 0} \mathbb{P}_{x,\theta}(S_0(t)) = 1$, we get the result. □

The following lemma is the stepping stone to prove Theorem 3.2.

LEMMA E.3. *Assume that the rates satisfy* (12), *Assumptions* 3.2, 3.3, 3.4 *and* 3.5 *hold and the* $(Z_t)_{t\geq 0}$ *is a SUZZ process with speed function* s. *Let* $(P^t)_{t\geq 0}$ *be the transition semi-group of a SUZZ process with speed function* s *and let* $\mathcal{A}$ *be the strong generator of the SUZZ process. If* $f \in C_c^1(E)$ *then for all* $t_0 > 0$, $P^{t_0} f \in \mathcal{D}(\mathcal{A})$ *and*

$$(75) \qquad \int_E \mathcal{A} P^{t_0} f(x, \theta) \mu(dx, d\theta) = 0.$$

If we could guarantee that for any $f \in C_c^1(E)$, for all $t_0 > 0$, $P^{t_0} f \in C_c^1(E)$, then Lemma E.3 would be easy to verify, using similar calculations to the proof of Proposition 5 of [10]. In our setting, due to the fact that we allow explosive deterministic dynamics, we cannot guarantee that $P^{t_0} f \in C_c^1$. However, Lemma E.2 guarantees that the function $P^{t_0} f$ must have a derivative along lines parallel to the vectors $\{-1, +1\}^d$. Therefore, the fundamental theorem of calculus and an integration by parts can be used along such lines. When we will integrate $\mathcal{A} P^{t_0} f$ over the ball O_m, we may do the integration over many different lines parallel to some vector $\{-1, +1\}^d$ and apply the integration by parts technique in each of these lines to get the result. This is the main idea of the following proof.

PROOF OF LEMMA E.3. We begin by noticing that, since $f \in \mathcal{D}(\mathcal{A})$, $P^{t_0} f \in \mathcal{D}(\mathcal{A})$ as well (see, e.g., Proposition 1.1.5 of [22]). If $E_m = O_m \times \{\pm 1\}^d$ then

$$\left| \int_{E_m} \mathcal{A} P^{t_0} f(x, \theta) \mu(dx, d\theta) - \int_E \mathcal{A} P^{t_0} f(x, \theta) \mu(dx, d\theta) \right| \xrightarrow{m \to \infty} 0,$$

since $E_m \nearrow E$ and $\mathcal{A} P^{t_0} f = P^{t_0} \mathcal{A} f = P^{t_0} \mathcal{L} f$ is bounded, due to that $\mathcal{L} f$ is bounded. Therefore, it suffices to prove that

$$(76) \qquad \lim_{m \to \infty} \int_{E_m} \mathcal{A} P^{t_0} f(x, \theta) \mu(dx, d\theta) = 0.$$

From now on, let us write $g = P^{t_0} f$ for notational convenience. From Lemma E.2 we get that g has a derivative along the deterministic dynamics of the SUZZ process. This means that there exists a function $Dg : E \to \mathbb{R}$ such that if X_t satisfies ODE (8) with starting point (x, θ) then for all $t \geq 0$

$$g(X_t, \theta) - g(x, \theta) = \int_0^t Dg(X_u, \theta) \, du.$$

Furthermore, using the same argument as in the proof of Proposition C.1 we get that for all $(x, \theta) \in E_m$,

$$\mathcal{A} g(x, \theta) = Dg(x, \theta) + \sum_{i=1}^d \lambda_i(x, \theta) \big(g(x, F_i(\theta)) - g(x, \theta) \big).$$

Our goal is to use an integration by parts technique to control the first part of the sum of the generator. We fix a $\theta \in \{-1, +1\}^d$. We use a linear, invertible transformation A on $\mathbb{R}^d$ such that $A\theta = \sqrt{d} e_1$, $A O_m = O_m$ and $|det A| = 1$ and let $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^d$. We use the transformation $y = (y_1, \ldots, y_d) = Ax$. Also, given $y_2, y_3, \ldots, y_d$ with $y_2^2 + \cdots + y_d^2 < m^2$ we write

$$y_1^* = \theta_1 \sqrt{m^2 - y_2^2 - \cdots - y_d^2},$$

and we omit the dependence on $y_2, \ldots, y_d$ and θ_1 for ease of notation. We also write $x_0 = A^{-1}(-y_1^*, y_2, \ldots, y_d)'$.

We further consider the solution X_t to the ODE (8) starting from (x_0, θ) and we write $Y_t = (Y_t^1, \ldots, Y_t^d) = A X_t$ so that Y_t starts from $(-y_1^*, y_2, \ldots, y_d)$ and solves the ODE $dY_t/dt = \sqrt{d}\, s(A^{-1} Y_t) e_1$. Also write t^* such that $Y_{t^*}^1 = y_1^*$. Then we can write,

$$\int_{O_m} Dg(x, \theta) \exp\{-U(x)\} \, dx$$

$$= \int_{O_m} Dg(A^{-1}y, \theta) \exp\{-U(A^{-1}y)\} \, dy$$

$$= \int_{-m}^m \int_{-\sqrt{m^2 - y_2^2}}^{\sqrt{m^2 - y_2^2}} \cdots \int_{-y_1^*}^{y_1^*} Dg(A^{-1}y, \theta) \exp\{-U(A^{-1}y)\} \, dy_1 \, dy_d \cdots dy_2$$

$$= \int_{-m}^m \int_{-\sqrt{m^2 - y_2^2}}^{\sqrt{m^2 - y_2^2}} \cdots \int_0^{t^*} Dg(A^{-1}Y_t, \theta) \exp\{-U(A^{-1}Y_t)\} s(A^{-1}Y_t) \sqrt{d} \, dt \, dy_d \cdots dy_2.$$

Having fixed $y_2, \ldots, y_d$, write $z_1 = (-y_1^*, y_2, \ldots, y_d)$, $z_2 = (y_1^*, y_2, \ldots, y_d) \in \partial O_m$ and $y = (y_1, \ldots, y_d)$ and using integration by parts we get

$$\int_0^{t^*} Dg(A^{-1}Y_t, \theta) \exp\{-U(A^{-1}Y_t)\} s(A^{-1}Y_t) \sqrt{d} \, dt$$

$$= g(A^{-1}z_2, \theta) \exp\{-U(A^{-1}z_2)\} s(A^{-1}z_2) \sqrt{d} - g(A^{-1}z_1, \theta)$$

$$\times \exp\{-U(A^{-1}z_1)\} s(A^{-1}z_1) \sqrt{d}$$

$$- \int_0^{t^*} g(A^{-1}Y_t, \theta) \frac{d}{dt} [\exp\{-U(A^{-1}Y_t)\} s(A^{-1}Y_t)] \sqrt{d} \, dt$$

$$= g(A^{-1}z_2, \theta) \exp\{-U(A^{-1}z_2)\} s(A^{-1}z_2) \sqrt{d} - g(A^{-1}z_1, \theta)$$

$$\times \exp\{-U(A^{-1}z_1)\} s(A^{-1}z_1) \sqrt{d}$$

$$- \int_0^{t^*} g(A^{-1}Y_t, \theta) \exp\{-U(A^{-1}Y_t)\}$$

$$\times \sum_{i=1}^d \{-\partial_i U(A^{-1}Y_t) \theta_i s(A^{-1}Y_t) + \partial_i s(A^{-1}Y_t) \theta_i\} s(A^{-1}Y_t) \sqrt{d} \, dt$$

$$= g(A^{-1}z_2, \theta) \exp\{-U(A^{-1}z_2)\} s(A^{-1}z_2) \sqrt{d} - g(A^{-1}z_1, \theta)$$

$$\times \exp\{-U(A^{-1}z_1)\} s(A^{-1}z_1) \sqrt{d}$$

$$- \int_{-y_1^*}^{y_1^*} g(A^{-1}y, \theta) \exp\{-U(A^{-1}y)\}$$

$$\times \sum_{i=1}^d (-\partial_i U(A^{-1}y) \theta_i s(A^{-1}y) + \partial_i s(A^{-1}y) \theta_i) \, dy_1.$$

Overall

$$\sum_{\theta \in \{\pm 1\}^d} \int_{O_m} Dg(x, \theta) \exp\{-U(x)\} \, dx$$

$$= \sum_{\theta \in \{\pm 1\}^d} \int_{-m}^m \int_{-\sqrt{m^2 - y_2^2}}^{\sqrt{m^2 - y_2^2}} \cdots \int_{-\sqrt{m^2 - y_2^2 - \cdots - y_{d-1}^2}}^{\sqrt{m^2 - y_2^2 - \cdots - y_{d-1}^2}} g(A^{-1}z_2, \theta)$$

$$\times \exp\{-U(A^{-1}z_2)\}s(A^{-1}z_2)\sqrt{d}\,dy_d\cdots dy_3\,dy_2$$

(77)
$$-\sum_{\theta\in\{\pm1\}^d}\int_{-m}^{m}\int_{-\sqrt{m^2-y_2^2}}^{\sqrt{m^2-y_2^2}}\cdots\int_{-\sqrt{m^2-y_2^2-\cdots-y_{d-1}^2}}^{\sqrt{m^2-y_2^2-\cdots-y_{d-1}^2}}g(A^{-1}z_1,\theta)$$

$$\times\exp\{-U(A^{-1}z_1)\}s(A^{-1}z_1)\sqrt{d}\,dy_d\cdots dy_3\,dy_2$$

$$+\sum_{\theta\in\{\pm1\}^d}\int_{O_m}g(x,\theta)\exp\{-U(x)\}\sum_{i=1}^{d}(\theta_i\partial_i U(x)s(x)-\theta_i\partial_i s(x))\,dx.$$

On the other hand, rearranging the sum over θ (see also the proof of Proposition 5 of [10]), we get

(78)
$$\sum_{i=1}^{d}\sum_{\theta\in\{\pm1\}^d}\int_{O_m}\lambda_i(x,\theta)(g(x,F_i(\theta))-g(x,\theta))\exp\{-U(x)\}\,dx$$

$$=-\sum_{i=1}^{d}\sum_{\theta\in\{\pm1\}^d}\int_{O_m}g(x,\theta)\exp\{-U(x)\}(\lambda_i(x,\theta)-\lambda_i(x,F_i(\theta)))\,dx.$$

Recall that since the rates satisfy (12), we have

$$\lambda_i(x,\theta)-\lambda_i(x,F_i(\theta))=\theta_i\partial_i U(x)s(x)-\theta_i\partial_i s(x).$$

When we integrate $\int_{E_m}\mathcal{A}g\,d\mu$, we get the sum of the RHS of equations (77) and (78). On this sum, only the boundary parts remain and we have

$$\left|2^d H\int_{E_m}\mathcal{A}g(x,\theta)\mu(dx,d\theta)\right|$$

$$=\left|\sum_{\theta\in\{\pm1\}^d}\int_{O_m}\exp\{-U(x)\}\mathcal{A}g(x,\theta)\,dx\right|$$

$$\leq\sum_{\theta\in\{\pm1\}^d}\int_{-m}^{m}\int_{-\sqrt{m^2-y_2^2}}^{\sqrt{m^2-y_2^2}}\cdots\int_{-\sqrt{m^2-y_2^2-\cdots-y_{d-1}^2}}^{\sqrt{m^2-y_2^2-\cdots-y_{d-1}^2}}|g(A^{-1}z_2,\theta)$$

$$\times\exp\{-U(A^{-1}z_2)\}s(A^{-1}z_2)\sqrt{d}|\,dy_d\cdots dy_3\,dy_2$$

$$+\sum_{\theta\in\{\pm1\}^d}\int_{-m}^{m}\int_{-\sqrt{m^2-y_2^2}}^{\sqrt{m^2-y_2^2}}\cdots\int_{-\sqrt{m^2-y_2^2-\cdots-y_{d-1}^2}}^{\sqrt{m^2-y_2^2-\cdots-y_{d-1}^2}}|g(A^{-1}z_1,\theta)$$

$$\times\exp\{-U(A^{-1}z_1)\}s(A^{-1}z_1)\sqrt{d}|\,dy_d\cdots dy_3\,dy_2$$

$$\leq 2\sqrt{d}\|g\|_\infty\sup_{x\in\partial O_m}\{\exp\{-U(x)\}s(x)\}\int_{x\in\partial O_m}1\,dx$$

$$\leq C2\sqrt{d}\|g\|_\infty\sup_{x\in\partial O_m}\{\exp\{-U(x)\}s(x)\}m^{d-1}\xrightarrow{m\to\infty}0,$$

where the convergence holds due to Assumption 3.5. Here $\|g\|_\infty$ is well-defined since $g=P^{t_0}f$ is bounded since f is bounded. This completes the proof. $\square$

Now, we can conclude with the proof of invariance.

PROOF OF THEOREM 3.2. Let $(P^t)_{t \geq 0}$ be the transition semigroup of the process and $\mathcal{L}$ the operator defined in (58). Let $f \in C_c^1(E)$. From Lemma E.1, $\mathcal{L}f = \mathcal{A}f$ is the strong generator of f. Because of Dynkin's formula, for any $t > 0$,

$$P^t f(x, \theta) - f(x, \theta) = \int_0^t \mathcal{A} P^{t_0} f(x, \theta) \, dt_0.$$

Since s and λ are bounded on compact sets, for any $f \in C_c^\infty(E)$ we have that $\mathcal{L}f$ is bounded and after integrating both sides over μ and using Fubini's theorem, we get

$$\int_E P^t f(x, \theta) \mu(dx, d\theta) - \int_E f(x, \theta) \mu(dx, d\theta)$$

$$= \int_E \int_0^t \mathcal{A} P^{t_0} f(x, \theta) \, dt_0 \mu(dx, d\theta)$$

$$= \int_0^t \int_E \mathcal{A} P^{t_0} f(x, \theta) \mu(dx, d\theta) \, dt_0 = 0,$$

where the last equality follows from Lemma E.3. Therefore, for all $f \in C_c^\infty$

$$(79) \qquad \int P^t f(x, \theta) \mu(dx, d\theta) = \int f(x, \theta) \mu(dx, d\theta).$$

Since, as a simple application of Stone–Weierstrass, C_c^∞ is dense in C_c, (79) holds for all $f \in C_c$. This further extends to all bounded measurable functions f from Lusin's theorem [24]. This proves the result. $\square$

REMARK E.1. It can be seen from the proof of Lemma E.3 that Assumption 3.5 was only used in order to ensure that the boundary terms appearing in the integration by parts will decay as $\|x\|$ goes to infinity. If the deterministic dynamics are nonexplosive, the path of the process until time t_0 has a bounded length, therefore the function $g = P^{t_0} f$ has compact support and all the boundary terms disappear as $\|x\| \to \infty$. This means that when the deterministic dynamics are nonexplosive we do not need to make Assumption 3.5, as long as we still impose Assumption 3.1.

APPENDIX F: PROOF OF THEOREM 3.3 (EXPONENTIAL ERGODICITY)

We first recall some stability notions of a Markov process. For more details see [41, 43].

DEFINITION F.1. A Markov Process X with state space E is ϕ-irreducible if there exists a nontrivial measure ϕ such that for any point z and any set A of positive ϕ-measure, there exists a t with $P_z(X_t \in A) > 0$. We call ϕ an irreducibility measure.

A set C is petite if there exists a probability measure v, a $c > 0$ and a distribution a on $\mathbb{R}_+$ such that for any $z \in C$ and $A \in \mathcal{B}(E)$

$$(80) \qquad K_a(z, A) = \int_0^{+\infty} \mathbb{P}_z(X_t \in A) a(dt) \geq c v(A).$$

A set C is called small for a continuous or a discrete time Markov process/chain X_t, if there exists a probability measure v, a $c > 0$ and $t > 0$ such that for any $z \in C$ and $A \in \mathcal{B}(E)$

$$(81) \qquad \mathbb{P}_z(X_t \in A) \geq c v(A).$$

Furthermore, the process is called strongly aperiodic if there exists a petite set C and a $T > 0$ such that for any $z \in C$ and $t \geq T$, $\mathbb{P}_z(X_t \in C) > 0$.

The process is called a T-*process* if there exists a probability density a on $[0, +\infty)$ and a kernel $K : E \times \mathcal{B}(E) \to [0, +\infty)$ such that for all $A \in \mathcal{B}(E)$ the function $z \to K(z, A)$ is lower semicontinuous and for all $z \in E$, $K(z, E) > 0$ and

$$\int_0^{+\infty} \mathbb{P}_z(X_t \in \cdot) a(dt) \geq K(z, \cdot).$$

Additionally, a ϕ-irreducible process is *Harris recurrent* if for all $z \in E$ and A such that $\phi(A) > 0$ we have $\mathbb{P}_z(\int_0^{+\infty} 1_A(X_t) = +\infty) = 1$. If the invariant measure of the process is finite the process is called *positive Harris recurrent*.

Finally the process with invariant measure μ is called *ergodic* if for all $z \in E$

$$\|\mathbb{P}_z(X_t \in \cdot) - \mu(\cdot)\|_{\mathrm{TV}} \xrightarrow{t \to +\infty} 0.$$

For the proof of Theorem 3.3 we will use the following result (Theorem 6.1 in [42]).

THEOREM F.1 (Meyn–Tweedie 1993). *Assume that a Markov process $(Z_t)_{t \geq 0}$ on E is càdlàg, and all compact subsets of E are petite for some skeleton chain of Z. Assume further that if $\mathcal{L}^m$ is the extended generator of the process $(Z_t^m)_{t \geq 0}$, which is the process Z, stopped upon exiting O_m, then there exists a function $V : E \to [1, +\infty)$ and $c, b > 0$ and a compact set C such that for all $m \in \mathbb{N}$ and $z \in E$,*

$$(82) \qquad\qquad \mathcal{L}^m V(z) \leq -cV(z) + b1_C(z).$$

Then there exists a constant $M > 0$ and $\rho < 1$ such that for all $z \in E$

$$\|\mathbb{P}_z(Z_t \in \cdot) - \pi(\cdot)\|_{\mathrm{TV}} \leq MV(z)\rho^t,$$

which means that the process Z is exponentially ergodic.

From the proof of nonexplosivity we have that the function V introduced in (64) satisfies the drift condition (82) for some compact set C. Therefore, in order to prove Theorem 3.3 we need to prove that the SUZZ process has all the compact sets as petite for some skeleton chain. The focus of this section is to prove this property.

In order to do this we need to establish the reachability property, introduced in [11].

Given a speed function s, generating the family of deterministic flows $\{\Phi_t(x, \theta), t \geq 0\}$ for every $(x, \theta) \in E$, we define as control sequence an object $u = (t, \iota)$, where $t = (t_0, \ldots, t_m) \in (0, +\infty)^{m+1}$, $\iota = (i_1, \ldots, i_m) \in \{1, \ldots, d\}^m$ for some $m \in \mathbb{N}$. Starting from $(x, \theta) \in E$, a control sequence u gives rise to a SUZZ trajectory (X_t, Θ_t) as follows: Start from (x, θ) and follow direction θ for t_0 time, that is, set $X_t = \Phi_t(x, \theta)$, $\Theta_t = \theta$ for $t \in [0, t_0)$. Then, switch the i_1th component of θ to $F_{i_1}(\theta)$ and follow that direction for t_1 time, that is, set $X_t = \Phi_{t - t_0}(\Phi_{t_0}(x, \theta), F_{i_1}(\theta))$, $\Theta_t = F_{i_1}(\theta)$ for $t \in [t_0, t_0 + t_1)$. Continue similarly until time $t_0 + \cdots + t_m$. Write $\tau_k = \sum_{i=0}^{k-1} t_i$ for the time of the kth switch and denote the final position $(X_{\tau_{m+1}}, \Theta_{\tau_{m+1}})$ of the path by $\Psi_u(x, \theta)$.

DEFINITION F.2. Given a starting point $(x, \theta) \in E$, a control sequence $u = (t, \iota)$ is admissible if for all $k \in \{1, \ldots, m\}$ we have $\lambda_{i_k}(X_{\tau_k}, \Theta_{\tau_k}) > 0$.

Given two points $(x, \theta), (y, \eta) \in E$ we say that (y, η) is reachable from (x, θ) and write $(x, \theta) \to (y, \eta)$ if there exists a control sequence u admissible from (x, θ) such that $\Psi_u(x, \theta) = (y, \eta)$.

We write $(x, \theta) \looparrowright (y, \eta)$ if $(x, \theta) \to (y, \eta)$ and for an admissible sequence $u = (t, \iota)$ connecting the two points, we have that every index of $\{1, \ldots, d\}$ appears in ι.

We now focus on proving that for any two points $(x, \theta), (y, \eta) \in E$ we have $(x, \theta) \rightarrow$ (y, η). Note that we can assume without loss of generality that the SUZZ has minimal rates (i.e., $\gamma_i \equiv 0$ for all $i \in \{1, \ldots, d\}$) as higher rates make admissible paths more likely. In the case of original Zig-Zag there is the following result (Theorem 4 of [11]).

THEOREM F.2 (Bierkens–Roberts–Zitt 2019). *Assume that $U \in C^3$, $\lim_{\|x\| \to \infty} U(x) =$ $+\infty$ and there exists an x_0 local minimum for U such that $\mathrm{Hess}(U)(x_0)$ is strictly positive definite. Then the original Zig-Zag process targeting the potential U satisfies that for all $(x, \theta), (y, \eta) \in E$, $(x, \theta) \looparrowright (y, \eta)$.*

We can then generalise these results on the SUZZ using the following Lemma.

LEMMA F.1. *Suppose $s \in C^2$ and $s(x) > 0$ for $x \in \mathbb{R}^d$. Then, for any $(x, \theta), (y, \eta) \in E$, $(x, \theta) \rightarrow (y, \eta)$ in a SUZZ with speed s, targeting a potential U with minimal rates if and only if $(x, \theta) \rightarrow (y, \eta)$ in an original Zig-Zag, targeting a potential $U - \log s$ with minimal rates.*

PROOF OF LEMMA F.1. Consider an original Zig-Zag process targeting the potential $U - \log s$ with minimal rates. The rates of this process for the i coordinate are $\lambda_i^0(x', \theta') = [\theta_i' \partial_i (U(x') - \log s(x'))]^+$.

On the other hand, a SUZZ process with minimal rates targeting the potential U has rates for the i coordinate given by

$$\lambda_i(x', \theta') = [\theta_i'(s(x')\partial_i U(x') - \partial_i s(x'))]^+ = s(x')\lambda_i^0(x', \theta').$$

Therefore for any $(x', \theta') \in E$ and any $i \in \{1, \ldots, d\}$

(83) $$\lambda_i(x', \theta') > 0 \quad \Longleftrightarrow \quad \lambda_i^0(x', \theta') > 0.$$

Assume $(x, \theta) \rightarrow (y, \eta)$ with some admissible control sequence $u = (t, \iota) = (t_0, \ldots, t_m, i_1, \ldots, i_m)$ for the original Zig-Zag process, targeting the potential $U - \log s$ with minimal rates. Let (X_t, Θ_t) be the configuration of that original Zig-Zag path and let $\tau_k = \sum_{i=0}^{k-1} t_i$ be the times of the switches. We have $\lambda_{i_k}^0(X_{\tau_k}, \Theta_{\tau_k}) > 0$ for all k.

Note that since s is continuous and strictly positive, for any $(x', \theta') \in E$, $\lim_{t \to +\infty} \|\Phi_t(x', \theta')\| = +\infty$. Therefore, there exists an $s_0 > 0$ such that $\Phi_{s_0}(x, \theta) = x + t_0\theta = X_{\tau_1}$. Likewise, there exists an $s_1 > 0$ such that $\Phi_{s_1}(X_{\tau_1}, \Theta_{\tau_1}) = X_{\tau_2}$ and via induction we can construct for all $k \in \{0, \ldots, m\}$ an s_k such that $\Phi_{s_k}(X_{\tau_k}, \Theta_{\tau_k}) = X_{\tau_{k+1}}$. Then the control sequence $\tilde{u} = (s, \iota) = (s_0, \ldots, s_m, i_1, \ldots, i_m)$ is an admissible sequence starting from (x, θ) for the SUZZ targeting the potential U with minimal rates. Furthermore, the ending point of $\tilde{u}$ starting from (x, θ) is (y, η).

The other way around, that is, that an admissible path for the SUZZ process targeting U implies existence of an admissible path for the ZZ process targeting $U - \log s$ follows using similar arguments. □

Combining Theorem F.2 and Lemma F.1 we can prove the following.

PROPOSITION F.1. *Assume that $s \in C^2$ is a strictly positive function such that Assumption 3.1 holds, $U - \log s \in C^3$ and there exists an $x_0 \in \mathbb{R}^d$ such that $U - \log s$ has a local minimum in x_0 with $\mathrm{Hess}(U - \log s)(x_0)$ being strictly positive definite. Then, for every $(x, \theta), (y, \eta) \in E$, $(x, \theta) \looparrowright (y, \eta)$.*

As in [11] we can use this to prove that from any starting point and given any other point $z \in E$, the process has a positive probability of visiting a neighbourhood of z. The following lemma is the same as Lemma 8 in [11].

LEMMA F.2 (Continuous component). *Assume that $(x, \theta) \looparrowright (y, \eta)$ and the rates λ_i are continuous. Then there exist $U_x, V_x \subset \mathbb{R}^d$ open with $x \in U_x$, $y \in V_x$ and $\epsilon, t_0, c > 0$ such that for all $x' \in U_x$, $t \in [t_0, t_0 + \epsilon]$*

$$(84) \qquad \mathbb{P}_{x',\theta}(X_t \in \cdot, \Theta_t = \eta) \geq c \operatorname{Leb}(\cdot \cap V_x),$$

where Leb is the Lebesgue measure on $\mathbb{R}^d$.

The proof is very similar in spirit to the proof of Lemma 8 (Continuous component) of [11], therefore we do not present it here but we present it in Section 2 of the Supplementary Material [54]). The main idea is that since there is an admissible path from $(x, \theta) \looparrowright (y, \eta)$, the process has a positive probability to follow some path very close to the admissible path. Therefore there is a positive probability that starting from somewhere close to (x, θ) the process ends up somewhere close to (y, η).

Lemma F.2 allows us to prove stability properties for the process.

PROPOSITION F.2. *Let $(Z_t)_{t \geq 0} = (X_t, \Theta_t)_{t \geq 0}$ be a SUZZ process with strictly positive speed function $s \in C^2$. Assume that the process is nonexplosive and has μ as invariant. Assume that for all $(x, \theta), (y, \eta) \in E$, $(x, \theta) \looparrowright (y, \eta)$. Then the process is T, ϕ-irreducible and strongly aperiodic. If in addition the invariant measure μ is a probability measure then all compact sets are petite, they are also small for some skeleton chain and the process is positive Harris recurrent and ergodic.*

PROOF OF PROPOSITION F.2. The fact that the SUZZ process under the assumptions of Proposition F.2 is T-process, ϕ-irreducible and strongly aperiodic can be proven using the same proof as in Theorem 5 of [11].

Now, we prove that every compact set is petite. A standard argument as in [11] shows that for μ-almost all starting points (x, θ) we have $\mathbb{P}_{x,\theta}(\lim_{t \to +\infty} \|X_t\| = +\infty) = 0$. Indeed, for any compact set K, $1_{\{X_t \text{ eventually leaves } K\}} = \liminf_{t \to +\infty} 1_{X_t \notin K}$ and by Fatou's lemma

$$\mathbb{P}_\mu(X_t \text{ eventually leaves } K) \leq \liminf_{t \to +\infty} \mathbb{P}_\mu(X_t \notin K) = 1 - \mu(K).$$

By exhausting E with compact sets we get $\mathbb{P}_\mu(\lim_{t \to +\infty} \|X_t\| = +\infty) = 0$. More specifically, there exists $(x, \theta) \in E$ such that $\mathbb{P}_{x,\theta}(\lim_{t \to +\infty} \|X_t\| = +\infty) < 1$. From Theorem 4.1 in [43] all compact sets are petite if and only if the process is T and ϕ-irreducible. The result follows.

Furthermore, the process is positive Harris recurrent from an application of Theorem 4.4 of [42].

The proof of the fact that some skeleton of the process is irreducible is the same as in Theorem 5 of [11].

From Theorem 6.1 of [43] we get that the process is ergodic.

Finally, all compacts are small for some skeleton chain from Proposition 6.1 in [43]. □

PROOF OF THEOREM 3.3. From Theorems 3.1 and 3.2 we know that the process is nonexplosive and μ introduced in (3) is invariant for the SUZZ. By Proposition F.1 for all $(x, \theta), (y, \eta) \in E$ we have $(x, \theta) \looparrowright (y, \eta)$ and therefore, by Proposition F.2 the process is ϕ-irreducible, aperiodic and all compact sets are small for some skeleton chain. Also, from Lemma D.1, V as in (64) satisfies the drift condition (66). All the conditions of Theorem F.1 are satisfied and the result follows. □

APPENDIX G: PROOF OF THEOREM 3.5

PROOF OF THEOREM 3.5.   The proof will rely on the following simple observation. Assume that $\mathcal{L}$ is as in (58) the extended generator of the SUZZ process targeting μ with refresh rate $\gamma(x)$ and $\mathcal{L}_{ZZ}^{\tilde{U}}$ is the strong generator of the original Zig-Zag process, with refresh rate $\frac{\gamma(x)}{s(x)}$ targeting the measure $\nu(dx, d\theta) = \frac{1}{H} \exp\{-\tilde{U}(x)\}$, where $\tilde{U}(x) = U(x) - \log s(x)$, that is, for all $f \in C^1$,

$$\mathcal{L}_{ZZ}^{\tilde{U}} f(x, \theta)$$
$$= \sum_{i=1}^{d} \theta_i \partial_i f(x, \theta) + \left([\theta_i \partial_i (U(x) - \log s(x))]^+ + \frac{\gamma(x)}{s(x)} \right)(f(x, F_i(\theta)) - f(x, \theta)),$$

then

$$\mathcal{L} f(x, \theta) = s(x) \mathcal{L}_{ZZ}^{\tilde{U}} f(x, \theta).$$

Let us consider the Zig-Zag process, having generator $\mathcal{L}_{ZZ}^{\tilde{U}}$ and targeting the measure ν. This is a special case of a SUZZ process where the speed function is equal to 1 everywhere and where the potential function U we have used throughout the document is replaced by $\tilde{U}$. Due to Assumption 3.6, the potential $\tilde{U}$ along with the constant speed function, equal to 1, satisfy all the assumptions of Lemma D.1. Furthermore, since by Assumption 3.6, $\gamma(x) \leq \tilde{M} s(x)$, the refresh rate of the Zig-Zag process is bounded. Therefore the function V introduced in (64) satisfies that there exists a compact set C and $c, b > 0$ such that

$$\mathcal{L}_{ZZ}^{\tilde{U}} V(x, \theta) \leq -c V(x, \theta) + b 1_{x \in C}.$$

Therefore,

$$(85) \qquad \mathcal{L} V(x, \theta) \leq -s(x) c V(x, \theta) + s(x) b 1_{x \in C} \leq -c' V(x, \theta) + b' 1_{x \in C}$$

since s is assumed to be bounded away from 0 and bounded on compact sets. Using Theorem 2.1 in [42], in the same way as in the conclusion of the proof of Theorem 3.1, we get that the process is nonexplosive.

For the other three bullet points, given that we have found a function V satisfying the drift condition (85), the proofs of Theorems 3.2 and 3.3 carry over here and we get the CLT result as a consequence of Theorem 2 of [14].   □

APPENDIX H: PROOF OF PROPOSITION 3.1

PROOF OF PROPOSITION 3.1.   We will consider the case where

$$s(x) = 1 + \|x\|_2^2.$$

The case $s(x) = \sqrt{1 + \|x\|_2^2}$ follows using a similar argument. We begin by noting that we can write

$$(86) \qquad A_i(x) = s(x) \partial_i U(x) - \partial_i s(x) = c(x) x_i,$$

where if π as in (24) then

$$c(x) = a\big(1 + \|x\|_2^2\big)^{\frac{a}{2}} - 2$$

and if π as in (25) then

$$c(x) = (\nu + d)\frac{1 + \|x\|_2^2}{\nu + \|x\|_2^2} - 2.$$

Given that $a > 0$ or ν satisfies (26), we see that for both targets, there exists a $K > 0$ such that for all $\|x\|_2 \geq K$

(87) $$c(x) \geq c > \frac{27}{2}d^3.$$

We set $b = (\frac{3}{2})^{1/\pi}$ and we consider the function

(88) $$V(x, \theta) = \sum_{i=1}^{d}(1 + x_i^2)b^{\arctan(\theta_i x_i)}.$$

First of all, $V \in C^1$ therefore $V \in D(\mathcal{L})$. Furthermore, $\lim_{\|x\| \to \infty} V(x, \theta) = +\infty$ and $V(x, \theta) \geq db^{-\pi/2} > 0$. Let $\|x\|_2 \geq K$. We write

$$\mathcal{L}V(x, \theta) = \sum_{i=1}^{d}\underbrace{2\theta_i x_i b^{\arctan(\theta_i x_i)}(1 + \|x\|_2^2)}_{\text{term } a_i} + \underbrace{(1 + \|x\|_2^2)\log(b)b^{\arctan(\theta_i x_i)}}_{\text{term } b_i}$$

$$+ \sum_{i=1}^{d}\underbrace{c(x)[\theta_i x_i]^+(1 + x_i^2)b^{\arctan(\theta_i x_i)}(b^{-2\arctan(\theta_i x_i)} - 1)}_{\text{term } c_i}.$$

Our goal is to show that V satisfies the drift condition (66). Let $\eta > 0$ be small enough (to be determined later), and having fixed η, let $\epsilon > 0$ be small enough (to be determined later). For any $\delta > 0$, consider the set $E_\delta = E_\delta(x) = \{i \in \{1, \ldots, d\} : \frac{|x_i|}{\|x\|_2} \geq \frac{\delta}{\sqrt{d}}\}$. Set

(89) $$\delta = \left(\frac{(2 + \eta)d}{c(1 - b^{-\pi})}\right)^{1/2} = \left(3\frac{(2 + \eta)d}{c}\right)^{1/2},$$

where c as in (87). We first note that for any $i \in \{1, \ldots, d\}$, if $\theta_i x_i \leq 0$ then $c_i = 0$ and $a_i \leq 0$ so $a_i + c_i \leq 0$.

If $\theta_i x_i > 0$ and $i \in E_\delta$, then for $\epsilon > 0$ small enough and assuming that K is large enough (given ϵ), for any $\|x\|_2 \geq K$,

$$a_i + c_i \leq |x_i|2b^{\frac{\pi}{2}}(1 + \epsilon)\|x\|_2^2 - c(x)|x_i||x_i|^2 b^{\frac{\pi}{2}-\epsilon}(1 - b^{-\pi+2\epsilon})$$

$$\leq |x_i|b^{\frac{\pi}{2}}\left(2(1 + \epsilon)\|x\|_2^2 - c(x)b^{-\epsilon}(1 - b^{-\pi+2\epsilon})\frac{\|x\|_2^2}{d}\delta^2\right)$$

$$\leq |x_i|b^{\frac{\pi}{2}}\|x\|_2^2\left(2(1 + \epsilon) - cb^{-\epsilon}(1 - b^{-\pi+2\epsilon})\frac{\delta^2}{d}\right) < 0,$$

from the definition of δ. Here we have used that if $i \in E_\delta$, then $|x_i| \geq \|x\|_2\frac{\delta}{\sqrt{d}}$.

On the other hand, if $\theta_i x_i > 0$ and $i \notin E_d$ then $c_i \leq 0$ so

$$a_i + c_i \leq a_i \leq |x_i|2b^{\frac{\pi}{2}}(1 + \epsilon)\|x\|_2^2 \leq 2(1 + \epsilon)b^{\frac{\pi}{2}}\frac{\delta}{\sqrt{d}}\|x\|_2^3,$$

where we have used the fact that $|x_i| \leq \|x\|_2\frac{\delta}{\sqrt{d}}$ when $i \notin E_\delta$.

Now consider $j = \arg\max\{|x_i|, i = 1, \ldots, d\}$. Combining all the results above, we then get

(90) $$\sum_{i \neq j}a_i + c_i \leq \sum_{i \neq j, i \notin E_\delta, \theta_i x_i > 0}a_i + c_i \leq (d - 1)2(1 + \epsilon)b^{\frac{\pi}{2}}\frac{\delta}{\sqrt{d}}\|x\|_2^3.$$

Now let us consider the quantity $a_j + c_j$. Since $|x_j| = \max\{|x_i|, i = 1, \ldots, d\}$ we have $|x_j| \geq \frac{\|x\|_2}{\sqrt{\delta}}$, that is, $j \in E_1$. We distinguish between the following cases.

Case 1: $\theta_j x_j > 0$. In this case, from the previous calculations

$$a_j + c_j \leq |x_j|\|x\|_2^2 b^{\frac{\pi}{2}}\left(2(1+\epsilon) - cb^{-\epsilon}(1 - b^{-\pi+2\epsilon})\frac{1}{d}\right)$$

$$\leq -\|x\|_2^3 \frac{1}{\sqrt{d}} b^{\frac{\pi}{2}}\left(cb^{-\epsilon}(1 - b^{-\pi+2\epsilon})\frac{1}{d} - 2(1+\epsilon)\right).$$

Case 2: $\theta_j x_j \leq 0$. In this case, $c_j = 0$ so

$$a_j + c_j = -2|x_j|b^{-\arctan(|x_j|)}(1 + \|x\|_2^2) \leq -2\|x\|_2^3 \frac{1}{\sqrt{d}}b^{-\frac{\pi}{2}},$$

where we have used that $x_j \geq \frac{\|x\|_2}{\sqrt{d}}$. In any case

$$(91) \qquad a_j + c_j \leq -\|x\|_2^3 b^{\frac{\pi}{2}} \frac{1}{\sqrt{d}} \min\left\{cb^{-\epsilon}(1 - b^{-\pi+2\epsilon})\frac{1}{d} - 2(1+\epsilon), 2b^{-\pi}\right\}.$$

Overall, combining (90) and (91) we have

$$(92) \qquad \sum_{i=1}^{d} a_i + c_i \leq \|x\|_2^3 \frac{b^{\frac{\pi}{2}}}{\sqrt{d}}\left[(d-1)2(1+\epsilon)\delta \right.$$

$$\left. - \min\left\{cb^{-\epsilon}(1 - b^{-\pi+2\epsilon})\frac{1}{d} - 2(1+\epsilon), 2b^{-\pi}\right\}\right],$$

where δ as in (89) for a small value of η. Our goal is to show that the term in the square bracket on the RHS is negative for small values of ϵ and η. To do this, it suffices to show that

$$(93) \qquad (d-1)2\delta < c(1 - b^{-\pi})\frac{1}{d} - 2$$

and that

$$(94) \qquad (d-1)2\delta < 2b^{-\pi},$$

where δ as in (89) for η small enough. We will first establish (93). To do this, it suffices to prove that

$$(d-1)2\left(\frac{2d}{c(1 - b^{-\pi})}\right)^{1/2} < c(1 - b^{-\pi})\frac{1}{d} - 2$$

$$\iff \quad 2^{3/2}(d-1)d^{1/2}c^{-1/2}3^{1/2} + 2 < \frac{c}{3d}$$

since $b^{\pi} = 3/2$. To prove the last inequality, it suffices to prove that

$$2 < \frac{1}{2}\frac{c}{3d}$$

and that

$$2^{3/2}d^{3/2}c^{-1/2}3^{1/2} < \frac{1}{2}\frac{c}{3d}.$$

The first equation is equivalent to asking that $c > 12d$ which holds due to (87). The second equation is equivalent to asking that $c > 2^{5/3}3d^{5/3}$ which also holds due to (87). This verifies

(93). In order to prove (94) for small enough η, it suffices to establish that

$$d\left(\frac{2d}{c(1-b^{-\pi})}\right)^{1/2} < b^{-\pi} \quad \iff \quad 2^{1/2}d^{3/2}c^{-1/2}3^{1/2} < \frac{2}{3}$$

$$\iff \quad c > \frac{27}{2}d^3,$$

which holds from (87). This confirms that the term in the square bracket on the RHS of (92) is negative for small values of ϵ and η. Overall, from (92) we get that there exist $K, k > 0$ and such that for any $\|x\|_2 \geq K$,

$$(95) \qquad \sum_{i=1}^{d} a_i + c_i \leq -k\|x\|_2^3.$$

Furthermore, one can bound for any $\|x\|_2 \geq K$,

$$\sum_{i=1}^{d} b_i \leq 2d \log(b) b^{\pi/2} \|x\|_2^2,$$

therefore, by increasing K appropriately, for any $\|x\|_2 \geq K$,

$$\mathcal{L}V(x,\theta) = \sum_{i=1}^{d} a_i + b_i + c_i \leq -\frac{k}{2}\|x\|_2^3.$$

Finally, one can bound V for all $\|x\|_2 \geq K$ by

$$V(x,\theta) \leq b^{\pi/2}\|x\|_2^2 + b^{\pi/2}d,$$

so overall, by further appropriately increasing K, we get that for all $\|x\|_2 \geq K$,

$$(96) \qquad \mathcal{L}V(x,\theta) \leq -\frac{k}{4}\|x\|_2 V(x,\theta),$$

while on $\{x : \|x\|_2 < K\}$, $\mathcal{L}V$ is bounded. This proves that V satisfies the drift condition (66). The rest of the proof can be concluded using the same arguments as in Appendices D, E and F.

The case where $s(x) = (1 + \|x\|_2^2)^{1/2}$ can be treated in a similar way.

Finally, one can prove Remark 3.6 by considering the function

$$V(x,\theta) = \sum_{i=1}^{d}(1 + (Bx)_i^2) b^{\arctan(\theta_i(Bx)_i)},$$

where B as in the Remark 3.6, and prove that V satisfies the drift condition (66) using similar arguments. $\square$

APPENDIX I: PROOF OF RESULTS IN SECTION 3.5

PROOF OF PROPOSITION 3.2. Introduce the space transformation $f(x) = \int_0^x 1/s(u)\, du$ and let $M^+ = \lim_{x \to +\infty} f(x) \in (0, +\infty]$ and $-M^- = \lim_{x \to -\infty} f(x) \in [-\infty, 0)$. Consider the process $(Y_t, \Theta_t)_{t \geq 0}$, where $Y_t = f(X_t)$, defined on $(-M^-, M^+) \times \{-1, +1\}$. When the process starts from $(y, \theta) = (f(x), \theta)$, it follows deterministic dynamics

$$\frac{dY_t}{dt} = f'(X_t)\frac{dX_t}{dt} = \frac{1}{s(X_t)}\Theta_t s(X_t) = \Theta_t,$$

and

$$\frac{d\Theta_t}{dt} = 0.$$

Furthermore, the random time T when the sign of Θ_t changes, is the same time when the direction changes for the $(X_t, \Theta_t)_{t \geq 0}$ process. Therefore,

$$\mathbb{P}(T \geq t) = \exp\left\{-\int_0^t \lambda(X_u, \theta)\,du\right\} = \exp\left\{-\int_0^t \lambda(f^{-1}(Y_u), \theta)\,du\right\}.$$

This proves that $(Y_t, \Theta_t)_{t \geq 0}$ is a one-dimensional original Zig-Zag with unit speed function and intensity rates λ_Y given by

$$
\begin{aligned}
\lambda_Y(y, \theta) &= \lambda(f^{-1}(y), \theta) \\
&= [\theta(s(f^{-1}(y))U'(f^{-1}(y)) - s'(f^{-1}(y)))]^+ \\
&= [\theta((U(f^{-1}(y)) - \log s(f^{-1}(y)))')]^+ = [\theta \tilde{U}'(y)]^+,
\end{aligned}
$$

where $\tilde{U}(y) = U(f^{-1}(y)) - \log s(f^{-1}(y))$. The result follows from standard results for the original Zig-Zag process (see [6]). $\square$

PROOF OF THEOREM 3.6. From Proposition 3.2 we know that if $f(x) = \int_0^x 1/s(u)\,du$, then the process $(Y_t, \Theta_t)_{t \geq 0} = (f(X_t), \Theta_t)_{t \geq 0}$ is an original Zig-Zag. From a separation of variables technique (see Appendix B), the solution of the ODE

(97)
$$\begin{cases} \dfrac{dX_t}{dt} = \theta s(X_t), \\ X_0 = x, \end{cases}$$

satisfies

(98)
$$X_t = f^{-1}(f(x) + \theta t).$$

Let $\lim_{x \to +\infty} f(x) = M^+ \in (0, +\infty]$ and $\lim_{x \to -\infty} f(x) = -M^- \in [-\infty, 0)$. Then the ODE has finite explosion time when starting from some $(x, +1)$ if and only if $M^+ < \infty$, whereas it has finite explosion time when starting from some $(x, -1)$ if and only if $M^- < \infty$. This further means that if the deterministic dynamics starting from some (x, θ) explode in finite time, then the deterministic dynamics starting from any other (x', θ) explode in finite time too. We also note that the Zig-Zag process $(Y_t, \Theta_t)_{t \geq 0}$ is defined on the space $(-M^-, M^+) \times \{\pm 1\}$.

We will begin by proving that the process (X_t, Θ_t) is nonexplosive.

Let us first consider the case where the deterministic flow explodes when starting both from $(x, +1)$ and $(x, -1)$. In that case the Zig-Zag process (Y_t, Θ_t) is defined on the bounded interval $(-M^-, M^+)$. Due to Assumption 3.1 and due to Proposition D.1 we see that the process a.s. cannot start from some point (x', θ) and reach infinity without the occurrence of a direction switch. This means that if ξ and ζ are as in (9) and (11), then $\xi \leq \zeta$. Therefore, the only way for the process to explode is for $\xi < \infty$. Let us consider the event that $\xi < \infty$ in order to show that it has probability zero. Let $T_1 < T_2 < \cdots$ be the switching times and let $\lim_{n \to \infty} T_n = \xi < \infty$. The T_n's are also the switching times of the Zig-Zag (Y_t, Θ_t) process, which is defined on the bounded interval $(-M^-, M^+)$. Let $Y_{T_1}, Y_{T_2}, \ldots$ be the switching points of the Zig-Zag process. Since these points lie in a bounded interval, there exists a subsequence of these points that converges to some point $z \in [-M^-, M^+]$. Assume for contradiction that there exists a second accumulation point $z' \neq z$ of the elements of $\{Y_{T_1}, Y_{T_2}, \ldots\}$. Since the Zig-Zag process moves with unit speed, it takes at least $\frac{|z-z'|}{2}$ time to get from a

small neighbourhood of one accumulation point to a neighbourhood of the other. On the other hand, we have assumed that the switching times T_n converge to a finite time ξ, therefore the difference between successive switching times must converge to zero and there is not enough time for the process to travel from one accumulation neighbourhood to the other. Therefore, there exists a unique accumulation point z, meaning that $\lim_{n\to\infty} Y_{T_n} = z$.

We distinguish between two cases for z. Assume first that $z \in (-M^-, M^+)$. Since $\lambda \in C^0$, it is bounded in a neighbourhood of z. Using the same argument as in the proof of Lemma C.1 we see that the probability that there are infinitely many switches in finite time in a small neighbourhood of z is zero. On the other hand, let us consider the case where $z \in \{-M^-, M^+\}$ and let us assume without loss of generality that $z = M^+$. Combining Assumptions 3.1 and 3.2, we have $\liminf_{x\to+\infty} (U(x) - \log s(x))' \geq 0$. This means that for large values of x,

$$\lambda(x, -1) = [s(x)U'(x) - s'(x)]^- + \gamma(x)$$
$$= s(x)[(U(x) - \log s(x))']^- + \gamma(x) = \gamma(x) \leq \bar{\gamma}.$$

Therefore, there exists an $\epsilon > 0$ such that if $y \in (M^+ - \epsilon, M^+)$ then

$$\lambda_Y(y, -1) = \lambda(f^{-1}(y), -1) \leq \bar{\gamma}.$$

On the event that M^+ is the accumulation point of switching times there are two possibilities. Either the Y-process switches from -1 to $+1$ infinitely many times in finite time inside the interval $(M^+ - \epsilon, M^+)$ or it does not. Since the process has a rate $\lambda_Y(y, -1)$ that is bounded above in $(M^- - \epsilon, M^+)$ the first event has probability zero. The other event is that only finitely many switches from -1 to $+1$ occur inside $(M^+ - \epsilon, M^+)$ until time ξ. In that case, since M^+ is an accumulation point for the switching times of Y, there are infinitely many switches from $+1$ to -1 inside $(M^+ - \frac{\epsilon}{2}, M^+)$ in finite time. Each of these switches has to be followed by a switch from -1 to $+1$. Since only finitely many will occur in $(M^+ - \epsilon, M^+)$, this means that in finite time there will be infinitely many switches from $+1$ to -1 inside $(M^+ - \frac{\epsilon}{2}, M^+)$ that have a successor switch from -1 to $+1$ outside $(M^+ - \epsilon, M^+)$. This means that between any of these two successive switches the process has travelled from $M^+ - \frac{\epsilon}{2}$ to $M^+ - \epsilon$ without switching and since the process moves with unit speed this takes at least $\frac{\epsilon}{2}$ time. Therefore, there cannot be infinitely many switches of this form in finite time. Therefore, we reach a contradiction assuming the second possibility. Overall this proves that the process is not explosive.

The case where the flow starting from $(x, +1)$ explodes in finite time but the one starting from $(x, -1)$ does not (and vice versa) can be handled in a similar way. In the case where the flow does not explode either in direction $+1$ or -1, the SUZZ algorithm is by definition nonexplosive. This finishes the proof of nonexplosivity.

We then observe that given nonexplosivity, for the proof of Theorem 3.2, only Assumption 3.5 (which is equivalent to Assumption 3.1 in dimension one) is needed. We therefore conclude that the process has μ as in (3) as invariant.

Let us now assume that the flow starting from any (x, θ) explodes in finite time. Following the same argument as in the proof of Lemma 15 of [10] for the one-dimensional original Zig-Zag, we get that the entire state space $(-M^-, M^+) \times \{-1, +1\}$ is petite for the transformed Zig-Zag process. Therefore, the transformed Zig-Zag is uniformly ergodic and we get that there exists a constant $M > 0$ and $\rho < 1$ and a measure ν such that for all $f(x) \in (-M^-, M^+)$ and any $\theta \in \{\pm 1\}$,

$$\|\mathbb{P}_{f(x),\theta}((f(X_t), \Theta_t) \in \cdot) - \nu(\cdot)\|_{TV} \leq M\rho^t.$$

Since f is $1-1$ we get that for all $x \in \mathbb{R}$, $\theta \in \{\pm 1\}$,

$$\|\mathbb{P}_{x,\theta}(Z_t \in \cdot) - \nu(f^{-1}(\cdot), \cdot)\|_{TV} \leq M\rho^t.$$

Since μ is invariant for the process Z_t, we get $\nu(f^{-1}(\cdot), \cdot) = \mu(\cdot)$. This means that for any $(x, \theta) \in E$,

$$\left\| \mathbb{P}_{x,\theta}(Z_t \in \cdot) - \mu(\cdot) \right\|_{\mathrm{TV}} \leq M\rho^t,$$

which proves the uniform ergodicity result.

Let us now assume that both the flows starting from $(x, +1)$ and $(x, -1)$ do not explode so the Zig-Zag Y-process is defined on $\mathbb{R} \times \{-1, +1\}$. Combining Assumptions 3.1 and 3.2, as previously done in this proof we get that

$$\liminf_{x \to +\infty} \lambda(x, +1) = \liminf_{x \to +\infty} |s(x)U'(x) - s'(x)| + \gamma(x) \geq A + \gamma(x),$$

where A as in (14). At the same time

$$\limsup_{x \to +\infty} \lambda(x, -1) = \gamma(x).$$

Therefore

$$\liminf_{x \to +\infty} \lambda(x, +1) > \limsup_{x \to +\infty} \lambda(x, -1),$$

and since $\lambda_Y(y, \theta) = \lambda(f^{-1}(y), \theta)$ we get

$$\liminf_{y \to +\infty} \lambda_Y(y, +1) > \limsup_{y \to +\infty} \lambda_Y(y, -1).$$

Using the same argument when $x \to -\infty$ we get that

$$\liminf_{y \to -\infty} \lambda_Y(y, -1) > \limsup_{y \to -\infty} \lambda_Y(y, +1).$$

From Lemma 16 of [10] the last two inequalities imply that the Zig-Zag process Y is exponentially ergodic. Using the same argument as when we concluded that the SUZZ process with explosive dynamics is uniformly ergodic, we conclude now that the SUZZ process with nonexplosive dynamics is exponentially ergodic.

The case where the flow $(x, +1)$ explodes and the one from $(x, -1)$ does not (and vice versa) can be treated using a similar argument. $\square$

APPENDIX J: PROOF OF PROPOSITION 4.1

PROOF OF PROPOSITION 4.1. The existence of ϕ such that $\phi(x, \theta) \leq CV(x, \theta)$ and such that $\mathcal{L}\phi(x, \theta) = -g(x, \theta)$ is guaranteed by Assumption 4.1. Consider

$$M_t = \phi(Z_t) - \phi(Z_0) + \int_0^t g(Z_s)\, ds,$$

which is a local martingale by Dynkin's formula. Assume that Z starts from the invariant measure μ. Under $\mathbb{P}_\mu$, M_t is also a martingale since for any $t > 0$ and any $s \leq t$, we have

$$\mathbb{E}_\mu[|M_s|] \leq \mathbb{E}_\mu[|\phi(Z_s)|] + \mathbb{E}_\mu[|\phi(Z_0)|] + \int_0^s \mathbb{E}_\mu[|g(Z_s)|]\, ds$$

$$\leq 2\mathbb{E}_\mu[|\phi|] + t\mathbb{E}_\mu[|g|] < \infty.$$

Furthermore M_t has stationary increments. From Theorem 2.1 of [34] we have

$$\frac{M_t}{\sqrt{t}} \xrightarrow{t \to \infty} N(0, \mathbb{E}[M_1^2])$$

in distribution under $\mathbb{P}_\mu$. Also, under $\mathbb{P}_\mu$

$$\frac{\phi(Z_t) - \phi(Z_0)}{\sqrt{t}} \xrightarrow{t \to +\infty} 0,$$

since $\phi(Z_t)$ has the same law as $\phi(Z_0)$ and $\mathbb{E}_\mu[\phi[Z_t]] = \mathbb{E}_\mu[\phi[Z_0]] < \infty$. Therefore

$$\frac{1}{\sqrt{T}} \int_0^T g(Z_s) \, ds \xrightarrow{T \to \infty} N(0, \mathbb{E}_{Z_0 \sim \mu}[M_1^2])$$

under $\mathbb{P}_\mu$. It suffices to prove that $\mathbb{E}_{Z_0 \sim \mu}[M_1^2]$ admits the expression in (42). Let K_t be the number of switches before time t and let $T_1, T_2, \ldots$ be the times of the switches. We write

$$M_t = \phi(Z_t) - \phi(Z_0) - \int_0^t \mathcal{L}\phi(Z_s) \, ds$$

$$= \int_0^t \Theta_s s(X_s) \phi'(Z_s) \, ds + \sum_{i=1}^{K_t} \phi(Z(T_i)) - \phi(Z(T_i^-))$$

$$- \int_0^t \Theta_s s(X_s) \phi'(Z_s) + \lambda(Z_s)(\phi(X_s, -\Theta_s) - \phi(X_s, \Theta_s)) \, ds$$

$$= \sum_{i=1}^{K_t} \phi(X_{T_i}, \Theta_{T_i}) - \phi(X_{T_i}, -\Theta_{T_i}) + \int_0^t \lambda(Z_s)(\phi(X_s, \Theta_s) - \phi(X_s, -\Theta_s)) \, ds$$

and therefore (see [33], Theorem 23.6) the predictable quadratic variation of M is

$$\langle M \rangle_t = \int_0^t \lambda(Z_t)(\phi(X_s, \Theta_s) - \phi(X_s, -\Theta_s))^2 \, ds = 4 \int_0^t \lambda(Z_t)(\psi(X_s))^2 \, ds,$$

where $\psi(x) = \frac{1}{2}(\phi(x, +1) - \phi(x, -1))$. Therefore, from the definition of predictable quadratic variation, under $\mathbb{P}_\mu$,

$$(99) \qquad \gamma_g^2 = \mathbb{E}_{Z_0 \sim \mu}[M_1^2] = \mathbb{E}_{Z_0 \sim \mu}\langle M \rangle_1 = 4 \int_E \lambda(x, \theta)\psi^2(x) \, d\mu(x, \theta).$$

It remains to write ψ in terms of g. This can be done since for all $(x, \theta) \in E$, $\mathcal{L}\phi(x, \theta) = -g(x, \theta)$ and therefore

$$\theta s(x)\phi'(x, \theta) + \lambda(x, \theta)(\phi(x, -\theta) - \phi(x, \theta)) = -g(x, \theta).$$

Writing down the two equations for $\theta = \pm 1$ and adding them up we get

$$s(x)(\phi'(x, +1) - \phi'(x, -1)) - (\lambda(x, +1) - \lambda(x, -1))(\phi(x, +1) - \phi(x, -1))$$
$$= -(g(x, +1) + g(x, -1))$$

and therefore

$$s(x)\psi'(x) - (s(x)U'(x) - s'(x))\psi(x) = -\frac{g(x, +1) + g(x, -1)}{2}.$$

Solving this first order linear ODE we get

$$\psi(x) = \frac{1}{2\exp\{-U(x)\}s(x)} \int_x^{+\infty} (g(y, +1) + g(y, -1)) \exp\{-U(y)\} \, dy,$$

which when combined with (99) gives (42). $\quad \square$

APPENDIX K: VERIFICATION OF ASSUMPTIONS FOR ONE-DIMENSIONAL SIMULATED TARGETS

Here we will check that using the speed function (46) leads to exponentially or uniformly ergodic algorithms when targeting some of the one-dimensional distributions in Section 4.1, where we calculate the inverse algorithmic efficiency, and in Section 5, where we present numerical simulations. In all cases we used SUZZ algorithms with refresh rate $\gamma(x) = 0$, therefore, the upper bound constant $\bar{\gamma}$ appearing in Assumption 3.2 is $\bar{\gamma} = 0$.

- $d = 1$, $\pi(x) = \frac{1}{H}(1 + \frac{1}{3}x^2)^{-2}$, with speed $s(x) = (1 + x^2)^{\frac{1+k}{2}}$, $k = 0, 1, 2$. This Student(3) target is used in Sections 4.1 and 5. Assumption 3.1 is easily verified since

$$\lim_{|x| \to \infty} \pi(x)s(x) = 0, \quad \text{for } k = 0, 1, 2.$$

Furthermore, here $U(x) = -\log \pi(x) = 2\log(1 + \frac{1}{3}x^2) + \log H$ and $U'(x) = 4\frac{x}{3+x^2}$. Straightforward calculations show that for $|x| > 1$ large enough such that $\frac{1+x^2}{3+x^2} > 1 - \frac{1}{10}$,

$$|A(x)| = |s(x)U'(x) - s'(x)|$$

$$= |x|(1 + x^2)^{\frac{k-1}{2}}\left(4\frac{1+x^2}{3+x^2} - 1 - k\right)$$

$$\geq |x|(1 + x^2)^{\frac{k-1}{2}}\left(4 - 4\frac{1}{10} - 1 - k\right) \geq \frac{6}{10\sqrt{2}} > 0$$

since $k \in [0, 2]$ and $|x|(1 + x^2)^{-1/2} \geq \frac{1}{\sqrt{2}}$. Therefore Assumption 3.2 is satisfied for $k = 0, 1, 2$, with $A = \frac{6}{10\sqrt{2}}$.

Therefore, the assumptions of Theorem 3.6 are verified. Using similar calculations, we conclude that the assumptions are verified for any Student target with ν degrees and speed function of the form $s(x) = (1 + x^2)^{\frac{1+k}{2}}$ as long as $k < \nu$. The SUZZ(0) algorithm is then exponentially ergodic and the SUZZ(k) algorithm with $k > 0$ is uniformly ergodic, due to Theorem 3.6.

- $d = 1$, $\pi(x) = \frac{1}{H}\exp\{-(1 + x^2)^{1/4}\}$, with speed $s(x) = (1 + x^2)^{\frac{1+k}{2}}$, $k = 0, 1, 2, 3$. This subexponential(0.5) target is used in Section 4.1. Assumption 3.1 is easily verified since

$$\lim_{|x| \to \infty} \pi(x)s(x) = 0, \quad \text{for } k = 0, 1, 2, 3.$$

Furthermore, here $U(x) = -\log \pi(x) = (1 + x^2)^{1/4} + \log H$ and $U'(x) = \frac{1}{2}x(1 + x^2)^{-3/4}$. Therefore,

$$|A(x)| = |s(x)U'(x) - s'(x)| = |x|(1 + x^2)^{\frac{k-1}{2}}\left(\frac{1}{2}(1 + x^2)^{\frac{1}{4}} - (1 + k)\right) \xrightarrow{|x| \to \infty} +\infty,$$

which verifies Assumption 3.2. Therefore, the assumptions of Theorem 3.6 are verified. The SUZZ(0) algorithm is then exponentially ergodic and the SUZZ(k) algorithm with $k = 1, 2, 3$ is uniformly ergodic.

- The $d = 1$ Exponential and Normal targets appearing in Section 4.1 can be shown to verify the assumptions of Theorem 3.6 in similar way.

SUPPLEMENTARY MATERIAL

Supplement to "Speed up Zig-Zag" (DOI: 10.1214/23-AAP1930SUPP; .pdf). Algorithmic description of the SUZZ process and proof of Lemma F.2.

REFERENCES

[1] ANDRIEU, C., DOBSON, P. and WANG, A. Q. (2021). Subgeometric hypocoercivity for piecewise-deterministic Markov process Monte Carlo methods. *Electron. J. Probab.* **26** 78. MR4269208 https://doi.org/10.1214/21-EJP643

[2] BESAG, J. (1994). Comments on "Representations of knowledge in complex systems" by Ulf Grenander and Michael I. Miller. *J. Roy. Statist. Soc. Ser. B* **56** 549–603.

[3] BIERKENS, J. (2016). Non-reversible Metropolis–Hastings. *Stat. Comput.* **26** 1213–1228. MR3538633 https://doi.org/10.1007/s11222-015-9598-x

[4] BIERKENS, J., BOUCHARD-CÔTÉ, A., DOUCET, A., DUNCAN, A. B., FEARNHEAD, P., LIENART, T., ROBERTS, G. and VOLLMER, S. J. (2018). Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains. *Statist. Probab. Lett.* **136** 148–154. MR3806858 https://doi.org/10.1016/j.spl.2018.02.021

[5] BIERKENS, J. and DUNCAN, A. (2017). Limit theorems for the Zig-Zag process. *Adv. in Appl. Probab.* **49** 791–825. MR3694318 https://doi.org/10.1017/apr.2017.22

[6] BIERKENS, J., FEARNHEAD, P. and ROBERTS, G. (2019). The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* **47** 1288–1320. MR3911113 https://doi.org/10.1214/18-AOS1715

[7] BIERKENS, J., GRAZZI, S., KAMATANI, K. and ROBERTS, G. O. (2020). The boomerang sampler. In *Proceedings of the 37th International Conference on Machine Learning* (H. D. III and A. Singh, eds.). *Proceedings of Machine Learning Research* **119** 908–918. PMLR, Virtual.

[8] BIERKENS, J., GRAZZI, S., VAN DER MEULEN, F. and SCHAUER, M. (2021). A piecewise deterministic Monte Carlo method for diffusion bridges. *Stat. Comput.* **31** 37. MR4249043 https://doi.org/10.1007/s11222-021-10008-8

[9] BIERKENS, J., NYQUIST, P. and SCHLOTTKE, M. C. (2021). Large deviations for the empirical measure of the Zig-Zag process. *Ann. Appl. Probab.* **31** 2811–2843. MR4350974 https://doi.org/10.1214/21-aap1663

[10] BIERKENS, J. and ROBERTS, G. (2017). A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *Ann. Appl. Probab.* **27** 846–882. MR3655855 https://doi.org/10.1214/16-AAP1217

[11] BIERKENS, J., ROBERTS, G. O. and ZITT, P.-A. (2019). Ergodicity of the zigzag process. *Ann. Appl. Probab.* **29** 2266–2301. MR3983339 https://doi.org/10.1214/18-AAP1453

[12] BIERKENS, J. and VERDUYN LUNEL, S. M. (2022). Spectral analysis of the zigzag process. *Ann. Inst. Henri Poincaré Probab. Stat.* **58** 827–860. MR4421609 https://doi.org/10.1214/21-aihp1188

[13] BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2018). The Bouncy Particle Sampler: A non-reversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113** 855–867. MR3832232 https://doi.org/10.1080/01621459.2017.1294075

[14] CHAN, K. S. and GEYER, C. J. (1994). Discussion: Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1747–1758.

[15] CHEN, T.-L. and HWANG, C.-R. (2013). Accelerating reversible Markov chains. *Statist. Probab. Lett.* **83** 1956–1962. MR3079029 https://doi.org/10.1016/j.spl.2013.05.002

[16] CORBELLA, A., SPENCER, S. E. F. and ROBERTS, G. O. (2022). Automatic Zig-Zag sampling in practice. *Stat. Comput.* **32** 107. MR4507158 https://doi.org/10.1007/s11222-022-10142-x

[17] COWLES, M. K. and CARLIN, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Statist. Assoc.* **91** 883–904. MR1395755 https://doi.org/10.2307/2291683

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

[19] DIACONIS, P., HOLMES, S. and NEAL, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.* **10** 726–752. MR1789978 https://doi.org/10.1214/aoap/1019487508

[20] DUNCAN, A. B., LELIÈVRE, T. and PAVLIOTIS, G. A. (2016). Variance reduction using nonreversible Langevin samplers. *J. Stat. Phys.* **163** 457–491. MR3483241 https://doi.org/10.1007/s10955-016-1491-2

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

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

[23] 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. MR3843382 https://doi.org/10.1214/18-STS648

[24] FOLLAND, G. B. (2013). *Real Analysis: Modern Techniques and Their Applications. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts*. Wiley, New York.

[25] FONTBONA, J., GUÉRIN, H. and MALRIEU, F. (2012). Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process. *Adv. in Appl. Probab.* **44** 977–994. MR3052846

[26] FONTBONA, J., GUÉRIN, H. and MALRIEU, F. (2016). Long time behavior of telegraph processes under convex potentials. *Stochastic Process. Appl.* **126** 3077–3101. MR3542627 https://doi.org/10.1016/j.spa.2016.04.002

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

[28] GOLDSTEIN, S. (1951). On diffusion by discontinuous movements, and on the telegraph equation. *Quart. J. Mech. Appl. Math.* **4** 129–156. MR0047963 https://doi.org/10.1093/qjmam/4.2.129

[29] HWANG, C.-R., HWANG-MA, S.-Y. and SHEU, S. J. (1993). Accelerating Gaussian diffusions. *Ann. Appl. Probab.* **3** 897–913. MR1233633

[30] JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030 https://doi.org/10.1016/S0304-4149(99)00082-4

[31] JOHNSON, L. T. and GEYER, C. J. (2012). Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *Ann. Statist.* **40** 3050–3076. MR3097969 https://doi.org/10.1214/12-AOS1048

[32] KAC, M. (1974). A stochastic model related to the telegrapher's equation. *Rocky Mountain J. Math.* **4** 497–509. MR0510166 https://doi.org/10.1216/RMJ-1974-4-3-497

[33] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications (New York)*. Springer, New York. MR1876169 https://doi.org/10.1007/978-1-4757-4015-8

[34] KOMOROWSKI, T., LANDIM, C. and OLLA, S. (2012). *Fluctuations in Markov Processes. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **345**. Springer, Heidelberg. Time symmetry and martingale approximation. MR2952852 https://doi.org/10.1007/978-3-642-29880-6

[35] LELIÈVRE, T., NIER, F. and PAVLIOTIS, G. A. (2013). Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *J. Stat. Phys.* **152** 237–274. MR3082649 https://doi.org/10.1007/s10955-013-0769-x

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

[37] LEWIS, P. A. W. and SHEDLER, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Nav. Res. Logist. Q.* **26** 403–413. MR0546120 https://doi.org/10.1002/nav.3800260304

[38] LIVINGSTONE, S. (2021). Geometric ergodicity of the random walk Metropolis with position-dependent proposal covariance. *Mathematics* **9**.

[39] LIVINGSTONE, S., BETANCOURT, M., BYRNE, S. and GIROLAMI, M. (2019). On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernoulli* **25** 3109–3138. MR4003576 https://doi.org/10.3150/18-BEJ1083

[40] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

[41] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253 https://doi.org/10.1017/CBO9780511626630

[42] MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548. MR1234295 https://doi.org/10.2307/1427522

[43] MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes. II. Continuous-time processes and sampled chains. *Adv. in Appl. Probab.* **25** 487–517. MR1234294 https://doi.org/10.2307/1427521

[44] NEKLYUDOV, K., BONDESAN, R. and WELLING, M. (2021). Deterministic gibbs sampling via ordinary differential equations.

[45] RIEDLER, M. G. (2011). Spatio-temporal stochastic hybrid models of biological excitable membranes Ph.D. thesis Heriot Watt Univ. United Kingdom.

[46] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273 https://doi.org/10.2307/3318418

[47] SUN, Y., SCHMIDHUBER, J. and GOMEZ, F. (2010). Improving the asymptotic performance of Markov chain Monte-Carlo by inserting vortices. In *Advances in Neural Information Processing Systems* (J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel and A. Culotta, eds.) **23** 2235–2243. Curran Associates, Red Hook.

[48] SUTTON, M. and FEARNHEAD, P. (2021). Concave-Convex PDMP-based sampling.

[49] TURITSYN, K. S., CHERTKOV, M. and VUCELJA, M. (2011). Irreversible Monte Carlo algorithms for efficient sampling. *Phys. D*, *Nonlinear Phenom.* **240** 410–414.

[50] VAN DYK, D. A. and PARK, T. (2011). Partially Collapsed Gibbs Sampling and Path-Adaptive Metropolis–Hastings in High-Energy Astrophysics. In *Handbook of Markov Chain Monte Carlo* CRC Press, Boca Raton.

[51] VANETTI, P., BOUCHARD-CÔTÉ, A., DELIGIANNIDIS, G. and DOUCET, A. (2017). Piecewise-Deterministic Markov Chain Monte Carlo.

[52] VASDEKIS, G. (2021). On Zig-Zag extensions and related ergodicity properties Ph.D. thesis Univ. Warwick United Kingdom.

[53] VASDEKIS, G. and ROBERTS, G. O. (2022). A note on the polynomial ergodicity of the one-dimensional Zig-Zag process. *J. Appl. Probab.* **59** 895–903. MR4480086 https://doi.org/10.1017/jpr.2021.97

[54] VASDEKIS, G. and ROBERTS, G. O (2023). Supplement to "Speed up Zig-Zag." https://doi.org/10.1214/23-AAP1930SUPP