Ergodicity of the zigzag process

Joris Bierkens, Gareth Roberts, Pierre-André Zitt

The Annals of Applied Probability
2019, Vol. 29, No. 4, 2266–2301
https://doi.org/10.1214/18-AAP1453

ERGODICITY OF THE ZIGZAG PROCESS[1]

BY JORIS BIERKENS, GARETH O. ROBERTS AND PIERRE-ANDRÉ ZITT

Delft University of Technology, University of Warwick and Université-Paris-Est-Marne-La-Vallée

The zigzag process is a piecewise deterministic Markov process which can be used in a MCMC framework to sample from a given target distribution. We prove the convergence of this process to its target under very weak assumptions, and establish a central limit theorem for empirical averages under stronger assumptions on the decay of the target measure. We use the classical "Meyn–Tweedie" approach (*Markov Chains and Stochastic Stability* (2009) Cambridge Univ. Press; *Adv. in Appl. Probab.* **25** (1993) 487–517). The main difficulty turns out to be the proof that the process can indeed reach all the points in the space, even if we consider the minimal switching rates.

1. Introduction.

1.1. *Motivation.* In recent years, there has been a growing interest in the use of Piecewise Deterministic Markov Process (PDMPs) within the field of Markov Chain Monte Carlo (MCMC). In MCMC, the objective is to simulate from a "target" probability distribution π by designing a Markov chain (or process) which is ergodic and has stationary distribution π. Although in principle MCMC, for example, in the form of the Metropolis–Hastings algorithm [27], can be used to sample from almost any probability distribution of interest, it can suffer from slow convergence as well as heavy computational cost per iteration.

It is for exactly these two reasons that PDMPs are so promising. First, PDMPs are nonreversible, and it is known that nonreversible Markov processes may offer faster convergence relative to reversible Markov processes (see, e.g., [5, 14, 16, 20, 22, 24, 37, 41]). Second, a remarkable feature of the simulation procedure of some PDMPs is that we can choose to use unbiased estimates of the "canonical" switching rate without affecting the stationarity of π. In settings in Bayesian statistics with large data sets (consisting of n observations, say), this offers significant benefits [7], reducing computational effort per iteration from $\mathcal{O}(n)$ to $\mathcal{O}(1)$. Similar

Received December 2017; revised October 2018.
[1]Supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement number 614492, the French National Research Agency under Grant ANR-12-JS01-0006 (PIECE) and the EPSRC Grants EP/D002060/1 (CRiSM) and EP/K014463/1 (iLike).
MSC2010 subject classifications. Primary 60F05; secondary 65C05.
Key words and phrases. Piecewise deterministic Markov process, irreducibility, ergodicity, exponential ergodicity, central limit theorem.

computational benefits can be obtained in systems in statistical physics consisting of many particles [32]. The use of PDMPs in sampling is a very active area of current research and (although it is not possible to give a complete list of references) we point the interested reader to [6, 10, 32–36, 39, 43, 44].

The zigzag process (ZZP) is an example of such a piecewise deterministic Markov process. As the name suggests, PDMPs follow deterministic dynamics, in between random times where they may jump or change to another deterministic dynamics (see [2, 25] for examples and additional references). For example, in the ZZP in $\mathbb{R}^d$, trajectories X_t have a piecewise constant velocity Θ_t belonging to the finite set $\{-1, 1\}^d$, with components of the direction changing at random times [7]. These random times are generated from inhomogeneous Poisson processes which have a space and direction dependent switching rate $\lambda_i(X_t, \Theta_t)$, for switching the ith component of Θ_t. Viewed as a process in the state space $E := \mathbb{R}^d \times \{-1, 1\}^d$, $(X_t, \Theta_t)_{t\geq 0}$ is a Markov process. The switching intensities λ_i can be chosen in such a way that the marginal density on $\mathbb{R}^d$ of the stationary probability distribution of (X_t, Θ_t) is equal to a prescribed density function $\overline{\pi}$. Other variants of PDMPs with similar properties exist, for example, the Bouncy Particle Sampler (BPS, [10]) which selects its direction from $\mathbb{R}^d$ or the unit sphere in $\mathbb{R}^d$.

In order for a Markov process to be useful in MCMC, it should have the prescribed stationary distribution and, furthermore, the process should be *ergodic*: the empirical time averages of a test function f along a trajectory should converge to the space average $\int f \, d\pi$, a property that usually follows from some kind of irreducibility, meaning roughly speaking that the process should be able to reach any point starting from any other point. The first requirement, stationarity, is relatively easy to satisfy. However, the second requirement is certainly nontrivial in the case of PDMPs. For example, it is known that without "refreshments" of the velocity, the BPS can be nonergodic, for instance, for any elliptically symmetric distribution such as a multivariate Gaussian [10]. In contrast, it is known that the ZZP is ergodic in certain cases in which the BPS is not ergodic [7], and computer experiments have suggested that in fact the ZZP is ergodic under only minimal assumptions. The main result of this paper is a proof of ergodicity for the ZZP under very mild and reasonable conditions, giving theoretical justification for its use in MCMC. This gives the ZZP a possible advantage over the BPS: the practitioner can be confident of the validity of the ZZP as MCMC algorithm and does not need to worry about tuning a refreshment parameter, which may slow down convergence to equilibrium if chosen suboptimally. However other aspects are also influential in determining speed of convergence and computational efficiency, and the relative merits of the ZZP versus the BPS is an area of challenging current and future research; see [1, 8, 13] for results in this direction.

Once ergodicity is established, one may look for estimates of rates of convergence to the invariant measure, in various senses. One of the possible approaches to establish such results is to find a Lyapunov function. For nonreversible processes with small noise, it is often very difficult to guess the form of a suitable Lyapunov function, and quite technical to prove that it indeed works; see, for example, [12, 17, 18]. In the zigzag case, it turns out that under a reasonable assumption on the decay of the target measure π at infinity, we are able to find a Lyapunov function in a quite simple form. Leveraging well-known results on long time convergence of processes, this proves in particular that the convergence towards the target measure π occurs exponentially fast, and we also get a central limit theorem for ergodic averages.

In [9], ergodicity of the one-dimensional zigzag process is established, which is significantly easier than the multidimensional case: for the one-dimensional process, it is always possible to switch the single direction component along a trajectory, so that irreducibility is relatively straightforward. The examples of Section 1.3 illustrate why proving ergodicity in the multidimensional case is fundamentally different. The conditions for exponential ergodicity in the one-dimensional case are weaker than those we impose for the multidimensional case, which is due to the fact that the one-dimensional Lyapunov function does not carry over to the multidimensional case; see Section 3.4 for a brief discussion. From a practical viewpoint, the slightly stronger conditions which we impose here are very reasonable.

1.2. *Preliminaries.* We briefly recall the construction of the zigzag process in $E = \mathbb{R}^d \times \{-1, 1\}^d$. For details, we refer to [7].

We equip E with its natural product topology, so that a function $(x, \theta) \mapsto f(x, \theta)$ is continuous if and only if $x \mapsto f(x, \theta)$ is continuous for every θ. Similarly, f is Lebesgue measurable if $x \mapsto f(x, \theta)$ is measurable for every θ.

For $i = 1, \dots, d$, introduce the mapping $F_i : \{-1, 1\}^d \to \{-1, 1\}^d$ which flips the ith component: For $j = 1, \dots, d$ and $\theta \in \{-1, 1\}^d$,

$$(F_i\theta)_j = \begin{cases} \theta_j & j \neq i, \\ -\theta_j & j = i. \end{cases}$$

Let $U : \mathbb{R}^d \to \mathbb{R}$ be a continuously differentiable *potential function*. We introduce continuous *switching intensities* (also referred to as *switching rates*) $\lambda_i : E \to [0, \infty)$, $i = 1, \dots, d$, and assume that they are linked with the potential through the relation

$$\lambda_i(x, \theta) - \lambda_i(x, F_i\theta) = \theta_i \partial_i U(x), \quad (x, \theta) \in E, i = 1, \dots, d. \tag{1}$$

An equivalent condition on the switching rates is the existence of a continuous function $\gamma : E \to [0, \infty)^d$ whose ith component does not depend on θ_i,

$$\gamma_i(x, F_i\theta) = \gamma_i(x, \theta), \quad (x, \theta) \in E, i = 1, \dots, d, \tag{2}$$

and which is related to the switching rate through

$$\lambda_i(x, \theta) = \big(\theta_i \partial_i U(x)\big)_+ + \gamma_i(x, \theta), \quad (x, \theta) \in E, i = 1, \dots, d. \tag{3}$$

Here, $(a)_+ := \max(0, a)$ is the positive part of $a \in \mathbb{R}$. We call γ the *excess switching intensity* and λ satisfying (3) with $\gamma \equiv 0$ the *canonical switching intensity*.

For $(x, \theta) \in E$, we construct a trajectory of (X, Θ) of the zigzag process with initial condition (x, θ) and switching intensities $\lambda(x, \theta)$ as follows. First, we construct a finite or infinite sequence of *skeleton points* (T^k, X^k, Θ^k) in $\mathbb{R}_+ \times E$ by the following iterative procedure:

- Let $(T^0, X^0, \Theta^0) := (0, x, \theta)$.
- For $k = 1, 2, \dots,$
  - Let $x^k(t) := X^{k-1} + \Theta^{k-1} t$, $t \geq 0$
  - For $i = 1, \dots, d$, let τ_i^k be distributed according to

$$\mathbb{P}(\tau_i^k \geq t) = \exp\left(-\int_0^t \lambda_i\big(x^k(s), \Theta^{k-1}\big)\, ds\right).$$

  - Let $i_0 := \operatorname{argmin}_{i \in \{1,\dots,d\}} \tau_i^k$ and let $T^k := T^{k-1} + \tau_{i_0}^k$. In principle, it is possible that $\tau_i^k = \infty$ for all i in which case the value of i_0 will turn out to be irrelevant and we set $T^k := \infty$.
  - If $T^k < \infty$, let $X^k := x^k(T^k)$ and $\Theta^k = F_{i_0} \Theta^{k-1}$ and repeat the steps. If $T^k = \infty$, terminate the procedure.

The piecewise deterministic trajectories (X_t, Θ_t) are now obtained as

$$(X_t, \Theta_t) := \big(X^k + \Theta^k(t - T^k), \Theta^k\big), \quad t \in [T^k, T^{k+1}), k = 0, 1, 2, \dots,$$

defining a process in E with the strong Markov property.

Informally, the process moves in straight lines, only changing velocities at the times T^k. In the case of canonical switching rates $\lambda_i(x, \theta) = (\theta_i \partial_i U(x))_+$, a change in the ith component θ_i of the velocity may only happen when in this direction, the process is going "uphill", that is, if $\theta_i \partial_i U(x) > 0$. Note, in particular, that if following the current velocity increases U, then $\langle \theta, \nabla U(x) \rangle > 0$ and at least one of the components has a positive rate of jump.

We further impose an integrability condition on the potential function

$$Z := \int_{\mathbb{R}^d} \exp\big(-U(x)\big)\, dx < \infty. \tag{4}$$

Under this condition, the zigzag process has a stationary probability distribution given by

$$\pi\big(A \times \{\theta\}\big) = \frac{1}{2^d Z} \int_A \exp\big(-U(x)\big)\, dx,$$
$$A \text{ Lebesgue measurable and } \theta \in \{-1, 1\}^d.$$

We will use the notation $\overline{\pi}(\cdot)$ for the marginal density function on $\mathbb{R}^d$, that is, $\overline{\pi}(x) = \exp(-U(x))/Z$, $x \in \mathbb{R}^d$.

1.3. *Why ergodicity of the ZZP is nontrivial.* First, consider a simple nonproblematic case, where at every point in space all switching rates λ_i are positive. This can be achieved by letting $\lambda_i(x,\theta) = \max(0, \theta_i \partial_i U(x)) + \gamma(x)$ where the excess switching rate $\gamma : \mathbb{R}^d \to (0,\infty)$ assumes only positive values. At an intuitive level, it is reasonable that such a process can reach any point in the state space, since by making a certain number of switches we can change direction to any direction in $\{-1,1\}^d$. These directions span $\mathbb{R}^d$. After reaching an arbitrary point in $\mathbb{R}^d$, we can switch to any desired final direction. Although we cannot change direction instantaneously but only over a time interval of positive length, the method above enables us to reach any point in $\mathbb{R}^d \times \{-1,1\}^d$ to arbitrary precision (and in fact, as will turn out, exactly).

However, having nonzero values for $\gamma(x,\theta)$ is not beneficial for efficiency: The zigzag process becomes more diffusive as γ_i increases which results in higher computational costs; see, for example, [6] for a detailed investigation of this phenomenon in the one-dimensional case. Therefore, we are mainly interested in the question of ergodicity for the case in which $\gamma_i(x,\theta) = 0$ for all i, x and θ, that is, for the canonical switching rates.

The expression for the canonical switching rates immediately tells us that one or more of the components of λ are zero in large parts of the state space. If the switching rate is zero on a set, it means that while the trajectory moves within this set, there is no freedom to switch the components of the direction vector. As a consequence, it is far from obvious how to construct trajectories between any two given points (x,θ) and (y,η) in the state space, which could be a realization of a canonical ZZP trajectory.

To illustrate the difficulties, let us discuss three examples highlighting what could go wrong with the zigzag process.

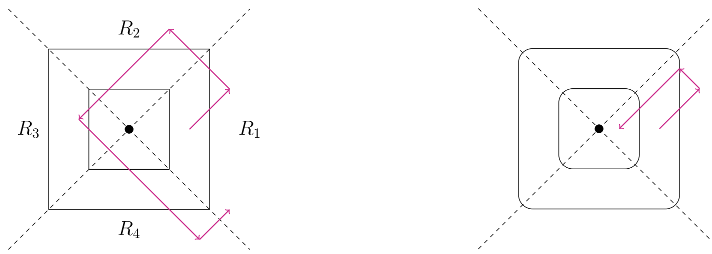
EXAMPLE 1 (A nonsmooth example). As an example of what can go wrong, consider the potential function $U : \mathbb{R}^2 \to \mathbb{R}$ given by $U(x) = \max(|x_1|, |x_2|)$. Having only a weak derivative, this example falls just outside the assumptions we will make in the formulation of the main results. Ignoring the diagonals $x_2 = x_1$ and $x_2 = -x_1$, divide the plane into four regions:

$$R_1 = \{(x_1,x_2) : x_1 > |x_2|\}, \qquad R_2 = \{(x_1,x_2) : x_2 > |x_1|\},$$
$$R_3 = \{(x_1,x_2) : x_1 < -|x_2|\}, \qquad R_4 = \{(x_1,x_2) : x_2 < -|x_1|\}.$$

The potential U is almost everywhere differentiable, with

$$\partial_1 U(x_1,x_2) = \begin{cases} 1 & \text{in } R_1, \\ -1 & \text{in } R_3, \\ 0 & \text{in } R_2 \cup R_4, \end{cases} \quad \text{and} \quad \partial_2 U(x_1,x_2) = \begin{cases} 1 & \text{in } R_2, \\ -1 & \text{in } R_4, \\ 0 & \text{in } R_1 \cup R_3, \end{cases}$$

and except for pathological initial values (along the diagonals), the switching rates are well defined (albeit discontinuous) and we can construct a zigzag process with these switching rates. Suppose we start a trajectory with initial condition



(a) Contour lines, the regions R_1, R_2, R_3 and R_4, and a typical trajectory for the potential function $U(x) = \max(|x_1|, |x_2|)$. From the displayed starting position it is impossible to reach a point in R_1 with direction $(-1, -1)$.

(b) Once we 'smoothen' the potential function slightly, it becomes possible to switch the second coordinate of the direction vector, making the process irreducible.

FIG. 1. *The canonical zigzag process for* $U(x) = \max(|x_1|, |x_2|)$ *and a smoothed version of* U.

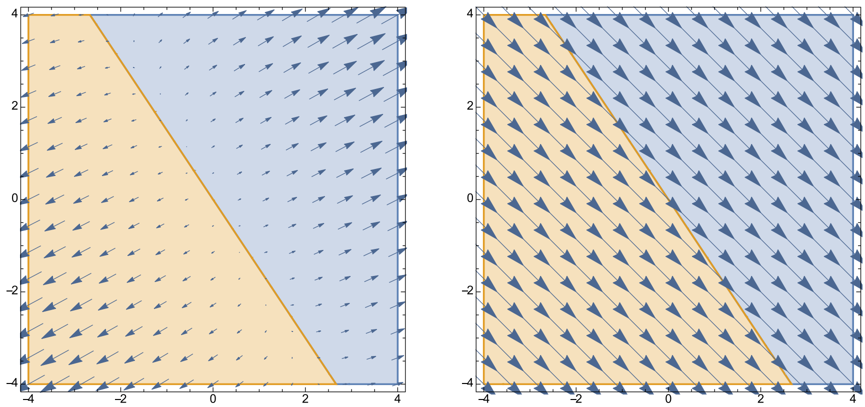
$(x_1, x_2) \in R_1$ and initial direction $\theta = (+1, +1)$. The trajectory will remain in R_1 at least until one of the components is switched. The only component which has a positive switching rate is the first component: $\lambda_1(x, \theta) = 1$ and $\lambda_2(x, \theta) = 0$ for $x \in R_1$ and $\theta = (+1, +1)$. Therefore, we will switch at some point to the direction $(-1, +1)$, after which we will eventually reach the region R_2. We can repeat this argument to find that, with full probability, we will subsequently enter the regions R_3, R_4 and R_1 with directions $(-1, -1)$, $(+1, -1)$ and $(+1, +1)$, respectively. In particular, from the given initial condition it is impossible to reach a point in R_1 with a direction θ for which $\theta_2 = -1$, and we conclude that the zigzag process is not irreducible. If we consider a slightly smoothed version of the potential function the associated zigzag process is irreducible on the combined position-momentum space $E = \mathbb{R}^2 \times \{-1, 1\}^2$. See Figure 1 for an illustration of this example.

EXAMPLE 2 (Gaussian distributions). In this example, we consider what may go wrong in the fundamental case of a Gaussian target distribution. Consider first the standard normal case, $U(x) = \frac{1}{2}\|x\|^2$, so that $\nabla U(x) = x$ and $\lambda_i(x, \theta) = \max(0, \theta_i x_i)$. As a result, starting from (x, θ),

$$\lambda_i(x + \theta t, \theta) = \big(\theta_i(x_i + \theta_i t)\big)_+ = (\theta_i x_i + t)_+.$$

We see that in this situation, as t increases, eventually the switching rate in any component becomes positive. This means that after travelling in a certain direction, we may switch any component of the direction vector. The same holds for Gaussian distributions with a diagonally dominant inverse covariance matrix. In our first attempts to prove irreducibility this provided us with a concrete way of building trajectories between any two points.

However, we should be careful since it is not always the case that, for large enough t, we can switch any component of the direction vector, even in ideal situations (e.g., with a strictly convex potential). For example, in a two-dimensional



(a) The gradient vector field ∇U (b) The constant vector field $(+1, -1)$

FIG. 2. *A non diagonally dominant Gaussian case. An example in the setting of Example 2 in which the switching rate in the second coordinate drops to zero after being non-zero initially. Consider a two-dimensional Gaussian target distribution, with potential function* $U(x) = \frac{1}{2}x^\top V x$, *where* $V = (6\ \ 3;\ \ 3\ \ 2)$ (*which is positive definite, but not diagonally dominant*). *In Figure* (a) *the gradient field of* U *is drawn. The region where* $\partial_2 U > 0$ *is shaded blue. In Figure* (b) *the constant vector field* $\theta = (+1, -1)$ *is superimposed over the division between regions. If a trajectory follows this vectorfield, coming from the yellow region where* $\partial_2 U < 0$, *at some point it enters the blue region. At this point the switching rate for* θ_2, *i.e.* $\lambda_2(x, \theta) = \max(0, -\partial_2 U(x))$, *drops to zero. The conclusion is that switching rates of individual components are not necessarily strictly increasing along the piecewise linear segments of the trajectory, contrary to what intuition may suggest.*

Gaussian case, it may happen that the switching rate in a certain component may drop from being positive to zero as time increases. See Figure 2 for an illustration of this phenomenon.

EXAMPLE 3 (Ridge). Consider a two-dimensional case in which $U(x_1, x_2) = |x_1 - x_2|^{2\alpha}(1 + |x_1 + x_2|^2)$, where $\frac{1}{2} < \alpha < 1$. Note that $U(x_1, x_2)$ is continuously differentiable and it can be seen that $\int_{\mathbb{R}} \int_{\mathbb{R}} \exp(-U(x_1, x_2))\, dx_1\, dx_2 < \infty$, so that U is (after normalization) the potential of a probability distribution on $\mathbb{R}^2$. However, a simple computation yields that the gradient ∇U vanishes along the diagonal $x_2 = x_1$, which is oriented with the directions $\pm(1, 1)$. As a consequence, starting from some initial condition (x_1, x_2) satisfying $x_2 = x_1$ in the direction $\pm(1, 1)$, it will be impossible to switch any component of the direction vector and inevitably we will drift off to infinity. The function $\exp(-U(x_1, x_2))$ corresponds to a narrow ridge, along which the derivative of U vanishes; see Figure 3. As we will see, it is essentially the fact that $U(x_1, x_2) \not\to \infty$ as $(x_1, x_2) \to \infty$ which results in this evanescent behaviour. The lack of a nondegenerate local minimum (our other fundamental assumption to prove irreducibility) is less problematic. This is because the shape of U can be modified smoothly around the origin to have a local nondegenerate minimum, without removing the possibility of drifting away to infinity.

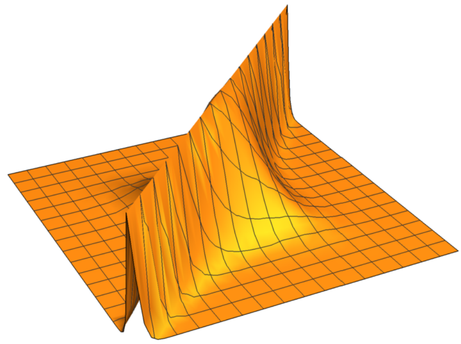


FIG. 3. *The "ridge" example. A continuously differentiable probability density function in two dimensions which has the property that along a narrow ridge the slope vanishes.*

1.4. *Main results.* We introduce three "growth conditions", that is, conditions on the tail behaviour of the potential function.

GROWTH CONDITION 1. $U \in \mathcal{C}^2$ and $\lim_{|x|\to\infty} U(x) = \infty$.

GROWTH CONDITION 2. $U \in \mathcal{C}^2$ and for some constants $c > d$, $c' \in \mathbb{R}$, $U(x) \geq c \ln(|x|) - c'$ for all $x \in \mathbb{R}^d$.

GROWTH CONDITION 3. $U \in \mathcal{C}^2$,

$$\lim_{|x|\to\infty} \frac{\max(1, \|\operatorname{Hess} U(x)\|)}{|\nabla U(x)|} = 0 \quad \text{and} \quad \lim_{|x|\to\infty} \frac{|\nabla U(x)|}{U(x)} = 0.$$

The following theorems are the main results of this paper.

THEOREM 1 (Ergodicity). *Suppose the potential function is $\mathcal{C}^3$, has a nondegenerate local minimum and satisfies Growth Condition 2. Then the zigzag process is* ergodic *in the sense that*

$$\lim_{t\to\infty} \|\mathbb{P}_{(x,\theta)}[(X_t, \Theta_t) \in \cdot] - \pi\|_{\mathrm{TV}} = 0 \quad \textit{for all } (x, \theta) \in E.$$

The proof of Theorem 1 also establishes that the process is positively Harris recurrent (see Section 3 below for a precise definition), so that the law of large numbers holds (see, e.g., [26]): for all initial conditions $(x, \theta) \in E$ and $g \in L^1(\pi)$ for which $s \mapsto g(X_s, \Theta_s)$ is almost surely locally integrable,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T g(X_s, \Theta_s)\, ds = \pi(g) \quad \text{almost surely.}$$

THEOREM 2 (Exponential ergodicity). *Suppose $U \in \mathcal{C}^3$, U has a nondegenerate local minimum and Growth Condition 3 is satisfied. Suppose the excess switching rates $(\gamma_i)_{i=1}^d$ are bounded. Then the zigzag process is* exponentially ergodic, *that is, there exists a function $M : E \to \mathbb{R}_+$ and a constant $c > 0$ such that*

$$\left\| \mathbb{P}_{(x,\theta)}\left[(X_t, \Theta_t) \in \cdot\right] - \pi \right\|_{\mathrm{TV}} \le M(x,\theta) e^{-ct} \quad \textit{for all } (x,\theta) \in E \textit{ and } t \ge 0.$$

In particular, the Theorem 2 allows for the case of canonical switching rates, that is, $\gamma \equiv 0$.

REMARK 1. Many target distributions which do not satisfy GC 3 can be transformed by a suitable change of variables after which GC 3 will be satisfied and exponential ergodicity can be obtained for the transformed distribution. The trajectories of the transformed process can then be used to compute ergodic averages approximating the intended target distribution. We refer to [12, 21] for details of this approach.

REMARK 2. Theorem 2 establishes exponential ergodicity under reasonable conditions (i.e., comparable to other sufficient conditions for establishing exponential ergodicity of other processes [12, 38, 40]) on the tails of the target distribution. For example, for potential functions of the form $U(x) = (1 + \|x\|^2)^{\alpha/2}$, Theorem 2 establishes exponential ergodicity for any $\alpha > 1$. For heavier tails, it is not yet clear what would be a suitable Lyapunov function and this remains a topic of current research.

REMARK 3. Although GC 3 does not seem to imply GC 2, it does imply nonevanescence through a Lyapunov argument [30], Theorem 3.1.

Under essentially the same conditions, we can also establish a functional central limit theorem. In the following theorem, we write $D[0,1]$ for the Skorohod space of cadlag functions on $[0,1]$.

THEOREM 3 (Functional central limit theorem). *Suppose that $U \in \mathcal{C}^3$, U has a nondegenerate local minimum, Growth Condition 3 is satisfied, and U satisfies the integrability condition $\int_{\mathbb{R}^d} \exp(-\eta U(x))\, dx < \infty$ for some $0 < \eta < 1$. Suppose the excess switching rates $(\gamma_i)_{i=1}^d$ are bounded.*

Let $g : E \to \mathbb{R}$ satisfy $|g(\cdot)| \le k \exp(\beta U(\cdot))$ on E for some $k > 0$ and $0 \le \beta < (1-\eta)/2$.

Define $Z_n(t) := \frac{1}{\sqrt{n}} \int_0^{nt} (g(X_s, \Theta_s) - \pi(g))\, ds$, $t \ge 0$.

There exists a $0 \le \sigma_g < \infty$ such that for any starting distribution, Z_n converges in distribution in $D[0,1]$ to $\sigma_g B$, where B is a standard Brownian motion.

In particular, under the conditions of Theorem 3 the central limit theorem of ergodic averages holds:

$$\frac{1}{\sqrt{T}} \int_0^T \big(g(X_s, \Theta_s) - \pi(g)\big)\, ds \xrightarrow{d} N\big(0, \sigma_g^2\big) \quad \text{as } T \to \infty.$$

REMARK 4. If U grows faster than a positive power of $|x|$, then the integrability condition will be satisfied for η arbitrarily small, and the CLT applies as soon as $|g(\cdot)| \le k \exp(\beta U)$ for some $\beta < 1/2$. In other words, it applies for "almost" all functions $g \in L^2(\pi)$.

REMARK 5. A CLT for the one-dimensional zigzag process was obtained earlier in [6].

1.5. *Strategy.* The diagram in Figure 4 illustrates how the different growth conditions of Section 1.4 are related to key properties of the zigzag process, which are crucial to establish the main results. As seen in the diagram, it is possible to distinguish between "deterministic" results and "probabilistic" results.

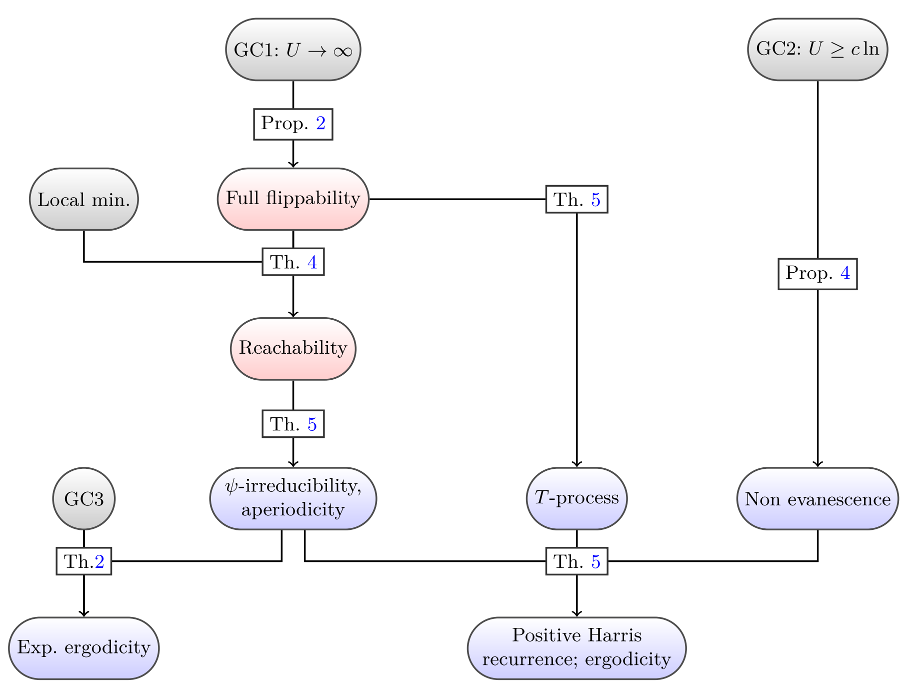


FIG. 4. *The key properties. Schematic overview of key properties of the zigzag process in relation to the Growth Conditions 1, 2 and 3. The grey nodes represent conditions on the potential U, the red nodes refer to deterministic "reachability" properties of trajectories, discussed in Section 2, and the blue nodes represent probabilistic properties discussed in Section 3.*

The "deterministic" results discussed in Section 2 concern the control theoretic aspects of zigzag trajectories. Here we are concerned with *reachability*: the existence of zigzag trajectories between any points in the state space such that, for a given potential function U, the trajectories are admissible: the switching intensities should be positive at the times at which the trajectory changes direction, even in the case of canonical switching rates. As a weaker notion, we are also interested in *full flippability*: can we, starting from any point in the state space, be certain that eventually all components of the direction vectors are switched at least once? This will all be made more precise in Section 2.

Next, in the "probabilistic" section, Section 3, the results of Section 2 are employed in order to establish several key properties (ψ-irreducibility, aperiodicity, the T-process property, nonevanescence and (positive) Harris recurrence) of the zigzag process as a Markov process, which finally result in proofs of the main theorems. The definitions of these probabilistic notions, which are standard in the Markov process literature [28, 31], are recalled in the introduction of Section 3. We conclude with proofs of the main results, located in Section 3.5.

2. Reachability.

2.1. *Admissible control sequences.* We define a *control sequence* to be a tuple $\mathbf{u} = (\mathbf{t}, \mathbf{i})$, where $\mathbf{t} = (t_0, \dots, t_m) \in (0, \infty)^{m+1}$ and $\mathbf{i} = (i_1, \dots, i_m) \in \{1, \dots, n\}^m$ for some $m \in \mathbb{N}$. Starting from (x, θ) at time 0, this sequence gives rise to a trajectory $(x(t), \theta(t))$ by: following θ for a time t_0, switching the i_1th component of θ, following the new velocity for a time t_1, etc.

More formally, writing $\tau_k = \sum_{i=0}^{k-1} t_i$ with the usual convention $\tau_0 = 0$, we define $(x(t), \theta(t))$ on $[0, \tau_{m+1}]$ by

$$\theta(t) = F_{(i_1,\dots,i_k)}\theta \quad \text{when } \tau_k \le t < \tau_{k+1} \text{ for } k = 0, \dots, m,$$
$$x(t) = x + \int_0^t \theta(s)\, ds.$$

Here, $F_{(i_1,\dots,i_k)} = F_{i_1} F_{i_2} \cdots F_{i_k}\theta$, that is, $F_I\theta$ flips all components of θ listed in the tuple $I = (i_1, \dots, i_k)$. This defines a piecewise constant trajectory $\theta(t)$ such that at at time τ_k, the i_kth component of $\theta(t)$ changes sign. The final position $(x(\tau_{m+1}), \theta(\tau_{m+1}))$ will be denoted by $\Phi_{\mathbf{u}}(x, \theta)$.

The following definitions apply for switching intensities $\lambda_i(x, \theta)$ satisfying (1).

DEFINITION 1 (Flippability). A component i of the velocity is *flippable* at a point $(x, \theta) \in E$ if the corresponding switching rate $\lambda_i(x, \theta)$ is strictly positive.

DEFINITION 2 (Admissible controls). Given a starting point (x, θ), a control sequence $(\mathbf{t}, \mathbf{i})$ is *admissible* if i_k is flippable at the point $(x(\tau_k), \theta(\tau_k))$, that is, if

$$\forall k \in \{1, \dots, m\}, \quad \lambda_{i_k}\big(x(\tau_k), \theta(\tau_k)\big) > 0.$$

DEFINITION 3 (Reachability). Given a starting point (x, θ) and an end point (x', θ'), we say that (x', θ') is *reachable* from (x, θ) and we write $(x, \theta) \rightsquigarrow (x', \theta')$ if there exists an admissible control sequence $\mathbf{u} = (\mathbf{t}, \mathbf{i})$ such that $\Phi_{\mathbf{u}}(x, \theta) = (x', \theta')$.

We write $(x, \theta) \looparrowright (x', \theta')$ if in addition, every index in $\{1, \dots, d\}$ appears at least once in $\mathbf{i}$, that is, all the components of the velocity are flipped at least once during the trajectory.

Our goal in this section is to prove that, under weak assumptions, any point is reachable from any other point. It is clear that if $(x, \theta) \rightsquigarrow (y, \eta)$ using the canonical, minimal switching rates $\lambda_i(x, \theta) = (\partial_i U(x)\theta_i)_+$, then the same is true for any choice of the switching rates. Consequently, we may and will assume in this section that the λ_i are the canonical switching rates.

REMARK 6. It follows immediately that if $(\mathbf{t}, \mathbf{i})$ is an admissible control sequence for some initial configuration, then by continuity of λ there exists an open environment U of $\mathbf{t} \in (0, \infty)^{m+1}$ such that $(\tilde{\mathbf{t}}, \mathbf{i})$ is admissible for the same initial configuration, for any $\tilde{\mathbf{t}} \in U$.

REMARK 7 (Reachability is transitive). Given two control sequences $\mathbf{u} = (s_0, \dots, s_p; i_1, \dots, i_p)$ and $\mathbf{v} = (t_0, \dots, t_q; j_1, \dots, j_q)$, we can concatenate them into

$$\mathbf{w} = (s_0, \dots, s_{p-1}, s_p + t_0, t_1, \dots, t_q; i_1, \dots, i_p, j_1, \dots, j_q).$$

If $\mathbf{u}$ is admissible starting from (x, θ) and $\mathbf{v}$ is admissible starting from $\Phi_{\mathbf{u}}(x, \theta)$, then $\mathbf{w}$ is admissible starting from (x, θ) and $\Phi_{\mathbf{w}}(x, \theta) = \Phi_{\mathbf{v}} \circ \Phi_{\mathbf{u}}(x, \theta)$.

REMARK 8 (Time reversal). If $(x, \theta) \rightsquigarrow (x', \theta')$, then $(x', -\theta') \rightsquigarrow (x, -\theta)$: indeed if $\lambda_i(x, \theta) > 0$, then

$$\lambda_i\big(x, -F_i(\theta)\big) = \big(\theta_i \partial_i U(x)\big)_+ = \lambda_i(x, \theta) > 0,$$

so if $(t_0, \dots, t_m; i_1, \dots, i_m)$ is an admissible control that sends (x, θ) to (x', θ'), then the reversed sequence $(t_m, \dots, t_0; i_m, \dots, i_1)$ is admissible and sends (x', θ') to (x, θ). (We thank the AE for pointing out that, without further conditions, this does not hold for noncanonical switching intensities.)

We will first establish reachability for the case where the potential U is quadratic, so that the target measure is Gaussian. We will use this in Section 2.3 to see that around a local minimum of the potential, we can reach any velocity. We will then show that, under Growth Condition 1, starting from any point, it is possible to switch all components of the velocity. All these results will be put together in Section 2.5 to prove reachability in the general case.

2.2. *Reachability for multivariate normal distributions.*

PROPOSITION 1. *Suppose that the target distribution is a nondegenerate Gaussian $U(x) = \langle x, Ax \rangle$, where A is a positive definite symmetric matrix. Then for any (x, θ), (x', θ'), $(x, \theta) \rightsquigarrow (x', \theta')$.*

Even for this simple case, the fact that the jump rates may be zero and that the process may be unable to jump for long stretches makes the proof quite involved. The main idea is to use the fact that by going in a straight line for a sufficiently long time, the process will always reach a region where it can switch some components of its velocity. Let us first define a useful notational shortcut.

DEFINITION 4 (Reachability for velocities). For any two velocities θ, θ', we say that θ' is *reachable* from θ, and we write $\theta \rightsquigarrow \theta'$ if for any x; there exists an x' such that $(x, \theta) \rightsquigarrow (x', \theta')$.

DEFINITION 5 (Asymptotic flippability). Let $\theta \in \{-1, 1\}^d$. If $\sum_j \theta_i A_{ij} \theta_j > 0$, we say that the ith component of θ is *asymptotically flippable*. The velocity θ itself is called asymptotically flippable if all its components are asymptotically flippable.

The above definition is explained by noting that in case of asymptotic flippability of the ith component, along any trajectory $x + \theta t$ the ith switching intensity will eventually become positive.

LEMMA 1. *If I is a sequence of asymptotically flippable components for θ, then $\theta \rightsquigarrow F_I(\theta)$. In particular, if η is asymptotically flippable, then for any θ, $\eta \rightsquigarrow \theta$.*

PROOF. Write $I = \{i_1, \dots, i_m\}$. Starting from x with velocity θ, after a large time t the components of $A(x + t\theta)$ will have the signs of the components of $A\theta$, so the ith component for $i \in I$ will all be flippable. The "pseudo"-control sequence $(t, 0, \dots, 0; i_1, \dots, i_m)$, would therefore bring (x, θ) to $(x', F_I\theta)$ for some x'. It is strictly speaking not a control sequence since its times between switches are zero. However, since the positivity of the jump rates is an open condition and the map $\mathbf{t} \mapsto \Phi_{(\mathbf{t},\mathbf{i})}(x, \theta)$ is continuous, this implies the existence of a $\mathbf{t}'$ with positive coefficients such that $(\mathbf{t}'; i_1, \dots, i_m)$ is admissible starting from (x, θ), proving that $\theta \rightsquigarrow F_I(\theta)$. $\square$

The usefulness of this definition is readily seen through the following result.

LEMMA 2 (Reachability for asymptotically flippable velocities). *If η is asymptotically flippable, then for any x and x', $(x, \eta) \rightsquigarrow (x', -\eta)$.*

Before proving this lemma, let us give a simple case where it is enough to conclude the argument.

COROLLARY 1. *If A is diagonally dominant, then every θ is asymptotically flippable, and $(x,\theta) \rightsquigarrow (x',\theta')$ for all pairs of states.*

PROOF. If A is diagonally dominant, then $\sum_j \theta_i A_{ij}\theta_j \geq A_{ii} - \sum_{j,j\neq i} |A_{ij}| > 0$ so all velocities are asymptotically flippable. Given (x,θ) and (x',θ'), we first use Lemma 1 to get the existence of x'' such that $(x,\theta) \rightsquigarrow (x'',-\theta')$. By Lemma 2, we can then reach (x',θ') from $(x'',-\theta')$, and we are done by transitivity. $\square$

PROOF OF LEMMA 2. Let η be an asymptotically flippable velocity, and x, x' be two arbitrary positions. To control the system from x to x', the idea is to go very far in the direction of η, to a region where all components of η are flippable, to flip them in a well- chosen order and with well-chosen time intervals between flips, so that when the last component is flipped, the system reaches x' after a long run in the direction $-\eta$.

To do this rigorously, define $d_i = (x'_i - x_i)/\eta_i$, and suppose first that the d_i are increasing: $d_1 < \cdots < d_n$. For $1 \leq i \leq n-1$, let $t_i = (d_{i+1} - d_i)/2$, and choose t_0 and t_n positive numbers such that $t_0 - t_n = \frac{d_1+d_n}{2}$.

Now let t be a large time to be chosen later, and consider the control

$$(\mathbf{t},\mathbf{i}) = (t+t_0, t_1, \dots, t_{n-1}, t_n + t; 1, 2, \dots, n).$$

Starting from (x,η), the ith component of the position will follow η_i for a time $t + t_0 + \cdots + t_{i-1}$, and $-\eta_i$ for the remaining time $t_i + \cdots + t_n + t$. Therefore, the ith component of the final position is

$$\begin{aligned}
&x_i + \eta_i\left(t + \sum_{j=0}^{i-1} t_j\right) - \eta_i\left(t + \sum_{j=i}^{n} t_j\right)\\
&\quad = x_i + \eta_i\left(t_0 - t_n + \frac{1}{2}\sum_{j=1}^{i-1}(d_{j+1} - d_j) - \frac{1}{2}\sum_{j=i}^{n-1}(d_{j+1} - d_j)\right)\\
&\quad = x_i + \frac{\eta_i}{2}(d_1 + d_n + d_i - d_1 - d_n + d_i)\\
&\quad = x_i + x'_i - x_i = x'_i.
\end{aligned}$$

If the d_i are not increasing but all distinct, we can reorder them by finding a permutation σ such that the $d_{\sigma(i)}$ increase, and perform the same argument using the control sequence $(t + t_0, t_1, \dots, t_{n-1}, t_n + T; \sigma(1), \dots, \sigma(n))$ where $t_i = (d_{\sigma(i+1)} - d_{\sigma(i)})$.

It remains to check that all the moves are admissible. By a computation similar to the one just above, the position $x^{(i)}$ just before the ith flip in the control sequence is given by

$$x_j^{(i)} = x_j + \eta_j \left(t + \sum_{k=0}^{i-1} (\mathbf{1}_{k \le j} - \mathbf{1}_{k>j}) t_k \right).$$

Once the t_k are fixed (by the given input of the starting and ending positions x and x'), one can always take t large enough so that $(Ax^{(i)})_i$ has the sign of η_i, which implies that the ith jump is indeed admissible.

Finally, if some of the d_i are equal, we may always introduce intermediary points y and y' such that the differences $(y_i - x_i)/\eta_i$ are distinct for all i, and likewise the differences $(y_i' - y_i)/(-\eta_i)$, and $(x_i' - y_i)/\eta_i$. Therefore, $(x,\eta) \rightsquigarrow (y,-\eta) \rightsquigarrow (y',\eta) \rightsquigarrow (x',-\eta)$, and we are done by transitivity. $\square$

We now tackle the general case when A is not diagonally dominant.

LEMMA 3 (All roads lead to an asymptotically flippable velocity). *For all θ, there exists an asymptotically flippable velocity η such that $\theta \rightsquigarrow \eta$.*

PROOF. To prove this result, it is useful to represent the matrix A as a Gramian matrix: as can be seen by an $LL^\top$ or a symmetric square root representation, there exists a family of vectors $(v_1, \dots, v_n)$ such that $A_{ij} = \langle v_i, v_j \rangle$. For a velocity θ, let $v(\theta) = \sum_i \theta_i v_i$. Using this representation, we have the equivalence:

$$\begin{aligned} i \text{ is asymptotically flippable for } \theta \quad &\iff \quad (A\theta)_i \theta_i > 0 \\ &\iff \quad \langle \theta_i v_i, v(\theta) \rangle > 0. \end{aligned}$$

Let θ be an arbitrary velocity, and suppose that θ is not asymptotically flippable. Denote by I the subset of asymptotically flippable indices:

$$i \in I \quad \iff \quad \langle \theta_i v_i, v(\theta) \rangle > 0.$$

Since $\sum_i \langle \theta_i v_i, v(\theta) \rangle = |v(\theta)|^2 > 0$ by positive definiteness of A, this set is nonempty; by hypothesis it is not equal to $\{1, \dots, n\}$. Let $F_I(\theta)$ be the velocity obtained by flipping all asymptotically flippable components. The key point is that this flip increases the norm of v:

$$|v(F_I(\theta))| > |v(\theta)|.$$

Indeed, let $v_+ = \sum_{i \in I} \theta_i v_i$ and $v_- = \sum_{i \notin I} \theta_i v_i$. Since $v(\theta) = v_+ + v_-$ and $v(F_I \theta) = v_- - v_+$,

$$|v(F_I \theta)|^2 - |v(\theta)|^2 = -4 \langle v_-, v_+ \rangle.$$

Now $\langle v(\theta), v_-\rangle$ must be nonpositive by definition of v_- and the set I, but this is $|v_-|^2 + \langle v_-, v_+\rangle$. The scalar product $\langle v_-, v_+\rangle$ is therefore negative, and

$$\left|v(F_I\theta)\right| > \left|v(\theta)\right|.$$

Now starting from θ, apply the following "algorithm":

- if θ is asymptotically flippable, stop.
- if it is not, move to $F_I\theta$ where I is the set of asymptotically flippable indices.

The fact that θ is not asymptotically flippable implies that v_- cannot be zero (because $I \neq \{1, \dots, d\}$ and the v_i are linearly independent because A is positive definite), so the norm will increase. Since along the algorithm, $|v(\theta)|$ is strictly increasing, it must stop at one time; at this time it has (by definition) reached an asymptotically flippable velocity. □

Now we have all the ingredients to prove the full reachability in the Gaussian case.

PROOF OF PROPOSITION 1. Let (x,θ) and (x',θ') be two points. By Lemma 3, there exists an asymptotically flippable velocity η' and a point y' such that $(x', -\theta') \rightsquigarrow (y', \eta')$. By the time-reversal property of Remark 8, $(y', -\eta') \rightsquigarrow (x', \theta')$. Now by Lemma 3 again, we get the existence of an asymptotically flippable velocity η and a point y such that $(x, \theta) \rightsquigarrow (y, \eta)$. Lemma 1 gives us a point z such that $(y, \eta) \rightsquigarrow (z, \eta')$, and Lemma 2 tells us that $(z, \eta') \rightsquigarrow (y', -\eta')$, which finishes the construction of an admissible trajectory. □

2.3. *Reachability around a local minimum.* As before $U : \mathbb{R}^d \to \mathbb{R}$ is the potential function of a probability density function $\overline{\pi}$, that is, $\overline{\pi}(x) \propto \exp(-U(x))$. We suppose that U has at least one nondegenerate local minimum, which we assume without loss of generality to be located in $x = 0$, that is, $\nabla U(0) = 0$ and $V := H_U(0)$ is positive definite. We will use the fact that all points in $\mathbb{R}^d$ are reachable through zigzag trajectories for the Gaussian density $\pi^V \propto \exp(-\frac{1}{2}x^T V x)$, to conclude that the same holds in a neighbourhood of 0 for the potential U.

LEMMA 4. *Suppose $U \in \mathcal{C}^3(\mathbb{R}^d)$, $\nabla U(0) = 0$ and $H_U(0)$ is positive definite. There exists a radius $\gamma > 0$ such that $(x, \theta) \looparrowright (y, \eta)$ for every (x, θ) and (y, η) satisfying $|x| < \gamma$, $|y| < \gamma$.*

PROOF. Let the switching rates for the Gaussian density π^V be denoted by (λ_i^V). For a given control sequence $(\mathbf{t}, \mathbf{i}) = (t_0, \dots, t_p; i_1, \dots, i_p)$ with associated switching points $(x(\tau_i), \theta(\tau_i))_{i=1}^p$ and final point $(x(\tau_{p+1}), \theta(\tau_{p+1}))$, define

$$\lambda_{\min}^V(\mathbf{t}, \mathbf{i}) = \min_{j=1,\dots,p} \lambda_{i_j}^V\big(x(\tau_j), \theta(\tau_j)\big) \quad \text{and} \quad r_{\max}(\mathbf{t}, \mathbf{i}) = \max_{j=0,\dots,p+1} \left|x(\tau_j)\right|,$$

for the minimum switching rate at a switching point and maximum distance from the origin for the associated trajectory, respectively. For $n \in \mathbb{N}$ and $\theta, \eta \in \{-1, 1\}^d$, define sets

$$\mathcal{U}_{n,\theta,\eta} := \{y \in \mathbb{R}^d : |y| < 2, (0, \theta) \looparrowright (y, \eta), \text{through a control } (\mathbf{t}, \mathbf{i}) \text{ such that}$$

$$\lambda^V_{\min}(\mathbf{t}, \mathbf{i}) > 1/n \text{ and } r_{\max}(\mathbf{t}, \mathbf{i}) < n\}. \tag{5}$$

Suppose $y \in \mathcal{U}_{n,\theta,\eta}$, so that there exists a control $(\mathbf{t}, \mathbf{i})$ taking $(0, \theta)$ to (y, η) by which every component of the direction vector is flipped. By perturbing the switching times $t_1, \dots, t_p$ in the control, we find that $(0, \theta) \rightsquigarrow (y', \eta)$ for all y' in a sufficiently small neighbourhood of y through a control $(\mathbf{t}', \mathbf{i}')$ such that $\lambda^V_{\min}(\mathbf{t}', \mathbf{i}') > 1/n$ and $r_{\max}(\mathbf{t}', \mathbf{i}') < n$. It follows that $\mathcal{U}_{n,\theta,\eta}$ is open for all n, θ, η. For a Gaussian density, we have $(x, \theta) \looparrowright (y, \eta)$ for all $(x, \theta), (y, \eta) \in E$ by a repeated use of Proposition 1. Thus for fixed θ, η we have the following open cover of the closed unit disc $D = \{y \in \mathbb{R}^d : |y| \le 1\}$:

$$D \subset \bigcup_{n \in \mathbb{N}} \mathcal{U}_{n,\theta,\eta}.$$

By compactness of D, for all θ, η, there exists an $N_{\theta,\eta} \in \mathbb{N}$ such that

$$\{y \in \mathbb{R}^d : |y| \le 1\} \subset \mathcal{U}_{N_{\theta,\eta},\theta,\eta}.$$

Let $N := \max_{\theta,\eta} N_{\theta,\eta}$. It follows that for every $\theta \in \{-1, 1\}^d$ and $(y, \eta) \in E$, $|y| \le 1$, we have $(0, \theta) \looparrowright (y, \eta)$ through trajectories with minimal switching rate larger than $1/N$ and a maximal distance from the origin smaller than N. By a Taylor expansion, we have that, for some constant c, which we may assume to satisfy $c > 1$,

$$\left|\nabla U(x) - Vx\right| \le c|x|^2 \quad \text{for } |x| \le 1. \tag{6}$$

Now let $\theta \in \{-1, 1\}^d$ and $(y, \eta) \in E$, such that $|y| < \gamma := \frac{1}{2cN^3}$. Let $z = y/\gamma$ so that $|z| < 1$. There exists a control sequence $(\mathbf{t}, \mathbf{i})$ for which $(0, \theta) \looparrowright (z, \eta)$ such that $\lambda^V_{\min}(\mathbf{t}, \mathbf{i}) > \frac{1}{N}$ and $r_{\max}(\mathbf{t}, \mathbf{i}) < N$. After a rescaling of $\mathbf{t}$ to $\mathbf{t}' = \gamma \mathbf{t}$, we obtain a control sequence for $(0, \theta) \looparrowright (y, \eta)$ such that $\lambda^V_{\min}(\mathbf{t}', \mathbf{i}) > \frac{\gamma}{N} = \frac{1}{2cN^4}$ (since the switching rates for the Gaussian potential scale linearly with distance from the origin), and such that the complete trajectory is contained within a ball of radius $\gamma N < \frac{1}{2cN^2} < 1$, so that we may apply (6) along the trajectory. Along the trajectory with switching times $(\tau_j)_{j=1}^p$ corresponding to the control sequence $(\mathbf{t}', \mathbf{i})$, we obtain

$$\left|\nabla U\big(x(\tau_j)\big) - Vx(\tau_j)\right| \le c\left|x(\tau_i)\right|^2 < c\gamma^2 N^2 = \frac{1}{4cN^4}, \quad j = 1, \dots, p,$$

so that, for all $j = 1, \dots, p$,

$$\begin{aligned} \lambda_{i_j}\big(x(\tau_j), \theta(\tau_j)\big) &= \big(\theta(\tau_j)\partial_{i_j} U\big(x(\tau_j)\big)\big)_+ \\ &\ge \lambda^V_{i_j}\big(x(\tau_j), \theta(\tau_j)\big) - \frac{1}{4cN^4} > \frac{1}{4cN^4} > 0, \end{aligned}$$

that is, the control sequence $(\mathbf{t}', \mathbf{i})$ is admissible for $(0, \theta) \looparrowright (y, \eta)$ with respect to the switching rates (λ_i).

By an analogous argument, there exists an admissible control sequence for $(y, \eta) \looparrowright (0, \theta)$. The statement of the proposition follows by concatenation of trajectories. □

2.4. *Flippability.* Recall that $(x, \theta) \looparrowright (y, \eta)$ if there is an admissible path from (x, θ) to (y, η) along which all components of the velocity are switched.

DEFINITION 6 (Full flippability). The process is *fully flippable* if for each (x, θ), there exists a (y, η) such that $(x, \theta) \looparrowright (y, \eta)$,

PROPOSITION 2. *If the potential U satisfies Growth Condition 1, then the process is fully flippable.*

PROOF. By definition, the process is fully flippable if for all points (x, θ), there exists an admissible control sequence $(\mathbf{i}, \mathbf{t})$ such that all indices appear in $\mathbf{i}$. Striving for a contradiction, suppose that there is an (x, θ) such that, for any admissible control sequence, there is an index in $\{1, \dots, d\}$ that does not appear in the indices sequence. Suppose that starting from (x, θ), we are able to construct, for any ε and any T, an admissible trajectory $(x(t), \theta(t))_{t \in [0,T]}$ along which the following bound holds:

$$\forall i, \forall t \in [0, T], \quad \theta(t) \partial_i U\big(x(t)\big) < \varepsilon. \tag{7}$$

Integrating U along this trajectory, we get $U(x(T)) \leq U(x) + \varepsilon d T$. However, by hypothesis, this trajectory leaves at least one index in the velocity unchanged, so $\|x(T) - x\|_\infty \geq T$. This shows that

$$\inf\big\{U(y) : y \text{ such that } \|y - x\|_\infty \geq T\big\} \leq U(x) + \varepsilon\, dT,$$

and is therefore not larger than $U(x)$ by taking ε to zero. This contradicts the hypothesis that U converges to infinity.

Let us now prove that such trajectories exist. Fix $\varepsilon > 0$, and say T is "nice" if there exists an admissible control sequence starting from (x, θ) such that the bound (7) holds. The set of nice T is clearly open in $[0, \infty)$, so it will be enough to check that it is closed.

To this end, suppose that the T_n are an increasing sequence of nice times converging to T. The natural idea to construct a nice trajectory of length T is to pick a trajectory of length T_n and continue it in the final direction θ_{T_n} until time T. The corresponding trajectory will be admissible, but it may fail to satisfy (7) if, during the interval $[T_n, T)$, one of the quantities $(\theta(t))_i \partial_i U(x(t))$ crosses the level ε. We will prove that by switching the corresponding indices, we can construct a nice trajectory.

Since the process moves at finite speed, we know that all admissible trajectories of length less than T starting from (x,θ) will lie in a bounded set, only depending on T. Let C_T be an upper bound on the Hessian of U on this bounded set. Let n be large enough so that $T - T_n < \varepsilon/2C_T$, and consider a "nice" trajectory of length T_n; we wish to continue it up to time T. Let $D = \{i_1,\dots,i_m\}$ be the set of "dangerous" indices, that is, indices for which $\theta_i \partial_i U(x(T_n)) > \varepsilon/2$. Consider the trajectory obtained by concatenating the nice T_n control sequence with the sequence $(i_1,\dots,i_m;\varepsilon',\dots,\varepsilon',T-T_n-m\varepsilon')$. If ε' is small enough, this trajectory will be both admissible and nice: all "dangerous" indices will be switched before the corresponding product reaches ε, and they will not have time to grow up to ε again. The set of nice T is therefore $[0,\infty)$ in its entirety. $\square$

2.5. *Reachability in the general case.*

LEMMA 5. *If $(x,\theta) \looparrowright (y,\eta)$, then there is an open neighbourhood U of (y,η) such that for all $(y',\eta') \in U$, $(x,\theta) \looparrowright (y',\eta')$.*

PROOF. By hypothesis, there is a sequence of times and indices such that

$$y = x + t_0\theta + t_1 F_{i_1}\theta + \cdots + t_n F_{i_1,\dots,i_n}\theta.$$

Define $\Phi : (s_0,\dots,s_n) \mapsto x + s_0\theta + s_1 F_{i_1}\theta + \cdots s_n F_{i_1,\dots,i_n}\theta$. Then $D\Phi = (\theta, F_{i_1}\theta,\dots,F_{i_1,\dots,i_n}\theta)$. Since the difference between two consecutive vectors in this family is $\pm 2e_{i_k}$, the map Φ has full rank if all components are switched at least once. Therefore, Φ is a submersion from a neighbourhood of $(t_0,\dots,t_n)$ to a neighbourhood of y. By continuity of the switching rates, we may assume without loss of generality that for all $(s_0,\dots,s_n)$ in this neighbourhood, the corresponding trajectory is admissible. Since the sequence of switches is the same as the original trajectory, we get the conclusion. $\square$

Say $(x,\theta) \sim (x',\theta')$ if they are equal or if $(x,\theta) \rightsquigarrow (x',\theta') \rightsquigarrow (x,\theta)$. Denote by $\mathrm{Cl}(x,\theta)$ the equivalence class of (x,θ) and by R the velocity reversal (applied to points in, or subsets of, $\mathbb{R}^d \times \{-1,1\}^d$).

LEMMA 6. *The equivalence classes of $\sim$ are either a single point or an open set in $\mathbb{R}^d \times \{-1,1\}^d$.*

For any (x,θ), $R(\mathrm{Cl}(x,\theta)) = \mathrm{Cl}(R(x,\theta))$. In particular, the classes of (x,θ) and $(x,-\theta)$ have the same type (open or singleton).

PROOF. Suppose that (x,θ) and (x',θ') are two different equivalent points. This means that there is an admissible loop starting from, and returning to, (x,θ). Along such a loop all components of the velocity must be flipped at least once: if the ith component of the velocity stays at 1 (resp., -1), then the ith component of the position strictly increases (resp., decreases) along the loop, a contradiction. Therefore, if $\mathrm{Cl}(x,\theta)$ is not a singleton, then $x \looparrowright x$.

Let us now prove openness. If (y,η) is in the nontrivial class of (x,θ), then $(x,\theta) \looparrowright (x,\theta) \rightsquigarrow (y,\eta)$, so (x,θ) leads to all points near (y,η). Similarly, $(x,-\theta) \looparrowright (y,-\eta)$, so $(x,-\theta)$ leads to all points in a neighbourhood of $(y,-\eta)$, and by reversal, all points near (y,η) must lead to (x,θ). Therefore, all points near (y,η) are in fact equivalent to (x,θ) and the class is open.

The reversal property is a consequence of the similar property for $\rightsquigarrow$. □

PROPOSITION 3 (Stability of open classes). *The open equivalent classes are "almost stable" under $\rightsquigarrow$ and its inverse, that is, if the class of (x,θ) is open, then for π-almost every (y,η), we have the equivalence $(y,\eta) \rightsquigarrow (x,\theta) \iff (x,\theta) \rightsquigarrow (y,\eta) \iff (x,\theta) \sim (y,\eta)$.*

If the process is fully flippable in the sense of Definition 6, then the open classes are of the form $\mathbb{R}^d \times V$, where V is a subset of the velocities $\{-1,1\}^d$.

REMARK 9 (Terminology). In the countable state setting, classes that are stable under the analogue of $\rightsquigarrow$ are called "essential" (see, e.g., [23]). In a general state space, it is known that the communication structures are more difficult to define and study; this has led in particular to the definition of ψ-irreducibility; see [28], Chapter 5. It turns out that in our particular case, the relation $\rightsquigarrow$ defines interesting equivalence classes that we can study before discussing ψ-irreducibility.

PROOF OF PROPOSITION 3. The first step is probabilistic.

Let O be an open class. Let O_+ be the "future" of O, that is, the set of (y,η) such that there exists $(x,\theta) \in O$ such that $(x,\theta) \rightsquigarrow (y,\eta)$. Note that since $(x,\theta) \looparrowright (x,\theta)$, O_+ is open, therefore measurable. Let $P^t((x,\theta),A)$ denote the Markov transition kernel of the zigzag process. Let us use the invariance of π through the resolvent kernel:

$$\begin{aligned}
\pi(O_+) &= \int_0^\infty e^{-t} \pi P^t(\cdot, O_+)\,dt \\
&= \int_0^\infty \int_E e^{-t} \mathbb{P}_{(x,\theta)}\big[(X_t,\Theta_t) \in O_+\big]\,d\pi(x,\theta)\,dt \\
&= \int_0^\infty \int_E e^{-t} \mathbf{1}_{(x,\theta)\in O_+} \mathbb{P}_{(x,\theta)}\big[(X_t,\Theta_t) \in O_+\big]\,d\pi(x,\theta)\,dt \\
&\quad + \int_0^\infty \int_E e^{-t} \mathbf{1}_{(x,\theta)\notin O_+} \mathbb{P}_{(x,\theta)}\big[(X_t,\Theta_t) \in O_+\big]\,d\pi(x,\theta)\,dt.
\end{aligned}$$

Since O_+ is stable by $\rightsquigarrow$, the probability in the first integral is 1, so the whole first integral is equal to $\pi(O_+)$. Therefore, the second integral must vanish: for all (x,θ) in some set A of full π-measure,

$$\mathbf{1}_{(x,\theta)\notin O_+} \int_0^\infty e^{-t} \mathbb{P}_{(x,\theta)}\big[(X_t,\Theta_t) \in O_+\big]\,dt = 0.$$

If (x, θ) is in A and leads to a point in O, then the probability above is strictly positive, so (x, θ) must be in O_+. Consequently, we can build a loop from (x, θ) that intersects O, so (x, θ) is in O.

In the other direction, we use reversal. Without loss of generality, we may assume A is stable by reversal of velocities. If (x, θ) in A is reachable from a point (y, η) in O, then $(x, -\theta) \rightsquigarrow (y, -\eta)$, so $(x, -\theta) \in RO$, and $(x, \theta) \in O$.

We now prove a stronger stability statement by getting rid of the "π-almost surely". Consider a point (x, θ) in an open class O and suppose that (y, η) is reachable from (x, θ). By the assumption, we can find a (z, ξ) such that $(y, \eta) \looparrowright (z, \xi)$. By Lemma 5, $(y, \eta) \rightsquigarrow (z', \xi')$ for all (z', ξ') in a neighbourhood of (z, ξ). By transitivity, (x, θ) itself leads to all points in this neighbourhood. Such a neighbourhood must have a positive π-measure, so at least one of the (z', ξ') leads back to (x, θ). Therefore, we have a loop $(x, \theta) \rightsquigarrow (y, \eta) \rightsquigarrow (z', \xi') \rightsquigarrow (x, \theta)$, so all three points are in the same class, so open classes are stable by $\rightsquigarrow$. Using reversal, it is easy to see that they are also stable in the other direction.

The third step of the proof is to use the stability to prove that nontrivial classes are closed, and must therefore consist of a certain number of copies of $\mathbb{R}^d$. Let O be a nontrivial class, and let (x, θ) be a point in the (topological) closure of O. By Lemma 5, there exists a (y, η) and an open set $\mathcal{U}$ such that (x, θ) leads to all points in $\mathcal{U}$. Write $y = x + t_0\theta + \cdots + F_{i_1,\dots,i_n}\theta$ for some sequence of times and indices. By continuity of the switching rates, the same control sequence will be admissible if x' is close to x, and will lead from (x', θ) to the point $(y', \eta) = (y + x' - x, \eta)$. Since x is in the closure of O, we can find x' in O such that $(x', \theta) \rightsquigarrow (y', \eta)$, and we may assume that (y', η) is in $\mathcal{U}$, so that $(x, \theta) \rightsquigarrow (y', \eta)$. Since (x', θ) is in O, (y', η) is also in O by forward stability, so (x, θ) is itself in O, proving that O is closed. $\square$

THEOREM 4. *If the potential U is $\mathcal{C}^3$, satisfies Growth Condition 1, and has a nondegenerate local minimum, then there is only one equivalence class. In particular $(x, \theta) \rightsquigarrow (y, \eta)$ for all $(x, \theta) \in E$ and $(y, \eta) \in E$.*

PROOF. By the local minimum approximation result (Lemma 4), we know that there exists an open set $\mathcal{U}$ such that all points in $\mathcal{U} \times \{-1, 1\}^d$ are in the same equivalence class, say O. By Lemma 6, O must then be open. Since the potential U goes to infinity, the process is fully flippable by Proposition 2, so we may apply Proposition 3 to see that O consists of copies of $\mathbb{R}^d$. Since O contains $\mathcal{U} \times \{-1, 1\}^d$, it follows that $O = E$. $\square$

3. Ergodicity and exponential ergodicity. To prove ergodicity and exponential ergodicity, we will use standard results from [15, 28, 29, 31, 42]. In order to show that they apply, we need to check a certain number of properties of the process. Some of these properties (aperiodicity, irreducibility) are analogues in the continuous time and continuous space setting of classical notions for Markov chains. In order to guarantee that the process does not behave too wildly with respect to the topology of the ambient space, Meyn and Tweedie have also introduced the notion of T-processes (where T stands for "topology"). We will first recall these here, phrased in terms of a general Markov process (Z_t) taking values in a space E, for completeness. For a more detailed overview of these notions, we refer to the aforementioned papers, in particular [31], and the reference book [28].

For a given measure ψ, a process is *ψ-irreducible* if for any starting point z and any set A of positive ψ-measure, $\mathbb{E}_z[\int_0^\infty \mathbf{1}_A(Z_t)\,dt] > 0$. It is a *$T$-process* if there exists a probability distribution a on $\mathbb{R}_+$ and a kernel $K(z, A)$ such that for fixed A, $z \mapsto K(z, A)$ is lower semi-continuous, and for fixed z, $K(z, E) > 0$ and we have the lower bound:

$$\int \mathbb{P}_z[Z_t \in A]\,da(t) \geq K(z, A).$$

A measurable set $C \subset E$ is called *petite* if there exists a probability distribution a, a constant $c > 0$ and a nontrivial measure ν on E such that

$$\int \mathbb{P}_z[Z_t \in \cdot]\,da(t) \geq c\nu(\cdot) \quad \text{for all } z \in C.$$

An irreducible process is called *aperiodic* if there exists a petite set C and a time T such that $\mathbb{P}_z[Z_t \in C] > 0$ for all starting points $z \in C$ and all times $t \geq T$. The process is called *Harris recurrent* if, for some σ-finite measure φ, $\mathbb{P}_z[\int_0^\infty \mathbf{1}_A(Z_t)\,dt = \infty] \equiv 1$ whenever $\varphi(A) > 0$. As discussed in [31], Harris recurrence implies existence of a unique (up to constant multiples) invariant measure. If, moreover, there is a finite invariant measure (which in this paper is always the case by assumption (4)), the process is called *positive Harris (recurrent)*.

In the next sections, we establish that the zigzag process is in fact an irreducible, aperiodic T-process; Section 3.4 is devoted to finding a suitable Lyapunov function.

3.1. *Continuous components*. In this section, we give two results on the existence of an absolutely continuous component in the distribution of the position of the process. We start with an easy result, expressed in terms of a certain stopping time.

LEMMA 7 (Absolute continuity from jumps). *Let (T_i) be the random times where the components of the velocity switch. Let N be the random integer such that T_N is the first time when $d-1$ components have switched; let $N = \infty$ if this does not occur. Let $\tau = T_{N+1}$ if T_N is finite, and $\tau = \infty$ otherwise.*

Then the distribution of X_τ (conditionally on $\tau < \infty$) is absolutely continuous with respect to the Lebesgue measure: if B is a Borel set in $\mathbb{R}^d$ of Lebesgue measure zero, then

$$\mathbb{P}[\tau < \infty, X_\tau \in B] = 0.$$

In particular, in case $d = 1$, then $N = 0$, $T_N = 0$ and τ is the time of the first switch.

PROOF OF LEMMA 7. Let B be a set of zero Lebesgue measure in $\mathbb{R}^d$ and t be arbitrary. It is enough to show that $\mathbb{P}_{(x,\theta)}[\tau \le t; X_\tau \in B] = 0$, since this implies $\mathbb{P}_{(x,\theta)}[X_\tau \in B, \tau < \infty] = 0$ by monotone convergence.

It is well known (see [4, 7]) that the law of (X_t, Θ_t) may be obtained by a thinning procedure. More precisely, let $\overline{\lambda}$ be an upper bound on the switching rates up to time t (such a bound exists since the process has finite speed and the switching rates are continuous). Then the process may be constructed on $[0, t]$ by running a Poisson clock with intensity $\overline{\lambda}d$, and, for each Poisson event, picking an index i uniformly, then accepting or rejecting the flip of the corresponding component of the velocity with a probability given by $\lambda_i(x, \theta)/\overline{\lambda}$.

Recall that $F_{i_1,\dots,i_k}\theta$ is the velocity obtained from θ by flipping, possibly many times, the components appearing in the sequence. For convenience, we extend this definition to allow zero values in the index sequence, which corresponds to no flipping. This allows us to write

$$X_\tau = x + E_1\theta + E_2 F_{I_1}\theta + \cdots + E_{M+1} F_{I_1,\dots,I_m}\theta,$$

where M is a random integer (larger than N), the (I_k) take values in $\{0, 1, \dots, d\}$ with $I_k = j$ for $j \neq 0$ indicating a proposed and accepted j flip, while $I_k = 0$ corresponding to all rejected flips, and the (E_i) are the interarrival times of the Poisson clock. We decompose over all possible index sequences:

$$\begin{aligned}
&\mathbb{P}[\tau \le t, X_\tau \in B] \\
&\quad = \sum_{m\in\mathbb{N}_0} \sum_{(i_1,\dots,i_m)\in\{0,\dots d\}^m} \mathbb{P}\big[\tau \le t, M = m, (I_1, \dots, I_M) = (i_1, \dots, i_m), \\
&\qquad (x + E_1\theta + \cdots + E_{m+1} F_{i_1,\dots,i_m}\theta) \in B\big].
\end{aligned}$$

If $M = m$, $N \le m$ so by definition, at least $d - 1$ different (nonzero) indices must appear in the sequence $(i_1, \dots, i_m)$, and

$$\begin{aligned}
&\mathbb{P}[\tau \le t, X_\tau \in B] \\
&\quad \le \sum_{m\in\mathbb{N}} \sum_{\substack{(i_1,\dots,i_m)\in\{0,\dots d\}^m \\ d-1 \text{ indices appear in } (i_1,\dots,i_m)}} \mathbb{P}\big[(x + E_1\theta + \cdots + E_{m+1} F_{i_1,\dots,i_m}\theta) \in B\big].
\end{aligned}$$

For each term in the sum, the vectors $(\theta, F_{i_1}\theta, \dots, F_{i_1,\dots,i_m}\theta)$ span $\mathbb{R}^d$, so the distribution of $x + E_1\theta + \cdots + E_{m+1} F_{i_1,\dots,i_m}\theta$ is absolutely continuous, and the probability that it falls in the set B is zero. $\square$

The proof of the existence of an absolutely continuous component at a fixed time is a bit more involved, but is the key ingredient to prove that the process behaves nicely.

LEMMA 8 (Continuous component). *If* $(x,\theta) \looparrowright (y,\eta)$, *then there exist open sets* U *and* V, *with* $x \in U$ *and* $y \in V$, *and constants* $\varepsilon > 0$, $t_0 > 0$, $c > 0$, *such that for any* $x' \in U$, *and all* $t \in (t_0, t_0+\varepsilon]$,

$$\mathbb{P}_{x',\theta}[X_t \in \cdot, \Theta_t = \eta] \geq c \cdot \mathrm{Leb}(\cdot \cap V).$$

REMARK 10. Similar results may be found in previous works, for example, [3], Lemmas 2 and 3, or [4], Section 6.5. In order to get the probabilistic consequences, we need the uniformity in the starting point that appears in [4]. Since our hypotheses here are slightly different, we include a proof for the sake of completeness. We also note that taking canonical switching rates leads to a degenerate situation where the local Hörmander type criteria of [3, 4] do not apply.

PROOF OF LEMMA 8. By hypothesis, there exists an admissible deterministic control sequence $\mathbf{u} = (\mathbf{t}, \mathbf{i}) = (t_0, \dots, t_m; i_1, \dots, i_m)$, such that all indices occur at least once in $\mathbf{i}$, and $\Phi_{\mathbf{u}}(x,\theta) = (y,\eta)$. Recall the notation $\tau_k = \sum_{j=0}^{k-1} t_j$ and let $t = \tau_{m+1} = \sum_{k=0}^{m} t_k$ be the final time of the trajectory.

We use the same thinning construction as in the proof of Lemma 7 above, with a Poisson clock of intensity $\overline{\lambda} d$, where $\overline{\lambda}$ is an upper bound on the switching rates up to time t.

For $j = 1, \dots, (m-1)$, let $\mathcal{U}_j$ be a bounded neighbourhood of τ_j; we may assume that the $\mathcal{U}_j$ do not intersect and, by continuity, that the control sequences $(\mathbf{s}, \mathbf{i}) = (s_0, \dots, s_{m-1}, \mathbf{i})$ satisfy $\lambda_{\min}(\mathbf{s}, \mathbf{i}) \geq \underline{\lambda} > 0$ for any $\mathbf{s}$ such that $\sum_{l=0}^{j-1} s_l \in \mathcal{U}_j$ for all j.

Now let f be an arbitrary nonnegative test function. Let A be the event that m Poisson events $T_1, \dots, T_m$ occur before time t, that $T_j \in \mathcal{U}_j$ for all j, that the indices are picked as in $\mathbf{i}$, and that all proposed switches are accepted. Then

$$\begin{aligned}
\mathbb{E}\big[f(X_t, \Theta_t)\big] &\geq \mathbb{E}\big[f(X_t, \Theta_t)\mathbf{1}_A\big] \\
&\geq \mathbb{E}\big[f\big(\Psi(x, t, T_1, \dots, T_m)\big)\mathbf{1}_A\big],
\end{aligned}$$

where the mapping Ψ is defined by

$$\Psi(x, t, \tau_1, \dots, \tau_m) = x + \tau_1\theta + (\tau_2 - \tau_1)F_{i_1}\theta + \cdots + (t - \tau_m)F_{i_1 \cdots i_m}\theta.$$

Since the choice of indices to switch and the acceptance/rejection tests are independent from the Poisson process, we get by conditioning:

$$\mathbb{E}\big[f(X_t, \Theta_t)\big] \geq \left(\frac{\underline{\lambda}}{\overline{\lambda} d}\right)^m \mathbb{E}\left[f\big(\Psi(x, t, T_1, \dots, T_m)\big)\mathbf{1}_{m \text{ events occur}} \prod_{j=1}^{m} \mathbf{1}_{T_j \in \mathcal{U}_j}\right].$$

Using classical properties of the Poisson process, this implies that for some positive constant c,

$$\mathbb{E}\big[f(X_t, \Theta_t)\big] \geq c\mathbb{E}\big[f\big(\Psi(x, t, U_1, \dots, U_m)\big)\big], \tag{8}$$

where the U_j are independent and U_j is uniformly distributed on $\mathcal{U}_j$.

The partial map $(u_1, \dots, u_m) \mapsto \Psi(x, t, u_1, \dots, u_m)$ has full rank: indeed, the image of its differential is spanned by the vectors

$$(\theta - F_{i_1}\theta, \dots, F_{i_1 \cdots i_{m-1}}\theta - F_{i_1 \cdots i_m}\theta) = (\pm 2e_{i_1}, \dots, \pm 2e_{i_m})$$

who span $\mathbb{R}^d$ since all indices in $\{1, \dots, d\}$ appear at least once in the sequence $\mathbf{i}$. This shows that $\Psi(x, t, \cdot)$ is a submersion. It follows that, $\Psi(x, t, \cdot)$ pushes the uniform distribution on $\prod \mathcal{U}_j$ to a measure which is absolutely continuous with respect to the Lebesgue measure, on an open set containing $y = \Phi_{\mathbf{u}}(x, \theta)$ (see [3], Lemmas 2 and 3, [4], Section 6, for related results and details). This proves a restricted form of the lemma, for the single starting point x and the single time t.

To prove the uniform version, we see x and t as a parameter and apply the uniform submersion lemma [4], Lemma 6.3, to get the result. $\square$

3.2. *Nonevanescence.* For classical Markov chains on countable spaces, it is well known that for any x and y, the following equivalence holds:

$$\mathbb{E}_x\left[\sum_n \mathbf{1}_{X_n = y}\right] = \infty \quad \iff \quad \sum_n \mathbf{1}_{X_n = y} = \infty, \quad \mathbb{P}_x\text{-a.s.}$$

For general chains and processes, this equivalence is no longer true: Starting from a point x, the time spent in a set A may be finite with positive probability, even when its expectation is infinite. This may essentially happen if the process has a positive probability of escaping to infinity when it starts in a particular set: This canonical counter-example is explained, for example, in [28], Section 9.1.2.

This equivalence is used to prove that a (classical) irreducible chain that admits an invariant probability measure is positive recurrent. To obtain the natural property of Harris recurrence for a general chain, (ψ-)irreducibility and the existence of the invariant probability are not enough, and we need to show additionally that the escaping to infinity does not happen.

In the context of the zigzag process, we refer to the "ridge", Example 3 in Section 1.3, which describes a smooth potential function with the property that for certain initial conditions the zigzag process will escape to infinity with full probability.

DEFINITION 7 (Nonevanescence). A point (x, θ) is said to be *nonevanescent* if $\mathbb{P}_{x,\theta}[|X_t| \to \infty] = 0$. It is *weakly nonevanescent* if this probability is strictly less than 1.

We start by showing how the deterministic statements on flippability may be used to prove probabilistic nonevanescence properties.

REMARK 11 (There are infinitely many switches). Note that the first growth condition $U \to \infty$ already has the probabilistic consequence that the process switches infinitely often. Indeed, for any (x,θ) and any n,

$$\begin{aligned}\mathbb{P}_{(x,\theta)}[\text{no switch before time } n] &= \exp\left(-\int_0^n \sum_{i=1}^d \big(\theta_i \partial_i U(x+\theta s)\big)_+ \, ds\right)\\ &\le \exp\left(-\int_0^n \sum_{i=1}^d \theta_i \partial_i U(x+\theta s)\, ds\right)\\ &= \exp\big(-U(x+\theta n)+U(x)\big) \to 0 \quad \text{as } n\to\infty,\end{aligned}$$

so $\mathbb{P}_{(x,\theta)}[T^1<\infty]=1$, where (T^i) are the switching times as introduced in Section 1.2. By the strong Markov property, this implies for all k

$$\mathbb{P}_{(x,\theta)}\big[T^{k+1}<\infty\big] = \mathbb{E}_{(x,\theta)}\big[\mathbf{1}_{T^k<\infty}\mathbb{P}_{(X_{T_k},\Theta_{T^k})}\big[T^1<\infty\big]\big] = \mathbb{P}_{(x,\theta)}\big[T^k<\infty\big],$$

proving the claim by recurrence.

LEMMA 9 (Two weak versions of nonevanescence). *If the invariant measure π is a probability measure* (*as it is assumed to be in this paper*), *then π-almost all points are nonevanescent*.

If additionally the process is fully flippable in the sense of Definition 6, *then all points are weakly nonevanescent*.

PROOF. The first statement is classical. For the sake of completeness, we include a proof. Let K be a compact set. Since $\liminf_{t\to\infty}\mathbf{1}_{X_t\notin K}=\{X_t \text{ eventually leaves } K\}$, we have by Fatou's lemma

$$\mathbb{P}_\pi[X_t \text{ eventually leaves } K] \le \liminf_{t\to\infty}\mathbb{P}_\pi[X_t\notin K] = 1-\pi(K).$$

Since $\{|X_t|\to\infty\} = \bigcap_K\{X_t \text{ eventually leaves } K\}$, we are done since $\{\pi\}$ is tight.

Let us now prove the second statement. Let $\mathcal{N}$ be the set of nonevanescent points: this set has full π-measure, so its complement is Lebesgue negligible. Let (x,θ) be an arbitrary starting point, and consider the stopping time τ introduced in Lemma 7. By the strong Markov property,

$$\begin{aligned}&\mathbb{P}_{(x,\theta)}\big[|X_t| \text{ does not go to infinity}\big]\\ &\quad\ge \mathbb{P}_{(x,\theta)}\big[\tau<\infty, |X_t| \text{ does not go to infinity}\big]\\ &\quad= \mathbb{E}_{(x,\theta)}\big[\mathbf{1}_{\tau<\infty}\mathbb{P}_{(X_\tau,\Theta_\tau)}\big[|X_t| \text{ does not go to infinity}\big]\big]\\ &\quad\ge \mathbb{E}_{(x,\theta)}[\mathbf{1}_{\tau<\infty}\mathbf{1}_{X_\tau\in\mathcal{N}}].\end{aligned}$$

Since $\mathbb{R}^d\setminus\mathcal{N}$ is Lebesgue negligible, $\mathbb{P}_{(x,\theta)}[\tau<\infty, X_\tau\notin\mathcal{N}]=0$, so

$$\mathbb{P}_{(x,\theta)}\big[|X_t| \text{ does not go to infinity}\big] \ge \mathbb{P}_{(x,\theta)}[\tau<\infty]. \tag{9}$$

If the process is fully flippable, this last probability is positive, proving the weak non-evanescence property. $\square$

If we add a slightly stronger hypothesis on the growth of the potential at infinity, namely Growth Condition 2, we get a stronger nonevanescence result. We start by saying that if the process is evanescent, it must go to infinity in a very particular way, by staying forever in an affine subspace.

LEMMA 10 (Two frozen directions). *Let $d \geq 2$. Suppose that there exists an invariant probability measure, and that (x, θ) satisfies $\mathbb{P}_{(x,\theta)}[|X_t| \to \infty] > 0$. Then there exist two indices i and j such that*

$$\mathbb{P}_{(x,\theta)}[\textit{the } i\textit{th and } j\textit{th components never switch}] > 0.$$

PROOF. We prove this statement by contraposition and assume that, with probability one, at most one component of the velocity does not switch. This implies that the time T_N defined in Lemma 7 is a.s. finite, and since there are infinitely many switches by Remark 11, the time $\tau = T_{N+1}$ of the same Lemma 7 is also finite. Reusing the bound (9) from the proof of Lemma 9, we immediately get that $\mathbb{P}_{(x,\theta)}[|X_t| \to \infty] = 0$, proving the lemma. □

Recall that Growth Condition 2 states, in dimension d, that

$$\exists c > d, \exists c', \forall x, \quad U(x) \geq c \ln\big(1 + |x|\big) - c'.$$

PROPOSITION 4 (Nonevanescence). *If the potential U satisfies Growth Condition 2, then the process is nonevanescent, that is, for any $(x, \theta) \in \mathbb{R}^d \times \{-1, 1\}^d$,*

$$\mathbb{P}_{(x,\theta)}\big[|X_t| \to \infty\big] = 0.$$

PROOF OF PROPOSITION 4. We wish to prove for all d the following statement:

$$(\mathcal{P}_d) \qquad \begin{aligned} &\forall U : \mathbb{R}^d \to \mathbb{R}, \quad U \text{ satisfies GC } 2 \\ &\quad \Longrightarrow \quad \text{the zigzag process for } U \text{ is nonevanescent.} \end{aligned}$$

If $d = 1$, by (9), with τ denoting the time of the first switch, and Remark 11, $(\mathcal{P}_d)$ follows.

For $d \geq 2$, the strategy is to prove this by induction. The form of the growth condition is tailored to this strategy: it clearly implies that $\int \exp(-U(x))\,dx$ is finite and may be normalized into a probability, but it crucially also implies that the same is true for all the conditional measures on affine subspaces. For the base case $d = 2$, using Lemma 10, we see that if $\mathbb{P}_{(x,\theta)}[|X_t| \to \infty] > 0$ then with positive probability the process never switches. Since $U \to \infty$, this is not possible (see Remark 11).

Let us now prove the induction step by contraposition. Assume that $(\mathcal{P}_{d+1})$ is false: there exists a potential U in dimension $d + 1$ that satisfies the growth condition, but for which the zigzag process is evanescent, that is, there is a point (x, θ)

such that $\mathbb{P}_{(x,\theta)}[|X_t| \to \infty] > 0$. Our goal is to define a potential in dimension d that also satisfies the growth condition and for which we also have evanescence.

By Lemma 10, there are two indices, say d and $d+1$ without loss of generality, such that

$$\mathbb{P}_{(x,\theta)}[d \text{ and } d+1 \text{ never switch}] > 0.$$

We may also assume without loss of generality that $\theta_d = \theta_{d+1} = 1$. Note that the process may be constructed by considering $d+1$ sequences of i.i.d. exponential random variables $(E_j^k)_{j=1,\dots,d+1;k\in\mathbb{N}}$ and saying that the kth jump of the jth component of Θ, say T_j^k, occurs when the accumulated jump rate $\int_{T_j^{k-1}}^t \lambda_j(X_s, \Theta_s)\,ds$ reaches E_j^k.

Consider now a second, d-dimensional zigzag process $(Y_1, \dots, Y_d; H_1, \dots, H_d)$ starting from $(x_1, \dots, x_d; \theta_1, \dots, \theta_d)$ in the potential $V(y_1, \dots, y_d) = U(y_1, \dots, y_d, y_d)$. Note that, since U satisfies the growth condition,

$$\begin{aligned} V(y_1, \dots, y_d) &\geq c \ln\bigl(1 + \bigl|(y_1, \dots, y_d, y_d)\bigr|_{\mathbb{R}^{d+1}}\bigr) - c' \\ &\geq c \ln\bigl(1 + \bigl|(y_1, \dots, y_d)\bigr|_{\mathbb{R}^d}\bigr) - c', \end{aligned}$$

where $c > d+1 > d$, so V satisfies the growth condition in dimension d. It remains to show that the zigzag process in V is evanescent.

We couple the process in V with the previous one, using the same randomness $(E_j^k)_{j=1,\dots,d-1,k\in\mathbb{N}}$ for the first $d-1$ coordinates, and an independent sequence $(\tilde{E}_d^k)_{k\in\mathbb{N}}$ for the last one. Let τ be the first time when one of Θ_d, Θ_{d+1} or H_d switches. For $t \leq \tau$, using the elementary bound $(a+b)_+ \leq a_+ + b_+$ and the fact that H_d, Θ_d and Θ_{d+1} are all equal to 1 up to time t, we get

$$\begin{aligned} &\int_0^t \bigl(\partial_d V(Y_s) H_d(s)\bigr)_+ \,ds \\ &\quad = \int_0^t \bigl(\partial_d U(Y_s) + \partial_{d+1} U(Y_s)\bigr)_+ \,ds \\ &\quad = \int_0^t \bigl(\partial_d U(X_s) + \partial_{d+1} U(X_s)\bigr)_+ \,ds \\ &\quad \leq \int_0^t \bigl(\Theta_d(s)\partial_d U(X_s)\bigr)_+ \,ds + \int_0^t \bigl(\Theta_{d+1}(s)\partial_{d+1} U(X_s)\bigr)_+ \,ds \\ &\quad \leq \int_0^\infty \bigl(\Theta_d(s)\partial_d U(X_s)\bigr)_+ \,ds + \int_0^\infty \bigl(\Theta_{d+1}(s)\partial_{d+1} U(X_s)\bigr)_+ \,ds. \end{aligned}$$

Now, the event $A = \{\tilde{E}_d^1 \geq E_d^1 + E_{d+1}^1\} \cap \{\Theta_d \text{ and } \Theta_{d+1} \text{ never switch}\}$ has positive probability, and on this event we can continue the bounds:

$$\begin{aligned} &\int_0^t \bigl(\partial_d V(Y_s) H_d(s)\bigr)_+ \,ds \\ &\quad \leq \int_0^\infty \bigl(\Theta_d(s)\partial_d U(X_s)\bigr)_+ \,ds + \int_0^\infty \bigl(\Theta_{d+1}(s)\partial_{d+1} U(X_s)\bigr)_+ \,ds \end{aligned}$$

$$< E_d^1 + E_{d+1}^1$$
$$\leq \tilde{E}_d^1.$$

This shows that on A, τ must be infinite, that is, H_d never switches either, and thus $|Y_t| \to \infty$. Since the growth hypothesis is satisfied for V, this concludes the proof of the induction step by contraposition. □

3.3. *Putting the pieces together.*

THEOREM 5. *If the zigzag process is fully flippable, then it is a weakly nonevanescent T-process.*

If in addition $(x, \theta) \rightsquigarrow (y, \eta)$ for all pairs of points, the process is ψ-irreducible and aperiodic, and all compact sets are petite.

If in addition the process is (strongly) nonevanescent, then it is positive Harris recurrent and ergodic.

PROOF. The fact that a fully flippable zigzag process is weakly nonevanescent is a consequence of Lemma 9.

We know that all points $(x, \theta) \in E$ lead to a different point by a sequence where all indices are switched. From Lemma 8 and a compactness argument, this implies that there exists a family $(\mathcal{U}_n)_{n\in\mathbb{N}}$ of open sets in E, a family $(\mathcal{V}_n)_{n\in\mathbb{N}}$ of open sets in $\mathbb{R}^d$, velocities $\eta_n \in \{-1, 1\}^d$ and numbers $(t_n, \varepsilon_n, c_n)$, such that:

- The $(\mathcal{U}_n)_{n\in\mathbb{N}}$ form a locally finite open cover: each $(x, \theta) \in E$ belongs to at least one, and at most a finite number of the $\mathcal{U}_n$.
- for all $(x, \theta) \in \mathcal{U}_n$, all $t \in [t_n, t_n + \varepsilon_n]$ and all positive measurable f,

$$\mathbb{E}_{(x,\theta)}\big[f(X_t, \Theta_t)\big] \geq c_n \int f(y, \eta_n)\mathbf{1}_{\mathcal{V}_n}(y)\,dy.$$

Define a kernel K by the formula

$$K\big((x, \theta), A \times \{\eta\}\big) = \int \mathbf{1}_A(y) \max_{n:(x,\theta)\in\mathcal{U}_n} \left(c_n \mathbf{1}_{\eta_n=\eta}\mathbf{1}_{\mathcal{V}_n}(y) \int_{t_n}^{t_n+\varepsilon_n} e^{-t}\,dt\right) dy.$$

By construction, the resolvent is bounded below by K. For all $(x, \theta) \in \mathcal{U}_n$, we have that $K((x, \theta), E) \geq c_n \operatorname{Leb}(\mathcal{V}_n) \int_{t_n}^{t_n+\varepsilon_n} e^{-t}\,dt > 0$, that is, K is nontrivial. Moreover, for any measurable set A and any η, $K((x, \theta), A \times \{\eta\})$ is lower semi-continuous in (x, θ): indeed, if (x_j) converges to x, then the x_j will eventually belong to all the $\mathcal{U}_n$ containing x, so $K((x_j, \theta), A) \geq K((x, \theta), A)$ for j large enough. To sum up, the resolvent kernel of the process is bounded below by a nontrivial lower semi-continuous kernel: the process is a T-process.

Suppose now that $(x, \theta) \rightsquigarrow (y, \eta)$ for all pairs of points. This implies that $(x, \theta) \looparrowright (y, \eta)$ for all pairs of points. For any such pair, and any neighbourhood $\mathcal{O} \times \{\eta\}$ of (y, η), another application of Lemma 8 yields $\mathbb{P}_{x,\theta}[\tau_{\mathcal{O}} < \infty] > 0$; this in turn implies that the process is open set irreducible in the sense of [42]. By [42], Theorem 3.2 (see also [28], Proposition 6.2.2, for the similar statement for discrete time chains), the process is then ψ-irreducible.

All compact sets are petite by an application [31], Theorem 4.1(i).

To prove aperiodicity, let (x,θ) be an arbitrary point. We know that $(x,\theta) \looparrowright (x,\theta)$, so by Lemma 8, there exists t_0, ε and two open neighbourhoods $\mathcal{U}$ and $\mathcal{V}$ of x such that

$$\mathbb{P}_{x',\theta}[X_t \in \cdot, \Theta_t = \theta] \geq c\,\mathrm{Leb}(\cdot \cap \mathcal{V}), \tag{10}$$

for all $x' \in \mathcal{U}$ and $t \in [t_0, t_0+\varepsilon]$. This shows that $\mathcal{U}$ is a petite set. Writing $\mathcal{W} = \mathcal{U} \cap \mathcal{V}$, we see that $\mathcal{W}$ is petite (as a subset of $\mathcal{U}$), and for all $x' \in \mathcal{W}$ and $t \in [t_0, t_0+\varepsilon]$,

$$\mathbb{P}_{x',\theta}[X_t \in \mathcal{W}, \Theta_t = \theta] \geq c',$$

where $c' = c\,\mathrm{Leb}(\mathcal{W})$. Let $N = \lceil t_0/\varepsilon \rceil$ and $T = Nt_0$. For any $t \geq T$, let $n = \lfloor t/t_0 \rfloor$ and $t_0' = t/n$. Then $t_0' \in [t_0, t_0+\varepsilon]$, so by iteration and the Markov property,

$$\mathbb{P}_{x',\theta}[X_t \in \mathcal{W}, \Theta_t = \theta] \geq (c')^n > 0,$$

proving the aperiodicity.

To prove Harris recurrence, we use the fact that for ψ-irreducible T-processes, it is in fact equivalent to nonevanescence ([31], Theorem 3.2), and the positivity follows from the fact that there is an invariant probability measure.

It remains to show that the process is ergodic. By [31], Theorem 6.1, it is enough to prove that some skeleton chain is irreducible. To this end, first take (x,θ) an arbitrary point: we reuse Lemma 8 to define $\mathcal{U}$, $\mathcal{V}$, t_0 and ε such that equation (10) holds; in words, it is possible to loop around (x,θ) and there is a little room ε in the looping time. Now let (y,η), (y',η') be two arbitrary points. By reachability, we can go from the first one to the second one with a visit to (x,θ) in between, and adding a loop around (x,θ) in the middle will give us what we need. More formally, using Lemma 8 twice more, there exists t_1, c_1 and a neighbourhood $\mathcal{V}_1$ of x such that

$$\mathbb{P}_{(y,\eta)}\big[(X_{t_1}, \Theta_{t_1}) \in \cdot \times \{\theta\}\big] \geq c_1\,\mathrm{Leb}(\cdot \cap \mathcal{V}_1),$$

and t_2, c_2 and two neighbourhoods $\mathcal{U}_2$ and $\mathcal{V}_2$ of x and y' such that

$$\mathbb{P}_{(x',\theta)}\big[(X_{t_2}, \Theta_{t_2}) \in \cdot \times \{\eta'\}\big] \geq c_2\,\mathrm{Leb}(\cdot \cap \mathcal{V}_2)$$

for all $x' \in \mathcal{U}_2$. Then for any $t \in [t_0+t_1+t_2, t_0+t_1+t_2+\varepsilon]$, applying the Markov property at the times t_1 and $t-t_2$ yields

$$\begin{aligned}
&\mathbb{P}_{(y,\eta)}\big[(X_t, \Theta_t) \in \mathcal{O} \times \{\eta'\}\big] \\
&\quad \geq \mathbb{P}_{(y,\eta)}\big[\Theta_{t_1} = \Theta_{t-t_2} = \theta, \Theta_t = \eta', X_{t_1} \in \mathcal{U} \cap \mathcal{V}_1, X_{t-t_2} \in \mathcal{V} \cap \mathcal{U}_2, X_t \in \mathcal{O}\big] \\
&\quad \geq \mathbb{P}_{(y,\eta)}[\Theta_{t_1} = \Theta_{t-t_2} = \theta, X_{t_1} \in \mathcal{U} \cap \mathcal{V}_1, X_{t-t_2} \in \mathcal{V} \cap \mathcal{U}_2] c_2\,\mathrm{Leb}(\mathcal{O} \cap \mathcal{V}_2) \\
&\quad \geq \mathbb{P}_{(y,\eta)}[\Theta_{t_1} = \theta, X_{t_1} \in \mathcal{U} \cap \mathcal{V}_1] c\,\mathrm{Leb}(\mathcal{V} \cap \mathcal{U}_2) c_2\,\mathrm{Leb}(\mathcal{O} \cap \mathcal{V}_2) \\
&\quad \geq c c_1 c_2\,\mathrm{Leb}(\mathcal{U} \cap \mathcal{V}_1)\,\mathrm{Leb}(\mathcal{V} \cap \mathcal{U}_2)\,\mathrm{Leb}(\mathcal{O} \cap \mathcal{V}_2),
\end{aligned}$$

since $(t - t_2) - t_1 \in [t_0, t_0 + \varepsilon]$. The time interval $[t_0 + t_1 + t_2, t_0 + t_1 + t_2 + \varepsilon]$ must contain a multiple of ε, proving that the ε-chain is open set irreducible and, therefore, irreducible. □

3.4. *Lyapunov function.* In order to establish exponential ergodicity, we have to establish contractivity in the tails for which a Lyapunov function argument is used. For this, we first require the notion of the generator of the zigzag process. We define the *generator* of the zigzag process in E with switching rates $(\lambda_i)_{i=1}^d$ as the operator L whose domain $\mathcal{D}(L)$ consists of continuous functions $f : E \to \mathbb{R}$, such that $t \mapsto f(x + \theta t, \theta)$ is absolutely continuous on $[0, \infty)$ for all $(x, \theta) \in E$. For such $f \in \mathcal{D}(L)$, the function Lf is defined as

$$Lf(x, \theta) = \langle \theta, \nabla f(x) \rangle + \sum_{i=1}^{d} \lambda_i(x, \theta)\big(f(x, F_i\theta) - f(x, \theta)\big), \quad (x, \theta) \in E.$$

The main result on exponential ergodicity (Theorem 2) will be proved using the following result from Down, Meyn and Tweedie ([15], Theorem 5.2).

THEOREM 6 (Drift criterion for exponential convergence). *Suppose that* (X_t, Θ_t) *is an irreducible aperiodic process, and suppose that there exists a Lyapunov function, that is, a function* $V \geq 1$ *such that*

$$LV \leq -\varepsilon V + c\mathbf{1}_K,$$

where K *is a petite set. Then* (X_t, Θ_t) *is exponentially ergodic*:

$$\big\| \mathbb{P}_{(x,\theta)}\big[(X_t, \Theta_t) \in \cdot\big] - \pi \big\|_{\mathrm{TV}} \leq M(x, \theta) e^{-ct},$$

for some positive constant c.

As discussed in [15], the function $M(x, \theta)$ may be taken to be a positive multiple of V. The approach in [15] does not yield quantitative results on the value of c. For estimates on the rate of convergence in a L^2-framework of the zigzag processes (and other piecewise deterministic process), we refer to [1].

REMARK 12. The continuity assumption on functions in the domain $\mathcal{D}(L)$ leads to a domain which is somewhat smaller than that of the *extended generator*, characterized in [11], Theorem 26.14. However, this definition is sufficient for our purposes.

In order to motivate our choice of Lyapunov function, first note that we are looking for a function that typically decreases along the dynamics. Since the velocity has a positive probability of switching whenever the process is going "uphill" (i.e., whenever $\langle \theta, \nabla U(x) \rangle > 0$, a first guess might be $V(x, \theta) = \exp(\alpha U(x))$ for some $\alpha > 0$. However, this velocity jump will not occur immediately, therefore, we wish to introduce a dependence on the partial derivatives of U and on the direction θ so that the effect of the switching intensity is to decrease V with sufficiently large probability while we are running uphill of the potential. For a zero excess switching rate, $\gamma(x,\theta) \equiv 0$, we could simply take $V(x,\theta) = \exp(\alpha U(x) + \beta\langle\theta, \nabla U(x)\rangle)$ but for nonzero excess switching rate we have to be more careful in dependence on the partial derivatives of U. The particular structure of the zigzag process enables us to work on each component of the gradient separately.

The Lyapunov function used for the one-dimensional zigzag process (see [9]) requires milder assumptions compared to Growth Condition 3: it only requires $|U'(x)|$ to be bounded away from zero for x outside of a compact set, without any conditions on the second derivative. However, it cannot be extended to the multidimensional case in a simple way. Indeed, the multidimensional generalization

$$V(x,\theta) = \exp\bigl(\alpha\|x\| + \beta\langle\theta, x/\|x\|\rangle\bigr)$$

fails to be contractive in, for example, the case of a nondiagonally dominant Gaussian target.

The Lyapunov function we will introduce in Lemma 11 may also be compared to the Lyapunov function for the Bouncy Particle Sampler [12],

$$V(x,v) = \exp\left(\frac{1}{2}U(x)\right) - \frac{1}{2}\ln\bigl(\lambda(x,-v)\bigr), \quad (x,v) \in \mathbb{R}^d \times S^{d-1}.$$

Note that this Lyapunov function is not well defined in our situation which should include the case of canonical switching rates, where $\gamma(\cdot) \equiv 0$.

LEMMA 11. *Suppose Growth Condition 3 is satisfied. Consider the process with a switching rate given by* $\lambda_i(x,\theta) = \gamma_i(x,\theta) + (\theta_i\partial_i U(x))_+$, *where* $\gamma : E \to [0,\infty)^d$ *is bounded: for some constant* $\overline{\gamma} \geq 0$,

$$\gamma_i(x,\theta) \leq \overline{\gamma}, \quad (x,\theta) \in E, i = 1,\dots,d.$$

Let $\delta > 0$ *and* $\alpha > 0$ *such that* $0 \leq \overline{\gamma}\delta < \alpha < 1$. *Define* $\phi(s) = \frac{1}{2}\operatorname{sign}(s)\ln(1 + \delta|s|)$. *Then the function*

$$V(x,\theta) = \exp\left(\alpha U(x) + \sum_i \phi\bigl(\theta_i\partial_i U(x)\bigr)\right) \tag{11}$$

is a Lyapunov function for (X_t,Θ_t), *that is*, $\lim_{|x|\to\infty} V(x) = \infty$ *and*

$$LV \leq -\varepsilon V + C\mathbf{1}_K,$$

where ε, C *are positive constants and* K *is a compact set in* E.

PROOF. It may be verified that $V \in \mathcal{D}(L)$. Using the expression of the generator,

$$\begin{aligned}(LV/V)(x,\theta) = {} & \alpha\langle\theta, \nabla U(x)\rangle + \sum_{i,j}\theta_i\partial_{ij}U(x)\theta_j\phi'\bigl(\theta_j\partial_j U(x)\bigr) \\ & + \sum_i\bigl(\gamma_i + (\theta_i\partial_i U)_+\bigr)\bigl(\exp\bigl(\phi(-\theta_i\partial_i U) - \phi(\theta_i\partial_i U)\bigr) - 1\bigr).\end{aligned}$$

For the ith component, if $s = \theta_i \partial_i U \geq 0$, then $\phi(-s) - \phi(s) = -\ln(1 + \delta s)$, so

$$\begin{aligned}&\alpha s + (\gamma_i + (s)_+)\big(\exp(\phi(-s) - \phi(s)) - 1\big)\\&\quad = (\alpha - 1)s + \frac{(1 - \delta\gamma_i)s}{1 + \delta s} \leq -(1 - \alpha)|s| + (1/\delta).\end{aligned}$$

When $s < 0$, we have $\phi(-s) - \phi(s) = \ln(1 + \delta|s|)$, so

$$\begin{aligned}&\alpha s + (\gamma_i + (s)_+)\big(\exp(\phi(-s) - \phi(s)) - 1\big)\\&\quad = \alpha s + \gamma_i\big(1 + \delta|s| - 1\big) \leq -(\alpha - \overline{\gamma}\delta)|s|.\end{aligned}$$

In either case,

$$\alpha s + (\gamma_i + (s)_+)\big(\exp(\phi(-s) - \phi(s)) - 1\big) \leq -\min(1 - \alpha, \alpha - \delta\overline{\gamma})|s| + (1/\delta).$$

Since $0 \leq \phi'(s) \leq \delta/2$,

$$(LV/V)(x, \theta) \leq -\min(1 - \alpha, \alpha - \overline{\gamma}\delta)\sum_i |\partial_i U| + d/\delta + \frac{\delta}{2}\sum_{i,j} |\partial_{ij} U|,$$

which is less than 1 outside a sufficiently large ball by our hypotheses. □

3.5. *Proofs of the main results.*

PROOF OF THEOREM 1. The steps of the proof are completely as depicted in Figure 4 and simply consist of combining Proposition 2, Theorem 4 and Theorem 5. □

PROOF OF THEOREM 2. By Lemma 11, there exists a Lyapunov function V such that for some $\varepsilon > 0$, $LV \leq -\varepsilon V$ outside a compact set, where L is the generator of the zigzag process; see Section 3.4. Since Growth Condition 3 implies Growth Condition 1, by Theorem 5, all compact sets are petite, and the process is ψ-irreducible and aperiodic, so that the conditions of Theorem 6 are satisfied, which establishes exponential ergodicity. □

PROOF OF THEOREM 3. By the growth condition, there exist $\alpha > 0$ such that $\alpha < \beta + \eta/4 < 1/2$ and $\delta > 0$ such that $0 < \delta < \alpha$ such that, for some $c > 0$, $g \leq cV$ with V given by (11). Furthermore, again by the growth condition, for x outside a bounded set, $V(x, \theta) \leq \exp((\beta + \eta/2)U(x))$. From the integrability assumption, $\pi(V^2) < \infty$. That all compact sets are petite follows from Theorem 5, whose conditions are satisfied by Theorem 4. The statement of the theorem then follows from Lemma 11 and [19], Theorem 4.3. □

Acknowledgements. We thank Tony Lelièvre, Paul Fearnhead and Eva Löcherbach for stimulating discussions, Pierre Monmarché for many exchanges on the merits of various Lyapunov functions, and Nikolas Nuesken and Julien Roussel for discussions on alternative approaches. We thank the Associate Editor and the anonymous referee for their comments which helped to correct and improve this manuscript.

REFERENCES

[1] ANDRIEU, C., DURMUS, A., NÜSKEN, N. and ROUSSEL, J. (2018). Hypercoercivity of piecewise deterministic Markov process-Monte Carlo. Available at https://arxiv.org/abs/1808.08592.

[2] AZAÏS, R., BARDET, J.-B., GÉNADOT, A., KRELL, N. and ZITT, P.-A. (2014). Piecewise deterministic Markov process—Recent results. In *Journées MAS* 2012. *ESAIM Proc*. **44** 276–290. EDP Sci., Les Ulis. MR3178622

[3] BAKHTIN, Y. and HURTH, T. (2012). Invariant densities for dynamical systems with random switching. *Nonlinearity* **25** 2937–2952. MR2979976

[4] BENAÏM, M., LE BORGNE, S., MALRIEU, F. and ZITT, P.-A. (2015). Qualitative properties of certain piecewise deterministic Markov processes. *Ann. Inst. Henri Poincaré Probab. Stat*. **51** 1040–1075. MR3365972

[5] BIERKENS, J. (2016). Non-reversible Metropolis–Hastings. *Stat. Comput*. **26** 1213–1228. MR3538633

[6] BIERKENS, J. and DUNCAN, A. (2017). Limit theorems for the zig-zag process. *Adv. in Appl. Probab*. **49** 791–825. MR3694318

[7] BIERKENS, J., FEARNHEAD, P. and ROBERTS, G. O. (2018). The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist*. To appear.

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

[9] 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

[10] BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2018). The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc*. **113** 855–867. MR3832232

[11] DAVIS, M. H. A. (1993). *Markov Models and Optimization. Monographs on Statistics and Applied Probability* **49**. CRC Press, London. MR1283589

[12] DELIGIANNIDIS, G., BOUCHARD-CÔTÉ, A. and DOUCET, A. (2017). Exponential ergodicity of the bouncy particle sampler. Preprint. Available at arXiv:1705.04579.

[13] DELIGIANNIDIS, G., PAULIN, D. and DOUCET, A. (2018). Randomized Hamiltonian Monte Carlo as scaling limit of the bouncy particle sampler and dimension-free convergence rates. Available at https://arxiv.org/abs/1808.04299.

[14] DIACONIS, P., HOLMES, S. and NEAL, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab*. **10** 726–752. MR1789978

[15] DOWN, D., MEYN, S. P. and TWEEDIE, R. L. (1995). Exponential and uniform ergodicity of Markov processes. *Ann. Probab*. **23** 1671–1691. MR1379163

[16] DUNCAN, A. B., LELIÈVRE, T. and PAVLIOTIS, G. A. (2016). Variance reduction using nonreversible Langevin samplers. *J. Stat. Phys*. **163** 457–491. MR3483241

[17] DURMUS, A., GUILLIN, A. and MONMARCHÉ, P. (2018). Geometric ergodicity of the bouncy particle sampler. Available at https://arxiv.org/abs/1807.05401.

[18] FETIQUE, N. (2017). Long-time behaviour of generalised zig-zag process. Available at https://arxiv.org/abs/1710.01087.

[19] GLYNN, P. W. and MEYN, S. P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931. MR1404536

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

[21] 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

[22] 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

[23] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR2466937

[24] MA, Y.-A., CHEN, T., WU, L. and FOX, E. B. (2016). A unifying framework for devising efficient and irreversible MCMC samplers. Available at https://arxiv.org/abs/1608.05973.

[25] MALRIEU, F. (2015). Some simple but challenging Markov processes. *Ann. Fac. Sci. Toulouse Math.* (6) **24** 857–883. MR3434260

[26] MARUYAMA, G. and TANAKA, H. (1959). Ergodic property of N-dimensional recurrent Markov processes. *Mem. Fac. Sci., Kyushu Univ., Ser. A* **13** 157–172. MR0112175

[27] 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.

[28] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253

[29] MEYN, S. P. and TWEEDIE, R. L. (1992). Stability of Markovian processes. I. Criteria for discrete-time chains. *Adv. in Appl. Probab.* **24** 542–574. MR1174380

[30] 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

[31] 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

[32] MICHEL, M., KAPFER, S. C. and KRAUTH, W. (2014). Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps. *J. Chem. Phys.* **140** 054116.

[33] MONMARCHÉ, P. (2016). Piecewise deterministic simulated annealing. *ALEA Lat. Am. J. Probab. Math. Stat.* **13** 357–398. MR3487077

[34] PAKMAN, A. (2017). Binary bouncy particle sampler. Available at https://arxiv.org/abs/1711.00922.

[35] PAKMAN, A., GILBOA, D., CARLSON, D. and PANINSKI, L. (2017). Stochastic bouncy particle sampler. In *Proc.* 34*th International Conference on Machine Learning* **70** 2741–2750.

[36] PETERS, E. A. J. F. and DE WITH, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E, Stat. Nonlinear Soft Matter Phys.* **85** 1–5.

[37] REY-BELLET, L. and SPILIOPOULOS, K. (2015). Irreversible Langevin samplers and variance reduction: A large deviations approach. *Nonlinearity* **28** 2081–2103. MR3366637

[38] ROBERTS, G. O. and ROSENTHAL, J. S. (1996). Quantitative bounds for convergence rates of continuous time Markov processes. *Electron. J. Probab.* **1** no. 9, approx. 21 pp. MR1423462

[39] SHERLOCK, C. and THIERY, A. H. (2017). A discrete bouncy particle sampler. Available at https://arxiv.org/abs/1707.05200.

[40] STRAMER, O. and TWEEDIE, R. L. (1999). Langevin-type models. I. Diffusions with given stationary distributions and their discretizations. *Methodol. Comput. Appl. Probab.* **1** 283–306. MR1730651

[41] TURITSYN, K. S., CHERTKOV, M. and VUCELJA, M. (2011). Irreversible Monte Carlo algorithms for efficient sampling. *Phys. D, Nonlinear Phenom.* **240** 410–414.
[42] TWEEDIE, R. L. (1994). Topological conditions enabling use of Harris methods in discrete and continuous time. *Acta Appl. Math.* **34** 175–188. MR1273853
[43] VANETTI, P., BOUCHARD-CÔTÉ, A., DELIGIANNIDIS, G. and DOUCET, A. (2017). Piecewise deterministic Markov chain Monte Carlo. Available at https://arxiv.org/abs/1707.05296.
[44] WU, C. and ROBERT, C. P. (2017). Generalized bouncy particle sampler. Available at https://arxiv.org/abs/1706.04781.

J. BIERKENS
DELFT INSTITUTE OF APPLIED MATHEMATICS
DELFT UNIVERSITY OF TECHNOLOGY
VAN MOURIK BROEKMANWEG 6
2628 XE DELFT
THE NETHERLANDS
E-MAIL: joris.bierkens@tudelft.nl

G. O. ROBERTS
DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
UNITED KINGDOM
E-MAIL: g.o.roberts@warwick.ac.uk

P.-A. ZITT
LABORATOIRE D'ANALYSE ET MATHÉMATIQUES APPLIQUÉES (UMR CNRS 8050)
UNIVERSITÉ-PARIS-EST-MARNE-LA-VALLÉE
5, BOULEVARD DESCARTES
CITÉ DESCARTES, CHAMPS-SUR-MARNE
77454 MARNE-LA-VALLÉE CEDEX 2
FRANCE
E-MAIL: pierre-andre.zitt@u-pem.fr