



## Kinetic walks for sampling

Pierre Monmarché

Sorbonne Université,  
Laboratoire Jacques-Louis Lions, LJLL,  
4 place Jussieu  
F-75005 Paris, France.

*E-mail address:* [pierre.monmarche@sorbonne-universite.fr](mailto:pierre.monmarche@sorbonne-universite.fr)

*URL:* <https://www.ljll.math.upmc.fr/~monmarche/>

**Abstract.** The persistent walk is a classical model in kinetic theory, which has also been studied as a toy model for Markov Chain Monte Carlo questions. Its continuous limit, the telegraph process, has recently been extended to various velocity jump processes (Bouncy Particle Sampler, Zig-Zag process, etc.) in order to sample general target distributions on  $\mathbb{R}^d$ . This paper studies, from a sampling point of view, general kinetic walks that are natural discrete-time (and possibly discrete-space) counterparts of these continuous-space processes. The main contributions of the paper are the definition and study of a discrete-space Zig-Zag sampler and the definition and time-discretisation of hybrid jump/diffusion kinetic samplers for multi-scale potentials on  $\mathbb{R}^d$ .

### 1. Introduction

The classical persistent walk on  $\mathbb{Z}$  is the Markov chain  $(X_k, V_k)_{k \in \mathbb{N}}$  on  $\mathbb{Z} \times \{-1, 1\}$  with transitions

$$(X_{k+1}, V_{k+1}) = \begin{cases} (X_k + V_k, V_k) & \text{with probability } 1 - \alpha \\ (X_k, -V_k) & \text{else,} \end{cases}$$

for some  $\alpha \in [0, 1]$ . It describes the constant-speed motion of a self-propelled particle,  $X_k$  denoting the position of the particle and  $V_k$  its velocity. Since the time between two changes of the velocity follows a geometric distribution with parameter  $\alpha$ ,  $(\alpha X_{\lfloor \alpha t \rfloor}, V_{\lfloor \alpha t \rfloor})_{t \geq 0}$  naturally converges as  $\alpha$  vanishes to the so-called telegraph process, for which the flips of the velocity are governed by a Poisson process (Herrmann and Vallois, 2010). From the seminal work of Goldstein (1951), these two

---

*Received by the editors May 23th, 2019; accepted April 4th, 2020.*

2010 *Mathematics Subject Classification.* 60J10, 65C40.

*Key words and phrases.* MCMC, non-reversible sampling, Zig-Zag process.

Research supported by ANR Project EFI (ANR-17-CE40-0030) of the French National Research Agency and ERC grants MSMATHS (Grant Agreement number 614492) and EMC2 (Grant Agreement number 810367).

processes, and various extensions, have been studied in details, in particular from the point of view of statistical physics and kinetic theory, or for other modelling motivations in physics, finance or biology (see for instance [Kac, 1974](#); [Zauderer, 1993](#); [Chen and Renshaw, 1994](#); [Hadeler, 1994](#); [Balakrishnan and Chaturvedi, 1988](#); [Rossetto, 2018](#); [Gruber and Schweizer, 2006](#) and references within).

Meanwhile, the search for efficient Markov Chain Monte Carlo (MCMC) methods led to the development of so-called rejection-free or lifted chains (see e.g. [Turitsyn et al., 2011](#); [Diaconis et al., 2000](#); [Bernard et al., 2009](#) and references within). In this context, the persistent walk has been a toy model to understand the efficiency of these algorithms, especially when compared to the reversible simple walk ([Diaconis et al., 2000](#); [Diaconis and Miclo, 2013](#); [Miclo and Monmarché, 2013](#)). For instance, correctly scaled, the persistent walk shows a ballistic behaviour, which means its expected distance to its initial position after  $K$  steps is of order  $K$ , while the simple walk shows a diffusive behaviour, moving to a distance  $\sqrt{K}$  after  $K$  steps. Since the efficiency of the MCMC schemes is related to the speed at which the space is explored, this is an argument in favour of non-reversible kinetic processes. The model being simple, it is even possible to determine the optimal  $\alpha$  (in the sense that it gives the maximal rate of convergence toward equilibrium on the periodic torus  $\mathbb{Z}/(N\mathbb{Z})$ ), which turns to be of order  $1/N$  ([Miclo and Monmarché, 2013](#)). This is consistent, as  $N$  goes to infinity, with the ballistic scaling that yields the telegraph process (by contrast, if  $\alpha$  is constant with  $N$  and if time is accelerated by  $N^2$ , the persistent walk converges to the Brownian motion).

Of course, both the persistent walk and the telegraph process sample the uniform measure in dimension one, which is not of practical interest. These last years, the telegraph process has been extended to several continuous-space processes, such as the Zig-Zag sampler ([Bierkens et al., 2019a](#); [Bierkens and Duncan, 2017](#); [Bierkens et al., 2019b](#); [Bierkens and Roberts, 2017](#)) or the Bouncy Particle Sampler ([Peters and de With, 2012](#); [Monmarché, 2016](#); [Durmus et al., 2018a](#); [Bouchard-Côté et al., 2018](#)), which are velocity jump processes designed to target any given distribution in any dimension. Many variants like randomized bounces ([Roberts, 2017](#); [Michel et al., 2017a](#)) are currently being developed and we refer to the review [Vanetti et al. \(2017\)](#) for more details, considerations and references on this vivid topic.

The present paper is concerned with similar extensions, but conducted at the level of the persistent walk rather than of the continuous kinetic process. Or, from another viewpoint, we are interested in persistent walks, but through the prism of MCMC sampling rather than kinetic theory. The motivations are the following: first, the discrete chain yields some insights on their continuous-time limits (for instance, we will see that the Zig-Zag process can be seen as the continuous limit of a Gibbs algorithm). Second, used in an MCMC scheme on  $\mathbb{Z}^d$ , a persistent walk shares, as will be detailed in this work, the following advantages with its continuous counterparts: thinning, factorization and ballistic behaviour. Finally, although continuous-time velocity jump processes can sometimes be sampled exactly thanks to thinning methods, it is not necessarily the case for mixed diffusion/jump kinetic samplers (see Section 5), in which case the corresponding chain obtained through an integration scheme (say, Euler scheme) is a discrete-time kinetic walk.

The rest of the paper is organized as follows. We start in Section 2 with the definition of an analogous on  $\mathbb{Z}$  of the Zig-Zag process on  $\mathbb{R}$  (or, equivalently, of the persistent walk but in a general potential landscape). The simplicity of the

chain allows an elementary study of its ergodicity, of its metastable behaviour at small temperature via an Eyring-Kramers formula and of its convergence toward the continuous Zig-Zag process on  $\mathbb{R}$  under proper scaling. Section 3 is a general and informal discussion about kinetic walks on  $\mathbb{R}^d$ , their simulation, invariant measures and continuous-time scaling limits. Finally, the last two sections present two particular applications, which are the main contributions of this work: the discrete Zig-Zag walk in a general potential in Section 4 and hybrid jump/diffusion kinetic samplers with a numerical integrator in Section 5. Although related by their motivation (the understanding of sampling with kinetic walks), the four sections are sufficiently independent to be read separately one from the other. The definition of the Zig-Zag walk associated to a general potential and all the results on this topic in Section 2 and Section 4 are new. The specific numerical scheme introduced in Section 5 is also new, although straightforwardly obtained from the well-known general method of Strang splittings.

**Notations.** If  $x, y \in \mathbb{R}^d$ , we denote  $x \cdot y$  their scalar product and  $|x| = \sqrt{x \cdot x}$ . The Dirac mass at  $x$  is denoted  $\delta_x$  and  $\mathbb{1}_A$  is 1 if  $A$  and 0 else. For  $r \in \mathbb{R}$ ,  $(r)_+ = \max(r, 0)$ . The set of  $\mathcal{C}^k$  functions on  $\mathbb{R}^d$  with compactly supported supported is denoted  $\mathcal{C}_c^k(\mathbb{R}^d)$ . The Gaussian distribution on  $\mathbb{R}^d$  with mean  $m$  and variance  $\Sigma^2$  is denoted  $\mathcal{N}(m, \Sigma^2)$ . We denote respectively  $\mathcal{P}(E)$ ,  $\mathcal{M}(E)$  and  $\mathcal{M}_b(E)$  the sets of probability measures, measurable functions and bounded measurable functions on a measurable space  $E$ , and for  $\mu \in \mathcal{P}(E)$  and  $f \in L^1(\mu)$  we write  $\mu f = \mu(f) = \int f d\mu$ . When  $(X_t^\varepsilon)_{t \geq 0}$  for  $\varepsilon > 0$  and  $(Y_t)_{t \geq 0}$  are càdlàg processes on  $\mathbb{R}^d$ , we write  $(X_t^\varepsilon)_{t \geq 0} \xrightarrow[\varepsilon \rightarrow 0]{law} (Y_t)_{t \geq 0}$  for the convergence in law in the Skorohod topology. Recall a sequence  $(x_n)_{n \in \mathbb{N}}$  of càdlàg functions from  $\mathbb{R}_+$  to  $\mathbb{R}^d$  is said to converge to  $x$  if, on all finite time interval, it converges uniformly up to a uniformly small change of time, i.e. if there exists a sequence  $(\gamma_n)_{n \in \mathbb{N}}$  with  $\gamma_n : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  increasing so that  $\sup_{t \in [0, T]} (|x_n(\gamma_n(t)) - x(t)| + |\gamma_n(t) - t|) \rightarrow 0$  as  $n \rightarrow \infty$  for all  $T > 0$ .

## 2. The Zig-Zag walk on $\mathbb{Z}$

Let  $U : \mathbb{Z} \rightarrow \mathbb{R}$  be such that  $\mathcal{Z} = \sum_{x \in \mathbb{Z}} \exp(-U(x)) < +\infty$ ,  $\pi(x) = \exp(-U(x))/\mathcal{Z}$  be the associated Gibbs distribution and  $\mu(x, v) = \pi(x)/2$  for  $v \in \{-1, 1\}$  and  $x \in \mathbb{Z}$ . We consider the Markov chain  $(X_k, V_k)_{k \in \mathbb{N}}$  on  $\mathbb{Z} \times \{-1, 1\}$  with transitions

$$(X_{k+1}, V_{k+1}) = \begin{cases} (X_k + V_k, V_k) & \text{with probability } \min\left(\frac{\pi(X_k + V_k)}{\pi(X_k)}, 1\right) \\ (X_k, -V_k) & \text{else,} \end{cases}$$

which we call the Zig-Zag walk on  $\mathbb{Z}$ . This transition can be seen as the composition of two Markov transitions. Indeed, consider on  $\mathbb{Z} \times \{-1, 1\}$  the Markov kernel given by  $x, v \mapsto \delta_{(x+v, -v)}$ . Since  $(y, w) = (x + v, -v)$  implies that  $(x, v) = (y + w, -w)$ , this kernel is symmetric. If a Metropolis accept/reject step with target measure  $\mu$  is added, the transition of the resulting chain is simply

$$(Y_{k+1}, W_{k+1}) = \begin{cases} (Y_k + W_k, -W_k) & \text{with probability } \min\left(\frac{\pi(Y_k + W_k)}{\pi(Y_k)}, 1\right) \\ (Y_k, W_k) & \text{else.} \end{cases}$$

By construction of the Metropolis-Hastings algorithm, this transition leaves  $\mu$  invariant. Now if we compose this transition with the deterministic transition

$(Y_{k+1}, W_{k+1}) = (Y_k, -W_k)$ , which obviously leaves  $\mu$  invariant, we obtain the initial chain. We have thus obtained that  $\mu$  is invariant for the Zig-Zag walk. Note however that, although both intermediate transition kernels are reversible with respect to  $\mu$ , their composition is not. Indeed,  $\mathbb{P}((X_2, V_2) = (X_0, V_0)) = 0$  for all initial condition.

The chain is clearly irreducible, and it is periodic. Indeed,  $(-1)^{X_k} V_k = (-1)^{k+X_0} V_0$  for all  $k \in \mathbb{N}$ , so that, if  $X_k$  is even with  $V_k = 1$  or  $X_k$  is odd with  $V_k = -1$ , then  $X_{k+1}$  is odd with  $V_{k+1} = 1$  or  $X_k$  is even with  $V_{k+1} = -1$ . In particular, the period is even. If  $U$  admits a strict local minimum  $x_0$  then there is a path of length 2 with strictly positive probability from  $(x_0, 1)$  to itself (which is  $(x_0, 1) \rightarrow (x_0, -1) \rightarrow (x_0, 1)$ ), so that the period is exactly 2, but this may not be the case in general. For instance, with  $U(k) = \lfloor k/2 \rfloor$ , the reader can check that the period is 4.

In the following, we prove an ergodic Law of Large Number and a Central Limit Theorem (CLT) in Theorem 2.1 (in the unimodal case; other cases are treated in any dimension in Section 4.4.4), an Eyring-Kramers formula in Theorem 2.2 and the convergence toward the continuous Zig-Zag process in Theorem 2.3.

2.1. *Asymptotic results.* Although the chain is already quite simple, let us focus for now on the case where  $\pi$  is unimodal. In that case, and similarly to the continuous-time case (Bierkens and Duncan, 2017), ergodicity can be established through elementary considerations on renewal chains.

**Theorem 2.1.** *Suppose that  $U$  is decreasing on  $]-\infty, 0]$  and increasing on  $[0, +\infty[$ , and let  $f \in L^1(\mu)$ . Then, for all initial conditions, almost surely,*

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k, V_k) \xrightarrow{n \rightarrow +\infty} \mu(f).$$

Moreover, denoting  $g(x) = f(x, 1) + f(x, -1)$  and

$$F(x) = \frac{1}{2}g(x) + \mathbb{1}_{x \geq 1} \sum_{i=1}^{x-1} g(i) + \mathbb{1}_{x \leq -1} \sum_{i=x+1}^{-1} g(i),$$

suppose that  $M_f := \mathbb{E}_\pi(g(X)F(X)) < \infty$  and that  $\mu(f) = 0$ . Then

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} f(X_k, V_k) \xrightarrow[n \rightarrow +\infty]{law} \mathcal{N}(0, \sigma_f^2),$$

with some explicit variance  $\sigma_f^2 \leq 3M_f$ .

*Proof:* Consider first the case where  $(X_0, V_0) = (0, 1)$  and denote  $T_1 = \inf\{n \in \mathbb{N} : X_{n+1} = X_n\}$ ,  $T_2 = \inf\{n \in \mathbb{N} : X_{2T_1+n+2} = X_{2T_1+1+n}\}$ . The monotonicities of  $U$  implies that almost surely  $X_n$  increases for  $n \in [0, T_1]$ , decreases for  $n \in [T_1 + 1, 2T_1 + 1 + T_2]$  with  $X_{2T_1+1} = 0$ , and finally, denoting  $S_1 = 2(T_1 + T_2 + 1)$ , increases for  $n \in [2T_1 + 2 + T_2, S_1]$  with  $(X_{S_1}, V_{S_1}) = (0, 1)$  (cf. Fig. 2.1). Remark that

$$\mathbb{P}(T_1 \geq k) = \prod_{j=0}^{k-1} e^{U(j)-U(j+1)} = e^{U(0)-U(k)},$$

so that  $T_1 < \infty$  almost surely (since we assumed that  $\mathcal{Z} < \infty$ ,  $U$  necessarily goes to  $\infty$  at  $\infty$ ). The same goes for  $T_2$ , hence for  $S_1$ . By the strong Markov

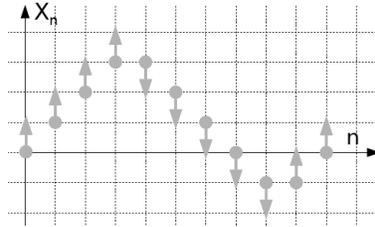


FIGURE 2.1. The trajectory between times  $S_0$  and  $S_1$ .

property,  $(X_n, V_n)_{n \geq S_1}$  has the same law as  $(X_n, V_n)_{n \in \mathbb{N}}$  and is independent from  $(X_n, V_n)_{n \in [0, S_1 - 1]}$ . Denote  $S_0 = 0$  and, for all  $n \in \mathbb{N}$ ,  $S_{n+1} = \inf\{k > S_n : (X_k, V_k) = (0, 1)\}$  and, given a function  $f \in L^1(\mu)$ ,

$$A_n = \sum_{k=S_n}^{S_{n+1}-1} f(X_k, V_k).$$

The  $A_n$ 's are i.i.d. and

$$\begin{aligned} \mathbb{E} \left( \left| \sum_{k=0}^{2T_1+1} f(X_k, V_k) \right| \right) &\leq \sum_{j \in \mathbb{N}} \sum_{k \leq j} \mathbb{P}(T_1 = j) (|f(k, 1)| + |f(k, -1)|) \\ &= \sum_{k \in \mathbb{N}} e^{U(0)-U(k)} (|f(k, 1)| + |f(k, -1)|) < +\infty. \end{aligned}$$

The sum for  $k \in [2T_1 + 2, S_1 - 1]$  is treated the same way, so that  $\mathbb{E}|A_0| < \infty$  and

$$\mathbb{E}(A_0) = \sum_{k \in \mathbb{Z}} e^{U(0)-U(k)} (f(k, 1) + f(k, -1)) = \lambda \mu(f)$$

with  $\lambda = 2e^{U(0)} \mathcal{Z}$ . The proof then follows from classical renewal arguments, which we recall for completeness. Considering the case  $f = 1$ , the law of large numbers implies that  $S_n/n$  converges almost surely toward  $\lambda$  as  $n$  goes to infinity. For  $n \in \mathbb{N}$  set  $K(n) = \sup\{k \in \mathbb{N} : S_k \leq n\}$ . If  $f$  is positive then for all  $n \in \mathbb{N}$ ,

$$\frac{1}{n} \sum_{j=0}^{K(n)} A_j \leq \frac{1}{n} \sum_{k=0}^{n-1} f(X_k, V_k) \leq \frac{1}{n} \sum_{j=0}^{K(n)+1} A_j.$$

Applied with  $f = 1$ , this reads

$$\frac{K(n)}{n} \times \frac{S_{K(n)}}{K(n)} \leq 1 \leq \frac{K(n)+1}{n} \times \frac{S_{K(n)+1}}{K(n)+1}.$$

Since  $K(n)$  almost surely goes to infinity with  $n$ , we get that  $K(n)/n$  almost surely converges to  $1/\lambda$ . Applied again with a general positive  $f$ , now,

$$\frac{K(n)}{n} \times \frac{1}{K(n)} \sum_{j=0}^{K(n)} A_j \leq \frac{1}{n} \sum_{k=0}^{n-1} f(X_k, V_k) \leq \frac{K(n)+1}{n} \times \frac{1}{K(n)+1} \sum_{j=0}^{K(n)+1} A_j,$$

and letting  $n$  go to infinity concludes. If  $f$  is not positive, the same conclusion follows from the decomposition  $f = (f)_+ - (-f)_+$ .

Now, consider the case of any general initial condition  $(X_0, V_0) = (x, v)$ , and let  $R = \inf\{n \in \mathbb{N} : (X_n, V_n) = (0, 1)\}$ . By similar arguments as above,  $R < \infty$  almost surely so that  $n^{-1} \sum_{k < R} f(X_k, V_k)$  almost surely goes to zero, while by the Markov property,  $n^{-1} \sum_{R \leq k \leq n} f(X_k, V_k)$  converges toward  $\mu(f)$ , which concludes.

The proof of the CLT is similar, and we refer to [Bierkens and Duncan \(2017, Lemma 4\)](#) to get that, if  $\mu(f) = 0$ ,

$$\sqrt{\frac{\lambda}{n}} \sum_{j=0}^{K(n)} A_j \xrightarrow[n \rightarrow \infty]{law} \mathcal{N}(0, \sigma_A^2),$$

provided that  $\sigma_A^2 := \mathbb{E}(A_0^2) < \infty$ . Now, even if  $(X_0, V_0) \neq (0, 1)$ , as before,

$$\frac{1}{\sqrt{n}} \left( \sum_{k=0}^{n-1} f(X_k, V_k) - \sum_{k=0}^{K(n)} A_k \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

hence, provided that  $\sigma_A^2 < \infty$ ,

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} f(X_k, V_k) \xrightarrow[n \rightarrow \infty]{law} \mathcal{N}(0, \sigma_A^2/\lambda).$$

Decompose  $A_0 = f(0, 1) + A'_0 + f(0, -1) + A''_0$  with

$$A'_0 = \sum_{k=1}^{2T_1} f(X_k, V_k), \quad A''_0 = \sum_{k=2T_1+2}^{S_1-1} f(X_k, V_k),$$

and remark that by the Markov property,  $A''_0$  is independent from  $A_0$ . Compute

$$\begin{aligned} \mathbb{E}((A'_0)^2) &= \sum_{k \in \mathbb{N}_*} \mathbb{P}(T_1 = k) \sum_{i=1}^k \sum_{j=1}^k g(i)g(j) \\ &= \sum_{i \in \mathbb{N}_*} \sum_{j \in \mathbb{N}_*} g(i)g(j) \mathbb{P}(T_1 \geq i \vee j) \\ &= \sum_{i \in \mathbb{N}_*} g(i) e^{U(0)-U(i)} \left( g(i) + 2 \sum_{j=1}^{i-1} g(j) \right) \\ &= \lambda \sum_{i \in \mathbb{N}_*} g(i) F(i) \pi(i). \end{aligned}$$

The case of  $A''_0$  is similar and, using that  $1/\lambda = \pi(0)/2$ , we get

$$\frac{1}{\lambda} \mathbb{E}(A_0^2) \leq \frac{3}{\lambda} \mathbb{E}((A'_0)^2 + (A''_0)^2 + g^2(0)) = 3 \sum_{x \in \mathbb{Z}} g(x) F(x) \pi(x),$$

which concludes. In fact  $\sigma_f^2 = \sigma_A^2/\lambda$  can be computed since  $\mathbb{E}(A'_0) = \lambda \sum_{x \in \mathbb{N}_*} g(x) \pi(x)$ , and similarly for  $A''_0$ .  $\square$

For  $N \in \mathbb{N}_*$  and  $t_1 < \dots < t_N$ , considering  $K(t_i) = \sup\{k \in \mathbb{N} : S_k \leq \lfloor nt_i \rfloor\}$  and decomposing  $\sum_{n=1}^{K(t_N)-1} A_n = \sum_{i=1}^N \sum_{n=K(t_i)}^{K(t_{i+1})-1} A_n$ , the previous elementary

proof is easily extended to obtain a functional CLT, namely the convergence

$$\left( \frac{1}{\sqrt{n}} \sum_{k=0}^{\lfloor nt \rfloor} (f(X_k, V_k) - \mu(f)) \right)_{t \geq 0} \xrightarrow[n \rightarrow +\infty]{law} (\sigma_f B_t)_{t \geq 0}$$

where  $(B_t)_{t \geq 0}$  is a one-dimensional Brownian motion. See also Section 4.4.4.

**2.2. Metastability.** We now consider the question of escape times from local minima at low temperature, as in [Monmarché \(2016\)](#) for the Zig-Zag process on  $\mathbb{R}$ . Recall that a random variable  $G$  on  $\mathbb{R}$  is said to be stochastically larger than a random variable  $F$  on  $\mathbb{R}$  if  $\mathbb{P}(G < t) \leq \mathbb{P}(F < t)$  for all  $t \in \mathbb{R}$ . In that case we write  $F \overset{sto}{\leq} G$ .

**Theorem 2.2.** *Let  $U : \mathbb{Z} \rightarrow \mathbb{R}$  and, for all  $\varepsilon > 0$ , let  $(X_k^\varepsilon, Y_k^\varepsilon)_{k \in \mathbb{N}}$  be the persistent walk on  $\mathbb{Z}$  associated to  $U/\varepsilon$  and with initial condition  $(0, 1)$ . Suppose that  $U(0) = 0$  and that  $U$  is decreasing on  $]]-\infty, 0]$  and increasing on  $[[0, +\infty[$ . Let  $a < \alpha \leq 0 \leq \beta < b$  be such that  $[[\alpha, \beta] = \{k \in \mathbb{Z}, U(k) = 0\}$ , and let*

$$\tau_\varepsilon = \inf\{n \in \mathbb{N}, X_n^\varepsilon \notin ]a, b[ \}.$$

Then

$$\begin{aligned} \mathbb{E}(\tau_\varepsilon) &= e^{E_1/\varepsilon} \left( \frac{2(\beta - \alpha + 1)}{1 + \mathbb{1}_{U(a)=U(b)}} + \mathcal{O}_{\varepsilon \rightarrow 0} \left( e^{-E_2/\varepsilon} \right) + \mathbb{1}_{U(a) \neq U(b)} \mathcal{O}_{\varepsilon \rightarrow 0} \left( e^{-E_3/\varepsilon} \right) \right) \end{aligned} \tag{2.1}$$

with  $E_1 = \min(U(a), U(b))$ ,  $E_2 = \min(U(\alpha - 1), U(\beta + 1))$  and  $E_3 = |U(a) - U(b)|$ . Moreover,  $\tau_\varepsilon/\mathbb{E}(\tau_\varepsilon)$  converges in law as  $\varepsilon$  vanishes to an exponential random variable with parameter 1, and

$$\mathbb{P}(X_{\tau_\varepsilon}^\varepsilon = a) \xrightarrow{\varepsilon \rightarrow 0} \frac{1}{2} (1 + \mathbb{1}_{U(a) \leq U(b)} - \mathbb{1}_{U(b) \leq U(a)}).$$

Finally, for all  $\varepsilon > 0$ ,

$$2(\beta - \alpha + 1)(G_\varepsilon - 1) \overset{sto}{\leq} \tau_\varepsilon \overset{sto}{\leq} 2(b - a + 1)G_\varepsilon$$

where  $G_\varepsilon$  is a geometric random variable with parameter

$$e^{-U(b)/\varepsilon} + e^{-U(a)/\varepsilon} - e^{-(U(b)+U(a))/\varepsilon}.$$

*Proof:* The proof is similar to the one appearing in [Monmarché \(2016\)](#). To alleviate notations, we only write  $X_k, V_k$  and  $\tau$  for  $X_k^\varepsilon, V_k^\varepsilon$  and  $\tau_\varepsilon$ . Like in the previous proof, set  $S_0 = 0$  and, by induction,  $S_{n+1} = \inf\{k > S_n, (X_k, V_k) = (0, 1)\}$ . For all  $n \in \mathbb{N}$ , let  $\tilde{S}_n = \inf\{k > S_n, (X_k, V_k) = (0, -1)\}$ , and let  $K = \inf\{n \in \mathbb{N}, S_n > \tau\}$ . Keep [Figure 2.1](#) in mind. By the strong Markov property,  $K$  follows a geometric distribution with parameter

$$\begin{aligned} p &:= \mathbb{P}(\tau < S_1) \\ &= \mathbb{P}(\tau < \tilde{S}_0) + \mathbb{P}(\tau > \tilde{S}_0) \mathbb{P}(\tau < S_1 | \tau > \tilde{S}_0) \\ &= e^{-U(b)/\varepsilon} + \left(1 - e^{-U(b)/\varepsilon}\right) e^{-U(a)/\varepsilon} \\ &= e^{-E_1/\varepsilon} \left(1 + \mathbb{1}_{U(a)=U(b)} + \mathcal{O}_{\varepsilon \rightarrow 0} \left(e^{-E_1/\varepsilon}\right) + \mathbb{1}_{U(a) \neq U(b)} \mathcal{O}_{\varepsilon \rightarrow 0} \left(e^{-E_3/\varepsilon}\right)\right), \end{aligned}$$

and

$$\mathbb{P}(X_\tau = b) = \mathbb{P}(X_\tau = b \mid \tau < S_1) = \frac{1}{p} e^{-U(b)/\varepsilon},$$

which indeed converges as  $\varepsilon$  vanishes to 0 if  $U(a) < U(b)$ , 1 if  $U(a) > U(b)$  and  $1/2$  if  $U(a) = U(b)$ . Decomposing  $\tau = \tau - S_{K-1} + \sum_{i=1}^{K-1} (S_i - S_{i-1})$ , remark that almost surely  $2(\beta - \alpha + 1) \leq S_i - S_{i-1} \leq 2(b - a + 1)$  for all  $i < K$  and  $\tau - S_{K-1} < 2(b - a + 1)$ , which proves the last claim of the theorem. Besides, again by the strong Markov property, conditionally to  $K$ ,  $(S_i - S_{i-1})_{i \in [1, K-1]}$  are i.i.d. random variables independent from  $K$ , so that

$$\mathbb{E}(\tau) = \mathbb{E}(\tau - S_{K-1}) + \mathbb{E}(K)\mathbb{E}(S_1 \mid \tau > S_1). \tag{2.2}$$

Since  $0 \leq \tau - S_{K-1} \leq 2(b - a)$  almost surely,

$$|\mathbb{E}(\tau - S_{K-1})| \leq 2(b - a) = e^{E_1/\varepsilon} \mathcal{O}_{\varepsilon \rightarrow 0} \left( e^{-E_2/\varepsilon} \right), \tag{2.3}$$

where we used that  $E_1 \geq E_2$ . For  $\varepsilon$  small, the most likely trajectory of the process between times 0 and  $S_1$  is the following: starting from  $(0, 1)$ , it deterministically goes to  $(\beta, 1)$ , then jumps to  $(\beta, -1)$  with high probability, then deterministically goes to  $(\alpha, -1)$  and jumps to  $(\alpha, 1)$  with high probability before going back to  $(0, 1)$  deterministically. More precisely,  $S_1 \geq 2(\beta - \alpha + 1)$  almost surely, and

$$\begin{aligned} \mathbb{P}(S_1 = 2(\beta - \alpha + 1)) &= \left( 1 - e^{(U(\alpha) - U(\alpha - 1))/\varepsilon} \right) \left( 1 - e^{(U(\beta) - U(\beta + 1))/\varepsilon} \right) \\ &= 1 + \mathcal{O}_{\varepsilon \rightarrow 0} \left( e^{-E_2/\varepsilon} \right). \end{aligned}$$

On the other hand, conditionally to  $\tau > S_1$ , almost surely,  $S_1 \leq 2(b - a)$ , so that

$$\mathbb{E}(S_1 \mathbb{1}_{\tau > S_1 > 2(\beta - \alpha + 1)}) \leq 2(b - a) \mathbb{P}(S_1 \neq 2(\beta - \alpha + 1)) = \mathcal{O}_{\varepsilon \rightarrow 0} \left( e^{-E_2/\varepsilon} \right).$$

Thus, we get that

$$\mathbb{E}(S_1 \mid \tau > S_1) = \frac{1}{1 - p} \mathbb{E}(S_1 \mathbb{1}_{\tau > S_1}) = 2(\beta - \alpha + 1) + \mathcal{O}_{\varepsilon \rightarrow 0} \left( e^{-E_2/\varepsilon} \right),$$

where we used again that  $E_1 \geq E_2$ . Using in (2.2) this estimate together with (2.3) and the fact that  $\mathbb{E}(K) = 1/p$  concludes the proof of the Eyring-Kramers formula (2.1).

Finally,  $K$  being an exponential random variable whose parameter vanishes with  $\varepsilon$ ,  $pK$  converges in law toward an exponential random variable with parameter 1. By the Markov inequality, for any  $\delta > 0$ ,

$$\begin{aligned} \mathbb{P} \left( \left| \frac{1}{K} \sum_{i=1}^{K-1} (S_i - S_{i-1}) - \mathbb{E}(S_1 \mid \tau > S_1) \right| > \delta \right) &\leq \frac{1}{\delta^2} \mathbb{E}(S_1^2 \mid \tau > S_1) \mathbb{E} \left( \frac{1}{K} \right) \\ &\leq \frac{4(b - a)^2}{\delta^2} \mathbb{E} \left( \frac{1}{K} \right) \xrightarrow{\varepsilon \rightarrow 0} 0. \end{aligned}$$

Since  $\mathbb{E}(S_1 \mid \tau > S_1)$  and  $p\mathbb{E}(\tau)$  both converges toward  $2(\beta - \alpha + 1)$  as  $\varepsilon$  vanishes,

$$\frac{1}{Kp\mathbb{E}(\tau)} \sum_{i=1}^{K-1} (S_i - S_{i-1}) \xrightarrow{\varepsilon \rightarrow 0} 1.$$

From Slutsky's theorem,  $\sum_{i=1}^{K-1} (S_i - S_{i-1})/\mathbb{E}(\tau)$  converges in law to an exponential random variable with parameter 1 as  $\varepsilon \rightarrow 0$ . Finally,  $|(\tau - S_{K-1})/\mathbb{E}(\tau)| \leq 2(b - a)/\mathbb{E}(\tau)$  almost surely goes to zero, which concludes.  $\square$

The simulated annealing chain obtained by considering a non-constant temperature  $(\varepsilon_k)_{k \in \mathbb{N}}$  and a potential  $U$  with possibly several local minima could also be studied by similar arguments as in [Monmarché \(2016, Theorem 3.1\)](#) to get a necessary and sufficient condition on the cooling schedule for convergence in probability toward the global minima of  $U$ .

**2.3. Continuous scaling limit.** The (continuous-time) Zig-Zag process on  $\mathbb{R}$  associated to a potential  $H \in C^1(\mathbb{R})$  (also known as the (integrated) telegraph or run-and-tumble process) is the Markov process on  $\mathbb{R} \times \{-1, 1\}$  with generator

$$Lf(y, w) = w\partial_y f(y, w) + (wH'(x))_+ (f(y, -w) - f(y, w)) .$$

In other words, it is a piecewise deterministic Markov process that, starting from an initial condition  $(y, w)$ , follows the flow  $(Y_t, W_t) = (y + tw, w)$  up to a random time  $T$  with distribution  $\mathbb{P}(T > t) = \exp(-\int_0^t (wH'(y + sw))_+ ds)$ , at which point  $(Y_T, W_T) = (y + Tw, -w)$ , after which it follows again the deterministic flow up to a new random jump time, etc.

**Theorem 2.3.** *For  $H \in C^2(\mathbb{R})$  that goes to infinity at infinity, for all  $\varepsilon > 0$ , define  $U_\varepsilon : \mathbb{Z} \mapsto \mathbb{R}$  by  $U_\varepsilon(k) = H(\varepsilon k)$  for all  $k \in \mathbb{Z}$ . Let  $(X_k^\varepsilon, V_k^\varepsilon)_{k \in \mathbb{N}}$  be the persistent walk on  $\mathbb{Z}$  associated to  $U_\varepsilon$  and with some initial condition  $(x_0^\varepsilon, v_0)$ . Suppose that  $\varepsilon x_0^\varepsilon$  converges to some  $x_0^* \in \mathbb{R}$  as  $\varepsilon$  vanishes. Then,*

$$\left( \varepsilon X_{\lfloor t/\varepsilon \rfloor}^\varepsilon, V_{\lfloor t/\varepsilon \rfloor}^\varepsilon \right)_{t \geq 0} \xrightarrow[\varepsilon \rightarrow 0]{law} (Y_t, W_t)_{t \geq 0} ,$$

where  $(Y_t, W_t)_{t \geq 0}$  is a Zig-Zag process on  $\mathbb{R}$  associated to  $H$  and with  $(Y_0, W_0) = (x_0^*, v_0)$ .

*Proof:* Denote  $T_1^\varepsilon = \varepsilon \inf\{n \in \mathbb{N} : X_{n+1}^\varepsilon = X_n^\varepsilon\}$ . Its cumulative function is

$$F_{x_0^\varepsilon, v_0}^\varepsilon(t) := \mathbb{P}(T_1^\varepsilon \leq t) = 1 - \prod_{k=1}^{\lfloor t/\varepsilon \rfloor} \exp\left(-\left(U_\varepsilon(x_0^\varepsilon + kv_0) - U_\varepsilon(x_0^\varepsilon + (k-1)v_0)\right)_+\right) .$$

From

$$\sum_{k=1}^{\lfloor t/\varepsilon \rfloor} \left( U_\varepsilon(x_0^\varepsilon + kv_0) - U_\varepsilon(x_0^\varepsilon + (k-1)v_0) \right)_+ \xrightarrow[\varepsilon \rightarrow 0]{} \int_0^t (v_0 H'(x_0^* + sv_0))_+ ds ,$$

we get that  $T_1^\varepsilon$  converges in law as  $\varepsilon$  vanishes to a random variable  $T_1^0$  with cumulative function

$$F_{x_0^*, v_0}^0(t) = 1 - \exp\left(-\int_0^t (v_0 H'(x_0^* + sv_0))_+ ds\right) .$$

Remark that

$$\int_0^t (v_0 H'(x_0^* + sv_0))_+ ds \geq \int_0^t v_0 H'(x_0^* + sv_0) ds = H(x_0^* + tv_0) - H(x_0^*) \xrightarrow[t \rightarrow \infty]{} +\infty ,$$

so that  $T_1^0$  is almost surely finite, and similarly for  $T_1^\varepsilon$  for all  $\varepsilon > 0$ . In particular,

$$\left( \varepsilon X_{T_1^\varepsilon}^\varepsilon, V_{T_1^\varepsilon}^\varepsilon, T_1^\varepsilon \right) = \left( \varepsilon x_0 + T_1^\varepsilon, -v_0, T_1^\varepsilon \right) \xrightarrow[\varepsilon \rightarrow 0]{law} \left( x_0^* + T_1^*, -v_0, T_1^* \right) .$$

Let  $(A_j)_{j \in \mathbb{N}}$  be an i.i.d. sequence of random variables uniformly distributed over  $[0, 1]$ . For all  $\varepsilon \geq 0$ , set  $(Z_0^\varepsilon, R_0^\varepsilon, S_0^\varepsilon) = (\varepsilon x_0^\varepsilon, v_0, 0)$  (with, in the case where  $\varepsilon = 0$ ,

$Z_0^0 = x_0^*$ ). Suppose by induction that, for some  $n \in \mathbb{N}$ ,  $(Z_n^\varepsilon, R_n^\varepsilon, S_n^\varepsilon)$  has been defined for all  $\varepsilon \geq 0$  and is independent from  $(A_j)_{j \geq n}$ . Then, for all  $\varepsilon \geq 0$ , set

$$\begin{aligned} S_{n+1}^\varepsilon &= S_n^\varepsilon + \left(F_{Z_n^\varepsilon, R_n^\varepsilon}^\varepsilon\right)^{-1}(A_n) \\ Z_{n+1}^\varepsilon &= Z_n^\varepsilon + (S_{n+1}^\varepsilon - S_n^\varepsilon)R_n^\varepsilon \end{aligned}$$

and  $R_{n+1}^\varepsilon = -R_n^\varepsilon$ . Remark that, for all  $t \geq 0$ ,  $(x, v) \mapsto F_{x,v}^\varepsilon(t)$  is continuous, uniformly in  $\varepsilon$ . As a consequence, for all  $N \in \mathbb{N}$ , by the previous result, almost surely,

$$(Z_n^\varepsilon, R_n^\varepsilon, S_n^\varepsilon)_{n \in [0, N]} \xrightarrow{\varepsilon \rightarrow 0} (Z_n^0, R_n^0, S_n^0)_{n \in [0, N]}. \quad (2.4)$$

By construction, for all  $\varepsilon > 0$ ,

$$(Z_n^\varepsilon, R_n^\varepsilon, S_n^\varepsilon)_{n \in [0, N]} \stackrel{\text{law}}{=} (\varepsilon X_{T_n^\varepsilon/\varepsilon}^\varepsilon, V_{T_n^\varepsilon/\varepsilon}^\varepsilon, T_n^\varepsilon)_{n \in [0, N]}$$

with  $T_0^\varepsilon = 0$  and, by induction,  $T_{k+1}^\varepsilon = \varepsilon \inf\{k > T_k^\varepsilon : X_{k+1}^\varepsilon = X_k^\varepsilon\}$ , and similarly

$$(Z_n^0, R_n^0, S_n^0)_{n \in [0, N]} \stackrel{\text{law}}{=} (Y_{J_n}, W_{J_n}, J_n)_{n \in [0, N]}$$

with  $J_0 = 0$  and by induction  $J_{k+1} = \inf\{t > J_k, W_t = -W_{J_k}\}$ . At this point, we have thus proved that the skeleton chain of the persistent walk (namely the persistent walk observed at its jump times, and those jump times) converges in law toward the skeleton chain of the Zig-Zag process (namely the process observed at its jump times, and those jump times). The convergence of the full chain is then a consequence from the fact that the latter is a deterministic function of its skeleton chain, as we detail now.

Note that  $(S_n^0)_{n \in \mathbb{N}}$  has the same distribution as  $(J_n)_{n \geq 0}$ . Moreover, for any  $t \geq 0$  and for all  $s \in [0, t]$ ,  $|Y_s - x_0^*| \leq t$ , so that the jump rate of the Zig-Zag process is bounded for times  $s \in [0, t]$  by  $\omega(t) = \sup_{x \in [x_0^* - t, x_0^* + t]} |H'(x)|$ , which is finite. In particular  $\sup\{n \in \mathbb{N}, J_n < t\}$  the number of jumps of the Zig-Zag process on  $[0, t]$  is stochastically smaller than a Poisson process with rate  $\omega(t)$ , hence is almost surely finite. As a consequence, almost surely  $S_n^0 \rightarrow +\infty$  as  $n \rightarrow \infty$ , and similarly for  $S_n^\varepsilon$  for all  $\varepsilon \geq 0$ .

Now the continuous-time processes are obtained by interpolating the skeleton chains. For all  $\varepsilon \geq 0$  and  $n \in \mathbb{N}$ , set  $\tilde{R}_t^\varepsilon = R_n^\varepsilon$  for all  $t \in [S_n^\varepsilon, S_{n+1}^\varepsilon[$ . For all  $t \geq 0$ , set  $\tilde{Z}_t^0 = x_0^* + \int_0^t \tilde{R}_s^0 ds$  and remark that, by construction,  $\tilde{Z}_{T_n^0}^0 = Z_n^0$  for all  $n \in \mathbb{N}$ . Finally, for all  $\varepsilon > 0$ , all  $n \in \mathbb{N}$  and all  $k \in [T_n/\varepsilon, T_{n+1}/\varepsilon]$ , set

$$\tilde{Z}_{k\varepsilon}^\varepsilon = Z_n^\varepsilon + \frac{k\varepsilon - T_n^\varepsilon}{T_{n+1}^\varepsilon - T_n^\varepsilon} (Z_{n+1}^\varepsilon - Z_n^\varepsilon)$$

and  $\tilde{Z}_t^\varepsilon = \tilde{Z}_{k\varepsilon}^\varepsilon$  for all  $t \in [k\varepsilon, (k+1)\varepsilon[$ . This construction ensures that, for all  $\varepsilon > 0$ ,

$$\left(\tilde{Z}_t^\varepsilon, \tilde{R}_t^\varepsilon\right)_{t \geq 0} \stackrel{\text{law}}{=} \left(\varepsilon X_{[t/\varepsilon]}^\varepsilon, V_{[t/\varepsilon]}^\varepsilon\right)_{t \geq 0} \quad \text{and} \quad \left(\tilde{Z}_t^0, \tilde{R}_t^0\right)_{t \geq 0} \stackrel{\text{law}}{=} (Y_t, W_t)_{t \geq 0}.$$

For all  $\varepsilon > 0$ , consider the increasing continuous change of time  $\gamma_\varepsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  given by: for all  $n \in \mathbb{N}$ ,  $\gamma_\varepsilon(T_n^\varepsilon) = T_n^0$  and  $\gamma_\varepsilon$  is linear on  $[T_n^\varepsilon, T_{n+1}^\varepsilon]$ . In particular,  $\tilde{R}_{\gamma_\varepsilon(t)}^\varepsilon = R_t^0$  for all  $t \geq 0$  (they start at the same value and both change sign at all times  $(T_n^0)_{n \geq 1}$ ). Moreover, from the convergence of the skeleton chains,  $\sup_{t \in [0, T_n^0]} \left(|\gamma_\varepsilon(t) - t| + |\tilde{Z}_{\gamma_\varepsilon(t)}^\varepsilon - Z_t^0|\right)$  almost surely goes to 0 as  $\varepsilon$  vanishes. Together with the fact  $T_n^0$  almost surely goes to  $+\infty$ , this means that for all fixed

$T > 0$ ,  $\sup_{t \in [0, T]} \left( |\gamma_\varepsilon(t) - t| + |\tilde{Z}_{\gamma_\varepsilon(t)}^\varepsilon - Z_t^0| \right)$  almost surely goes to 0, which concludes.  $\square$

### 3. Kinetic walks on $\mathbb{R}^d$

In the following, we will be interested in a class of Markov chains  $(X_n, V_n)_{n \in \mathbb{N}}$  on  $\mathbb{R}^d \times \mathbb{R}^d$  for  $d \in \mathbb{N}_*$ , with transitions given by

$$V_1 \sim p(X_0, V_0; \cdot), \quad X_1 = X_0 + \frac{\delta}{2}(V_0 + V_1) \quad (3.1)$$

for some  $\delta > 0$  and some kernel  $p : (x, v) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto p(x, v; \cdot) \in \mathcal{P}(\mathbb{R}^d)$ . We call such a chain the kinetic walk on  $\mathbb{R}^d$  associated to  $p$  with timestep  $\delta$ . Up to a rescaling of the velocities, we can always consider that  $\delta = 1$ .

This definition is close to – but distinct from – the definition of second-order Markov chains on  $\mathbb{R}^d$  (sometimes also called correlated random walks like in [Gruber and Schweizer \(2006\)](#)). Indeed,  $(X_n, X_{n-1})_{n \in \mathbb{N}}$  is a Markov chain if and only if  $(X_n, X_n - X_{n-1})_{n \in \mathbb{N}}$  is, with a simple way to express the transition of one of these chains from the transition of the other. Denoting  $V_n = (X_n - X_{n-1})/\delta$  would yield  $X_1 = X_0 + \delta V_0$ . On the contrary, consider the chain defined in Section 2, which satisfies (3.1) with  $\delta = 1$ . For this chain,  $(X_n, X_{n-1})_{n \in \mathbb{N}}$  is not Markovian: if  $X_n = X_{n-1}$  is at a strict local minimum of the potential  $U$ , then that only means that the velocity  $V_n$  has changed between times  $n-1$  and  $n$ , but it could be from 1 to  $-1$  or the converse (which we could know by looking farther in the past trajectory, for instance with the fact that  $(-1)^{X_k} V_k = (-1)^{k+X_0} V_0$  for all  $k \in \mathbb{N}$ ), and this affects the law of  $X_{n+1}$ . Our present definition is only motivated by the fact it gives a simple and unified framework for the cases studied in Sections 2, 4 and 5. Second-order Markov or related chains (like the discrete-time bounce sampler of [Roberts, 2017](#)) may be studied with the same arguments (especially concerning their continuous-time scaling limits). We use the term *kinetic* rather than *persistent* in order to keep the latter for cases where the velocity is typically constant for some times, and this is not always the case for the different kinetic walks we will be interested in.

Note that discrete-space walks can be seen as particular cases of walks on  $\mathbb{R}^d$  as follows. Let  $(X_n, V_n)_{n \in \mathbb{N}}$  be a kinetic walk on  $\mathbb{Z}^d \times \mathbb{Z}^d$  with transitions given by (3.1) with  $\delta > 0$  and  $p : \mathbb{Z}^d \times \mathbb{Z}^d \rightarrow \mathcal{P}(\mathbb{Z}^d)$  and let  $\eta, \kappa > 0$ . Consider the chain  $(\tilde{X}_n, \tilde{V}_n)_{n \in \mathbb{N}}$  on  $\mathbb{R}^d \times \mathbb{R}^d$  with transitions given by

$$\mathbb{P}_{x_0, v_0} \left( (\tilde{X}_1, \tilde{V}_1) = (x_1, v_1) \right) = \delta_{\eta \lfloor \frac{x_0}{\eta} + \frac{\delta}{2\kappa} (v_0 + v_1) \rfloor} (x_1) p \left( \left\lfloor \frac{x_0}{\eta} \right\rfloor, \left\lfloor \frac{v_0}{\kappa} \right\rfloor; \frac{v_1}{\kappa} \right)$$

if  $v_1 \in \kappa \mathbb{Z}^d$  and zero else. In particular, whatever the initial condition,  $\tilde{X}_n \in \eta \mathbb{Z}^d$  and  $\tilde{V}_n \in \kappa \mathbb{Z}^d$  for all  $n \geq 1$ . If  $(\tilde{X}_0, \tilde{V}_0) = (\eta X_0, \kappa V_0)$ , then  $(\eta X_n, \kappa V_n)_{n \in \mathbb{N}}$  and  $(\tilde{X}_n, \tilde{V}_n)_{n \in \mathbb{N}}$  have the same law. For this reason, in the rest of this section, only kinetic walks on  $\mathbb{R}^d$  will be considered. See Section 4 for an example of kinetic walk on  $\mathbb{Z}^d$ .

This section is more concerned with a general and informal discussion than with rigorous results, the latter possibly requiring technical details that can be checked on explicit examples (see Section 4 in particular). In particular the results of this section (Proposition 3.1 and Theorem 3.2) are not new results.

3.1. *First examples.* **Example 1.** Let  $U \in \mathcal{C}^1(\mathbb{R}^d)$ . Then the Hamiltonian dynamics  $\partial_t(x_t, v_t) = (v_t, -\nabla U(x_t))$  can be discretized as

$$V_{(n+1)\delta} = V_{n\delta} - \delta \nabla U \left( X_{n\delta} + \frac{\delta}{2} V_{n\delta} \right), \quad X_{(n+1)\delta} = X_{n\delta} + \frac{\delta}{2} (V_{n\delta} + V_{(n+1)\delta})$$

for some time-step  $\delta > 0$ . This is a slight modification of the classical velocity Verlet integrator. It is a second-order scheme and, contrary to the basic Euler scheme, it is symplectic, like the Hamiltonian dynamics. From KAM theory and backward error analysis, it can be shown to conserve up to a high precision an approximate Hamiltonian, which ensures long-time stability, see [Hairer et al. \(2003\)](#) and in particular [Hairer et al. \(2003, Theorem 5.1\)](#) for long-time energy conservation.

**Example 2.** The Langevin diffusion (or sometimes underdamped Langevin diffusion)

$$dX_t = V_t dt, \quad dV_t = -(\nabla U(X_t) - \gamma V_t) dt + \sqrt{2\gamma} dB_t,$$

where  $\gamma > 0$  and  $(B_t)_{t \geq 0}$  is a standard Brownian motion on  $\mathbb{R}^d$ , can be approximated by similar second-order schemes (see [Leimkuhler and Matthews, 2013](#); [Bou-Rabee, 2014](#), references within and Section 5 for more details on this topic). For instance, the Ricci-Ciccotti scheme ([Ricci and Ciccotti, 2003](#)) reads

$$\begin{aligned} V_{(n+1)\delta} &= e^{-\gamma\delta} V_{n\delta} - (1 - e^{-\gamma\delta}) \nabla U \left( X_{n\delta} + \frac{\delta}{2} V_{n\delta} \right) + \sqrt{(1 - e^{-\gamma\delta})} G_n \\ X_{(n+1)\delta} &= X_{n\delta} + \frac{\delta}{2} (V_{n\delta} + V_{(n+1)\delta}), \end{aligned}$$

where  $(G_n)_{n \in \mathbb{N}}$  is an i.i.d. sequence with law  $\mathcal{N}(0, I_d)$ .

3.2. *Sampling by thinning.* The continuous-time thinning and superposition method for sampling inhomogeneous Poisson processes, hence piecewise-deterministic Markov processes, is detailed e.g. in [Lemaire et al. \(2018\)](#). See Section 5.2 for an example of application. This section is concerned with its discrete analogous, which is essentially a rejection method applied to Bernoulli random variables (see also [Mandelbaum et al., 2007](#); [Michel et al., 2017b](#) on similar topics).

Suppose that the transition  $p$  can be decomposed as

$$p(x, v; \cdot) = q(x, v) p_1(x, v; \cdot) + (1 - q(x, v)) p_2(x, v; \cdot)$$

where, from a numerical point of view, computing  $q$  and sampling according to  $p_1$  is expensive, and sampling according to  $p_2$  is not (for instance,  $p_2(x, v; \cdot) = \delta_v(\cdot)$  for persistent chains). Suppose moreover that  $q(x, v) \leq \tilde{q}(x, v)$  where  $\tilde{q}$  is cheaper to compute than  $q$ . Then, for  $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ , a random variable  $V \sim p(x, v; \cdot)$  can be sampled as follows. Draw two independent random variables  $U_1$  and  $U_2$  uniformly distributed over  $[0, 1]$ . If  $U_1 \leq \tilde{q}(x, v)$  and  $U_2 \leq q(x, v)/\tilde{q}(x, v)$ , draw  $V$  according to  $p_1(x, v; \cdot)$  else draw  $V$  according to  $p_2(x, v; \cdot)$ . That way, obviously,  $V \sim p(x, v; \cdot)$ . The trick is that if  $U_1 > \tilde{q}(x, v)$  then we already know that  $V$  has to be drawn according to  $p_2(x, v; \cdot)$  and in that case we don't even have to compute  $q(x, v)$ . The smaller is  $\tilde{q}$ , the higher is the computational gain.

We can go a bit further in two cases for which the first step  $K$  where  $q$  has to be computed, i.e. where  $U_1 < \tilde{q}(X_K, V_K)$ , can be computed more efficiently than with Bernoulli variables at each step.

- If  $\tilde{q}(x, v) = \tilde{q} \in (0, 1)$  is constant. In that case,  $K+1$  follows a geometric law with parameter  $\tilde{q}$  and can be sampled through the inverse transformation method i.e. by setting  $K = \lfloor \ln U / \ln \tilde{q} \rfloor$  with  $U$  uniformly distributed over  $[0, 1]$  (which is particularly more efficient than with Bernoulli variables when  $\tilde{q}$  is small).
- If  $p_2(x, v; \cdot) = \delta_{f(x, v)}(\cdot)$  is deterministic, for some  $f : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  (typically, for a persistent walk,  $f(x, v) = v$ ). In that case,  $K$  follows the distribution

$$\mathbb{P}(K > n) = \prod_{k=0}^n (1 - \tilde{q}(\varphi_k(x, v)))$$

where  $\varphi_0(x, v) = (x, v)$ ,  $\varphi_1(x, v) = (x + \delta(v + f(x, v))/2, f(x, v))$  and  $\varphi_{k+1} = \varphi_k \circ \varphi_1$  for all  $k \in \mathbb{N}$ . In particular cases, depending on  $f$  and  $\tilde{q}$ , this distribution may again be sampled through the inverse transformation method.

In both cases, the algorithm is thus the following: draw  $K$  as above and an independent  $U_2$  uniformly distributed over  $[0, 1]$ . Sample  $(X_n, V_n)_{n \in \llbracket 0, K \rrbracket}$  as a kinetic chain associated to the transition  $p_2$ . If  $U_2 \leq q(X_K, V_K)/\tilde{q}(X_K, V_K)$ , draw  $V_{K+1}$  according to  $p_1(X_K, V_K; \cdot)$ , else according to  $p_2(X_K, V_K; \cdot)$ , and in both cases set  $X_{K+1} = X_K + \delta(V_K + V_{K+1})/2$ . Then, draw a new  $K'$  in a similar way as  $K$ , etc.

In the general case, of course each of the kernels  $p_1$  and  $p_2$  may also be decomposed in a similar way as  $p$ , so that we end up for some  $N \geq 1$  with a decomposition  $p = \sum_{i=1}^N p_n q_n$  where for each  $n \in \llbracket 1, N \rrbracket$ ,  $p_n$  is a transition kernel and  $q_n \in [0, 1]$  with some  $\sum_{i=1}^N q_n = 1$ . Similarly, if  $\tilde{q}(x, y) \leq \hat{q}(x, y)$  with  $\hat{q}(x, v)$  cheaper to compute than  $\tilde{q}(x, v)$ , then we can sample a Bernoulli variable with parameter  $q(x, v)$  as the product of three Bernoulli variables with respective parameters  $\hat{q}(x, v)$ ,  $\tilde{q}(x, v)/\hat{q}(x, v)$  and  $q(x, v)/\tilde{q}(x, v)$ . In other words, in the general decomposition we can decompose each weight  $q_n$  as a product  $\prod_{k=1}^{r_n} q_{n,k}$  for some  $r_n \geq 1$  and  $q_{n,k} \in [0, 1]$  for each  $k \in \llbracket 1, r_n \rrbracket$ . At the end of the day we get a representation of the form

$$p(x, v; \cdot) = \sum_{n=1}^N p_n(x, v; \cdot) \prod_{k=1}^{r_n} q_{n,k}(x, v)$$

that we can use to sample according to  $p$  in such a way that the average cost of computation is minimized. See Sections 4 and 5 for examples and related questions, in particular the link with factorization for Metropolis acceptance probabilities in Section 4.3. See also [Monmarché et al. \(2020\)](#) for an application in molecular dynamics.

**3.3. Invariant measure.** For MCMC applications, usual continuous-time kinetic Markov processes are designed to sample according to a given probability measure of the form  $\mu(dx, dv) = \pi(dx) \otimes \nu(dv)$  on  $\mathbb{R}^d \times \mathbb{R}^d$ . The target is the position marginal  $\pi$  and the velocity marginal  $\nu$  can be chosen by the user, usual choices being Gaussian or uniform (over the sphere or a discrete set of velocities) distributions. By definition,  $\mu$  is invariant for the kinetic walk on  $\mathbb{R}^d$  associated to some kernel  $p$  and time-step  $\delta$  if

$$\int f \left( x + \frac{\delta}{2}(v + w), w \right) p(x, v; dw) \mu(dx, dv) = \int f(x, v) \mu(dx, dv)$$

for all  $f \in \mathcal{M}_b(\mathbb{R}^{2d})$ . Nevertheless, there are cases (such as Examples 1 and 2 of Section 3.1) where this condition is only approximately satisfied, typically  $\mu$  is invariant only for the continuous-time process that is approximated by the discrete walk. In that case, for a fixed  $\delta > 0$ , the kinetic walk, with timestep  $\delta$  and associated to some transition  $p_\delta$ , typically admits some invariant measure  $\mu_\delta$ , which is also an invariant measure for the continuous-time Markov chain with generator

$$L_\delta f(x, v) = \int \left( f \left( x + \frac{\delta}{2}(v + w), w \right) - f(x, v) \right) p_\delta(x, v; dw).$$

In the typical case where  $L_\delta$  converges, as  $\delta$  vanishes, to some  $L$  for which  $\mu$  is invariant, it is possible to obtain explicit error bounds between  $\mu_\delta$  and  $\mu$ . We recall here a general argument based on Stein's method (Mattingly et al., 2010, Section 6.2).

**Proposition 3.1.** *Let  $L$  and  $L_\delta$  be two Markov generators on some Polish space  $E$  and let  $\mu, \mu_\delta \in \mathcal{P}(E)$  be invariant measures of, respectively,  $L$  and  $L_\delta$ . Denote  $(P_t)_{t \geq 0}$  the semi-group associated to  $L$ . Suppose that there exist  $C, \rho, h_\delta > 0$  and two norms  $N_1$  and  $N_2$  on a subspace  $\tilde{\mathcal{M}}$  of  $M_b(E)$  such that the following holds: 1)  $\tilde{\mathcal{M}}$  is dense in  $(M_b(\mathbb{R}^{2d}), \|\cdot\|_\infty)$  and is contained by the domains of  $L$  and  $L_\delta$ ; 2) for all  $f \in \tilde{\mathcal{M}}$  and all  $t \geq 0$ ,  $N_1(P_t f - \mu f) \leq C e^{-\rho t} N_2(f)/(1 \vee \sqrt{t})$ ; 3) for all  $f \in \tilde{\mathcal{M}}$ ,  $\|(L - L_\delta)f\|_\infty \leq h_\delta N_1(f)$ . Then, denoting  $N_2(\mu - \mu_\delta) = \sup\{\mu(f) - \mu_\delta(f), N_2(f) \leq 1\}$ , it holds:*

$$N_2(\mu - \mu_\delta) \leq C(2/3 + 1/\rho)h_\delta.$$

*Proof:* Let  $f \in \tilde{\mathcal{M}}$  and  $g = \int_0^\infty P_t(f - \mu f)dt$ , which is well defined and satisfies  $N_1(g) \leq C(2/3 + 1/\rho)N_2(f)$  and solves the Poisson equation  $Lg = \mu f - f$ . Using that  $\int L_\delta g \mu_\delta = 0$  by invariance of  $\mu_\delta$ , we get that

$$\mu(f) - \mu_\delta(f) = \mu_\delta(Lg) = \mu_\delta((L - L_\delta)g) \leq h_\delta C(2/3 + 1/\rho)N_2(f).$$

□

Remark that, in Proposition 3.1,  $C$  and  $\rho$  only depend on the limit process  $L$ . For particular processes, their existence usually follows from regularization and ergodicity results for  $L$  (with  $N_1$  and  $N_2$ , typically,  $L^2(\mu)$ ,  $H^1(\mu)$  or  $V$ -norms associated to some Lyapunov function), see e.g. Section 4.4.4, Glynn and Meyn (1996); Mattingly et al. (2010) or, for velocity jump processes, Monmarché (2017, Section 3). Then, if we are given a family of generators  $L_\delta$  for all  $\delta > 0$  such that  $\|(L - L_\delta)f\|_\infty \leq h_\delta N_1(f)$  with  $h_\delta \rightarrow 0$  for  $\delta \rightarrow 0$ , we get a quantitative estimate on the convergence  $\mu_\delta \rightarrow \mu$  (remark that, though the uniqueness of the invariant measure is ensured for  $L$  from the geometric ergodicity assumption, we haven't assumed the uniqueness of the invariant measure for  $L_\delta$ ).

In fact, reiterating this argument, following the Talay-Tubaro method Talay and Tubaro (1990), an expansion of the bias  $\mu_\delta f - \mu f$  in term of powers of  $\delta$  can be computed and a Romberg extrapolation (or related methods) can be used to kill the first order terms, see Talay and Tubaro (1990, Section 2.3) and Lelièvre and Stoltz (2016, Section 3.3.4).

Proposition 3.1 can be shown to apply in Example 2 of Section 3.1, i.e. the Langevin dynamics, see for instance Leimkuhler et al. (2016). On the contrary, it does not apply to the Hamiltonian dynamics of Example 1. Indeed, in that case, the limit process admits many invariant measures (because of energy conservation) and thus  $C$  and  $\rho$  cannot exist.

With Proposition 3.1, we have seen that the convergence of the invariant measures is related to the convergence of the generators. Now, the latter is classically related to the convergence of the processes and we address this question in the next section. See Theorem 3.2 below for some considerations on the convergence of the generators in the specific case of kinetic walks.

**3.4. Scaling limits.** In this section, we consider for all  $\varepsilon \in (0, 1]$  a kinetic walk  $(X_n^\varepsilon, V_n^\varepsilon)_{n \in \mathbb{N}}$  on  $\mathbb{R}^d \times \mathbb{R}^d$  with timestep 1, kernel  $p_\varepsilon$  and initial condition  $(x_0^\varepsilon, v_0^\varepsilon)$ . We are interested in the possible convergence of this chain, possibly rescaled as  $\varepsilon$  vanishes, toward a continuous-time process. The regime for which  $(X_n^\varepsilon)_{n \in \mathbb{N}}$  converges toward an elliptic diffusion has been abundantly studied for second-order chains, see Gruber (2004, Section 5) and references therein, and for this reason we will mostly focus on the cases where the full system  $(X_n^\varepsilon, V_n^\varepsilon)_{n \in \mathbb{N}}$  converges toward a continuous-time kinetic process  $(Y_t, W_t)_{t \geq 0}$ , where kinetic means that  $Y_t = Y_0 + \int_0^t W_s ds$ .

To alleviate notations, unless otherwise specified, we drop the  $\varepsilon$  superscript in all the rest of the section and simply write  $(X_n, V_n)$ . We start with an informal discussion on the scaling in the simple case where the dynamics are homogeneous with respect to the space variable  $x$ . Since, in order to expect a continuous-time limit,  $X_n$  should be nearly constant over a large number of steps as  $\varepsilon$  goes to zero, this homogeneous case should be expected to describe the short time dynamics of the general case.

**3.4.1. The space homogeneous case.** Consider the case where  $p_\varepsilon(x, v; \cdot) = h_\varepsilon(v; \cdot)$  for some transition kernel  $h_\varepsilon : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . In that case,  $(V_n)_{n \in \mathbb{N}}$  is a Markov chain by itself. Since

$$X_n = X_0 + \sum_{k=1}^n V_k + \frac{V_0 - V_n}{2}, \quad (3.2)$$

then, provided that, say, the variance of the last term is bounded uniformly in  $n$ , the situation is essentially the same as the case of correlated random walks. In term of time/space scaling, different cases may be distinguished concerning the limit of  $(X_n)_{n \in \mathbb{N}}$ :

- If  $(V_k)_{k \in \mathbb{N}}$  is in fact an i.i.d. sequence – namely if  $v \mapsto h_\varepsilon(v, \cdot)$  is constant – then, up to a vanishing term,  $(X_n)_{n \in \mathbb{N}}$  is a simple random walk. If it admits a continuous-time limit, then the latter is necessarily a Levy process, and conversely any Levy process  $(L_t)_{t \geq 0}$  may be obtained as a scaling limit of such a walk, even if we restrict the question to walks on  $\mathbb{Z}^d$ : indeed, considering  $V_k = \lfloor (L_{(k+1)\varepsilon} - L_{k\varepsilon}) / \varepsilon \rfloor$ , then  $(\varepsilon X_{\lfloor t/\varepsilon \rfloor})_{t \geq 0} \rightarrow (L_t)_{t \geq 0}$ .

Let us consider two particular cases. First, suppose that there exist  $\eta : (0, 1] \rightarrow (0, +\infty)$  such that

$$\begin{aligned} \frac{\eta(\varepsilon)}{\varepsilon} \mathbb{E}(V_1) &\xrightarrow{\varepsilon \rightarrow 0} \mu \in \mathbb{R}^d \\ \frac{\eta^2(\varepsilon)}{\varepsilon} \text{Var}(V_1) &\xrightarrow{\varepsilon \rightarrow 0} \Sigma \in \mathcal{M}_{d \times d}^{sym \geq 0}(\mathbb{R}) \end{aligned}$$

(with possibly  $\mu = 0$  or  $\Sigma = 0$ ). Then, from Donsker’s Theorem, provided that  $\eta(\varepsilon)x_0^\varepsilon$  converges to some  $x_0^* \in \mathbb{R}^d$ , we get the drifted Brownian motion

$$(\eta(\varepsilon)X_{[t/\varepsilon]})_{t \geq 0} \xrightarrow[\varepsilon \rightarrow 0]{law} (x_0^* + t\mu + \Sigma^{\frac{1}{2}}B_t)_{t \geq 0},$$

where  $(B_t)_{t \geq 0}$  is a standard Brownian motion. Second, suppose that there exist  $\lambda > 0$ ,  $\mu \in \mathbb{R}^d$ ,  $\nu \in \mathcal{P}(\mathbb{R}^d)$ ,  $\eta : (0, 1] \rightarrow (0, +\infty)$  and  $\mu : (0, 1] \rightarrow \mathbb{R}^d$  such that

$$\begin{aligned} \mathbb{P}(V_1 = \mu(\varepsilon)) &= 1 - \lambda\varepsilon + o_{\varepsilon \rightarrow 0}(\varepsilon) \\ \mu(\varepsilon) &\xrightarrow[\varepsilon \rightarrow 0]{} \mu \\ \mathcal{L}aw(\eta(\varepsilon)V_1 \mid V_1 \neq \mu(\varepsilon)) &\xrightarrow[\varepsilon \rightarrow 0]{} \nu. \end{aligned}$$

Then the cardinality  $N_t^\varepsilon$  of  $\{n \in \mathbb{N} : \varepsilon n \leq t, V_k \neq \mu(\varepsilon)\}$  converges as  $\varepsilon$  vanishes to a Poisson process  $N_t$  with intensity  $\lambda$ . As a consequence, considering an i.i.d. sequence  $(W_k)_{k \in \mathbb{N}}$  with law  $\nu$  independent from  $(N_t^\varepsilon)_{t \geq 0}$  for all  $\varepsilon$  and from  $(N_t)_{t \geq 0}$ , provided that  $\eta(\varepsilon)x_0^\varepsilon$  converges to some  $x_0^* \in \mathbb{R}^d$ , we get the drifted compound Poisson process

$$(\eta(\varepsilon)X_{[t/\varepsilon]})_{t \geq 0} \xrightarrow[\varepsilon \rightarrow 0]{law} x_0^* + t\mu + \left( \sum_{n=0}^{N_t} W_k \right)_{t \geq 0}.$$

In the more general situation where  $(V_k)_{k \in \mathbb{N}}$  is not necessarily an i.i.d. sequence, if there exists a  $\kappa : (0, 1] \rightarrow (0, +\infty)$  such that  $(\kappa(\varepsilon)V_{[t/\varepsilon]})_{t \geq 0}$  converges in distribution toward a continuous-time Markov process  $(W_t)_{t \geq 0}$  then, denoting  $\eta(\varepsilon) = \varepsilon\kappa(\varepsilon)$ , (3.2) reads

$$\eta(\varepsilon)X_{[t/\varepsilon]} = \eta(\varepsilon)x_0^\varepsilon + \int_0^t \kappa(\varepsilon)V_{[s/\varepsilon]} ds + \frac{\varepsilon}{2} (\kappa(\varepsilon)V_{[t/\varepsilon]} - \kappa(\varepsilon)V_0).$$

Integration being continuous with respect to Skorohod convergence, provided that  $\eta(\varepsilon)x_0^\varepsilon \rightarrow x_0^*$  as  $\varepsilon$  vanishes, this yields

$$(\eta(\varepsilon)X_{[t/\varepsilon]}, \kappa(\varepsilon)V_{[t/\varepsilon]})_{t \geq 0} \xrightarrow[\varepsilon \rightarrow 0]{} \left( x_0^* + \int_0^t W_s ds, W_t \right)_{t \geq 0}.$$

Remark that, of course, the scaling factor for the space variable is fixed by the scaling factors of the time and velocity variables.

- For instance, if  $(V_k)_{k \in \mathbb{N}}$  is a random walk on  $\mathbb{Z}^d$ , then as seen before it can converge toward a drifted Brownian motion, in which case the scaling limit of  $(X_n, V_n)_{n \in \mathbb{N}}$  is the Langevin diffusion, i.e. the solution of the SDE

$$\begin{cases} dX_t &= V_t dt \\ dV_t &= \mu + \Sigma^{1/2} dB_t, \end{cases}$$

where  $(B_t)_{t \geq 0}$  is a standard Brownian motion on  $\mathbb{R}^d$ .

- Alternatively, if there exist  $\lambda > 0$ ,  $\nu \in \mathcal{P}(\mathbb{R}^d)$  and  $\kappa : (0, 1] \rightarrow (0, +\infty)$  such that

$$\begin{aligned} \mathbb{P}(V_{k+1} = V_k) &= 1 - \lambda\varepsilon + o_{\varepsilon \rightarrow 0}(\varepsilon) \\ \mathcal{L}aw(\kappa(\varepsilon)V_{k+1} \mid V_{k+1} \neq V_k) &\xrightarrow[\varepsilon \rightarrow 0]{} \nu, \end{aligned}$$

then  $(\kappa(\varepsilon)V_{\lfloor t/\varepsilon \rfloor})_{t \geq 0}$  converges as  $\varepsilon$  vanishes toward  $(Y_{N_t})_{t \geq 0}$ , where  $(Y_k)_{k \in \mathbb{N}}$  is an i.i.d. sequence with law  $\nu$  and  $(N_t)_{t \geq 0}$  is a Poisson process with intensity  $\lambda$ , independent from  $(Y_k)_{k \in \mathbb{N}}$ . In that case,  $(X_t, V_t)_{t \geq 0}$  is the velocity jump process associated to the linear Boltzmann (or BGK) equation (Bouin et al., 2017).

These different examples highlighted three cases: if there is no inertia, the velocity tends to mix fast and  $(X_n)_{n \in \mathbb{N}}$  is Markovian. If there is some inertia in the sense that the velocity tends to stay aligned from one step to the other but with possible small fluctuations, the limit is a kinetic diffusion. If the velocity is rigorously constant for large times, the chain converges toward a velocity jump process. Of course this is a non-exhaustive list: in general, drift, diffusion, Poisson or  $\alpha$ -stable jumps can all be present in the limit, either kinetic or not (see in particular Section 5). But with these three regimes we cover the cases of the processes classically used in MCMC sampling.

3.4.2. *The case of kinetic limits.* We keep the notations of the beginning of Section 3.4. In particular, for all  $\varepsilon \in (0, 1]$ ,  $(X_n^\varepsilon, V_n^\varepsilon)_{n \in \mathbb{N}}$  is a kinetic walk on  $\mathbb{R}^d \times \mathbb{R}^d$  with transition  $p_\varepsilon$  and time-step 1. Let  $\kappa : (0, 1] \rightarrow (0, +\infty)$  and, for all  $\varepsilon \in (0, 1]$ , set  $\eta(\varepsilon) = \varepsilon\kappa(\varepsilon)$ . For any fixed  $x \in \mathbb{R}^d$ , let  $(\tilde{V}_k^{\varepsilon, x})_{k \in \mathbb{N}}$  be the Markov chain on  $\mathbb{R}^d$  with transitions

$$\tilde{V}_{k+1}^{\varepsilon, x} / \kappa(\varepsilon) \sim p_\varepsilon(x / \eta(\varepsilon), \tilde{V}_k^{\varepsilon, x} / \kappa(\varepsilon); \cdot),$$

and let  $N_t^\varepsilon$  be a Poisson process with intensity  $1/\varepsilon$ . Denote  $\tilde{L}_\varepsilon$  the operator defined on  $\mathcal{M}_b(\mathbb{R}^{2d})$  by

$$\tilde{L}_\varepsilon f(x, v) = \frac{1}{\varepsilon} \int [f(x, \kappa(\varepsilon)w) - f(x, v)] p_\varepsilon(x / \eta(\varepsilon), v / \kappa(\varepsilon); dw),$$

which is the infinitesimal generator of the Feller process  $(\tilde{X}_t^x, \tilde{V}_{N_t^\varepsilon}^{\varepsilon, x})_{t \geq 0}$ , where we simply set  $\tilde{X}_t^x = x$  for all  $t \geq 0$  (hence the link with the space homogeneous case). A direct corollary of Kallenberg (1997, Theorem 17.28) is the following:

**Theorem 3.2.** *Suppose that there exists a Feller generator  $\tilde{L}$  on  $\mathbb{R}^{2d}$  with domain containing  $\mathcal{C}_c^2(\mathbb{R}^{2d})$  and such that  $\|\tilde{L}_\varepsilon f - \tilde{L}f\|_\infty \rightarrow 0$  as  $\varepsilon \rightarrow 0$  for all  $f \in \mathcal{C}_c^2(\mathbb{R}^{2d})$ . Define the operator  $L$  by*

$$Lf(x, v) = v \cdot \nabla_x f(x, v) + \tilde{L}f(x, v).$$

*Suppose that  $L$  is the infinitesimal generator of a Feller process  $(Y_t, W_t)_{t \geq 0}$  and that  $\mathcal{C}_c^2(\mathbb{R}^{2d})$  is a core of  $L$ . Suppose that for all compact set  $\mathcal{K} \subset \mathbb{R}^d$ ,*

$$\sup_{x \in \mathcal{K}} \sup_{v \in \mathbb{R}^d} \int |v - \kappa(\varepsilon)w|^2 p_\varepsilon \left( \frac{x}{\eta(\varepsilon)}, \frac{v}{\kappa(\varepsilon)}; dw \right) \xrightarrow{\varepsilon \rightarrow 0} 0. \tag{3.3}$$

*Finally, suppose that  $(\eta(\varepsilon)X_0^\varepsilon, \kappa(\varepsilon)V_0^\varepsilon)$  converges in law toward  $(Y_0, W_0)$  as  $\varepsilon$  vanishes. Then*

$$\left( \eta(\varepsilon)X_{\lfloor t/\varepsilon \rfloor}^\varepsilon, \kappa(\varepsilon)V_{\lfloor t/\varepsilon \rfloor}^\varepsilon \right)_{t \geq 0} \xrightarrow[\varepsilon \rightarrow 0]{law} (Y_t, W_t)_{t \geq 0}.$$

See some applications with Proposition 4.2, Section 5.2 or Monmarché et al. (2020).

*Proof:* Denoting  $(Y_t^\varepsilon, W_t^\varepsilon) = (\eta(\varepsilon)X_{N_t^\varepsilon}^\varepsilon, \kappa(\varepsilon)V_{N_t^\varepsilon}^\varepsilon)$ , the generator  $L_\varepsilon$  of  $(Y_t^\varepsilon, W_t^\varepsilon)_{t \geq 0}$  is defined on  $\mathcal{M}_b(\mathbb{R}^{2d})$  by

$$L_\varepsilon f(x, v) = \frac{1}{\varepsilon} \int [f(x + \varepsilon(v + w)/2, \kappa(\varepsilon)w) - f(x, v)] p_\varepsilon(x/\eta(\varepsilon), v/\kappa(\varepsilon); dw).$$

Note that  $L_\varepsilon f \in \mathcal{M}_b(\mathbb{R}^{2d})$  for all  $f \in \mathcal{M}_b(\mathbb{R}^{2d})$ , so that if  $(X_0^\varepsilon, V_0^\varepsilon) = (x, v)$ ,

$$\begin{aligned} \mathbb{E}(f(Y_t^\varepsilon, W_t^\varepsilon)) &= \sum_{k \in \mathbb{N}} \mathbb{P}(N_t^\varepsilon = k) \mathbb{E}(f(\eta(\varepsilon)X_k^\varepsilon, \kappa(\varepsilon)V_k^\varepsilon)) \\ &= f(x, v) + tL_\varepsilon f(x, v) + o_{t \rightarrow 0}(t), \end{aligned}$$

with a negligible term uniform in  $(x, v)$  (which will be the case of all the negligible terms in the rest of the proof). This means that  $\mathcal{M}_b(\mathbb{R}^{2d})$ , hence  $\mathcal{C}_c^2(\mathbb{R}^{2d})$ , is included in the strong domain of  $L_\varepsilon$  for all  $\varepsilon \in (0, 1]$ . From [Kallenberg \(1997, Theorem 17.28\)](#) and the assumption that  $\mathcal{C}_c^2(\mathbb{R}^{2d})$  is a core for  $L$ , it only remains to check that  $\|L_\varepsilon f - Lf\|_\infty$  vanishes with  $\varepsilon$  for all  $f \in \mathcal{C}_c^2(\mathbb{R}^{2d})$ . Now, indeed, for  $f \in \mathcal{C}_c^2(\mathbb{R}^{2d})$ ,

$$\begin{aligned} L_\varepsilon f(x, v) &= \tilde{L}_\varepsilon f(x, v) \\ &+ \frac{1}{\varepsilon} \int \left[ f\left(x + \frac{\varepsilon}{2}(v + \kappa(\varepsilon)w), \kappa(\varepsilon)w\right) - f(x, \kappa(\varepsilon)w) \right] p_\varepsilon\left(\frac{x}{\eta(\varepsilon)}, \frac{v}{\kappa(\varepsilon)}; dw\right) \\ &= \tilde{L}f(x, v) + \int \frac{v + \kappa(\varepsilon)w}{2} \cdot \nabla_x f(x, \kappa(\varepsilon)w) p_\varepsilon\left(\frac{x}{\eta(\varepsilon)}, \frac{v}{\kappa(\varepsilon)}; dw\right) + o_{t \rightarrow 0}(1). \end{aligned}$$

Considering a ball  $\mathcal{B} \subset \mathbb{R}^d$  of some radius  $R$  such that the support of  $f$  is included in  $\mathcal{B}^2$ , we bound

$$\begin{aligned} \left| \left( v - \frac{v + \kappa(\varepsilon)w}{2} \right) \cdot \nabla_x f(x, \kappa(\varepsilon)w) \right| &\leq \frac{1}{2} \|\nabla_x f\|_\infty |v - \kappa(\varepsilon)w| \mathbb{1}_{\mathcal{B}}(x) \mathbb{1}_{\mathcal{B}}(\kappa(\varepsilon)w) \\ |\nabla_x f(x, \kappa(\varepsilon)w) - \nabla_x f(x, v)| &\leq |v - \kappa(\varepsilon)w| \|\nabla^2 f\|_\infty \mathbb{1}_{\mathcal{B}}(x) (\mathbb{1}_{\mathcal{B}}(v) + \mathbb{1}_{\mathcal{B}}(\kappa(\varepsilon)w)) \\ |v| (\mathbb{1}_{\mathcal{B}}(v) + \mathbb{1}_{\mathcal{B}}(\kappa(\varepsilon)w)) &\leq 2R + |v - \kappa(\varepsilon)w|. \end{aligned}$$

As a consequence, for some  $C_f > 0$ , for all  $(x, v) \in \mathbb{R}^{2d}$ ,

$$\begin{aligned} |L_\varepsilon f(x, v) - Lf(x, v)| &\leq \\ \mathbb{1}_{\mathcal{B}}(x) C_f \int (|v - \kappa(\varepsilon)w| + |v - \kappa(\varepsilon)w|^2) p_\varepsilon\left(\frac{x}{\eta(\varepsilon)}, \frac{v}{\kappa(\varepsilon)}; dw\right) &+ o_{t \rightarrow 0}(1). \end{aligned}$$

Condition [\(3.3\)](#) concludes.  $\square$

#### 4. The discrete Zig-Zag walk

This section is devoted to the definition and study of a discrete-space analogous of the Zig-Zag process on  $\mathbb{R}^d$ .

**4.1. Definition.** Let  $d \in \mathbb{N}_*$ ,  $U : \mathbb{Z}^d \rightarrow \mathbb{R}$  be such that  $\mathcal{Z} = \sum_{x \in \mathbb{Z}^d} \exp(-U(x)) < +\infty$ ,  $\pi(x) = \exp(-U(x))/\mathcal{Z}$  be the associated Gibbs distribution and  $\mu(x, v) =$

$\pi(x)/2^d$  for  $v \in \{-1, 1\}^d$  and  $x \in \mathbb{Z}^d$ . For  $i \in \llbracket 1, d \rrbracket$ , denote  $e_i$  the  $i^{\text{th}}$  vector of the canonical basis of  $\mathbb{R}^d$  and let

$$\begin{aligned} q_i(x, v_i) &= \min\left(\frac{\pi(x + v_i e_i)}{\pi(x)}, 1\right) = e^{-(U(x + v_i e_i) - U(x))_+} \\ p_i(x, v_i; w_i) &= q_i(x, v_i)\delta_{v_i}(w_i) + (1 - q_i(x, v_i))\delta_{-v_i}(w_i) \\ p(x, v; w) &= \prod_{i=1}^d p_i\left(x + \sum_{j=1}^{i-1} \frac{v_j + w_j}{2} e_j, v_i, w_i\right). \end{aligned} \tag{4.1}$$

That way,  $p : \mathbb{Z}^d \times \{-1, 1\}^d \rightarrow \mathcal{P}(\{-1, 1\}^d)$ . We call Zig-Zag walk on  $\mathbb{Z}^d$  the kinetic walk  $(X_n, V_n)_{n \in \mathbb{N}}$  associated to this kernel  $p$  with timestep  $\delta = 1$ , i.e. the Markov chain on  $\mathbb{Z}^d \times \{-1, 1\}^d$  whose transitions are given by (3.1). Remark that, for  $d = 1$ , we retrieve the chain studied in Section 2.

A random variable  $V \sim p(x, v; \cdot)$  in  $\{-1, 1\}^d$  can be sampled as follows. Set  $Y_0 = x$ , and suppose by induction that  $Y_{k-1} \in \mathbb{Z}^d$  has been defined for some  $k \in \llbracket 1, d \rrbracket$ . Set  $V_k = v_k$  with probability  $q_k(Y_{k-1}, v_k)$  and  $V_k = -v_k$  else, and in either case set  $Y_k = Y_{k-1} + (v_k + V_k)/2$ . Then  $V$  is distributed according to  $p(x, v; \cdot)$ , and  $X := Y_d = x + (v + V)/2$ . In other words, this is a Gibbs algorithm based on the Zig-Zag walk on  $\mathbb{Z}$ : one step of the Zig-Zag walk in  $\mathbb{Z}^d$  is the result of  $d$  successive one-dimensional Zig-Zag steps on each coordinate, the others being fixed.

If we want the coordinates to play a symmetric role in the transition, for  $\sigma$  a permutation of  $\llbracket 1, d \rrbracket$  we can define  $p_\sigma(x, v; \cdot)$  to be the law of  $W_{\sigma^{-1}}$  when  $W \sim p(x_\sigma, v_\sigma; \cdot)$ , where  $u_\sigma$  for  $u \in \mathbb{Z}^d$  and  $\sigma \in \mathfrak{S}_d$  denotes  $(u_{\sigma(1)}, \dots, u_{\sigma(d)})$ . This accounts to use the order given by  $\sigma$  to update the coordinates. Then

$$p_{sym}(x, v; \cdot) = \frac{1}{d!} \sum_{\sigma \in \Sigma_d} p_\sigma(x, v; \cdot)$$

corresponds to a transition where the order is sampled at random at each step of the Zig-Zag walk. There is no particular practical interest to consider  $p_{sym}$  rather than  $p$ , and moreover any result on  $p$  can straightforwardly be adapted to  $p_\sigma$  by renumbering of the coordinates, and then to  $p_{sym}$ .

#### 4.2. Equilibrium and scaling limit.

**Proposition 4.1.** *The probability distribution  $\mu$  is invariant for the Zig-Zag walk on  $\mathbb{Z}^d$ .*

*Proof:* As proven in Section 2, for all fixed  $(x_j, v_j)_{j \in \llbracket 2, d \rrbracket}$ , the transition on  $\mathbb{Z} \times \{-1, 1\}$  defined by

$$W_{n+1} \sim p_1([Y_n, x_2, \dots, x_d], [V_n, v_2, \dots, v_d]; \cdot), \quad Y_{n+1} = \frac{W_n + W_{n+1}}{2}$$

admits the conditional law  $(y, w) \mapsto \pi(y, x_2, \dots, x_d)/2$  as an invariant measure. As a consequence, the transition of the Markov chain  $(\tilde{X}_n, \tilde{V}_n)$  on  $\mathbb{Z}^d \times \{-1, 1\}^d$  with  $(\tilde{X}_{n,1}, \tilde{V}_{n,1}) = (Y_n, W_n)$  and  $(\tilde{X}_{n,j}, \tilde{V}_{n,j}) = (\tilde{X}_{0,j}, \tilde{V}_{0,j})$  for  $j \neq 1$  also fixes  $\mu$ . Since the transition of the Zig-Zag walk is the composition of  $d$  such transitions, it fixes  $\mu$ .  $\square$

Remark that, if the target law is of a tensor form  $\pi(x) = \prod_{i=1}^d \pi_i(x_i)$ , then the coordinates of a Zig-Zag walk are just  $d$  independent one-dimensional Zig-Zag walks (which is similar to the continuous-time process).

Recall that the continuous-time Zig-Zag process on  $\mathbb{R}^d$  associated to a potential  $H$  is the Markov process on  $\mathbb{R}^d \times \{-1, 1\}^d$  with generator

$$Lf(x, v) = v \cdot \nabla_x f(x, v) + \sum_{i=1}^d (v_i \partial_{x_i} H(x))_+ (f(x, v_{-i}) - f(x, v)),$$

where we denote by  $v_{-i}$  the vector of  $\{-1, 1\}^d$  obtained from  $v$  by multiplying its  $i^{\text{th}}$  coordinate by  $-1$ . The following is the extension of Theorem 2.3 in larger dimension.

**Theorem 4.2.** *For  $H \in \mathcal{C}^2(\mathbb{R}^d)$  that goes to infinity at infinity, for all  $\varepsilon > 0$ , define  $U_\varepsilon : \mathbb{Z}^d \mapsto \mathbb{R}$  by  $U_\varepsilon(x) = H(\varepsilon x)$  for all  $x \in \mathbb{Z}^d$ . Let  $(X_k^\varepsilon, V_k^\varepsilon)_{k \in \mathbb{N}}$  be the Zig-Zag walk on  $\mathbb{Z}^d$  associated to  $U_\varepsilon$  and with some initial condition  $(x_0^\varepsilon, v_0)$ . Suppose that  $\varepsilon x_0^\varepsilon$  converges to some  $x_0^* \in \mathbb{R}^d$  as  $\varepsilon$  vanishes. Then*

$$\left( \varepsilon X_{\lfloor t/\varepsilon \rfloor}^\varepsilon, V_{\lfloor t/\varepsilon \rfloor}^\varepsilon \right)_{t \geq 0} \xrightarrow[\varepsilon \rightarrow 0]{\text{law}} (Y_t, W_t)_{t \geq 0},$$

where  $(Y_t, W_t)_{t \geq 0}$  is a Zig-Zag process on  $\mathbb{R}^d$  associated to  $H$  and with  $(Y_0, W_0) = (x_0^*, v_0)$ .

*Proof:* Let us show that Theorem 3.2, or rather directly Kallenberg (1997, Theorem 17.28), applies. First, following Durmus et al. (2018a), we can see that the continuous-time process can be smoothly and compactly approximated (in the sense of Durmus et al., 2018b, Definition 20) by replacing its continuous jump rates by  $\mathcal{C}^\infty$  jump rates (see the case of the BPS in Durmus et al., 2018b, Proposition 23 for details). From Durmus et al. (2018b, Theorem 21), this proves that  $\mathcal{C}_c^1(\mathbb{R}^d \times \{-1, 1\}^d)$  is a core for the strong generator  $L$  of the Zig-Zag process. Denote  $L_\varepsilon$  the generator of  $(\varepsilon X_{N_t}^\varepsilon, V_{N_t}^\varepsilon)_{k \in \mathbb{N}}$  where  $(N_t)_{t \geq 0}$  is a Poisson process with intensity  $1/\varepsilon$ . Then all bounded measurable functions  $f$  are in the domain of  $L_\varepsilon$  and

$$L_\varepsilon f(x, v) = \frac{1}{\varepsilon} \sum_{w \in \{-1, 1\}^d} \left( f\left(x + \varepsilon \frac{v+w}{2}, w\right) - f(x, v) \right) p_\varepsilon\left(\frac{x}{\varepsilon}, v; w\right),$$

where  $p_\varepsilon$  is given by (4.1) with  $U = U_\varepsilon$ . Next, for all  $i \in \llbracket 1, d \rrbracket$ ,

$$\exp\left(-\left(U(x + \varepsilon v_i e_i) - U(x)\right)_+\right) = 1 - \varepsilon (v_i \partial_{x_i} U(x))_+ + o_{\varepsilon \rightarrow 0}(\varepsilon),$$

where the negligible term is uniform over all compact set of  $\mathbb{R}^d \times \{-1, 1\}^d$  since  $H$  is  $\mathcal{C}^2$  (and this will be the case for all the negligible terms below). More generally,

$$\exp\left(-\left(U\left(x + \varepsilon \sum_{j=1}^i v_j e_j\right) - U\left(x + \varepsilon \sum_{j=1}^{i-1} v_j e_j\right)\right)_+\right) = 1 - \varepsilon (v_i \partial_{x_i} U(x))_+ + o_{\varepsilon \rightarrow 0}(\varepsilon),$$

from which

$$\begin{aligned} p_\varepsilon\left(\frac{x}{\varepsilon}, v; w\right) &= \prod_{i=1}^d \left( \delta_{v_i}(w_i) + \varepsilon (v_i \partial_{x_i} U(x))_+ (\delta_{-v_i}(w_i) - \delta_{v_i}(w_i)) + o_{\varepsilon \rightarrow 0}(\varepsilon) \right) \\ &= \delta_v(w) + \varepsilon \sum_{i=1}^d (v_i \partial_{x_i} U(x))_+ (\delta_{v-2v_i e_i}(w) - \delta_v(w)) + o_{\varepsilon \rightarrow 0}(\varepsilon). \end{aligned}$$

On the other hand, if  $f \in \mathcal{C}_c^2(\mathbb{R}^d \times \{-1, 1\}^d)$ ,

$$f\left(x + \varepsilon \frac{v+w}{2}, w\right) = f(x, w) + \varepsilon \frac{v+w}{2} \cdot \nabla_x f(x, w) + \underset{\varepsilon \rightarrow 0}{o}(\varepsilon),$$

so that

$$\|L_\varepsilon f - Lf\|_\infty \xrightarrow{\varepsilon \rightarrow 0} 0,$$

and [Kallenberg \(1997, Theorem 17.28\)](#) concludes.  $\square$

Note that the space/time scaling in [Theorem 4.2](#) is ballistic. It means in particular that, in  $n = \lfloor 1/\varepsilon \rfloor$  steps, the Zig-Zag walk with potential  $U_\varepsilon$  is at distance of order  $n$  (and not  $\sqrt{n}$  as in the diffusive case) from its starting point.

**4.3. Thinning and factorization.** In order to sample the Zig-Zag walk, a priori, at each time step,  $U(x)$  has to be computed for  $d+1$  values of  $x$ . However, thanks to the thinning method recalled in [Section 3.2](#), this computational cost may drop if simple bounds are known on the increments of  $U$ .

Moreover, the factorization principle used for the continuous-time Zig-Zag process in [Bierkens et al. \(2019a\)](#) to do subsampling is still available here. Suppose that we can decompose  $U(x+v) - U(x) = \sum_{j=1}^M f_j(x, v)$  for all  $(x, v) \in \mathbb{Z}^d \times \{-1, 1\}^d$  for some  $M \in \mathbb{N}_*$  and  $f_j : \mathbb{Z}^d \times \{-1, 1\}^d \rightarrow \mathbb{R}$ . This is for instance the case if  $U = \sum_{j=1}^M U_j$ , in which case we can take  $f_j(x, v) = U_j(x+v) - U_j(x)$ , but in general the  $f_j$ 's are not required to be discrete gradients. Consider the Zig-Zag walk as defined above except that the probability  $q_i(x, v_i)$  is replaced by

$$\tilde{q}_i(x, v_i) = \prod_{j=1}^M e^{-(f_j(x, v_i e_i))_+}.$$

**Proposition 4.3.** *This Zig-Zag walk with  $\tilde{q}_i$  still admits  $\mu$  as an invariant measure.*

*Proof:* As in the proof of [Proposition 4.1](#), we just have to prove the result for  $d=1$ . Following [Section 2](#), this stems from the same result applied to classical Metropolis-Hastings algorithms. Indeed, let  $q$  be a symmetric Markov kernel on a space  $E$  and let  $\alpha : E \times E \rightarrow [0, 1]$  be the acceptance probability of a Metropolis-Hastings chain with proposal  $q$ , namely a chain with transition kernel  $p(x, y) = q(x, y)\alpha(x, y)$  for  $y \neq x$ . Then this chain is reversible with respect to a probability  $\mu$  if and only if

$$\frac{\alpha(x, y)}{\alpha(y, x)} = \frac{\mu(x)}{\mu(y)} \quad \forall x, y \in E.$$

In particular, if  $\mu(x) = C \prod_{i=1}^M g_i(x)$  for some positive  $g_i$ 's and a normalization constant  $C$  then, setting  $\alpha_i(x, y) = \min(1, g_i(x)/g_i(y))$  for all  $i \in \llbracket 1, M \rrbracket$  and  $x, y \in E$  ensures that

$$\frac{\alpha_i(x, y)}{\alpha_i(y, x)} = \frac{g_i(x)}{g_i(y)} \quad \forall i \in \llbracket 1, M \rrbracket, x, y \in E.$$

Taking the product over  $i \in \llbracket 1, M \rrbracket$  and noting that  $\min(1, a/b) = \exp(-(\ln(a/b))_+)$  for  $a, b > 0$  concludes.  $\square$

The bad side of factorization is that it increases the number of rejections (for the Metropolis-Hastings algorithm, hence of collisions for the Zig-Zag process). On the other hand, if  $U(x+v) - U(x)$  can be decomposed in a part that is cheap to compute and a part that may be expensive to compute but is small and for

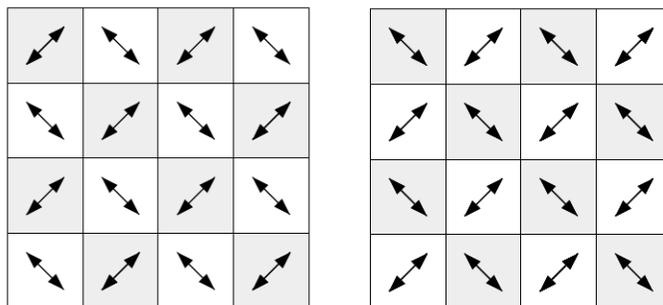


FIGURE 4.2. Representation of the two irreducible classes of the Zig-Zag walk in dimension 2. For instance, if we start in a grey box with velocity  $(1, 1)$ , then whenever we will be in a grey box the velocity will be either  $(1, 1)$  or  $(-1, -1)$  and whenever we will be in a white box the velocity will be  $(1, -1)$  or  $(-1, 1)$ . This corresponds to the array on the left.

which an efficient bound is available for thinning, then it may give a significant computational gain. This is particularly well-adapted for multi-scale potentials, as we can see in Section 5 on a similar problem.

#### 4.4. Irreducibility, Ergodicity, CLT.

4.4.1. *Irreducibility.* Irreducibility is a delicate question for the continuous Zig-Zag process, see Bierkens et al. (2019b). Here, for the discrete Zig-Zag walk, we will only tackle the restrictive case where, following the definitions of Bierkens et al. (2019b), all velocities are asymptotically flippable<sup>1</sup> and thus the proof is similar to the Gaussian case with dominant diagonal of Bierkens et al. (2019b, Corollary 1). Anyway we are interested in the exponentially fast convergence toward equilibrium under the assumption of Proposition 4.5 below, which is even stronger. Moreover, with this restriction, we can focus on the specificities of the discrete realm.

Indeed, let us call  $\sigma(x, v) := ((-1)^{x_i v_i})_{i \in \llbracket 1, d \rrbracket} \in \{-1, 1\}^d$  the signature of  $(x, v) \in \mathbb{Z}^d \times \{-1, 1\}^d$ . If  $(X_n, V_n)_{n \in \mathbb{N}}$  is a Zig-Zag walk on  $\mathbb{Z}^d$  then, like in the one-dimensional case,  $\sigma(X_n, V_n) = -\sigma(X_{n-1}, V_{n-1})$ . In particular, denoting  $\mathcal{A}_s = \{(x, v) \in \mathbb{Z}^d \times \{-1, 1\}^d : \sigma(x, v) = s\}$  then  $\mathcal{A}_s \cup \mathcal{A}_{-s}$  is fixed by the Zig-Zag walk (see Figure 4.2). Therefore, in dimension larger than 1, the Zig-Zag walk is not irreducible on  $\mathbb{Z}^d \times \{-1, 1\}^d$ .

**Proposition 4.4.** *Suppose that there exist  $R > 0$  such that  $U(x + v_i e_i) > U(x)$  for all  $i \in \llbracket 1, d \rrbracket$ ,  $x \in \mathbb{Z}^d$ ,  $v \in \{-1, 1\}^d$  with  $x_i v_i > R$ . Then for all  $s \in \{-1, 1\}^d$  the Zig-Zag walk on  $\mathbb{Z}^d$  associated to  $U$  is irreducible on  $\mathcal{A}_s \cup \mathcal{A}_{-s}$ .*

*Proof:* Let  $s \in \{-1, 1\}^d$  be fixed, and let  $(x, v) \in \mathcal{A}_s \cup \mathcal{A}_{-s}$ . Remark that  $(x, w) \in \mathcal{A}_s \cup \mathcal{A}_{-s}$  if and only if  $w \in \{-v, v\}$ . We say that we can reach  $(y, w)$  from  $(x, v)$

<sup>1</sup>The nice proof of irreducibility of Bierkens et al. (2019b) under weaker conditions on  $U$  may possibly be partially adapted for the discrete Zig-Zag walk. Nevertheless, note that the smoothness condition on  $U$  has no discrete counterpart. In any case, the potential  $U(x) = \|x\|_\infty$  would still be a counter-example for which the conclusion of Proposition 4.4 would not hold.

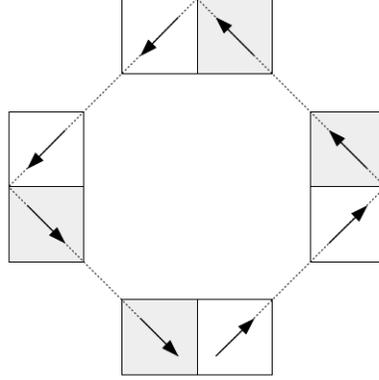


FIGURE 4.3. An admissible path from  $(x, (1, 1))$  to  $(x - e_1, (1, -1))$ .

if there is a path from  $(x, v)$  to  $(y, w)$  that has a non-negative probability for the Zig-Zag walk. Starting from  $(x, v) \in \mathbb{Z}^d \times \{-1, 1\}^d$ , we can reach all the points  $(x + nv, v)$  with  $n \in \mathbb{N}$ . For  $n$  large enough,  $v_i(x_i + nv_i) > R$  for all  $i \in \llbracket 1, d \rrbracket$  so that each coordinate has a non-negative probability to flip its velocity in the next step. As a consequence, from  $(x + nv, v)$  with such a  $n$ ,  $(x + nv + (v + w)/2, w)$  can be reached for all  $w \in \{-1, 1\}^d$ . In particular, if  $w = -v$ , since  $(x, -v)$  can be reached from  $(x + nv, -v)$ , we see by transitivity that  $(x, -v)$  can be reached from  $(x, v)$ .

Second, let us show that for all  $j \in \llbracket 1, d \rrbracket$  and all  $a \in \{-1, 1\}$ ,  $(x + ae_j, -v + 2v_j e_j)$  can be reached from  $(x, v)$ . Remark that this will conclude the proof: indeed, repeating this, then for all  $x' \in \mathbb{Z}^d$ , there will exist  $w' \in \{-1, 1\}^d$  such that  $(x', w')$  (and then  $(x', -w')$  by the previous result) can be reached from  $(x, v)$ . Since  $\mathcal{A}_s \cup \mathcal{A}_{-s}$  is fixed by the Zig-Zag walk, if  $(x', z) \in \mathcal{A}_s \cup \mathcal{A}_{-s}$  then necessarily  $z \in \{w', -w'\}$ , and thus we will have obtained that all points of  $\mathcal{A}_s \cup \mathcal{A}_{-s}$  will be reachable from  $(x, v)$ .

Hence, fix  $j \in \llbracket 1, d \rrbracket$ ,  $a \in \{-1, 1\}$  and set  $(x', v') = (x + ae_j, -v + 2v_j e_j)$ . Since  $(x, -v)$  can be reached from  $(x, v)$  we can suppose that  $v_j = -a$ . Let  $n_1$  and  $n_2$  be large enough so that, for all  $i \in \llbracket 1, d \rrbracket$ ,

$$v_i(x_i + n_1 v_i) > R, \quad -v'_i(x'_i - n_2 v'_i) > R, \quad -v'_i(x_i + (n_1 + 1)v_i - v_j e_j - n_2 v'_i) > R.$$

Consider the following path: from  $(x, v)$ , go to  $(x + n_1 v, v)$ , flip the  $j^{\text{th}}$  velocity, which gives  $(x + (n_1 + 1)v - v_j e_j, -v')$ , go straight to  $(x + (n_1 + 1)v - v_j e_j - n_2 v', -v')$ , flip all the velocities but the  $j^{\text{th}}$ , which gives  $(x + (n_1 + 1)v - 2v_j e_j - n_2 v', -v)$ , go straight to  $(x + v - 2v_j e_j - n_2 v', -v)$ , flip the  $j^{\text{th}}$  velocity, which gives  $(x - v_j e_j - n_2 v', v')$  and go straight to  $(x', v')$  (see Figure 4.3). It is clear that, in view of the conditions on  $n_1$  and  $n_2$  the two first flips have a non-negative probability. For the third one, remark that the  $j^{\text{th}}$  coordinate of  $x + v - 2v_j e_j - n_2 v'$  is  $x_j - n_2 v'_j$  and that  $v'_j = v_j$ . Hence, the condition that  $-v'_i(x'_i - n_2 v'_i) > R$  ensures that the third flip, hence the whole path, has a non-negative probability, which concludes.  $\square$

4.4.2. *A Lyapunov function.* For some fixed  $a, b > 0$  and for  $x \in \mathbb{Z}^d, v \in \{-1, 1\}^d$  and  $i \in \llbracket 1, d \rrbracket$ , denote

$$\mathcal{V}_i(x, v) = e^{a|x_i| + b\mathbb{1}_{\{x_i v_i > 0\}}},$$

and  $\mathcal{V}(x, v) = \sum_{i=1}^d \mathcal{V}_i(x, v)$ . Consider  $(X_n, V_n)_{n \in \mathbb{N}}$  the Zig-Zag walk on  $\mathbb{Z}^d$  associated to  $U : \mathbb{Z}^d \rightarrow \mathbb{R}$  and denote  $Q$  its transition operator, namely

$$Qf(x, v) = \sum_{w \in \{\pm 1\}^d} f\left(x + \frac{v+w}{2}, w\right) p(x, v; w).$$

**Proposition 4.5.** *Suppose that there exist  $R, h > 0$  such that for all  $i \in \llbracket 1, d \rrbracket, x \in \mathbb{Z}^d$  and  $v \in \{-1, 1\}^d$ ,*

$$(U(x + v_i e_i) - U(x))_+ \geq h \mathbb{1}_{\{v_i x_i > R\}}. \tag{4.2}$$

Then, for all choice of  $a, b > 0$  and for all  $x \in \mathbb{Z}^d$  and  $v \in \{-1, 1\}^d$ ,

$$Q\mathcal{V}(x, v) \leq \max(e^{-h+a} + (1 - e^{-h})e^{-b}, e^{-a}) \mathcal{V}(x, v) + de^{a(R+1)+b}. \tag{4.3}$$

*Proof:* For all  $x \in \mathbb{Z}^d, v \in \{-1, 1\}^d$ ,

$$Q\mathcal{V}_1(x, v) = q_1(x, v_1) e^{a|x_1+v_1| + b\mathbb{1}_{\{(x_1+v_1)v_1 > 0\}}} + (1 - q_1(x, v_1)) e^{a|x_1| + b\mathbb{1}_{\{x_1 v_1 < 0\}}}.$$

If  $x_1 v_1 > R$ ,  $q_1(x, v_1) \leq e^{-h}$ , so that

$$\begin{aligned} Q\mathcal{V}_1(x, v) &= q_1(x, v_1) e^{a|x_1| + a + b} + (1 - q_1(x, v_1)) e^{a|x_1|} \\ &\leq e^{a|x_1|} + e^{-h} (e^{a|x_1| + a + b} - e^{a|x_1|}) \\ &= (e^{-b} + e^{-h+a} - e^{-h-b}) \mathcal{V}_1(x, v). \end{aligned}$$

If  $x_1 v_1 < -R$  then  $(x_1 + v_1)v_1 \leq 0$  and  $q_1(x, v_1) = 1$ , and thus

$$Q\mathcal{V}_1(x, v) = e^{a|x_1| - a} = e^{-a} \mathcal{V}_1(x, v).$$

If  $|x_1 v_1| \leq R$  then

$$Q\mathcal{V}_1(x, v) \leq e^{a(R+1)+b}.$$

Since  $Q$  is the result of  $d$  consecutive and identical one-dimensional transitions, we get the result by summing over  $i \in \llbracket 1, d \rrbracket$ .  $\square$

Taking  $a = h/2$  and  $b$  arbitrarily large we get that  $Q\mathcal{V} \leq \gamma \mathcal{V} + C$  with  $\gamma < 1$  (in fact  $\gamma$  arbitrarily close to  $e^{-h/2}$ ), which means that  $\mathcal{V}$  is a Lyapunov function for  $Q$ .

Remark that similar computations in the case of the continuous-time Zig-Zag process on  $\mathbb{R}^d$  show that if  $(v_i \partial_{x_i} U(x))_+ \geq h \mathbb{1}_{\{x_i v_i > 0\}}$  for all  $x \in \mathbb{R}^d$  such that  $|x_i| > R$  then

$$\tilde{\mathcal{V}}(x, v) := \sum_{i=1}^d e^{a|x_i| + b\varphi(v_i x_i)},$$

where  $\varphi(s)$  is some smooth approximation of  $sign(s)$ , is a Lyapunov function for the continuous-time process. Note that this condition on  $U$  is not covered by Bierkens et al. (2019b, Condition 3) since the latter constrains  $|\nabla U(x)|$  to go to infinity at infinity, excluding Laplace-tail distributions. Hence, our computations extends the scope of Lemma 2 (hence Theorem 2) of Bierkens et al. (2019b). Note that condition (4.2) holds when  $U(x) = \sum_{i=1}^d |x_i|$  but not when  $U(x) = |x|$ . Of course, condition (4.2) roughly means that the different coordinates are more or less independent at

infinity and thus it is not surprising that we recover, in a discrete-space case, the one-dimensional computations of [Fontbona et al. \(2016, Proposition 2.8\)](#).

In the following, under condition (4.2) we fix  $a = h/2$  and  $b$  such that  $e^{-b} = e^{-h/4} - e^{-h/2}$ , in which case (4.3) implies

$$Q\mathcal{V}(x, v) \leq e^{-h/4}\mathcal{V}(x, v) + \left(1 - e^{-h/4}\right) d \frac{e^{h(R/2+1)}}{(1 - e^{-h/4})^2}.$$

This classically yields uniform in time exponential moment bounds on the Zig-Zag walk, since for all  $n \in \mathbb{N}$ ,

$$Q^n \mathcal{V}(x, v) \leq e^{-nh/4} \mathcal{V}(x, v) + \left(1 - e^{-nh/4}\right) d \frac{e^{h(R/2+1)}}{(1 - e^{-h/4})^2}$$

and thus, if  $(X_0, V_0) = (x, v)$ , we bound

$$\mathbb{E} \left( e^{\frac{h}{2d} |X_n|} \right) \leq \mathbb{E} \left( e^{\frac{h}{2d} \sum_{i=1}^d |X_{n,i}|} \right) \leq \frac{1}{d} \mathbb{E} (\mathcal{V}(X_n, V_n)) = \frac{1}{d} Q^n \mathcal{V}(x, v).$$

**4.4.3. Long-time convergence.** For  $x \in \mathbb{Z}^d$  and  $s \in \{-1, 1\}^d$ , let  $v(x, s)$  be the (unique) vector of  $\{-1, 1\}^d$  such that  $\sigma(x, v(x, s)) = s$ . Then  $\mathcal{A}_s = \{(x, v(x, s)), x \in \mathbb{Z}^d\}$  and  $\mathcal{A}_{-s} = \{(x, -v(x, s)), x \in \mathbb{Z}^d\}$ . Consider on  $\mathcal{A}_s$  the probability measure

$$\mu_s(x, v) = \pi(x) \delta_{v(x, s)}(v) = \frac{\mu(x, v) \mathbb{1}_{(x, v) \in \mathcal{A}_s}}{\mu(\mathcal{A}_s)}.$$

For a given  $\mathcal{W} : \mathbb{Z}^d \times \{-1, 1\}^d \rightarrow [1, +\infty)$ , we endow  $\mathcal{P}_{\mathcal{W}} := \{\mu \in \mathcal{P}(\mathbb{Z} \times \{-1, 1\}^d) : \mu(\mathcal{W}) < +\infty\}$  with the norm

$$\|\mu - \nu\|_{\mathcal{W}} := \sup_{|f| \leq \mathcal{W}} (\mu(f) - \nu(f)),$$

which makes it complete.

**Theorem 4.6.** *Suppose that  $U$  admits a strict local minimum at some  $x_* \in \mathbb{Z}^d$  and that there exist  $R, h > 0$  such that (4.2) holds for all  $i \in \llbracket 1, d \rrbracket$ ,  $x \in \mathbb{Z}^d$  and  $v \in \{-1, 1\}^d$ . Set  $\mathcal{W}(x, v) = \sum_{i=1}^d e^{h|x_i|/2}$ . Then, there exist  $C > 0$  and  $\rho \in (0, 1)$  such that for all  $x \in \mathbb{Z}^d$ ,  $v \in \{-1, 1\}^d$  and  $n \in \mathbb{N}_*$ ,*

$$\|\delta_{(x, v)} Q^{2n} - \mu_{\sigma(x, v)}\|_{\mathcal{W}} \leq C \rho^n \mathcal{W}(x, v).$$

*Proof:* Let  $(x, v), (x', v') \in \mathcal{A}_s$  for some  $s \in \{-1, 1\}^d$ . Proposition 4.4 gives a path from  $(x, v)$  to  $(x', v')$  whose transitions are non-negative under  $Q$ . Since  $Q(\mathcal{A}_s) \subset Q(\mathcal{A}_{-s})$ , the length of such a path is necessarily even, from which  $Q^2$  is irreducible on  $\mathcal{A}_s$ . The path  $(x_*, v(x_*, s)) \rightarrow (x_*, -v(x_*, s)) \rightarrow (x_*, v(x_*, s))$  having a non-negative probability under  $Q$ ,  $Q^2$  is aperiodic on  $\mathcal{A}_s$ . As a consequence, for all  $(x, v), (x', v') \in \mathcal{A}_s$ , there exist  $n_0$  such that  $Q^{2n}((x, v), (x', v')) > 0$  for all  $n \geq n_0$ . From Proposition 4.1,  $\mu_s$  is invariant for  $Q^2$ .

Consider  $\mathcal{V}$  as defined in Section 4.4.2 with  $a = h/2$  and  $b$  large enough, so that

$$Q^{2n} \mathcal{V}(x, v) \leq \gamma^{2n} \mathcal{V}(x, v) + (1 - \gamma^{2n}) C \tag{4.4}$$

for some  $\gamma \in (0, 1)$ ,  $C > 0$ . The set  $\{\mathcal{V} \leq 4C\}$  is finite. Fix any point  $x' \in \mathbb{Z}^d$  (say,  $x' = x_*$ ) and let

$$m := \max_{\mathcal{V}(x, v) \leq 4C} \min\{k \in \mathbb{N} : Q^{2r}((x, v), (x', v(x', \sigma(x, v)))) > 0 \forall r \geq k\}.$$

Then

$$\min\{Q^{2m}((x, v), (x', s)) : (x, v) \in \mathcal{A}_s, \mathcal{V}(x, v) \leq 4C\} > 0. \tag{4.5}$$

From Hairer and Mattingly (2011, Theorem 1.2) applied to  $Q^{2m}$ , the Foster-Lyapunov condition (4.4) (applied with  $n = m$ ) and the Doeblin condition (4.5) imply the existence of  $\rho \in (0, 1)$  and  $C' > 0$  such that for all  $x \in \mathbb{Z}^d, v \in \{-1, 1\}^d$  and  $k \in \mathbb{N}_*$ ,

$$\|\delta_{(x,v)}Q^{2km} - \mu_{\sigma(x,v)}\|_{\mathcal{V}} \leq C'\rho^k\mathcal{V}(x, v).$$

In fact, even without the Doeblin condition, following the proof of Hairer and Mattingly (2011, Theorem 1.2), we also get that the Lyapunov condition given by Proposition 4.5 alone implies the following: there exist  $\tilde{C} > 0$  such that, for all probability measures  $\nu, \nu' \in \mathcal{P}_{\mathcal{V}}$ ,

$$\|\nu Q - \nu' Q\|_{\mathcal{V}} \leq \tilde{C}\|\nu - \mu_s\|_{\mathcal{V}}, \tag{4.6}$$

and in particular

$$\|\nu Q^2 - \mu_s\|_{\mathcal{V}} = \|\nu Q^2 - \mu_s Q^2\|_{\mathcal{V}} \leq \tilde{C}^2\|\nu - \mu_s\|_{\mathcal{V}}.$$

Then for all  $n \in \mathbb{N}$ , considering the Euclidian division  $n = km + r$  we get that

$$\|\delta_{(x,v)}Q^{2n} - \mu_{\sigma(x,v)}\|_{\mathcal{V}} = \|\delta_{(x,v)}Q^{2r}Q^{2km} - \mu_{\sigma(x,v)}\|_{\mathcal{V}} \leq \tilde{C}^{2m}C'\rho^{n/m-1}\mathcal{V}(x, v).$$

The equivalence between  $\mathcal{V}$  and  $\mathcal{W}$ , hence between  $\|\cdot\|_{\mathcal{V}}$  and  $\|\cdot\|_{\mathcal{W}}$ , concludes.  $\square$

4.4.4. Asymptotic theorems.

**Theorem 4.7.** *Suppose that  $U$  admits a strict local minimum and that there exist  $R, h > 0$  such that (4.2) holds for all  $i \in \llbracket 1, d \rrbracket, x \in \mathbb{Z}^d$  and  $v \in \{-1, 1\}^d$ . Let  $f : \mathbb{Z}^d \rightarrow \mathbb{R}$  be such that  $\|f/\mathcal{W}\|_{\infty} < \infty$ , where  $\mathcal{W}$  is defined in Theorem 4.6 (here and below we identify  $f$  with the function  $(x, v) \in \mathbb{Z}^d \times \{-1, 1\}^d \mapsto f(x)$ ). Consider the Zig-Zag walk on  $\mathbb{Z}^d \times \{-1, 1\}^d$  associated to  $U$  with some initial condition  $(x, v)$ . Then, almost surely,*

$$\frac{1}{n} \sum_{k=0}^n f(X_k) \xrightarrow{n \rightarrow \infty} \pi(f).$$

If, moreover,  $\|f/\sqrt{\mathcal{W}}\|_{\infty} < \infty$ , then

$$\left( \frac{1}{\sqrt{n}} \left( \sum_{k=0}^{\lfloor nt \rfloor} f(X_k) - \pi(f) \right) \right)_{t \geq 0} \xrightarrow[n \rightarrow +\infty]{law} (\sigma_f B_t)_{t \geq 0}$$

for some  $\sigma_f \geq 0$ , where  $(B_t)_{t \geq 0}$  is a one-dimensional Brownian motion.

*Proof:* For the first part of the Theorem, simply decompose

$$\frac{1}{n} \sum_{k=0}^n f(X_k) = \frac{1}{n} \sum_{k=1}^{\lfloor n/2 \rfloor} f(X_{2k}) + \frac{1}{n} \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} f(X_{2k+1})$$

From Proposition 4.6 and the law of large numbers for  $\mathcal{W}$ -regular ergodic Markov chains, these terms almost surely converge respectively to  $\mu_{\sigma(x,v)}(f)/2$  and  $\mu_{-\sigma(x,v)}(f)/2$  (since  $(X_1, V_1) \in \mathcal{A}_{-\sigma(x,v)}$  almost surely), which are both equal to  $\pi(f)/2$ .

For the second part, note that by the Jensen inequality and Proposition 4.5,

$$Q\sqrt{\mathcal{V}} \leq \sqrt{Q\mathcal{V}} \leq \sqrt{\gamma\mathcal{V} + C} \leq \sqrt{\gamma}\sqrt{\mathcal{V}} + \sqrt{C}$$

for some  $\gamma \in (0, 1)$  and  $C > 0$ . Hence,  $\sqrt{\mathcal{V}}$  is still a Lyapunov function for  $Q$  and the results established with  $\mathcal{V}$  and  $\mathcal{W}$  in the previous section also hold with  $\sqrt{\mathcal{V}}$  and  $\sqrt{\mathcal{W}}$ . Let  $f : \mathbb{Z}^d \times \{-1, 1\}^d \rightarrow \mathbb{R}$  with  $\|f/\sqrt{\mathcal{V}}\|_\infty < \infty$ . From Theorem 4.6 (applied with  $\sqrt{\mathcal{W}}$ ), for all  $(x, v) \in \mathbb{Z}^d \times \{-1, 1\}^d$  and all  $n \in \mathbb{N}$ ,

$$|Q^{2n}f(x, v) - \mu_{\sigma(x,v)}(f)| \leq C\rho^n \|f/\sqrt{\mathcal{W}}\|_\infty \sqrt{\mathcal{W}(x)}$$

and, using (4.6),

$$\begin{aligned} |Q^{2n+1}f(x, v) - \mu_{-\sigma(x,v)}(f)| &= |Q^{2n}Qf(x, v) - \mu_{\sigma(x,v)}Q(f)| \\ &\leq \tilde{C}C\rho^n \|f/\sqrt{\mathcal{W}}\|_\infty \sqrt{\mathcal{W}(x)}. \end{aligned}$$

Here we used that  $\mu_s Q = \mu_{-s}$ , which can be obtained from

$$\mu_s Q = \mu_s Q^{2n} Q = \mu_s Q Q^{2n} \xrightarrow{n \rightarrow \infty} \mu_{-s},$$

where the limit holds in  $\mathcal{P}_{\mathcal{W}}$  thanks to Proposition 4.6 together with the fact that the support of  $\mu_s Q$  is included in  $\mathcal{A}_{-s}$ . We have thus obtained that

$$g(x, v) := \sum_{n \in \mathbb{N}} (Q^n f(x, v) - \mu_{(-1)^n \sigma(x,v)}(f))$$

is well-defined and satisfies  $\|g/\sqrt{\mathcal{W}}\|_\infty \leq C\|f/\sqrt{\mathcal{W}}\|_\infty$  for some  $C > 0$  independent from  $f$ . Now suppose that in fact  $f$  is a function of space alone, i.e.  $f(x, v) = f(x)$ . In that case  $\mu_s(f) = \mu(f)$  for all  $s \in \{-1, 1\}^d$  and thus

$$Qg(x, v) = \sum_{n \in \mathbb{N}} (Q^{n+1}f(x, v) - \mu(f)) = g(v, x) - f(x) + \mu(f),$$

in other words  $g$  is the Poisson solution associated to  $Q$  and  $f$ . Since  $\sqrt{\mathcal{W}} \in L^2(\mu)$ , so does  $g$ , and Maigret (1978, Theorem 3.1) concludes.  $\square$

### 5. Numerical scheme for hybrid kinetic samplers

5.1. *The continuous-time processes.* In this section we consider a class of kinetic processes for MCMC that can have a jump, drift and/or diffusion component at the same time in their dynamics. They are defined as follows.

Let  $U \in \mathcal{C}^\infty(\mathbb{T}^d)$  (where  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$  is the 1-periodic torus) and denote by  $\mu$  the Gibbs measure on  $E = \mathbb{T}^d \times \mathbb{R}^d$  associated to the Hamiltonian  $H(x, v) = U(x) + |v|^2/2$ , namely the probability law on  $E$  with density proportional to  $\exp(-H)$ . Suppose that  $\nabla U(x) = \sum_{i=0}^N F_i(x)$  where  $N \in \mathbb{N}$  and  $F_i \in \mathcal{C}^\infty(\mathbb{T}^d, \mathbb{R}^d)$  for all  $i \in \llbracket 0, N \rrbracket$ . Consider the operator  $L$  defined for all  $f \in \mathcal{C}^2(E)$  by

$$L = A_1 + A_2 + \sum_{i=1}^N A_{3,i} + \gamma A_4 + \lambda A_5, \tag{5.1}$$

where

$$\begin{aligned} A_1 f(x, v) &= v \cdot \nabla_x f(x, v) \\ A_2 f(x, v) &= -F_0(x) \cdot \nabla_v f(x, v) \\ A_{3,i} f(x, v) &= (v \cdot F_i(x))_+ (f(x, R_i(x, v)) - f(x, v)) \quad \forall i \in \llbracket 1, N \rrbracket \\ A_4 f(x, v) &= v \cdot \nabla_v f(x, v) + \Delta_v f(x, v) \\ A_5 f(x, v) &= \int_{\mathbb{R}^d} (f(x, w) - f(x, v)) \nu_d(dw) \end{aligned}$$

and where  $\gamma, \lambda \geq 0$ ,  $\nu_d$  is the standard  $d$ -dimensional Gaussian distribution and

$$R_i(x, v) = v - 2 \left( \frac{F_i(x) \cdot v}{|F_i(x)|^2} F_i(x) \right) \mathbb{1}_{F_i(x) \neq 0}$$

is the orthogonal reflection of  $v$  with respect to  $F_i(x)$ . We call respectively  $A_1$  the transport operator,  $A_2$  the drift one,  $A_{3,i}$  the  $i^{\text{th}}$  bounce one,  $A_4$  the Ornstein-Uhlenbeck (or friction/dissipation) one and  $A_5$  the refreshment one.

As particular cases, many usual kinetic processes used in MCMC algorithms can be recovered:

- $F_0 = \nabla_x U$  and  $\lambda = \gamma = 0$  corresponds to the Hamiltonian dynamics.
- $F_0 = \nabla_x U$ ,  $\gamma > 0$  and  $\lambda = 0$  to the Langevin diffusion.
- $F_0 = \nabla_x U$ ,  $\gamma = 0$  and  $\lambda > 0$  to the Hybrid Monte Carlo (HMC) algorithm.
- $F_0 = 0$ ,  $N = 1$ ,  $F_1 = \nabla_x U$ ,  $\gamma = 0$  and  $\lambda > 0$  to the Bouncy Particle sampler.
- $F_0 = 0$ ,  $N = d$ ,  $F_i = \nabla_{x_i} U e_i$  (recall  $e_i$  denotes the  $i^{\text{th}}$  vector of the canonical basis of  $\mathbb{R}^d$ ),  $\gamma = \lambda = 0$  to the Zig-Zag process.

We could also consider other kinds of jump mechanisms, like the randomized bounces of [Roberts \(2017\)](#); [Michel et al. \(2017a\)](#), or different kinds of relaxation operators in the velocity operator rather than  $A_4$  and  $A_5$ , for instance refreshment of velocities coordinate by coordinate, or partial refreshments for which, at exponential random times with parameter  $\lambda$ , the velocities  $v$  jumps to  $(1 - \alpha)v + \sqrt{\alpha}G$  where  $G \sim \nu_d$  (varying  $\alpha \in (0, 1]$  and  $\lambda$  interpolates between  $A_5$  and  $A_4$ , the latter being the limit  $\alpha \rightarrow 0$  and  $\lambda \rightarrow +\infty$  with  $\lambda\alpha = 1$ ). However, this would just make the notations heavier and the presentation more confused, without adding any particularly new idea with respect to the discussion to come, and thus we stick to (5.1).

For MCMC purposes, the law of the process associated to a generator  $L$  of the form above should converge in large times toward the target measure  $\mu$ . This is established in the next two results. However, this is not the main motivation of this section, which is the study of the numerical sampling of such a process, and thus we only give sketches of proof and references for these results.

**Proposition 5.1.** *The set of compactly supported smooth functions  $C_c^\infty(E)$  is a core for  $L$  and  $\mu$  is invariant for  $L$ .*

*Proof:* The proof that  $C_c^\infty(E)$  is a core for  $L$  is similar to the case of the Bouncy Particle Sampler in [Durmus et al. \(2018b\)](#), Theorem 21 and Proposition 23). From that, the invariance of  $\mu$  straightforwardly follows from the fact that  $\mu(Lf) = 0$  for all compactly supported  $C^2$  functions, which is easily checked through integration by part (see also [Monmarché, 2016](#), Section 1.4).  $\square$

Similarly to Section 4.4.3, we denote  $\mathcal{W}(x, v) = 1 + |v|^2$ ,  $\mathcal{P}_{\mathcal{W}} = \{\nu \in \mathcal{P}(E), \nu(\mathcal{W}) < \infty\}$ ,  $\|\nu_1 - \nu_2\|_{\mathcal{W}} = \sup_{|f| \leq \mathcal{W}} |\nu_1 f - \nu_2 f|$ .

**Proposition 5.2.** *Suppose that either  $\gamma > 0$  or  $\lambda > 0$ . Then there exist  $\kappa, C > 0$  such that the semi-group  $(P_t)_{t \geq 0}$  associated to  $L$  satisfies, for all  $t \geq 0$  and  $\nu \in \mathcal{P}_{\mathcal{W}}$ ,*

$$\|\nu P_t - \mu\|_{\mathcal{W}} \leq C e^{-\kappa t} \|\nu - \mu\|_{\mathcal{W}}$$

*Proof:* Similarly to Theorem 4.6, this is classically obtained by combining a Doeblin and a Foster-Lyapunov conditions ([Hairer and Mattingly, 2011](#)).

When  $\gamma > 0$  or  $\lambda > 0$ , the Doeblin condition is obtained through controllability arguments (a notable fact is that under quite general conditions on  $U$  the Zig-Zag process is irreducible even if  $\lambda = \gamma = 0$ , see [Bierkens et al., 2019b](#)). Remark that the bounce operators play no role here since if, uniformly in  $(x, v)$  in a compact of  $E$ ,  $\delta_{(x,v)} e^{t_0(A_1+A_2+\gamma A_4+\lambda A_5)}$  is bounded below at some time  $t_0 > 0$  by  $c_0$  times the uniform measure on some compact set with some  $c_0 > 0$ , then  $\delta_{(x,v)} e^{t_0 L}$  is bounded below by  $c_0 e^{-t_0 \sum_{i=1}^N \|F_i\|_\infty}$  times this uniform measure, where we have used that the total bounce rate is bounded above by  $\sum_{i=1}^N \|F_i\|_\infty$  so that there is a probability at least  $e^{-t_0 \sum_{i=1}^N \|F_i\|_\infty}$  that no bounce occurs on the time interval  $[0, t_0]$ .

The Lyapunov condition stems from the fact that for all  $(x, v) \in E$ ,

$$\begin{aligned} A_1 \mathcal{W}(x, v) &= 0 \\ A_2 \mathcal{W}(x, v) &\leq 2 \|F_0\|_\infty |v| \\ A_{3,i} \mathcal{W}(x, v) &= 0 \quad \forall i \in \llbracket 1, N \rrbracket \\ A_4 \mathcal{W}(x, v) &= -2|v|^2 + 2d \\ A_5 \mathcal{W}(x, v) &= -|v|^2 + d, \end{aligned}$$

and thus  $L\mathcal{W} \leq -\min(2\gamma, \lambda)/2\mathcal{W} + C$  for some  $C > 0$ . □

As announced, from now and in the rest of Section 5 we focus on the question of the numerical sampling of the process with generator  $L$ .

When  $F_0 = 0$ , and  $\gamma = 0$ , the process is a piecewise deterministic velocity jump process. Between two random jumps, the process simply follows the flow  $(x, v) \mapsto (x + tv, v)$ , so that the jump time  $T_i$  associated to the vector field  $F_i$  follows the law

$$\mathbb{P}(T_i > t) = \exp\left(-\int_0^t (v \cdot F_i(x + sv))_+ ds\right).$$

Provided that for all  $x, v \in \mathbb{R}^d$  and  $s > 0$ ,  $(v \cdot F_i(x + sv))_+ \leq \varphi_{x,v}(s)$  for some function  $\varphi_{x,v}$  such that  $\int_0^t \varphi_{x,v}(s) ds$  can be computed, the continuous-time thinning algorithm allows for an exact simulation of the jump times, hence of the process ([Lemaire et al., 2018](#)). The absence of discretization bias on the invariant measure is an argument in favour of these kinetic processes. This is no longer the case when  $F_0 \neq 0$ , except in very particular cases (e.g. harmonic oscillators) where the ODE  $\partial_t(x, v) = (v, -F_0(x))$  can be explicitly solved. In most cases, this ODE is solved numerically and thus the simulation for the stochastic process is not exact. In order to conserve the invariant measure and suppress the bias, a Metropolis step can be added ([Bou-Rabee, 2014](#)), but this slows down the motion of the process, increases the variance and may thus be counter-productive, as observed when comparing Metropolis Adjusted Langevin algorithm and Unadjusted Langevin Algorithm ([Durmus and Moulines, 2017](#)).

So why mix a deterministic drift and jump mechanisms if this prohibits exact simulation? The motivation is given by the possible numerical gain given by thinning, as presented in Section 3.2. Indeed, as said before, exact simulation requires a bound on the jump rate, and a poor (i.e. large) bound leads to many jump proposals per time unit and a very low efficiency. In particular, jump mechanisms are not adapted for fastly-varying potentials, like potentials used in molecular dynamics with a singularity at zero (Lennard-Jones, Coulomb... ). On the other hand, jump

mechanisms are very efficient with Lipschitz potentials. So, a mixed drift/jump part is interesting as soon as the potential exhibits different scales, like fast-varying but numerically cheap parts together with Lipschitz but numerically intensive parts. This is similar to the idea of multi-time-steps algorithms (Tuckerman et al., 1991; Gibson and Carter, 1993) with somehow random adaptive time-steps, except that now we are simply going to discretize (with a unique time-step, no subtlety here) a continuous-time process which is ergodic with respect to the target law and thus there shouldn't be any resonance problem as exhibited by multi-time-steps algorithms. Besides, for the applications in molecular dynamics that have motivated this question (Monmarché et al., 2020), due to stability issues raised by very fast oscillations in some parts of the system, the time-step is anyway constrained to be very small, as compared to a high variance that comes from the problem of exploring a complex, high-dimensional, multi-scale, metastable landscape. So, exact simulation is not necessarily our objective.

5.2. *A Strang splitting scheme.* A Strang splitting scheme to compute the evolution given by a generator  $L = L_1 + L_2$  is based on the fact that, formally,

$$e^{t(L_1+L_2)} = e^{tL_1/2}e^{tL_2}e^{tL_1/2} + o_{t \rightarrow 0}(t^2).$$

Hence, if we can simulate exactly a process with generator  $L_1$  and  $L_2$ , or more generally if we have second-order approximations of those, we get a second-order scheme for  $L_1 + L_2$ . Using twice this fact,

$$\begin{aligned} e^{t(L_1+L_2+L_3)} &= e^{tL_1/2}e^{t(L_2+L_3)}e^{tL_1/2} + o_{t \rightarrow 0}(t^2) \\ &= e^{tL_1/2}e^{tL_2/2}e^{tL_3}e^{tL_2/2}e^{tL_1/2} + o_{t \rightarrow 0}(t^2). \end{aligned}$$

In particular, from the considerations developed in Section 3.3 (and Theorem 3.2), the equilibrium of the corresponding Markov chain is close (in some senses) to the Gibbs measure  $\mu$  at order  $\delta^2$ , where  $\delta$  is the timestep.

For instance, for the Langevin diffusion, which corresponds to  $L$  given by (5.1) with  $\gamma > 0$ ,  $\lambda = 0$ ,  $N = 0$  and  $F_0 = \nabla_x U$ , many splitting schemes can be considered, see e.g. the discussions in Leimkuhler and Matthews (2013); Bou-Rabee (2014). A very precise study, both theoretical and empirical, of all the possible schemes for mixed jump/diffusion processes is beyond the scope of the present paper, and we will only consider one particular choice. Consider the splitting

$$L_1 = A_1, \quad L_2 = A_2 + \gamma A_4, \quad L_3 = \sum_{i=1}^N A_{3,i} + \lambda A_5.$$

Each of the three evolutions corresponding to  $e^{tL_i}$ ,  $i = 1, 2, 3$  can be sampled exactly (remark that, in particular cases, the velocity jump process corresponding to  $e^{t(L_1+L_3)}$  could also be sampled exactly). As a consequence, for a given time-step  $\delta > 0$ , we consider the kinetic walk  $(X_n, V_n)_{n \in \mathbb{N}}$  whose transition is defined as follows:

- (1) Set  $\tilde{X}_n = X_n + \delta V_n/2$ .
- (2) If  $\gamma > 0$ , set

$$\tilde{V}_n = e^{-\gamma\delta/2}V_n - \left(1 - e^{-\gamma\delta/2}\right)F_0(\tilde{X}_n) + \sqrt{(1 - e^{-\gamma\delta/2})}G$$

with  $G$  a standard Gaussian random variable. If  $\gamma = 0$ , set  $\tilde{V}_n = V_n - \delta F_0(\tilde{X}_n)/2$ .

- (3) Set  $\tilde{V}_n = W_\delta$  where  $(Y_t, W_t)_{t \in [0, \delta]}$  is a Markov chain with generator  $L_3$  and initial condition  $(Y_0, W_0) = (\tilde{X}_n, \tilde{V}_n)$  (so that  $Y_t = \tilde{X}_n$  for all  $t \in [0, \delta]$ ).
- (4) If  $\gamma > 0$ , set

$$V_{n+1} = e^{-\gamma\delta/2} \hat{V}_n - (1 - e^{-\gamma\delta/2}) F_0(\tilde{X}_n) + \sqrt{(1 - e^{-\gamma\delta/2})} G'$$

with  $G'$  a standard Gaussian random variable. If  $\gamma = 0$ , set  $V_{n+1} = \hat{V}_n - \delta F_0(\tilde{X}_n)/2$ .

- (5) Set  $X_{n+1} = \tilde{X}_n + \delta V_{n+1}/2 = X_n + \delta(V_n + V_{n+1})/2$ .

(Of course, implicitly, the variables  $G, G'$  and  $(W_t)_{t \in [0, \delta]}$  are independent one from the other and from the past trajectory). Then, denoting  $Q_\delta$  the transition operator of  $(X_n, V_n)$  defined by  $Q_\delta f(x, v) = \mathbb{E}(f(X_1, V_1) \mid (X_0, V_0) = (x, v))$  for all measurable bounded function  $f$ , one has

$$Q_\delta = e^{\delta L_1/2} e^{\delta L_2/2} e^{\delta L_3} e^{\delta L_2/2} e^{\delta L_1/2}.$$

To make the link with the discussion of Section 3, for all measurable bounded  $f$ ,

$$Q_\delta f(x, v) = \int f\left(x + \frac{v+w}{2}, w\right) p_\delta(x, v; dw),$$

where  $p_\delta$  is defined by the fact that, in the algorithm above, for all  $n \in \mathbb{N}$  the law of  $V_{n+1}$  is  $p_\delta(X_n, V_n; \cdot)$ .

**Proposition 5.3.** *Let  $(Y_t, W_t)_{t \geq 0}$  be a Markov process associated to  $L$  and, for all  $\delta > 0$ , let  $(X_n^\delta, V_n^\delta)_{n \geq 0}$  be a Markov chain with transition operator  $Q_\delta$  and initial condition  $(X_0^\delta, V_0^\delta) = (Y_0, W_0)$ . Then*

$$\left(X_{\lfloor t/\delta \rfloor}^\delta, V_{\lfloor t/\delta \rfloor}^\delta\right)_{t \geq 0} \xrightarrow[\delta \rightarrow 0]{law} (Y_t, W_t)_{t \geq 0}.$$

*Proof:* We start by a smoothing/truncation step similar to the study in [Durmus et al. \(2018b, Theorem 21\)](#). For all  $\varepsilon > 0$  small enough and all  $i \in \llbracket 1, N \rrbracket$ , we consider the jump rate on  $E$  given by

$$\lambda_{i,\varepsilon}(x, v) = \frac{(v \cdot F_i(x) - \varepsilon)_+^2}{\varepsilon + (v \cdot F_i(x) - \varepsilon)_+}$$

and the associated regularized bounce operator

$$A_{3,i,\varepsilon} f(x, v) = \lambda_{i,\varepsilon}(x, v) (f(x, R_i(x, v)) - f(x, v)).$$

We also consider  $\nu_{d,\varepsilon}$  the standard Gaussian law conditioned on  $\{|v| \leq 1/\varepsilon\}$  and the corresponding truncated refreshment operator

$$A_{5,\varepsilon} f(x, v) = \int_{\mathbb{R}^d} (f(x, w) - f(x, v)) \nu_{d,\varepsilon}(dw).$$

Then,  $L_\varepsilon := A_1 + A_2 + \sum_{i=1}^N A_{3,i,\varepsilon} + \gamma A_4 + \lambda A_{5,\varepsilon}$  is such that

$$\sup_{\|f\|_\infty \leq 1} \|L f - L_\varepsilon f\|_\infty \leq C \varepsilon$$

for some  $C > 0$ . The two processes  $(Y_t, W_t)_{t \geq 0}$  and  $(Y_t^\varepsilon, W_t^\varepsilon)_{t \geq 0}$  with respective generators  $L$  and  $L_\varepsilon$  can be defined simultaneously following the synchronous coupling detailed in [Durmus et al. \(2018b, Section 6\)](#), in such a way that

$(Y_t, W_t) = (Y_t^\varepsilon, W_t^\varepsilon)$  up to a random time that is stochastically bounded above by an exponential variable with parameter  $C\varepsilon$ . In other words, for all  $t \geq 0$ ,

$$\mathbb{P}((Y_s, W_s) = (Y_s^\varepsilon, W_s^\varepsilon) \forall s \in [0, t]) \geq 1 - e^{-C\varepsilon t}.$$

In particular  $(Y_t^\varepsilon, W_t^\varepsilon)_{t \geq 0} \xrightarrow[\varepsilon \rightarrow 0]{law} (Y_t, W_t)_{t \geq 0}$ . The same coupling argument holds for a chain  $(X_n^{\delta, \varepsilon}, Y_n^{\delta, \varepsilon})_{n \in \mathbb{N}}$  with transition operator  $Q_{\delta, \varepsilon}$  defined like  $Q_\delta$  but after smoothing/truncation. It is thus sufficient to prove Proposition 5.3 in the case where  $A_5$  and  $(A_{3,i})_{i \in \llbracket 1, N \rrbracket}$  are replaced by the smooth and truncated operators.

Denote  $\mathcal{B} = \text{vect}(\{A_1, A_2, A_4, A_{5, \varepsilon}\} \cup \{A_{3,i, \varepsilon} : i \in \llbracket 1, N \rrbracket\})$ . If  $f \in \mathcal{C}_c^\infty(E)$  then  $Bf \in \mathcal{C}_c^\infty(E)$  for all  $B \in \mathcal{B}$  (this is false without the smoothing/truncation step, which is the reason why it has been added). Since  $\mathcal{C}_c^\infty(E)$  is in the strong domain of all operators of  $\mathcal{B}$ , we can use for  $B_1, B_2 \in \mathcal{B}$  the decomposition

$$\begin{aligned} e^{tB_1} e^{tB_2} f - f - t(B_1 + B_2)f &= e^{tB_1} (e^{tB_2} f - f - tB_2 f) + e^{tB_1} f - f - tB_1 f \\ &\quad + t(e^{tB_1} B_2 f - B_2 f) \end{aligned}$$

and the fact  $\|e^{tB} f\|_\infty \leq \|f\|_\infty$  for all  $B \in \mathcal{B}$  to get that

$$\|e^{tB_1} e^{tB_2} f - f - t(B_1 + B_2)f\|_\infty = o_{t \rightarrow 0}(t^2)$$

for all  $f \in \mathcal{C}_c^\infty(E)$ . Using this repeatedly, we get that

$$\left\| \frac{1}{\delta} (Q_{\delta, \varepsilon} f - f) - L_\varepsilon f \right\|_\infty \xrightarrow{\delta \rightarrow 0} 0$$

for all  $f \in \mathcal{C}_c^\infty(E)$ , and [Kallenberg \(1997, Theorem 17.28\)](#) concludes.  $\square$

**5.3. Computational complexity.** Let us informally discuss the efficiency of the algorithm introduced in the previous section that defines the transition associated with the operator  $Q_\delta$ . It should be compared with the transition of a numerical scheme for classical processes, like Langevin or Hamiltonian dynamics. The numerical cost mainly comes from the computation of the forces. For the Langevin or Hamiltonian dynamics, at each time step,  $\nabla U$  is computed once. Similarly, for a transition associated with  $Q_\delta$ ,  $F_0$  is evaluated once per step, at  $x = \tilde{X}_n$ . What about  $(F_i)_{i \in \llbracket 1, N \rrbracket}$ ?

First, consider a naive construction of the third step of the algorithm, namely the construction of  $W_\delta$  when  $(Y_t, W_t)_{t \geq 0}$  is a continuous-time Markov chain with generator  $L_3$  and initial condition  $(x, w) \in \mathbb{R}^d \times \mathbb{R}^d$ . For all  $t \geq 0$ ,  $Y_t = x$ . Start by sampling a Poisson process with intensity  $\lambda$  on  $[0, \delta]$ , consider  $T_0$  its last jump in  $[0, \delta]$  (with  $T_0 = 0$  if there is no jump, in particular if  $\lambda = 0$ ). If  $T_0 = 0$ , set  $W_{T_0} = w$ , else draw  $W_{T_0}$  according to the standard Gaussian distribution. Suppose by induction that  $T_n$  and  $W_{T_n}$  have been defined for some  $n \in \mathbb{N}$ . Let  $(E_{n,i})_{i \in \llbracket 1, N \rrbracket}$  be i.i.d. exponential random variables and let

$$\forall i \in \llbracket 1, N \rrbracket, T_{n,i} = \frac{E_{n,i}}{(W_{T_n} \cdot F_i(x))_+}, \quad T_{n+1} = \min_{i \in \llbracket 1, N \rrbracket} T_{n,i}.$$

Let  $i_n \in \llbracket 1, N \rrbracket$  be such that  $T_{n,i_n} = T_{n+1}$  ( $i_n$  is almost surely unique), set  $W_t = W_{T_n}$  for all  $t \in [T_n, T_{n+1})$  and  $W_{T_{n+1}} = R(x, W_{T_n})$ . Remark that the norm of  $W_t$  is conserved at a jump time, so that the jump rate  $\sum_{i=1}^N (W_t \cdot F_i(x))_+$  is bounded and  $\max\{n \in \mathbb{N} : T_n < \delta\}$  is almost surely finite, and thus  $W_\delta$  is defined after a finite number of jumps.

This is a correct construction, but it relies on the computation of  $F_i(x)$  for all  $i \in \llbracket 1, N \rrbracket$ . In other words, with this naive construction, there is no numerical gain:  $\nabla U(x) = \sum_{i=0}^N F_i(x)$  is computed at each time step. As a consequence, the algorithm is only useful if a relevant thinning procedure such as discussed in Section 4.3 is available, to avoid the systematic computation of  $(F_i(x))_{i \in \llbracket 1, N \rrbracket}$  at each step. Such a thinning relies on suitable bounds on the vector fields  $(F_i)_{i \in \mathbb{N}}$ , that depends on the form of  $U$  and on the way  $\nabla U$  is splitted. For this reason, in the rest of this section, we won't address this question in a general case, but rather focus on a particular example. This will illustrate the fact that there exist cases where splitting forces between drift and jump mechanisms can decrease the total computational cost of the simulation.

5.3.1. *A motivating example.* To fix ideas, consider a system of  $M$  particles in the torus  $(a\mathbb{T})^3$  interacting through truncated a Lennard-Jones potential, in other words, given some parameters  $U_0, r, a > 0$ , the total energy of the configuration  $x \in ((a\mathbb{T})^3)^M$  is

$$U(x) = U_0 \sum_{i=1}^M \sum_{j \neq i} W(|x_i - x_j|)$$

where  $W(h) = [(r/h)^{12} - (r/h)^6]\chi(h/a)$  with  $\chi$  a  $\mathcal{C}^2$  positive function with values in  $[0, 1]$  such that  $\chi(s) = 1$  for  $s \leq 1/2$  and  $\chi(s) = 0$  for  $s \geq 1$  (in particular, a particle doesn't interact with its periodic image, or with several copies of the same other particle). Remark that strictly speaking this problem doesn't enter the framework considered above where, for the sake of simplicity,  $U$  was supposed smooth, but this is typically the kind of singular potentials met in molecular dynamics simulations.

For  $i \in \llbracket 1, M \rrbracket$ , denote  $J_i$  the  $3M \times 3$  matrix with zeros everywhere except  $J_i(3(i-1)+1, 1) = J_i(3(i-1)+2, 2) = J_i(3(i-1)+3, 3) = 1$ , so that  $\nabla U = \sum_{i=1}^M J_i \nabla_{x_i} U$ . For some  $R \ll a$ , we decompose  $\nabla U = F_0 + \sum_{i=1}^M \sum_{j \neq i} J_i G_{i,j}$  with

$$\begin{aligned} F_0(x) &= U_0 \sum_{i=1}^M J_i \sum_{j \neq i} \frac{x_i - x_j}{|x_i - x_j|} W'(|x_i - x_j|) \chi(|x_i - x_j|/R) \\ G_{i,j}(x) &= U_0 \frac{x_i - x_j}{|x_i - x_j|} W'(|x_i - x_j|) (1 - \chi(|x_i - x_j|/R)). \end{aligned}$$

Then  $F_0$  gathers the (singular) short-range forces, and the  $G_{i,j}$ 's the (bounded) long-range ones. Set

$$F_{i,j} = J_i G_{i,j} \quad \forall i, j \in \llbracket 1, M \rrbracket, i \neq j. \quad (5.2)$$

In particular the number of jump mechanisms is  $N = M(M-1)$ . Remark that the jump rate associated with the vector field  $F_{i,j}$  can be bounded as  $(v \cdot J_i G_{i,j}(x))_+ \leq |v_i| U_0 \sup\{|W'(s)|, s \geq R/2\} := |v_i| C_R$ . As  $R$  increases (keeping  $R \ll a$ ),  $C_R$  decays very fast toward zero (more precisely, as  $R^{-5}$ ).

Conversely, computing  $F_0(x)$  involves computing, for each  $i \in \llbracket 1, N \rrbracket$ , a sum over all the particles  $j$  such that  $\{|x_i - x_j| \leq R\}$ . The numerical cost of this computation increases with  $R$ , but is small with respect to the computation of  $\nabla U(x)$  as long as  $R \ll a$ .

5.3.2. *Thinning.* For the Lennard-Jones system with the decomposition  $\nabla U = F_0 + \sum_{i=1}^M \sum_{j \neq i} F_{i,j}$  introduced above, the third step of the algorithm of Section 5.2 can be achieved as follows:

- Set  $W = \tilde{V}_n$ .
- For all  $i \in \llbracket 1, M \rrbracket$ ,
- Draw  $K_i$  according to a Poisson law with parameter  $|W_i|C_R M \delta$ .
- For all  $k \in \llbracket 1, K_i \rrbracket$ ,
- Draw  $j$  uniformly over  $\llbracket 1, M \rrbracket$  and  $U_{i,k}$  uniformly over  $[0, 1]$ .
- If  $j \neq i$  and  $U_{i,k} \leq \left( W_i \cdot G_{i,j}(\tilde{X}_n) \right)_+ / |W_i|C_R$ , do

$$W_i \leftarrow R_{i,j}(x, W_i) := W_i - 2 \left( \frac{G_{i,j}(\tilde{X}_n) \cdot W_i}{|G_{i,j}(\tilde{X}_n)|^2} G_{i,j}(\tilde{X}_n) \right),$$

else do nothing.

- end for all  $k$
- end for all  $i$ .
- Set  $\hat{V}_n = W$ .

We call this the thinned algorithm. Remark that  $|W_i|$  is unchanged by the reflection  $R_{i,j}$ , which allows to draw  $K_i$ , the number of jump proposals for the  $i^{\text{th}}$  particle, at the beginning of the loop (but in practice this is not important). More crucially, note also that each particle  $i \in \llbracket 1, M \rrbracket$  can be treated in parallel, and there is no need for time synchronization: indeed, two particles  $i$  and  $j$  only interact through  $G_{i,j}(\tilde{X}_n)$ , and the positions  $\tilde{X}_n$  are fixed at this step, so that the velocity jumps of each particle doesn't affect the law of the other.

5.3.3. *Numerical efficiency.* Computing  $\nabla U$  has a numerical cost of order  $\mathcal{O}(M^2)$ , and it is computed  $T/\delta$  times if we sample a trajectory of a classical process (Langevin dynamics, etc.) in a time interval  $[0, T]$  with a usual integrator with time-step  $\delta$ . Let us compare this with the cost of computing gradients in the thinned algorithm for the hybrid process.

First, denote  $\mathcal{N}(R)$  the average number of particles that are at distance less than  $R$  from a given particle. By using a Verlet list of neighbors, computing  $F_0$  has an average cost of  $\mathcal{O}(MN(R))$ .

Second, denote  $D_T$  the number of times that  $G_{i,j}$  has been evaluated for some  $i, j \in \llbracket 1, M \rrbracket$  in a trajectory of length  $T$  with time-step  $\delta$ . Then

$$D_T = \sum_{n=1}^{T/\delta} \sum_{i=1}^M S_{n,i}$$

where, conditionally to the  $\tilde{V}_{n,i}$ 's the  $S_{n,i}$ 's are independent Poisson random variables with parameter  $|\tilde{V}_{n,i}|C_R M \delta$ . By the law of large numbers for ergodic Markov chains,

$$D_T \underset{T \rightarrow \infty}{\simeq} TC_R M^2 H$$

where  $H$  is the average of  $|w_i|$  with respect to the equilibrium distribution of the chain. When  $\delta$  is small, the velocity distribution at equilibrium is close to a standard Gaussian one of dimension 3, so that  $H \leq \sqrt{3} + \mathcal{O}(\delta)$ .

Computing  $G_{i,j}(x)$  being of order  $\mathcal{O}(1)$ , the total cost is  $\mathcal{O}(M^2 C_R T + M \mathcal{N}(R) T / \delta)$ . As noted previously, as  $R$  increases,  $C_R$  decays but  $\mathcal{N}(R)$  increases. In the regime where  $\mathcal{N}(R) \ll M$ , the gain in the thinned algorithm, with respect to an integrator where the full gradient  $\nabla U$  is computed at each step, is the ratio of  $M^2 C_R T$  and  $M^2 T / \delta$ , namely  $C_R \delta$ . In other words, the long-range forces are only evaluated at an average time-step  $1/C_R$ . Depending on the parameters, this can be a significant speed-up. Indeed,  $\delta$  is typically constrained to be small in order to handle the singular behaviour of short-range forces while, for intermediate values of  $R$ ,  $C_R$  can already be small. We refer to [Monmarché et al. \(2020\)](#) for a practical case where  $\delta C_R \simeq 10^{-6}$ .

**5.3.4. Sampling rate: the mean-field regime.** The numerical efficiency discussed above is encouraging, but one should be careful when comparing two Markov processes for sampling. Indeed, suppose we are given two kinetic processes  $(X_t, V_t)_{t \in [0, T]}$  and  $(\tilde{X}_t, \tilde{V}_t)_{t \in [0, T]}$ , both  $\mu$ -ergodic, such that simulating a trajectory of length  $T$  for the first one is 10 times faster than for the second one. This is useless if the first process explores the space 100 times slower than the second one so that, to achieve a similar quality of sampling, much longer trajectories are required.

Such a situation may be feared for the hybrid drift/jump process introduced above in the Lennard-Jones model. Indeed, under the effect of many collisions, the process may show as  $M \rightarrow \infty$  a diffusive behaviour, namely the jumps may average and give a velocity close to zero. In that case, the exploration of the space (hence the convergence toward equilibrium) would be very slow. Let us (informally) check that it is not the case here, at least in the mean-field regime: in the following, we suppose that  $U_0 = 1/M$  (in which case the constant  $C_R$  is of order  $1/M$ ).

Consider the Langevin dynamics with generator  $L = A_1 + A_2 + A_4$  with  $F_0 = \nabla U$ . Suppose that the initial conditions  $(x_i, v_i)_{i \in \llbracket 1, M \rrbracket}$  are i.i.d. with the  $x_i$ 's distributed according to some law  $\nu$ . Conditionally to  $(x_1, v_1)$ , by the law of large numbers, the force felt by the first particle at time 0 satisfies

$$\nabla_{x_1} U(x) = \nabla_{x_1} \frac{1}{M} \sum_{j=2}^M W(|x_i - x_j|) \xrightarrow{M \rightarrow +\infty} \nabla_{x_1} \int_{(a\mathbb{T})^3} W(|x_1 - z|) \nu(dz).$$

Classical propagation of chaos results show that, as  $M \rightarrow +\infty$ , the particles behaves approximately as  $M$  independent processes on  $(a\mathbb{T})^3 \times \mathbb{R}^3$  with generator

$$L_t f = v \cdot \nabla_x f - \nabla_x \left( \int_{(a\mathbb{T})^3} W(|x - z|) \nu_t(dz) \right) \cdot \nabla_v f - v \cdot \nabla_v f + \Delta_v f,$$

where  $\nu_t$  is the law of the process at time  $t$ , solution of the non-linear equation  $\partial_t(\nu_t f) = \nu_t L_t f$  for all nice  $f$ .

Similarly, for the hybrid jump/diffusion process obtained from the splitting of the forces introduced in Section 5.3.1, conditionally to  $(x_1, v_1)$ , by the law of large numbers, denoting  $g$  the function such that  $G_{i,j}(x) = g(x_i - x_j)/M$  for  $i, j \in \llbracket 1, M \rrbracket$ , the jump rate at time 0 for the first particle satisfies

$$\frac{1}{M} \sum_{j=2}^M (v_1 \cdot G_{1,j}(x))_+ \xrightarrow{M \rightarrow \infty} \int_{(a\mathbb{T})^3} (v_1 \cdot g(|x_1 - z|))_+ \nu(dz) := \lambda_\nu(x_1, v_1),$$

and a similar limit  $Q_\nu$  is obtained for the jump kernel at time 0. From the propagation of chaos phenomenon, the system is expected to behave as  $M \rightarrow \infty$  as  $M$  independent non-linear processes on  $(a\mathbb{T})^3 \times \mathbb{R}^3$  with non-homogeneous generators given by

$$L_t f = v \cdot \nabla_x f - \left( \int_{(a\mathbb{T})^3} \tilde{g}(x-z) \nu_t(dz) \right) \cdot \nabla_v f + \lambda_{\nu_t} (Q_{\nu_t} f - f) - v \cdot \nabla_v f + \Delta_v f$$

with a suitable function  $\tilde{g}$ , and  $\nu_t$  the law of the process (see [Monmarché, 2018](#) and references within). The non-homogeneous Markov process with generator  $L_t$  is not degenerated in the sense that its velocity is not averaged to zero, which means that for large  $M$  the system of interacting particles is not expected to have a diffusive behaviour, nor to converge to equilibrium with a rate that vanishes as  $M \rightarrow +\infty$ .

Of course the dynamics (Langevin versus hybrid) are different and thus they may converge toward equilibrium at different speed. Nevertheless, the ratio of the convergence rate may be expected to be of order independent from  $M$ , since the mean-field limits as  $M \rightarrow \infty$  are obtained with the same scaling (namely it was not necessary to accelerate time). In other words, in order to achieve the same convergence as the Langevin process in a time  $T$  (for a cost of order  $\mathcal{O}(M^2 T / \delta)$ ), we may need to sample the hybrid process up to a time  $T' > T$  (for a cost of order  $\mathcal{O}(MN(R)T' / \delta + M^2 C_R T')$ ) but with a ratio  $T'/T$  bounded below independently from  $M$ . For  $R$  small enough (independent from  $M$ ) so that  $\mathcal{N}(R) < (T'/T)M$ , using that  $C_R$  is  $\mathcal{O}(1/M)$ , we see that the hybrid process appears as the most efficient for large  $M$ .

More generally, if  $U_0$  is not of order  $1/M$ , the conclusion may be unclear, but the informal discussion above shows that there exist cases where the parameters of the problem are such that a suitable factorization can lower the numerical cost of the simulation at constant long-time convergence quality. We refer to [Monmarché et al. \(2020\)](#) for a practical case where the diffusion constant, used as an indicator of the sampling rate, is of the same order for the two processes.

**5.3.5. An alternative splitting of the forces.** For the Lennard-Jones model, consider the splitting of the forces  $\nabla U = \sum_{i=0}^M F_i$  where  $F_0$  is as above and  $F_i(x) = J_i \nabla_{x_i} U = J_i \sum_{j \neq i} G_{i,j}$  for  $i \in \llbracket 1, M \rrbracket$ . Since  $(v_i \cdot \sum_{j \neq i} G_{i,j})_+ \leq \sum_{j=1}^M (v_i \cdot G_{i,j})_+$ , this alternative splitting reduces the total jump rate of the process, by comparison with the previous splitting. Increasing the jump rate is known to increase the diffusive behaviour and the asymptotical variance for the Zig-Zag process, at least in dimension 1 ([Bierkens and Duncan, 2017](#)). Since the present case is similar, we may expect the alternative splitting to yield a process that converges faster toward equilibrium than the initial one. Nevertheless, in the mean-field case, similarly to the previous section, this speed-up should be independent from  $M$ .

On the other hand, let us estimate the numerical cost (in term of computations of forces) of the alternative process. We can bound  $|F_i(x)| \leq M C_R$  (with the same  $C_R$  as previously), and thus a thinning algorithm yields again an average of  $M C_R T$  jump proposals per particle in a time period  $[0, T]$ . But now, at each jump proposal,  $F_i(x)$  has to be computed, which has a cost  $\mathcal{O}(M)$ , so the total cost is  $\mathcal{O}(M^3 C_R T)$ .

This little computation indicates that, in the case of forces that come from pairwise interactions between particles, rather than considering one jump mechanisms per particle (as in the usual Zig-Zag process, and in the alternative splitting of

this section), it may be better to split  $\nabla_{x_i} U$  so that each jump mechanism is only associated to a single interaction  $(i, j)$  (as in the initial splitting introduced in Section 5.3.1).

### Acknowledgements

The author would like to thank Gabriel Stoltz and Louis Lagardère for fruitful discussions about Strang splitting schemes, and the anonymous referee for his useful remarks. He acknowledges partial financial support from the ERC grant MS-MATHS (European Union’s Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement number 614492) and the ERC grant EMC2 (European Union’s Horizon 2020 Research and Innovation Programme, Grant Agreement number 810367), and from the French ANR grant EFI (Entropy, flows, inequalities, ANR-17-CE40-0030).

### References

- Balakrishnan, V. and Chaturvedi, S. Persistent diffusion on a line. *Phys. A*, **148** (3), 581–596 (1988). [MR939154](#).
- Bernard, E. P., Krauth, W., and Wilson, D. B. Event-chain Monte Carlo algorithms for hard-sphere systems. *Phys. Rev. E*, **80**, 056704 (2009). DOI: [10.1103/PhysRevE.80.056704](#).
- Bierkens, J. and Duncan, A. Limit theorems for the zig-zag process. *Adv. in Appl. Probab.*, **49** (3), 791–825 (2017). [MR3694318](#).
- Bierkens, J., Fearnhead, P., and Roberts, G. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.*, **47** (3), 1288–1320 (2019a). [MR3911113](#).
- Bierkens, J. and Roberts, G. A piecewise deterministic scaling limit of lifted Metropolis-Hastings in the Curie-Weiss model. *Ann. Appl. Probab.*, **27** (2), 846–882 (2017). [MR3655855](#).
- Bierkens, J., Roberts, G. O., and Zitt, P.-A. Ergodicity of the zigzag process. *Ann. Appl. Probab.*, **29** (4), 2266–2301 (2019b). [MR3983339](#).
- Bou-Rabee, N. Time integrators for molecular dynamics. *Entropy*, **16** (1), 138–162 (2014). DOI: [10.3390/e16010138](#).
- Bouchard-Côté, A., Vollmer, S. J., and Doucet, A. The bouncy particle sampler: a nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.*, **113** (522), 855–867 (2018). [MR3832232](#).
- Bouin, E., Dolbeault, J., Mischler, S., Mouhot, C., and Schmeiser, C. Hypocoercivity without confinement. *ArXiv Mathematics e-prints* (2017). [arXiv: 1708.06180](#).
- Chen, A. Y. and Renshaw, E. The general correlated random walk. *J. Appl. Probab.*, **31** (4), 869–884 (1994). [MR1303919](#).
- Diaconis, P., Holmes, S., and Neal, R. M. Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.*, **10** (3), 726–752 (2000). [MR1789978](#).
- Diaconis, P. and Miclo, L. On the spectral analysis of second-order Markov chains. *Ann. Fac. Sci. Toulouse Math. (6)*, **22** (3), 573–621 (2013). [MR3113027](#).
- Durmus, A., Guillin, A., and Monmarché, P. Geometric ergodicity of the bouncy particle sampler. *ArXiv Mathematics e-prints* (2018a). [arXiv: 1807.05401](#).

- Durmus, A., Guillin, A., and Monmarché, P. Piecewise Deterministic Markov Processes and their invariant measure. *ArXiv Mathematics e-prints* (2018b). [arXiv: 1807.05421](https://arxiv.org/abs/1807.05421).
- Durmus, A. and Moulines, E. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, **27** (3), 1551–1587 (2017). [MR3678479](https://doi.org/10.1214/15-AAP1187).
- Fontbona, J., Guérin, H., and Malrieu, F. Long time behavior of telegraph processes under convex potentials. *Stochastic Process. Appl.*, **126** (10), 3077–3101 (2016). [MR3542627](https://doi.org/10.1016/j.spa.2016.07.007).
- Gibson, D. A. and Carter, E. A. Time-reversible multiple time scale ab initio molecular dynamics. *J. Phys. Chem.*, **97** (51), 13429–13434 (1993). [DOI: 10.1021/j100153a002](https://doi.org/10.1021/j100153a002).
- Glynn, P. W. and Meyn, S. P. A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.*, **24** (2), 916–931 (1996). [MR1404536](https://doi.org/10.1214/aop/1023619407).
- Goldstein, S. On diffusion by discontinuous movements, and on the telegraph equation. *Quart. J. Mech. Appl. Math.*, **4**, 129–156 (1951). [MR47963](https://doi.org/10.1093/qjmath/4.3.129).
- Gruber, U. *Convergence of binomial large investor models and general correlated random walks*. Ph.D. thesis, Technical University of Berlin (2004). [http://www.edocs.tu-berlin.de/diss/2004/gruber\\_urs.htm](http://www.edocs.tu-berlin.de/diss/2004/gruber_urs.htm).
- Gruber, U. and Schweizer, M. A diffusion limit for generalized correlated random walks. *J. Appl. Probab.*, **43** (1), 60–73 (2006). [MR2225050](https://doi.org/10.1111/j.1467-9892.2006.01505.x).
- Hadeler, K. P. Travelling fronts for correlated random walks. *Canad. Appl. Math. Quart.*, **2** (1), 27–43 (1994). [MR1271437](https://doi.org/10.1002/cma.1001).
- Hairer, E., Lubich, C., and Wanner, G. Geometric numerical integration illustrated by the Störmer-Verlet method. *Acta Numer.*, **12**, 399–450 (2003). [MR2249159](https://doi.org/10.1017/S1446788703000017).
- Hairer, M. and Mattingly, J. C. Yet another look at Harris’ ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63 of *Progr. Probab.*, pp. 109–117. Birkhäuser/Springer Basel AG, Basel (2011). [MR2857021](https://doi.org/10.1007/978-3-03-061221-1_10).
- Herrmann, S. and Vallois, P. From persistent random walk to the telegraph noise. *Stoch. Dyn.*, **10** (2), 161–196 (2010). [MR2652885](https://doi.org/10.1142/S021959981000161).
- Kac, M. A stochastic model related to the telegrapher’s equation. *Rocky Mountain J. Math.*, **4**, 497–509 (1974). [MR510166](https://doi.org/10.1215/RMJ.1974.4.497).
- Kallenberg, O. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York (1997). ISBN 0-387-94957-7. [MR1464694](https://doi.org/10.1007/978-1-4612-0000-9).
- Leimkuhler, B. and Matthews, C. Rational construction of stochastic numerical methods for molecular sampling. *Appl. Math. Res. Express. AMRX*, (1), 34–56 (2013). [MR3040887](https://doi.org/10.1016/j.amrx.2013.06.001).
- Leimkuhler, B., Matthews, C., and Stoltz, G. The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics. *IMA J. Numer. Anal.*, **36** (1), 13–79 (2016). [MR3463433](https://doi.org/10.1093/iman/36.1.13).
- Lelièvre, T. and Stoltz, G. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numer.*, **25**, 681–880 (2016). [MR3509213](https://doi.org/10.1017/S1446788716000017).
- Lemaire, V., Thieullen, M., and Thomas, N. Exact simulation of the jump times of a class of piecewise deterministic Markov processes. *J. Sci. Comput.*, **75** (3), 1776–1807 (2018). [MR3798122](https://doi.org/10.1007/s11222-018-0812-2).

- Maigret, N. Théorème de limite centrale fonctionnel pour une chaîne de Markov récurrente au sens de Harris et positive. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, **14** (4), 425–440 (1979) (1978). [MR523221](#).
- Mandelbaum, M., Hlynka, M., and Brill, P. H. Nonhomogeneous geometric distributions with relations to birth and death processes. *TOP*, **15** (2), 281–296 (2007). [MR2361318](#).
- Mattingly, J. C., Stuart, A. M., and Tretyakov, M. V. Convergence of numerical time-averaging and stationary measures via Poisson equations. *SIAM J. Numer. Anal.*, **48** (2), 552–577 (2010). [MR2669996](#).
- Michel, M., Durmus, A., and Sénécal, S. Forward Event-Chain Monte Carlo: Fast sampling by randomness control in irreversible Markov chains. *ArXiv Mathematics e-prints* (2017a). [arXiv: 1702.08397](#).
- Michel, M., Tan, X., and Deng, Y. Clock Monte Carlo methods. *ArXiv Mathematics e-prints* (2017b). [arXiv: 1706.10261](#).
- Miclo, L. and Monmarché, P. Étude spectrale minutieuse de processus moins indécis que les autres. In *Séminaire de Probabilités XLV*, volume 2078 of *Lecture Notes in Math.*, pp. 459–481. Springer, Cham (2013). [MR3185926](#).
- Monmarché, P. Piecewise deterministic simulated annealing. *ALEA Lat. Am. J. Probab. Math. Stat.*, **13** (1), 357–398 (2016). [MR3487077](#).
- Monmarché, P. Weakly self-interacting velocity jump processes for bacterial chemotaxis and adaptive algorithms. *Markov Process. Related Fields*, **23** (4), 609–659 (2017). [MR3754143](#).
- Monmarché, P. Elementary coupling approach for non-linear perturbation of Markov processes with mean-field jump mechanisms and related problems. *ArXiv Mathematics e-prints* (2018). [arXiv: 1809.10953](#).
- Monmarché, P., Weisman, J., Lagardère, J.-P., and Piquemal, L. Velocity jump processes : an alternative to multi-timestep methods for faster and accurate molecular dynamics simulations. *ArXiv Mathematics e-prints* (2020). [arXiv: 2002.07109](#).
- Peters, E. A. J. F. and de With, G. Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E*, **85**, 026703 (2012). DOI: [10.1103/PhysRevE.85.026703](#).
- Ricci, A. and Ciccotti, G. Algorithms for Brownian dynamics. *Molecular Physics*, **101** (12), 1927–1931 (2003). DOI: [10.1080/0026897031000108113](#).
- Roberts, B. Partite saturation problems. *J. Graph Theory*, **85** (2), 429–445 (2017). [MR3647819](#).
- Rossetto, V. The one-dimensional asymmetric persistent random walk. *J. Stat. Mech. Theory Exp.*, (4), 043204, 21 (2018). [MR3801956](#).
- Talay, D. and Tubaro, L. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, **8** (4), 483–509 (1991) (1990). [MR1091544](#).
- Tuckerman, M. E., Berne, B. J., and Rossi, A. Molecular dynamics algorithm for multiple time scales: Systems with disparate masses. *The Journal of Chemical Physics*, **94** (2), 1465–1469 (1991). DOI: [10.1063/1.460004](#).
- Turitsyn, K. S., Chertkov, M., and Vucelja, M. Irreversible Monte Carlo algorithms for efficient sampling. *Physica D: Nonlinear Phenomena*, **240** (4), 410–414 (2011). DOI: [10.1016/j.physd.2010.10.003](#).
- Vanetti, P., Bouchard-Côté, A., Deligiannidis, G., and Doucet, A. Piecewise Deterministic Markov Chain Monte Carlo. *ArXiv Mathematics e-prints* (2017). [arXiv: 1707.05296](#).

---

Zauderer, E. Correlated random walks, hyperbolic systems and Fokker-Planck equations. *Mathematical and Computer Modelling*, **17** (10), 43 – 47 (1993). DOI: [10.1016/0895-7177\(93\)90116-G](https://doi.org/10.1016/0895-7177(93)90116-G).