Piecewise deterministic simulated annealing

Pierre Monmarché

Université de Neuchâtel
11 rue Emile Argand,
Neuchâtel, Switzerland.
E-mail address: `pierre.monmarche@unine.ch`
URL: `http://members.unine.ch/pierre.monmarche/`

Abstract. Given an energy potential on the Euclidian space, a piecewise deterministic Markov process is designed to sample the corresponding Gibbs measure. In dimension one an Eyring-Kramers formula is obtained for the exit time of the domain of a local minimum at low temperature, and a necessary and sufficient condition is given on the cooling schedule in a simulated annealing algorithm to ensure the process converges to the set of global minima. This condition is similar to the classical one for diffusions and involves the critical depth of the potential. In higher dimensions a non optimal sufficient condition is obtained.

1. Introduction and main results

1.1. *Simulated annealing.* The simulated annealing algorithm is a classical stochastic optimization algorithm, which can be seen as a descent algorithm perturbed by random locally counter-productive moves to escape from non-global minima. More precisely, consider an energy potential U on the Euclidian space such that $e^{-U} \in L^1(\mathbb{R}^d)$. The mass of the Gibbs law associated to $\frac{1}{\varepsilon}U$,

$$\mu_\varepsilon(dx) = \frac{e^{-\frac{U(x)}{\varepsilon}}}{\int_{\mathbb{R}^d} e^{-\frac{U(y)}{\varepsilon}} dy} dx,$$

concentrates on the set of global minima of U as $\varepsilon > 0$, called the temperature, goes to zero. At fixed temperature, μ_ε can be sampled by a Markov Monte-Carlo procedure, namely it can be approximated by the law of X_T^ε where $(X_t^\varepsilon)_{t \geq 0}$ is an ergodic Markov process whose invariant law is μ_ε and T is large enough so that the process is close to equilibrium at time T. A usual choice for X^ε would be the

Fokker-Planck diffusion, that is the solution of

$$dX_t^\varepsilon \quad = \quad -\nabla U\left(X_t^\varepsilon\right) dt + \sqrt{2\varepsilon} dB_t, \tag{1.1}$$

where B is a Brownian motion. In the simulated annealing algorithm the temperature ε_t decays over time, so that the solution of this stochastic differential equation becomes an inhomogeneous Markov process $(X_t^{\varepsilon_t})_{t\geq 0}$. If the system cools slowly, *i.e.* if $t \mapsto \varepsilon_t$ goes to zero slowly enough, then the process has enough time to explore the space and approach equilibrium, so that the law of $X_t^{\varepsilon_t}$ gets and stays close to its "instantaneous" invariant law, μ_{ε_t}. In particular, the mass of the law of $X_t^{\varepsilon_t}$ goes to the set of the global minima of U as t goes to infinity. However, if the system is abruptly frozen, in other word if the decay $t \mapsto \varepsilon_t$ is too fast, the process will have a non-zero probability to be trapped in local minima.

There is a broad literature on this question, both theoretical and practical, and we refer to Kalivas et al. (1995) for an introduction. A key phenomenon in the analysis of this algorithm (and of many stochastic algorithms, indeed) is metastability (cf. Lelièvre, 2013): at low temperature, the process spends a lot of time trapped in a neighbourhood of local minima, so that mixing - *i.e.* convergence to equilibrium (at fixed temperature) - is very slow. This yields extremely slow theoretical cooling schedules $t \mapsto \varepsilon_t$ to ensure the process converges in probability to the set of global minima of U. For instance in the Fokker-Planck diffusion case, it is well-known (see Holley et al. (1989) among others) that in order for the process to converge in probability to any neighbourhood of the global maxima of U, ε_t should be of order at least $\frac{E^*}{\ln(1+t)}$ where E^* is the critical depth of the potential, a constant that depends on U which will be defined below.

One line of inquiry to improve the algorithm is then to look for other dynamics than (1.1), which would have more inertia and thus would escape more easily from local traps. One of the main example is the kinetic Langevin dynamics, studied in Scemama et al. (2006); Monmarché (2015). Among other possibilities, the reversible dynamics (1.1) is perturbed in Lelièvre et al. (2013) with a divergence-free drift; or processes with more general memories than kinetic ones are considered in Gadat and Panloup (2014).

In this work we propose a Markov process $(X_t, Y_t)_{t\geq 0}$ on $\mathbb{R}^d \times \mathbb{S}^{d-1}$ with the following properties :

- $dX_t = Y_t dt$.
- the process is ergodic (in the sense that for all initial condition, its law converges to a unique invariant law μ), and the first marginal of μ is prescribed as $e^{-U(x)} dx$.
- Y_t is a jump process.

The first property means (X_t, Y_t) is a kinetic process, such as the kinetic Langevin one, and we call X the position and Y the velocity. The second one means it serves its intended purpose. Finally the fact that Y_t is a jump process on the sphere makes the whole process very easy to sample on a computer. More precisely between two times of jump of Y_t, (X_t, Y_t) is completely deterministic. Such piecewise deterministic Markov processes (PDMP) have recently attracted much attention in various fields, since they are a simple alternative to diffusions to model stochastic systems (see Azaïs et al. (2013) and references within for an overview).

In dimension greater than 1 there are many ways to meet the above requirements and thus in a first instance we will focus on the case $d = 1$, for which the possibilities are more limited.

1.2. *The one-dimensional process.* Our aim is to define a Markov process (X_t, Y_t) with the properties above. In dimension one, the velocity is either 1 or -1. If the process goes twice through the same state (x, y), necessarily it has had to make a U-turn in the meanwhile and come back the other way, hence to visit $(x, -y)$. On average (in time), $(x, 1)$ and $(x, -1)$ are thus equally visited. Ergodicity then implies the invariant law is necessarily a product measure whose second marginal is uniform on $\{\pm 1\}$.

Recall that the semi-group $(P_{s,t})_{t \geq s \geq 0}$ and the infinitesimal generator $(L_t)_{t \geq 0}$ associated to the process are defined as

$$P_{s,t} f(x, y) = \mathbb{E}\left[f(X_t, Y_t) \mid (X_s, Y_s) = (x, y) \right]$$

for all bounded f and

$$L_t f = \lim_{h \to 0} \frac{P_{t,t+h} f - f}{h}$$

whenever this limit (say in the uniform norm sense) exists. When the process is homogeneous (in our case, it means when the temperature is constant, which is the case for now), we only write $P_t = P_{0,t}$ and $L = L_t$. We will describe the dynamics through the generator.

In dimension one, the only possibility when a jump occurs is to transform the velocity into its opposite. This yields an infinitesimal generator of the form

$$L f(x, y) = y f'(x, y) + \lambda(x, y) \left(f(x, -y) - f(x, y) \right) \tag{1.2}$$

where the rate of jump λ is a non-negative function. In this case, the law μ is invariant iff $\int L f \, d\mu = 0$ for all f. This is equivalent to

$$y U'(x) = \lambda(x, y) - \lambda(x, -y),$$

which is satisfied if and only if λ is of the form $\lambda(x, y) = \frac{1}{2} \left(y U'(x) + a(x) \right)$ for some function $a(x)$ (note that necessarily $a(x) = \lambda(x, y) + \lambda(x, -y)$). The non-negativity of λ implies $a(x) \geq |U'(x)|$. When this is indeed an equality, $\lambda(x, y) = (y U'(x))_+$ (where $(g)_+$ denotes the positive part of g, equal to g if $g > 0$ and 0 else): this is the choice that minimizes the rate of jump, namely the dissipative behaviour of the process. On the other hand it is convenient from the simulation viewpoint. Indeed, it implies that while $Y_t U'(X_t) \leq 0$, or in other words while the process is going down the potential, no jump is allowed. On the contrary if $Y_t U'(X_t) > 0$ the next time T of jump will be such that

$$E := \int_0^T y U'(x + ys) ds = U(x + yT) - U(x)$$

has an exponential law with mean 1, which we denote by $\mathcal{E}(1)$. Thus we only need to compute the potential along the trajectory, and to simulate a Poisson process. More precisely, let $(E_k)_{k \geq 0}$ be a family of i.i.d. variables with exponential law and set $(X_0, Y_0) = (x_0, y_0)$ and $T_0 = 0$. Suppose the process is already constructed up to a jump time T_i, $i \geq 0$, and is independent from $(E_k)_{k \geq i}$ up to T_i. Let

$$T_{i+1} = \inf \left\{ t > T_i, \int_{T_i}^t \lambda \left(X_{T_i} + Y_{T_i}(s - T_i), Y_{T_i} \right) ds \geq E_i \right\} \tag{1.3}$$

$$\begin{aligned}
X_s &= X_{T_i} + Y_{T_i}(t - T_i) && \text{for } s \in [T_i, T_{i+1}] \\
Y_s &= Y_{T_i} && \text{for } s \in [T_i, T_{i+1}) \\
Y_{T_{i+1}} &= -Y_{T_i}.
\end{aligned}$$

This define a Markov process (X_t, Y_t) with generator (1.2).

In the literature such a process, which belongs to the larger class of switched PDMP (Benaïm et al., 2012) goes sometimes by the name of (integrated) telegraph process (Ratanov, 2015; Monmarché, 2014). It can be seen as the continuous limit of persistent walks (Diaconis and Miclo, 2013; Miclo and Monmarché, 2013), which were already studied as a possible alternative to reversible walks to sample discrete Gibbs measures, and it is reminiscent of the so-called Hit-and-Run sampler (Bélisle et al., 1993). It has already been studied to model the motion of the bacterium *Escheria coli* (Calvez et al., 2015; Fontbona et al., 2012; Erban and Othmer, 2004/05) and is called in this context a velocity jump process, which is the name we are going to use since it still makes sense in a metastable context and in dimension greater than 1 (see Section 1.4).

1.3. *Main results in dimension one.* As a first step we will consider the velocity jump process $(X_t^\varepsilon, Y_t^\varepsilon)_{t \geq 0}$ on $\mathbb{R} \times \{\pm 1\}$ at low (but fixed) temperature $\varepsilon > 0$, namely with generator

$$Lf(x, y) = yf'(x, y) + \frac{1}{\varepsilon}(yU'(x))_+ (f(x, -y) - f(x, y)). \qquad (1.4)$$

We want to understand how long it takes for the process to escape from a local minimum. Let U be a double-well potential, namely a Morse function with two local minima, denoted by $x_0 \leq x_2$, a local maximum $x_1 \in (x_0, x_2)$, and which is convex outside (x_0, x_2) and goes to infinity as $|x| \to \infty$. Recall that U is said to be a Morse function if all its critical points are non-degenerate; in other words if $(U'(x) = 0) \Rightarrow (U''(x) \neq 0)$. Suppose $(X_0^\varepsilon, Y_0^\varepsilon) = (x_0, -1)$, and let

$$\tau = \inf\{t > 0, X_t^\varepsilon = x_1\}$$

be the first hitting time of x_1 (note that, contrary to a diffusion which may fall back, when X^ε reaches x_1 it deterministically leaves $[x_0, x_1]$ and falls down to x_2). Then the energy barrier to overcome in order to leave $[x_0, x_1]$ is $U(x_1) - U(x_0)$. We will prove what is usually called an Eyring-Kramers formula (or an Arrhenius law):

Theorem 1.1. *For the velocity jump process with generator* (1.4) *starting at* $(X_0^\varepsilon, Y_0^\varepsilon) = (x_0, -1)$ *in the double-well potential* U,

$$\mathbb{E}[\tau] = \sqrt{\frac{8\pi\varepsilon}{U''(x_0)}} e^{\frac{U(x_1) - U(x_0)}{\varepsilon}} \left(1 + \underset{\varepsilon \to 0}{o}(1)\right),$$

$$\mathbb{P}(\tau \geq t\mathbb{E}[\tau]) \underset{\varepsilon \to 0}{\longrightarrow} e^{-t}.$$

This can be compared to the case of a Fokker-Planck diffusion Z_t^ε with generator

$$L_{di}f(x) = -U'(x)f'(x) + \varepsilon f''(x),$$

which has been studied in much more general settings. Let $\eta > 0$ be small and

$$\tau_{di} = \inf\{t > 0, Z_t^\varepsilon = x_1 + \eta\}.$$

The work of Bovier et al. (2004, 2005) applies here and yields:

Theorem 1.2 (from Bovier et al., 2004, 2005). *For the Fokker-Planck diffusion starting at $Z_0^\varepsilon = x_0$ in the double-well potential U,*

$$\mathbb{E}\left[\tau_{di}\right] \quad = \quad \frac{2\pi e^{\frac{U(x_1)-U(x_0)}{\varepsilon}}}{\sqrt{|U''(x_1)|U''(x_0)}}\left(1 + \underset{\varepsilon \to 0}{o}(1)\right),$$

$$\mathbb{P}\left(\tau_{di} \geq t\mathbb{E}\left[\tau_{di}\right]\right) \quad \underset{\varepsilon \to 0}{\longrightarrow} \quad e^{-t}.$$

Remarks:

- Both processes samples the same Gibbs law. To sample the diffusion, one needs to generate a Brownian motion, while the PDMP can be constructed from a sequence of independent exponential variables $(E_k)_{k \geq 0}$.
- The velocity jump process moves at constant speed. In particular it takes a constant time (constant in the sense it does not depends on ε) to pass through the interval $(x_1 - \eta, x_1 + \eta)$, whose probability under the law μ_ε is of order

$$\exp\left(-\frac{1}{\varepsilon}\left(\min(U(x_1 - \eta), U(x_1 + \eta)) - \min\left(U(x_0), U(x_2)\right)\right)\right).$$

  Since the ratio between the average times spent in this interval and outside of it should be of this order, it means the time between two crossing of $(x_1 - \eta, x_2 + \eta)$ needs to be of order the inverse of this probability, which explains the exponential factor of $\mathbb{E}\left[\tau\right]$ had to be expected.
- The fact $U''(x_1)$ does not appear in the PDMP case is also natural. Indeed, the probability that the process starting at $(x_0, 1)$ reaches x_1 in one shot (*i.e.* before coming back to $(x_0, -1)$) depends only on $U(x_1) - U(x_0)$, and not on the local geometry of the potential near x_1. On the contrary the process stays mainly in the neighbourhood of x_0, so that $U''(x_0)$ does intervene. If U were flat, for instance $U = d\left(1 - \mathbb{1}_{[x_0-1,x_0+1]}\right)$ for some $d > 0$, an adaptation of the proof of Theorem 1.1 would yield $\mathbb{E}\left[\tau\right] \simeq 2e^{\frac{d}{\varepsilon}}$.
- Both $\mathbb{E}\left[\tau\right]$ and $\mathbb{E}\left[\tau_{di}\right]$ have the same exponential order.

With Theorem 1.1 in mind we then turn to the study of the inhomogeneous process $(X_t^{\varepsilon_t}, Y_t^{\varepsilon_t})$ with generator

$$L_t f(x,y) \quad = \quad yf'(x,y) + \frac{1}{\varepsilon_t}\left(yU'(x)\right)_+\left(f(x,-y) - f(x,y)\right). \tag{1.5}$$

Assumption 1.

- The cooling schedule $t \mapsto \varepsilon_t > 0$ is non-increasing and goes to 0.
- The potential U on $\mathbb{R}$ is a smooth Morse function with a finite number of local extrema (one of which at least is a non-global minimum), unbounded and convex at infinity.

We say that z is reachable from x at height V if $\max\{U\left(x + t(z - x)\right), \ t \in [0,1]\} \leq V$ and we call the depth of a local minimum x the smallest V such that there exist a z with $U(z) < U(x)$ which is reachable from x at height $U(x) + V$ (the depth of a global minimum is set to $+\infty$). The critical depth of U, denoted by E^*, is then defined as the maximal among the depths of all local minima of U which are not global minima (see Fig. 1.1).

Adapting to our settings the work of Hajek (1988) on simulated annealing on a discrete space, we will prove the following :
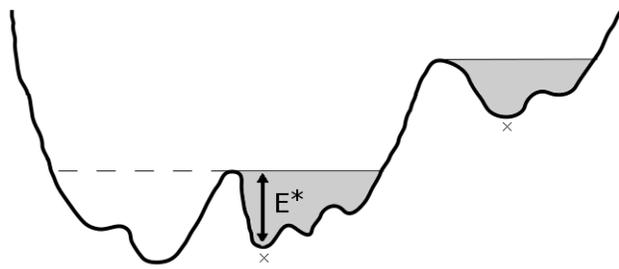
FIGURE 1.1. The depth of a local minima, and the critical depth.

Theorem 1.3. *Let U and $(\varepsilon_t)_{t\geq 0}$ satisfy Assumption 1 and consider the process $(X_t^{\varepsilon_t}, Y_t^{\varepsilon_t})$ with generator (1.5) and any initial condition. We have the following:*

(1) *If S is a neighbourhood of all local minima of U, then*

$$\lim_{t\to\infty} \mathbb{P}\left(X_t^{\varepsilon_t} \in S\right) = 1.$$

(2) *If S is a neighbourhood of all local minima of depth E, and its complementary S^c is a neighbourhood of all other local minima,*

$$\lim_{t\to\infty} \mathbb{P}\left(X_t^{\varepsilon_t} \in S\right) = 0 \qquad \Leftrightarrow \qquad \int_0^\infty (\varepsilon_s)^{-\frac{1}{2}}\, e^{-\frac{E}{\varepsilon_s}}\, ds = \infty$$

(3) *As a consequence,*

$$\forall \delta > 0 \ \lim_{t\to\infty} \mathbb{P}\left(U\left(X_t^{\varepsilon_t}\right) < \min_{\mathbb{R}} U + \delta\right) = 1 \qquad \Leftrightarrow \qquad \int_0^\infty (\varepsilon_s)^{-\frac{1}{2}}\, e^{-\frac{E^*}{\varepsilon_s}}\, ds = \infty$$

In particular if the cooling schedule is of the parametric form

$$\varepsilon_t = \frac{c}{\ln(1+t)}$$

with $c > 0$, the algorithm succeed (*i.e.* the process converges in probability to any neighbourhood of the global minima of U) if and only if $c \geq E^*$.

It is somehow a negative result: it means this velocity jump process does not allow faster cooling schedules than the classical reversible diffusion one. On the other hand it is somehow a positive result, since it allows cooling schedule as fast as the diffusion *and* it is easier and faster to compute numerically. Finally, positive or negative, it is above all a theoretical result. In practice the simulation is done in a finite time horizon; a context where the theoretical logarithmic schedules are far from efficient. The next step of the study should thus be to give non-asymptotic results in the spirit of the work of Catoni (1992).

Before proceeding to the proofs, here is a remark about our method. The extreme simplicity of the motion permits an elementary analysis. We could have tried to use functional inequalities tools, such as in Holley et al. (1989); Miclo (1992) (see also Ané et al. (2000) for a general introduction). Indeed, since U is assumed to be convex at infinity, it is known the associated Gibbs measure satisfies a spectral gap (or Poincaré) inequality, with a well-understood asymptotic of the constant. But since the *carré du champs* of the velocity jump process is not the square of the gradient, it is not clear whether it gives any information on the way the process

relaxes to equilibrium. This is a typical problem in the field of hypocoercivity, and indeed our PDMP has been studied in Calvez et al. (2015); Monmarché (2014) from this viewpoint. But in both works the rate of jump is assumed to be bounded from below by a positive constant, which means there is at all time a residual randomness. This is not the case with our minimal choice $\lambda(x, y) = (yU'(x))_+$ and in this sense our process is quite degenerate among degenerate processes.

However, if on the one hand some ideas are missing to treat this degenerate situation with hypocoercive tools, on the other hand, hopefully, once precisely understood thanks to elementary analysis, this process may be a good benchmark to investigate several hypocoercive questions, such as the relationships between functional inequalities, gradient estimates and Wasserstein convergence.

Finally note that, even if the pathwise strategy we will adapt from Hajek is very close to the Freidlin and Wentzell (2012) approach, the latter theory would yield slightly less precise results since for continuous-time processes it only deals with the large deviation scaling (namely with the asymptotic of quantities such as $\frac{1}{t} \ln \mathbb{P}(X_t^\varepsilon \in A)$ rather than $\mathbb{P}(X_t^\varepsilon \in A)$).
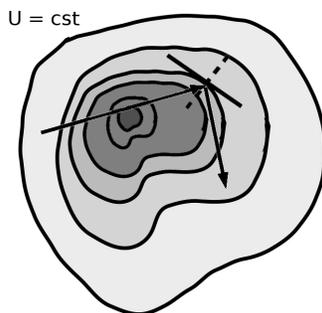

FIGURE 1.2. At a jump time the process bounces off the level sets of U.

1.4. *Definition and results in any dimension.* The interest of the simulated annealing algorithm appears in large dimension, and so we now define a suitable piecewise deterministic Gibbs sampler in this context. We call velocity jump process the Markov process on $\mathbb{R}^d \times \mathbb{S}^{d-1}$ whose generator is

$$
\begin{aligned}
Lf(x, y) = \ & y.\nabla_x f(x, y) + (y.\nabla_x U(x))_+ (f(x, y^*) - f(x, y)) \\
& + r \left(\int f(x, z) dz - f(x, y) \right),
\end{aligned}
$$

where $r > 0$ is a parameter, dz denotes the uniform measure on $\mathbb{S}^{d-1}$ and

$$
y^* = y - 2 \left(y.\frac{\nabla_x U(x)}{|\nabla_x U(x)|} \right) \frac{\nabla_x U(x)}{|\nabla_x U(x)|}
$$

(the explanations of this definition is postponed to the end of this section). As we will see, the measure $e^{-U(x)} dx \times dy$ is invariant for this process. A trajectory is defined in a similar manner that in dimension 1, except that there are now two

different clocks:

$$T = \inf\left\{t > 0, \int_0^t \left(Y_0.\nabla U\left(X_0 + Y_0 s\right)\right)_+ ds \geq E_1\right\},$$

$$S = \frac{1}{r}E_2,$$

where E_1, E_2 are independent standard exponential random variables. The process evolves deterministically according to $dX = Y$ and $dY = 0$ up to time $T \wedge S$, at which its velocity jumps from Y to Y' uniformly sampled on the sphere if $T \wedge S = S$, or to Y^* if $T \wedge S = T$. This latter case means that only the part of Y which is parallel to $\nabla_x U(X)$ jumps (to its opposite), while the orthogonal part is left untouched. An interpretation is that when it jumps, the process is deterministically reflected according to optical laws – or as a billiard – on the level set of U it has reached (see Fig 1.2).

We restrict the study to a compact case: suppose U is smooth and 1-periodic in the sense $x - z \in \mathbb{Z}^d \Rightarrow U(x) = U(z)$, let $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$, consider a cooling-schedule $(\varepsilon_t)_{t \geq 0}$ and the inhomogeneous Markov process $Z^\varepsilon = (X_t^{\varepsilon_t}, Y_t^{\varepsilon_t})_{t \geq 0}$ on $\mathbb{T}^d \times \mathbb{S}^{d-1}$ with generator

$$L_t f(x, y) = y.\nabla_x f(x, y) + \frac{(y.\nabla_x U(x))_+}{\varepsilon_t}\left(f(x, y^*) - f(x, y)\right)$$

$$+ r\left(\int f(x, z)dz - f(x, y)\right). \tag{1.6}$$

Recall the associated semi-group acts on probability laws $\mathcal{P}\left(\mathbb{T}^d \times \mathbb{S}^{d-1}\right)$ by

$$\mu P_{s,t} = \mathcal{L}\left((Z_t) \mid \mathcal{L}((Z_s) = \mu)\right).$$

Let $\nu_t = e^{-\frac{1}{\varepsilon_t}U(x)}dx \times dy$ and

$$\|\mu_1 - \mu_2\|_{TV} = \inf_{Z_1 \sim \mu_1, Z_2 \sim \mu_2} \mathbb{P}\left(Z_1 \neq Z_2\right)$$

be the total variation distance.

First we establish an exponential relaxation to equilibrium when the temperature is fixed:

Theorem 1.4. *There exist $c, \theta > 0$ that depend only on d, r and U such that if $\varepsilon_t = \varepsilon > 0$ is constant (so that $P_{0,t} = P_t$ is homogeneous and $\nu_t = \nu_0$ for all t), for all $t \geq 0$ and all initial law μ*

$$\|\mu P_t - \nu_0\|_{TV} \leq e^{-ce^{-\frac{\theta}{\varepsilon}}\left(t - \sqrt{d}\right)}\|\mu - \nu_0\|_{TV}.$$

Second we obtain a sufficient condition on a cooling schedule for the inhomogeneous process with generator given by (1.6) to converge on probability to any neighbourhood of the set of global maxima of U:

Theorem 1.5. *There exists $\theta > 0$ that depends only on d and U such that the following hold: let $(\varepsilon_t)_{t \geq 0}$ be any positive decreasing cooling schedule that goes to 0 as t goes to infinity, and such that moreover for t large enough*

$$\partial_t\left(\frac{1}{\varepsilon_t}\right) \leq \frac{1}{(\theta + \eta)t}$$

for some $\eta > 0$. Then, for any $h > 0$ and for any initial point $(X_0, Y_0) \in \mathbb{T}^d \times \mathbb{S}^{d-1}$,

$$\mathbb{P}\left(U\left(X_t^{\varepsilon_t}\right) > \min U + h\right) \quad \xrightarrow[t \to \infty]{} \quad 0.$$

In particular if $\varepsilon_t = \frac{\theta + \eta}{\ln t}$ for t large enough, then there exists $C > 0$ such that for all $t, h > 0$ and $(X_0, Y_0) \in \mathbb{T}^d \times \mathbb{S}^{d-1}$,

$$\mathbb{P}\left(U\left(X_t^{\varepsilon_t}\right) > \min U + h\right) \quad \leq \quad C\left(\frac{1}{t}\right)^{\frac{\min(h/2, \eta)}{\theta + \eta}}.$$

There is no reason not to expect that the smallest θ such that Theorems 1.4 and 1.5 hold is E^* the critical depth of the potential U (see the remark after the proof of Theorem 1.4 concerning the explicit bound we get). We believe that a sharper analysis, similar to our study in dimension 1 or to classical studies of the simulated annealing based on the Fokker-Planck diffusion, together with some ideas from the proof of Theorem 1.5 which are specific to the present process, could in fact enable us to reach $\theta = E^*$ (see the conclusive remark at the end of the paper). However we also think that somehow the additional technicalities is not worth the improvement of the result, at least in a first instance: indeed, in an applied problem, E_* is anyway unknown and out of reach.

The restriction to a compact space in Theorem 1.5, which simplifies the study, should not be necessary. If for instance we suppose U goes to infinity as $|x| \to \infty$ and $\varepsilon_t \to 0$ then at some time the process will stay forever trapped in a compact set, and the same arguments will work (the constant C however will depend on the initial position). In Theorem 1.4 if U goes to infinity as $|x| \to \infty$ we can prove via Lyapunov techniques that the process always come back to a given compact set, but since it moves at constant speed, $\|\mu P_t - \nu_0\|_{TV}$ will not be controlled by $\|\mu - \nu_0\|_{TV}$ uniformly on μ; some moments of μ will be involved to account for the first hitting time of a given compact set.

Remark on the definition of the multidimensional process. We want to define a process with the specifications of Section 1.2: (X, Y) has to be a kinetic Markov process on $\mathbb{R}^d \times \mathbb{S}^{d-1}$ with a piecewise constant velocity. It remains to choose the jump rate and kernel of Y with the constraint that the first marginal of the invariant measure has to be the Gibbs law associated to some potential U.

Recall that in dimension 1 an asset of the velocity jump process with minimal jump rate $(yU'(x))_+$ is that, in order to determine the next jump time, it is only necessary to compute $U(X_t)$ along the trajectory. In any dimension, this is still true with the rate $\lambda(x, y) = (y.\nabla U(x))_+$. Indeed in that case, while there is no jump,

$$\int_0^t \lambda(X_s, Y_s) ds \quad = \quad \int_0^t \left(\partial_s\left(U(X_s)\right)\right)_+ ds$$

$$= \quad \begin{cases} 0 & \text{while } U(X_t) \text{ decreases with } t \\ U(X_t) - U(X_0) & \text{while } U(X_t) \text{ increases with } t, \end{cases}$$

and more generally it is the cumulated increases of U along the trajectory since the last jump. Note that on the contrary, in the case of the Fokker-Planck diffusion, $\nabla U(X_t)$ should be estimated all the time, possibly at a huge numerical cost in large dimension.

This choice of jump rate yields a generator of the form

$$Lf(x,y) = y.\nabla_x f(x,y) + (y.\nabla_x U(x))_+ \left(\int_{\mathbb{S}^{d-1}} f(x,z) p_{x,y}(z)\sigma(dz) - f(x,y) \right)$$

with a transition kernel $p_{x,y}(z)$ on $\mathbb{S}^{d-1}$ which is still to be determined. Let $\nu = e^{-U(x)}dx \times dy$ where dy is the uniform law on the sphere. In dimension greater than 1, this is an arbitrary - albeit the simplest - choice to ensure the first marginal of ν is the Gibbs law associated to U. Then ν is invariant for L iff $p_{x,y}(z)$ is a (weak) solution of

$$y.\nabla_x U(x) \quad = \quad (y.\nabla_x U(x))_+ - \int_{\mathbb{S}^{d-1}} (z.\nabla_x U(x))_+ \, p_{x,z}(y)\sigma(dz)$$

for a.e. $(x,y) \in \mathbb{R}^d \times \mathbb{S}^{d-1}$, or in other words

$$\int_{\mathbb{S}^{d-1}} (z.\nabla_x U(x))_+ \, p_{x,z}(y)\sigma(dz) \quad = \quad (-y.\nabla_x U(x))_+ \,.$$

This is true if when Y jumps, the part $Y.\nabla_x U(X)$ is changed to its opposite. We have complete freedom for the choice of what happens to the part of Y which is orthogonal to $\nabla_x U(X)$, as long as it stays orthogonal.

In particular the condition would be satisfied with $p_{x,y} = \delta_{-y}$, namely Y jumps to $-Y$. But then X_t would be forever trapped on the same line, and the process would not be ergodic. This could be fixed by adding uniform jumps of Y at a constant rate, which leaves ν invariant, but then the trajectories would still seem rather inefficient from a mixing point of view.

The choice $p_{x,y} = \delta_{y^*}$ seems more natural. From a practical point of view, it means that the process has more inertia than if we had chosen the kernel $Y \leftarrow -Y$. However to sample the process, it is then necessary to estimate $\nabla_x U$ at each jump time; but this is still less than for a diffusion.

It is not exactly clear whether it is ergodic: if U is infinite outside an ellipse and vanishes inside, then there is no randomness, the process is a deterministic billiard and so even if its initial speed is uniformly distributed, it will stay in an area defined by some caustic. If we add uniform jumps on the velocity at constant rate $r > 0$, then obviously the process may reach any open set of $\mathbb{R}^d \times \mathbb{S}^{d-1}$ with positive probability.

All these considerations lead to the definition (1.6).

Outline of the paper. In the rest of the paper are proven the above results. Section 2 is concerned with the low-temperature regime for a double-well potential; Theorem 1.1 is proven, which introduces a discussion on the time scale of total variation and Wasserstein convergence to equilibrium. The proof of Theorem 1.3, in inhomogeneous setting, is addressed in Section 3. Section 4 investigates the non-minimal rate case, namely the case $\lambda(x,y) > (yU'(x))_+$. Finally, the multi-dimensional case is addressed in Section 5, where Theorems 1.4 and 1.5 are proven.

2. Escape time at low temperature

In this section the dimension is 1, the temperature $\varepsilon > 0$ is fixed through time, $(X_t^\varepsilon, Y_t^\varepsilon)$ is the velocity jump process with generator L defined by (1.4); the double-well potential U has three local extrema $x_0 < x_1 < x_2$ such as described in the

previous section in the settings of Theorem 1.1. Suppose $(X_0^\varepsilon, Y_0^\varepsilon) = (x_0, -1)$ and recall that

$$\tau = \inf\{t > 0, X_t^\varepsilon = x_1\}.$$

We start the proof of Theorem 1.1 with the following Lemma:

Lemma 2.1. *for $\delta > 0$ small enough,*

$$\int_0^\delta \frac{t}{\varepsilon} \left(-U'(x_0 - t)\right) e^{-\frac{U(x_0 - t) - U(x_0)}{\varepsilon}} dt = \sqrt{\frac{\pi\varepsilon}{2U''(x_0)}} \left(1 + \underset{\varepsilon \to 0}{o}(1)\right).$$

Proof:

$$\int_0^\delta \frac{t}{\varepsilon} \left(-U'(x_0 - t)\right) e^{-\frac{U(x_0 - t) - U(x_0)}{\varepsilon}} dt$$

$$= \int_0^{\frac{\delta}{\sqrt{\varepsilon}}} s \left(-U'(x_0 - \sqrt{\varepsilon}s)\right) e^{-\frac{U(x_0 - \sqrt{\varepsilon}s) - U(x_0)}{\varepsilon}} ds$$

$$= \sqrt{\varepsilon} \int_0^\infty f_\varepsilon(s) ds$$

with

$$f_\varepsilon(s) = \frac{s\left(-U'(x_0 - \sqrt{\varepsilon}s)\right)}{\sqrt{\varepsilon}} e^{-\frac{U(x_0 - \sqrt{\varepsilon}s) - U(x_0)}{\varepsilon}} \mathbb{1}_{s \leq \delta/\sqrt{\varepsilon}}.$$

On the one hand

$$f_\varepsilon(s) \underset{\varepsilon \to 0_+}{\longrightarrow} s^2 U''(x_0) e^{-\frac{U''(x_0)}{2} s^2}$$

and on the other hand, writing $M = \underset{t \in [0,\delta]}{\sup} |U^{(3)}(t)|,$

$$|f_\varepsilon(s)| \leq s^2 \left(U''(x_0) + \frac{\delta M}{2}\right) e^{-\frac{U''(x_0)}{2} s^2 + \frac{\delta s^2 M}{6}}.$$

As M decreases with δ, for δ small enough, $\delta M < 3U''(x_0)$. Thus by the dominated convergence theorem

$$\int_0^\infty f_\varepsilon(s) ds \underset{\varepsilon \to 0}{\longrightarrow} \int_0^\infty s^2 U''(x_0) e^{-\frac{U''(x_0)}{2} s^2} ds$$

$$= \sqrt{\frac{\pi}{2U''(x_0)}}.$$

$\square$

Proof of Theorem 1.1: The strategy of this proof is quite simple: starting at $(x_0, -1)$, the process starts to climb to the left. After a time with bounded expectation (since the potential is increasing and convex along the trajectory), it turns back, reaches $(x_0, 1)$ and start to climb toward x_1. If it reaches x_1 then this attempt to escape is a success. If not, it goes back to $(x_0, -1)$, and starts anew. Since the duration of the last (and successful) attempt is negligible with respect to the expected duration of a failed attempt (of order $\sqrt{\varepsilon}$ according to Lemma 2.1) times the expected number of failed attempts, conclusion follows from the usual convergence of geometrical laws to the exponential one.

More precisely let $\theta_0 = 0$, and for all $k \in \mathbb{N}$ set

$$\tilde{\theta}_k = \min\left(\inf\{t > \theta_k, (X_t, Y_t) = (x_0, 1)\}, \tau\right),$$

$$\theta_{k+1} \;=\; \min\left(\inf\left\{t > \tilde{\theta}_k,\; (X_t, Y_t) = (x_0, -1)\right\},\; \tau\right).$$

We also define $N = \max\{k,\; \theta_k < \tau\}$, so that

$$\tau \;=\; \tau - \theta_N + \sum_{k=1}^{N}(\theta_k - \theta_{k-1}).$$

Conditionally on $N \geq k$, $\theta_k - \theta_{k-1}$ is the length of a failed attempt to reach x_1; it has the same law as θ_1, given $N \geq 1$, and is independent from N. In particular

$$\mathbb{E}\left[\tau\right] \;=\; \mathbb{E}\left[\tau - \theta_N\right] + \mathbb{E}\left[N\right]\mathbb{E}\left[\theta_1 \mid N \geq 1\right], \tag{2.1}$$

and N is a geometric variable with parameter

$$\begin{aligned}
q_\varepsilon \;&=\; \mathbb{P}\left(\varepsilon E > U(x_1) - U(x_0)\right) \\
&=\; e^{-\frac{U(x_1) - U(x_0)}{\varepsilon}}
\end{aligned}$$

(and so with expectation $\mathbb{E}\left[N\right] = q_\varepsilon^{-1}$). Now we decompose

$$\mathbb{E}\left[\theta_1 \mid N \geq 1\right] \;=\; 2\mathbb{E}\left[\frac{\tilde{\theta}_0}{2}\right] + 2\mathbb{E}\left[\left.\frac{\theta_1 - \tilde{\theta}_0}{2}\right| N \geq 1\right]. \tag{2.2}$$

Notice that $\frac{\tilde{\theta}_0}{2}$ is the jump time starting from $(x_0, -1)$, which is independent from N and whose law has density $\frac{1}{\varepsilon}\left(-U'(x_0 - t)\right)e^{-\frac{U(x_0 - t) - U(x_0)}{\varepsilon}}$. If $\varepsilon < \frac{1}{2}$, then for any $\delta > 0$

$$\begin{aligned}
\mathbb{E}\left[\frac{\tilde{\theta}_0}{2}\mathbb{1}_{\frac{\tilde{\theta}_0}{2} > \delta}\right] \;&=\; \int_\delta^\infty \frac{t}{\varepsilon}\left(-U'(x_0 - t)\right)e^{-\frac{U(x_0 - t) - U(x_0)}{\varepsilon}}\,dt \\
&\leq\; \int_\delta^\infty \frac{t}{\varepsilon}\left(-U'(x_0 - t)\right)e^{-\frac{U(x_0 - t) - U(x_0)}{2\varepsilon} - \frac{U(x_0 - \delta) - U(x_0)}{2\varepsilon}}\,dt \\
&\leq\; \frac{e^{-\frac{U(x_0 - \delta) - U(x_0)}{2\varepsilon}}}{\varepsilon}\int_\delta^\infty t\left(-U'(x_0 - t)\right)e^{-(U(x_0 - t) - U(x_0))}\,dt \\
&=\; \underset{\varepsilon \to 0}{o}\left(\sqrt{\varepsilon}\right).
\end{aligned}$$

On the other hand thanks to Lemma 2.1, for δ small enough,

$$\mathbb{E}\left[\frac{\tilde{\theta}_0}{2}\mathbb{1}_{\frac{\tilde{\theta}_0}{2} < \delta}\right] \;=\; \sqrt{\frac{\pi\varepsilon}{2U''(x_0)}}\left(1 + \underset{\varepsilon \to 0}{o}(1)\right).$$

If $\delta < x_1 - x_0$, Lemma 2.1 also applies to $\frac{\theta_1 - \tilde{\theta}_0}{2}$, and since $\mathbb{P}(N \geq 1)$ goes to 1,

$$\mathbb{E}\left[\left.\frac{\theta_1 - \tilde{\theta}_0}{2}\mathbb{1}_{\frac{\theta_1 - \tilde{\theta}_0}{2} < \delta}\right| N \geq 1\right] \;=\; \sqrt{\frac{\pi\varepsilon}{2U''(x_0)}}\left(1 + \underset{\varepsilon \to 0}{o}(1)\right).$$

Furthermore

$$\begin{aligned}
\mathbb{E}\left[\left.\frac{\theta_1 - \tilde{\theta}_0}{2}\mathbb{1}_{\frac{\theta_1 - \tilde{\theta}_0}{2} \geq \delta}\right| N \geq 1\right] \;&\leq\; (x_1 - x_0)\frac{\mathbb{P}\left(\frac{\theta_1 - \tilde{\theta}_0}{2} \geq \delta\right)}{\mathbb{P}(N \geq 1)} \\
&=\; \underset{\varepsilon \to 0}{o}\left(\sqrt{\varepsilon}\right).
\end{aligned}$$

By combining the estimates above with (2.2) we get

$$\mathbb{E}\left[\theta_1 \mid N \geq 1\right] = \sqrt{\frac{8\pi\varepsilon}{U''(x_0)}} \left(1 + \underset{\varepsilon \to 0}{o}(1)\right).$$

Besides, note that $\frac{\tilde{\theta}_N - \theta_N}{2}$ has the same law as $\frac{\tilde{\theta}_0}{2}$, so that

$$\mathbb{E}\left[\tau - \theta_N\right] = 2\mathbb{E}\left[\frac{\tilde{\theta}_N - \theta_N}{2}\right] + x_1 - x_0 = \underset{\varepsilon \to 0}{O}(1) = \underset{\varepsilon \to 0}{o}\left(\sqrt{\varepsilon}q_\varepsilon^{-1}\right),$$

and finally Equality (2.1) becomes

$$\mathbb{E}\left[\tau\right] = \sqrt{\frac{8\pi\varepsilon}{U''(x_0)}} e^{\frac{U(x_1) - U(x_0)}{\varepsilon}} \left(1 + \underset{\varepsilon \to 0}{o}(1)\right).$$

Now as far as the second assertion of the theorem is concerned,

$$\mathbb{P}\left(\tau \geq t\mathbb{E}\left[\tau\right]\right)$$

$$=\mathbb{P}\left(\tau - \theta_N + \sum_{k=1}^{N}(\theta_k - \theta_{k-1}) \geq t\mathbb{E}\left[\tau\right]\right)$$

$$=\mathbb{P}\left(\frac{\mathbb{E}\left[N\right]\mathbb{E}\left[\theta_1 \mid N \geq 1\right]}{\mathbb{E}\left[\tau\right]} \frac{N}{\mathbb{E}\left[N\right]} \left(\frac{\tau - \theta_N}{N\mathbb{E}\left[\theta_1 \mid N \geq 1\right]} + \sum_{k=1}^{N}\frac{\theta_k - \theta_{k-1}}{N\mathbb{E}\left[\theta_1 \mid N \geq 1\right]}\right) \geq t\right).$$

First, $\frac{N}{\mathbb{E}[N]}$ converges in law to an exponential variable of parameter 1. Second, if the times of jump of the process are defined by the same sequence $(E_k)_{k \geq 0}$ of exponential variables (according to Equality (1.3)) for all ε, then the processes at different temperatures are all defined at once on the same probability space, and in this case

$$N = \min\{k \geq 0, \varepsilon E_{2k} > U(x_1) - U(x_0)\}$$

almost surely goes to infinity. Thanks to the law of large numbers,

$$\frac{\tau - \theta_N}{N\mathbb{E}\left[\theta_1 \mid N \geq 1\right]} + \sum_{k=1}^{N}\frac{\theta_k - \theta_{k-1}}{N\mathbb{E}\left[\theta_1 \mid N \geq 1\right]} \underset{\varepsilon \to 0}{\longrightarrow} 1 \quad a.s.$$

At last $\frac{\mathbb{E}[N]\mathbb{E}[\theta_1 \mid N \geq 1]}{\mathbb{E}[\tau]} \underset{\varepsilon \to 0}{\longrightarrow} 1$. As a conclusion $\frac{\tau}{\mathbb{E}[\tau]}$ converges to an exponential law. $\qquad \square$

This result yields a lower bound on the total variation distance between

$$\nu_t = \mathcal{L}\left(X_t^\varepsilon, Y_t^\varepsilon \mid X_0^\varepsilon = x_0, \ Y_0^\varepsilon = -1\right)$$

and the invariant measure μ_ε. We recall that the total variation distance between two laws $\nu, \tilde{\nu}$ on a topological space W is defined by

$$\|\nu - \tilde{\nu}\|_{TV} = \sup_{A \in \mathcal{B}(W)} |\nu(A) - \tilde{\nu}(A)|$$

where $\mathcal{B}(W)$ is the Borel σ-algebra on W.

Corollary 2.2. *In the setting of Theorem 1.1, assume furthermore $U(x_2) < U(x_0)$. For any $t > 0$, writing $t_\varepsilon = t\mathbb{E}\left[\tau\right]$,*

$$\liminf_{\varepsilon \to 0} \|\nu_{t_\varepsilon} - \mu_\varepsilon\|_{TV} \geq e^{-t}.$$

Proof: The assumption $U(x_2) < U(x_0)$ implies that the first marginal of μ_ε concentrates near the unique global minimum of U, x_2. In particular

$$\mu_\varepsilon\left[(x_1, \infty) \times \{\pm 1\}\right] \underset{\varepsilon \to 0}{\longrightarrow} 1.$$

However

$$\nu_t\left[(x_1, \infty) \times \{\pm 1\}\right] \quad \leq \quad \mathbb{P}\left(\tau \leq t\right).$$

As a consequence

$$\begin{aligned}
\|\nu_{t_\varepsilon} - \mu_\varepsilon\|_{TV} \quad &\geq \quad (\mu_\varepsilon - \nu_{t_\varepsilon})\left[(x_1, \infty) \times \{\pm 1\}\right] \\
&\geq \quad \mu_\varepsilon\left[(x_1, \infty) \times \{\pm 1\}\right] - \mathbb{P}\left(\tau \leq t\mathbb{E}\left[\tau\right]\right) \\
&\underset{\varepsilon \to 0}{\longrightarrow} \quad e^{-t}.
\end{aligned}$$

$\square$

We will now give a more precise result concerning the first marginal and the Wasserstein distance

$$\mathcal{W}\left(\nu_1, \nu_2\right) = \sup_\pi \mathbb{E}_\pi\left[|Z_1 - Z_2|\right],$$

where the infimum is taken over all probability measures π on $\mathbb{R}^2$ whose first (resp. second) marginal is ν_1 (resp. ν_2). We start with the following lemma (recall that we fixed $(X_0^\varepsilon, Y_0^\varepsilon) = (x_0, -1)$) :

Lemma 2.3. *Let* $t > 0$. *Then*

$$\limsup_{\varepsilon \to 0} \mathbb{E}\left[(X_{t_\varepsilon}^\varepsilon)^2\right] < \infty.$$

Proof: Let $M > U(x_1)$, and for all $k \geq 0$ let

$$s_k = \inf\{s > 0,\ U(X_s^\varepsilon) > M + k\}.$$

From Theorem 3.2 (stated and proved in the next section) we get that for all $k \geq 0$, there exist $\Gamma, \varepsilon_0 > 0$ such that for all $\varepsilon \leq \varepsilon_0$ and all $t > 0$,

$$\begin{aligned}
\mathbb{P}\left(s_k \leq t_\varepsilon\right) \quad &\leq \quad 1 - \exp\left(-\Gamma\left(e^{-\frac{M+k}{\varepsilon}} + t_\varepsilon \frac{e^{-\frac{M+k-U(x_2)}{\varepsilon}}}{\sqrt{\varepsilon}}\right)\right) \\
&\leq \quad 1 - \exp\left(-\tilde{\Gamma}e^{-\frac{M+k}{\varepsilon_0}}\right),
\end{aligned}$$

where $\tilde{\Gamma}$ depends on t but not on ε. In fact, since we consider here escape times from nested intervals, and from the convexity of U outside of a compact, Γ (and hence $\tilde{\Gamma}$) and ε_0 can be chosen uniformly over k (see the remark at the end of the proof of Theorem 3.2). Noticing that $U^{-1}(v) = \sup\{|x|,\ U(x) \leq v\}$ is sub-linear, we write

$$\begin{aligned}
\mathbb{E}\left[(X_{t_\varepsilon}^\varepsilon)^2\right] \quad &\leq \quad \left(U^{-1}(M)\right)^2 + \sum_{k \geq 0}\left(U^{-1}(M + k + 1)\right)^2 \mathbb{P}\left(s_k \leq t_\varepsilon\right) \\
&\leq \quad \left(U^{-1}(M)\right)^2 + \sum_{k \geq 0}\left(U^{-1}(M + k + 1)\right)^2 \left(1 - \exp\left(-\tilde{\Gamma}e^{-\frac{M+k}{\varepsilon_0}}\right)\right) \\
&\leq \quad \left(U^{-1}(M)\right)^2 + \tilde{\Gamma}\sum_{k \geq 0}\left(U^{-1}(M + k + 1)\right)^2 e^{-\frac{M+k}{\varepsilon_0}}
\end{aligned}$$

which is finite.

$\square$

Let $m_t = e^{-t}\delta_{x_0} + (1 - e^{-t})\delta_{x_2}$ and h_t be the first marginal of ν_t, namely

$$h_t = \mathcal{L}\left(X_t^\varepsilon \mid X_0^\varepsilon = x_0,\ Y_0^\varepsilon = -1\right).$$

Theorem 2.4. *In the setting of Theorem 1.1, assume moreover $U(x_2) < U(x_0)$. Then for any $t > 0$, writing $t_\varepsilon = t\mathbb{E}[\tau]$,*

$$\mathcal{W}\left(h_{t_\varepsilon}, m_t\right) \underset{\varepsilon \to 0}{\longrightarrow} 0.$$

Proof: Consider $(X_t^\varepsilon, Y_t^\varepsilon)_{t \geq 0}$ a trajectory of the process, from which we will define a variable W of law m_t. In the first instance assume $p_\varepsilon := \mathbb{P}(\tau \geq t_\varepsilon) \leq e^{-t}$. If $\tau \geq t_\varepsilon$, set $U = 1$. Else set $U = 1$ with probability $\frac{e^{-t} - p_\varepsilon}{1 - p_\varepsilon}$, and else set $U = 0$. Similarly in the case where $p_\varepsilon \geq e^{-t}$, if $\tau \leq t_\varepsilon$, set $U = 0$, and else set $U = 0$ with probability $\frac{p_\varepsilon - e^{-t}}{p_\varepsilon}$, and else set $U = 1$.

Either ways, U is a Bernoulli variable with parameter e^{-t} such that

$$\mathbb{P}\left(U = \mathbb{1}_{\tau \geq t_\varepsilon}\right) = 1 - |e^{-t} - p_\varepsilon| \underset{\varepsilon \to 0}{\longrightarrow} 1.$$

We naturally set $W = Ux_0 + (1 - U)x_2$. The Cauchy-Schwarz inequality

$$\mathbb{E}\left[|X_{t_\varepsilon}^\varepsilon - W|\mathbb{1}_{U \neq \mathbb{1}_{\tau \geq t_\varepsilon}}\right] \leq \sqrt{\mathbb{E}\left[|X_{t_\varepsilon}^\varepsilon - W|^2\right]\mathbb{P}(U \neq \mathbb{1}_{\tau \geq t_\varepsilon})},$$

together with Lemma 2.3, yields

$$\mathbb{E}\left[|X_{t_\varepsilon}^\varepsilon - W|\right] = \mathbb{E}\left[|X_{t_\varepsilon}^\varepsilon - W|\mathbb{1}_{U = \mathbb{1}_{\tau \geq t_\varepsilon}}\right] + \underset{\varepsilon \to 0}{o}(1).$$

We write

$$\tau^{(2)} = \inf\left\{t > \tau,\ X_t^\varepsilon < x_1\right\},$$

and remark that $\tau^{(2)} - \tau$, which is independent from τ, is the first hitting time of $(x_1, -1)$ starting from $(x_2, 1)$: from Theorem 1.1, and since $U(x_2) < U(x_1)$, we see that $\mathbb{E}[\tau]$ is negligible with respect to $\mathbb{E}[\tau^{(2)} - \tau]$, that $\frac{\tau^{(2)} - \tau}{\mathbb{E}[\tau^{(2)} - \tau]}$ converges in law to an exponential variable and so that $\frac{\tau^{(2)} - \tau}{\mathbb{E}[\tau]}$ diverges in probability to infinity. In particular

$$\mathbb{P}\left(\tau^{(2)} \geq t_\varepsilon\right) \geq \mathbb{P}\left(\tau^{(2)} - \tau \geq t_\varepsilon\right)$$
$$\underset{\varepsilon \to 0}{\longrightarrow} 1.$$

Using again Lemma 2.3, we get

$$\mathbb{E}\left[|X_{t_\varepsilon}^\varepsilon - W|\right] = \mathbb{E}\left[|X_{t_\varepsilon}^\varepsilon - W|\ \mathbb{1}_{U = \mathbb{1}_{\tau \geq t_\varepsilon},\ \tau^{(2)} > t_\varepsilon}\right] + \underset{\varepsilon \to 0}{o}(1)$$
$$= \mathbb{E}\left[|X_{t_\varepsilon}^\varepsilon - x_0|\ \mathbb{1}_{t_\varepsilon < \tau}\right] + \mathbb{E}\left[|X_{t_\varepsilon}^\varepsilon - x_2|\ \mathbb{1}_{\tau < t_\varepsilon < \tau^{(2)}}\right] + \underset{\varepsilon \to 0}{o}(1).$$

Both expectations are treated the same way, let us focus on the first one. Suppose $t_\varepsilon < \tau$ and let

$$t' = \sup\{s < t_\varepsilon,\ X_s^\varepsilon = x_0\}, \qquad t'' = \min\left(\inf\{s > t_\varepsilon,\ X_s^\varepsilon = x_0\}, \tau\right),$$

and $I = t'' - t'$. Note that $|X_{t_\varepsilon}^\varepsilon - x_0| < I$ and for $0 \leq a \leq b$ the following events are included:

$$\{a \leq I < b\} \subset A,$$

where

$$A = \left\{\exists s_1, s_2 \in [t_\varepsilon - b, t_\varepsilon + b],\ s_1 < s_2,\ X_{s_1}^\varepsilon = x_0,\ |X_{s_2}^\varepsilon - x_0| = \frac{a}{2}\right\}.$$

Since the process starts anew when it reaches x_0,

$$
\begin{aligned}
\mathbb{P}(A) &\leq \mathbb{P}_{X_0=x_0}\left(\exists s < 2b, \ |X_s^\varepsilon - x_0| = \frac{a}{2}\right) \\
&\leq \mathbb{P}\left(\theta(a) \leq 2b\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\theta(a) &= \inf\{s > 0, \ U(X_s^\varepsilon) > U(x_0) + \delta(a)\} \\
\text{and} \quad \delta(a) &= \min\left(U\left(x_0 - \frac{a}{2}\right), U\left(x_0 + \frac{a}{2}\right)\right) - U(x_0) \text{ if } x_0 + \frac{a}{2} \leq x_1 \\
&= U\left(x_0 - \frac{a}{2}\right) - U(x_0) \text{ else.}
\end{aligned}
$$

In other words, $\delta(a)$ is the minimal energy barrier to overcome to be at distance at least $\frac{a}{2}$ from x_0 (when the process is still in the catchment area of x_0). If $a \leq I < b$, such a distance has been attained, and it has been so in a prescribed time window of length at most $2b$. As in the proof of Lemma 2.3 we use Theorem 3.2 to control the probability to reach a given energy level in a given time: for all $k \geq 0$ and $\varepsilon < \varepsilon_0$,

$$
\begin{aligned}
\mathbb{P}\left(\theta(a) \leq 2b\right) &\leq 1 - \exp\left(-\Gamma\left(e^{-\frac{\delta(a)}{\varepsilon}} + 2b\frac{e^{-\frac{\delta(a)}{\varepsilon}}}{\sqrt{\varepsilon}}\right)\right) \\
&\leq \Gamma\left(1 + 2b\varepsilon^{-\frac{1}{2}}\right)e^{-\frac{\delta(a)}{\varepsilon}}.
\end{aligned}
$$

By combining the estimates above, when $\varepsilon < 1$,

$$
\begin{aligned}
&\mathbb{E}\left[|X_{t_\varepsilon} - x_0| \ \mathbb{1}_{t_\varepsilon < \tau}\right] \\
&\leq \mathbb{E}\left[I \ \mathbb{1}_{t_\varepsilon < \tau}\right] \\
&\leq \sum_{0 \leq k \leq \lceil \varepsilon^{-\frac{1}{2}} \rceil} \sqrt{\varepsilon}(k+1)\mathbb{P}\left(k\sqrt{\varepsilon} \leq I < (k+1)\sqrt{\varepsilon}\right) + \sum_{k \geq 1}(k+1)\mathbb{P}\left(k \leq I < (k+1)\right) \\
&\leq 4\Gamma\sqrt{\varepsilon}\sum_{0 \leq k \leq \lceil \varepsilon^{-\frac{1}{2}} \rceil}(k+1)^2 e^{-\frac{\delta(k\sqrt{\varepsilon}/2)}{\varepsilon}} + 4\Gamma\sum_{k \geq 1}(k+1)^2 \varepsilon^{-\frac{1}{2}} e^{-\frac{\delta(k)}{\varepsilon}}.
\end{aligned}
$$

(2.3)

From the convexity of U, $\delta(k)$ grows faster than linearly, so that, if ε_0 is such that

$$
\varepsilon \mapsto \varepsilon^{-\frac{1}{2}} e^{-\frac{\delta(1)}{\varepsilon}}
$$

is non-increasing on $(0, \varepsilon_0)$, for all $\varepsilon < \varepsilon_0$,

$$
\sum_{k \geq 1}(k+1)^2 \varepsilon^{-\frac{1}{2}} e^{-\frac{\delta(k)}{\varepsilon}} < \sum_{k \geq 1}(k+1)^2 \varepsilon_0^{-\frac{1}{2}} e^{-\frac{\delta(k)}{\varepsilon_0}} < \infty.
$$

Thus, by the dominated convergence Theorem, the second sum in (2.3) vanishes as ε goes to 0. As far as the first sum is concerned, note that, since x_0 is a non-degenerated minimum of U, there exist $\eta > 0$ such that $\delta(s) \geq \eta s^2$ for all $s \in [0,1]$. This implies

$$
\sum_{0 \leq k \leq \lceil \varepsilon^{-\frac{1}{2}} \rceil}(k+1)^2 e^{-\frac{\delta(k\sqrt{\varepsilon}/2)}{\varepsilon}} \leq \sum_{k \geq 0}(k+1)^2 e^{-\eta k^2} < \infty.
$$

By inserting this estimate in (2.3), $\mathbb{E}\left[\left|X_{t_\varepsilon}^\varepsilon - x_0\right| \mathbb{1}_{t_\varepsilon < \tau}\right]$ goes to 0; the same goes for $\mathbb{E}\left[\left|X_{t_\varepsilon}^\varepsilon - x_2\right| \mathbb{1}_{\tau < t_\varepsilon < \tau^{(2)}}\right]$ by a similar argument, and we conclude that $\mathbb{E}\left[\left|X_{t_\varepsilon}^\varepsilon - W\right|\right]$ goes to 0. For ε small enough one can thus find a coupling (X, W) with marginal laws $\mathcal{L}\left(X_{t_\varepsilon}^\varepsilon\right)$ and $e^{-t}\delta_{x_0} + (1 - e^{-t})\delta_{x_2}$ such that $\mathbb{E}\left[\left|X - W\right|\right]$ is arbitrarily small. $\square$

In particular since $\mu_\varepsilon^1 = e^{-\frac{1}{\varepsilon}U(x)}dx$, the first marginal of the invariant measure μ_ε, converges to the Dirac mass on x_2, and

$$\mathcal{W}\left(e^{-t}\delta_{x_0} + (1 - e^{-t})\delta_{x_2}, \delta_{x_2}\right) = e^{-t}|x_2 - x_0|,$$

we get

Corollary 2.5. *In the setting of Theorem 2.4,*

$$\mathcal{W}\left(h_{t_\varepsilon}, \mu_\varepsilon^1\right) \xrightarrow[\varepsilon \to 0]{} e^{-t}|x_2 - x_0|.$$

3. NS conditions for the cooling schedule

We now turn to the study of the inhomogeneous velocity jump process $(X_t^{\varepsilon_t}, Y_t^{\varepsilon_t})_{t \geq 0}$ on $\mathbb{R} \times \{-1, +1\}$ with generator

$$L_t f(x, y) \quad = \quad yf'(x, y) + \frac{1}{\varepsilon_t}\left(yU'(x)\right)_+ \left(f(x, -y) - f(x, y)\right),$$

when Assumption 1 holds, which is supposed in the whole Section 3. To lighten the notations, in this section we drop the ε_t exponent and only call the process $(X_t, Y_t)_{t \geq 0}$.

It can be explicitly constructed in the same manner as the homogeneous process of Section 1.2, except that the definition (1.3) of the jump times is now replaced by

$$T_{i+1} \quad = \quad \inf\left\{t > T_i, \int_{T_i}^t \frac{\left(Y_{T_i}U'\left(X_{T_i} + Y_{T_i}(s - T_i)\right)\right)_+}{\varepsilon_s}ds \geq E_i\right\}.$$

In a finite time interval $[0, T]$, ε_t is bounded by ε_T and thus there cannot be an infinite number of jumps in a finite time. In particular the sequence of $(T_i)_{i \geq 0}$ goes to infinity and the process is well defined at all times.

Our method to prove Theorem 1.3 follows the work of Hajek (1988) on simulated annealing on a discrete space. We consider a smooth Morse potential U on $\mathbb{R}$ with a finite number of local extrema, and convex at infinity. If x is a minimum of depth E, denote by C_x the set of all points which are reachable from x at height strictly less than E. We call C_x the cusp of x (see the grey area of Figure 1.1).

More generally we call cusp an interval $C = (z_l, z_r)$ with $U(z_l) = U(z_r)$ and for all $x \in C$, $U(x) < U(z_l)$. The depth d of C is defined as

$$d = U(z_l) - \min\{U(x), \ x \in C\}.$$

We denote by $B = \{z \in C, U(z) = \min\{U(x), \ x \in C\}\}$ the bottom of the cusp. Note that a minimum x is always in the bottom of C_x, and that conversely if z is in the bottom of C_x then $C_x = C_z$. Obviously the depth of C_x equals the depth of x.

We want to bound the time the process spends in a cusp C, depending on the depth d of the latter. Nevertheless it is impossible to do so if we only assume the initial position is in C: we should put aside the cases where the process starts near the boundary of C, with a velocity directed toward the exit.

We denote by $\mathcal{N}$ the set of local extrema of U in C, and

$$u = U(z_l) - \max\{U(z), \ z \in \mathcal{N}\}.$$

The maximum over $\mathcal{N}$ is necessarily attained on a local maximum, except when there is no local maximum between z_l and z_r, in which case $\mathcal{N}$ is a single point, which is a local minimum, and $u = d$. We define

$$x_l = \inf\{x \in C, U(z_l) - U(x) \geq u\} \quad x_r = \sup\{x \in C, U(z_r) - U(x) \geq u\},$$

and

$$A_C = (z_l, x_l) \times \{-1\} \cup (x_r, z_r) \times \{1\} \subset \mathbb{R} \times \{\pm 1\}.$$

The process can only leave C from A_C, and can only enter it through $\bar{A}_C = (C \times \{\pm 1\}) \setminus A_C$. We will note $A_C(t)$ (resp. $\bar{A}_C(t)$) the event $(X_t, Y_t) \in A_C$ (resp. $\bar{A}_C$).

Let $t_0 > 0$ and

$$\tau_C = \inf\{t > t_0, \ X_t \notin C\}.$$

In order to prove Theorem 1.3, we will establish the two following intermediate results:

Theorem 3.1. *Let C be a cusp of depth d. There exist ε_0, $c > 0$ such that for all time $t_0 > 0$, any $z \in \{z_l, z_r\}$ and all cooling schedule $t \mapsto \varepsilon_t$, if $\bar{A}_C(t_0)$ holds, $\varepsilon_{t_0} \leq \varepsilon_0$ and*

$$\int_{t_0}^{\infty} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds = \infty,$$

then

$$\mathbb{E}\left[\int_{t_0}^{\tau_C} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds\right] \leq c, \tag{3.1}$$

$$\mathbb{P}(X_{\tau_C} = z) \geq \frac{1}{c}. \tag{3.2}$$

Theorem 3.2. *Let C be a cusp of depth d. There exist $\Gamma, \varepsilon_0 > 0$ such that for all times $t_0 > 0$, $r \geq t_0$ and all cooling schedule $t \mapsto \varepsilon_t$, if $\bar{A}_C(t_0)$ holds and $\varepsilon_{t_0} \leq \varepsilon_0$ then*

$$\mathbb{P}(\tau_C \geq r) \geq \exp\left(-\Gamma\left(\frac{e^{-\frac{d}{\varepsilon_{t_0}}}}{\sqrt{\varepsilon_{t_0}}} + \int_{t_0}^r \frac{e^{-\frac{d}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du\right)\right).$$

Note that contrary to Theorem 1.3, these intermediate results do not require the temperature to go to zero. In particular they hold for constant ε.

3.1. *Theorem 3.1.* The proof is based on an induction over the number of local minima that are contained in the cusp C. Thus we start with a cusp that contains only one minimum, and so no maximum.

3.1.1. *A simple cusp.* We fix throughout this section a cusp C of depth d with only one local minimum x_0 of U. Note that in this case $u = d$, and $x_l = x_r = x_0$. We will prove Theorem 3.1 in this situation, which is pretty similar to the settings of Section 2. Let $\tau_0 = t_0$ and

$$\tau_{i+1} = \min\left(\inf\{t > \tau_i,\ X_t = x_0\},\ \tau_C\right).$$

We first prove:

Lemma 3.3. *There exists c_1 and $\varepsilon_0 > 0$ such that for all $\varepsilon_{t_0} \le \varepsilon_0$, for all $i \ge 1$,*

$$\mathbb{E}\left[\tau_{i+1} - \tau_i \middle|\, \bar{A}_C(\tau_i), \tau_i\right] \le c_1 \sqrt{\varepsilon_{\tau_i}}.$$

Proof: At time τ_i, $i \ge 1$, given $\bar{A}_C(\tau_i)$, the position X_{τ_i} is x_0 and the velocity Y_{τ_i} is either 1 or -1. Writing

$$\mathbb{E}\left[\tau_{i+1} - \tau_i \middle|\, \bar{A}_C(\tau_i), \tau_i\right] = \mathbb{E}\left[\mathbb{E}\left[\tau_{i+1} - \tau_i \middle|\, \bar{A}_C(\tau_i), \tau_i, Y_{\tau_i}\right] \middle|\, \bar{A}_C(\tau_i), \tau_i\right],$$

both cases $Y_{\tau_i} = -1$ or 1 are treated the same way. For instance

$$\mathbb{E}\left[\tau_{i+1} - \tau_i \middle|\, \tau_i, (X_{\tau_i}, Y_{\tau_i}) = (x_0, -1)\right] =$$
$$\mathbb{P}\left(\tau_{i+1} = \tau_C \middle|\, \tau_i, (X_{\tau_i}, Y_{\tau_i}) = (x_0, -1)\right) |x_0 - z_l|$$
$$+ \mathbb{E}\left[(\tau_{i+1} - \tau_i)\mathbb{1}_{\tau_{i+1} \ne \tau_C} \middle|\, \tau_i, (X_{\tau_i}, Y_{\tau_i}) = (x_0, -1)\right].$$

Since the cooling schedule is non-increasing, the probability to reach the boundary and the expectation of the time of jump are bounded by the corresponding quantities at constant temperature ε_{τ_i}. Indeed, if $E \sim \mathcal{E}(1)$, conditionally on τ_i and to $(X_{\tau_i}, Y_{\tau_i}) = (x_0, -1)$,

$$\mathbb{P}\left(\tau_{i+1} = \tau_C\right) = \mathbb{P}\left(E \ge \int_0^{x_0 - z_l} \frac{U'(x_0 - s)}{\varepsilon_{\tau_i + s}} ds\right)$$
$$\le \mathbb{P}\left(E \ge \frac{1}{\varepsilon_{\tau_i}} \int_0^{x_0 - z_l} U'(x_0 - s) ds\right)$$
$$= e^{-\frac{d}{\varepsilon_{\tau_i}}}$$

and

$$\mathbb{E}\left[(\tau_{i+1} - \tau_i)\mathbb{1}_{\tau_{i+1} \ne \tau_C}\right] = 2 \int_0^{x_0 - z_l} \mathbb{P}\left(\frac{\tau_{i+1} - \tau_i}{2} \ge u\right) du$$
$$= 2 \int_0^{x_0 - z_l} \mathbb{P}\left(E \ge \int_0^u \frac{U'(x_0 - s)}{\varepsilon_{\tau_i + s}} ds\right) du$$
$$\le 2 \int_0^{x_0 - z_l} \mathbb{P}\left(E \ge \frac{1}{\varepsilon_{\tau_i}} \int_0^u U'(x_0 - s) ds\right) du$$
$$= 2 \int_0^{x_0 - z_l} \frac{s U'(x_0 - s)}{\varepsilon_{\tau_i}} e^{-\frac{U(x_0 - s) - U(x_0)}{\varepsilon_{\tau_i}}} ds.$$

Thanks to Lemma 2.1 there exist ε_0 and c_1 such that for all $\varepsilon \le \varepsilon_0$,

$$e^{-\frac{d}{\varepsilon}} |x_0 - z_l| + 2 \int_0^{x_0 - z_l} \frac{s U'(x_0 - s)}{\varepsilon} e^{-\frac{U(x_0 - s) - U(x_0)}{\varepsilon}} ds \le c_1 \sqrt{\varepsilon},$$

which concludes the proof since $\varepsilon_{\tau_i} \le \varepsilon_{t_0} \le \varepsilon_0$. $\qquad\qquad\square$

Proof of Part (3.1) *of Theorem* 3.1 *for a simple cusp:*
By choosing ε_0 small enough, for all cooling schedule $\varepsilon_t \leq \varepsilon_0$, the map

$$s \mapsto (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}}$$

is non-increasing, so that

$$\mathbb{E}\left[\int_{\tau_i}^{\tau_{i+1}} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \,\middle|\, \tau_i,\ \bar{A}_C(\tau_i)\right] \leq (\varepsilon_{\tau_i})^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_{\tau_i}}} \mathbb{E}\left[\tau_{i+1} - \tau_i | \tau_i,\ \bar{A}_C(\tau_i)\right]$$

$$\leq c_1 e^{-\frac{d}{\varepsilon_{\tau_i}}}.$$

We define

$$\Phi(i,j) = \mathbb{E}\left[\int_{\tau_i}^{\tau_j} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} \mathbb{1}_{s < \tau_C} ds \,\middle|\, \tau_{i-1},\ \bar{A}_C(\tau_{i-1})\right].$$

Obviously $\Phi(j,j) = 0$; suppose $\Phi(i+1,j) \leq c_1$ has already been proved for some $i < j$. Under the event $\bar{A}_C(\tau_{i-1})$, there are two possibilities: either the process escapes the cusp C between the times τ_{i-1} and τ_i, in which case $\tau_i = \tau_C = \tau_j$ and the integral appearing in the definition of $\Phi(i,j)$ vanishes; or the attempt to exit C between τ_{i-1} and τ_i fails and the process returns to $X_{\tau_i} = x_0$. To sum up,

$$\Phi(i,j) = \mathbb{P}\left(\tau_i < \tau_C \mid \tau_{i-1},\ \bar{A}_C(\tau_{i-1})\right) \mathbb{E}\left[\int_{\tau_i}^{\tau_j} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} \mathbb{1}_{s < \tau_C} ds \,\middle|\, \tau_{i-1},\ \bar{A}_C(\tau_i)\right]$$

$$\leq \left(1 - e^{-\frac{d}{\varepsilon_{\tau_{i-1}}}}\right) \times$$

$$\mathbb{E}\left[\mathbb{E}\left[\int_{\tau_i}^{\tau_{i+1}} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} \mathbb{1}_{s < \tau_C} ds \,\middle|\, \tau_i,\ \bar{A}_C(\tau_i)\right] + \Phi(i+1,j) \,\middle|\, \tau_{i-1}\right]$$

$$\leq \left(1 - e^{-\frac{d}{\varepsilon_{\tau_{i-1}}}}\right) \mathbb{E}\left[c_1 e^{-\frac{d}{\varepsilon_{\tau_i}}} + c_1 \,\middle|\, \tau_{i-1}\right]$$

$$\leq \left(1 - e^{-\frac{d}{\varepsilon_{\tau_{i-1}}}}\right) \left(c_1 e^{-\frac{d}{\varepsilon_{\tau_{i-1}}}} + c_1\right) \leq c_1.$$

This proves by induction that $\Phi(1,j) \leq c_1$ for all $j \geq 1$. As was already pointed out there can only be a finite number of jumps in a finite time, so that the sequence $(\tau_i)_{i \geq 0}$ almost surely converges to τ_C (which, at this point, may be infinite). The monotone convergence theorem yields

$$\mathbb{E}\left[\int_{\tau_1}^{\tau_C} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \,\middle|\, \bar{A}_C(t_0)\right] \leq c_1.$$

On the other hand, as soon as $\varepsilon_t \leq \varepsilon_0$,

$$\mathbb{E}\left[\int_{t_0}^{\tau_1} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \,\middle|\, \bar{A}_C(t_0)\right] \leq (\varepsilon_0)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_0}} \max\left(|x_0 - z_l|,\ |z_r - x_0|\right).$$

Bringing all the pieces together, there exists a constant $c > 0$ such that

$$\mathbb{E}\left[\int_{t_0}^{\tau_C} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \,\middle|\, \bar{A}_C(t_0)\right] \leq c.$$

$\square$

Note that if the cooling schedule is such that $\int_{t_0}^{\infty} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds = \infty$, this first result implies that τ_C is almost surely finite, in which case (X_{τ_C}, Y_{τ_C}) is well defined, and the second half of Theorem 3.1 makes sense.

Proof of Part (3.2) *of Theorem* 3.1 *for a simple cusp:* We note p_i^l (resp. p_i^r) the probability to exit C in one shot, meaning before reaching again x_0, starting at $(X_{\tau_i}, Y_{\tau_i}) = (x_0, -1)$ (resp. $(x_0, 1)$), with $X_{\tau_C} = z_l$ (resp. z_r). Namely, if $E \sim \mathcal{E}(1)$,

$$p_i^l = \mathbb{P}\left(\int_0^{x_0-z_l} \frac{1}{\varepsilon_{\tau_i+s}} U'(x_0 - s) ds \leq E \right)$$

$$p_i^r = \mathbb{P}\left(\int_0^{z_r-x_0} \frac{1}{\varepsilon_{\tau_i+s}} U'(x_0 + s) ds \leq E \right).$$

Since the cooling schedule is not-increasing,

$$e^{-\frac{d}{\varepsilon_{\tau_{i+1}}}} \leq p_i^l, p_i^r \leq e^{-\frac{d}{\varepsilon_{\tau_i}}}.$$

In particular $p_i^l \geq p_{i+1}^r$, which means that between two consecutive attempts to exit the cusp, the second one is always less likely to succeed than the first one, and thus if $(X_{\tau_i}, Y_{\tau_i}) = (x_0, -1)$ then the probability that the exit point will be z_l is greater than $\frac{1}{2}$. On the other hand,

$$\mathbb{P}\left(X_{\tau_C} = z_l \mid (X_{\tau_1}, Y_{\tau_1}) = (x_0, 1) \right) = (1 - p_1^r)\mathbb{P}\left(X_{\tau_C} = z_l \mid (X_{\tau_2}, Y_{\tau_2}) = (x_0, -1) \right)$$

$$\geq \frac{1}{2}\left(1 - e^{-\frac{d}{\varepsilon_{t_0}}} \right).$$

Under the event $\bar{A}_C(t_0)$, necessarily $\tau_1 < \tau_C$, in other words $X_{\tau_1} = x_0$ and Y_{τ_1} is either 1 or -1. Since the previous arguments cover both cases, as soon as $\varepsilon_{t_0} \leq \varepsilon_0$,

$$\mathbb{P}\left(X_{\tau_C} = z_l \mid \bar{A}_C(t_0) \right) \geq \frac{1}{2}\left(1 - e^{-\frac{d}{\varepsilon_0}} \right).$$

The case of z_r is symmetric. □

3.1.2. *Induction.* In this section we consider a general cusp of depth d and we prove Theorem 3.1 under the induction assumption that it holds for any cusp with strictly fewer local maxima than C. There is a finite number of maxima z in C for which $u = U(z_l) - U(z)$, and each connected component of $(x_l, x_r) \setminus \{z, u = U(z_l) - U(z)\}$ is a cusp of depth at most $g = d - u$ (cf. Fig. 3.3). We call $(C_i)_{i=1..n}$ these cusps of depth g. We can consider ε_0 and M such that for all $\varepsilon_t \leq \varepsilon_0$, for all $i = 1, \ldots, n$, for all z_i which is an end of C_i and for all t_0,

$$\mathbb{P}\left(X_{\tau_{C_i}} = z_i \mid \bar{A}_{C_i}(t_0) \right) \geq \frac{1}{M}$$

$$\text{and} \quad \mathbb{E}\left[\int_{t_0}^{\tau_{C_i}} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{g}{\varepsilon_s}} ds \,\middle|\, \bar{A}_{C_i}(t_0) \right] \leq M$$

$$\Rightarrow \quad \mathbb{E}\left[\int_{t_0}^{\tau_{C_i}} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \,\middle|\, \bar{A}_{C_i}(t_0) \right] \leq Me^{-\frac{u}{\varepsilon_{t_0}}}.$$

Proof of Theorem 3.1, *part* (3.1)*:* Set $J_0 = t_0$ and define K_i, J_i for $i \geq 1$ by

$$K_i = \min\left(\inf\{t > J_{i-1}, A_C(t)\}, \tau_C \right)$$

$$J_i = \min\left(\inf\{t > K_i, \bar{A}_C(t)\}, \tau_C \right).$$

At time K_i, the process is ready to escape: it is located near the boundary of C, and its velocity is directed toward the exit. At time J_i, if it failed to leave the cusp, it goes back to the C_j's (see Fig. 3.3) and, as previously,
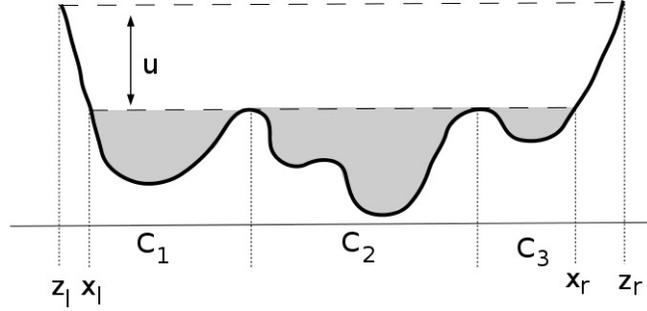
$$\Phi(i, j)$$

FIGURE 3.3. The cusp C is divided in several smaller cusps.

$$
\begin{aligned}
&:= \quad \mathbb{E}\left[\left.\int_{K_i}^{K_j}\left(\varepsilon_s\right)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds\,\right|\, K_{i-1},\ \tau_C > K_{i-1}\right] \\[2mm]
&= \quad \mathbb{P}\left(\tau_C > K_i \mid K_{i-1},\ \tau_C > K_{i-1}\right) \mathbb{E}\left[\left.\int_{K_i}^{K_j}\left(\varepsilon_s\right)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds\,\right|\, K_{i-1},\ \tau_C > K_i\right] \\[2mm]
&\leq \quad \left(1 - e^{-\frac{u}{\varepsilon_{K_{i-1}}}}\right) \mathbb{E}\left[\left.\int_{K_i}^{J_i} + \int_{J_i}^{K_{i+1}} + \int_{K_{i+1}}^{K_j}\left(\varepsilon_s\right)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds\,\right|\, K_{i-1},\ \tau_C > K_i\right].
\end{aligned}
$$

We treat the three parts of the integral one after another. From Lemma 3.3, there exists a constant c_1 (depending on C but not on $(\varepsilon_t)_{t\geq t_0}$) such that

$$
\mathbb{E}\left[\left.\int_{K_i}^{J_i}\left(\varepsilon_s\right)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds\,\right|\, K_i,\ \tau_C > K_i\right] \leq c_1 e^{-\frac{u}{\varepsilon_{K_i}}}.
$$

As far as the time interval $[J_i, K_{i+1}]$ is concerned, we distinguish the following cases: if $J_i = \tau_C$, the process has escaped from C, hence $K_{i+1} = \tau_C$ and the integral vanishes. Otherwise the process goes back down to a cusp C_i, and then for a while pass from a C_j to another, until it reaches A_C again. The contribution of the time passed in each C_i to the integral can be controlled thanks to the induction assumption. It remains to make sure the number of transitions between the C_j's before existing is not too large. When the process exits C_i, the probability it does so through a given end is bounded below thanks to the induction assumption. Thus the expectation of the number N of jumps from a C_i to another C_j before the process reaches A_C is bounded by a constant D which does not depend on the cooling schedule. Hence, denoting by $\tau_{C_i}^*$ an entry time of the process in C_i,

$$
\begin{aligned}
&\mathbb{E}\left[\left.\int_{J_i}^{K_{i+1}}\left(\varepsilon_s\right)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds\,\right|\, K_i,\ \tau_C > K_i\right] \\[2mm]
&\leq \quad \mathbb{E}\left[\left. N \max_i \mathbb{E}\left[\left.\int_{\tau_{C_i}^*}^{\tau_{C_i}}\left(\varepsilon_s\right)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds\,\right|\, \tau_{C_i}^* > K_i,\ \bar{A}_{C_i}\left(\tau_{C_i}^*\right)\right]\,\right|\, K_i,\ \tau_C > J_i\right] \\[2mm]
&\leq \quad DM e^{-\frac{u}{\varepsilon_{K_i}}}.
\end{aligned}
$$

Since $\Phi(j,j) = 0$, as an induction assumption we can suppose

$$\Phi(i+1,j) \le DM + c_1,$$

and thus we get

$$
\begin{aligned}
\Phi(i,j) &\le \left(1 - e^{-\frac{u}{\varepsilon_{K_{i-1}}}}\right) \mathbb{E}\left[e^{-\frac{u}{\varepsilon_{K_i}}}(DM + c_1) + \Phi(i+1,j)\Big| K_{i-1},\ \tau_C > K_i\right]. \\
&\le \left(1 - e^{-\frac{u}{\varepsilon_{K_{i-1}}}}\right)\left(e^{-\frac{u}{\varepsilon_{K_{i-1}}}} + 1\right)(DM + c_1) \\
&\le DM + c_1.
\end{aligned}
$$

The monotone convergence Theorem yields

$$\mathbb{E}\left[\left.\int_{K_1}^{\tau_C}(\varepsilon_s)^{-\frac{1}{2}}e^{-\frac{d}{\varepsilon_s}}ds\right|\ \bar{A}_C(t_0)\right] \le DM + c_1.$$

On the other hand in the event $\bar{A}_C(t_0)$, either the process start in a C_i, or it reaches a C_i in a time bounded by $\max(|x_l - z_l|,\ |z_r - x_r|)$. In both cases with the previous argument we used for $[J_i, K_{i+1}]$,

$$\mathbb{E}\left[\left.\int_{t_0}^{K_1}(\varepsilon_s)^{-\frac{1}{2}}e^{-\frac{d}{\varepsilon_s}}ds\right|\ \bar{A}_C(t_0)\right] \le DM + (\varepsilon_{t_0})^{-\frac{1}{2}}e^{-\frac{d}{\varepsilon_{t_0}}}\max\left(|x_l - z_l|,\ |z_r - x_r|\right),$$

and ultimately part (3.1) of Theorem 3.1 is proved with

$$c = 2DM + c_1 + (\varepsilon_{t_0})^{-\frac{1}{2}}e^{-\frac{d}{\varepsilon_{t_0}}}\max\left(|x_l - z_l|,\ |z_r - x_r|\right).$$

$\square$

The same remark as in Section $3.1.1$ holds: when $\int_{t_0}^{\infty}(\varepsilon_s)^{-\frac{1}{2}}e^{-\frac{d}{\varepsilon_s}}ds = \infty$, the first part of Theorem 3.1 implies that τ_C is almost surely finite, and X_{τ_C} well-defined.

Proof of Theorem 3.1, part (3.2): The situation is very similar to the simple cusp one. Take $M_0' = t_0$ and

$$
\begin{aligned}
M_i &= \min\left(\inf\{t > M_i',\ A_C(t)\},\ \tau_C\right) \\
M_{i+1}' &= \min\left(\inf\{t > M_i,\ \bar{A}_C(t)\},\ \tau_C\right).
\end{aligned}
$$

Given the sequence $(M_i)_{i\ge 1}$, $(X_{M_i})_{i\ge 1}$ is an inhomogeneous Markov chain on the set $\{z_l, x_l, x_r, z_r\}$. Let

$$
\begin{aligned}
p_i^l &= \mathbb{P}\left(\int_0^{x_l - z_l}\frac{1}{\varepsilon_{M_i + s}}U'(x_0 - s)ds \le E\right), \\
p_i^r &= \mathbb{P}\left(\int_0^{z_r - x_r}\frac{1}{\varepsilon_{M_i + s}}U'(x_0 + s)ds \le E\right).
\end{aligned}
$$

The induction assumption on the C_j's implies that the transitions of this chain from x_r to x_r and vice-versa are bounded below by a constant $h > 0$ which does not depend on the cooling schedule. Hence $(X_{M_i})_{i\ge 1}$ is more likely to reach z_l before z_r than the chain $(\tilde{X}_{M_i})_{i\ge 1}$ with transition

$$
\begin{array}{llll}
\mathbb{P}(x_l \to x_r) &= 1 - p_i^l & \mathbb{P}(x_r \to x_l) &= h \\
\mathbb{P}(x_l \to z_l) &= p_i^l & \mathbb{P}(x_r \to z_r) &= p_i^r \\
\mathbb{P}(x_l \to x_l) &= 0 & \mathbb{P}(x_r \to x_r) &= 1 - h - p_i^r.
\end{array}
$$

Similarly to the situation in Section 3.1.1, $p_i^l \geq p_j^r$ if $j > i$, and $p_1^r \leq e^{-\frac{u}{\varepsilon_0}}$. The probability that $\tilde{X}$, starting from $\tilde{X}_{M_{i+1}} = x_r$, hits z_r before x_l is

$$\sum_{j \geq 1} p_{i+j}^r \prod_{1 \leq k < j} (1 - h - p_k^r) \leq \sum_{j \geq 1} p_{i+1}^r (1-h)^{j-1} \leq \frac{p_i^l}{1-h}.$$

In particular this is less than $\frac{e^{-\frac{u}{\varepsilon_0}}}{1-h}$, so that the probability that there exist $k_0 \geq 0$ such that $\tilde{X}_{M_{k_0}} = x_l$ is bounded below. Hence to bound the probability to reach z_l before z_r we can assume the initial point is x_l. Let

$$k_{i+1} = \inf \{n > k_i, \ \tilde{X}_{M_n} \neq x_r\}.$$

Given the times (M_{k_i}), the chain $\tilde{X}$ is more likely to reach z_l before z_r than the chain X' with transition

$$\mathbb{P}(x_l \to z_l) = p_{k_i}^l \qquad \mathbb{P}(x_l \to z_r) = \frac{p_{k_i}^l}{1-h} \qquad \mathbb{P}(x_l \to x_l) = 1 - p_{k_i}^l - \frac{p_{k_i}^l}{1-h}.$$

Finally, X' goes to z_l rather than z_r with a probability $\frac{1}{1+\frac{1}{1-h}}$, which does not depend on the cooling schedule, which concludes.   □

3.2. *Theorem 3.2.* We start by some preliminary lemmas. Let C be a cusp of depth d, recall that its bottom is $B = \{x \in C, U(z) \geq U(x) \ \forall z \in C\}$ and let $x_b = \min B$. We will give an upper bound of the probability that the process, starting at time t_0 at point $(x_b, -1)$, reaches z_l before $(x_b, 1)$. Obviously, if there were no local maximum between x_b and z_l, this probability would be bounded by $e^{-\frac{d}{\varepsilon_{t_0}}}$. To prove a similar bound in more general cases, we decompose C in the following way: let

$$J = \{x \in [z_l, x_b], \text{ s.t. } \forall z \in (x, x_b], \ U(z) < U(x)\}.$$

On the set J, U is non-increasing, and the connected component of the closure of $[z_l, x_b] \setminus J$ are cusps whose right end are local maxima. We call $u_1 > u_2 > \cdots > u_q$ this local maxima, $C_1, \ldots, C_{q-1}$ the corresponding cusps, $u_0 = x_b$, $C_0 = \{x_b\}$, $u_{q+1} = z_l$ and $C_{q+1} = \{z_l\}$. The point of taking the connected component of *the closure* of $[z_l, x_b] \setminus J$ was to ensure $U(u_i) > U(u_j)$ if $i > j$. We say $U(u_i)$ is the energy level of C_i, denoted by E_i, and we note

$$\delta_i = E_{i+1} - E_i.$$

Note that $\sum_{i=0}^q \delta_i = E_{q+1} - E_0 = d$. The situation is represented in Figure 3.4. Let $t_0 > 0$, $1 \leq i < q$ and suppose $\bar{A}_{C_i}(t_0)$ holds. Let

$$s = \inf\{t > t_0 \text{ s.t. } \exists j \neq i, \ X_t \in C_j\}.$$

Lemma 3.4. *There exists $c_2 > 0$ which depends on C but not on the cooling schedule so that if $\bar{A}_{C_i}(t_0)$ holds,*

$$\mathbb{P}\left(U(X_s) = E_{i+1}\right) \quad \leq \quad c_2 e^{-\frac{\delta_i}{\varepsilon_{t_0}}}. \tag{3.3}$$

Proof: Since there is a finite number of cusps C_i, it is enough to show such a constant exists for any one of them. Starting from A_{C_i}, if the process exits C_i to the right, it will deterministically falls down to C_{i-1}, which means $U(X_s) = E_{i-1}$. If the process exits to the left, it reaches C_{i+1} in one shot with probability less than
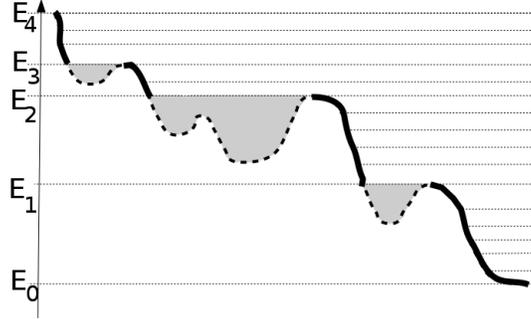
FIGURE 3.4. If the light comes from the right, J, in bold line, is the sunny part, and the C_i's are in the shadow.

$e^{-\frac{\delta_i}{\varepsilon_{t_0}}}$, else goes back down to $\bar{A}_{C_i}$. Since the probability the process escapes to the left, starting from $\bar{A}_{C_i}$, is bounded by a constant $h > 0$ thanks to Theorem 3.1,

$$\mathbb{P}\left(U(X_s) = E_{i+1}\right) \leq \sum_{k \geq 0} (1-h)^k e^{-\frac{\delta_i}{\varepsilon_{t_0}}} = \frac{e^{-\frac{\delta_i}{\varepsilon_{t_0}}}}{1-h}.$$

$\square$

Remark: *The constant c_2 is defined from h, which depends only on the C_i's. Thus If $(D_l)_{l \geq 0}$ is a family of nested cusps such that for all $k \geq 0$ all the minima of D_k belongs to $\bigcap_{l \geq 0} D_l$, the constant c_2 may be defined uniformly on k so that (3.3) holds for all D_k's. Since q (the number of small cusps) is also the same for all D_k's, this remark will extend to the next result.*

This lemma implies the process, starting from $\bar{A}_{C_i}$, is less likely to hit $(z_l, -1)$ before $(x_0, 1)$ than the birth-death process $(W_n)_{n \geq 0}$ on $[0, q+1]$ with transition probabilities

$$\mathbb{P}(j \to j+1) = c_2 e^{-\frac{\delta_j}{\varepsilon_{t_0}}}, \qquad \mathbb{P}(j \to j-1) = \left(1 - c_2 e^{-\frac{\delta_j}{\varepsilon_{t_0}}}\right), \qquad j \in \{1, ..., q\},$$

is to hit $q+1$ before 0, starting at i.

Lemma 3.5. *For ε_{t_0} small enough,*

$$\mathbb{P}\left(W \text{ hits } q+1 \text{ before } 0 \mid W_0 = 1\right) \leq (2c_2)^q e^{-\frac{E_{q+1} - E_1}{\varepsilon_{t_0}}}.$$

As a consequence,

$$\mathbb{P}\left(X \text{ hits } z_l \text{ before } x_b \mid \bar{A}_{C_1}(t_0)\right) \leq (2c_2)^q e^{-\frac{E_{q+1} - E_1}{\varepsilon_{t_0}}}$$

and

$$\mathbb{P}\left((X, Y) \text{ hits } (z_l, -1) \text{ before } (x_b, 1) \mid (X_{t_0}, Y_{t_0}) = (x_b, -1)\right) \leq (2c_2)^q e^{-\frac{d}{\varepsilon_{t_0}}}.$$

Proof: More generally, let

$$r_i = \mathbb{P}\left(W \text{ hits } q+1 \text{ before } i-1 \mid W_0 = i\right).$$

If W hits q before $i-1$, necessarily $W_1 = i+1$ (which occurs with probability $c_2 e^{-\frac{\delta_i}{\varepsilon_{t_0}}}$) and either the chain stays above i (with probability r_{i+1}), either it goes back at some point to i and then we are back to the initial problem. Thus

$$r_i = c_2 e^{-\frac{\delta_i}{\varepsilon_{t_0}}} \left(r_{i+1} + (1 - r_{i+1}) r_i \right).$$

Suppose ε_{t_0} is small enough to have $c_2 e^{-\frac{\delta_i}{\varepsilon_{t_0}}} \leq \frac{1}{2}$ for all $i = 1, .., q$, so that

$$r_i \leq 2 c_2 e^{-\frac{\delta_i}{\varepsilon_{t_0}}} r_{i+1},$$

which concludes since $r_{q+1} = 1$.   □

We only considered here the left part of C, but the same goes for the right one. If $x_b' = \max B$ then there exist c_2' and q' such that

$$\mathbb{P}\left((X, Y) \text{ hits } (z_r, 1) \text{ before } (x_b', -1) \mid (X_{t_0}, Y_{t_0}) = (x_b', 1)\right) \leq (2c_2')^{q'} e^{-\frac{d}{\varepsilon_{t_0}}},$$

and we write $c_3 = \max\left((2c_2)^q, \; (2c_2')^{q'}\right)$.

Proof of Theorem 3.2: Let $a > 0$ be small enough such that U is non-decreasing (resp. non-increasing) on $\{x_b - a, x_b\}$ (resp. $\{x_b', x_b' + a\}$) and that $U(z_l) - \max\left(U(x_b - a), U(x_b' + a)\right) \geq \frac{d}{2}$. That way,

$$\mathbb{P}\left((X, Y) \text{ hits } (z_l, 1) \text{ before } (x_b, 1) \mid (X_{t_0}, Y_{t_0}) = (x_b - a, -1)\right) \leq c_3 e^{-\frac{d}{2\varepsilon_{t_0}}}$$

(and similarly in the right part of C). It means if at some time t_1 the temperature has been divided by two since t_0, then from t_1 we have the same bound on the probability of success of an attempt to leave C starting from $(x_b - a, -1)$ than we had for an attempt to leave C starting from $(x_b, -1)$ at initial temperature.

Let $t_j = t_0 + 2aj$ and $K_0 = t_0$, suppose K_j has been defined for some $j \in \mathbb{N}$, let $S_{0,j} = K_j$ and for $i \geq 0$

$$\widetilde{S}_{i+1,j} = \inf\left\{t > S_{i,j} \text{ s.t. } \left((X_t, Y_t) \in \{(x_b, -1), (x_b', 1)\} \text{ and } \varepsilon_t \geq \frac{1}{2}\varepsilon_{t_j}\right)\right\}$$

$$\overline{S}_{i+1,j} = \inf\left\{t > S_{i,j} \text{ s.t. } \left((X_t, Y_t) \in \{(x_b - a, -1), (x_b' + a, 1)\} \text{ and } \varepsilon_t < \frac{1}{2}\varepsilon_{t_j}\right)\right\}$$

$$S_{i+1,j} = \min\left(\tau_C, \widetilde{S}_{i+1,j}, \overline{S}_{i+1,j}\right).$$

Let $N_j = 0$ if $K_j = \tau_C$ and else let

$$\begin{aligned}
N_j &= \inf\{n \geq 1, \; S_{n,j} \geq K_j + 2a \text{ or } S_{n,j} = \tau_C\} \\
K_{j+1} &= S_{N_j, j}.
\end{aligned}$$

The situation is the following: while the process has not escaped C yet, the $S_{i,j}$'s are starting times of new attempts to leave with, for all $i, j \in \mathbb{N}^2$

$$\mathbb{P}\left(S_{i+1,j} = \tau_C \mid S_{i,j}, \; S_{i,j} < \tau_C\right) \leq c_3 e^{-\frac{d}{2\varepsilon S_{i,j}}}.$$

A problem is that we don't control the way the temperature evolves with i, j, and this is why we introduced the K_j's, which are the times at which we update the

temperature. Indeed $K_j \geq t_j$ so that for all $i, j \in \mathbb{N}^2$, $\varepsilon_{S_{i,j}} \leq \varepsilon_{t_j}$, which implies $\left(\mathbb{1}_{S_{i \wedge N_j,j} < \tau_C} + (i \wedge N_j) c_3 e^{-\frac{d}{2\varepsilon_{t_j}}} \right)_{i \geq 0}$ is a submartingale and

$$
\begin{aligned}
\mathbb{P}\left(K_{j+1} < \tau_C \mid K_j, \ K_j < \tau_C\right) &= 1 + \mathbb{E}\left(\mathbb{1}_{K_{j+1} < \tau_C} - \mathbb{1}_{K_j < \tau_C} \mid K_j, \ K_j < \tau_C \right) \\
&\geq 1 - c_3 e^{-\frac{d}{2\varepsilon_{t_j}}} \mathbb{E}\left(N_j \mid K_j, \ K_j < \tau_C \right).
\end{aligned}
$$

In the first instance suppose there exists m (which does not depend on the cooling schedule) such that for all $j \geq 0$

$$
\mathbb{E}\left(N_j \mid K_j, \ K_j < \tau_C \right) \leq \frac{m}{\sqrt{\varepsilon_{t_j}}}. \tag{3.4}
$$

In that case by induction we get

$$
\begin{aligned}
\mathbb{P}\left(K_{j+1} < \tau_C\right) &= \mathbb{E}\left(\mathbb{P}\left(K_{j+1} < \tau_C \mid K_j, \ K_j < \tau_C\right) \mathbb{P}\left(K_j < \tau_C\right) \right) \\
&\geq \left(1 - \frac{mc_3}{\sqrt{\varepsilon_{t_j}}} e^{-\frac{d}{2\varepsilon_{t_0+j}}} \right) \mathbb{P}\left(K_j < \tau_C\right) \\
&\geq \prod_{k=0}^{j} \left(1 - \frac{mc_3}{\sqrt{\varepsilon_{t_k}}} e^{-\frac{d}{2\varepsilon_{t_k}}} \right).
\end{aligned}
$$

Suppose ε_{t_0} is small enough so that $mc_3 \left(\varepsilon_{t_0}\right)^{-\frac{1}{2}} e^{-\frac{d}{2\varepsilon_{t_0}}}$ is less than the positive solution of $1 - z = e^{-2z}$, and so that $\varepsilon \mapsto \varepsilon^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon}}$ is increasing on $(0, \varepsilon_{t_0})$. Hence for $j \geq 1$

$$
\begin{aligned}
\mathbb{P}\left(t_j < \tau_C\right) &\geq \mathbb{P}\left(K_{j+1} < \tau_C\right) \\
&\geq \exp\left(-2mc_3 \sum_{k=0}^{j-1} \frac{e^{-\frac{d}{2\varepsilon_{t_k}}}}{\sqrt{\varepsilon_{t_k}}} \right) \\
&\geq \exp\left(-\frac{mc_3}{a} \int_{t_0}^{t_j} \frac{e^{-\frac{d}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du \right),
\end{aligned}
$$

and for all $r > 0$,

$$
\begin{aligned}
\mathbb{P}\left(r < \tau_C\right) &\geq \mathbb{P}\left(t_0 + 2a \left\lceil \frac{r - t_0}{2a} \right\rceil < \tau_C \right) \\
&\geq \exp\left(-2mc_3 \left(\frac{e^{-\frac{d}{\varepsilon_{t_0}}}}{\sqrt{\varepsilon_{t_0}}} + \frac{1}{2a} \int_{t_0}^{r} \frac{e^{-\frac{d}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du \right) \right).
\end{aligned}
$$

It remains to prove (3.4), which states that between two consecutive updates of the temperature, there are not too many attempts to leave. Intuitively, if the temperature decays slowly then the inter-jump times are of order $\sqrt{\varepsilon_{t_j}}$ and so the number of attempts in a fixed duration should of the order $\varepsilon_{t_j}^{-\frac{1}{2}}$. On the other hand if the temperature falls rapidly then we only take into account attempts that starts from $(x_b - a, -1)$ or $(x_b + a, 1)$, and between two such attempts the process has to cover a distance of at least $2a$.

More precisely, since the process moves at constant speed, for all i, j, $\overline{S}_{i+1,j} - \overline{S}_{i,j} > 2a$, so that

$$N_j \leq 1 + \inf\left\{ n \geq 1, \ \widetilde{S}_{n,j} \geq K_j + 2a \ \text{or} \ \varepsilon_{\widetilde{S}_{n,j}} \leq \frac{1}{2}\varepsilon_{t_j} \right\}.$$

Let $\rho > 0$ be such that $U(x_b - s) \leq U(x_b) + \rho(x_b - s)^2$ for $s \leq |x_b - z_l|$ (and similarly in the right part of C; the situation being the same, we only consider the left part in the following). Starting from $(x_b, -1)$ at time $S_{i,j}$ the next jump time T of the process, defined by (1.3) from an exponential r.v. E, satisfies,

$$T \geq \inf\left\{ s > S_{i,j}, \ E < \frac{\rho s^2}{\frac{1}{2}\varepsilon_{t_j}} \right\} \wedge \inf\left\{ s > S_{i,j}, \ \varepsilon_s \leq \frac{1}{2}\varepsilon_{t_j} \right\}$$

and thus $\widetilde{S}_{i+1,j} - S_{i,j} \geq 2T$ and

$$N_j \leq 1 + \inf\left\{ n \geq 1, \sqrt{\frac{\varepsilon_{t_j}}{2\rho}} \sum_{i=1}^{n} \sqrt{E_i} \geq 2a \right\}$$

where $(E_i)_{i \geq 1}$ is a sequence of i.i.d. random variables with law $\mathcal{E}(1)$, independent from the past $(t \leq K_j)$. Now

$$N_j \ \leq \ 1 + \left\lceil \frac{4a\sqrt{\rho}}{\mathbb{E}\left(\sqrt{E}\right)\sqrt{\varepsilon_{t_j}}} \right\rceil + \inf\left\{ n \geq 1, \frac{1}{n}\sum_{i=1}^{n}\sqrt{E_i} \geq \frac{1}{2}\mathbb{E}\left(\sqrt{E}\right) \right\}.$$

Note that $Z_k = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{\sqrt{E_i}}{\mathbb{E}(\sqrt{E})} - 1 \right)$ satisfies a Large Deviation Principle, so that

$$\mathbb{E}\left(N_j \mid K_j, \ K_j < \tau_C \right) \ \leq \ 1 + \left\lceil \frac{4a\sqrt{\rho}}{\mathbb{E}\left(\sqrt{E}\right)\sqrt{\varepsilon_{t_j}}} \right\rceil + \sum_{k \geq 1}\mathbb{P}\left(Z_k \leq -\frac{1}{2} \right)$$

$$\leq \ \frac{m}{\sqrt{\varepsilon_{t_j}}}$$

for some $m < \infty$ as long as ε_{t_0} is small enough.   □

Remark: *Here the constant c_4 only depends on $U''(x_b)$ and $U''(x_b')$, and m only depends on c_4. Furthermore, as has already been noticed, c_3 only depends on the internal sub-cusps of C. Thus, if $(D_l)_{l \geq 0}$ is a family of cusps so that for all $k \geq 0$ all the minima of D_k belong to $\bigcap_{l \geq 0} D_l$, the constant Γ in Theorem 3.2 may be chosen uniformly over k. And if U is a potential with a finite number of local minima, going to $+\infty$ at $\pm\infty$, Γ may be chosen uniformly over all cusps of U.*

3.3. *Proof of the NS condition.* Now that Theorems 3.1 and 3.2 are established, we recall (and slightly adapt) the arguments from Hajek (1988) to prove Theorem 1.3.

Let $E \geq 0$ and

$$W_E \ = \ \{x \in \mathbb{R}, \ x \text{ is a local minimum of depth strictly larger than } E\}$$
$$R_E \ = \ \{x \in \mathbb{R}, \ x \text{ is reachable from } y \text{ at height } V(y) + E \text{ for some } y \in W_E\}$$
$$J \ = \ \{x \in \mathbb{R}, \ W_E \text{ is reachable from } x \text{ at height } V(x)\}.$$

If $x \notin J$, then the set of points which are reachable from x at height $V(x)$ is a cusp of depth at most E (else its bottom would be in W_E and reachable from x at height $V(x)$) which does not intersect J (if a point y were in the intersection, we could reach y from x at height $V(x)$ and then reach a point in W_E from y at height $V(y) \le V(x)$). Thus the connected component of $\mathbb{R} \setminus J$ which contains x is itself a cusp of depth at most E. Moreover one of its end is a local maximum (else it could be thicken without intersecting J) which means there are finitely many connected components of $\mathbb{R} \setminus J$; we call them $C_1, \ldots, C_n$.

Lemma 3.6. *Let $E > 0$. If*

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds = \infty$$

then

$$\lim_{t \to \infty} \mathbb{P}(X_t \in R_E) = 1.$$

Proof: Note that there are a finite number of minima in W_E, so that there exists $\gamma_0 > 0$ such that all those are of depth larger than $E + \gamma_0$. We thicken R_E, for $\gamma < \gamma_0$, as

$$R_{E,\gamma} = \{x \in \mathbb{R}, \ x \text{ is reachable from } y \text{ at height } V(y) + E + \gamma \text{ for some } y \in W_E\}.$$

Since $R_E = \bigcap_{\gamma > 0} R_{E,\gamma}$, we only need to prove the result for an arbitrary $\gamma < \gamma_0$. Let $t_0 \ge 0$, $A_0 = t_0$ and

$$
\begin{aligned}
B_i &= \inf\{t > A_i, \ X_t \notin J\}, \\
A_{i+1} &= \min\{t > B_i, \ X_t \in J\}, \\
\alpha &= \inf\{k \ge 0, \ X_{A_k} \in W_E\}, \\
\beta &= \inf\{k > \alpha, \ X_{B_k} \notin R_{E,\gamma}\}.
\end{aligned}
$$

Since there is only a finite number of C_j, we can consider ε_0 and $c > 0$ in Theorem 3.1 such that if $\varepsilon_{t_0} \le \varepsilon_0$, for all $i \ge 0$,

$$\mathbb{E}\left[\int_{B_i}^{A_{i+1}} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds \,\middle|\, B_i \right] \le c.$$

(Note that this would be false for $E = 0$).

On the other hand, note that on each connected component of $J \setminus W_E$, U is monotone. There are two such components which are infinite, and where the potential is convex, so that the expected time the process stays there is bounded by a constant which only depends on ε_0. The time the process stays in a compact connected component of $J \setminus W_E$ is bounded by twice the length of the component, and there is a finite number of such components. Thus there exist c' such that if $\varepsilon_{t_0} \le \varepsilon_0$, for all $i \ge 0$,

$$\mathbb{E}\left[\int_{A_i}^{B_i} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds \,\middle|\, A_i \right] \le c'.$$

Moreover as a consequence of part (3.2) of Theorem 3.1, $\mathbb{E}(\alpha)$ is bounded by a constant which does not depend on the cooling schedule, and finally there exists $\tilde{c} > 0$ such that

$$\mathbb{E}\left[\int_{t_0}^{A_\alpha} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds \right] \le \tilde{c},$$

so that

$$\mathbb{P}\left(A_\alpha \geq r\right) \;\; \leq \;\; \frac{\tilde{c}}{\int_{t_0}^r (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds}.$$

At time A_α, the process attains the bottom of a cusp C (and therefore is in $\bar{A}_C$) of depth $E + \gamma$ which is included in $R_{E,\gamma}$. Thanks to Theorem 3.2,

$$\mathbb{P}\left(A_\beta \geq r\right) \;\; \geq \;\; \exp\left(-\Gamma\left(\frac{e^{-\frac{E+\gamma}{\varepsilon_{t_0}}}}{\sqrt{\varepsilon_{t_0}}} + \int_{t_0}^r \frac{e^{-\frac{E+\gamma}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du\right)\right)$$

$$\geq \;\; \exp\left(-\Gamma e^{-\frac{\gamma}{\varepsilon_{t_0}}}\left(\frac{e^{-\frac{E}{\varepsilon_{t_0}}}}{\sqrt{\varepsilon_{t_0}}} + \int_{t_0}^r \frac{e^{-\frac{E}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du\right)\right).$$

Thus, for any t_0 and $r \geq t_0$,

$$\mathbb{P}\left(X_r \in R_{E,\gamma}\right) \geq \mathbb{P}\left(A_\beta \geq r \geq A_\alpha\right)$$

$$\geq \exp\left(-\Gamma e^{-\frac{\gamma}{\varepsilon_{t_0}}}\left(\frac{e^{-\frac{E}{\varepsilon_{t_0}}}}{\sqrt{\varepsilon_{t_0}}} + \int_{t_0}^r \frac{e^{-\frac{E}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du\right)\right) - \frac{\tilde{c}}{\int_{t_0}^r (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds}.$$

Let $h(t)$ be defined for any $t \geq 0$ by

$$\int_t^{h(t)} \frac{e^{-\frac{E}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du = \frac{1}{\varepsilon_t}.$$

As a strictly increasing function it is invertible, and in particular $(t \to \infty) \Leftrightarrow (h(t) \to \infty)$.

$$\mathbb{P}\left(X_{h(t)} \in R_{E,\gamma}\right) \quad \geq \quad \exp\left(-\Gamma e^{-\frac{\gamma}{\varepsilon_t}}\left(\frac{e^{-\frac{E}{\varepsilon_t}}}{\sqrt{\varepsilon_t}} + \frac{1}{\varepsilon_t}\right)\right) - \tilde{c}\varepsilon_t$$

$$\underset{h(t)\to\infty}{\longrightarrow} \quad 1.$$

$\square$

Proof of Theorem 1.3: We treat first the case of fast cooling, namely we assume that for all $\delta > 0$,

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{\delta}{\varepsilon_s}} ds < \infty.$$

Let x be a local minimum of U, $\delta > 0$ and $C_\delta(x)$ be the set of points which are reachable from x at height $U(x) + \delta$. Any neighbourhood of x contains $C_\delta(x)$ for δ small enough. If at some time t_0 the process enters $C_\delta(x)$, Theorem 3.2 yields

$$\mathbb{P}\left(X_t \in C_\delta(x) \,\forall t \geq t_0\right) \geq \exp\left(-\Gamma\left(\frac{e^{-\frac{\delta}{\varepsilon_{t_0}}}}{\sqrt{\varepsilon_{t_0}}} + \int_{t_0}^\infty \frac{e^{-\frac{\delta}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du\right)\right) > 0.$$

Thus, each time the process reaches a local minimum x, it has a positive probability to stay trapped forever in a neighbourhood of x. If it escapes, almost surely it will reach another local minimum later. Thus, the probability that it get trapped at some time is 1, and the probability that it's already been trapped at time t goes to 1 as t goes to infinity, so that, if S is a neighborhood of all local minima of U,

$$\lim_{t\to\infty} \mathbb{P}\left(X_t \in S\right) = 1.$$

Moreover, for any local minimum x, as the temperature is not allowed to vanish, there is a non-zero probability to reach x, and so to stay trapped in $C_\delta(x)$, which yields

$$\liminf_{t\to\infty} \mathbb{P}\left(X_t \in C_\delta(x)\right) > 0.$$

Thus, we have proved the part 1 and the "if" ($\Rightarrow$) half of part 2 of Theorem 1.3 in the case of fast cooling. Finally, the "only if" half of part 2 is tautological in this case.

Concerning the part 3, recall that we have supposed there is at least one non-global minimum $\tilde{x}$, near which the process has a non-zero probability to stay forever. In this event $U(X_t) \geq U(\tilde{x}) > \min_{\mathbb{R}} U + \delta$ for δ small enough, so that

$$\lim_{t\to\infty} \mathbb{P}\left(U(X_t) < \min_{\mathbb{R}} U + \delta\right) \quad \leq \quad 1 - \mathbb{P}\left(X_t \text{ gets trapped near } \tilde{x}\right) \quad < \quad 1.$$

Now we turn to slow cooling, namely we suppose there exists $F > 0$ such that $\forall \delta \neq F$,

$$\left(\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{\delta}{\varepsilon_s}} ds = \infty\right) \quad \Leftrightarrow \quad (\delta < F).$$

Then, for all $\delta < F$, according to Lemma 3.6,

$$\lim_{t\to\infty} \mathbb{P}\left(X_t \in R_\delta\right) = 1.$$

Since any neighbourhood of all local minima contains R_δ for δ small enough, part 1 of Theorem 1.3 is proved.

Let $E > 0$ be such that

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds = \infty,$$

and S be a neighbourhood of all minima of depth E such that S^c is a neighbourhood of all other minima. We want to prove

$$\lim_{t\to\infty} \mathbb{P}\left(X_t \in S\right) = 0.$$

Lemma 3.6 together with the first part of Theorem 1.3 we have just proven implies that for any neighbourhood $\mathcal{U}$ of the minima in R_E,

$$\lim_{t\to\infty} \mathbb{P}\left(X_t \in \mathcal{U}\right) = 1.$$

Thus, since S^c is a neighbourhood of all the minima it contains, it is enough to prove that R_E does not contain any minimum of depth exactly E.

Let x be such a minimum, and let z be such that $U(z) < U(x)$ and z is reachable from x at height $U(x) + E$. Suppose $x \in R_E$, and let $y \in W_E$ be such that x is reachable from y at height $U(y) + E$.

If $U(y) < U(x)$, since y is reachable from x at depth $U(y) + E < U(x) + E$, by definition of the depth of a local minimum, it means x is of depth strictly less than E, which is a contradiction.

On the other hand if $U(y) \geq U(x)$, then z is reachable from y at height $U(y) + E$, while $U(z) < U(y)$, which is contradictory with the fact that y is of depth strictly larger than E.

This means R_E does not contain any local minimum of depth E.

At this point we have proven part 1 and implication $\Leftarrow$ of part 2 of Theorem 1.3, which we will use to prove the converse.

Now suppose

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds < \infty.$$

In particular $E \geq F$. Let x be a minimum of depth E. As a first step, assume

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{F}{\varepsilon_s}} ds < \infty.$$

Let C be the set of all points which are reachable from x at height strictly less than $U(x) + F$ (if $F = E$, $C = C_x$). Then C is a cusp of depth F, whose bottom B is constituted of minima of depth exactly E, and the depth of any other minimum in C is strictly less than F (since B is reachable from them without leaving C).

From Theorem 3.2, the process has a non-zero probability to stay trapped forever in C, so that

$$\liminf_{t \to \infty} \mathbb{P}(X_t \in C) > 0.$$

On the other hand, if $S \subset C$ is a neighbourhood of B, since the depth d of any minimum in $C \setminus S$ satisfies $d < F$ and so

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds = \infty,$$

from part 1 and implication $\Leftarrow$ of part 2 of Theorem 1.3

$$\lim_{t \to \infty} \mathbb{P}(X_t \in C \setminus S) = 0.$$

Hence,

$$\liminf_{t \to \infty} \mathbb{P}(X_t \in S) > 0$$

which ends the proof of part 2.

A slight adaptation is needed when

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{F}{\varepsilon_s}} ds = \infty.$$

In this case, necessarily $F < E$. Let $\eta \in (0, F - E)$ and C be the set of all points which are reachable from x at height $U(x) + F + \eta$. Then C is a cusp of depth $F + \eta$ whose bottom B is constituted of minima of depth exactly E, and whose other minima are all of depth strictly less than $F + \eta$ (since B is reachable from them without leaving C). Since there is only a finite number of such minima, in fact if η is small enough these non-global minima in C are even of depth less than F (possibly equal).

Since $F + \eta > F$, the process has a non-zero probability to stay trapped forever in C. On the other hand, since all non-global minima in C are of depth less than F,

$$\lim_{t \to \infty} \mathbb{P}(X_t \in C \setminus S) = 0.$$

as soon as S is a neighbourhood of B. Thus the same conclusion holds.

As in the fast cooling case, part 3 is a direct consequence of parts 1 and 2 and of the presence of at least one non-global minimum.

$\square$

4. Non-minimal rate

As have been seen in Section 1.2, the measure $\mu_\varepsilon = e^{-\frac{U(x)}{\varepsilon}} dx \otimes \frac{\delta_1 + \delta_{-1}}{2}$ is invariant for the Markov process with generator

$$Lf(x, y) \quad = \quad yf'(x, y) + \lambda(x, y)\left(f(x, -y) - f(x, y)\right) \tag{4.1}$$

if and only if $\lambda(x, y) = \frac{(yU'(x))_+}{\varepsilon} + r(x)$, where r can be any non-negative function. A positive r is a residual rate of jump which brings randomness in the system at any time. The previous works on explicit estimations of convergence to equilibrium for the velocity jump process (Fontbona et al., 2012; Calvez et al., 2015; Monmarché, 2014) all assume r is bounded below by a positive constant r_*.

If r is bounded from above by a constant r^* uniformly in x and ε, it is expected that he behaviour of the process does not change. Indeed, at low temperature, it only adds random jumps at exponential (macroscopic, in the sense: of order of magnitude independent from ε) times to the minimal-rate dynamics, and the latter accounts both for the way the process overcome an energy barrier, and the metastable behaviour in the vicinity of a local minimum, namely: many microscopic excursions of length of order $\sqrt{\varepsilon}$. In the rest of this section, we will make this statement more precise, and prove it.

We will consider the context of Section 2, that is a uni-dimensional double-well potential U with its three local extrema $x_0 < x_1 < x_2$ and an homogeneous Markov process starting at $(X_0, Y_0) = (x_0, 1)$; but now the generator is given by (4.1) with $r \neq 0$. We are interested in the following:

$$\begin{aligned} \eta &= \inf\left\{t > 0, \ X_t \in \{x_0, x_1\}\right\}, \\ p_{x_0} &= \mathbb{P}\left(X_\eta = x_1\right). \end{aligned} \tag{4.2}$$

Proposition 4.1. *Whatever the residual rate of jump $x \mapsto r(x)$,*

$$p_{x_0} \quad = \quad \frac{e^{-\frac{U(x_1) - U(x_0)}{\varepsilon}}}{1 + \int_{x_0}^{x_1} r(z) e^{-\frac{U(x_1) - U(z)}{\varepsilon}} dz}$$

Remarks :

- A positive residual rate of jump can only worsen the probability to overcome an energy barrier in one shot.
- A positive residual rate of jump in the neighbourhood of x_1 does more harm than the same rate near x_0.

Proof: If $x \in (x_0, x_1)$, let p_x be the probability that the process $(X_t, Y_t)_{t \geq 0}$, starting from $(x, 1)$, reaches $(x_1, 1)$ before $(x, -1)$. Suppose the process starts at $(x - s, 1)$ for some small $s > 0$. The probability the process goes from $x - s$ to x without any jump is $1 - s\left(\frac{U'(x)}{\varepsilon} + r(x)\right) + \underset{s \to 0}{o}(s)$. The probability it reaches $(x, 1)$ before $(x - s, -1)$ but with at least one jump (and so with at least two jumps) is of order s^2 (when $s \to 0$). Once the process has reached $(x, 1)$, it has a probability p_x to reach $(x_1, 1)$ before having fallen back to $(x, -1)$. If nevertheless it has fallen back to $(x, -1)$ - which occurs with probability $(1 - p_x)$ - it has a probability $sr(x) + \underset{s \to 0}{o}(s)$ to jump before reaching $(x - s, -1)$, in which case it reaches again $(x, 1)$ with probability $1 + \underset{s \to 0}{o}(1)$. In this latter case, it reaches $(x_1, 1)$ before $(x - s, -1)$ with

probability $p_x + \underset{s\to 0}{o}(1)$. Thus everything boils down to

$$
\begin{aligned}
p_{x-s} &= \left(1 - s\left(\frac{U'(x)}{\varepsilon} + r(x)\right)\right)(p_x + s(1 - p_x)p_x r(x)) + \underset{s\to 0}{o}(s) \\
&= p_x - sp_x\left(\frac{U'(x)}{\varepsilon} + p_x r(x)\right) + \underset{s\to 0}{o}(s).
\end{aligned}
$$

We shall solve the differential equation, for all $x \in (x_0, x_1)$,

$$
\partial_x p_x = p_x\left(\frac{U'(x)}{\varepsilon} + p_x r(x)\right),
$$

by looking for a solution of the form $p_x = h(x)e^{-\frac{U(x_1)-U(x)}{\varepsilon}}$. This yields

$$
\left(\frac{1}{h}\right)'(x) = -r(x)e^{-\frac{U(x_1)-U(x)}{\varepsilon}},
$$

and we conclude with $h(x_1) = p_{x_1} = 1$. $\square$

To study η, the length of an excursion away from x_0, we construct the process with generator (4.1) in the following way: let $(E_k)_{k\geq 1}$ and $(F_k)_{k\geq 1}$ be two independent sequences of independent r.v. with law $\mathcal{E}(1)$. Let V_1 and W_1 be defined by

$$
\varepsilon E_1 = \int_0^{V_1}(Y_0 U'(X_0 + Y_0 s))_+\, ds, \qquad F_1 = \int_0^{W_1} r(X_0 + Y_0 s)ds.
$$

Let $S_1 = \min(V_1, W_1)$. From $T_0 = 0$ to $T_1 = S_1$ the process goes on deterministically, in the sense that $Y_s = Y_0$ and $X_s = X_0 + sY_0$. At time T_1 the process jumps, so that $Y_{T_1} = -Y_0$. If $S_1 = V_1$, the process jumps due to the minimal rate; we will call this a first type jump. If $S_1 = W_1$ the jump is due to the residual rate and we will call this a second type jump. When the process has been defined up to a jump time T_j, we start the same procedure again, replacing E_1 by E_{j+1} and F_1 by F_{j+1}.

Since U is supposed smooth, $U^{(3)}(x)$ is bounded on $[x_0, x_1]$, interval on which U is strictly increasing. Thus it exists $\rho > 0$ such that for all $x \in [x_0, x_1)$, for all positive $s \leq x_1 - x$,

$$
U(x + s) \geq U(x) + \rho s^2.
$$

We define

$$
D_k = \min\left(|x_1 - x_0|, \sqrt{\frac{\varepsilon E_{2k-1}}{\rho}}\right),
$$

which is independent from the sequence $(F_k)_{k\geq 0}$. On the other hand, we consider the events

$$
A_k = \left\{F_k < \int_0^{x_1 - x_0} r(z)dz\right\}
$$

and the geometric variable

$$
N = \max\{k \geq 1, A_{2k}\},
$$

which is independent from the sequence $(E_k)_{k\geq 0}$.

Lemma 4.2. *Whatever the residual rate of jump $x \mapsto r(x)$,*

$$
\eta \leq 2\sum_{k=1}^{N} D_k,
$$

where η is defined by (4.2).

Proof: From time 0 to η, the process alternates between ascending and downward moves. If $X_\eta = x_0$, the process spends as much time going up as going down, and if $X_\eta = x_1$ then it spends more time going up than down; in either case, η is less than twice the cumulated ascending time. Let n be the number of ascending moves before the time η. At the end of the kth ascending move, the next time of jump will be defined thanks to the variable F_{2k} (as long as the process does not reach $(x_0, -1)$). In the event $\bar{A}_{2k}$ (namely the negation of A_{2k}) there is no second type jump before the process reaches $(x_0, -1)$. This means $n \leq N$.

Let $(d_k)_{1 \leq k \leq n}$ be the duration of the n ascending moves, and $(Z_k)_{1 \leq k \leq n}$ be the starting point of these moves (for instance $Z_1 = x_0$).

$$
\begin{aligned}
d_k &\leq V_{2k-1} \\
&= \sup\{t < (x_1 - x), \ \varepsilon E_{2k-1} > U(Z_k + t) - U(Z_k)\} \\
&\leq \sup\{t < (x_1 - x_0), \ \varepsilon E_{2k-1} > \rho t^2\} \\
&= D_k.
\end{aligned}
$$

Finally,

$$
\eta \leq 2 \sum_{k=1}^n d_k \leq 2 \sum_{k=1}^N D_k.
$$

$\square$

Now if r is supposed bounded above by a constant r^*,

$$
\mathbb{E}[N] = 1 + \frac{1}{\mathbb{P}(\bar{A}_2)} \leq 1 + e^{(x_1 - x_0)r^*},
$$

and thus, N and the D_k's being independent, $\mathbb{E}[\eta] \leq c\sqrt{\varepsilon}$ for some constant c. On the other hand,

$$
\mathbb{E}[\eta] \geq \mathbb{P}\left(\bar{A}_1, \bar{A}_2\right) \mathbb{E}\left[\eta \mid \bar{A}_1, \bar{A}_2\right] \geq e^{-2(x_1 - x_0)r^*} \mathbb{E}[\min(V_1, x_1 - x_0)].
$$

From Lemma 2.1 we get that $\mathbb{E}[\eta] \geq c'\sqrt{\varepsilon}$ for some $c' > 0$. But if $\mathbb{E}[\eta]$ is of order $\sqrt{\varepsilon}$, it means when ε goes to 0 it is unlikely that a second type jump occurs during an excursion, so that only the asymptotic of $\mathbb{E}[V_1]$ should intervene. Indeed, we can prove the following:

Proposition 4.3. *If $r \leq r^*$,*

$$
\mathbb{E}[\eta] = \sqrt{\frac{2\pi\varepsilon}{U''(x_0)}} \left(1 + \underset{\varepsilon \to 0}{o}(1)\right)
$$

Proof: When $r \leq r^*$, $F_k \leq r^* W_k$. Since $V_k \leq D_k$, the event $B = \{r^* D_1 \leq \min(F_1, F_2, \delta)\}$ (for some $\delta > 0$) implies that $V_1 \leq \min(W_1, W_2)$: there is no second type jump during the excursion. We decompose

$$
\mathbb{E}[\eta] = \mathbb{E}[\eta \mathbb{1}_B] + \mathbb{E}[\eta \mathbb{1}_{\bar{B}}].
$$

First, note that $\mathbb{P}(B)$ goes to 1 when ε goes to 0. Second,

$$
\begin{aligned}
\mathbb{E}[\mathbb{1}_{\bar{B}}\eta] &\leq 2\mathbb{E}\left[\mathbb{1}_{\bar{B}} \sum_{k=1}^N D_k\right] \\
&= 2\mathbb{E}[\mathbb{1}_{\bar{B}} D_1] + 2\mathbb{P}\left(\bar{B}\right) \mathbb{E}[N-1] \mathbb{E}[D_2] \\
&\leq 2\sqrt{\mathbb{P}\left(\bar{B}\right) \mathbb{E}[D_1^2]} + 2\mathbb{P}\left(\bar{B}\right) \mathbb{E}[N-1] \mathbb{E}[D_2]
\end{aligned}
$$

$$= \underset{\varepsilon \to 0}{o} \left(\sqrt{\varepsilon} \right).$$

Suppose that B holds and $\delta < r^*|x_1 - x_0|$. Then the process goes up until a first time jump occurs, then it falls back to x_0, and no second type jump occurs: $\eta = 2V_1$, and

$$
\begin{aligned}
\varepsilon E_1 &= U(x_0 + V_1) - U(x_0) \\
&= \frac{1}{2} U''(x_0) \left(\frac{V_1}{\sqrt{\varepsilon}} \right)^2 \varepsilon + \underset{\varepsilon \to 0}{o}(\varepsilon) \\
\Rightarrow \quad \frac{V_1}{\sqrt{\varepsilon}} &\underset{\varepsilon \to 0}{\longrightarrow} \sqrt{\frac{2E_1}{U''(x_0)}} \quad a.s.
\end{aligned}
$$

Moreover $\mathbb{1}_{\bar{B}} \to \mathbb{1}$ almost surely, and

$$\frac{V_1}{\sqrt{\varepsilon}} \mathbb{1}_{\bar{B}} \leq \frac{D_1}{\sqrt{\varepsilon}} \leq \sqrt{\frac{E_1}{\rho}}.$$

By the dominated convergence theorem,

$$2\mathbb{E}\left[V_1 \mathbb{1}_{\bar{B}} \right] = 2\sqrt{\varepsilon}\mathbb{E}\left[\sqrt{\frac{2E_1}{U''(x_0)}} \right] \left(1 + \underset{\varepsilon \to 0}{o}(1) \right),$$

which concludes this proof. $\qquad \square$

As a conclusion, note than when $r \leq r^*$,

$$e^{-\frac{U(x_1) - U(x_0)}{\varepsilon}} \geq p_{x_0} \geq \frac{e^{-\frac{U(x_1) - U(x_0)}{\varepsilon}}}{1 + (x_1 - x_0)r^*}.$$

From these bounds, following the proof of Theorem 1.3, we can see the same result holds even if the non-minimal rate is not zero (as long as it is bounded above uniformly in time). On the other hand if r goes to infinity as ε goes to 0, the velocities at two different instants are more and more decorrelated, and we may think the (suitably rescaled) process ends up with a genuine diffusive part. We recall this has been proved at least for the case of a constant potential on the torus, in Miclo and Monmarché (2013).

5. The multidimensional process

In this section we consider an inhomogeneous process Z^ε on $\mathbb{T}^d \times \mathbb{S}^{d-1}$ with generator $(L_t)_{t \geq 0}$ given by (1.6), with a positive cooling schedule $(\varepsilon_t)_{t \geq 0}$, and its associated semi-group $(P_{s,t})_{t \geq s \geq 0}$. To lighten the notations, in this section we drop the ε exponent and only write $Z = (X, Y)$. In order to prove Theorems 1.4 and 1.5 we will establish the following intermediary result:

Theorem 5.1. *There exist $c, \theta > 0$ that depend only on d, r, U such that for all cooling schedule, all initial law μ, and all $t \geq 0$,*

$$\|\mu P_{t,t+\sqrt{d}} - \nu_t\|_{TV} \leq \left(1 - ce^{-\frac{\theta}{\varepsilon_t}} \right) \|\nu - \nu_t\|_{TV} + \theta \left(\frac{1}{\varepsilon_{t+\sqrt{d}}} - \frac{1}{\varepsilon_t} \right).$$

In the first instance we consider the case $U = 0$:

Lemma 5.2. *Let* $\left(P_t^0\right)_{t\geq 0}$ *be the semi-group on* $\mathcal{P}\left(\mathbb{T}^d \times \mathbb{S}^{d-1}\right)$ *generated by*

$$Lf(x,y) \;=\; y.\nabla_x f(x,y) + r\left(\int f(x,z)dz - f(x,y)\right). \qquad (5.1)$$

Then, there exists $c > 0$ *which depends only on* d *and* r *such that for all* $\mu_1, \mu_2 \in \mathcal{P}\left(\mathbb{T}^d \times \mathbb{S}^{d-1}\right),$

$$\left\|\mu_1 P_{\sqrt{d}}^0 - \mu_2 P_{\sqrt{d}}^0\right\|_{TV} \;\leq\; (1-c)\,\|\mu_1 - \mu_2\|_{TV}$$

Proof: Let (Z_0, Z_0') be an optimal coupling of μ_1 and μ_2, in the sense that the law of Z_0 (resp. Z_0') is μ_1 (resp. μ_2) and that

$$\|\mu_1 - \mu_2\|_{TV} = \mathbb{P}\left(Z_0 \neq Z_0'\right).$$

We aim to construct a coupling $(Z_t, Z_t')_{t\in[0,\sqrt{d}]}$ for which each marginal is a Markov process with generator (5.1) and such that $Z_{\sqrt{d}}$ and $Z_{\sqrt{d}}'$ are likely to be equal. If $Z_0 = Z_0'$ we define Z up to time $\sqrt{d}$ and then let $Z_t' = Z_t$ for all $t \in (0, \sqrt{d}]$.

Now if $Z_0 \neq Z_0'$ we will show that the (density part of the) laws of $Z_{\sqrt{d}}$ and $Z_{\sqrt{d}}'$ are bounded below by a constant time the uniform density on $\mathbb{T}^d \times \mathbb{S}^{d-1}$, which yields a lower bound on their total variation distance. Indeed, let $(E_i)_{i\geq 0}$ be a sequence of i.i.d. exponential variables and defined $S_0 = 0$ and $S_{i+1} = S_i + \frac{1}{r}E_i$ to be the jump times of Z. Under the event $\{E_3 > r\sqrt{d}\}$ there are at most two jumps before time $\sqrt{d}$. More precisely consider the event

$$B = \left\{E_3 > r\sqrt{d},\; \frac{r\sqrt{d}}{2} < E_1 + E_2 < r\sqrt{d}\right\}$$

so that if f is any positive function, starting from $Z_0 = (x_0, y_0)$,

$$\mathbb{E}\left(f(Z_{\sqrt{d}})\right) \;\geq\; \mathbb{E}\left(f(Z_{\sqrt{d}})\mathbb{1}_B\right)$$

$$\geq\; \mathbb{P}\left(E_3 > r\sqrt{d}\right)\int_{t=\frac{1}{2}\sqrt{d}}^{\sqrt{d}}\int_{s=0}^{t}\int f\left(\tilde{x}, y_2\right)r^2 e^{-rt}dy_1 dy_2 ds dt,$$

where dy_1 and dy_2 stand for the uniform law on $\mathbb{S}^{d-1}$ and

$$\tilde{x} = x_0 + sy_0 + (t-s)y_1 + (\sqrt{d} - t)y_2.$$

Suppose $t > \frac{1}{2}\sqrt{d}$ and $y_2 \in \mathbb{S}^{d-1}$ (and thus $x_0' := x_0 + (\sqrt{d} - t)y_2$) are fixed. We want to prove

$$\int_0^t\int_{\mathbb{S}^{d-1}} f\left(\tilde{x}, y_2\right)e^{-rs}ds dy_1 \;\geq\; c_1\int_{\mathbb{T}^d} f\left(u, y_2\right)du \qquad (5.2)$$

for some $c_1 > 0$. Let $u \in \mathbb{T}^d$ and $\delta > 0$, S be a r.v. uniform on $[0,1]$ and Y be uniformly distributed on $\mathbb{S}^{d-1}$ and independent from S, so that (5.2) reads

$$\mathbb{P}\left(|x_0' + tSy_0 + t(1-S)Y - u| \leq \delta\right) \geq c_2\delta^d$$

with a constant c_2 that depends neither on x_0', t, u, δ nor y_0. Since the diameter of $\mathbb{T}^d$ is $\frac{1}{2}\sqrt{d} < t$, there exists $s_* \in (0,1)$ and $y_* \in \mathbb{S}^{d-1}$ such that $u = x_0' + ts_*y_0 + t(1-s_*)y_*$.

$$\mathbb{P}\left(|x_0' + tSy_0 + t(1-S)Y - u| \leq \delta\right) \;\geq\; \mathbb{P}\left(|S - s_*| + |Y - y_*| \leq \frac{\delta}{2\sqrt{d}}\right)$$

$$\geq \quad \mathbb{P}\left(|S - s_*| \wedge |Y - y_*| \leq \frac{\delta}{4\sqrt{d}}\right)$$

$$\geq \quad c_2 \delta^d$$

for some $c_2 > 0$ since S and Y are uniformly drawn on compact spaces with respective dimensions 1 and $d - 1$. Bringing the pieces together we have obtained

$$\mathbb{E}\left(f(Z_{\sqrt{d}})\right) \quad \geq \quad r^2 e^{-2r\sqrt{d}} c_1 \int_{\mathbb{T}^d \times \mathbb{S}^{d-1}} f(u, y_2)\, du dy_2$$

for any $f \geq 0$. In other words for any Borel set $\mathcal{A} \subset \mathbb{T}^d \times \mathbb{S}^{d-1}$,

$$\mathbb{P}\left(Z_{\sqrt{d}} \in \mathcal{A}\right) \quad \geq \quad c\mathbb{P}\left(H \in \mathcal{A}\right)$$

where $c = r^2 e^{-2r\sqrt{d}} c_1$ and H is a r.v. with uniform law $dx \times dy$ on $\mathbb{T}^d \times \mathbb{S}^{d-1}$. This means there exists a probability measure m (that depends on Z_0) such that the law of $Z_{\sqrt{d}}$ is $c(dx \times dy) + (1-c)m$. The same goes for Z' with a law m'. With probability c, set $Z_{\sqrt{d}} = H = Z'_{\sqrt{d}}$, and with probability $(1-c)$, draw $Z_{\sqrt{d}}$ (resp. Z') according to the law m (resp. m'). with this coupling,

$$\mathbb{P}\left(Z_{\sqrt{d}} \neq Z'_{\sqrt{d}}\right) \quad \leq \quad (1-c)\mathbb{P}\left(Z_0 \neq Z'_0\right).$$

$\square$

Proof of Theorem 5.1: Let (Z_0, Z'_0) be an optimal coupling of μ and ν_t, in the sense that the law of Z_0 (resp. Z'_0) is μ (resp. ν_t) and that

$$\|\mu - \nu_t\|_{TV} = \mathbb{P}\left(Z_0 \neq Z'_0\right).$$

If $Z_0 = Z'_0$ let $Z' = (X', Y')$ evolve as a Markov process with generator L_t (here t is fixed). Since ν_t is invariant for L_t, $Z'_{\sqrt{d}} \sim \nu_t$. Let E be an exponential r.v. independent from $(Z'_s)_{s \in [0, \sqrt{d}]}$ and

$$S \quad = \quad \inf\left\{s \geq t,\ E < \int_t^s (Y'_u \nabla_x U(X'_u))_+ \left(\frac{1}{\varepsilon_u} - \frac{1}{\varepsilon_t}\right) du\right\}.$$

Define $Z_s = Z'_s$ up to time $T = S \wedge \sqrt{d}$. If $T = S$, the process Z is reflected at time T (in the sense $Y \leftarrow Y^*$) and then we let Z evolve as an inhomogeneous Markov process with generator $(L_s)_{s \geq 0}$ independently from Z'. Note that $T = \sqrt{d}$ implies $Z_{\sqrt{d}} = Z'_{\sqrt{d}}$, and that

$$\mathbb{P}\left(T < S\right) \quad \geq \quad \mathbb{P}\left(E > \sqrt{d}\|\nabla_x U\|_\infty \left(\frac{1}{\varepsilon_{t+\sqrt{d}}} - \frac{1}{\varepsilon_t}\right)\right). \tag{5.3}$$

This means

$$\|\mu P_{t, t+\sqrt{d}} - \nu_t\|_{TV} \quad \leq \quad \mathcal{P}\left(Z_{\sqrt{d}} \neq Z'_{\sqrt{d}} \mid Z_0 \neq Z'_0, T < S\right) \|\mu - \nu_t\|_{TV}$$

$$+ \left(1 - e^{-\sqrt{d}\|\nabla_x U\|_\infty \left(\frac{1}{\varepsilon_{t+\sqrt{d}}} - \frac{1}{\varepsilon_t}\right)}\right)$$

$$\leq \quad \mathcal{P}\left(Z_{\sqrt{d}} \neq Z'_{\sqrt{d}} \mid Z_0 \neq Z'_0, T < S\right) \|\mu - \nu_t\|_{TV}$$

$$+ \sqrt{d}\|\nabla_x U\|_\infty \left(\frac{1}{\varepsilon_{t+\sqrt{d}}} - \frac{1}{\varepsilon_t}\right). \tag{5.4}$$

In the following, suppose $T < S$ and $Z_0 \neq Z_0'$. Let $W = (R, Q)$ and $W' = (R', Q')$ be Markov processes associated to the generator (5.1) with respectively $W_0 = Z_0$ and $W_0' = Z_0'$ and such that

$$\mathbb{P}\left(W_{\sqrt{d}} \neq W'_{\sqrt{d}}\right) = \|\mu P^0_{\sqrt{d}} - \nu_t P^0_{\sqrt{d}}\|_{TV}.$$

Let F be an exponential r.v. which is independent from all the other variables and

$$
\begin{aligned}
S_1 &= \inf\left\{s \geq t_0, \ \varepsilon_t F < \int_t^s (Q_u \nabla_x U(R_u))_+ \, du\right\} \\
S_2 &= \inf\left\{s \geq t_0, \ \varepsilon_t F < \int_t^s (Q'_u \nabla_x U(R'_u))_+ \, du\right\}
\end{aligned}
$$

Define $Z = W$ up to time $T_1 = S_1 \wedge \sqrt{d}$ and $Z' = W'$ up to time $S_2 \wedge \sqrt{d}$. If $T_1 = S_1$ the process Z is reflected at time T_1 and then let it evolve independently from W, and similarly for Z'. Note that the event $\{\varepsilon_t F > \kappa\}$ is independent from $\{W_{\sqrt{d}} = W'_{\sqrt{d}}\}$, so that

$$
\begin{aligned}
&\mathbb{P}\left(Z_{\sqrt{d}} = Z'_{\sqrt{d}} \mid Z_0 \geq Z_0', T < S\right) \\
\geq \ & \mathbb{P}\left(\varepsilon_t F > \sqrt{d}\|\nabla_x U\|_\infty\right) \mathbb{P}\left(W_{\sqrt{d}} = W'_{\sqrt{d}}\right) \qquad (5.5) \\
\geq \ & c e^{-\frac{\sqrt{d}\|\nabla_x U\|_\infty}{\varepsilon_t}}
\end{aligned}
$$

where c is given by Lemma 5.2.   □

Remarks:

- Concerning the constants involved in Theorem 5.1, with the previous proofs we obtained that it is possible to choose $c \leq r^2 e^{-2\sqrt{d}} c_1$, where c_1 only depends on d, and $\theta \leq \sqrt{d}\|\nabla_x U\|_\infty$. Nevertheless in the proof of Lemma 5.2, notice that the bound c on the probability of success for the coupling is obtained by considering only the situation where both processes jump twice and only twice. Hence, the bound (5.3) may be improved to

$$\mathbb{P}\left(T < S\right) \ \geq \ \mathbb{P}\left(E > \kappa(\sqrt{d})\left(\frac{1}{\varepsilon_{t+\sqrt{d}}} - \frac{1}{\varepsilon_t}\right)\right),$$

  where we define $\kappa(t)$ as the supremum of $\int_0^t (q(u)\nabla_x U(r(u)))_+ \, du$ over all the (r, q) which are the trajectory of a velocity jump process that jumps only twice during time t. The same goes at line (5.5) and thus Theorem 5.1 holds with $\theta \leq \kappa_2(\sqrt{d})$.
- when we proved (5.4) we hadn't use the fact $\frac{1}{2}\sqrt{d}$ is the diameter of $\mathbb{T}^d$ yet, which means that in fact for all $s \leq \sqrt{d}$

$$\|\mu P_{t,t+s} - \nu_t\|_{TV} \ \leq \ \|\mu - \nu_t\|_{TV} + \sqrt{d}\|\nabla U\|_\infty \left(\frac{1}{\varepsilon_{t+s}} - \frac{1}{\varepsilon_t}\right).$$

Proof of Theorem 1.4: As a direct consequence of the previous remark and of Theorem 5.1 we obtain, if $\varepsilon_t = \varepsilon$ is constant (so that $P_{0,t} = P_t$ is homogeneous),

$$
\begin{aligned}
\|\mu P_t - \nu_\infty\|_{TV} \ &\leq \ e^{-ce^{-\frac{\theta}{\varepsilon}}\left\lfloor\frac{t}{\sqrt{d}}\right\rfloor} \|\mu - \nu_\infty\|_{TV} \\
&\leq \ e^{-\frac{c}{\sqrt{d}}e^{-\frac{\theta}{\varepsilon}}\left(t-\sqrt{d}\right)} \|\mu - \nu_\infty\|_{TV}.
\end{aligned}
$$

$\square$

Proof of Theorem 1.5: Let t_0 be such that $\partial_s \left(\frac{1}{\varepsilon_s} \right) \leq \frac{1}{(\theta + \eta)s}$ holds for $s \geq t_0$, let $t > 2t_0 + \sqrt{d}$ and $n = \left\lfloor \frac{t - t_0}{\sqrt{d}} \right\rfloor$. For $k \in [\![0, n]\!]$ let $t_{n-k} = t - k\sqrt{d}$,

$$
\begin{aligned}
d_k &= \|\mu P_{0, t_k} - \nu_{t_k}\|_{TV}, \\
a_k &= c e^{-\frac{\theta}{\varepsilon_{t_k}}}, \\
b_k &= \theta \left(\frac{1}{\varepsilon_{t_{k+1}}} - \frac{1}{\varepsilon_{t_k}} \right),
\end{aligned}
$$

so that Theorem 5.1 reads $d_{k+1} - d_k \leq -a_k d_k + b_k$. This situation is a discrete analogous of Miclo (1992, Lemme 6). From $\left(\frac{1}{\varepsilon_t} \right)' \leq \frac{1}{(\theta + \eta)t}$,

$$
b_k \leq \frac{\theta \sqrt{d}}{(\theta + \eta)t_k}, \qquad a_k \geq m \left(\frac{1}{t_k} \right)^{\frac{\theta}{\theta + \eta}}, \qquad c_k := \frac{b_k}{ak} \leq m' \left(\frac{1}{t_k} \right)^{\frac{\eta}{\theta + \eta}}
$$

for some $m, m' > 0$. For a fixed $l \in \mathbb{N}$ and for all $k \geq l$, $f_k := d_k - m' t_l^{-\frac{\eta}{\theta + \eta}}$ satisfies

$$
\begin{aligned}
f_{k+1} - f_k &\leq -a_k f_f \\
\Rightarrow \quad f_k &\leq f_l \prod_{j=l}^{k-1} (1 - a_j) \\
&\leq \exp \left(-m \sum_{j=l}^{k-1} \left(\frac{1}{t_j} \right)^{\frac{\theta}{\theta + \eta}} \right) \\
&\leq \exp \left(-\frac{m}{\sqrt{d}} \int_{t_l}^{t_k} \left(\frac{1}{s} \right)^{\frac{\theta}{\theta + \eta}} ds \right).
\end{aligned}
$$

Since we choose $t > 2t_0 + \sqrt{d}$ there exists $l \in \mathbb{N}$ such that $t_l \geq \frac{t}{2} \geq t_l - \sqrt{d}$, hence

$$
\begin{aligned}
\|\mu P_{0, t} - \nu_t\|_{TV} &= d_n \\
&\leq m' t_l^{-\frac{\eta}{\theta + \eta}} + \exp \left(-m'' \left(t^{\frac{\eta}{\theta + \eta}} - t_l^{\frac{\eta}{\theta + \eta}} \right) \right) \\
&\leq M t^{-\frac{\eta}{\theta + \eta}}
\end{aligned}
$$

for some $m'', M > 0$.

Let (X, Y) be a velocity jump process with generator $(L_s)_{s \geq 0}$ and initial law μ and (V, W) be a r.v. with law ν_t. By Laplace method for any $\alpha > 1$ there exists a constant C such that for all $h > 0$,

$$
\mathbb{P} \left(U(V) > \min U + h \right) \leq C e^{-\frac{h}{\alpha \varepsilon_t}}.
$$

The conclusion follows from

$$
\begin{aligned}
\mathbb{P} \left(U(X_t) > \min U + h \right) &\leq \mathbb{P} \left(U(V) > \min U + h \right) + \|\mu P_{0, t} - \nu_t\|_{TV} \\
&\leq C e^{-\frac{h}{\alpha \varepsilon_t}} + M t^{-\frac{\eta}{\theta + \eta}}.
\end{aligned}
$$

$\square$

Remark: Here is an heuristic to adapt the previous arguments to prove Theorem 1.5 holds with $\theta = E_*$: given any smooth path $\gamma : [0,1] \to \mathbb{T}^d$ with $|\gamma'| = 1$ and any $\delta > 0$, a process W with generator (5.1) (namely for which $U = 0$) starting from $\gamma(0)$ has a positive probability (depending on r, d and γ) to reach $\gamma(1)$ in a given time (say $t = 2$) staying in a tube of diameter δ around the path γ. Here, a tube is defined by

$$\mathcal{T}_{\gamma,\delta} = \{(x,y) \in \mathbb{T}^d \times \mathbb{S}^{d-1} \text{ s.t. } \exists s \in [0,1], \ |x - \gamma(s)| + |y - \gamma'(s)| \leq \delta\}.$$

Then we can define a process Z with generator (1.6) to be equal to W up to a time at which a clock depending on the cooling schedule (similar to S_1 and S_2 in the proof of Theorem 5.1) rings, at which point Z is reflected. If γ is an optimal path to leave a cusp C, in the sense a path with $\max \gamma - \min \gamma$ equal to D the depth of C, the process is not reflected during this climb with probability of order $\exp(-(D + f(\delta))/\varepsilon)$ where $f(\delta)$ goes to 0 with δ.

This leads to an Eyring-Kramers formula (with a non explicit prefactor since the probability for W to approximatively follow γ is not easily tractable) and more generally to an adaptation of all the arguments from Section 3. Note however that such a study in the regime $\varepsilon \to 0$ only quantifies metastability due to energy barriers, so that E_* naturally appears, but not entropic barriers (see Lelièvre (2013) for the distinction). The latter are indeed related to the very probability for W to follow a particular path, and a really relevant question to compare the velocity Gibbs sampler to the Fokker-Planck diffusion is: is it easier (i.e. faster) to find a particular path, in some applied problem, with broken lines rather than with a Brownian motion? When there are both energy and entropic barriers, is there an optimal amount of inertia to overcome both?

Acknowledgements

References

C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris (2000). ISBN 2-85629-105-8. With a preface by Dominique Bakry and Michel Ledoux. MR1845806.

R. Azaïs, J.-B. Bardet, A. Genadot, N. Krell and P.-A. Zitt. Piecewise deterministic markov process - recent results. *ArXiv Mathematics e-prints* (2013). arXiv: 1309.6061.

C. J. P. Bélisle, H. E. Romeijn and R. L. Smith. Hit-and-run algorithms for generating multivariate distributions. *Math. Oper. Res.* **18** (2), 255–266 (1993). MR1250117.

M. Benaïm, S. Le Borgne, F. Malrieu and P.-A. Zitt. Quantitative ergodicity for some switched dynamical systems. *Electron. Commun. Probab.* **17**, no. 56, 14 (2012). MR3005729.

A. Bovier, M. Eckhoff, V. Gayrard and M. Klein. Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc. (JEMS)* **6** (4), 399–424 (2004). MR2094397.

A. Bovier, V. Gayrard and M. Klein. Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues. *J. Eur. Math. Soc. (JEMS)* **7** (1), 69–99 (2005). MR2120991.

V. Calvez, G. Raoul and C. Schmeiser. Confinement by biased velocity jumps: aggregation of *escherichia coli*. *Kinet. Relat. Models* **8** (4), 651–666 (2015). MR3375486.

O. Catoni. Rough large deviation estimates for simulated annealing: application to exponential schedules. *Ann. Probab.* **20** (3), 1109–1146 (1992). MR1175253.

P. Diaconis and L. Miclo. On the spectral analysis of second-order Markov chains. *Ann. Fac. Sci. Toulouse Math. (6)* **22** (3), 573–621 (2013). MR3113027.

R. Erban and H. G. Othmer. From individual to collective behavior in bacterial chemotaxis. *SIAM J. Appl. Math.* **65** (2), 361–391 (electronic) (2004/05). MR2123062.

J. Fontbona, H. Guérin and F. Malrieu. Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process. *Adv. in Appl. Probab.* **44** (4), 977–994 (2012). MR3052846.

M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third edition (2012). ISBN 978-3-642-25846-6. MR2953753.

S. Gadat and F. Panloup. Long time behaviour and stationary regime of memory gradient diffusions. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** (2), 564–601 (2014). MR3189085.

B. Hajek. Cooling schedules for optimal annealing. *Math. Oper. Res.* **13** (2), 311–329 (1988). MR942621.

R. A. Holley, S. Kusuoka and D. W. Stroock. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.* **83** (2), 333–347 (1989). MR995752.

J. H. Kalivas et al. *Adaption of simulated annealing to chemical optimization problems*, volume 15 of *Data Handling in Science and Technology*. Elsevier (1995). DOI: 10.1016/S0922-3487(06)80001-8.

T. Lelièvre. Two mathematical tools to analyze metastable stochastic processes. In *Numerical mathematics and advanced applications 2011*, pages 791–810. Springer, Heidelberg (2013). MR3307611.

T. Lelièvre, F. Nier and G. A. Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *J. Stat. Phys.* **152** (2), 237–274 (2013). MR3082649.

L. Miclo. Recuit simulé sur $\mathbf{R}^n$. Étude de l'évolution de l'énergie libre. *Ann. Inst. H. Poincaré Probab. Statist.* **28** (2), 235–266 (1992). MR1162574.

L. Miclo and P. Monmarché. Étude spectrale minutieuse de processus moins indécis que les autres. In *Séminaire de Probabilités XLV*, volume 2078 of *Lecture Notes in Math.*, pages 459–481. Springer, Cham (2013). MR3185926.

P. Monmarché. Hypocoercive relaxation to equilibrium for some kinetic models. *Kinet. Relat. Models* **7** (2), 341–360 (2014). MR3195078.

P. Monmarché. Hypocoercivity in metastable settings and kinetic simulated annealing. *ArXiv Mathematics e-prints* (2015). arXiv: 1502.07263.

N. Ratanov. Telegraph processes with random jumps and complete market models. *Methodol. Comput. Appl. Probab.* **17** (3), 677–695 (2015). MR3377855.

A. Scemama, T. Lelièvre, G. Stoltz, E. Cancès and M. Caffarel. An efficient sampling algorithm for variational monte carlo. *The Journal of Chemical Physics* **125** (11), 114105–10 (2006). DOI: 10.1063/1.2354490.