Cooling down Lévy flights

2007 J. Phys. A: Math. Theor. 40 12299

(http://iopscience.iop.org/1751-8121/40/41/003)

IOP Publishing Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **40** (2007) 12299–12313 doi:10.1088/1751-8113/40/41/003

Cooling down Lévy flights

Ilya Pavlyukevich

Institut für Mathematik, Humboldt-Universität zu Berlin, Rudower Chaussee 25, 12489 Berlin, Germany

E-mail: pavljuke@mathematik.hu-berlin.de

Received 7 November 2006, in final form 10 September 2007
Published 25 September 2007
Online at stacks.iop.org/JPhysA/40/12299

Abstract
Let $L(t)$ be a Lévy flights process with a stability index $\alpha \in (0, 2)$, and U be an external multi-well potential. A jump diffusion Z satisfying a stochastic differential equation $\mathrm{d}Z(t) = -U'(Z(t-))\,\mathrm{d}t + \sigma(t)\,\mathrm{d}L(t)$ describes an evolution of a Lévy particle of an 'instant scale' $\sigma(t)$ in an external force field. The scale is supposed to decrease polynomially fast, i.e. $\sigma(t) \approx t^{-\theta}$ for some $\theta > 0$. We discover two different decrease regimes. If $\theta < 1/\alpha$ (slow cooling), the jump diffusion $Z(t)$ has a non-trivial limiting distribution as $t \to \infty$, which is concentrated at the potential's local minima. If $\theta > 1/\alpha$ (fast cooling), the Lévy particle gets trapped in one of the potential wells.

PACS numbers: 05.40.Fb, 02.50.Ey, 02.50.Fz, 02.50.Ga

1. Introduction

In this paper, we study a Lévy flights dynamics in an external multi-well potential in the annealed regime. We are motivated by the problem of random search of the global minimum of an unknown function U with the help of simulated annealing. For simplicity, we consider a one-dimensional case. Let U be a multi-well potential satisfying some regularity conditions. Classical continuous time simulated annealing consists in an examination of a time non-homogeneous Smoluchowski diffusion process $\hat{Z} = (\hat{Z}(t))_{t\geqslant 0}$, satisfying the stochastic differential equation

$$\mathrm{d}\hat{Z}(t) = -U'(\hat{Z}(t))\,\mathrm{d}t + \hat{\sigma}(t)\,\mathrm{d}W(t) \tag{1}$$

with some positive decreasing temperature function $\hat{\sigma}(t)$, $\hat{\sigma}(t) \to 0$ as $t \to +\infty$. For small values of $\hat{\sigma}(t)$, the process $\hat{Z}$ spends most of the time in small neighbourhoods of the potential's local minima and makes occasional transitions between the adjacent wells. It is possible to choose an appropriate cooling schedule $\hat{\sigma}(t)$, such that the diffusion settles down near the global maximum of U. Indeed, one should take $\hat{\sigma}^2(t) \approx \frac{\theta}{\ln(\lambda+t)}$, the parameter $\theta > 0$ being a *cooling rate* and $\lambda > 1$ parameterizing the initial temperature. Then there is a critical value $\hat{\theta} > 0$, such that the marginals $\hat{Z}(t)$ converge in probability to the global minimum of U if $\theta > \hat{\theta}$, and the convergence fails if $0 < \theta < \hat{\theta}$. Moreover, the critical value $\hat{\theta}$ is a logarithmic growth rate of the principal non-zero eigenvalue $\lambda^1(\sigma)$ of the infinitesimal generator $A_\sigma f = \frac{\sigma^2}{2}\Delta f - U'f'$ of the time-homogeneous diffusion $\hat{X} = (\hat{X}(t))_{t\geqslant 0}$ satisfying the stochastic differential equation

$$\mathrm{d}\hat{X}(t) = -U'(\hat{X}(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t), \tag{2}$$

i.e. $\hat{\theta} = -\lim_{\sigma\to 0}\sigma^2 \ln|\lambda^1(\sigma)|$. Heuristic justification for the convergence is as follows. The principal non-zero eigenvalue $\lambda^1(\sigma)$ determines the convergence rate of $\hat{X}$ to its invariant measure $\mu_\sigma(\mathrm{d}x) = c_\sigma \exp(-2U(x)/\sigma^2)\,\mathrm{d}x$, c_σ being a normalizing factor. Thus, for any continuous positive function f we have an estimate

$$|\mathbf{E}_x f(\hat{X}(t)) - \int f(x)\mu_\sigma(\mathrm{d}x)| \leqslant C\,\mathrm{e}^{-|\lambda^1(\sigma)|t}. \tag{3}$$

The weak limit of the invariant measures $\mu_\sigma(\mathrm{d}y)$ as $\sigma \to 0$ is a Dirac mass at the potential's global minimum. For small values of $\hat{\sigma}(t)$, the dynamics of $\hat{Z}$ reminds of a dynamics of $\hat{X}$. Thus, $\hat{Z}(t)$ has enough time to settle down in the deepest potential well if $\hat{\sigma}(t)$ is such that

$$t|\lambda^1(\hat{\sigma}(t))| \to \infty \quad\Leftrightarrow\quad \frac{t}{(\lambda+t)^{\hat{\theta}/\theta}} \to \infty \quad\Leftrightarrow\quad \theta > \hat{\theta}, \qquad t \to +\infty. \tag{4}$$

It was Kirkpatrick *et al* [1] and Černy [2] who generalized the seminal paper [3] by Metropolis *et al*, bridged the statistical mechanics with combinatorial optimization problems and gave rise to the extensive physical and mathematical study of simulated annealing. We mention here the work by Geman and Geman [4], who firstly obtained the logarithmic decrease rate of $\hat{\sigma}(t)$. Vanderbilt and Louie [5] applied simulated annealing to optimization problems over continuous variables, and Aluffi-Pentini *et al* [6], Geman and Hwang [7] and Gidas [8] considered simulated annealing of Gaussian diffusions, i.e. studied the process $\hat{Z}$ from (1).

Further mathematical results on Gaussian continuous time simulated annealing can be found in works by Chiang *et al* [9], Hajek [10], Holley and Stroock [11], Holley *et al* [12], and Hwang and Sheu [13, 14]. We also refer the reader to the review paper [15] by Gidas, to Chapter 11 of [16] by Hartmann and Rieger and further references therein.

Our present research is motivated by the paper [17] by Szu and Hartley, where they introduced the so-called *fast simulated annealing* which allows us to perform a non-local search of the deepest well. The fast simulated annealing process in the sense of [17] is a discrete time Markov chain, where the states are obtained from the Euler approximation of (1) driven not by Gaussian noise but by *Cauchy* noise. The new state is accepted according to the Metropolis algorithm with acceptance probability which equals 1, if the potential value in this state is smaller, i.e. the new position is 'lower' in the potential landscape. If the new position is 'higher', it is accepted with the probability $\sim \exp(-\Delta U/\sigma)$, where ΔU is the difference of the potential values in the new and the old states and σ is a decreasing noise amplitude. The advantage of this method consists in faster transitions between the potential wells due to the heavy tails of Cauchy distribution. Moreover, the authors claim that the optimal cooling rate is algebraic, $\sigma(t) \approx t^{-1}$, which also accelerates convergence.

In the present paper, we consider a continuous-time Lévy flights counterpart of the diffusion (1). Our goal is to study the asymptotic properties of the system in dependence of a cooling schedule. We notify the reader that in regimes where a Lévy flights process converges to some limiting distribution, it does not locate the global minimum of U, but reveals the *spatial* structure of the potential. However, the results presented in this paper allow us to design a class of non-local search algorithms which successfully locate the global minimum of the potential. We address the reader to our recent paper [18] on a theoretical and numerical description of these algorithms.

The present paper contains a heuristic derivation of results, which are proved rigorously in [19]. It can be seen as a sequel of [20–22] where a small-noise dynamics of Lévy flights in external potentials was studied. We emphasize that our methods are purely probabilistic.

2. Object of study and results

2.1. Lévy flights

Throughout this paper we understand a Lévy flights (LF) process $L = (L(t))_{t\geqslant 0}$ of stability index $\alpha \in (0, 2)$ as a Lévy process (i.e. a stochastically continuous process with independent stationary increments and sample paths being right continuous and having left limits) whose marginals have the Fourier transform

$$\mathbf{E}\, e^{i\omega L(t)} = e^{-c(\alpha)t|\omega|^\alpha}, \qquad c(\alpha) = 2\int_0^\infty \frac{1-\cos y}{y^{1+\alpha}}\,dy = 2\left|\cos\left(\frac{\pi\alpha}{2}\right)\Gamma(-\alpha)\right|. \tag{5}$$

There exists a broad physical and mathematical literature on Lévy flights. We refer the reader to monographs by Sato [23] and Uchaikin and Zolotarev [24] and a topical review [25] by Metzler and Klafter.

We emphasize that our definition of Lévy flights slightly differs from that used by other authors. For example, Eliazar and Klafter [26, 27], Chechkin *et al* [28–31], Sokolov [32], Brockmann and Sokolov [33] and Ditlevsen [34] consider the Lévy flights process $L'(t) = c(\alpha)^{-1/\alpha}L(t)$ with the Fourier transform $\mathbf{E}\, e^{i\omega L'(t)} = e^{-t|\omega|^\alpha}$. This difference is not essential for the analysis as long as the stability index $\alpha \in (0, 2)$ is fixed. However, one should be careful when deriving theoretical and numerical results for α varying over the interval $(0, 2)$, especially for α from small neighbourhoods of 0 and 2. We shall comment on this subject in section 2.3.

In our analysis we shall use the Lévy–Khinchin representation of the characteristic function of $L(t)$, namely

$$\mathbf{E}\, e^{i\omega L(t)} = \exp\left\{t\int_{\mathbb{R}\backslash\{0\}} [e^{i\omega y} - 1 - i\omega y \mathbb{I}\{|y| \leqslant 1\}]\frac{dy}{|y|^{1+\alpha}}\right\}, \tag{6}$$

where $\mathbb{I}\{A\}$ denotes the indicator function of a set A. The most important ingredient of the representation (6) is the so-called *Lévy* (*jump*) *measure* of the random process L, given by

$$\nu(A) = \int_{A\backslash\{0\}} \frac{dy}{|y|^{1+\alpha}}, \qquad A \text{ is a Borel set in } \mathbb{R}. \tag{7}$$

It is easy to see that the Lévy measure ν' of the Lévy flights process L' has the density $\nu'(dy) = c(\alpha)^{1/\alpha}|y|^{-1-\alpha}\mathbb{I}(y \neq 0)\,dy$.

The measure ν controls the intensity and sizes of the jumps of the Lévy flights process. Let $\Delta L(t) = L(t) - L(t-)$ be the random jump size of L at a time instance t (here we use the existence of the left limit $L(t-)$), $t > 0$, and the number of jumps belonging to the set A on the time interval $(0, t]$ be denoted by $N(t, A)$, i.e.

$$N(t, A) = \sharp\{s : (s, \Delta L(s)) \in (0, t] \times A\}. \tag{8}$$

Then the random variable $N(t, A)$ has a Poisson distribution with mean $t\nu(A)$ (which can possibly be infinite or zero), see e.g. Sato [23, chapter 4]. It is helpful to note that for any stability index $\alpha \in (0, 2)$, the Lévy measure of any neighbourhood of 0 is infinite; hence LFs make infinitely (countably) many very small jumps on any time interval. The tails of the density

$\nu(\mathrm{d}y) = |y|^{-1-\alpha}\mathbb{I}(y \neq 0)\,\mathrm{d}y$ determine big jumps of LFs. Thus, $\mathbf{E}|L(t)|^\delta < \infty, t > 0$, iff $\int_{|y|\geqslant 1} |y|^\delta \nu(\mathrm{d}y) < \infty$ iff $\delta < \alpha$.

2.2. *External potential*

We assume that the external potential U is smooth and has n local minima m_i and $n-1$ local maxima s_i enumerated in the increasing order, i.e.

$$-\infty = s_0 < m_1 < s_1 < \cdots < s_{n-1} < m_n < s_n = +\infty. \tag{9}$$

We also assume that local extrema are non-degenerate, i.e. $U''(m_i) > 0$ and $U''(s_i) < 0$, and the potential increases fast at infinity, i.e. $|U'(x)| > |x|^{1+c}$, $|x| \to \infty$ for some $c > 0$.

Under the assumptions on U, the deterministic dynamical system

$$X_x^0(t) = x - \int_0^t U'\big(X_x^0(u)\big)\,\mathrm{d}u \tag{10}$$

has n domains of attraction $\Omega_i = (s_{i-1}, s_i)$ with asymptotically stable attractors m_i. We note that if $x \in \Omega_i$ then $X_x^0(t) \in \Omega_i$ for all $t \geqslant 0$, i.e. the deterministic trajectory cannot pass between different domains of attraction. Denote $B_i = \{x : |m_i - x| \leqslant \Delta\}$ as a Δ-neighbourhood of the attractor m_i. We suppose that Δ is small enough, so that $B_i \subset \Omega_i$, $1 \leqslant i \leqslant n$. Due to the rapid increase of U' at infinity, the return of $X_x^0(t)$ from $\pm\infty$ to B_1 or B_n occurs in finite time.

2.3. *Small constant noise amplitude*

First, we consider the dynamics of the LFs εL in the external potential U in the limit of a small time-independent scale $\varepsilon \to 0$. This dynamics is described by the stochastic differential equation

$$X_x^\varepsilon(t) = x - \int_0^t U'\big(X_x^\varepsilon(u-)\big)\,\mathrm{d}u + \varepsilon L(t), \qquad x \in \mathbb{R}, \quad t \geqslant 0. \tag{11}$$

In the case of arbitrary fixed noise amplitude, equation (11) was studied analytically and numerically by Jespersen *et al* [35], Chechkin *et al* [28, 29], and Eliazar and Klafter [26, 27]. Small noise asymptotics $\varepsilon \to 0$ of X^ε, in particular the barrier crossing problem and the asymptotics of Kramers' times, were studied by Chechkin *et al* in [30, 31].

In our previous papers [20–22], we developed a new purely probabilistic approach to the LF dynamics in the limit of small ε. In particular, we obtained a law and the mean value of the first exit time of a LF process from a potential well (Kramers' time) and studied the meta-stable behaviour of X^ε in a multi-well potential. Below we formulate the results.

For any $\Delta > 0$ sufficiently small, in the limit $\varepsilon \to 0$, the process X^ε spends an overwhelming proportion of time in the set $\cup_{i=1}^n B_i$ making occasional abrupt jumps between different neighbourhoods B_i. Thus, the knowledge of the transition times and probabilities is essential for understanding the asymptotic properties of X^ε. Let $T_x^i(\varepsilon) = \inf\big\{t \geqslant 0 : X_x^\varepsilon(t) \in \cup_{j\neq i} B_j\big\}$. For $x \in B_i$, the stopping time $T_x^i(\varepsilon)$ denotes the first transition time to a Δ-neighbourhood of a minimum of a different well. Then we proved the following result.

Theorem 2.1 (Transitions, [22])**.** *For $x \in B_i$, $1 \leqslant i \leqslant n$, the following estimates hold in the limit $\varepsilon \to 0$:*

$$\mathbf{P}_x(X^\varepsilon(T^i(\varepsilon)) \in B_j) \to q_{ij}q_i^{-1}, \qquad i \neq j, \tag{12}$$

$$\mathbf{P}_{(x)}(\varepsilon^\alpha T^i(\varepsilon) \geqslant U) \to \mathrm{e}^{-q_i U}, \qquad U \geqslant 0, \tag{13}$$

$$\varepsilon^{\alpha}\mathbf{E}_x T^i(\varepsilon) \to q_i^{-1}, \tag{14}$$

where

$$q_{ij} = \alpha^{-1}||s_{j-1} - m_i|^{-\alpha} - |s_j - m_i|^{-\alpha}|, \qquad i \neq j, \tag{15}$$

$$q_i = \sum_{j\neq i} q_{ij} = \alpha^{-1}(|s_{i-1} - m_i|^{-\alpha} + |s_i - m_i|^{-\alpha}). \tag{16}$$

As we see, the transition times between the wells of X^{ε} are asymptotically exponentially distributed in the limit of small noise, and hence *unpredictable*, due to the memoryless property of the exponential law. The transition probabilities between the wells are noise independent and strictly positive. Thus, X^{ε} reminds of a Markov process on a finite state space. Indeed, the following theorem holds.

Theorem 2.2 (Metastability, [22])**.** *If $x \in \Omega_i$, $1 \leqslant i \leqslant n$, then for $t > 0$*

$$X_x^{\varepsilon}(t\varepsilon^{-\alpha}) \to Y_{m_i}(t), \qquad \varepsilon \to 0, \tag{17}$$

in the sense of finite-dimensional distributions, where $Y = (Y_y(t))_{t\geqslant 0}$ is a continuous time Markov chain on a state space $\{m_1, \ldots, m_n\}$ with the infinitesimal generator $Q = (q_{ij})_{i,j=1}^n$, q_{ij} being defined in (15) and (16), $q_{ii} = -q_i$.

Since none of the entries q_{ij} vanishes, the limiting Markov process Y has a unique invariant distribution $\pi = (\pi_1, \ldots, \pi_n)^T$, which can be calculated from the matrix equation $Q^T\pi = 0$.

For example, in the case of a double-well potential, $n = 2$, with local minima at $m_1 < 0 < m_2$ and a saddle point s_1 at the origin, the generator matrix Q from theorem 2.2 takes the form

$$Q = \frac{1}{\alpha}\begin{pmatrix} -|m_1|^{-\alpha} & |m_1|^{-\alpha} \\ m_2^{-\alpha} & -m_2^{-\alpha} \end{pmatrix}. \tag{18}$$

Solving the Fokker–Planck equation $Q^T\pi = 0$ for the invariant measure $\pi = (\pi_1, \pi_2)^T$ with the normalizing condition $\pi_1, \pi_2 \geqslant 0$, $\pi_1 + \pi_2 = 1$ one finds the limiting distribution of $X^{\varepsilon}(t)$ as $t \to +\infty$ and $\varepsilon \to 0$:

$$\pi_1 = \frac{|m_1|^{\alpha}}{|m_1|^{\alpha} + m_2^{\alpha}}, \qquad \pi_2 = \frac{m_2^{\alpha}}{|m_1|^{\alpha} + m_2^{\alpha}}. \tag{19}$$

Now we comment on the parameterization of LFs (5) and (6). As it is seen from the proof of theorems 2.1 and 2.2, only the weight of the tails of the jump measure ν is important for the barrier crossing. In other words, the Lévy particle does not climb up the potential barrier but penetrates it at one big jump. Consider the tails of the jump measures ν and ν', namely

$$T_{\alpha}(x) = \int_{|x|\leqslant y} \frac{\mathrm{d}y}{|y|^{1+\alpha}} = \frac{1}{\alpha x^{\alpha}}, \tag{20}$$

$$T'_{\alpha}(x) = c(\alpha)^{1/\alpha} \int_{|x|\leqslant y} \frac{\mathrm{d}y}{|y|^{1+\alpha}} = \frac{c(\alpha)^{1/\alpha}}{\alpha x^{\alpha}}, \qquad x > 0. \tag{21}$$

The tail function $T_{\alpha}(x)$ is well defined for all $\alpha > 0$, whereas T'_{α} can be considered only for $\alpha \in (0, 2)$. The reason for this is the discontinuity of the prefactor $c(\alpha)$ as $\alpha \uparrow 2$. On the other hand, $c(\alpha)$ describes small jumps of the LF process which play no important role for the small noise asymptotics $\varepsilon \to 0$. Thus, the parametrization of the process L' mixes small and big jump features, whereas the parametrization of the process L allows us to take into account only its heavy-tail dynamics.

A further discussion on different parametrizations can be found in appendix C of [31] by Chechkin *et al*, where the authors showed that the results of theorem 2.1 are in good agreement with numerical experiments.

In fact, theorems 2.1 and 2.2 formulated in this paper for LFs are proven in [22] in a much more general setting which allows general Lévy noises with power tails and a Gaussian component. An interesting class of such processes is constituted by the so-called *weakly tempered LFs* $H = (H(t))_{t\geqslant 0}$, which are defined via their Fourier transform

$$\mathbf{E}\, \mathrm{e}^{\mathrm{i}\omega H(t)} = \exp\left\{ t\int_{\mathbb{R}\setminus\{0\}} [\mathrm{e}^{\mathrm{i}\omega y} - 1 - \mathrm{i}\omega y \mathbb{I}\{|y|\leqslant 1\}] \frac{\mathrm{d}y}{|y|^{1+\alpha}(1+y^2)^{\beta/2}} \right\}. \tag{22}$$

The jump measure of the process H has a density $\mu(\mathrm{d}y) = |y|^{-1-\alpha}(1+y^2)^{-\beta/2}\mathbb{I}(y\neq 0)\,\mathrm{d}y$ with parameters $\alpha\in(0,2)$ and $\beta\geqslant 0$. It is clear that if $\beta = 0$, the process H is just an α-stable Lévy process.

Since the values of the jump measure in the vicinity of the origin control *small jumps* of H, and since $|y|^{-1-\alpha}(1+y^2)^{-\beta/2}\approx |y|^{-1-\alpha}$ as $|y|\to 0$, we deduce that small jumps of H have the same size and intensity as of a LF process with the stability index α. On the other hand, since $|y|^{-1-\alpha}(1+y^2)^{-\beta/2}\approx |y|^{-1-\alpha-\beta}$ as $|y|\to\infty$, the big, *extremal jumps* of H, which are responsible for the exit from the potential well, have power tails of the order $\alpha+\beta$. Thus, $\mathbf{E}|H(t)|^{\delta}<\infty$ iff $\delta<\alpha+\beta$, and H has finite variance if $\alpha+\beta>2$. Writing $\alpha+\beta$ instead of α in theorems 2.1 and 2.2, we obtain Kramers' times and metastability results for the jump diffusions driven by the process H.

We emphasize that no explicit formula for the Fourier transform (22) is needed for the proof: it is enough to know the tail behaviour of the process's jump measure.

Finally, we note that theorems 2.1 and 2.2 also hold for asymmetric LFs with the jump measure $\nu(\mathrm{d}y) = c_1|y|^{-1-\alpha}\mathbb{I}\{y<0\} + c_2 y^{-1-\alpha}\mathbb{I}\{y>0\}$, $c_1, c_2\geqslant 0$, $c_1+c_2>0$, and provide a theoretic background for the numerical experiments by Dybiec *et al* [36].

2.4. Decreasing noise amplitude

In the annealed regime, the dynamics of Lévy flights is characterized by the time non-homogeneous equation

$$Z^{\lambda}_{s,z}(t) = z - \int_s^t U'\big(Z^{\lambda}_{s,z}(u-)\big)\,\mathrm{d}u + \int_s^t \frac{\mathrm{d}L(u)}{(\lambda+u)^{\theta}}, \qquad z\in\mathbb{R},\quad 0\leqslant s\leqslant t, \tag{23}$$

where a positive parameter θ denotes the *cooling rate* and $\lambda>0$ determines the initial scale, which equals to $(\lambda+s)^{-\theta}$.

It is easily seen from (23) that the evolution of the process starting at time $s\geqslant 0$ is the same as that of the process starting at time zero with a different initial scale, namely

$$\big(Z^{\lambda}_{s,z}(s+t)\big)_{t\geqslant 0} \overset{d}{=} \big(Z^{\lambda+s}_{0,z}(t)\big)_{t\geqslant 0}, \tag{24}$$

and thus the particular values of s or λ do not influence the asymptotic properties of the process in the limit $t\to\infty$. However, since our theory will work for small scales, it is often convenient to study the dynamics not for large values of s and t but for large values of λ.

The goal of this paper is to study the limiting behaviour of $Z^{\lambda}_{0,z}(t)$ as $t \to \infty$ in dependence on the cooling rate θ, inverse initial scale λ and initial point z.

Similar to the classical Gaussian case discussed in the introduction, the candidate for the limiting law of $Z^{\lambda}_{0,z}(t)$ is the invariant distribution π of the Markov chain Y from theorem 2.2. Furthermore, we have to distinguish between two different cooling regimes.

As in the previous section, for $1 \leqslant i \leqslant n$, consider the stopping times

$$\tau^{i,\lambda}_{s,z} = \inf\big\{u \geqslant s : Z^{\lambda}_{s,z}(u) \in \cup_{j\neq i} B_j\big\}. \tag{25}$$

If $z \in B_i$, then $\tau^{i,\lambda}_{s,z}$ denotes the *transition time* from a Δ-neighbourhood of m_i to a Δ-neighbourhood of some other potential's minimum. For all $j \neq i$, we also consider the corresponding *transition probabilities* $\mathbf{P}_{s,z}(Z^{\lambda}(\tau^{i,\lambda}) \in B_j)$. Then the following analogue of theorem 2.1 holds.

Theorem 2.3 (Slow cooling, transitions)**.** *Let $\theta < 1/\alpha$. For $z \in B_i$, $1 \leqslant i \leqslant n$, the following estimates hold in the limit $\lambda \to +\infty$:*

$$\mathbf{P}_{0,z}(Z^{\lambda}(\tau^{i,\lambda}) \in B_j) \to q_{ij}q_i^{-1}, \qquad i \neq j, \tag{26}$$

$$\lambda^{-\alpha\theta}\mathbf{E}_{0,z}\tau^{i,\lambda} \to q_i^{-1}, \tag{27}$$

q_i and q_{ij} being defined in (15) and (16) respectively.

Theorem 2.4 (Slow cooling, convergence)**.** *Let $\theta < 1/\alpha$. Then for any $\lambda > 0$, $z \in \mathbb{R}$, the law of $Z^{\lambda}_{0,z}(t)$ converges weakly to the measure π, i.e. for any continuous and bounded function f, we have*

$$\mathbf{E}_{0,z} f(Z^{\lambda}(t)) \to \sum_{j=1}^{n} f(m_j)\pi_j, \qquad t \to \infty. \tag{28}$$

As we see, the difference between the Gaussian and the Lévy dynamics is huge. In the appropriate annealing regime, the Gaussian diffusion $\hat{Z}$ converges with probability 1 to the coordinate of the global minimum of U. In contrast, the limiting law of the Lévy flights jump diffusion is the measure π, which has positive masses at all local minima of U. Moreover, it is easy to see that the masses $\pi_1, \ldots, \pi_n$ depend not on the heights of the potential barriers of U, but only on the distances between its local minima and maxima.

If the cooling rate θ is above the threshold $1/\alpha$, the solution Z^{λ} gets trapped in one of the wells and thus the convergence fails. Consider the first exit time from the ith well

$$\sigma^{i,\lambda}_{s,z} = \inf\big\{t \geqslant 0 : Z^{\lambda}_{s,z} \notin \Omega_i\big\}. \tag{29}$$

Then the following trapping result holds.

Theorem 2.5 (Fast cooling, trapping)**.** *Let $\theta > 1/\alpha$. For $z \in B_i$, $1 \leqslant i \leqslant n$,*

$$\mathbf{P}_{0,z}(\sigma^{i,\lambda} < \infty) = \mathcal{O}(\lambda^{1-\alpha\theta}), \qquad \lambda \to \infty. \tag{30}$$

Consequently, $\mathbf{E}_{0,z}\sigma^{i,\lambda} = \infty$.

In the subsequent section, we sketch the proof of theorems 2.3–2.5 and discuss the results.

3. Predominant behaviour of the annealed process

3.1. Big and small jumps of a Lévy flights process

Our study of the random process Z^λ is based on the probabilistic analysis of its sample paths. We use the decomposition of the process L into small- and big-jump parts similar to that used in [20–22]. Thus, we refer the reader to these papers for details and sketch the idea briefly.

With the help of the Lévy–Khinchin formula (6), we decompose the process L into a sum of two independent Lévy processes with relatively small and big jumps. For any cooling rate $\theta > 0$, we introduce two new jump measures by setting

$$\nu_\xi^\lambda(A) = \nu(A \cap \{x : |x| \leqslant \lambda^{\theta/2}\}), \tag{31}$$

$$\nu_\eta^\lambda(A) = \nu(A \cap \{x : |x| > \lambda^{\theta/2}\}), \tag{32}$$

and two Lévy processes ξ^λ and η^λ with the corresponding Fourier transforms:

$$\mathbf{E}\, \mathrm{e}^{\mathrm{i}\omega\xi_t^\lambda} = \exp\left\{t \int_{\mathbb{R}\setminus\{0\}} [\mathrm{e}^{\mathrm{i}\omega y} - 1 - \mathrm{i}\omega y \mathbb{I}\{|y| \leqslant 1\}]\nu_\xi^\lambda(\mathrm{d}y)\right\}, \tag{33}$$

$$\mathbf{E}\, \mathrm{e}^{\mathrm{i}\omega\eta_t^\lambda} = \exp\left\{t \int_{\mathbb{R}\setminus\{0\}} [\mathrm{e}^{\mathrm{i}\omega y} - 1 - \mathrm{i}\omega y \mathbb{I}\{|y| \leqslant 1\}]\nu_\eta^\lambda(\mathrm{d}y)\right\}. \tag{34}$$

It is clear that the processes ξ^λ and η^λ are independent, and $L \stackrel{d}{=} \xi^\lambda + \eta^\lambda$.

Since $\nu_\xi^\lambda(\mathbb{R}) = \infty$, the process ξ^λ makes infinitely many jumps on each time interval. Its jumps are, however, bounded by the threshold $\lambda^{\theta/2}$, i.e. $\left|\Delta\xi_t^\lambda\right| \leqslant \lambda^{\theta/2}$. Thus, ξ_t^λ has a finite variance and more generally moments of all orders.

In contrast, the Lévy measure of the process η^λ is finite, and its mass equals

$$\beta_\lambda = \nu_\eta^\lambda(\mathbb{R}) = \int_{-\infty}^{-\lambda^{\theta/2}} \frac{\mathrm{d}y}{|y|^{1+\alpha}} + \int_{\lambda^{\theta/2}}^{\infty} \frac{\mathrm{d}y}{y^{1+\alpha}} = 2\int_{\lambda^{\theta/2}}^{\infty} \frac{\mathrm{d}y}{y^{1+\alpha}} = \frac{2}{\alpha}\lambda^{-\alpha\theta/2}. \tag{35}$$

Hence, η^λ is a compound Poisson process with jumps of absolute value larger than $\lambda^{\theta/2}$. Let τ_k^λ and W_k^λ, $k \geqslant 0$, respectively, be the jump arrival times and jump sizes of η^λ under the convention $\tau_0^\lambda = W_0^\lambda = 0$. The inter-arrival times $T_k^\lambda = \tau_k^\lambda - \tau_{k-1}^\lambda$, $k \geqslant 1$, are independent and exponentially distributed with mean β_λ^{-1}. The jump sizes W_k^λ are also independent random variables with the probability distribution function given by

$$\mathbf{P}\big(W_k^\lambda < u\big) = \frac{\nu_\eta^\lambda(-\infty, u)}{\nu_\eta^\lambda(\mathbb{R})} = \frac{1}{\beta_\lambda}\int_{-\infty}^{u} \mathbb{I}\{|y| > \lambda^{\theta/2}\}\nu_\eta^\lambda(\mathrm{d}y). \tag{36}$$

Finally, we can represent the random perturbation in (23) as a sum of two processes, namely

$$\int_0^t \frac{\mathrm{d}L(u)}{(\lambda+u)^\theta} = \int_0^t \frac{\mathrm{d}\xi_u^\lambda}{(\lambda+u)^\theta} + \sum_{k=1}^{\infty} \frac{W_k^\lambda}{\big(\lambda+\tau_k^\lambda\big)^\theta}\mathbb{I}\{t \geqslant \tau_k\}. \tag{37}$$

3.2. Predominant behaviour

Consider now the process $Z_{0,z}^\lambda$ given by equation (23). On the inter-arrival intervals $\big[\tau_{k-1}^\lambda, \tau_k^\lambda\big)$, $k \geqslant 1$, it is driven only by the process $\varphi_t^\lambda = \int_0^t (\lambda+u)^{-\theta}\, \mathrm{d}\xi_u^\lambda$, and at the time instants τ_k^λ it makes jumps of the size $W_k^\lambda \big/ \big(\lambda + \tau_\lambda^k\big)^\theta$. Recall that the jumps of ξ^λ are bounded by $\lambda^{\theta/2}$;

hence the jumps sizes of φ^λ tend to zero as $\lambda \to \infty$ for all $t \geqslant 0$, i.e.

$$\left|\Delta\varphi_t^\lambda\right| \leqslant \frac{\lambda^{\theta/2}}{(\lambda+t)^\theta} \leqslant \frac{1}{\lambda^{\theta/2}}. \tag{38}$$

The variance of φ_t^λ tends to zero in the limit of large λ, and the random trajectory $Z_{0,z}^\lambda(t)$ can be seen as a small random perturbation of the deterministic trajectory $X_z^0(t)$ of the underlying dynamical system on the intervals $\left[\tau_{k-1}^\lambda, \tau_k^\lambda\right)$. Consider a well Ω_i with a minimum m_i. Let initial points z be away from the unstable points s_{i-1} and s_i, namely $z \in (s_{i-1}+\lambda^{-\gamma}, s_i-\lambda^{-\gamma})$ for some positive γ. Then the deterministic trajectory $X_z^0(t)$ reaches a $\lambda^{-\gamma}$-neighbourhood of m_i in at most logarithmic time $\mathcal{O}(\ln\lambda)$. Since the periods between the big jumps are essentially longer, i.e.

$$\mathbf{E}T_k^\lambda = \mathbf{E}\left(\tau_k^\lambda - \tau_{k-1}^\lambda\right) = \frac{\alpha}{2}\lambda^{\alpha\theta/2} \gg \mathcal{O}(\ln\lambda), \qquad \lambda \to +\infty, \tag{39}$$

we can show that with probability close to 1 the random trajectory Z^λ is located in a small neighbourhood of m_i before the big jump.

Thus, we can summarize the pathwise behaviour of $Z_{0,z}^\lambda$ for large values of λ as follows:

$$\begin{aligned}
&Z_{0,z}^\lambda(0) = z \in (s_{i-1}+\lambda^{-\gamma}, s_i-\lambda^{-\gamma}),\\
&Z_{0,z}^\lambda\left(\tau_\lambda^1-\right) \approx m_i,\\
&Z_{0,z}^\lambda\left(\tau_\lambda^1\right) \approx m_i + \frac{W_1^\lambda}{\left(\lambda+\tau_1^\lambda\right)^\theta} \in (s_{i-1}+\lambda^{-\gamma}, s_i-\lambda^{-\gamma})\\
&Z_{0,z}^\lambda\left(\tau_\lambda^2-\right) \approx m_i,\\
&\ldots\\
&Z_{0,z}^\lambda\left(\tau_\lambda^k-\right) \approx m_i,\\
&Z_{0,z}^\lambda\left(\tau_\lambda^k\right) \approx m_i + \frac{W_k^\lambda}{\left(\lambda+\tau_k^\lambda\right)^\theta} \in (s_{j-1}+\lambda^{-\gamma}, s_j-\lambda^{-\gamma}), \qquad j \neq i,\\
&Z_{0,z}^\lambda\left(\tau_\lambda^{k+1}-\right) \approx m_j,\\
&\ldots,
\end{aligned} \tag{40}$$

whereas on the intervals $\left[\tau_{k-1}^\lambda, \tau_k^\lambda\right)$ the process Z^λ follows the deterministic trajectory X^0. Thus, since we know the initial location of the particle, as well as the jump sizes W_k^λ and jump times τ_k^λ, we can catch the essential features of the random path Z^λ.

Of course, we have to be careful when dealing with trajectories which occasionally enter the $\lambda^{-\gamma}$-neighbourhoods of the saddle points s_i, where the force field U' becomes insignificant. In these neighbourhoods, the Lévy particle has no strong deterministic drift which brings it to a certain well's minimum. Thus, we cannot decide whether Z^λ converges to m_i or to m_{i+1}. However, in the limit $\lambda \to \infty$, the probability that Z^λ jumps from a neighbourhood of m_i to a $\lambda^{-\gamma}$-neighbourhood of some s_j is negligible. In our further exposition, we do not consider the unstable dynamics in these $\lambda^{-\gamma}$-neighbourhoods and assume that (40) holds for all $z \in \Omega_i$. Interested readers can find rigorous arguments in [19].

4. Transitions between the wells in the slow cooling regime. Proof of theorem 2.3

4.1. Mean transition time

Let us obtain the mean value of the first exit time $\sigma^{i,\lambda}$ form the well Ω^i in the limit $\lambda \to \infty$. Indeed, Z^λ can roughly leave Ω_i only at one of the time instants τ_k^λ when

$m_i + W_k^\lambda/\left(\lambda+\tau_k^\lambda\right)^\theta \notin \Omega_i$. We can therefore calculate the mean value of $\sigma^{i,\lambda}$ using the total probability formula:

$$\begin{aligned}\mathbf{E}_{0,z}\sigma^{i,\lambda} &\approx \sum_{k=1}^\infty \mathbf{E}\left[\tau_k^\lambda \mathbb{I}\{\sigma^{i,\lambda}=\tau_k^\lambda\}\right]\\ &\approx \sum_{k=1}^\infty \mathbf{E}\Bigg[\tau_k^\lambda\cdot\mathbb{I}\Bigg\{m_i+\frac{W_1^\lambda}{\left(\lambda+\tau_1^\lambda\right)^\theta}\in\Omega_i,\ldots,m_i\\ &\quad+\frac{W_{k-1}^\lambda}{\left(\lambda+\tau_{k-1}^\lambda\right)^\theta}\in\Omega_i, m_i+\frac{W_k^\lambda}{\left(\lambda+\tau_k^\lambda\right)^\theta}\notin\Omega_i\Bigg\}\Bigg].\end{aligned}\tag{41}$$

Since the arrival times $\tau_1^\lambda, \tau_2^\lambda, \ldots, \tau_k^\lambda$, are dependent, no straightforward calculation of the expectations in the latter sum seems possible. However, we can estimate these expectations from above and below. Our argument is based on the inequalities $0<\tau_1^\lambda<\tau_2^\lambda<\cdots<\tau_k^\lambda$, and the obvious inclusions

$$\left\{m_i+\frac{W_j^\lambda}{\lambda^\theta}\in\Omega_i\right\}\subseteq\left\{m_i+\frac{W_j^\lambda}{\left(\lambda+\tau_j^\lambda\right)^\theta}\in\Omega_i\right\}\subseteq\left\{m_i+\frac{W_j^\lambda}{\left(\lambda+\tau_k^\lambda\right)^\theta}\in\Omega_i\right\},\quad 1\leqslant j\leqslant k-1,\tag{42}$$

where the probability of these events can be calculated explicitly from (36), to yield the formula

$$\mathbf{P}\left(m_i+\frac{W_1^\lambda}{(\lambda+t)^\theta}\notin\Omega_i\right)=\frac{1}{\beta_\lambda}\left(\int_{-\infty}^{-|m_i-s_{i-1}|(\lambda+t)^\theta}+\int_{(s_i-m_i)(\lambda+t)^\theta}^\infty\right)\frac{\mathrm{d}y}{|y|^{1+\alpha}}=\frac{q_i}{\beta_\lambda(\lambda+t)^{\alpha\theta}}.\tag{43}$$

Let us obtain the estimate for $\mathbf{E}_{0,z}\sigma^{i,\lambda}$ from above. Note that for each $k\geqslant 1$, the arrival time τ_k^λ is a sum of k independent exponentially distributed random variables T_j^λ and thus has a Gamma(k,β_λ) distribution with a probability density $\beta_\lambda\,\mathrm{e}^{-\beta_\lambda t}(\beta_\lambda t)^{k-1}/(k-1)!$, $t\geqslant 0$. Then, applying the second inclusion in (42) we obtain

$$\begin{aligned}&\mathbf{E}\left[\tau_k^\lambda\cdot\mathbb{I}\left\{m_i+\frac{W_1^\lambda}{\left(\lambda+\tau_1^\lambda\right)^\theta}\in\Omega_i,\ldots,m_i+\frac{W_{k-1}^\lambda}{\left(\lambda+\tau_{k-1}^\lambda\right)^\theta}\in\Omega_i, m_i+\frac{W_k^\lambda}{\left(\lambda+\tau_k^\lambda\right)^\theta}\notin\Omega_i\right\}\right]\\ &\quad\leqslant\mathbf{E}\left[\tau_k^\lambda\cdot\mathbb{I}\left\{m_i+\frac{W_1^\lambda}{\left(\lambda+\tau_k^\lambda\right)^\theta}\in\Omega_i,\ldots,m_i+\frac{W_{k-1}^\lambda}{\left(\lambda+\tau_k^\lambda\right)^\theta}\in\Omega_i, m_i+\frac{W_k^\lambda}{\left(\lambda+\tau_k^\lambda\right)^\theta}\notin\Omega_i\right\}\right]\\ &\quad=\int_0^\infty\beta_\lambda t\,\mathrm{e}^{-\beta_\lambda t}\frac{(\beta_\lambda t)^{k-1}}{(k-1)!}\mathbf{P}\left(m_i+\frac{W_1^\lambda}{(\lambda+t)^\theta}\in\Omega_i\right)^{k-1}\mathbf{P}\left(m_i+\frac{W_1^\lambda}{(\lambda+t)^\theta}\notin\Omega_i\right)\mathrm{d}t\\ &\quad=\int_0^\infty\beta_\lambda t\,\mathrm{e}^{-\beta_\lambda t}\frac{(\beta_\lambda t)^{k-1}}{(k-1)!}\left[1-\frac{q_i}{\beta_\lambda(\lambda+t)^{\alpha\theta}}\right]^{k-1}\frac{q_i}{\beta_\lambda(\lambda+t)^{\alpha\theta}}\,\mathrm{d}t.\end{aligned}\tag{44}$$

Summation over k yields

$$\begin{aligned}\mathbf{E}_{0,z}\sigma^{i,\lambda}&\lesssim\int_0^\infty\beta_\lambda t\,\mathrm{e}^{-\beta_\lambda t}\frac{q_i}{\beta_\lambda(\lambda+t)^{\alpha\theta}}\sum_{k=1}^\infty\frac{(\beta_\lambda t)^{k-1}}{(k-1)!}\left[1-\frac{q_i}{\beta_\lambda(\lambda+t)^{\alpha\theta}}\right]^{k-1}\mathrm{d}t\\ &=\int_0^\infty\frac{q_i t}{(\lambda+t)^{\alpha\theta}}\exp\left(-\frac{q_i t}{(\lambda+t)^{\alpha\theta}}\right)\mathrm{d}t,\end{aligned}\tag{45}$$

where '$\lesssim$' emphasizes that the inequality holds in the limit $\lambda\to+\infty$. Since $\alpha\theta<1$, the latter integral converges for all $\lambda>0$, and it is possible to estimate its asymptotic value. Introducing a new variable $u = \frac{\lambda+t}{\lambda}$, we transform the integral to the so-called Laplace-type integral with a big parameter, which can be evaluated asymptotically (see [37, chapter 3]), i.e.

$$\begin{aligned} \mathbf{E}_{0,z}\sigma^{i,\lambda} &\lesssim \lambda^{2-\alpha\theta}\int_1^\infty \frac{q_i(u-1)}{u^{\alpha\theta}}\exp\left(-\frac{q_i(u-1)\lambda^{1-\alpha\theta}}{u^{\alpha\theta}}\right)\mathrm{d}u \\ &\approx \lambda^{2-\alpha\theta}\int_1^\infty q_i(u-1)\exp(-q_i(u-1)\lambda^{1-\alpha\theta})\,\mathrm{d}u = q_i^{-1}\lambda^{\alpha\theta}. \end{aligned} \tag{46}$$

Applying analogously the first inclusion from (42) to the first $k-1$ jumps, we obtain the estimate from below:

$$\begin{aligned} \mathbf{E}_{0,z}\sigma^{i,\lambda} &\gtrsim \sum_{k=1}^\infty \mathbf{E}\left[\tau_k^\lambda\cdot\mathbb{I}\left\{m_i+\frac{W_1^\lambda}{\lambda^\theta}\in\Omega_i,\ldots,m_i+\frac{W_{k-1}^\lambda}{\lambda^\theta}\in\Omega_i, m_i+\frac{W_k^\lambda}{(\lambda+\tau_k^\lambda)^\theta}\notin\Omega_i\right\}\right] \\ &\geqslant \sum_{k=1}^\infty\int_0^\infty t\beta_\lambda\,\mathrm{e}^{-\beta_\lambda t}\frac{(\beta_\lambda t)^{k-1}}{(k-1)!}\left[1-\frac{q_i}{\beta_\lambda\lambda^{\alpha\theta}}\right]^{k-1}\frac{q_i}{\beta_\lambda(\lambda+t)^{\alpha\theta}}\,\mathrm{d}t \\ &= \int_0^\infty\frac{q_i t}{(\lambda+t)^{\alpha\theta}}\exp\left(-\frac{q_i t}{\lambda^{\alpha\theta}}\right)\mathrm{d}t = \lambda^{2-\alpha\theta}\int_1^\infty\frac{q_i(u-1)}{u^{\alpha\theta}}\exp(-q_i(u-1)\lambda^{1-\alpha\theta}) \\ &\approx \lambda^{2-\alpha\theta}\int_1^\infty q_i(u-1)\exp(-q_i(u-1)\lambda^{1-\alpha\theta})\,\mathrm{d}u = q_i^{-1}\lambda^{\alpha\theta}. \end{aligned} \tag{47}$$

Fortunately, the estimates from below and above coincide, and thus give the asymptotic value of the mean life time of the slowly cooled Lévy particle in a potential well.

To obtain the limit (27) for the mean transition time $\tau^{i,\lambda}$ between the sets B_i and $\cup_{j\neq i}B_j$, we note that at the exit time $\sigma^{i,\lambda}$ the process Z^λ enters some of the wells Ω_j, $j\neq i$, with high probability following the deterministic trajectory, and reaches a Δ-neighbourhood of a well's minimum in a time of the order $\mathcal{O}(\ln\lambda)$, which is negligible in comparison with $\lambda^{\alpha\theta}$. Thus, the limit (27) holds.

4.2. Transition probability

To calculate the transition probability between the wells, it suffices to obtain an estimate from below. Similar to the estimate of the mean exit time, we have

$$\begin{aligned} \mathbf{P}_{0,z}(Z^\lambda(\sigma^{i,\lambda})\in\Omega_j) &\gtrsim \sum_{k=1}^\infty\mathbf{P}\left(m_i+\frac{W_1^\lambda}{\lambda^\theta}\in\Omega_i,\ldots,m_i+\frac{W_{k-1}^\lambda}{\lambda^\theta}\in\Omega_i, m_i+\frac{W_k^\lambda}{(\lambda+\tau_k^\lambda)^\theta}\in\Omega_j\right) \\ &\geqslant \sum_{k=1}^\infty\int_0^\infty\beta_\lambda\,\mathrm{e}^{-\beta_\lambda t}\frac{(\beta_\lambda t)^{k-1}}{(k-1)!}\left[1-\frac{q_i}{\beta_\lambda\lambda^{\alpha\theta}}\right]^{k-1}\frac{q_{ij}}{\beta_\lambda(\lambda+t)^{\alpha\theta}}\,\mathrm{d}t \\ &= \int_0^\infty\frac{q_{ij}}{(\lambda+t)^{\alpha\theta}}\exp\left(-\frac{q_i t}{\lambda^{\alpha\theta}}\right)\mathrm{d}t = \lambda^{1-\alpha\theta}\int_1^\infty\frac{q_{ij}}{u^{\alpha\theta}}\exp(-q_i(u-1)\lambda^{1-\alpha\theta}) \\ &\approx \lambda^{1-\alpha\theta}\int_1^\infty q_{ij}\exp(-q_i(u-1)\lambda^{1-\alpha\theta})\,\mathrm{d}u = q_{ij}q_i^{-1}. \end{aligned} \tag{48}$$

With the help of the equality $\sum_{j\neq i}q_{ij}q_i^{-1}=1$, we conclude that $\mathbf{P}_{0,z}(Z^\lambda(\sigma^{i,\lambda})\in\Omega_j)\to q_{ij}q_i^{-1}$. Finally, since after entering Ω_j, the process Z^λ reaches B_j with high probability, $\mathbf{P}_{0,z}(Z^\lambda(\sigma^{i,\lambda})\in\Omega_j)\approx\mathbf{P}_{0,z}(Z^\lambda(\tau^{i,\lambda})\in B_j)$, and (26) holds.

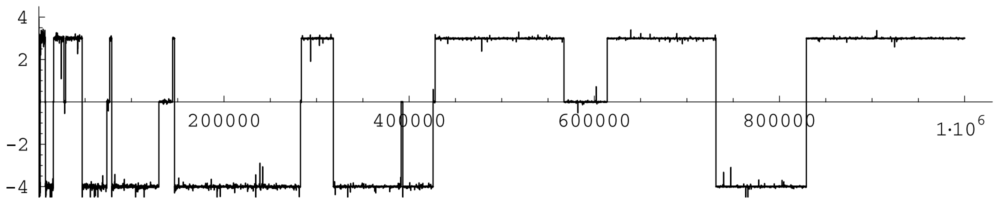

Figure 1. A slow cooling of a Lévy particle in a potential with local minima at −4, 0 and 3.

5. Convergence in the slow cooling regime. Proof of theorem 2.4

Figure 1 illustrates the typical behaviour of Z in the slow cooling regime, $\alpha\theta < 1$. Roughly speaking, we can distinguish two different behaviours when the scale of the random perturbation is big or small.

(1) In a general case, the initial scale $\lambda^{-\theta}$ can be relatively big, so that the asymptotics of theorem 2.3 do not hold. Thus, we have no theory for the transitions of Z^λ between the wells. At some time instance T which depends on the potential U and other parameters of the system, the instant scale $(\lambda + T)^{-\theta}$ becomes low enough so that theorem 2.3 starts working. Moreover, choosing the time T sufficiently large, we make the transition probabilities of Z between the neighbourhoods B_i close to q_{ij}/q_i with any prescribed precision. For brevity, we can also assume that $Z^\lambda_{0,z}(T) \in B_i$ for some i.

(2) Denote $\tau(k)$, $k \geqslant 0$, successive transition times after T between different B_j, $1 \leqslant j \leqslant n$, with $\tau(0) = T$ by convention. The mean values $\mathbf{E}_{0,z}\tau(k)$ are finite and can be calculated from theorem 2.3. Indeed, if $\tau(k-1) = t_{k-1}$ and $Z^\lambda(\tau(k-1)) = z_{k-1} \in B_i$, then the conditional expectation of $\mathbf{E}_{0,z}\tau(k)$ equals

$$\begin{aligned}\mathbf{E}_{0,z}[\tau(k)|\tau(k-1) = t_{k-1}, Z^\lambda(\tau(k-1)) = z_{k-1}] &= t_{k-1} + \mathbf{E}_{t_{k-1},z_{k-1}}\tau^{i,\lambda}\\ &= t_{k-1} + \mathbf{E}_{0,z_{k-1}}\tau^{i,\lambda+t_{k-1}} \approx t_{k-1} + q_i^{-1}(\lambda + t_{k-1})^{\alpha\theta} < \infty. \qquad (49)\end{aligned}$$

From the time instance T on, Z^λ makes transitions between the wells with probabilities close to $p_{ij} = q_{ij}/q_i$, $i \neq j$, where $p_{ii} = 0$ by convention. These probabilities determine a discrete time Markov chain $V(k)$ on $\{m_1, \ldots, m_n\}$, such that $\mathbf{P}(V(k) = m_j|V(k-1) = m_i) = p_{ij}$. It is clear that V has the unique invariant distribution π. Moreover, $V(k)$ converges to the invariant distribution geometrically fast, i.e. there is $0 < \rho < 1$ such that for all $1 \leqslant i, j \leqslant n$ and $k \geqslant 0$,

$$\left|\mathbf{P}_{m_i}(V(k) = m_j) - \pi_j\right| = \mathcal{O}(\rho^k), \qquad k \to \infty. \qquad (50)$$

With the help of the asymptotic relation,

$$\mathbf{P}(Z^\lambda(\tau(k)) \in B_j|Z^\lambda(\tau(k-1)) \in B_i) \approx p_{ij} = \mathbf{P}(V(k) = m_j|V(k-1) = m_i). \qquad (51)$$

one can show that the distributions of $Z(\tau(k))$ and $V(k)$ are also close for $k \geqslant 1$, i.e.

$$\mathbf{P}\big(Z^\lambda_{0,z}(\tau(k)) \in B_j\big|Z^\lambda_{0,z}(T) \in B_i\big) \approx \mathbf{P}_{m_i}(V(k) = m_j). \qquad (52)$$

Hence, with the help of (50) for any prescribed accuracy level we can find $k_0 \geqslant 1$ such that for $k \geqslant k_0$ we have

$$\mathbf{P}_{0,z}(Z^\lambda(\tau(k)) \in B_j) \approx \pi_j, \qquad (53)$$

independent of the initial point z.

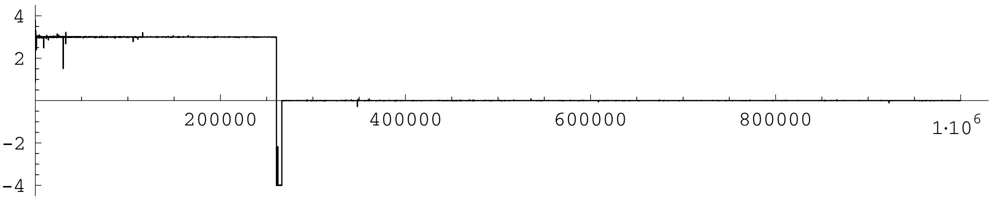

Figure 2. A fast cooling of a Lévy particle in a potential with local minima at −4, 0 and 3.

Finally, we note that after time T, the process Z^λ spends most of the time in the neighbourhoods B_i,

$$Z^\lambda_{0,z}(t) \approx Z^\lambda_{0,z}(\tau(k)) \qquad \text{for} \quad t \in [\tau(k), \tau(k+1)), \tag{54}$$

and thus if $t \geqslant \tau(k_0)$ then $\mathbf{P}_{0,z}(Z^\lambda(t) \in B_j) \approx \pi_j$. This proves the statement of theorem 2.4.

As we see, if $\alpha\theta < 1$, the process Z^λ reminds of a piece-wise constant jump process on the state space $\{m_1, \ldots, m_n\}$. It never stops jumping between the wells of U, and the random sequence $Z^\lambda(\tau(k))$, $k \geqslant k_0$, behaves like a stationary discrete time Markov chain with a distribution π.

6. Trapping in the fast cooling regime. Proof of theorem 2.5

The regime of fast cooling $\theta > 1/\alpha$ is simple. We estimate the probability of the exit from a well. Since the exit occurs with high probability only at the arrival times of the big jump process η^λ, we estimate

$$\begin{aligned}
\mathbf{P}_{0,z}(\sigma^{i,\lambda} < \infty) &\approx \sum_{k=1}^\infty \mathbf{P}_{0,z}\big(\sigma^{i,\lambda} = \tau_k^\lambda\big) \\
&\lesssim \sum_{k=1}^\infty \mathbf{P}\left(m_i + \frac{W_1^\lambda}{\big(\lambda+\tau_1^\lambda\big)^\theta} \in \Omega_i, \ldots, m_i + \frac{W_{k-1}^\lambda}{\big(\lambda+\tau_{k-1}^\lambda\big)^\theta} \in \Omega_i, m_i + \frac{W_k^\lambda}{\big(\lambda+\tau_k^\lambda\big)^\theta} \notin \Omega_i\right) \\
&\leqslant \sum_{k=1}^\infty \mathbf{P}\left(m_i + \frac{W_k^\lambda}{\big(\lambda+\tau_k^\lambda\big)^\theta} \notin \Omega_i\right) = \int_0^\infty \beta_\lambda\, \mathrm{e}^{-\beta_\lambda t} \frac{q_i}{\beta_\lambda(\lambda+t)^{\alpha\theta}} \sum_{k=1}^\infty \frac{(\beta_\lambda t)^{k-1}}{(k-1)!}\, \mathrm{d}t \\
&= \int_0^\infty \frac{q_i}{(\lambda+t)^{\alpha\theta}}\, \mathrm{d}t = \frac{q_i}{\alpha\theta - 1} \frac{1}{\lambda^{\alpha\theta-1}} \to 0, \qquad \lambda \to +\infty.
\end{aligned} \tag{55}$$

As a consequence, we have infinite mean exit times $\mathbf{E}_{0,z}\sigma^{i,\lambda} = \infty$. In other words, if $\theta > 1/\alpha$, the dynamics of Z^λ has two qualitatively different regimes as well. First, for big instant scales, we have no theory for the transitions between the wells. Second, when the scale becomes small enough, the Lévy particle gets trapped in one of the wells; see figure 2. In this case, there is no convergence to the invariant measure π.

7. Conclusion and discussion

In this paper we studied the large time dynamics of a Lévy particle in a multi-well external potential, with the instant scale decreasing with time as $1/t^\theta$.

We discovered that if the cooling is slow, i.e. $\theta < 1/\alpha$, then the system reaches a quasi-stationary regime where the transition probabilities between the wells converge to certain values which are explicitly determined in terms of the potential's spatial geometry. Moreover, the mean transition times between the wells are finite, and between the transitions the process lives in a small neighbourhood of wells' minima. As opposed to the Gaussian simulated annealing, the Lévy flights process does not settle down near the global maximum of U. However, our results can be applied for a search for the global minimum of the potentials, which possess the so-called 'large-rims-have-deep-wells' property, see [38, 39], i.e. when the spatially largest well is at the same time the deepest. Then, having empirical estimates of the local minima locations m_i and the invariant distribution π, we can derive the coordinates of the saddle points s_i, and thus reconstruct the sizes of the wells.

Further, if the cooling is fast, i.e. $\theta > 1/\alpha$, the Lévy particle gets trapped in one of the wells when the scale decreases below some critical level.

In this paper we do not answer the most interesting question: is it possible to detect a global minimum of U with the help of a non-local search and Lévy flights? The answer to this question is affirmative, and the algorithm and simulations are presented in our paper [18].

Acknowledgments

This work was supported by the DFG Research Project 'Stochastic Dynamics of Climate States'. The author is grateful to anonymous referees for careful reading and helpful suggestions which lead to a significant improvement of the manuscript.

References

[1] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 *Science* **220** 671–80
[2] Černý V 1985 *J. Optim. Theory Appl.* **45** 41–51
[3] Metropolis N, Rosenbluth A W, Rosenbluth M N and Teller E 1953 *J. Chem. Phys.* **21** 1087–92
[4] Geman S and Geman D 1984 *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–41
[5] Vanderbilt D and Louie S G 1984 *J. Comput. Phys.* **56** 259–71
[6] Aluffi-Pentini F, Parisi V and Zirilli F 1985 *J. Optim. Theory Appl.* **47** 1–16
[7] Geman S and Hwang C R 1986 *SIAM J. Control Optim.* **24** 1031–43
[8] Gidas B 1985 *24th Conf. Decision and Control (Fort Lauderdale, FL, USA, 11–13 Dec.)*
[9] Chiang T S, Hwang C R and Sheu S J 1987 *SIAM J. Control Optim.* **25** 737–53
[10] Hajek B 1988 *Math. Oper. Res.* **13** 311–29
[11] Holley R and Stroock D 1988 *Commun. Math. Phys.* **115** 553–69
[12] Holley R, Kusuoka S and Stroock D 1989 *J. Funct. Anal.* **83** 333–47
[13] Hwang C R and Sheu S J 1990 *Acta Appl. Math.* **19** 253–95
[14] Hwang C R and Sheu S J 1992 *J. Theor. Probab.* **5** 223–49
[15] Gidas B 1995 *Topics in Contemporary Probability and its Applications (Probability and Stochastics Series)* ed J L Snell (Boca Raton, FL: CRC Press) pp 159–232
[16] Hartmann A K and Rieger H 2002 *Optimization Algorithms in Physics* (Berlin: Wiley)
[17] Szu H and Hartley R 1987 *Phys. Lett.* A **122** 157–62
[18] Pavlyukevich I 2007 *J. Comput. Phys.* doi:10.1016/j.jcp.2007.06.008 *at press*
[19] Pavlyukevich I 2007 *Stoch. Process. Appl.* doi:10.1016/j.spa.2007.07.012 *at press*
[20] Imkeller P and Pavlyukevich I 2006 *Stoch. Process. Appl.* **116** 611–42
[21] Imkeller P and Pavlyukevich I 2006 *J. Phys. A: Math. Gen.* **39** L237–46
[22] Imkeller P and Pavlyukevich I 2007 *at press ESAIM Probab. Stat.* (*Preprint* math/0601771)
[23] Sato K I 1999 Lévy processes and infinitely divisible distributions *Cambridge Studies in Advanced Mathematics* vol 68 (Cambridge: Cambridge University Press)
[24] Uchaikin V V and Zolotarev V M 1999 Chance and stability. Stable distributions and their applications *Modern Probability and Statistics* (Utrecht: VSP)
[25] Metzler R and Klafter J 2004 *J. Phys. A: Math. Gen.* **37** R161–208
[26] Eliazar K and Klafter J 2003 *J. Stat. Phys.* **111** 739–68
[27] Eliazar K and Klafter J 2005 *J. Stat. Phys.* **119** 165–96
[28] Chechkin A V, Klafter J, Gonchar V Y, Metzler R and Tanatarov L V 2003 *Phys. Rev.* E **67** 010102

[29] Chechkin A V, Gonchar V Y, Klafter J, Metzler R and Tanatarov L V 2004 *J. Stat. Phys.* **115** 1505–35
[30] Chechkin A V, Gonchar V Y, Klafter J and Metzler R 2005 *Europhys. Lett.* **72** 348–54
[31] Chechkin A, Sliusarenko O, Metzler R and Klafter J 2007 *Phys. Rev.* E **75** 041101
[32] Sokolov I M 2001 *Phys. Rev.* E **63** 011104
[33] Brockmann D and Sokolov I M 2002 *Chem. Phys.* **284** 409–21
[34] Ditlevsen P D 1999 *Phys. Rev.* E **60** 172–9
[35] Jespersen S, Metzler R and Fogedby H C 1999 *Phys. Rev.* E **59** 2736–45
[36] Dybiec B, Gudowska-Nowak E and Hänggi P 2007 *Phys. Rev.* E **75** 021109
[37] Olver F W J 1974 *Asymptotics and Special Functions (Computer Science and Applied Mathematics)* (New York: Academic)
[38] Schön J C 1997 *J. Phys. A: Math. Gen.* **30** 2367–89
[39] Locatelli M 2002 Handbook of global optimization II *Nonconvex Optimization and Its Applications* vol 62 ed P M Pardalos and H E Romeijn (Dordrecht: Kluwer) pp 179–229