

# Lévy flights, non-local search and simulated annealing

Ilya Pavlyukevich \*

*Department of Mathematics, Humboldt University of Berlin, Rudower Chaussee 25, 12489 Berlin, Germany*

Received 22 January 2007; received in revised form 11 June 2007; accepted 11 June 2007

Available online 16 June 2007

## Abstract

We solve a problem of non-convex stochastic optimisation with help of simulated annealing of Lévy flights of a variable stability index. The search of the ground state of an unknown potential is non-local due to big jumps of the Levy flights process. The convergence to the ground state is fast due to a polynomial decrease rate of the temperature.

© 2007 Elsevier Inc. All rights reserved.

*Keywords:* Lévy flights; Simulated annealing; Non-local search; Heavy-tails; Variable stability index; Stable-like process; Global optimisation

## 1. Introduction

Let  $U$  be a potential function in  $\mathbb{R}^d$  having several local minima and increasing fast at infinity. We look for a global minimum of  $U$ . Classical continuous-time simulated annealing (Boltzmann machine) (see [28,2,9,8,10]) consists in running a diffusion process

$$\begin{aligned}\widehat{Z}_{0,z}(t) &= z - \int_0^t \nabla U(\widehat{Z}_{0,z}(u)) du + \int_0^t \widehat{\sigma}(u) dW(u), \\ \widehat{\sigma}(t) &= \left( \frac{\theta}{\ln(\lambda + t)} \right)^{1/2},\end{aligned}\tag{1.1}$$

where  $W$  is a standard Brownian motion,  $\theta > 0$  denotes the cooling rate and  $\lambda > 1$  parametrises the initial temperature, which equals  $\sqrt{\theta/\ln(\lambda)}$  at time  $t = 0$ . It is known that there is a critical value  $\widehat{\theta}$  such that the diffusion  $\widehat{Z}(t)$  converges in distribution to the *global* minimum of  $U$  if  $\theta > \widehat{\theta}$  and the convergence fails otherwise. Moreover, the critical value  $\widehat{\theta}$  is the logarithmic rate of the principal non-zero eigenvalue  $\lambda_1(\sigma)$  of a time homogeneous diffusion generator  $A_\sigma f = \frac{\sigma^2}{2} \Delta f - \langle \nabla U, \nabla f \rangle$ , i.e.,

$$\widehat{\theta} = - \lim_{\sigma \rightarrow 0} \sigma^2 \ln |\lambda_1(\sigma)|.\tag{1.2}$$

\* Fax: +49 30 2093 5848.

E-mail address: [pavljuke@mathematik-hu-berlin.de](mailto:pavljuke@mathematik-hu-berlin.de)

The value of  $\hat{\theta}$  can be calculated explicitly, if one knows the heights of potential barriers between different wells of  $U$  (see [29,30,18] for precise results). Rigorous results on optimal cooling rate in simulated annealing algorithms can be found in [11,5,12,13].

In order to accelerate the search, Szu and Hartley in [26] suggested the so-called *fast simulated annealing* (Cauchy machine), which is a combination of a classical Metropolis algorithm introduced in [20] and a concept of non-local search due to the heavy-tail Cauchy visiting distribution. The authors claimed that in the Cauchy machine the temperature can be chosen decreasing as a power of time, namely  $\sigma(t) \sim t^{-1}$ , and applied the algorithm in image processing, see [27].

Motivated by [26], in our papers [22,21] we considered a continuous-time counterpart of the process  $\hat{Z}$  driven by Lévy flights of stability index  $\alpha \in (0,2)$  and temperature  $\sigma(t) \sim t^{-\theta}$ ,  $\theta > 0$ . We discovered that such a jump process never settles in the neighbourhood of a global minimum, but can be used to reveal a spatial structure of the potential  $U$ . The dynamics of Lévy flights with constant small noise was studied in our previous papers [14–16].

In the present paper we solve the problem of global optimisation with help of *state-dependent* Lévy flights in a multi-well potential  $U$ . We show, that in certain annealing regimes, the global minimum is localised always, as in the classical Gaussian case. For simplicity, we restrict our theoretical argument in Sections Section 2–5 to one-dimensional potentials. However, it will be clear from the presentation, that the algorithm also works in a multi-dimensional setting. In our numerical examples in Section 6, we compare our algorithm with related Gaussian and Cauchy optimisation techniques in case of a two-dimensional potential with five local minima, four-dimensional Shekel  $S_{10,4}$  function with ten local minima and six-dimensional Hartmann  $H_{4,6}$  function with four local minima.

## 2. Results on the cooled down Lévy flights

In [22,21] we considered a one-dimensional Lévy flights process in an external potential  $U$ . The process is determined by the stochastic differential equation

$$Z_{0,z}^{(\alpha)}(t) = z - \int_0^t U'(Z_{0,z}^{(\alpha)}(u-)) du + \int_0^t \frac{dL^{(\alpha)}(u)}{(\lambda + u)^\theta}. \tag{2.1}$$

We understand a Lévy flights process  $L^{(\alpha)}$  as a symmetric stable Lévy process with stability index  $\alpha \in (0,2)$ , whose marginal distributions have the Fourier transform

$$\mathbf{E}e^{i\omega L^{(\alpha)}(t)} = e^{-c(\alpha)t|\omega|^\alpha}, \quad c(\alpha) = 2 \left| \cos\left(\frac{\pi\alpha}{2}\right) \Gamma(-\alpha) \right|. \tag{2.2}$$

Although the Fourier transform has a very simple form, we shall use another representation due to the Lévy–Hinchin formula:

$$\mathbf{E}e^{i\omega L^{(\alpha)}(t)} = \exp\left(t \int_{\mathbb{R} \setminus \{0\}} [e^{i\omega y} - 1 - i\omega y \mathbb{1}_D(y)] \frac{dy}{|y|^{1+\alpha}}\right), \tag{2.3}$$

where  $\mathbb{1}_D(y)$  is the indicator function of the unit disk  $D = \{y : |y| \leq 1\}$ . Note that some authors prefer to use a different scaling of Lévy flights with  $c(\alpha) \equiv 1$  in the Fourier transform (2.2). We do not choose such a parametrisation in order to have a simple form of the Lévy–Hinchin formula, where the integral in the exponent is calculated w.r.t. the measure  $\nu(dy) = |y|^{-1-\alpha} dy$ . We refer the reader to Appendix C in [6] for a discussion on this subject.

The measure  $\nu$  is also called a jump measure of the Lévy process  $L^{(\alpha)}$ . It controls the intensity and sizes of its jumps. Indeed, let  $\Delta L^{(\alpha)}(t) = L^{(\alpha)}(t) - L^{(\alpha)}(t-)$  denote the jump size of  $L^{(\alpha)}$  at time instance  $t > 0$ . Then the number of jumps on the time interval  $(0, t]$  with values in a set  $J \subseteq \mathbb{R}$  is a Poissonian random variable with the mean  $t\nu(J)$  (which can be possibly zero or infinite).

The process  $Z^{(\alpha)}$  is a Markov process with a non-local generator

$$Af(x) = \int_{\mathbb{R} \setminus \{0\}} [f(x+y) - f(x) - yf'(x)\mathbb{1}_D(y)] \frac{dy}{|y|^{1+\alpha}}, \tag{2.4}$$

which is also referred to as a fractional Laplacian,  $A = -(-\Delta)^{\alpha/2}$ .

We direct reader’s attention to the books [17,24,1,23] on a rigorous mathematical theory of Lévy processes and stochastic differential equations. Physical results on the subject can be found in [19,4,3].

We assume that the potential  $U$  has  $n$  local minima  $m_i$  and  $n - 1$  local maxima  $s_i$  such that  $-\infty = s_0 < m_1 < s_1 < \dots < m_n < s_{n+1} = +\infty$ . The extrema are non-degenerate, i.e.,  $U''(m_i) > 0$  and  $U''(s_i) < 0$ . Moreover we demand that  $|U'(x)| > |x|^{1+c}$  as  $|x| \rightarrow +\infty$  for some positive  $c$ .

In the small temperature limit, i.e., when  $\lambda \rightarrow +\infty$  or  $t \rightarrow +\infty$ , the process  $Z^{(\alpha)}$  can be seen as a random perturbation of a deterministic dynamical system

$$X_x^0(t) = x - \int_0^t U'(X_x^0(s)) ds. \tag{2.5}$$

We denote  $\Omega_i = (s_{i-1}, s_i)$ ,  $1 \leq i \leq n$ , the domains of attraction of the stable points  $m_i$ . The positive parameter  $\theta$  is called the cooling rate and  $\lambda > 0$  determines the initial temperature of the system which equals to  $\lambda^{-\theta}$  at  $t = 0$ .

Eq. (2.1) describes a non-linear dynamics of an overdamped Lévy particle, whose temperature is being decreased at a polynomial rate as  $t \rightarrow \infty$ . In [22,21], we discovered two cooling regimes – slow cooling  $\theta < 1/\alpha$  and fast cooling  $\theta > 1/\alpha$  – in which the transitions of a particle between the wells of  $U$  have different asymptotic properties.

Let  $\Delta > 0$  be a small number and let  $B_i = \{y : |y - m_i| \leq \Delta\}$  denote a  $\Delta$ -neighbourhood of a local minimum  $m_i$ . Consider transition times

$$T_{s,z}^i = \inf \left\{ u \geq s : Z_{s,z}^{(\alpha)}(u) \in \cup_{j \neq i} B_j \right\} \tag{2.6}$$

between different neighbourhoods  $B_i$  and the corresponding transition probabilities  $\mathbf{P}_{s,z}(Z^{(\alpha)}(T^{i,\lambda}) \in B_j)$ ,  $i \neq j$ . Then for  $\theta < 1/\alpha$  and  $z \in B_i$  we have:

$$\frac{\mathbf{E}_{0,z} T^i}{\lambda^{\alpha\theta}} \rightarrow [q_i^{(\alpha)}]^{-1}, \quad \lambda \rightarrow +\infty, \tag{2.7}$$

$$\mathbf{P}_{0,z}(Z^{(\alpha)}(T^i) \in B_j) \rightarrow q_{ij}^{(\alpha)} [q_i^{(\alpha)}]^{-1}, \quad i \neq j, \tag{2.8}$$

where

$$q_{ij}^{(\alpha)} = \int_{\Omega_j} \frac{dy}{|m_i - y|^{1+\alpha}} = \frac{1}{\alpha} \left| \frac{1}{|s_{j-1} - m_i|^\alpha} - \frac{1}{|s_j - m_i|^\alpha} \right|, \quad i \neq j, \tag{2.9}$$

$$q_i^{(\alpha)} = \int_{\mathbb{R} \setminus \Omega_i} \frac{dy}{|m_i - y|^{1+\alpha}} = \sum_{j \neq i} q_{ij}^{(\alpha)} = \frac{1}{\alpha} \left( \frac{1}{|s_{i-1} - m_i|^\alpha} + \frac{1}{|s_i - m_i|^\alpha} \right).$$

We have also shown that in the limit  $t \rightarrow +\infty$ ,  $Z_{0,z}^{(\alpha)}(t)$  has a distribution:

$$\pi^{(\alpha)}(dy) = \sum_{i=1}^n \pi_i^{(\alpha)} \delta_{m_i}(dy) \tag{2.10}$$

where the vector  $\pi^{(\alpha)} = (\pi_1^{(\alpha)}, \dots, \pi_n^{(\alpha)})^T$ , solves the equation  $Q^T \pi^{(\alpha)} = 0$ ,  $Q = (q_{ij}^{(\alpha)})_{i,j=1}^n$ ,  $q_{ii}^{(\alpha)} = -q_i^{(\alpha)}$ . It is clearly seen, that all  $\pi_i^{(\alpha)} > 0$  and  $Z^{(\alpha)}$  does not settle down near the global minimum of  $U$ . However, the values  $\pi_i^{(\alpha)}$ , which can be estimated from the Monte Carlo simulations, reveal the spatial structure of  $U$ , e.g. the sizes of the domains  $\Omega_i$ .

If the cooling rate  $\theta$  is above the threshold  $1/\alpha$ , the Lévy particle  $Z^{(\alpha)}$  gets trapped in one of the wells and thus the convergence fails. Consider the first exit time from the  $i$ th well

$$S_{s,z}^i = \inf \left\{ u \geq s : Z_{s,z}^{(\alpha)}(u) \in \Omega_i \right\}. \tag{2.11}$$

Then, for  $z \in B_i, 1 \leq i \leq n$ ,

$$\mathbf{P}_{0,z}(\mathcal{S}^i < \infty) = \mathcal{O}\left(\frac{1}{\lambda^{2\theta-1}}\right), \quad \lambda \rightarrow \infty, \tag{2.12}$$

and consequently,  $\mathbf{E}_{0,z}\mathcal{S}^i = \infty$ .

For practical simulations, solutions of (2.1) can be approximated by Euler approximations

$$z_{kh} = z_{(k-1)r} - U'(z_{(k-1)r})r + \frac{I_k^r}{(\lambda + (k-1)r)^\theta}, \quad z_0 = z, \quad k \geq 1, \tag{2.13}$$

with a time step  $r > 0$  and a random input  $(I_k^r)_{k \geq 1}$ . The random variables are independent identically distributed stable symmetric random variables, what follows from the independence and stationarity of increments of  $L^{(\alpha)}$  and its scaling property:

$$I_k^r = L^{(\alpha)}(kr) - L^{(\alpha)}((k-1)r) \stackrel{d}{=} L^{(\alpha)}(r) \stackrel{d}{=} r^{1/\alpha}L^{(\alpha)}(1). \tag{2.14}$$

We refer the reader to the book [17] for more information on simulation and approximation of jump-diffusions driven by Lévy flights.

### 3. Lévy flights with variable stability index (stable-like processes)

In order to take into account the energy geometry of the potential, we have to make the Lévy flights process depend on its current position. Thus instead of Lévy flights  $L^{(\alpha)}$  defined in (2.2), we consider now the so-called *stable-like process*  $H = (H(t))_{t \geq 0}$ , which is a Markov process defined by the non-local generator

$$Bf(x) = \int_{\mathbb{R} \setminus \{0\}} [f(x+y) - f(x) - yf'(x)\mathbb{1}_D(y)] \frac{dy}{|y|^{1+\alpha(x)}}, \tag{3.1}$$

with a function  $\alpha(x)$  taking values in the interval  $(0, 2)$ . Sometimes, the notation  $B = -(-\Delta)^{\alpha(\cdot)/2}$  is used. The crucial difference between  $L^{(\alpha)}$  and  $H$  consists in a dependence of a stable-like jump measure  $\nu_x(dy) = |y|^{-1-\alpha(x)}dy$  on the spatial coordinate  $x$ . Thus, if  $H(t_0) = x_0$ , the instant jump distribution of  $H$  at time  $t_0$  is governed by a stable measure  $\nu_{x_0}(dy)$ .

Of course, the dynamics of  $H$  is completely determined by a variable stability index  $\alpha(x)$ . For example, if  $\alpha(x) = \alpha_0 \in (0, 2)$ , then  $H$  is just a usual Lévy flights process of index  $\alpha_0$ . From now on, we assume that  $\alpha(x)$  takes values strictly between 0 and 2, i.e.,  $0 < a \leq \alpha(x) \leq A < 2$ , to exclude degeneration of the jump measure.

We are going to study the dynamics of a stochastic differential equation with the driving process  $H$ , namely

$$Y_{0,y}(t) = y - \int_0^t U'(Y_{0,y}(u-))du + \int_0^t \frac{dH(Y_{0,y}(u-), u)}{(\lambda + u)^\theta}. \tag{3.2}$$

For a better understanding of the process  $Y$ , it is instructive to consider a discrete time analogue of (3.2) given by the recurrent formula:

$$y_{kr} = y_{(k-1)r} - U'(y_{(k-1)r})r + \frac{h_k^r(y_{(k-1)r})}{(\lambda + (k-1)r)^\theta}, \quad k \geq 1. \tag{3.3}$$

The discrete time dynamical system (3.3) is obtained from the Euler approximation of (3.2) with the time step  $r$  and can be used for simulations. (However, one should be careful when  $U'$  is not globally Lipschitz.) The random input is determined by the random variables  $h_k^r(y)$  such that

$$h_k^r(y) \stackrel{d}{=} L^{(\alpha(y))}(r) \stackrel{d}{=} r^{1/\alpha(y)}L^{(\alpha(y))}(1), \tag{3.4}$$

where  $L^{(\alpha(y))}(1)$  has a standard symmetric  $\alpha(y)$ -stable distribution with the Fourier transform (2.2).

#### 4. One-well dynamics: transitions and trapping

The dynamics  $Y$  of the Lévy flights with variable stability index in the force field  $U'$  is a result of an interplay of two independent effects. First, for small temperatures, i.e., when  $\lambda \rightarrow +\infty$  or  $t \rightarrow +\infty$ ,  $Y$  is close to the underlying deterministic trajectory  $X^0$ . Starting from any point of  $\Omega_i$ , it follows  $X^0$  with the same initial point and with high probability reaches a small neighbourhood of  $m_i$  in relatively short time. On the other hand,  $Y$  tries to deviate from  $X^0$  making jumps controlled by the jump measure  $\nu_x(dy)$ . Finally, if  $Y$  is in the well  $\Omega_i$ , it spends most of the time in a neighbourhood of  $m_i$  and thus has jumps approximately governed by the *stable jump measure*  $\nu_{m_i}(dy) = |y|^{-1-\alpha(m_i)} dy$ .

Thus, the exit time and the exit probability from the well  $\Omega_i$  of the process  $Y$  are approximately the same as for the process  $Z^{(\alpha(m_i))}$ . This approximation becomes exact if we consider a piece-wise constant stability index  $\alpha(x)$ ,  $\alpha(x) = \sum_{i=1}^n \alpha_i \mathbb{1}\{x \in \Omega_i\}$ ,  $0 < \alpha_i < 2$ . With this choice of  $\alpha(x)$ , the process  $Y$  is just driven by the Eq. (2.1) until it exits the well. (We skip a discussion on the behaviour of the process in the small neighbourhoods of the saddle points.)

Let us introduce the following transition and exit times for the process  $Y$ :

$$\tau_{s,y}^i = \inf\{u \geq s : Y_{s,y}(u) \in \cup_{j \neq i} B_j\}, \tag{4.1}$$

$$\sigma_{s,y}^i = \inf\{u \geq s : Y_{s,y}(u) \notin \Omega_i\}. \tag{4.2}$$

It follows from (2.7) and (2.8), that if  $y \in B_i$  and  $\alpha(m_i)\theta < 1$ , then the following relations hold as  $\lambda \rightarrow +\infty$ :

$$\frac{\mathbf{E}_{0,y} \tau^i}{\lambda^{\alpha(m_i)\theta}} \rightarrow [q_i^{(\alpha(m_i))}]^{-1}, \tag{4.3}$$

$$\mathbf{P}_{0,y}(Y(\tau^i) \in B_j) \rightarrow q_{ij}^{(\alpha(m_i))} [q_i^{(\alpha(m_i))}]^{-1}, \quad i \neq j. \tag{4.4}$$

On the other hand, if  $\alpha(m_i)\theta > 1$ , the Lévy particle gets trapped in the well due to (2.12), i.e.,

$$\mathbf{P}_{0,y}(\sigma^i < \infty) = \mathcal{O}\left(\frac{1}{\lambda^{\alpha(m_i)\theta-1}}\right), \quad \lambda \rightarrow +\infty, \tag{4.5}$$

and consequently,  $\mathbf{E}_{0,y} \sigma^{i,\lambda} = \infty$ .

Relations 4.3, 4.4 and 4.5 are crucial for our analysis.

#### 5. Non-local random search and simulated annealing

##### 5.1. Getting trapped in an assigned well

We demonstrate now how to drive a Lévy particle  $Y$  to an assigned well  $\Omega_i$ , if the approximate location of its minimum  $m_i$  is known.

Indeed, the function  $\alpha(x)$  given, the limiting dynamics of  $Y$  is determined by the values  $\alpha(m_i)$ ,  $1 \leq i \leq n$  and the cooling rate  $\theta$ . Moreover, for our analysis we can freely choose both  $\alpha(x)$  and  $\theta$ .

Let  $\alpha(x)$  be smooth and attain its unique global maximum at  $m_i$ . Then we have:

$$\alpha(m_i) > \max_{j \neq i} \alpha(m_j). \tag{5.1}$$

For instance, one can take  $\alpha(x) = a + (A - a)/(1 + (x - m_i)^2)$  for some  $0 < a < A < 2$ . Then we can choose  $\theta > 0$ , such that

$$\alpha(m_i)\theta > 1, \quad \text{whereas } \alpha(m_j)\theta < 1 \quad \text{for } j \neq i. \tag{5.2}$$

With this choice of parameters, as  $t \rightarrow +\infty$ , the particle leaves any well  $\Omega_j$ ,  $j \neq i$ , in finite time according to (4.3). Moreover, since all transition probabilities in (4.4) are strictly positive, the probability to enter the well  $\Omega_i$  after a finite number of transitions between the wells  $\Omega_j$ ,  $j \neq i$ , equals 1. Finally, upon entering  $\Omega_i$ , the particle gets trapped there due to (4.5).

### 5.2. Looking for the global minimum

Let  $M$  be the (unique) unknown global minimum of the potential  $U$ . To make a Lévy particle settle near  $M$ , we have to determine the appropriate variable stability index  $\alpha(x)$  and the cooling rate  $\theta$  such that

$$\alpha(M)\theta > 1, \quad \text{whereas } \alpha(m_i)\theta < 1, m_i \neq M. \tag{5.3}$$

Let  $\phi(u)$  be an arbitrary smooth monotone decreasing function on  $\mathbb{R}$ ,  $0 < a < \phi(u) < A < 2$ . Then we set

$$\alpha(y) = \phi(U(y)). \tag{5.4}$$

It is clear that

$$\alpha(M) > \alpha(m_i) \quad \text{for all } m_i \neq M, \tag{5.5}$$

and we choose  $\theta$  to satisfy relations (5.3). The trapping of the particle near the global minimum  $M$  follows from the argument of the preceding section.

### 5.3. The local minimum with maximal energy

Analogously, we can determine the coordinate of the (unique) ‘highest’ local minimum  $m$ , i.e., such that

$$U(m) = \max_{1 \leq i \leq n} U(m_i). \tag{5.6}$$

In this case, we should take  $\alpha(y) = \psi(U(y))$  with an arbitrary smooth monotone increasing function  $\psi$ ,  $0 < a < \phi(y) < A < 2$ , which leads to the inequality

$$\alpha(m) > \alpha(m_i) \quad \text{for all } m_i \neq m, \tag{5.7}$$

and the arguments of the previous sections justify the success of the search.

### 5.4. A local minimum with certain energy

Finally, we can perform a search not for the global minimum of  $U$  but for a local minimum  $m_E$  satisfying the condition  $U(m_E) \leq E$  for some energy level  $E$ . In such a case we take a piece-wise constant stability index

$$\alpha(y) = \begin{cases} A, & U(y) \leq E, \\ a, & U(y) \geq E, \quad 0 < a < A < 2, \end{cases} \tag{5.8}$$

and a cooling rate  $\theta$  satisfying conditions  $A\theta > 1$  and  $a\theta < 1$ . Consequently, if for some  $i$ , the minimum  $U(m_i)$  lies below the threshold  $E$ , then near this minimum  $Y$  behaves like a jump-diffusion  $Z^{(A)}$  driven by a Lévy flights process  $L^{(A)}$  and thus the Lévy particle gets trapped due to (4.5). If the well minimum lies above the level  $E$ , a Lévy particle behaves like a process  $Z^{(a)}$  and leaves this well in finite time due to relations (4.3) and (4.4).

If there are several wells with minima below  $E$ , the Lévy particle settles down near one of them.

Running the search for decreasing energies  $E_1 > E_2 > \dots$ , we can also estimate the ground energy level  $E^* = U(M)$  and to determine the global minimum  $M$ . Analogously, one can determine a local minimum  $m_E$  satisfying conditions  $U(m_E) \geq E$  or  $E_1 \leq U(m_E) \leq E_2$ .

We emphasise that the search algorithm described in this section requires no particular information about the potential  $U$ : to determine the local minimum  $m_E$  we use three numbers  $a$ ,  $A$  and  $\theta$ , the energy level  $E$  and values of  $U(x)$ .

## 6. Numerical examples

### 6.1. Simulation of random increments

In the preceding sections we gave a theoretic justification of the search algorithm in dimension one. It is clear that a  $d$ -dimensional case,  $d > 1$ , does not differ much. It suffices to consider isotropic (spherically

symmetric)  $d$ -dimensional Lévy flights with a jumping measure  $\nu_x(d\mathbf{y}) = \|\mathbf{y}\|^{-d-\alpha(x)}d\mathbf{y}$  and to calculate new values of transition probabilities writing  $d$  instead of 1 in the integrals in (2.9). (Here  $\|\cdot\|^2 = \langle \cdot, \cdot \rangle$  denotes the usual Euclidean norm in  $\mathbb{R}^d$ .)

In this section we illustrate our optimisation method by three numerical examples. However, first it would be helpful to give some details on computer simulation of stochastic differential equations driven by non-Gaussian stable noises.

For our examples we consider the Euler discretisation of a  $d$ -dimensional stochastic differential equation

$$\mathbf{Y}_{0,\mathbf{y}}(t) = \mathbf{y} - \int_0^t \nabla U(\mathbf{Y}_{0,\mathbf{y}}(u-))du + \int_0^t \frac{d\mathbf{H}(\mathbf{Y}_{0,\mathbf{y}}(u-), u)}{(\lambda + u)^\theta}, \quad t \geq 0, \tag{6.1}$$

i.e., we consider a  $d$ -dimensional Markov chain:

$$\mathbf{y}_{kr} = \mathbf{y}_{(k-1)r} - \nabla U(\mathbf{y}_{(k-1)r})r + \frac{\mathbf{h}_k^r(\mathbf{y}_{(k-1)r})}{(\lambda + (k-1)r)^\theta}, \quad k \geq 1, \tag{6.2}$$

with the initial value  $\mathbf{y}_0 = \mathbf{y}$  and a (small) time step  $r > 0$ .

As we have seen in Sections 2 and 3, the random vectors  $\mathbf{h}_k^r(\mathbf{y})$  have isotropic  $\alpha$ -stable distribution, namely  $\mathbf{h}_k^r(\mathbf{y}) \stackrel{d}{=} r^{1/\alpha(\mathbf{y})} \mathbf{L}^{\alpha(\mathbf{y})}(1)$ , where  $(\mathbf{L}^{\alpha}(t))_{t \geq 0}$  is a  $d$ -dimensional Lévy flights process with a Fourier transform

$$\begin{aligned} \mathbf{E} \exp(i\langle \boldsymbol{\omega}, \mathbf{L}^{\alpha}(t) \rangle) &= e^{-tc(\alpha,d)\|\boldsymbol{\omega}\|^\alpha} = \exp\left(t \int_{\mathbb{R}^d \setminus \{0\}} [e^{i\langle \boldsymbol{\omega}, \mathbf{y} \rangle} - 1 - i\langle \boldsymbol{\omega}, \mathbf{y} \rangle \mathbb{1}_D(\mathbf{y})] \frac{d\mathbf{y}}{\|\mathbf{y}\|^{d+\alpha}}\right), \quad \boldsymbol{\omega} \in \mathbb{R}^d, \\ c(\alpha, d) &= \frac{\pi^{d/2}}{2^\alpha} \frac{|\Gamma(-\frac{\alpha}{2})|}{\Gamma(\frac{d+\alpha}{2})}. \end{aligned} \tag{6.3}$$

Note that  $c(\alpha, 1) = c(\alpha)$ ,  $c(\alpha)$  being defined in (2.2).

To simulate samples of random vectors  $\mathbf{h}_k^r$  we exploit a well-known fact that an isotropic Lévy flights process can be obtained by a random time change (subordination) of a standard Brownian motion (see Examples 24.12 and 30.6 in [24] or Algorithm 6.10 in [7]).

More precisely, for  $\alpha \in (0,2)$  let  $S^{(\alpha/2)} = (S^{(\alpha/2)}(t))_{t \geq 0}$  be an  $\alpha/2$ -stable Lévy subordinator, i.e., a one-dimensional positive increasing Lévy process with the Fourier transform:

$$\mathbf{E} \exp(i\xi S^{(\alpha/2)}(t)) = \exp\left(-t|\xi|^{\alpha/2} \left(1 - \text{isgn}(\xi) \tan\left(\frac{\pi\alpha}{4}\right)\right)\right), \quad \xi \in \mathbb{R}. \tag{6.4}$$

Since  $S^{(\alpha/2)}(t) \geq 0$  for all  $t \geq 0$  it is more convenient to consider its Laplace transform:

$$\mathbf{E} \exp(-uS^{(\alpha/2)}(t)) = \exp\left(-t \frac{u^{\alpha/2}}{\cos \frac{\pi\alpha}{4}}\right), \quad u \geq 0. \tag{6.5}$$

Let now  $\mathbf{W}$  be a standard  $d$ -dimensional Brownian motion, independent of  $S^{(\alpha/2)}$ , with a characteristic function:

$$\mathbf{E} \exp(i\langle \boldsymbol{\omega}, \mathbf{W}(t) \rangle) = \exp\left(-t \frac{\|\boldsymbol{\omega}\|^2}{2}\right), \quad \boldsymbol{\omega} \in \mathbb{R}^d. \tag{6.6}$$

Let us consider the Brownian motion  $\mathbf{W}$  on a random time scale determined by the increasing process  $S^{(\alpha/2)}$ , i.e., consider a superposition  $(\mathbf{W}(S^{(\alpha/2)}(t)))_{t \geq 0}$ . Then it is known (see, e.g. Theorem 30.1 in [24]) that this process is a  $d$ -dimensional Lévy process and for any  $\mu > 0$  its a characteristic function is given by

$$\mathbf{E} \exp(i\langle \boldsymbol{\omega}, \mu \mathbf{W}(S^{\alpha/2}(t)) \rangle) = \exp\left(-t \frac{1}{\cos \frac{\pi\alpha}{4}} \left(\frac{\|\mu \boldsymbol{\omega}\|^2}{2}\right)^{\alpha/2}\right) = \exp\left(-t \frac{\mu^\alpha}{2^{\alpha/2} \cos \frac{\pi\alpha}{4}} \|\boldsymbol{\omega}\|^\alpha\right), \quad \boldsymbol{\omega} \in \mathbb{R}^d, \tag{6.7}$$



i.e., it is a  $d$ -dimensional Lévy flights process, maybe with a scale parameter different from those of  $\mathbf{L}^{(\alpha)}$  in (6.3). Recalling that  $S^{(\alpha/2)}(t) \stackrel{d}{=} t^{2/\alpha} S^{(\alpha/2)}(1)$  and  $\mathbf{W}(t) \stackrel{d}{=} \sqrt{t} \mathbf{W}(1)$  and comparing the constants in the final exponents in (6.3) and (6.7) we can determine  $\mu = \mu(\alpha, d)$ , such that for any  $t \geq 0$  the following equality holds:

$$\begin{aligned} \mathbf{L}^{(\alpha)}(t) &\stackrel{d}{=} \mu(\alpha, d) \mathbf{W}(S^{(\alpha/2)}(t)) \stackrel{d}{=} \mu(\alpha, d) \sqrt{S^{(\alpha/2)}(t)} \mathbf{W}(1) \stackrel{d}{=} \mu(\alpha, d) t^{1/\alpha} \sqrt{S^{(\alpha/2)}(1)} \mathbf{W}(1), \\ \mu(\alpha, d) &= \frac{1}{\sqrt{2}} \left[ \pi^{d/2} \cos \frac{\pi\alpha}{4} \cdot \frac{|\Gamma(-\frac{\alpha}{2})|}{\Gamma(\frac{d+\alpha}{2})} \right]^{1/\alpha}. \end{aligned} \tag{6.8}$$

Consequently, the vectors  $\mathbf{h}_k^r(\mathbf{y})$  in (6.2) have the following distribution:

$$\mathbf{h}_k^r(\mathbf{y}) \stackrel{d}{=} r^{1/\alpha(\mathbf{y})} \mu(\alpha(\mathbf{y}), d) \sqrt{S^{(\alpha/2)}(1)} \mathbf{W}(1). \tag{6.9}$$

Thus from the computational point of view, to simulate Lévy flights increments we need a standard Gaussian random vector and a standard positive  $\alpha/2$ -stable random variable. The latter can be obtained as a function of two independent random variables  $V$  and  $E$ , where  $V$  is uniformly distributed on  $(-\frac{\pi}{2}, \frac{\pi}{2})$ , and  $E$  has a standard exponential distribution,  $\mathbf{P}(E > t) = e^{-t}$  (e.g. see, [17, §3.5]):

$$S^{(\alpha/2)}(1) \stackrel{d}{=} \frac{\sin(\frac{\alpha}{2}V + \frac{\pi\alpha}{4}) \cdot \cos((1 - \frac{\alpha}{2})V - \frac{\pi\alpha}{4})}{(\cos \frac{\pi\alpha}{4} \cdot \cos V)^{2/\alpha} \cdot E^{2/\alpha-1}}. \tag{6.10}$$

In our simulations we used a function `gsl_ran_levy_skew` from the *GNU Scientific Library (GSL)*.

### 6.2. Optimisation methods

In what follows we illustrate our method by three numerical examples. We consider a five-well potential in  $\mathbb{R}^2$ , a ten-well potential in  $\mathbb{R}^4$  (Shekel function  $S_{10,4}$ ) and a four-well potential in  $\mathbb{R}^6$  (Hartman function  $H_{4,6}$ ).

For each function we apply four different optimisation methods: our stable-like simulated annealing, the fast simulated annealing in sense of Szu and Hartley [26] with Barker’s acceptance probability, simulated annealing obtained by solving a Gaussian SDE in sense of [2] and a Gaussian simulated annealing with Metropolis’ acceptance probability. For the sake of exactness we give their complete description.

*Our stable-like simulated annealing.* We simulate a Markov chain:

$$\begin{aligned} \mathbf{y}_{kr} &= \mathbf{y}_{(k-1)r} - \nabla U(\mathbf{y}_{(k-1)r})r + \sigma((k-1)r) \mathbf{h}_k^r(\mathbf{y}_{(k-1)r}), \quad k \geq 1, \\ \sigma(t) &= \frac{1}{(\lambda + t)^\theta}, \end{aligned} \tag{6.11}$$

with random inputs  $\mathbf{h}_k^r$  generated from formula (6.9). According to the Section 5.4, the stability index takes two values

$$\alpha(\mathbf{y}) = \begin{cases} A, & \text{if } U(\mathbf{y}) < E_s, \\ a, & \text{if } U(\mathbf{y}) \geq E_s, \end{cases} \tag{6.12}$$

with some energy level  $E_s$  (here ‘s’ stands for success) which will be specified later and  $0 < a \leq A < 2$ . The set  $\{\mathbf{x} : U(\mathbf{x}) \leq E_s\}$  will be the *trapping domain* for the Markov chain  $\mathbf{y}_{kr}$ .

*Fast simulated annealing.* The states are obtained from the relation:

$$\begin{aligned} \mathbf{y}_{kr} &= \mathbf{y}_{(k-1)r} - \nabla U(\mathbf{y}_{(k-1)r})r + \sigma((k-1)r) \mathbf{h}_k^r, \quad k \geq 1, \\ \sigma(t) &= \frac{\lambda}{1+t}, \end{aligned} \tag{6.13}$$

where random inputs  $\mathbf{h}_k^r$  are Cauchy-distributed according to (6.9) with  $\alpha(\cdot) \equiv 1$ . The state  $\mathbf{y}_{kr}$  is accepted with the Barker acceptance probability:



$$P_B = \left[ 1 + \exp \left( \frac{U(\mathbf{y}_{kr}) - U(\mathbf{y}_{(k-1)r})}{\sigma((k-1)r)} \right) \right]^{-1}. \tag{6.14}$$

Gaussian stochastic differential equation. We simulate a Markov chain with Gaussian increments:

$$\begin{aligned} \mathbf{y}_{kr} &= \mathbf{y}_{(k-1)r} - \nabla U(\mathbf{y}_{(k-1)r})r + \sigma((k-1)r)\mathbf{g}_k^r, \quad k \geq 1, \\ \sigma(t) &= \sqrt{\frac{\theta}{\ln(\lambda + t)}}, \end{aligned} \tag{6.15}$$

where  $\mathbf{g}_k^r$  are identically distributed independent Gaussian vectors,  $\mathbf{g}_k^r \stackrel{d}{=} \sqrt{r}\mathbf{W}(1)$ .

Gaussian simulated annealing. The states are obtained from the relation (6.15). The state  $\mathbf{y}_{kr}$  is accepted with the Metropolis acceptance probability

$$P_M = 1 \wedge \exp \left( -\frac{U(\mathbf{y}_{kr}) - U(\mathbf{y}_{(k-1)r})}{\sigma((k-1)r)} \right). \tag{6.16}$$

Let  $U^*$  denote the value of  $U$  in its global minimum and  $u^*, u^* > U^*$ , its value in the second-deep local minimum. Consider the energy level:

$$E_s = \frac{U^* + u^*}{2}. \tag{6.17}$$

We say that the search on step  $k$  is *successful* if  $U(\mathbf{y}_{kr}) \leq E_s$ , i.e., the observed energy value is well separated from the second best solution.

To determine the speed and the precision of the algorithms, we choose another small energy value  $E_p$  ('p' stands for *precision*) and consider the step number

$$N_{\text{first}} = 2 \cdot 10^6 \wedge \min\{k \geq 0 : U(\mathbf{y}_{kr}) \leq U^* + E_p\}, \tag{6.18}$$

i.e., the first step when the minimal energy value is determined within the accuracy  $E_p$ . We bound the maximal value of  $N_{\text{first}}$  by  $2 \cdot 10^6$  to exclude very long simulation times.

To test the speed and stability of searches, we perform simulations of the Markov chains  $\mathbf{y}_{kr}$ , calculate the mean time  $\langle N_{\text{first}} \rangle$ , the number of successful searches  $N_k$  on the steps  $k = 5 \cdot 10^4$  and  $k = 5 \cdot 10^5$  and the mean deviation  $\Delta_k = U^* - \langle U(\mathbf{y}_{kr}) \rangle$  where the average is taken over the successful searches. We make simulations for different values of  $\lambda$  and compare the best results.

### 6.3. Five-well potential in $\mathbb{R}^2$

To discuss the geometrical features of the searches, we start with a simple a two-dimensional potential function  $U$  given by the formula

$$\begin{aligned} U(\mathbf{y}) &= \left[ 1 - \frac{1}{1 + 0.05(y_1^2 + (y_2 - 10)^2)} - \frac{1}{1 + 0.05((y_1 - 10)^2 + y_2^2)} - \frac{1.5}{1 + 0.03((y_1 + 10)^2 + y_2^2)} \right. \\ &\quad \left. - \frac{2}{1 + 0.05((y_1 - 5)^2 + (y_2 + 10)^2)} - \frac{1}{1 + 0.1((y_1 + 5)^2 + (y_2 + 10)^2)} \right] (1 + 0.0001(y_1^2 + y_2^2)^{1.2}). \end{aligned} \tag{6.19}$$

The function  $U$  has five local minima. Its global minimum  $\mathbf{m}^* = (4.92, -9.89)$  has the energy value  $U^* = -1.4616$ , and the second-deep minimum  $\mathbf{m}_1 = (-9.73, -0.11)$  has the energy  $u^* = -0.8532$ . We set  $E_s = -1.1574$  and  $E_p = 0.0616$ . The step size is  $r = 0.1$ . The initial values  $\mathbf{y}_0$  are uniformly distributed in the square  $[-20, 20]^2$ . The results of simulations are presented in the tables below. On Fig. 1 we show typical random paths  $(\mathbf{y}_{rk})$  on the plain for all methods.

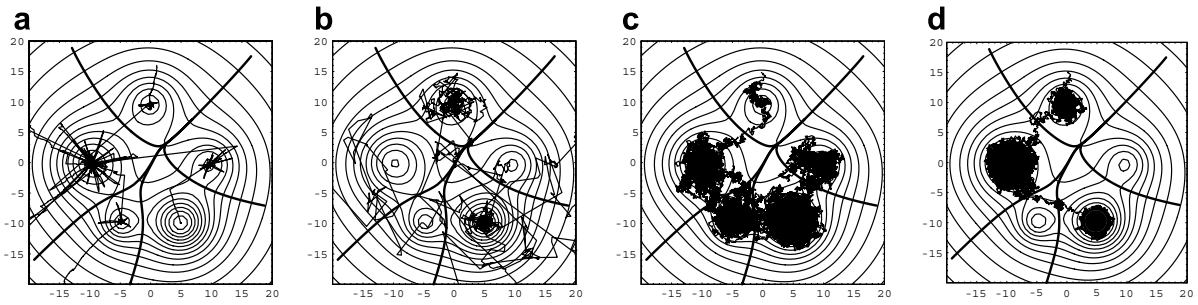


Fig. 1. Typical random search paths  $(\mathbf{y}_k)_{0 \leq k \leq n}$  for our stable-like simulated annealing (a), FSA (b), Gaussian SDE (c) and Gaussian Metropolis SA (d). Thick lines denote the boundaries of the attraction domains of the dynamical system  $\dot{x}_t = -\nabla U(x_t)$ .

$\lambda$	$\langle N_{\text{first}} \rangle$	$k = 5 \cdot 10^4$		$k = 5 \cdot 10^5$	
		$N_k$	$\Delta_k$	$N_k$	$\Delta_k$
<i>Stable-like simulated annealing</i>					
$a = 1.1, A = 1.8, \theta = 0.75$					
$1 \cdot 10^2$	175458	61	0.0004	90	0.0004
$5 \cdot 10^2$	93273	75	0.0004	96	0.0004
$1 \cdot 10^3$	135081	62	0.0004	93	0.0004
$5 \cdot 10^3$	148972	60	0.0004	93	0.0004
$1 \cdot 10^4$	264070	47	0.0004	85	0.0004
<i>Fast simulated annealing</i>					
$10^1$	586029	67	0.0005	69	0.0004
$5 \cdot 10^1$	5787	97	0.0019	98	0.0006
$1 \cdot 10^2$	1887	100	0.0035	100	0.0008
$5 \cdot 10^2$	4477	99	0.0294	100	0.0038
$1 \cdot 10^3$	7552	70	0.0600	100	0.0112
<i>Gaussian SDE</i>					
$\theta = 3$					
$10^3$	8749	62	0.1211	76	0.1086
$10^4$	12768	71	0.1170	79	0.1109
$10^5$	34961	61	0.1204	81	0.1154
$10^6$	37262	66	0.0888	86	0.0923
<i>Gaussian simulated annealing</i>					
$\theta = 3$					
$10^3$	81943	67	0.0786	93	0.0804
$10^4$	113516	62	0.0715	94	0.0756
$10^5$	301035	44	0.0685	79	0.0643
$10^6$	573990	50	0.0662	63	0.0633

As we see, the fast simulated annealing shows the best results in the present example. The search is very fast and samples taken on steps  $k = 5 \cdot 10^4$  and  $k = 5 \cdot 10^5$  with high precision estimate the minimal energy. However, it is seen from Fig. 1b that the method uses the geometry of the potential not very efficiently and visits many states away from the local minima of  $U$ .

Our algorithm estimates the ground state stably but takes essentially longer time for its first detection. Its search paths (Fig. 1a) efficiently use the geometry of  $U$ , i.e., coming to the domain of attraction of a local

minimum, they visit its neighbourhood and either leave it in finite time, or get trapped if the local minimum has energy below  $E_s$ .

Both Gaussian methods are characterised by relatively poor precision due to the big variance of random increments.

6.4. Four-dimensional Shekel's function

$$S_{10,4}(\mathbf{y}) = - \sum_{i=1}^{10} \frac{1}{c_i + \|\mathbf{y} - \mathbf{a}_i\|^2},$$

$$\begin{aligned} \mathbf{a}_1 &= (4.0, 4.0, 4.0, 4.0), & \mathbf{a}_6 &= (2.0, 9.0, 2.0, 9.0), \\ \mathbf{a}_2 &= (1.0, 1.0, 1.0, 1.0), & \mathbf{a}_7 &= (5.0, 5.0, 3.0, 3.0), \\ \mathbf{a}_3 &= (8.0, 8.0, 8.0, 8.0), & \mathbf{a}_8 &= (8.0, 1.0, 8.0, 1.0), \\ \mathbf{a}_4 &= (6.0, 6.0, 6.0, 6.0), & \mathbf{a}_9 &= (6.0, 2.0, 6.0, 2.0), \\ \mathbf{a}_5 &= (3.0, 7.0, 3.0, 7.0), & \mathbf{a}_{10} &= (7.0, 3.6, 7.0, 3.6), \\ (c_i)_{i=1}^{10} &= (0.1, 0.2, 0.2, 0.4, 0.4, 0.6, 0.3, 0.7, 0.5, 0.5). \end{aligned} \tag{6.20}$$

The function  $S_{10,4}$  has ten local minima. Its global minimum has the energy value  $U^* = -10.5364$  and the second-deep minimum has the energy  $u^* = -5.1285$ . We set  $E_s = -7.8560$  and  $E_p = 0.5364$ . The step size is  $r = 0.01$ . The initial values  $\mathbf{y}_0$  are uniformly distributed in the cube  $[0, 10]^4$ . Since the gradient of  $S_{10,4}$  vanishes for big values of  $\|\mathbf{y}\|$ , in our simulations we consider the penalised function:

$$S_{10,4}^p(\mathbf{y}) = S_{10,4}(\mathbf{y}) + \sum_{j=1}^4 [(y_j - 10)^{2.2} \mathbb{1}(y_j > 10) + |y_j|^{2.2} \mathbb{1}(y_j < 0)]. \tag{6.21}$$

Simulation results are presented below:

$\lambda$	$\langle N_{\text{first}} \rangle$	$k = 5 \cdot 10^4$		$k = 5 \cdot 10^5$	
		$N_k$	$\Delta_k$	$N_k$	$\Delta_k$
<i>Stable-like simulated annealing</i>					
$a = 1.2, A = 1.9, \theta = 0.6$					
100	50 198	39	0.3598	92	0.0887
500	46 768	52	0.2456	92	0.0866
1000	43 951	63	0.2072	95	0.0810
5000	63 960	56	0.0848	95	0.0595
10000	77 239	53	0.0572	97	0.0515
<i>Fast simulated annealing</i>					
50	612 902	67	0.1385	70	0.0179
100	354 639	81	0.2499	82	0.0324
500	57 214	7	0.8834	99	0.1470
1000	99 846	0	–	99	0.2644
1500	139 922	0	–	100	0.4101
<i>Gaussian SDE</i>					
$\theta = 10$					
$10^3$	37 089	1	1.5448	7	1.7751
$10^4$	26 749	10	1.5667	18	1.6223
$10^5$	24 657	52	1.5839	45	1.6339
$10^6$	26 611	59	1.5631	67	1.3583
$10^7$	53 874	54	1.4889	69	1.5545

*Gaussian simulated annealing*

$\theta = 10$

$10^3$	16889	69	1.4155	79	1.2417
$10^4$	18722	78	1.2062	90	1.3376
$10^5$	59266	68	1.0551	89	1.1888
$10^6$	183003	74	0.9401	84	1.0355
$10^7$	536875	67	1.1161	72	1.0520

For this test function, our algorithm detects the global minimum about 20% faster than other methods and also demonstrates better precision.

6.5. Six-dimensional Hartman function

In this section we apply our algorithm to the six-dimensional Hartman function:

$$H_{4,6}(\mathbf{y}) = - \sum_{i=1}^4 c_i \exp \left( - \sum_{j=1}^6 a_{ij} (y_j - p_{ij})^2 \right), \tag{6.22}$$

with

$$(p_{ij})_{i,j=1}^{4,6} = \begin{pmatrix} 0.1312 & 0.1696 & 0.5569 & 0.0124 & 0.8283 & 0.5886 \\ 0.2329 & 0.4135 & 0.8307 & 0.3736 & 0.1004 & 0.9991 \\ 0.2348 & 0.1451 & 0.3522 & 0.2883 & 0.3047 & 0.6650 \\ 0.4047 & 0.8828 & 0.8732 & 0.5743 & 0.1091 & 0.0381 \end{pmatrix}, \tag{6.23}$$

$$(a_{ij})_{i,j=1}^{4,6} = \begin{pmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{pmatrix}, \quad (c_i)_{i=1}^4 = \begin{pmatrix} 1 \\ 1.2 \\ 3 \\ 3.2 \end{pmatrix}.$$

The function  $H_{4,6}$  has four local minima. Its global minimum has the energy value  $U^* = -3.3224$  and the second-deep minimum has the energy  $u^* = -3.2032$ . We set  $E_s = -3.2628$  and  $E_p = 0.0324$ . The step size is  $r = 0.01$ . The initial values  $\mathbf{y}_0$  are uniformly distributed in the cube  $[-1, 1]^6$ . Since the gradient of  $H_{4,6}$  vanishes for big values of  $\|\mathbf{y}\|$ , in our simulations we consider the penalised function:

$$H_{4,6}^p(\mathbf{y}) = H_{4,6}(\mathbf{y}) + \frac{1}{6} \sum_{i=1}^6 (|y_i| - 1)^{2.2} \mathbb{1}(|y_i| > 1). \tag{6.24}$$

$\lambda$	$\langle N_{\text{first}} \rangle$	$k = 5 \cdot 10^4$		$k = 5 \cdot 10^5$	
		$N_k$	$\Delta_k$	$N_k$	$\Delta_k$

*Stable-like simulated annealing*

$a = 1.5, A = 1.9, \theta = 0.6$

100	20070	13	0.0295	84	0.0044
250	18492	18	0.0257	84	0.0043
500	12841	43	0.0264	80	0.0049
1000	18624	42	0.0155	83	0.0042
5000	29495	73	0.0043	83	0.0025

*Fast simulated annealing*

$a = 1.5, A = 1.9, \theta = 0.6$

10	1 302 701	35	0.0044	35	0.0001
50	1 525 483	17	0.0323	23	0.0029
100	1 451 393	2	0.0428	26	0.0049

500	1 348 318	0	–	11	0.0323
1000	865 728	0	–	7	0.0388
<i>Gaussian SDE</i>					
$\theta = 6$					
1 000 000	360 969	0	–	0	–
10 000 000	191 639	0	–	0	–
100 000 000	597 533	0	–	0	–
<i>Gaussian simulated annealing</i>					
$\theta = 6$					
10 000	668 062	0	–	0	–
100 000	193 789	0	–	0	–
1 000 000	466 592	0	–	1	0.0430

In  $\mathbb{R}^6$ , our method shows essentially better results. The success ratio of the fast simulated annealing seems to relate to the geometric probability to jump to the domain of attraction of the global minimum. Gaussian methods fail to converge to the global minimum due to the big values of the temperature parameter  $\sigma(\cdot)$ . Making  $\sigma(\cdot)$  smaller leads to much longer computation times.

## 7. Conclusion and discussion

The goal of this paper is to present and theoretically justify a new stochastic algorithm for global optimisation. The algorithm allows to determine a global minimum of an unknown potential  $U$  with help of simulated annealing of non-Gaussian jump-diffusions driven by the so-called stable-like processes, or Lévy flights with a variable stability index  $\alpha(x)$ . We have shown that choosing  $\alpha(x)$  in an appropriate way, we can force the Lévy particle to settle in a neighbourhood of the global maximum of  $U$ . We note, that the non-constant behaviour of the stability index is crucial for the success of the search and a similar algorithm with usual spatially homogeneous Lévy flights, i.e., when  $\alpha(x) = \alpha_0$ , leads to quite different results, see [21].

Our method has the following advantages in comparison with the Gaussian simulated annealing considered in the introduction and in Section 6. First, the search of the global minimum is non-local, i.e., when the annealed process leaves a potential well, it does not necessarily pass to one of the neighbouring wells, but with strictly positive probability can jump to any well. Moreover, the probability to jump into the deepest well is maximal, if this well is also spatially the biggest, which is observed in typical potential landscapes, see [25]. We do not expect that our algorithm would effectively detect the so-called ‘golf-hole’ wells. Mean transition times between the wells increase as a power of the large parameter  $\lambda$  or, equivalently, the current time  $t$ . We can easily obtain theoretic estimates for a number of transitions between the wells before settling in the deepest well. These estimates follow from the analysis of a discrete time Markov chain with transition probabilities  $p_{ij} = q_{ij}^{(\alpha(m_i))} [q_i^{(\alpha(m_i))}]^{-1}$ ,  $p_{ii} = 0$ .

In comparison to the Cauchy machine by Szu and Hartley, our method better uses the geometry properties of the potential function, which is especially important in higher dimensions.

Further, we have more freedom to choose the parameters of the system. Indeed, if the values of  $U(m_i)$  are not known, there is no method which helps to determine the cooling rate  $\theta$ . (One has the same problem to determine  $\hat{\theta}$  in (1.2) in Gaussian case.) However, in our algorithm,  $\theta$  is chosen together with a variable stability index  $\alpha(x)$ .

Our method also allows to drive the Lévy particle into any well whose location is approximately known. We can also determine a local minimum with energy below a certain given value. The choice of parameters in these regimes is independent of the geometry of the potential. Such search regimes are not possible in the Gaussian and Cauchy settings.

Finally, the temperature decreases polynomially fast in time, i.e.,  $\sim t^{-\theta}$  and not logarithmic. Moreover, depending on the choice of  $\alpha(\cdot)$ , we can choose  $\theta > 1$ , i.e., cool the system down faster than in the Cauchy machine, what significantly increases the accuracy of empirical estimates for the local minima locations.

Although the theoretic basis for the success of the search is established, many questions are still open. For example, we have to understand how to choose the optimal triple  $\alpha(x)$ ,  $\theta$  and  $\lambda$  which minimises the search time. Indeed, if  $m_i$  is not a global minimum, we can reduce the life time of the particle in the neighbourhood of  $m_i$  making  $\alpha(m_i)$  small. On the other hand, in this case, the process  $Y$  will tend to make very big jumps, and thus jump out to one of the peripheral wells. As a consequence, the search would be slow, if the global minimum of  $U$  is attained in one of the inner wells. Thus, the value of  $\alpha(m_i)$  should not be very small to exclude very big jumps and should be well separated from  $\alpha(M)$  to block trapping in the false well. The problem of very big jumps can be also avoided by consideration of truncated Lévy flights with maximal jump size not exceeding the size of the search domain. However, in this case the simulation of random input can be more complicated.

Another question addresses the proper choice of  $\alpha(\cdot)$ . If the stability index takes two values as in (5.8), it could be helpful to decrease  $E$  gradually in time, or make some adaptive search, with  $E$  being dependent from the observed energy values.

We shall address these and other questions in our further research.

## Acknowledgments

This work was supported by the DFG research project *Stochastic Dynamics of Climate States*. The author thanks P. Imkeller for stimulating discussions and two unknown referees for their helpful comments.

## References

- [1] D. Applebaum, Lévy Processes and Stochastic Calculus, Cambridge Studies in Advanced Mathematics, 93, Cambridge University Press, Cambridge, 2004.
- [2] F. Aluffi-Pentini, V. Parisi, F. Zirilli, Global optimization and stochastic differential equations, *Journal of Optimization Theory and Applications* 47 (1) (1985) 1–16.
- [3] D. Brockmann, I.M. Sokolov, Lévy flights in external force fields: from models to equations, *Chemical Physics* 284 (2002) 409–421.
- [4] A.V. Chechkin, V. Yu. Gonchar, J. Klafter, R. Metzler, L.V. Tanatarov, Lévy flights in a steep potential well, *Journal of Statistical Physics* 115 (5–6) (2004) 1505–1535.
- [5] T.-S. Chiang, C.-R. Hwang, S.-J. Sheu, Diffusion for global optimization in  $\mathbb{R}^n$ , *SIAM Journal on Control and Optimization* 25 (3) (1987) 737–753.
- [6] A. Chechkin, O. Sliusarenko, R. Metzler, J. Klafter, Barrier crossing driven by Lévy noise: Universality and the role of noise intensity, *Physical Review E* 75 (2007) 041101.
- [7] R. Cont, P. Tankov, *Financial modelling with jump processes*, Chapman & Hall/CRC, 2004.
- [8] S. Geman, C.-R. Hwang, Diffusions for global optimization, *SIAM Journal on Control and Optimization* 24 (5) (1986) 1031–1043.
- [9] B. Gidas. Global optimization via the Langevin equation, in: *Proceedings of Conference on Decision and Control*, 24th, Fort Lauderdale, FL, USA, 11–13 December, 1985.
- [10] S.B. Gelfand, S.K. Mitter, Metropolis-type annealing algorithms for global optimization in  $\mathbb{R}^d$ , *SIAM Journal on Control and Optimization* 31 (1) (1993) 111–131.
- [11] B. Hajek, Cooling schedules for optimal annealing, *Mathematics of Operations Research* 13 (2) (1988) 311–329.
- [12] R. Holley, S. Kusuoka, D. Stroock, Asymptotics of the spectral gap with applications to the theory of simulated annealing, *Journal of Functional Analysis* 83 (2) (1989) 333–347.
- [13] C.-R. Hwang, S.-J. Sheu, Large-time behavior of perturbed diffusion Markov processes with applications to the second eigenvalue problem for Fokker–Planck operators and simulated annealing, *Acta Applicandae Mathematicae* 19 (3) (1990) 253–295.
- [14] P. Imkeller, I. Pavlyukevich, First exit times of SDEs driven by stable Lévy processes, *Stochastic Processes and their Applications* 116 (4) (2006) 611–642.
- [15] P. Imkeller, I. Pavlyukevich, Lévy flights: transitions and meta-stability, *Journal of Physics A: Mathematical and General* 39 (2006) L237–L246.
- [16] P. Imkeller, I. Pavlyukevich, Metastable behaviour of small noise Lévy-driven diffusions, *math.PR/0601771*, 2006.
- [17] A. Janicki, A. Weron, *Simulation and chaotic behaviour of  $\alpha$ -stable stochastic processes*, Pure and Applied Mathematics, 178, Marcel Dekker, Inc., New York, Basel, 1994.
- [18] V.N. Kolokol'tsov, K.A. Makarov, Asymptotic spectral analysis of a small diffusion operator and the life times of the corresponding diffusion process, *Russian Journal of Mathematical Physics* 4 (3) (1996) 341–360.

- [19] R. Metzler, J. Klafter, The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics, *Journal of Physics A: Mathematical and General* 37 (2004) R161–R208.
- [20] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, E. Teller, Equation of state calculations by fast computing machines, *The Journal of Chemical Physics* 21 (5) (1953) 1087–1092.
- [21] I. Pavlyukevich, Cooling down Lévy flights. cond-mat/0701651, submitted.
- [22] I. Pavlyukevich, Simulated annealing of Lévy-driven jump-diffusions, submitted.
- [23] Ph. E. Protter, Stochastic integration and differential equations, *Applications of Mathematics*, 2nd ed., 21, Springer, Berlin, 2004.
- [24] K.-I. Sato, Lévy processes and infinitely divisible distributions, *Cambridge Studies in Advanced Mathematics*, 68, Cambridge University Press, Cambridge, 1999.
- [25] J.C. Schön, Preferential trapping on energy landscapes in regions containing deep-lying minima: The reason for the success of simulated annealing? *Journal of Physics A: Mathematical and General* 30 (7) (1997) 2367–2389.
- [26] H. Szu, R. Hartley, Fast simulated annealing, *Physics Letters A* 122 (3,4) (1987) 157–162.
- [27] H. Szu, Automated fault recognition by image correlation neural network technique, *IEEE Transactions on Industrial Electronics* 40 (2) (1993) 197–208.
- [28] D. Vanderbilt, S.G. Louie, A Monte Carlo simulated annealing approach to optimization over continuous variables, *Journal of Computational Physics* 56 (1984) 259–271.
- [29] A.D. Wentzell, On the asymptotic behavior of the greatest eigenvalue of a second-order elliptic differential operator with a small parameter in the higher derivatives, *Soviet Math Doklady* 13 (1) (1972) 13–17.
- [30] A.D. Wentzell, On the asymptotic of eigenvalues of matrices with elements of order  $\exp\{-V_{ij}/(2\epsilon^2)\}$ , *Soviet Math Doklady* 13 (1) (1972) 65–68.