J Optim Theory Appl
DOI 10.1007/s10957-013-0394-5

Interactive Diffusions for Global Optimization

Yue Sun · Alfredo Garcia

Received: 1 March 2012 / Accepted: 5 August 2013

Abstract We present a novel approach, in which parallel annealing processes interact in a manner that expedites the identification of a globally optimal solution. A first annealing process operates at a faster time scale and has a drift function that converges to a non-zero (but relatively small) noise level. A second annealing process (operating at a slower time scale) is subject to a modified drift term in which the steepest descent direction is perturbed with the first annealing process density gradient. This additional term ensures that the second process is "repelled" from regions already explored. As a result, the first annealing process (which quickly identifies locally optimal solutions) allows the second annealing process to *bypass* locally optimal solutions recently identified, so that it can be made to converge to global optima at a faster rate. We show that, when compared to independent annealing processes, the proposed interactive diffusions can increase the speed of convergence at the expense of minimal additional computational overhead.

Keywords Diffusions · Simulated annealing · Parallel computing

1 Introduction

Simulated annealing for global optimization (see [1, 2]) can be modeled in a continuous domain as a diffusion process, in which the drift term is equal to the steepest descent direction and the drift function (i.e. the "cooling schedule") approaches zero

Communicated by Panos M. Pardalos.

Y. Sun · A. Garcia (✉)
Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA, USA
e-mail: agarcia@virginia.edu

Y. Sun
e-mail: ys6dn@virginia.edu

Published online: 04 September 2013

at a suitable rate so that weak convergence to a limiting distribution (concentrating almost all mass in the set of globally optimal solutions) is guaranteed (see, for example, [3–5]). The main drawback of this class of methods relates to the speed of convergence as the cooling schedule cannot be "too fast" in order to guarantee convergence to globally optimal solutions. In other words, the initial emphasis on "exploration" must slowly give way to emphasis on "exploitation" in order to guarantee globally optimal solutions are identified. Thus, many research efforts have been devoted to accelerating the convergence rate for certain classes of smooth objective functions (see, e.g., [6, 7] and, more recently, [8]). Given the technological challenges currently facing speed-up of processing,[1] parallel implementation of simulated annealing appears as a sensible approach for speeding up convergence to globally optimal solutions (see, for instance, [9] for an application to molecular clustering).

In this paper, we present a novel approach, in which parallel annealing processes interact in a manner that enables a faster identification of a globally optimal solution. A first annealing process operates at a faster time scale and has a drift function that converges to a non-zero (but relatively small) noise level. A second annealing process (operating at a slower time scale) is subject to a modified drift term, in which the steepest descent direction is perturbed with the density gradient associated to the first annealing process. In other words, a repulsive potential is added in order to ensure the second annealing process does not duplicate the first annealing process search effort. We show that this repulsive potential (based upon the first annealing process, which quickly identifies locally optimal solutions) allows the second annealing process to *bypass* locally optimal solutions in a faster fashion.

Our work is related to the class of smoothing methods (see [10, 11]), in which original objective function evolves into a (smoothed) function possessing far fewer local minima. A potential drawback of these methods pertains to the possibility that the minimum of the modified objective function may shift away from the global minimum of the original objective function (see [11] for a smoothing method designed to avoid this difficulty). The approach proposed in this paper can be seen as a time-varying smoothing of the original objective function, which is related to the first diffusion density. The computational work by the first diffusion is used to construct a smoother objective function governing the second (slower) diffusion's search. Building on [12], we show that it is easier for the second diffusion to escape the attraction basins of locally optimal solutions.

The structure for this paper is as follows. In Sect. 2, we formalize the proposed approach in the context of a global optimization problem. In Sect. 3 we study the "escape time" of interactive annealing processes. It is shown that due to interaction, one of the annealing processes is able to escape the basins of attraction of local minima in less time (in expectation). The discussion on this speed-up effect is followed by the analysis of global asymptotic behavior in Sect. 4, in which convergence to global minima is proven. Building up on the results in Sects. 3 and 4, we show that the speed of convergence is improved in Sect. 5. Finally in Sect. 6, we provide a limited numerical testbed to illustrate the merits of the proposed approach (which are established

[1] As chips are clocked at higher speeds, it is increasingly difficult to control their temperature and they become much less energy-efficient.

in Sects. 2 and 3 from a theoretical standpoint). Finally, in Sect. 7, we offer some conclusions and briefly comment upon future research.

2 Framework

Consider the optimization problem $\min\{H(x) : x \in \mathbb{R}^n\}$, where $H : \mathbb{R}^n \to \mathbb{R}$ is assumed to be continuously differentiable. As in [4], we make the following regularity assumption throughout the paper:

$$\lim_{\|x\|\uparrow+\infty} H(x) \uparrow +\infty, \tag{1}$$

$$\lim_{\|x\|\uparrow+\infty} \left\|\nabla H(x)\right\| \uparrow +\infty, \tag{2}$$

$$\lim_{\|x\|\uparrow+\infty} \left\|\nabla H(x)\right\| - \Delta H(x) > -\infty. \tag{3}$$

Let $y_l := \min\{H(x) : x \in B_l \subset \mathbb{R}^n\}$ for $l = 1, 2, \ldots, m$ denote the collection of local minima values, where $B_l \subset \mathbb{R}^n$ is the basin of attraction for the lth local minimum. Let $y_* := \min\{y_l : l = 1, \ldots, m\}$. Assume there exists a positive number M such that

$$y_l - y_* > M \gg 0, \quad \forall y_l \neq y_*, l = 1, 2, \ldots, m. \tag{4}$$

2.1 Interactive Diffusions

We start by introducing the dynamic system for a diffusion in $\mathbb{R}^n$ as follows:

$$\begin{cases} \dot{x}_p(t) = v_p(t) \\ \dot{v}_p(t) = \dfrac{1}{M}\left[-\dfrac{\partial H(x)}{\partial x_p} - \gamma_0 v_p(t) + \sigma_0(t)\dfrac{dW_p(t)}{dt}\right], \end{cases}$$

according to the Fokker–Planck equation (see [13]), where $p = 1, \ldots, n$ and M is particle mass, γ_0 is damping coefficient, $\sigma_0(\cdot)$ is noise intensity and $W_p(t)$ is pth component of a standard n-dimensional Brownian motion, $W(t)$. Assume the particle mass M is large enough relative to $-\frac{\partial H(x)}{\partial x_p} - \gamma_0 v_p(t) + \sigma_0(t)\frac{dW_p(t)}{dt}$, so that $\dot{v}(t) \approx 0$ and the system can be reduced to

$$\frac{dX(t)}{dt} = \frac{1}{\gamma_0}\left[-\nabla H(x) + \sigma_0(t)\frac{dW(t)}{dt}\right].$$

Let $U(x) = \frac{H(x)}{\gamma_0}$, $\sigma(t) = \frac{\sigma_0(t)}{\gamma_0}$. We obtain a diffusion model for simulated annealing as

$$\frac{dX(t)}{dt} = -\nabla U(x) + \sigma(t)\frac{dW(t)}{dt}. \tag{5}$$

The partial differential equation of this diffusion process is given by

$$\frac{\partial}{\partial t}V(t,x) = \nabla \cdot \big(V(t,x)\nabla U(x)\big) + \frac{1}{2}\sigma^2(t)\Delta V(t,x),$$

where $V(\cdot,\cdot)$ is the density of particles. In our method, we shall make use of the notion of different time scales for interacting diffusions. In what follows, we shall consider a first diffusion (with subindex) i in a time scale t/ϵ^2 where $\epsilon > 0$ with constant noise intensity σ_i, and a second diffusion j in a slower time scale t, with annealing noise intensity $\sigma_j(t) \sim C/\log(t)$, interacting with diffusion i. The dynamic system for these interactive diffusions is

$$\begin{cases} dX_i(t) = -\dfrac{1}{\epsilon^2}\nabla U\big(X_i(t)\big)\,dt + \dfrac{1}{\epsilon}\sigma_i\,dW(t) \\ dX_j(t) = -\nabla U\big(X_j(t)\big)\,dt - k\nabla V_i\big(t, X_j(t)\big)\,dt + \sigma_j(t)\,dW(t) \\ X_i(0) = \hat{x}_i, \quad X_j(0) = \hat{x}_j, \end{cases} \tag{6}$$

where V_i is density for diffusion i, $\hat{x}_i$ and $\hat{x}_j$ are the initial conditions for diffusion i and j, $k > 0$ is a parameter that controls the strength of the "repulsive" potential, i.e. $-\nabla V_i(t, X_j(t))$ affecting diffusion j. From the Feynman–Kac formula [14], the associated Fokker–Planck type equation is

$$\begin{cases} \dfrac{\partial}{\partial t}V_i(t,x) = \dfrac{1}{2\epsilon^2}\sigma_i^2\Delta V_i(t,x) + \dfrac{1}{\epsilon^2}\nabla \cdot \big(V_i(t,x)\nabla U(x)\big) \\ \dfrac{\partial}{\partial t}V_j(t,x) = \dfrac{1}{2}\sigma_j^2(t)\Delta V_j(t,x) + \nabla \cdot \big[V_j(t,x)\big(\nabla U(x) + k\nabla V_i(t,x)\big)\big] \\ V_i(0,x) = \delta_{\hat{x}_i}(x), \quad V_j(0,x) = \delta_{\hat{x}_j}(x). \end{cases} \tag{7}$$

3 Local Escape Time Problem

In this section, we analyze the "escape times" for the interactive diffusions using the theory of large deviations introduced in [5]. We then proceed to compare the expected escape time properties for the interactive diffusions and the original simulated annealing. Finally, we argue that such comparison is fair even when different time scales (used in the interactive diffusions approach) are taken into account.

3.1 Escape Time Theory

For the dynamical system (5), define L as follows:

$$L(\beta, x) := \sup_{\alpha}\big[\alpha'\beta - \alpha'\big(-\nabla U(x)\big)\big], \quad \alpha, \beta \in \mathbb{R}^n.$$

The action functional $S_x(T,\phi)$ for point x is defined as

$$S_x(T,\phi) := \int_0^T L\big(\dot{\phi}(s), \phi(s)\big)\,ds, \quad \phi(0) = x.$$

Finally, the action function between two points x and y is

$$S(x, y) = \inf_{\phi, T}\{S_x(T, \phi) : \phi(T) = y\}.$$

Let K_0 be the set of local minima as

$$K_0 = \{x_l \in \mathbb{R}^n :| H(x_l) = y_l, l = 1, \ldots, m\}.$$

Let G be a bounded open set containing K_0, with a piecewise differentiable boundary ∂G and $\bar{G}$ in the domain of attraction of K_0, i.e. $\bar{G} \subseteq \bigcup_{l=1}^{m} B_l$, for $x \in G$ define

$$S_G(x) := \inf_{y \in \partial G} S(x, y) = \inf_{\phi, T}\{S_x(T, \phi) : \phi(T) \in \partial G\}.$$

Define the action function for any open set $B \subset G$ as follows:

$$S_G(B) := \inf_{x \in B} S_G(x).$$

The action function being the integral of difference between the potential gradient direction $-\nabla U(\phi(s))$ and the path moving direction $\dot{\phi}(s)$, can be understood as the "energy" needed to resist the potential force through the path ϕ.

Assumption 3.1 (See [5], A3.1) *For $\delta > 0$, there is a ρ-neighborhood $N_\rho(K_0)$ of K_0 and $\delta_\rho > 0$, $T_\rho < +\infty$, such that for each $x, y \in N_\rho(K_0)$, there is a path $\phi(\cdot)$:*

$$\phi(0) = x, \phi(T) = y,$$

where $T_y \leq T_\rho$ and $S_x(T_\rho, \phi) \leq \delta$.

Lemma 3.1 *Let $X^\sigma(t)$ be the solution of dynamic system (5) with noise intensity $\sigma(t) = \sigma$ and escape time $\tau^\sigma = \min\{t : X^\sigma(t) \notin G\}$. Under Assumption 3.1, the expected escape time can be written as*

$$\lim_{\sigma} \sigma \log E_x \tau^\sigma = S_G(K_0),$$

which can be written as

$$E_x \tau^\sigma \sim \exp\left(\frac{S_G(K_0)}{\sigma}\right).$$

Proof See proof of [5], p. 174, Theorem 1. □

3.2 Escape Time Comparison

In this section, we will compare the local minima escape time of standard simulated annealing and that of the interactive diffusions.

Theorem 3.1 *Let* $\bar{\tau}_j^\sigma = \min\{t : X_j^\sigma(t) \notin G\}$, *i.e. the escape time of the* j*-diffusion process defined in* (6) *when the noise intensity for the* i*- diffusion process is* $\sigma_i = \sigma$. *When* $\epsilon \downarrow 0$ (*the time scale ratio for diffusion* i *in* (6)), *we have*

$$\lim_\sigma \sigma \log E_x \bar{\tau}_j^\sigma \le \lim_\sigma \sigma \log E_x \tau^\sigma.$$

Proof Consider the behavior of fluid i of (7) in fast time scale t/ϵ^2 with constant noise intensity σ_i. let $\hat{t} = t/\epsilon^2$, the dynamic for fluid i is

$$\epsilon^2 \frac{\partial}{\partial t} V_i(t,x) = \frac{\partial}{\partial \hat{t}} V_i(t,x) = \frac{1}{2}\sigma_i^2 \Delta V_i(t,x) + \nabla \cdot \big(V_i(t,x)\nabla U(x)\big).$$

When $\epsilon \downarrow 0$, we have $\hat{t} \uparrow +\infty$, from standard results in the theory of diffusions (see [15], p. 147), we have the fluid i stationary density $\bar{V}_i(x) = \pi_\sigma(x)$, where

$$\pi_\sigma(x) = C_0 \exp\left(-\frac{2U(x)}{\sigma^2}\right), \tag{8}$$

and

$$C_0 = \left(\int_{\mathbb{R}^n} \exp\left(-\frac{2U(x)}{\sigma^2}\right) dx\right)^{-1},$$

which is the Gibbs density.

For any K_0 and $\epsilon \downarrow 0$, the fluid j can be seen as a diffusion governed by the modified potential function

$$\bar{U}(x) = U(x) + k\bar{V}_i(x) = U(x) + kV_0 \exp\left(-\frac{2U(x)}{\sigma^2}\right).$$

The corresponding L-function $\bar{L}$ can be written as

$$\bar{L}(\beta, x) = \sup_\alpha \left[\alpha'\beta + \alpha'\nabla U(x)\left(1 - \frac{2k}{\sigma^2}\exp\left(-\frac{2U(x)}{\sigma^2}\right)\right)\right].$$

We have, for all $x \in G$, $\|\nabla \bar{U}(x)\| \le \|\nabla U(x)\|$,

$$\bar{S}_G(K_0) = \inf_{x\in K}\inf_{y\in\partial G}\inf_{\phi,T}\left\{\int_0^T \bar{L}\big(\dot{\phi}(s), \phi(s)\big)\, ds : \phi(0) = x, \phi(T) = y\right\} < S_G(K_0).$$

From Lemma 3.1, we have

$$\lim_\sigma \sigma \log E_x \bar{\tau}_j^\sigma = \bar{S}_G(K_0) \le S_G(K_0) = \lim_\sigma \sigma \log E_x \tau^\sigma. \qquad \square$$

3.3 Discussion

According to the previous results, in expectation, diffusion j is able to escape the basin of attraction of locally optimal solutions in less time than a standard simulated

annealing type diffusion. However, in this comparison the standard simulated annealing is assumed to diffuse in slow time scale t. In order to show that this comparison is fair, we need to consider the standard simulated annealing method in a faster time scale.

Consider the standard simulated annealing method in different time scale: a first diffusion $\{X_a(t) : t > 0\}$ (with density V_a) in slow time scale t and a second diffusion $\{X_b(t) : t > 0\}$ (with density V_b) in faster time scale t/ϵ^2 so that

$$\frac{\partial}{\partial t}V_a(t,x) = \frac{1}{2}\frac{c}{\log t}\Delta V_a(t,x) + \nabla\cdot\big(V_a(t,x)\nabla U(x)\big)$$
$$\frac{\partial}{\partial t}V_b(t,x) = \frac{1}{2}\frac{c/\epsilon^2}{\log t/\epsilon^2}\Delta V_b(t,x) + \nabla\cdot\left(V_b(t,x)\frac{\nabla U(x)}{\epsilon^2}\right).$$

Let t_a and t_b be the time required to reach a noise intensity $\sigma_0 > 0$. It follows that

$$\sigma_0 = \frac{C/\epsilon^2}{\log t_b/\epsilon^2} = \frac{C}{\log t_a}.$$

From this we infer that

$$t_b = \epsilon^2\, t_a^{1/\epsilon^2}.$$

For general situation, for $t_a \gg 1$ and $\epsilon \downarrow 0$, we have $t_b \gg t_a$. That is, in a faster time scale it takes longer to arrive a given level of noise intensity. Using a faster time scale cannot accelerate convergence to global minima for the standard simulated annealing method.

4 Global Asymptotic Behavior

In this section, we will analyze the global asymptotic behavior of "slow" diffusion process, i.e. $X_j(t)$. As in [4], let S denote the set of all local minima of U and $S(\eta) = \{x | d(x,S) < \eta\}$. Let

$$J(t,\eta) := \sup_{x,y\in S(\eta)}\big(I(t,x,y) - 2U(y)\big),$$

where

$$I(t,x,y) := \inf_{\substack{\psi(0)=x\\ \psi(t)=y}}\frac{1}{2}\int_0^t \big|\dot\psi(s) + \nabla U\big(\psi(s)\big)\big|^2\,ds.$$

Define

$$c_0 := \inf_{\eta}\limsup_{t\uparrow+\infty} J(t,\eta) \tag{9}$$

Lemma 4.1 *Under assumption* (1), (2) *and* (3), *and* $c > c_0$,

$$V(t,x) \to \pi(x) \quad \text{as } t\uparrow+\infty,$$

where $V(\cdot,\cdot)$ is the solution for the equation

$$\frac{\partial}{\partial t}V(t,x)=\frac{c}{2\log t}\Delta V(t,x)+\nabla\cdot\big(V(t,x)\nabla U(x)\big),\qquad V(0,x)=\delta_{x_0}(x)$$

and $\pi(x):=\lim_{\sigma\downarrow 0}\pi_\sigma(x)$ concentrate on the global minimum x_* of $U(x)$, with $\pi_\sigma(x)$ defined as (8).

Proof See proof of [4], p. 739, theorem. □

For interactive diffusion (7), define the modified objective function

$$\tilde{U}(t,x):=U(x)+kV_i(t,x);$$

for a slow diffusion process j, define the constant $\tilde{c}_0$ similar as (9)

$$\tilde{c}_0:=\inf_{\eta}\limsup_{t\uparrow+\infty}\tilde{J}(t,\eta),$$

where

$$\tilde{J}(t,\eta):=\sup_{x,y\in S(\eta)}\big(\tilde{I}(t,x,y)-2\tilde{U}(y)\big),$$

and

$$\tilde{I}(t,x,y):=\inf_{\substack{\psi(0)=x\\ \psi(t)=y}}\frac{1}{2}\int_0^t\left|\dot{\psi}(s)+\nabla U\big(\psi(s)\big)\left(1-k\frac{V_i(x)}{\sigma^2}\right)\right|^2 ds.$$

Theorem 4.1 *For fluid j with dynamic (7), and for fluid i, and time scale ratio $\epsilon\downarrow 0$, let $\sigma_j=c/\log t$ as $c>\tilde{c}_0$; under assumption (4), we have*

$$V_j(t,x)\to\tilde{\pi}(x),$$

where

$$\tilde{\pi}(x)=\lim_{\sigma\downarrow 0}\tilde{\pi}_\sigma(x)=\lim_{\sigma\downarrow 0}\left[\tilde{C}_0\exp\left(-\frac{2\tilde{U}_0(x)}{\sigma^2}\right)\right].$$

This $\tilde{\pi}(x)$ is concentrated on the global minimum $\tilde{x}_*$ of the modified objective function $\tilde{U}(x)$. Moreover $\tilde{x}_*$ is in the attraction basin of the global minimum x_* of the original objective function $U(x)$.

Proof From (8), with time scale ratio $\epsilon\downarrow 0$, $V_i(t,x)\to\bar{V}_i(x)$, where

$$\bar{V}_i(x)=V_0\exp\left(-\frac{2U(x)}{\sigma^2}\right).$$

Let $\tilde{U}_0(x)=U(x)+k\bar{V}_i(x)$, we have

$$\tilde{U}(t,x)\to\tilde{U}_0(x)\quad\text{as } t\uparrow+\infty.$$

Since $U(x)$ is twice continuous differentiable, then $\bar{V}_i(x)$ is at least twice differentiable. By assumption, $U(x) \uparrow +\infty$ as $\|x\| \uparrow +\infty$, hence, $\bar{V}_i(x) \downarrow 0$ as $U(x) \uparrow +\infty$ and $\lim_{\|x\|\uparrow+\infty} \bar{V}_i(x) = 0$. We conclude that the stationary modified objective function $\tilde{U}_0(x)$ satisfies assumptions (1), (2), and (3). From Lemma 4.1 for fluid j we have

$$V_j(t,x) \to \tilde{\pi}(x),$$

where

$$\tilde{\pi}(x) = \lim_{\sigma\downarrow 0} \tilde{\pi}_\sigma(x) = \lim_{\sigma\downarrow 0}\left[\tilde{C}_0 \exp\left(-\frac{2\tilde{U}_0(x)}{\sigma^2}\right)\right];$$

and fluid j will concentrate on the global minima of the modified objective function $\tilde{U}_0(x)$, which is different from the global minima of the original objective function $U(x)$.

Now, we show the global minimum of $\tilde{U}_0(x)$ is in the attraction basin of the global minimum of $U(x)$ under assumption (4) with $k < M$. Let x_* be the global minimum for $U(x)$ with value y_*, x_i, $i = 1, 2, \ldots, m$, be the local minima with values y_i. Assume $\tilde{x}_*$ be a global minimum for the modified objective function $\tilde{U}_0(x)$ with value $\tilde{y}_*$. Note that $\bar{V}_j(t,x) \le 1, \forall x \in \mathbb{R}^n$. From $k < M$, we have

$$\tilde{U}_0(x_*) = y_* + k\bar{V}_i(x) \le y_* + k \le y_* + M.$$

From assumption (4), for all $y_l \ne y_*$, $l = 1, 2, \ldots, m$, $y_* + M < y_i$, and $\tilde{y}_*$ is the global minimum value for $\tilde{U}_0(x)$, we get

$$\tilde{y}_* \le \tilde{U}_0(x_*) \le y_* + M < y_l \quad \forall y_l \ne y_*, l = 1, 2, \ldots, m.$$

From

$$\tilde{y}_* = \tilde{U}_0(\tilde{x}_*) = U(\tilde{x}_*) + k\bar{V}(x) \ge U(\tilde{x}_*),$$

we have

$$U(\tilde{x}_*) \le \tilde{y}_* \le y_* + M < y_l \quad \forall y_l \ne y_*, l = 1, 2, \ldots, m.$$

This means that $U(\tilde{x}_*)$ is less than all the local minima for original objective function. It follows that the global minimum $\tilde{x}_*$ for the modified objective function is in the attraction basin of the global minimum of the original objective function $U(x)$. □

5 Improved Speed of Convergence

In Sects. 3 and 4, we have shown that interactive annealing processes have shorter local escape time and converge to global minima attraction basin. To complement these results, in this section, we show that interactive annealing also exhibits improved speed of convergence.

The standard simulated annealing method requires an annealing schedule $c_0/\log t$ in order to guarantee the fluid converge to global minima (see Lemma 4.1), where c_0

is defined as follows:

$$c_0 := \inf_{\eta} \limsup_{t\uparrow+\infty} \sup_{x,y\in S(\eta)} \big(I(t,x,y) - 2U(y)\big)$$

and

$$I(t,x,y) := \inf_{\substack{\psi(0)=x\\ \psi(t)=y}} \frac{1}{2}\int_0^t \big|\dot\psi(s) + \nabla U\big(\psi(s)\big)\big|^2 ds.$$

In Theorem 4.1, it is shown that the interactive diffusion approach also requires an annealing schedule $\tilde c_0/\log t$ in order to converge to global minima,

$$\tilde c_0 := \inf_{\eta} \limsup_{t\uparrow+\infty} \sup_{x,y\in S(\eta)} \big(\tilde I(t,x,y) - 2\tilde U(y)\big),$$

where

$$\tilde I(t,x,y) := \inf_{\substack{\psi(0)=x\\ \psi(t)=y}} \frac{1}{2}\int_0^t \left|\dot\psi(s) + \nabla U\big(\psi(s)\big)\left(1 - k\frac{V_i(\psi(s))}{\sigma^2}\right)\right|^2 ds.$$

From (8), for fluid i, we have

$$V_i(y,t) = \pi_{\sigma_i}(y) + \delta(y,\epsilon), \quad \text{as } \lim_{\epsilon\downarrow 0}\delta(y,\epsilon) = 0.$$

For the Gibbs density

$$\pi_{\sigma_i}(x) := \frac{\exp(-\frac{2U(x)}{\sigma_i^2})}{\int_{R^n}\exp(-\frac{2U(x)}{\sigma_i^2})\,dx},$$

we have

$$\nabla V_i(y,t) = -\nabla U(y)\cdot\frac{2\pi_{\sigma_i}(y)}{\sigma^2} + \nabla\delta(y,\epsilon).$$

It follows that

$$\begin{aligned}
\tilde I(t,x,y) &= \inf_{\substack{\psi(0)=x\\ \psi(t)=y}} \frac{1}{2}\int_0^t \left|\dot\psi(\tau) + \nabla U\big(\psi(\tau)\big)\left(1 - K\frac{2\pi_\sigma(\psi(\tau))}{\sigma^2}\right) + \nabla\delta(\psi(\tau),\epsilon)\right|^2 d\tau\\
&= \inf_{\substack{\psi(0)=x\\ \psi(t)=y}} \left\{\frac{1}{2}\int_0^t \left|\dot\psi(\tau) + \nabla U\big(\psi(\tau)\big)\left(1 - K\frac{2\pi_\sigma(\psi(\tau))}{\sigma^2}\right)\right|^2 d\tau\right.\\
&\quad + \frac{1}{2}\int_0^t \big|\nabla\delta(\psi(\tau),\epsilon)\big|^2 d\tau\\
&\quad \left. + \int_0^t \left|\dot\psi(\tau) + \nabla U\big(\psi(\tau)\big)\left(1 - K\frac{2\pi_\sigma(\psi(\tau))}{\sigma^2}\right)\right|\cdot\big|\nabla\delta(\psi(\tau),\epsilon)\big| d\tau\right\}\\
&= \inf_{\substack{\psi(0)=x\\ \psi(t)=y}} \frac{1}{2}\int_0^t \left|\dot\psi(\tau) + \nabla U\big(\psi(\tau)\big)\left(1 - K\frac{2\pi_\sigma(\psi(\tau))}{\sigma^2}\right)\right|^2 d\tau + C(\epsilon)\\
&\le I(t,x,y) + C(\epsilon),
\end{aligned}$$

where $C(\epsilon) \downarrow 0$ as $\epsilon \downarrow 0$. If $\epsilon \downarrow 0$, then

$$\begin{aligned}\tilde{c}_0 &= \inf_{\eta} \limsup_{t\uparrow+\infty} \sup_{x,y\in S(\eta)} \big(\tilde{I}_j(t,x,y) - 2\big(U(y) + K \cdot V_i(y,t)\big)\big) \\ &< \inf_{\eta} \limsup_{t\uparrow+\infty} \sup_{x,y\in S(\eta)} \big(I(t,x,y) - 2U(y)\big) = c_0,\end{aligned}$$

where the inequality follows from the fact that

$$2\big(U(y) + K \cdot V_i(y,t)\big) > 2U(y).$$

In other words, the cooling or annealing schedule for interactive diffusion can be decreased in constant level from $c_0/\log t$ to $\tilde{c}_0/\log t$. We now restate a result ([4], p. 740, Lemma 3) that establishes the relationship between the cooling schedule and the speed of convergence for an annealing process.

Let $V^\sigma(x,t)$ be the solution of dynamic system

$$\frac{\partial}{\partial t} V(t,x) = \nabla \cdot \big(V(t,x)\nabla U(x)\big) + \frac{1}{2}\sigma^2 \Delta V(t,x).$$

Lemma 5.1 *For noise intensity σ*

$$\big\| V^\sigma(x,t) - \pi_\sigma(x) \big\| \le \exp\big(t\lambda_2(\sigma)\big),$$

where $\lambda_2(\sigma)$ is the second eigenvalue of $L_\sigma = \frac{1}{2}\sigma^2\Delta - \nabla U \cdot \nabla$ and

$$\lambda_2(\sigma) \sim -\exp\left(-\frac{c_0}{\sigma}\right),$$

where c_0 is the previous constant.

Let $V^\sigma(x,t)$ denote the density associated with the standard simulated annealing method. We recall that $V_j^\sigma(x,t)$ is the j fluid solution for the dynamic system associated to interactive annealing, i.e.,

$$\begin{cases} \dfrac{\partial}{\partial t} V_i(t,x) = \dfrac{1}{2\epsilon^2}\sigma_i^2 \Delta V_i(t,x) + \dfrac{1}{\epsilon^2}\nabla \cdot \big(V_i(t,x)\nabla U(x)\big) \\ \dfrac{\partial}{\partial t} V_j(t,x) = \dfrac{1}{2}\sigma^2 \Delta V_j(t,x) + \nabla \cdot \big[V_j(t,x)\big(\nabla U(x) + k\nabla V_i(t,x)\big)\big]. \end{cases} \tag{10}$$

Proposition 5.1 *For cooling schedule $\sigma \sim c/\log t$, with $c > c_0 > \tilde{c}_0$, fluid j of interactive diffusion $V_j^\sigma(x,t)$ has faster speed of convergence to global optima than standard annealing $V^\sigma(x,t)$.*

Proof Let S^* and $\tilde{S}^*$ denote the set of global minima of U and $\tilde{U}$, respectively. Let $B_1(\epsilon) = \{x \in \mathbb{R}^n \mid d(x,S^*) < \epsilon\}$ and $B_2(\epsilon) = \{x \in \mathbb{R}^n \mid d(x,\tilde{S}^*) < \epsilon\}$. For $\delta > 0$ and $\epsilon > 0$ with $\int_{B_1(\epsilon)} dx = \int_{B_2(\epsilon)} dx > e\delta/2$, there exists $\sigma > 0$ such that

$$\min\left\{\int_{B_1(\epsilon)} \pi_\sigma(x)\,dx,\ \int_{B_2(\epsilon)} \tilde{\pi}_\sigma(x)\,dx\right\} > 1 - \delta/2.$$

From Lemma 5.1, we have

$$\left\| V^{\sigma}(x,t) - \pi_{\sigma}(x) \right\| \le \exp\left[-t^{(1-\frac{c_0}{c})}\right], \qquad \left\| V_j^{\sigma}(x,t) - \tilde{\pi}_{\sigma}(x) \right\| \le \exp\left[-t^{(1-\frac{\tilde{c}_0}{c})}\right].$$

It follows that

$$\begin{aligned}\int_{B_1(\epsilon)} \pi_{\sigma}(x)\,dx - \int_{B_1(\epsilon)} V^{\sigma}(x,t)\,dx &\le \int_{B_1(\epsilon)} \left\| V^{\sigma}(x,t) - \pi_{\sigma}(x) \right\| dx \\ &\le \exp\left[-t^{(1-\frac{c_0}{c})}\right] \int_{B_1(\epsilon)} dx,\end{aligned}$$

hence

$$\int_{B_1(\epsilon)} V^{\sigma}(x,t)\,dx \ge \int_{B_1(\epsilon)} \left(\pi_{\sigma}(x) - \exp\left[-t^{(1-\frac{c_0}{c})}\right]\right) dx.$$

Similarly, we obtain

$$\int_{B_2(\epsilon)} V_j^{\sigma}(x,t)\,dx \ge \int_{B_2(\epsilon)} \left(\tilde{\pi}_{\sigma}(x) - \exp\left[-t^{(1-\frac{\tilde{c}_0}{c})}\right]\right) dx.$$

Define $t_1(\delta)$ and $t_2(\delta)$ as

$$t_1(\delta) := \left[-\log\left(\frac{\delta/2}{\int_{B_1(\epsilon)} dx}\right)\right]^{\frac{c}{c-c_0}}, \qquad t_2(\delta) := \left[-\log\left(\frac{\delta/2}{\int_{B_2(\epsilon)} dx}\right)\right]^{\frac{c}{c-\tilde{c}_0}}.$$

It follows that, for $t > t_1(\delta)$,

$$\int_{B_1(\epsilon)} V^{\sigma}(x,t)\,dx \ge 1 - \delta,$$

and, for $t > t_2(\delta)$,

$$\int_{B_2(\epsilon)} V_j^{\sigma}(x,t)\,dx \ge 1 - \delta.$$

Finally, from $c > c_0 > \tilde{c}_0 > 0$, we get $\frac{c}{c-c_0} > \frac{c}{c-\tilde{c}_0}$. Together with

$$\int_{B_1(\epsilon)} dx = \int_{B_2(\epsilon)} dx > e\delta/2,$$

which implies

$$\left[-\log\left(\frac{\delta/2}{\int_{B_1(\epsilon)} dx}\right)\right] = \left[-\log\left(\frac{\delta/2}{\int_{B_2(\epsilon)} dx}\right)\right] > 1,$$

we have

$$t_1(\delta) = \left[-\log\left(\frac{\delta/2}{\int_{B_1(\epsilon)} dx}\right)\right]^{\frac{c}{c-c_0}} > \left[-\log\left(\frac{\delta/2}{\int_{B_2(\epsilon)} dx}\right)\right]^{\frac{c}{c-\tilde{c}_0}} = t_2(\delta). \qquad \square$$

6 Illustration: Numerical Experiments

In order to complement the theoretical results obtained, we now present an illustration of the application of interactive diffusions to global optimization. We use Ackley's problem [16] and the extended Michalewicz function [17] to compare the performance of interactive annealing with the standard simulated annealing method. It can be verified that these two problems satisfy our standing assumptions (1) to (4).

6.1 Ackley Problem

Ackley's problem [16] is to find $x \in \mathbb{R}^n$, with $x_i \in (-32.768, 32.768)$, that minimizes the following function:

$$H(x) = -20 \cdot \exp\left(-0.2 \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n} \cdot \sum_{i=1}^{n} \cos(2\pi x_i)\right) + 20 + \exp(1).$$

When $n = 3$, the global minimum for the Ackley function is $x_* = (0, 0, 0)$ with function value $y_* = 0$, There are many local minima for this problem. Moreover, all local minima have function values $y_i > 2.1$. Hence, the margin between global and local minima satisfies $M \geq 2.5$, and for any x with function value $H(x) < 2.5$, x is in the attraction basin of the global minimum x_*.

We set $kC_0 = 1 < M$ for 3-dimension Ackley function, use constant temperature $T = 5$ a ball of radius $\rho = 0.118$ to ensure that $H(x) \leq 0.5$ for all $x \in B_\rho(x_*)$.

We compare the performance of three methods:

(1) standard simulated annealing (single thread) annealing;
(2) independent diffusion (two independent, parallel threads, one running the standard simulated annealing and another with constant temperature diffusion); and
(3) interactive diffusion (i.e. parallel annealing processes interacting with constant temperature diffusion in fast time scale).

We set constant temperature $T = 5$ for both the constant temperature diffusion in the second method (independent diffusion) and fluid i in the interactive diffusions approach.

To compare the speed of convergence, we report the average time (over 500 runs) to enter the ρ-neighborhood of the global minima for different temperature settings (see Table 1). For all three methods, we fixed initial point at (20, 20, 20). The reported number of iterations for independent diffusion is the minimum number of iterations

Table 1 Average iterations before entering ρ-neighborhood of x_* for 3-D Ackley function

Initial Temperature c	7.5	9	10.5	12	13.5	15
Standard Annealing	4696	5118	6150	6862	8336	9587
Independent Diffusion	4555	4908	5843	6392	7603	8842
Interactive Diffusion	4025	4281	4596	5188	6092	6699

Table 2 Probability for reaching global optima within 10^4 iterations for 3-D Ackley function

Initial Temperature c	7.5	9	10.5	12	13.5	15
Standard Annealing	93.4 %	91.6 %	84.0 %	77.6 %	68.4 %	63.0 %
Independent Diffusion	93.8 %	93.0 %	85.8 %	80.6 %	73.6 %	67.0 %
Interactive Diffusion	95.8 %	97.2 %	93.2 %	89.4 %	83.2 %	80.6 %

Table 3 Average iterations before entering ρ-neighborhood of x_* for 5-D Ackley function

Initial Temperature c	3	4.5	6	7.5	9	10.5
Standard Annealing	7279	18129	35733	52951	72908	79332
Independent Diffusion	7270	18089	35484	52418	72019	78404
Interactive Diffusion	3879	10401	21111	34508	52557	61078

Table 4 Probability for reaching global optima within 10^5 iterations for 5-D Ackley function

Initial Temperature c	3	4.5	6	7.5	9	10.5
Standard Annealing	100.0 %	100.0 %	97.0 %	83.0 %	56.2 %	42.0 %
Independent Diffusion	100.0 %	100.0 %	97.0 %	84.0 %	57.8 %	43.2 %
Interactive Diffusion	100.0 %	100.0 %	99.8 %	97.6 %	83.6 %	71.6 %

needed to enter the ρ-neighborhood of global minima by either the standard simulated annealing or the constant temperature annealing thread. The data supports the theoretical results regarding a speedier identification of global minima by the interactive diffusion method. This effect is more pronounced with lower initial temperatures. However, as we shall see below, the finite-time performance of the interactive annealing method is superior for higher initial temperature.

To evaluate finite-time performance (in this 3-dimension Ackley function) we run the three methods with a limit of 10^4 iterations. We report the probability (i.e. the empirical frequency over 500 runs) with which the three methods reached the ρ-neighborhood of the global minima (see Table 2)

We now test the implications of increasing the dimensionality of the problem. When $n = 5$, the global minimum for the Ackley function is $x_* = (0, 0, 0, 0, 0)$ with function value $y_* = 0$, and all local minima have function value $y_i > 1.64$. We set $kC_0 = 1.2 < M$ for 5-dimension Ackley function, use constant temperature $T = 5$ a ball of radius $\rho = 0.176$ (again to ensure that $H(x) \leq 0.6$ for all $x \in B_\rho(x_*)$). We report the average time needed (over 500 runs) to enter the ρ-neighborhood of the global minima for different temperature settings (see Table 3). For all processes of three methods, we fixed initial point at $(5, 5, 5, 5, 5)$.

Clearly, the gains in convergence speed by the interactive diffusion method are made even more apparent in higher dimensions. To evaluate finite-time performance (in this 5-dimension Ackley function), we report the probability (i.e. the empirical frequency over 500 runs), with which the three methods reached the ρ-neighborhood of the global minima (see Table 4).

6.2 Rastrigin Problem

The Rastrigin problem is the minimization of the function defined as follows:

$$F(x) := 10n + \sum_{i=1}^{n} \left[x_i^2 - 10\cos(2\pi x_i)\right] \quad x_i \in [-5.12, 5.12], i = 1, 2, \ldots, n.$$

When $n = 3$, the global minimum for the Rastrigin function is $x_* = (0, 0, 0)$ with function value $y_* = 0$. All the local minima have function values $y_i > 0.95$. The margin M (difference between function value of global minima and local minima) verifies $M \geq 0.95$. In our implementation of the interactive diffusions approach, we have fluid i diffusing with constant temperature $T = 5$ in a fast time scale.

Here again, as in the previous section we compare the speed of convergence of

(1) standard simulated annealing (single thread) annealing;
(2) independent diffusion (two independent, parallel threads, one running the standard simulated annealing and another with constant temperature diffusion); and
(3) interactive diffusion (i.e. parallel annealing processes interacting with constant temperature diffusion in fast time scale).

We set the same temperature $T = 5$ for both constant temperature diffusions in methods (2) and (3). We choose a ball of radius $\rho = 0.05$ that $H(x) \leq 0.5$ for all $x \in B_\rho(x_*)$. We report the average time (over 500 runs) to enter the ρ-neighborhood of the global minima for different temperature settings (see Table 5). For all processes of three methods, we fixed initial point at $(5, 5, 5)$.

To evaluate finite-time performance (in this 3-dimension Rastrigin function), we report the probability that process reached global minima ρ-neighborhood within 10^5 (see Table 6).

Table 5 Average iterations before entering ρ-neighborhood of x_* for 3-D Rastrigin function

Initial Temperature c	7.5	9	10.5	12	13.5	15
Standard Annealing	42170	11900	18458	39301	68866	77824
Independent Diffusion	39785	11658	18260	38177	64837	73342
Interactive Diffusion	23251	8885	14675	30855	54646	64104

Table 6 Probability for reaching global optima within 10^5 iterations for 3-D Rastrigin function

Initial Temperature c	7.5	9	10.5	12	13.5	15
Standard Annealing	62.2 %	99.6 %	100.0 %	99.6 %	70.6 %	33.4 %
Independent Diffusion	66.0 %	99.8 %	100.0 %	99.6 %	73.8 %	40.6 %
Interactive Diffusion	82.8 %	100.0 %	100.0 %	100.0 %	90.6 %	53.2 %

Table 7 Probability for two trivial paralleled annealing threads reaching global optima within iterations cap

Initial Temperature c	7.5	9	10.5	12	13.5	15
3-D Ackley Func., cap 10^4	99.6 %	99.3 %	97.4 %	95.0 %	90.0 %	86.3 %
Initial Temperature c	3	4.5	6	7.5	9	10.5
5-D Ackley Func., cap 10^5	100.0 %	100.0 %	99.9 %	97.1 %	80.8 %	66.4 %
Initial Temperature c	7.5	9	10.5	12	13.5	15
3-D Rastrigin Func., cap 10^5	85.7 %	100.0 %	100.0 %	100.0 %	91.4 %	55.6 %

6.2.1 Discussion

The limited computational evidence presented in this section confirms the theoretical prediction of the interactive diffusion approach's faster speed of convergence. The results suggest that the relative gains in convergence speed are increasing with decreasing initial temperatures. It is often the case that practitioners using simulated annealing would attempt to reduce the value of the initial temperature as this parameter serves as a proxy for available computational budget. In this sense, the proposed method works best when there is a limited computational budget (in the form of a relatively low value for initial temperature). The finite-time performance evaluation is conducted by imposing a cap on the number of iterations. Here again, the interactive annealing method proposed exhibits a higher probability of finding the global optima.

6.3 Comparison with Parallel Simulated Annealing

We now compare the performance of independent parallel implementations of simulated annealing suggested in [18–20]. Let P_{c_i} be probability of single annealing threads reaching global optima within iteration cap for initial temperature c_i. Consider two annealing threads, let Q_{c_i} be the probability of one of paralleled annealing threads entering global optima. We have

$$Q_{c_i} = 1 - (1 - P_{c_i})^2,$$

the result are shown in Table 7.

With only two threads, independent simulated annealing processes may outperform the interactive diffusions approach in certain cases. This can be explained as follows. In the interactive diffusions approach with two threads, one thread is "sacrificed" by having constant diffusion in order to speed up the interactive annealing thread. With only two threads, the opportunity cost of running a constant diffusion thread to speed up a second diffusion is too high. However, with more reactive threads this cost is diluted. To make this point we now present the simulation results with five (5) threads (one (1) constant diffusion and four (4) interactive diffusion threads) and compare with the performance of five (5) independent simulated annealing processes. The performance is evaluated on the basis of (i) average iterations before first thread enters ρ-neighborhood of x_* in Tables 8, 9 and 10 and (ii) the probability of 1 of 5

Table 8 Average iterations before one of five threads enter ρ-neighborhood of x_* for 3-D Ackley function

Initial Temperature c	7.5	9	10.5	12	13.5	15
Paralleled Annealing	1916	2093	2207	2364	2806	3138
Interactive Diffusion	1770	1887	1971	2270	2429	2559

Table 9 Average iterations before one of five threads enter ρ-neighborhood of x_* for 5-D Ackley function

Initial Temperature c	3	4.5	6	7.5	9	10.5
Paralleled Annealing	2399	5881	11049	17640	27869	41975
Interactive Diffusion	1645	4072	7539	13222	20178	27819

Table 10 Average iterations before one of five threads enter ρ-neighborhood of x_* for 3-D Rastrigin function

Initial Temperature c	7.5	9	10.5	12	13.5	15
Paralleled Annealing	3403	4450	8911	17767	32681	30926
Interactive Diffusion	2251	3860	8124	15718	22697	20783

Table 11 Probability for one of five threads reaching global optima within 5000 iterations for 3-D Ackley function

Initial Temperature c	7.5	9	10.5	12	13.5	15
Paralleled Annealing	98.8 %	98.6 %	96.8 %	94.6 %	88.8 %	84.8 %
Interactive Diffusion	100.0 %	99.4 %	97.6 %	96.8 %	93.2 %	91.8 %

Table 12 Probability for one of five threads reaching global optima within 10^4 iterations for 5-D Ackley function

Initial Temperature c	3	4.5	6	7.5	9	10.5
Paralleled Annealing	100.0 %	83.6 %	51.8 %	33.6 %	19.0 %	12.2 %
Interactive Diffusion	100.0 %	96.2 %	73.4 %	45.8 %	29.6 %	23.0 %

threads entering global optima within the total number of iterations in Tables 11, 12 and 13.

With five (5) threads, we observe that the interactive diffusions approach exhibits better performance than independent simulated annealing processes. This is further indication of the proposed method's improved speed of convergence.

Table 13 Probability for one of five threads reaching global optima within 10^4 iterations for 3-D Rastrigin function

Initial Temperature c	7.5	9	10.5	12	13.5	15
Paralleled Annealing	96.7 %	98.0 %	61.3 %	21.3 %	31.0 %	57.0 %
Interactive Diffusion	100.0 %	99.7 %	69.3 %	26.0 %	44.7 %	67.7 %

7 Conclusions

Simulated annealing for global optimization is a well-known method for global optimization. Its main drawback pertains to the speed at which emphasis on "exploration" gives way to "exploitation". This transition cannot be "too fast" in order to guarantee convergence to globally optimal solutions. Evidently, with faster computing this drawback may become less critical. However, computing speed (which used to double every couple of years) has stopped increasing because as chips are clocked at higher speeds they become difficult to cool and much less energy-efficient. Parallel implementation of annealing-like search for global optima appears as a sensible approach for speeding up convergence to globally optimal solutions.

In this paper, we have introduced a novel approach, in which parallel annealing processes interact in a manner that expedites the identification of a globally optimal solution. A first annealing process operates at a faster time scale and has a drift function that converges to a non-zero (but relatively small) noise level. A second annealing process (operating a slower time scale) is subject to a modified drift term, in which the steepest descent direction is perturbed with the first process density gradient. This additional term ensures that the second annealing process is "repelled" from areas explored by the first process. As a result, the second annealing process is able to *bypass* locally optimal solutions that its own "cooling schedule" can be decreased. We have shown that when compared to independent diffusions, the proposed interactive diffusions approach can increase the speed of convergence at the expense of the minimal additional computational overhead.

In limited computational testbed, we provide numerical illustration of the (theoretical) speed-up effect. These numerical experiments suggest that the relative gains in convergence speed are increasing with decreasing initial temperatures. Since the initial temperature often serves as proxy for the available computational budget, the proposed method works best when there is a limited computational budget in the form of a relatively low value for initial temperature. The finite-time performance evaluation also consistently indicates that the interactive annealing method exhibits a higher probability of finding the global optima in finite time. Finally, the numerical illustrations suggest that the relative performance of interactive annealing improves with higher dimensions.

There are a number of outstanding research questions, which we continue to explore. First, in this paper we have limited our analysis to the case of *two* interacting annealing processes. It would be desirable to characterize the extent to which faster convergence to globally optimal solutions is obtained as a function of *several* interacting annealing processes. Given the computational overhead involved, this type of analysis could help identify the "optimal" number of interacting annealing processes.

Secondly, in this paper, we have only considered one form of interaction amongst annealing processes, namely, one in which a second (slower) annealing process is repelled by the density of the first (faster) process. In addition to a repulsive potential, the incorporation of an attractive potential in the differential equations governing the interaction across annealing process, could conceivably yield a speedier exploration of neighborhoods of promising solutions. In a manner analogous to the results obtained in this paper, this may yield faster identification of globally optimal solutions.

Acknowledgements This work was partially funded by Air Force Office of Scientific Research (AFOSR) through grant FA9550-12-1-0163.

References

1. Cerný, V.: Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. J. Optim. Theory Appl. **45**(1), 41–51 (1985)
2. Kirkpatrick, S., Gelatt, C.D. Jr., Vecchi, M.P.: Optimization by simulated annealing. Science **220**, 621–680 (1983)
3. Geman, S., Hwang, C.R.: Diffusions for global optimization. SIAM J. Control Optim. **24**, 1031–1043 (1986)
4. Chiang, T.S., Hwang, C.R., Sheu, S.J.: Diffusion for global optimization in R^n. SIAM J. Control Optim. **25**, 737–752 (1987)
5. Kushner, H.J.: Asymptotic global behavior for stochastic approximation and diffusions with slowly decreasing noise effects: global minimization via Monte Carlo. SIAM J. Appl. Math. **47**, 169–185 (1987)
6. Hwang, C.R., Hwang-Ma, S.Y., Sheu, S.J.: Accelerating Gaussian diffusions. Ann. Appl. Probab. **3**(3), 897–913 (1993)
7. Yin, G., Yin, K.: Global optimization using diffusion perturbations with large noise intensity. Acta Math. Appl. Sin., Engl. Ser. **22**(4), 529–542 (2006)
8. Poliannikov, O., Zhizhina, E., Krim, H.: Global optimization by adapted diffusion. IEEE Trans. Signal Process. **58**(12), 6119–6125 (2010)
9. Coleman, T., Shalloway, D., Wu, Z.: A parallel build-up algorithm for global energy minimizations of molecular clusters using effective energy simulated annealing. J. Glob. Optim. **4**, 171–186 (1994)
10. Wu, Z.: The effective energy transformation scheme as a special continuation approach to global optimization with application to molecular conformation. SIAM J. Optim. **6**, 748–768 (1996)
11. Lau, M., Kwong, C.P.: A smoothing method of global optimization that preserves global minima. J. Glob. Optim. **34**, 369–398 (2006)
12. Khasminskii, R.Z., Yin, G.: Limit behavior of two-time scale diffusions revisited. J. Differ. Equ. **212**, 85–113 (2005)
13. Risken, H.: The Fokker–Planck Equation (Methods of Solution and Applications), 2nd edn. Springer Series in Synergetics. Springer, Berlin (1989). 1989
14. Pham, H.: Continuous-Time Stochastic Control and Optimization with Financial Applications. Stochastic Modelling and Applied Probability, vol. 61. Springer, Berlin (2009)
15. Gardiner, C.W.: Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences, 2nd edn. Springer Series in Synergetics, vol. 13. Springer, Berlin (1985)
16. Ackley, D.H.: A Connectionist Machine for Genetic Hillclimbing. Kluwer Academic Publishers, Boston (1987)
17. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. Springer, Berlin (1992)
18. Gomes, C.P., Selman, B., Kautz, H.: Boosting combinatorial search through randomization. In: Proceedings of the National Conference on Artificial Intelligence, pp. 431–437. Wiley, New York (1998)
19. Luby, M., Ertel, W.: Optimal Parallelization of Las Vegas Algorithms. Springer, Berlin (1994)
20. Shylo, O.V., Middelkoop, T., Pardalos, P.M.: Restart strategies in optimization: parallel and serial cases. Parallel Comput. **37**(1), 60–68 (2011)