

On Simulated Annealing with Temperature-Dependent Energy and Temperature-Dependent Communication

Marc C. Robini^a, Pierre-Jean Reissman^b

^aCREATIS (CNRS research unit UMR 5220; INSERM research unit U1044), INSA Lyon, 69621 Villeurbanne, France

^bSales and e-commerce platforms division, Amadeus SAS, 485 route du Pin Montard, 06902 Sophia Antipolis, France

Abstract

Simulated annealing (SA) is a generic optimization method that is quite popular because of its ease of implementation and its optimal convergence properties. Still, SA is widely reported to converge very slowly and it is common practice to allow extra freedom in its design at the expense of losing global convergence guarantees.

In this paper, we derive simple sufficient conditions for the global convergence of SA when the cost function and the candidate solution generation mechanism are temperature-dependent. These conditions are surprisingly weak — they do not involve the variations of the cost function with temperature — and exponential cooling makes it possible to be arbitrarily close to the best possible convergence exponent of standard SA.

Keywords: Combinatorial optimization, simulated annealing, Markov chains, Monte Carlo methods

2010 MSC: 82C80, 65C05, 60J10, 90C27

1. Introduction

We consider the problem of finding a global minimum of an arbitrary real-valued energy function U defined on a general but finite state space E . We denote by U_{\inf} the ground state energy and we let $E_{\inf}(U)$ be the set of global minima of U , that is, $U_{\inf} = \inf_{x \in E} U(x)$ and $E_{\inf}(U) = \{x \in E \mid U(x) = U_{\inf}\}$. A simulated annealing (SA) algorithm is a Markov chain $(X_n)_{n \in \mathbb{N}}$ on E whose transitions are guided by a communication mechanism q and controlled by a sequence of temperatures $(\tau_n)_{n \in \mathbb{N}^*}$ called a cooling schedule; the communication mechanism is a symmetric and irreducible Markov matrix on E which specifies how to generate a new candidate solution from the current solution, and the cooling schedule is decreasing and converges to zero. The transitions of $(X_n)_{n \in \mathbb{N}}$ are defined by $P(X_n = y \mid X_{n-1} = x) = P_{\tau_n}(x, y)$ with

$$P_{\tau}(x, y) = \begin{cases} q(x, y) \exp\left(-\tau^{-1}(U(y) - U(x))^+\right) & \text{if } y \neq x \\ 1 - \sum_{z \in E \setminus \{x\}} P_{\tau}(x, z) & \text{if } y = x, \end{cases} \quad (1)$$

where $a^+ := \sup\{a, 0\}$. The key feature of the Markov matrix P_{τ} is that its stationary distribution π_{τ} tends to the uniform distribution on $E_{\inf}(U)$ as the temperature τ decreases to zero (π_{τ} is the Gibbs distribution with energy U at temperature τ). Consequently, the law of X_n should stay close to π_{τ_n} if the cooling schedule decreases slowly enough, and hence we can expect that the convergence measure

$$\mathcal{M}(n) = \sup_{x \in E} P(X_n \notin E_{\inf}(U) \mid X_0 = x) \quad (2)$$

goes to zero as $n \rightarrow +\infty$. It is well known (see Hajek (1988)) that this is the case for logarithmic cooling schedules of the form $\tau_n = \tau_0 / \ln(n + 1)$ provided τ_0 is larger than the critical height H_U of the energy landscape. Formally, H_U is the maximum energy barrier separating a non-optimal state from a ground state, that is,

$$H_U = \sup_{x \in E \setminus E_{\inf}(U)} h_U(x), \quad (3)$$

where $h_U(x)$ — the depth of x — is defined as follows:

$$h_U(x) = \inf_{y \in E_{\text{inf}}(U)} \eta_U(x, y) - U(x) \quad \text{with} \quad \eta_U(x, y) = \inf_{(x_i)_{i=1}^m \in \Pi_q(x, y)} \sup_{1 \leq i \leq m} U(x_i), \quad (4)$$

where $\Pi_q(x, y)$ denotes the set of paths from x to y in the digraph

$$(E, \Delta(q)), \quad \Delta(q) = \{(x, y) \in E^2 \mid y \neq x \text{ and } q(x, y) > 0\}. \quad (5)$$

Logarithmic cooling, however, is inefficient for most practical problems: H_U is generally too large to reach the low temperature regime in a reasonable amount of computation time, while the annealing process gets easily stuck in poor local minima for feasible values of τ_0 .

The most significant advance in SA theory beyond the asymptotic properties in Hajek (1988) is due to Catoni. He showed in Catoni (1992) that exponential cooling must be preferred over logarithmic cooling when the available computing time is bounded (as is the case in practice), and that the convergence measure (2) cannot decrease faster than some optimal power of $1/n$. More precisely, the optimal convergence speed exponent is $1/D_U$, where D_U — the difficulty of the energy landscape — is the maximum ratio of the depth to the energy level above the ground state energy:

$$D_U = \sup_{x \in E \setminus E_{\text{inf}}(U)} \frac{h_U(x)}{U(x) - U_{\text{inf}}}. \quad (6)$$

Besides, the upper bound $1/D_U$ is sharp, since it is possible to construct a family $\{(\tau_n^N)_{1 \leq n \leq N}; N \in \mathbb{N}^*\}$ of finite exponential cooling sequences such that $\ln \mathcal{M}(N) \sim \ln(N^{-1/D_U})$; such families are of the form $\tau_n^N = \tau_0 \exp(-n \xi_N)$, where τ_0 does not depend on the horizon N and $\xi_N \sim N^{-1} \ln N$. These results provide the first theoretical justification for the widely used exponential cooling schedules. Still, although successfully applied to many difficult combinatorial optimization problems, SA is often criticized for converging very slowly. In fact, it is common practice to allow extra freedom in the design of SA algorithms, but such variations on the theme generally come without optimal convergence guarantees.

A natural generalization of SA is to allow the energy function and the communication mechanism to be temperature-dependent. We call this class of algorithms stochastic continuation (SC) by extension of the stochastic optimization processes studied in Robini et al. (2007). The first idea is to ease the annealing process by gradually revealing the complexity of the optimization problem, which is obtained by replacing the energy U by the elements of a family $(U_\tau)_{\tau \in \mathbb{R}_+^*}$ of functions whose difficulty D_{U_τ} increases with decreasing τ . The second idea is to facilitate the exploration of the state space by adapting the communication mechanism to the temperature regime. SC belongs to the general class of Markov processes studied in Del Moral and Miclo (1999); it includes SA with temperature-dependent energy, which is studied in Robini et al. (2007) for the finite-time case and in Frigerio and Grillo (1993) and Löwe (1996) for the asymptotic case. The convergence results in Robini et al. (2007) and Del Moral and Miclo (1999) require that

$$\sup_{(x, \tau) \in E \times \mathbb{R}_+^*} \tau^{-1} |U_\tau(x) - U(x)| < +\infty, \quad (7)$$

while it is assumed in Frigerio and Grillo (1993) and Löwe (1996) that there exists $a > 0$ such that

$$\sup_{(x, n) \in E \times \mathbb{N}^*} n^a |U_{\tau_n}(x) - U_{\tau_{n-1}}(x)| < +\infty. \quad (8)$$

Both conditions (7) and (8) significantly limit the freedom in parameterizing the energy with temperature, and, in addition, the conditions for convergence in Del Moral and Miclo (1999), Frigerio and Grillo (1993) and Löwe (1996) involve impractical logarithmic cooling sequences. We show here that these limitations can be overcome while allowing the communication mechanism to vary with temperature.

The formalism and the basic ideas of SC are presented in Section 2. Our starting point for studying the convergence of SC is the observation that its transitions obey a large deviation principle with speed τ^{-1} , which suggests to appeal to generalized SA (GSA) theory (see, e.g., Trouvé (1996) and Catoni (1999)). The convergence properties of GSA are outlined in Section 3, and those of SC are then derived in Section 4. Our main result states that SC with suitably adjusted exponential cooling can have a convergence speed exponent arbitrarily close to the optimal exponent $1/D_U$ of SA. Moreover, the conditions for this to happen are weak and only involve the communication mechanism (putting aside the obvious necessary condition that $(U_\tau)_{\tau \in \mathbb{R}_+^*}$ converges pointwise to the target energy U as $\tau \rightarrow 0$).

2. Definition and basic ideas

We define an SC process with target energy landscape (E, U, q) to be a family $(Q_\tau)_{\tau \in \mathbb{R}_+^*}$ of Markov matrices on E of the form

$$Q_\tau(x, y) = \begin{cases} q_\tau(x, y) \exp(-\tau^{-1}(U_\tau(y) - U_\tau(x))^+) & \text{if } y \neq x \\ 1 - \sum_{z \in E \setminus \{x\}} Q_\tau(x, z) & \text{if } y = x, \end{cases} \quad (9)$$

with $\lim_{\tau \rightarrow 0} U_\tau(x) = U(x)$ and $\lim_{\tau \rightarrow 0} q_\tau(x, y) = q(x, y)$.

Given such a family together with a cooling sequence $(\tau_n)_{n \in \mathbb{N}^*}$, we call a Markov chain $(X_n)_{n \in \mathbb{N}}$ on E with transitions $P(X_n = y | X_{n-1} = x) = Q_{\tau_n}(x, y)$ an SC algorithm, and we denote it by $\text{SC}(E, (U_\tau), (q_\tau), (\tau_n))$ — we use the notation $(E, (Q_\tau))$ or $(E, (U_\tau), (q_\tau))$ for the underlying SC process. The family of energy functions $(U_\tau : E \rightarrow \mathbb{R})_\tau$ is called the continuation scheme and the family of Markov matrices $(q_\tau : E^2 \rightarrow [0, 1])_\tau$ is called the communication scheme. The limit communication matrix q is assumed to be irreducible, as there is otherwise no guarantee to reach a ground state of the target energy U .

The basic idea of SC is quite easy to explain if q_τ is symmetric for all τ (we will relax this assumption in Section 4). Indeed, Proposition 1 below states that in this case the invariant measure θ_τ of Q_τ is a Gibbs distribution which concentrates on the set of global minima of U as $\tau \rightarrow 0$. Consequently, similarly to SA, if the cooling sequence does not decrease too fast, the law of X_n should stay close enough to θ_{τ_n} to expect convergence to an optimum.

Proposition 1. *Let $(E, (Q_\tau))$ be an SC process with q irreducible and q_τ symmetric for all τ . Then, there exists $\tau^* > 0$ such that for any $\tau \in (0, \tau^*]$, Q_τ is irreducible and its unique invariant measure θ_τ satisfies $\theta_\tau(x) \propto \exp(-\tau^{-1}U_\tau(x))$ and $\lim_{\tau \rightarrow 0} \theta_\tau(E_{\text{inf}}(U)) = 1$.*

PROOF. For τ sufficiently small, q_τ inherits the irreducibility of q and hence Q_τ is irreducible. By the symmetry of q_τ , we have

$$\forall (x, y) \in E^2, \quad \exp(-\tau^{-1}U_\tau(x))Q_\tau(x, y) = \exp(-\tau^{-1}U_\tau(y))Q_\tau(y, x),$$

and therefore $\theta_\tau(x) \propto \exp(-\tau^{-1}U_\tau(x))$.

Let $Z_\tau(F) = \sum_{z \in F} \exp(-\tau^{-1}U_\tau(z))$ and let $\omega_\tau = \sup_{z \in E_{\text{inf}}(U)} U_\tau(z)$. We have

$$\lim_{\tau \rightarrow 0} \frac{Z_\tau(E \setminus E_{\text{inf}}(U))}{Z_\tau(E_{\text{inf}}(U))} = 0,$$

and it follows that

$$\forall x \in E, \quad \theta_\tau(x) \in O\left(\exp(-\tau^{-1}(U_\tau(x) - \omega_\tau))\right).$$

Then, since $\lim_{\tau \rightarrow 0} (U_\tau(x) - \omega_\tau) = U(x) - U_{\text{inf}}$, we have $\lim_{\tau \rightarrow 0} \theta_\tau(x) = 0$ for any $x \notin E_{\text{inf}}(U)$. \square

Proposition 1 gives the go-ahead for studying the global convergence properties of SC. To do so, we start from the basic observation that SC and SA behave similarly at low temperatures in the sense that

$$\forall (x, y) \in \Delta(q), \quad \lim_{\tau \rightarrow 0} -\tau \ln Q_\tau(x, y) = \lim_{\tau \rightarrow 0} -\tau \ln P_\tau(x, y) = (U(y) - U(x))^+. \quad (10)$$

In other words, for any $(x, y) \in E^2$ such that $y \neq x$ and $q(x, y) > 0$, $Q_\tau(x, y)$ obeys a large deviation principle with speed τ^{-1} and rate $(U(y) - U(x))^+$, which suggests to appeal to the GSA theory developed in Trounev (1996) and Catoni (1999).

3. Generalized simulated annealing

A GSA process on E is defined by a family $(\Theta_\tau)_{\tau \in \mathbb{R}_+^*}$ of Markov matrices on E satisfying a large deviation assumption with speed τ^{-1} and with irreducible rate function $J : E^2 \rightarrow \mathbb{R}_+ \cup \{+\infty\}$; that is,

$$\forall (x, y) \in E^2, \quad \lim_{\tau \rightarrow 0} -\tau \ln \Theta_\tau(x, y) = J(x, y) \quad (11)$$

(with the convention that $\ln 0 = -\infty$), and the digraph

$$(E, \Delta(J)), \quad \Delta(J) = \{(x, y) \in E^2 \mid y \neq x \text{ and } J(x, y) < +\infty\} \quad (12)$$

is strongly connected. Given a GSA process $(E, (\Theta_\tau))$ and a cooling sequence $(\tau_n)_n$, we call a Markov chain $(X_n)_{n \in \mathbb{N}}$ on E with transitions $P(X_n = y \mid X_{n-1} = x) = \Theta_{\tau_n}(x, y)$ a GSA algorithm; we denote it by $\text{GSA}(E, (\Theta_\tau), (\tau_n))$.

The Markov matrix Θ_τ has a unique invariant measure ϑ_τ for τ sufficiently small, and it is shown in Catoni (1999) that $(\vartheta_\tau)_\tau$ satisfies a large deviation principle: there is a function $V : E \rightarrow \mathbb{R}_+$, called the virtual energy, such that

$$\forall x \in E, \quad \lim_{\tau \rightarrow 0} -\tau \ln \vartheta_\tau(x) = V(x) - V_{\inf}, \quad (13)$$

where $V_{\inf} = \inf_{x \in E} V(x)$. Clearly, ϑ_τ concentrates on the set of global minima of V as $\tau \rightarrow 0$, which is to say that the virtual energy plays the role of the objective function. It is defined by

$$\forall x \in E, \quad V(x) = \inf_{T \in \mathcal{T}(x)} \sum_{(z,t) \in \mathcal{E}_T} J(z, t), \quad (14)$$

where $\mathcal{T}(x)$ is the set of directed trees $T = (E, \mathcal{E}_T)$, $\mathcal{E}_T \subset E^2$, with root x and whose edges are directed toward x (i.e., $d^+(x) = 0$ and $d^+(y) = 1$ for all $y \neq x$, where $d^+(z)$ is the outdegree of vertex z in T). Similarly to standard SA theory, the triplet (E, V, J) defines an energy landscape and is accompanied with a critical depth H_V and a difficulty D_V ; these two constants are defined by

$$H_V = \sup_{x \in E \setminus E_{\inf}(V)} h_V(x) \quad \text{and} \quad D_V = \sup_{x \in E \setminus E_{\inf}(V)} \frac{h_V(x)}{V(x) - V_{\inf}}, \quad (15)$$

where $h_V(x)$ denotes the depth of x in (E, V, J) :

$$h_V(x) = \inf_{y \in E_{\inf}(V)} \eta_V(x, y) - V(x) \quad \text{with} \quad \eta_V(x, y) = \inf_{(x_i)_{i=1}^m \in \Pi_J(x, y)} \sup_{1 \leq i \leq m-1} (V(x_i) + J(x_i, x_{i+1})), \quad (16)$$

where $\Pi_J(x, y)$ is the set of paths from x to y in $(E, \Delta(J))$.

The main convergence result for GSA is stated in the following theorem; it gives an asymptotic bound on the probability of error for suitably adjusted piecewise-constant exponential cooling sequences.

Theorem 1 (Catoni (1999)). *Let $(E, (\Theta_\tau))$ be a GSA process with virtual energy landscape (E, V, J) . For any positive reals A, β and γ , consider the family of finite-time algorithms $\{(X_n)_{0 \leq n \leq \sigma K} = \text{GSA}(E, (\Theta_\tau), (\tau_n^{\sigma, K}))\}; \sigma, K \in \mathbb{N}^*\}$ with cooling sequence of the form*

$$\tau_n^{\sigma, K} = \frac{A}{\ln K} \left(\frac{\gamma}{\beta} \right)^{\frac{1}{\sigma} \lfloor \frac{n-1}{K} \rfloor}. \quad (17)$$

If $A > H_V$, $\beta > H_V/D_V$ and $\gamma < \beta$, then, for any σ ,

$$\liminf_{K \rightarrow +\infty} - \frac{\ln \sup_{x \in E} P(V(X_{\sigma K}) \geq V_{\inf} + \gamma \mid X_0 = x)}{\ln(\sigma K)} \geq \frac{1}{D_V} \left(\frac{\gamma}{\beta} \right)^{\frac{1}{\sigma}}. \quad (18)$$

4. Finite-time convergence of stochastic continuation

To apply Theorem 1 to optimization by SC, we need to find sufficient conditions for an SC process with target energy U to be a GSA process with virtual energy V such that $E_{\inf}(V) \subset E_{\inf}(U)$. This is the subject of the two propositions below. The conditions for an SC process to fit into the GSA framework only involve the communication scheme; they are given in Proposition 2. Proposition 3 gives a simple additional condition for the virtual energy to be equal to the target energy plus a constant.

Proposition 2. Let $(E, (Q_\tau))$ be an SC process with communication scheme $(q_\tau)_\tau$ and with target energy landscape (E, U, q) . Assume that

(A1) q is irreducible,

(A2) $\forall (x, y) \in E^2, q(x, y) = 0 \implies \lim_{\tau \rightarrow 0} (q_\tau(x, y))^\tau = \mathbf{1}_{\{x=y\}}$.

Then $(E, (Q_\tau))$ is a GSA process with rate function

$$J(x, y) = \begin{cases} (U(y) - U(x))^+ & \text{if } q(x, y) > 0 \text{ or } x = y \\ +\infty & \text{otherwise.} \end{cases} \quad (19)$$

PROOF. For any $(x, y) \in E^2$,

$$(x \neq y \text{ and } q(x, y) > 0) \implies -\tau \ln Q_\tau(x, y) = -\tau \ln q_\tau(x, y) + (U(y) - U(x))^+ \xrightarrow{\tau \rightarrow 0} (U(y) - U(x))^+,$$

$$(x \neq y \text{ and } q(x, y) = 0) \implies -\tau \ln Q_\tau(x, y) \geq -\tau \ln q_\tau(x, y) \xrightarrow{\tau \rightarrow 0} +\infty,$$

$$x = y \implies 0 \leq -\tau \ln Q_\tau(x, x) \leq -\tau \ln \left(1 - \sum_{z \neq x} q_\tau(x, z)\right) = -\tau \ln q_\tau(x, x) \xrightarrow{\tau \rightarrow 0} 0.$$

The irreducibility of J follows from the irreducibility of q . \square

Proposition 3. Let $(E, (Q_\tau))$ be an SC process with target energy landscape (E, U, q) and satisfying assumptions (A1) and (A2). If

(A3) $\Delta(q)$ is symmetric (i.e., $\forall (x, y) \in E^2, q(x, y) > 0 \iff q(y, x) > 0$),

then the virtual energy V of $(E, (Q_\tau))$ satisfies $V - V_{\inf} = U - U_{\inf}$.

PROOF. Let $(x, y) \in E^2$ such that $x \neq y$. Let $T \in \mathcal{T}(x)$ and let $\pi_T(x, y) = (x_i)_{i=1}^m$ be the (unique) path from x to y in T . Assume that $\pi_T(x, y)$ is J -admissible, that is, $J(x_i, x_{i+1}) < +\infty$ for all $i \in \{1, \dots, m-1\}$. Then, by (19) and (A3), $J(x_i, x_{i+1}) - J(x_{i+1}, x_i) = U(x_{i+1}) - U(x_i)$ for all i , and it follows that

$$\sum_{(z,t) \in \mathcal{E}_T: z \in \pi_T(x,y)} J(z, t) = U(x) - U(y) + \sum_{(z,t) \in \mathcal{E}_T: z \notin \pi_T(x,y)} J(z, t). \quad (20)$$

Since J is irreducible, there exists $T \in \mathcal{T}(x)$ such that $J(T) := \sum_{(z,t) \in \mathcal{E}_T} J(z, t) < +\infty$ (and hence such that $\pi_T(x, y)$ is J -admissible); consequently, using (20),

$$V(x) + U(y) = \inf_{T \in \mathcal{T}(x): J(T) < +\infty} \left(\sum_{(z,t) \in \mathcal{E}_T: z \notin \pi_T(x,y)} J(z, t) + \sum_{(z,t) \in \mathcal{E}_T: z \in \pi_T(x,y)} J(t, z) + U(x) \right).$$

Letting $\zeta: \mathcal{T}(x) \rightarrow \mathcal{T}(y)$ be the one-to-one mapping that reverses the orientation of each edge in the path from y to x , we obtain

$$V(x) + U(y) = \inf_{T \in \mathcal{T}(x): J(T) < +\infty} J(\zeta(T)) + U(x) = V(y) + U(x).$$

In particular, given $y_0 \in E_{\inf}(U)$, we have $V(x) = V(y_0) + U(x) - U_{\inf} \geq V(y_0)$ for all x , and hence $V(y_0) = V_{\inf}$. \square

Our main result is given by Theorem 2. Putting it simply, it states that increasing the number of temperature stages of piecewise-constant exponential cooling makes it possible for SC to have a convergence speed exponent arbitrarily close to the optimal convergence speed exponent of SA.

Theorem 2. Let $(E, (U_\tau), (q_\tau))$ be an SC process with target energy landscape (E, U, q) and satisfying assumptions (A1)–(A3), and let \mathcal{M} be the convergence measure defined in (2). For any $\varepsilon > 0$, there is a family of piecewise-constant cooling sequences $(\tau_n^{\sigma, K})_{1 \leq n \leq \sigma K}$ with σ stages of length K such that the family of finite-time algorithms

$$\{(X_n)_{0 \leq n \leq \sigma K} = \text{SC}(E, (U_\tau), (q_\tau), (\tau_n^{\sigma, K}))\}; \sigma, K \in \mathbb{N}^* \quad (21)$$

satisfies

$$\lim_{K \rightarrow +\infty} -\frac{\ln \mathcal{M}(\sigma K)}{\ln(\sigma K)} \geq \frac{1}{(1 + \varepsilon)D_U}. \quad (22)$$

These cooling sequences are of the form

$$\tau_n^{\sigma, K} = \frac{A}{\ln K} \exp\left(-\frac{B}{\sigma} \left\lfloor \frac{n-1}{K} \right\rfloor\right) \quad (23)$$

with $A > H_U$, $B > \ln\left(\frac{H_U}{D_U \inf_{x \in E \setminus E_{\text{inf}}(U)} U(x) - U_{\text{inf}}}\right)$, and $\sigma \geq \frac{B}{\ln(1 + \varepsilon)}$.

PROOF. Let (E, V, J) be the virtual energy landscape associated with $(E, (U_\tau), (q_\tau))$. From Propositions 2 and 3, we have $E_{\text{inf}}(V) = E_{\text{inf}}(U)$, $\Delta(J) = \Delta(q)$, and for any $(z, t) \in \Delta(J)$,

$$V(z) + J(z, t) = U(z) - U_{\text{inf}} + V_{\text{inf}} + (U(t) - U(z))^+ = \sup\{U(z), U(t)\} - U_{\text{inf}} + V_{\text{inf}}.$$

It follows that $\eta_V - V_{\text{inf}} = \eta_U - U_{\text{inf}}$, and thus $h_V = h_U$, $H_V = H_U$, and $D_V = D_U$. Consequently, for a family of finite-time SC algorithms of type (21) with cooling sequence of the form (17), Theorem 1 gives that if $A > H_U$, $\beta > H_U/D_U$, and $\gamma < \beta$, then

$$\forall \varepsilon > 0, \quad \forall \sigma \geq \frac{\ln(\beta/\gamma)}{\ln(1 + \varepsilon)}, \quad \liminf_{K \rightarrow +\infty} -\frac{\ln \sup_{x \in E} \mathbb{P}(U(X_{\sigma K}) \geq U_{\text{inf}} + \gamma \mid X_0 = x)}{\ln(\sigma K)} \geq \frac{1}{(1 + \varepsilon)D_U}.$$

Taking $\gamma < \inf_{x \in E \setminus E_{\text{inf}}(U)} U(x) - U_{\text{inf}}$ gives the asymptotic bound in (22). \square

The conditions for Theorem 2 to apply are weak; they only concern the communication scheme $(q_\tau)_\tau$. Assumptions (A1) and (A3) are standard in SA theory: the irreducibility of the limit q and the symmetry of its support $\Delta(q)$ ensure that the target energy landscape can be fully explored and that any path in this landscape can be traveled in the opposite direction. The meaning of (A2) is less clear, but a simple sufficient condition for it to hold is that $q(x, x) > 0$ for all x and that $\Delta(q_\tau) = \Delta(q)$ for τ small enough. In other words, it suffices to allow the limit communication mechanism to rest anywhere and to “freeze” the set of possible moves at low temperatures. Quite interestingly, there is no condition on the continuation scheme $(U_\beta)_\beta$ apart from pointwise convergence to the target energy.

Acknowledgements

This work was partly supported by the French National Research Agency under grant ANR-09-BLAN-0372-01. The authors would like to thank the referee for helpful comments and suggestions.

References

- Catoni, O., 1992. Rough large deviation estimates for simulated annealing: application to exponential schedules. *Ann. Probab.* 20 (3), 1109–1146.
- Catoni, O., 1999. Simulated annealing algorithms and Markov chains with rare transitions. In: *Séminaire de probabilités XXXIII*. Vol. 1709 of *Lecture Notes in Math*. Springer, New York, pp. 69–119.
- Del Moral, P., Miclo, L., 1999. On the convergence and applications of generalized simulated annealing. *SIAM J. Control Optim.* 37 (4), 1222–1250.
- Frigerio, A., Grillo, G., 1993. Simulated annealing with time-dependent energy function. *Math. Z.* 213, 97–116.
- Hajek, B., 1988. Cooling schedules for optimal annealing. *Math. Oper. Res.* 13 (2), 311–329.
- Löwe, M., 1996. Simulated annealing with time-dependent energy function via Sobolev inequalities. *Stochastic Process. Appl.* 63 (2), 221–233.
- Robini, M. C., Lachal, A., Magnin, I. E., 2007. A stochastic continuation approach to piecewise constant reconstruction. *IEEE Trans. Image Process.* 16 (10), 2576–2589.
- Trounev, A., 1996. Cycle decompositions and simulated annealing. *SIAM J. Control Optim.* 34 (3), 966–986.