Journal of Global Optimization **18:** 219–234, 2000.

Convergence of a Simulated Annealing Algorithm for Continuous Global Optimization

M. LOCATELLI
Dipartimento di Informatica, Universitá di Torino, Corso Svizzera, 185, 10149 Torino, Italy
(e-mail: locatell@di.unito.it)

(Received 14 October 1998; accepted 19 January 2000)

Abstract. In this paper a simulated annealing algorithm for continuous global optimization will be considered. The algorithm, in which a cooling schedule based on the distance between the function value in the current point and an estimate of the global optimum value is employed, has been first introduced in Bohachevsky, Johnson and Stein (1986) [2], but without any proof of convergence. Here it will be proved that, under suitable assumptions, the algorithm is convergent

Key words: Simulated annealing; Continuous global optimization; Cooling schedule

1. Introduction

In recent years many simulated annealing algorithms have been proposed for the solution of the continuous global optimization problem, i.e. for determining, or at least approximating, the value f^* defined as follows:

$$f^* = \min_{x \in X} f(x) ,$$

where f is the objective function and $X \subset R^d$ is the feasible region.

The name simulated annealing comes from a physical process called annealing, the process for growing crystals, which can be simulated by the Metropolis Monte Carlo method (see [13]). The simulated annealing idea was first introduced in the field of combinatorial optimization independently in [4] and [9] and, lately, it has been applied to continuous global optimization.

The general scheme of a simulated annealing algorithm is the following.

1. Let $Y_0 \in X$ be a given starting point, c_0 a nonnegative constant, $z_0 = \{(Y_0, f(Y_0))\}$ and set $k := 0$.
2. Sample a point X_{k+1} from a given distribution $D(\cdot; Y_k, z_k)$.
3. Sample a uniform random number p in $[0, 1]$ and set

$$Y_{k+1} := \begin{cases} X_{k+1} & \text{if } p \leqslant A(Y_k, X_{k+1}, c_k) \\ Y_k & \text{otherwise} \end{cases}$$

where A is a function with values in $[0, 1]$ and c_k is a parameter called the temperature at iteration k.

4. Set $z_{k+1} := z_k \cup \{(X_{k+1}, f(X_{k+1}))\}$ (the set z_k is the set of information collected by the algorithm up to iteration k, i.e. the set of points at which the function has been evaluated together with the corresponding function values).
5. Set $c_{k+1} := U(z_{z_{k+1}} \to z_{k+1})$, where U is a nonnegative function.
6. Check a stopping criterion and if it fails set $k := k + 1$ and go back to Step 2.

In order to define a particular simulated annealing algorithm one must specify the stopping criterion and the three functions D, A and U, which define respectively the distribution of the next candidate point, the probability of accepting it as the next iterate and the cooling schedule, i.e. the temperature, which is a parameter through which the acceptance of the candidate points is controlled. A common choice for the acceptance function A is the so-called Metropolis function

$$A(Y_k, X_{k+1}, c_k) = \min\left\{1, \exp\left\{\frac{f(Y_k) - f(X_{k+1})}{c_k}\right\}\right\}, \tag{1}$$

which always accepts descent steps, but also accepts with a positive probability (unless $c_k = 0$) ascent steps in order to avoid getting trapped in a local minimum which is not a global minimum. Note how the parameter c_k controls the acceptance of ascent steps: by decreasing c_k also the acceptance probability of ascent steps is decreased. Different choices for D and U have been proposed and computationally tested in the literature (see e.g. [2, 3, 5, 6, 8, 15, 16]). Moreover, different conditions under which convergence in probability to the global optimum of simulated annealing algorithms can be established, have been presented (see e.g. [1, 7, 10 and 11]).

In this paper a convergence result is presented for a simulated annealing algorithm based on a cooling schedule presented and computationally tested in [2], but for which, to the author’s knowledge, no convergence result has appeared in the literature. In Section 2 the details of the above mentioned algorithm will be discussed, and the assumptions under which the convergence result can be established will be presented and commented upon. In Section 3 the convergence in probability of the algorithm will be proved. The convergence result is based on two lemmas and one theorem whose proofs will be given in the Appendix.

2. A particular simulated annealing algorithm

In [2] a simulated annealing algorithm with the following characteristics has been introduced. The next candidate point distribution D generates random points over the boundary of a sphere of radius R with center in the current point Y_k; the acceptance function A is the Metropolis function (1); the cooling schedule U defines the temperature at iteration k as follows

$$c_k = \alpha[f(Y_k) - \hat{f}_k]^g ,$$

where α and g are positive constants and $\hat{f}_k$ is an estimate of f^* at iteration k. Note that the closer $f(Y_k)$ is to the estimate $\hat{f}_k$ of the global optimum value at iteration k, the smaller is the temperature and, consequently, the probability of accepting ascent steps. Such a cooling schedule has been also studied in [14] and [15].

The algorithm considered in this paper is the same as the one in [2] with the only difference that the next candidate point distribution D is assumed to generate random points over $S(Y_k, R)$, i.e. over the whole sphere with radius R and center Y_k, and not only on its boundary. Moreover, as an estimate $\hat{f}_k$ of the global optimum value, the value $f_k^* - t_k$ will be employed, where $f_k^* = \min\{f(Y_1), \ldots, f(Y_k)\}$ is the record value, and $\{t_k\}$ is a deterministic nonincreasing sequence converging to 0. Therefore, the cooling schedule will be the following

$$c_k = \alpha[f(Y_k) - f_k^* + t_k]^g . \tag{2}$$

We will discuss the convergence in probability of such an algorithm, i.e. we will prove that, under suitable assumptions

$$\forall \epsilon > 0 \lim_{k\to\infty} P[Y_k \in B_\epsilon] = 1 , \tag{3}$$

where

$$B_\epsilon = \{x \in X : f(x) \leq f^* + \epsilon\} .$$

Before introducing the assumptions under which the convergence result will be established, we compare (2) with the following cooling schedule, introduced in [10] and also employed in [11]

$$c_k = \begin{cases} f(Y_k) - f_k^* & \text{if } f(Y_k) - f_k^* > \bar{\epsilon} \\ t_k & \text{otherwise} \end{cases} \tag{4}$$

It has been proved that, under suitable assumptions, (3) holds when (4) is employed. Both in (2) and in (4) the temperature depends on how close the function value is to the current record f_k^*. In (4) we switch to a deterministic temperature when we are too close, within a distance $\bar{\epsilon}$, to the current record. A consequence of this fact is that, while we can always prove that

$$\lim_{k\to\infty} P[Y_k \in B_{\bar{\epsilon}}] = 1 ,$$

for the fixed value $\bar{\epsilon}$ employed in (4), in order to prove (3) it is necessary to assume that $\bar{\epsilon}$ is below a threshold which is strictly problem dependent and cannot generally be known in advance.

By employing (2) such a problem does not arise because the distance of the function value in the current point from the current record and the deterministic temperature t_k are merged into a unique function returning the value of the temperature at iteration k. We will show how to prove (3) when the cooling schedule (2) is employed. It will actually turn out that (3) is true only if the parameter g is appropriately chosen. However, there is an important difference with respect to the choice of the parameter $\bar{\epsilon}$ in (4): while, as already commented, it is generally not possible to know in advance how to choose $\bar{\epsilon}$ in order to guarantee (3), an appropriate choice of g seems to be much easier. In particular, we will see that in many cases any $g > 1$ can be chosen.

Now we introduce the assumptions under which the convergence result will be established. The first assumption is about the objective function f and the feasible region X.

ASSUMPTION 1. *X is convex, compact and full-dimensional, and the objective function f is continuous.*

This assumption is more restrictive than needed. In particular, the assumption of convexity of X is too strong and may be relaxed, but, on the other hand, it will help in proving the result without introducing too many technicalities.

The second assumption requires that f has only a finite number of global minima over X.

ASSUMPTION 2. *Let $X^* = \{x \in X : f(x) = f^*\}$, then $|X^*| = n < \infty$.*

Now let $x_i^* \in X^*$, $i \in \{1, \ldots, n\}$ be the ith global minimum of f over X. Then, for ϵ small enough, the set B_ϵ is partitioned into n disjoint connected components $C_\epsilon^i \ni x_i^*$, $i \in \{1, \ldots, n\}$. Let

$$s(\epsilon) = \min_{i \in \{1,\ldots,n\}} m(C_\epsilon^i), \tag{5}$$

where m denotes the Lebesgue measure. Then we introduce the following assumption which restricts the possible choices for the constant g in the cooling schedule (2).

ASSUMPTION 3. *The constant value g in the cooling schedule (2) is such that*

$$\exists \delta > 0 : m(B_{\epsilon + \epsilon^{g-\delta}} \backslash B_\epsilon) + e^{-(1/\alpha\epsilon^\delta)} = o(s(\epsilon)) \quad \textit{as } \epsilon \to 0 .$$

Assumption 3 is an important one because it gives a condition which must be satisfied by the constant g employed in the cooling schedule. In the next observation we prove that, under some regularity conditions in the regions around the global optima, any value $g > 1$ can be chosen. For simplicity, we only present the case in which only one global optimum exists and belongs to the interior of the feasible region X. However, the result can be extended to the case of multiple global optima, possibly belonging to the boundary of X.

OBSERVATION 1. *Assume that f has a unique global optimum x^* over X belonging to the interior of X. Moreover, assume that f is twice continuously* *differentiable in some neighbourhood of* x^* *and that sufficient second order optimality conditions are satisfied in* x^*. *Then, Assumption 3 is satisfied by any* $g > 1$.

Proof. Let us consider the second-order Taylor expansion of f around x^*

$$f(x) = f^* + (x - x^*)H(x^*)(x - x^*) + o(\|x - x^*\|^2)\,, \tag{6}$$

where $H(x^*)$ denotes the Hessian of f in x^*. Consider the set

$$\{x : (x - x^*)H(x^*)(x - x^*) \leqslant \epsilon\}\,.$$

Since, by assumption, $H(x^*)$ is positive definite, the Lebesgue measure of the set above is $\rho_d \epsilon^{d/2}$ for some positive constant ρ_d. Moreover, in view of (6), it follows that

$$s(\epsilon) = m(B_\epsilon) = \rho_d \epsilon^{d/2} + o(\epsilon^{d/2})\,. \tag{7}$$

Now let $g > 1$ and $\delta = (g - 1)/2 > 0$. We immediately note that for any $\alpha > 0$

$$e^{-(1/\alpha\epsilon^\delta)} = o(s(\epsilon)) \quad \text{as } \epsilon \to 0\,.$$

Therefore, we need only to prove that the Lebesgue measure of the set $B_{\epsilon+\epsilon^{g-\delta}} \backslash B_\epsilon$ is also $o(s(\epsilon))$. From (7) and the choice for the value of δ it follows that

$$\begin{aligned} m(B_{\epsilon+\epsilon^{g-\delta}} \backslash B_\epsilon) &= m(B_{\epsilon+\epsilon^{g+1/2}}) - m(B_\epsilon) \\ &= \rho_d[(\epsilon + \epsilon^{g+1/2})^{d/2} - \epsilon^{d/2}] + o(\epsilon^{d/2})\,. \end{aligned} \tag{8}$$

Therefore

$$m(B_{\epsilon+\epsilon^{g-\delta}} \backslash B_\epsilon) = \rho_d \epsilon^{d/2}[(1 + \epsilon^\delta)^{d/2} - 1] + o(\epsilon^{d/2}) = o(s(\epsilon))\,,$$

as we wanted to prove.

From the proof of the observation we notice that, for $g > 1$, we can choose a value $\delta = (g - 1)/2 > 0$ so that the proof that Assumption 3 holds basically reduces to prove that the difference between the Lebesgue measures of a level set $B_{\epsilon+o(\epsilon)}$ and of the level set B_ϵ decreases to 0 faster than the measure of B_ϵ itself (see (8)). This has been proven to be true under the conditions of the observation but it is also true under much more general conditions.

Finally, we introduce the following assumption on the distribution D of the next candidate point.

ASSUMPTION 4. *For any* $x \in X$, *for any* k *and for any* z_k, $\exists q, Q > 0$ *such that* $\forall C \subseteq S(x, R) \cap X$

$$qm(C) \leqslant D(C; x, z_k) \leqslant Qm(C)\,.$$

In the next section the convergence in probability of the algorithm under the above assumptions will be proved.

3. Convergence in probability of the algorithm

The proof that (3) holds under the cooling schedule (2) will follow the scheme employed in Theorem 3 in [10] to prove (3) under the cooling schedule (4), but the single steps of the proof will need to be adapted to the different cooling schedule.

In order to prove that (3) holds under the given assumptions and under the cooling schedule (2), two lemmas are needed. The first lemma gives a lower bound for the probability that Y_{k+T} is inside the set B_ϵ given that Y_k belongs to this set, for any fixed positive integer T. From now on the event $\{Y_k \in B_\epsilon\}$ will be denoted by V_k^ϵ and its complement by $\overline{V}_k^\epsilon$.

LEMMA 1. *Let Assumption 4 hold. Let*

$$v(\epsilon) = e^{-(1/\alpha\epsilon^\delta)} + Qm(B_{\epsilon+\epsilon^{g-\delta}} \backslash B_\epsilon) ,$$

where δ is some positive constant. Then, given a positive integer T, there exists $\epsilon_1 > 0$ such that $\forall \epsilon \leqslant \epsilon_1$, there exists a positive integer $K_1 = K_1(\epsilon, T)$ such that $\forall k \geqslant K_1$

$$P[V_{k+T}^\epsilon \,|\, V_k^\epsilon] \geqslant 1 - 2Tv(\epsilon) .$$

Proof. See the Appendix.

The second lemma gives a lower bound for the probability that, for a given constant N, Y_{k+N} belongs to the set B_ϵ given that Y_k does not belong to this set and that the set B_ϵ has been visited at least once in the first k iterations, i.e. the event E_k defined as follows

$$E_k = \{\exists h \leqslant k : Y_h \in B_\epsilon\} , \tag{9}$$

holds.

LEMMA 2. *Let Assumptions 1, 2 and 4 hold. Let*

$$N = \left\lfloor \frac{4\,\mathrm{diam}(X)}{R} \right\rfloor + 1 , \tag{10}$$

where diam(X) *denotes the diameter of the convex set X. Then there exists $\epsilon_2 > 0$ such that $\forall \epsilon \leqslant \epsilon_2$, there exists a positive integer $K_2 = K_2(\epsilon)$ such that $\forall k \geqslant K_2$*

$$P[V_{k+N}^\epsilon \,|\, \overline{V}_k^\epsilon, E_k] \geqslant \gamma s(\epsilon) ,$$

where $s(\epsilon)$ is defined in (5) and γ is a positive constant.

Proof. See the Appendix.

Before the convergence proof, we need a further result proving that under the given assumptions and under a suitable choice for the sequence $\{t_k\}$, the infinite sequence $\{Y_k\}$ generated by the algorithm will visit the set B_ϵ for any $\epsilon > 0$ with probability 1, i.e. the event E_k, defined in (9), has a probability converging to 1 as $k \to \infty$.

THEOREM 1. *Let*

$$\Delta F = \max_{x \in X} f(x) - f^* , \tag{11}$$

and N be defined by (10). If Assumptions 1, 2 and 4 hold and

$$t_k \geqslant \left[\left(\frac{1}{\alpha} \right) \frac{N \, \Delta F}{\log k} \right]^{1/g} , \tag{12}$$

then

$$\lim_{k \to \infty} P[E_k] = 1 ,$$

where E_k is the event defined in (9).

Proof. See the Appendix.

Now we are ready for the proof of convergence.

THEOREM 2. *Let Assumptions 1, 2, 3 and 4 hold. Let t_k satisfy (12). Then*

$$\forall \epsilon > 0 \lim_{k \to \infty} P[Y_k \in B_\epsilon] = 1 .$$

Proof. First we note that for any positive integer H and any $k \geqslant H$

$$P[V_k^\epsilon] \geqslant P[V_k^\epsilon | E_H] P[E_H] .$$

Since, in view of Theorem 1, $P[E_H] \to 1$ as $H \to \infty$, in order to prove (3) it is enough to prove that

$$a_k(\epsilon) = P[V_k^\epsilon | E_H] \to 1 \text{ as } k \to \infty .$$

We note that $a_{k+N}(\epsilon) = P[V_{k+N}^\epsilon | E_H]$ is equal to

$$a_k(\epsilon) P[V_{k+N}^\epsilon | V_k^\epsilon, E_H] + [1 - a_k(\epsilon)] P[V_{k+N}^\epsilon | \overline{V}_k^\epsilon, E_H] ,$$

where N is the same as in Lemma 2. Therefore, it follows from Lemmas 1 and 2 that for ϵ small enough and k big enough

$$a_{k+N}(\epsilon) \geqslant [1 - 2N\nu(\epsilon)] a_k(\epsilon) + [1 - a_k(\epsilon)] \gamma s(\epsilon) ,$$

or, equivalently

$$a_{k+N}(\epsilon) - a_k(\epsilon) \geqslant \gamma s(\epsilon) - [\gamma s(\epsilon) + 2N\nu(\epsilon)] a_k(\epsilon) . \tag{13}$$

It also follows that

$$a_{k+N}(\epsilon) < a_k(\epsilon) \;\Rightarrow\; a_k(\epsilon) > \frac{\gamma s(\epsilon)}{\gamma s(\epsilon) + 2N\nu(\epsilon)} = \xi(\epsilon) ,$$

and, in view of (13), the difference $[a_k(\epsilon) - a_{k+N}(\epsilon)]$ is bounded from above by $[\gamma s(\epsilon) + 2N\nu(\epsilon)]$. Two cases are possible.

1. $\exists \bar{k}$ such that $a_{\bar{k}}(\epsilon) \geqslant \xi(\epsilon)$: then the sequence $\{a_k(\epsilon)\}_{k \geqslant \bar{k}}$ will never fall below the following limit

$$\xi(\epsilon) - [\gamma s(\epsilon) + 2N\nu(\epsilon)] ;$$

2. the sequence $\{a_k(\epsilon)\}$ is always below $\xi(\epsilon)$: then the sequence is increasing and converges to a limit. By taking the limit for $k \to \infty$ of both sides in (13), it follows that

$$0 \geqslant \gamma s(\epsilon) - [\gamma s(\epsilon) + 2N\nu(\epsilon)] \lim_{k \to \infty} a_k(\epsilon) ,$$

or, equivalently

$$\lim_{k \to \infty} a_k(\epsilon) \geqslant \xi(\epsilon) .$$

In both cases we have that

$$\underline{\lim}_{k \to \infty} a_k(\epsilon) \geqslant \xi(\epsilon) - [\gamma s(\epsilon) + 2N\nu(\epsilon)] .$$

Since, obviously, $\forall \epsilon > 0$ and $\forall k$ it holds that $a_k(\epsilon) \leqslant 1$, and since $a_k(\epsilon)$ is a nondecreasing function of ϵ, it follows that

$$1 \geqslant \underline{\lim}_{k \to \infty} a_k(\epsilon) \geqslant \lim_{\epsilon \to 0} \underline{\lim}_{k \to \infty} a_k(\epsilon) \geqslant \lim_{\epsilon \to 0} \xi(\epsilon) - [\gamma s(\epsilon) + 2N\nu(\epsilon)] ,$$

where the limit on the right-hand side is equal to 1 under Assumption 3. It follows that

$$\lim_{k \to \infty} a_k(\epsilon) = 1 ,$$

as we wanted to prove.

4. Conclusion

In this paper a simulated annealing algorithm for continuous global optimization, first introduced in [2], has been considered. It has been proved that, under suitable assumptions, the algorithm is convergent. A cooling schedule, which depends on the distance of the function value of the current iterate from an estimate of the global optimum value f^*, is employed. Here at each iteration an estimate has been proposed based on a suitable decrease of the record value. One of the assumptions introduced in order to guarantee convergence restricts the choice of the values of the parameter g, appearing in the cooling schedule and controlling the speed at which the temperature varies. It has been shown (see Observation 1) that, under some regularity conditions, any value $g > 1$ can be chosen.

Appendix

A. PROOF OF LEMMA 1

Let A_{k+i}, $i = 1, \ldots, T$, denote the event $\{f(Y_{k+i}) \leqslant f(Y_{k+i-1})\}$. It holds that

$$\bigcap_{i=1}^{T} A_{k+i} | V_k^{\epsilon} \ \Rightarrow \ V_{k+T}^{\epsilon} | V_k^{\epsilon} .$$

Therefore

$$P[V_{k+T}^{\epsilon} | V_k^{\epsilon}] \geqslant P\left[\bigcap_{i=1}^{T} A_{k+i} | V_k^{\epsilon} \right] .$$

The right-hand side of this inequality is equal to

$$\prod_{i=1}^{T} P\left[A_{k+i} \,\middle|\, \bigcap_{j=1}^{i-1} A_{k+j}, V_k^{\epsilon} \right] . \tag{14}$$

By definition the term in the product is equal to

$$1 - P\left[f(Y_{k+i}) > f(Y_{k+i-1}) \,\middle|\, \bigcap_{j=1}^{i-1} A_{k+j}, V_k^{\epsilon} \right] , \tag{15}$$

where the probability can be split, for any $\delta > 0$, in the sum of

$$P\left[f(Y_{k+i}) > f(Y_{k+i-1}) + \epsilon^{g-\delta} \,\middle|\, \bigcap_{j=1}^{i-1} A_{k+j}, V_k^{\epsilon} \right] , \tag{16}$$

and

$$P\left[f(Y_{k+i-1}) < f(Y_{k+i}) \leqslant f(Y_{k+i-1}) + \epsilon^{g-\delta} \,\middle|\, \bigcap_{j=1}^{i-1} A_{k+j}, V_k^{\epsilon} \right] . \tag{17}$$

We note that

$$\bigcap_{j=1}^{i-1} A_{k+j}, V_k^{\epsilon} \ \Rightarrow \ V_{k+i-1}^{\epsilon} \ \Rightarrow \ c_{k+i} \leqslant \alpha[\epsilon + t_{k+i}]^g \leqslant \alpha[\epsilon + t_k]^g ,$$

where the right-hand side of the last inequality follows from the nonincreasingness of $\{t_k\}$. Therefore, an upper bound for the probability in (16) is

$$e^{-\epsilon^{g-\delta}/\alpha[\epsilon + t_k]^g} .$$

In view of Assumption 4, the probability in (17) can be bounded from above by $Qm(B_{\epsilon+\epsilon^{g-\delta}} \backslash B_{\epsilon})$. Therefore, an upper bound for the probability in (15) is given by

$$v_k(\epsilon) = e^{-\epsilon^{g-\delta}/\alpha[\epsilon + t_k]^g} + Qm(B_{\epsilon+\epsilon^{g-\delta}} \backslash B_{\epsilon}) ,$$

and, from (14), it follows that

$$P[V^{\epsilon}_{k+T} | V^{\epsilon}_{k}] \geqslant [1 - v_k(\epsilon)]^T .$$

By observing that, since $t_k \downarrow 0$, $\forall \epsilon > 0 : v_k(\epsilon) \downarrow v(\epsilon)$ as $k \to \infty$, and that $v(\epsilon) \to 0$ as $\epsilon \to 0$, it follows that we can choose a small enough $\epsilon_1 > 0$ such that $\forall \epsilon \leqslant \epsilon_1$, there exists a positive integer $K_1 = K_1(\epsilon, T)$ such that $\forall k \geqslant K_1$

$$[1 - v_k(\epsilon)]^T \geqslant 1 - 2Tv(\epsilon) ,$$

from which the result of the lemma follows.

B. PROOF OF LEMMA 2

In order to prove Lemma 2 we first need some remarks. Let us consider $y \in X \backslash B_{\epsilon}$ and let $x^*_v = x^*_{v(y)} \in X^*$ be the global minimum closest to y, i.e.

$$d(y, x^*_v) = \min_{i \in \{1,\ldots,n\}} d(y, x^*_i) .$$

Consider the segment $[y, x^*_v]$, which belongs to X in view of its convexity. Define the following sequence

$$w_0 = y , \quad w_i = w_{i-1} + \frac{R}{4} \frac{(x^*_v - y)}{\|x^*_v - y\|} , \quad i = 1, \ldots, M(y) , \tag{18}$$

where

$$M(y) = \min\left\{ j : d(y, x^*_v) - \frac{R}{4} j < \frac{3}{4} R \right\} ,$$

and the corresponding sequences of spheres $S(w_i, R/8)$, $i = 1, \ldots, M(y)$. Note that each point w_i in the sequence defined by (18) belongs to the segment $[y, x^*_v]$. The following remark gives a bound from above for the length of the sequence.

REMARK 1. For any $y \in X$

$$M(y) \leqslant \left\lfloor \frac{4\,\mathrm{diam}(X)}{R} \right\rfloor = N - 1 , \tag{19}$$

where N is the same as in (10).

Proof. The inequality (19) follows by the observation that

$$M(y) \leqslant \min\left\{ j : \mathrm{diam}(X) - \frac{R}{4} j < \frac{3}{4} R \right\} .$$

The next remark shows that a sphere with center in some point $\omega \in S(w_i, R/8)$ and radius R completely contains the sphere $S(w_{i+1}, R/8)$.

REMARK 2. For any $i = 1, \ldots, M(y) - 1$, and for any $\omega \in S(w_i, R/8)$

$$S(\omega, R) \supseteq S\left(w_{i+1}, \frac{R}{8}\right).$$

Proof. The remark immediately follows by the observation that the maximum distance between a point $\omega \in S(w_i, R/8)$ and any point belonging to $S(w_{i+1}, R/8)$ is not greater than $d(w_i, w_{i+1}) + \frac{2}{8}R = R/2 < R$.

The following remark introduces a lower bound for the distance between any point in $S(w_i, R/8)$, $i \in \{1, \ldots, M(y)\}$ and x_v^*, and an upper bound for the distance between any point in $S(w_{M(y)}, R/8)$ and x_v^*.

REMARK 3. For any $\omega \in S(w_i, R/8)$, $i \in \{1, \ldots, M(y)\}$

$$d(\omega, x_v^*) \geqslant \frac{3}{8} R,$$

and for any $\omega \in S(w_{M(y)}, R/8)$

$$d(\omega, x_v^*) < \frac{7}{8} R.$$

Proof. Note that, by definition of $M(y)$, for any $i \in \{1, \ldots, M(y)\}$

$$d(w_i, x_v^*) \geqslant \frac{R}{2}. \tag{20}$$

Therefore, for any $\omega \in S(w_i, R/8)$, $i \in \{1, \ldots, M(y)\}$

$$d(\omega, x_v^*) \geqslant d(w_i, x_v^*) - \frac{R}{8} \geqslant \frac{3}{8} R.$$

Moreover, again by definition of $M(y)$, it follows that $d(w_{M(y)}, x_v^*) < \frac{3}{4}R$, and for any $\omega \in S(w_{M(y)}, R/8)$

$$d(\omega, x_v^*) \leqslant d(w_{M(y)}, x_v^*) + \frac{R}{8} < \frac{3}{4} R + \frac{R}{8} = \frac{7}{8} R.$$

In the following remark it is stated that any point belonging to one of the spheres $S(w_i, R/8)$ has a distance from any global optimum not smaller than $R/8$.

REMARK 4. For any $\omega \in S(w_i, R/8)$, $i \in \{1, \dots, M(y)\}$, and for any $j \in \{1, \dots, n\}$

$$d(\omega, x_j^*) \geqslant \frac{R}{8} .$$

Proof. In view of Remark 3, $d(\omega, x_v^*) \geqslant \frac{3}{8}R$, thus we only need to prove the remark for $j \neq v$. By contradiction, we assume that there exist $j \neq v$, $t \in \{1, \dots, M(y)\}$ and $\omega \in S(w_t, R/8)$ such that $d(\omega, x_j^*) < R/8$. It follows from the triangular inequality that

$$d(y, x_j^*) \leqslant d(y, w_t) + d(w_t, \omega) + d(\omega, x_j^*) < d(y, w_t) + \frac{R}{4} .$$

Since from (20) it follows that

$$d(y, w_t) \leqslant d(y, x_v^*) - \frac{R}{2} ,$$

then

$$d(y, x_j^*) \leqslant d(y, x_v^*) - \frac{R}{4} ,$$

thus contradicting the fact that x_v^* is the global minimum closest to y.

Then we need a remark which shows that the Lebesgue measure of the intersection $S(x, R) \cap X$ is, for any $x \in X$, a positive fraction of the Lebesgue measure of the sphere $S(x, R)$.

REMARK 5. Let X be convex, compact and full-dimensional. Then $\forall x \in X$, $\exists \beta > 0$ such that

$$\inf_{r \in (0, \mathrm{diam}(X)]} \frac{m(S(x, r) \cap X)}{m(S(x, r))} \geqslant \beta .$$

Proof. See Lemma 1 in [12].

Finally, we need a remark which gives a lower bound for the probability of moving from any point outside B_ϵ at iteration k to the set B_ϵ at iteration $k + N$, given any past history z_{k-1}.

REMARK 6. There exists a positive constant ϵ_3 such that $\forall \epsilon \leqslant \epsilon_3$ there exists a positive integer $K_3 = K_3(\epsilon)$ such that $\forall k \geqslant K_3$, $\forall y \in X \setminus B_\epsilon$ and $\forall z_{k-1}$

$$P[V^\epsilon_{k+N} \,|\, Y_k = y,\, z_{k-1}] \geqslant \eta s(\epsilon) \prod_{i=1}^{M(y)} e^{-\Delta F / lb(c_{k+i})}\,, \tag{21}$$

where ΔF is defined in (11), N is defined in (10), η is a positive constant and $lb(c_{k+i})$ denotes a lower bound for the temperature c_{k+i}.

Proof. Let ϵ be small enough so that

$$x \in X,\ d(x, x^*_v) \leqslant \frac{7}{8} R \ \Rightarrow\ S(x, R) \supseteq C^v_\epsilon\,. \tag{22}$$

A lower bound for $P[V^\epsilon_{k+N} \,|\, Y_k = y, z_{k-1}]$ is given by

$$P[V^\epsilon_{k+M(y)+1} \,|\, Y_k = y, z_{k-1}] P[V^\epsilon_{k+N} \,|\, V^\epsilon_{k+M(y)+1}, Y_k = y, z_{k-1}]\,,$$

which, on the other hand, in view of Lemma 1, can be bounded from below, for ϵ small enough and k big enough, by

$$P[V^\epsilon_{k+M(y)+1} \,|\, Y_k = y, z_{k-1}][1 - 2Nv(\epsilon)]\,. \tag{23}$$

For any $i \in \{1, \ldots, M(y)\}$, let F_{k+i} denote the event $\{Y_{k+i} \in S(w_i, R/8)\}$, where the points w_i are those defined in (18). Then, we obviously have that a lower bound for (23) is

$$[1 - 2Nv(\epsilon)] P\left[V^\epsilon_{k+M(y)+1}, \bigcap_{i=1}^{M(y)} F_{k+i} \,\middle|\, Y_k = y, z_{k-1} \right], \tag{24}$$

where the probability is equal to the product of

$$P\left[V^\epsilon_{k+M(y)+1} \,\middle|\, \bigcap_{i=1}^{M(y)} F_{k+i}, Y_k = y, z_{k-1} \right] \tag{25}$$

and

$$\prod_{i=1}^{M(y)} P\left[F_{k+i} \,\middle|\, \bigcap_{j=1}^{i-1} F_{k+j}, Y_k = y, z_{k-1} \right]. \tag{26}$$

From Remark 3, (22), Assumption 4 and by definition of $s(\epsilon)$, it follows that

$$P\left[V^{\epsilon}_{k+M(y)+1} \,\middle|\, \bigcap_{i=1}^{M(y)} F_{k+i}, Y_k = y, z_{k-1}\right] \geqslant qm(C^{v}_{\epsilon}) \geqslant qs(\epsilon)\,. \tag{27}$$

In view of Assumption 4 and Remarks 2 and 5

$$P\left[F_{k+i} \,\middle|\, \bigcap_{j=1}^{i-1} F_{k+j}, Y_k = y\right] \geqslant q\beta m\left(S\left(w_i, \frac{R}{8}\right)\right) e^{-\Delta F/lb(c_{k+i})}\,. \tag{28}$$

Therefore, from (24)–(28) and recalling that $v(\epsilon) \to 0$ as $\epsilon \to 0$, it follows that for ϵ small enough and k big enough

$$P[V^{\epsilon}_{k+N} | Y_k = y, z_{k-1}] \geqslant \eta s(\epsilon) \prod_{i=1}^{M(y)} e^{-\Delta F/lb(c_{k+i})}\,,$$

for some positive constant η, as we wanted to prove.

Now we are ready for the proof of Lemma 2.
Proof. Let

$$\Delta = \bar{f} - f^* = \min_{x \in X \setminus \cup_{i=1}^{n} S(x_i^*, R/8)} f(x) - f^* > 0\,. \tag{29}$$

Note that in order to prove the lemma it is enough to prove that $\forall y \in X \setminus B_\epsilon$

$$P[V^{\epsilon}_{k+N} | Y_k = y, E_k] \geqslant \gamma s(\epsilon)\,. \tag{30}$$

From Remark 4 and (29), it follows that for any $i \in \{1, \ldots, M(y)\}$

$$F_{k+i}, E_k \;\Rightarrow\; c_{k+i} \geqslant \Delta - \epsilon + t_{k+i}\,.$$

Therefore, for $\epsilon \leqslant \Delta/2$, a lower bound $lb(c_{k+i})$ for c_{k+i}, $i \in \{1, \ldots, M(y)\}$, is given by $\Delta/2$ and, in view of Remark 6

$$P[V^{\epsilon}_{k+N} | Y_k = y, E_k] \geqslant \gamma s(\epsilon)\,,$$

for some positive constant γ, as we wanted to prove.

C. PROOF OF THEOREM 1

Obviously

$$P[E_k] \geqslant P\left[V^{\epsilon}_{pN}, \text{ for some } p \in \left\{0, \ldots, \left\lfloor \frac{k}{N} \right\rfloor \right\} \right],$$

where N is defined in (10). Equivalently

$$P[E_k] \geqslant 1 - \prod_{p=0}^{\lfloor k/N \rfloor} P\left[\overline{V}^{\epsilon}_{pN} \,\middle|\, \bigcap_{j=0}^{p-1} \overline{V}^{\epsilon}_{jN} \right]. \tag{31}$$

We note that an upper bound for the term in the productory is given by

$$\sup_{y \in X \setminus B_\epsilon} P\left[\overline{V}^{\epsilon}_{pN} \,\middle|\, Y_{(p-1)N} = y, \bigcap_{j=0}^{p-2} \overline{V}^{\epsilon}_{jN} \right].$$

Note that, for any $i \in \{1, \ldots, N\}$

$$c_{(p-1)N+i} \geqslant \alpha(t_{(p-1)N+i})^g \geqslant \frac{N\,\Delta F}{\log(pN)},$$

where the last inequality follows from (12) and the nonincreasingness of $\{t_k\}$. Therefore, in view of Remark 6, it holds that

$$P\left[\overline{V}^{\epsilon}_{pN} \,\middle|\, Y_{(p-1)N} = y, \bigcap_{j=0}^{p-2} \overline{V}^{\epsilon}_{jN} \right] \leqslant 1 - \eta s(\epsilon) \frac{1}{pN}.$$

By substitution in (31), it follows that

$$P[E_k] \geqslant 1 - e^{\sum_{p=0}^{\lfloor k/N \rfloor} \log\left(1 - \frac{\eta s(\epsilon)}{pN}\right)},$$

where the term in the summatory is asymptotic to $-\eta s(\epsilon) 1/(pN)$ as $p \to \infty$ and, consequently, the summatory diverges to $-\infty$ as $k \to \infty$. Therefore

$$1 \geqslant \lim_{k\to\infty} P[E_k] \geqslant \lim_{k\to\infty} 1 - e^{\sum_{p=0}^{\lfloor k/N \rfloor} \log\left(1 - \frac{\eta s(\epsilon)}{pN}\right)} = 1,$$

from which the result of the theorem follows.

References

1. Belisle, C.J.P. (1992), Convergence theorems for a class of simulated annealing algorithms on $\mathscr{R}^d$, *Journal of Applied Probability* 29: 885–892.
2. Bohachevsky, I.O., Johnson, M.E. and Stein, M.L. (1986), Generalized simulated annealing for function optimization, *Technometrics* 28: 209–217.
3. Brooks, D.G. and Verdini, W.A. (1988), Computational experience with generalized simulated annealing over continuous variables, *Amer. J. of Mathematical and Management Sciences* 8: 425–449.
4. Černy, V. (1985), Thermodynamical approach to the travelling salesman problem: An efficient simulation algorithm, *J. Optim. Theory Appl.* 45: 41–51.
5. Coleman, T., Shalloway, D. and Zhijun Wu (1994), A parallel build-up algorithm for global energy minimizations of molecular clusters using effective energy simulated annealing, *J. of Global Optimization* 4: 171–185.
6. Corana, A., Marchesi, M., Martini, C. and Ridella, S. (1987), Minimizing multimodal functions of continuous variables with the 'Simulated Annealing' algorithm, *ACM Trans. Math. Software* 13: 262–280.
7. Gelfand, S.B. and Mitter, S.K. (1993), Metropolis-type annealing algorithms for global optimization in $\mathscr{R}^d$, *SIAM J. of Control and Optimization* 31(1): 111–131.
8. Jones, A.E.W. and Forbes, G.W. (1995), An adaptive simulated annealing algorithm for global optimization over continuous variables, *J. of Global Optimization* 6: 1–37.
9. Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983), Optimization by simulated annealing, *Science* 220(4598): 671–680.
10. Locatelli, M. (1996), Convergence properties of simulated annealing for continuous global optimization, *J. of Applied Probability* 33: 1127–1140.
11. Locatelli, M. (2000), Simulated annealing algorithms for continuous global optimization: convergence conditions, *J. of Optimization Theory and Applications* 104: 121–133.
12. Locatelli, M. (1999), On convergence of simulated annealing algorithms for continuous global optimization, Technical Report 01-99, Dip. di Sistemi e Informatica, Universitá di Firenze, also available at the web site www.dsi.unifi.it/users/locatell/dist4.ps
13. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. and Teller, A.H. (1953), Equation of state calculations by fast computer machines, *J. Chem. Phys.* 21: 1087.
14. Romeijn, H.E. and Smith, R.L. (1994), Simulated annealing and adaptive search in global optimization, *Probability in the Engineering and Informational Sciences* 8: 571–590.
15. Romeijn, H.E. and Smith, R.L. (1994), Simulated annealing for constrained global optimization, *J. of Global Optimization* 5: 101–126.
16. Vanderbilt, D. and Louie, S.G. (1984), A Monte Carlo simulated annealing approach to optimization over continuous variables, *J. of Computational Physics* 56: 259–271.