JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 104, No. 1, pp. 121–133, JANUARY 2000

Simulated Annealing Algorithms for Continuous Global Optimization: Convergence Conditions[1]

M. LOCATELLI[2]

Communicated by G. Di Pillo

Abstract. In this paper, simulated annealing algorithms for continuous global optimization are considered. After a review of recent convergence results from the literature, a class of algorithms is presented for which strong convergence results can be proved without introducing assumptions which are too restrictive. The main idea of the paper is that of relating both the temperature value and the support dimension of the next candidate point, so that they are small at points with function value close to the current record and bounded away from zero otherwise.

Key Words. Global optimization, simulated annealing, convergence conditions.

1. Introduction

Let f be a continuous objective function defined over a compact feasible set X. We face the problem of determining

$$f^* = \min_{x \in X} f(x), \qquad X \subseteq R^d,$$

called a global optimization problem. In particular, we consider the use of simulated annealing algorithms for the solution of this problem. Simulated annealing algorithms have received a great deal of attention in the last years. The name "simulated annealing" comes from a physical process called annealing, the process for growing crystals, which can be simulated by the

[1]This research was partially supported by Project MOST. The author thanks two anonymous referees for their helpful suggestions.

[2]Postdoctoral Fellow, Universitá di Firenze, Dipartimento di Sistemi ed Informatica, Firenze, Italy.

0022-3239/00/0100-0121$18.00/0

Metropolis Monte Carlo method (see Ref. 1). It was applied first to combinatorial global optimization independently in Refs. 2 and 3, and lately to continuous global optimization. In the field of continuous global optimization, remarkable works from the point of view of applications are Refs. 4–10. The general structure of a simulated annealing algorithm is the following.

Step 1. Let $Y_0 \in X$ be a given starting point, let $z_0 = \{(Y_0, f(Y_0))\}$, $f_0^* = f(Y_0)$, and $k = 0$.

Step 2. Sample a point X_{k+1} from a distribution $D(\cdot; z_k)$.

Step 3. Sample a uniform random number p in [0, 1] and set

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

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

Step 4. Set $z_{k+1} = z_k \cup \{(X_{k+1}, f(X_{k+1}))\}$; the set z_k contains the 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.

Step 5. Set $t_{k+1} = U(z_{k+1})$, where U is a function with nonnegative values.

Step 6. Set $f_{k+1}^* = \min\{f_k^*, f(X_{k+1})\}$, where f_k^* is the record value at iteration k, i.e., the best observed value up to iteration k.

Step 7. Check a stopping criterion; if it fails, set $k := k + 1$ and go back to Step 2.

A particular simulated annealing algorithm is specified by choosing the stopping criterion and the functions D, A, 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 as a consequence of the dependency of A on it. In what follows, the acceptance probability is the so-called Metropolis function,

$$A(Y_k, X_{k+1}, t_k) = \min\{1, \exp((f(Y_k) - f(X_{k+1}))/t_k)\}, \tag{1}$$

which accepts always descent steps and accepts ascent steps with a positive probability (unless $t_k = 0$) in order to avoid getting trapped in a local minimum which is not a global minimum. The parameter t_k controls the acceptance of the ascent steps: decreasing t_k also decreases their acceptance probability.

An interesting theoretical issue is to establish conditions under which we can guarantee convergence in probability to the global optimum of a

simulated annealing algorithm for continuous global optimization. This problem is faced in the following sections. In Section 2, recent results from the literature are presented. In Section 3, the assumptions under which new results can be derived are introduced. Finally, in Section 4, the new convergence results are presented. The proofs of the convergence results are rather technical and have been omitted in the final version of the paper due to space reasons; they can be found in detail in Ref. 11.

2. Convergence Conditions in the Literature

In this section, we introduce briefly some recent convergence results for simulated annealing algorithms for continuous global optimization. First, we recall the convergence result presented in Ref. 12. For any $\epsilon > 0$, let

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

In Ref. 12, the following assumptions are introduced:

(A1) $D(\cdot; z_k)$ has the form $D(Y_k, \cdot)$, where $D(\cdot, \cdot)$ is a Markov kernel, absolutely continuous with respect to the Lebesgue measure m and with density β bounded away from zero, i.e.,

$$D(x, B) = \int_B \beta(x, y)\, dy \quad \text{and} \quad \inf_{x,y \in X} \beta(x, y) = \rho > 0, \qquad \forall B \in \mathcal{B}, \tag{2}$$

where $\mathcal{B}$ is a σ-algebra over X.

(A2) For any $G \subseteq X$ open, $D(x, G)$ is continuous in x.

(A3) For any initial state Y_0 and any initial temperature t_0,

$$t_k = U(z_k) \to 0, \tag{3}$$

with probability 1.

(A4) X is a compact set, f is a continuous function, and $m(B_\epsilon) > 0$, $\forall \epsilon > 0$.

Under the above assumptions, it has been proved that

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

Note that (3) is much less restrictive than the usual conditions for the temperature in the combinatorial case, where it is required that the temperature decreases to zero not faster than the inverse of the logarithm of the iteration counter k times a constant (see Ref. 13). The reason for this is that, while

in combinatorial optimization we perform only local steps (i.e., steps in the neighborhood of the current point), here (2) says that every step is a global one. The negative consequence of this fact is that, for any $A \subseteq X \setminus B_\epsilon$, we have that

$$P[X_k \in A] \geq \rho m(A), \qquad \forall k.$$

Therefore, at any iteration, there is a probability bounded away from zero of sampling points in regions far from the global optimum region. Instead, we would like to be able to perform steps which are only local in order to explore more deeply the most promising parts of the feasible region.

Another interesting convergence result has been given in Ref. 14. In this paper, a class of simulated annealing algorithms is presented for which convergence to the global optimum not only of the sequence $\{Y_k\}$ but also of the sequence of candidate points $\{X_k\}$ is guaranteed, so that asymptotically we have a local exploration in the region of the global optimum. On the other hand, the conditions under which these results are proved are restrictive. In particular it is required that:

(B1) $f \in C^2$.

(B2) The distribution D of the next candidate point is Gaussian.

In this paper, we introduce a class of algorithms for which convergence to the global optimum of the sequence of candidate points $\{X_k\}$ is guaranteed (as in Ref. 14), by introducing assumptions which are not too restrictive (as in Ref. 12). A first step in this direction has been made in Ref. 15. In that paper, a class of algorithms has been introduced for which the support of $D(\cdot; z_k)$ is not the whole feasible set but its intersection with the sphere $S(Y_k, R)$ around the current point (R is a constant value). The following cooling schedule has been introduced:

$$t_k \geq \delta_1 > 0, \qquad \text{if } f(Y_k) - f_k^* > \bar{\epsilon}, \tag{4a}$$

$$t_k = c_k, \qquad \text{otherwise}, \tag{4b}$$

where $\{c_k\}$ is a deterministic nonincreasing sequence converging to 0 and $\bar{\epsilon} > 0$ is a constant such that

$$m(S(x, R) \cap B_\epsilon) > 0, \qquad \forall x \in B_{2\bar{\epsilon}}, \forall \epsilon > 0, \tag{5}$$

Basically, condition (5) requires that, once we are in the set $B_{2\bar{\epsilon}}$, we can reach the set B_ϵ in one single step for any $\epsilon > 0$. The meaning of (4) is that, if we are in a point whose function value is poor with respect to the current record, then we guarantee a probability bounded away from 0 of accepting ascent steps, while if we are close to the record, we decrease the probability

of accepting ascent steps. Under suitable mild assumptions, the following theorem has been proved.

Theorem 2.1. Let N be a sufficiently large integer and let

$$\Delta F = \max_{x \in X \setminus B_\epsilon} \max_{y \in S(x,R) \cap X} [f(y) - f(x)].$$

Under the cooling schedule (4), if

$$c_k \geq (1+\mu)(N\Delta F / \log k), \qquad \mu > 0,$$

then

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

Proof. See Theorem 3 in Ref. 15. □

The integer N depends on R and the shape of X. If X is a convex set, N can be chosen approximately equal to the integer part of the ratio between the diameter of X and R. If $\bar{\epsilon}$ does not satisfy (5), it is still possible to prove that

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

From Theorem 2.1, it follows that

$$\lim_{k \to \infty} P[d(X_k, B_\epsilon) \leq R] = 1, \qquad \forall \epsilon > 0.$$

Therefore, asymptotically, we cannot sample too far from the region of the global optima, i.e., at a distance greater than R. The natural development at this point is to try to see whether it is possible to obtain that

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

We will consider this in the following sections.

3. Assumptions for the Convergence Results

In Ref. 15, the next candidate point X_{k+1} is generated inside the intersection of the feasible set X with a sphere with center in the current point Y_k and radius R. Here, we consider algorithms which satisfy the following

condition:

$$\text{supp}(D(\cdot, z_k)) = S(Y_k, R_k) \cap X,$$

where $\text{supp}(\cdot)$ denotes the support of a distribution. Therefore, the new candidate point X_{k+1}, generated by the distribution $D(\cdot; z_k)$, must belong to the intersection of the feasible set X with the sphere whose center is the current point Y_k and whose radius is R_k. The basic idea is that not only should the temperature t_k decrease to 0 when we are close to the current record [see (4)], but also that the same should be true for the radius R_k of the support sphere for the distribution $D(\cdot; z_k)$; i.e., we should have

$$R_k \geq \delta_2 > 0, \qquad \text{if } f(Y_k) - f_k^* > \epsilon, \tag{6a}$$

$$R_k = r_k, \qquad \text{otherwise}, \tag{6b}$$

where $\{r_k\}$ is a deterministic nonincreasing sequence converging to 0. The meaning of (6) is that, if we are in a point whose function value is poor with respect to the current record, then we guarantee the possibility of performing steps $X_{k+1} - Y_k$ whose length is bounded away from 0, while if we are close to the record, we decrease to 0 the maximum allowed steplength. We need to introduce some assumptions, more restrictive than the ones presented in Ref. 15, but still quite general. The first assumption restricts the possible choices for the distribution D of the next candidate point.

Assumption 3.1. For any z_k, the distribution $D(\cdot; z_k)$ of the next candidate point X_{k+1} is equivalent to the uniform distribution over $S(Y_k, R_k) \cap X$; i.e., there exist constants g, $G > 0$, $g \leq G$ such that

$$g[m(B)/m(S(Y_k, R_k) \cap X)] \leq D(B; z_k) \leq G[m(B)/m(S(Y_k, R_k) \cap X)],$$
$$\forall B \subseteq S(Y_k, R_k) \cap X, \forall Y_k \in X. \tag{7}$$

The second assumption introduces restrictions on the objective function f and the feasible set X.

Assumption 3.2. The feasible set X is compact, convex, and full-dimensional, and the objective function f is continuous; moreover, f has a unique global optimum over X (the extension to the case of finitely many global optima is trivial).

Note that the continuity of f and the compactness of X imply that f is uniformly continuous over X; i.e.,

$$\forall \delta > 0, \exists \gamma = \gamma(\delta) > 0 : d(x, y) \leq \gamma \Rightarrow |f(x) - f(x)| \leq \delta. \tag{8}$$

Next, we introduce an assumption about the speed of convergence to 0 for the sequences $\{c_k\}$ and $\{r_k\}$. First, we need to introduce some notation. Let

$$s_k = \min\left\{s: \sum_{i=0}^{s} r_{k+i} \geq 2\,\mathrm{diam}(X)\right\}, \qquad \mathrm{diam}(X) = \max_{x,y\in X} d(x,y),$$

and let

$$\Delta F_k = \max_{x\in X} \max_{y\in S(x,r_k)\cap X} [f(y) - f(x)].$$

Moreover, we define the following sequence of integers:

$$q_0 = 1, \qquad q_{j+1} = q_j + s_{q_j}, \qquad \forall j \in \{0, 1, \ldots\}.$$

Then, the assumption is the following.

Assumption 3.3. The sequences $\{c_k\}$ and $\{r_k\}$ decrease to 0 slowly enough; specifically, the following condition must be satisfied:

$$\sum_{j=1}^{\infty} \psi^{s_{q_j}} \exp\{-[s_{q_j}\Delta F_{q_j}/c_{q_{j+1}}]\} = \infty, \tag{9}$$

where $\psi \in (0, 1)$ is a given constant.

While condition (9) is quite general, it is quite difficult to appreciate its dependency on the two sequences $\{c_k\}$ and $\{r_k\}$. In particular, while the sequence $\{c_k\}$ appears explicitly in (9), the sequence $\{r_k\}$ appears only implicitly through s_{q_j} and ΔF_{q_j}. In order to make the last sequence appear explicitly, we can introduce two further requirements. The first one is that the sequence $\{r_k\}$ is updated only at iterations q_j, $j = 1, 2, \ldots$; i.e.,

$$r_k = r_{q_j}, \qquad \forall k \in [q_j, q_{j+1}).$$

Then, it follows that

$$s_{q_j} = \min\{s: sr_{q_j} \geq 2\,\mathrm{diam}(X)\} = \lceil 2\,\mathrm{diam}(X)/r_{q_j} \rceil.$$

The second requirement is that the function f is Lipschitz with Lipschitz constant L. Then, an upper bound for ΔF_{q_j} is given by Lr_{q_j}. If both these requirements hold, (9) is implied by the following condition:

$$\sum_{j=1}^{\infty} \psi^{2\,\mathrm{diam}(X)/r_{q_j}} \exp\{-2L\,\mathrm{diam}(X)/c_{q_{j+1}}\} = \infty,$$

where also the sequence $\{r_k\}$ appears explicitly.

Under the above assumptions, the following result can be proved.

Theorem 3.1. Let t_k and R_k be given by (4) and (6), respectively, and let Assumptions 3.1–3.3 hold. Then,

$$P[\exists k: Y_k \in B_\epsilon] = 1, \qquad \forall \epsilon > 0. \tag{10}$$

Proof. See Theorem 2 in Ref. 11. □

Basically, Theorem 3.1 states that, under the given assumptions, the record value f_k^* converges to the global optimum value f^* with probability 1. However, Assumptions 3.1–3.3 are not enough to ensure the stronger result

$$\lim_{k \to \infty} P[Y_k \in B_\epsilon] = 1, \qquad \forall \epsilon > 0. \tag{11}$$

In order to get this result, further assumptions have to be introduced. The first one imposes restrictions on the possible choices of the value $\bar{\epsilon}$ which appears in (4) and (6).

Assumption 3.4. The value $\bar{\epsilon}$ satisfies

$$m(S(x, \rho_2) \cap B_\epsilon) > 0, \qquad \forall x \in B_{2\bar{\epsilon}}, \qquad \forall \epsilon > 0.$$

This assumption is the correspondent of condition (5) with R replaced by ρ_2 and can be interpreted in the same way; see the comment after (5). We point out that a drawback of the above assumption is that generally it is not possible to know in advance a suitable value for $\bar{\epsilon}$. We delay the discussion on this difficulty to the following section.

Next, we need another assumption, basically requiring that the function f is locally Lipschitz in a neighborhood of the global optimum.

Assumption 3.5. There exists $v_1 > 0$ such that, $\forall x \in B_{\bar{\epsilon}}$ and $\forall y \in S(x, r_k)$,

$$f(y) \leq f(x) + v_1 r_k.$$

While Assumption 3.5 requires that the function does not increase too fast in a neighborhood of the global optimum, the next assumption requires that the function is not too flat. Indeed, it is required that, while we are in a neighborhood of the global optimum, it is always possible, with a strictly positive probability, to get close enough to the global optimum, both from the point of view of the function value and from the point of view of the distance.

Assumption 3.6. Let x^* be the global optimizer, and let $\epsilon \in (0, \bar{\epsilon}]$. For some $v_2 = v_2(\epsilon) > 0$, let D_k denote the event

$$\{f(Y_{k+1}) \le f(Y_k) - v_2 r_k, d(Y_{k+1}, x^*) \le d(Y_k, x^*) - (1/2) r_k\},$$

i.e., the event of getting close enough to x^*. Then, there exist $p > 0$ and a positive integer K_2 such that

$$P[D_k \mid Y_k = x, z_k] \ge p, \qquad \forall x \in B_{2\epsilon} \backslash B_{\epsilon/2}, \forall k \ge K_2, \forall z_k.$$

It can be proved that for instance Assumption 3.6 holds if f is twice continuously differentiable in a neighborhood of the global minimum and if the sufficient optimality conditions of the second-order are satisfied in the global optimum. However, the assumption is satisfied also for more general functions.

The final assumption requires that, in a neighborhood of the global optimum, the probability of accepting an ascent step decreases to 0 as the iteration counter k increases.

Assumption 3.7. There exists a sequence $b_k = b_k(\epsilon) \to 0$ such that

$$P[f(Y_{k+1}) > f(Y_k) \mid Y_k = x, R_k = r_k, z_k] \le b_k, \qquad \forall x \in B_{2\epsilon} \backslash B_{\epsilon/2}, \forall z_k. \tag{12}$$

While Assumption 3.7 is what we need for the proof of (11), we would like to introduce an alternative assumption, which is less general than Assumption 3.7, but gives a better indication about how $\{c_k\}$ and $\{r_k\}$ should be chosen in order to enforce it.

Assumption 3.8. There exists a sequence $\{u_k\}$ such that $c_k/u_k \to 0$ as $k \to \infty$, and a sequence $\rho_k = \rho_k(\epsilon) \to 0$ such that

$$m(X \cap S(x, r_k) \cap \{y : f(x) < f(y) < f(x) + u_k\}) / m(S(x, r_k) \cap X) \le \rho_k, \qquad \forall x \in B_{2\epsilon} \backslash B_{\epsilon/2}.$$

Intuitively, Assumption 3.8 requires that r_k decreases to 0 at a rate slower than u_k, and consequently at a rate also slower than c_k. Indeed, if u_k is smaller than r_k, the measure of the set of points in $S(x, r_k) \cap X$ with function value bigger than $f(x)$, but by not more than u_k, is typically also small with respect to the measure of the whole set $S(x, r_k) \cap X$. Under Assumption 3.1, Assumption 3.7 is always implied by Assumption 3.8. Indeed, for any $x \in B_{2\epsilon} \backslash B_{\epsilon/2}$, the probability in (12) can be split in the following sum

$$P[f(Y_{k+1}) \ge f(Y_k) + u_k \mid Y_k = x, R_k = r_k, z_k] + P[f(Y_k) < f(Y_{k+1}) < f(Y_k) + u_k \mid Y_k = x, R_k = r_k, z_k]. \tag{13}$$

By observing the Metropolis acceptance function (1), it follows that an upper bound for the first probability in (13) is the value $\exp\{-(u_k/c_k)\}$, while from Assumption 3.8 and (7) with

$$B = X \cap S(x, r_k) \cap \{y : f(x) < f(y) < f(x) + u_k\},$$

it follows that an upper bound for the second probability in (13) is $G\rho_k$. Therefore,

$$P[f(Y_{k+1}) > f(Y_k) \mid Y_k = x, R_k = r_k, z_k]$$
$$\leq b_k = \exp\{-(u_k/c_k)\} + G\rho_k \to 0,$$

i.e., Assumption 3.7 holds.

4. Convergence Results

Now, we are ready to present the new convergence results. The first one states the convergence with probability 1 of the sequence $\{Y_k\}$ to the global optimum.

Theorem 4.1. If Assumptions 3.1–3.7 hold, then there exists a value $\bar{\epsilon} > 0$ such that, if in (4) and (6) $\epsilon \in (0, \bar{\epsilon}]$ is chosen, then

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

Proof. See Theorem 3 in Ref. 11.

The next theorem states that, under the introduced assumptions, it is possible to guarantee also that the sequence $\{X_k\}$ of candidate points converges with probability 1 to the global optimum.

Theorem 4.2. Let $\bar{\epsilon}$ be the same as in Theorem 4.1. If Assumptions 3.1–3.7 hold, and if in (4) and (6) $\epsilon \in (0, \bar{\epsilon}]$ is chosen, then

$$\lim_{k \to \infty} P[X_k \in B_\epsilon] = 1, \qquad \forall \epsilon > 0. \tag{14}$$

Proof. See Theorem 4 in Ref. 11. □

As already remarked in the previous section, a drawback of the above result is that generally it is not possible to know in advance a suitable value for $\bar{\epsilon}$. However, it is possible to prove that, for any fixed value $\bar{\epsilon}$, a weaker

result holds, namely that, with probability 1, the sequence $\{X_k\}$ will lie in the limit in the set $B_{2\bar{\epsilon}}$.

Theorem 4.3. If Assumptions 3.1–3.3 and Assumption 3.7 hold, then

$$\lim_{k\to\infty} P[X_k \in B_{2\bar{\epsilon}}] = 1.$$

Proof. See Theorem 5 in Ref. 11. □

In this case, the value $\bar{\epsilon}$ acts like a predefined accuracy fixed by the user. We also note that the choice of $\bar{\epsilon}$ is an important task. By choosing it too small, the algorithm may perform too many free steps (i.e., steps with $t_k \geq \delta_1$ and $R_k \geq \delta_2$), without exploring locally the feasible region. On the other hand, by choosing it too big, the algorithm may perform small steps even when the current iterate is far away from the current record.

There is still a lot of space for further research in this field. In particular, it could be interesting to see whether it is possible to relax some of the assumptions. A partial answer is given in the following theorem, whose result is weaker than (11) and (14), but has been obtained by removing completely Assumptions 3.5–3.7.

Theorem 4.4. If Assumptions 3.1–3.4 hold, then

$$\lim_{i\to\infty} (1/i) \sum_{j=1}^{i} P[Y_j \notin B_{2\epsilon}] = 0.$$

Proof. See Theorem 6 in Ref. 11. □

5. Conclusions

This paper has dealt with the convergence problem of simulated annealing algorithms for continuous global optimization. After a short introduction about the simulated annealing approach to continuous global optimization, some recent convergence results from the literature have been presented. It has been noted that either these results are strong [both (11) and (14) hold], but proved under restrictive assumptions, or they are weak [only (11) holds], but proved under quite general assumptions. In this paper, strong convergence results have been derived without introducing assumptions which are too restrictive. The main idea of the paper is that the temperatures as well as the support of the distribution of the next candidate points should depend on the current value of the iterate. In particular, they

should be decreased to zero if the current iterate has a value close to the current record, while both can be arbitrarily chosen (but bounded away from zero) when the value of the current iterate is sufficiently worse than the record. While exploited only for the proof of convergence results, this idea seems to be reasonable for the development of cooling schedules for the application of simulated annealing algorithms. Besides the cooling schedule, some other things have to be specified in order to define a practical simulated annealing algorithm belonging to the class for which convergence, under the given assumptions, has been proved above. In particular, the distribution of the next candidate point has to be specified. A possible idea is to employ the collected information z_k in order to build a model of the function over the support sphere and then to define a distribution with a density connected to the model (with higher values where the model has lower values). In this way, the algorithm acts like a randomized local search similarly, but without the differentiability requirement, to algorithms based on stochastic differential equations (see e.g. Refs. 14, 16, 17), where each step is represented by the sum of an antigradient step and a random perturbation.

Finally, we remark that there are still many open questions in the field of convergence results for simulated annealing algorithms for continuous global optimization. In particular, relaxations of the assumptions under which the results have been proved could be searched for.

References

1. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., and Teller, A. H., *Equation of State Calculations by Fast Computer Machines*, Journal of Chemical Physics, Vol. 21, pp. 1087–1091, 1953.
2. Černy, V., *Thermodynamical Approach to the Travelling Salesman Problem: An Efficient Simulation Algorithm*, Journal of Optimization Theory and Applications, Vol. 45, pp. 41–51, 1985.
3. Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P., *Optimization by Simulated Annealing*, Science, Vol. 220, pp. 671–680, 1983.
4. Bohachevsky, I. O., Johnson, M. E., and Stein, M. L., *Generalized Simulated Annealing for Function Optimization*, Technometrics, Vol. 28, pp. 209–217, 1986.
5. Brooks, D. G., and Verdini, W. A., *Computational Experience with Generalized Simulated Annealing over Continuous Variables*, American Journal of Mathematical and Management Sciences, Vol. 8, pp. 425–449, 1988.
6. Coleman, T., Shalloway, D., and Wu, Z. J., *A Parallel Buildup Algorithm for Global Energy Minimizations of Molecular Clusters Using Effective Energy*

Simulated Annealing, Journal of Global Optimization, Vol. 4, pp. 171–185, 1994.

7. CORANA, A., MARCHESI, M., MARTINI, C., and RIDELLA, S., *Minimizing Multimodal Functions of Continuous Variables with the Simulated Annealing Algorithm*, ACM Transactions on Mathematical Software, Vol. 13, pp. 262–280, 1987.
8. JONES, A. E. W., and FORBES, G. W., *An Adaptive Simulated Annealing Algorithm for Global Optimization over Continuous Variables*, Journal of Global Optimization, Vol. 6, pp. 1–37, 1995.
9. ROMEIJN, H. E., and SMITH, R. L., *Simulated Annealing for Constrained Global Optimization*, Journal of Global Optimization, Vol. 5, pp. 101–126, 1994.
10. VANDERBILT, D., and LOUIE, S. G., *A Monte Carlo Simulated Annealing Approach to Optimization over Continuous Variables*, Journal of Computational Physics, Vol. 56, pp. 259–271, 1984.
11. LOCATELLI, M., *On the Convergence of Simulated Annealing Algorithms*, Technical Report 01-99, Dipartimento di Sistemi ed Informatica, Universitá di Firenze, 1999; also available at the web site http://www.dsi.unifi.it/users/locatell/dist4.ps
12. BELISLE, C. J. P., *Convergence Theorems for a Class of Simulated Annealing Algorithms on* R^d, Journal of Applied Probability, Vol. 29, pp. 885–892, 1992.
13. HAJEK, B., *Cooling Schedules for Optimal Annealing*, Mathematics of Operations Research, Vol. 13, pp. 311–329, 1988.
14. GELFAND, S. B., and MITTER, S. K., *Metropolis-Type Annealing Algorithms for Global Optimization in* R^d, SIAM Journal on Control and Optimization, Vol. 31, pp. 111–131, 1993.
15. LOCATELLI, M., *Convergence Properties of Simulated Annealing for Continuous Global Optimization*, Journal of Applied Probability, Vol. 33, pp. 1127–1140, 1996.
16. GELFAND, S. B., and MITTER, S. K., *Recursive Stochastic Algorithms for Global Optimization in* R^d, SIAM Journal of Control and Optimization, Vol. 29, pp. 999–1018, 1991.
17. ALUFFI-PENTINI, F., PARISI, V., and ZIRILLI, F., *Global Optimization and Stochastic Differential Equations*, Journal of Optimization Theory and Applications, Vol. 47, pp. 1–16, 1985.