INFORMATICA, 1997, Vol. 8, No. 3, 425–430

ON ONE OPTIMIZATION ALGORITHM OF SIMULATED ANNEALING WITH NOISE

Elvyra SENKIENĖ

Institute of Mathematics and Informatics
Akademijos 4, 2600 Vilnius, Lithuania

Abstract. In this paper we are concerned with global optimization, which can be defined as the problem of finding points on a bounded subset of R^m, in which some real-valued function $f(x)$ assumes its optimal value. We consider here a global optimization algorithm. We present a stochastic approach, which is based on the simulated annealing algorithm. The optimization function $f(x)$ here is discrete and with noise.

Key words: global optimization, simulated annealing.

1. Introduction. We consider one simulated annealing algorithm to search for the global extremum of the function in the discrete optimization problem. Simulated annealing is a stochastic method of finding a global extremum with an asymptotic convergence guarantee in probability. A global minimization problem can be formulated as a pair S, f, where $S = [A, B]^m \subset R^m$ is a bounded set on R^m and $f(x)$: $S \to R$ is an m-dimensional real-valued function, i.e., $x = (x^1, \ldots, x^m) \in S = [A, B]^m \subset R^m$, $A = (A^1, \ldots, A^m)$, $B = (B^1, \ldots, B^m)$, where $A^i \leqslant x^i \leqslant B^i$ and x^i, A^i, B^i, $i = \overline{1, m}$, take integer values. The problem now is to find a point $x_{\min} \in S$, such that $f(x_{\min})$ be globally minimal on S.

Let us denote the set $N(x_j) \subset S$ as a set of neighbors of $x_j = (x_j^1, \ldots, x_j^m)$, $(x_j \notin N(x_j))$ and $N(x_j^k) \subset S$ as a set of neighbors of x_j^k, $\overline{k = 1, m}$, $(x_j^k \notin N(x_j^k))$.

The search for the global minimum of $f(x)$ can be performed in such a manner: the n-th step of the algorithm is as follows:

$$x_n^i = x_{n-1}^i + \xi^i, \quad n = 1, 2, \ldots, \; i = \overline{1, m}, \tag{1}$$

where ξ^i, $i = \overline{1, m}$, are integers taking values with some probabilities:

1) ξ^i, $i = \overline{1,m}$, are random numbers taking integer values in the set $\{-1, 0, 1\}$; $P\{\xi^i = 0,\ i = \overline{1,m}\} = 0$, and the probability for any other combination ξ^i, $i = \overline{1,m}$, to appear is equal to $1/(3^m - 1)$;

2) $\xi^i \in S^i = \{A^i - x^i_{n-1}, A^i + 1 - x^i_{n-1}, \dots, B^i - x^i_{n-1}\} - \{0\}$, $i = \overline{1,m}$, with the same probability $p_i = 1/(B^i - A^i)$, i.e. $N(x_{n-1}) = S\backslash\{x_n\}$;

3) $P\{\xi^k = -1\} = P\{\xi^k = 1\} = \frac{1}{2}$, $\xi^i = 0$, $i = 1, 2, \dots, k-1, k+1, \dots, m$, i.e. we describe the transition to the next (neighboring) point along the coordinate k;

4) $\xi^k \in S^k = \{A^k - x^k_{n-1}, A^k + 1 - x^k_{n-1}, \dots, B^k - x^k_{n-1}\} - \{0\}$ with the probability $p_k = 1/(B^k - A^k)$, $\xi^i = 0$, $i = 1, 2, \dots, k-1, k+1, \dots, m$, i.e., we describe the transition to any point of the set $S = [A, B]^m$ with the same probability, i.e., all the points of set S are the neighbors along the given coordinate k (see Dzemyda *et al.*, 1990).

The probability of transition to the point x_n is defined by the formula:

$$P\{x_n\} = \begin{cases} 1, & \text{as } f(x_n) < f(x_{n-1}), \\ \exp\left\{-\frac{f(x_n) - f(x_{n-1})}{T_n}\right\}, & \text{as } f(x_n) \geqslant f(x_{n-1}), \end{cases} \tag{2}$$

and as $x_n \in N(x_{n-1})$ in the cases 1) and 2); and x_n such that $x^k_n \in N(x^k_{n-1})$ in the cases 3) and 4). $P\{x_n\} = 0$, as $x_n \notin N(x_{n-1})$ and $x^k_n \notin N(x^k_{n-1})$.

Equality (2) means that $P\{x_n\} = 1$ for $f(x_n) < f(x_{n-1})$; in the other case, as $f(x_n) \geqslant f(x_{n-1})$, a random number $\eta \in [0, 1]$ is generated, and as $\eta < \exp\left\{-\dfrac{f(x_n) - f(x_{n-1})}{T_n}\right\}$, we take a new point x_n; as $\eta \geqslant \exp\left\{-\dfrac{f(x_n) - f(x_{n-1})}{T_n}\right\}$, we stay at the point x_{n-1}.

Note that $x^i_n = A^i$, as $x^i_{n-1} = B^i$, $\xi^i = 1$; and $x^i_n = B^i$, as $x^i_{n-1} = A^i$, $\xi^i = -1$. $T_n = c/\ln[\ln(1 + n_0 + n)]$, $n = 1, 2, \dots$, is the number of a step, c is a positive constant, n_0 is a constant from $[1, \infty)$.

Algorithm (1), (2) is a special case of algorithms, described by Metropolis *et al.* (1953) and Mitra *et. al.* (1986).

Theorem 1. *If* $T_n \leqslant T_{n-1}$, $\lim\limits_{n\to\infty} T_n = 0$, *where* $T_n = c/\ln[\ln(1+n_0+n)]$, *and* $c \geqslant r \cdot L$, *then the simulated annealing algorithm* (1), (2) *converges in probability to the global minimum of* $f(x)$, *i.e.,* $\lim\limits_{n\to\infty} P\{|x_n - q| < \varepsilon\} = 1$, *where* q *is in the set of all the points which are the global minima of* $f(x)$.

Here in the cases 1) and 2) of algorithm (1), (2)

$$r = \min_{x_i \in (S \setminus S_m)} \max_{x_j \in S} \left[\sum_{k=1}^{m} \left(x_i^k - x_j^k \right)^2 \right]^{1/2},$$
$$S_m = \left\{ x_i \in S \mid f(x_j) \leqslant f(x_i),\ \forall x_j \in N(x_i) \right\},$$
$$L = \max_{x_i \in S} \max_{x_j \in N(x_i)} \left| f(x_i) - f(x_j) \right|;$$

and in the cases 3) and 4)

$$r = \min_{x_i \in (S \setminus S_m)} \max_{x_j \in S} \left[\sum_{k=1}^{m} \left(x_i^k - x_j^k \right)^2 \right]^{1/2},$$
$$S_m = \left\{ x_i \in S \mid f(x_j) \leqslant f(x_i),\ \forall x_j : x_j^k \in N(x_i^k) \right\},$$
$$L = \max_{x_i \in S} \max_{x_j : x_j^k \in N(x_i^k)} \left| f(x_i) - f(x_j) \right|.$$

The proof of Theorem 1 is presented by Senkienė (1994).

2. Theoretical knowledge. Simulated annealing algorithm (1), (2) is defined as a Markov chain $\{x_n\}$, $n = 1, 2, \ldots$, with the probability of transition (2). Usually a simulated annealing algorithm is defined as a Markov chain $\{x_n\}$ with the probability of transition:

$$P\{x_{n+1} = x_j \mid x_n = x_i\} = \begin{cases} \dfrac{q_{ij}}{q_i}, & \text{as } f(x_j) - f(x_i) < 0, \\ \dfrac{q_{ij}}{q_i} \exp\left\{ -\dfrac{f(x_j) - f(x_i)}{T_n} \right\}, & \text{as } f(x_j) - f(x_i) \geqslant 0, \end{cases} \tag{3}$$

where $x_i, x_j \in S,\ i \neq j,\ x_j \in N(x_i)$, $\dfrac{q_{ij}}{q_i}$ is a probability of generating a point $x_j \in N(x_i)$ from the point $x_i \in S$ $\left(\dfrac{1}{q} \sum_{x_j \in N(x_i)} \dfrac{q_{ij}}{q_i} = 1 \right)$ (see Mitra *et al.*, 1986; Gelfand and Mitter, 1989).

In some physical problems the difference of energy $f(x_j) - f(x_i)$ can be calculated only with noise η_n (see Gelfand and Mitter, 1989). Then the simulated annealing algorithm is defined as a Markov chain with the following transition probability:

$$P\{x_{n+1} = x_j \mid x_n = x_i\}$$

$$= \begin{cases} \dfrac{q_{ij}}{q_i}, & \text{as } f(x_j) - f(x_i) + \eta_n < 0, \\ \dfrac{q_{ij}}{q_i} \exp\left\{\dfrac{f(x_j) - f(x_i) + \eta_n}{T_n}\right\}, & \text{as } f(x_j) - f(x_i) + \eta_n \geqslant 0. \end{cases} \tag{4}$$

Denote that the noise η_n is random variables of normal distribution with mean 0 and variance σ^2. Then (see Gelfand and Mitter, 1989), if $T_n \to 0$ and $\sigma_n = o(T_n)$ as $n \to \infty$, in both cases the denote Markov chains of simulated annealing are equivalent and the theorem of convergence of simulated annealing algorithm in probability to the global minimum of $f(x)$ with noise is correct only if this convergence to the global minimum of the function $f(x)$ without noise is correct.

3. Fundamental results. Let the optimized function $f(x)$ can be measured with noise, i.e. the difference $f(x_n) - f(x_{n-1})$, $n = 1, 2, \ldots$, in (2) can be calculated only with noise η_n, where η_n is random variables of normal distribution with mean 0 and variance σ_n^2. Then the presented simulated annealing algorithm (1) is defined as a Markov chain $\{x_n\}$, $n = 1, 2, \ldots$, with the following transition probability:

$$P\{x\} = \begin{cases} 1, \quad \text{as} \quad f(x_n) - f(x_{n-1}) + \eta_n < 0, \\ \exp\left\{-\dfrac{f(x_n) - f(x_{n-1}) + \eta_n}{T_n}\right\}, \\ \qquad\qquad \text{as} \quad f(x_n) - f(x_{n-1}) + \eta_n \geqslant 0, \end{cases} \tag{5}$$

and as $x_n \in N(x_{n-1})$. $P\{x_n\} = 0$, as $x_n \notin N(x_{n-1})$.

We formulate a theorem analogous to Theorem 1, where the difference of the function $f(x_n) - f(x_{n-1})$, $n = 1, 2, \ldots$ is measured with noise η_n, $n = 1, 2, \ldots$.

Theorem 2. *If $T_n \leqslant T_{n-1}$, $\lim_{n\to\infty} T_n = 0$, where $T_n = c/\ln[\ln(1 + n_0 + n)]$, $c \geqslant r \cdot L$ and $\sigma_n = o(T_n)$, as $n \to \infty$, then simulated annealing algorithm* (1), (5) *converges in probability to the global minimum of $f(x)$ only if simulated annealing algorithm* (1), (2) *converges in probability to the global minimum of $f(x)$.* (Constants r and L are in the definition of Theorem 1).

The proof of Theorem 2 follows from the papers of Senkienė (1994), Senkienė (1996) and Gelfand and Mitter (1989).

REFERENCES

Dzemyda G., E. Senkienė, J. Valevičienė (1990). Simulated annealing for parameter grouping. *Informatica,* **1**(1), 20–39.

Gelfand S.B, S.K. Mitter (1989). Simulated annealing with noise or in precise energy measurements. *J. Optimiz. Theory and Appl.,* **62**(1), 49–62.

Metropolis N., A.W. Rosenbluth, A.H. Teller, E. Teller (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics,* **21**(6), 1087–1092.

Mitra D., F. Romeo, A. Sangiovanni-Vincentelli (1986). Convergence and finite-time behaviour of simulated annealing. *Adv. Appl. Prob.,* **18**(3), 717–771.

Senkienė E. (1994). On the convergence of one global optimization algorithm. *Problemy upravlenija i informatiki,* **4**, 81–88.

Senkienė E. (1996). On one algorithm of the global optimization. *Problems of Management and Informatics,* **6**, 31–38 (in Russian).

Received August 1997

E. Senkienė graduated from Vilnius University, Lithuania, in 1966. She received a Ph.D. degree of mathematical sciences from Vilnius University, Vilnius, Lithuania, in 1974. She is a senior researcher at the Department of Optimization, Institute of Mathematics and Informatics, Vilnius, Lithuania. Her research interests include stochastic processes and optimization problems.

APIE VIENĄ FUNKCIJOS, STEBIMOS SU TRIUKŠMU, OPTIMIZACIJOS ALGORITMĄ

Elvyra SENKIENĖ

Straipsnyje nagrinėjamas vienas globalinės optimizacijos algoritmas funkcijos minimumui surasti vadinamas "simulated annealing" algoritmu. Optimizuojama funkcija čia yra diskretinė ir stebima su triukšmu.