ISSN 1064-5624, Doklady Mathematics, 2011, Vol. 83, № 2, pp. 1–4. © Pleiades Publishing, Ltd., 2011.
Original Russian Text © Yu.G. Evtushenko, M.A. Posypkin, 2011, published in Doklady Akademii Nauk, 2011, Vol. 437, № 2, pp. 168-172.

COMPUTER SCIENCE

Versions of a Method of Nonuniform Coverings for Global Optimization of Partial-Integer Nonlinear Problems

Academician Yu.G. Evtushenko[a] and M.A. Posypkin[b]

[a] *Computing Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119333 Russia*
e-mail:evt@ccas.ru
[b] *Institute of System Analysis, Russian Academy of Sciences, pr. 60-letiya Oktyabrya 9, Moscow, 117312 Russia*
e-mail:mposypkin@gmail.com

DOI: 10.1134/S1064562411020074

The method of nonuniform coverings as applied to the global optimization of functions of several variables was proposed in 1971 in [1] and was further developed in numerous works, e.g., in [2]–[5]. Various versions of the method were implemented as software codes and were used for computations on multiprocessor systems [4, 5].

This paper gives a more general treatment of the method than in [1, 2]. The method is applied to the simplest nonlinear programming problem of finding a global isolated minimum. The computations were performed with and without using the integer-valuedness condition. The introduction of this condition led to a considerable reduction in the computation time.

Given a continuous function $f : \mathbb{R}^n \to \mathbb{R}$, the problem is to find its global minimum on the feasible set $X \subseteq \mathbb{R}^n$:

$$f_* = \text{glob} \min_{x \in X} f(x) = f(x_*), \tag{1}$$

where x_* is any global minimizer point yielding the global minimum f_*. For this problem, the solution set X_* and the ϵ-optimal solution set X_ϵ are defined as

$$\begin{aligned} X_* &= \{x \in X : f(x) = f_*\}, \\ X_\epsilon &= \{x \in X : f(x) \le f_* + \epsilon\}, \qquad \epsilon > 0. \end{aligned} \tag{2}$$

Assume that X_* is not empty. The goal is to find at least one point of X_ϵ.

Given a set $Z \subseteq \mathbb{R}^n$, a function $f(x) : \mathbb{R}^n \to \mathbb{R}$, and a constant $\lambda \in \mathbb{R}$, we define the Lebesgue set $S(Z, f(x), \lambda) = \{x \in Z : f(x) \ge \lambda\}$ and the open Lebesgue set $S'(Z, f(x), \lambda) = \{x \in Z : f(x) > \lambda\}$.

Consider the collection of sets $\{X_i\}$, $X_i \subseteq \mathbb{R}^n$, $i = 1, 2, \ldots, k$. On X_i, we define a minorant $\mu_i(x)$ such that $f(x) \ge \mu_i(x)$ for all $x \in X_i$. Given a feasible point $x_r \in X$ and a collection of sets $\{S_i\}$ satisfying the conditions

$$S_i \subseteq S(X_i, \mu_i(x), f(x_r) - \epsilon), \qquad i = 1, 2, \ldots, k, \tag{3}$$

we say that $\{S_i\}$ covers the set X if

$$X \subseteq \bigcup_{i=1}^{k} S_i. \tag{4}$$

Theorem 1. *A sufficient condition for a feasible point x_r to be an ϵ-optimal solution of problem* (1) *that satisfies the estimate*

$$f(x_r) \geq f_* \geq f(x_r) - \epsilon \tag{5}$$

on the set X is that there exists a collection of sets $\{X_i\}$ and a collection of minorants $\{\mu_i(x)\}$ for which the covering condition (4) *holds.*

Theorem 1 is a theoretical basis for implementing various computational schemes for the nonuniform covering method. The cardinality and structure of the set X can be arbitrary. To use Theorem 1 in global minimization, we need a method for finding a feasible point x_r and verifying condition (4). Consider the case when the feasible set in problem (1) is given by functional constraints:

$$X = \{x \in \mathbb{R}^n : g(x) \leq 0_m\}, \tag{6}$$

where $g(x) : \mathbb{R}^n \to \mathbb{R}^m$ is a continuous vector function.

Assume that the feasible set X is bounded and $X \subseteq P$, where P is an n-dimensional parallelepiped.

Given $\delta \in \mathbb{R}$, the δ-feasible set is defined as

$$X^\delta = \{x \in \mathbb{R}^n : g^j(x) \leq \delta,\ j = 1, 2, \dots, m\}. \tag{7}$$

The set X^δ can be equivalently defined using the function

$$\phi(x) = \max\left(g^1(x), \dots, g^m(x)\right) : X^\delta = \{x \in P : \phi(x) \leq \delta\}.$$

Since $g(x)$ is a continuous function, $\phi(x)$ is continuous as well.

Define $\underline{\delta} = \inf\{\delta : X^\delta \neq \emptyset\}$, $\bar{\delta} = \sup\{\delta : X^\delta \subseteq P\}$.

For $\underline{\delta} < \delta < \bar{\delta}$, we set

$$f_*^\delta = \min_{x \in X^\delta} f(x).$$

The function f_*^δ is called the sensitivity function. It has the following well-known properties [6, 7]:

1. $f_*^0 = f_*$.

2. The set X^δ does not decrease and the sensitivity function monotonically does not increase as δ grows:

$$X^{\delta_1} \subseteq X^{\delta_2}, \quad f_*^{\delta_1} \geq f_*^{\delta_2} \quad \text{for any} \quad \delta_1, \delta_2, \qquad \underline{\delta} < \delta_1 \leq \delta_2 < \bar{\delta}. \tag{8}$$

3. For any δ_0 such that $\underline{\delta} < \delta < \bar{\delta}$, the sensitivity function is right continuous:

$$\lim_{\delta \to \delta_0 + 0} f_*^\delta = f_*^{\delta_0}. \tag{9}$$

For $\epsilon \in \mathbb{R}_+$ and δ_1, δ_2, $\underline{\delta} < \delta_1 \leq \delta_2 < \bar{\delta}$, define the set

$$X_\epsilon^{\delta_1, \delta_2} = \left\{x \in X^{\delta_2} : f(x) \leq f_*^{\delta_1} + \epsilon\right\}.$$

In the special case of $\delta_1 = 0$ and $\delta_2 = \delta$, it is called the set of ϵ, δ-optimal solutions and is denoted by X_ϵ^δ:

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

Consider the collection of sets $P_1, P_2, \dots, P_k$, $P_i \subseteq P$. Define minorants $\mu_i(x)$ such that $f(x)(x) \geq \mu_i(x)$ for all $x \in P_i$ and minorants $\nu_i(x)$ such that $\phi(x) \geq \nu_i(x)$ for all $x \in P_i$. Let $\underline{\delta} < \delta_1 \leq \delta_2 < \bar{\delta}$, $x_r \in X^{\delta_2}$, and the sets S_i satisfy the relation

$$S_i \subseteq S(P_i, \mu_i(x), f(x_r) - \epsilon) \cup S'(P_i, \nu_i(x), \delta_1). \tag{10}$$

In this case, the covering condition has the form

$$P = \bigcup_{i=1}^{k} S_i. \tag{11}$$

Theorem 2. *Let $\underline{\delta} < \delta_1 \leq \delta_2 < \bar{\delta}$. Let a point $x_r \in X^{\delta_2}$ and a collection of sets $\{P_i\}$ be given such that the covering condition* (11) *holds. Then $x_r \in X_\epsilon^{\delta_1,\delta_2}$ we have the estimate*

$$f_*^{\delta_1} + \epsilon \geq f(x_r) \geq f_*^{\delta_2}. \tag{12}$$

Consider the following two special cases of Theorem 2, which are the most important in practice. In the first case, $\delta_1 < 0$ and $\delta_2 = 0$, the point x_r is an approximate feasible solution of the problem that satisfies the inequality $f_*^{\delta_1} + \epsilon \geq f(x_r) \geq f_*$. Since the sensitivity function is not left continuous, the difference between the approximate and exact solutions can be arbitrarily large. Moreover, this approach is inapplicable to problems with $\delta = 0$.

In the second case, $\delta_1 = 0$ and $\delta_2 > 0$, the point x_r is an approximate but possibly infeasible solution that satisfies the inequality $f_* + \epsilon \geq f(x_r) \geq f_*^{\delta_2}$. According to Proposition (9) and Theorem 2, as ϵ and δ_2 decrease, the value of $f(x_r)$ tends to f_*.

Consider a possible scheme for the implementation of the nonuniform covering method, namely, the bisection algorithm proposed in [2, 3]. The consideration is restricted to the special case of $\delta_1 = 0$ and $\delta_2 = \delta > 0$. The other important case ($\delta_1 < 0$, $\delta_2 = 0$) is considered in an analogous fashion. As before, assume that we are given an n-dimensional parallelepiped P such that its faces are parallel to the coordinate planes and it contains the set X^δ. In the bisection method, the sets P_i are also parallelepipeds obtained by partitioning the original one.

Let $\{x_i\}$ be a set of points that belong to P. The value of $f(x_i)$ is calculated at every point x_i. The record point x_r in the set and the record point x'_r in X are determined using the following result rule.

REC-UPDATE Rule. If $x_i \in X^\delta$ and $f(x_i) < f(x_r)$ or x_r is not defined, then execute $x_r := x_i$. If $x_i \in X$ and $f(x_i) < f(x'_r)$ or x'_r is not defined, then execute $x'_r := x_i$.

In addition to the record point x_r, we also determine the record value

$$f_r = \begin{cases} f(x_r), & \text{if } x_r \text{ is defined,} \\ \infty & \text{otherwise.} \end{cases}$$

Let $f(x)$ and $\phi(x)$ satisfy the Lipschitz condition with constants ℓ_1 and ℓ_2, respectively. Let x_i be the center of the parallelepiped P_i. The support minorants $\mu_i(x)$ and $\nu_i(x)$ for $f(x)$ and $\phi(x)$ are defined as

$$\mu_i(x) = f(x_i) - \ell_1 \|x - x_i\|, \qquad \nu_i(x) = \phi(x_i) - \ell_2 \|x - x_i\|. \tag{13}$$

Theorem 3. *Let* ℓ_1 *and* ℓ_2 *be the Lipschitz constants for the functions* $f(x)$ *and* $\phi(x)$. *Then the following two assertions hold:*

1. The set $S_i = \{x \in P_i : \|x - x_i\| < \rho_i\}$, *where*

$$\rho_i = \max\left\{\frac{f(x_i) - f_r + \epsilon}{\ell_1}, \frac{\phi(x_i)}{\ell_2}\right\} \tag{14}$$

satisfies (10).

2. If the REC-UPDATE rule is used, then $\rho_i \geq \rho_*$, *where* $\rho_* = \min\left\{\frac{\epsilon}{\ell_1}, \frac{\delta}{\ell_2}\right\}$.

Below is the simplest version of the method based on longest-edge bisections of parallelepipeds.

Algorithm BISECT

Input parameters:

P	is the original parallelepiped;
ϵ and δ	are the prescribed accuracy of the objective function and the constraints;
ℓ_1 and ℓ_2	are the Lipschitz constants for the objective function and the constraints.

Output parameters:

x_r	is a record point;
x'_r	is a feasible record point.

Internal variables:

$\mathcal{L}$ is the list of covered parallelepipeds.

Step 1. Place P in the list $\mathcal{L}$: $\mathcal{L} = \{P\}$. Set $f_r = \infty$.

Step 2. Choose the current parallelepiped $P_i \in \mathcal{L}$.

Step 3. Calculate the function value at the center x_i of P_i and update the record points x_r and x'_r according to the REC-UPDATE rule.

Step 4. Use x_i as the center of a covering ball, and calculate the radius ρ_i of the covering ball by formula (14).

Step 5. If the covering ball contains P_i, then go to **Step 2**. Otherwise, divide P_i in the longest-edge direction into two identical new parallelepipeds, which are added to the list $\mathcal{L}$.

Step 6. If the list is empty, then the algorithm terminates; otherwise, go to **Step 2**.

Steps 2–6 in the algorithm BISECT are cyclically repeated. Theorem 4 gives an upper bound for the number of iterations in this cycle.

Theorem 4. *The algorithm BISECT terminates after a finite number of iterations not exceeding the values*

$$4\left(\frac{\rho_*}{\rho}\right)^{\theta(n)} - 1, \tag{15}$$

where

$$\theta(n) = \frac{2}{\log_2\left(1 - \frac{3}{4n}\right)}$$

and ρ *is the radius (half the distance between two most distant points) of the parallelepiped* P.

Theorem 5. *If the algorithm BISECT terminates in a finite number of steps, then the following two assertions hold:*

1. *If no record point x_r is found, then the feasible set X of problem* (6) *is empty. Otherwise, the record point x_r found is an ϵ, δ-optimal solution of problem* (6) *that satisfies the inequality*

$$f_* + \epsilon \geq f_r \geq f_*^\delta. \tag{16}$$

2. *If a feasible record point x_r' is found, then the feasible set of solutions to problem* (6) *is not empty and*

$$f(x_r') \geq f_*. \tag{17}$$

This algorithm can be successfully used to solve partially integer programming problems. For this purpose, the parallelepiped-partitioning method and the record-updating rule in BISECT have to be modified. Consider problem (1) in which constraints (6) are supplemented with the integer-valuedness conditions $x_j \in \mathbb{Z}$, $j \in J$, where J is the index subset of integer variables.

In the bisection of the parallelepiped, the new boundaries of the resulting parallelepipeds are rounded off to the nearest integers if partition is in the direction of an integer variable. Let $Q = [a^1, b^1] \times [a^2, b^2] \times \cdots \times [a^n, b^n]$ be the parallelepiped to be partitioned, and let

$$s = \arg \min_{i=1,\ldots,n} |a^i - b^i|$$

be the index of the longest edge in the parallelepiped. The partition procedure yields two parallelepipeds

$$\begin{aligned} Q_1 &= [a^1, b^1] \times \cdots \times [a^s, c^s] \times \cdots \times [a^n, b^n], \\ Q_2 &= [a^1, b^1] \times \cdots \times [d^s, b^s] \times \cdots \times [a^n, b^n], \end{aligned}$$

where

$$c^s = \begin{cases} \left\lfloor \dfrac{a^s + b^s}{2} \right\rfloor, & \text{if } \ s \in J, \\ \dfrac{a^s + b^s}{2}, & \text{if } \ s \notin J, \end{cases} \qquad d^s = \begin{cases} \left\lceil \dfrac{a^s + b^s}{2} \right\rceil, & \text{if } \ s \in J, \\ \dfrac{a^s + b^s}{2}, & \text{if } \ s \notin J. \end{cases}$$

Before applying the REC-UPDATE rule, the point x_i is transformed as follows:

$$x_i^j := \begin{cases} \lfloor x_i^j \rfloor, & \text{if } \ j \in J, \\ x_i^j, & \text{if } \ j \in J. \end{cases}$$

The other steps in BISECT remain unchanged.

In practice, various acceleration techniques are added to the above basic implementation of the bisection method. Specifically, for differentiable functions, it is more effective to use the minorant следующей миноранты:

$$\mu_i(x) = f(x_i) + \langle f_x(x_i), x - x_i \rangle + \frac{k_i}{2} \|x - x_i\|^2, \tag{18}$$

where k_i is a lower bound for the Hessian spectrum of $f(x)$ on P_i. Additionally, the computations can be accelerated if the search domain is reduced by eliminating the part of the current parallelepiped that does not contain any optimum in the case when it cannot be entirely eliminated. The results concerning the finite number of steps in the basic BISECT version and the properties of approximate solutions remain valid in the case of more sophisticated versions.

Variants of the nonuniform covering method have been successfully used to solve large-scale problems. For example, the interaction energy of a molecular cluster was minimized in [4]. Specifically, the nonuniform covering method was used to minimize a function of 255 variables.

We consider only the following simplest example (see [8]):

$$\begin{aligned} f(x) &= x^1 \to \min, \\ g^1(x) &= (x^1-5)^2 + 2(x^2-5)^2 + (x^3-5)^2 - 18 \le 0, \\ g^2(x) &= 100 - (x^1+7-2x^2)^2 - 4(2x^1+x^2-11)^2 - 5(x^3-5)^2 \le 0. \end{aligned} \tag{19}$$

Numerical results for various δ

δ	Record x_r	$\phi(x_r)$	Computation time, s	Number of iterations
-0.01	$(3.722, 7.276, 7.452)$	-0.00204	0.61	10602
-0.0001	$(3.721, 7.15, 2.331)$	-0.00001	24.57	506351
0.0001	$(0.997, 4.007, 4.999)$	0.00007	7.94	165547
0.01	$(0.965, 4.071, 4.994)$	0.00993	0.16	2671

A feature of this problem is that the global minimum is reached at the isolated feasible point $x_* = (1,4,5)$ with $f(x_*) = 1$, $g^1(x_*) = 0$, and $g^2(x_*) = 0$ (the point (1,3,5)) was mistakenly indicated in [8]). The global minimization method used in [8] produced the feasible point $\bar{x} = (3.747692, 7.171420, 2.362317)$ with $f(\bar{x}) = 3.747692$ in 33.703 seconds on a Pentium IV 2.53 GHz computer.

The set P in the nonuniform covering method was specified as the parallelepiped $-10 \le x^i \le 10$, $i = 1,2,3$, which contains the feasible set of problem (19). The minorants were defined by (18), and $\epsilon = |\delta|$ was used. The experiments were performed on an Intel Core 2 Quad 2.33 GHz personal computer with a single processor core. The numerical results are presented in the table. The value of $f(x_r)$ coincides with the first component of the vector x_r given in the second column.

For $\delta > 0$, the approximate solutions were found to be close to the optimal solution x_*. For $\delta > 0$, the points found are feasible and close to $\bar{x}$ with the resulting objective function value being smaller in the third digit than that found in [8].

This problem was used to explore the possibility of finding an integer solution. For this purpose, all the variables were assumed to be integers. The computations were performed for three versions at $\epsilon = \delta = 0$. The first version relied on the basic BISECT scheme with minorants (13) and the Lipschitz constants estimated from above by using interval analysis on the current parallelepipeds. Minorants (18) were used in the second version. The part of the parallelepiped that does not contain the optimal solution was additionally eliminated in the third version. The global minimum was found after 585, 121, and 55 iterations in the first, second, and third versions, respectively. Thus, the computations in this problem are considerably accelerated by introducing the integer-valuedness condition. Therefore, for complicated problems, it seems reasonable to introduce this condition deliberately in order to find an approximate solution of the original continuous problem.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 09-01-12098-ofi_m, 10-01-90011-BEL_a), by Program P-14 of the Presidium of the Russian Academy of Sciences, and by the program “Leading Scientific Schools” (project no. NSh-4096.2010.1).

Литература

1. Yu.G. Evtushenko, Zh. Vychisl. Mat. Mat. Fiz. **11**, 1390–1403 (1971).
2. Yu.G. Evtushenko and V.A. Rat’kin, Tekh. Kibern., No. 1, 119–127 (1987).
3. H. Ratchek and J. Rokne, *New Computer Methods for Global Optimization* (Ellis Horwood, Chichester, 1988).
4. Yu.G. Evtushenko, V.U. Malkova, and A.A. Stanevichyus, Comput. Math. Math. Phys. **49**, 246–260 (2009) [Zh. Vychisl. Mat. Mat. Fiz. **49**, 255–269 (2009)].
5. Y. Evtushenko, M. Posypkin, and I. Sigal, Comput. Sci. Res. Develop. **23**, 211–215 (2009).
6. J.P. Evans and F.J. Gould, Oper. Res. **18**, 107–118 (1970).
7. A.I. Golikov and G.G. Kotkin, Zh. Vychisl. Mat. Mat. Fiz. **28**, 1461–1474 (1988).
8. H. Tuy, J. Glob. Optim. **47**, 485–501 (2010).