DOI 10.1007/s10559-015-9753-6
Cybernetics and Systems Analysis, Vol. 51, No. 4, July, 2015

SYNTHESIS OF CUTTING AND SEPARATING PLANES IN A NONSMOOTH OPTIMIZATION METHOD[1]

E. A. Vorontsova[a] **and E. A. Nurminski**[b]

UDC 519.8

Abstract. *A solution algorithm is proposed for problems of nondifferentiable optimization of a family of separating plane methods with additional clippings generated by the solution of an auxiliary problem of the cutting plane method. The convergence of this algorithm is proved, and the results of computational experiments are given that demonstrate its overall computational efficiency compared to that of well-known leaders in this field. Transportation-type problems with constraints on flows are reduced to problems of projection of a sufficiently remote point onto an admissible set.*

Keywords: *convex optimization, separating plane method, cutting plane method.*

INTRODUCTION

Let $x=(x^1, \ldots, x^n)$ be a vector of an n-dimensional Euclidean space $\Re^n$ with the usual scalar product xy. In $\Re^n$, the following problem of unconstrained convex nondifferentiable optimization (NDO) is considered: $\min\limits_{x \in \Re^n} f(x)$, where $f(x)$ is a convex nondifferentiable function. We consider that this problem is solvable.

Such problems arise in various scientific and technical fields, for example, in solving problems of continuum mechanics with allowance for friction [1], control theory [2], economy [3–5], etc. Moreover, the progress in the field of development and implementation of NDO methods makes it possible to construct more efficient methods for solving high-dimensional optimization problems.

This article is devoted to the further investigation of the efficient method from [21] for solving problems of multidimensional convex NDO without constraints that does not require additional information on the internal structure of the function being optimized and is a representative of the so-called black-box optimization. We suppose that the entire accessible information on the objective function $f(x)$ of a problem is provided by a subgradient oracle, and, at an arbitrary point $\bar{x} \in \Re^n$, only the value of the function $f(\bar{x})$ and a subgradient $g \in \partial f(\bar{x})$ arbitrarily chosen from the subdifferential $\partial f(\bar{x})$ of the function $f(x)$ can be found.

Oracle-type schemes for minimizing smooth and nonsmooth functions have essential distinctions. Oracles of differentiable functions allow one to construct convergent minimizing relaxation sequences (see, for example, [6]). For nonsmooth functions, this possibility is inapplicable in principle since an arbitrary chosen subgradient does not determine the relaxation direction [7]. In actual fact, methods of convex NDO use only separability properties, which considerably decreases their convergence rate.

[1]This work is carried out within the framework of the Federal Target Program "Research and Development in Priority Fields of S&T Complex of Russia for 2014–2020," agreement 14.604.21.0052 of June 30, 2014 with the Russian Ministry of Education and Science. Unique project identifier RFMEFI60414X0052.

[a]Schools of Natural Sciences of the Far East Federal University, Vladivostok, Russia, *vorontsovaea@gmail.com*. [b]Far East Federal University, Vladivostok, Russia, *nurminskiy.ea@dvfu.ru*. Translated from Kibernetika i Sistemnyi Analiz, No. 4, pp. 137–150, July–August, 2015. Original article submitted October 27, 2014.

In [8], lower bounds are found for complexity estimates of the methods using only oracles for different classes of optimization problems. It turned out that, for the class of convex NDO problems being considered, any method whose convergence rate is larger than $O(q^{k/n})$ is absent, where $q<1$ is an absolute constant and k is the number of iterations of the corresponding method. In any of such methods, any convergence rate estimate uniform with respect to the space dimension n cannot be better than a rate estimate of order of $O(k^{-1/2})$. It is proved that the center-of-gravity method [10, 11] and subgradient method [32] have such convergence rates. Thus, the two first methods developed for solving nonsmooth minimization problems turned out to be unimprovable with respect to their convergence characteristics. In practice, the center-of-gravity method is inapplicable since the operation of finding the center of gravity of a convex set in a multidimensional space is a very complicated problem. The subgradient method proposed for the first time by N. Z. Shor [32] has the simplest computational scheme

$$x_{k+1} = x_k - \lambda_k g_k, \; g_k \in \partial f(x_k), \; k = 0, 1, \ldots,$$

and converges under very acceptable conditions. Its practical computational efficiency depends on methods of control of stepsizes λ. The best choice law for a stepsize λ is recognized to be the B. T. Polyak law [12] but under the condition that the optimum value is known in advance. Among other methods of step adjustment, the technique from [13] can also be mentioned. The subgradient method with this step regulation has shown a practically linear convergence rate in the case of well-known tests. However, it should be noted that this statement about the convergence rate of NDO methods is uniformly valid with respect to the dimension of the space of variables. More efficient schemes can be developed for moderate-dimensional problems [9].

The next stage of the development of NDO methods became the emergence of the so-called bundle methods [14–16] whose representative is the level method [17] developed in 1995. Among recent publications in the field of bundle methods, mention may be made of the idea of splitting into subspaces of smoothness–nonsmoothness, which is called the VU-algorithm [18]; in the field of NDO methods without oracles, we note a special technique of smoothing with the subsequent application of gradient schemes of smooth minimization [19] for nonsmooth functions with a definite structure [20].

The projection algorithm SPACLIP for solving problems of minimization of nonsmooth functions that is proposed in this article is a result of the further development and improvement of separating plane methods [21–23] that have a number of important theoretical and computing features.

In computational experiments, the SPACLIP algorithm was used for solving the transportation problem that not only is one of the most widespread in economic applications but also has definite symbolic importance since the development of NDO began exactly with problems of this type [37].

In the matrix statement of this problem, which is considered in this article, upper and lower bounds are imposed on volumes of deliveries, which complicates the application of the method of potentials and simplex method to the solution of such problems and especially high-dimensional problems. Of definite interest is the reduction of the transportation problem to the problem of finding the projection onto a shift of the admissible set.

PROJECTION ALGORITHM SPACLIP

Separating plane methods [21] are based on the idea of replacement of the initial minimization problem by the problem of computing the Fenchel–Moreau conjugate at zero,

$$f(x^*) = \min_{x \in \Re^n} f(x) = -\sup_x \{0 \cdot x - f(x)\} = -f^*(0), \; x^* \in \partial f^*(0), \qquad (1)$$

where the function $f^*(g) = \sup_x \{gx - f(x)\}$ is the Fenchel–Moreau conjugate of $f(x)$. Problem (1) can be interpreted as the problem of finding the point of intersection of the plot of the conjugate with the vertical line $0 \times \Re_+$ (Fig. 1). We consider that $f(0) = 0$ and that the origin of coordinates of the space of primal variables is not the solution of the minimization problem, i.e., $f(x^*) < 0$. In Fig. 1, the dotted line represents the plot of the conjugate $f^*(g)$. The support vector of the tangent hyperplane at the point $(0, f^*(0))$ to the epigraph of the conjugate determines the solution of problem (1) up to normalization.

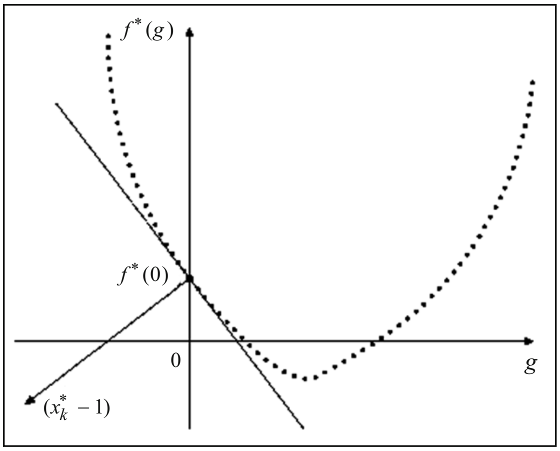

Fig. 1. Graphical interpretation of problem (1).

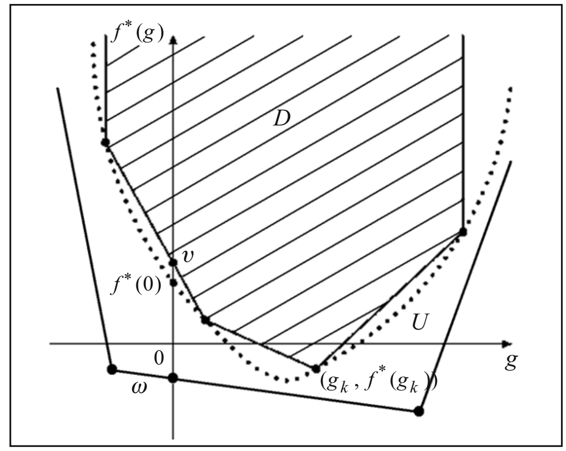

Fig. 2. Illustration of the standard separating plane method.

In the separating plane method (SPM), the epigraph of the conjugate f^* is approximated by internally and externally convex polyhedral sets D and U (Fig. 2). Refining these approximations at each iteration in the neighborhood of the vertical line $0\times\Re_+$, we obtain convergent lower w and upper υ estimates for $f^*(0)$.

The sets D and U are modified by the addition of clippings (for U) or new points $(\varepsilon, g(\varepsilon))$ located on the plot of f^*.

The basic version of the SPM does not guarantee monotonicity, especially in the case of nearing an extremum. To improve the monotonicity property of the method, we introduce the following additional clipping with respect to the upper-bound estimate υ obtained from the internal approximation of the epigraph (see Fig. 2) for the value of $f^*(0)$:

$$\upsilon = \min_{(0,\varepsilon)\in D} \varepsilon \ge \min_{(0,\varepsilon)\in \operatorname{epi} f^*} \varepsilon = f^*(0) \ge \min_{(0,\varepsilon)\in U} \varepsilon.$$

This clipping more precisely localizes a priori potential points of the epigraph of f^* that are added later on in refining its approximation.

In the computational aspect, the value of υ can be obtained as the solution of a linear programming problem and is an estimate found by the Kelly cutting plane method [36],

$$\upsilon = \min_{\substack{(0,\tau)\in \operatorname{co}((g_k, f^*(g_k)),\\ k=1,2,\ldots,m)+0\times\Re_+}} \tau = \min_{\substack{\tau=\sum_{k=1}^m \lambda_k f^*(g_k),\\ 0=\sum_{k=1}^m \lambda_k g_k,\\ \lambda\in\Delta_m}} \tau = \min_{\substack{0=\sum_{k=1}^m \lambda_k g_k,\\ \lambda\in\Delta_m}} \sum_{k=1}^m \lambda_k f^*(g_k)$$

$$= \max_x \min_{\lambda\in\Delta_m} \sum_{k=1}^m \lambda_k (f^*(g_k) - x g_k) = \max_x \min_k \{f^*(g_k) - x g_k\}$$

$$= \max_x \min_k \{x_k g_k - f(x_k) - x g_k\} = -\min_x \max_k \{f(x_k) + (x - x_k) g_k\}, \tag{2}$$

where $\lambda = (\lambda_1, \ldots, \lambda_m)$ and Δ_m is the standard simplex, $\Delta_m = \left\{ \lambda_k \ge 0, k = 1, \ldots, m; \sum_{k=1}^m \lambda_k = 1 \right\}$.

The addition of the next point to the approximation of epi f^* in this clipping lies in the solution of the problem of construction of the supporting hyperplane to the truncated epigraph of f^*,

$$\sup_{\substack{(g,\varepsilon)\in \operatorname{epi} f^*\\ \varepsilon\le\upsilon}} \{gx - \varepsilon\}. \tag{3}$$

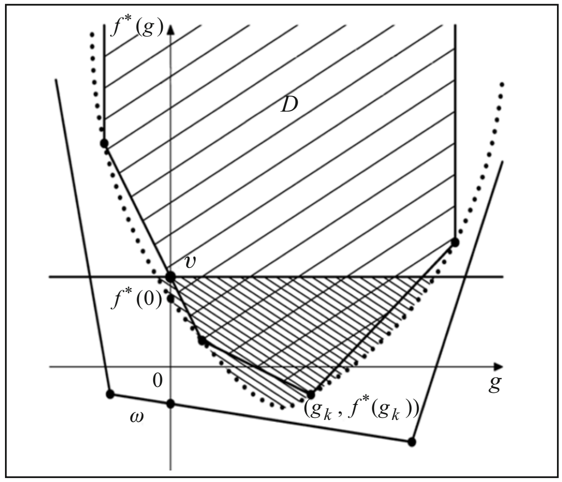

Fig. 3. Illustration of the SPACLIP–SPM algorithm with additional clippings.

This problem differs from a similar problem of refining an approximation of the epigraph of a conjugate function in the standard SPM by the presence of the additional constraint $\varepsilon \le v$. In this respect, the algorithm is similar to the level method [17] but differs from it in that approximations are constructed in the extended space $\Re^{n+1}$. Fig. 3 illustrates the idea of this additional clipping.

As well as in the standard SPM, problem (3) can be easily transferred to the space of primal variables $x \in \Re^n$, but, in this case, an auxiliary one-dimensional minimization problem arises with a rather unexpected objective function. In fact, introducing a dual variable λ for an additional constraint, we obtain

$$\sup_{(g,\varepsilon)\in \operatorname{epi} f^*;\ \varepsilon\le v} \{gx-\varepsilon\} = \sup_g \inf_{\lambda\ge 0} \{gx - f^*(g) + \lambda(v - f^*(g))\}$$

$$= \inf_{\lambda\ge 0} \{\lambda v + \sup_g \{gx - (\lambda+1) f^*(g)\}\} = \inf_{\lambda\ge 0} \left\{ \lambda v + (1+\lambda) f\left(\frac{x}{1+\lambda}\right)\right\}$$

$$= -v + \inf_{\lambda\ge 1} \{\lambda(v + f(\lambda^{-1}x))\} = -v + \inf_{\lambda \ge 1} \varphi(\lambda, x). \tag{4}$$

It is easy to show that the function $\varphi(\lambda,x) = \lambda(f(\lambda^{-1}x)+v)$ is convex with respect to the collection of variables (λ, x). An upper estimate for $\varphi(\alpha\lambda_1 + (1-\alpha)\lambda_2, \alpha x + (1-\alpha)y)$, where $0 \le \alpha \le 1$, can be obtained using the following Jensen inequality for f:

$$\varphi(\alpha\lambda_1 + (1-\alpha)\lambda_2, \alpha x + (1-\alpha)y) = (\alpha\lambda_1 + (1-\alpha)\lambda_2)\left(f\left(\frac{\alpha x + (1-\alpha)y}{\alpha\lambda_1 + (1-\alpha)\lambda_2}\right) + v\right)$$

$$= (\alpha\lambda_1 + (1-\alpha)\lambda_2)\left(f\left(\frac{\alpha\lambda_1}{\alpha\lambda_1 + (1-\alpha)\lambda_2}\cdot\frac{x}{\lambda_1} + \frac{(1-\alpha)\lambda_2}{\alpha\lambda_1 + (1-\alpha)\lambda_2}\cdot\frac{y}{\lambda_2}\right) + v\right)$$

$$\le \alpha\lambda_1\left(f\left(\frac{x}{\lambda_1}\right) + v\right) + (1-\alpha)\lambda_2\left(f\left(\frac{x}{\lambda_2}\right) + v\right) = \alpha\varphi(\lambda_1, x) + (1-\alpha)\varphi(\lambda_2, y).$$

Thus, the function $\varphi(\lambda, x)$ is convex with respect to the set of variables (λ, x) and, hence, also with respect to the variable λ. ■

It is proposed to solve one-dimensional NDO problem (4) with the help of the fast one-dimensional search algorithm [27, 28] since it can achieve superlinear or even quadratic convergence rates under favorable conditions. A special implementation of the fast one-dimensional search algorithm was created for the SPACLIP algorithm.

As a result, the algorithm of the separating plane method with clippings (SPACLIP) is of the following form.

Step 0. Initialization. Set the counter of iterations $k = 0$ and determine the initial point x_0 of the minimizing sequence.

Step 1. Find $\inf_{0 \in U_k(\omega)} \omega = \omega_k$, where U_k is the kth external approximation of the epigraph of the conjugate f^*. This problem can be solved recursively as follows:

$$\omega_k = \inf_{0 \in U_k(\omega)} \omega = \inf_{0 \in U_{k-1}(\omega) \cap (g, \omega) \mid gx_{k-1} - \omega \le f(x_{k-1})} \omega$$

$$= \max\left\{ \inf_{0 \in U_{k-1}(\omega)} \omega, \inf_{\omega \ge -f(x_{k-1})} \omega \right\} = \max\{\omega_{k-1}, -f(x_{k-1})\}, \ k \ge 1. \tag{5}$$

In this case, we can consider that $\omega_0 = -\infty$. In fact, $-\omega_k$ is the record of the function f.

Step 2. Find the vector $\bar{z}^k = (z^k, \xi_k)$, i.e., the projection of the point $(0, \omega_k)$ onto the polyhedron D_k of internal approximation of the epigraph of the conjugate,

$$\bar{z}^k = P_{D_k}((0, \omega_k)),$$

where $P_X(a)$ is the solution of the problem of projection of a point a onto a set X.

To solve this problem, the suitable affine subspace method from [26] is used. This finite method solves the problem of finding the vector of minimum length in a polyhedron of a finite-dimensional Euclidean space and possesses the global convergence rate that is "better than linear."

Step 3. Compute the next approximation to the solution of the problem $\min_{x \in \Re^n} f(x)$ as follows:

$$x_k = -z^k / \xi_k.$$

As a result of this normalization, the last coordinate in $\bar{x}_k = -\bar{z}^k / \xi_k = (x_k, -1)$ will be equal to -1, which is the required result (see Fig. 1).

Step 4. Find υ, i.e., the level of clipping the upper part of the epigraph epi f^*; to this end, linear programming problem (2) should be solved. Note that this clipping does not prevent the solution of the problem of projection onto D_k at step 2 and, hence, the construction of an approximate solution x_k.

If problem (2) has no solution, go to step 7.

Step 5. Solve one-dimensional NDO problem (4). Let λ_k be a solution found at the kth iteration of solving this problem.

Step 6. Modify the approximation x_k by the formula $x_k = \lambda_k^{-1} x_k$.

In contrast to the SPM, when the next point is added to an approximation epi f^* during the operation of the SPACLIP algorithm, the subgradient of the function being optimized is computed not at the point x being tested but at the point $\lambda_k^{-1} x$ scaled with respect to x.

Step 7. Add the pair $(g_k \in \partial f(x_k), f^*(g_k))$ to the polyhedron D_k.

Step 8. If any of the completion conditions of the algorithm is fulfilled, then complete its operation. Otherwise, increment the counter of iterations k by one and go to step 1.

The next sections contain the proof of convergence of this algorithm and computational results.

CONVERGENCE OF THE METHOD

The convergence of the projective separating plane algorithm with clippings is substantiated by the following theorem.

THEOREM 1. Let $f(x)$ be a finite convex function, let $f(0) = 0$, and let $\omega_* = -\min f(x) < \Omega < \infty$. Then $\lim_{k \to \infty} \omega_k = \omega_*$.

Proof. We can prove by induction that $\omega_k \le f^*(0)$ for any k.

In fact, this inequality is satisfied by $k=0$. According to recursion (5), we have $\omega_k = \max\{\omega_{k-1}, -f(x_{k-1})\}$. Making the inductive step and taking into account that $-f(x_{k-1}) = 0\cdot x_{k-1} - f(x_{k-1}) \le \sup_x \{0\cdot x - f(x)\} = f^*(0)$, we obtain $\omega_k \le \max\{f^*(0), f^*(0)\} = f^*(0)$, which is what had to be proved.

Since $\omega_k \le f^*(0)$ and $\bar{z}_k + (0, \omega_k) \in \mathrm{co}\{D_k\} = D_k$, to prove the statement of the theorem, it suffices to show that $\|\bar{z}_k\| \underset{k\to\infty}{\to} 0$ since this means that $-f(x_k) \underset{k\to\infty}{\to} f^*(0)$.

To prove the monotone decrease in the norm of vectors $\|\bar{z}_k\|$, we consider several cases.

1. At step 4, problem (2) has no solution for all k. Then we can consider that $\upsilon \equiv \infty$ and the SPM with additional clippings is transformed into the standard SPM whose convergence is proved in [25].

2. Problem (2) has a solution at step 4. Denote by $\bar{x}_k$ the solution to problem (3), $\bar{x}_k = \lambda_k^{-1} x_k$, $\lambda_k \ge 1$. Then the vector $\bar{z}_k$ is representable in the form $\bar{z}_k = -r_k(\lambda_k^{-1} x_k, -1)$.

Depending on whether the current record of the objective function ω_k changes at the kth iterations, two versions are possible.

Version 1. $\omega_k = \omega_{k-1}$. Then the projection is carried out from the same point $\|\bar{z}_k\|^2 = \min_{z+(0,\omega_k)\in \mathrm{co}D_k'} \|z\|^2 = \min_{z+(0,\omega_{k-1})\in \mathrm{co}D_{k-1}'} \|z\|^2$. Here, D_k' is a polytope obtained at the kth iterations after clipping (3). In this case, the following inequality holds:

$$\|\bar{z}_k\|^2 \le \min_{\lambda\in[0,1]} \|\bar{z}_{k-1} + \lambda(g_k, f^*(g_k)) - \bar{z}_{k-1}\|^2.$$

The solution of the problem $\min_{\lambda\in[0,1]} \|\bar{z}_{k-1} + \lambda(g_k, f^*(g_k)) - \bar{z}_{k-1}\|^2$ is the projection of the minimum of a one-dimensional quadratic function onto the interval $[0,1]$,

$$\lambda^* = \min\{(\bar{z}_{k-1} - (g_k, f^*(g_k)))\bar{z}_{k-1} / \|(g_k, f^*(g_k)) - \bar{z}_{k-1}\|^2, 1\}.$$

Then, for any

$$\lambda \le \frac{(\bar{z}_{k-1} - (g_k, f^*(g_k)))\bar{z}_{k-1}}{\|(g_k, f^*(g_k)) - \bar{z}_{k-1}\|^2}, \tag{6}$$

the inequality $\|\bar{z}_k\|^2 \le \|\bar{z}_{k-1} + \lambda((g_k, f^*(g_k)) - \bar{z}_{k-1})\|^2$ holds.

After exponentiation, we obtain

$$\|\bar{z}_k\|^2 \le \|\bar{z}_{k-1}\|^2 - 2\lambda((\bar{z}_{k-1} - (g_k, f^*(g_k)))\ \bar{z}_{k-1} - \frac{\lambda}{2}\|(g_k, f^*(g_k)) - \bar{z}_{k-1}\|^2). \tag{7}$$

According to inequality (6),

$$(\bar{z}_{k-1} - (g_k, f^*(g_k)))\ \bar{z}_{k-1} - \frac{\lambda}{2}\|(g_k, f^*(g_k)) - \bar{z}_{k-1}\|^2$$

$$\ge \frac{\lambda}{2}\|(g_k, f^*(g_k)) - \bar{z}_{k-1}\|^2 > 0 \text{ when } \lambda \ne 0.$$

To prove the monotone decrease in the norm of vectors $\|\bar{z}_k\|$ for version 1, it suffices to take into account in inequality (7) that the subtrahend in the right side is positive with respect to the last inequality, $\|\bar{z}_k\|^2 < \|\bar{z}_{k-1}\|^2$.

Version 2. $\omega_k = -f(x_{k-1}) > \omega_{k-1}$. Then

$$\|\bar{z}_k\|^2 = \min_{z+(0,\omega_k)\in \mathrm{co}D_k'} \|z\|^2 \le \|\bar{z}_\lambda\|^2, \tag{8}$$

where $\bar{z}_\lambda = \dfrac{\Omega - \omega_k}{\Omega - \omega_{k-1}} \bar{z}_{k-1} < \bar{z}_{k-1}$.

As is easily seen, the last inequality in formula (8) is fulfilled when $\bar{z}_\lambda$ is defined in this manner. Hence, $\|\bar{z}_k\|^2 < \|\bar{z}_{k-1}\|^2$.

The monotonicity of the norm of vectors $\|\bar{z}_k\|$ implies the existence of the limit $\lim_{k\to\infty}\|\bar{z}_k\| = \rho$.

To prove that $\rho = 0$, suppose the contrary. Let $\|\bar{z}_k\| \ge \tau r_k$ for some $\tau > 0$.

The following equality holds:

$$\lim_{k\to\infty} \lfloor ((g_{k+1}, f^*(g_{k+1})) - (0, \omega_k))\,\bar{z}_k - \|\bar{z}_k\|^2 \rfloor = 0.$$

Suppose the contrary. Let, for some subsequence, we have $((g_{k'+1}, f^*(g_{k'+1})) - (0,\omega_{k'}))\,\bar{z}_{k'} \le \|\bar{z}_{k'}\|^2 - \gamma,\ \gamma > 0$. Then

$$\|\bar{z}_{k'+1}\|^2 = \min_{z+(0,\,\omega_{k'}) \in \operatorname{co}\{\{(g_i,\, f^*(g_i)),\ i=0,1,\ldots,k'+1\},\,(0,\Omega)\}} \|z\|^2$$

$$\le \min_{z+(0,\,\omega_{k'}) = \lambda\tilde{g}_{k'} + (1-\lambda)(g_{k'+1},\, f^*(g_{k'+1}))} \|z\|^2,\ \lambda \in [0,1].$$

Taking into account that $\tilde{g}_{k'} = (0, \omega_{k'}) + \bar{z}_{k'}$, we obtain

$$\|\bar{z}_{k'+1}\|^2 \le \min_{\lambda\in[0,1]} \|\lambda\,\bar{z}_{k'} + (1-\lambda)((g_{k'+1}, f^*(g_{k'+1})) - (0,\omega_{k'}))\|^2. \tag{9}$$

After squaring the right side of inequality (9), we obtain

$$\|\bar{z}_{k'+1}\|^2 \le \min_{\lambda\in[0,1]} \{\lambda^2 \|\bar{z}_{k'}\|^2 + 2\lambda(1-\lambda)\,\bar{z}_{k'}((g_{k'+1}, f^*(g_{k'+1})) - (0,\omega_{k'}))$$

$$+ (1-\lambda)^2\ \|(g_{k'+1}, f^*(g_{k'+1})) - (0,\omega_{k'})\|^2\}. \tag{10}$$

We continue transformations in the right side of inequality (10) as follows:

$$\|\bar{z}_{k'+1}\|^2 \le \min_{\lambda\in[0,1]} \{\lambda^2 \|\bar{z}_{k'}\|^2 + 2\lambda(1-\lambda)\,\|\bar{z}_{k'}\|^2 - 2\gamma\,\lambda(1-\lambda)$$

$$+ (1-\lambda)^2\ \|(g_{k'+1},\ f^*(g_{k'+1})) - (0,\omega_{k'})\|^2\}$$

$$= \min_{\lambda\in[0,1]} \{(2\lambda - \lambda^2)\,\|\bar{z}_{k'}\|^2 - 2\gamma\,\lambda(1-\lambda) + (1-\lambda)^2\,\|(g_{k'+1},\ f^*(g_{k'+1})) - (0,\omega_{k'})\|^2\}. \tag{11}$$

From inequality (11), the following estimate is obtained:

$$\|\bar{z}_{k'+1}\|^2 \le \min_{\lambda\in[0,1]} \{\|\bar{z}_{k'}\|^2 - 2\gamma\,\lambda(1-\lambda) + ((1-\lambda)^2\,\|(g_{k'+1},\ f^*(g_{k'+1})) - (0,\omega_{k'})\|^2 - \ \|\bar{z}_{k'}\|^2\ )\}$$

$$\le \|\bar{z}_{k'}\|^2 - 2\gamma\,\lambda(1-\lambda) + (1-\lambda)^2\,\delta^2 \ \text{ for any } \lambda\in[0,1].$$

Substituting the expression $\lambda = (\delta^2 + \gamma)/(\delta^2 + 2\gamma) > 0$ in this inequality, we obtain

$$\|\bar{z}_{k'+1}\|^2 \le \|\bar{z}_{k'}\|^2 - \gamma^2/(\delta^2 + 2\gamma). \tag{12}$$

Proceeding to the limit in inequality (12) as $k' \to \infty$, we obtain a contradiction. Hence,

$$\lim_{k\to\infty} \lfloor ((g_{k+1}, f^*(g_{k+1})) - (0, \omega_k))\,\bar{z}_k - \|\bar{z}_k\|^2 \rfloor = 0.$$

Since $0 \le (\bar{z}_k + (0,\omega_k))\,\bar{z}_k - (g_{k+1}, f^*(g_{k+1}))\bar{z}_k \to 0$ as $k\to\infty$, for a sufficiently large k, the inequality $(\bar{z}_k + (0,\omega_k))\,\bar{z}_k - (g_{k+1}, f^*(g_{k+1}))\bar{z}_k \le r_k^2\varepsilon^2$ holds for any $\varepsilon > 0$. Hence,

$$(0,\omega_{k+1})\bar{z}_k \geq (g_{k+1}, f^*(g_{k+1}))\bar{z}_k \geq (0,\omega_k)\bar{z}_k + \|\bar{z}_k\|^2 - r_k\varepsilon$$

$$\geq (0,\omega_k)\bar{z}_k + r_k^2\tau^2 - r_k^2\varepsilon^2 \geq (0,\omega_k)\bar{z}_k + r_k^2\tau^2 / 2$$

for $\varepsilon \leq \tau/\sqrt{2}$, i.e., $r_k(0,\omega_{k+1}) \geq r_k(0,\omega_k) + r_k^2\tau^2/2$.

Next, $f^*(0) \geq (0, \omega_{k+1}) \geq (0,\omega_k) + r_k\tau^2 \geq (0,\omega_k) + \delta$, where $\delta \geq r_k\tau^2 \geq 0$, which is impossible as $k \to \infty$. This contradiction proves the equality $\lim_{k\to\infty} \|\bar{z}_k\| = 0$. The theorem is proved. ■

USING THE ALGORITHM SPACLIP TO SOLVE PROBLEMS OF TRANSPORTATION TYPE

Consider the transportation problem with constraints on volumes of delivery in its standard statement [30, 33]. Assume that m_1 is the number of consumers, n_1 is the number of suppliers, A_i $(i = 1, \ldots, n_1)$ are supplies of products delivered by suppliers, B_j $(j = 1, \ldots, m_1)$ are consumer needs for a product, c_{ij} is the price of delivery of a product unit from the *i*th supplier to the *j*th consumer, and x_{ij}^{low} and x_{ij}^{up} are lower and upper bounds on volumes of deliveries x_{ij}. We consider that the problem is balanced, $\sum_{i=1}^{n_1} A_i = \sum_{j=1}^{m_1} B_j$.

The minimization of transportation expenditures with establishing balances among suppliers and consumers and imposing two-sided bounds on variables leads to the classical problem of linear programming

$$\min_{x_{ij} \in X_{ij}} \sum_{i=1}^{n_1} \sum_{j=1}^{m_1} c_{ij} x_{ij}, \tag{13}$$

where a set X_{ij} imposes the constraints

$$\sum_{j=1}^{m_1} x_{ij} = A_i,\ i = 1, \ldots, n_1; \quad \sum_{i=1}^{n_1} x_{ij} = B_j,\ j = 1, \ldots, m_1; \tag{14}$$

$$x_{ij}^{low} \leq x_{ij} \leq x_{ij}^{up},\ \ i = 1, \ldots, n_1;\ \ j = 1, \ldots, m_1, \tag{15}$$

or, in matrix-vector form, $\min_{x\in X} cx$, $X = \{x \mid Ax = b,\ x^{low} \leq x \leq x^{up},\ b = (A_i, B_j)^{\mathrm{T}}\}$.

Instead of problem (13)–(15), we will consider the following equivalent quadratic problem under the same constraints:

$$\min_{x_{ij} \in X_{ij}} \sum_{i=1}^{n_1} \sum_{j=1}^{m_1} \varepsilon \left(x_{ij} + \frac{c_{ij}}{2\varepsilon} \right)^2 = \min_{x \in X} \left\| x + \frac{c}{\varepsilon} \right\|^2 = \left\| P_X \left(-\frac{c}{\varepsilon} \right) + \frac{c}{\varepsilon} \right\|^2, \tag{16}$$

where $P_X(a)$ is the result of solving the problem of projection of a point a onto the set X.

Developing the results of [30], it may be proved that there really is some $\bar{\varepsilon} > 0$ such that, for all $0 < \varepsilon < \bar{\varepsilon}$, linear programming problem (13)–(15) and projection problem (16) are equivalent, and $P_X(-c\varepsilon^{-1})$ solves transportation problem (13)–(15).

To formulate this result, we will introduce some additional definitions. Let $\hat{x} \in X$ be some fixed point. Denote by $K_{\hat{x}} = \{z \mid \hat{x} + \lambda z \in X, \lambda > 0\}$ the cone of admissible directions at the point $\hat{x}$. If the set X is polyhedral, then the cone $K_{\hat{x}}$ is closed. Denote by $K_{\hat{x}}^*$ the cone dual to the cone $K_{\hat{x}}$, $K_{\hat{x}}^* = \{u \mid uz \leq 0, z \in K_{\hat{x}}\}$.

Note that if $y = \hat{x} + u, u \in K_{\hat{x}}^*$, then, for any $x \in X$, the inequality $(y - \hat{x})(x - \hat{x}) = u(x - \hat{x}) \leq 0$ holds since $x - \hat{x} \in K_{\hat{x}}$. Therefore, $\hat{x} = P_X(y)$, i.e., $\hat{x}$ is the projection of y onto X.

In terms of the introduced notations, this result can be formulated in the following form.

THEOREM 2. Let transportation problem (13)–(15) have a unique solution $\hat{x}$. Then there is some $\bar{\varepsilon} > 0$ such that

$$\left\| P_X\left(-\frac{c}{\varepsilon}\right)+\frac{c}{\varepsilon}\right\|^2 = \min_{x\in X}\left\| x+\frac{c}{\varepsilon}\right\|^2 = \left\| \hat{x}+\frac{c}{\varepsilon}\right\|^2 \text{ for all } 0<\varepsilon\le\bar{\varepsilon}.$$

Proof. For $\hat{x}$, we determine the corresponding cones $K_{\hat{x}}$ and $K_{\hat{x}}^*$. By virtue of the uniqueness of the solution, we have $-c(x-\hat{x})<0$ for any $x\in X$. Hence, $-c\in \operatorname{int} K_{\hat{x}}^*$, which also proves the nonemptiness of the interior $\operatorname{int} K_{\hat{x}}^*$. Then $-\frac{c}{\varepsilon}=\hat{x}-\frac{c}{\varepsilon}-\hat{x}=\hat{x}+\frac{1}{\varepsilon}(-\varepsilon\hat{x}-c)=\hat{x}+\frac{1}{\varepsilon}z_\varepsilon$, where $z_\varepsilon=-c-\varepsilon\hat{x}\in\operatorname{int} K_{\hat{x}}^*$ for a sufficiently small ε. Accordingly, $\frac{1}{\varepsilon}z_\varepsilon\in K_{\hat{x}}^*$ and $\hat{x}=P_X\left(\hat{x}+\frac{1}{\varepsilon}z_\varepsilon\right)=P_X\left(-\frac{c}{\varepsilon}\right)$, which is what had to be proved. ■

The above theorem allows one, under uniqueness conditions imposed on the solution of a linear programming problem, to reduce it to a special quadratic programming problem with the simplest quadratic form, which considerably simplifies its solution.

COMPUTATIONAL EXPERIMENTS

Transportation problem with three suppliers. The projection algorithm SPACLIP was implemented on the freely distributable system Octave of matrix-vector computations [35]. The syntax of Octave is very close to MATLAB, and this system is a convenient tool for developing pilot versions of computational algorithms and fast prototyping.

After verifying the operability of the implemented algorithm as applied to a number of NDO problems, this method was applied to the solution of transportation logistics optimization problems with the help of the expedients described in the previous section.

The transportation problem with three suppliers and four consumers (respectively, $n_1=3$, $m_1=4$, and the vector x for this problem had the dimension 3∗4 = 12), which is presented in [30], was first solved. The specified tariff matrix $\{c_{ij}\}_{i,\,j=1}^{3,\,4}$ was transformed into the vector $c=(7,8,1,2,4,5,9,8,9,2,3,6)$. Under the condition of the problem, the supplies of products amounted to $A_i=[200,\ 180,\ 190]$ and needs for products amounted to $B_j=[150,\ 130,\ 150, 140]$. The lower bound x_{ij}^{low} was set to zero, and the upper bound $x_{ij}^{up}=200$ was set to $i=1,\dots,n_1$, $j=1,\dots,m_1$.

The results of solution of this transportation problem are given in Table 1. In solving the problem, both methods completed their operation under the condition of closeness to the value of the argument. The condition of closeness of the gradient norm of the optimal solution to zero was not fulfilled. The value of ε from problem (16) coincides with the accuracy prescribed for the chosen SPM and SPACLIP nonsmooth optimization methods and is presented in the last column of the table. A decline in the results for $\varepsilon=10^{-8}$ is connected with the constraints of the implementation of the Octave language on digit capacity. On the whole, both methods coped with the solution of this model problem and correctly computed the optimal distribution of product volumes. The comparison of these results with the results of solution of the problem from [30] showed that the methods of the family of separating planes are more sensitive to the closeness of the upper bound x_{ij}^{up} to the solution than Shor's r-algorithm [32]. The upper bound was taken equal to 200 since, for smaller values, the algorithms of the family of separating plane algorithms yield a vector combination of given numbers-constraints as the optimal solution. Such problems did not arise in the case of the r-algorithm. Note that the table contain components of the optimal solution x whose values are rounded to the nearest integer and, hence, the presented expenditures for the transportation of products can insignificantly differ from the product $c\cdot x$.

Transportation problems of small and large dimensions. After the solution of the model problem, mass testing of the projection algorithm was performed using a series of small- and high-dimensional transportation problems with randomly generated data.

TABLE 1. Results of Solution of a Transportation Problem with Three Suppliers with the Help of the SPM and SPACLIP

Method	Components of the Vector x												Number of Iterations	Transportation Expenditures (13)	ε
	x_{11}	x_{12}	x_{13}	x_{14}	x_{21}	x_{22}	x_{23}	x_{24}	x_{31}	x_{32}	x_{33}	x_{34}			
SPM	0	0	61	136	150	27	0	0	0	97	90	0	75	1532	10^{-4}
SPACLIP	0	0	56	141	146	28	0	0	0	101	87	0	63	1527	10^{-4}
SPM	0	0	57	136	145	28	0	0	0	104	83	0	60	1506	10^{-5}
SPACLIP	0	0	55	145	146	24	0	0	0	99	90	0	54	1514	10^{-5}
SPM	0	0	54	148	148	24	0	0	0	101	83	0	46	1514	10^{-6}
SPACLIP	0	0	96	167	123	61	0	0	0	166	154	0	53	2024	10^{-6}
SPM	0	0	65	154	166	19	0	0	0	106	95	0	41	1628	10^{-7}
SPACLIP	0	0	30	59	81	54	0	0	0	108	200	0	100	1557	10^{-7}
SPM	0	0	61	152	138	62	0	0	0	103	83	0	35	1680	10^{-8}
SPACLIP	0	0	200	100	100	0	0	0	0	0	200	0	38	1401	10^{-8}

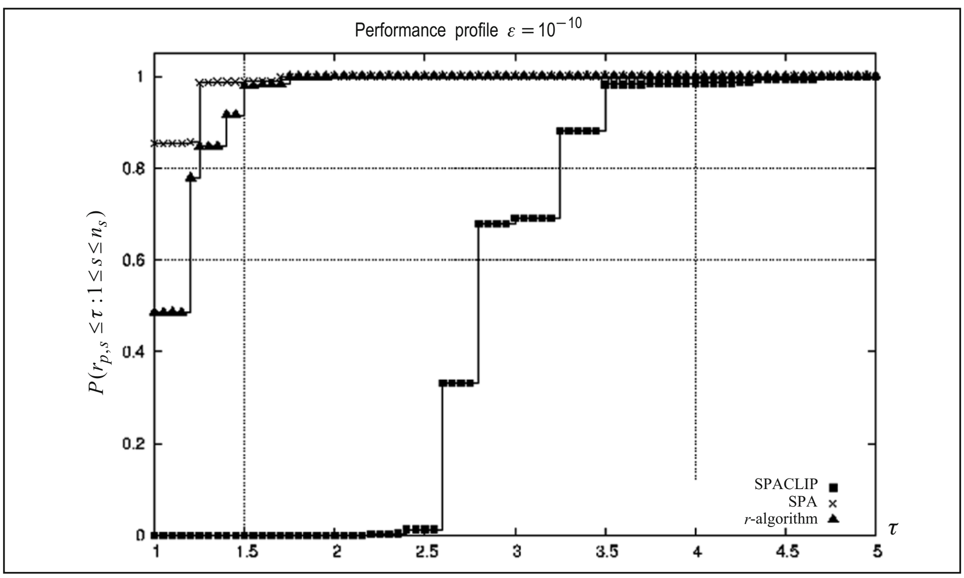

Fig. 4. Performance profiles for the standard SPM (SPA), SPACLIP algorithm, and r-algorithm for a transportation problem with 100 variables.

For each of these problems, $n_1 = m_1 = n$; the tariff vector c was randomly generated as a vector of length n^2 and consisted of random numbers belonging to the interval [1, 101]. Next, a vector x_{opt} of dimension n^2, i.e., a feasible solution of the problem (with components from the interval [1, 1001]) was (also randomly) generated. Then the vector b equal to $[A_i, B_j]^{\text{T}}$ was computed by the formula $b = A \cdot x_{\text{opt}}$. The lower and upper bounds on volumes of delivery were specified by multiplying the feasible solution by 0.1 and 20, respectively. The value of ε from problem (16) was equated to the required accuracy prescribed by the user for the solution of the problem. A random vector of length n^2 with components from the interval [0, 10] was taken as the initial approximation. Then tests were performed using thousands of problems of the same type and the obtained results were processed according to the methodology from [34] with constructing performance profiles.

The performance profile [34] ρ_s for a method of solving an optimization problem is understood to be the distribution function of some measurable performance index. The computation of performance profiles makes it possible to visualize distinctions in the efficiency of several optimization methods.

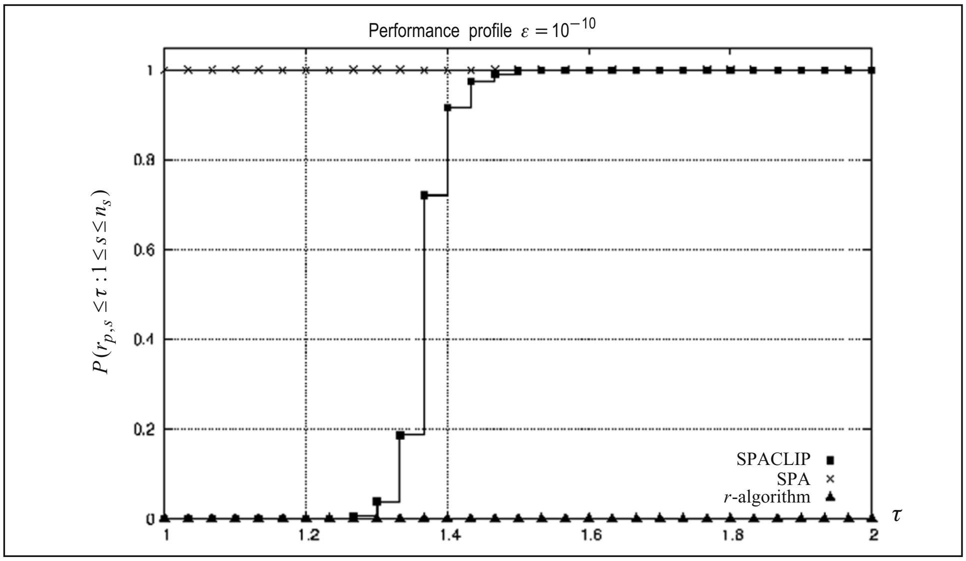

Fig. 5. Performance profiles for the standard SPM (SPA), SPACLIP algorithm, and *r*-algorithm for a transportation problem with 10000 variables.

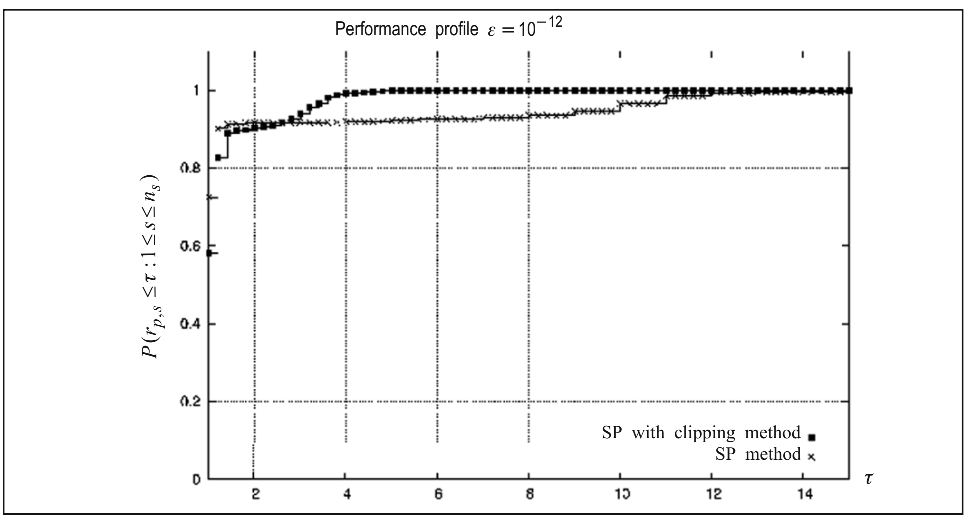

Fig. 6. Performance profiles for the standard SPM and SPACLIP algorithm for a transportation problem with 10000 variables.

Such a distribution function is found as follows:

$$\rho_s(\tau) = \frac{1}{n_p} \,|\, p \in P : r_{p,s} \le \tau \,|; \;\; r_{p,s} = \frac{t_{p,s}}{\min \{t_{p,s} : s \in S\}}.$$

Here, S is the set of the methods being compared and P is the set of problems solved with the help of these methods. The number of elements in P is denoted by n_p, and n_s is the number of elements in S. In this case, $n_s = 3$ (three methods are compared including the *r*-algorithm) or $n_s = 2$ (standard SPM and SPACLIP are compared) and $n_p = 5000$. As the measurable performance index $t_{p,s}$, the processor time spent for the solution of a problem was estimated.

Testing was performed on a computer under the control of Linux OS, distribution kit openSUSE 12.3 Dartmouth, AMD Athlon 64 3500+, 2 Gb; interpreter version Octave 3.6.4. The test results are given in Figs. 4–7. Problems were solved with 100 variables ($n = 10$), 10000 variables ($n = 100$), and 40000 variables ($n = 200$). For each dimension, 5000 transportation problems were solved.

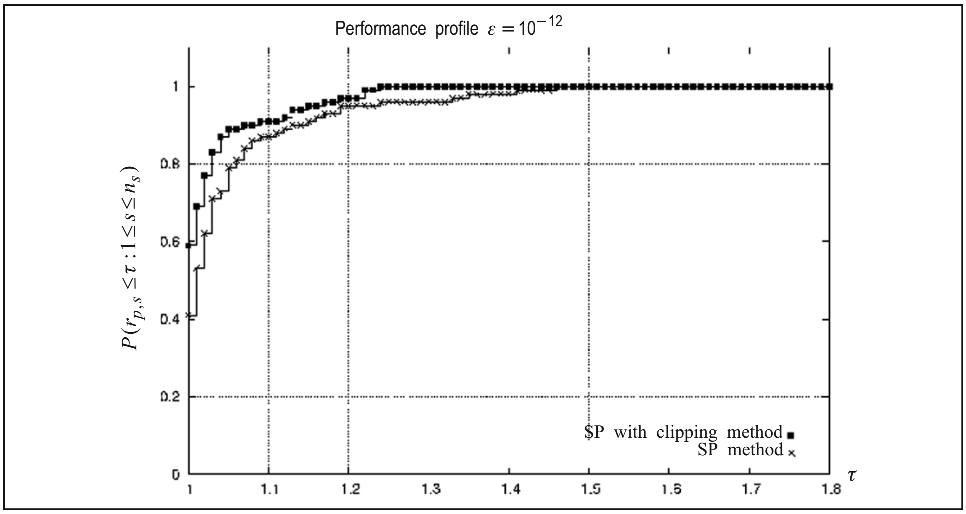

Fig. 7. Performance profiles for the standard SPM and SPM with additional clippings for a transportation problem with 40000 variables.

The analysis of the obtained performance profiles shows that, in solving transportation problems of small dimensions, the r-algorithm operates faster than the projection algorithm SPACLIP. The standard SPM for problems of such dimensions also outruns the projection algorithm SPACLIP. It is connected with the lack of need for solving linear programming problem (2). But the larger the dimension of a problem and required solution accuracy, the greater the advantage of SPACLIP over other methods.

CONCLUSIONS

In this article, the algorithm SPACLIP for solving convex NDO problems is proposed that belongs to the family of separating plane methods with additional clippings generated as a result of solution of an auxiliary problem of the cutting plane method and its convergence is proved. A practical application of the algorithm is considered to be the solution of high-dimensional transportation problems. Under conditions of a unique solution of a transportation problem, the equivalence of problems of the projection of a sufficiently remote point onto an admissible set and problems of transportation type is proved. Computational experiments demonstrate a rather high performance of SPACLIP in solving projection problems.

REFERENCES

1. R. V. Namm and G. Woo, Introduction to the Theory and Solution Method for Variational Inequalities, Changwon National University Press, Changwon (2002).
2. F. N. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski, Nonsmooth Analysis and Control Theory, Springer, New York (1998).
3. I. V. Sergienko, M. V. Mikhalevich, P. I. Stetsyuk, and L. B. Koshlai, "Models and information technologies for decision support during structural and technological changes," Cybernetics and Systems Analysis, **45**, No. 2, 204–213 (2009).
4. I. V. Sergienko, Methods of Optimization and Systems Analysis for Problems of Transcomputational Complexity [in Ukrainian], Akademperiodika, Kyiv (2010).
5. N. Z. Shor and S. I. Stetsenko, Quadratic Extremal Problems and Nondifferentiable Optimization [in Russian], Naukova Dumka, Kyiv (1989).
6. J. Nocedal and S. Wright, Numerical Optimization, Springer Series in Operations Research and Financial Engineering, XXII, 2nd Ed., Springer, New York (2006).
7. B. T. Polyak, Introduction to Optimization [in Russian], 2nd Ed., LENAND, Moscow (2014).

8. A. S. Nemirovskii and D. B. Yudin, Complexity of Problems and Efficiency of Optimization Methods [in Russian], Nauka, Moscow (1979).
9. Yu. E. Nesterov, Introduction to Convex Optimization [in Russian], MTsNMO, Moscow (2010).
10. A. Yu. Levin, "On an algorithm for the minimization of convex functions," DAN SSSR, **160**, No. 6, 1244–1247 (1965).
11. D. J. Newman, "Location of maximum on unimodal surfaces," Journal of ACM, **12**, 395–398 (1965).
12. B. T. Polyak, "Minimization of nonsmooth functionals," Computational Mathematics and Mathematical Physics, **9**, No. 3, 509–521 (1969).
13. E. A. Nurminski, "Envelope stepsize control for iterative algorithms based on Fejer processes with attractants," Optimization Methods and Software, **25**, No. 1, 97–108 (2010).
14. K. C. Kiwiel, "An aggregate subgradient method for nonsmooth convex minimization," Mathematical Programming, No. 27, 320–341 (1983).
15. C. Lemarechal, "An extension of Davidon methods to nondifferentiable problems," Mathematical Programming Study, No. 3, 95–109 (1975).
16. P. Wolfe, "A method of conjugate subgradients for minimizing nondifferentiable functions," Mathematical Programming Study, No. 3, 145–173 (1975).
17. C. Lemarechal, A. Nemirovskii, and Yu. Nesterov, "New variants of bundle methods," Mathematical Programming, No. 69, 111–148 (1995).
18. R. Mifflin and C. Sagastizabal, "A VU-algorithm for convex minimization," Mathematical Programming, No. 104, 583–608 (2005).
19. Yu. Nesterov, Introductory Lectures on Convex Optimization: A Basic Course, Kluwer Academic Publishers, Boston (2004).
20. Yu. Nesterov, "Smooth minimization of non-smooth functions," Mathematical Programming, No. 103, 127–152 (2005).
21. E. A. Nurminski, "Separating plane algorithms for convex optimization," Mathematical Programming, No. 76, 373–391 (1997).
22. E. A. Nurminski, "Separating plane method with bounded memory for the solution of convex nonsmooth optimization problems," Computational Methods and Programming, **7**, 133–137 (2006).
23. E. A. Vorontsova, "A projective separating plane method with additional clipping for non-smooth optimization," WSEAS Transactions on Mathematics, **13**, 115–121 (2014).
24. V. F. Demyanov and L. V. Vasilyev, Nondifferentiable Optimization [in Russian], Nauka, Moscow (1981).
25. E. A. Nurminski, Numerical Methods of Convex Optimization [in Russian], Nauka, Moscow (1991).
26. E. A. Nurminski, "Convergence of the suitable affine subspace method for finding the least distance to a simplex," Computational Mathematics and Mathematical Physics, **45**, No. 11, 1996-2004 (2005).
27. E. Nurminski, "A quadratically convergent line-search algorithm for piecewise smooth convex optimization," Optimization Methods and Software, No. 6, 59–80 (1995).
28. E. A. Vorontsova, "A quickly convergent linear search algorithm in nondifferentiable optimization," Informatics and Control Systems, No. 2, 39–48 (2012).
29. P. I. Stetsyuk and E. A. Nurminski, "Nonsmooth penalty and subgradient algorithms to solve the problem of projection onto a polytope," Cybernetics and Systems Analysis, **46**, No. 1, 51–55 (2010).
30. P. I. Stetsyuk, E. A. Nurminski, and D. I. Solomon, "Transportation problem and orthogonal projection onto linear manifolds," in: Proc. Vth Intern. Sci. Conf. "Transportation systems and logistics," Eureka, Chisinau (2013), pp. 251–263.
31. B. N. Pshenichnyi, Convex Analysis and Extremal Problems [in Russian], Nauka, Moscow (1980).
32. N. Z. Shor, Methods of Nondifferentiable Optimization and Complicated Extremal Problems [in Russian], Eureka, Chisinau (2008).
33. D. B. Yudin and E. G. Holstein, Problems and Methods of Linear Programming: Transportation Type Problems [in Russian], 3rd Ed., Librokom, Moscow (2010).
34. E. Dolan and J. More, "Benchmarking optimization software with performance profiles," Mathematical Programming, No. 91, 201–213 (2002).
35. Octave Page [Electronic Resources], http://www.gnu.org/software/octave/.
36. J. E. Kelley, "The cutting plane method for solving convex programs," Journal of the SIAM, **8** (4), 703–712 (1960).
37. N. Z. Shor, "Application of the method of gradient descent to the solution of the network transportation problem," in: Proc. Sci. Seminar on Theoretical and Applied Problems of Cybernetics and Operations Research, Scientific Board of AS of UkrSSR on Cybernetics, Issue 1, Kiev (1962), pp. 9–17.