Implementation of Non-Smooth Optimization Algorithms for Logistics Problems

Mykola Zhurbenko

V.M. Glushkov Institute of Cybernetics of the NASU, Kyiv, Ukraine
zhurbnick@gmail.com

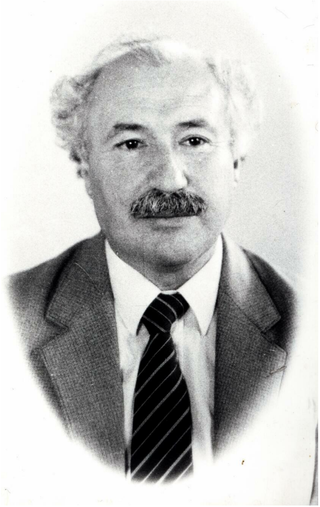
3rd International Conference on Problems of Logistics
PLMO 2024

Baku, May 15-17, 2024

Outline

- Introduction
- Decomposition in block structure problems
- Planning Multi-Product Flows
- Conclusions
- References

Naum Zuselevich Shor



Academician Naum Zuselevich Shor (1937 - 2006)

Introduction

Founded by academician N.Z. Shor department of non-smooth optimization methods is known not only for the original algorithms developed, but also for the experience in using them for a wide range of applied problems.

Subgradient optimization algorithms are used to solve a wide class of applied problems.

1962. The numerical subgradient algorithm was proposed by Shor for solving a network transport problem [1].

1969. Shor proposed using the space transformation operation [2]. Based on this idea, subgradient algorithms were developed with space dilation along the subgradients.

1976, 1977. The ellipsoid method (Yudin and Nemirovskii, Shor) [3], [4].

1971. The algorithm was developed with space dilation along the difference of two successive subgradients [5] (r-algorithm).

Introduction

Research on the development and use of new variants of non-smooth optimization algorithms continues intensively. Modifications of the r-algorithm have been developed.

$r(\sigma)$-algorithms: algorithms with program control of the values of space dilation coefficients.

r_n-algorithms: algorithms based on the operator of space dilation in the direction of the difference of normalized subgradients.

$\alpha(\varepsilon)$-algorithms: algorithms based on the use of ε-subgradients. In some sense, the $\alpha(\varepsilon)$-algorithm combines the Shor's algorithms with dilation of the space along the subgradient and the difference of two successive subgradients. An estimate of the complexity of the algorithm for solving the ε-optimization problem with guaranteed accuracy is obtained.

Introduction

The paper presents experience in using non-smooth optimization algorithms to solve problems of long-term planning of the functioning and development of transport systems.

Mathematical models are characterized by high dimensionality and block structure of constraints. The solution method is based on the use of effective non-smooth optimization algorithms in combination with decomposition methods. As a rule, solving a problem consists of two stages. At the first stage, a dual problem is solved regarding general constraints that contain variables from various blocks of constraints. At the second stage, the solution to the initial problem is determined.

Numerical aspects of this approach to solving applied logistics problems are discussed.

A brief description of recently developed subgradient algorithms is given.

Decomposition in block structure problems

We write the block convex programming problem in the following form:

$$\min\{f(x) = \sum \{f_\ell(\bar{x}_\ell) | l = \overline{1, N}\}\}, \tag{1}$$

$$\bar{x}_\ell \in D_\ell, \quad \ell = \overline{1, N} \tag{2}$$

$$g_i(x) = \sum g_i^l(\bar{x}_\ell) \le 0, \quad i = \overline{1, m}, \tag{3}$$

where $\bar{x}_\ell \in E_{n_\ell}$, $D_\ell \subset R^{n_\ell}$, $D = \bigcup D_\ell$.
Vector of problem variables (1)-(3) $x = \{\bar{x}_1, \dots, \bar{x}_N\} \in R^n$,

$$n = \sum \{n_\ell | l = 1, N\}.$$

We will consider problem (1)-(3) solvable, functions $f_\ell(\bar{x}_\ell)$, $g_\ell^i(\bar{x}_\ell)$ are convex, the sets D_ℓ are convex, closed and bounded.

Decomposition in block structure problems

The dual problem to (1)-(3) with respect to constraints (3) will be as follows:

$$\max\{\psi(u) = \sum_{\ell=1}^{N} \psi_\ell(u) \mid u \geq 0\}, \tag{4}$$

$$\psi_\ell(u) = \min_{\bar{x}_\ell \in D_\ell} L_\ell(\bar{x}_\ell, u), \quad \ell = \overline{1, N}, \tag{5}$$

$L_\ell(\bar{x}_\ell, u) = f_\ell(\bar{x}_\ell) + \sum\limits_{i=1}^{m} u_i g_i^l(\bar{x}_\ell), \quad \ell = \overline{1, N}$, $u \in R^v$, and the Lagrange function $L(x, u)$ of problem (1)-(3) is defined by the equality $L(x, u) = \sum\limits_{\ell=1}^{N} L_\ell(\bar{x}_\ell, u)$. Thus, the peculiarity of the block problem is that the problem $\psi(u) = \min\{L(x, u) | x \in D\}$ is reduced to solving N independent problems (5).

Decomposition in block structure problems

Two-stage solution scheme.
At the first stage, the dual problem (4) is solved. As a rule, the function $\psi(u)$ is nonsmooth.
Problem (4) is preliminary reduced to an unconditional maximization problem. The latter is achieved by a simple change of variables: $u_i = |\tilde{u}_i|$, $i = \overline{1, m}$.
Let $\bar{x}_\ell(u)$ be the solution to problems (5) .
If the solution to problems (5) is not unique, then any of them is used. Note that the uniqueness of the solution to problems (5) determines the smoothness of the function $\psi(u)$ at the point u.
The subgradient $g(u)$ of the function $\psi(u)$ at the point u is determined by the following formula:
$g(u) = (g_1(x(u)), g_2(x(u)), ...g_m(x(u)))^T$.

Decomposition in block structure problems

Let u^* be the solution to the dual problem (4), $\tilde{u}$ be the approximation for u^* obtained at the first stage. Then the second stage will be as follows.

We select those blocks ℓ for which $\bar{x}_\ell(\tilde{u})$ is an approximation to $\bar{x}_\ell^*$ - components of the solution to problem (1)-(3) x^*. Such blocks include ℓ blocks for which the solution to problems (5) for $u = u^*$ is unique (it can be shown that in this case the uniqueness of the solution is also true for an arbitrary point from a certain neighborhood optimal point u^*).

The next step will be to solve the problem obtained after deriving from the original problem (1)-(3) those blocks for which, using the described analysis of the solution to the dual problem, the approximations $\bar{x}_\ell(\tilde{u})$ to the solution of problem $\bar{x}_\ell^*$. Irrelevant restrictions (3) are also removed from problem (1)-(3) (for these restrictions $u^* = 0$). The generated in this way task will be denoted as task (1*)-(3*).

Decomposition in block structure problems

Quadratic perturbation. It can be shown that if the solution to the original problem is unique, then the number of blocks of problem (1*)-(3*) is no more than m.

In this regard, when constructing a mathematical model of an applied optimization problem, it is necessary, if possible, to ensure the uniqueness of its solution. For this, a quadratic perturbation of the objective function is often used: to $f_\ell(\bar{x}_\ell)$ we add a quadratic perturbation term $\varepsilon_\ell(\bar{x}_\ell - \bar{z}_\ell)^2$, where $\bar{z}_\ell$ is an apriori estimate of the solution $\bar{x}^*_\ell$, $\varepsilon_\ell > 0$ – quadratic perturbation parameter. The permissible degree of disturbance is determined based on trial solutions of the problem. It is advisable to use the quadratic perturbation procedure at the second stage when solving problem (1*)-(3*).

Decomposition in block structure problems

Averaging according to Cesaro. The solution to problem (1*)-(3*) can be determined based on the following procedure for averaging the sequence $x(u_k)$ (Cesaro averaging). The sequence of points u_k is determined by the classical subgradient algorithm in relation to problem (4):

$u_{k+1} = u_k + h_k g(x(u_k))$,
$h_k > 0,\ h_k \to 0,\ \sum h_k = \infty$.

Then any limit point of the sequence $z_K = \sum_{k=0}^{K} h_k x(u_k) / \sum_{k=0}^{K} h_k$ is a solution to problem (1*)-(3*).

The solution to the dual problem (4) obtained at the first stage is used as the starting point u_0.

In practice, you can use subgradient descent with a constant step:

$u_{k+1} = u_k + hg(x(u_k))$, then $z_K = \sum_{k=0}^{K} x(u_k)/K$

Decomposition in block structure problems

Evaluation of solution accuracy.
Let $\tilde{u}$, $\tilde{x}$ be the obtained approximate solutions to problems (4) and (1)-(3). Assessing the accuracy of the solution is quite simple.
$\tilde{x}$ satisfies conditions (2).
The accuracy of the fulfillment of constraints (3) is checked directly by calculating $g^i(\tilde{x})$.
The optimality conditions $f(x^*) = \psi(u^*)$ are checked by calculating the value $f(\tilde{x}) - \psi(\tilde{u})$.

Decomposition in block structure problems

Procedure of "holding zero".
It is advisable to assume that the starting point u^0 for algorithms is equal to 0.
In this case, use the following procedure of "holding zero". If at iteration k $u_i^k \approx 0$ and $g_i(x(u^k)) < 0$, then we set $g_i(x(u^k)) = 0$.
The meaning of such a procedure is that for an insignificant constraint i (that is, $g_i(x(u^k)) < 0$), movement along the component i from 0 to the invalid region of the variable u_i is not allowed.

Planning Multi-Product Flows

Transport network. The transport network is represented by a directed graph $G = \{I, J\}$. I (data) is the set of vertices (nodes, stations) of the graph, J (data) is a set of oriented edges (arcs, sections) of the graph.

Transportation correspondence. L(data) is the set of types of products (cargo). Q (data) is a set of correspondences. For the correspondence $q \in Q$ the following data are defined: cargo type l_q(data), projected volume of transportation b_q(data).

For correspondence q a number of ways of its implementation are given Ω_q(data). For $\omega \in \Omega_q$ the following data are defined: route $\Re_q(\omega)$(data), the tariff for the cost of transportation of a cargo unit $c_q^+(\omega)$(data).

$x_q^+(\omega) \geq 0$(result) is the volume of correspondence transportation q according to the variant of its implementation ω. The total volume of realized transportation of correspondence should not be greater than the specified projected volume of correspondence b_q.

Planning Multi-Product Flows

Denote as $x_q^- \geq 0$(result) unrealized volume of the correspondence q. c_q^-(data) is a penalty factor for the presence of a unit of unrealized volume of correspondence q. Variables $x_q^+(\omega)$ and x_q^- satisfy the balance ratios:

$$\sum \{x_q^+(\omega) | \omega \in \Omega_q\} + x_q^- = b_q,\ q \in Q, x_q^+(\omega) \geq 0,\ x_q^- \geq 0. \quad (6)$$

X_q^- is the set of all variables x_q^- : $X_q^- = \{x_q^- \,|\, q \in Q\}$. Total income $F^+(X_q^+)$ from the implementation of correspondence and a general penalty $F_b^-(X_q^-)$ for the presence of unrealized shipments are determined by equalities:
$F^+(X_q^+) = \sum \sum \{c_q^+(\omega) x_q^+(\omega) | \omega \in \Omega_q; q \in Q\}$,
$F_b^-(X_q^-) = \sum \{c_q^- x_q^- | q \in Q\}$. $F_c^-(X_q^+)$(result) is operating costs for the transportation of all correspondence:
$F_c^-(X_q^+) = \sum \{c_{qj}^- w^q(j) | q \in Q; j \in J\}$, where c_{qj}^- is costs for transportation of a unit volume of correspondence q on the edge j.

Planning Multi-Product Flows

Bandwidth of sections. J^* (data) is a set of sections of the transport network with limited bandwidth ($J^* \subset J$). r_j(data) is a bandwidth of the edge j. Let $w^q(j)$(result) be the volume of correspondence transported on the edge j. The conditions for limiting the bandwidth of the sections, taking into account their reconstruction:

$$\sum \{w^q(j) | q \in Q\} \leq r_j + x_j^r,\ j \in J^*, \tag{7}$$

$$\bar{x}_j^r \geq x_j^r \geq 0,\ j \in J^*, \tag{8}$$

where x_j^r(result) is the increase in capacity of the edge's throughput due to its reconstruction, $\bar{x}_j^r$(data) is the upper limit for the variable x_j^r, X_r is the set of all variables x_j^r : $X_r = \{x_j^r \,|\, j \in J^*\}$. $F_r^-(X_r)$ is the total amount of funds spent on the reconstruction of sections of the transport network: $F_r^-(X_r) = \sum \{c_j^r x_j^r | j \in J^*\}$, where c_j^r(data) means specific financial costs of edge reconstruction (costs for increasing throughput by one unit).

Planning Multi-Product Flows

B^f(data) is the maximum amount of finance for the reconstruction of sections of the transport network:

$$F_r^-(X_r) \leq B^f. \tag{9}$$

Let $X = \{X_q^+, X_q^-, X_r\}$. The task is to maximize the function $F(X)$:

$$\max F(X) = F^+(X_Q^+) - F_B^-(X_Q^-) - F_C^-(X_Q^+) - F_R^-(X_R), \tag{10}$$

under constraints (6), (7), (8), (9). The main variables of the problem are the following: X_q^+ is volumes of realized transportation of correspondence, X_q^- is unrealized volumes of correspondence transportation, X_r is variable increases in the bandwidth of edges. The rest of the variables and the above ratios are auxiliary and are introduced for the compactness of the record and ease of interpretation of the model.

Planning Multi-Product Flows

Solution method. The main feature of real problems (6)–(10) is their large dimension: the number of correspondences ($|Q|$) and network sections ($|J|$) can reach hundreds of thousands. The constraints (7), (9) are the "binding constraints"of task (6)–(10). The constraints (7) are constraints on the bandwidth of individual sections of the network. The constraint (9) corresponds to the financial limit for the reconstruction of the transport network. The number of constraints (7), as a rule, is significantly less than the number of all sections of the network and is of the order of several dozen.

The method of solving problem (6)–(10) is based on the use of non-smooth optimization algorithms in decomposition schemes according to binding constraints (7), (9).

For this model, the solution to the internal problems of the decomposition method (minimization of the Lagrange function) is determined analytically.

Planning Multi-Product Flows

Software. The software is developed in the C++ language in the style of object-oriented programming. A new modification $r(\sigma)$ of the r-algorithm is used to solve the dual problem. Unlike the r-algorithm, the values of the dilation coefficients of the space of variables in the proposed modification are determined programmatically during the operation of the algorithm. As basic elements, the software uses the following algorithms:

- the algorithm for finding the shortest paths on a graph,
- the algorithm for solving the transport problem on the graph,
- non-smooth optimization algorithm.

Conclusions

The developed non-smooth optimization algorithms can be effectively used to solve a wide class of problems of long-term planning of material and technical supply and development of transport systems. The paper presents numerical aspects of the use of these algorithms for solving high-dimensional problems based on decomposition schemes.

REFERENCES

1. N. Z. Shor, “Minimization methods for non-differentiable functions”, Springer-Verlag, Berlin, 1985, 162 p.
2. V. S. Mikhalevich, V. A. Trubin, and N. Z Shor,. “Optimization problems of production and transport planning: Models, methods, algorithms”. Science, Moscow, 1986, 264 p..
3. N.G. Zhurbenko, “Towards a two-stage scheme of the decomposition method for solving block linear programming problems”, Theory of optimal solutions, Kyiv, 2001, pp. 22–26.
4. L. V. Belyaeva, N. G. Zhurbenko, and N. Z. Shor. "On a method for solving one class of dynamic distribution problems Economics and Mathematical Methods, 1983. Vol. 14 (11), pp. 137-146.
5. N. G. Zhurbenko “Numerical efficiency of a modification of the r-algorithm”, Theory of optimal solutions, Optimal Theory. Decisions, Kyiv, 2017, pp. 33–38.

ACKNOWLEDGEMENTS

The author is grateful to Bardadym T.A. for help with the preparation of the article.
The work was supported by a grant from the Volkswagen Foundation (grant № 97775).

Thank you for attention!

zhurbnick@gmail.com