ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 8, pp. 947–955.

Generalization of the Carathéodory Theorem and the Maximum Principle in Averaged Problems of Non-Linear Programming

A. M. Tsirlin

Ailamazyan Program Systems Institute, Russian Academy of Sciences, Pereslavl-Zalessky, Russia
e-mail: tsirlin@sarc.botik.ru

Received June 9, 2022
Revised March 7, 2023
Accepted June 9, 2023

Abstract—The relationship between the averaging of functions over time and its averaging over the set of values of the required variables is considered. Optimization problems are studied, the criterion and constraints of which include the averaging of functions or functions of the average values of variables. It is shown that the optimality conditions for these problems have the form of the maximum principle, and their optimal solution in the time domain is a piecewise constant function. A generalization of Carathéodory's theorem on convex hulls of a function is proved. Optimality conditions are obtained for non-linear programming problems with averaging over a part of the variables and functions depending on the average values of the variables.

Keywords: averaged constraints, sliding modes, convex hulls of functions, reachability function, maximum principle in averaged problems

DOI: 10.25728/arcRAS.2023.17.50.001

1. INTRODUCTION

For a wide class of problems, the optimality criterion and all or part of the constraints averagely depend on all or part of the variables. Such problems arise when, in technological processes, some variables to be selected must be unchanged (design parameters), while others may change over time, and the presence of devices that smooth out fluctuations, e.g. capacitances, leads to the average influence of these changes [1]. Such problems arise in the optimal control of macrosystems (systems consisting of a set of individually uncontrollable elements), in which it is possible to control only the average parameters of the set of these elements. All such problems are called averaged optimization problems.

In systems, whose set of admissible controls is non-convex (e.g. relay systems), the optimal solution is often a sliding mode, in which the change of the object state depends averagely on any frequently switching control [2–5]. Averaged problems also arise as auxiliary estimation problems in the optimization of cyclic modes, when the introduction of averaging expands the set of admissible solutions and simplifies the solution, allowing to obtain an estimate of the efficiency of the cyclic mode without finding the form of optimal cycles. The value of such an estimation problem is known to be "not worse" than the value of the initial one, and its optimal solution contains useful information about the nature of the optimal solution of the initial one. For definiteness, we will consider problems for the maximum of the optimality criterion.

In the first section of this paper, we will discuss the relationship between the averaging of functions whose argument varies in time over a set of values of that argument and over time, and define what is sought as a solution to the averaged problem and how this solution can be implemented. In the second section, we will formulate the theorem on the optimality conditions of the non-linear programming problem with averaging of the optimality criterion and constraints and give its proof based on Carathéodory's theorem on convex hulls of functions. In the third section, we will consider possible generalizations of the proved theorem.

2. ON THE RELATIONSHIP BETWEEN TIME AVERAGING AND SET AVERAGING

The mean value of the continuous scalar function $f(x(t))$, $t \in [0, \tau]$, $x \in V \subset R^n$ can be calculated on time as

$$\overline{f_t(x)} = \frac{1}{\tau} \int_0^\tau f(x(t))dt \tag{1}$$

or on set as

$$\overline{f_p(x)} = \int_V f(x)p(x)dx. \tag{2}$$

The function $p(x)$ is called the distribution density. When $x(t)$ is a random function, $p(x)$ is the distribution density of the random variable. It is non-negative and its integral on V is equal to one. In particular, the set V can be a parallelepiped in R^n. In our case, $x(t)$ is a determined function, so let us focus more on the properties of $p(x)$ such that the results of averaging by formulas (1) and (2) are the same.

Let us consider the variable x as scalar, the set V here and below as bounded and closed, and introduce the function $\theta(x_0)$, $x_0 \in V$, equal to the total duration of those time intervals t, for which $x(t) \leqslant x_0$. It is obvious that this function does not exceed τ. Through $P(x_0)$, let us denote the ratio $\frac{\theta(x_0)}{\tau}$, i.e., the fraction of the interval $[0, \tau]$, for which $x(t) \leqslant x_0$. This function grows monotonically as x_0 increases, varying from zero to one. It is similar to the distribution function of a random variable.

The distribution density is equal to

$$p(x_0) = \frac{dP(x_0)}{dx_0} = \frac{1}{\tau}\frac{d\theta(x_0)}{dx_0} = \frac{1}{\tau}\frac{1}{\sum\limits_\nu \left|\frac{dx_\nu}{dt}\right|_{x_\nu = x_0}}. \tag{3}$$

The interval θ increases as x_0 increases for any sign of the derivative at those values x_ν of the function $x(t)$, in which it is equal to x_0.

If at some value of x_0 the function $x(t)$ is constant over a fraction γ of the interval $[0, \tau]$, then the function $P(x_0)$ experiences a jump of magnitude γ at that point, and the distribution density at that point is equal to $\gamma\delta(x - x_0)$.

Examples

1. **Linear functions.** Let $x(t) = \frac{ht}{\tau}$. Then, according to formula (3), we get $p(x) = \frac{1}{h} = \text{const}$. The same distribution density corresponds to all triangles with base $[0, \tau]$ and height h.

2. **Piecewise constant functions.** These functions take discrete values of x_i, each within a fraction γ_i of the interval $[0, \tau]$. Any such function, according to formula (3), corresponds to the distribution density function (3)

$$p(x_0) = \sum_i \gamma_i \delta(x - x_i), \quad \gamma_i > 0, \quad \sum_i \gamma_i = 1. \tag{4}$$

The order, in which the piecewise constant function takes one or another of the possible values, does not matter.

From these examples we see that *every function $x(t)$ corresponds to the distribution density of its values $p(x)$ defined on V, and every distribution density corresponds to any number of functions $x(t)$, for which $\overline{f_p(x)} = \overline{f_t(x)}$.* An exception is the distribution density of the form

$p(x) = \delta(x - x_1)$. In this case, the corresponding function is $x(t) = x_1 = \text{const}$ over the entire interval $[0, \tau]$, and it is unique.

Let us consider the case when the function f depends on several variables (e.g., for the sake of simplicity, on two variables, $x_1(t)$ and $x_2(t)$). In this case, the distribution function $P(x^0)$ of the values of vector x represents the fraction of the interval $[0, \tau]$, for which the two following inequalities are satisfied: $x_1(t) \leqslant x_1^0$ and $x_2(t) \leqslant x_2^0$. This function grows monotonically with the growth of each of the arguments. When the first of the components of the vector x^0 is at the maximum $(p_1(x_1) = 1)$, it is equal to and its derivative is equal to the distribution density $p(x_1^{\max}, x_2) = p_2(x_2)$. Similarly, when $x_2 = x_2^{\max}$, $p(x_2^{\max}, x_1) = p_1(x_1)$. The functions $x_1(t)$ and $x_2(t)$ are independent of each other, so $p(x_1, x_2) = p_1(x_1)p_2(x_2)$.

The sought solution to the averaged optimization problem is the distribution density $p^*(x)$ of the vector x on the set V of its admissible values. To implement this solution over time, we need to find one of the possible functions $x(t)$ having the distribution $p^*(x)$. The solution of this last problem is greatly facilitated by the peculiarities of optimal solutions of $p^*(x)$ proved in the next section.

3. ON THE OPTIMAL SOLUTION OF AVERAGED OPTIMIZATION PROBLEMS

We will denote the averaging operation by a line drawn over the function or vector to be averaged. Thus,

$$\overline{x} = \int_V x p(x) dx, \quad \overline{f(x)} = \int_V f(x) p(x) dx.$$

The simplest problem of averaged optimization is the problem of maximizing the average value of a scalar function f(x) at a given average value of its argument:

$$\overline{f(x)} \to \max \Big/ \overline{x} = x_0, \quad x \in V \subset R^n. \tag{5}$$

Or in a more detailed form

$$\int_V f(x) p(x) dx \to \max \Big/ \int_V x p(x) dx = x_0, \quad p(x) \geqslant 0, \quad \int_V p(x) dx = 1. \tag{6}$$

The sought function in this problem is $p(x)$ (the distribution density of the vector of sought variables). This function is non-negative and its integral on the set V is equal to one.

4. CARATHÉODORY'S THEOREM ON CONVEX HULLS OF FUNCTIONS

Carathéodory's theorem [3, 4, 6] on convex hulls of sets states that any element of a convex hull CoD of a compact set D in Euclidean space of dimension n can be represented as an element of a simplex having at most $n + 1$ vertices (base points), each of which belongs to D.

In particular, a subgraph of function $f(x)$ can be the set D. The convex hull of a function is the convex hull of a subgraph. A function depending on n variables is the boundary of a set in R^{n+1} space of dimension n. The basis points are known to lie on this boundary, and hence their number does not exceed $n + 1$. Below we will call Carathéodory's theorem the theorem on convex hulls of functions.

The ordinate of the convex hull of the function $f_0(x)$ at the point x_0 belonging to the convex hull of the set of the function definition is the value of the problem

$$\overline{f_0(x)} \to \max_{p(x)} \Big/ \begin{array}{l} \overline{x_i} = x_{i0}, \quad i = \overline{1, n}, \\ x \in V \subset R^n, \end{array} \tag{7}$$

where V is the compact space.

According to Carathéodory's theorem, the optimal solution of this problem is

$$p^*(x) = \sum_{j=0}^{n} \gamma_j \delta(x - x^j), \quad \gamma_j \geqslant 0, \quad \sum_{i=0}^{n} \gamma_j = 1.$$

That is, the optimal allocation is concentrated in at most $(n+1)$ base points.

This fact allows us to rewrite the problem (7) as a non-linear programming problem

$$\sum_{j=0}^{n} \gamma_j f_0(x^j) \to \max \Bigg/ \begin{array}{l} \sum\limits_{j=0}^{n} \gamma_j x^j = x_0, \\ x^j \in V \subset R^m, \quad \sum\limits_{j=0}^{n} \gamma_j = 1, \quad \gamma_j \geqslant 0, \end{array} \tag{8}$$

whose variables are the basis vectors x^j and the vector of weight coefficients γ, and use the Kuhn–Tucker theorem [7] to solve it:

If y^* is a solution to a non-linear programming problem

$$f(y) \to \max \big/ \varphi_i(y) \leqslant 0, \quad y_j \geqslant 0, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n, \tag{9}$$

then there is such a non-zero vector of multipliers

$$\lambda = \lambda_0, \ldots, \lambda_m \ \ (\lambda_0 \textit{ equal to } 0 \textit{ or } 1, \quad \lambda_i \leqslant 0 \quad \textit{when} \quad i > 0),$$

that for the Lagrangian function

$$R = \lambda_0 f(y) + \sum_{i=1}^{m} \lambda_i \varphi_i(y)$$

the following conditions are true:

$$\left(\frac{\partial R}{\partial y_j}\right)_{y=y^*} = 0, \text{ if } y_j^* > 0; \quad \left(\frac{\partial R}{\partial y_j}\right)_{y=y^*} \leqslant 0, \text{ if } y_j^* = 0; \tag{10}$$

$$\lambda_i = 0, \text{ if } \varphi_i(y^*) < 0; \quad \lambda_i \leqslant 0, \text{ if } \varphi_i(y^*) = 0. \tag{11}$$

For problem (8) the Lagrangian function takes the form

$$R = \sum_{j=0}^{n} \gamma_j \left[f_0(x^j) + \sum_{i=1}^{n} \lambda_i x_i^j - \Lambda \right], \tag{12}$$

where Λ is the Lagrange multiplier corresponding to the condition of equality of the sum of weight coefficients to one.

Kuhn–Tucker conditions on weighting factors lead to requirements:

$$R^0(x_j, \lambda) = f_0(x^j) + \sum_{i=1}^{n} \lambda_i x_i^j < \Lambda, \text{ if } \gamma_j = 0, \tag{13}$$

$$R^0(x_j, \lambda) = f_0(x^j) + \sum_{i=1}^{n} \lambda_i x_i^j = \Lambda, \text{ if } \gamma_j > 0, \quad j = 0, \ldots, n+1.$$

Here, R^0 is the Lagrangian function of problem (8) without averaging. Hereafter such a problem will be called the *initial* one.

Thus, *for all base values of x included in the optimal solution of the convex hull problem of the function f_0 with non-zero weight, the Lagrangian function of the original problem is maximal. The number of such points does not exceed $n + 1$.*

5. PROBLEM WITH BOND AVERAGING, GENERALIZATION OF CARATHÉODORY'S THEOREM

In the non-linear programming problem with averaging functions defining relations between variables, it is required to maximize the average value of the function $f_0(x)$ on the set V of admissible values of x, provided that the average value of the vector function $f(x) = (f_1(x), \dots, f_i(x), \dots, f_m(x))$ is equal to zero. Formally,

$$\overline{f_0(x)} \to \max \Big/ \overline{f_i(x)} = 0, \quad i = 1, \dots, m, \quad x \in V \in R^n. \tag{14}$$

Theorem 1. 1. *The optimal distribution density in problem* (14) *has the form*

$$p^*(x) = \sum_{j=0}^{m} \gamma_j \delta(x - x^j), \quad \gamma_j \geqslant 0, \ \sum_{j=0}^{m} \gamma_j = 1. \tag{15}$$

2. *There is such a non-zero vector*

$$\lambda = \lambda_0, \dots, \lambda_i, \dots, \lambda_m, \quad \lambda_0 = (0; 1),$$

that, at each base point x^j, the Lagrangian function of the original problem

$$R = \sum_{i=0}^{m} \lambda_i f_i(x) \tag{16}$$

is maximal over $x \in V$.

Proof. To prove this statement, we will introduce the concept of the *reachability function* of the problem (14):

$$f_0^*(C) = \max f_0(x) \Big/ f_k(x) = C_k, \quad k = 1, \dots, m, \quad x \in V. \tag{17}$$

This function is defined algorithmically on the set

$$V_c = \{C \in R^m : f(x) = C, x \in V \subset R^n\}.$$

It may be non-smooth and semi-continuous on top.

The following statement is true.

Statement. *For those values of x, for which $f(x) = C$, $p^*(x)$ is deliberately equal to zero if $f_0(x) \neq f_0^*(C)$.*

Thus, only those values $x = x^*(C)$, for which the value of $f_0(x)$ coincides with the ordinate of the reachability function, can be included in the solution of the averaged problem with non-zero weight. If this statement were not true, it would be possible to change the density of the distribution so that the average value of $f_0(x)$ would increase.

Since for each C the value of f_0 coincides with the ordinate of the reachability function, the problem (14) can be rewritten as

$$\overline{f_0^*(C)} \to \max \Big/ \overline{C_k} = 0, \quad k = 1, \dots, m, \quad C \in V_c \subset R^m. \tag{18}$$

This is the problem on the ordinate of the convex hull of the reachability function at zero. According to Carathéodory's theorem, its optimal solution is equal to

$$p^*(C) = \sum_{j=0}^{m} \gamma_j \delta(C - C^j), \quad \gamma_j \geqslant 0, \ \sum_{j=0}^{m} \gamma_j = 1. \tag{19}$$

Since each base value of C^j corresponds to the value of $x^{j*}(C^j)$, the optimal distribution density in problem (14) is of the form (15). The first statement of Theorem 1 is proved.

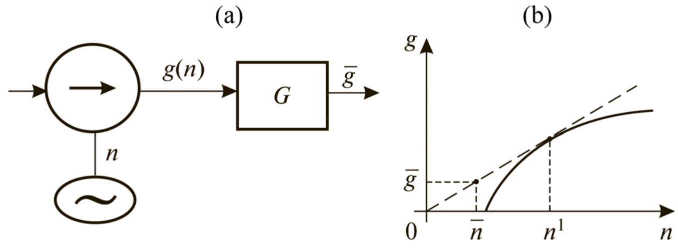

Fig. 1. System consisting of a pump and a smoothing tank (a); relation between flow rate and power input (b).

The proof of the second statement completely repeats the analogous proof for the problem on the ordinate of the convex hull of a function with the difference that the Lagrangian function of the non-averaged problem has the form (16). We emphasize that the number of base points does not depend on the dimensionality of the vector x, but is determined by the dimensionality m of the vector function f.

Note that here and below conditions in the form of the maximum principle do not require the functions defining the averaged problem to be smooth on x, the set V can be non-contiguous [8–10].

Example 1. Let us consider the system consisting of an electric motor, a pump rotated by it and a vessel in Fig. 1a. The motor consumes power n, on which depends the pump capacity g. The dependence of $g(n)$ is shown in Fig. 1b. It is required to find the mode for which, for a given average power input $\overline{n}$, the average pumping capacity $\overline{g}$ is maximized. This is the problem of the ordinate of the convex hull of the function $g(n)$ at the point $\overline{n}$. The number of base points is two, one of them is the origin of coordinates, and the second one, n^1, is defined by the condition that the Lagrangian function $R = g(n) + \lambda n$ reaches the maximum in it (the same as at $n = 0$). Excluding λ from the conditions for the maximum of the Lagrangian function and the requirement that this maximum be zero, we reach the equation for n^1:

$$\frac{g(n)}{n} = \frac{dg(n)}{dn}.$$

There are many optimal implementations of this solution over time, and for each of them the pump power takes values zero and n^1, and the fraction of the interval τ, for which $n = n^1$, is equal to $1 - \frac{\overline{n}}{n^1}$. The maximum value of the interval τ is determined by the value of capacitance G, it is equal to

$$\tau_{\max} = \frac{2G}{g(n^1)}.$$

The value of the problem is equal to

$$\overline{g}^* = g(n^1)\left(1 - \frac{\overline{n}}{n^1}\right).$$

It does not depend on G, and the sliding mode is the optimal solution when the capacitance goes down to zero.

6. GENERALIZATIONS OF THE AVERAGED NON-LINEAR PROGRAMMING PROBLEM

6.1. Averaged Problem with Deterministic Variables

As mentioned in the introduction, there can be two types of variables in averaged problems: randomized and deterministic. There is no averaging for variables of the second type. Let us consider a non-linear programming problem, in which some variables are not averaged.

The problem with averaging over a part of variables will take the form:

$$\overline{f_0(x,y)} \to \max \Big/ \overline{f_j(x,y)} = 0, \quad x \in V \subset R^n, \quad y \in V_y \subset R^K, \quad j = 1, \dots, m, \tag{20}$$

functions $f_0, \dots, f_m$ are continuous and continuously differentiable over the set of arguments, the line corresponds to averaging over $x \in V$, the sets V and V_y are closed and bounded.

For any y, this problem is an averaged non-linear programming problem (14), and hence, due to the theorem, the optimal distribution density x is concentrated in at most $(m+1)$ base points, so that $p^*(x) = \sum_0^m \gamma_j \delta(x - x^j)$ and there exists such a non-zero vector λ that, at each of these points, the Lagrangian function of the original problem

$$R = \sum_{j=0}^{m} \lambda_j f_j(x,y), \quad x \in V \subset R^n, \quad y \in V_y \subset R^K \tag{21}$$

is maximal on x.

The Lagrangian function of the problem (20), in which the distribution density x is equal to $p^*(x)$, has the form

$$\overline{R^*} = \sum_{j=0}^{m} \lambda_j \sum_{i=0}^{m} \gamma_i f_j(x^i, y), \quad x^i \in V \subset R^n, \quad y \in V_y \subset R^K. \tag{22}$$

For any distribution density of randomized variables $p(x)$, the problem (20) is a non-linear programming problem and, according to Kuhn–Tucker theorem, there is such a non-zero vector λ with components $\lambda_0 = (0;1)$, λ_j, $j = 1, \dots, m$, that the conditions of local non-improvability on y are satisfied for the function (21) at the optimal solution

$$\frac{\partial \overline{R^*}}{\partial y_l} \delta y_l \leqslant 0, \quad l = 1, \dots, K. \tag{23}$$

Here, δy_l is the acceptable variation of y_l.

The non-linear programming problem with averaging over a part of variables has much in common with the optimal control problem with links in the form of differential equations. There, the control actions enter the problem in such a way that their fast changes are averaged in the neighborhood of each time instant, which cannot be said about the phase coordinates. That is why the conditions in the form of Pontryagin's maximum principle are valid for the control actions.

6.2. Problem Containing Functions of Mean Values of Variables

This problem has the form

$$\overline{f_0(x, \overline{x_l})} \to \max \Big/ \overline{f_j(x, \overline{x_l})} = 0, \quad x \in V \subset R^n, \quad l = 1, \dots, \quad K \leqslant n. \tag{24}$$

Let us introduce the notation: $y_l = \overline{x_l}$. The variable y_l belongs to the convex hull CoV_{x_l} of the set of admissible values x_l. Given the introduced notations, the problem (24) can be rewritten as

$$\overline{f_0(x,y)} \to \max \Big/ \overline{f_j(x,y)} = 0, \quad \overline{x_l} - y_l = 0, \quad x \in V \subset R^n, \quad y_l \in CoV_{x_l} \subset R^K. \tag{25}$$

When written in this form, problem (25) differs from problem (20) only by additional averaged conditions $\overline{x_l} - y_l = 0$. The Lagrangian function of the original problem will take the form

$$R = \sum_{j=0}^{m} \lambda_j f_j(x,y) + \sum_{l=1}^{K} \lambda_l (x_l - y_l), \quad x \in V \subset R^n, \quad y_l \in CoV_{x_l}. \tag{26}$$

From the optimality conditions (21), (23) it follows that the maximum number of base values of x in problem (24) is $m + K + 1$ and that there is such a non-zero vector λ that at each of the base points the function R appearing in (26) reaches a maximum on the optimal solution on x, while, on y, the function (22) is locally non-improvable.

When solving averaged problems, the Lagrange multipliers are expressed through the base values x^j and y from the condition of maximum of the Lagrangian function on x and equality of these maxima to each other, as well as the conditions of non-improvability on y are written down. After that, from the averaged conditions, the weight coefficients for each of the base points are found, given that the sum of these weight coefficients is equal to one.

Example 2. As an illustrative example, let us consider the following problem

$$\overline{(x-\overline{x})^2} \to \min \Big/ \overline{\left(\frac{1}{x+\overline{x}}\right)} = 1, \quad x = -1; 0; 1. \tag{27}$$

The Lagrangian function for this problem is equal to

$$L = (x-y)^2 + \lambda\left(\frac{1}{x+y-1}\right) + \mu(y-x). \tag{28}$$

The number of averaged conditions is two, hence all three admissible values of x are basic, and the uncertain multipliers must be chosen so that the maximum of the function L^* is the same at these points, which leads to the following conditions:

$$\begin{aligned} L^* = (1+y)^2 + \lambda\left(\frac{1}{y-1} - 1\right) - \mu(1+y) = y^2 + \lambda\left(\frac{1}{y} - 1\right) - \mu y \\ = (1-y)^2 + \lambda\left(\frac{1}{y+1} - 1\right) + \mu(1-y). \end{aligned} \tag{29}$$

Thus,

$$\lambda = \frac{2y}{2+y}(1-y-2y^2), \quad \mu = \frac{y(2y-1)}{2+y}. \tag{30}$$

After substituting these expressions into L^* and differentiating the resulting expression by y, we come to the equation

$$3y^3 + 17y^2 + 20y - 10 = 0. \tag{31}$$

To the second decimal place, $y = 0.37$. Weight multipliers $\gamma_1, \gamma_2, \gamma_3$ for $x = -1$, $x = 0$, $x = 1$, respectively, can now be found from the following conditions

$$\sum_{i=1}^{3} \gamma_i = 1, \quad \sum_{i=1}^{3} \gamma_i x_i = 0.37, \quad \sum_{i=1}^{3} \gamma_i \frac{1}{x_i + 0.37} = 1. \tag{32}$$

Then we get $\gamma_1 = 0.155$, $\gamma_2 = 0.320$, $\gamma_3 = 0.525$.

7. CONCLUSION

Various formulations of non-linear programming problems with averaging were considered. It is shown that, with the introduction of the concept of reachability function of non-linear programming problems, the problems containing averaging of functions from a vector of randomized variables x can be reduced to extremal problems on convex hulls of sets and functions. The optimal distribution in all these problems is centered in discrete "base" points of the compact set V of admissible values of x. The maximum principle for such problems is proved. It is shown that the number of base points does not exceed the number of averaged conditions in the problem by more than one. The criterion and constraints of averaged non-linear programming problems may depend on time. If these dependencies are continuous, then the above optimality conditions are valid for each moment of time and determine the time variation of the coordinates of the base points and their weights. The vector of Lagrange indeterminate multipliers corresponding to the optimal solution delivers the minimum to the maximum value of the maximized function with respect to the sought variables, which serves as a basis for computational algorithms.

FUNDING

This work was supported by the Russian Science Foundation, project no. 20-61-46013.

REFERENCES

1. Tsirlin, A.M., *Optimal'nye tsikly i tsiklicheskie rezhimy* (Optimal Cycles and Cycling Modes), Moscow: Energoatomizdat, 1983.
2. Rozonoer, L.I., Pontryagin Maximum Principle in the Theory of Optimal Systems, *Autom. Remote Control*, 1959, no. 10, pp. 1320–1334; no. 11, pp. 1441–1458; no. 12, pp. 1561–1578.
3. Yudin, D.B., *Matematicheskie metody upravleniya v usloviyakh nepolnoi informatsii* (Mathematical Methods of Management under Conditions of Incomplete Information), Moscow: Sovetskoe Radio, 1974.
4. Yang, L., *Lectures on the Calculus of Variations and Optimal Control Theory*, London: Saunders, 1969. Translated under the title *Lektsii po variatsionnomu ischisleniyu i optimal'nomu upravleniyu*, Moscow: Mir, 1974.
5. Tsirlin, A.M., Optimization in Mean and Sliding Modes in Optimal Control Problems, *Izv. Akad. Nauk SSSR. Tekhn. Kibernetika*, 1974, no. 2, pp. 143–151.
6. Polovinkin, E.S. and Balashov, M.V., *Elementy vypuklogo i sil'no vypuklogo analiza* (Elements of Convex and Strongly Convex Analysis), Moscow: Fizmatlit, 2004. ISBN 5-9221-0499-3.
7. Himmelblau, D.M., *Applied Nonlinear Programming*, New York: McGraw-Hill, 1972. Translated under the title *Prikladnoe nelineinoe programmirovanie*, Moscow: Mir, 1975.
8. Afanas'ev, A.P., Dikusar, V.V., Milyutin, A.A., and Chyukanov, S.A., *Neobkhodimoe uslovie v optimal'nom upravlenii* (Necessary Condition in Optimal Control), Moscow: Nauka, 1990.
9. Dubovitskii, A.Ya. and Milyutin, A.A., *Teoriya printsipa maksimuma. Metody teorii ekstremal'nykh zadach v ekonomike* (Theory of the Maximum Principle. Methods of the Theory of Extreme Problems in Economics), Moscow: Nauka, 1981.
10. Tsirlin, A.M., Optimality Conditions of Sliding Modes and the Maximum Principle for Control Problems with the Scalar Argument, *Autom. Remote Control*, 2009, no. 5, pp. 839–854.

This paper was recommended for publication by P.S. Shcherbakov, a member of the Editorial Board