Koba Gelashvili

ABOUT A SIMPLE ANALYTIC PROOF
OF PONTRYAGIN'S MAXIMUM PRINCIPLE

Abstract. Using compact expressions for optimal and varied trajectories based on the representation of the Cauchy problem solution as evolution system of functions of initial values, a simple analytic proof of Pontryagin's Maximum Principle for fixed endpoints is obtained. Some non-standard problems of optimal control theory are considered shortly, too.

რეზიუმე. გამოყენებულია რა კომპაქტური გამოსახულებები ოპტიმალური და ვარირებული ტრაექტორიებისთვის, რომლებიც ეფუძნება კოშის ამოცანის ამონახსნის წარმოდგენას საწყის მნიშვნელობათა ფუნქციების ევოლუციური სისტემის სახით, მიღებულია პონტრიაგინის მაქსიმუმის პრინციპის მარტივი ანალიზური დამკიცება ფიქსირებული ბოლოების შემთხვევაში. ოპტიმალური მართვის თეორიის ზოგიერთი არასტანდარტული ამოცანა აგრეთვე არის მოკლედ განხილული.

Statement of the problem. Pontryagin's Maximum Principle. Usually, by $\mathbb{R}_k$ we denote the set of collections of k real numbers $(x^1, \ldots, x^k)$, endowed with natural algebraic operations and Euclidean norm. For $g = (g_1, \ldots, g_m) : \mathbb{R}_k \to \mathbb{R}_m$, Jacobi's matrix at the point $x = (x^1, \ldots, x^k)$ is denoted by

$$
g'(x) = \begin{bmatrix} \frac{\partial g_1(x)}{\partial x^1} & \cdots & \frac{\partial g_1(x)}{\partial x^k} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m(x)}{\partial x^1} & \cdots & \frac{\partial g_m(x)}{\partial x^k} \end{bmatrix},
$$

and for a vector function $g = (g_1, \ldots, g_m) : \mathbb{R}_k \times \mathbb{R}_l \to \mathbb{R}_m$, Jacobi's matrix with respect to $x = (x^1, \ldots, x^k)$ at the point $(x, y) \in \mathbb{R}_k \times \mathbb{R}_l$ is denoted by

$$
\frac{\partial g(x, y)}{\partial x} = \begin{bmatrix} \frac{\partial g_1(x,y)}{\partial x^1} & \cdots & \frac{\partial g_1(x,y)}{\partial x^k} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m(x,y)}{\partial x^1} & \cdots & \frac{\partial g_m(x,y)}{\partial x^k} \end{bmatrix}.
$$

For a vector or matrix function $x(t)$, $t \in \mathbb{R}$, $\dot{x}(t)$ denotes the componentwise derivative of $x(t)$; for example, if $x(t) = (x^1(t), \ldots, x^k(t))$, then

$$
\dot{x}(t) = (\dot{x}^1(t), \ldots, \dot{x}^k(t)).
$$

Let us formulate the main problem of the optimal control theory.

Let $U \subset \mathbb{R}_k$ be the control area, the functions $f_i(x^1, \ldots, x^n, u_1, \ldots, u_k)$ and $\frac{\partial f_i}{\partial x^j}$ continuously map $\mathbb{R}_n \times U$ into $\mathbb{R}$, $i = 0, 1, \ldots, n$, $j = 1, \ldots, n$; $[t_0, t_1]$ be a fixed non-trivial segment (i.e., $t_0 < t_1$); Ω be the set of admissible controls consisting the functions $u(t) = (u_1(t), \ldots, u_k(t))$ such that $u(\cdot) : [t_0, t_1] \to U$ is continuous from the left, continuous at the points t_0, t_1, and $u(\cdot)$ can have but a finite number of points of discontinuity, and all these points must be of the first kind.

Let $(x_0^1, \ldots, x_0^n)$ and $(x_1^1, \ldots, x_1^n)$ be given points in $\mathbb{R}_n$. Denote by $\prod$ a line, which passes trough the point $(0, x_1^1, \ldots, x_1^n)$ in $\mathbb{R}_{n+1}$ and is parallel to the axis x^0. In the sequel, the elements of $\mathbb{R}_{n+1}$ are denoted by $(x^0, x^1, \ldots, x^n)$.

Define the map $f : \mathbb{R}_{n+1} \times U \to \mathbb{R}_{n+1}$ by the rule:

$$
f(x, u) = (f_0(x^1, \ldots, x^n, u_1, \ldots, u_k), \ldots, f_n(x^1, \ldots, x^n, u_1, \ldots, u_k)).
$$

Obviously, f does not depend on the values of the component x^0.

Definition 1. We say that the admissible control $u(t)$, $t_0 \leq t \leq t_1$, moves the point $x_0 = (0, x_0^1, \ldots, x_0^n)$ into some point of the line $\prod$, if the trajectory $\varphi(\cdot)$ corresponding to $u(\cdot)$, which is a solution of the Cauchy problem

$$
\dot{x}(t) = f(x(t), u(t)), \quad x(t_0) = x_0, \tag{1}
$$

is defined on the entire $[t_0, t_1]$ and $\varphi(t_1) \in \prod$.

Definition 2. The extremal problem

$$x^0(t_1) \to \min, \tag{2}$$

$$\dot{x}(t) = f(x(t), u(t)), \quad x(t_0) = x_0, \quad x(t_1) \in \prod, \quad u(\cdot) \in \Omega, \tag{3}$$

is said to be the main problem of the optimal control theory. The admissible control $u(\cdot)$, which is the solution of (2)–(3), and the trajectory $x(\cdot)$ corresponding to $u(\cdot)$, are said to be optimal.

For any $(\psi, x, u) \in \mathbb{R}_{n+1} \times \mathbb{R}_{n+1} \times U$ suppose as in [1]:

$$\mathcal{H}(\psi, x, u) = \sum_{i=0}^{n} \psi_i f_i(x, u),$$

$$\mathcal{M}(\psi, x) = \sup_{u \in U} \mathcal{H}(\psi, x, u).$$

Theorem 1 (Pontryagin's Maximum Primciple)**.** *Let $u(t)$, $t_0 \le t \le t_1$, be an admissible control, which moves x_0 into some point of the line $\prod$. For optimality of the admissible control $u(\cdot)$ and its corresponding trajectory $\varphi(\cdot)$ it is necessary the existence of a non-zero continuous vector function $\psi(t) = (\psi_0(t), \psi_1(t), \ldots, \psi_n(t))$ such that*

1) $\psi(\cdot)$ corresponds to $u(\cdot)$ and $x(\cdot)$ by the following rule:

$$\dot{\psi}_i(t) = -\sum_{\alpha=0}^{n} \frac{\partial f_\alpha(\varphi(t), u(t))}{\partial x^i} \psi_\alpha(t), \tag{4}$$

$$0 \le i \le n; \quad t_0 \le t \le t_1;$$

2) For each $t \in [t_0, t_1]$ there holds the maximum condition:

$$\mathcal{H}(\psi(t), x(t), u(t)) = \mathcal{M}(\psi(t), x(t)).$$

3) $\psi_0(t)$ is constant with respect to t and $\psi_0(t_1) \le 0$.

Auxiliary lemmas. By $\mathbb{R}^m$ we denote the set of m - dimensional vector columns

$$y = \begin{pmatrix} y^1 \\ \vdots \\ y^m \end{pmatrix} = (y^1, \ldots, y^m)^T.$$

If $x = (x^1, \cdots, x^m) \in \mathbb{R}_m$ and $y = (y^1, \cdots, y^m)^T \in \mathbb{R}^m$, then we use the notation xy for the inner product:

$$xy = \sum_{i=1}^{m} x^i y^i.$$

Lemma 1. *Let $Y \subset \mathbb{R}^m$ be a convex set, $y_0 \in \mathbb{R}^m$, $y_0 \ne 0 \in Y$ and the system of inequalities*

$$\begin{cases} 0 \le \psi y_0, \\ 0 \ge \psi y, \quad \forall y \in Y, \end{cases} \tag{5}$$

have only the null-solution with respect to $\psi \in \mathbb{R}_m$.

Then there exists a subset $\{y_1, \ldots, y_m\} \subset Y$ such that:

1) *$\{y_1, \ldots, y_m\}$ is a basis of $\mathbb{R}^m$.*

2) *There exist numbers $\alpha_i > 0$, $i = 1, \ldots, m$, such that $y_0 = \sum\limits_{i=1}^{m} \alpha_i y^i$.*

Proof. Devide the proof into 3 parts.

A) Let us show that the set of interior points of Y is not empty, i.e., $\mathrm{inter}\, Y \neq \varnothing$.

Y contains a basis of $\mathbb{R}^m$. In the contrary case, Y will turn out to be in the space having dimension less then m and there will exist $\psi \neq 0$ such that $\pm\psi y = 0$, $\forall y \in Y$. Obviously, either ψ or -ψ satisfies (5).

Y contains some basis of $\mathbb{R}^m$, say $\{z_1, \ldots, z_m\}$, and $0 \in Y$, so the simplex S spanned on $\{0, z_1, \ldots, z_m\}$ is such that $S \subset Y$ (as Y is convex) and S have interior points. Hence, $\mathrm{inter}\, Y \neq \varnothing$.

B) Prove the existence of $\lambda_0 > 0$ such that $\lambda_0 y_0 \in \mathrm{inter}\, Y$.

In the contrary case, $\exists \psi \neq 0$ which separates the convex, open and nonempty sets $\mathrm{inter}\, Y$ and $\{\lambda y_0\}_{\lambda > 0}$, and separates their closures as well, i.e.,

$$\psi(\lambda y_0) \geq \psi y, \quad \forall y \in Y, \quad \forall \lambda \geq 0. \tag{6}$$

From (6) it follows $0 \geq \psi y$, $\forall y \in Y$. Also from (6) it follows $\psi y_0 \geq 0$, since in the contrary case it would be $\psi(\lambda y_0) \to -\infty$ as $\lambda \to +\infty$, which contradicts to (6). Finally, (6) means that the system (5) has a non-vanishing solution, which contradicts to the condition of Lemma 1.

C) Construct a basis $\{y_1, \ldots, y_m\}$ satisfying the conclusions of Lemma 1.

As $\lambda_o y_0 \in \mathrm{inter}\, Y$, for some $\varepsilon > 0$ we have:

$$\|y - \lambda_o y_0\| < \varepsilon \;\; \Rightarrow y \in Y.$$

Let L be an $(m-1)$-dimensional subspace in $\mathbb{R}^m$ which does not contain $\lambda_o y_0$; S be a simplex in L with the set of vertices $\{z_1, \ldots, z_m\}$ such that $0 \in \mathrm{inter}\, S$ (i.e., $\{z_m - z_1, \ldots, z_2 - z_1\}$ is a basis of L). Obviously, we can take $\|z_i\| < \varepsilon$, $i = 1, \ldots, m$. Then $0 = \sum\limits_{i=1}^{m} \beta_i z_i$ for some numbers $\beta_i > 0$ with $\sum\limits_{i=1}^{m} \beta_i = 1$, i.e.,

$$\lambda_o y_0 = \sum\limits_{i=1}^{m} \beta_i(z_i + \lambda_o y_0) \equiv \sum\limits_{i=1}^{m} \beta_i y_i, \tag{7}$$

where $y_i = z_i + \lambda_o y_0$, $i = 1, \ldots, m$. As $\|y_i - \lambda_o y_0\| < \varepsilon$, we have $y_i \in Y$, $\forall i$. As $y_i - y_1 = z_i - z_1$, every element of L is a linear combination of the set $\{y_1, \ldots, y_m\}$, which together with $\lambda_o y_0 \notin L$ and (7) gives that $\{y_1, \ldots, y_m\}$ is a basis of $\mathbb{R}^m$. Finally, we can assume $\alpha_i = \frac{\beta_i}{\lambda_0}$. $\qquad\square$

Lemma 2. *Let $0 \in Y \subset \mathbb{R}^m$, $0 \neq y_0 \in \mathbb{R}^m$ and the system of inequalities*

$$\begin{cases} 0 \leq \psi y_0, \\ 0 \geq \psi y, \quad \forall y \in Y, \end{cases} \tag{8}$$

have only the null-solution with respect to $\psi \in \mathbb{R}_m$.

Then there exist: vectors $y_{ij} \in Y$ and numbers α_i and γ_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, m+1$, such that:

1) *for every index i we have: $\alpha_i > 0$, $\gamma_{ij} \in [0,1]$, and $\sum\limits_{j=1}^{m+1} \gamma_{ij} = 1$;*

2) *$\left\{ \sum\limits_{j=1}^{m+1} \gamma_{1j} y_{1j}, \ldots, \sum\limits_{j=1}^{m+1} \gamma_{mj} y_{mj} \right\}$ is a basis of $\mathbb{R}^m$;*

3) *$y_0 = \sum\limits_{i=1}^{m} \alpha_i \sum\limits_{j=1}^{m+1} \gamma_{ij} y_{ij}$.*

Proof. It is easy to see that (8) is equivalent to the following system:

$$\begin{cases} 0 \leq \psi y_0, \\ 0 \geq \psi y, \quad \forall y \in \operatorname{co} Y, \end{cases}$$

where $\operatorname{co} Y$, as usually, denotes the convex hull of Y. Now, using of Lemma 1 gives the existence of vectors $y_i \in Y$ and numbers α_i such that $y_0 = \sum\limits_{i=1}^{m} \alpha_i y_i$. By virtue of a well - known theorem (see [2], p. 9), every point of $\operatorname{co} Y$ is representable as a convex combination of no more than $(m+1)$ points of Y:

$$y_i = \sum_{j=1}^{m+1} \gamma_{ij} y_{ij},$$

$$y_{ij} \in Y, \quad \sum_{j=1}^{m+1} \gamma_{ij} = 1, \quad 0 \leq \gamma_{ij} \leq 1. \qquad \square$$

The proof of the theorem.

Proof. We prove Theorem 1 in several steps.

1. *Variations of optimal control.* Everywhere in the proof, $u(t)$, $t_0 \leq t \leq t_1$ denotes the optimal control, and $\varphi(\cdot)$ denotes the optimal trajectory. Consider the set:

$$\operatorname{var} = \{(s_i, \sigma_i, v_i)\}_{i=1}^{m} \subset (t_0, t_1] \times \mathbb{R}_+ \times U.$$

If $\{(s_i - \sigma_i, s_i)\}_{i=1}^{m}$ is a subset of pairwise disjoint intervals in $[t_0, t_1]$, then var is said to be the variation of the control $u(\cdot)$. Besides, the admissible

control $u_{\text{var}}(\cdot)$, which is defined by formula:

$$
u_{\text{var}}(t) = \begin{cases}
v_1, & t \in (s_1 - \sigma_1, s_1], \\
\vdots & \vdots \\
v_m, & t \in (s_m - \sigma_m, s_m], \\
u(t), & t \notin \bigcup_{i=1}^m (s_i - \sigma_i, s_i],
\end{cases}
$$

is said to be a varied control. When m=1, i.e., var $= \{(s, \sigma, v)\}$, then var is said to be a simple variation and for the sake of simplicity is identified with its element: var $= (s, \sigma, v)$.

2. *Varied trajectories.* Denote by $\Phi_{t,s}(x)$ the value of the solution of the Cauchy problem:

$$
\dot{x}(t) = f(x(t), u(t)), \quad x(s) = x, \tag{9}
$$

at the point t, $\forall s \in [t_0, t_1]$, $\forall x \in \mathbb{R}_{n+1}$. Note that under the term "solution" we mean non-extendable (maximal) solution. Obviously,

$$
\varphi(t) = \Phi_{t,t_0}(x_0), \quad t_0 \le t \le t_1.
$$

For every fixed parameter $v \in U$, denote by $\Phi_{t-s}^v(x)$ the value of the solution of the Cauchy problem:

$$
\dot{x}(t) = f(x(t), v), \quad x(s) = x, \tag{10}
$$

at the point t. $\Phi_{t-s}^v(x)$ depends on $t - s$ only, as (10) is autonomous.

For a given variation var $= \{(s_i, \sigma_i, v_i)\}_{i=1}^m$, denote by $\varphi_{\text{var}}(t)$ the value of the solution of the Cauchy problem:

$$
\dot{x}(t) = f(x(t), u_{\text{var}}(t)), \quad x(t_0) = x_0,
$$

at the point t.

Determine the form of the trajectory $\varphi_{\text{var}}(\cdot)$.

For each simple variation (s, σ, v), when σ is small enough, the following representation

$$
\varphi_{(s,\sigma,v)}(t) = \left(\Phi_{t,s} \circ \Phi_\sigma^v \circ \Phi_{s-\sigma,t_0} \right)(x_0), \quad s \le t \le t_1, \tag{11}
$$

is valid.

In particular, for every $s \in [t_0, t_1]$ there holds

$$
\varphi(t_1) = \Phi_{t_1,s}(\varphi(s)) = x_1
$$

(to show this, it is enough to take $\sigma = 0$), and by the theorem on continuous dependence on initial data for every $\varepsilon > 0$ there exists a neighborhood Δ of $\varphi(s)$ such that for $\forall x \in \Delta$ and $\forall t \in [t_0, t_1]$, $\Phi_{t,s}(x)$ is correctly defined and there holds

$$
\| \Phi_{t,s}(x) - \varphi(t) \| < \varepsilon \quad (\varphi(t) = \Phi_{t,s}(\varphi(s))).
$$

Now, let us take arbitrarily a variation var $= \{(s_i, \sigma_i, v_i)\}_{i=1}^m$, for which $\sum_{i=1}^m \sigma_i$ is small enough. Then

$$\varphi_{\text{var}}(t) = \left(\Phi_{t,\max\{s_i\}_i} \circ \prod_{s_i}^{\leftarrow} \left[\Phi_{\sigma_i}^{v_i} \circ \Phi_{s_i-\sigma_i, s_i-\Delta s_i} \right] \right)(x_0),$$

$$\max\{s_i\}_i \leq t \leq t_1. \tag{12}$$

In (12), $\Delta s_i = s_i - s$, where $s = \max\{s_k \mid s_k < s_i\}$ is the moment of time in the variation previous to s_i; $\Delta \min\{s_i\}_i = \min\{s_i\}_i - t_0$; $\prod_{s_i}^{\leftarrow}$ means that the multipliers are ordered chronologically with respect to s_i: $\min\{s_i\}_i$ is placed at the right, then s_i monotonically increases and finally $\max\{s_i\}_i$ is placed at the left.

Note, that the right-hand side of (12) is defined for σ_i of arbitrary sign, when $\sum_{i=1}^m \sigma_i$ is small enough and $\max\{s_i\}_i \leq t \leq t_1$.

3. *Influence of duration of variation on the position of the end of the trajectory.* Let var $= (s, \sigma, v)$ be a simple variation. For sufficiently small $\sigma \geq 0$ we can consider a curve $\sigma \mapsto \varphi_{\text{var}}(t)$, defined for every $t \in [t_0, t_1]$. For this curve, t presents itself as a parameter. When $t \geq s$, for this curve the representation (11) is valid, which allows us to calculate $\frac{\partial \varphi_{(s,\sigma,v)}(t)}{\partial \sigma}\big|_{\sigma=0}$. Using (11), $\varphi_{(s,\sigma,v)}(t)$ can be represented in the form:

$$\varphi_{(s,\sigma,v)}(t) = g(h(\sigma)),$$

where $g : \mathbb{R}_{n+1} \to \mathbb{R}_{n+1}$ acts by the rule $g(x) = \Phi_{t,s}(x)$, $h(\sigma)$ is defined in the small neighborhood of $\sigma = 0$ and $h(\sigma) = \Phi_\sigma^v(\Phi_{s-\sigma,t_0}(x_0)) \in \mathbb{R}_{n+1}$. Since $h(0) = \varphi(s)$, we have:

$$\frac{\partial \varphi_{(s,\sigma,v)}(t)}{\partial \sigma}\bigg|_{\sigma=0} = g'(h(0)) \cdot h'(0) = \frac{\partial \Phi_{t,s}(\varphi(s))}{\partial x} h'(0).$$

Now, represent $h(\sigma)$ in the form:

$$h(\sigma) = G(\xi(\sigma), \sigma),$$

where $G : \mathbb{R}_{n+1} \times \mathbb{R} \to \mathbb{R}_{n+1}$ acts by the rule $G(x, \sigma) = \Phi_\sigma^v(x)$, and $\xi(\sigma) = \Phi_{s-\sigma, t_0}(x_0)$.

$$h'(0) = \frac{d}{d\sigma}(G \circ (\xi, id_{\mathbb{R}}))(\sigma)|_{\sigma=0} = G'(\xi(0), 0) \cdot (\xi, id_{\mathbb{R}})'(0) =$$

$$= \begin{bmatrix} \frac{\partial G^0(\xi(0),0)}{\partial x^0} & \cdots & \frac{\partial G^0(\xi(0),0)}{\partial x^n} & \frac{\partial G^0(\xi(0),0)}{\partial \sigma} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial G^{n-1}(\xi(0),0)}{\partial x^0} & \cdots & \frac{\partial G^{n-1}(\xi(0),0)}{\partial x^n} & \frac{\partial G^{n-1}(\xi(0),0)}{\partial \sigma} \\ \frac{\partial G^n(\xi(0),0)}{\partial x^0} & \cdots & \frac{\partial G^n(\xi(0),0)}{\partial x^n} & \frac{\partial G^n(\xi(0),0)}{\partial \sigma} \end{bmatrix} \cdot \begin{bmatrix} \dot{\xi}^0(0) \\ \vdots \\ \dot{\xi}^n(0) \\ 1 \end{bmatrix} =$$

$$= \frac{\partial G(\xi(0), 0)}{\partial x} \xi'(0) + \frac{\partial G(\xi(0), \sigma)}{\partial \sigma}\bigg|_{\sigma=0}; \tag{13}$$

here $(id_\mathbb{R})\alpha = \alpha, \ \forall \alpha \in \mathbb{R}$. Since $G(x,0) = x$, we have that $\frac{\partial G(\xi(0),0)}{\partial x} = E$ is the identity matrix,

$$\xi'(0) = \frac{\partial \Phi_{s-\sigma,t_0}(x_0)}{\partial \sigma}\Big|_{\sigma=0} = -\frac{\partial \Phi_{s-\sigma,t_0}(x_0)}{\partial(s-\sigma)}\Big|_{\sigma=0} =$$

$$= -\Big(f(\Phi_{s-\sigma,t_0}(x_0), u(s-\sigma))|_{\sigma=o}\Big)^T = -\Big(f(\varphi(s), u(s))\Big)^T$$

(note that $\xi' = (\dot{\xi})^T$, by definition), and

$$\frac{\partial G(\xi(0),\sigma)}{\partial \sigma}\Big|_{\sigma=0} = \frac{\partial \Phi^v_\sigma(\varphi(s))}{\partial \sigma}\Big|_{\sigma=0} =$$

$$= \Big(f(\Phi^v_\sigma(\varphi(s)), v)\Big|_{\sigma=o}\Big)^T = \Big(f(\varphi(s)), v)\Big)^T.$$

Finally, for every pair $(s,v) \in [t_0, t_1] \times U$ we have

$$\frac{\partial \varphi_{(s,\sigma,v)}(t)}{\partial \sigma}\Big|_{\sigma=0} =$$

$$= \frac{\partial \Phi_{t,s}(\varphi(s))}{\partial x}\Big[f(\varphi(s), v) - f(\varphi(s), u(s))\Big]^T. \tag{14}$$

Define the set $Y \in \mathbb{R}^{n+1}$ as follows:

$$Y = \Big\{\frac{\partial \Phi_{t_1,s}(\varphi(s))}{\partial x}\Big[f(\varphi(s), v) - f(\varphi(s), u(s))\Big]^T \ \Big| \ \forall(s,v) \in [t_0, t_1] \times U\Big\}.$$

4. *The main fact and its proof.* Suppose:

$$y_0 = (-1, 0, \ldots, 0)^T \in \mathbb{R}^{n+1}.$$

Let us prove the existence of a non-zero $\psi \in \mathbb{R}_{n+1}$ such that

$$\psi \, y_0 \geq 0 \ \text{ and } \ \psi \, y \leq 0, \ \forall y \in Y.$$

y_0 has a special simple form, consequently the above condition is the same as

$$\psi_o \leq 0 \ \text{ and } \ \psi \, y \leq 0, \ \forall y \in Y. \tag{15}$$

(15) is the main fact in the proof of Maximum Principle.

Suppose the contrary, that only $\psi = 0$ satisfies (15), so we can use Lemma 2 with $m = n + 1$. Thus there exist: vectors

$$y_{ij} = \frac{\partial \Phi_{t_1,s_{ij}}(\varphi(s_{ij}))}{\partial x}\Big[f(\varphi(s_{ij}), v_{ij}) - f(\varphi(s_{ij}), u(s_{ij}))\Big]^T$$

and numbers $\alpha_i, \gamma_{ij}, i = 1, \ldots, n+1, j = 1, \ldots, n+2$, such that for every i: $\alpha_i > 0, \gamma_{ij} \in [0,1], \sum\limits_{j=1}^{n+2} \gamma_{ij} = 1$;

$$\Big\{\sum_{j=1}^{n+2} \gamma_{ij} y_{ij}\Big\}_{i=1}^{n+1} \tag{16}$$

is a basis of $\mathbb{R}^{n+1}$;

$$y_0 = (-1, 0, \ldots, 0)^T = \sum_{i=1}^{n+1} \alpha_i \sum_{j=1}^{n+2} \gamma_{ij} y_{ij}. \tag{17}$$

By virtue of continuity of the corresponding functions we can assume that all s_{ij} are pairwise different and belong to (t_0, t_1). Without loss of generality we can assume that for each $i \in \{1, \ldots, n+1\}$

$$s_{i1} < s_{i2} < \cdots < s_{i(n+2)} \tag{18}$$

(in the contrary case, we can number the indices over again).

If $\sigma_1, \ldots, \sigma_{n+1}$ are small enough and non-negative numbers, then the variation

$$\mathrm{var}_\sigma = \left\{ (s_{ij}, \sigma_i \gamma_{ij}, v_{ij}) \,\big|\, i = 1, \ldots, n+1, \ \ j = 1, \ldots, n+2 \right\}$$

depending on the vector $\sigma = (\sigma_1, \ldots, \sigma_{n+1})$ is correctly defined. When the norm of the vector σ is small enough, the corresponding to $u_{\mathrm{var}_\sigma}(\cdot)$ trajectory is defined on $[t_0, t_1]$ and a representation similar to (12) holds:

$$\varphi_{\mathrm{var}_\sigma}(t_1) =$$

$$= \left(\Phi_{t_1, \max\{s_{ij}\}_{i,j}} \circ \overset{\leftarrow}{\underset{s_{ij}}{\prod}} \left[\Phi_{\sigma_i \gamma_{ij}}^{v_{ij}} \circ \Phi_{s_{ij} - \sigma_i \gamma_{ij}, s_{ij} - \Delta s_{ij}} \right] \right)(x_0). \tag{19}$$

The right - hand side of (19) is defined for every $\sigma \in \mathbb{R}_{n+1}$, if its norm is small enough, regardless of the sign of σ.

Let V be a small enough neighborhood of 0 in $\mathbb{R}_{n+1}$. Define map $F : V \to \mathbb{R}_{n+1}$ as follows:

$$F(\sigma_1, \sigma_2 \ldots, \sigma_{n+1}) =$$

$$= \left(\Phi_{t_1, \max\{s_{ij}\}_{i,j}} \circ \overset{\leftarrow}{\underset{s_{ij}}{\prod}} \left[\Phi_{\sigma_i \gamma_{ij}}^{v_{ij}} \circ \Phi_{s_{ij} - \sigma_i \gamma_{ij}, s_{ij} - \Delta s_{ij}} \right] \right)(x_0). \tag{20}$$

By construction, if $\sigma = (\sigma_1, \sigma_2 \ldots, \sigma_{n+1}) \geq 0$ componentwise, then

$$F(0) = \varphi(t_1), \quad F(\sigma_1, \sigma_2 \ldots, \sigma_{n+1}) = \varphi_{\mathrm{var}_\sigma}(t_1). \tag{21}$$

F has partial derivatives, as it is a composition of continuously differentiable mappings.

Let us calculate $\left. \frac{\partial F(\sigma)}{\partial \sigma_k} \right|_{\sigma=0}$, $k = 1, \ldots, n+1$.

Let us fix $k \in \{1, \ldots, n+1\}$. By virtue of (20) and (18),

$$F(0, \ldots, 0, \sigma_k, 0, \ldots, 0) =$$

$$= \left(\Phi_{t_1, s_{k(n+2)}} \circ \overset{\leftarrow}{\underset{j}{\prod}} \left[\Phi_{\sigma_k \gamma_{kj}}^{v_{kj}} \circ \Phi_{s_{kj} - \sigma_k \gamma_{kj}, s_{kj} - \Delta s_{kj}} \right] \right)(x_0) =$$

$$= \left(\Phi_{t_1, s_{k(n+2)}} \circ G_{\sigma_k}^{n+2} \circ \cdots \circ G_{\sigma_k}^1 \right)(x_0), \tag{22}$$

where, since by virtue of (18) $s_{k(j-1)} = s_{kj} - \Delta s_{kj}$,

$$G^j_{\sigma_k} = \left(\Phi^{v_{kj}}_{\sigma_k \gamma_{kj}} \circ \Phi_{s_{kj} - \sigma_k \gamma_{kj}, s_{k(j-1)}} \right), \tag{23}$$

and $G^j_{\sigma_k}(x)$ is differentiable, at least in some small neighborhood of $(\sigma_k, x) = (0, (G^{j-1}_0 \circ \cdots \circ G^1_0)(x_0))$. Exactly the same method which was used in the proof of (13) gives:

$$\frac{\partial}{\partial \sigma_k} G^2_{\sigma_k}(G^1_{\sigma_k}(x_0)) \Big|_{\sigma_k=0} = \frac{\partial G^2_0(G^1_0(x_0))}{\partial x} \frac{\partial G^1_{\sigma_k}(x_0)}{\partial \sigma_k} \Big|_{\sigma_k=0} +$$

$$+ \frac{\partial G^2_{\sigma_k}(G^1_0(x_0))}{\partial \sigma_k} \Big|_{\sigma_k=0}. \tag{24}$$

Using estimations of the type (24) n-times and taking into consideration that $(G^j_0 \circ \cdots \circ G^1_0)(x_0) = \Phi_{s_{kj}, t_0}(x_0) = \varphi(s_{kj})$, we obtain:

$$\frac{\partial}{\partial \sigma_k}(G^{n+2}_{\sigma_k} \circ \cdots \circ G^1_{\sigma_k})(x_0) \Big|_{\sigma_k=0} =$$

$$= \sum_{j=1}^{n+2} \frac{\partial G^{n+2}_0((G^{n+1}_0 \circ \cdots \circ G^1_0)(x_0))}{\partial x} \cdots \frac{\partial G^{j+1}_0((G^j_0 \circ \cdots \circ G^1_0)(x_0))}{\partial x} \cdot$$

$$\cdot \frac{\partial}{\partial \sigma_k} G^j_{\sigma_k}((G^{j-1}_0 \circ \cdots \circ G^1_0)(x_0)) \Big|_{\sigma_k=0} =$$

$$= \sum_{j=1}^{n+2} \frac{\partial G^{n+2}_0(\varphi(s_{k(n+1)}))}{\partial x} \cdots \frac{\partial G^{j+1}_0(\varphi(s_{kj}))}{\partial x} \cdot \frac{\partial}{\partial \sigma_k} G^j_{\sigma_k}(\varphi(s_{k(j-1)})) =$$

$$= \sum_{j=1}^{n+2} \frac{\partial \Phi_{s_{k(n+2)}, s_{k(n+1)}}(x)}{\partial x} \Big|_{x=\varphi(s_{k(n+1)})} \cdots \frac{\partial \Phi_{s_{k(j+1)}, s_{kj}}(x)}{\partial x} \Big|_{x=\varphi(s_{kj})} \cdot$$

$$\cdot \frac{\partial}{\partial \sigma_k} G^j_{\sigma_k}(\varphi(s_{k(j-1)})) \Big|_{\sigma_k=0}; \tag{25}$$

in (25) we assume $s_{k0} = t_0$ and $G^{-1}_0 \circ G^1_0 = id$ – identity.

By virtue of (25), the formula (22) gives:

$$\frac{\partial F(0)}{\partial \sigma_k} \Big|_{\sigma_k=0} = \sum_{j=1}^{n+2} \frac{\partial \Phi_{t_1, s_{k(n+2)}}(x)}{\partial x} \Big|_{x=\varphi(s_{k(n+2)})} \cdot$$

$$\cdot \frac{\partial \Phi_{s_{k(n+2)}, s_{k(n+1)}}(x)}{\partial x} \Big|_{x=\varphi(s_{k(n+1)})} \cdots \frac{\partial \Phi_{s_{k(j+1)}, s_{kj}}(x)}{\partial x} \Big|_{x=\varphi(s_{kj})} \cdot$$

$$\cdot \frac{\partial}{\partial \sigma_k} \left(\Phi^{v_{kj}}_{\sigma_k \gamma_{kj}} \circ \Phi_{s_{kj} - \sigma_k \gamma_{kj}, s_{k(j-1)}} \right)(\varphi(s_{k(j-1)})) \Big|_{\sigma_k=0}. \tag{26}$$

In (26), we must take into account the following two facts. First, the derivatives of both sides of the following equality

$$\Phi_{t_1, s_{kj}}(x) = \left(\Phi_{t_1, s_{k(n+2)}} \circ \Phi_{s_{k(n+2)}, s_{k(n+1)}} \circ \cdots \circ \Phi_{s_{k(j+1)}, s_{kj}} \right)(x)$$

exist at $x = (\varphi(s_{kj}))$ and are equal. Secondly, reasoning as in point 3, it is easy to see that

$$\frac{\partial}{\partial \sigma_k} \left(\Phi_{\sigma_k \gamma_{kj}}^{v_{kj}} \circ \Phi_{s_{kj} - \sigma_k \gamma_{kj}, s_{k(j-1)}} \right)(\varphi(s_{k(j-1)})) \Big|_{\sigma_k = 0} =$$
$$= \gamma_{kj} \left[f(\varphi(s_{kj}), v_{kj}) - f(\varphi(s_{kj}), u(s_{kj})) \right]^T.$$

Indeed, in the non-trivial case where $\gamma_{kj} \neq 0$, we have

$$\frac{\partial}{\partial \sigma_k} \Phi_{\sigma_k \gamma_{kj}}^{v_{kj}}(\varphi(s_{kj})) \Big|_{\sigma_k = 0} =$$
$$= \gamma_{kj} \frac{\partial}{\partial \sigma_k \gamma_{kj}} \Phi_{\sigma_k \gamma_{kj}}^{v_{kj}}(\varphi(s_{kj})) \Big|_{\sigma_k = 0} = \gamma_{kj} \left[f(\varphi(s_{kj}), v_{kj}) \right]^T.$$

Besides, $\Phi_0^{v_{kj}}(x) \equiv x$, therefore for $\gamma_{kj} \neq 0$ we have:

$$\frac{\partial}{\partial \sigma_k} \varphi(s_{kj} - \sigma_k \gamma_{kj}) \Big|_{\sigma_k = 0} =$$
$$= -\gamma_{kj} \frac{\partial}{\partial(s_{kj} - \sigma_k \gamma_{kj})} \varphi(s_{kj} - \sigma_k \gamma_{kj}) \Big|_{\sigma_k = 0} =$$
$$= -\gamma_{kj} \left[f(f(\varphi(s_{kj}), u(s_{kj})) \right]^T;$$

obviously, the same result remains true when $\gamma_{kj} = 0$. Finally, we obtain:

$$\frac{\partial F(0)}{\partial \sigma_k} = \sum_{j=1}^{n+2} \gamma_{kj} y_{kj}.$$

Since $\left\{ \sum_{j=1}^{n+2} \gamma_{ij} y_{ij} \right\}_{i=1}^{n+1}$ is a basis, $F'(0)$ is an invertible matrix. Thus, in some neighborhood W of the point $F(0) = \varphi(t_1)$ there exists the inverse to the mapping F.

For sufficiently small $\tau \in \mathbb{R}$ we have: $\varphi(t_1) - \tau(1, 0, \ldots, 0) \in W$. For τ of such type denote: $\sigma(\tau) = F^{-1}(\varphi(t_1) - \tau(1, 0, \ldots, 0))$. Obviously, $\sigma(\tau) = (\sigma_1(\tau), \ldots, \sigma_{n+1}(\tau))$ is continuously differentiable, $\sigma(0) = 0$ and

$$\varphi(t_1) - \tau(1, 0, \ldots, 0) = F(\sigma_1(\tau), \ldots, \sigma_{n+1}(\tau)). \tag{27}$$

Differentiating both sides of (27) with respect to τ, we obtain:

$$y_0 = \sum_{i=1}^{n+1} \dot{\sigma}_i(0) \sum_{j=1}^{n+2} \gamma_{ij} y_{ij}. \tag{28}$$

The expansion of the vector y_0 with respect to the basis (16) is unique, therefore (17) and (28) give

$$\dot{\sigma}_i(0) = \alpha_i > 0, \quad \forall i.$$

Thus, for sufficiently small and positive numbers τ we have:

$$\sigma_i(\tau) > 0, \quad \forall i. \tag{29}$$

Now, for sufficiently small and positive numbers τ, (21) and (29) give

$$\varphi(t_1) - \tau(1, 0, \ldots, 0) = \varphi_{\mathrm{var}_{\sigma(\tau)}}(t_1). \tag{30}$$

(30) means that $u(\cdot)$ is not optimal.

5. *Necessary conditions of optimality.* As we have proved in 4, there exists $\widehat{\psi} \neq 0$ such that

$$\widehat{\psi}_0 \leq 0, \quad \widehat{\psi}\, y \leq 0, \quad \forall y \in Y. \tag{31}$$

There exists a continuous and continuously differentiable at the points of continuity of $u(\cdot)$ function $\psi(t)$, $t_0 \leq t \leq t_1$, such that the first conclusion of the maximum principle holds:

$$\begin{cases} \dot{\psi}(t) = -\psi(t)\frac{\partial f(\varphi(t), u(t))}{\partial x}, \\ \psi(t_1) = \widehat{\psi}. \end{cases} \tag{32}$$

From (32) it is easily seen that $\psi_0(t) \equiv const$, i.e., $\psi_0(t) \equiv \widehat{\psi}_0$ and the third conclusion holds, too.

For arbitrarily given $(s, v) \in [t_0, t_1] \times U$ denote:

$$\Psi_t = \left.\frac{\partial \Phi_{t,s}(x)}{\partial x}\right|_{x = \varphi(s)}.$$

It is well - known from the standard course of ordinary differential equations that Ψ_t satisfies the variational equation:

$$\dot{\Psi}_t = \frac{\partial f(\varphi(t), u(t))}{\partial x}\Psi_t, \tag{33}$$

and Ψ_s is the unit matrix. Obviously, from (32) and (33) it follows:

$$\frac{d}{dt}\left(\psi(t)\Psi_t\Big[f(\varphi(s), v) - f(\varphi(s), u(s))\Big]^T\right) =$$
$$= \left(\dot{\psi}(t)\Psi_t\Big[f(\varphi(s), v) - f(\varphi(s), u(s))\Big]^T\right) +$$
$$+ \left(\psi(t)\dot{\Psi}_t\Big[f(\varphi(s), v) - f(\varphi(s), u(s))\Big]^T\right) = 0, \tag{34}$$

i.e., the function $t \mapsto \psi(t)\Psi_t\Big[f(\varphi(s), v) - f(\varphi(s), u(s))\Big]^T$ is constant. In particular, when $t = s$ and $t = t_1$, we have:

$$\widehat{\psi}\,\frac{\partial \Phi_{t_1, s}(\varphi(s))}{\partial x}\Big[f(\varphi(s), v) - f(\varphi(s), u(s))\Big]^T =$$

$$= \psi(s) \Big[f(\varphi(s), v) - f(\varphi(s), u(s)) \Big]^T.$$

The last identity, (31) and (34) together give that

$$\psi(s) \Big[f(\varphi(s), v) - f(\varphi(s), u(s)) \Big]^T \le 0, \quad \forall (s, v) \in [t_0, t_1] \times U,$$

i.e., the second conclusion of the maximum principle holds, too.   □

Features of the applied method. As we have seen above, using of evolution systems considerably simplifies the proof of the maximum principle as compared with other proofs. Applying this method, the same success can be achieved in other standard problems of the optimal control theory, but now we consider some non-standard problems.

1. *The optimal problem with variable control area.* Consider the case where the control area depends on time by the following rule:

$$\text{if} \quad s_1 \le s_2 \quad \text{then} \quad U(s_2) \subset U(s_1). \tag{35}$$

To deal with this case, first modify some definitions. Let $U = U(t_0)$, the functions

$$f_i(x^1, \dots, x^n, u_1, \dots, u_k), \quad \frac{\partial f_i}{\partial x^j}$$

continuously map $\mathbb{R}_n \times U$ in $\mathbb{R}$, $i = 0, 1, \dots, n$, $j = 1, \dots, n$; $[t_0, t_1]$ be a fixed non-trivial segment (i.e., $t_0 < t_1$); Ω_0 be the set of admissible controls, consisting of the functions $u(t) = (u_1(t), \dots, u_k(t))$ such that $u(t) \in U(t) \subset \mathbb{R}_k$, $\forall t \in [t_0, t_1]$, the function $t \mapsto U(t)$ has the property (35), $u(\cdot)$ is continuous from the left, continuous at the points t_0, t_1, and $u(\cdot)$ can have but a finite number of points of discontinuity, and all these points must be of the first kind. For any $(\psi, x, u) \in \mathbb{R}_{n+1} \times \mathbb{R}_{n+1} \times U$ put

$$\mathcal{H}(\psi, x, u) = \sum_{i=0}^n \psi_i f_i(x, u),$$

and for any $(\psi, x) \in \mathbb{R}_{n+1} \times \mathbb{R}_{n+1}$, $t \in [t_o.t_1]$:

$$\mathcal{M}_t(\psi, x) = \sup_{u \in U(t)} \mathcal{H}(\psi, x, u).$$

Pontryagin's Maximum Principle takes the following natural form.

Theorem 2. *Let $u(t)$, $t_0 \le t \le t_1$, be an admissible control, which moves x_0 into some point of the line $\prod$. For optimality of the control $u(\cdot)$ and its corresponding trajectory $\varphi(\cdot)$ it is necessary the existence of a non-zero continuous vector function $\psi(t) = (\psi_0(t), \psi_1(t), \dots, \psi_n(t))$ such that*

1) $\psi(\cdot)$ *corresponds to $u(\cdot)$ and $x(\cdot)$ by the following rule:*

$$\dot{\psi}_i(t) = -\sum_{\alpha=0}^n \frac{\partial f_\alpha(\varphi(t), u(t))}{\partial x^i} \psi_\alpha(t),$$
$$0 \le i \le n; \quad t_0 \le t \le t_1;$$

2) *For each $t \in [t_0, t_1]$ there holds the maximum condition:*
$$\mathcal{H}(\psi(t), x(t), u(t)) = \mathcal{M}_t(\psi(t), x(t)).$$

3) $\psi_0(t)$ *is constant with respect to t and $\psi_0(t_1) \leq 0$.*

In the proof, we have to change the definition of the variation of the control $u(\cdot)$. Consider the set var $= \{(s_i, \sigma_i, v_i)\}_{i=1}^m$, where for every i: $(s_i, \sigma_i) \in (t_0, t_1] \times \mathbb{R}_+$, $v_i \in U(s_i)$. If $\{(s_i - \sigma_i, s_i)\}_{i=1}^m$ is a subset of pairwise disjoint intervals in $[t_0, t_1]$, then var is said to be the variation of the control $u(\cdot)$. Varied controls $u_{\mathrm{var}}(\cdot)$ can be defined usually and after this the proof of Theorem 2 can be repeated without other changes.

If instead of (35) there holds:
$$\text{if} \quad s_1 \leq s_2, \quad \text{then} \quad U(s_1) \subset U(s_2), \tag{36}$$
then we have to take $U = U(t_1)$ and change the definition of varied controls, too. Consider the set: var $= \{(s_i, \sigma_i, v_i)\}_{i=1}^m$, where for every i: $(s_i, \sigma_i) \in [t_0, t_1) \times \mathbb{R}_+$, $v_i \in U(s_i)$. If $\{(s_i, s_i + \sigma_i)\}_{i=1}^m$ is a subset of pairwise disjoint intervals in $[t_0, t_1]$, then var is said to be the variation of the control $u(\cdot)$.

An admissible control $u_{\mathrm{var}}(\cdot)$ defined by the formula:
$$u_{\mathrm{var}}(t) = \begin{cases} v_1, & t \in [s_1, s_1 + \sigma_1), \\ \vdots & \vdots \\ v_m, & t \in [s_m, s_m + \sigma_m), \\ u(t), & t \notin \bigcup_{i=1}^m [s_i, s_i + \sigma_i), \end{cases}$$

is said to be a varied control. Now it is easy to see that Theorem 2 is valid, too.

Of course, choosing definitions of varied controls properly, we can cover more complicated cases where $U(t)$ depends on t. Unfortunately, we can not develop this theme now, since it is far from purposes of the present paper. Note only that even simple cases considered above have important applications in the theory of economic growth (see [3] and [4]).

2. *An optimal problem in the class of controls having uniformly limited number of points of discontinuity.* The optimization of many real economic controllable processes takes place in the class of admissible controls, which consists of piecewise continuous functions, whose number of points of discontinuity is less than some preliminary given natural number. Such class of admissible controls has "good mathematical properties". In applications, such class appears naturally and frequently, as in many real processes discontinuity of an admissible control means to spend some fixed dose of an exhaustible resource. Thus, in such problems, setting of the upper limit for the number of discontinuities is a natural thing.

For every $l \in \{1, 2, \ldots, +\infty\}$ denote by Ω_l the set of admissible controls consisting of the functions $u(t) = (u_1(t), \ldots, u_k(t))$ such that $u(\cdot) : [t_0, t_1] \to U$ is continuous from the left, continuous at the points t_0, t_1, and $u(\cdot)$ can

have points of discontinuity, but the number of points of discontinuity is less or equal to l and all these points must be of the first kind.

Consider the optimal problem:

$$x^0(t_1) \to \min, \tag{37}$$

$$\dot{x}(t) = f(x(t), u(t)), \quad x(t_0) = x_0, \quad x(t_1) \in \prod, \quad u(\cdot) \in \Omega_l. \tag{38}$$

It is easy to see that the above - given proof of Theorem 1 is valid also for the following modification of Pontryagin's Maximum Principle, since in our method we use only varied controls having no more than $2(n+1)(n+2)$ "new" points of discontinuity as compared with the optimal control.

Theorem 3. *Let $u(t)$, $t_0 \le t \le t_1$, be an admissible control, which moves x_0 into some point of the line $\prod$, and number of points of discontinuity is less than or equal to $l - 2(n+1)(n+2)$. In the problem (37)–(38), for optimality of the control $u(\cdot)$ and its corresponding trajectory $\varphi(\cdot)$ it is necessary the existence of a non-zero continuous vector function $\psi(t) = (\psi_0(t), \psi_1(t), \dots, \psi_n(t))$ such that*

1). *$\psi(\cdot)$ corresponds to $u(\cdot)$ and $x(\cdot)$ by the following rule:*

$$\dot{\psi}_i(t) = -\sum_{\alpha=0}^{n} \frac{\partial f_\alpha(\varphi(t), u(t))}{\partial x^i} \psi_\alpha(t),$$

$$0 \le i \le n; \quad t_0 \le t \le t_1;$$

2). *For each $t \in [t_0, t_1]$ there holds the maximum condition:*

$$\mathcal{H}(\psi(t), x(t), u(t)) = \mathcal{M}(\psi(t), x(t)).$$

3). *$\psi_0(t)$ is constant with respect to t and $\psi_0(t_1) \le 0$.*

As far as we know, other methods used to prove Pontryagin's Maximum Principle in its standard formalization are not applicable to Theorem 3.

REFERENCES

1. L. PONTRYAGIN, V. BOLTIANSKIĬ, R. GAMKRELIDZE, AND E. F. MISHCHENKO, The Mathematical theory of optimal processes. *Nauka, Moscow,* 1976.
2. B. N. PSHENICHNIĬ, Convex analysis and extremal problems. Russian *Nauka, Moscow,* 1980.
3. M. INTRILIGATOR, Mathematical optimization and economic theory. *Prentice-Hall, Inc., Englewood Cliffs, N.J.,* 1971.
4. A. CHIANG, Elements of dynamic optimization. *McGaw-Hill, Inc.,* 1992.

Author's address:

Department of Applied Mathematics and Computer Sciences
I. Javakhishvili Tbilisi State University
2, University St., Tbilisi 0143
Georgia