Computational Mathematics and Modeling, Vol. 22, No. 2, April, 2011

MAXIMUM PRINCIPLE FOR OPTIMAL CONTROL PROBLEMS WITH INTERMEDIATE CONSTRAINTS

A. V. Dmitruk and A. M. Kaganovich

We consider optimal control problems with constraints at intermediate points of the trajectory. A natural technique (propagation of phase and control variables) is applied to reduce these problems to a standard optimal control problem of Pontryagin type with equality and inequality constraints at the trajectory endpoints. In this way we derive necessary optimality conditions that generalize the Pontryagin classical maximum principle. The same technique is applied to so-called variable structure problems and to some hybrid problems. The new optimality conditions are compared with the results of other authors and five examples illustrating their application are presented.

1. Statement of the Problem

Let $t_0 < t_1 < \ldots < t_\nu$ be real numbers. For every n-dimensional continuous function $x(t)$ on the interval $[t_0, t_\nu]$ define the vector

$$p = \big((t_0, x(t_0)),\, (t_1, x(t_1)),\, \ldots,\, (t_\nu, x(t_\nu))\big).$$

On the interval $[t_0, t_\nu]$ consider the optimal control problem

$$\textbf{Problem } A: \qquad \begin{cases} \dot{x} = f(t, x, u), \quad u \in U, \\ \eta_j(p) = 0, \quad j = 1, \ldots, q, \\ \varphi_i(p) \le 0, \quad i = 1, \ldots, m, \\ J = \varphi_0(p) \to \min, \end{cases}$$

where $t_0, t_1, \ldots, t_\nu$ are not fixed, $x \in \mathbb{R}^n$, $u \in \mathbb{R}^r$, the function $x(\cdot)$ is absolutely continuous, the function $u(\cdot)$ is measurable bounded.

Thus, Problem A contains equality and inequality constraints that depend on the values of the phase variable not only at the endpoints of $[t_0, t_\nu]$, but also at the intermediate points $t_1, t_2, \ldots, t_{\nu-1}$. If $\nu = 1$, i.e, there are no intermediate points, Problem A is the well-known classical problem of Pontryagin type. In our formulation this problem has been considered in [3, 4, 6, 22]. Particular cases of the problem, without endpoint inequalities and with the endpoint equalities specified separately for the left and right endpoints, have been considered in [1, 5] and in many other books and articles.

The objective of the present study is to generalize the Pontryagin maximum principle to this class of problems. We will show that Problem A can be reduced to a standard optimal control problem without intermediate constraints.

Let the following assumptions hold.

(A1) the function f is defined and continuous on the open set $\mathcal{Q} \subset \mathbb{R}^{n+r+1}$, the partial derivatives f_t, f_x, exist on this set and are jointly continuous in all the arguments;

Translated from Nelineinaya Dinamika i Upravlenie, No. 6, pp. 101–136, 2008.

 1046–283X/11/2202–0180

(A2) the functions $\varphi_i(p)$ and $\eta_j(p)$ are defined on the open set $\mathcal{P} \subset \mathbb{R}^{(\nu+1)(n+1)}$ and have continuous derivatives on this set;

(A3) U is an arbitrary set in $\mathbb{R}^r$.

Definition 1. *The triple $w = (x(t), u(t), p)$ is called an admissible process if it satisfies all the constraints and there exists a compactum $\Omega \subset \mathcal{Q}$ such that $(t, x(t), u(t)) \in \Omega$ almost everywhere on $\Delta = [t_0, t_\nu]$.*

Definition 2. *The admissible process $w^0 = (x^0(t), u^0(t), p^0)$ is called optimal (attaining the global minimum) in Problem A if $J(w^0) \le J(w)$ for every admissible process w.*

We assume that the Pontryagin maximum principle (MP) is known for the following canonical autonomous optimal control problem of Pontryagin type on a fixed time interval $[0, T]$:

$$\textbf{Problem } \boldsymbol{K}: \qquad \begin{cases} \dot{x} = f(x, u), \\ u \in U, \quad (x, u) \in \mathcal{Q}, \\ \eta_j(p) = 0, \quad j = 1, \dots, q, \\ \varphi_i(p) \le 0, \quad i = 1, \dots, m, \\ J = \varphi_0(p) \to \min. \end{cases}$$

Here $p = (x(0),\, x(T)) \in \mathbb{R}^{2n}$ is the vector of endpoint values of the trajectory $x(t)$ and $\mathcal{Q}$ is an open set in $\mathbb{R}^{n+r}$. The condition $(x, u) \in \mathcal{Q}$ should be treated not as a constraint, but as a definition of the open region in the (x, u) space where the problem is considered (see [3, 4, 7, 22]).

Note that the endpoint equalities and inequalities are not separated by x_0 and x_T; they depend on these points in an arbitrary smooth manner. The importance of these endpoint constraints was understood already in classical variational calculus (see, e.g., [2]), but this understanding has been lost in a certain sense during the transition to optimal control. In optimal control problems, non-separated endpoint constraints $\varphi(x_0, x_T) \le 0$ and $\eta(x_0, x_T) = 0$ have been introduced and systematically studied by Dubovitskii and Milyutin [3, 4]. The advantages of this approach will be elucidated below.

Recall the following two concepts formalizing the concept of optimality.

Definition 3. *Problem K attains a strong minimum on the process $w^0 = (x^0(t),\, u^0(t))$ if there exists $\varepsilon > 0$ such that for every admissible process $w = (x(t), u(t))$ satisfying the inequality $\|x - x^0\|_C < \varepsilon$ we have $J(w^0) \le J(w)$.*

Definition 4. *Problem K attains a Pontryagin minimum on the process $w^0 = (x^0(t), u^0(t))$ if for every constant N there exists $\varepsilon > 0$ such that for every admissible process $w = (x(t), u(t))$ satisfying the inequalities*

$$\|x - x^0\|_C < \varepsilon, \qquad \|u - u^0\|_1 < \varepsilon, \qquad \|u - u^0\|_\infty \le N$$

we have $J(w^0) \le J(w)$.

The latter concept has been proposed by Dubovitskii and Milyutin (in their investigation of the Pontryagin MP and its generalization to problems with phase and mixed constraints); it arose as a natural formalization of the concept of minimum in the class of uniformly small and spike control variations. For more details see [7, Sec. 2; 22, Sec. 1.8].

For Problem $A,$ where the time interval and the intermediate points t_k are not fixed, the definition of strong and Pontryagin minimum is modified in the following natural manner.

Definition 3′. *Problem A attains a strong minimum on the process $w^0 = (x^0(t), u^0(t), p^0),$ defined on the interval $\Delta^0 = [t_0^0, t_\nu^0],$ if there exists $\varepsilon > 0$ such that for every admissible process $w = (x(t), u(t), p)$ defined on the interval $\Delta = [t_0, t_\nu]$ and satisfying the inequalities*

$$\|x - x^0\|_C < \varepsilon, \qquad |t_k - t_k^0| < \varepsilon \qquad \textit{for all } k = 0, \ldots, \nu,$$

we have $J(w^0) \le J(w)$.

Definition 4′. *Problem A attains a Pontryagin minimum on the process $w^0 = (x^0(t), u^0(t), p^0)$ if for every constant N there exists $\varepsilon > 0$ such that for every admissible process $w = (x(t), u(t), p)$ satisfying the inequalities*

$$\|x - x^0\|_C < \varepsilon, \quad |t_k - t_k^0| < \varepsilon \qquad \textit{for all } k = 0, \ldots, \nu,$$

$$\|u - u^0\|_1 < \varepsilon, \qquad \|u - u^0\|_\infty \le N$$

we have $J(w^0) \le J(w)$.

In both definitions the process w is defined on the interval Δ which, in general, is different from the interval Δ^0. Therefore all the norms should be considered on the joint definition interval of the processes, i.e., on the intersection $\Delta \cap \Delta^0$.

The concepts of strong and Pontryagin minimum are sometimes conveniently defined in an equivalent manner in terms of sequences.

Problem A attains a strong minimum on the process $w^0 = (x^0(t), u^0(t),\, p^0)$ if for every sequence of admissible processes $w^i = (x^i(t), u^i(t), p^i),\ \ i = 1, 2, \ldots,$ such that as $i \to \infty$

$$\|x^i - x^0\|_C \to 0, \qquad |t_k^i - t_k^0| \to 0 \qquad \text{for all } k = 0, \ldots, \nu, \tag{1}$$

we have $J(w^0) \le J(w^i)$ for all sufficiently large i.

Problem A attains a Pontryagin minimum on the process $w^0 = (x^0(t), u^0(t),\, p^0)$ if for every sequence of admissible processes $w^i = (x^i(t), u^i(t), p^i),\ \ i = 1, 2, \ldots,$ such that as $i \to \infty$

$$\begin{gathered} \|x^i - x^0\|_C \to 0, \qquad |t_k^i - t_k^0| \to 0 \qquad \text{for al } k = 0, \ldots, \nu, \\ \|u^i - u^0\|_1 \to 0, \qquad \|u^i - u^0\|_\infty \le \mathcal{O}(1), \end{gathered} \tag{2}$$

we have $J(w^0) \le J(w^i)$ for all sufficiently large i.

From these definitions it obviously follows that if Problem A attains a strong minimum on the process w^0, then it also attains a Pontryagin minimum on this process. Thus, these minima and ordinary optimality (global minimum) are related in the following way:

optimality $\Rightarrow$ strong minimum $\Rightarrow$ Pontryagin minimum.

Both implications, in general, are not reversible (counterexamples exist).

The following definition is useful in the context of Pontryagin minimum.

Definition 5. *The sequence of admissible processes* $w^i = (x^i(t), u^i(t), p^i)$, $i = 1, 2, \ldots,$ *is Pontryagin convergent to the process* $w^0 = (x^0(t), u^0(t), p^0)$ *if conditions (2) hold; this sequence is "strongly" convergent if conditions (1) hold. (Quotation marks around "strongly" have been added because neither strong nor any other convergence by control is required.)*

Let us now state the Pontryagin MP for Problem K, given that Assumptions A1–A3 hold.

Theorem 1. *Assume that Problem* K *attains a Pontryagin minimum on the process* $w^0 = (x^0(t), u^0(t))$. *Then there exists a tuple* $\lambda = (\alpha, \beta, \sigma, \psi(\cdot))$, *where* $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_m) \in \mathbb{R}^{m+1}$, $\beta = (\beta_1, \beta_2, \ldots, \beta_q) \in \mathbb{R}^q$, $\sigma \in \mathbb{R}^1$, $\psi(\cdot)$, *is an* n*-dimensional Lipschitz function on* $[0, T]$, *for which we can construct*

– *the Pontryagin function*

$$H(\psi, x, u) = \langle \psi, f(x, u) \rangle,$$

– *the endpoint Lagrange function*

$$l(p) = \sum_{i=0}^{m} \alpha_i \varphi_i(p) + \sum_{j=1}^{q} \beta_j \eta_j(p),$$

and the following conditions are satisfied:

(a) *nontriviality condition:* $(\alpha, \beta) \neq (0, 0)$, *i.e., the tuple is not identically zero;*

(b) *nonnegativity conditions:* $\alpha_i \geq 0$, $i = 0, \ldots, m$;

(c) *complementary slackness conditions:* $\alpha_i \varphi_i(p^0) = 0$, $i = 1, \ldots, m$;

(d) *conjugate equation:* $\dot{\psi}(t) = -H_x^0 = -\psi(t) f_x(x^0(t), u^0(t))$ *a.e. on* $[0, T]$;

(e) *transversality conditions:* $\psi(0) = l_{x(0)}(p^0)$, $\psi(T) = -l_{x(T)}(p^0)$;

(f) *constancy of the function H: for almost all* $t \in [0, T]$

$$H(\psi(t), x^0(t), u^0(t)) = \sigma;$$

(g) *maximum condition: for all* $t \in [0, T]$

$$\max_{u \in C(t)} H(\psi(t), x^0(t), u) = \sigma,$$

where $C(t) = \big\{ u \in U \,\big|\, (x^0(t), u) \in \mathcal{Q} \big\}$.

Proof is given, e.g., in [1, 6, 22].

Remark 1. Alongside the Pontryagin function, we can also consider the Hamilton function or the Hamiltonian

$$\mathcal{H}(\psi, x) = \max_{u \in C(t)} H(\psi, x, u).$$

These functions are identical on an optimal process, although in general they are two different functions — they even depend on different sets of arguments (the Hamilton function does not contain any control!). In mechanics the Hamiltonian $\mathcal{H}$ is the total energy of the system, and therefore for optimal control problems condition (f) is sometimes called the law of energy conservation. Note that condition (f) containing a measurable function $u^0(t)$ holds for almost all t, whereas condition (g), where all functions are continuous, holds for all $t \in [0, T]$.

2. Reduction of Problem A to Problem K

We will now pass from Problem A with intermediate constraints to canonical Problem K and establish the correspondence between admissible and optimal processes in these problems. The idea behind this transition is quite natural: we have to "propagate" both the phase and the control variables to the intervals $\Delta_k = [t_{k-1}, t_k]$, $k = 1, \ldots, \nu$, created by the intermediate points partitioning the full interval $[t_0, t_\nu]$ and then reduce all the new variables to one joint time interval, for instance $[0, 1]$.

Let $(x(t), u(t), p)$ be an arbitrary admissible process in Problem A. Define the new time $\tau \in [0, 1]$ and the functions $\rho_k \colon [0, 1] \to \Delta_k$, $k = 1, \ldots, \nu$, from the equations

$$\frac{d\rho_k}{d\tau} = z_k(\tau), \quad \rho_k(0) = t_{k-1},$$

where $z_k(\tau) > 0$ are arbitrary measurable bounded functions on $[0, 1]$ such that $\rho_k(1) = t_k$, i.e., $\int_0^1 z_k(\tau)\, d\tau = |\Delta_k|$. The functions ρ_k play the role of the time t on the interval Δ_k. Define the functions $y_k(\tau) = x(\rho_k(\tau))$ and $v_k(\tau) = u(\rho_k(\tau))$, $k = 1, \ldots, \nu$, $\tau \in [0, 1]$. These functions obviously satisfy the following relationships:

$$\begin{cases} \dfrac{dy_k}{d\tau} = z_k f(\rho_k, y_k, v_k), \\[2ex] \dfrac{d\rho_k}{d\tau} = z_k, \quad k = 1, \ldots, \nu, \end{cases} \tag{3}$$

$$v_1, v_2, \ldots, v_\nu \in U, \tag{4}$$

$$\begin{cases} y_2(0) - y_1(1) = 0, \\ y_3(0) - y_2(1) = 0, \\ \quad \cdots\cdots\cdots\cdots\cdots\cdots \\ y_\nu(0) - y_{\nu-1}(1) = 0, \end{cases} \tag{5}$$

$$\begin{cases} \rho_2(0) - \rho_1(1) = 0, \\ \rho_3(0) - \rho_2(1) = 0, \\ \quad \cdots\cdots\cdots\cdots\cdots\cdots \\ \rho_\nu(0) - \rho_{\nu-1}(1) = 0, \end{cases} \tag{6}$$

$$\begin{cases} \eta_j(\hat{p}) = 0, \quad j = 1, \ldots, q, \\ \varphi_i(\hat{p}) \le 0, \quad i = 1, \ldots, m, \end{cases} \tag{7}$$

where for simplicity we denote

$$\hat{p} = \hat{p}(\rho, y) = \big((\rho_1(0), y_1(0)),\ (\rho_2(0), y_2(0)),\ \ldots,\ (\rho_\nu(0), y_\nu(0)),\ (\rho_\nu(1), y_\nu(1))\big).$$

This vector is clearly identical with the original vector $p(t,x)$.

The full vector of endpoint values in this case has the form

$$\tilde{p} = \big((\rho_1(0), y_1(0)),\ (\rho_2(0), y_2(0)),\ \ldots,\ (\rho_\nu(0), y_\nu(0)),$$

$$(\rho_1(1), y_1(1)),\ (\rho_2(1), y_2(1)),\ \ldots,\ (\rho_\nu(1), y_\nu(1))\big).$$

For brevity we introduce the notation $\rho = (\rho_1, \rho_2, \ldots, \rho_\nu)$, $y = (y_1, y_2, \ldots, y_\nu)$, $v = (v_1, v_2, \ldots, v_\nu)$, $z = (z_1, z_2, \ldots, z_\nu)$.

On the set of all possible processes $\tilde{w} = (\rho(\tau), y(\tau), v(\tau), z(\tau))$, satisfying constraints (3)–(7) we minimize the functional

$$\tilde{J}(\tilde{w}) = \varphi_0(\hat{p}) \to \min.$$

This optimal control problem is called Problem $\tilde{A}$. Here the phase variables are ρ_k and y_k, the controls are v_k and z_k, $k = 1, \ldots, \nu$, and the time interval $[0,1]$ is fixed. The open set $\widetilde{\mathcal{Q}}$ in this case consists of all tuples (ρ_k, y_k, v_k, z_k), where $(\rho_k, y_k, v_k) \in \mathcal{Q}$, $z_k > 0$. The open set $\widetilde{\mathcal{P}}$ consists of all the vectors $\tilde{p}$ for which the “truncated” vector $\hat{p} \in \mathcal{P}$.

The following two correspondences F and G can be established between the admissible processes of Problems A and $\tilde{A}$. We have shown that to each admissible process $w = (x(t), u(t), p)$ of Problem A we can associate an admissible process $\tilde{w} = (\rho(\tau), y(\tau), v(\tau), z(\tau))$ of Problem $\tilde{A}$. This transition from process w to process $\tilde{w}$ is not unique: it is determined by the choice of the functions $z_k(\tau)$. To make this transition unique, we set these functions, e.g., equal to constants, $z_k(\tau) = |\Delta_k|$. The resulting map is denoted by F.

Let us now define the map G. Let $\tilde{w} = (\rho(\tau), y(\tau), v(\tau), z(\tau))$ be an admissible process in Problem $\tilde{A}$. Then by definition of admissible process $\exists\, c > 0$ such that all $z_k(\tau) \geq c > 0$ almost everywhere. The functions $\rho_k(\tau)$ are strictly increasing on $[0,1]$, the inverse function $\tau(t) = \rho_k^{-1}(t)$ is bounded, and

$$\frac{d\tau(t)}{dt} = \frac{1}{z_k(\tau(t))}.$$

Define the points $t_0 = \rho_1(0)$, $t_1 = \rho_2(0)$, $\ldots$, $t_{\nu-1} = \rho_\nu(0)$, $t_\nu = \rho_\nu(1)$, and on each interval $\Delta_k = [t_{k-1}, t_k]$ define the functions $x(t) = y_k(\rho_k^{-1}(t))$, $u(t) = v_k(\rho_k^{-1}(t))$. Then by (4) $u(t) \in U$. Moreover,

$$x(t_0) = y_1(0), \quad x(t_1) = y_2(0), \quad \ldots, \quad x(t_{\nu-1}) = y_\nu(0), \quad x(t_\nu) = y_\nu(1),$$

and thus $p(t,x) = \hat{p}(\rho, y)$. Hence by (7) all $\varphi_i(p) \leq 0$, $\eta_j(p) = 0$.

The function $x(t)$ is absolutely continuous on each interval Δ_k and almost everywhere on this interval it satisfies the differential equation

$$\frac{dx}{dt} = \frac{dy_k}{d\tau}\frac{d\tau}{dt} = f(\rho_k, y_k, v_k) = f(t, x, u)$$

and is continuous at each intermediate point $t_k,$ because

$$x(t_k - 0) = y_k(\rho_k^{-1}(t_k)) = y_k(1) = y_{k+1}(0) = y_{k+1}(\rho_{k+1}^{-1}(t_k)) = x(t_k + 0).$$

Therefore $x(t)$ is absolutely continuous on the entire interval $\Delta = [t_0, t_\nu]$ and almost everywhere on Δ satisfies the same differential equation

$$\frac{dx}{dt} = f(t, x, u).$$

Thus, all the constraints of Problem A are satisfied for the process $w = (x(t),\ u(t), p)$ Therefore, w is an admissible process in Problem $A.$ The mapping from $\tilde{w}$ to w is denoted by $G.$

Remark 2. When investigating the maps F and G it is helpful to consider, in addition to the functions $\rho_k,\ y_k,\ v_k$ defined for $\tau \in [0,1],$ also the auxiliary "long" functions $P,\ Y,\ V,$ defined for $\tau \in [0,\nu]$ in the following way: on each interval $[k-1,\, k],\ \ k = 1,\dots,\nu,$

$$P(\tau) = \rho_k(\tau - k + 1) = t_{k-1} + |\Delta_k|\tau,$$

$$Y(\tau) = y_k(\tau - k + 1) = x(\rho_k(\tau - k + 1)),$$

$$V(\tau) = v_k(\tau - k + 1) = u(\rho_k(\tau - k + 1)).$$

Since the intervals Δ_k are joined end to end, the functions $P(\tau)$ and $Y(\tau)$ are continuous, P is piecewise-linear and $\dfrac{dP}{d\tau} \geq \text{const} > 0.$ Then the inverse of $P(\tau),\ \ \theta\colon \Delta \to [0,\nu],$ is also defined: it is piecewise-linear and strictly increasing.

Using these functions, we can represent the process $w = G(\tilde{w})$ in the form $w = (x(t), u(t), p),$ where

$$x(t) = Y(\theta(t)), \qquad u(t) = V(\theta(t)). \tag{8}$$

Note that the maps F and G and not mutually invertible $(GF$ is the identity map, while FG is not). Nevertheless, the very existence of two transformations mapping any admissible point (in out case, process) of one problem to some admissible point of another problem while preserving the functional value immediately leads to the following proposition.

Theorem 2. *If the process w^0 is optimal (i.e., attains the global minimum) in Problem $A,$ then the process $\tilde{w}^0 = F(w^0)$ is optimal in Problem $\tilde{A},$ and conversely, if some process $\tilde{w}^0$ is optimal in Problem $\tilde{A},$ then the process $w^0 = G(\tilde{w}^0)$ is optimal in Problem $A.$*

Indeed, let us prove the first implication. Assume that the process w^0 is optimal in Problem $A.$ If the process $\tilde{w}^0 = F(w^0)$ is not optimal in Problem $\tilde{A},$ then there exists another admissible process $\tilde{w},$ in this problem for which $\tilde{J}(\tilde{w}) < \tilde{J}(\tilde{w}^0).$ Then the corresponding process $w = G(\tilde{w})$ is admissible in Problem A and for this process

$$J(w) = \tilde{J}(\tilde{w}) < \tilde{J}(\tilde{w}^0) = J(w^0),$$

which contradicts the optimality of the process $w^0.$ The second implication is proved similarly. Q.E.D.

Theorem 2 deals with the conservation of a very crude property (global minimality); it does not allow for any specific features of the problems and the maps $F,\ G$. For our Problems $A,\ \tilde{A}$, and the maps $F,\ G$ we have a sharper proposition.

Theorem 3. *If Problem A attains a strong (a Pontryagin) minimum on the process w^0, then Problem $\tilde{A}$ attains a strong (respectively a Pontryagin) minimum on the process $\tilde{w}^0 = F(w^0)$, and conversely, if Problem $\tilde{A}$ attains a strong (a Pontryagin) minimum on some process $\tilde{w}^0$, then Problem A attains a strong (respectively a Pontryagin) minimum on the process $w^0 = G(\tilde{w}^0)$.*

Proof. We are primarily interested in the direct implication. We will show that if Problem A attains a strong (a Pontryagin) minimum on the process $w^0 = (x^0(t), u^0(t), p^0)$, then Problem $\tilde{A}$ attains a strong (respectively a Pontryagin) minimum on the process $\tilde{w}^0 = F(w^0) = (\rho^0(\tau), y^0(\tau), v^0(\tau), z^0(\tau))$. To this end, it suffices to show that if the sequence of admissible processes $\tilde{w}^i = (\rho^i(\tau), y^i(\tau), v^i(\tau), z^i(\tau))$ in Problem $\tilde{A}$ is "strongly" (Pontryagin) convergent to the process $\tilde{w}^0 = (\rho^0(\tau), y^0(\tau), v^0(\tau), z^0(\tau))$ as $i \to \infty$, then the corresponding sequence of processes $w^i = G(\tilde{w}^i) = (x^i(t), u^i(t), p^i)$ is "strongly" (respectively Pontryagin) convergent to the process $w^0 = G(\tilde{w}^0) = (x^0(t), u^0(t), p^0)$. (This would imply that violation of a strong or Pontryagin minimum in Problem $\tilde{A}$ leads to violation of the corresponding minimum in Problem A.)

First consider the "strong" convergence case. For every element of the sequence, as noted previously, we have the intervals Δ^i_k and $\Delta^i = \bigcup\limits_{k=1}^{\nu} \Delta^i_k$. The convergence of $t^i_k = \rho^i_{k+1}(0)$ to $t^0_k = \rho^0_{k+1}(0)$ for all $k = 0, 1, \ldots, \nu-1$ and of $t^i_\nu = \rho^i_\nu(1)$ to $t^0_\nu = \rho^0_\nu(1)$ follows from uniform convergence $\rho^i_k(\tau) \rightrightarrows \rho^0_k(\tau)$.

Let us show that $\|x^i - x^0\|_C \to 0$ on $\Delta^i \cap \Delta^0$. As noted in Remark 2, with each process $\tilde{w}^i$ and the process $\tilde{w}^0$ we can associate respectively the functions $\theta^i(t)$ and $Y^i(\tau)$ and the functions $\theta^0(t)$ and $Y^0(\tau),\ \tau \in [0, \nu]$. Furthermore, by (8) the proof of "strong" convergence $x^i \rightrightarrows x^0$ on $\Delta^i \cap \Delta^0$ is equivalent to the proof of uniform convergence $Y^i(\theta^i) \rightrightarrows Y^0(\theta^0)$.

By uniform convergence $y^i_k \rightrightarrows y^0_k$ on $[0, 1]$, the functions $Y^i(\tau)$ obviously converge uniformly to $Y^0(\tau)$ on each interval $[k-1, k]$ and thus on the entire interval $[0, \nu]$. Since $\rho^i \rightrightarrows \rho^0$ on $[0, \nu]$, we have $\theta^i(t) \rightrightarrows \theta^0(t)$ on $\Delta^i \cap \Delta^0$. The required fact follows from the next lemma.

Lemma 1. *Let $\{y^i(\tau)\}$ be a sequence of continuous functions defined on some interval I and uniformly convergent on I to the function $y^0(\tau)$ as $i \to \infty$. Given is a sequence of continuous functions $s^i\colon \Delta^i \to I$ and the function $s^0\colon \Delta^0 \to I$ such that $\Delta^i \to \Delta^0$ and $s^i(t) \rightrightarrows s^0(t)$ uniformly on $\Delta^i \cap \Delta^0$. Then $y^i(s^i(t)) \rightrightarrows y^0(s^0(t))$ on $\Delta^i \cap \Delta^0$.*

Proof. We have the obvious bound

$$\|y^i(s^i) - y^0(s^0)\|_C = \max_{t \in \Delta^i \cap \Delta^0} \left|y^i(s^i(t)) - y^0(s^0(t))\right|$$

$$\leq \max_{t \in \Delta^i \cap \Delta^0} \left|y^i(s^i(t)) - y^0(s^i(t))\right| + \max_{t \in \Delta^i \cap \Delta^0} \left|y^0(s^i(t)) - y^0(s^0(t))\right|.$$

Consider each term in the right-hand side of this inequality. The first term obviously goes to zero, because $y^i(\tau) \rightrightarrows y^0(\tau)$ on I. Now, since $|s^0(t) - s^i(t)| \rightrightarrows 0$ on $\Delta^i \cap \Delta^0$, the second term also goes to zero by uniform continuity of $y^0(\tau)$ on I. Q.E.D.

Let us apply this lemma to our functions $Y^i(\tau),\ Y^0(\tau)$ and intervals Δ^i and Δ^0, setting $s^i(t) = \theta^i(t)$, $s^0(t) = \theta^0(t),\ I = [0, \nu]$. We obtain uniform convergence $x^i(t) = Y^i(\theta^i(t)) \rightrightarrows Y^0(\theta^0(t)) = x^0(t)$ on $\Delta^i \cap \Delta^0$. Thus, the process $(x^i(t), u^i(t), p^i)$ "strongly" converges to the process $(x^0(t), u^0(t), p^0)$. Q.E.D.

Let us now consider the Pontryagin minimum. Let $\tilde{w}^i = (\rho^i(\tau), y^i(\tau), v^i(\tau), z^i(\tau))$ be a sequence of admissible processes of Problem $\tilde{A}$ Pontryagin-convergent to the process $\tilde{w}^0 = (\rho^0(\tau), y^0(\tau), v^0(\tau), z^0(\tau))$. We will show that the sequence $w^i = G(\tilde{w}^i)$ is Pontryagin-convergent to $w^0 = G(\tilde{w}^0)$.

We have just proved the uniform convergence $x^i \rightrightarrows x^0$, and it is therefore sufficient to check that Pontryagin convergence of the controls $v^i \to v^0$ and $z^i \to z^0$ implies Pontryagin convergence $u^i \to u^0$. Note that since v^i and v^0 are uniformly bounded (by definition of Pontryagin convergence), then u^i and u^0 are also uniformly bounded and it remains to prove the convergence $u^i \to u^0$ in the norm L_1. To this end we need two lemmas. The first is a slight modification of Lemma 42 [22].

Lemma 2. *Let $v(\tau) \in L_1$ on some interval I and let the sequence of absolutely continuous functions $s^i \colon \Delta^i \to I$ and the absolutely continuous function $s^0 \colon \Delta^0 \to I$ be given such that $\Delta^i \to \Delta^0$ and $s^i(t) \rightrightarrows s^0(t)$ uniformly on $\Delta^i \cap \Delta^0$. Moreover, let $\dfrac{ds^i(t)}{dt} \geq \mathrm{const} > 0$ on Δ^i. Then*

$$\int\limits_{\Delta^i \cap \Delta^0} \left| v(s^i(t)) - v(s^0(t)) \right| dt \to 0. \tag{9}$$

Proof. This is obviously true for the characteristic functions of intervals (follows from pointwise convergence $s^i(t) \to s^0(t)$), and is therefore true for arbitrary linear combinations of these functions. Hence, it is also true for any continuous $v(t)$ (because they are uniformly approximated by such combinations).

Now let $v \in L_1(I)$. Then $\forall\, \varepsilon > 0$ there exists a continuous function v_ε such that $\|v_\varepsilon - v\|_1 < \varepsilon$. Here

$$\int \left| v(s^i(t)) - v(s^0(t)) \right| dt \leq \int \left| v(s^i(t)) - v_\varepsilon(s^i(t)) \right| dt + \int \left| v_\varepsilon(s^i(t)) - v_\varepsilon(s^0(t)) \right| dt + \int \left| v_\varepsilon(s^0(t)) - v(s^0(t)) \right| dt. \tag{10}$$

(All integrals are over the interval $\Delta^i \cap \Delta^0$.)

Taking ε fixed, we bound each integral in the right-hand side of this inequality. Since $s^i(t) \rightrightarrows s^0(t)$ on $\Delta^i \cap \Delta^0$ and the function v_ε is uniformly continuous on $\Delta^i \cap \Delta^0$, the second integral $< \varepsilon$ for sufficiently large i. In the first integral we change the variable to $s = s^i(t)$ and set $\omega^i = s^i(\Delta^i \cap \Delta^0)$. By assumption $\dfrac{dt}{ds^i} \leq C = \mathrm{const}$, and this integral is bounded by

$$\int\limits_{\omega^i} \left| v(s) - v_\varepsilon(s) \right| \frac{dt}{ds^i}\, ds \leq C \|v_\varepsilon - v\|_1 < C\varepsilon.$$

Similarly the third integral $< C\varepsilon$. Summing, we obtain that for sufficient large i the right-hand side of inequality (10) is less than $(1 + 2C)\varepsilon$. Since $\varepsilon > 0$ is arbitrary, this proves (9). Q.E.D.

We now generalize this lemma to the case of a sequence of functions v^i.

Lemma 3. *Let $v^i(\tau) \in L_1(I)$ and $||v^i - v^0||_1 \to 0$. Given is the sequence of absolutely continuous functions $s^i \colon \Delta^i \to I$ and the absolutely continuous function $s^0 \colon \Delta^0 \to I$ such that $\Delta^i \to \Delta^0$ and $s^i(t) \rightrightarrows s^0(t)$*

uniformly on $\Delta^i \cap \Delta^0$. *Moreover, let* $\dfrac{ds^i(t)}{dt} \geq \mathrm{const} > 0$ *on* Δ^i. *Then*

$$\int\limits_{\Delta^i\cap\Delta^0} \left|v^i(s^i(t)) - v^0(s^0(t))\right| dt \to 0.$$

Proof. We have the obvious bound

$$\int\limits_{\Delta^i\cap\Delta^0} \left|v^i(s^i(t)) - v^0(s^0(t))\right| dt \leq \int\limits_{\Delta^i\cap\Delta^0} \left|v^i(s^i(t)) - v^0(s^i(t))\right| dt + \int\limits_{\Delta^i\cap\Delta^0} \left|v^0(s^i(t)) - v^0(s^0(t))\right| dt.$$

Let us bound each integral in the right-hand side of this inequality. The second integral goes to zero by Lemma 2. In the first integral we change the variable to $s = s^i(t)$ and set $\omega^i = s^i(\Delta^i \cap \Delta^0)$. Since by assumption $\dfrac{dt}{ds^i} \leq C = \mathrm{const},$ we have

$$\int\limits_{w^i} \left|v^i(s) - v^0(s)\right| \frac{dt}{ds^i}\, ds \leq C\|v^i - v^0\|_1 \to 0,$$

and this integral also goes to zero. Q.E.D.

Applying Lemma 3 to our functions $v^i(\tau)$, $v^0(\tau)$ and setting $s^i(t) = \theta^i(t)$, $s^0(t) = \theta^0(t)$, $I = [0, \nu]$, we obtain the convergence $u^i(t) = V^i(\theta^i(t)) \to V^0(\theta^0(t)) = u^0(t)$ in the norm L_1 on $\Delta^i \cap \Delta^0$. We have thus proved the Pontryagin convergence $u^i \to u^0$.

We have shown that convergence of the processes $\tilde{w}^i \to \tilde{w}^0$ in Problem $\tilde{A}$ implies corresponding convergence of the processes $w^i = G(\tilde{w}^i) \to w^0 = G(\tilde{w}^0)$ in Problem A. Similarly we can prove the converse proposition, i.e., if the sequence of admissible processes $w^i = (x^i(t), u^i(t), p^i)$ of Problem A is "strongly" (Pontryagin) convergent to the process $w^0 = (x^0(t), u^0(t), p^0)$, then the corresponding sequence of processes $\tilde{w}^i = F(w^i) = (\rho^i(\tau), y^i(\tau), v^i(\tau), z^i(\tau))$ of Problem $\tilde{A}$ is "strongly" (respectively, Pontryagin) convergent to the process $\tilde{w}^0 = F(\tilde{w}^0) = (\rho^0(\tau), y^0(\tau), v^0(\tau), z^0(\tau))$. This completes the proof of Theorem 2. Q.E.D.

Thus, optimality of the process w^0 in Problem A in each of the three senses indicated above (Pontryagin, strong, and global) corresponds to optimality of the process $\tilde{w}^0$ in Problem $\tilde{A}$.

3. Pontryagin Maximum Principle for Problem $\tilde{A}$ and Its Analysis

Let $\tilde{w}^0 = (\rho^0(\tau),\, y^0(\tau),\, v^0(\tau),\, z^0(\tau))$ be an arbitrary process on which Problem $\tilde{A}$ attains a Pontryagin minimum. Then, by Theorem 1, this process satisfies the Pontryagin MP, which can be formulated as follows.

There exists a tuple

$$\lambda = (\alpha,\, \beta,\, \gamma,\, \delta,\, \sigma,\, \psi_y(\cdot),\, \psi_\rho(\cdot)),$$

where $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_m) \in \mathbb{R}^{m+1}$, $\beta = (\beta_1, \beta_2, \ldots, \beta_q) \in \mathbb{R}^q$, $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_{\nu-1}) \in \mathbb{R}^{\nu-1}$, $\delta = (\delta_1,\, \delta_2,\, \ldots,\, \delta_{\nu-1}) \in \mathbb{R}^{\nu-1}$, $\sigma \in \mathbb{R}^1$, $\psi_y = (\psi_{y_1},\, \psi_{y_2},\, \ldots,\, \psi_{y_\nu})$, $\psi_\rho = (\psi_{\rho_1}, \psi_{\rho_2}, \ldots, \psi_{\rho_\nu})$, the functions ψ_{y_k}, and ψ_{ρ_k} are Lipschitzian on $[0, 1]$, for which we construct

– the Pontryagin function

$$\tilde{H}(\psi_\rho, \psi_y, \rho, y, v, z) = \sum_{k=1}^{\nu} z_k\big(\psi_{y_k} f(\rho_k, y_k, v_k) + \psi_{\rho_k}\big) = \sum_{k=1}^{\nu} z_k\, \Pi_k(\psi_{\rho_k}, \psi_{y_k}, \rho_k, y_k, v_k),$$

where $\Pi_k(\psi_{\rho_k}, \psi_{y_k}, \rho_k, y_k, v_k) = \psi_{y_k} f(\rho_k, y_k, v_k) + \psi_{\rho_k}$,

– the endpoint Lagrange function

$$\tilde{l}(\tilde{p}) = l(\hat{p}) + \sum_{k=1}^{\nu-1} \gamma_k (y_{k+1}(0) - y_k(1)) + \sum_{k=1}^{\nu-1} \delta_k (\rho_{k+1}(0) - \rho_k(1)),$$

where

$$l(\hat{p}) = \sum_{i=0}^{m} \alpha_i \varphi_i(\hat{p}) + \sum_{j=1}^{q} \beta_j \eta_j(\hat{p}),$$

and the following conditions hold:

(a) nontriviality condition: $(\alpha, \beta, \gamma, \delta) \neq (0, 0, 0, 0)$;

(b) nonnegativity conditions: $\alpha_i \geq 0,\ \ i = 0, \ldots, m$;

(c) complementary slackness conditions: $\alpha_i \varphi_i(\hat{p}^0) = 0,\ \ i = 1, \ldots, m$;

(d) conjugate equations

$$\dot{\psi}_{y_k}(\tau) = -\tilde{H}^0_{y_k} = -z^0_k(\tau) \psi_{y_k}(\tau) f_x(\rho^0_k(\tau), y^0_k(\tau), v^0_k(\tau)),$$

$$\dot{\psi}_{\rho_k}(\tau) = -\tilde{H}^0_{\rho_k} = -z^0_k(\tau) \psi_{y_k}(\tau) f_t(\rho^0_k(\tau), y^0_k(\tau), v^0_k(\tau)), \quad k = 1, \ldots, \nu;$$

(e) transversality conditions:

$$\begin{cases} \psi_{y_1}(0) = l_{y_1(0)}, & \psi_{y_1}(1) = \gamma_1, \\ \psi_{y_2}(0) = l_{y_2(0)} + \gamma_1, & \psi_{y_2}(1) = \gamma_2, \\ \dots\dots\dots\dots\dots\dots & \\ \psi_{y_{\nu-1}}(0) = l_{y_{\nu-1}(0)} + \gamma_{\nu-2}, & \psi_{y_{\nu-1}}(1) = \gamma_{\nu-1}, \\ \psi_{y_\nu}(0) = l_{y_\nu(0)} + \gamma_{\nu-1}, & \psi_{y_\nu}(1) = -l_{y_\nu(1)}; \end{cases}$$

$$\begin{cases} \psi_{\rho_1}(0) = l_{\rho_1(0)}, & \psi_{\rho_1}(1) = \delta_1, \\ \psi_{\rho_2}(0) = l_{\rho_2(0)} + \delta_1, & \psi_{\rho_2}(1) = \delta_2, \\ \dots\dots\dots\dots\dots\dots & \\ \psi_{\rho_{\nu-1}}(0) = l_{\rho_{\nu-1}(0)} + \delta_{\nu-2}, & \psi_{\rho_{\nu-1}}(1) = \delta_{\nu-1}, \\ \psi_{\rho_\nu}(0) = l_{\rho_\nu(0)} + \delta_{\nu-1}; & \psi_{\rho_\nu}(1) = -l_{\rho_\nu(1)}, \end{cases}$$

where all derivatives of the function $l(p)$ are at the point $\hat{p}^0$;

(f) constancy of $\tilde{H}$: for almost all $\tau \in [0,1]$,

$$\tilde{H}(\psi_\rho(\tau), \psi_y(\tau), \rho^0(\tau), y^0(\tau), v^0(\tau), z^0(\tau)) = \sigma;$$

(g) maximum condition: for all $\tau \in [0,1]$

$$\max_{(v,z)\in\tilde{C}(\tau)} \tilde{H}(\psi_\rho(\tau), \psi_y(\tau), \rho^0(\tau), y^0(\tau), v, z) = \sigma,$$

where

$$\tilde{C}(\tau) = V(\tau) \times \{z = (z_1, \dots, z_\nu) \mid \text{all } z_k > 0\},$$

$$V(\tau) = V_1(\tau) \times \dots \times V_\nu(\tau), \quad V_k(\tau) = \{v_k \in U \mid (\rho_k^0(\tau), y_k^0(\tau), v_k) \in \mathcal{Q}\}.$$

Let us analyze conditions (a)–(g).

1. From conditions (f), (g) it follows that $\forall\, k,$ for almost all $\tau \in [0,1]$ the function $\tilde{H}$ attains at the point $z_k^0(\tau)$ its maximum on the open set $z_k > 0.$ Since $\tilde{H}$ is linear in all $z_k,$ we have

$$\frac{\partial \tilde{H}}{\partial z_k}(\psi_\rho(\tau), \psi_y(\tau), \rho^0(\tau), y^0(\tau), v^0(\tau), z^0(\tau)) = \Pi_k^0 = 0 \tag{11}$$

for almost all $\tau \in [0,1].$ Hence $\sigma = 0.$ This combined with condition (g) obviously implies (all the controls v_k enter $\tilde{H}$ separately) that for all $\tau \in [0,1]$

$$\max_{v_k \in V_k(\tau)} \Pi_k\,(\psi_{\rho_k}, \psi_{y_k}, \rho_k^0(\tau), y_k^0(\tau), v_k) = 0. \tag{12}$$

2. Let us simplify the nontriviality condition.

Lemma 4. *The inequality*

$$\sum_{i=0}^{m} \alpha_i + \sum_{j=1}^{q} |\beta_j| + \sum_{k=1}^{\nu-1} |\gamma_k| + \sum_{k=1}^{\nu-1} |\delta_k| > 0$$

is equivalent to the inequality

$$\sum_{i=0}^{m} \alpha_i + \sum_{j=1}^{q} |\beta_j| > 0. \tag{13}$$

Proof. The implication $(\Leftarrow)$ is obvious. The implication $(\Rightarrow)$ is proved by contradiction. Let $\alpha_i = \beta_j = 0.$ Then by the transversality condition $\psi_{y_1}(0) = 0$ and by linearity of the conjugate equation in ψ_{y_1} we obtain $\psi_{y_1}(\tau) \equiv 0,$ which again by the transversality condition gives $\gamma_1 = 0.$ Thus $\psi_{y_2}(0) \equiv 0.$ Then we successively find that all the remaining $\psi_{y_k} \equiv 0$ and $\gamma_k = 0.$

We similarly show that all $\psi_{\rho_k}(\tau) \equiv 0,$ $\delta_k = 0.$ Q.E.D.

The nontriviality condition (a) thus can be replaced with the equivalent condition (13).

3. Note that the Lagrange multipliers $\gamma_k,\ \delta_k$ are present only in the transversality conditions (e) and they can be eliminated by rewriting these conditions in the following form:

– the transversality conditions for ψ_y :

$$\begin{aligned}
&\psi_{y_1}(0) = l_{y_1(0)}(\hat p^0),\\
&\psi_{y_{k+1}}(0) - \psi_{y_k}(1) = l_{y_{k+1}(0)}(\hat p^0), \quad k = 1,\dots,\nu-1,\\
&\psi_{y_\nu}(1) = -l_{y_\nu(1)}(\hat p^0);
\end{aligned} \tag{14}$$

– the transversality conditions for ψ_ρ :

$$\begin{aligned}
&\psi_{\rho_1}(0) = l_{\rho_1(0)}(\hat p^0),\\
&\psi_{\rho_{k+1}}(0) - \psi_{\rho_k}(1) = l_{\rho_{k+1}(0)}(\hat p^0), \quad k = 1,\dots,\nu-1,\\
&\psi_{\rho_\nu}(1) = -l_{\rho_\nu(1)}(\hat p^0).
\end{aligned} \tag{15}$$

We assert that these conditions are equivalent to the original transversality conditions. Indeed, conditions (14), (15) follow from transversality conditions (e), and conversely, if conditions (14), (15) are satisfied, the transversality condition (e) will hold when $\gamma_k = \psi_{y_k}(1),\ \delta_k = \psi_{\rho_k}(1),\ k = 1,\dots,\nu-1.$ Since the multipliers γ_k, δ_k do not affect nontriviality, this operation does not lead to any loss of information, and these multipliers can be excluded from the MP for Problem $\tilde A$.

4. Now assume that our process $\tilde w^0$ has been obtained from some process w^0 of Problem A by the mapping F. To reconstruct the process w^0, we apply the transformation G to $\tilde w^0$ (since $\tilde w^0 = F(w^0)$ implies $G(\tilde w^0) = GF(w^0) = w^0$). To this end, following the above procedure, we define the time instants $t^0_{k-1} = \rho^0_k(0)$, $k = 1,\dots,\nu,\ t^0_\nu = \rho^0_\nu(1)$, and the intervals $\Delta^0_k = [t^0_{k-1},\, t^0_k],\ k = 1,\dots,\nu.$. On each interval we take $\theta^0_k(t) = (\rho^0_k)^{-1}(t)$ and define the functions

$$x^0(t) = y^0_k(\theta^0_k(t)), \qquad u^0(t) = v^0_k(\theta^0_k(t)), \qquad \psi_t(t) = \psi_{\rho_k}(\theta^0_k(t)), \qquad \psi_x(t) = \psi_{y_k}(\theta^0_k(t)).$$

These functions are thus defined on the entire interval $\Delta^0 = [t^0_0,\, t^0_\nu]$ and we have reconstructed the process $w^0 = (x^0(t), u^0(t), p^0)$, where $p^0 = \big((t^0_0, x^0(t^0_0)), (t^0_1, x^0(t^0_1)), \dots, (t^0_\nu, x^0(t^0_\nu))\big)$.

Since all $\psi_{\rho_k}(\tau)$ and $\psi_{y_k}(\tau)$ are Lipschitzian on $[0,1]$, $\psi_t(t)$ and $\psi_x(t)$ are Lipschitzian on each Δ^0_k with their own Lipschitz constants and may have discontinuities at the points t^0_k (such functions are called piecewise-Lipschitzian). Conditions (14), (15) are rewritten in terms of $\psi_x,\ \psi_t$ as transversality conditions at the endpoints of the interval Δ^0 and discontinuity conditions at intermediate points. On each Δ^0_k we have the equations

$$\frac{d\psi_t}{dt} = \frac{d\psi_{\rho_k}}{d\tau}\frac{d\tau}{dt} = -\psi_x\, f_t(t,x,u), \qquad \frac{d\psi_x}{dt} = \frac{d\psi_{y_k}}{d\tau}\frac{d\tau}{dt} = -\psi_x\, f_x(t,x,u).$$

We now introduce the function $H(\psi_t, \psi_x, t, x, u) = \psi_x\, f(t,x,u) + \psi_t$. On the process $x^0(t),\ u^0(t)$ with the above $\psi_x(t),\ \psi_t(t)$ it is obviously identical with

$$\Pi_k(\psi_{\rho_k}(\tau), \psi_{y_k}(\tau), \rho^0_k(\tau), y^0_k(\tau), v^0_k(\tau)) \qquad \text{for } \tau = \theta^0_k(t),$$

and therefore from (11) we obtain that $H(\psi_t(t), \psi_x(t), t, x^0(t), u^0(t)) = 0$ almost everywhere on each Δ_k^0, and thus on the entire Δ^0. From (12) we obtain

$$\max_{u \in C(t)} H(\psi_t(t), \psi_x(t), t, x^0(t), u^0(t)) = 0,$$

where $C(t) = \{u \in U \,|\, (t, x^0(t), u) \in \mathcal{Q}\}$.

If Problem A attains a Pontryagin minimum on the process w^0, then Theorem 3 and our analysis of the MP for the process $\tilde{w}^0$ lead to the following theorem.

Theorem 4 (Maximum principle for Problem A)**.** *Assume that Problem A attains a Pontryagin minimum on the process $w^0 = (x^0(t), u^0(t), p^0)$, $t \in \Delta^0$. Then there exists a tuple $\lambda = (\alpha, \beta, \psi_x(\cdot), \psi_t(\cdot))$, where $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_m) \in \mathbb{R}^{m+1}$, $\beta = (\beta_1, \beta_2, \ldots, \beta_q) \in \mathbb{R}^q$, ψ_x and ψ_t are piecewise-Lipschitzian functions on $\Delta^0 = [t_0^0, t_\nu^0]$, for which we construct*

– *the Pontryagin function*

$$H(\psi_t, \psi_x, t, x, u) = \langle \psi_x, f(t, x, u) \rangle + \psi_t,$$

– *the endpoint Lagrange function*

$$l(p) = \sum_{i=0}^{m} \alpha_i \varphi_i(p) + \sum_{j=1}^{q} \beta_j \eta_j(p),$$

and the following conditions are satisfied:

(a) *nontriviality condition:* $(\alpha, \beta) \neq (0, 0)$;

(b) *nonnegativity conditions:* $\alpha_i \geq 0$, $i = 0, \ldots, m$;

(c) *complementary slackness conditions:* $\alpha_i \varphi_i(p^0) = 0$, $i = 1, \ldots, m$;

(d) *conjugate equations: almost everywhere on* Δ^0,

$$\dot{\psi}_x(t) = -H_x^0 = -\psi_x(t) f_x(t, x^0(t), u^0(t)),$$

$$\dot{\psi}_t(t) = -H_t^0 = -\psi_x(t) f_t(t, x^0(t), u^0(t));$$

(e) *transversality conditions at the interval endpoints:*

$$\psi_x(t_0^0) = l_{x(t_0)}(p^0), \quad \psi_x(t_\nu^0) = -l_{x(t_\nu)}(p^0),$$

$$\psi_t(t_0^0) = l_{t_0}(p^0); \qquad \psi_t(t_\nu^0) = -l_{t_\nu}(p^0);$$

(f) *discontinuity conditions for* ψ_x *and* ψ_t *at intermediate points:*

$$\triangle \psi_x(t_k^0) = \psi_x(t_k^0 + 0) - \psi_x(t_k^0 - 0) = l_{x(t_k)}(p^0),$$

$$\triangle \psi_t(t_k^0) = \psi_t(t_k^0 + 0) - \psi_t(t_k^0 - 0) = l_{t_k}(p^0);$$

(g) for almost all $t \in \Delta^0$ $H(\psi_t(t), \psi_x(t), t, x^0(t), u^0(t)) = 0;$

(h) maximum condition: for all $t \in \Delta^0,$

$$\max_{u\in C(t)} H(\psi_t(t), \psi_x(t), t, x^0(t), u) = 0,$$

where $C(t) = \{u \in U \mid (t, x^0(t), u) \in \mathcal{Q}\}.$

Discussion of Theorem 4. It is easy to see that the Pontryagin MP for Problem K (Theorem 1) follows from Theorem 4. Indeed, apply Theorem 4 to Problem $K,$ noting that $\nu = 1.$ Conditions (a)–(e) of both theorems are identical. Since there are no intermediate points, we have no discontinuity conditions for the conjugate variables at the intermediate points; and since Problem K is autonomous, the conjugate variable $\psi_t = \mathrm{const}$ on the entire $\Delta^0.$ Denoting this constant by $-\sigma,$ we obtain conditions (f)–(g) of Theorem 1.

Problem A is thus a generalization of Pontryagin Problem $K,$ and Theorem 4 generalizes the classical Pontryagin MP (even in its strong form, as a necessary condition of Pontryagin minimum, not global optimality). However, this generalization concerns only the form of the MP and in fact does not touch its essence.

As we have seen, the proof of Theorem 4 does not require the implementation of the full complex procedure that directly proves the MP through the introduction of a special class of variations (e.g., spike variations), examination of adjoint problems in this class of variations, writing out of time-independence conditions for each such problem, or approximation of the reachability set by tangent cones, expansion of the functional on these variations, and so on. None of this needs to be done if we assume that the MP is known for Problem K with unseparated endpoint equalities (and also inequalities). Then, as we have shown, the MP for Problem A is derived from the MP for Problem K by a simple change of variables.

Thus, if the Pontryagin MP is stated for the "correct" Problem $K,$ which we accept as canonical, then it is easily extended to the formally more general Problem A (and, as we shall see below, not only to Problem A). In our view, the possibility of this extension to more general classes of problems is evidence that the potential of the Pontryagin MP has not been fully explored and recognized.

It is interesting to note that in classical variational calculus (CVC) the totally natural technique, used above, of propagation of variables and their reduction to a single time interval has been known for decades: already after writing this article the authors discovered in [21, pp. 57–62] a reference to the dissertation of the American mathematician C. H. Denbow [9] from 1937, where this technique was described in detail and applied to various CVC problems. However, by the 1960 s, when optimal control theory was being developed, virtually the entire CVC theory from the first half of the 20th century had sunk into oblivion and many of its results were simply forgotten.

B. M. Miller brought to our attention the fact that in optimal control theory this technique had been applied by Volin and Ostrovskii [10] for phase-constrained problems (t_k were the points where the system hit or left the phase boundary) and also for Problem $C',$ which is considered below (but certainly not with the same elaboration of all details as in the present study). However, Volin and Ostrovskii's article escaped the attention of experts in optimization theory, the present authors included, because it had been published in a journal very far from the relevant field and also because this technique had not been clearly identified among other constructs. This neglect is confirmed by much later publications [11, 15, 18–20], where the MP for problems of type A is derived as an independent result, by implementing the entire variation procedure, or remains without proof [14], despite the fact that it easily follows from the classical Pontryagin MP.

Ashchepkov [11] in his book considers Problem A with supplementary parameters (and also Problems B and C — see below); Arutyunov and Okulevich [15] in their article allow mixed constraints, but do not introduce the constraint $u \in U.$ On the class of "general applicability" problems our results are identical with the results of these studies (mixed constraints can obviously be allowed in Problem A). Other instances of earlier work will be mentioned later.

Note that conditions (f) of Theorem 4 concerning discontinuities of conjugate variables constitute a generalization of the Weierstrass–Erdmann conditions for extremals with corners in CVC problems. Under standard smoothness assumptions for the function f, which are always present in CVC, an extremal with corners implies that the phase variable $x(t)$ may experience a break at finitely many points $t_k \in (t_0, t_\nu)$ (i.e., the control $u(t)$ may have a discontinuity of the first kind at these corner points), while between these corner points $x(t)$ is smooth (i.e., $u(t)$ is continuous). The functional depends only on the trajectory endpoints $x(t_0)$, $x(t_\nu)$, and is independent of the break instants t_k and the values $x(t_k)$. (Recall that the integral part of the functional is reducible to an endpoint part.) The Weierstrass–Erdmann conditions indicate, in our terms, that at the points t_k^0 the conjugate functions $\psi_x(t)$ and $\psi_t(t) = -\psi_x(t)\, f(t, x^0(t), u^0(t))$ remain continuous (outside the corner points their continuity follows from continuity of $u^0(t)$).

In this case the endpoint Lagrange function is independent of t_k and $x(t_k)$, and conditions (f) therefore give $\Delta\psi_x(t_k^0) = 0$, $\Delta\psi_t(t_k^0) = 0$ — which is precisely the Weierstrass–Erdmann conditions.

Finally note that the Pontryagin function can be defined not in the form $H = \psi_x f + \psi_t$, but simply in the form $H = \psi_x f$ (as in Pontryagin's original work). Then condition (g) takes the form $H^0 + \psi_t = 0$, which in fact provides a definition of the function ψ_t, while condition (h) should be written in the form $\max_u H + \psi_t = 0$.

4. Some Generalizations of Problem A

I. Assume, as previously, that $t_0 < t_1 < \ldots < t_\nu$ are real numbers (not fixed) and for every n-dimensional continuous function $x(t)$ on the interval $[t_0, t_\nu]$ we determine the vector $p = \big((t_0, x(t_0)), (t_1, x(t_1)), \ldots, (t_\nu, x(t_\nu))\big)$.

Consider the following optimal control problem:

$$\textbf{Problem } B\textbf{:} \quad \begin{cases} \dot{x} = f_k(t, x, u), \quad u \in U_k, \qquad \text{for } t \in \Delta_k, \quad k = 1, \ldots, \nu, \\ \eta_j(p) = 0, \quad j = 1, \ldots, q, \\ \varphi_i(p) \le 0, \quad i = 1, \ldots, m, \\ J = \varphi_0(p) \to \min. \end{cases}$$

Unlike Problem A, here the trajectory $x(t)$ should satisfy on each interval Δ_k a differential equation with its own function f_k and its own control set U_k. Each function f_k is defined and satisfies Assumption A1 on its own open set $\mathcal{Q}_k \subset \mathbb{R}^{n+r+1}$.

The triple $w = (x(t), u(t), p)$ is called an admissible process if it satisfies all the constraints and for every $k = 1, \ldots, \nu$ there exists a compactum $\Omega_k \subset \mathcal{Q}_k$ such that $(t, x(t), u(t)) \in \Omega_k$ almost everywhere on Δ_k.

Note that the proof of Theorem 4 never assumed that the control system was the same on all intervals Δ_k. Therefore, as previously, we can reduce Problem B to a Pontryagin problem $\tilde{B}$, which differs from Problem $\tilde{A}$ only in one respect: the equations for $y_k(\tau)$, $\tau \in [0, 1]$, now have the form

$$\frac{dy_k}{d\tau} = z_k f_k(\rho_k, y_k, v_k), \quad v_k \in U_k, \quad k = 1, \ldots, \nu.$$

Writing the MP for Problem $\tilde{B}$ and analyzing it along the same lines as the MP for Problem $\tilde{A}$, we obtain the following theorem.

Theorem 5 (Maximum principle for Problem B)**.** *Assume that Problem B attains a Pontryagin maximum on the process $w^0 = (x^0(t), u^0(t),\, p^0)$. Then there exists a tuple $\lambda = (\alpha, \beta, \psi_x(\cdot), \psi_t(\cdot))$, where $\alpha = (\alpha_0, \alpha_1, \ldots, \lambda_m) \in \mathbb{R}^{m+1}$, $\beta = (\beta_1, \beta_2, \ldots, \beta_q) \in \mathbb{R}^q$, ψ_x and ψ_t are piecewise-Lipschitzian on Δ^0, for which we construct*

– *the Pontryagin function*

$$H(\psi_t, \psi_x, t, x, u) = \langle \psi_x, f_k(t, x, u) \rangle + \psi_t \qquad \textit{for } t \in \Delta_k^0,$$

– *the endpoint Lagrange function*

$$l(p) = \sum_{i=0}^{m} \alpha_i \varphi_i(p) + \sum_{j=1}^{q} \beta_j\, \eta_j(p),$$

and the following conditions are satisfied:

(a) nontriviality condition: $(\alpha, \beta) \neq (0, 0)$;

(b) nonnegativity conditions: $\alpha_i \geq 0$, $i = 0, \ldots, m$;

(c) complementary slackness conditions: $\alpha_i \varphi_i(p^0) = 0$, $i = 1, \ldots, m$;

(d) conjugate equations:

$$\begin{aligned} \dot{\psi}_x(t) &= -H_x^0 = -\psi_x(t) f_{kx}(t, x^0(t), u^0(t)), \\ \dot{\psi}_t(t) &= -H_t^0 = -\psi_x(t) f_{kt}(t, x^0(t), u^0(t)) \end{aligned} \qquad \textit{a.e. on } \Delta_k^0;$$

(e) transversality conditions at the interval endpoints:

$$\psi_x(t_0^0) = l_{x(t_0)}(p^0), \quad \psi_x(t_\nu^0) = -l_{x(t_\nu)}(p^0),$$

$$\psi_t(t_0^0) = l_{t_0}(p^0), \qquad \psi_t(t_\nu^0) = -l_{t_\nu}(p^0);$$

(f) discontinuity conditions for ψ_x and ψ_t at intermediate points:

$$\triangle\psi_x(t_k^0) = \psi_x(t_k^0 + 0) - \psi_x(t_k^0 - 0) = l_{x(t_k)}(p^0),$$

$$\triangle\psi_t(t_k^0) = \psi_t(t_k^0 + 0) - \psi_t(t_k^0 - 0) = l_{t_k}(p^0);$$

(g) for almost all $t \in \Delta^0$,

$$H(\psi_t(t), \psi_x(t), t, x^0(t), u^0(t)) = 0;$$

(h) maximum condition: for all $t \in \Delta_k^0$, $k = 1, \ldots, \nu$,

$$\max_{u \in C_k(t)} H(\psi_t(t), \psi_x(t), t, x^0(t), u) = 0,$$

where $C_k(t) = \big\{ u \in U_k \,\big|\, (t, x^0(t), u) \in \mathcal{Q}_k \big\}$.

Remark 3. Problem B encompasses very important classes of problems, which have attracted the attention of many experts. These include so-called multistage problems [14] and also variable structure problem, when a finite number of hyperplanes of the form $g_s(x) = 0$ are defined in the phase space, separating it into regions G_μ, each with its own controlled system $\dot{x} = f_\mu(x, u),\ \ u \in U_\mu$ (see, e.g., the recent work of Boltyanski [20]). In this setting it is assumed that the given trajectory crosses this surface transversally, i.e., at the time instants $t_k,$ when the trajectory goes from region G_{μ_k} to region G_{ν_k} crossing the surface $g_{s_k}(x) = 0$ we have the inequalities

$$\begin{aligned} &\dot{g}_{s_k}(x^0(t_k - 0)) = (g'_{s_k}(x^0(t_k)),\ f_{\mu_k}(x^0(t_k),\ u^0(t_k - 0)) > 0, \\ &\dot{g}_{s_k}(x^0(t_k + 0)) = (g'_{s_k}(x^0(t_k)),\ f_{\nu_k}(x^0(t_k),\ u^0(t_k + 0)) > 0, \end{aligned} \tag{16}$$

or both these quantities < 0 (the case when $\mu_k = \nu_k$ and the quantities (16) have opposite signs corresponds to the reflection of the trajectory from the given surface). It is easy to see that this case fits completely within Problem B (the constraints should be augmented with the equalities $g_{s_k}(x(t_k)) = 0$ for fixed time instants t_k), and there is no need to carry out directly the full variation procedure: it is sufficient to apply Theorem 5. Note that this transition to Problem B can be accomplished without requiring condition (16); the role of this condition is to ensure equivalence of small variations of the points t_k to small variations of the trajectory $x(t),$ as otherwise the latter are poorer than the former.

II. As further generalization of Problem $A,$ we consider Problem $C,$ which has the same form as Problem $B,$ the points t_k are again not fixed, but the trajectory may have discontinuities at these points t_k. The endpoint vector now has the form

$$\begin{aligned} p = \big(&(t_0, x(t_0)),\ (t_1, x(t_1 - 0), x(t_1 + 0)),\ (t_2,\ x(t_2 - 0), x(t_2 + 0)), \dots, \\ &(t_{\nu-1}, x(t_{\nu-1} - 0), x(t_{\nu-1} + 0)),\ (t_\nu, x(t_\nu))\big), \end{aligned}$$

where $x(t_k - 0),\ \ x(t_k + 0)$ are the left and the right values of the piecewise continuous function $x(t)$ at the points $t_k,\ \ k = 1, \dots, \nu - 1.$

Using a technique similar to that described above, we can derive a necessary condition of optimality for this problem also. Indeed, with Problem C on a free interval $[t_0,\ t_\nu]$ we can associate a standard optimal control Problem $\tilde{C}$ on a fixed time interval $[0, 1]$. Since the trajectory $x(t)$ may have discontinuities at intermediate points $t_k,$ there are no joining conditions for the phase variable $x(t)$ at time instants t_k and the Lagrange multipliers corresponding to these constraints are missing. Having established, as previously, the correspondence between Problems C and $\tilde{C},$ we write the Pontryagin MP for Problem $\tilde{C}$. Analysis of this MP leads to the following theorem.

Theorem 6 (Maximum principle for Problem C)**.** *Assume that Problem C attains a Pontryagin minimum on the process $w^0 = (x^0(t), u^0(t),\ p^0)$. Then there exists a tuple $\lambda = (\alpha, \beta, \psi_x(\cdot), \psi_t(\cdot)),$ where $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_m) \in \mathbb{R}^{m+1},\ \ \beta = (\beta_1, \beta_2, \dots, \beta_q) \in \mathbb{R}^q,\ \ \psi_x$ and ψ_t are piecewise-Lipschitzian on Δ^0, for which we construct*

- *the Pontryagin function*

$$H(\psi_t, \psi_x, t, x, u) = \langle \psi_x, f_k(t, x, u)\rangle + \psi_t, \quad t \in \Delta^0_k,$$

- *the endpoint Lagrange function*

$$l(p) = \sum_{i=0}^{m} \alpha_i \varphi_i(p) + \sum_{j=1}^{q} \beta_j \eta_j(p),$$

and the following conditions are satisfied:

(a) *nontriviality condition:* $(\alpha, \beta) \neq (0,0)$;

(b) *nonnegativity conditions:* $\alpha_i \geq 0,\ \ i = 0, \ldots, m$;

(c) *complementary slackness conditions:* $\alpha_i\, \varphi_i(p^0) = 0,\ \ i = 1, \ldots, m$;

(d) *conjugate equations:*

$$\left.\begin{aligned} \dot\psi_x(t) &= -H_x^0 = -\psi_x(t)\, f_{kx}(t, x^0(t), u^0(t)), \\ \dot\psi_t(t) &= -H_t^0 = -\psi_x(t)\, f_{kt}(t, x^0(t), u^0(t)) \end{aligned}\right. \qquad \text{a.e. on } \Delta_k^0;$$

(e) *transversality conditions at the interval endpoints:*

$$\psi_x(t_0^0) = l_{x(t_0)}(p^0), \quad \psi_x(t_\nu^0) = -l_{x(t_\nu)}(p^0),$$

$$\psi_t(t_0^0) = l_{t_0}(p^0), \qquad \psi_t(t_\nu^0) = -l_{t_\nu}(p^0);$$

(f) *transversality conditions for* ψ_x *and discontinuity conditions for* ψ_t *at intermediate points:*

$$\text{f1:} \quad \begin{cases} \psi_x(t_k^0 + 0) = l_{x(t_k+0)}(p^0), \\ \psi_x(t_k^0 - 0) = -l_{x(t_k-0)}(p^0), \quad k = 1, \ldots, \nu - 1, \end{cases}$$

$$\text{f2:} \quad \triangle\psi_t(t_k^0) = l_{t_k}(p^0);$$

(g) *for almost all* $t \in \Delta^0$

$$H(\psi_t(t), \psi_x(t), t, x^0(t), u^0(t)) = 0;$$

(h) *maximum condition: for all* $t \in \Delta_k^0,\ \ k = 1, \ldots, \nu$,

$$\max_{u \in C_k(t)} H(\psi_t(t), \psi_x(t), t, x^0(t), u) = 0,$$

where $C_k(t) = \big\{ u \in U_k \,\big|\, (t, x^0(t), u) \in \mathcal{Q}_k \big\}$.

The conditions on the conjugate variables ψ_x and ψ_t at intermediate points are different because the variable x may have discontinuities at these points, while the variable t is continuous.

Remark 4. The possibility of discontinuities of the trajectory $x(t)$ at intermediate points characterizes admissibility of impulse controls on the trajectory at time instants t_k. Problems with impulse controls are quite common in applications; see, e.g., [8, 16, 17]. The case when the impulses are allowed at the end points of the interval $[t_0, t_\nu]$, is also reduced to Problem C by introduction of two supplementary vector parameters equal to these impulses; these supplementary parameters, like any other parameters q, may be interpreted as phase variables subject to the equation $\dot q = 0$. These simple technical manipulations are omitted.

III. In the proof of Theorem 6 we nowhere assumed that the phase variable $x(t)$ and the control $u(t)$ in Problem C were of "constant" dimension on the entire interval Δ. Let, as previously, $t_0 < t_1 < \ldots < t_\nu$ be real numbers and $\Delta_k = [t_{k-1}, t_k]$. For every set of continuous functions $x_k \colon \Delta_k \to \mathbb{R}^{n_k}$, $k = 1, \ldots, \nu$, define the vector

$$p = \big(t_0,\, (t_1, x_1(t_0), x_1(t_1)),\, (t_2, x_2(t_1), x_2(t_2)),\, \ldots,\, (t_\nu, x_\nu(t_{\nu-1}),\, x_\nu(t_\nu))\big).$$

On the interval $[t_0, t_\nu]$ consider the following optimal control problem:

$$\textbf{Problem } C'\textbf{:} \qquad \begin{cases} \dot{x}_k = f_k(t, x_k, u_k), \quad u_k \in U_k, \qquad \text{for } t \in \Delta_k,\ k = 1, \ldots, \nu, \\ \eta_j(p) = 0, \quad j = 1, \ldots, q, \\ \varphi_i(p) \le 0, \quad i = 1, \ldots, m, \\ J = \varphi_0(p) \to \min, \end{cases}$$

where t_0, $t_1, \ldots, t_\nu$ are not fixed, $x_k \in \mathbb{R}^{n_k}$, $u_k \in \mathbb{R}^{r_k}$, the functions $x_k(t)$ are absolutely continuous, $u_k(t)$ are measurable bounded on their Δ_k. Each function f_k here satisfies Assumption A1 on the entire open set $\mathcal{Q}_k \subset \mathbb{R}^{n_k + r_k + 1}$ and takes values in $\mathbb{R}^{n_k}$.

If the dimensions of x and u are "constant" over the entire interval Δ, Problem C' clearly reduces to Problem C.

Applying the same procedure to prove the MP as in Problem C (here there is no need to propagate the variables x, u, because from the very beginning each Δ_k is assigned its own variables and it only remains to reduce them all to a single time interval), we obtain the MP for Problem C' (Theorem 7). The difference between Theorem 7 and Theorem 6 is that the set of Lagrange multipliers, instead of an n-dimensional piecewise-Lipschitzian conjugate variable $\psi_x \colon \Delta \to \mathbb{R}^n$, now contains a set of Lipschitz functions $\psi_{x_k} \colon \Delta_k \to \mathbb{R}^{n_k}$ that satisfy the conjugate equation

$$\dot{\psi}_{x_k}(t) = -\psi_{x_k}(t)\, f_{kx}(t, x_k^0(t), u_k^0(t)), \quad t \in \Delta_k^0,$$

and the transversality conditions for ψ_{x_k} take the form

$$\psi_{x_k}(t_{k-1}^0) = l_{x_k(t_{k-1})}(p^0), \qquad \psi_{x_k}(t_k^0) = -l_{x_k(t_k)}(p^0), \quad k = 1, \ldots, \nu.$$

All other conditions migrate without change from Theorem 6 to Theorem 7.

Problem C' encompasses problems for so-called hybrid control systems of a certain type (more precisely, this refers to the investigation of the given process for optimality in these problems). The maximum principle for such problems derived in [18] (see also [19]) follows from Theorem 7 for smooth control systems (see [23]). Hybrid systems of another type have been considered in [17].

IV. Finally, consider Problem D, which also has the same form as Problem C, but discontinuities are allowed for both the phase variable and the time. Here the vector p has the form

$$p = \big((t_0, t_1^{\mathrm{L}}, x(t_0), x(t_1^{\mathrm{L}})),\, (t_1^{\mathrm{R}}, t_2^{\mathrm{L}}, x(t_1^{\mathrm{R}}), x(t_2^{\mathrm{L}})),$$

$$\ldots,\ (t^{\mathrm{R}}_{(\nu-2)}, t^{\mathrm{L}}_{(\nu-1)}, x(t^{\mathrm{R}}_{(\nu-2)}), x(t^{\mathrm{L}}_{(\nu-1)})),\ (t^{\mathrm{R}}_{(\nu-1)}, t_\nu, x(t^{\mathrm{R}}_{(\nu-1)}), x(t_\nu))\big),$$

where $t_0 < t^{\mathrm{L}}_1 < t^{\mathrm{R}}_1 < \ldots < t^{\mathrm{L}}_{(\nu-1)} < t^{\mathrm{R}}_{(\nu-1)} < t_\nu$ are real numbers, the function $x(t)$ is defined on the union of the intervals $\Delta_k = [t^{\mathrm{R}}_{(k-1)},\, t^{\mathrm{L}}_k]$, $k = 1, \ldots, \nu$, and

$$x(t_0), x(t^{\mathrm{L}}_1),\ x(t^{\mathrm{R}}_1), \ldots, x(t^{\mathrm{L}}_{(\nu-1)}), x(t^{\mathrm{R}}_{(\nu-1)}), x(t_\nu) \in \mathbb{R}^n$$

are its values at the endpoints of these intervals.

We thus have a set of processes on nonintersecting intervals $\Delta_k = [t^{\mathrm{R}}_{(k-1)}, t^{\mathrm{L}}_k]$, $k = 1, \ldots, \nu$, and a functional that depends on the entire vector p. It is easy to see that the same technique as previously can be applied to derive the MP for Problem D (Theorem 8), which generalizes the MP for Problem C, i.e., Theorem 6. The difference between Theorem 8 and Theorem 6 is in conditions (f), which now take the following form:

(f′) transversality conditions for ψ_x and ψ_t at intermediate points:

$$\begin{aligned} &\psi_x(t^{\mathrm{R}}_k) = l_{x(t^{\mathrm{R}}_k)}(p^0), \quad \psi_x(t^{\mathrm{L}}_k) = -l_{x(t^{\mathrm{L}}_k)}(p^0), \\ &\psi_t(t^{\mathrm{R}}_k) = l_{t^{\mathrm{R}}_k}(p^0), \qquad \psi_t(t^{\mathrm{L}}_k) = -l_{t^{\mathrm{L}}_k}(p^0), \end{aligned} \qquad k = 1, \ldots, \nu - 1.$$

Here t^{L}_k and t^{R}_k are the optimal values; the index 0 has been omitted to avoid unnecessary clutter.

V. Let us consider the relationship between Problem D and the preceding problems. Problem D is obviously the most general of all the "constant" dimension problems considered above. Indeed, for $t^{\mathrm{L}}_k = t^{\mathrm{R}}_k$, $k = 1, \ldots, \nu-1$ (i.e., when the intervals are joined) Problem D goes into Problem C, which in turn goes into Problem B for $x(t^{\mathrm{L}}_k) = x(t^{\mathrm{R}}_k)$, $k = 1, \ldots, \nu - 1$, and the latter goes into Problem A when $f_k = f$ and $U_k = U$ for all $k = 1, \ldots, \nu$.

We will show that in applications to Problems C, B, and A Theorem 8 generalizes the MP for Problems C, B, and A, respectively. To this end it suffices to show that the assertion is true for Problems C and B, because we know that the MP for Problem B in application to Problem A generalizes the MP for Problem A.

Note that all the MP differ only by the conditions on conjugate variables at intermediate points: in some cases we have discontinuities for ψ_x or ψ_t, while in other cases separate conditions have to be written out for right and left limits.

We will show that the application of Theorem 8 to Problem C produces Theorem 6, i.e., the conditions (f) of Theorem 8 imply the conditions (f) of Theorem 6. Note that in this case only the conditions for ψ_t are different.

The vector p of Problem C includes the "continuous" time instants t_k, whereas the vector p of Problem D includes the "discontinuous" time instants t^{L}_k and t^{R}_k. Therefore, to write Problem C in the form of Problem D, we have to take $t_k = t^{\mathrm{R}}_k$ for all k in the vector p of Problem C. This vector will be denoted by $\hat{p}$, while p denotes the full vector of Problem D.

Let $l^C(\hat{p})$ be the endpoint Lagrange function for Problem C. Since the intervals Δ_k in Problem C are joined, the reformulation of Problem C in the form of Problem D gives rise to additional constraints $t^{\mathrm{R}}_k - t^{\mathrm{L}}_k = 0$, $k = 1, \ldots, \nu - 1$. These equalities correspond to Lagrange multipliers γ_k, $k = 1, \ldots, \nu - 1$, so that the endpoint Lagrange function of Problem D takes the form

$$l^D(p) = l^C(\hat{p}) + \sum_{k=1}^{\nu-1} \gamma_k (t^{\mathrm{R}}_k - t^{\mathrm{L}}_k).$$

Conditions (f) of Theorem 8 for ψ_t thus take the form

$$\psi_t(t_k^{\mathrm{R}}) = l_{t_k}^C(\hat{p}^0) + \gamma_k, \qquad \psi_t(t_k^{\mathrm{L}}) = \gamma_k, \quad k = 1, \dots, \nu - 1. \tag{17}$$

Hence we obtain

$$\triangle\psi_t(t_k^0) = \psi_t(t_k^{\mathrm{R}}) - \psi_t(t_k^{\mathrm{L}}) = l_{t_k}^C(\hat{p}^0), \quad k = 1, \dots, \nu - 1. \tag{18}$$

Conversely, conditions (18) imply conditions (17). Indeed, if (18) hold, we can satisfy (17) by setting $\gamma_k = \psi_t(t_k^{\mathrm{L}}),\ k = 1, \dots, \nu - 1.$ Thus, both conditions are equivalent, and therefore (17) can be replaced with (18). The only danger with this replacement is that the disappearing multipliers γ_k may affect the nontriviality of the entire set of Lagrange multipliers. We will show, however, that there is no such danger.

Lemma 5. *The inequality*

$$\sum_{i=0}^{m} \alpha_i + \sum_{j=1}^{q} |\beta_j| + \sum_{k=1}^{\nu-1} |\gamma_k| > 0$$

is equivalent to the inequality

$$\sum_{i=0}^{m} \alpha_i + \sum_{j=1}^{q} |\beta_j| > 0.$$

Proof. It is sufficient to prove the implication $(\Rightarrow)$. The proof is by contradiction. Let all $\alpha_i = \beta_j = 0.$ Then by transversality conditions $\psi_x(t_0) = 0,$ and by linearity of the conjugate equation in $\psi_x(t)$ it follows that $\psi_x(t) \equiv 0$ on $\Delta_1^0,$ whence again invoking the transversality conditions we obtain $\psi_x(t_1^{\mathrm{R}}) = 0,$ and therefore $\psi_x(t) \equiv 0$ on $\Delta_2^0.$ Then we successively obtain that $\psi_x(t) \equiv 0$ on the entire $\Delta^0.$ Hence it follows that in the interior of each interval Δ_k^0 $\dot{\psi}_t = 0,$ i.e., $\psi_t = \mathrm{const}\,.$

From the transversality conditions it follows that $\psi_t(t_0) = 0,$ whence we obtain that $\psi_t(t) \equiv 0$ on $\Delta_1^0,$ and again by the transversality conditions we obtain $\gamma_1 = 0$ and thus $\psi_t(t_1^{\mathrm{R}}) = 0.$ This leads to $\psi_t(t) \equiv 0$ on $\Delta_2^0,$ and then we successively obtain that $\psi_t(t) \equiv 0$ on the entire Δ^0 and all $\gamma_k = 0$ — a contradiction. Q.E.D.

Thus, the Lagrange multipliers γ_k do not affect nontriviality, and replacement of conditions (17) with conditions (18) does not lead to any loss of information. These multipliers may be simply excluded from the MP.

Conditions (18) are identical with the discontinuity conditions for ψ_t in the MP for Problem $C.$ Therefore, in application to Problem $C,$ Theorem 8 is equivalent to Theorem 6. We can similarly prove that in application to Problem B Theorem 8 is equivalent to Theorem 5, and thus the MP for Problem D is the most general of all previously given MP.

These assertions are particular cases of the following theorem. Assume in Problem D that not all phase components and time may have discontinuities, and discontinuities are possible only at some of the points $t_k.$ In other words, part of these variables are by definition continuous at some points. In fact, this is an intermediate problem between C and $D,$ but we will regard it primarily as Problem $D.$ In the vector p of this problem we should naturally retain only one of the endpoint values of the corresponding component (left or right). For this problem we have the following theorem.

Theorem 7. *If in Problem D the time t (and possibly also some component x_i) are continuous at the intermediate point t_k for some $k,$ then the transversality conditions for the conjugate variable ψ_t (and ψ_{x_i}) at the optimal point t_k^0 are equivalent to the discontinuity condition $\triangle\psi_t(t_k^0) = l_{t_k}(p^0)$ (and respectively the discontinuity condition $\triangle\psi_{x_i}(t_k^0) = l_{x_i(t_k)}(p^0)$).*

Proof is similar to the proof of equivalence of the MP of Problem D in application to Problem C and the MP of Problem C itself, but not for all intermediate points at once — only for the given point t_k^0.

Corollary 1. *If in Problem C (or Problem D with discontinuous time) part of the phase variables are continuous and part are discontinuous, then in this case in order to apply Theorem 6 (or Theorem 8) there is no need to partition the values of the continuous components at the intermediate point into left and right endpoint values and add a condition of their equality to the problem constraints. Instead we need simply replace in Theorem 8 the transversality conditions at the intermediate points for the corresponding conjugate variable with discontinuity conditions.*

Remark 5. Similarly to Problem C we can consider Problem D' with "variable" dimensions of $x,\ u$ and discontinuous time. This is a generalization of Problem D. In the proof of the MP for Problem D we never assumed disjointness of the intervals Δ_k. Therefore, in Problem D' we can assume that these intervals overlap and still derive the corresponding MP (Theorem 10). In this MP, each $\Delta_k^0 = [t^{\mathrm{R}}_{(k-1)},\, t^{\mathrm{L}}_k]$ has its own time variable t_k, and therefore in addition to its own ψ_{x_k} each Δ_k^0 will also have its own ψ_{t_k} and thus its own function $H_k(\psi_{t_k}, \psi_{x_k}, t_k, x_k, u_k)$. A complete formulation of this theorem is left as a simple exercise to the interested reader.

Problem D' for nonsmooth controlled systems (and also for differential inclusions) has been considered by Clarke and Vinter [12]. For a smooth controlled system the MP obtained in [12] is identical with our Theorem 10. However, it was derived in [12] as an independent result, and not as a simple corollary of the known MP.

5. Examples

We present five examples. In Examples 1–4 we assume convexity and compactness in u, and therefore passing to Problems K and applying standard existence theorems (see, e.g., [5]), we can show that the problems are always solvable.

For purposes of further discussion, it is helpful to define a many-valued mapping $\mathrm{Sign}\,(\cdot)\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$,

$$\mathrm{Sign}\,\psi = \begin{cases} \dfrac{\psi}{|\psi|} & \text{for } \psi \neq 0, \\[2ex] B_1(0) & \text{for } \psi = 0, \end{cases}$$

where $|\psi|$ is the standard Euclidean length of the vector ψ and $B_1(0)$ is the closed unit ball in $\mathbb{R}^n$ centered at zero. ($\mathrm{Sign}\,\psi$ is exactly the subdifferential of the function $|\psi|$.)

Example 1 (overcoming an obstacle)**.** Consider the following problem on a fixed time interval $[0,T]$ with a fixed intermediate point $t_1 \in (0,T)$:

$$\dot{x} = u, \quad x \in \mathbb{R}^1, \quad |u| \le b,$$

$$t_0 = 0, \quad t_1 = 1, \quad t_2 = T,$$

$$x(t_0) = 0, \quad x(t_1) \ge m, \quad x(t_2) = 0,$$

$$J = \int\limits_0^T x(t)\,dt \to \min.$$

Starting from $x = 0,$ it is required to return back to zero after overcoming an obstacle of height $m > 0$ at the point t_1 while minimizing the area under the trajectory. We assume that all the parameters are chosen so that the obstacle can be overcome with a margin:

$$b > m, \quad m - b(T-1) < 0. \tag{19}$$

(If at least one of the inequalities reduces to an equality, the solution is substantially simplified.)

To reduce this problem to Problem $A,$ we introduce a supplementary phase variable that follows the equation $\dot{y} = x,$ and write the functional in terminal form:

$$J = y(t_2) - y(t_0) \to \min .$$

The open set $\mathcal{Q}$ is the entire space. Write out the MP for this problem. The Pontryagin function is $H = \psi_x u + \psi_y x + \psi_t,$ the endpoint Lagrange function is

$$l(p) = \alpha_0(y(t_2) - y(t_0)) + \alpha_1(m - x(t_1)) + \beta_0 x(t_0) + \beta_2 x(t_2) + \delta_0 t_0 + \delta_1(t_1 - 1) + \delta_2(t_2 - T).$$

For the optimal process $\alpha_0 \geq 0,$ $\alpha_1 \geq 0,$ $(\alpha_0, \alpha_1, \beta_0, \beta_2, \delta_0, \delta_1, \delta_2) \neq 0,$

- conjugate system:

$$\dot{\psi}_x = -\psi_y, \qquad \dot{\psi}_y = 0, \qquad \dot{\psi}_t = 0;$$

- transversality conditions:

  left endpoint: $\psi_x(0) = \beta_0,$ $\psi_y(0) = -\alpha_0,$ $\psi_t(0) = \delta_0;$

  right endpoint: $\psi_x(T) = -\beta_2,$ $\psi_y(T) = -\alpha_0,$ $\psi_t(T) = -\delta_2;$

- discontinuity conditions for the functions $\psi_x,$ $\psi_y,$ and ψ_t at the point $t_1:$

$$\triangle\psi_x(t_1) = -\alpha_1, \qquad \triangle\psi_y(t_1) = 0, \qquad \triangle\psi_t(t_1) = \delta_1;$$

- complementary slackness conditions: $\alpha_1(m - x(t_1)) = 0;$
- for almost all $t \in [0, T]$ $H = \psi_x u + \psi_y x + \psi_t = 0;$
- maximum condition: for all $t \in [0, T]$

$$\max_{|v| \leq b} (\psi_x v + \psi_y x + \psi_t) = 0, \qquad \text{whence } u(t) \in b \cdot \operatorname{Sign} \psi_x(t).$$

From the conjugate system and the transversality conditions we obtain that $\psi_y(t) \equiv -\alpha_0,$ and almost everywhere $\dot{\psi}_x = \alpha_0,$ i.e., to the left and to the right of $t_1 = 1$ the function ψ_x is linear and $\Delta\psi_x(t_1) = -\alpha_1:$

$$\psi_x(t) = \begin{cases} \alpha_0 t + \beta_0, & t \in [0, 1], \\ \alpha_0 t - \alpha_0 T - \beta_2, & t \in [1, T]. \end{cases}$$

Let us prove that $\alpha_1 > 0$.

Assume that $\alpha_1 = 0$. Then ψ_x is without discontinuity, and thus $\psi_x(t) = \alpha_0 t + \beta_0$ on the entire interval $[0,T]$. If additionally $\alpha_0 = 0$ and $\beta_0 \neq 0$, then $\psi_x = \beta_0 \neq 0$, and therefore $u = \text{const} \neq 0,$. Hence, from $x(0) = 0$ we obtain $x(T) \neq 0$, i.e., the right-hand boundary condition is violated.

If $\alpha_0 = \beta_0 = 0$, then $\psi_y \equiv \psi_x \equiv 0$ and from the equality $H = 0$ we have $\psi_t = 0$. Then $\delta_0 = \delta_1 = \delta_2 = 0$, i.e., all the multipliers vanish — a contradiction.

Thus, $\alpha_0 > 0$, i.e., ψ_x linearly increases on the entire interval $[0,T]$. Then either $\psi_x < 0$, $u = -b$ on $(0,1)$, or $\psi_x > 0$, $u = b$ on $(1,T)$. Both cases are inconsistent with the requirement $x(0) = x(T) = 0$, $x(1) \geq m$.

The assumption $\alpha_1 = 0$ thus leads to a contradiction. Hence, $\alpha_1 > 0$ and from the complementary slackness conditions we obtain $x(t_1) = m$.

We will now show that $\alpha_0 > 0$. If $\alpha_0 = 0$, then the function ψ_x is constant to the left and to the right of the point $t_1 = 1$, and since it has a discontinuity at $t_1 = 1$, then either to the left or to the right of $t_1 = 1$ we have $\psi_x \neq 0$, $u = b$ or $u = -b$. But by (19) neither case is consistent with the conditions

$$x(0) = x(T) = 0, \quad x(1) = m. \tag{20}$$

Thus, $\alpha_0 > 0$, and ψ_x linearly increases both to the right and to the left of the point $t_1 = 1$, experiencing a discontinuity $\Delta\psi_x(1) = -\alpha_1 < 0$ at this point. Furthermore, $\psi_x = 0$ at some points $t' \in (0,1)$ and $t'' \in (1,T)$, where it reverses its sign from minus to plus. (If the sign of ψ_x is constant in at least one of these intervals, then again $u = \pm b$, which, as we know, contradicts conditions (20).)

The switch points t', t'' are easily found from conditions (20). Obviously,

$$t' = \frac{b-m}{2b}, \qquad t'' = \frac{bT+m+b}{2b},$$

the control $u = -b, b, -b, b$ sequentially on the intervals $(0,t')$, $(t',1)$, $(1,t'')$, (t'',T), and the corresponding trajectory $x(t)$ is a piecewise-linear function with corners at the points t', 1, t''. Since this is the only process that satisfies the MP, it is optimal.

Example 2 (maximum velocity at intermediate point)**.** It is required to control the acceleration of a point mass so as to achieve a maximum velocity as some intermediate (not fixed) time instant subject to the given boundary conditions:

$$\ddot{x} = u, \qquad x \in \mathbb{R}^1,$$

$$x(t_0) = 0, \qquad \dot{x}(t_0) = 0, \qquad x(t_2) = 0,$$

$$t_0 = 0, \qquad t_2 - T = 0,$$

$$u \in U = [-1,1],$$

$$\dot{x}(t_1) \to \max .$$

We rewrite this problem as Problem A for $(x,y) \in \mathbb{R}^2$:

$$\dot{x} = y, \qquad x(t_0) = 0, \qquad x(t_2) = 0,$$

$$\dot{y} = u, \qquad y(t_0) = 0,$$

$$t_0 = 0, \qquad t_2 - T = 0,$$

$$u \in U = [-1, 1],$$

$$J = -y(t_1) \to \min .$$

The set $\mathcal{Q}$ is the entire space, the Pontryagin function is $H = \psi_x y + \psi_y u + \psi_t$, the endpoint Lagrange function is

$$l(p) = -\alpha_0 y(t_1) + \beta_1 x(t_0) + \beta_2 y(t_0) + \beta_3 x(t_2) + \gamma_0 t_0 + \gamma_1 (t_2 - T).$$

For the optimal process $\alpha_0 \geq 0$, $(\alpha_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1) \neq 0$,

- conjugate system: $-\dot{\psi}_x = 0$, $-\dot{\psi}_y = \psi_x$, $-\dot{\psi}_t = 0$;
- transversality conditions:

  left endpoint: $\psi_x(t_0) = \beta_1$, $\psi_y(t_0) = \beta_2$, $\psi_t(t_0) = \gamma_0$;
  right endpoint: $\psi_x(t_2) = -\beta_3$, $\psi_y(t_2) = 0$, $\psi_t(t_2) = -\gamma_1$;
- discontinuity conditions: $\triangle\psi_x(t_1) = 0$, $\triangle\psi_y(t_1) = -\alpha_0$, $\triangle\psi_t(t_1) = 0$;
- almost everywhere on $[0, T]$, $H = \psi_x y + \psi_y u + \psi_t = 0$;
- maximum condition: for all $t \in [0, T]$,

$$\max_{|v| \leq 1} (\psi_x y + \psi_y v + \psi_t) = 0, \qquad \text{whence } u(t) \in \operatorname{Sign} \psi_y(t).$$

From the conjugate system and the transversality conditions we obtain

$$\psi_x(t) \equiv \beta_1 = -\beta_3, \qquad \psi_t(t) \equiv \gamma_0 = -\gamma_1 \qquad \text{on the entire interval } [0, T],$$

$$\psi_y(t) = \begin{cases} -\beta_1 t + \beta_2, & t \in [0, t_1], \\ -\beta_1 (t - T), & t \in [t_1, T]. \end{cases}$$

This combined with the discontinuity condition for ψ_y implies that $\beta_2 = \beta_1 T + \alpha_0$, and so

$$\psi_y(t) = \begin{cases} \beta_1 (T - t) + \alpha_0, & t \in [0, t_1], \\ \beta_1 (T - t), & t \in [t_1, T]. \end{cases}$$

We will show that $\beta_1 < 0$. Indeed, if $\beta_1 > 0$, then $\psi_y(t) > 0$, $u(t) \equiv 1$ on the interval $(0, T)$, and thus $y(t) > 0$ everywhere for $t > 0$. Hence $x(T) \neq 0$ — a contradiction.

If $\beta_1 = 0$, then $\psi_x(t) \equiv 0$ on the entire interval $[0, T]$,

$$\psi_y(t) = \begin{cases} \alpha_0, & t \in [0, t_1], \\ 0, & t \in [t_1, T]. \end{cases}$$

Then on the optimal trajectory

$$0 = H = \begin{cases} \alpha_0 + \gamma_0, & t \in [0, t_1], \\ \gamma_0, & t \in [t_1, T], \end{cases}$$

whence $\gamma_0 = \alpha_0 = 0$, and then all the remaining Lagrange multipliers also vanish — a contradiction.

Hence, $\beta_1 < 0$, and then $u(t) \equiv -1$ on $[t_1, T]$. Furthermore, ψ_y linearly increases on $[0, t_1]$, whence it follows that on this interval the control may only switch from -1 to $+1$ and at most once.

We will now show that $\alpha_0 > 0$. Indeed, if $\alpha_0 = 0$, then $\psi_y < 0$ everywhere on $[0, T]$, which given the maximum condition leads to $u \equiv -1$. But then $y < 0$ everywhere on $[0, T]$, and thus $x(T) \neq 0$, a contradiction. Thus, $\alpha_0 > 0$ and therefore ψ_y has a negative discontinuity at the point t_1.

Since $\beta_1 < 0$, we may set $\beta_1 = -1$. Then on the optimal trajectory

$$H = -y + |\psi_y| + \psi_t = 0,$$

whence by constancy of ψ_t on $[0, T]$ and continuity of $y(t)$ at the point t_1 we obtain that $\psi_y(t_1 - 0) = -\psi_y(t_1 + 0)$. Hence it follows that $\psi_y(t_1 - 0) > 0$ and then by linearity of the function $\psi_y \neq \text{const}$ there exists a unique time instant $\theta \in (-\infty, t_1)$, where $\psi_y(\theta) = 0$.

Let us consider two cases:

(i) $\theta \in (0, t_1)$. Set $T - t_1 = \Delta$; then the equality $\psi_y(t_1 - 0) = -\psi_y(t_1 + 0)$ implies that $t_1 - \theta = \Delta$, whence, writing the condition $\int_0^T y(t)\, dt = 0$ for a piecewise-linear $y(t)$ with two corner points, we obtain the equation

$$\Delta^2 + \frac{1}{2}\theta^2 - \theta(\theta + 2\Delta) = 0,$$

i.e., $\theta^2 + 4\Delta\theta - 2\Delta^2 = 0$. Setting $\lambda = \theta/\Delta$, we obtain $\lambda > 0$ and $\lambda^2 + 4\lambda - 2 = 0$, so that $\lambda = -2 + \sqrt{6}$ (the second root $\lambda < 0$ is dropped).

The control has two switching points θ and $t_1 = \theta + \Delta$, and the corresponding process satisfies the MP (a fairly unexpected result!). However, we can show that this process is not optimal in second order by varying the points θ, t_1.

(ii) $\theta \leq 0$. Setting $T - t_1 = \Delta$ we again obtain $T = 2\Delta + \theta$, $t_1 - \theta = \Delta$, but the piecewise-linear $y(t)$ now has only one corner point. For convenience denote $\theta = -\omega < 0$. Since $t_1 + \omega = \Delta$, we have $\omega < \Delta$ and the condition $\int_0^T y(t)\, dt = 0$ is transformed to the equation

$$\Delta^2 - \frac{1}{2}\theta^2 + \theta(\theta + 2\Delta) = 0,$$

i.e., $2\Delta^2 + \omega^2 - 4\omega\Delta = 0$. Setting $\mu = \omega/\Delta$, we obtain $\mu \in (0, 1)$ and $\mu^2 - 4\mu + 2 = 0$, so that $\mu = 2 - \sqrt{2}$ (the second root is $\mu = 2 + \sqrt{2} \notin (0, 1)$). Hence

$$t_1 = T\left(1 - \frac{\sqrt{2}}{2}\right).$$

The control has a single switch point: it switches from $+1$ to -1 at t_1. The corresponding process satisfies the MP and is optimal.

Example 3 (transmission of a light ray through an interface of two media)**.** Two optically different media in the space $\mathbb{R}^n$ are separated by a smooth surface $S = \{x \in \mathbb{R}^n\colon g(x) = 0\}$ without singular points. The two media are isotropic, i.e., the velocity of light at each point depends only on the location of the point and is independent of the direction; we assume that it is defined by a positive differentiable function of x. A ray of light issuing from the point x_0 which is contained in the first medium but does not lie on the surface S eventually hits some point x_2 in the second medium. By Fermat's principle, the ray completes the transition from point x_0 to point x_2 in minimum time. Since the media are heterogeneous, the ray is refracted when it crosses the surface S. It is required to determine to refraction law. (Sussmann [18] examines reflection of light from a hypersurface.) This is a classical problem of geometrical optics that has been considered in many studies by variational calculus techniques. Recently the corresponding models have begun to be treated as variable structure problems. Let us check what our results produce.

We have the following problem:

$$\dot{x} = c_1(x)u \qquad \text{on } \Delta_1 = [t_0, t_1],$$

$$\dot{x} = c_2(x)u \qquad \text{on } \Delta_2 = [t_1, t_2],$$

$$t_0 = 0, \qquad x(t_0) = x_0, \qquad g(x(t_1)) = 0, \qquad x(t_2) = x_2,$$

$$u \in U = B_1(0),$$

$$J = t_2 \to \min,$$

where $g(x_0) < 0$, $g(x_2) > 0$, the time instants $0 < t_1 < t_2$ are not fixed, $c_k(x) > 0$ is the velocity of light in the medium $k = 1, 2$ at the point x. To ensure solvability, we introduce the convex constraint $|u| \le 1$, i.e., the velocity is subject to a "weak" constraint $|\dot{x}| \le c_k(x)$, but for optimal motion the constraint still holds as an equality (we will demonstrate this shortly).

This is a problem of type B. The set $\mathcal{Q}$ is the entire space, the Pontryagin function is $H = c_k(x)(\psi_x, u) + \psi_t$ on Δ_k, the endpoint Lagrange function is

$$l(p) = \alpha_0 t_2 + \gamma_0 t_0 + \beta_0(x(t_0) - x_0) + \beta_1 ga(x(t_1)) + \beta_2(x_2(t_2) - x_2).$$

For the optimal process $\alpha_0 \ge 0$, $(\alpha_0, \gamma_0, \beta_0, \beta_1, \beta_2) \ne 0$,

- conjugate system: $-\dot{\psi}_x = (\psi_x, u)c_k'(x)$ on Δ_k, $-\dot{\psi}_t = 0$;
- transversality conditions:

  at the left endpoint: $\psi_x(0) = \beta_0$, $\psi_t(0) = \gamma_0$;

  at the right endpoint: $\psi_x(t_2) = -\beta_2$, $\psi_t(t_2) = -\alpha_0$;
- discontinuity conditions: $\triangle\psi_{x_1}(t_1) = \beta_1 ga'(x(t_1))$, $\triangle\psi_t(t_1) = 0$;
- almost everywhere on Δ_k, $H = c_k(x)(\psi_x, u) + \psi_t = 0$;

– maximum condition: for all $t \in \Delta_k,\ \ k = 1, 2,$

$$\max_{|v|\le 1} \left(c_k(x)(\psi_x, v) + \psi_t\right) = 0,$$

whence by positivity of $c_k(x)$

$$u(t) \in \operatorname{Sign} \psi_x(t).$$

From the transversality conditions and the discontinuity conditions for ψ_t we obtain that $\psi_t(t) \equiv -\alpha_0$ on $[0, t_2]$, and then the maximum condition gives $c_k(x)|\psi_x| = \alpha_0$ on Δ_k. It is easy to see that $\alpha_0 \ne 0$, as otherwise $\psi_x \equiv 0$ on $[0, t_2]$ and all the Lagrange multipliers vanish—a contradiction. Moreover, $\psi_x(t) \ne 0\ \ \forall t,$ since otherwise $\alpha_0 = 0$, which contradicts the previously established fact $\alpha_0 > 0$.

Setting $\alpha_0 = 1$, we obtain $c_k(x)|\psi_x| = 1$ on each Δ_k, and in particular for $t = t_1$

$$c_1(x_1)|\psi_1| = c_2(x_1)|\psi_2| = 1, \tag{21}$$

where ψ_1 and ψ_2 stand for $\psi_x(t_1 - 0)$ and $\psi_x(t_1 + 0)$ respectively, $x_1 = x(t_1)$, and the optimal control $u = \psi_x/|\psi_x|$ is a piecewise-continuous function with a possible discontinuity at time t_1. Note that $|u| = 1$, and the velocity $\dot{x} = c(x)u$ is always collinear with ψ_x.

Let h be the unit vector codirectional with the gradient $g'(x_1)$ (the gradient is nonzero, because the surface is without singular points), α_1 is the angle between the velocity $\dot{x}(t_1 - 0)$ and the vector $-h$ (incidence angle), and α_2 is the angle between the velocity $\dot{x}(t_1 + 0)$ and the vector h (refraction angle). Let us find the relationship between these angles.

Let $L = \{\bar{x} \in \mathbb{R}^n \colon (h, \bar{x}) = 0\}$ be the subspace orthogonal to h (then $x_1 + L$ is the tangent hyperplane to the surface S at the point x_1). The discontinuity condition for ψ_x at time t_1 implies that $\psi_2 = \psi_1 + \beta h$ for some β (the case $\beta = 0$ is not excluded). The projections of ψ_1 and ψ_2 on the subspace L have the form

$$\operatorname{pr}_L \psi_1 = \psi_1 - (\psi_1, h)h,$$

$$\operatorname{pr}_L \psi_2 = \psi_1 + \beta h - (\psi_1 + \beta h, h)h = \psi_1 - (\psi_1, h)h.$$

They should be identical, because the vector $\psi_2 - \psi_1$ is orthogonal to L. Hence

$$\sin\alpha_1 = \frac{|\operatorname{pr}_L \psi_1|}{|\psi_1|} = \frac{|\psi_1 - (\psi_1, h)h|}{|\psi_1|}, \qquad \sin\alpha_2 = \frac{|\operatorname{pr}_L \psi_2|}{|\psi_2|} = \frac{|\psi_1 - (\psi_1, h)h|}{|\psi_2|}.$$

Note that $\sin\alpha_1$ and $\sin\alpha_2$ are both zero or not zero simultaneously. The fact that they vanish simultaneously corresponds to the case of "through" transmission of the ray through the surface S orthogonally, without refraction. If $\sin\alpha_1 \ne 0$, then the refracted ray is in the plane formed by the incident ray and the normal to the surface, and allowing for (21) we obtain

$$\frac{\sin\alpha_1}{\sin\alpha_2} = \frac{|\psi_2|}{|\psi_1|} = \frac{c_1(x_1)}{c_2(x_1)}, \tag{22}$$

i.e., the ratio of the sine of the incidence angle to the sine of the refraction angle is proportional to the ratio of the velocity of light in the first medium at the given point of the interface to the velocity of light in the second medium at the same point. This is the well-known light refraction law. These conditions make it possible to uniquely determine the velocity vector of the refracted ray given the velocity vector of the incident ray. (Clarke and Vinter [13] considered light refraction with nonsmooth velocities $c_1,\ c_2$ and nonsmooth surface S. The nonsmoothness made their analysis more complex, while the motivation for this generality was not entirely clear.)

Within each medium we have the system of equations

$$\dot{x} = c(x)u, \qquad u = \frac{\psi}{|\psi|}, \qquad c(x)|\psi| = 1, \qquad -\dot{\psi} = c'(x)(\psi, u).$$

Whence

$$|\psi| = \frac{1}{c(x)}, \qquad u = c(x)\psi, \qquad -\dot{\psi} = c'(x)c(x)|\psi|^2 = \frac{c'(x)}{c(x)},$$

and so

$$\dot{x} = c^2(x)\psi, \qquad \ddot{x} = 2c(x)(c'(x), \dot{x})\psi - c^2(x)\frac{c'(x)}{c(x)}.$$

The motion of light in an isotropic medium thus follows the equation

$$\ddot{x} = \frac{2}{c(x)}(c'(x), \dot{x})\, \dot{x} - c(x)c'(x). \tag{23}$$

In the particular case when $c(x)$ is independent of x (a homogeneous medium), we have $c'(x) = 0,$ and then $\ddot{x} = 0,$ i.e., the motion is along straight lines.

Example 4 (fastest traversal of specified points)**.** A point mass is moving on the plane under the action of a bounded force (acceleration). Starting from the point x_0 with the velocity $y_0,$ it has to traverse in the shortest time a given set of s points and arrive at the terminal point x_T with a given velocity y_T. If there are no intermediate points $(s = 0),$ the problem becomes the well-known Feldbaum problem, which was one of the origins of optimal control theory and served as one of the first illustrations of the Pontryagin MP [1]. We thus have the problem

$$\begin{aligned}
&\dot{x} = y, \qquad \dot{y} = u, \qquad x,\, y \in \mathbb{R}^2,\\
&t_0 = 0, \qquad 0 < t_1 < \ldots < t_s < t_{s+1} = T \qquad \text{not fixed},\\
&x(t_0) = x_0, \qquad x(t_k) = x_k, \qquad k = 1, \ldots, s; \quad x(T) = x_T,\\
&y(t_0) = y_0, \qquad y(T) = y_T,\\
&u \in B_1(0) \subset \mathbb{R}^2,\\
&J = T \to \min.
\end{aligned}$$

Here the vector $p = \big((t_0, x_0, y_0),\, (t_1, x_1, y_1),\, \ldots,\, (t_s, x_s, y_s),\, (T, x_T, y_T)\big),$ the open set $\mathcal{Q}$ is the entire space. Let us write the Pontryagin MP for this problem.

The Pontryagin function is $H = \psi_x y + \psi_y u + \psi_t,$ the endpoint Lagrange function is

$$l(p) = \alpha_0 T + \delta t_0 + (\beta_{x_0}, x_0) + (\beta_{y_0}, y_0) + (\beta_{x_T}, x_T) + (\beta_{y_T}, y_T) + \sum_{k=1}^{s} (\sigma_k, x_k).$$

For the optimal process $\alpha_0 \geq 0$, $(\alpha_0, \delta, \sigma_1, \ldots, \sigma_s, \beta_{x_0}, \beta_{y_0}, \beta_{x_T}, \beta_{y_T}) \neq 0$,

– the conjugate system:

$$\dot{\psi}_x = 0, \qquad \dot{\psi}_y = -\psi_x, \qquad \dot{\psi}_t = 0;$$

– the transversality conditions:

at the left endpoint: $\psi_x(t_0) = \beta_{x_0}$, $\psi_y(t_0) = \beta_{y_0}$, $\psi_t(t_0) = \delta$;

at the right endpoint: $\psi_x(T) = -\beta_{x_T}$, $\psi_y(T) = -\beta_{y_T}$, $\psi_t(T) = -\alpha_0$;

– the discontinuity conditions at intermediate points:

$$\triangle\psi_x(t_k) = \sigma_k, \qquad \triangle\psi_y(t_k) = 0, \qquad \triangle\psi_t(t_k) = 0, \qquad k = 1, \ldots, s;$$

– for almost all $t \in [0, T]$, $H = \psi_x y + \psi_y u + \psi_t = 0$;

– the maximum condition: for all $t \in [0, T]$

$$\max_{|v|\leq 1} (\psi_x y + \psi_y v + \psi_t) = 0, \qquad \text{whence } u(t) \in \operatorname{Sign} \psi_y(t).$$

From the conjugate system it follows that for each $\Delta_k = [t_{k-1}, t_k]$, $k = 1, \ldots, s+1$, we have

$$\psi_x(t) = a_k, \quad \psi_y(t) = -a_k t + b_k,$$

where a_k, $b_k \in \mathbb{R}^2$. Moreover, from the discontinuity conditions it follows that the function ψ_y is continuous on the entire interval $[0, T]$, while the transversality and discontinuity conditions for ψ_t give $\psi_t(t) = \delta = -\alpha_0$ on the entire $[0, T]$.

Let us find all the extremals that satisfy the following supplementary assumption:

There exists a time instant $t' \in [0, T]$ *such that* $(\psi_x(t'), y(t')) > 0$.

When this assumption holds, $\alpha_0 = (\psi_x(t'), y(t')) + |\psi_y(t')| > 0$ and we may set $\alpha_0 = 1$. Hence it follows that $\psi_y(t)$ is not identically zero on any of the Δ_k (otherwise it would follow from the conjugate system that $\psi_x(t) \equiv 0$ on some Δ_k, but then $\alpha_0 = 0$ — a contradiction), and using the continuity of ψ_y on the entire $[0, T]$ we obtain that the control

$$u(t) = \frac{\psi_y(t)}{|\psi_y(t)|}$$

is a piecewise-continuous function and the entire process is determined by finitely many parameters: $s + 1$ time instants $t_1, \ldots, t_s$, T and $4(s + 1)$ two-dimensional vectors a_k and b_k, $k = 1, \ldots, s + 1$, defining $\psi_y(t)$. The total number of unknowns is thus $5s + 5$.

Given the MP conditions, we construct the system of nonlinear equations to find these parameters. Assuming a priori that the trajectory originates from the point (x_0, y_0), we count the number of equations: $2s$ joining equations for $\psi_y(t)$ at interior points, $2(s + 1)$ conditions for the traversal of the trajectory through the point x_k, two conditions for the value of $y(T)$, and $s + 1$ conditions that set H^0 equal to zero at the points $t_0, \ldots, t_s$. It is easy to see that when these relationships are satisfied, all the MP conditions are also satisfied. We thus have a total of $5s+5$ equations.

The number of equations is equal to the number of unknowns, and therefore in a typical case the system of equations can be used to find all the Lagrange multipliers and the corresponding extremal.

This system of equations is conveniently written using the vector functions

$$\xi_k(t) = \int\limits_{t_{k-1}}^{t} \frac{-a_k\tau + b_k}{|-a_k\tau + b_k|}\, d\tau, \qquad \eta_k(t) = \int\limits_{t_{k-1}}^{t} \xi_k(\tau)\, d\tau, \quad k = 1, \dots, s+1.$$

Denote $y_k = y(t_k),\ k = 1, \dots, s.$ Then y_s is determined from the recurrences $y_k = y_{k-1} + \xi_k(t_k),$ where y_0 is known.

We obtain the following system of equations:

– joining conditions for ψ_y at the interior points ($2s$ equations);

$$-a_k t_k + b_k = -a_{k+1} t_k + b_{k+1}, \quad k = 1, \dots, s;$$

– arrival condition of $y(t)$ at the point y_T at the time instant T (two equations):

$$y_s + \xi_{s+1}(T) - y_T = 0;$$

– traversal conditions of the trajectory $x(t)$ through the points x_k at times t_k ($2(s+1)$ equations):

$$x_k + y_k(t_{k+1} - t_k) + \eta_{k+1}(t_{k+1}) - x_{k+1} = 0, \quad k = 0, \dots, s;$$

– the conditions $H^0 = 0$ at the points $t_0, \dots, t_s$ ($s+1$ equations).

Note that the last group of equations is equivalent to the following:

$$H^0(t_0) = 0, \qquad H^0(t_k + 0) - H^0(t_k - 0) = 0, \quad k = 1, \dots, s.$$

By continuity of the functions $\psi_y(t)$ and $y(t)$ on $[0, T]$ this system of equations has the form

$$(\psi_x(t_0), y_0) + |\psi_y(t_0)| - 1 = 0, \qquad (\psi_x(t_k + 0) - \psi_x(t_k - 0), y(t_k)) = 0,$$

or in the original variables

$$(a_1, y_0) + |b_1| - 1 = 0, \qquad (a_{k+1} - a_k, y_k) = 0, \quad k = 1, \dots, s.$$

Thus, the problem of finding the MP extremals has been reduced to solving a system of nonlinear equations. The general case when the new assumption does not hold requires a more detailed analysis.

Example 5 (optimal preparation for exams)**.** A student has to sit for ν exams scheduled at specific dates. In each exam the student will receive a grade that reflects his level of preparation for the particular exam; in the end all the grades are summed. The level of preparation for each exam depends both on the effort that the student put into studying the particular subject and on the general level of culture, which in turn depends on the cumulative effort up to the given time instant. The total effort for the entire exam period is limited. It is required to distribute the effort in a way that will maximize the sum of grades in all the exams. This problem can be formalized in the following way.

Given are the time instants t_k, $k = 0, \dots, \nu$. Denote by Δ_k the interval $[t_{k-1}, t_k]$, $k = 1, \dots, \nu$, and consider the following optimal control problem for $(x, z) \in \mathbb{R}^2$:

$$\dot{x} = -r_k x + \sqrt{w} \qquad \text{on } \Delta_k,$$

$$\dot{z} = -\rho z + \varepsilon_k \sqrt{w} \qquad \text{on } \Delta_k, \quad k = 1, \dots, \nu,$$

$$x(t_0) = x_0, \qquad z(t_0) = z_0, \qquad x(t_k + 0) = x_k^0 + z(t_k), \quad k = 1, \dots, \nu - 1,$$

$$w(t) \geq 0, \qquad \int_{t_0}^{t_\nu} w(t)\, dt \leq E, \qquad J = \sum_{k=1}^{\nu} x(t_k - 0) \to \max,$$

where $r_k > 0$, $\varepsilon_k > 0$; $k = 1, \dots, \nu$, $\rho > 0$, $x_0 > 0$, $z_0 > 0$, $E > 0$.

Here x represents the level of knowledge in the subject of the nearest exam, z is the general level of knowledge, and the control w is the effort in preparing for the nearest exam. The variable $x(t)$ may have discontinuities at the exam times t_k, while the function $z(t)$ is continuous on the entire interval $[t_0, t_\nu]$. Given the effort w, both knowledge levels increase in proportion to $\sqrt{w}$; according to neoclassical views, this reflects decreasing marginal efficiency of each additional unit of effort.

Introducing for convenience the control $u(t) = \sqrt{w(t)}$ and rewriting the functional in terminal form, we obtain a canonical Problem C with phase variables $(x, z, y) \in \mathbb{R}^3$:

$$\begin{cases} \dot{x} = -r_k x + u \qquad \text{on } \Delta_k, \\ \dot{z} = -\rho z + \varepsilon_k u \qquad \text{on } \Delta_k, \quad k = 1, \dots, \nu, \\ \dot{y} = u^2, \qquad u \in U = [0, +\infty), \end{cases}$$

$$\begin{cases} x(t_0) = x_0, \qquad z(t_0) = z_0, \\ t_k = t_k^0, \quad k = 0, 1, \dots, \nu, \\ x(t_k + 0) = x_k^0 + z(t_k), \quad k = 1, \dots, \nu - 1, \\ y(t_\nu) - y(t_0) \leq E, \\ J = -\sum_{k=1}^{\nu} x(t_k - 0) \to \min . \end{cases}$$

The vector p has the form

$$p = \big((t_0, x_0, z_0, y_0),\ (t_1, x(t_1 - 0), x(t_1 + 0), z(t_1), y(t_1)), \ldots,$$

$$(t_{\nu-1}, x(t_{\nu-1} - 0), x(t_{\nu-1} + 0), z(t_{\nu-1}), y(t_{\nu-1})),\ (t_\nu, x(t_\nu - 0), z(t_\nu), y(t_\nu))\big).$$

The existence of a solution in this problem follows from linearity of the controlled system in the phase variables (x, z), convexity of the control set, and boundedness of the squared control integral (see [5]). Let us write the MP. The Pontryagin function is

$$H = \psi_x(-r_k x + u) + \psi_z(-\rho z + \varepsilon_k u) + \psi_y u^2 + \psi_t \qquad \text{on } \Delta_k,$$

the endpoint Lagrange function is

$$l(p) = \alpha_0 \left(-\sum_{k=1}^{\nu} x(t_k - 0) \right) + \alpha_1\big(y(t_\nu) - y(t_0) - E\big) + \sigma_0(x(t_0) - x_0)$$

$$+ \beta_{z_0}(z(t_0) - z_0) + \sum_{k=0}^{\nu} \beta_{t_k}(t_k - t_k^0) + \sum_{k=1}^{\nu-1} \sigma_k\big(x(t_k + 0) - x_k^0 - z(t_k)\big).$$

By the maximum principle, the tuple

$$\lambda = (\alpha_0, \alpha_1, \beta_{z_0}, \beta_{t_k},\ k = 0, \ldots, \nu; \sigma_k,\ k = 0, \ldots,\ \nu - 1) \neq 0,$$

exists for the optimal process and satisfies the following conditions:

– nonnegativity conditions: $\alpha_0 \geq 0,\ \ \alpha_1 \geq 0$;

– complementary slackness conditions: $\alpha_1\big(y(t_\nu) - y(t_0) - E\big) = 0$;

– conjugate equations:

$$\dot{\psi}_x = r_k \psi_x, \quad \dot{\psi}_z = \rho \psi_z, \quad \dot{\psi}_y = 0, \quad \dot{\psi}_t = 0 \qquad \text{on } \Delta_k;$$

– transversality conditions at the endpoints:

$$\psi_x(t_0 + 0) = \sigma_0, \quad \psi_z(t_0 + 0) = \beta_{z_0}, \quad \psi_y(t_0 + 0) = -\alpha_1, \quad \psi_t(t_0 + 0) = \beta_{t_0};$$

$$\psi_x(t_\nu - 0) = \alpha_0, \quad \psi_z(t_\nu - 0) = 0, \qquad \psi_y(t_\nu - 0) = -\alpha_1, \quad \psi_t(t_\nu - 0) = -\beta_{t_\nu};$$

– discontinuity conditions for $\psi_z,\ \psi_y,\ \psi_t$ and transversality conditions for ψ_x at intermediate points:

$$\triangle\psi_z(t_k) = -\sigma_k, \qquad \triangle\psi_y(t_k) = 0, \qquad \triangle\psi_t(t_k) = \beta_{t_k}, \quad k = 1, \ldots, \nu - 1;$$

$$\psi_x(t_k - 0) = \alpha_0, \qquad \psi_x(t_k + 0) = \sigma_k, \quad k = 1, \ldots, \nu - 1;$$

– for every $k = 1, \dots, \nu$ for almost all $t \in \Delta_k$

$$H = \psi_x(-r_k x + u) + \psi_z(-\rho z + \varepsilon_k u) + \psi_y u^2 + \psi_t = 0;$$

– maximum condition: for every $k = 1, \dots, \nu$ and all $t \in \Delta_k$

$$\max_{v \geq 0} \big(\psi_x(-r_k x + v) + \psi_z(-\rho z + \varepsilon_k v) + \psi_y v^2 + \psi_t\big) = 0.$$

From the conjugate system, the transversality conditions, and the discontinuity conditions for ψ_y we obtain that $\psi_y(t) \equiv -\alpha_1$. The maximum condition on Δ_k thus takes the form

$$\max_{v \geq 0} \big(-\alpha_1 v^2 + (\psi_x + \varepsilon_k \psi_z)v - r_k \psi_x x - \rho \psi_z z + \psi_t\big) = 0.$$

From the conjugate system we also obtain that on each $\Delta_k = [t_{k-1},\, t_k]$

$$\psi_x(t) = \psi_x(t_{k-1} + 0)e^{r_k t}, \qquad \psi_z(t) = \psi_z(t_{k-1} + 0)e^{\rho t}.$$

Since $\psi_x(t_k - 0) = \alpha_0 \geq 0$, we have $\psi_x(t) \geq 0$ on each Δ_k and thus on the entire interval $[t_0, t_\nu]$. Hence it follows that all $\sigma_k = \psi_x(t_k + 0) \geq 0$. Then from the endpoint condition $\psi_z(t_\nu - 0) = 0$ and the discontinuity conditions $\Delta\psi_z(t_k) = -\sigma_k \leq 0$, we obtain that everywhere $\psi_z(t) \geq 0$.

We will now prove that $\alpha_1 \neq 0$. Indeed, if $\alpha_1 = 0$, then $\psi_y \equiv 0$ and from the maximum condition we obtain $\psi_x + \varepsilon_k \psi_z \leq 0$. But we have proved that $\psi_x \geq 0$ and $\psi_z \geq 0$, whence $\psi_x = \psi_z = 0$, and from $H^0 = 0$ we obtain that $\psi_t = 0$ and all Lagrange multipliers vanish — a contradiction.

Thus, $\alpha_1 > 0$ and by the complementary slackness conditions we obtain

$$y(t_\nu) - y(t_0) = E. \tag{24}$$

From the maximum condition, noting that $\alpha_1 > 0$ and $\psi_x + \varepsilon_k \psi_z \geq 0$, we obtain

$$u = \frac{\psi_x + \varepsilon_k \psi_z}{2\alpha_1} \qquad \text{on } \Delta_k.$$

We will now show that $\alpha_0 \neq 0$. Let $\alpha_0 = 0$. Then $\psi_x(t_k - 0) = 0$ for all $k = 1, \dots, \nu$, and therefore $\psi_x(t) \equiv 0$ and all $\sigma_k = 0$. Using the endpoint condition $\psi_z(t_\nu - 0) = 0$ and the discontinuity conditions for ψ_z, we obtain $\psi_z(t) \equiv 0$. Then $u(t) \equiv 0$ and thus $y(t_\nu) - y(t_0) = 0$, which contradicts equality (24).

Thus, $\alpha_0 > 0$ and we may take $\alpha_0 = 1$. Then on all Δ_k we uniquely determine the function $\psi_x(t)$ and all the numbers σ_k; from them we determine $\psi_z(t)$ on all the intervals Δ_k (starting with the last one and moving in reverse direction), and the control $u(t)$ is thus determined apart from a multiplier α_1. This multiplier is uniquely obtained from condition (24). The extremal is thus completely determined.

We would like to thank B. M. Miller and V. A. Dykhta for useful discussions and suggestions regarding some publications on this subject.

The study has been supported by the Russian Foundation for Basic Research (grant 04-01-00482) and by the Program for Leading Scientific Schools (grant NSh-304.2003.1).

REFERENCES

1. L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *Mathematical Theory of Optimal Processes* [in Russian], Nauka, Moscow (1961).
2. G. A. Bliss, *Lectures in Variational Calculus* [Russian translation], IL, Moscow (1950).
3. A. Ya. Dubovitskii and A. A. Milyutin, "The theory of maximum principle," in: *Methods of the Theory of Extremum Problems in Economics* [in Russian], Nauka, Moscow (1981).
4. A. A. Milyutin, "Maximum principle for regular systems," in: *Necessary Condition in Optimal Control* [in Russian], Nauka, Moscow (1990).
5. A. D. Ioffe and V. M. Tikhomirov, *Theory of Extremum Problems* [in Russian], Nauka, Moscow (1974).
6. V. M. Alekseev, V. M. Tikhomirov, and S. V. Fomin, *Optimal Control* [in Russian], Nauka, Moscow (1979).
7. A. A. Milyutin and N. P. Osmolovskii, *Calculus of Variations and Optimal Control*, AMS (1998).
8. M. I. Kamien and N. L. Schwartz, *Dynamic Optimization. The Calculus of Variations and Optimal Control in Economics and Management*, North-Holland (1981).
9. C. H. Denbow, *A Generalized Form of the Problem of Bolza*, Dissertation, University of Chicago Press (1937).
10. Yu. M. Volin and G. M. Ostrovskii, "The maximum principle for discontinuous systems and its application to problems with phase constraints," *Izv. Vuzov, Radiofizika*, **12**, No. 11, 1609–1621 (1969).
11. L. T. Ashchepkov, *Optimal Control of Discontinuous Systems* [in Russian], Nauka, Novosibirsk (1987).
12. F. H. Clarke and R. B. Vinter, "Optimal multiprocesses," *SIAM J. Control Optimization*, **27**, No. 5, 1072–1091 (1989).
13. F. H. Clarke and R. B. Vinter, "Application of optimal multiprocesses," *SIAM J. Control Optimization*, **27**, No. 5, 1048–1071 (1989).
14. K. Tomiyama and R. J. Rossana, "Two-stage optimal control problems with an explicit switch point dependence," *J. Econ. Dynam. Control*, **13**, No. 3, 319–337 (1989).
15. A. V. Arutyunov and A. I. Okoulevich, "Necessary optimality conditions for optimal control problems with intermediate constraints," *J. Dynam. Control Sys.*, **4**, No. 1, 49–58 (1998).
16. V. A. Dykhta and O. N. Samsonyuk, *Optimal Impulse Control with Applications* [in Russian], Fizmatlit, Moscow (2003).
17. B. M. Miller and E. Ya. Rubinovich, *Optimization of Dynamical Systems with Impulse Controls* [in Russian], Fizmatlit, Moscow (2005).
18. H. J. Sussmann, "A maximum principle for hybrid optimal control problems," *Proc. 38th IEEE Conf. Decision and Control*, Phoenix (1999).
19. M. Garavello and B. Piccoli, "Hybrid necessary principle," *SIAM J. Control Optimization*, **43**, No. 5, 1867–1887 (2005).
20. V. G. Boltyanski, "The maximum principle for variable structure systems," *Int. J. Control*, **77**, No. 17, 1445–1451 (2004).
21. E. R. Smol'yakov, *Unknown Pages in the History of Optimal Control* [in Russian], URSS, Moscow (2002).
22. A. A. Milyutin, A. V. Dmitruk, and N. P. Osmolovskii, *Maximum Principle in Optimal Control* [in Russian], Izd. MGU, Moscow (2004).
23. A. V. Dmitruk and A. M. Kaganovich, "The hybrid maximum principle is a consequence of Pontryagin maximum principle," *Syst. Contr. Lett.*, **57**, 964–970 (2008).