Control and Cybernetics

vol. **38** (2009) No. 4B

Lagrange principle and necessary conditions*

by

Vladimir Tikhomirov

Department of Mechanics and Mathematics, Moscow State University
Vorobyevy Gory, Moscow 119899, Russia
e-mail: vmtikh@googlemail.com

Abstract: Necessary conditions of extremum (from the times of Fermat and Lagrange till our times) for extremal problems where smoothness is interlaced with convexity, and some type of regularity takes place, correspond to a unique general principle, which is due to Lagrange. This report is devoted to the Lagrange principle in the theory of optimization.

Keywords: mathematical programming, convex analysis, calculus of variations, optimal control, Lagrange multiplier rule.

Introduction

This paper[1] represents a fragment of my lecture "On some methods, calculi, phenomena, principles and results in the theory of extremum" given at the Conference in Będlewo (Poland) "50 years of Optimal Control, September 2008".

An idea of the way to solve problems with equality constraints was expressed by Lagrange. He wrote:

Here we only sketch these procedures and it will be easy to apply them, but one can reduce them to this general principle: If a function of several variables should be maximum or minimum, and there are between these variables one or several equations, then it will suffice to add to the proposed function the functions that should be zero, each multiplied by an undetermined quantity and then to look for the maximum or the minimum as if the variables were independent; the equations that one will find, combined with the given equations, will serve to determine all the unknowns.

J.-L. Lagrange

"Théorie des fonctions analytique", Paris, 1797

*Submitted: January 2009; Accepted: November 2009.

[1]It has been written in collaboration with G. Magaril-Il'yaev. The research was carried out with the financial support of the Russian Foundation for Basic Research (Grants 08-01-00450 and 08-01-90001) and the President Grant for State Support of Leading Scientific Schools in Russian Federation (Grant NSH-3233.2008.1).

The Lagrange principle of eliminating the constraints:

According to this principle, when one searches a necessary condition of an extremal problem with equality constraints in which smoothness is interlaced with convexity, it is sufficient to construct the Lagrange function of the problem and then to apply necessary conditions for a minimum of the Lagrange function **"as if the variables were independent"**.

Now we will translate this heuristic idea to the mathematical language.

1. The Lagrange principle for smooth-convex problems

Definition Let X and Z be normed spaces and V be a neighborhood of a point $\widehat{x} \in X$[2]. A mapping $F : V \to Z$ is said to be *strictly differentiable at a point* $\widehat{x}$ (and we write in this case $F \in SD^1(\widehat{x})$) if there is a linear continuous operator $M : X \to Z$ such that for any $\varepsilon > 0$ there exists $\delta > 0$, such that $\|F(x_2) - F(x_1) - M(x_2 - x_1)\|_Y \leq \varepsilon\|x_2 - x_1\|_Z$ whenever $\|x_i - \widehat{x}\|_X < \delta,\ i = 1, 2$. The operator M is the Fréchet derivative of the mapping F at the point $\widehat{x}$, denoted $F'(\widehat{x})$. The mapping F is said to be *regular (weakly regular)*, if the operator $F'(\widehat{x})$ is surjective (if $F'(\widehat{x})X$ is a closed subspace of finite codimension).

Statement of the smooth-convex problem

Let X and Y be normed spaces, V neighborhood of a point $\widehat{x} \in X$, $\mathcal{U}$ a set; and let a functional $f_0 : V \to \mathbb{R}$ and a mapping $F : V \times \mathcal{U} \to Y$ be given. Consider the problem

$$f_0(x) \to \min; \quad F(x, u) = 0. \tag{P}$$

The constraint $F(x, u) = 0$ is said to be *smooth-convex* at $(\widehat{x}, \widehat{u})$, if the mapping $x \to F(x, \widehat{u})$ is strictly differentiable at the point $\widehat{x}$, and subsets $F(x,\mathcal{U}) \subset Y$ are convex for all $x \in V$. If the functional f_0 in (P) is smooth and the mapping F is a smooth-convex, then (P) we call a *smooth-convex problem.*

If the constraint $F = 0$ is absent, the problem (P) is called *the problem without constraints* or *an elementary smooth problem.*

Proposition 1 *If $\widehat{x}$ is a local extremum in the problem (P) without constraints and f_0 is differentiable at $\widehat{x}$, then $f_0'(\widehat{x}) = 0$.*

This result (which trivially follows from definitions) is called *Fermat theorem.*

A pair $(\widehat{x}, \widehat{u})$ is said to afford a *strong local minimum* in the problem (P), if there exists $\varepsilon > 0$ such that for any pair$(x, u) \in V \times \mathcal{U}$ satisfying the condition $F(x, u) = 0$ and $\|x - \widehat{x}\|_X < \varepsilon$ the inequality $f_0(x) \geq f_0(\widehat{x})$ holds.

[2]Sometimes we shall write in this case $V \in \mathcal{O}(\widehat{x}, X)$

The function $\mathcal{L}((x,u),\bar{\lambda}) = \lambda_0 f_0(x) + \langle \lambda, F(x,u)\rangle$ is referred to as the *Lagrange function of problem (P),* and $\bar{\lambda} = (\lambda_0, \lambda) \in \mathbb{R}_+ \times Y^*$ is referred to as *the collection of Lagrange multipliers.*

According to the Lagrange's idea, if $(\widehat{x}, \widehat{u})$ is a strong local minimum in (P), then we have to consider two problems: $\mathcal{L}((x,\widehat{u}),\overline{\lambda}) \to \min$ and $\mathcal{L}((\widehat{x},u),\overline{\lambda}) \to \min, u \in \mathcal{U}$. The first one is a smooth problem without constraints and according to Fermat theorem at the point $\widehat{x}$ of local minimum the equality

$$\mathcal{L}_x((\widehat{x},\widehat{u}),\overline{\lambda}) = 0 \Leftrightarrow \lambda_0 f_0'(\widehat{x}) + (F_x(\widehat{x},\widehat{u}))^* \lambda = 0, \qquad (a)$$

holds. The equality (a) is called *a stationarity condition.*

The second problem is in fact a convex problem $\langle \lambda, y\rangle \to \min$, $y = F(\widehat{x}, u)$, $u \in \mathcal{U}$. The condition of minimum at the point $\widehat{u}$ can be expressed analytically, but we prefer to write the following tautological condition

$$\min_u \mathcal{L}((\widehat{x},u),\overline{\lambda}) = \mathcal{L}((\widehat{x},\widehat{u}),\overline{\lambda}) \Leftrightarrow \min_u \langle \lambda, F(\widehat{x},u)\rangle = 0. \qquad (b)$$

We call it *a minimum condition.*

Now we formulate a necessary condition of minimum in the problem (P).

The main theorem (Lagrange principle for smooth-convex problem) (see Ioffe and Tikhomirov, 1979). *Let X and Y be Banach spaces, the function f_0 be differentiable at $\widehat{x}$, $F = 0$ a smooth-convex constraint and the mapping $x \to F(x,\widehat{u})$ weakly regular at the point $(\widehat{x},\widehat{u})$. Then the necessary condition of a strong local minimum in (P) at the point $(\widehat{x},\widehat{u})$ corresponds to the Lagrange principle, i.e. there exists a nontrivial collection of Lagrange multipliers $\overline{\lambda} = (\lambda_0,\lambda) \in \mathbb{R}_+ \times Y^*$, for which the stationarity condition (a) and the minimum condition (b) are satisfied.*

2. Proof

Preliminaries from functional analysis. In the proof we will use the Banach open mapping principle and separation theorems (see Ioffe and Tikhomirov, 1979). The set of all linear continuous operators from a normed space X to a normed space Y is denoted $\mathcal{L}(X,Y)$.

Lemma 1 (on right inverse mapping) *Let X and Y be Banach spaces and $\Lambda \in \mathcal{L}(X,Y)$ be a surjective operator. Then there exists an operator $R\colon Y \to X$ (right inverse of Λ) and a constant γ such that $\Lambda R(y) = y$ and $\|R(y)\|_X \le \gamma \|y\|_Y$ for all $y \in Y$.*

Proof of Lemma 1. Let $U_X(0,1) = \{x \in X \mid \|x\|_X < 1\}$ be an open unit ball. According to Banach open mapping principle $\Lambda U_X(0,1)$ contains a ball $U_Y(0,\delta)$. Thus, for each element $y \in U_Y(0,\delta)$ there exists an element $x(y) \in U_X(0,1)$, such that $\Lambda x(y) = y$. Hence, the mapping $R(y) = \frac{2\|y\|_Y}{\delta} x(\frac{\delta y}{2\|y\|_Y})$ satisfies the conditions of the lemma. ∎

Lemma 2 (on nontriviality of the annihilator) *Let X be a normed space and L be its closed proper subspace. Then the annihilator $L^{\perp}$ of L has a nonzero element.*

This lemma trivially follows from the separation theorem.

Lemma 3 (on closedness of the image) *Let X and Y be Banach spaces, $M \in \mathcal{L}(X, \mathbb{R}^n)$, $\Lambda \in \mathcal{L}(X, Y)$, $\Lambda X = Y$ $Nx = (Mx, \Lambda x)$. Then* ImN *is closed in* $\mathbb{R}^n \times Y$.

Proof of Lemma 3. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence such that $Mx_n \to \xi,\ \Lambda x_n \to \eta$. Denote $x'_n = R(\Lambda x_n - \eta)$, where R is the right-inverse mapping for Λ. Then $\|x'_n\|_X \le \gamma\|\Lambda x_n - \eta\|_Y \to 0$. We see that $\Lambda(x_n - x'_n) = \eta$ and at the same time $\xi = \lim\limits_{n\to\infty} M(x_n - x'_n)$. Thus, ξ belongs to the affine manifold $A = \{Mx \mid \Lambda x = \eta\} \in \mathbb{R}^n$, which is a closed set. Thus, there exists $\overline{x}$ such that $M\overline{x} = \xi$, i.e. $(\xi, \eta) = (M\overline{x}, \Lambda\overline{x})$. ■

Lemma 4 (on annihilator of a kernel of a regular operator) *The annihilator of a linear continuous surjective operator from one Banach space to another coincides with the image of the conjugate operator.*

Proof of Lemma 4. The inclusion Im$\Lambda^* \subset (\text{Ker}\Lambda)^{\perp}$ follows from definition. Let $\Lambda \in \mathcal{L}(X, Y)$ and $x^* \in (\text{Ker}\,\Lambda)^{\perp}$. Consider the operator $Mx = (\langle x^*, x\rangle, \Lambda x) \in \mathcal{L}(X, \mathbb{R}\times Y)$. It follows from Lemma 3 that MX is a closed subset in $\mathbb{R}\times Y$. It is the proper subspace (because $(1, 0) \notin MX$). From Lemma 2 there exists an element $(\lambda_0, \lambda) \in \mathbb{R}\times Y^* \setminus (0, 0)$ such that $\lambda_0\langle x^*, x\rangle + \langle\lambda, \Lambda x\rangle = 0\ \forall x \in X$. The operator Λ is a surjective operator, thus $\lambda_0 \ne 0$, so we obtain: $x^* = -\Lambda^*\frac{\lambda}{\lambda_0}$. ■

Modified Newton method and a theorem on right-inverse mapping. Let X and Y be Banach spaces, $V \in \mathcal{O}(x_0, X)$, $F : V \to Y$, $\Lambda \in \mathcal{L}(X, Y)$, $\Lambda X = Y$. The sequence

$$x_n = x_{n-1} + R(y - F(x_{n-1})),\ n \in \mathbb{N} \tag{A}$$

where R is a right-inverse mapping to Λ, is called *a modified Newton's sequence,* and application of it is called *a modified Newton method.*

Theorem (on right-inverse mapping). *Let V be a neighborhood of $\widehat{x}$ in X and $F : V \to Y$. If there exist a linear continuous surjective operator Λ from X onto Y and a number θ, $0 < \theta < 1$, such that for all pairs $(x', x) \in V$*

$$\|F(x') - F(x) - \Lambda(x' - x)\|_Y \le \frac{\theta}{\gamma}\|x' - x\|_X \tag{B}$$

holds where γ is a constant from the right-inverse map lemma, then there exists an open neighborhood W of $F(\widehat{x})$ in Y and a map $\varphi\colon W \to V$, a constant $K > 0$ such that $F(\varphi(y)) = y$ for all $y \in W$ and $\|\varphi(y) - \widehat{x}\|_X \le K\|y - F(\widehat{x})\|_Y$ for all $y \in W$.

Proof. Let $\delta > 0$ be so small that the closed ball $B_X(\widehat{x}, \delta) = \{x \in X \mid \|x - \widehat{x}\|_X \le \delta\}$ (with center $\widehat{x}$ and radius δ) belongs to V and $y \in B_Y(F(\widehat{x}), \delta_0)$, $\delta_0 \le \frac{(1-\theta)\delta}{\gamma}$. Let us prove that a) all elements $\{x_k\}_{k\in\mathbb{N}}$ of the modified Newton's sequence ($x_0 = \widehat{x}$) belong to $B_X(x_0, \delta)$ and b) that this sequence is fundamental. We prove proposition $a)$ by induction over n. The element x_0 belongs to $B_X(x_0, \delta)$ by definition. Let $x_k \in B_X(x_0, \delta)$, $1 \le k \le n$. The equality $\Lambda(x_k - x_{k-1}) - y + F(x_{k-1}) = 0$ $1 \le k \le n$, (i), follows from (A) and, besides, the equality $\Lambda R(y) = y$ holds. Thus ("An" means here "analogously") $\|x_{n+1} - x_n\|_X \overset{(A)}{=} \|R(y - F(x_n))\|_Y \overset{(i),\text{Lemma1}}{\le} \gamma\|y - F(x_n) - y + F(x_{n-1}) + \Lambda(x_n - x_{n-1})\|_Y \overset{(B)}{\le} \theta\|x_n - x_{n-1}\|_X \overset{An}{\le} \theta^2\|x_{n-1} - x_{n-2}\|_X \overset{An}{\le} \dots \overset{An}{\le} \theta^n\|x_1 - x_0\|_X$ (ii). In the subsequent calculations (iii) means the triangle inequality, (iv) is the formula for the sum of geometrical progression: $\|x_{n+1} - x_0\|_X \overset{(iii)}{\le} \|x_{n+1} - x_n\|_X + \ldots + \|x_1 - x_0\|_X \overset{(ii)}{\le} (\theta^n + \theta^{n-1} + \ldots + 1)\|x_1 - x_0\|_X \overset{(A),(iv)}{\le} \gamma/(1-\theta)\|y - F(x_0)\|_Y \overset{\text{def}\delta_0}{\le} \delta$ (v).

We see that elements x_n are defined for all n.

Let us prove b). We have for all $n, m \in \mathbb{N}$: $\|x_{n+m} - x_n\|_X \overset{(iii)}{\le} \|x_{n+m} - x_{n+m-1}\|_X + \ldots + \|x_{n+1} - x_n\|_X \overset{(ii)}{\le} (\theta^{n+m-1} + \ldots + \theta^n)\|x_1 - x_0\|_X \overset{(iv),(A)}{\le} \frac{\gamma\theta^n}{1-\theta}\|y - F(x_0)\|_Y \overset{\text{def}\delta_0}{\le} \delta\theta^n$ (vi).

Consequently, $\{x_n\}_{n\in\mathbb{N}}$ is a fundamental sequence. Denote $\lim_{n\in\mathbb{N}} x_n = \varphi(y)$. From (i) and continuity of F in $B_X(x_0, \delta))$ we obtain the equality $F(\varphi(y)) = y$. The inequality $\|\varphi(y) - x_0\|_X \le K\|y - F(x_0)\|_Y$ with $K = \frac{\gamma}{1-\theta}$ follows from (vi). ∎

Proof of the main theorem. The notion of smoothness is related to analysis, whereas that of convexity is related to geometry. The proof of the Lagrange principle will consists of three parts, one of which is analytic, based on the theorem on right-inverse mapping and the other two are geometric, based on separation theorems.

Denote $\Lambda := F_x(\widehat{x}, \widehat{u})$, $Y_1 := \text{Im}\Lambda$, $Z = Y/Y_1$. By condition, $\dim Z < \infty$. Let $\pi : Y \to Y/Y_1$ be the canonical projection, $C := Y_1 + F(\widehat{x}, \mathcal{U})$.

We will distinguish between two cases: degenerate, where either $\text{int}\,\pi(C) = \emptyset$ or $\text{int}\pi(C) \ne \emptyset$, $0_Z \notin \text{int}\,\pi(F(\widehat{x}, \mathcal{U})$ and nondegenerate, where $\text{int}\pi(C) \ne \emptyset$ and $0_Z \in \text{int}\,\pi(F(\widehat{x}, \mathcal{U})$.

Degenerate case: Here we use geometry. From the finite-dimensional separation theorem it follows that there exists a vector $z^* \in Z^* \setminus 0_{Z^*}$ such that $\langle z^*, z\rangle \ge 0$ $\forall z \in \pi(F(\widehat{x}, \mathcal{U}))$ (i).

Denote by π^* the conjugate operator $\pi^* : Z^* \to Y^*$ and $\lambda = \pi^* z^*$. It is evident that $\lambda \ne 0$ (because π is a surjective operator) and then $\langle \lambda, \Lambda x + F(\widehat{x}, \mathcal{U})\rangle \overset{\text{Id}}{=} \langle \pi^* z^*, \Lambda x + F(\widehat{x}, \mathcal{U})\rangle \overset{\text{Id}}{=} \langle z^*, \pi(\Lambda x + F(\widehat{x}, \mathcal{U}))\rangle \overset{\text{(i)}}{\ge} 0$ (ii).

From this inequality we obtain that $\Lambda^*\lambda = 0$ and $\langle \lambda, F(\widehat{x}, u)\rangle \geq 0\ \forall u \in \mathcal{U})$, i.e. the stationarity condition a) and minimum condition b) with $\overline{\lambda} = (0, \lambda)$ hold true.

Nondegenerate case: $0_Z \in \operatorname{int} \pi C$. Here we use differential calculus. From the equality $\operatorname{span} \pi\, C = Z$ it follows that there exist $m \in \mathbb{N}$ and m elements $\{z_j\}_{j=1}^m$, $z_j \in \pi F(\widehat{x}, v_j)$ such that the conic hull of $\{z_j\}_{j=1}^m$ is Z and there exist m positive numbers $\{\bar{\beta}_i\}_{i=1}^m$ such that $\sum_{i=1}^m \bar{\beta}_i z_i = 0 \Rightarrow \exists \xi :$ $\Lambda\xi + \sum_{i=1}^m \bar{\beta}_i F(\widehat{x}, v_i) = 0$ (iii).

Let v be an element of $\mathcal{U}$ such that $F(\widehat{x}, v) \in Y_1$. Then, there exists $x_v \in X$ such that $\Lambda x_v + F(\widehat{x}, v) = 0$ (iv).

Define the mapping $\Phi : (V \setminus \widehat{x}) \times \mathbb{R} \times \mathbb{R}^m \to Y$ by the formula: $\Phi(x, \alpha, \beta) = (1 - \alpha - \alpha\varepsilon \sum_{j=1}^m \bar{\beta}_j - \sum_{j=1}^m \beta_j))F(\widehat{x} + x, \widehat{u}) + \alpha F(\widehat{x} + x, v) + \alpha\varepsilon \sum_{j=1}^m \bar{\beta}_j F(\widehat{x} + x, v_j) + \sum_{j=1}^m \beta_j F(\widehat{x} + x, v_j)$, where $\beta = (\beta_1, \dots, \beta_m)$. It is obvious that if the coefficient at $F(\widehat{x} + x, \widehat{x})$ is nonnegative, $\alpha > 0$ and $\alpha\varepsilon\bar{\beta}_j + \beta_j \geq 0$, $1 \leq j \leq m$, then this expression is a convex combination of vectors $F(\widehat{x} + x, \widehat{u})$, $F(\widehat{x} + x, v)$, $\{F(\widehat{x} + x, v_j)\}_{j=1}^m$. From the condition of smoothness it follows that $\Phi \in SD^1(0,0,0)$ and $\Phi'(0,0,0)[(x, \alpha, \beta)] = \Lambda(x - \alpha\varepsilon\xi) + \alpha F(\widehat{x}, v) + \sum_{j=1}^m \beta_j F(\widehat{x}, v_j)$. Let us be convinced, that $\Phi'(0,0,0)(X \times \mathbb{R} \times \mathbb{R}^m) = Y$. In fact, let y be in Y. From $\{\operatorname{cone}(\pi(F(\widehat{x}, v_i))\}_{i=1}^m = Z$, it follows that there exist numbers $(\beta_1(y), \dots, \beta_m(y))$ such that $\sum_{i=1}^m \beta_i(y)\pi(F(\widehat{x}, v_i) = \pi y$, i.e. $y - \sum_{i=1}^m \beta_i(y)(F(\widehat{x}, v_i)) \in Y_1$. Consequently there exists an element $x(y)$ such than $\Lambda x(y) = y - \sum_{i=1}^m \beta_i(y)(F(\widehat{x}, v_i))$, i.e. $\Phi'(0,0,0)[x(y), 0, \beta(y)] = y$.

From the formula for $\Phi'(0,0,0)$ and equality (iv) we obtain that $\Phi'(0,0,0)$ $[x_v + \varepsilon\xi, 1, 0] = 0$, hence starting from the point $\Phi((t(x_v + \varepsilon\xi), t, 0)$, and using the modified Newton method, we shall find $r_i(t)$, $i = 1, 2, 3$, such that $\Phi(\widehat{x} + t(x_v + \varepsilon\xi) + r_1(t), t + r_2(t), r_3(t)) = 0$ with the estimate $\|r_1(t)\|_X + |r_2(t)| + |r_3(t)| \leq K\|\Phi(t(x_v + \varepsilon\xi), t, 0)\|_Y = \|\Phi(0,0,0) + t\Phi'(0,0,0)[x_v + \varepsilon\xi, 1, 0] + o(t)\|_y = o(t)$. Thus (from the condition about the convexity), for some $u(t) \in \mathcal{U}$ the equality $F(\widehat{x} + t(x_v + \varepsilon\xi) + r(t), u(t)) = 0$ holds true. We construct an admissible element $(\widehat{x} + t(x_v + \varepsilon\xi) + r(t), u(t))$ in the problem (P). We supposed that $(\widehat{x}, \widehat{u})$ is a local minimum of (P), consequently (because ε is an arbitrary number) $\langle f_0'(\widehat{x}), x_v\rangle \geq 0$. Thus, implication $\Lambda x_v + F(\widehat{x}, v) = 0 \Rightarrow \langle f_0'(\widehat{x}), x_v\rangle \geq 0$ (v) is proved.

Let v in (iv) be $\widehat{u}$. Then $f_0'(\widehat{x}) \in (\operatorname{Ker}\Lambda)^\perp$. From Lemma 4 on a kernel of a regular operator $f_0'(x) + \Lambda^* y_1^* = 0$ (vi) for some $y_1^* \in Y_1^*$. Let us take now $v \in Y_1 \bigcap F(\widehat{x}, \mathcal{U})$. Then we will have

$$\langle y_1^*, F(\widehat{x}, v)\rangle \overset{(iv)}{=} -\langle y_1^*, \Lambda x_v\rangle \overset{\text{Id}}{=} -\langle \Lambda^* y_1^*, x_v\rangle \overset{(vi)}{=} \langle f_0'(\widehat{x}), x_v\rangle \overset{(v)}{\geq} 0. \qquad (v)$$

Now we shall show that there exists an extension λ of the functional y_1^* to the entire Y such that $\langle \lambda, F(\widehat{x}, u)\rangle \geq 0$ or all $u \in \mathcal{U}$ (which will mean that (b) holds with $\lambda_0 = 1$). To this end we again use geometry.

Consider the subspace $Y_0 \subset Y_1 = \{y \in Y_1 \mid \langle y_1^*, y\rangle = 0\}$. It is a hyperplane in the space Y_1. The factor-space $Z_0 = Y/Y_0$ is equal to $Y/Y_1 \times Y_1/Y_0$. The canonical projection $\pi_0 : Y \to Z_0$ maps Y_1 into one-dimensional space so that the image of $\Pi = \{y \in Y_1 \mid \langle y_1^*, y\rangle < 0\}$ is a ray $(0,\zeta)$, $\zeta < 0$. From (v) it follows that this ray does not intersect $\pi_0(F(\widehat{x},\mathcal{U})$. Finite-dimensional separation theorem allows for separating them by a functional η^*. Then, the functional $y^* = \pi_0^*\eta^*$ has the property

$$\langle y^*, F(\widehat{x},\mathcal{U})\rangle = \langle \pi_0^*\eta^*, F(\widehat{x},\mathcal{U})\rangle = \langle \eta^*, \pi_0 F(\widehat{x},\mathcal{U})\rangle \geq 0, \qquad (vi)$$

and, besides, if $y \in \mathrm{Ker} y_1^*$, then $\langle y^*, y\rangle = \langle \pi_0^*\eta^*, y\rangle = \langle \eta^*, \pi_0 y\rangle = 0$. So $\beta y^*|_{Y_1} = y_1^*$, $\beta > 0$. If we put $\beta y^* := \lambda$ we obtain the stationarity condition (see (v)) and the condition of minimum (see (vi)). The Lagrange principle is proven. ■

3. Applications (the Lagrange principle for particular classes of extremal problems)

3.1. Problems of mathematical programming.

Let X, Y be normed spaces, $V \in \mathcal{O}(\widehat{x}, X)$, $f_i\colon V \to \mathbb{R}$, $0 \leq i \leq m$, $\mathcal{F} : V \to Y$. The problem

$$f_0(x) \to \min,\ \ f_i(x) \leq 0,\ 1 \leq i \leq m,\ \mathcal{F}(x) = 0 \qquad (P_1)$$

is called *the problem of mathematical programming*. The function $\mathcal{L}(x, \bar\lambda) = \mathcal{L}(x, \lambda_0, \ldots, \lambda_m, \lambda) = \sum_{i=0}^m \lambda_i f_i(x) + \langle \lambda, \mathcal{F}(x)\rangle$, where $\lambda_i \in \mathbb{R}$, $\lambda \in Y^*$, is called *the Lagrange function* of the problem (P_1). The vector $\overline{\lambda} = (\lambda_0, \ldots, \lambda_m, \lambda)$ is called *a collection of Lagrange multipliers*.

Proposition 2 *The vector $\widehat{u} = (\widehat{u}_1, \ldots, \widehat{u}_m)$ affords an absolute minimum of the problem $\varphi(u) = \sum_{i=1}^m \lambda_i u_i \to \min$, $u_i \geq 0$ (which we call an elementary problem of linear programming) iff the conditions of nonnegativity $\lambda_i \geq 0$, $1 \leq i \leq m$, and complementary slackness $\lambda_i \widehat{u}_i = 0$, $1 \leq i \leq m$, hold.*

The proof of this proposition is evident.

Theorem 1 (the Lagrange principle for problems of mathematical programming) *If in the problem (P_1) the following smoothness conditions: $f_0 \in D^1(\widehat{x})$, $\mathcal{F} \in SD^1(\widehat{x})$, and regularity conditions: $\mathcal{F}'(\widehat{x})X$ is a closed subspace in Y are satisfied, then necessary conditions for a local minimum in the problem (P_1) at the point $\widehat{x}$ coincide with the Lagrange principle, i.e., there exists a collection of Lagrange multipliers $\overline{\lambda} = (\lambda_0, \ldots, \lambda_m, \lambda) \in \mathbb{R}^{m+1} \times Y^*$ such that the stationarity condition*

$$\mathcal{L}_x(\widehat{x}, \overline{\lambda}) = 0 \ \Leftrightarrow\ \sum_{i=0}^m \lambda_i f_i'(\widehat{x}) + \mathcal{F}'(\widehat{x})^*\lambda = 0, \qquad (1)$$

the nonnegativity condition $\lambda_i \geq 0,\ 0 \leq i \leq m$, *and the condition of complementary slackness* $\lambda_i f_i(\widehat{x}) = 0,\ 1 \leq i \leq m$, *are satisfied. In the regular case (when* $\mathcal{F}'(\widehat{x})X = Y$*) of the problem without inequalities the multiplier* $\lambda_0 \neq 0$.

Proof. If $\mathcal{F}'(\widehat{x})X \neq Y$, then (by Lemma 2 on nontriviality of annihilator) there exists an element $\bar{\lambda} \in Y^*$ such that $\langle \bar{\lambda}, \mathcal{F}(\widehat{x})x \rangle = 0\ \forall x \in X$; i.e., $\mathcal{L}_x(\widehat{x}, \overline{\lambda}) = 0$ for $\overline{\lambda} = (0, \dots, 0, \bar{\lambda})$.

If $\mathcal{F}'(\widehat{x})X = Y$, we put $\widetilde{Y} = Y \times \mathbb{R}^m$, $F(x,u) = (f_1(x)+u_1, \dots, f_m(x)+u_m)$, $u = (u_1, \dots, u_m)$, $\mathcal{U} = \mathbb{R}^m_+$ and apply the main theorem (together with Lemma 3 on closedness of the image). The stationarity condition of the main theorem together with Proposition 1 lead to the stationarity condition in Theorem 1. Condition of minimum of the main theorem together with Proposition 2 lead to nonnegativity of Lagrange multipliers and conditions of complementary slackness. ∎

3.2. Problems of the Calculus of Variations

Let $\Delta = [t_0, t_1]$, $-\infty < t_0 < t_1 < \infty$, $L_i \colon \Delta \times \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$, $0 \leq i \leq m$, $l_i \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $0 \leq i \leq m$, $\varphi \colon \Delta \times \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$, $\Xi = \{\xi = (x(\cdot), u(\cdot))\} = C^1(\Delta, \mathbb{R}^n) \times C(\Delta, \mathbb{R}^r)$, $f_i(\xi) = \int_\Delta L_i(t, x(t), u(t))\,dt + l_i(x(t_0), x(t_1))$, $0 \leq i \leq m$.

The problem

$$f_0(\xi) \to \min,\ f_i(\xi) \leq 0,\ 1 \leq i \leq m',\ f_i(\xi) = 0,\ m'+1 \leq i \leq m, \\ \dot{x} = \varphi(t, x, u) \qquad (P_2)$$

is called the *Lagrange problem of the Calculus of Variations.* The problem $\mathcal{B}(x(\cdot)) = \int_\Delta L(t, x(t), \dot{x}(t))\,dt + l(x(t_0), x(t_1)) \to \min$ is called *an elementary problem of the Calculus of Variations* or *the Bolza problem.*

A local extremum in the space $C^1(\Delta, \mathbb{R}^n)$ for the Bolza problem, and a local minimum for the Lagrange problem in the space Ξ are called *weak extrema.*

The Lagrange function for the problem (P_2) has the following form:

$$\begin{aligned} \mathcal{L}(\xi, \overline{\lambda}) &= \mathcal{L}(\xi, \lambda_0, \dots, \lambda_m, p(\cdot)) \\ &= \sum_{i=0}^{m} \lambda_i f_i(\xi) + \int_\Delta p(t) \cdot (\dot{x}(t) - \varphi(t, x(t), u(t))\,dt \\ &= \int_\Delta (\widetilde{L}(t, x(t), u(t)) + p(t) \cdot (\dot{x}(t) - \varphi(t, x(t), u(t)))\,dt + l(x(t_0), x(t_1)), \end{aligned}$$

where $\widetilde{L} = \sum_{i=0}^m \lambda_i L_i$, $l = \sum_{i=0}^m \lambda_i l_i$, $\overline{\lambda} = (\lambda_0, \dots, \lambda_m, p(\cdot)) \in \mathbb{R}^{m+1} \times C^1(\Delta, \mathbb{R}^{n*})$.

PROPOSITION 3 *Let* $\widehat{x}(\cdot) \in C^1(\Delta, \mathbb{R}^n)$*, let* $L = L(t, x, y) \colon \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *be a continuous function continuously differentiable in* x *and* y *in a neighborhood of the graph* $\{(t, x, y) \mid t \in \Delta, x = \widehat{x}(t), y = \widehat{y}(t)\}$ *and let the function* $l = l(\xi_0, \xi_1) \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *be continuously differentiable in a neighborhood of the* *point* $(\widehat{x}(t_0), \widehat{x}(t_1))$. *If* $\widehat{x}(\cdot)$ *is a weak minimum of the Bolza problem in the space* $C^1(\Delta, \mathbb{R}^n)$, *then* $\widehat{L}_{\dot{x}}(\cdot) \in C^1(\Delta)$ *and the Euler equation* $-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0$, *and the transversality conditions* $\widehat{L}_{\dot{x}}(t_i) = (-1)^{i+1}\widehat{l}_{\xi_i}$, $i = 0, 1$, *are satisfied.*

Proof of Proposition 3. It can be proved that if smoothness conditions of L and l are satisfied, then $\mathcal{B} \in D^1(\widehat{x}(\cdot))$ in the space $C^1([t_0, t_1], \mathbb{R}^n)$ and

$$\mathcal{B}'(\widehat{x}(\cdot))[x(\cdot)] = \int_{t_0}^{t_1} (\widehat{L}_{\dot{x}}(t) \cdot \dot{x}(t) + \widehat{L}_x(t) \cdot x(t))\, dt + \widehat{l}_{\xi_0} \cdot x(t_0) + \widehat{l}_{\xi_1} \cdot x(t_1). \quad (i)$$

By solving the Cauchy problem $\dot{p} = \widehat{L}_x(t), p(t_1) = -\widehat{l}_{\xi_1}$ (in other words, by denoting $p(t) = -(\widehat{l}_{\xi_1} + \int_t^{t_1} \widehat{L}_x(\tau)\, d\tau)$), substituting to (i) and integrating by parts, we obtain (from Fermat's theorem)

$$\int_{t_0}^{t_1} (\widehat{L}_{\dot{x}}(t) - p(t)) \cdot \dot{x}(t)\, dt + (\widehat{l}_{\xi_0} - p(t_0)) \cdot x(t_0) = 0 \quad \forall x(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n). \ (ii)$$

By solving the Cauchy problem $\dot{x} = \widehat{L}_x(t) - p(t), x(t_0) = l_{\xi_0} - p(t_0)$ (in other words, by denoting $x(t) := l_{\xi_0} - p(t_0) + \int_{t_0}^{t} \widehat{L}_{\dot{x}}(\tau) d\tau - p(\tau))$ and substituting to (ii), we obtain the required relations. ∎

Theorem 2 (the Lagrange principle for the Lagrange problem of the Calculus of Variations) *Let in the Lagrange problem* (P_2) *the functions* $(\widehat{x}(\cdot), \widehat{u}(\cdot)) \in C^1(\Delta, \mathbb{R}^n) \times C(\Delta, \mathbb{R}^r)$, $L_i \colon \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ *and* $\varphi \colon \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ *be continuous and continuously differentiable in* x *and* u *in a neighborhood of the graph* $\{(t, x, y) \mid t \in \Delta, x = \widehat{x}(t), y = \widehat{u}(t)\}$, *and let the functions* l_i *be continuously differentiable in a neighborhood of the point* $(\widehat{x}(t_0), \widehat{x}(t_1))$. *Then, the necessary conditions of a weak minimum in the problem* (P_2) *at the point* $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ *coincide with the Lagrange principle, i.e., there exists a nontrivial collection of Lagrange multipliers* $\overline{\lambda} = (\lambda_0, \ldots, \lambda_m, p(\cdot))$ *such that for* $L(t, x, \dot{x}, u) = \widetilde{L}(t, x, u) + p(t) \cdot (\dot{x} - \varphi(t, x, u))$ *the Euler equations for* $x(\cdot)$ *and* $u(\cdot)$*:*

$$-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0, \ \widehat{L}_u(t) = 0, \tag{2a}$$

the transversality conditions:

$$\widehat{L}_{\dot{x}}(t_i) = (-1)^{i+1}\widehat{l}_{\xi_i}, \quad i = 0, 1, \tag{2b}$$

together with nonnegativity conditions $(\lambda_i \geq 0,\ 0 \leq i \leq m')$ *and conditions of complementary slackness:* $(\lambda_i f_i(\widehat{\xi}) = 0,\ 1 \leq i \leq m')$ *hold.*

Proof of Theorem 2.

Preliminaries: global existence theorem for linear systems. Let $D = [t_0 - a, t_0 + a] \times B_{\mathbb{R}^n}(x_0, b)$, $f\colon D \to \mathbb{R}^n$, $x_0 \in \mathbb{R}^n$. Consider the problem:

$$\dot{x} = f(t, x), \;\; x(t_0) = x_0. \tag{3}$$

We call it *the Cauchy problem for the differential equation* $\dot{x} = f(t, x)$.

Theorem on global existence of solution of the Cauchy problem for linear systems. *Let* $\Delta = [t_0, t_1]$*, let functions* $A\colon \Delta \mapsto \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ *and* $b\colon \Delta \mapsto \mathbb{R}^n$ *be continuous in the segment* Δ*, let* $\tau \in \Delta$ *and* $\xi \in \mathbb{R}^n$*. Then there exists a unique solution on* Δ *of the problem:*

$$\dot{x} = A(t)x + b(t), \quad x(\tau) = \xi.$$

This result is a simple corollary of the theorem on right-inverse mapping.

Let X be $\Xi \times \mathbb{R}^m$. Denote $\mathcal{F}(\xi) := \dot{x}(\cdot) - \varphi(\cdot, x(\cdot), u(\cdot))$, $\mathcal{U} = \mathbb{R}^{m'}_+$, $F(\xi, u) = (\mathcal{F}(\xi), f_1(\xi) + u_1, \dots, f_{m'}(\xi) + u_{m'}, f_{m'+1}(\xi), \dots, f_m(\xi))$. Let us verify conditions of the main theorem. Conditions of smooth-convexity are fulfilled trivially, the condition of weak regularity follows from the global existence theorem for linear systems and Lemma 3 on closedness of the image. Differentiating in $x(\cdot)$ leads to the identity:

$$\int_\Delta \widehat{\widehat{L}}_x(t) \cdot x(t)\, dt + \langle \lambda, \dot{x}(\cdot) - \widehat{\varphi}_x(\cdot)x(\cdot)) \rangle = 0 \quad \forall x(\cdot). \tag{i}$$

Theorem on existence of solution of linear systems allows us to solve the problems: $\dot{x}(\cdot) - \widehat{\varphi}_x(\cdot)x(\cdot) = y(\cdot)$, $x(t_0) = x_0$, (ii), and $-\dot{p}(\cdot) = p(\cdot)\widehat{\varphi}_x(\cdot) - \widehat{\widehat{L}}_x(\cdot)$, $p(t_1) = -\widehat{l}_{\xi_1}$ (iii). Thus, the Euler equation and the second transversality conditions are satisfied. Substituting in (i) the expression (iii) for $\widehat{\widehat{L}}_x$, integrating by parts and using (ii), we obtain the equality $\int_\Delta p(t) \cdot y(t)\, dt = \langle \lambda, y(\cdot) \rangle$. (iv). Differentiation in $u(\cdot)$ leads to the identity: $\int_\Delta \widehat{\widehat{L}}_u(t) \cdot u(t)\, dt + \langle \lambda, -\widehat{\varphi}_u(\cdot)u(\cdot)) \rangle = 0\ \forall u(\cdot)$ (v).

Using (iv) we obtain the Euler equation for $u(\cdot)$ and the first transversality condition. ■

3.3. Convex and Lyapunov problems

The set of all convex functions in a vector space Z will be denoted by $\mathrm{Co}^f(Z)$.

Let $\Delta = [t_0, t_1]$, $-\infty < t_0 < t_1 < \infty$, $U \subset \mathbb{R}^r$, $L_i\colon \Delta \times U \to \mathbb{R}$, $0 \le i \le m$, Z be a vector space, $\varphi_i \in \mathrm{Co}^f(Z)$, $0 \le i \le m$, A be a convex subset of Z. The problem

$$\varphi_0(z) \to \min, \;\; \int_\Delta L_i(t, u(t))\, dt + \varphi_i(z) \le 0, \; 1 \le i \le m, \; z \in A, \tag{P_3}$$

where $u(\cdot)$ is a measurable vector-function, is called a *problem of convex programming.*

The problem of the Calculus of Variations of the form

$$\begin{aligned} &J_0(u(\cdot)) = \int_\Delta L_0(t,u(t))\,dt \to \min, \\ &J_i(u(\cdot)) = \int_\Delta L_i(t,u(t))\,dt \le 0,\ 1 \le i \le m', \\ &J_i(u(\cdot)) = \int_\Delta L_i(t,u(t))\,dt = 0,\ m'+1 \le i \le m,\ u(t) \in U, \end{aligned} \qquad (P_3')$$

is called the *Lyapunov problem.*

The problem $\mathcal{J}(u(\cdot)) = \int_\Delta L(t,u(t))\,dt \to \min,\ u(t) \in U$ is called *an elementary Lyapunov's problem or an elementary problem of Optimal Control.*

The Lagrange function for the problem (P_2) has the following form:

$$\mathcal{L}((z,u(\cdot)),\overline{\lambda}) = \sum_{i=0}^{m}(\lambda_i\varphi_i(z) + \lambda_i \int_\Delta L_i(t,u(t))dt).$$

Proposition 4 *A function $\widehat{u}(\cdot)$ is a solution of an elementary problem of optimal control iff $f(t,\widehat{u}(t)) = \min_{u\in U} f(t,u)$ for a.a. t of $\widehat{u}(\cdot)$. We call it the minimum condition.*

The proof of this proposition is evident.

Theorem 3 *(the Lagrange principle for problems of convex programming) Let in the problem (P_3) $L_i\colon \Delta \times U$ be continuous functions and $\varphi_i\colon Z \to \mathbb{R}$ be convex functions. Then, the necessary conditions for an absolute minimum in the problem (P_3) at the point $(\widehat{z},\widehat{u}(\cdot))$ coincide with the Lagrange principle, i.e., there exists a nontrivial collection of Lagrange multipliers $\lambda_0,\dots,\lambda_m$ such that*

$$\sum_{i=0}^{m}\lambda_i\varphi_i(z) \ge \sum_{i=0}^{m}\lambda_i\varphi_i(\widehat{z})\ \forall z \in A. \qquad (3a)$$

If t is a point of continuity of $\widehat{u}(\cdot)$, then

$$\sum_{i=0}^{m}\lambda_i L_i(t,u) \ge \sum_{i=0}^{m}\lambda_i L_i(t,\widehat{u}(t)),\ \forall u \in U. \qquad (3b)$$

If the Lagrange multiplier $\lambda_0 \ne 0$, then the absolute minimum in the problem (P_3) attains.

Proof of Theorem 3.

Preliminaries: phenomenon of convexity of finite-dimensional integral mappings.

LYAPUNOV THEOREM. *Let Δ be a segment in $\mathbb{R}$, and let $p(\cdot) = (p_1(\cdot), \dots, p_n(\cdot))$ be an integrable vector-function. Then the set $M = \{x \in \mathbb{R}^n \mid x = \int\limits_A p(t)\,dt,\ A \in \mathcal{A}\}$, where $\mathcal{A}$ is the σ-algebra of all Lebesgue measurable sets, is a convex compact set in $\mathbb{R}^n$* (Alekseev, Tikhomirov and Fomin, 1987).

The problem (P_3) can be reduced to the problem (P) by denoting $X := \mathbb{R}$, $\mathcal{U} := \mathcal{A} \times A \times \mathbb{R}^m_+$ (where $\mathcal{A}$ is the set of measurable functions such that $t \mapsto L_i(t, u(t)) \in L_1(\Delta)$), $u = (u(\cdot), z, \alpha) \in \mathcal{A} \times A \times \mathbb{R}^m_+$, $f_0(x) = x$, $\phi_0(x.u) = \alpha_0 - x$, $\phi_i(u) = \int_\Delta L_i(t, u(t))dt + \varphi_i(z) + \alpha_i$, $1 \le i \le m$. Conditions of smoothness and weak regularity are satisfied trivially, the convexity condition follows from Lyapunov's theorem. Together with Proposition 4 this implies the theorem. ∎

4. Problems of Optimal Control

Let $\Delta = [t_0, t_1]$, $-\infty < t_0 < t_1 < \infty$, $U \subset \mathbb{R}^r$, $L_i \colon \Delta \times \mathbb{R}^n \times U \to \mathbb{R}$, $1 \le i \le m$, $l_i \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $0 \le i \le m$, $\varphi \colon \Delta \times \mathbb{R}^n \times U \to \mathbb{R}^n$, $\Xi_1 = \{\xi = (x(\cdot), u(\cdot))\} = PC^1(\Delta, \mathbb{R}^n) \times PC(\Delta, \mathbb{R}^r)$ (piecewise continuously differentiable and piecewise continuous functions), $f_i(\xi) = \int_\Delta L_i(t, x(t), u(t))\,dt + l_i(x(t_0), x(t_1))$, $0 \le i \le m$.

The problem

$$\begin{aligned} &f_0(\xi) \to \min,\ f_i(\xi) \le 0,\ 1 \le i \le m',\ f_i(\xi) = 0,\ m'+1 \le i \le m, \\ &\dot{x} = \varphi(t, x, u),\ u \in U \end{aligned} \tag{P_4}$$

is called *the problem of Optimal Control.* We call the problem

$$\begin{aligned} &J(x(\cdot), u(\cdot)) = \int_\Delta f(t, x(t), u(t))\,dt \to \min, \\ &\dot{x} = \varphi(t, x, u),\ x(t_0) = x_0\ , (x(t_0) = x_0,\ x(t_1) = x_1),\ u \in U, \end{aligned} \tag{P_4'}$$

the problem of Optimal Control in Pontryagin's form with a free boundary condition (with fixed boundary conditions).

The Lagrange function of the problem (P_4) has the following form:

$$\begin{aligned} &\mathcal{L}(\xi, \bar{\lambda}) = \mathcal{L}(\xi, \lambda_0, \dots, \lambda_m, p(\cdot)) \\ &= \sum_{i=0}^m \lambda_i f_i(\xi) + \int_\Delta p(t) \cdot (\dot{x}(t) - \varphi(t, x(t), u(t))\,dt \\ &= \int_\Delta (\widetilde{L}(t, x(t), u(t)) + p(t) \cdot (\dot{x}(t) - \varphi(t, x(t), u(t)))\,dt + l(x(t_0), x(t_1)), \end{aligned}$$

where $\widetilde{L} = \sum_{i=0}^m \lambda_i L_i$, $l = \sum_{i=0}^m \lambda_i l_i$.

A pair $\widehat{\xi} = (\widehat{x}(\cdot), \widehat{u}(\cdot))$ is said to be an *optimal process* or to afford a *strong local minimum* in problem (P_4), if there exists a $\varepsilon > 0$ such that for any admissible pair(x, u) such that $\|x - \widehat{x}\|_{C(\Delta,\mathbb{R}^n)} < \varepsilon$ the inequality $f_0(\xi) \geq f_0(\widehat{\xi})$ holds.

THEOREM 4 (THE LAGRANGE PRINCIPLE FOR PROBLEMS OF OPTIMAL CONTROL) *Let in the problem (P_4) of Optimal Control the functions $(\widehat{x}(\cdot), \widehat{u}(\cdot)) \in PC^1(\Delta, \mathbb{R}^n) \times PC(\Delta, \mathbb{R}^r)$, $L_i \colon \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ and $\varphi \colon \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ be continuous and continuously differentiable in x in a neighborhood of the graph $\{(t, x, y) \mid t \in \Delta, x = \widehat{x}(t), y = \widehat{u}(t)\}$, and let the functions l_i be continuously differentiable in a neighborhood of the point $(\widehat{x}(t_0), \widehat{x}(t_1))$. Then, necessary conditions of a strong local minimum in the problem (P_4) at the point $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ coincide with the Lagrange principle, i.e., there exists a nontrivial collection of Lagrange multipliers $\overline{\lambda} = (\lambda_0, \ldots, \lambda_m, p(\cdot))$ such that for $L(t, x, \dot{x}, u) = \widetilde{L}(t, x, u) + p(t) \cdot (\dot{x} - \varphi(t, x, u))$ the Euler equations for $x(\cdot)$, minimum condition for $u(\cdot)$:*

$$-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0, \tag{4a}$$

$$\min_{u \in U} L(t, \widehat{x}(t), \dot{\widehat{x}}(t), u) = \widehat{L}(t) \tag{4a'}$$

for all t of continuity $\widehat{u}(\cdot)$, the transversality conditions:

$$\widehat{L}_{\dot{x}}(t_i) = (-1)^{i+1}\widehat{l}_{\xi_i}, \quad i = 0, 1, \tag{4b}$$

together with nonnegativity conditions ($\lambda_i \geq 0,\ 0 \leq i \leq m'$) and conditions of complementary slackness: ($\lambda_i f_i(\widehat{\xi}) = 0,\ 1 \leq i \leq m'$) hold.

THEOREM 4′ (THE LAGRANGE PRINCIPLE FOR THE PROBLEM OF OPTIMAL CONTROL IN PONTRYAGIN'S FORM). *Let the functions f and φ in the problem (P_4') be continuous together with their derivatives f_x and φ_x. Then, the necessary conditions for strong minimum in the problem (P_4') at the point $(\widehat{x}(\cdot), \widehat{u}(\cdot))$ coincide with the Lagrange principle, i.e., for the Lagrange function*

$$\mathcal{L} = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t), u(t))\, dt,$$

where $L = \lambda_0 f + p(t)(\dot{x} - \varphi)$, the Euler equation in x:

$$-\frac{d}{dt}\widehat{L}_{\dot{x}}(t) + \widehat{L}_x(t) = 0 \quad \Leftrightarrow \quad -\dot{p} = p\widehat{\varphi}_x(t) - \lambda_0 f_x(t) \tag{4'a}$$

and b) the minimum condition in u:

$$\min_u L(t, \widehat{x}(t), u) = \widehat{L}(t) \tag{4'b}$$

are satisfied. In the problem with free end the transversality condition $p(t_1) = 0$ *is satisfied.*

Outline of the proof of Theorem 4'. We restrict ourself to the problems of Optimal Control in Pontryagin's form with free end (with fixed end):

$$f_0(x(\cdot)) = \int_\Delta f(t, x(t))dt \to \min,\ x(t_0) = x_0 \quad (x(t_i) = x_i,\ i = 1, 2),$$
$$\dot{x} = \varphi(t, x, u),\ u(t) \in U. \qquad (P_4'')$$

To prove Theorem 4 one has to overcome only some technicalities.

Preliminaries: the phenomenon of "almost convexity"of infinite-dimensional integral mappings. Consider the mapping $F(x(\cdot), u(\cdot)) = x_0 + \int_{t_0}^t \varphi(s, x(s), u(s))ds$. Let $\mathcal{U}(x(\cdot)) \subset PC(\Delta, \mathbb{R}^n)$ denote the set of functions $F(x(\cdot),\ u(\cdot))$, for which $u(t) \in U$ at continuity points of $u(\cdot)$. The image $u(\cdot) \mapsto F(x(\cdot), u(\cdot))$ is not convex: if we take $u_i(\cdot) \in \mathcal{U}(x(\cdot))$, $i = 1, 2$, and a number $\beta \in [0, 1]$, then there need not exist a function $u_\beta(\cdot)$, such that $(1 - \beta)F(x(\cdot), u_1(\cdot)) + \beta F(x(\cdot), u_2(\cdot) = F(x(\cdot), u_\beta(\cdot))$. But it is possible to construct a function $u_{\beta\delta}(\cdot)$, which almost satisfies this equality. This means that for any $\varepsilon > 0$ there exists a number $\delta > 0$ and a function $u_{\beta\delta}(\cdot)$ such that $\|(1 - \beta)F(x(\cdot), \widehat{u}_1(\cdot)) + \beta F(x(\cdot), u_2(\cdot)) - F(x(\cdot), u_{\beta\delta}(\cdot))\|_{C(\Delta, \mathbb{R}^n)} < \varepsilon$.

To do this, divide the segment Δ into segments Δ_i of length δ, and then divide each segment Δ_i into two parts, of length $(1 - \beta)|\Delta_i|$ and $\beta|\Delta_i|$. Then we put $u_{\beta\delta}(t) = u_1(t)$ on the first parts of the segments and $u_{\beta\delta}(t) = u_2(t)$ on the second parts. Denote this function by $M_{\beta,\delta}(u_1(\cdot), u_2(\cdot))$. It is easy to prove that $\lim\limits_{\delta\to 0} F(x(\cdot), M_{\beta,\delta}(u_1(\cdot), u_2(\cdot))) = (1 - \beta)F(x(\cdot), u_1(\cdot)) + \beta F(x(\cdot), u_2(\cdot))$, $\lim\limits_{\beta\to 0} F(x(\cdot), M_{\beta,\delta}(u_1(\cdot), u_2(\cdot))) = F(x(\cdot), u_1(\cdot))$.

Such construction is called a "mix of control functions".

Now, we shall apply the method of proof of the main theorem to our problem (P_4') with free end. Here $X = Y = C(\Delta, \mathbb{R}^n)$, $\mathcal{U}$ is the set of piecewise-continuous functions $u(\cdot)\colon \Delta \to U$, $F(x(\cdot), u(\cdot)) = x_0 + \int_{t_0}^t \varphi(s, x(s), u(s))\, ds$, $\Lambda = F_x(\widehat{x}(\cdot), \widehat{u}(\cdot))$. By the theorem on existence of solution, Λ is a surjective operator from X onto Y.

Let $v(\cdot)$ be an admissible control function and let $x_{v(\cdot)}(\cdot) \in X$ be a function such that $\Lambda x_{v(\cdot)} + F(\widehat{x}(\cdot), v(\cdot)) = 0$. Define the mapping Φ by the formula $\Phi(x(\cdot), \theta) = (1-\theta)F(\widehat{x}(\cdot)+x(\cdot), \widehat{u}) + \theta F(\widehat{x}(\cdot)+x(\cdot), v(\cdot))$. Applying the theorem on the right-inverse mapping we obtain the equality $(1-\theta-\rho(\theta))F(\widehat{x}(\cdot)+x_{v(\cdot)}(\cdot)+ r(\theta), \widehat{u}(\cdot)) + (\theta+\rho(\theta))F(\widehat{x}(\cdot)+x_{v(\cdot)}(\cdot)+r(\theta), v(\cdot)) = 0$, where $r(\theta)$ and $\rho(\theta)$ are $o(\theta)$. Consider a mix $M_{\theta+\rho(\theta),\delta}(\widehat{u}(\cdot), v(\cdot))$ with δ so small that the modified Newton's sequence starting from the point $F(\widehat{x}(\cdot)+x_{v(\cdot)}(\cdot)+r(\theta), M_{\theta+\rho(\theta),\delta}(\widehat{u}(\cdot), v(\cdot)))$ converges to the solution of the equation $F(\widehat{x}(\cdot)+\theta x_{v(\cdot)}(\cdot)+o(\theta), M_{\theta+\rho(\theta),\delta}(\widehat{u}(\cdot), v(\cdot))) = 0$. Then one can finish the proof as it was done in the main theorem.

In the case of the problem with fixed ends it is necessary also to realize the plan of proof of the main theorem.

At first, it is necessary to consider the degenerate case and to reject it. Then we have to prove the equality $\Phi(\theta(x_{v(\cdot)}+\varepsilon\xi(\cdot))+r_1(\theta), \theta+r_2(\theta), r_3(\theta)) = 0,\ r_i(\theta)) = o(\theta),\ i=1,2,3$. This means that $(1-\theta-\rho(\theta)-\sum_{j=1}^{N}(\beta_j+\varepsilon\sum_{j=1}^{m}\bar{\beta}_j+\rho_j(\theta)))F(\widehat{x}(\cdot)+\theta x_{v(\cdot)}(\cdot), \widehat{u}(\cdot))+(\theta+\rho(\theta)F(\widehat{x}(\cdot)+\theta x_{v(\cdot)}(\cdot)+r(\theta), v(\cdot)+(\varepsilon\sum_{j=1}^{m}\bar{\beta}_j+\sum_{j=1}^{m}\beta_j+\rho_j(\theta))F(\widehat{x}(\cdot)+\theta x_{v(\cdot)}(\cdot)+r_j(\theta), v_j(\cdot)) = 0$.

And then we must mix the control functions proportionally to $(\theta+\rho(\theta), \varepsilon\bar{\beta}+\beta+\overline{\rho}(\theta))$, i.e., construct the following mix:

$$M_{(\theta+\rho(\theta),\varepsilon\bar{\beta}+\beta+\overline{\rho}(\theta),\delta)}(\widehat{u}(\cdot), v_0(\cdot), v_1(\cdot), \ldots, v_m(\cdot)).$$

Then, we have to choose δ so small that the modified Newton's sequence starting from the point $(\widehat{x}+\theta x_{v_0(\cdot)}(\cdot)+r(\theta), M_{(\theta+\rho(\theta),\varepsilon\bar{\beta}+\beta+\overline{\rho}(\theta),\delta)}(\widehat{u}(\cdot), v_0(\cdot), v_1(\cdot), \ldots, v_m(\cdot))$ and varying $x(\cdot)$ and β converges to the solution of the equation $F(\widehat{x}(\cdot)+\theta x_{v(\cdot)}(\cdot)+o(\theta), M_{\theta+\rho(\theta),\delta}(\widehat{u}(\cdot), v(\cdot))) = 0$. Then one can finish the proof as it was done in the main theorem.

5. Comments

The notion of differentiability appears for the first time in the papers by Newton (in the 1660s) (Newton, 1736) and Leibniz (1684). The modern definition was introduced by Cauchy (1823). The concept of differentiability of functions of many variables is due to Weierstrass (1880s, see Weierstrass, 1903). Derivatives for functions of infinitely many variables were defined by Frèchet (1912). The idea of strong differentiability is due to Leach (1961).

Newton's method (and inverse function theorem for one variable) goes back to Newton (dated 1676), the implicit function theorem goes back to Dini. Infinite-dimensional versions of the inverse map theorem go back to Lyusternik (1934), Graves (1950) and Robinsson (1976), among others.

A necessary condition of extremum for smooth problems without constraints goes back to Fermat (1638) (see Fermat, 1891). Actually, the theory of extremal problems was born with the letter of Fermat. Stationary conditions in terms of derivatives appear in the first papers on calculus (see Newton, 1736, and Leibniz, 1684). In the infinite-dimensional case, a stationarity condition goes back to Frèchet (1912).

A classical calculus of variations was born in 1696 when John Bernoulli (1696) posed the brachistochrone problem. The Euler equation for the simplest problem of calculus of variations was obtained by Euler (1744). He also deduced a necessary condition for isoperimetric problem. Euler's equation for multidimensional problems was derived by Gauss and Ostrogradsky.

Lagrange started the study of problems with constraints. The Lagrange multipliers rule was formulated in the book Lagrange (1797). He used it as a heuristic method from the beginning of the 1770s. Rigorous (in the sense of the end of the 19th century) proofs of the Lagrange multipliers rule were given for the finite-dimensional case after proving the inverse map and implicit function theorems, when finite-dimensional linear algebra was created. For the first time a rigorous proof of the Lagrange multipliers method was given by Mayer. (Hilbert also proposed a rigorous proof of this result.) Before that, many particular cases have been treated. Necessary conditions for isoperimetric problems were given by Weierstrass, necessary conditions for the derivatives of higher order were given by Poisson; moreover, necessary conditions for the problems with non-stationarity ends have been found, and so on. In the 19th century this was the main subject of all textbooks on the calculus of variation.

Convex problems appeared at the end of the 1930s: Kantorovich (1939), Karush (1939), John (1948), Kuhn and Tucker (1951) and others. The phenomenon of the convexity of integral maps was discovered by Lyapunov (1940).

Optimal control appeared in 1687 (before the classical calculus of variations) when Newton, in his *"Principia"* posed and solved the problem of minimal resistance of a solid of revolution in the discrete space, but nobody noted that.

The optimal control theory was elaborated in 1950s by Pontryagin and his collaborators. The results of the first stage of the theory were summarized in Pontryagin (1959). This paper stimulated a great growth of the extremum theory.

The paper by Dubovitsky and Milyutin (1965) made an impact on the extremum theory. It may be regarded as the first paper on extremum problems theory. The developments of this theory for ten years were set out in the book by Ioffe and Tikhomirov (1979). The results of Milyutin's work on Pontryagin's maximum principle were partially exposed in the book by Milyutin, Dmitruk and Osmolovsky (2004).

Almost all results on necessary extremum conditions contained in the papers referred to correspond to the Lagrange principle given in this paper. The majority of the results on necessary conditions for extremum, in particular, all the results on necessary conditions contained in Ioffe and Tikhomirov (1979), Alekseev, Tikhomirov and Fomin (1987), Magaril-Ilyaev and Tikhomirov (2003), Brinkhuis and Tikhomirov (2005) and Arutyunov, Magaril-Ilyaev and Tikhomirov (2006), follow directly from the main theorem.

References

Alekseev, V.M., Tikhomirov, V.M. and Fomin, S.V. (1987) *Optimal Control.* Consultants Bureau.

Arutyunov, A.I., Magaril-Ilyaev, G.G. and Tikhomirov, V.M. (2006) *Pontryagin maximum principle. Proofs and applications* (in Russian). Factorial, Moscow.

Bernoulli, J. (1696) Problema novum, ad cuius solutionem Mathematici invitatur. *Acta Eruditorum.* (Opera tom. I).

BRINKHUIS, J. and TIKHOMIROV, V.M. (2005) *Optimization: Insight and Applications*. Princeton.
CAUCHY, A.L. (1823) Resumé des Lecons Données ā L'École Royale Polytechnique sur le Calcul Infinitésimal. Reprinted in: A.L. Cauchy, *Oeuvres, Series 2,* **4** (1899).
DUBOVITSKY, A.YA. and MILUTIN, A.A. (1965) Problems for extremum under constraints (in Russian). *Zh. Vichisl. Mat. i Mat. Fiz.*, **5**,3, 395–453.
EULER, L. (1744) *Methodus inveniendi* ... Lausanne.
FERMAT, P. (1891) *Oevres de Fermat*, **1**, Gauthier-Villars, Paris.
FRÈCHET, V. (1912) Sur la notion de differentielle. *Nouvelle annale de mathematique*, Ser. 4, V.XII.1912, p. 845.
GRAVES, L.M. (1950) Some mapping theorems. *Duke Math. J.,* **17**, 111–114.
IOFFE, A.D. and TIKHOMIROV, V.M. (1979) *Theory of Extremal Problems*. North-Holland.
JOHN, F. (1948) Extremum problems with inequalities as subsidary conditions. *Studies an Essays. Courant Anniversary volume*. Interscience, NY, 187–204.
KANTOROVICH, L.V.(1939) *Mathematical methods of production management and planning* (in Russian). Leningrad Univ., Leningrad.
KARUSH, W.E. (1939) *Minima of Functions of Several Variables with Inequalities as Side Conditions*. Univ. of Chicago Press.
KUHN, H.W. and TUCKER, A.W. (1951) *Nonlinear Programming*. Univ. of California Press, Berkeley, 481–482.
LAGRANGE, L. (1797) *Théorie des fonctions analytiques*. Paris.
LEACH, E. (1961) A note on inverse function theorem. *Proc. AMS*, **12**, 694–697.
LEIBNIZ, G. (1684) Nova methodus pro maximis et minimis... *Acta Eruditorum*, 467–473.
LYAPUNOV, A.A. (1940) On totally additive functions (in Russian). *Izv. Akad. Nauk SSSR, Ser. Mat.,* **4**, 465–478.
LYUSTERNIK, L.A. (1934) Sur les extrémés relatifs des fonctionalles (in Russian). *Matem. Sbornik*, **41**, 390–400.
MAGARIL-ILYAEV, G.G. and TIKHOMIROV, V.M. (2003) *Convex Analysis: Theory and Applications*. AMS, Providence.
MILYUTIN, A.A., DMITRUK, A.V. and OSMOLOVSKY, N.P. (2004) *Pontryagin maximum principle in optimal control* (in Russian). MGU.
NEWTON, I. (1736) *The Method of Fluxions and infinite series*. London.
PONTRYAGIN, L.S. (1959) Optimal control process (in Russian). *Uspekhi mat. nauk*, **14**.
ROBINSON, S.M. (1976a) Regularity and stability for convex multivalued functions. *Math. Oper. Res.* **1**, 130–143.
ROBINSON, S.M. (1976b) Stability theory for systems of inequalities. Part II: differentiable nonlinear systems. *SIAM J. Numer. Anal.*, **13**, 497–513.
WEIERSTRASS, K. (1930) *Mathematische Werke*. Mayer & Müller, Berlin.