Control and Cybernetics

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

Degeneracy and generalized solutions of optimal control problems*

by

Vladimir I. Gurman

Program Systems Institute, RAS, Pereslavl-Zalesskii, Russia

Abstract: Degenerate optimal control problems are defined as those which have some latent passive differential or finite-difference constraints. They are typical for applications, but irregular for general methods. In this paper an analytical survey is presented of special methods developed starting from the 1960s and based, in essence, on detection and elimination of passive constraints, and some important results including most recent ones are given. Illustrative examples are considered.

Keywords: optimal control, degenerate problem, pulse mode, turnpike solution.

1. Introduction. Basic definitions

Degenerate optimal control problems are often met in applications and bring serious difficulties for general methods. They were studied systematically beginning from the 1960s (see Gurman, 1965, 1972, 1977, 1985) and defined (generalizing the classical notion of degeneracy) as those which have some latent passive differential constraints. More precisely, consider the following abstract problem.

There is a functional $I : \mathbb{M} \to \mathbb{R}$, a mapping of some set $\mathbb{M}$ (called *basic set*) onto numerical axis. A subset $\mathbf{D} \subset \mathbb{M}$ is given, called the *admissible set*. It is required to find a minimizing sequence $\{m_s\}$ of the functional I at the set $\mathbf{D}$:

$$I(m_s) \to \inf_{\mathbf{D}} I = I_*.$$

Let a certain class $\mathcal{E}$ of extensions $\mathbf{E} \supset \mathbf{D}$ be given.

Definition 1 *The problem* $(\mathbf{D}, I)$ *is called degenerate with respect to the class* $\mathcal{E}$ *if there is an extension* $\mathbf{E}_* \in \mathcal{E}$, $\mathbf{E}_* \neq \mathbf{D}$, *such that* $\inf_{\mathbf{E}_*} I = \inf_{\mathbf{D}} I$.

*Submitted: January 2009; Accepted: August 2009.

For the optimal control problem

$$\dot{x} = f(t, x, u), \quad t \in \mathbf{T} = [t_I, t_F],\ x \in \mathbb{R}^n,\ u \in \mathbf{U}(t, x) \subset \mathbb{R}^p, \tag{1}$$

$x(t_I) = x_I$ is fixed,

$$x \in \mathbf{X}(t), \quad x(t_F) \in \mathbf{\Gamma} \subset \mathbb{R}^n, \quad I = F(x(t_F)) \to \inf$$

(with continuous $f(t, x, u)$ and $F(x)$, piecewise smooth $x(t)$, and piecewise continuous $u(t)$) the basic set $\mathbb{M}$ is formed by all arbitrary pairs of functions $(x(t), u(t))$, the set $\mathbf{D}$ is selected from $\mathbb{M}$ by the above constraints, and class $\mathcal{E}$ is generated by transformations of differential constraint (1)

$$\dot{y} = \eta_x f\,(t, x, u) + \eta_t, \quad u \in \mathbf{U}(t, x), \quad y = \eta\,(t, x) \tag{2}$$

on subsets $\mathbf{T}' \subset \mathbf{T}$ $(\operatorname{mes} \mathbf{T}' > 0)$ via smooth irreversible mappings

$$\eta\colon\ R^{n+1} \to \mathbb{R}^m, \quad m < n.$$

In what follows we give a survey of special methods for the degenerate problems based, in essence, on detection and elimination of passive constraints.

2. Transformation of unbounded control system

The core of this theory is a certain transformation of system (1) with unbounded velocity set $\mathbf{V}(t, x) = f(t, x, \mathbf{U}(t, x))$. Namely, we introduce an auxiliary system, which describes asimptotically the behavior of the original system for large velocities (called *limit system*):

$$\frac{dx}{d\tau} = lw, \quad l \in \mathbf{K}(t, x), \quad w \in [0, \infty), \tag{3}$$

where t is a parameter and $\mathbf{K}$ is the union of all limits $l = \lim v_q |v_q|^{-1}$ as $|v_q| \to \infty$, $\{v_q\} \subset \mathbf{V}$.

Let $\mathbf{Q}(t, y) = \{x : y = \eta(t, x)\}$, $y \in \mathbb{R}^m$, $0 < m < n$, be the full controllability manifold of (3). Then, (2) will be called the *derived system*. The set $\mathbf{E}_x$ of piecewise continuous functions $\hat{x}(t)$ satisfying (2), is an extension of the set $\mathbf{D}_x$ of piecewise smooth $x(t)$ satisfying (1). The following statement is true (under some natural assumptions).

THEOREM 1 *For any $\hat{x}(t) \in \mathbf{E}_x$ there exists a sequence $\{x_s(t)\} \subset \mathbf{D}_x$ converging to $\hat{x}(t)$ in measure on a prescribed bounded interval $\mathbf{T}$ with $x_s(t_\alpha) \to \hat{x}(t_\alpha)$ for any prescribed finite set of values $\{t_\alpha\} \subset \mathbf{T}$.*

This sequence is constructed as follows. Introduce an intermediate class of piecewise continuous functions $\tilde{\mathbf{D}}_x$, $\mathbf{D}_x \subset \tilde{\mathbf{D}}_x \subset \mathbf{E}_x$, with the following properties: a function $x(t)$ from $\tilde{\mathbf{D}}$ satisfies system (1) in any interval of continuity, and its left and right limits lie in the same set $\mathbf{Q}(t', \eta(t', x(t-0))) = \mathbf{Q}(t', y(t'))$ at any discontinuity point t':

$$x(t'+0) \in \mathbf{Q}(t', y(t')), \quad x(t'-0) \in \mathbf{Q}(t', y(t')).$$

A sequence $\{x_q(t)\}$ approximating $x(t)$ is constructed of elements from $\tilde{\mathbf{D}}_x$. And then each element $x_q(t)$ is approximated by a sequence from $\mathbf{D}_x$ (Fig. 1).

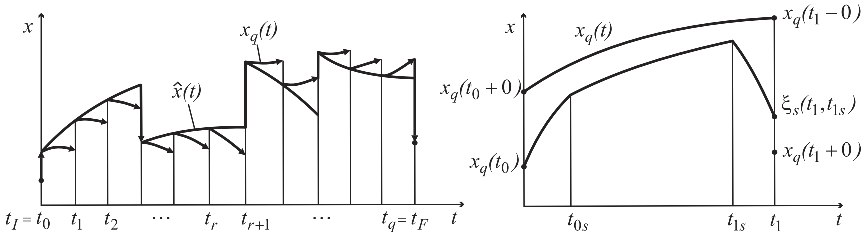

Figure 1.

So, let $(y(t), x(t), u(t))$ be a solution of the derived system (2). Divide the interval $\mathbf{T}$ by a system of points $\{t_i\}$, $i = 0, 1, \ldots, q$, $t_0 = t_I$, $t_q = t_F$, which includes all discontinuity points of the pair $(x(t), u(t))$ and points from the given set $\{t_\alpha\}$. We construct a function $x_q(t) \in \tilde{\mathbf{D}}$ as follows. Set $x_q(t_0) = x(t_0)$. In each interval (t_i, t_{i+1}) define $x_q(t)$ as the solution of system (1) for $u = u(t)$ starting at the point $x_q^*(t_i) \in \mathbf{Q}(t_i, y_q(t_i)) = \mathbf{Q}(t_i, \eta(t_i, x_q(t_i - 0))$. We choose this point for each division $\{t_i\}$ so that for $x = x(t_i + 0)$

$$\frac{\rho(x, x_q^*) - \rho(x, \mathbf{Q})}{(\Delta t_i)^2} \to 0 \quad \text{as} \quad \Delta t_i = t_{i+1} - t \to 0.$$

Taking into account the properties of system (3) we state that there is a sequence $\{x_{qs}(t)\} \subset \mathbf{D}_x$ converging to $x_q(t)$ in measure with $x_{qs}(t_F) \to x_q(t_F)$. We construct it as follows. Let $\xi_s(t, \theta, a) \in \mathbf{D}_x$ be a term of a sequence of solutions from $\mathbf{D}_x$ that approximates the corresponding limit trajectory, which is a solution of system (3) passing through a point $a \in \mathbb{R}^n$ for $t = \theta$.

1) In the initial interval $[t_0, t_1]$ set $x_{qs}(t_0) = x_q(t_0) = x(t_0)$,
$x_{qs}(t) = \xi_s(t, t_0, x_{qs}(t_0))$ for $t \in [t_0, t_{0s})$, where t_{0s} is given by the condition

$$\min_{t_{0s} \in [t_0, t_1]} \rho(\xi_s(t_{0s}, t_0, x_{qs}(t_0)), x_q).$$

2) Then, define $x_{qs}(t)$ as the solution of system (1) for $u = u(t)$ in the interval $[t_{0s}, t_{1s})$, where t_{1s} is chosen from the condition:

$$\min_{t_{1s} \in [t_{0s}, t_1]} \rho(\xi_s(t_1, t_{1s}, x_{qs}(t_{1s})), x_q(t_1 + 0))$$

and in the interval $[t_{1s}, t_1)$ set

$$x_{qs}(t) = \xi_s(t, t_{1s}, x_{qs}(t_{1s})).$$

3) The second and all the next intervals $[t_i, t_{i+1}]$ are divided by a moment $t_{(i+1)s}$ into two intervals

$$[t_i, t_{(i+1)s}), \quad (t_{(i+1)s}, t_{i+1}]$$

and the constructions of item 2) are repeated starting from the points $(t_i, x_{qs}(t_i))$. Denote

$$\delta_q = \sup_T |x_q(t) - x(t)|, \quad \delta_{qs} = \sup_{T \setminus T_{qs}} |x_{qs}(t) - x_q|, \quad \delta_s = \sup_{T \setminus T_{qs}} |x_{qs} - x|$$

$$\Delta_{qs} = \text{mes}\, \mathbf{T}_{qs}, \quad \mathbf{T}_{qs} \bigcup_{0 \le i < q-1} [t_i, t_{i+1}], \quad d_{\alpha s} = \max_\alpha \delta x_{qs}(t_\alpha + 0).$$

Define $q(s)$ so that

$$\max(\delta_q, \delta_{qs}, \delta_s, \Delta_{qs}) \to 0.$$

Thus, we obtain a sequence $\{x_s(t)\} = \{x_{q(s)s}(t)\}$, which converges to $x(t)$ in measure with $x_s(t_\alpha) \to x_s(t_\alpha + 0)$, including $x_s(t_F) \to x(t_F)$.

The complete proof of the theorem is given in Gurman (1998).

Any solution of the derived system $x(t)$ is viewed as a generalized solution of the original system called the *pulse sliding mode*. In particular cases it may be realized also as *turnpike solution* (see Section 6).

Originally, the pulse sliding modes have been revealed by Krotov (1961) when investigating the basic problem of Variational Calculus systematically for the case of linear and asimptotically linear integrant w.r.t. the derivative. For the case of control system with scalar linear control, the transformation approach was proposed independently in Kelley (1964) and Gurman (1965). The above general transformation was proposed in Gurman (1972). Later, when developing constructive procedures (see Gurman, 1977; Baturin, Dykhta, Gurman, and others, 1987; Kolokolnikova, 1996; Gurman and Ukhin, 2005), commutativity was assumed of basis vectors $\{h_q(t,x)\}$ of $\mathbf{K}(t,x)$ linear envelope: $[h_p, h_q] = 0,\ p, q = 1, \ldots, k$, which is also traditional for other works on pulse control. However, it is not necessary. Recently, in Gurman and Sachkov (2008), the following results of geometrical control theory were applied to describe full controllability sets of the limit system without commutativity assumption.

Let **Lie** $(h_1, \ldots, h_k)$ be Lie algebra, generated by addition to $h_1, \ldots, h_k$ of all commutators $[h_i, h_j]$, $[[h_i, h_j], h_l], \ldots$, and their linear combinations, and let $\mathbb{L}(t,x)$ be its linear envelope. If $\dim \mathbb{L}(t,x) = l \le n$ for all $t \in \mathbf{T}$ and $x \in \mathbb{R}^n$, then system (3) is fully controllable on some l-dimensional manifold $\mathbf{Q}(t, x_0) \subset \mathbb{R}^n$, called the *orbit* of the $(h_1, \ldots, h_k)$ family, containing some initial point $x_0 \in \mathbb{R}^n$ (Nagano-Sussman orbit theorem, see Agrachev and Sachkov, 2004), and (3) can be equivalently replaced by the following "virtual" limit system:

$$\frac{dx}{d\tau} = \sum_1^l h_j(t,x)u^j. \tag{4}$$

Example 1

$$\begin{aligned} &\dot{x} = g(t,x) + h_1(x)u^1 + h_2(x)u^2, \quad x \in \mathbb{R}^4, \\ &h_1 = (-x^3, -x^4, x^1, x^2)^{\mathrm{T}}, \quad h_2 = (x^2, -x^1, -x^4, x^3)^{\mathrm{T}}. \end{aligned} \tag{5}$$

The related limit system is

$$\frac{dx}{d\tau} = h_1(x)u^1 + h_2(x)u^2. \tag{6}$$

The pair $h_1(x), h_2(x)$ generates 3-dimensional Lie algebra: $(h_1,h_2)=(h_1,h_2,h_3)$, $h_3 = \frac{1}{2}[h_1,h_2] = (-x^4, x^3, -x^2, x^1)^{\mathrm{T}}$. Then, (6) is fully controllable on any sphere $\mathbf{S}(y) = \{x\colon |x|^2 = y\}$ (but not in $\mathbb{R}^4$). Indeed, $|x|^2 = y$ is an integral of this system:

$$\frac{dy}{d\tau} = 2x^{\mathrm{T}}h_1(x)u^1 + 2x^{\mathrm{T}}h_2(x)u^2 = 0.$$

This means that any trajectory of (6), beginning in the spere $\mathbf{S}(y)$ does not leave it. It follows from $\det(h_1(x), h_2(x), h_3(x), x) = -y^2 \neq 0$ on $\mathbf{S}(y)$, that $\mathbb{L}(x) = \operatorname{span}(h_1(x), h_2(x), h_3(x))$ is the touching space to the sphere at the point x. Thus, due to orbit theorem, $\mathbf{S}(y)$ is the orbit, the greatest fully controllable set of (6). Consider the derived system $\dot{y} = 2x^{\mathrm{T}}g(t,x)$, $x \in \mathbf{S}(y)$. It is a control system of **first** order, where y is the phase variable, and x is control vector belonging to $\mathbf{S}(y)$ as the control set.

Compare this transformation with analogous one for the case of "commutative" limit system

$$\dot{x} = g(t,x) + h_1u^1 + h_2u^2, \quad h_1 = (-1,-1,1,1)^{\mathrm{T}}, \quad h_2 = (1,-1,-1,1)^{\mathrm{T}}.$$

Here, h_i are constants, so that commutativity condition is fulfilled a priori and this is the only difference from (5). The limit system integral is

$$y = \nu x, \quad \nu = (\nu_1\ \nu_2), \quad (\nu_1\ \nu_2)^{\mathrm{T}}(h_1,h_2) = 0.$$

The derived system $\dot{y} = \nu^{\mathrm{T}}g(t,x), \quad y = \nu x$, is of **second** order. Thus, one can see that "noncommutative" limit system leads to the derived system that is simpler (of lesser order) than the "commutative" one.

Of course, the first representatives of degenerate problems are free-end optimal control problems ($\mathbf{X}(t) = \mathbf{\Gamma} = \mathbb{R}^n$), stated for such systems with an unbounded velocity set. We eliminate the passive constraints, replacing the original differential constraint by the derived system and obtain an equivalent *derived problem* of reduced order.

A series of important theoretical results have been obtained in this way, including the generalized Pontryagin principle and other optimality conditions, and nonlocal iterative algorithms for the search of singular and generalized solutions (see Gurman, 1985; Kolokolnikova, 1996; Dykhta and Samsoniuk, 2000).

3. Optimality of singular and sliding modes

The standard transformation of (1) to the relaxed system (see Warga, 1962)

$$\dot{x} \in \mathbf{V}_C(t, x) = \overline{\text{co}}\, \mathbf{V}(t, x), \tag{7}$$

allowes one to describe in terms of piecewise smooth solutions all regular, singular and sliding modes of (1). Recall that a smooth $x(t)$ is called an $\mathbf{F}$-singular mode of (1) on $\mathbf{T} = [t_I, t_F]$ if $\dot{x} \in \text{int}\, \mathbf{F}(t, x(t))$ (Fig. 2), where $\mathbf{F}$ is a face of $\mathbf{V}_C$ (see Webster, 1994). In a series of works (see Kelley, 1964b, Kelley, Kopp, and Moyer, 1967, Gabasov and Kirillova, 1973) some necessary optimality conditions for such modes have been derived as additional to degenerate classical Euler-Lagrange equations. The above results can be applied to obtain global sufficient conditions for such types of solutions. This is attained by preliminary extension of the original velocity set $\mathbf{V}(t, x) = f(t, x, \mathbf{U}(t, x))$ to an unbounded one. Then, the corresponding derived problem is solved, and one checks whether the solution belongs to the original admissible set $\mathbf{D}$.

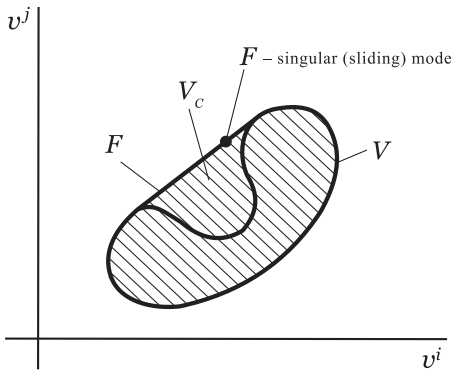

Figure 2.

Let $\mathbf{F} \subset \mathbf{V}_C$ be a face of dimension $k > 0$ and $\mathbf{P}^{(k)}$ be its carrying plane with a basis $h = (h_1, h_2, \ldots, h_k)$. Consider the following auxiliary unbounded system:

$$\dot{x} \in \mathbf{V}_E(t,x) = \mathbf{V}_C(t,x) \cup \mathbf{P}^{(k)}(t,x). \tag{8}$$

Let the system $dx/d\tau_j = h(t,x)w$, $w \in \mathbb{R}^k$, be a limit system having the integral $y = \eta(t,x) = (\eta^1(t,x)\ldots,\eta^{n-k}(t,x))$ as a continuous and smooth function of t and x, so that $\eta_x h(t,x) = 0$. Then, according to Section 2, (8) can be equivalently transformed to the derived system

$$\dot{y} = \eta_x v + \eta_t, \quad v \in \mathbf{V}_E(t,x), \quad x \in \mathbf{Q}(t,y) \tag{9}$$

where $\mathbf{Q}(t,y) = \{x\colon y = \eta(t,x)\}$. We call (9) the **F**-derived system. It has the following important property (A):

$$\dot{x}(t) \in \mathbf{P}^{(k)}(t,x(t))$$

for any smooth solution $x(t)$ of (9). Indeed, any $v \in \mathbf{P}^{(k)}$ can be represented as $v = v_0 + hw$, $w \in \mathbb{R}^k$, where v_0 is some arbitrary constant point of $\mathbf{P}^{(k)}$. Hence, from the equality $\eta_x h(t,x) = 0$ it follows that

$$\dot{y} = \eta_x v + \eta_t = \eta_x(v_0 + hw) + \eta_t = \eta_x v_0 + \eta_t. \tag{10}$$

This is constant for any constant t and $x(t)$. This means that (11) can be considered (in statics) as the equation of $\mathbf{P}^{(k)}$. On the other hand, $x(t)$ satisfies the equation $y(t) = \eta(t,x(t))$, so that $\dot{y} = \eta_t + \eta_x\dot{x}$, that is, $\dot{x}$ satisfies the equation of $\mathbf{P}^{(k)}$.

It follows from the property A that the set of pairs $(y(t),x(t))$ of all solutions of (10) with smooth $x(t)$ is invariant for replacing $\mathbf{V}_E$ by $\mathbf{V}_C$ in (9). In these terms, global sufficient optimality conditions for singular and sliding modes of (1) are proposed.

Let $\mathbf{F}(t,x)$ be a k-dimensional face of $\mathbf{V}_C(t,x)$ that generates the corresponding **F**-derived system. Define an extension of the admissible set for the problem, taking the derived system (10) instead of (1), and call the result the **F**-derived problem. Any solution of this problem should satisfy the Pontryagin maximum principle, and in particular, the condition of maximum of the linear form $p^T v$ on $\mathbf{V}_C$ where $p^T = q^T\eta_x$, and q is an ajoint $(n-k)$-vector for the derived problem. The equality $\eta_x h(t,x) = 0$ implies that if for some q the maximum is reached at some point $v_0 \in \mathbf{F}$, then all the face $\mathbf{F}$ is the maximum set because it is a priori orthogonal to p. In this case a solution to the **F**-derived problem is a triple $(y_*(t), x_*(t), \mathbf{F}(t,x_*(t)))$. Thus, the following statement holds.

Theorem 2 *Let $(y_*(t), x_*(t), \mathbf{F}(t,x_*(t)))$ be a solution of the corresponding* **F***-derived problem, $x_*(t)$ be smooth, and let $\dot{x}_* \in \mathbf{F}(t,x_*(t))$. Then $x_*(t)$ is a solution of the original problem, an optimal* **F***-singular or* **F***-sliding mode.*

Note that $\dot{x}_* \in \mathbf{P}^{(k)}(t,x_*(t))$ a priori due to the property A, so $(\dot{x}_* \in \mathbf{F}(t,x_*(t)))$ is the condition for some point of a plane to be in the domain of this plane which is a “nonzero probability event” when checking a posteriori.

4. Connection with Krotov sufficient conditions. The method of multiple maxima

Recall the most popular version of Krotov theorem on global sufficient optimality conditions for the problem stated in Section 1 (see Krotov, 1996). We introduce a functional parameter $\varphi(t,x)$, which is a scalar function everywhere continuous and smooth, and the following functions are constructed:

$$R(t,x,u)=\varphi_x^T f(t,x,u)+\varphi_t,\quad \mu(t)=\sup\{R(t,x,u):\ u\in \mathbf{U}(t,x), x\in \mathbf{X}(t)\},$$
$$G(x)=F(x)+\varphi(t_F,x)-\varphi(t_I,x_I),\quad l=\inf\{G(x):\ x\in \Gamma\cap \mathbf{X}(t_F),$$

where (φ_x,φ_t) is the vector of partial derivatives,

Theorem 3 *Let there be a sequence* $\{m_s\}=\{x_s(t),u_s(t)\}\subset \mathbf{D}$ *and a function* φ *such that*

1) $\mu(t)$ *is piecewise continuous;*

2) $\int_{\mathbf{T}}(\mu(t))-R(t,x_s(t),u_s(t))\,dt\to 0$;

3) $G(x_s(t_F))\to l$.

Then this sequence is minimizing, and any minimizing sequence satisfies conditions 2) and 3).

Any function φ, satisfying these conditions is called *resolving (Krotov) function.*

The following assertions are direct corollaries of Theorem 1.

Theorem 4 *Let* $\varphi^{\mathrm{I}}(t,y)$ *be a resolving (Krotov) function for the derived problem obtained via the mapping* $y=\eta(t,x)$ *applied to the free-end problem. Then the superposition* $\varphi(t,x)=\varphi^{\mathrm{I}}(t,\eta(t,x))$ *is a Krotov function for the original problem.*

In particular, if $\varphi^{\mathrm{I}}(t,y)$ is Krotov-Bellman function for the derived problem, satisfying the relations

$$\sup_{\substack{u\in\mathbf{U}(t,x)\\ x\in\mathbf{Q}(t,y)}}\left(\varphi_y^{\mathrm{I}}(\eta_x f(t,x,u)+\eta_t)\right)+\varphi_t^{\mathrm{I}}=0,\quad \varphi^{\mathrm{I}}(t,y)=-F^I(y),\tag{11}$$

then the original Krotov-Bellman function satisfies the relations

$$\sup_{\substack{u\in\mathbf{U}(t,x)\\ x\in\mathbf{Q}(t,y)}}\left(\varphi_x^{\mathrm{T}} f(t,x,u)-\varphi_t\right)=0,\qquad \sup_{x\in\mathbf{Q}(t_F,y)}(\varphi(t_F,x))-F(x))=0\tag{12}$$

and is first integral of the limit system. One can consider these relations as *generalized Bellman conditions* whereas the ordinary ones are not valid in this case. They can be used effectively to construct the feedback optimal control laws (see Gurman and Ukhin, 2005), which is illustrated in Section 5.

There is a way to specify Krotov function for the degenerate problems (*method of multiple maxima*), which is tightly connected with the above considered transformations. It is based on a special system of partial equations and inequalities as follows.

Let for every (t,x) a family of faces $\{\mathbf{F}(t,x)\}$ of $\mathbf{V}_C(t,x)=\overline{co}\mathbf{V}(t,x)$ be given. Consider the following condition

$$p^T\mathbf{F}=\sup_{v\in\mathbf{V}_C(t,x)} p^T v,\quad \mathbf{F}\in\{\mathbf{F}_\alpha(t,x)\}, \tag{13}$$

which defines the conjugate cone $\mathbf{L}(t,x)$ for this family. Require that function $\varphi(t,x)$ satisfy the condition $\varphi_x\in\mathbf{L}(t,x)$. Then, (13) will be called the *multiple maxima system (MMS)* w.r.t function $\varphi(t,x)$ and the family $\{\mathbf{F}_\alpha\}$.

Select $k+1$ base points $\{v_0,\ldots,v_k\}$ out of the face $\mathbf{F}$ of the base set $\mathbf{V}^{(k)}$. Then, (13) reduces to the following system

$$\begin{aligned}&p^T\left(v_l(t,x,p)-v_0(t,x,p)\right)=0,\\&p^T\left(v_0(t,x,p)-v\right)>0,\quad v\in\mathbf{V}_C(t,x,p)\backslash\mathbf{F}(t,x,p).\end{aligned} \tag{14}$$

By denoting $H_l(t,x,p)=p^Tv_l(t,x,p)=p^Tf(t,x,u_l(t,x,p))$, the system (14) can be rewritten in the form (which explains its name):

$$\mathcal{F}_l(t,x,p)=H_l(t,x,p)-H_0(t,x,p)=p^Th_l(t,x,p)=0,\quad l=1,\ldots,k. \tag{15}$$
$$H_0(t,x,p)=\max_{u\in\mathbf{U}(t,x)}H(t,x,p,u)=\mathcal{H}(t,x,p),$$

$h_l=v_l-v_0$, $h=(h_1,\ldots,h_k)$ (a matrix) is an internal basis of $\mathbf{F}$. These conditions mean that the face $\mathbf{F}(t,x,p)$ gives the strict maximum p^Tv w.r.t. all points of $\mathbf{V}_C(t,x)$ not belonging to $\mathbf{F}$.

Let function $\varphi(t,x)$ in sufficient conditions of Section 3 be specified by the linear involute MMS

$$h^T(t,x)\varphi_x=0,\quad \left(\varphi_x^Th_l(t,x)=0,\quad l=1,2,\ldots,k\right), \tag{16}$$

for which (14) reduces to when $\{\mathbf{F}_\alpha\}$ consists of one face. Then, $\varphi(t,x)=\varphi^{\mathrm{I}}(t,\eta(t,x)$ where $\varphi^{\mathrm{I}}(t,y)$ is an arbitrary smooth function, $\eta(t,x)$ is an integral of the system

$$\frac{\partial x}{\partial z}=h(t,x)\quad\left(\frac{\partial x}{\partial z^l}=h_l(t,x)\right),$$

and R and G are transformed automatically to corresponding constructions for the derived problem:

$$\begin{aligned}&R^{\mathrm{I}}(t,x,y,u)=\left(\varphi_y^{\mathrm{IT}}\right)^T(\eta_xf(t,x,u)+\eta_t)-f^0+\varphi_t^{\mathrm{I}},\quad y=\eta(t,x),\\&G^{\mathrm{I}}(y)=F^{\mathrm{I}}(y)+\varphi^{\mathrm{I}}(t_F,y)-\varphi^{\mathrm{I}}(t_I,y_I),\\&F^{\mathrm{I}}=\inf\{F(x):\ x\in\mathbf{\Gamma}\cap\mathbf{X}(t_F)\cap\mathbf{Q}(t_F,y)\},\end{aligned}$$

(with φ^{I} playing the same part as φ in the original problem).

The MMS is not always compatible. The general compatibility condition is that it admits a completion. In Antipina and Dykhta (2004) it is shown that the completion procedure leads to an optimal problem for the derived system in the above considered "noncommutative" case of the limit system.

In general case the following implicit scheme is proposed. Consider the equality conditions in (14) as involute partial differential equations. The corresponding characteristic system is

$$\begin{aligned} \frac{\partial x}{\partial z} &= \mathcal{F}_p(t,x,p), \quad \frac{\partial p}{\partial z} = -\mathcal{F}_x^T(t,x,p), \\ \frac{\partial q}{\partial z} &= -\mathcal{F}_t^T(t,x,p), \quad z = z^1,\ldots,z^k. \end{aligned} \tag{17}$$

Consider Cauchy problem for system (17). Specify an $(n-k)$-parametric initial manifold $\mathbf{S}(t)$:

$$x = \kappa(t,y), \quad y = y^1,\ldots,y^{n-k}$$

and an initial function $\varphi^{\mathrm{I}}(t,y) = \varphi(t,\kappa(t,y))$, so that

$$\varphi_y^{\mathrm{I}} = p^{\mathrm{I}} = \kappa_y^T\varphi_x = \kappa_y^T p, \quad \varphi_t^{\mathrm{I}} = q^{\mathrm{I}} = \varphi_t + \kappa_t^T\varphi_x = q + \kappa_t^T p. \tag{18}$$

At the points of $\mathbf{S}(t)$ the variables x, p, q are expressed directly from (16), (18). Consider the resulting solution of (17)

$$\begin{aligned} &x = x\left(t, x_S\left(t,y,p^{\mathrm{I}}\right), z\right) = x\left(t,y,p^{\mathrm{I}},z\right), \\ &p = p\left(t, x_S\left(t,y,p^{\mathrm{I}}\right), p_S\left(t,y,p^{\mathrm{I}}\right), z\right) = p\left(t,y,p^{\mathrm{I}},z\right), \\ &q = q\left(t, x_S\left(t,y,p^{\mathrm{I}}\right), p_S\left(t,y,p^{\mathrm{I}}\right), z\right) = q^{\mathrm{I}} - \int_0^z h_t d\xi, \end{aligned}$$

where initial value $z = 0$ corresponds to $\mathbf{S}(t)$. Substitution of these functions into original sufficient conditions by taking $p = \varphi_x$, $q = \varphi_t$, leads to the constructions quite similar to original ones. Subsequent investigation with their help can be carried out by general methods. For example, for the free-end problem it is possible to apply Bellman scheme to determine $\varphi^{\mathrm{I}}(t,y)$, and hence $\varphi(t,x)$. In Gurman (1977) this implicit scheme was used to obtain concrete strong local optimality conditions for the degenerate problems.

5. Applications to periodical systems

There exist problems especially characteristic for the space applications where the time of control is not strictly limited, the limited resource being represented by the stock of the jet working fluid. In this connection we consider the design of the optimal control of a sufficiently general periodic system over an unlimited time interval using the criterion of the norm-minimal control in $\mathbb{L}^1$, considered as the control resource. Consideration is given to the control system obeying the following sufficiently general model:

$$\dot{x} = h(x) + \sum_j B_j u_j, \quad x \in \mathbb{R}^n, \quad u_j \in \mathbf{U}_j \subset \mathbb{R}^{p_j}, \tag{19}$$

$$\dot{z} = \sum_j |u_j|, \tag{20}$$

where u_j are the vector controls, whose available sets $\mathbf{U}_j$ have zero values. The value z is called the control resource. Equation (1) for $u_j = 0$, that is, $\dot{x} = h(x)$, describes the periodic motions.

This model embraces various special cases and applications, such as control of elastic oscillations, Gurman and Znamenskaya (2001), spacecraft orbital and orientation maneuvering, and control of the "predator-prey" system, Krotov and Gurman (1973), Gurman (1977), Gurman and Ukhin (2006). The following problem of optimal control is formulated on the basis of model (19), (20): it is required to establish feedback control law (positional control) $u(x,z) = \{u_j(x,z)\}$ driving the system from any initial state $(x(0), z(0))$ to the final state $(x(t_F), z(t_F))$ with the least value of the functional $I = |(x(t_F) - \bar{x}(t_F)|^2$, $z(0) = 0$, $z(t_F)$, and $\bar{x}(t_F)$ being given values. The finite time t_F is not fixed. If, in particular, $\inf I = 0$, then the problem of driving the system from $(x(0), z(0))$ to $(\bar{x}(t_F), z(t_F))$ with the least $z(t_F)$ is solved.

The approach consists of the following stages:

(I) The argument in the form of time variable is replaced by a new argument, the control resource (nondecreasing time function by virtue of (20)). As the result, time is eliminated, and a system with unbounded-control $v = \left(\sum_j |u_j|\right)^{-1}$ is obtained:

$$x' = vh(x) + \sum_j B_j l_j, \quad l = \{l_j\} \in \mathbf{\Lambda}$$

where $l_j = \left(\sum_j |u_j|\right)^{-1} u_j$, $v_{\min} \le v < \infty$, and the bounded set $\mathbf{\Lambda}$ is obtained by direct recalculation of the collection $\{\mathbf{U}_j\}$, $v_{\min} = \min v / u_j \in \mathbf{U}_j$.

(II) The set of $n-1$ independent first integrals $y = \eta(x)$, $\left(y^k = \eta^k(x)\right)$ of the periodic passive motion is determined according to the above theory, and the transition is made to the derived system

$$y' = \eta_x \sum_j B_j l_j, \quad \eta(x) = y, \quad l \in \mathbf{\Lambda}, \quad \eta_x = [\eta^k_{x^j}]$$

and, correspondingly, to the derived problem of the minimum of the same functional I.

(III) The derived problem is solved in the form of (exact or approximate) optimal control design.

(IY) The procedure is applied of constructing the minimizing sequence of the positional controls in the original problem, Gurman and Ukhin (2005), on the basis of information from item (III) with the upper estimate of the accuracy of its elements (because there exist no optimal solution in the traditional sense of the word).

Having carried out the obvious preliminary operation of partial minimization in the original expression of the functional, we get a new minimized functional of the derived problem

$$J = F^{\mathrm{I}}\left(y(z_F)\right) = \min_t |(\xi(y,t) - \bar{x}_F|^2, \tag{21}$$

where $\xi(y,t)$ is the trajectory of passive periodic motion. It is possible to solve it by means of generalized Bellman conditions

$$\max_{l,t} \varphi_y^{\mathrm{IT}} \eta_x \sum_j B_j l_j + \varphi_{y^z}^{\mathrm{I}} = 0, \quad \varphi_y^{\mathrm{I}}(z_F, y) = F^{\mathrm{I}}(y), \tag{22}$$

where $\varphi_y^{\mathrm{I}}(z,y)$ is the Krotov-Bellman function.

Let us assume that Eq. (22) was solved precisely or approximately and provided, along with (21), the functions $l_*\left(z,x\right)$ and $x_*\left(z,y\right)$. Let manifold $\mathbf{S}_* = \{(z,x) : \ x = x_*\left(y\right), \ y \in \mathbb{R}^{n-1}\}$ be the graph of the latter function. We assume for simplicity that it is continuous. The function $v_*\left(z,y\right)$ remains indefinite and may be defined arbitrarily within its constraints $v_{\min} \leq v < \infty$. In that way we get a solution of the derived problem, defining the minimizing sequence of the original problem (with the argument z) constructed according to the general rule, described above. In this case, it can be interpreted in model (19)–(20) with original argument t, which simplifies the rule.

Let us define the $(1/s)$-neighborhood $\mathbf{S}_*$ ($(1/s)$-layer) (where s is the number of the term in the sequence):

$$\mathbf{G}\left(\mathbf{S}_*, 1/s\right) = \{(z,x) : |x - x_*\left(z,y\right)| < 1/s\}.$$

For $(z,x) \in \mathbf{G}\left(\mathbf{S}_*, 1/s\right)$, we assume for definiteness that $l_s\left(z,x\right) = l_*\left(t,x\right)$, $v = v_{\min}$. Otherwise, a point (z, x_*) on $\mathbf{S}_*$, is determined such that for a given z the points x_* and x are connected by a passive-motion trajectory. As the result, we obtain for a given s the following control law $u_s(z,x) = \{u_{js}\}(z,x)$:

$$u_{js}(z,x) = \begin{cases} 0, & |x - x_*\left(z,y\right)| > 1/s, \\ l_{j*}\left(z,x\right)/v_{\min}, & |x - x_*\left(z,y\right)| \leq 1/s. \end{cases}$$

The time t_F grows unlimitedly with s, and for a given s the greater $v_{\min}$, the smaller t_F.

Despite some unhandiness of the above construction of $u_s(z,x)$, it reflects sufficiently simple and obvious rule: from any point (z,x) lying beyond the manifold $\mathbf{S}_*$ one should move to the manifold along the trajectory of passive motion and use the control $l_*(t,x)$ $v = v_{\min}$ in its $(1/s)$-neighborhood.

Example 2 Optimal damping of satellite oscillations.

Consider the equations of angular motion of a gravitationally stabilized satellite on a circular orbit under the gravity and control torques (see Beletskii, 1966)

$$\dot{\theta} = q, \quad \dot{q} = -3(\Omega)^2\beta\sin\theta\cos\theta + la, \quad \dot{z} = a, \tag{23}$$

where θ is the angle between satellite axis and the local vertical, Ω is the constant orbit angular speed of the satellite w.r.t. the planet center, β is the gravity constant, l and a are direction and value of the angular acceleration developed by the engine ($l = \pm 1,\ a \le a \le a_{\max}$). The problem is to damp the angular oscillations with the minimum fuel consumption, proportional to z.

Transform the system (23) to the new argument z:

$$\frac{d\theta}{dz} = \frac{1}{a}q, \quad \frac{dq}{dz} = -\frac{1}{a}3(\Omega)^2\beta\sin\theta\cos\theta + l.$$

The corresponding limit system trajectories are the same as of (23) when $a = 0$,

$$y = \frac{1}{2}(q)^2 + \frac{3}{2}(\Omega)^2\beta(\sin\theta)^2 \tag{24}$$

(closed curves for constant y). When $y = 0$, then there are no oscillations ($\vartheta = q = 0$). The corresponding derived problem is

$$\frac{dy}{dz} = ql, \quad y(z_I) = y_I(q_I,\theta_I), \quad I = z_F,$$

under condition (24). The solution is found evidently by minimizing the right hand side of above differential equation. After integration one obtains

$$y = (\sqrt{y_I} - z/\sqrt{2})^2,$$

and $z_F = \sqrt{2y_I}$ when $y = 0$. This solution does not satisfy directly the original system and could be realized as an impulse sliding mode by a sequence of $\mathbf{D}$, as shown in Fig. 3. In this case it is active braking (for example at $a = a_{\max}$) in the vicinity of equilibrium point. When decreasing this vicinity, the functional goes to $\sqrt{2y_I}$, and the number of active modes grows infinitely.

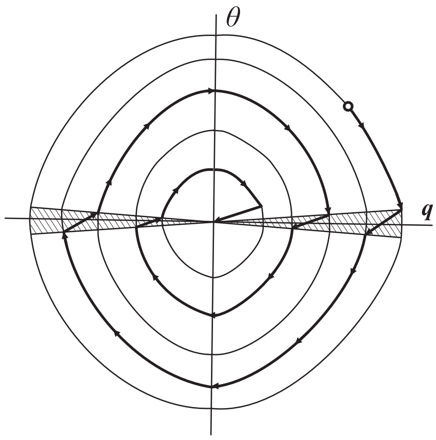

Figure 3.

6. Turnpike solutions and approximate schemes

For the class of systems with unbounded linear control, typical generalized solutions have piecewise continuous trajectories $x(t)$ that contain a finite number of impulses on any bounded time interval.

Consider as a special case of system (1) the following control system

$$\dot{x} = g\,(t, x, u_1) + h\,(t, x)\,u_2, \quad u_1 \in \mathbf{U_1}(t, x), \quad u_2 \in \mathbb{R}^k,$$

where $k \leq n$ and h, is an $n \times k$ matrix of rank k. Its limit system

$$\frac{dx}{d\tau} = h\,(t, x)\,u_2, \quad u_2 \in \mathbb{R}^k,$$

is assumed to have $(n-k)$-vector integral $y = \eta(t, x)$, leading to the derived system

$$\dot{y} = \eta_x g\,(t, x, u_1) + \eta_x, \quad u \in \mathbf{U}(t, x), \quad x \in \mathbf{Q}(t, y) = \{x : y = \eta\,(t, x)\}.$$

Reduction of the order means that in the general case solution of the derived problem cannot satisfy regularly all the boundary conditions of the original problem. The trajectory turns out to be invariant to a certain set of these conditions so that they are satisfied by discontinuously jumping in time, which in practice is realized by great control actions. We borrow a term from the theory of economic growth and call such a solution of the original problem the *turnpike solution.* Each continuous section of $x(t)$ satisfying the original systems is called a *turnpike*. Stated differently, this solution can be treated as a motion along turnpikes with fast (instantaneous in the limit terms) transitions from one turnpike to another (Fig. 4).

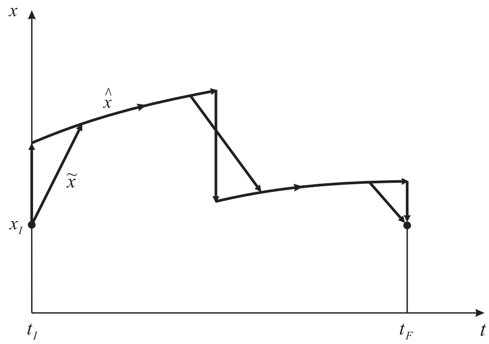

Figure 4.

The connection between original and derived systems can be expressed explicitly by passing to new variables $y = \eta(t, x)$, $z = \zeta(t, x)$ (it is known from Gaishun, 1983, that there exists a one-to-one transformation of this kind). In doing so, the original system (1) comes to $\dot{y} = g^y(t, y, z, u_1)$, $\quad u \in \mathbf{U}_1(t, y, z)$ and

$$\dot{z} = g^z(t, y, z, u) + h^z(t, y, z)\, u_2, \tag{25}$$

and in the new variables the derived system is obtained by eliminating Eq. (25).

Under sufficiently great u_2, the minimizing sequence is constructed as a standard approximation of the piecewise continuous function $z(t)$ by a sequence of piecewise smooth trajectories of the original system in the neighborhood of each discontinuity point. Outside these neighborhoods, u_2 is established from the condition for strict satisfaction of the original differential relation, that is,

$$u_2(t) = (h^z){-}1\,(t, x(t))\,(\dot{z}(t) - g^z(t, x(t), u(t)))\,.$$

Let now u_2 be bounded: $u_2 \in \mathbf{U}_2(t, x)$, but at the same time $\mathbf{U}_2(t, x)$ be a domain in $\mathbb{R}^k$ admitting passages between the boundary values and their closest turnpikes and between the turnpikes (Fig. 4) in a time that is small enough as compared with the given time $t_F - t_I$. Then, one can consider an approximate turnpike solution $m \in \mathbf{D}$ with the upper bound

$$I(m) - d \le \Delta = I(m) - e,$$

which is the more precise, the smaller the total number of transitions. Such a solution can be further refined, if necessary, by one or another iterative procedure. Such an approximate approach proved to be very useful in applications (see Krotov, Bukreev and Gurman, 1971; Gurman, 1985).

On the whole, order reduction is bigger for higher numbers of linear controls (k). Moreover, if it turns out that the derived system has the specific structure of the original system, that is, is linear relative to one or another set of new controls, then it can be transformed to a new derived system by the same scheme. This multistage procedure may result in a dramatic reduction of the problem order right down to one or zero. In this connection, the class of control problems associated with contemporary sustainable development paradigm (see Gurman, 1981, 2006, Gurman and Ryumina, 2001) are representative because of presence great amount of linear controls comparable with state dimensions.

Analogues for discrete-time systems have been also developed. In particular, for systems with the unbounded linear control

$$x(t+1) = f(t, x(t), u) + h(t)v, \quad t \in \mathbf{T} = \{t_I, t_I + 1, \ldots, t_F\} \quad u \in \mathbf{U}(t, x),$$

transition to the derived problem is done by the linear transformation $y = \Lambda(t)x$, where $\Lambda(t)$ is defined so that $\Lambda(t+1)h(t) = 0$. The derived system is as follows:

$$y(t+1) = \Lambda(t+1)g(t, x, u), \quad u \in \mathbf{U}(t, x), \quad x \in \{x : y = \Lambda(t)x\}.$$

7. Conclusions

As follows from the aforementioned, degeneracy of the optimal control problems and the turnpike type of solutions related with it can be used successfully to construct control optimization procedures, especially at the stage of seeking good initial approximations, to be refined later by regular iterative algorithms that work effectively near the desired solution. The proposed approach lies in seeking and eliminating passive differential relations or discrete chains, existing in the formulations of degenerate problems, which — as attested by rich practical experience — leads to their regularization and simplification owing to the lower, sometimes radically lower, order of the problem. This refers first to the relatively new application areas such as socio-ecologico-economic applications. Thus, the "evil" of degeneracy for the general methods turns into a "blessing" for special methods based on the possibility to lower the order of the problem with its simultaneous regularization.

This work is financially supported by the Russian Fund of Fundamental Research, project 09-01-00170.

References

Agrachev, A.A. and Sachkov, Yu.L. (2004) *Control Theory From the Geometric Viewpoint*. Springer-Verlag, New York.

Antipina, N.V. and Dykhta, V.A. (1969) Sufficient Optimality Conditions for Impulse Control Problems. *Journal of Computer and Systems Sciences International* c/c of *Tekhnicheskaia kibernetika* **43** (4), 566-573.

Baturin V.A., Dykhta V.A., Gurman V.I. et al. (1987) *New control improvement methods*. Nauka, Novosibirsk (in Russian).

Beletskii, V.V. (1966) *Motion of an artifical satellite about its center of masses.* Israel Program for Scientific Translations, Jerusalem.

Dykhta, V. A. (1981) Conditions of local minimum for singular modes in the systems with linear control. *Aut. Remot. C.* **12**, 5-10.

Dykhta, V.A. and Samsoniuk, O.N. (2000) *Optimal impulse control with applications.* Fizmatlit, Moscow (in Russian).

Gabasov, R. Kirillova F.M. and Kirillova, F.M. (2000) . *Optimal impulse control with applications.* Fizmatlit, Moscow (in Russian).

Gaishun, I.V. (1983) *Completely resolvable differential equations.* Nauka i tekhnika. Minsk (in Russian).

Gurman, V.I. (1965) On optimal processes of singular control. *Aut. Remot. C.* **26** (5), 782-791.

Gurman, V.I. (1967) Method of multiple maxima and conditions of relative optimality of degenerate regimes. *Aut. Remot. C.* **12**, 1845-1852.

Gurman, V.I. (1972) On optimal processes with unbounded derivatives. *Aut. Remot. C.* **12**, 14-21.

Gurman, V.I. (1977) *Degenerate optimal control problems.* Nauka, Moscow (in Russian).

Gurman, V.I., ed. (1981) *Models of Managing Natural Resources.* Nauka, Moscow (in Russian).

Gurman, V.I. (1985) *The extension principle in control problems.* Nauka, Moscow (in Russian).

Gurman, V.I. (1998) *The extension principle in control problems. General theory and learning examples.* Nauka, Fizmatlit, Moscow.

Gurman, V.I. and Dykhta, V.A. (1977) Singular problems of optimal control and the multiple maxima method. *Aut. Remot. C.* **3**, 343-350.

Gurman, V.I. and Ryumina, E.V., eds. (2001) *Modeling of the Regional Socio-ecology-economic systems.* Nauka, Moscow (in Russian).

Gurman, V.I. and Sachkov, Yu.L. (2008) Representation and realization of generalized solutions of control systems with unbounded hodograph. *Aut. Remot. C.* **4**, 72-80.

Gurman, V.I. and Ukhin, M.Yu. (2005) Design of Optimal Control in Systems with Unbounded Set of Velocities. *Diff. Uravn.* **49** (11), 1482-1490.

Gurman, V.I. and Ukhin, M.Yu. (2006) *The extension principle in control problems. Constructive methods and applied problems.* Fizmatlit, Moscow.

Gurman, V.I. and Znamenskaya, L.N. (2005) Oscillation control under limited control resource. *Izv. Ross. Akad. Nauk, Teoriya i Sist. Upravlen.* **1**, 41-49.

Kelley, H.J. (1964a) A transformation approach to singular subarcs in optimal trajectory. *SIAM J. Control* **2** (2), 234-240.

Kelley, H.J. (1964b) A second variation test for singular extremals. *AIAA J.* **2** (1) 26-29.

KELLEY, H.J., KOPP, R.E. and MOYER, H.G. (1967) Singular extremals. In: G. Leitman, ed., *Topics in Optimization.* Academic Press, New York-London, 63-101.

KOLOKOLNIKOVA, G.A. (1996) Discontinuous trajectories optimality in the nonlinear optimal control problems. *Prepr. 13th World Congress IFAC. Nonlinear Systems I* **E**, 353-357.

KROTOV, V.F. (1996) *Global Methods in Optimal Control.* Marcel Dekker, New York.

KROTOV, V.F. (1961) Discontinuous solutions of variational problems. *Izv. Vuzov. Matematika* **1**, 86-98 (in Russian).

KROTOV, V.F., V.I. BUKREEV, V.Z. and GURMAN, V.I. (1971) *New variational methods in flight dynamics. NASA Transl. TTF-657.* Israel Program for Scientific Translations, Jerusalem.

KROTOV, V.F. and GURMAN, V.I. (1973) *Methods and problems of optimal control.* Nauka, Moscow (in Russian).

WARGA, J. (1962) Relaxed variational problems. *J. Math. Anal. and Applic.* **4** (1), 111-128.

WEBSTER R. (1994) *Convexity.* Oxford University Press, New York-Tokyo.