On the development of Pontryagin's Maximum Principle in the works of A.Ya. Dubovitskii and A.A. Milyutin*

by

Andrei V. Dmitruk

Russian Academy of Sciences,
Central Economics & Mathematics Institute
e-mail: dmitruk@member.ams.org

Abstract: We give a short review of the development and generalizations of the Pontryagin Maximum Principle, provided in the studies of Dubovitskii and Milyutin in the 1960s and later years.

Keywords: Euler–Lagrange equation, Maximum Principle, classes of variations, Pontryagin minimum, state constraints, measure in adjoint equation, regular and nonregular mixed constraints, phase points, closure with respect to measure, three-storey theorem.

1. Introduction

The discovery of Maximum Principle (MP) by L.S. Pontryagin and his students V.G. Boltyanskii and R.V. Gamkrelidze in 1956–58 (see a description in Gamkrelidze, 1999, and the preceding history in Sussmann and Willems, 1997, and in Pesch and Plail, 2009), and especially the publication of the book by Pontryagin, Boltyanskii, Gamkrelidze and Mischchenko (1961), gave a powerful impetus to an explosive development of the theory both of the optimal control itself, and of extremum problems in general. An intriguing thing was that, although the optimal control problems generalize the problems of classical calculus of variations (CCV), the proposed method of their analysis seemed quite different from the methods of CCV, and the relation between them was not clear. The common opinion was that the authors created a "new calculus of variations".

Since the original proof proposed in Pontryagin et al. (1961) was quite uneasy both in its ideas and technique (at least requiring the university level of mathematical background, and unaccessible for those having just engineering level of mathematical education), a large number of attempts were made to simplify the proof. On the other hand, many attempts were made to generalize the MP to broader classes of optimal control problems.

Not pretending to review the entire wealth of papers devoted to investigation and development of MP, which is hardly a realizable task, we consider here its central, to our opinion, line elaborated by Abram Yakovlevich Dubovitskii and Alexei Alexeevich Milyutin during the 1960s and later years (mentioning just a few papers of other authors, closely related to this line).

In their works, the following results were obtained:

a) A general approach to deriving necessary optimality conditions of the first order in a very broad class of abstract extremum problems with constraints (the so-called *Dubovitskii–Milyutin scheme*), which results in a *stationarity condition*, called the (abstract) Euler–Lagrange equation. In particular, for smooth problems with constraints of equality and inequality type, this condition readily yields the Lagrange multipliers rule, and for problems of CCV it yields the classical Euler–Lagrange (EL) equation. This scheme turned out to be universal and, because of its simplicity and transparency, very popular.

b) To analyze the optimality of a given process, they proposed to consider a family of *associated smooth problems*, in each of which one should write out the corresponding stationarity condition (EL equation), and then in a sense "press" all these conditions into one mutual condition, the MP.

Thus, the basis of all necessary conditions of optimality is the stationarity condition — the Euler–Lagrange equation, while the MP is obtained as a result (squeezing extract) of a family of these equations. In this sense, the relation between MP and CCV was restored and clarified.

c) The above approach was applied in Dubovitskii and Milyutin (1965) to a more general class of optimal control problems, those including *state constraints* $\Phi(x, t) \leq 0$, which made it possible to obtain a generalization of Pontryagin's Maximum Principle.[1] Its formulation was essentially new, because the adjoint equation involved a measure in the right hand side.

d) Almost simultaneously with the state constraints, Dubovitskii and Milyutin considered also the *mixed state-control constraints* $\varphi(x, u, t) \leq 0$ and $g(x, u, t) = 0$ under the assumption of their *regularity*, and also obtained the corresponding generalization of Pontryagin MP.

e) Finally, they also considered the general case of mixed constraints, without the regularity assumption. However, this turned out to be a much more difficult problem than all the preceding ones. Suffice to say that up till now (by 2009), as far as we know, apart from Dubovitskii and Milyutin, only very few authors have had the courage to consider such problems: Ter-Krikorov (1976), Dyukalov (1977), Chukanov (1977, 1990), Yakubovich and Matveev (1994), probably no one else.

---

[1] In earlier works of Gamkrelidze, based on another approach, only partial result was obtained. In particular, the sign of the jump of the costate variable at junction points was not specified; see Pontryagin et al. (1961).

The matter is that the Lagrange multipliers at the mixed constraints are linear functionals on the space L_∞, and it is well known that the space L_∞^* of such functionals is "very bad": its elements can contain *singular components*, which do not admit conventional description in terms of functions.

Nevertheless, even for such problems, Dubovitskii and Milyutin, after several years of intensive study, and using a very nontrivial technique, obtained a generalization of MP. However, the result was quite unexpected: it turned out that, for a given optimal process, there is a family of "partial" MPs, partially ordered in their "power", which does not reduce to a usual "common" (the most powerful) MP. The last one can exist only as an exception, in some specific cases. This multiplicity (or uncertainty) is a "price" for the nonregular mixed constraints.

It is worth noting that this uncertainty does not concern the so-called *local maximum principle* (or Euler–Lagrange equation), which is the condition of stationarity in the class of uniformly small variations.

Due to the works of Dubovitskii and Milyutin one can say that the question of obtaining first order optimality conditions (MP) in optimal control problems with ODEs is now completely solved. What remains to do for those problems is to study the MP itself, i.e. to analyze its conditions and relations between them, to specify it in different particular cases, etc., thus transforming it from the goal of investigation into a working tool, and apply it to concrete problems.

(Note that both authors themselves were extremely skillful in the usage of MP in various problems: in "deciphering" its conditions and extracting nontrivial information.)

f) In the process of studying the extremum problems, both by Dubovitskii and Milyutin, and by many other authors, the fundamental role of convex structures for the theory of extremum was revealed, which led to a rapid development of the *convex analysis* (the name was proposed by the book of Rockafellar, 1970), that earlier had been usually considered as just a special branch of geometry, and since that time have been "promoted" to be *the base of optimization theory.*

From the very beginning of their research, Dubovitskii and Milyutin discovered and repeatedly used convex structures. They made an essential contribution to convex analysis, including the famous theorem on nonintersection of the convex cones, the theorem on the subdifferential of the maximum of convex functions, the formulas for the subdifferential of max $x(t)$ in the space $C[0,1]$ and vraimax $u(t)$ in the space $L_\infty[0,1]$, the notion of upper convex approximation of a function at a point, etc.

Let us now pass to deeper details.

2. General abstract problem with equality and inequality constraints

In a Banach space X consider the problem

$$f_0(x) \longrightarrow \min, \qquad f_i(x) \le 0, \quad i = 1, ..., k, \qquad g(x) = 0, \tag{1}$$

with scalar functions $f_i : X \to \mathbf{R}$, and an operator $g : X \to Y$ mapping to a Banach space Y. Let be given an open set $\mathcal{O} \subset X$, and a point $x_0 \in \mathcal{O}$ analyzed for a local minimum. Suppose that

(a) all f_i are Lipschitzian in $\mathcal{O}$ and have directional derivatives $f_i'(x_0, \bar{x})$ for any $\bar{x} \in X$, that are sublinear functions in $\bar{x}$;

(b) the operator g has a strict derivative at x_0, whose image $g'(x_0) X$ is a closed subspace in Y.

Without loss of generality assume that $f_0(x_0) = 0$. Introduce the set of active indices $I = \{ i \ge 0 \mid f_i(x_0) = 0 \}$. Note that automatically $0 \in I$. Consider first the main, nondegenerate case, when $g'(x_0) X = Y$ (the Lyusternik condition).

The Dubovitskii–Milyutin scheme for the analysis of problem (1) consists in the following two steps (and two theorems).

THEOREM 1   *If x_0 is a local minimum, then there can be no $\bar{x}$ such that*

$$f_i'(x_0, \bar{x}) < 0 \qquad for~all \quad i \in I, \tag{2}$$
$$g'(x_0) \bar{x} = 0. \tag{3}$$

Proof. Suppose such an $\bar{x}$ does exist. By the Lyusternik theorem it is tangent to the level set $g(x) = 0$ at x_0, i.e. there is a correction vector $\tilde{x}_\varepsilon$ with $||\tilde{x}_\varepsilon|| = o(\varepsilon)$ as $\varepsilon \to 0$, such that the corrected point $x_\varepsilon = x_0 + \varepsilon \bar{x} + \tilde{x}_\varepsilon$ satisfies $g(x_\varepsilon) = 0$. Then, (2) implies that for small $\varepsilon > 0$

$$f_i(x_\varepsilon) = \varepsilon f_i'(x_0, \bar{x}) + o(\varepsilon) \le -c\varepsilon, \qquad \forall\, i \in I,$$

where $c > 0$ is some constant. Thus, the point x_ε satisfies all equality and inequality constraints, and gives a smaller value of the cost, a contradiction with local minimality of x_0. ■

The system (2), (3) is the intersection of a finite number of open convex cones $f_i'(x_0, \bar{x}) < 0$ and the subspace $g'(x_0) \bar{x} = 0$. This situation is a particular case of a general one considered in the following

THEOREM 2   *Let $\Omega_1, \ldots, \Omega_m$ be nonempty open convex cones, and H be a nonempty convex cone. Then $\Omega_1 \bigcap \ldots \bigcap \Omega_m \bigcap H = \emptyset \iff$ there exists a collection $p_i \in \Omega_i^*$, $i = 1, \ldots, m$, and $q \in H^*$ (from the dual cones), not all zero, such that*

$$p_1 + \ldots + p_m + q = 0. \tag{4}$$

The authors called the last relation the Euler equation, but perhaps, as we will soon see, a more proper name for it would be the (abstract) Euler–Lagrange equation.

Thus, analyzing a local minimum in Problem (1), we made two steps: 1) we passed to conical approximations of the cost and constraints (more precisely, of the sublevel sets of the cost and inequality constraints and of the level set of equality constraint), that should not intersect, and 2) rewrote equivalently the nonintersection of these cones in the primal space as some equation in the dual space.

The advantage of writing equation (4) in dual variables is that, for $\Omega_i = \{\bar{x} \mid f_i'(x_0, \bar{x}) < 0\}$, if it is nonempty, one has $p_i = -\alpha_i x_i^*$, where $\alpha_i \geq 0$ and $x_i^* \in \partial f_i'(x_0, \cdot)$; in the smooth case simply $p_i = -\alpha_i f_i'(x_0)$; and for $H = \ker g'(x_0)$ with surjective $g'(x_0)$ one has $q = -y^* g'(x_0)$ with some $y^* \in Y^*$, so we obtain

$$\sum_{i \in I} \alpha_i x_i^* + y^* g'(x_0) = 0$$

(the stationarity condition). These α_i, y^* are *Lagrange multipliers*.

Summing up we get the following

THEOREM 3   *Let x_0 be a local minimum in Problem (1). Then there exists a collection $(\alpha_0, ..., \alpha_k, x_0^*, ..., x_k^*, y^*)$, where all $\alpha_i \geq 0$, $x_i^* \in \partial f_i'(x_0, \cdot)$ and $y^* \in Y^*$, such that the following conditions hold:*

(i) nontriviality: $\sum_{i=0}^{k} \alpha_i + \|y^*\| > 0$;

(ii) complementary slackness: $\alpha_i f_i(x_0) = 0$, $i = 1, ..., k$;

(iii) Euler–Lagrange equation: $\sum_{i=0}^{k} \alpha_i x_i^* + y^* g'(x_0) = 0$.

(In the degenerate case, when $g'(x_0) X \neq Y$, these conditions obviously hold with all $\alpha_i = 0$ and some $y^* \neq 0$. The case when some $\Omega_{i_0} = \emptyset$, is also trivial with $\alpha_{i_0} = 1$ and $x_{i_0}^* = 0$.)

In case of smooth functions one gets the "classical" *Lagrange multipliers rule:*

$$\sum_{i=0}^{k} \alpha_i f_i'(x_0) + y^* g'(x_0) = 0. \tag{5}$$

Saying here "classical" we mean that this condition is a natural generalization of the classical Lagrange multipliers rule for problems on conditional extremum (with equality constraints) in the finite-dimensional spaces to smooth problems with equality and inequality constraints in spaces of infinite dimension.

Unfortunately and really strange, equation (5) was not explicitly written in Dubovitskii and Milyutin (1965), though, as we just have seen, for smooth problems it immediately follows from the Dubovitskii–Milyutin scheme. Formally, the Lagrange multipliers rule for problems with both equality and inequality constraints, but only in $\mathbf{R}^n$, was first published (two years later) in Mangasarian and Fromovitz (1967).

REMARK 1 An important feature of Problem (1) is that it admits only a finite number of inequality constraints, but arbitrary number, even infinite, of equality constraints that are considered as one united equality constraint. In view of this, inequality constraints can be studied separately and independently one from another and arbitrarily added to the problem (it is even possible to create a "catalog" of inequality constraints), whereas any collection of equality constraints should be studied as a joint unit that can not be split into separate constraints (so, creation of a "catalog" of equality constraints is a much more difficult task because of combinatorial obstacles).

Note also that inequality constraints and the cost come similarly into the necessary conditions (except for the complementary slackness, which is just a formal notation for eliminating the inactive constraints), unlike the equality constraint. Moreover, Assumption a) is the same for the inequality constraints and the cost, while Assumption b) for the equality constraint is quite different. For these reasons, the inequality constraints and the cost are denoted by the same letter f (with corresponding indices), while the equality constraint is denoted by another letter g.

REMARK 2 The Dubovitskii–Milyutin scheme works also (and, in fact, was originally developed) for a more abstract problem of the form

$$f_0(x) \to \min, \qquad x \in Q_i, \quad i = 1, \ldots, m, \qquad x \in M,$$

where all the sets Q_i have nonempty interior and are considered as "inequality constraints", while M is considered as an "equality constraint". Here we do not dwell on this case, referring the reader to the original paper of Dubovitskii and Milyutin (1965) or to the book of Girsanov (1970).

For problems of CCV, the Dubovitskii–Milyutin scheme, complemented with a standard technique like the DuBois-Reymond lemma, gives the classical EL equation, which is a first order necessary condition for a weak minimum.

3.   Canonical optimal control problem of the Pontryagin type

On a time interval $\Delta = [t_0, t_1]$ (a priori non-fixed), consider a control system

$$\dot{x} = f(t, x, u), \qquad u \in U, \tag{6}$$

where $x(t) \in AC^n(\Delta)$, $u(t) \in L^r_\infty(\Delta)$, and denote for brevity the vector of endpoint values $p = (t_0, x(t_0), t_1, x(t_1))$.

Among all trajectories of system (6) satisfying the endpoint constraints

$$F(p) \leq 0, \qquad K(p) = 0 \tag{7}$$

(of arbitrary finite dimensions), minimize the cost functional

$$J = F_0(p) \rightarrow \min. \tag{8}$$

All the functions F_0, F, K are assumed to be smooth, f be smooth in t, x and continuous in u, the set $U \subset \mathbf{R}^r$ be arbitrary.

As we see, the cost (8) and constraints (7) are in the Mayer form. They are called *the endpoint block* of the problem. Relations (6) that include all the pointwise constraints, are called *the control system*.

To make the problem statement more precise, one should also indicate the domains of the admissible variables:

$$p \in \mathcal{P}, \qquad (t, x, u) \in \mathcal{Q}, \tag{9}$$

where $\mathcal{P} \subset \mathbf{R}^{2n+2}$ and $\mathcal{Q} \subset \mathbf{R}^{1+n+r}$ are open sets. These inclusions are not constraints, they are a kind of "lebensraum", domain of the problem, usually not directly indicated, but just implied. The explicit addition of inclusions (9) in the statement of canonical problem was proposed by A.Ya. Dubovitskii.

This class of problems, on the one hand, is quite general — it includes, e.g. both the time-optimal problem and problems with integral cost and integral constraints (the integral terms can be reduced to endpoint terms by introducing additional state variables), and on the other hand, it is very convenient for theoretical studies.

Maximum principle for the canonical problem of Pontryagin type

Introduce the endpoint Lagrange function

$$l(p) = \alpha_0 F_0 + \alpha F + \beta K,$$

where $\alpha_0 \in \mathbf{R}$, $\alpha \in \mathbf{R}^{d(F)}$, $\beta \in \mathbf{R}^{d(K)}$, and the Pontryagin function

$$H(\psi_x, t, x, u) = \psi_x f(t, x, u),$$

where $\psi_x \in \mathbf{R}^n$ is the adjoint vector for the state x. (For any vector a we use the notation $d(a) = \dim a$ proposed by Dubovitskii and Milyutin as a convenient tool for saving letters.)

If a process $(x^0(t), u^0(t))$ defined on $\Delta^0 = [t_0^0, t_1^0]$ provides a strong minimum, then there exists a number α_0, vectors α, β, and Lipschitzian functions $\psi_x(t)$, $\psi_t(t)$ (adjoint, or costate variables) such that the following conditions hold:

a) nonnegativity $\alpha_0 \geq 0, \quad \alpha \geq 0$;

b) nontriviality $\alpha_0 + |\alpha| + |\beta| > 0$;

c) complementary slackness $\alpha_i\, F_i(p^0) = 0, \qquad i = 1, \ldots, d(F),$

d) adjoint (costate) equations

$$-\dot\psi_x(t) = H_x(t, x^0(t), u^0(t)),$$
$$-\dot\psi_t(t) = H_t(t, x^0(t), u^0(t));$$

e) transversality

$$\psi_x(t_0) = l_{x_0}(p^0), \qquad \psi_x(t_1) = -l_{x_1}(p^0);$$
$$\psi_t(t_0) = l_{t_0}(p^0), \qquad \psi_t(t_1) = -l_{t_1}(p^0),$$

f) "energy evolution law"

$$H(t, x^0(t), u^0(t)) + \psi_t(t) = 0 \qquad \text{for almost all } t \in \Delta^0,$$

g) maximality

$$H(t, x^0(t), u) + \psi_t(t) \leq 0 \quad \text{ for all } t \in \Delta^0, \quad u \in U.$$

The two last conditions yield: for almost all $t \in \Delta^0$

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

the maximality condition in the standard form.

Introduce the Hamiltonian

$$\mathcal{H}(t, \psi_x, x) \;=\; \max_{u \in U} H(t, \psi_x, x, u).$$

This one and the Pontryagin function are two different functions, coinciding only along the optimal trajectory. They even have different sets of arguments. (The Hamiltonian does not involve the control!) In classical mechanics, $\mathcal{H}$ is the total energy of the system. In autonomous problems $\psi_t = \text{const}$, so $H = \text{const}$, and ψ_x can be simply denoted by ψ.

Following in fact Milyutin (1990a), condition f) is called here "energy evolution law", because together with the adjoint equation for ψ_t it gives the equation $\dot H = H_t$, which in classical mechanics describes the evolution of the energy of the system. The adjoint equation for ψ_x describes the evolution of the impulse of the system.

REMARK 3 Notation ψ_x and ψ_t, where x and t do not indicate the derivatives, but are just subscripts, were proposed by Dubovitskii and Milyutin. Though seem unusual, they are actually very convenient, especially in problems with many state variables. Moreover, if the problem admits a smooth Bellman function $V(x, t)$, then $\psi_x(t) = V_x(x^0(t), t)$ and $\psi_t(t) = V_t(x^0(t), t)$, so the subscripts indeed turn into the derivatives.

By analogy with the concept of extremals in CCV, Dubovitskii and Milyutin proposed the concept of *Pontryagin extremals*, those satisfying all the pointwise conditions of MP regardless the endpoint ones. Milyutin (1990c) showed that these extremals are invariant w.r.t. a broad class of reformulations of the problem (see also Milyutin and Osmolovskii, 1998 (Part 1, Ch.5)). The optimal process should be found among the Pontryagin extremals (and afterwards, the endpoint and transversality conditions should be taken into account).

Moreover, Milyutin showed that, for problems of CCV, the whole theory of sufficient conditions of a strong minimum, based on the construction of a field of extremals, can be as well developed with the use of Pontryagin extremals (see Milyutin and Osmolovskii, 1998 (Part 1, Ch.4)). An interesting and still open question is whether it is possible to develop a similar theory for optimal control problems.

Note that MP is also invariant w.r.t. reformulations of the problem, whereas the Euler–Lagrange equation is not.

REMARK 4 Recall that the principle of stationary action (less strictly, of minimal action) in the classical mechanics says that the mechanical or physical system moves along extremals of some functional (called action). Milyutin (1990a) proposed to strengthen this principle: the motion should go along *Pontryagin extremals* of the corresponding functional of action. At least, to his knowledge, there were no counterexamples.

4.   Classes of variations in optimal control problems

Most part of necessary optimality conditions in extremum problems are obtained by varying the given point, and the result essentially depends on the chosen class of variations. In optimal control, the following classes of variations have been known and used up till now:

a) uniformly small variations: $||u - u^0||_\infty \to 0$, appeared in CCV; in the abstract problem they correspond to the local minimality w.r.t. the norm of the given Banach space;

b) needle-type variations (or pack of needles), introduced by Weierstrass, and used also by McShane, Graves, Boltyansky, and many others;

c) sliding mode variations;

d) $v-$change of time.

The two first classes are widely known, so we turn to the two last classes.

c) Sliding mode variations consist in passing from the initial control system $\dot{x} = f(x, u, t)$ to its extension (relaxation) of the form

$$\dot{x} = \sum_{i=0}^{N} \alpha_i(t)\, f(x, u^i, t), \qquad \alpha_i(t) \geq 0, \qquad \sum_{i=0}^{N} \alpha_i(t) = 1, \tag{10}$$

where $u^0(t)$ is the optimal control. System (10) was introduced by Gamkrelidze
(1962) for proving the existence of optimal trajectory. As a tool for proving
MP it was first used by Dubovitskii and Milyutin (1965) (for a problem of the
so-called Lyapunov type, as brought to the authors' attention by V.I. Arkin).
The idea here is briefly as follows.

Let $(x^0(t), u^0(t))$ be an optimal process. Fix arbitrary functions $u^1(t), \ldots,$
$u^N(t) \in U$ and consider system (10). For $\alpha^0(t) = (1, 0, \ldots, 0)$ its solution is the
optimal $x^0(t)$. Now, let the vector-function $\alpha = (\alpha_0, \alpha_1 \ldots, \alpha_N)$ be uniformly
sufficiently close to α^0, and let $\alpha^{(n)} \overset{\text{weak}-*}{\longrightarrow} \alpha$ in such a way that each its
component $\alpha_i^{(n)}(t) = 0$ or 1. Then the corresponding solution $x^{(n)}(t)$ of system
(10) is also a solution of the initial system for the combined ("mixed") control
$u^{(n)}(t) = \sum \alpha_i^{(n)}(t) u^i(t)$, and the latter can be "far" from $u^0(t)$ only on a set of
a small measure, which tends to zero if $||\alpha - \alpha^0||_\infty \to 0$ (hence $||x - x^0||_C \to 0$
too). We thus obtain a kind of needle variations in the initial system. Note that
the extended system (10) is smooth (even linear) w.r.t. controls α_i, so it is very
convenient for study.

The main obstacle in this approach is that the approximating trajectory
$x^{(n)}(t)$ may not satisfy the terminal constraints, and so, one should find con-
ditions under which it is possible to satisfy them. This is rather a nontrivial
question. However, it can be settled, e.g. in the case when all individual con-
trols $u^i(t) \in int\, U$, $i = 0, 1, \ldots, N$, they are also variable together with α_i,
and the linear approximation of system (10) is controllable (see the details in
Milyutin, Dmitruk, Osmolovskii, 2004, and Dmitruk, 2007).

d) $v-$**change of time.** Instead of the original time $t \in [t_0, t_1]$ introduce a
new time $\tau \in [\tau_0, \tau_1]$ (usually, on a fixed interval), and consider the old time as
an additional state variable $t(\tau)$ subject to

$$\frac{dt}{d\tau} = v(\tau), \tag{11}$$

where $v(\tau) \geq 0$ is a new control, and so

$$\frac{dx}{d\tau} = v(\tau) f(t, x, u), \qquad u \in U. \tag{12}$$

Thus, now the state variables are $x(\tau)$, $t(\tau)$, and the controls are $u(\tau)$, $v(\tau)$.

Let $(x^0(t), u^0(t))$ be an optimal process in the initial problem. Obviously,
any (measurable bounded) function $v^0(\tau) \geq 0$ generates a function $t^0(\tau)$ and
the corresponding process in the new problem

$$\tilde{x}^0(\tau) = x^0(t^0(\tau)), \qquad \tilde{u}^0(\tau) = u^0(t^0(\tau)).$$

On the other hand, any process in the τ-problem generates a unique process
in the $t-$problem; this is due to the specific form of the new control system:
both equations for t and x contain the factor $v(\tau) \geq 0$, so when $t(\tau)$ does not
increase, $x(\tau)$ stays constant.

   What is the advantage of this passage to the new problem? Take as $v^0(\tau)$ any piecewise constant nonnegative function. Define the set $M_0 = \{\, \tau \mid v^0(\tau) = 0 \,\}$. It is a union of a finite number of intervals $\Delta_1, \ldots, \Delta_m$. Outside M_0 we have a one-to-one correspondence between t and τ, whereas each interval $\Delta_k \subset M_0$ is mapped into a single point $t^0(\Delta_k) = t_k$. On each interval Δ_k let us set $\tilde{u}^0(\tau) = u_k$, where $u_k \in U$ are arbitrary fixed values. The choice of these values does not impact the trajectory $\tilde{x}^0(\tau)$ and hence $x^0(t)$. But this is so only for $v^0(\tau)$!

   Now, let us slightly vary it, i.e. take a piecewise constant $v(\tau) \geq 0$ close to $v^0(\tau)$, having the same intervals of constancy. Then, according to (11), we get a new function $t(\tau)$, for which $t(\Delta_k)$ is not a single point but a small interval, on which $u(t) = u_k$.

   Thus, *small variations of $v^0(\tau)$* in the $\tau-$problem generate *needle variations of $u^0(t)$* in the $t-$problem! Note that, since the control $\tilde{u}^0(\tau)$ in the $\tau-$problem is not varied, only the control $v(\tau)$ remains variable, which smoothly (even linearly) comes into the control system and the constraint $(v \geq 0)$. For a fixed $u = \tilde{u}^0(\tau)$, the system

$$\frac{dx}{d\tau} = v(\tau)\, f(t, x, \tilde{u}^0), \qquad \frac{dt}{d\tau} = v(\tau),$$

is smooth in t, x, v *regardless of the sign of $v(\tau)$*, and therefore, it can be varied by the standard technique of the theory of ODEs. The sign of v is taken into account only by an independent constraint $v \geq 0$, which is also given by a smooth function.

   So, the $v-$change of time makes it possible to obtain needle variations of the control in the original problem by passing to a *smooth control system,* thus avoiding the "damnation" of nonnegativity of the needle width, that inevitably appears in the usual construction of the needles and essentially harms the application of standard technique of ODEs and the calculus (see e.g. a recent book by Arutyunov, Magaril-Ilyaev and Tikhomirov, 2006). For details, see Dubovitskii and Milyutin (1965), Girsanov (1970), Milyutin, Dmitruk, Osmolovskii (2004).

   An interesting fact is that the class of $v-$variations turned out to be richer than the class of sliding variations. If a process is stationary in all associated problems obtained by $v-$variations, then it is also stationary in all associated problems obtained by sliding mode variations. On the other hand, Dubovitskii and Milyutin (1981) provide an example where a process is stationary with respect to all sliding variations, but is not stationary with respect to some $v-$variations.

5.   Scheme of obtaining the Maximum Principle

For a given optimal process, one should construct a family of associated smooth problems parametrized by an index θ, in each of which the corresponding process is a local minimum. In each of these associated problems one should write out

the "standard" first order necessary condition for a local minimum — the Euler–Lagrange equation (stationarity condition), the family of which should then be "pressed" into one mutual condition. How to do this?

For each index θ, the stationarity condition gives a *nonempty* set $M^\theta = \{ \lambda^\theta \}$ of normalized collections of Lagrange multipliers. It turned out to be a *compact set* w.r.t. an appropriate topology.

The set of indices should have an ordering making it a net, i.e. for any two indices there should exist a third one greater than both of these. Moreover, if $\theta' \prec \theta''$, then $M^{\theta'} \supset M^{\theta''}$, which implies that the family of compacta M^θ is a centered (i.e. Alexandrov type) system: any finite number of these compacta do intersect. Therefore, its total intersection $\bigcap_\theta M^\theta$ is nonempty. Any element λ (a collection of Lagrange multipliers) of this intersection gives the desired "pressed" optimality condition. By definition, this condition is a Maximum Principle, see Milyutin (1970).

For example, the proof of MP for the canonical problem of Pontryagin type can be carried out by using the above piecewise constant $v-$changes of time, which lead to a family of smooth finite-dimensional associated problems, in each of which one should use just the standard Lagrange multipliers rule(!). As the index θ, one can take here the collection of values (t_k, u_k) (i.e., the parameters of the corresponding pack of needle variations) with its natural ordering by inclusion (one index follows another if the first collection of values includes the second one). For the sliding mode variations, the index θ is the set of all corresponding base controls $\{u^1(t), \ldots, u^N(t)\}$, again with its natural ordering by inclusion; see Dubovitskii and Milyutin (1981), Milyutin (2001), Milyutin, Dmitruk, Osmolovskii (2004).

6. Pontryagin minimum

To what type of minimum the MP corresponds? Usually, MP is stated as a necessary condition for a strong minimum, which is a minimum on the set of admissible processes (x, u) satisfying the additional restriction $||x - x^0||_C < \varepsilon$ for some $\varepsilon > 0$. However, the analysis of the proof of MP allows to weaken the notion of minimum (and so, to strengthen the assertion of MP). Dubovitskii and Milyutin proposed the following notion. (For simplicity, here we give it for problems on a fixed time interval $\Delta = [t_0, t_1]$.)

DEFINITION 1 *A process* $(x^0(t), u^0(t))$ *provides the Pontryagin minimum if for any number* N, *there exists an* $\varepsilon > 0$ *such that, for any admissible process* $(x(t), u(t))$ *satisfying the restrictions*

$$|x(t) - \hat{x}(t)| < \varepsilon, \quad |u(t) - \hat{u}(t)| \leq N \quad on \ \Delta, \quad and \quad \int_\Delta |u(t) - \hat{u}(t)| \, dt < \varepsilon,$$

one has $J(x, u) \geq J(x^0, u^0)$.

In other words, for any N the process (x^0, u^0) provides a local minimum w.r.t. the norm $\|x\|_C + \|u\|_1$ in the problem with the additional restriction $\|u - \hat{u}\|_\infty \leq N$.

This type of minimum includes both the uniformly small and needle variations of the control, and so, it occupies an intermediate position between the classical weak and strong minima. The relation between different types of minima is as follows:

$$\text{global} \implies \text{strong} \implies \text{Pontryagin} \implies \text{weak.}$$

One can also introduce the corresponding convergence: a sequence $u^k(t)$ *converges in the Pontryagin sense* to $u^0(t)$, if

$$\|u^k - u^0\|_1 \to 0 \quad \text{and} \quad \|u^k - u^0\|_\infty \leq \mathcal{O}(1).$$

An interesting fact is that this convergence does not correspond to any topology in $L_\infty(\Delta)$ of the Frechet–Uryson type (when the closure of a set can be obtained by sequences): if one constructs the closure operator corresponding to the Pontryagin convergence, then it would not satisfy the axioms of Kuratowski: the double closure would not always coincide with the simple closure. (A good exercise for students — to find an example of such a set.)

In spite of this "negative" fact, the Pontryagin minimum has that advantage against the weak minimum, that it is invariant with respect to a broad class of transformations (reformulations) of the problem, whereas the weak minimum is not (Milyutin, 1990c). This corresponds to the invariance of MP. Note that the Euler–Lagrange equation (necessary condition for the weak minimum) does not enjoy such a broad invariance.

Moreover, Dubovitskii and Milyutin discovered that MP is a *necessary and sufficient* condition (i.e., criterion!) for the *stationarity* of the given process in the class of all variations of the Pontryagin type.

On the other hand, unlike the strong minimum, which does not require restrictions on the control and hence does not allow to estimate the increment of the cost (or the Lagrange function) by using its expansion at the given process, the Pontryagin minimum requires some, though mild, control restrictions, which make it possible to use such expansions.

The practice of research performed up till now shows that the concept of Pontryagin minimum is a convenient and effective working tool. It admits a rich theory not only of the first, but also of higher order optimality conditions; see e.g. Osmolovskii (1988, 1993, 1995), Milyutin and Osmolovskii (1998), V.A. Dubovitskii (1982), Dmitruk (1987, 1999), Dykhta (1994) and references therein.

7. Problems with state constraints

As was shown by Dubovitskii and Milyutin (1965), for problems with state constraints $\Phi(t, x(t)) \leq 0$ only adjoint equations in MP are modified; they are now

$$
\begin{aligned}
\dot{\psi}_x &= -H_x + \dot{\mu}\,\Phi_x\,, \\
\dot{\psi}_t &= -H_t + \dot{\mu}\,\Phi_t\,,
\end{aligned}
$$

where $\dot{\mu}$ is the generalized derivative of a nondecreasing function $\mu(t)$. These equations can be also written as equalities between measures:

$$
\begin{aligned}
d\psi_x &= -H_x\, dt + \Phi_x\, d\mu\,, \\
d\psi_t &= -H_t\, dt + \Phi_t\, d\mu\,,
\end{aligned}
$$

or, in the integral form,

$$
\begin{aligned}
\psi_x(t) - \psi_x(t_0) &= -\int_{t_0}^{t} H_x\, d\tau \; + \int_{t_0}^{t} \Phi_x\, d\mu(\tau), \\
\psi_t(t) - \psi_t(t_0) &= -\int_{t_0}^{t} H_t\, d\tau \; + \int_{t_0}^{t} \Phi_t\, d\mu(\tau),
\end{aligned}
$$

where $d\mu \in C^*[t_0,\, t_1]$, and the last integrals in both formulas are taken in the Riemann–Stiltjes sense.

The corresponding measure satisfies the complementary slackness

$$
\Phi\left(t, x^0(t)\right) d\mu(t) \equiv 0\,,
$$

which means that on any interior interval, where $\Phi\left(t, x^0(t)\right) < 0$, the measure vanishes: $d\mu(t) = 0$.

When it appeared, the MP with these new adjoint equations seemed rather unusual, especially for engineers, but actually the formulation of this result is quite natural: since the function $\Phi(t, x(t)) \in C[t_0,\, t_1]$, then its Lagrange multiplier should be nothing else but a measure $d\mu \in C^*[t_0,\, t_1]$. However, this "obvious" observation appeared only after a deep mathematical understanding of the essence of the problem, and, of course, the proof of MP (based on $v-$change of time) was not obvious at all. (Later, another proof, based, in fact, on sliding variations, was given in Ioffe and Tikhomirov, 1974).

State constraints often appear in applied optimal control problems. However, solution of such problems is rather difficult, because of a nonstandard form of the adjoint equation, in which almost nothing is known about the measure. Here we give a simple example, where state constraints are added to a well known classical problem.

Example: isoperimetric problem

For $x \in \mathbf{R}^2$, $u \in \mathbf{R}^2$, and a fixed time interval $[0, T]$ consider the problem:

$$\dot{x} = u, \qquad x(0) = x(T), \qquad |u| \leq 1,$$

$$J = \int_0^T \left(-x_1 u_2 + x_2 u_1\right) dt = \int_0^T (Px, u)\, dt \to \max,$$

where P is the matrix of rotation in $90°$. This is the classical isoperimetric Dido's problem in the optimal control setting. Here T is the upper bound of the length of the rope, J is the oriented square within the rope times 2.

Here $H = \psi u + \frac{1}{2}(Px, u)$, $\dot{\psi} = \frac{1}{2} Pu$, hence $\psi = \frac{1}{2} Px - Pa$ with some $a \in \mathbf{R}^2$, and so $H = (P(x - a), u)$, and since it should attain its maximum over the unit ball $|u| \leq 1$, we get

$$u = \frac{P(x - a)}{|P(x - a)|} = P \frac{x - a}{|x - a|}.$$

Moreover, since the problem is time-independent, $\max H = |P(x - a)| = |x - a| = \text{const} = r \geq 0$, hence, excluding the trivial case $r = 0$, we have a uniform motion along a circumference of radius r centered at an arbitrary point $a \in \mathbf{R}^2$ with velocity $u = P(x - a)/r$. Since the problem is invariant w.r.t. translations, we can set $a = 0$. Condition $x(0) = x(T)$ yield $2\pi r\, k = T$, where k is the number of rotations, so $r = T/(2\pi k)$, and we obtain a countable set of extremals. By a simple calculation we find that the optimal extremal has just one rotation: $k = 1$.

(By the way, compare the above with the statement and solution of this problem in CCV — square root, singularities at the boundary, etc. A verification question: where is the square root now?)

All the above is well known (see, e.g., Ioffe and Tikhomirov, 1974). Now, consider the same problem in the presence of state constraints, say, as was proposed by Dubovitskii and Milyutin, in the form of a triangle. If the length of the rope is within appropriate bounds, the rope, obviously, should partially lie on the sides of the triangle (the boundary subarcs), whereas the subarcs lying inside the triangle should obviously be pieces of circumferences. What is not obvious is that these circumferences must have *the same radius*! This immediately follows from the MP, because on any interior subarc we still have the same adjoint equation $\dot{\psi} = \frac{1}{2} Pu$, hence, as before, $H = (P(x - a), u)$, but now the point a depends on the given subarc. Since the condition $H = \text{const} = r$ is still valid on the whole interval $[0, T]$, any interior subarc is a piece of circumference of the same radius r around its own center a.

Moreover, the measure has no jumps (atoms) at the junction points. To show this, let the triangle be given by inequalities $(a_i, x) \leq b_i$, $i = 1, 2, 3$ with all $a_i \neq 0$ and, say, $b_i > 0$. Here the adjoint equation is

$$\dot\psi \;=\; \frac{1}{2}\,Pu \;+\; \sum \dot\mu_i\,a_i\,.$$

Suppose that $(a_1\,,x(t)) \,=\, b_1$ on an interval $I_1 \,=\, [t_1\,,s_1]$ (boundary subarc) and $(a_1\,,x(t)) \,<\, b_1$ in a right neighborhood $\mathcal{O}_+(s_1)$ (interior subarc). Then $u(t) \,=\,$ const $\,=\, u_1 \perp a_1$ on I_1, and since $H \,=\, (\psi + \frac{1}{2}\,Px)\,u \,\equiv\, r$, we get $\psi + \frac{1}{2}\,Px = r u_1$ on I_1. Differentiating this, we obtain $Pu_1 + \dot\mu_1\,a_1 = 0$, whence $\dot\mu_1$ on I_1 is uniquely determined. If the jump $\Delta\mu_1(s_1) > 0$, then the jump of $\psi + \frac{1}{2}\,Px$ is $\Delta\psi(s_1) = \Delta\mu_1(s_1)\,a_1$, which implies that in $\mathcal{O}_+(s_1)$ the maximum of $H(x,u)$ over $|u| \le 1$ is attained at some $u(t)$ such that $(a_1,u(t)) > 0$, whence $(a_1,x(t)) > b_1$, a violation of the state constraint. (Note that here the sign of $\Delta\psi$ is important!)

Thus, the boundary and interior subarcs are joined *tangentially*, and hence, the total trajectory is determined uniquely (modulo number of rotations).

Another generalization is when the control constraint $|u| \le 1$ is replaced by $u \in U$, where U is a convex compact set containing the origin in its interior. Introduce its support function $\varphi(z) = \max\,(z, U)$. Since $H = (P(x-a), u)$ attains its maximum over $u \in U$, we immediately have $\varphi(P(x-a)) = $ const $= r$, whence $P(x-a) \in r \cdot \partial U^0$, where U^0 is the polar of U, and therefore, $(x-a) \in -r \cdot P(\partial U^0)$, $\quad u \in \partial\varphi(P(x-a))$. As before, the motion is determined uniquely (modulo number of rotations).

Again, one can add state constraints to this problem. Then, like before, the interior subarcs are pieces of the boundary of U^0 with different "centers" a, but with the same "radius" (homothety coefficient) r, and the measure still has no jumps.

Measure in the adjoint equations

As is well known, any nondecreasing function $\mu(t)$ is the sum of three components:

$$\mu = \mu_a + \mu_s + \mu_d\,,$$

which are absolute continuous, singular, and a jump functions, respectively. When analyzing the MP, it is very desirable to have an a priori information about the two last components, since they are in a sense less convenient than the first one. In some cases these components are simply absent. Milyutin (1990b) revealed, in particular, the following cases.

Condition for the absence of jump component. Let at some point (t,x), the convex hull of the velocity set $f(t,x,U)$ have a smooth boundary, and the time derivative of the state constraint

$$\frac{d\Phi}{dt} \;=\; \Phi_x(t,x)\,f(t,x,u) + \Phi_t(t,x)$$

can be both negative and positive for different $u \in U$. Then, the measure μ cannot have a jump at this point.

Condition for the absence of singular component. Let the control system have the form

$$\dot{x} = u, \qquad \dot{y} = Y(t, x, u), \qquad \Phi(t, x) \le 0$$

(which corresponds to CCV with a state constraint). Let along a given extremal the control $u^0(t)$ be piecewise continuous, the strengthened Legendre condition hold, and $(\Phi_x, \Phi_t) \ne (0, 0)$. Then the measure μ cannot have a singular component.

(Other results concerning properties of the measure are obtained, e.g., in Maurer, 1979, and Malanowski, 2003).

REMARK 5 The above formulated MP is a nontrivial optimality condition only under the assumption that the endpoints of the optimal trajectory do not lie on the boundary of state constraint. Otherwise, the measure may have atoms at the endpoints, while all other multipliers vanish, which is a trivial result. To guarantee the nontriviality of MP in this special case, one should impose some regularity conditions on the joint behavior of the endpoint and state constraints with the control system; see Dubovitskii and Dubovitskii (1987, 1988, 1995), Matveev (1987), Arutyunov and Aseev (1997).

Junction of different regimes

An important issue in determining optimal processes is the question: how can consecutive subarcs presenting regimes of different kind be joined? The first example of nontrivial junction was proposed by Fuller (1961):

$$\int_0^T x^2 \, dt \to \min, \qquad \ddot{x} = u, \qquad |u| \le 1,$$

$(x(0), \dot{x}(0)) \ne (0, 0)$ are given, and T is sufficiently large. Here the optimal trajectory should likely be zero after some $T_0 > 0$, which is a singular regime. The first part of trajectory, on $[0, T_0]$, is a nonsingular (bang-bang) regime, and it turns out that the junction between these regimes can occur only through a *countable number of switchings*, except for the case when the initial conditions lie on a specific line in $\mathbf{R}^2$. Such a phenomenon is called *chattering regime.* (By the way, this example justifies the choice of control space $u \in L_\infty$ in optimal control problems, instead of piecewise controls.) A further study of this phenomenon is given in Zelikin and Borisov (1994).

Note, however, that here the control u is scalar. What is the analog of this chattering phenomenon for a multidimensional control, when $u \in U$ with a strictly convex compact set $U \subset \mathbf{R}^r$? In this case, the optimal control generically has no switchings, but can have points of *discontinuity of the second kind.* Milyutin proposed to take this as the definition of multidimensional chattering. (In the one-dimensional case, discontinuity of the second kind reduces to an infinite number of switchings.) Milyutin (1993), Milyutin and Chukanov (1993),

Chukanov and Milyutin (1994), and Chukanov (1993a, 1993b) investigated this phenomenon more thoroughly and described the typical behavior of extremals. See also the related papers by Chukanov (1977) and Milyutin (1990e, 1994) concerning one-dimensional case.

For problems with state constraints the typical regimes are boundary and interior ones, and the junctions between them are not always trivial. In his doctoral dissertation, Milyutin (1966) proposed the following

Example

$$\int_0^T y \, dt \to \min, \qquad \frac{d^3y}{dt^3} = u, \qquad y \geq 0, \qquad \int_0^T u^2 \, dt \leq 1,$$

$(y, \dot{y}, \ddot{y})$ at 0 and T are given. Here, except the case of special endpoint conditions, the junction of interior and boundary regimes occurs through a countable number of *jumps of the measure* corresponding to the state constraint $y \geq 0$, while the control is continuous. (Later, this example was independently given also by Robbins, 1980).

In Dikusar and Milyutin (1989, Ch.3), and Milyutin (2000), Milyutin studied this phenomenon more thoroughly. He proposed the notion of *depth* (sometimes called *order*) of the state constraint, which is the number of its differentiations until the control appears explicitly, and, analyzing linear control systems with linear state constraints, discovered that, if the depth is 1, the measure usually (but not always) has no jumps; if the depth is 2, the measure can have only a finite number of jumps, and if the depth is 3 and greater, the measure typically has a countable number of jumps. An interesting fact is that the jumps can occur not only at the entry and exit points, but also at intermediate points of boundary intervals; a corresponding example is proposed in Milyutin (1990b).

8.   Nonsmooth functionals

Since the general Dubovitskii–Milyutin scheme admits only a finite number of inequality constraints, the constraint $\Phi(t, x(t)) \leq 0$, $t \in \Delta = [t_0, t_1]$, is represented by the nonsmooth functional $\max_{t \in \Delta} \Phi(t, x(t)) \leq 0$, and the constraint $\varphi(t, x(t), u(t)) \leq 0$ by the functional $\operatorname*{vraimax}_{t \in \Delta} \varphi(t, x(t), u(t)) \leq 0$. Both these functionals are Lipschitz continuous and have sublinear directional derivatives. Recall that, for any $v \in L_\infty(\Delta)$,

$$\operatorname*{vraimax}_{t \in \Delta} v(t) \; = \; \min \{ b : \; v(t) \leq b \quad \text{a.e. on} \quad \Delta \}.$$

The following results were obtained in Dubovitskii and Milyutin (1965).

a) Consider the functional $F : C(\Delta) \to \mathbf{R}$,

$$F(x) = \max_{t \in \Delta} \Phi(t, x(t)).$$

Denote $F(x^0) = a$ and introduce the set (nonempty and closed)

$$M_0 = \{\, t \in \Delta \mid \Phi(t, x^0(t)) = a \,\}.$$

THEOREM 4   *The derivative of F along any direction $\bar{x}(t)$ is*

$$F'(x^0, \bar{x}) = \max_{t \in M_0} \left(\Phi'_x(t, x^0(t))\, \bar{x}(t) \right);$$

it is a sublinear functional in $\bar{x}$, which subdifferential consists of all linear functionals $l \in C^*(\Delta)$ of the form

$$l(\bar{x}) = \int\limits_{\Delta} \Phi'_x(t, x^0(t))\, \bar{x}(t)\, d\mu(t),$$

where $d\mu$ is a normalized nonnegative Radon measure on M_0.

b) Consider the functional $P : L_\infty(\Delta) \to \mathbf{R}$,

$$P(v) = \operatorname*{vraimax}_{t \in \Delta} v(t).$$

Denote $P(v^0) = a$, and for any $\delta > 0$ introduce the set (obviously, of positive measure)

$$M_\delta = \{\, t \in \Delta \mid v^0(t) \geq a - \delta \}.$$

THEOREM 5   *For any $\bar{v}(t) \in L_\infty(\Delta)$, the directional derivative of P is*

$$P'(v^0, \bar{v}) = \lim_{\delta \to +0} \operatorname*{vraimax}_{t \in M_\delta} \bar{v}(t),$$

it is a sublinear functional in $\bar{v}$, which subdifferential consists of all functionals $\lambda \in L_\infty^*(\Delta)$ satisfying the following three conditions:
  (i) λ is supported on M_δ for any $\delta > 0$,
  (ii) $\lambda \geq 0$,
  (iii) $\lambda(\mathbf{1}) = 1$, where $\mathbf{1}(t) \equiv 1$ on Δ.

Condition (i) does not mean that λ is supported on the intersection of all $M_\delta(!)$. The set $M_0 = \bigcap\limits_{\delta > 0} M_\delta$ can have zero measure, and then a functional $\lambda \in L_\infty^*$ supported on M_0 is identically zero, whereas λ satisfying (i) can be nonzero (a purely singular functional).

Recall here the classical

IOSIDA–HEWITT THEOREM.   Any functional $l \in L_\infty^*$ is the sum $l = l_a + l_s$ of an absolute continuous functional $l_a \in L_1$ and a singular functional l_s supported on each set of some sequence $\{E_n\}$ with $mes\, E_n \to 0$.

c) Consider now the functional $G:\ C(\Delta) \times L_\infty(\Delta) \to \mathbf{R}$,

$$G(x,u) = \operatorname*{vraimax}_{t\in\Delta} \varphi(t,x(t),u(t)).$$

As before, denote $G(x^0,u^0) = a$, and introduce the sets

$$M_\delta = \{t\in\Delta \mid \varphi(t,x^0(t),u^0(t)) \geq a-\delta\}.$$

Theorem 5 readily yields the following result. (To shorten the formulas, we use the notation $w = (x,u)$ and similar to it.) For any $\bar{w} = (\bar{x},\bar{u})$

$$G'(w^0,\bar{w}) = \lim_{\delta\to+0} \operatorname*{vraimax}_{t\in M_\delta} \langle \varphi'_w(t,w^0(t)), \bar{w}(t)\rangle,$$

and its subdifferential consists of all linear functionals l of the form

$$l(\bar{w}) = \lambda\,(\varphi'_w(t,w^0(t))\,\bar{w}),$$

where the functional $\lambda \in L_\infty^*(\Delta)$ satisfies the above properties $(i)-(iii)$.

9. Problems with mixed constraints

Considering the mixed constraints

$$\varphi_i(t,x,u) \leq 0,\ \ i=1,\ldots,d(\varphi), \qquad g_j(t,x,u)=0,\ \ j=1,\ldots,d(g), \quad (13)$$

one should distinguish two essentially different cases of regular and nonregular constraints.

According to Dubovitskii and Milyutin, constraints (13) are *regular* if for any point $(t,x,u)\in\mathcal{Q}$ satisfying these constraints, the gradients in u

$$\varphi'_{iu}(t,x,u),\ \ i\in I(t,x,u), \qquad g'_{ju}(t,x,u),\ \ j=1,\ldots,d(g),$$

where $I(t,x,u) = \{i\mid \varphi_i(t,x,u)=0\}$ is the set of active indices for inequality $\varphi\leq 0$ at the given point, are *positive-linear independent*, i.e., there are no coefficients

$$a_i \geq 0,\ \ i\in I(t,x,u), \quad \text{and} \quad b_j,\ \ j=1,\ldots,d(g),$$

which are not all zero, such that

$$\sum_{i\in I} a_i\,\varphi'_{iu}(t,x,u) \ + \ \sum_{j=1}^{d(g)} b_j\,g'_{ju}(t,x,u) \ = 0.$$

(Here the word "positive" relates to φ'_{iu}, while the word "linear" to g'_{ju}.)

An equivalent requirement: the gradients $g'_{ju}(t,x,u)$ are linear independent, and $\exists\,\bar{u}\in\mathbf{R}^r$ such that

$$\varphi'_{iu}(t,x,u)\,\bar{u} < 0, \quad \forall\, i\in I, \quad \text{and} \quad g'_{ju}(t,x,u)\,\bar{u} = 0, \quad \forall\, j$$

(the so-called Mangasarian–Fromovitz condition).

Since the functions $\varphi_i(t, x(t), u(t))$ and $g_j(t, x(t), u(t))$ belong to $L_\infty(\Delta)$, then, a priori, the corresponding Lagrange multipliers k_i, m_j $\in L_\infty^*(\Delta)$. However, for the regular mixed constraints they all are elements of $L_1(\Delta)$, and this is a crucial fact that essentially simplifies the formulation and proof of MP for problems with regular mixed constraints as compared with that in the general case.

THEOREM 6 *(on the absence of singular components). Let $r-$vector functions $A_i(t)$, $B_j(t)$ with L_∞ components be positive-linear independent uniformly on Δ, and let functionals k_i, m_j $\in L_\infty^*(\Delta)$, where all $k_i \geq 0$, be such that*

$$\sum k_i A_i(t) + \sum m_j B_j(t) = \lambda(t) \in L_1^r(\Delta)$$

(i.e. the left hand linear functional on $L_\infty^r(\Delta)$ belongs to $L_1^r(\Delta)$). Then all k_i, m_j actually belong to $L_1(\Delta)$, i.e., they have no singular components.

This fact was discovered by Dubovitskii and Milyutin in the end of the 1960s, but was not explicitly published. The proof can be found in Dmitruk (1990), and Milyutin, Dmitruk, Osmolovskii (2004).

Theorem 6 should be applied to $A_i(t) = \varphi'_{iu}$ and $B_j(t) = g'_{ju}$ calculated along the process $(x^0(t), u^0(t))$.

Canonical problem C with smooth regular mixed constraints

$$J = F_0(p) \longrightarrow \min,$$
$$F(p) \leq 0, \qquad K(p) = 0,$$
$$\dot{x} = f(t, x, u),$$
$$g(t, x, u) = 0, \qquad \varphi(t, x, u) \leq 0,$$
$$\Phi(t, x) \leq 0;$$
$$p \in \mathcal{P}, \quad (t, x, u) \in \mathcal{Q} \qquad \text{(open sets)}.$$

As before, here $p = (t_0, x(t_0), t_1, x(t_1))$, the interval $\Delta = [t_0, t_1]$ is a priori nonfixed. We assume that f, g, φ and Φ are smooth in t, x, u. Note that here the inclusion constraint $u \in U$ is not allowed, and the state constraint $\Phi(t, x) \leq 0$ cannot be considered as a special case of the mixed constraint $\varphi(t, x, u) \leq 0$, because of regularity assumption.

In order to formulate the corresponding MP, we need, as before, the endpoint Lagrange function $l(p) = \alpha_0 F_0(p) + \alpha F(p) + (\beta, K(p))$, the Pontryagin function $H = (\psi_x, f(t, x, u))$, and also the "extended" Pontryagin function

$$\overline{H}(t, x, u, v) = (\psi_x, f) - (k, \varphi) - (m, g) - (\dot{\mu}, \Phi),$$

where the multipliers k, m and μ are of dimensions $d(\varphi)$, $d(g)$ and $d(\Phi)$, respectively.

Maximum Principle for Problem C

Let a process $(x^0(t), u^0(t))$, $t \in \Delta^0 = [t^0_0, t^0_1]$ provide a Pontryagin minimum in Problem C. Then there exist a collection of numbers $\alpha = (\alpha_0, \ldots, \alpha_{d(F)}) \geq 0$, a vector $\beta \in \mathbf{R}^{d(K)}$, functions of bounded variation ψ_x, ψ_t, nondecreasing functions $\mu_s(t)$, $s = 1, \ldots, d(\Phi)$ (generating measures $d\mu_s(t)$), a vector-function $k(t) \geq 0$ from $L_\infty(\Delta)$ and a vector-function $m \in L_\infty(\Delta)$, such that the following conditions hold:

a) nontriviality

$$|\alpha| + |\beta| + \sum_s |\mu_s(t_1) - \mu_s(t_0)| + \int_{\Delta^0} |k(t)|\, dt > 0,$$

b) complementary slackness

$$\alpha_\nu\, F_\nu(p^0) = 0, \qquad \nu = 1, \ldots, d(F),$$
$$d\mu_s(t)\, \Phi_s(t, x^0(t)) \equiv 0, \qquad \text{on} \quad \Delta^0 \quad \forall\, s,$$
$$k_i(t)\, \varphi_i(t, x^0(t), u^0(t)) = 0 \qquad \text{a.e. on} \quad \Delta^0 \quad \forall\, i,$$

c) adjoint equations

$$-\dot{\psi}_x = \overline{H}_x = \psi_x\, f_x - k\, \varphi_x - m\, g_x - \dot{\mu}\, \Phi_x,$$
$$-\dot{\psi}_t = \overline{H}_t = \psi_x\, f_t - k\, \varphi_t - m\, g_t - \dot{\mu}\, \Phi_t$$

(here, again, $\dot{\mu}$ is the generalized derivative of the function $\mu(t)$),

d) transversality

$$\psi_x(t_0) = l_{x_0}(p^0), \qquad \psi_x(t_1) = -l_{x_1}(p^0),$$
$$\psi_t(t_0) = l_{t_0}(p^0), \qquad \psi_t(t_1) = -l_{t_1}(p^0),$$

e) stationarity in u:

$$\overline{H}_u = \psi_x\, f_u - k\, \varphi_u - m\, g_u = 0 \tag{14}$$

f) "energy evolution law"

$$H(t, x^0(t), u^0(t)) + \psi_t(t) = 0 \qquad \text{for a.a. } t \in \Delta^0,$$

g) maximality: for all $t \in \Delta^0$ and all $u \in C(t)$

$$H(t, x^0(t), u) + \psi_t(t) \leq 0,$$

where $C(t) = \{\, u \mid (t, x^0(t), u) \in \mathcal{Q},\ \varphi(t, x^0(t), u) \leq 0, \quad g(t, x^0(t), u) = 0\,\}$,

i.e., $C(t)$ is the set of control values admissible to comparison with $u^0(t)$ at time t for the optimal trajectory $x^0(t)$.

The two last conditions yield: for a.a. $t \in \Delta^0$

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

The proof can be made by using either $v-$variations (Milyutin, 1990d), or sliding mode variations (Dmitruk, 1990; Milyutin, Dmitruk, Osmolovskii, 2004).

REMARK 6 The presence of state and mixed constraints provides for broad possibilities to reformulate the problem, much broader than in CCV! For example, the so-called *minimax problem* with the nonsmooth cost

$$J = \max_{t \in \Delta} \Phi(t, x(t)) \to \min$$

can be easily reformulated as a smooth problem with a state constraint:

$$J = z(0) \to \min, \qquad \dot{z} = 0, \qquad \Phi(t, x(t)) \leq z.$$

Similarly, a problem with the cost

$$J = \operatorname*{vraimax}_{t \in \Delta} \varphi(t, x(t), u(t)) \to \min$$

can be easily reformulated as a problem with a mixed constraint:

$$J = z(0) \to \min, \qquad \dot{z} = 0, \qquad \varphi(t, x(t), u(t)) \leq z.$$

Another example: a problem with the nonsmooth cost

$$J = \int_0^T |L(t, x, u)| \, dt \to \min$$

can be reduced to a standard "smooth" problem:

$$J = \int_0^T v(t) \, dt \to \min, \qquad (v \text{ is a new control}),$$

$$\pm L(t, x(t), u(t)) \leq v(t) \qquad \text{(regular mixed constraints)}.$$

Reformulations of problems essentially extend the area of MP. This idea has not yet been properly exploited.

One more example is *the problem of S. Ulam on matching the segments.* Given two segments of the same length in a plane, it is required to move one of them onto the position of another in such a way that its endpoints describe a minimal total path.
Dubovitskii (1976) , V.A. Dubovitskii (1985), and Milyutin in Dikusar and Milyutin (1989, Ch.2) stated this problem as a time-optimal control problem for the system

$$\dot{x} = u, \qquad \dot{y} = v, \qquad (x - y)(u - v) = 0, \qquad |u| + |v| \leq 1,$$

where $x \in \mathbf{R}^2$ and $y \in \mathbf{R}^2$ are the endpoints of the moving segment, $x(0)$, $y(0)$ and $x(T)$, $y(T)$ are given. The mixed constraints here are regular. This problem can be considered as the problem about geodesics on the surface $|x-y| = 1$ in $\mathbf{R}^4$ with the Finsler metric $\rho(dx, dy) = |dx|+|dy|$. It was shown that any Pontryagin extremal in this problem is a combination of a finite number of elementary motions — translations, rotations, etc, and all the possible combinations were described.

General problem D with regular mixed constraints

Recall that problem C does not contain the inclusion constraint (because it would prevent to obtain the stationarity condition (14)). In order to allow it in the problem, Dubovitskii and Milyutin proposed to consider two control vectors:

$$u \in L_\infty(\Delta, \mathbf{R}^{r_u}) \quad \text{and} \quad v \in L_\infty(\Delta, \mathbf{R}^{r_v}).$$

(We use u and v instead of the authors' original notation u_1 and u_2.)

Let the time interval $\Delta = [t_0, t_1]$ be fixed. The problem is:

$$
\begin{aligned}
&J = F_0(p) \longrightarrow \min, \\
&F(p) \leq 0, \qquad K(p) = 0, \\
&\dot{x} = f(t, x, u, v), \\
&g(t, x, u, v) = 0, \qquad \varphi(t, x, u, v) \leq 0, \qquad\qquad (15) \\
&\Phi(t, x) \leq 0, \\
&v(t) \in V(t) \qquad \text{a.e.,}
\end{aligned}
$$

where $V(t)$ is a measurable set-valued mapping $\Delta \to \mathbf{R}^{r_v}$,

$$p = (x(t_0), x(t_1)) \in \mathcal{P}, \qquad (t, x, u, v) \in \mathcal{Q}.$$

All the data functions are smooth in x, u, and just continuous in t, v but v is subject to the inclusion constraint.

The assumption on regularity of mixed constraints is meant here with respect only to the "smooth" control u : at any point $(t, x, u, v) \in \mathcal{Q}$ satisfying (15), the gradients φ'_{iu}, $i \in I(t, x, u, v)$ and g'_{ju}, $j = 1, \ldots, d(g)$ are positive-linear independent.

Maximum Principle for Problem D

Let a process $(x^0(t), u^0(t), v^0(t))$ provide a Pontryagin minimum. Then there exist a collection of numbers $\alpha = (\alpha_0, \ldots, \alpha_{d(F)}) \geq 0$, a vector $\beta \in \mathbf{R}^{d(K)}$, an $n-$vector function of bounded variation $\psi_x(t)$, nondecreasing functions $\mu_s(t)$, $s = 1, \ldots, d(\Phi)$ (generating measures $d\mu_s(t)$), a vector-function $k(t) \geq 0$ from $L_\infty(\Delta)$ of dimension $d(\varphi)$, and a vector-function $m(t) \in L_\infty(\Delta)$ of dimension $d(g)$, such that the following conditions hold:

a) nontriviality,

b) complementary slackness,

c) adjoint equation

$$-\dot{\psi}_x = \overline{H}_x = \psi_x\, f_x - k\,\varphi_x - m\,g_x - \dot{\mu}\,\Phi_x\,,$$

d) transversality for ψ_x,

e) stationarity in u:

$$\overline{H}_u = \psi_x\, f_u - k\,\varphi_u - m\,g_u = 0,$$

g) maximality: for almost all $t \in \Delta$

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

where $C(t) = \{\, (u,v) \in \mathbf{R}^{r_u + r_v} \mid (t, x^0(t), u, v) \in \mathcal{Q},$

$$\varphi(t, x^0(t), u, v) \le 0, \quad g(t, x^0(t), u, v) = 0, \quad v \in V(t)\,\}.$$

If $V(t) \equiv V$ (constant), and all data functions are smooth in t, then the time interval can be variable, and again, the costate variable $\psi_t = -H$ satisfies the adjoint equation $-\dot{\psi}_t = \overline{H}_t$ and the transversality conditions.

REMARK 7 An interesting question is whether the equation $-\dot{\psi}_t = \overline{H}_t$ follows from the other conditions of MP. Milyutin (1990a) showed that if the gradients in u of the mixed constraints are *linearly independent*, then the answer is positive, and if those gradients are just positive-linearly independent (as in the general regular case), the answer is negative; he gave corresponding examples.

REMARK 8 Numerical methods for problems with state and mixed constraints based on the MP were proposed in Smoljakov (1968), Afanasjev (1990), Ilyutovich (1993), Dikusar (1990), Dikusar and Milyutin (1989), and in papers by J.F. Bonnans, A. Hermant, H. Maurer, H.J. Oberle, H.J. Pesch, and others.

10.   General problem G with nonregular mixed constraints

As was already said, the problems with nonregular mixed constraints are much more difficult than problems with regular ones. Here we give just a very brief look at these problems and the corresponding results obtained by Dubovitskii and Milyutin. More details can be found in Dubovitskii and Milyutin (1968, 1969, 1971, 1981), Dubovitskii (1975), and Milyutin (2001).

Consider the following problem on a fixed interval $\Delta = [t_0, t_1]$ with $p = (x(t_0), x(t_1))$:

$$J = F_0(p) \longrightarrow \min,$$
$$F(p) \le 0, \qquad K(p) = 0,$$
$$\dot{x} = f(t, x, u),$$
$$g(t, x, u) = 0, \qquad \varphi(t, x, u) \le 0,$$
$$p \in \mathcal{P}, \qquad (t, x, u) \in \mathcal{Q}.$$

The pure state constraint $\Phi(t, x) \le 0$ is included here as a particular case of the mixed inequality constraint. The matrix g'_u is assumed to have a full rank on the surface $g(t, x, u) = 0$.

In the study of this problem, the following notion proposed by Dubovitskii and Milyutin is useful.

CLOSURE WITH RESPECT TO MEASURE. For any set $E \subset \mathbf{R}$ define $clm\, M$ as the set of all points $t \in \mathbf{R}$ such that $\forall \varepsilon > 0$ the set $B_\varepsilon(t) \cap M$ has a positive measure. Obviously, this is a closed set contained in the usual closure of M. (We do not use the authors' notation because of typesetting problems.)

Obviously, this definition can be given in any finite-dimensional space.

The importance of this notion is justified e.g. by the fact that, if a functional $l \in L_\infty^*(\Delta)$ is supported on a measurable set $E \subset \Delta$, then its restriction $l|_C$ to the space $C(\Delta)$ is supported on $clm\, E$.

Similarly, for any set $F \subset \mathbf{R}^{1+r}$ considered as the graph of a set-valued mapping $\mathbf{R} \to \mathbf{R}^r$, define $clmg\, F$ as the set of all points (t, u) such that $\forall \varepsilon > 0$ the projection of $B_\varepsilon(t, u) \cap F$ on t has a positive measure.

Given a measurable function $u : \mathbf{R} \to \mathbf{R}^r$, let us denote by $clmg\, u$ the set-valued mapping, whose graph is the mes-closure of graph u, so that $(clmg\, u)(t) \subset \mathbf{R}^r$ is its value at t.

THEOREM 7 (Dubovitskii and Milyutin, 1981). *Let $F : \mathbf{R} \to \mathbf{R}^r$ be a set-valued mapping with a closed graph. Then there exists a measurable selection $f(t) \in F(t)$ such that $clmg\, f = clmg\, F$.*

DEFINITION 2 *A triple $(t, x, u) \in Q$ is called a phase point of the mixed constraints if $\exists\, a \in \mathbf{R}^{d(\varphi)}$, $a \ge 0$, and $b \in \mathbf{R}^{d(g)}$ such that $\sum a_i = 1$ and*

$$a\, \varphi'_u(t, x, u) + b\, g'_u(t, x, u) = 0, \qquad a\, \varphi(t, x, u) = 0,$$

i.e. the positive–linear independence fails to hold.

The term *phase point* is explained by the fact that near such a point the mixed constraints are much like a state (i.e. phase) constraint. The corresponding vector $s = a\, \varphi'_x + b\, g'_x$ is called *a phase jump*.

Note that the set of all phase points is determined only by the mixed constraints and does not depend on the control system nor the endpoint block of the problem.

Denote by $S(t, x, u) = \{\, s \,\}$ the set of all phase jumps at a given point. Obviously, it is a compact set, maybe empty. If the mixed constraints are regular, $S(t, x, u)$ is empty for any point. For any set $A \subset \mathbf{R}^r$ denote $S(t, x, U) = \bigcup_{u \in U} S(t, x, u)$.

Note that nonregular mixed constraints are not exotic; a simple example is $\varphi(x, u) = x^2 + u^2 - 1 \le 0$. Here $(x, u) = \pm\, (1, 0)$ are phase points with the phase jump $s = \pm 2$ respectively.

Conditions of weak stationarity in the nonregular problem G

Here, as before, we construct the endpoint Lagrange function $l(p) = \alpha_0 F_0(p) + \alpha F(p) + (\beta, K(p))$, the Pontryagin function $H = (\psi, f(t, x, u))$, and the "extended" Pontryagin function, which is now

$$\overline{H}(t, x, u, v) = (\psi, f) - (k, \varphi) - (m, g).$$

The process (x^0, u^0) is stationary in the class of all uniformly small variations iff the following conditions hold: $\exists\, \alpha = (\alpha_0, \ldots, \alpha_{d(F)}) \ge 0$, $\beta \in \mathbf{R}^{d(K)}$, integrable functions $k(t) \ge 0$, $m(t)$ of dimensions $d(\varphi)$, $d(g)$ respectively, a measure $d\nu \in C^*(\Delta)$, $d\nu \ge 0$, a "jump function" $s \in L_\infty(\Delta, d\nu)$ of dimension n (measurable and bounded with respect to $d\nu$), and a function of bounded variation $\psi(t)$, such that

a) α and $k(t)$ satisfy the corresponding complementary slackness,

b) ψ satisfies the transversality conditions,

c) $\quad |\alpha| + |\beta| + \displaystyle\int_\Delta |k(t)|\, dt \, + \int_\Delta d\nu > 0 \qquad$ (normalization),

d) $s(t) \in \operatorname{conv} S(t, x^0(t), (\operatorname{clmg} u^0)(t))$ a.e. on Δ with respect to $d\nu$,

e) $\qquad -\dot\psi_x = \overline{H}_x - s(t)\dfrac{d\nu}{dt} \qquad$ (costate equation),

or, equivalently,

$$-d\psi \;=\; \overline{H}_x\, dt - s(t)\, d\nu \quad \text{(an equality between measures)},$$

f) $\qquad \overline{H}_u \;=\; 0 \qquad$ (stationarity in u).

Let us give some remarks. Since the set-valued mapping $\operatorname{clmg} u^0$ has a compact graph, the set M_0 of all t, where $S(t, x^0(t), (\operatorname{clmg} u^0)(t)) \ne \emptyset$, is closed. If the measure $d\nu$ of its complement is positive, condition d) would fail to hold. Hence, $d\nu$ is supported on M_0, and actually, $s \in L_\infty(M_0, d\nu)$. Equation e) is

essential on M_0, while outside M_0 it holds in a truncated form with $d\nu = 0$:
$-\dot{\psi}_x = \overline{H}_x$.

These conditions of weak stationarity were obtained in Dubovitskii and Milyutin (1968, 1971) (see also Milyutin, 2001, Ch.3) and called the *local maximum principle*. (Note that the term is a bit confusing, because no maximality condition is here.) They can be also called Euler–Lagrange equation, since they play the role similar to that of EL equation in CCV.

The proof is essentially based on the following generalization of the Dubovitskii–Milyutin theorem on nonintersecting cones; see Dubovitskii and Milyutin (1971, 1981), Dubovitskii (1975), and Milyutin (2001).

Three storey theorem

Let Y, X, Z be Banach spaces, $Y^* = X$ and $X^* = Z$. (So, Y is a first storey, X a second storey, and Z a third storey).

In the space X let be given nonempty convex cones $\Omega_1, \ldots, \Omega_m, \Omega_{m+1}$, first m of which are open. Suppose $\forall i$ there is a convex cone $H_i \subset \Omega_i^* \cap Y$ such that

$$(x, y) \geq 0 \quad \forall y \in H_i \quad \implies \quad x \in \overline{\Omega}_i.$$

(Such a cone H_i is called *thick* on Ω_i, and in general it is not unique.)

Fix any points $x_i^0 \in \Omega_i, \quad i = 1, \ldots, m$ from the open cones.

THEOREM 8 $\Omega_1 \cap \ldots \cap \Omega_m \cap \Omega_{m+1} = \emptyset \iff \forall \varepsilon > 0 \quad \exists h_i \in H_i, \quad i = 1, \ldots, m+1$, *such that*

$$(x_1^0, h_1) + \ldots + (x_m^0, h_m) = 1 \qquad (normalization),$$

and

$$\| h_1 + \ldots + h_m + h_{m+1} \| \leq \varepsilon$$

(Euler–Lagrange equation with $\varepsilon-$accuracy).

In optimal control we have $Y = L_1, \quad X = L_\infty, \quad Z = L_\infty^*$, and this theorem allows for avoiding singular Lagrange multipliers from L_∞^* by taking approximate solutions to the EL equation with multipliers from L_1 and then by passing to a limit, in which one should use the so-called

BITING LEMMA.   Let $\varphi_n(t) \in L_1(\Delta)$ be a bounded sequence: $||\varphi_n||_1 \leq const$. Then there is a subsequence $n_k \to \infty$ and measurable sets $E_k \subset \Delta$ such that $mes\, E_k \to 0$ and the sequence of functions

$$\varphi_{n_k}(t) = \begin{cases} \varphi_{n_k}(t), & t \notin E_k, \\ 0, & t \in E_k, \end{cases}$$

is uniformly integrable, and so, it is a weakly precompact family in $L_1(\Delta)$.

This lemma (and its various versions) was proved by many authors, and it is not clear where this was done for the first time. I.V. Evstigneev gave me a number of references, the earliest ones being Kadec and Pelczynski (1961/62) and Gaposhkin (1972). In the first one the formulation was not explicitly given and the proof was hidden in the proof of another result. Dubovitskii and Milyutin (1971) gave a proof not being aware of that paper. The biting lemma is now very popular among the specialists in probability and stochastics. More about this lemma see in Saadoune and Valadier (1995).

Maximum principle for the nonregular problem G

DEFINITION 3 *A set-valued mapping* $t \mapsto Z(t)$ *with convex images and a compact graph is called associated with an admissible process* $(x^0(t), u^0(t))$, *if* $\forall\, t \in \Delta$

$$S(t, x^0(t), (clmg\, u^0)(t)) \subset [a, b]\, Z(t)$$

for some $0 < a \le b$ *(i.e.,* $Z(t)$ *majorates essentially the phase jumps of all points that are "stuck" to the graph of the reference process).*

Let a process $(x^0(t), u^0(t))$ provide a Pontryagin minimum. Then, for any associated mapping $Z(t)$ there exist a collection of Lagrange multipliers including a measure $d\nu \in C^*(\Delta)$, $d\nu \ge 0$, and a "jump function" $s(t) \in Z(t)$ a.e. in $d\nu$, such that the adjoint equation

$$\dot\psi_x = -H_x + k(t)\,\varphi_x + m(t)\,g_x + s(t)\,\frac{d\nu}{dt}$$

holds, and for any "test" control function $u(t)$ that satisfies the conditions

$$(t, x^0(t), u(t)) \in Q, \quad \varphi(t, x^0(t), u(t)) \le 0, \quad g(t, x^0(t), u(t)) = 0 \quad \text{a.e. on } \Delta,$$

and generates phase jumps majorated by $Z(t)$:

$$S(t, x^0(t), (clmg\, u)(t)) \subset [a', b']\, Z(t), \qquad \text{with some } 0 < a' \le b',$$

the maximality condition holds:

$$H(t, x^0(t), u(t)) \le H(t, x^0(t), u^0(t)) \quad \text{for almost all } t \in \Delta.$$

The other conditions of MP are the same as earlier. If the data functions are smooth in t, the state jumps should be defined in $\mathbf{R}^{n+1}$, and the costate function $\psi_t = -H$ should satisfy the corresponding adjoint equation.

In some "good" cases (but not always) the maximality condition can be written in a convenient pointwise form.

This is the formulation of MP corresponding to sliding mode variations. (The formulation of MP corresponding to v−variations is more complicated.)

Thus, the adjoint equation and maximality condition depend here on the choice of associated mapping $Z(t)$. Note that the larger $Z(t)$, the broader the class of compared $u(t)$ in maximality condition, but more uncertain $s(t)$ in the adjoint equation. So, we come to a quite unexpected fact that there is a family of "partial" MPs (each of which corresponds to the stationarity in a certain associated problem), without a "common" MP (that would guarantee the stationarity in all associated problems). This family can be partially ordered, so that there is an "hierarchy" of partial MPs. However, the above "pressing" procedure cannot be perfectly accomplished in the case of nonregular mixed constraints. These results obviously require further and more in-depth analysis.

REMARK 9 Like the regular problem C, problem G can be also generalized to involve the inclusion constraint. To this end, one should again consider two control vectors u and v, etc. The details can be found in Dubovitskii and Milyutin (1981), Dubovitskii (1975), and Milyutin (2001).

REMARK 10 Some examples of the usage of the "nonregular" MP are given in Milyutin and Dubovitskii (1981), Dubovitskii and Milyutin (1985) and Milyutin (2001, Ch.2). One more example (given as an answer to a question of F. Clarke) is the construction of a problem with a Hölder continuous differential inclusion whose extremals have a discontinuous costate function (Milyutin, 1999). It was made by constructing first an optimal control problem with a smooth nonregular mixed constraint, whose associated measure has a jump, and then by reformulating this constraint as a differential inclusion.

The nonregular MP has also practical applications. Dikusar and Shilov (1970), Klumov and Merkulov (1984), Dikusar in Dikusar and Milyutin (1989, Ch.3), and Dikusar (1990) used it for numerical solution and investigation of some problems of aerospace navigation.

REMARK 11 Chukanov (1990) showed that the Dubovitskii–Milyutin scheme of obtaining MP can be extended from control systems with ODEs defined on a time interval to a very general control system governed by integral equations of the second kind defined on an arbitrary metric compact set with a measure and subjected to state and nonregular mixed constraints. This class includes both ODE systems, systems with delays, with intermediate constraints, and some PDE systems (e.g. heating equation). Using the sliding mode variations, he obtained an MP for such problems.

Acknowledgments

References

AFANASJEV, A.P. (1990) Construction of extremal trajectories in problems with a free terminal point on the base of continuation procedure. *Necessary condition in optimal control.* Nauka, Moscow (in Russian), Ch. 7.

ARUTYUNOV, A.V. and ASEEV, S.M. (1997) Investigation of the degeneracy phenomenon of the maximum principle for optimal control problems with state constraints. *SIAM J. Control Optimization*, **35** (3), 930–952.

ARUTYUNOV, A.V., MAGARIL-ILYAEV, G.G. and TIKHOMIROV, V.M. (2006) *Pontryagin Maximum Principle: Proofs and Applications.* Factorial Press, Moscow.

BOLTYANSKY, V.G. (1969) *Mathematical Methods of Optimal Control.* Nauka, Moscow.

CHUKANOV, S.V. (1977) Examples of optimal trajectories in continuous-time input-output models of economical dynamics. *Autom. Remote Control*, **38** (1), 69-78.

CHUKANOV, S.V. (1990) Maximum principle for optimal control problems governed by integral equations. *Necessary condition in optimal control.* Nauka, Moscow (in Russian), Ch. 6.

CHUKANOV, S.V. (1993a) A problem with a strictly positive functional and a control of full dimension. *Optimal control in linear systems.* Nauka, Moscow (in Russian), Ch. 4.

CHUKANOV, S.V. (1993b) Construction of integral funnels at interior points of the singular manifold. *ibid.,* Ch. 7.

CHUKANOV, S.V. and MILYUTIN, A.A. (1994) Qualitative study of singularities for extremals of quadratic optimal control problem. *Russian Journal of Math. Physics,* **2** (1), 31–48.

DIKUSAR, V.V. (1990) Numerical determination of optimal trajectories in optimal control problems with mixed constraints. Optimization of the flight distance for an apparatus in the atmosphere. *Necessary condition in optimal control.* Nauka, Moscow (in Russian), Ch. 8–10.

DIKUSAR, V.V. and MILYUTIN, A.A. (1989) *Qualitative and numerical methods in Maximum principle.* Nauka, Moscow (in Russian).

DIKUSAR, V.V. and SHILOV, A.A. (1970) Nonregular optimal trajectories of the apparatus flying in the atmosphere. *Uchenye zapiski TsAGI (Scientific notes of Central Aero-Hydrodynam. Inst.)*, **4**, 73–83.

DMITRUK, A.V. (1987) Quadratic conditions for a Pontryagin minimum in an optimal control problem, linear in the control. *Mathematics of the USSR, Izvestiya,* **28** (2), 275–303.

DMITRUK, A.V. (1990) Maximum principle for a general optimal control problem with state and regular mixed constraints. In: *Optimal'nost' upravlyaemyh dinamicheskih sistem.* Nauka, Moscow, **14**, 26–42; English translation in *Computat. Math. and Modeling,* 1993, **4** (4), 364–377.

DMITRUK, A.V. (1999) Quadratic order conditions of a local minimum for singular extremals in a general optimal control problem. In: G. Ferreyra et al., eds., *Proc. of Symposia in Pure Mathematics,* **64***, "Diff. Geometry and Control".* American Math. Society, 163–198.

DMITRUK, A.V. (2007) Approximation theorem for a nonlinear control system with sliding modes. *Proc. of Steklov Math. Institute,* **256**, 102–114.

DUBOVITSKII, A.YA. (1975) *Integral Maximum principle in the general optimal control problem.* Doctoral Dissertation, Computing Center of the USSR Academy of Sciences, Moscow (in Russian).

DUBOVICKII, A.J. (DUBOVITSKII, A.YA.) (1976) Solution of a problem of S. Ulam on optimal matching of segments. *Mathematics of USSR-Izvestiya,* **10** (3), 639–651.

DUBOVITSKII, A.YA. (1978) The separation, and translations of the Euler equation in linear topological spaces. *Mathematics of the USSR-Izvestiya,* **12** (1), 194–204.

DUBOVITSKIJ, A.YA. and DUBOVITSKIJ, V.A. (1987) The maximum principle in regular optimal control problems with phase trajectory endpoints lying at the boundary of the phase constraint. *Autom. Remote Control,* **48** (12), 1578–1585.

DUBOVITSKIJ, A.YA. and DUBOVITSKIJ, V.A. (1988) On pointwise nontriviality of the maximum principle in problems with phase constraints. *Zeitschrift für Analysis und ihre Anwendungen,* **7** (5), 387–404.

DUBOVITSKII, A.YA. and DUBOVITSKIJ, V.A. (1995) Existence criterion for a significant maximum principle for a problem with phase restrictions. *Differ. Equations,* **31** (10), 1595–1602.

DUBOVITSKII, A.YA. and MILYUTIN, A.A. (1963) Extremum problems in the presence of restrictions. *Soviet Math. Doklady,* **4**, 425–455.

DUBOVITSKII, A.YA. and MILYUTIN, A.A. (1965) Extremum problems in the presence of restrictions. *USSR Comput. Math. and Math. Phys.,* **5** (3), 1–80.

DUBOVITSKII, A.YA. and MILYUTIN, A.A. (1968) Necessary conditions for a weak extremum in optimal control problems with mixed constraints of the inequality type. *ibid.,* **8** (4), 24–98.

DUBOVITSKII, A.YA. and MILYUTIN, A.A. (1969) Translations of Euler's equation. *ibid.,* **9** (6), 37–64.

DUBOVITSKII, A.YA. and MILYUTIN, A.A. (1971) *Necessary conditions of a weak minimum in the general optimal control problem.* Nauka, Moscow (in Russian).

DUBOVITSKII, A.YA. and MILYUTIN, A.A. (1981) Theory of the Maximum principle. In: V.L. Levin, ed., *Methods of the theory of extremal problems in economics.* Nauka, Moscow, 6–47 (in Russian).

DUBOVITSKII, A.YA and MILYUTIN, A.A. (1985) Maximum Principle in linear control problems with convex mixed restrictions. *Zeitschrift für Analysis und ihre Anwendungen,* **4** (2), 133–191.

DUBOVITSKIJ, V.A. (1982) Necessary and sufficient conditions for a Pontryagin minimum in problems of optimal control with singular conditions and generalized controls. *Russ. Math. Surveys,* **37** (3), 202–203.

DUBOVITSKIJ, V.A. (1985) *The Ulam problem of optimal motion of line segments.* Translation Series in Mathematics and Engineering. Optimization Software, Inc., Springer-Verlag, New York.

DYKHTA, V.A. (1994) Variational maximum principle and quadratic optimality conditions for impulse and singular processes. *Siberian Math. Journal,* **35** (1), 65–76.

DYUKALOV, A.N. (1977) An optimally criterion in linear dynamic optimal control problems with mixed constraints. *USSR Comput. Math. Math. Phys.* **16** (4), 31–47.

FULLER, A.T. (1961) Relay control systems optimized for various performance criteria. *Automatic and remote control* (Proc. First IFAC Congress, Moscow, 1960), Butterworth, London (1961), 510–519.

GAMKRELIDZE, R.V. (1962) Optimal sliding states. *Soviet Math. Doklady,* **3** (2), 559–562.

GAMKRELIDZE, R.V. (1977) *Foundations of Optimal Control.* Metsniereba Tbilisi (in Russian).

GAMKRELIDZE, R.V. (1999) Discovery of the maximum principle. *J. of Dynamical and Control Systems,* **5** (4), 437–551.

GAPOSHKIN, V.F. (1972) Convergence and limit theorems for sequences of random variables. *Theory of Probability and Appl.,* **17-3**, 379–400.

GIRSANOV, I.V. (1970) *Lectures on the theory of extremal problems.* Moscow State University Press, Moscow (in Russian).

ILYUTOVICH, A.E. (1993) Numerical methods for optimal control problems, based on the Maximum principle. *Optimal control in linear systems.* Nauka, Moscow (in Russian), Ch. 8.

IOFFE, A.D. and TIKHOMIROV, V.M. (1974) *Theory of extremal problems.* Nauka, Moscow.

KADEC, M. and PELCZYNSKI, A. (1961/62) Bases, lacunary sequences and complemented subspaces in the spaces Lp. *Studia Mathematica* **21**, 161—176.

KLUMOV, A.S. and MERKULOV, A.E. (1984) Speed-optimal control of an astatic second-order loop with nonregular mixed constraints on the control. *Autom. Remote Control,* **45** (12), 1544–1552.

MALANOWSKI, K. (2003) On normality of Lagrange multipliers for state constrained optimal control problems. *Optimization,* **52** (1), 75–91.

MANGASARIAN, O.L. and FROMOVITZ, S. (1967) The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *J. Math. Anal. Appl.,* **17**, 37–47.

MATVEEV, A.S. (1987) Necessary conditions for an extremum in an optimal-control problem with phase restrictions. *Diff. Equations,* **23**, 427–436.

MAURER, H. (1979) On the minimum principle for optimal control problems with state constraints. Preprint no. 41, University of Münster.

MILYUTIN, A.A. (1966) Extremum problems in the presence of constraints. Doctoral Dissertation, Institute of Applied Mathematics, Moscow.

MILYUTIN, A.A. (1970) General schemes of necessary conditions for extrema and problems of optimal control. *Russ. Math. Surveys,* **25** (5), 109–115.

MILYUTIN, A.A. (1990a) Extremals and their properties. *Necessary condition in optimal control.* Nauka, Moscow (in Russian), Ch. 2.

MILYUTIN, A.A. (1990b) Properties of the measure – the Lagrange multiplier at the state constraint. *ibid.,* Ch. 3.

MILYUTIN, A.A. (1990c) A theory of invariance of extremals. *ibid.,* Ch. 4.

MILYUTIN, A.A. (1990d) Maximum principle for the regular systems. *ibid.,* Ch. 5.

MILYUTIN, A.A. (1990e) On the equation $\psi^{(4)} = sign\,\psi$. In: *Optimization of controllable dynamical systems.* VNIISI, Moscow, **1**, 76–84 (in Russian).

MILYUTIN, A.A. (1993) Quadratic optimization. *Optimal control in linear systems.* Nauka, Moscow (in Russian), Chapters 3 and 5.

MILYUTIN, A.A. (1994) An example of an optimal control problem whose extremals possess a continual set of discontinuities of the control function. *Russian Journal of Math. Physics,* **1** (3), 397–402.

MILYUTIN, A.A. (1999) Calculus of variations and optimal control. In: *Calculus of Variations and Optimal Control.* Chapman & Hall/CRC Res. Notes in Math. Series, Boca Raton, FL, **411**, 159–172.

MILYUTIN, A.A. (2000) On a certain family of optimal control problems with a phase constraint. *J. of Math. Sciences,* **100** (5), 2564–2571.

MILYUTIN, A.A. (2001) *Maximum principle in the general optimal control problem.* Fizmatlit, Moscow (in Russian).

MILYUTIN, A.A. and CHUKANOV, S.V. (1993) The problem $\int x^2\,dt \to \min$, $\ddot{x} = u,\ \ u \in U$ and similar problems. *Optimal control in linear systems,* Nauka, Moscow (in Russian), Ch. 6.

MILYUTIN, A.A., DMITRUK, A.V., OSMOLOVSKII, N.P. (2004) *Maximum Principle in Optimal Control.* Mech-Math. Faculty of Moscow State University, Moscow (in Russian).

MILYUTIN, A.A. and OSMOLOVSKII, N.P. (1998) *Calculus of Variations and Optimal Control.* American Math. Society, Providence, RI.

OSMOLOVSKII, N.P. (1988) Necessary and sufficient conditions of a high order for Pontryagin and bounded-strong minima in an optimal control problem. *Soviet Phys. Doklady,* **33** (12), 883–885.

OSMOLOVSKII, N.P. (1993) Quadratic order conditions of extremum in a canonical optimal control problem. *Optimal control in linear systems.* Nauka, Moscow (in Russian), Ch. 1.

OSMOLOVSKII, N.P. (1995) Quadratic conditions for nonsingular extremals in optimal control (a theoretical treatment). *Russian Journal of Math. Physics,* **2** (4), 487–516.

PESCH, H.J. and PLAIL, M. (2009) The Maximum principle in optimal control: a story of ingenious ideas and missed opportunities. In this issue.

PONTRYAGIN, L.S., BOLTYANSKY, V.G., GAMKRELIDZE, R.V. and MISHCHENKO, E.F. (1961) *Mathematical Theory of Optimal Processes.* Moscow, Nauka.

ROBBINS, H.M. (1980) Junction phenomena for optimal control problems with state variable inequality constraints of third order. *JOTA,* **31** (1), 85–99.

ROCKAFELLAR, R.T. (1970) *Convex Analysis.* Princeton Univ. Press.

SAADOUNE, M. and VALADIER, M. (1995) Extraction of a "good" subsequence from a bounded sequence of integrable functions. *J. of Convex Analysis*, **2**, 345–-357.

SMOLJAKOV, E.R. (1968) Optimization of corridor of entering the atmosphere. *Kosmicheskie Issledovaniya (Cosmic Research)*, **1**, 49–57.

SUSSMANN, H.J. AND WILLEMS, J.C. (1997) 300 years of optimal control: From the brachystochrone to the Maximum principle. *IEEE Control Systems Magazine*, **17**, 32–44.

TER-KRIKOROV, A.M. (1976) Some linear problems of optimal control theory with phase constraints. *USSR Comput. Math. Math. Phys.* **15** (1), 51–63.

YAKUBOVICH, V.A. and MATVEEV, A.S. (1994) *Abstract theory of optimal control.* St.-Petersburg University Press, St.-Petersburg (in Russian).

ZELIKIN, M.I. and BORISOV, V.F. (1994) *Theory of Chattering Controls.* Birkhäuser, Boston-Basel-Berlin.

Some of papers by Dubovitskii and Milyutin are available at www.milyutin.ru.