PROOF OF THE MAXIMUM PRINCIPLE
FOR A PROBLEM WITH STATE CONSTRAINTS
BY THE V-CHANGE OF TIME VARIABLE

ANDREI V. DMITRUK
Russian Academy of Sciences, Central Economics and Mathematics Institute
Russia 117418, Moscow, Nakhimovskii prospekt, 47
and Lomonosov Moscow State University, Russia

NIKOLAI P. OSMOLOVSKII
University of Technology and Humanities in Radom, Poland
26-600 Radom, ul. Malczewskiego 20A, Poland
Systems Research Institute, Polish Academy of Sciences, Warszawa
and Moscow State University of Civil Engineering, Russia

ABSTRACT. We give a new proof of the maximum principle for optimal control problems with running state constraints. The proof uses the so-called method of $v-$change of the time variable introduced by Dubovitskii and Milyutin. In this method, the time $t$ is considered as a new state variable satisfying the equation $dt/d\tau = v$, where $v(\tau) \geq 0$ is a new control and $\tau$ a new time. Unlike the general $v-$change with an arbitrary $v(\tau)$, we use a piecewise constant $v$. Every such $v-$change reduces the original problem to a problem in a finite dimensional space, with a continuum number of inequality constraints corresponding to the state constraints. The stationarity conditions in every new problem, being written in terms of the original time $t$, give a weak* compact set of normalized tuples of Lagrange multipliers. The family of these compacta is centered and thus has a nonempty intersection. An arbitrary tuple of Lagrange multipliers belonging to the latter ensures the maximum principle.

1. Introduction. In this paper we propose a new proof of the maximum principle (MP) for optimal control problems with state constraints. There is a vast literature on the subject. The very first results were obtained by R.V. Gamkrelidze at the dawn of the optimal control theory (see e.g. [16, Ch.6]), under some special assumptions on the structure of optimal process. These assumptions were avoided somewhat later in the seminal work [18] by A.Ya. Dubovitskii and A.A. Milyutin, containing a complete form of the MP for state constraints together with a new, quite universal method of its proof. An important part of this method is the $v-$change of time variable and related $v-$variations of the control. Later, several other methods of the proof were used by different authors, such as the so-called sliding mode regimes [13, 20, 14, 15], the smoothly-convex structure of the controlled system [11], the Ekeland variational principle [9, 17, 1], etc. However, all these

2010 Mathematics Subject Classification. Primary: 49K15; Secondary: 49K27, 46N10.
Key words and phrases. Pontryagin maximum principle, $v-$change of time variable, semi-infinite problem, Lagrange multipliers, Lebesgue–Stieltjes measure, function of bounded variation, finite-valued maximality condition, centered family of compacta.
proofs are technically rather complex and not completely mastered even by specialists, not to mention a wider range of readers, albeit with a good mathematical background.

The idea of $v$-change of the time variable, proposed in [18], is as follows. Consider the time $t$, varying in an interval $[t_0, t_1]$, as a new state variable $t = t(\tau)$ that depends on a new time $\tau$, varying in an interval $[\tau_0, \tau_1]$. Let the function $t(\tau)$ satisfy the relations

$$\frac{dt(\tau)}{d\tau} = v(\tau), \quad t(\tau_0) = t_0, \quad t(\tau_1) = t_1, \quad v(\tau) \geq 0 \text{ a.e. on } [\tau_0, \tau_1],$$

where the function $v(\tau)$ (hence the name of the change) is a new control, measurable and essentially bounded. It follows that $t(\tau)$ is a nondecreasing function, mapping a time interval $[\tau_0, \tau_1]$ onto the original time interval $[t_0, t_1]$. The function $t(\tau)$ enables to match for any control $u(t)$ a new control $\tilde{u}(\tau) = u(t(\tau))$ on the set $M_u := \{ \tau \in [\tau_0, \tau_1] : v(\tau) > 0 \}$, while on the set $M_0 := \{ \tau \in [\tau_0, \tau_1] : v(\tau) = 0 \}$ the new control can be defined arbitrarily, with the only condition of belonging $u(\tau) \in U$. Assume e.g. that $v(\tau)$ vanishes on an interval $[\tau', \tau''] \subset [\tau_0, \tau_1]$ of a positive measure. Then we can put $\tilde{u}(\tau) \equiv u_*$ on this interval, where $u_* \in U$ is an arbitrary fixed value, while on the complement to $[\tau', \tau'']$ we can use the formula $\tilde{u}(\tau) = u(t(\tau))$. Thus, we obtain a new control $\tilde{u}(\tau)$ and a new trajectory $\tilde{x}(\tau)$ defined as $\tilde{x}(\tau) = x(t(\tau))$ for all $\tau \in [\tau_0, \tau_1]$. It is easy to transform the initial optimal control problem with the independent variable $t$ into a new problem, corresponding to this change, with the independent variable $\tau$. Moreover, one can easily show that, under such a transformation, any optimal process of the initial problem transforms into an optimal process of the new problem.

Furthermore, it is possible to return from the variable $\tau$ to the original variable $t$ using the “inverse” (to be more precise, right inverse) change. Namely, for any $t \in [t_0, t_1]$, let $\tau(t)$ be the smallest root of the equation $t(\tau) = t$. Then, obviously, $t(\tau(t)) \equiv t$, $\tilde{u}(\tau(t)) = u(t)$, and $\tilde{x}(\tau(t)) = x(t)$ on $[t_0, t_1]$.

It could seem that the selected value $u_*$ does not play any role in such transition, since the interval $[\tau', \tau'']$ maps into a single point $t_* \in [t_0, t_1]$. However, this is true only for the given control $v(\tau)$. But what happens if the function $v(\tau)$ is perturbed by a uniformly small variation $\bar{v}(\tau)$ such that still $v(\tau) + \bar{v}(\tau) \geq 0$ a.e. on $[\tau_0, \tau_1]$? Then the interval $[\tau', \tau'']$ maps onto a small interval $[t', t'']$ with $u(t) = u_*$ on it, so we obtain a needle shape variation of the original control! It is this fact that was used by Dubovitskii and Milyutin to obtain the MP by means of the $v$-change. They systematically used it; see e.g. [18, 20, 14] and other works. The $v$-change turned out to be a very powerful tool for obtaining necessary optimality conditions in the form of MP.

Note that though the $v$-variations are much alike the convenient needle variations, they have advantages. Whereas the usage of needle variations requires the assumption of piecewise continuity of the optimal control, the usage of $v$-variations does not need it, and the optimal control can be an arbitrary measurable bounded function. Moreover, the $v$-variations generate a smooth control system, well defined for $v(\tau)$ of arbitrary sign, where the requirement $v(\tau) \geq 0$ can be regarded as a separate standard constraint, while the needle variations can be considered only for nonnegative widths of the needles, so the terminal value of the state is a function of the needles widths defined just on the positive cone, which is not convenient to differentiate. Thus, the $v$-change of time is a more preferable tool in research of optimality than the needle variations.
Note however, that for an arbitrary nonnegative $v(t)$, the $v$-change is rather complicated technically. To make the proof of MP more simple, A.A. Milyutin proposed in 2001 to use the $v$-change with a piecewise constant function $v(t)$ (obviously, it then suffices to allow only two values: 0 and 1). This idea was realized in [15, 21] for a general optimal control problem of the Pontryagin type (i.e. without state constraints), where a quite simple proof of MP was obtained. Such a primitive $v$-change allowed to pass to a family of smooth optimization problems in finite dimensional spaces, i.e. to problems of mathematical programming, and then to use the well-known necessary optimality conditions in each problem. Proper arranging of the obtained family of these optimality conditions made it possible to pass from them to one universal condition, which had the form of MP. Later it turned out that this approach works as well for problems with state constraints, but then one obtains a family of “semi-infinite” problems in finite dimensional spaces, optimality conditions for which are also known. In the case of one state constraint the passage from a family of finite dimensional optimality conditions to the universal MP is rather simple, see [6], while the case of several state constraints, considered in this article, requires an additional nontrivial trick.

The paper is organized as follows. In Section 2 we give the problem statement and formulate the MP. Section 3 is devoted to the proof of MP. The proof uses some auxiliary facts given in Appendix (Sections A–C). In particular, Section A gives an abstract Lagrange multipliers rule for a nonsmooth optimization problem in a Banach space, Section B presents a generalized formula of integration by parts (obviously, it then suffices to allow only two values: 0 and 1). MP is rather simple, see [6], while the case of several state constraints, considered in this article, requires an additional nontrivial trick.

2. Statement of the problem and the maximum principle. We consider the following optimal control problem on a variable time interval $[t_0, t_1]$:  

$$J := J(t_0, x(t_0), t_1, x(t_1)) \to \min,$$  

(1)  

$$F(t_0, x(t_0), t_1, x(t_1)) \leq 0, \quad K(t_0, x(t_0), t_1, x(t_1)) = 0,$$  

(2)  

$$x(t) = f(t, x(t), u(t)), \quad u(t) \in U \text{ a.e. in } [t_0, t_1],$$  

(3)  

$$\Phi(t, x(t)) \leq 0 \quad \forall t \in [t_0, t_1].$$  

(4)  

Here the function $x(\cdot) : [t_0, t_1] \to \mathbb{R}^n$ is absolutely continuous, $u(\cdot) : [t_0, t_1] \to \mathbb{R}^r$ is measurable and essentially bounded; the mappings $F_0 : \mathbb{R}^{2n+2} \to \mathbb{R}$, $F : \mathbb{R}^{2n+2} \to \mathbb{R}^d(F)$, $K : \mathbb{R}^{2n+2} \to \mathbb{R}^d(K)$, and $\Phi : \mathbb{R}^{n+1} \to \mathbb{R}^d(\Phi)$ are of class $C^1$, while $f : \mathbb{R}^{n+r+1} \to \mathbb{R}^n$ is continuous together with its first derivatives with respect to $t$ and $x$. Here $d(F)$, $d(K)$, and $d(\Phi)$ are the dimensions of $F$, $K$, and $\Phi$, respectively. The set $U \subset \mathbb{R}^r$ is arbitrary. For brevity, the problem (1)–(4) will be called problem A.

A pair of functions $w(\cdot) = (x(\cdot), u(\cdot))$, together with the interval of their definition $[t_0, t_1]$, is called the process of the problem. A process is called admissible if it satisfies all the constraints of the problem. We say that an admissible process $\hat{w}(t) = (\hat{x}(t), \hat{u}(t))$, $t \in [\hat{t}_0, \hat{t}_1]$ is a strong minimum if there exists $\varepsilon > 0$ such that  

$$J(w) \geq J(\hat{w}) \quad \text{for all admissible processes } w(t) = (x(t), u(t)), \quad t \in [t_0, t_1]$$

satisfying the conditions  

$$|t_0 - \hat{t}_0| < \varepsilon, \quad |t_1 - \hat{t}_1| < \varepsilon, \quad \text{and} \quad |x(t) - \hat{x}(t)| < \varepsilon \quad \forall t \in [t_0, t_1] \cap [\hat{t}_0, \hat{t}_1].$$

To ensure nontriviality of MP for a process $\hat{w}(\cdot)$, we assume that  

$$\Phi(t_0, \hat{x}(t_0)) < 0 \quad \text{and} \quad \Phi(t_1, \hat{x}(t_1)) < 0.$$

(5)
In order to formulate the maximum principle for problem A, introduce the Pontryagin function

$$H(t, x, u, p_x) = p_x f(t, x, u),$$

where $p_x$ is a row vector of the dimension $n$, and the endpoint Lagrange function

$$l(t_0, x_0, t_1, x_1) = (\alpha_0 F_0 + \alpha F + \beta K)(t_0, x_0, t_1, x_1),$$

where $\alpha_0$ is a real number, and $\alpha$, $\beta$ are row vectors of the same dimensions as $F$, $K$, respectively (the dependence of $l$ on $\alpha_0$, $\alpha$, $\beta$ is omitted).

We say that, that an admissible process $(x(t), u(t))$, $t \in [t_0, t_1]$ satisfies the maximum principle (MP) if there exist a real number $\alpha_0$, row vectors $\alpha \in \mathbb{R}^d(F)$, $\beta \in \mathbb{R}^d(K)$, nondecreasing functions $\mu_j(\cdot) : [t_0, t_1] \to \mathbb{R}$, $j = 1, \ldots, d(\Phi)$, left continuous functions of bounded variation $p_x(\cdot) : [t_0, t_1] \to \mathbb{R}^n$ and $p_t(\cdot) : [t_0, t_1] \to \mathbb{R}$ such that

1. $\alpha_0 \geq 0$, $\alpha \geq 0$;
2. $\alpha_0 + |\alpha| + |\beta| + \sum_{j=1}^{d(\Phi)} \int_{[t_0, t_1]} d\mu_j(t) > 0$,
3. $\alpha F(t_0, x(t_0), t_1, x(t_1)) = 0$, $\Phi_j(t, x(t)) \cdot d\mu_j(t) = 0$, $j = 1, \ldots, d(\Phi)$,
4. $dp_x(t) = H_x(t, x(t), u(t), p_x(t)) dt - \sum_{j=1}^{d(\Phi)} \Phi_{jx}(t, x(t)) \cdot d\mu_j(t),$
5. $dp_t(t) = H_t(t, x(t), u(t), p_x(t)) dt - \sum_{j=1}^{d(\Phi)} \Phi_{jt}(t, x(t)) \cdot d\mu_j(t),$
6. $p_x(t_0) = I_{x_0}(t_0, x(t_0), t_1, x(t_1))$, $p_x(t_1) = -I_{x_1}(t_0, x(t_0), t_1, x(t_1));$
7. $p_t(t_0) = I_{t_0}(t_0, x(t_0), t_1, x(t_1))$, $p_t(t_1) = -I_{t_1}(t_0, x(t_0), t_1, x(t_1)),
8. $H(t, x(t), u(t), p_x(t)) + p_t(t) = 0$ for a.a. $t \in [t_0, t_1],$
9. $H(t, x(t), u', p_x(t)) + p_t(t) \leq 0$ for all $t \in [t_0, t_1]$ and all $u' \in U$.

Here $p_x$ and $p_t$ are costate (adjoint) variables corresponding to the state variables $x$ and $t$, respectively (so, the subindices $x$ and $t$ do not denote the derivatives)

The adjoint equations $(iv_x) - (iv_t)$ can be understood as equalities between measures. From the second complementarity condition (iii) and assumption (5) it follows that all measures $d\mu_j$ vanish in some neighborhoods of the ends of the interval $[t_0, t_1]$, and therefore, in view of the adjoint equations, the functions $p_x(t)$ and $p_t(t)$ are continuous in these neighborhoods.

Observe that conditions (vi) and (vii) imply the maximality condition for the Pontryagin function:

$$\max_{u' \in U} H(t, x(t), u', p_x(t)) = H(t, x(t), u(t), p_x(t)) \text{ for a.a. } t \in [t_0, t_1].$$

Finally note that, in this definition of MP, we may do not care about the left or right continuity of the adjoint variables $p_x(t)$ and $p_t(t)$, considering them as arbitrary functions of bounded variation, but then condition (vii) should be replaced by the two conditions:

$$p_x(t-0) f(t, \dot{x}(t), u') + p_t(t-0) \leq 0 \text{ and } p_x(t+0) f(t, \dot{x}(t), u') + p_t(t+0) \leq 0 \forall u' \in U, \forall t \in (t_0, t_1).$$

The necessary conditions for a strong minimum are given in the following theorem, first proved in [18] for a somewhat less general problem.

---

1This notation was proposed by Dubovitskii and Milyutin; its convenience is readily revealed when analyzing the MP in problems with several state variables.
Theorem 2.1. If a process \( \hat{w} = (\hat{x}(t), \hat{u}(t)) \mid t \in [\hat{t}_0, \hat{t}_1] \) is a strong minimum in Problem A and satisfies assumption (5), then it satisfies the maximum principle (i)–(vii).

3. Proof of the maximum principle. It is convenient to carry out the proof first for the time-independent case of problem A.

3.1. The autonomous problem B. Consider the following time-independent problem B (a special case of Problem A), still on a non-fixed interval \([t_0, t_1]\):

\[
J := F_0(x(t_0), x(t_1)) \to \min,
\]

\[
F(x(t_0), x(t_1)) \leq 0, \quad K(x(t_0), x(t_1)) = 0,
\]

\[
\dot{x}(t) = f(x(t), u(t)), \quad u(t) \in U \quad \text{a.e. in } [t_0, t_1],
\]

\[
\Phi(x(t)) \leq 0 \quad \forall t \in [t_0, t_1].
\]

Assumption (5) reads here

\[
\Phi(\hat{x}(t_0)) < 0 \quad \text{and} \quad \Phi(\hat{x}(t_1)) < 0.
\]

For this problem, the adjoint equation (vi) \( p(t) = p(x(t), u(t)) = 0 \) for a.a. \( t \in [t_0, t_1] \),

\[
\text{(vi')} \quad p(t)f(x(t), u(t)) \leq 0 \quad \text{for all } t \in [t_0, t_1] \text{ and all } u' \in U.
\]

In the rest, the formulation of the MP remains unchanged.

3.2. Index \( \theta \). Let a process \( (\hat{x}(t), \hat{u}(t)), \ t \in [\hat{t}_0, \hat{t}_1] \) be a strong minimum in Problem B. With this process, we associate a family of finite-dimensional problems \( B^\theta \) and their optimal solutions, numbered by a certain index \( \theta \). By the index, we mean a collection of time instants and control values

\[
\theta = \{(t^1, u^1), \ldots, (t^s, u^s)\},
\]

where \( \hat{t}_0 < t^1 \leq \ldots \leq t^s < \hat{t}_1 \), and the values of \( u^k \in U, \ k = 1, \ldots, s \) are arbitrary. The length \( s = s(\theta) \) of the index depends on \( \theta \).

Next, define an interval \([\tau_0, \tau_1]\) as follows: take the interval \([\hat{t}_0, \hat{t}_1]\), and at the points \( t^1, \ldots, t^s \) insert segments of unit length, always preserving the position of the point \( \hat{t}_0 \). As a result, we obtain the interval \([\tau_0, \tau_1]\) with endpoints \( \tau_0 = \hat{t}_0 \), \( \tau_1 = \hat{t}_1 + s \), and the inserted intervals

\[
\Delta^1 = [t^1, t^1 + 1], \quad \Delta^2 = [t^2 + 1, t^2 + 2], \ldots, \quad \Delta^s = [t^s + (s - 1), t^s + s].
\]

Set

\[
E_0 = \bigcup_{k=1}^s \Delta^k, \quad E_+ = [\tau_0, \tau_1] \setminus E_0,
\]

and define

\[
v^\theta(\tau) = \begin{cases} 0, & \tau \in E_0, \\ 1, & \tau \in E_+ \end{cases}, \quad t^\theta(\tau) = \hat{t}_0 + \int_{\tau_0}^\tau v^\theta(r)dr; \quad \tau \in [\tau_0, \tau_1].
\]

Then

\[
\frac{dt^\theta(\tau)}{d\tau} = v^\theta(\tau), \quad t^\theta(\tau_0) = \hat{t}_0, \quad t^\theta(\tau_1) = \hat{t}_1.
\]
Recall that \( \hat{t} \) is the solution to the Cauchy problem (15) with the function \( v(\cdot) \); therefore, at such a point \( t \), the value \( u^\theta \) of the index \( \theta \) may coincide: \( t^{k'} = \ldots = t^{k''} = t^\theta \), therefore, at such a point \( t \), several consecutive intervals \( \Delta^k \) are inserted in succession, on each of which we have \( v^\theta = 0 \) and \( u^\theta = u^k \), where the value \( u^k \in U \) corresponds to the interval \( \Delta^k \).

So, the set \( E_0 \) consists of a finite number of nonoverlapping intervals \( \Delta^k \), \( k = 1, \ldots, s \), and the set \( E_+ \) also consists of a finite number of some nonoverlapping intervals. (It does not really matter, whether or not they include their endpoints.) Let us unite all these intervals of the sets \( E_0 \) and \( E_+ \) into a general collection, denoted by \( \sigma_k \), \( k = 1, \ldots, m \). So, \( [\tau_0, \tau_1] = \sigma_1 \cup \ldots \cup \sigma_m \), and the different \( \sigma_k \) do not overlap. Denote by \( \chi_k(\cdot) \) the characteristic function of the interval \( \sigma_k \), \( k = 1, \ldots, m \).

### 3.3. Problem \( B^\theta \) of the index \( \theta \)

For a given index \( \theta \), we fix the constructed interval \([\tau_0, \tau_1]\) and fix the function \( u^\theta(\cdot) \). Consider the space \( \mathbb{R}^{m+n} \) with the variables \((z, x_0)\), where \( z = (z_1, \ldots, z_m) \in \mathbb{R}^m \) and \( x_0 = x(\tau_0) \in \mathbb{R}^n \). Set

\[
v(\tau) = \sum_{k=1}^m z_k \chi_k(\tau).\tag{13}
\]

Define the following problem \( B^\theta \) in the space \( \mathbb{R}^{m+n} \):

\[
F_0(x_0, x(\tau_1)) \rightarrow \min,
\]

\[
F(x_0, x(\tau_1)) \leq 0, \quad K(x_0, x(\tau_1)) = 0,\tag{14}
\]

\[
\frac{dz}{d\tau} = v(\tau) f(x(\tau), u^\theta(\tau)), \quad x(\tau_0) = x_0,\tag{15}
\]

\[
-\ z \leq 0, \quad \Phi_j(x(\tau)) \leq 0 \quad \text{on} \quad [\tau_0, \tau_1], \quad j = 1, \ldots, d(\Phi),\tag{16}
\]

where \( x(\cdot) \) is the solution to the Cauchy problem (15) with the function \( v(\cdot) \), defined by (13). We call it the associated problem corresponding to the index \( \theta \). In view of (13), the control system of this problem has actually the form:

\[
\frac{dz}{d\tau} = \sum_{k=1}^m z_k \chi_k(\tau) f(x, u^\theta(\tau)).\tag{17}
\]

Let \( \hat{z}_k \) be the value of \( v^\theta(\cdot) \) on \( \sigma_k \), \( k = 1, \ldots, m \), i.e. \( v^\theta(\cdot) = \sum \hat{z}_k \chi_k(\cdot) \). Recall that \( \hat{z}_k = 0 \) if \( \sigma_k \subset E_0 \), and \( \hat{z}_k = 1 \) if \( \sigma_k \subset E_+ \). We set \( \hat{z} = (\hat{z}_1, \ldots, \hat{z}_m) \).
and \( \hat{x}_0 = \hat{x}(t_0) \). Obviously, for each pair \((z, x_0)\) in a neighborhood of \((\hat{z}, \hat{x}_0)\), the value \( x_1 = x(\tau_1) \) is uniquely determined by equation (17), and moreover, depends smoothly on this pair.

Since the process \((\hat{x}(t), \hat{u}(t)), t \in [\hat{t}_0, \hat{t}_1] \) is a strong minimum in Problem \( B \), it is easy to show that the point \((\hat{z}, \hat{x}_0)\) is a local minimum in Problem \( B^\theta \).

All the data of Problem \( B^\theta \) is smooth, but this problem cannot be regarded as smooth, since it contains a continuum of inequality constraints \( \Phi(x(\tau)) < 0, \tau \in [\tau_0, \tau_1] \). In fact, this is a standard problem of “semi-infinite” mathematical programming, for which one can use the known necessary conditions for a local minimum (see Theorem A.1 in Section A). For the point \((\hat{z}, \hat{x}_0)\) in Problem \( B^\theta \), these conditions are as follows.

Theorem 3.1. If a point \((\hat{z}, \hat{x}_0)\) is a local minimum in problem \( B^\theta \), then there exist a real number \( \alpha_0 \), row vectors \( \alpha \in \mathbb{R}^{d(F)}, \beta \in \mathbb{R}^{d(K)}, \gamma \in \mathbb{R}^m \), and nondecreasing functions \( \mu_j(\cdot) : [\tau_0, \tau_1] \to \mathbb{R}, j = 1, \ldots, d(\Phi) \), such that

\[
\begin{align*}
(i) & \quad \alpha_0 \geq 0, \quad \alpha \geq 0, \quad \gamma \geq 0, \\
(ii) & \quad \alpha_0 + |\alpha| + |\beta| + |\gamma| + \sum_{j=1}^{d(\Phi)} \int_{[\tau_0, \tau_1]} d\mu_j(\tau) > 0, \\
(iii) & \quad \alpha \Phi(\hat{z}_0, \hat{x}_1) = 0, \quad \gamma \hat{z} = 0, \quad \Phi_j(x^\theta(\tau)) d\mu_j(\tau) \equiv 0, \quad j = 1, \ldots, d(\Phi),
\end{align*}
\]

and moreover, the Lagrange function of problem \( B^\theta \)

\[
L(z, x_0) = (\alpha_0 F_0 + \alpha F + \beta K) - \gamma z + \sum_{j=1}^{d(\Phi)} \int_{[\tau_0, \tau_1]} \Phi_j(x^\theta(\tau)) d\mu_j(\tau)
\]

is stationary at the point \((\hat{z}, \hat{x}_0)\), i.e. \( L'(\hat{z}, \hat{x}_0) = 0 \).

Our nearest goal is to decipher these conditions. Note that in view of (12) and (iii), all \( d\mu_j(\tau) = 0 \) in some neighborhoods of \( \tau_0 \) and \( \tau_1 \).

3.4. Stationarity conditions in Problem \( B^\theta \). As above, we use the endpoint Lagrange function \( l = \alpha_0 F_0 + \alpha F + \beta K \) and set for brevity \( F^\theta = f(x^\theta, u^\theta), f_\theta^\theta = f_\theta(x^\theta, u^\theta), \) and \( \mu^\theta(\cdot) = (\mu^\theta_1(\cdot), \ldots, \mu^\theta_{d(\Phi)}(\cdot)) \). Recall that the function \( x(\cdot) \) and its terminal value \( x_1 = x(\tau_1) \) are smoothly determined by the pair \((z, x_0)\) by virtue of equation (17) and the initial condition \( x(\tau_0) = x_0 \). In other words, there is an operator

\[
P : (z, x_0) \in \mathbb{R}^{m+n} \rightarrow x(\cdot) \in C[\tau_0, \tau_1],
\]

whose derivative at the point \((\hat{z}, \hat{x}_0)\) is a linear mapping \( P'(\hat{z}, \hat{x}_0) : (\bar{z}, \bar{x}_0) \rightarrow \bar{x}(\cdot) \), where the function \( \bar{x}(\cdot) \) is the solution to the Cauchy problem for the equation in variations:

\[
\frac{d\bar{x}}{d\tau} = \epsilon^\theta f_\theta(x^\theta, u^\theta) \bar{x} + \sum \bar{z}_k \chi_k f(x^\theta, u^\theta), \quad \bar{x}(\tau_0) = \bar{x}_0.
\]

(18)

The condition \( L'(\hat{z}, \hat{x}_0) = 0 \) means that for any \((\bar{z}, \bar{x}_0)\)

\[
L'(\hat{z}, \hat{x}_0)(\bar{z}, \bar{x}_0) = l_{x_0} \bar{x}_0 + l_x \bar{x}_1 - \gamma \bar{z} + \sum_{j=1}^{d(\Phi)} \Phi_j'(x^\theta(\tau)) \bar{x}(\tau) d\mu_j(\tau) = 0,
\]

(19)

where \( \bar{x}_1 := \bar{x}(\tau_1) \), and \( \bar{x}(\cdot) \) is the solution to (18). (Here, the derivatives of the function \( l(x_0, x_1) \) are taken at the optimal point \((\hat{x}_0, \hat{x}_1)\).)
Now, rewrite this relation in terms of independent variables \((\bar{z}, \bar{x}_0)\). To this end, introduce the function of bounded variation \(p^\theta(\cdot)\) which is the solution to equation

\[
dp^\theta = -v^\theta p^\theta f^\theta_x \, d\tau + \sum_{j=1}^{d(\Phi)} \Phi_j'(x^\theta) \, d\mu_j^\theta, \quad p^\theta(\tau_1) = -l_{x_1}. \tag{20}
\]

Then, by Lemma B.1 (see Appendix), condition (19) takes the form:

\[
(l_{x_0} - p^\theta(\tau_0)) \bar{x}_0 - \sum_k \bar{z}_k \int_{\sigma_k} p^\theta f^\theta \, d\tau = \sum_k \gamma_k \bar{z}_k. \tag{21}
\]

This equality holds for all \(\bar{x}_0 \in \mathbb{R}^n\) and all \(\bar{z}_k\), \(k = 1, \ldots, s\). Hence \(p^\theta(\tau_0) = l_{x_0}\), and for each \(k\)

\[
\int_{\sigma_k} p^\theta f^\theta \, d\tau = -\gamma_k. \tag{22}
\]

Recall that all \(\gamma_k \geq 0\), and according to the complementary slackness condition,
\(\gamma \bar{z} = \sum_k \gamma_k \bar{z}_k = 0\), whence \(\gamma_k \bar{z}_k = 0\) for all \(k\). If \(\sigma_k \subset E_+\), then \(\bar{z}_k = 1\), and then \(\gamma_k = 0\). If \(\sigma_k \subset E_0\), then \(\bar{z}_k = 0\), and then we only know that \(\gamma_k \geq 0\).

Finally, note that one can eliminate \(\gamma\) from the nontriviality condition \((ii)\) of Theorem 3.1. Indeed, if \(\alpha_0 + |\alpha| + |\beta| = 0\) and \(d\mu^\theta \equiv 0\), then \(l_{x_1} = 0\), and equation (20) implies \(p^\theta \equiv 0\), and then \(\gamma = 0\) from (22), which contradicts \((ii)\).

Thus, the preliminary deciphering of the stationarity conditions in Problem \(B^\theta\) results in the following

**Theorem 3.2.** For any index \(\theta\), there exists a tuple \((\alpha_0, \alpha, \beta, \mu^\theta(\cdot))\) with nondecreasing functions \(\mu_j^\theta(\cdot), \, j = 1, \ldots, d(\Phi)\), and the corresponding function of bounded variation \(p^\theta(\cdot)\), both continuous at \(\tau_0, \tau_1\), such that the following conditions hold:

\begin{enumerate}
  \item[(i)] \(\alpha_0 \geq 0\), \(\alpha \geq 0\), \(d\mu_j^\theta \geq 0\), \(j = 1, \ldots, d(\Phi)\),
  \item[(ii)] \(\alpha_0 + |\alpha| + |\beta| + \sum_{j=1}^{d(\Phi)} [\mu_j^\theta(\tau_1) - \mu_j^\theta(\tau_0)] = 1\),
  \item[(iii)] \(\alpha F(\bar{x}_0, \bar{x}_1) = 0\), \(\Phi_j(x^\theta(\tau)) \, d\mu_j^\theta(\tau) = 0\), \(j = 1, \ldots, d(\Phi)\),
\end{enumerate}

\[
dp^\theta = -v^\theta p^\theta f^\theta_x \, d\tau + \sum_{j=1}^{d(\Phi)} \Phi_j'(x^\theta) \, d\mu_j^\theta, \quad p^\theta(\tau_0) = l_{x_0}, \quad p^\theta(\tau_1) = -l_{x_1}. \tag{23}
\]

\[
\int_{\sigma_k} p^\theta f^\theta \, d\tau = \begin{cases} 0 & \text{if } \sigma_k \subset E_+, \\ \leq 0 & \text{if } \sigma_k \subset E_0, \quad k = 1, \ldots, m. \end{cases} \tag{24}
\]

Consider in more detail the second condition in (24). Let \(\sigma_k = [\tau', \tau''] \subset E_0\). Then \(v^\theta(\tau) = v^k\) on \(\sigma_k\). Further, let \(\bar{\sigma}_k = [\tau'_*, \tau''_*]\) be the union of \(\sigma_k\) with all segments of \(E_0\) adjoining \(\sigma_k\) (if any). Then \(v^\theta(\tau) = 0\) on \(\bar{\sigma}_k\), hence \(v^\theta(\tau) = t^k\) is constant on \(\bar{\sigma}_k\), and the value \(x^\theta(\tau)\) is also constant on \(\bar{\sigma}_k\); we denote it by \(\bar{x}_*\). Then, according to (23),

\[
dp^\theta(\tau) = \sum_{j=1}^{d(\Phi)} \, d\mu_j^\theta(\tau) \Phi_j'(\bar{x}_*) \quad \text{on } \bar{\sigma}_k.
\]
hence, for any $\tau \in [\tau'_*, \tau''_*]$  
\[ p^0(\tau) - p^0(\tau'_* - 0) = \sum_{j=1}^{d(\Phi)} [\mu^0_j(\tau) - \mu^0_j(\tau'_* - 0)] \Phi_j(\hat{x}_*). \]  
(25)

The second condition in (24) says that  
\[ \int_{\tau}^{\tau''} p^0(\tau)f(\hat{x}_*, u^k) d\tau \leq 0. \]  
(26)

Substituting here the value $p^0(\tau)$ from (25), we obtain  
\[ p(\tau'_* - 0)f(\hat{x}_*, u^k)(\tau'' - \tau') + \sum_{j=1}^{d(\Phi)} \Phi_j(\hat{x}_*)f(\hat{x}_*, u^k) \int_{\tau}^{\tau''} [\mu^0_j(\tau) - \mu^0_j(\tau'_* - 0)] d\tau \leq 0. \]  
(27)

Since $\mu^0_j(\tau) \leq \mu^0_j(\tau''_* + 0)$ a.e. in $[\tau', \tau'']$ for all $j$, we get  
\[ \int_{\tau}^{\tau''} [(\mu^0_j(\tau) - \mu^0_j(\tau'_* - 0)] d\tau \leq [(\mu^0_j(\tau''_* + 0) - \mu^0_j(\tau'_* - 0)] (\tau'' - \tau'). \]

Hence there exist real numbers $0 \leq \varepsilon_j \leq 1, \ j = 1, \ldots, d(\Phi)$ such that  
\[ \int_{\tau}^{\tau''} [(\mu^0_j(\tau) - \mu^0_j(\tau'_* - 0)] d\tau = \varepsilon_j [(\mu^0_j(\tau''_* + 0) - \mu^0_j(\tau'_* - 0)] (\tau'' - \tau'). \]

Then condition (27) implies  
\[ \left( p(\tau'_* - 0) + \sum_{j=1}^{d(\Phi)} \varepsilon_j [\mu^0_j(\tau''_* + 0) - \mu^0_j(\tau'_* - 0)] \Phi_j(\hat{x}_*) \right) f(\hat{x}_*, u^k) \leq 0. \]  
(28)

Let us now summarize our findings for the index $\theta$, returning back to the original time $t$.

3.5. The finite-valued maximum principle of the index $\theta$. For every $t \in [\hat{t}_0, \hat{t}_1]$ define the value $\tau^\theta(t)$ as the smallest root of the equation $t^\theta(\tau) = t$. Clearly, the function $\tau^\theta : [\hat{t}_0, \hat{t}_1] \rightarrow [\tau_0, \tau_1]$ is surjective, nondecreasing, and left continuous. It has discontinuity at each point $t^k, \ k = 1, \ldots, s$ with the jump $\Delta \tau(t^k) = \tau''_* - \tau'_*$, where $[\tau'_*, \tau''_*]$ is the above maximal interval corresponding to the point $t^k$.

Define the functions  
\[ \mu_j(t) = \mu_j^\theta(\tau^\theta(t)), \quad j = 1, \ldots, s, \quad p(t) = p^\theta(\tau^\theta(t)), \quad t \in [\hat{t}_0, \hat{t}_1]. \]

Obviously, each $\mu_j(t)$ is a nondecreasing function having at every point $t^k$ a jump equal to the measure of the singleton $\{t^k\}$:  
\[ \Delta \mu_j(t^k) := \mu_j(t^k + 0) - \mu_j(t^k - 0) = \mu^\theta(t''_* + 0) - \mu^\theta(t'_* - 0), \]

while $p(t)$ is a function of bounded variation satisfying the equation  
\[ dp(t) = -p(t) f_\theta(\hat{x}(t), \hat{u}(t)) dt + \sum_{j=1}^{d(\Phi)} \Phi_j(\hat{x}(\tau)) d\mu_j(t), \]

and having the same endpoint values as $p^\theta(\tau)$. Set $\mu(t) = (\mu_1(t), \ldots, \mu_{d(\Phi)}(t))$. 

PROOF OF THE MAXIMUM PRINCIPLE BY THE V-CHANGE 2197
Recall that by (12) all \( d\mu_j(\tau) = 0 \) in some neighborhoods of \( \tau_0 \) and \( \tau_1 \), hence, all \( d\mu_j(t) = 0 \) in some neighborhoods of \( \hat{t}_0 \) and \( \hat{t}_1 \), so the function \( p(t) \) is continuous at these points.

Theorem 3.2 can be rewritten in the original time \( t \in [\hat{t}_0, \hat{t}_1] \) as follows:

**Theorem 3.3 (maximum principle of the index \( \theta \)).** For any index \( \theta \), there exist a tuple of multipliers \( (\alpha_0, \alpha, \beta, \mu(t)) \) with nondecreasing functions \( \mu_j(t), \ j = 1, \ldots, d(\Phi) \), and a function of bounded variation \( p(t) \) such that the following conditions hold

1. \( \alpha_0 \geq 0, \quad \alpha \geq 0, \quad d\mu_j \geq 0, \quad j = 1, \ldots, d(\Phi) \),
2. \( \alpha_0 + |\alpha| + |\beta| + \sum_{j=1}^{d(\Phi)} [\mu_j(\hat{t}_1) - \mu_j(\hat{t}_0)] = 1 \),
3. \( \alpha F(\hat{x}_0, \hat{x}_1) = 0, \quad \Phi_j(\hat{x}(t)) d\mu_j(t) \equiv 0, \quad j = 1, \ldots, d(\Phi) \),
4. \( dp(t) = -p(t) f_x(\hat{x}(t), \hat{u}(t)) dt + \sum_{j=1}^{d(\Phi)} \Phi'_j(\hat{x}(t)) d\mu_j(t) \),
5. \( p(\hat{t}_0) = l_{x_0}, \quad p(\hat{t}_1) = -l_{x_1} \),
6. For any neighboring points \( t_k < t_{k+1} \) from the index \( \theta \)

\[
\int_{(t_k, t_{k+1})} p(t) f(\hat{x}(t), \hat{u}(t)) \, dt = 0,
\]

7. For any pair \( (t^k, u^k) \) of the index \( \theta \) there exist real numbers \( \varepsilon_j \in [0, 1], \ j = 1, \ldots, d(\Phi) \), such that

\[
\left( p(t^k - 0) + \sum_{j=1}^{d(\Phi)} \varepsilon_j \Delta \mu_j(t^k) \Phi'_j(\hat{x}(t^k)) \right) f(\hat{x}(t^k), u^k) \leq 0. \tag{29}
\]

Condition (vi) is obtained here from the first condition in (24), taking into account that on each \( \Delta \subset E_x \) the map \( \tau \to t \) is one-to-one, \( \delta^\theta(\tau) = 1 \), and therefore \( dt = dr \). Condition (vii) follows from (28).

(Note that if all \( \mu_j \) are continuous at some \( t^k \), then (29) reduces to a simple inequality \( p(t^k - 0) f(\hat{x}(t^k), u^k) \leq 0 \).)

Thus, for any given index \( \theta \), we obtain a tuple \( (\alpha_0, \alpha, \beta, \mu(\cdot)) \) of Lagrange multipliers and the corresponding function \( p(\cdot) \) such that conditions (i) – (vii) of Theorem 3.3 are satisfied. These multipliers, in general, depend on the index \( \theta \). Conditions (i) – (v) are the same for all indices, while conditions (vi) – (vii) are associated with each individual index. Our goal now is to pass to conditions that do not depend on the index \( \theta \).

**Remark 1.** The above technically simple but quite nontrivial in its idea trick with the replacement of condition (27) in the time \( \tau \) by condition (28) including some unknown numbers \( \varepsilon_j \), that made it possible to pass to the corresponding condition (29) in the time \( t \), was proposed by A.A. Milyutin in his unpublished lectures at Moscow State University in the 1970-s. It is this trick that allows to make the following key step in the proof of MP for problems with several state constraints. In the case of a single state constraint one do not need this trick, since the function \( p^\theta(\tau)f(\hat{x}_*, u^k) \) is monotone, see [6].
3.6. **Transition to the universal MP.** Like in some other proofs of MP [15, 5, 6],
we perform the following procedure. For a given index $\theta$, denote by $\Lambda^\theta$ the set of all tuples
$\lambda = (\alpha_0, \alpha, \beta, \mu(t))$ with nondecreasing functions $\mu_j(t)$, for which there
exists a function of bounded variation $p(\cdot)$ such that conditions $(i) - (vii)$ hold
with some $\varepsilon_j \in [0, 1]$, $j = 1, \ldots, d(\Phi)$. This is a set in the space
$$Y^* = \mathbb{R}^{1+d(F)+d(K)} \times (C^*[t_0, \hat{t}_1]d(\Phi),$$
dual to the space $Y = \mathbb{R}^{1+d(F)+d(K)} \times (C[\hat{t}_0, \hat{t}_1])d(\Phi)$. A key fact is that $\Lambda^\theta$ is a
compact set in some topology (see Lemma C.1 in Appendix).

Thus, taking all possible indices $\theta$, we obtain, for each of them, a nonempty compact set $\Lambda^\theta$. Let us show that the family of all these compacta form a centered
system (i.e., have the finite intersection property). To this end, we introduce a
partial order in the set of all indices. We say that $\theta_1 \preceq \theta_2$ if each pair $(t^k, u^k)$ of
$\theta_1$ is contained in $\theta_2$. It is clear that, for any two indices $\theta_1$ and $\theta_2$, there is a third one containing each of them, for example, their union. Obviously, when the index $\theta$
expands, the set $\Lambda^\theta$ narrows, i.e., the inclusion $\theta_1 \preceq \theta_2$ implies the inverse inclusion
$\Lambda^{\theta_1} \supset \Lambda^{\theta_2}$. Now, let be given any finite collection of compacta $\Lambda^{\theta_1}, \ldots, \Lambda^{\theta_j}$. Take
any index $\theta$ containing all indices $\theta_1, \ldots, \theta_j$. Then the nonempty compact set $\Lambda^\theta$
is contained in each of the compacta $\Lambda^{\theta_1}, \ldots, \Lambda^{\theta_j}$, and therefore, is contained in
their intersection. Hence, the finite intersection property of the family $\{\Lambda^\theta\}$ holds,
and then its total intersection is nonempty:
$$\Lambda_* := \bigcap_{\theta} \Lambda^\theta \neq \emptyset.$$

Now, take an arbitrary tuple of multipliers $\lambda = (\alpha_0, \alpha, \beta, \mu) \in \Lambda_*$ and let $p(\cdot)$ be the
adjoint function corresponding to this tuple. By definition, conditions $(i) - (v)$
hold for this tuple. The fulfillment of condition $(vi)$ in any index $\theta$ means that
for any interval $(t', t'')$,
$$\int_{t'}^{t''} p(t) f(\hat{x}(t), \hat{u}(t)) \, dt = 0,$$
(because there is an index containing the points $t'$, $t''$), which is equivalent to the
fulfillment of equality
$$p(t) f(\hat{x}(t), \hat{u}(t))) = 0 \quad \text{a.e. in } [\hat{t}_0, \hat{t}_1].$$

Finally, the fulfillment of $(vii)$ for the chosen “universal” tuple $\lambda$ implies that
for any $u \in U$ and any point $t \in (\hat{t}_0, \hat{t}_1)$ of continuity of all functions $\mu_j$ (i.e. any
point in $(\hat{t}_0, \hat{t}_1)$ except for a countable set), we have
$$p(t-0) f(\hat{x}(t), u) \leq 0.$$
Then this inequality holds for any $u \in U$ and any interior point of the interval
$[\hat{t}_0, \hat{t}_1]$, and then also for its boundary points, since the function $p(t)$ is continuous
at these points. Assuming, without loss of generality, that the function $p(t)$ is left
continuous, we get: $p(t) f(\hat{x}(t), u) \leq 0$ for all $u \in U$ and all $t \in [\hat{t}_0, \hat{t}_1]$.

Thus, for the chosen tuple $\lambda$ and the corresponding $p$ all conditions $(i) - (vii)$
of the maximum principle are satisfied. Theorem 2.1 for the autonomous Problem
$B$ is proved. 

---

This procedure is in fact the simplest case of an essentially more general one proposed in [19].
3.7. MP for problem A. It remains to show that the MP for problem A can be obtained from the MP for problem B, despite the fact that problem B is a particular case of problem A. Let us show that problem A can be reduced to the form (6)–(9). This is achieved by the following simple trick. To the controlled part of problem B, be obtained from the MP for problem A.

2200 ANDREI V. DMITRUK AND NIKOLAI P. OSMOLOVSKII

where \( \tilde{x} \) represents the "autonomous" conditions. The Pontryagin function for the "autonomous" system (30) is

\[
\tilde{\tilde{A}}(\tau, x) = 0 \quad \text{and} \quad \tilde{\tilde{A}}(\tau, u) = 0,
\]

where \( \tilde{\tilde{A}} \) are given functionals, may also be presented in the form \( f(x, u, \tau) \). We study the local minimality of an admissible point \( x \) and its dual space \( Z \).

It is easy to see that both the admissible and optimal processes of Problems A and A' are in one-to-one correspondence. Therefore, obtaining necessary optimality conditions in Problem B, one can also obtain necessary conditions in Problem A. The Pontryagin function for the "autonomous" system (30) is \( \tilde{H} = p_x f + p_t \), the "autonomous" conditions \( \tilde{H}(x, u) = 0 \) and \( \tilde{H}(x, u') \leq 0 \) take the form \( p_x f(x, u) + p_t = 0 \) and \( p_x f(x, u') + p_t \leq 0 \), respectively, that is, exactly the conditions (vi) and (vii) of MP. The details of this correspondence are left to the reader.

Appendix A. A general Lagrange multipliers rule.

Statement of the problem. Let \( X, Y, Z \) and \( Z_i, \ i = 1, \ldots, \nu \) be Banach spaces, \( D \subset X \) an open set, and \( K_i \subset Z_i, \ i = 1, \ldots, \nu \) closed convex cones with nonempty interiors. Let \( F_0 : D \to R, \ g : D \to Y \), and \( f_i : D \to Z_i, \ i = 1, \ldots, \nu \), be given mappings. Consider the following optimization problem:

\[
\begin{align*}
F_0(x) & \to \text{min}, \\
\{ & f_i(x) \in K_i, \ i = 1, \ldots, \nu, \\
g(x) = 0. 
\}
\end{align*}
\]

Let \( K_i^0 := \{ z_i^* : \langle z_i^*, z_i \rangle \leq 0 \} \) for every \( z_i \in K_i \), \( i = 1, \ldots, \nu \). Here \( \langle z_i^*, z_i \rangle \) is the duality pairing between \( Z_i \) and its dual space \( Z_i^* \).

We study the local minimality of an admissible point \( x^0 \in D \).

It is worth noting that the inequality constraints \( f_i(x) \leq 0 \), where \( f_i : D \to R \) are given functionals, may also be presented in the form \( f_i(x) \in K_i \). If we put \( K_i = R_- := (-\infty, 0] \). Then \( K_i^0 = R_+ := [0, \infty) \). On the other hand, all the inequality constraints \( f_i(x) \in K_i \) can be written as one constraint \( f(x) \in K \) if we define a mapping \( f : X \to Z = Z_1 \times \ldots \times Z_\nu \) by \( f(x) = (f_1(x), \ldots, f_\nu(x)) \), and a cone \( K = K_1 \times \ldots \times K_\nu \) in \( Z \). However, we chose the form (31) to keep a visual relation with convenient statements.

We impose the following

Assumptions. 1) The objective function \( F_0 \) and the mappings \( f_i \) are Fréchet differentiable at \( x_0 \); the operator \( g \) is strictly differentiable at \( x_0 \) (smoothness of
2) the image of the derivative $g'(x_0)$ is closed in $Y$ (weak regularity of the equality constraint).

Though all the mappings in the problem are differentiable, it is not a standard smooth problem, because the cones $K_i$ can be given, in general, by infinite numbers of linear inequalities. In this sense, problem (31) can be regarded as semi-smooth. (If the basic space $X$ is finite-dimensional and at least one of $Z_i$ is infinite-dimensional, problem (31) is usually called semi-infinite.)

The following theorem gives necessary optimality conditions in problem (31).

**Theorem A.1.** Let $x_0$ be a local minimum in problem (31). Then there exist Lagrange multipliers $\alpha_i \geq 0$, $z_i^* \in K_i^0$, $i = 1, \ldots, \nu$, and $y^* \in Y^*$, not all equal zero, satisfying the complementary slackness conditions

$$
\langle z_i^*, f_i(x_0) \rangle = 0, \quad i = 1, \ldots, \nu,
$$

(i.e., each $z_i^*$ is an outer normal to the cone $K_i$ at the point $f_i(x_0)$), and such that the Lagrange function

$$
\mathcal{L}(x) = \alpha_0 F_0(x) + \sum_{i=1}^{\nu} \langle z_i^*, f_i(x) \rangle + \langle y^*, g(x) \rangle
$$

is stationary at $x$: $\mathcal{L}'(x_0) = 0$, i.e.,

$$
\alpha_0 F_0'(x_0) + \sum_{i=1}^{\nu} z_i^* f_i'(x_0) + y^* g'(x_0) = 0.
$$

The proof is based on well-known facts from the functional and convex analysis (see [4, Theorem 4], [7, 8], and references therein). In our opinion, this theorem is reasonably general and, at the same time, most simple and convenient both in the proof and in the practical usage in many situations.

**Remark 2.** The functional $F_0$ in problem (31) can be taken not necessarily smooth; one can take it in the form $F_0(x) = \varphi_0(f_0(x))$, where $f_0 : X \to Z_0$, similarly to other $f_i$, is a smooth mapping, while $\varphi_0$ is a sublinear functional on $Z_0$ such that $\varphi_0(f_0(x_0)) = 0$ and the cone $K_0 = \{ z : \varphi_0(z) \leq 0 \}$ has a nonempty interior. Then the first terms in the functions $\mathcal{L}$ and $\mathcal{L}'(x_0)$ should be replaced, respectively, by $\langle z_i^*, f_i(x) \rangle$ and $z_i^* f_i'(x_0)$, where $z_i^* \in K_i^0$ with $\langle z_i^*, f_0(x_0) \rangle = 0$, which makes them symmetric with respect to all indices $i = 0, 1, \ldots, \nu$. More details can be found in [8].

**Appendix B.** A generalized integration by parts.

**Lemma B.1.** Let an absolutely continuous function $\bar{x}(\tau)$ and a function of bounded variation $p(\tau)$ (both $n$-dimensional), defined on an interval $[\tau_0, \tau_1]$, satisfy equations:

$$
\dot{\bar{x}} = A\bar{x} + \bar{v}, \quad \bar{x}(\tau_0) = \bar{x}_0, \\
dp = -p A d\tau + (d\mu) \varphi, \quad p(\tau_1) = -l_1,
$$

where the matrix $A(\cdot)$ and the function $\bar{v}(\cdot)$ are measurable and bounded, $\varphi(\cdot) : [\tau_0, \tau_1] \to \mathbb{R}^{d(\nu)}$ is continuous, $\mu(\cdot) : [\tau_0, \tau_1] \to \mathbb{R}^{d(\nu)}$ is nondecreasing (for every component), continuous at $\tau_0, \tau_1$, and $l_1$ is a vector from $\mathbb{R}^n$. Then

$$
l_1 \bar{x}_1 + \int_{\tau_0}^{\tau_1} \bar{x} \varphi \, d\mu = -p_0 \bar{x}_0 - \int_{\tau_0}^{\tau_1} p \bar{v} \, d\tau.
$$
First of all, we show that $p$ bounded variation. Here we will assume for convenience that it is left continuous.

**Appendix C. The weak* compactness of the set $\Lambda^\theta$.** Fix any index $\theta$. Recall that the set $\Lambda^\theta$ is defined by the conditions $(i)-(vii)$ of Theorem 3.3.

**Lemma C.1.** The set $\Lambda^\theta$ is weak* compact (that is, a compact set with respect to the usual convergence of finite-dimensional vectors $(\alpha_0, \alpha, \beta)$ and weak* convergence of the measures $\mu$ in the space $C^*$).

**Proof.** Let $\lambda = (\alpha_0, \alpha, \beta, \mu(\cdot)) \in \Lambda^\theta$, and $p(\cdot)$ be the corresponding function of bounded variation. Here we will assume for convenience that it is left continuous.

First of all, we show that $p$ is uniquely determined by the tuple $\lambda$ from the adjoint equation $(iv)$:

$$dp(t) = -p(t) f_x(\hat{x}(t), \hat{u}(t))) dt + \sum_{j=1}^{d(\Phi)} \Phi'_j(\hat{x}(t)) d\mu_j(t), \quad (36)$$

and by any of the endpoint conditions $(v)$, for example, the left one.

Let $W$ be the fundamental matrix of the homogeneous equation $W = -W f_x$ with $W(t_0) = I$. Then for all $t$ the Cauchy formula gives

$$p(t) = p(t - 0) = \left( p(t_0) + \sum_{j=1}^{d(\Phi)} \int_{t_0}^{t-0} \Phi'_j(\hat{x}(s)) W^{-1}(s) d\mu_j(s) \right) W(t). \quad (37)$$

Further, since the weak* topology on the bounded subsets of the space $Y^*$ is metrizable (because $Y$ is separable), to establish the compactness of $\Lambda^\theta$ it suffices to consider an arbitrary sequence of its elements $\lambda^n$. We can assume that the vectors $(\alpha_0, \alpha, \beta)^n$ converge to some $(\alpha_0, \alpha, \beta)$.

By condition $(ii)$, the norms of all vector measures $\mu^n = (d\mu^n_1, \ldots, d\mu^n_{d(\Phi)})$ are uniformly bounded, and all the initial values $p^n(t_0) = (l_{x_0})^n$ are bounded too, hence $(37)$ implies that the functions $p^n(t)$ are uniformly bounded by the common constant. Then, according to the Helly theorems (see, e.g., [12, chapter 6]), there is a subsequence of $p^n(t)$ converging for each $t$ to a function of bounded variation $p(t)$, while the measures $dp^n$ weakly* converge to the measure $dp$. We retain the previous numbering for this subsequence, and also assume that the vector measures $d\mu^n$ weakly* converge to some vector measure $d\mu$ connected with $dp$ by relations $(36)$ and $(37)$.

It is clear that the conditions $(i), (ii), (iv), (v)$ and the first equality in $(iii)$ remain valid under this transition to the limit. Take any $j \in \{1, \ldots, d(\Phi)\}$. The second condition in $(iii)$ means that, on any interval lying between neighboring points $t^k < t^{k+1}$, on which $\Phi'_j(\hat{x}(t)) < 0$, we have $d\mu_j(t) = 0$. This is equivalent to the fact that, for any continuous function $\eta(t)$, whose support is contained in this interval, we have $\int \eta(t) d\mu_j(t) = 0$. Obviously, this property is preserved when passing to the weak* limit.
Since \( p^n(t) \to p(t) \) for all \( t \), by the Lebesgue theorem condition (vi) also withstands the passage to the limit. Therefore, it remains to consider only the last condition (vii).

Fix any \( k \) and consider the pair \((t^k, u^k)\) of the index \( \theta \). To avoid confusion with the sequence numbering, denote here \( t^k = t_* \) and \( u^k = u \). (Recall that \( u^k = u \) is any preassigned point in \( U \).

According to condition (vii), for the pair \((t_*, u)\) there exist numbers \( \epsilon_j^n \in [0, 1], j = 1, \ldots, d(\Phi) \), such that

\[
\left( p^n(t_*) + \sum_{j=1}^{d(\Phi)} \epsilon_j^n \Delta \mu_j^n(t_*) \Phi'_j(\hat{x}(t_*)) \right) f(\hat{x}(t_*), u) \leq 0, \quad n = 1, 2, \ldots \quad (38)
\]

Without loss of generality can we assume that

\[
\epsilon_j^n \to \epsilon_j \quad \text{and} \quad \Delta \mu_j^n(t_*) \to m_j \quad \text{as} \quad n \to \infty, \quad j = 1, \ldots, d(\Phi).
\]

Clearly, \( \epsilon_j \in [0, 1] \) and \( m_j \geq 0 \) for all \( j = 1, \ldots, d(\Phi) \). Recall that \( p^n(t_*) \to p(t_*) \) since the sequence \( p^n \) converges to \( p \) pointwise. Therefore, passing to the limit in (38) as \( n \to \infty \), we get

\[
\left( p(t_*) + \sum_{j=1}^{d(\Phi)} \epsilon_j m_j \Phi'_j(\hat{x}(t_*)) \right) f(\hat{x}(t_*), u) \leq 0. \quad (39)
\]

An important fact is that \( m_j \leq \Delta \mu_j(t_*) \) for all \( j \), since the measures \( d\mu_j^n \) converge to the measure \( d\mu_j \) as \( n \to \infty \). Then there exist reals \( \epsilon_j \in [0, \epsilon_j] \subset [0, 1] \) such that \( \epsilon_j m_j = \epsilon_j \Delta \mu_j(t_*) \), \( j = 1, \ldots, d(\Phi) \), whence condition (39) gives

\[
\left( p(t_*) + \sum_{j=1}^{d(\Phi)} \epsilon_j \Delta \mu_j(t_*) \Phi'_j(\hat{x}(t_*)) \right) f(\hat{x}(t_*), u) \leq 0.
\]

Lemma C.1 is proved. \( \square \)

Acknowledgments. This research was partially supported by the Russian Foundation for Basic Research under grants 16-01-00585 and 17-01-00805.

REFERENCES

[1] L. Bourdin, Note on Pontryagin maximum principle with running state constraints and smooth dynamics – Proof based on the Ekeland variational principle, arXiv:1604.04051 [math.OC].
[2] A. V. Dmitruk, A. A. Milyutin and N. P. Osmolovsky, Lyusternik’s theorem and the theory of extrema, Russian Math. Surveys, 35 (1980), 11–46.
[3] A. V. Dmitruk, On the development of Pontryagin’s Maximum principle in the works of A.Ya. Dubovitskii and A.A. Milyutin, Control and Cybernetics, 38 (2009), 923–957.
[4] A. V. Dmitruk and N. P. Osmolovskii, Necessary conditions for a weak minimum in optimal control problems with integral equations subject to state and mixed constraints, SIAM J. on Control and Optimization, 52 (2014), 3437–3462.
[5] A. V. Dmitruk and N. P. Osmolovskii, On the proof of Pontryagin’s Maximum principle by means of needle variations, Journal of Mathematical Sciences, 218 (2016), 581–598.
[6] A. V. Dmitruk and N. P. Osmolovskii, Variations of the type of \( v- \) change of time in problems with state constraints, Proc. of the Institute of Mathematics and Mechanics, the Ural Branch of Russian Academy of Sciences, 24 (2018), 76–92 (in Russian).
[7] A. V. Dmitruk and N. P. Osmolovskii, A General Lagrange Multipliers Theorem, Constructive Nonsmooth Analysis and Related Topics (CNSA-2017), IEEE Xplore Digital Library, 2017.
[8] A. V. Dmitruk and N. P. Osmolovskii. A General Lagrange Multipliers Theorem and Related Questions Control Systems and Math. Methods in Economics, (G. Feichtinger et al. eds.), Lecture Notes in Economics and Math. Systems, 687 (2018), 165–194.
[9] I. Ekeland, Nonconvex minimization problems, Bull. of American Math. Society (New Series), 1 (1979), 443–474.
[10] I. V. Girsanov, Lectures on Mathematical Theory of Extremum Problems, Lecture Notes in Economics and Mathematical Systems, 1972.
[11] A. D. Ioffe and V. M. Tikhomirov, Theory of Extremal Problems, North-Holland Publishing Company, Amsterdam, 1979, Russian Edition: Nauka, Moscow, 1974.
[12] A. N. Kolmogorov and S. V. Fomin. Elements of the Theory of Functions and Functional Analysis, Dover Books on Mathematics, 1999; Russian 4th Edition: Nauka, Moscow, 1976.
[13] A. A. Milyutin, General schemes of necessary conditions for extrema and problems of optimal control, Russian Mathematical Surveys, 25 (1970), 110–116.
[14] A. A. Milyutin, Maximum Principle in the General Optimal Control Problem [in Russian], Fizmatlit, Moscow, 2001.
[15] A. A. Milyutin, A. V. Dmitruk and N. P. Osmolovsky, Maximum Principle in Optimal Control, Moscow State University, Moscow, 2004 (in Russian).
[16] L. S. Pontryagin, V. G. Boltyanski, R. V. Gamkrelidze and E. F. Mishchenko, The Mathematical Theory of Optimal Processes, Fitzmatgiz, Moscow; English translation: Pergamon Press, New York, 1964.
[17] R. Vinter, Optimal Control, Birkhauser, Boston, 2000.
[18] A. Ya. Dubovitskii and A. A. Milyutin, Extremum problems in the presence of restrictions, USSR Comput. Math. and Math. Physics, 5 (1965), 1–80.
[19] A. Ya. Dubovitskii and A. A. Milyutin, Translation of Euler’s equations, USSR Comput. Math. and Math. Physics, 9 (1969), 37–64.
[20] A. Ya. Dubovitskii and A. A. Milyutin, Theory of the maximum principle, Methods of the Theory Of Extremal Problems in Economics (V.L. Levin ed.), Nauka, Moscow, (1981), 6–47 (in Russian, see http://www.milyutin.ru/).
[21] Optimal’noe upravlenie [Optimal Control], (N.P. Osmolovskii and V.M. Tikhomirov eds.), Moscow Center for Continuous Mathematical Education (MCCME), Moscow, Russia, 2008 (in Russian).

Received December 2017; revised January 2019.

E-mail address: avdmi@cemi.rssi.ru
E-mail address: osmolovski@uph.edu.pl