Dedicated to
Professor Czesław Olech

Control and Cybernetics

vol. **34** (2005) No. 3

On necessary conditions in variable end-time optimal control problems

by

A. Ioffe[1]

Department of Mathematics, The Technion
Haifa 32000, Israel

Abstract: The paper contains a new necessary optimality condition for optimal control problems with free terminal time and discontinuous time dependence, which actually is a family of conditions, each corresponding to a continuation of optimal trajectory beyond the optimal time for an arbitrary small interval.

Keywords: optimal control, necessary optimality condition, differential inclusion, variable end-time.

1. Introduction

We prove here a new necessary optimality condition for optimal control problems with variable end-time and measurable dependence of data on time. This is actually not one but a family of necessary conditions, each one associated with an extension of the optimal trajectory beyond the optimal time. This family includes the earlier necessary condition of Clarke and Vinter which is associated with some special extension (naturally connected with the maximum principle), as is explicit in the proof given by Vinter (2005). It has to be emphasized that scanning through all possible extensions may give more information. An illustrative example will be given in the next section.

In the problem we consider, the dynamics is described by a differential inclusion. Existence theorems for differential inclusions therefore play a crucial role in the proof. These are the classical Filippov existence theorem (Filippov, 1967) and its extension to unbounded differential inclusion recently given by Ioffe (2005). The other main instrument of the proof is the "optimality alternative" - a general principle whose embryonic version was used already in Ioffe

[1]The research was supported in part by the US-Israel Binational Fund under the grant 2000157

(1997) and later in Vinter (2000) to prove necessary conditions for fixed end-time problems. In the proof here it is used twice, each time in a different way, once through reduction to a fixed-time problem depending on a parameter.

The technical machinery is standard for modern non-smooth analysis. We refer to Rockafellar and Wets (1998) for all necessary details. Here we just note that ∂ always means the limiting subdifferential in $I\!R^n$ and $N(S,\cdot)$ and $D^*G(\cdot)$ stand for the associated normal cone and coderivative.

2. Statement of the problem and the main result

We shall consider the following problem:

$$(\mathbf{P}) \qquad \begin{array}{ll} \text{minimize} & \varphi(T, x(T)), \\ \text{s.t.} & \dot{x} \in F(t,x) \quad \text{a.e. on } [0,T]; \\ & (T, x(T)) \in S,\ x(0) = x_0. \end{array}$$

Here $S \subset I\!R_+ \times I\!R^n$, $x_0 \in I\!R^n$ is a fixed point and F is a set-valued mapping from $I\!R_+ \times I\!R^n$ into $I\!R^n$. Thus, the variables in the problem are pairs $(T, x(\cdot))$, where $T > 0$ and $x(\cdot)$ is an absolutely continuous $I\!R^n$-valued function defined on $[0,T]$.

In what follows we fix some admissible $(\overline{T}, \overline{x}(\cdot))$ and assume the following about the data of the problem:

$(\mathbf{A}_1)$ φ satisfies the Lipschitz condition near $(\overline{T}, \overline{x}(\overline{T}))$ and S is a closed set;

$(\mathbf{A}_2)$ there are: (a) an open set $\mathcal{O} \subset I\!R \times I\!R^n$ containing the graph of $\overline{x}(\cdot)$, (b) a function $r(t) \geq 0$ defined and summable on the projection of $\mathcal{O}$ to the t-axis and (c) $\beta > 0$ such that for any $N \geq 0$

$$F(t,x) \subset F(t,x') + (r(t) + \beta N)\|x - x'\|B, \quad \text{whenever } (t,x) \in \mathcal{O},\ (t,x') \in \mathcal{O}.$$

$(\mathbf{A}_3)$ There are a constant $K > 0$ and a neighborhood $\mathcal{U} \subset I\!R \times I\!R^n$ of $(\overline{T}, \overline{x}(\overline{T}))$ such that the sets $F(t,x)$ are uniformly bounded for $(t,x) \in U$ and

$$F(t,x) \subset F(t,x') + K\|x - x'\|B, \quad \text{whenever } (t,x) \in \mathcal{U},\ (t,x') \in \mathcal{U}.$$

As follows from $(\mathbf{A}_2)$ in view of the main theorem of Ioffe (2005) it is possible to extend $\overline{x}(t)$ as a solution to the differential inclusion beyond $\overline{T}$. ($(\mathbf{A}_3)$ allows to use a slightly stronger Fillippov's existence theorem (see e.g. Aubin, Cellina, 1984; Filippov, 1967). However, certain statements in this paper rely solely on $(\mathbf{A}_2)$ and do not depend on $(\mathbf{A}_3)$.) To avoid confusion, we would like once again emphasize that by "extension of a solution of the differential inclusion" we always mean a solution of the inclusion which coincides with the original solution on its domain.

Finally, let

$$H(t,x,p) = \sup_{u \in F(t,x)} \langle p, u \rangle$$

be the Hamiltonian of the system.

THEOREM 2.1 *Assume* ($\mathbf{A}_1$)-($\mathbf{A}_2$). *If* $(\overline{T}, \overline{x}(\cdot))$ *solves the problem, then for any subflow* Φ *of the inclusion containing* $\overline{x}(\cdot)$ *there are a* $p(\cdot) \in W^{1,\infty}$ $\lambda \geq 0$ *and* $\mu \in I\!R$ *such that the following relations are satisfied*

(a) $\lambda + \|p(\cdot)\| > 0$;

(b) $(0, -p(\overline{T})) \in \lambda\partial\varphi((\overline{T}, \overline{x}(\overline{T}))) + N(S, (\overline{T}, \overline{x}(\overline{T}))) + Q(\overline{T}, \overline{x}(\overline{T}))(-p(\overline{T}))$;

(c) $\dot{p}(t) \in \text{conv}\,\{w : \ (w, p(t)) \in N(\text{Graph}\, F(t,\cdot), (\overline{x}(t), \dot{\overline{x}}(t)))\}$ a.e. on $[0, \overline{T}]$;

(d) $\langle p(t), \dot{\overline{x}}(t)\rangle = H(t, \overline{x}(t), p(t))$, a.e. on $[0, \overline{T}]$.

Moreover, in the normal case (when no $p(\cdot)$) *satisfying (a)-(d) with* $\lambda = 0$ *may exist), the conclusion of the theorem holds only under* ($\mathbf{A}_1$) *and* ($\mathbf{A}_2$).

Here $D^*\overline{y}(T)(\cdot)$ stands for the coderivative of $\overline{y}(\cdot)$, associated with the *limiting proximal subdifferential* (see e.g. Rockafellar and Wets, 1998).

REMARK 2.1 The Clarke-Vinter condition (Clarke and Vinter, 1989; Vinter, 2000) differs from Theorem 2.1 at one point: instead of the second part of (b), they claim that

$$0 \in \mu + \operatorname*{ess}_{t \to \overline{T}} H(t, \overline{x}(\overline{T}), p(\overline{T}))$$

where the set on the right is the set of essential values (see the definition in the next section) of the function $H(\cdot, \overline{x}(\overline{T}), p(\overline{T}))$ at $\overline{T}$. [1] The following example shows that the condition of Theorem 2.1 is stronger.

Consider first a fixed time autonomous problem

$$\begin{array}{ll} \text{minimize} & \varphi(x(1)), \\ \text{s.t.} & \dot{x} \in F(x) \quad \text{a.e. on } [0,1]; \\ & x(1) \in S,\ x(0) = x_0, \end{array}$$

with a Lipschitz bounded-valued F. Let $\overline{x}(\cdot)$ be a solution. We assume that the problem is normal at $\overline{x}(\cdot)$ which means that the collection $\mathcal{P}$ of $p(\cdot)$ satisfying the conditions (c) and (d) of the theorem (with $\overline{T}$ replaced by 1) along with

$$-p(1) \in \partial\varphi(\overline{x}(1)) + N(S, \overline{x}(1));$$

is bounded. Assume also that $\inf\{\|p(1)\| : \ p(\cdot) \in \mathcal{P}\} = \beta > 0$.

As F does not depend on t, for any q the Hamiltonian H is continuous and its essential value at any point is unique and equal to its value at the point. Set $\alpha = \inf\{H(\overline{x}(\overline{T}), p(\overline{T})) : \ \ p(\cdot) \in \mathcal{P}\}$. Take a $\xi < \min\{\alpha, 0\}$ and consider the problem

$$\begin{array}{ll} \text{minimize} & \varphi(x(T)) + \xi T, \\ \text{s.t.} & \dot{x} \in F(t,x) \quad \text{a.e. on } [0,T]; \\ & x(T) \in S,\ x(0) = x_0 \end{array}$$

[1] As we have already mentioned, this condition corresponds to a specific choice of the continuation $y(t)$, namely such that $\langle p(\overline{T}), \dot{y}(t)\rangle$ well approximates $H(t, \overline{x}(\overline{T}), p(\overline{T}))$ near $\overline{T}$.

with the same $\varphi(x)$ and $F(t,x)$ equal to $F(x)$ for $t \le 1$ and for $t > 1$ defined as follows:

$$F(t,x) = r|\sin\frac{1}{t-1}|B,$$

where $r > |\xi|/\beta$. Then for any $p(\cdot) \in \mathcal{P}$ we have (for $t > 1$): $H(t, x(1), p(1)) = r\|p(1)\||\sin(\frac{1}{t-1})|$ which means that

$$\operatorname*{ess}_{t\to 1} H(t, \overline{x}(1), p(1)) \supset [0, \beta r]$$

and therefore $(1, \overline{x}(\cdot))$ satisfies the necessary condition of Clarke-Vinter.

On the other hand, let $\overline{y}(t)$ be the extension of $\overline{x}(\cdot)$ identically equal to $\overline{x}(1)$ for $t > 1$. Then $\overline{y}(t)$ is Lipschitz, so by the well known scalarization formula, $D^*\overline{y}(\overline{T})(q)$ is equal to the (limiting proximal) subdifferential of $\langle q, \overline{y}(t)\rangle$ at $\overline{T}$. If $p(\cdot) \in \mathcal{P}$, then in view of the condition (d)

$$\langle p(t), \overline{y}(t)\rangle = \int_0^t H(\overline{x}(s), p(s))ds + \text{const},$$

if $t \le 1$ and $\langle p(\overline{T}), \overline{y}(t)\rangle = \text{const}$ if $t > 1$ according to the definition of $\overline{y}$. It follows that the subdifferential of the function $t \to -\langle p(\overline{T}), \overline{y}(t)\rangle$ at $\overline{T}$ lies within conv $\{-p(\overline{T}), 0\}$.

If the condition (b) of the theorem was satisfied at $\overline{y}$, then we would have $\mu = \xi$ and $-\xi \in$ conv $\{-p(\overline{T}), 0\}$ which does not hold as by definition $-\xi > \max\{0, -\alpha\} \ge \max\{0, -p(\overline{T})\}$.

Remark 2.2 Condition ($\mathbf{A}_3$) introduced in Clarke and Vinter (1989), although indispensable in the proof of the "singular part" of Theorem 2.1, looks somewhat artificial in combination with ($\mathbf{A}_2$). It seems natural to ask what kind of a result can be obtained if this condition is dropped. It turns out that a certain result can be obtained which, however, is less precise than Theorem 2.1 because instead of a single continuation of a single trajectory, we have to deal with some "subflows" of the differential inclusion.

Let us agree to call a set-valued mapping $\Phi(t,s,x)$ into $\mathbb{R}^n$ ($0 \le t \le s \le \overline{T}$) a *subflow* of the inclusion $F(t,x)$ if

(i) for any t the graph of $\Phi(t,\cdot,\cdot)$ is closed;

(ii) if $z \in \Phi(\tau, s, x)$, then there is a solution $x(t)$ of the inclusion such that $x(\tau) = x$ and $x(s) = z$;

(iii) $\Phi(t,t,x) = \{x\}$ and if $0 < t < \tau < s \le \overline{T}$, then

$$\Phi(t,s,x) = \bigcup_{u\in\Phi(t,\tau,x)} \Phi(\tau, s, u).$$

A trivial example of a subflow is the set-valued mapping generated by all solutions of the differential inclusion. The existence theorem of Ioffe (2005) allows for construction of the nontrivial subflows as follows: take any solution $\overline{x}(t)$ defined e.g. on $[0, T]$. Then by the theorem (under $(\mathbf{A}_2)$) there is an $\varepsilon > 0$ such that for any $t \in [0, T]$ and any x with $\|x - \overline{x}(t)\| \leq \varepsilon$ there is a solution $u_t(\cdot)$ defined on $[t, T]$ and satisfying

$$u(t) = x, \quad \int_t^T \|\dot{u}(s) - \dot{\overline{x}}(s)\| ds \leq N \int_t^T d(\dot{\overline{x}}(s), F(s, \overline{x}(s) + x - \overline{x}(t)))ds$$

with the constant N depending only on ε. Take all such solutions and set

$$\Phi(t, s, x) = \{u : \exists\, u_t(\cdot),\ u_t(t) = x,\ u_t(s) = u\}.$$

Given a solution $x(t)$ of the differential inclusion defined on $[0, T]$, we say that the subflow Φ *contains* $x(\cdot)$ if $x(s) \in \Phi(t, s, x(t))$ for all $0 \leq t \leq s \leq T$. Finally, set $G(s, x) = \Phi(t, s, x)$ and $Q(s, x(\cdot))(p) = \limsup_{t \to s} D^* G(s, x(t))(p)$.

THEOREM 2.2 *Assume* $(\mathbf{A}_1)$*-*$(\mathbf{A}_2)$*. If* $(\overline{T}, \overline{x}(\cdot))$ *solves the problem, then for any subflow* Φ *of the inclusion containing* $\overline{x}(\cdot)$ *there are a* $p(\cdot) \in W^{1,\infty}$ $\lambda \geq 0$ *and* $\mu \in I\!R$ *such that the following relations are satisfied*

(a) $\lambda + \|p(\cdot)\| > 0$;

(b) $(0, -p(\overline{T})) \in \lambda \partial \varphi((\overline{T}, \overline{x}(\overline{T}))) + N(S, (\overline{T}, \overline{x}(\overline{T}))) + Q(\overline{T}, \overline{x}(\overline{T}))(-p(\overline{T}))$;

(c) $\dot{p}(t) \in \text{conv}\, \{w : (w, p(t)) \in N(\text{Graph } F(t, \cdot)(\overline{x}(t), \dot{\overline{x}}(t)))\}$ a.e. on $[0, \overline{T}]$;

(d) $\langle p(t), \dot{\overline{x}}(t) \rangle = H(t, \overline{x}(t), p(t))$, a.e. on $[0, \overline{T}]$.

We omit the proof of Theorem 2.2 as it follows the same pattern as the proof of Theorem 2.1 and is actually even simpler in certain respects.

3. An auxiliary problem

As in Ioffe (1997) we start with a generalized Bolza problem:

$$(\mathbf{BP}) \quad \text{minimize} \quad J(T, x(T)) = l(T, x(\cdot)) + \int_0^T L(t, x(t), \dot{x}(t))dt, \quad x(0) = x_0$$

in which the length of the time interval is not fixed. But first we consider another problem with integration over a fixed time interval, namely:

$$(\mathbf{BPP}) \qquad \text{minimize} \quad l(a, x(1)) + \int_0^1 L(t, x(t), \dot{x}(t))dt, \quad x(0) = x_0,$$

which differs from the standard formulation by the presence of a parameter a in the terminal part of the functional. Here $x(\cdot)$ is of the same kind as above, absolutely continuous with values in $I\!R^n$ and $a \in I\!R^m$. About the components of the problem we assume the following (see Ioffe, 1997; Vinter, 2000):

$(\mathbf{A}_4)$ $l(a, x)$ is l.s.c., with values in $(-\infty, \infty]$ and $l(\overline{a}, \overline{x}(1))$ finite;

($\mathbf{A}_5$) there is an $\varepsilon > 0$ such that whenever $x(t)$ is continuous and satisfies $\|x(t) - \overline{x}(t)\| < \varepsilon$, the value $L(t, x(t), y)$ is finite for all y and for each y the function $t \to L(t, x(t), y)$ is measurable;

($\mathbf{A}_6$) for any N there are a summable $k(t)$ and $c(t)$ such that the inequalities

$$L(t,x,y) - L(t,x',y) \le k(t)\|x - x'\|, \quad L(t,x,y) \ge c(t)$$

hold almost everywhere on $(0,1)$ whenever $\|x - \overline{x}(t)\| < \varepsilon$, $\|x' - \overline{x}(t)\| < \varepsilon$ and $\|y\| \le N$.

PROPOSITION 3.1 *Suppose that $(\overline{a}, \overline{x}(\cdot))$ furnishes a local minimum in (**BPP**) (with respect to the norm topology of $I\!R^m \times W^{1,1}$) and that ($\mathbf{A}_4$)-($\mathbf{A}_6$) hold. Then there is a $p(\cdot) \in W^{1,1}$ such that following three relations are satisfied:*

(a) $\dot{p}(t) \in \mathrm{conv}\{w : (w, p(t)) \in \partial L(t, \overline{x}(t), \dot{\overline{x}}(t))\}$ *a.e. in t;*

(b) $L(t, \overline{x}(t), u) \ge L(t, \overline{x}(t), \dot{\overline{x}}(t)) + \langle p(t),\, u - \dot{x}_*(t)\rangle$ *for all u, a.e. in t;*

(c) $(0, -p(1)) \in \partial l(\overline{a}, \overline{x})$.

The proof is straightforward: we treat a as, say the value at zero of an additional state variable, say $x_0(t)$ and apply the necessary optimality condition for the standard generalized Bolza problem (Ioffe and Rockafellar, 1996; Vinter, 2000) to the following problem:

$$\text{minimize} \quad m(z(0), z(1)) + \int_0^1 M(t, z(t), \dot{z}(t))dt,$$

where $z(t) = (x_0(t), x(t))$, $m(z(0), z(1)) = l(x_0(0), x(1))$ and $M(t, z, \dot{z}) = L(t, x, \dot{x})$.

We can now state and prove a first order necessary optimality condition for (**BP**). To begin with, we have to explain in which sense we shall understand local minimum in this problem. We shall fix a certain sufficiently big time interval, say $(0,1)$ and assume that the integrand L is defined on $[0,1] \times I\!R^n \times I\!R^n$. If now $x(t)$ is an absolutely continuous function defined on $[0,T]$ with $T < 1$, we extend $x(t)$ to $[0,1]$ by setting $x(t) = x(T)$ for $t > T$. The distance between $(T, x(\cdot))$ and $(T', x'(\cdot))$ will now be measured by the $I\!R \times W^{1,1}$-norm of the difference $(T - T', x(\cdot) - x'(\cdot))$, that is, it is equal to

$$|T - T'| + \int_0^1 \|\dot{x}(t) - \dot{x}'(t)\| dt.$$

We denote by $\mathcal{X}$ the collection of all pairs $(T, x(\cdot))$, where $0 \le T \le 1$, $x(\cdot)$ is absolutely continuous, equal to x_0 at zero and constant on $[T, 1]$. With the above defined distance, $\mathcal{X}$ is a complete metric space. We can now reformulate our problem as the problem of minimizing J on $\mathcal{X}$.

Let $f(t)$ be a summable real-valued function and

$$F(T) = \int_a^T f(t)dt.$$

Elementary arguments show that

$$\partial F(T) \subset \operatorname*{ess}_{t\to T} f(t). \tag{1}$$

PROPOSITION 3.2 *Assume that*

- $l(T,x)$ *satisfies the Lipschitz condition near* $(\overline{T}, \overline{x}(\overline{T}))$*;*
- $L(t,x,u)$ *satisfies* $(\mathbf{A}_5)$*-*$(\mathbf{A}_6)$ *with* $\overline{x}(\cdot)$ *replaced by* $\overline{y}(\cdot)$*;*
- $L(t,\overline{y}(t),\dot{\overline{y}}(t))$ *is essentially bounded near* $\overline{T}$*.*

Suppose $(\overline{T},\overline{x}(\cdot))$ *furnishes a local minimum in (*$\mathbf{BP}$*). Then for any absolutely continuous extension* $\overline{y}(t)$ *of* $\overline{x}(t)$ *beyond* $\overline{T}$ *there are a* $p(\cdot) \in W^{1,\infty}$ *and a* $\mu \in I\!R$ *such that following three relations are satisfied:*

(a) $\dot{p}(t) \in \operatorname{conv}\{w : (w,p(t)) \in \partial L(t,\overline{x}(t),\dot{\overline{x}}(t))\}$ *a.e. in* $(0,\overline{T})$*;*

(b) $L(t,\overline{x}(t),u) \geq L(t,\overline{x}(t),\dot{\overline{x}}(t)) + \langle p(t),\, u - \dot{x}_*(t)\rangle$ *for all* u*, a.e. in* $(0,\overline{T})$*;*

(c) $(\mu, -p(\overline{T}) \in \partial l(\overline{T},\overline{x}(\overline{T}));\quad 0 \in \mu + D^*\overline{y}(\overline{T})(-p(\overline{T})) + \operatorname*{ess}_{t\to\overline{T}} L(t,\overline{x}(t),\dot{\overline{x}}(t)).$

Proof. Fix a certain absolutely continuous extension $\overline{y}(t)$ of $\overline{x}(t)$ from $[0,\overline{T}]$ to all of $[0,1]$ (not necessarily constant). For any $\tau \in [0,\overline{T}]$ we define the mapping

$$G_\tau(T) = \overline{y}(T) - \overline{y}(\tau) = \int_\tau^T \dot{\overline{y}}(t)dt$$

from R into $I\!R^n$ and set $G = G_{\overline{T}}$. Finally, let us recall that the interval of *essential values* of a function $f(t)$ at T is defined as (see Vinter, 2000)

$$\operatorname*{ess}_{t\to T} f(t) = [\operatorname{ess} - \liminf_{t\to T} f(t), \operatorname{ess} - \limsup_{t\to T} f(t)].$$

Observe that G differs from $\overline{y}$ by a constant and therefore coderivatives of G and $\overline{y}$ coincide.

For any positive $\tau < \overline{T}$, let $A(\tau)$ be the collection of pairs $(T,x(\cdot))$ such that $T > \tau$ and $\dot{x}(t) = \dot{\overline{y}}(t)$ for $t \in (\tau, T)$. For $(T,x(\cdot)) \in A(\tau)$

$$\begin{aligned} J(T,x(\cdot)) =& l(T, x(\tau) + G_\tau(T)) + \int_0^\tau L(t,x(t),\dot{x}(t))dt \\ &+ \int_\tau^T L(t, x(\tau) + \overline{y}(t) - \overline{y}(\tau), \dot{\overline{y}}(t))dt. \end{aligned}$$

By $(\mathbf{A}_6)$

$$\int_\tau^T L(t,x(\tau)+\overline{y}(t)-\overline{y}(\tau),\dot{\overline{y}}(t)) \leq \int_\tau^T L(t,\overline{y}(t),\dot{\overline{y}}(t))dt + K(T-\tau)\|x(\tau)-\overline{x}(\tau)\|,$$

where $K = \int_\tau^T k(t)dt$ (of course $\overline{x}(\tau) = \overline{y}(\tau)$) with the equality if $(T, x(\cdot)) = (\overline{T}, \overline{x}(\cdot))$

Set

$$\varphi(T,x) = l(T, x + G_\tau(T)) + \int_\tau^T L(t, \overline{y}(t), \dot{\overline{y}}(t))dt + (T-\tau)\|x - \overline{x}(\tau)\|.$$

Then

$$I(T, x(\cdot)) = \varphi(T, x(\tau)) + \int_0^\tau L(t, x(t), \dot{x}(t))dt$$

attains at $(\overline{T}, \overline{x}(\cdot))$ a local minimum in $I\!R \times W^{1,1}(0,\tau)$. The problem of minimizing I on $I\!R \times W^{1,1}(0,\tau)$ is the Bolza problem on fixed interval with a parameter T, that is a problem of the type considered in Proposition 3.1. It is clear that all assumptions of the proposition are satisfied in our case (moreover φ is even a continuous function), so the proposition can be applied. It follows that there is a summable $p_\tau(t)$ on $[0,\tau]$ such that

$$\dot{p}_\tau(t) \in \mathrm{conv}\big\{w : (w, p_\tau(t)) \in \partial L(t, \overline{x}(t), \dot{\overline{x}}(t))\big\}\text{a.e. in}(0,\tau), \tag{2}$$

$$(0, -p_\tau(\tau)) \in \partial\varphi(\overline{T}, \overline{x}(\tau)). \tag{3}$$

and the condition (b) of the proposition (with $p(\cdot)$ replaced by $p_\tau(\cdot)$ is satisfied for all almost $t \in (0,\tau)$.

The function φ is the sum of three continuous terms two of which satisfy the Lipchitz condition (the integral term thanks to the essential boundedness assumption). Therefore the limiting subdifferential of φ is a subset of the sum of limiting subdifferentials of the terms in the same point. The subdifferential of the integral term is estimated by the essential value of the integrand (as was mentioned prior the statement of the proposition). The subdifferential of the first term at $(\overline{T}, \overline{x}(\tau))$ by the standard chain rule of the subdifferential calculus belongs to the set of all $(\lambda, q) \in I\!R \times I\!R^n$ such that $\lambda \in \mu + D^*G_\tau(\overline{T})(q)$ and $(\mu, q) \in \partial l(\overline{T}, \overline{x}(\overline{T}))$. Thus by (3) we have to conclude that there is a $(\mu_\tau, q_\tau) \in \partial l(\overline{T}, \overline{x}(\tau))$ such that

$$0 \in \mu_\tau + D^*G(\overline{T})(q_\tau) + \operatorname*{ess}_{t\to\overline{T}} L(t, \overline{x}(t), \dot{\overline{x}}(t)); \quad -p_\tau(\tau) \in q_\tau + K|\overline{T} - \tau|B. \tag{4}$$

If we extend $p_\tau(\cdot)$ to the entire $(0,\overline{T})$ by $p_\tau(t) = p_\tau(\tau)$ for $t > \tau$, then these extensions form a weakly precompact set in in $W^{1,1}(0,\overline{T})$ (due to the Lipschitz property ($\mathbf{A}_6$)), so we may assume that, when $\tau \to \overline{T}$ they converge to some $p(t)$ for which we get from (2) using the standard argument (see Ioffe and Rockafellar,1996) the inclusion

$$\dot{p}(t) \in \mathrm{conv}\big\{w : (w, p(t)) \in \partial L(t, \overline{x}(t), \dot{\overline{x}}(t))\big\}\text{a.e. in } (0,\overline{T})$$

which is (a). Finally, as l satisfies the Lipschitz condition, the set of (μ_τ, q_τ) is bounded and we get (c) as a consequence of (4). As to (b), it is straightforward. ■

4. Proof of the main theorem

As was mentioned in Section 2, there is a solution $\overline{y}(t)$ of the differential inclusion defined on a certain longer segment $[0,\tilde{T}]$ with $\tilde{T} > \overline{T}$ and coinciding with $\overline{x}(t)$ on $[0,\overline{T}]$. Without loss of generality we may assume that $\tilde{T} = 1$. Of course, there maybe many such extensions. We shall fix one throughout the proof assuming that the graph of the extension is also contained in $\mathcal{O}$. Then the standard transformation $w \to \dot{\overline{x}}(t) + w$ allows to rewrite $(\mathbf{A}_2)$ as follows:

$(\mathbf{A}_2')$ there are: an $\varepsilon > 0$, a function $r(t) \geq 0$ (other than in $(\mathbf{A}_3)$ but still summable) on $[0,1]$ and a $\beta > 0$ such that for any $N \geq 0$ and almost every $t \in [0,1]$

$$F(t,x)\bigcap B(\dot{\overline{y}}(t),N) \subset F(t,x') + (r(t)+\beta N)\|x-x'\|B, \quad \text{if } \|x-\overline{y}(t)\| < \varepsilon\,.$$

Step 1: Optimality alternative. Consider an abstract minimization problem

$$\text{minimize} \quad f(x), \text{ s.t. } x \in M \subset X,$$

where X is the domain (metric) space, M is the constraint set and f is the cost function.

Theorem 4.1 (Optimality alternative) *Suppose that (X,d) is a metric space, $\overline{x} \in M$ and f satisfies the Lipschitz condition near $\overline{x}$. Let further $\Phi(x)$ be a nonnegative extended-real-valued function equal to zero at $\overline{x}$. Suppose finally that $\overline{x}$ is a local solution of the problem. Then the following alternative holds:*

– either there is a $\lambda > 0$ such that the function $\lambda f + \Phi$ attains an unconditional minimum at $\overline{x}$;

– or there is a sequence (x_n) converging to $\overline{x}$ such that $\Phi(x_n) < n^{-1}d(x_n,M)$. In particular, if X is complete, M is closed and Φ is lower semicontinuous, then there is a sequence (z_n) converging to $\overline{x}$ such that $z_n \notin M$ and each of the function $\Phi(x) + n^{-1}d(x,z_n)$ attains an absolute minimum at z_n.

Proof. Indeed, either there are a neighborhood of $\overline{x}$ and $R > 0$ such that $R\Phi(x) \geq d(x,M)$ for all x of the neighborhood, or there is a sequence (x_n) converging to $\overline{x}$ with $2n\Phi(x_n) < d(x_n,M)$. In the first case, as f is Lipschitz (e.g. with constant L), we can choose for any x close to $\overline{x}$ a $u \in M$ such that, say $d(x,u) \leq 2d(x,M)$. Then

$$f(x) \geq f(u) - Ld(x,u) \geq f(\overline{x}) - 2LR\Phi(x).$$

In the second case, if X is complete, M is closed and Φ l.s.c., we apply Ekeland's variational principle to the function $\Phi(x)$. As it is nonnegative, we have $\Phi(x_n) \leq \inf\Phi + (2n)^{-1}d(x_n,M)$, so Ekeland's principle guarantees the existence of (z_n) such that $d(x_n,z_n) \leq d(x_n.M)/2$ and $\Phi(x)+n^{-1}d(x,z_n)$ attains an absolute minimum at z_n. ∎

Step 2: Problems without end-point constraints. We first apply the optimality alternative to the problem

$$(\mathbf{P}_1) \qquad \begin{array}{ll} \text{minimize} & l(T, x(T)), \\ \text{s.t.} & \dot{x} \in F(t, x) \quad \text{a.e. on } [0, T]; \\ & x(0) = x_0 \end{array}$$

in which the end-point constraint $(T, x(T)) \in S$ is absent.

As the domain metric space we take the same $\mathcal{X}$ that was introduced in the preceding section. The set $\mathcal{M}$ of admissible elements consists of elements $(T, x(\cdot))$ of $\mathcal{X}$ whose second components are solutions to the differential inclusion on $[0, T]$ with the initial condition $x(0) = x_0$. We define the perturbation function $\Phi(t, x(\cdot))$ as follows:

$$\Phi(T, x(\cdot)) = \int_0^T d(\dot{x}(t), F(t, x(t)))dt$$

Here $d(x, Q)$ stands for the distance from x to Q.

Let $(T, w(\cdot)) \in \mathcal{M}$. It follows from the main theorem of Ioffe (2005) and $(\mathbf{A}_2')$ that for $(T, x(\cdot)) \in \mathcal{X}$ such that the $\max_{0 \le t \le T} \|x(t) - y(t)\|$ is sufficiently small

$$d((T, x(\cdot)), \mathcal{M}) \le R \int_0^T d(\dot{x}(t), F(t, x(t)))dt$$

with constant R not depending on $x(\cdot)$. This means (see the proof of the optimality alternative) that only the regular case may take place if we take

$$\Phi(T, x(\cdot)) = \int_0^T d(\dot{x}(t), F(t, x(t)))dt.$$

In other words, if $(\overline{T}, \overline{x}(\cdot))$ is a local minimum in the problem, then there is a $\lambda > 0$ such that $(\overline{T}, \overline{x}(\cdot))$ gives an unconditional local minimum to

$$\lambda l(T, x(T)) + \int_0^T d(\dot{x}(t), F(t, x(t)))dt.$$

Thus we get a Bolza problem of the same type as $(\mathbf{BP})$ as considered in the preceding section with the function $d(\dot{x}(t), F(t, x(t))$ playing the role of the integrand L.

This means that we can apply Proposition 3.2 to get necessary optimality condition in the problem. So let $\overline{y}(t)$ be any absolutely continuous extension of $\overline{x}(t)$ beyond $\overline{T}$, and let $p(\cdot)$ and μ be the corresponding multipliers. As the subdifferential of the distance function to a set at a point of the set belongs to the normal cone to the set at the point, the part (a) of the proposition implies that

$$\dot{p}(t) \in \operatorname{conv}\{w : (w, p(t)) \in N(\operatorname{Graph} F(t, \cdot), (\overline{x}(t), \dot{\overline{x}}(t)))\} \quad \text{a.e. in } (0, \overline{T}).$$

The Weierstrass condition (b) for L being the distance function to $F(t,x)$ leads to

$$\langle p(t), \dot{\overline{x}}(t)\rangle = H(t, \overline{x}(t), p(t)), \quad \text{a.e. on } [0, \overline{T}].$$

Finally, the last condition (c) of Proposition 3.2 gives

$$(\mu, -p(\overline{T}) \in \partial l(\overline{T}, \overline{x}(\overline{T})); \quad 0 \in \mu + D^*G(\overline{T})(-p(\overline{T}))$$

as the integrand is identical to zero along $\overline{x}(t)$.

This proves the theorem for problems without end-point constraints.

Step 3: General case. Here we shall apply the optimality alternative in a different way. As the domain space X we shall consider the collection of all pairs $(T, x(\cdot))$ such that $x(t)$ is a solution to the differential inclusion defined on $[0,T]$ and starting from $(0, x_0)$. In other words, the domain space now is precisely the collection of admissible elements at the previous step. The metric in X is the same as above (induced from $\mathcal{X}$), so X is a complete metric space. As to the admissible set, it now consists of all elements of X satisfying the terminal condition $(T, x(T)) \in S$.

We apply optimality alternative with

$$\Phi(T, x(\cdot)) = d((T, x(T)), S).$$

It follows that

– either there is a $\lambda > 0$ such that $(\overline{T}, \overline{x}(\cdot))$ is an unconditional local minimum on X of

$$\lambda\varphi(T, x(T)) + d((T, x(T), S)$$

(regular case),

– or there is a sequence $(T_k, x_k(\cdot))$ converging to $(\overline{T}, \overline{x}(\cdot))$ and *not admissible in the problem* and a sequence (ε_k) of positive numbers converging to zero such that for any n the function

$$\Phi_n(T, x(\cdot)) = d((T, x(T)), S) + \varepsilon_n(|T - T_k| + \int_0^T |\dot{x}(t) - \dot{x}_k(t)| dt)$$

attains on X its absolute minimum at $(T_k, x_k(\cdot))$ (singular case).

In the regular case $(\overline{T}, \overline{x}(\cdot))$ is a local solution of $(\mathbf{P}_1)$ with $l(t,x) = \lambda\varphi(t,x) + d((t,x), S)$. As this function satisfies the Lipschitz condition near $(\overline{T}, \overline{x}(\overline{T}))$, the result of Step 2 applies and immediately leads to the the proof of the theorem.

The singular case is more complicated, and is actually the only case in which the boundedness condition $(\mathbf{A}_3)$ must be put at work. By the extension theorem for differential inclusions for any k there is an extension $y_k(t)$ of $x_k(t)$ beyond T_k satisfying the differential inclusion and

$$\int_{T_k}^t \|\dot{y}_k(s) - \dot{\overline{y}}(s)\| ds \le R \int_{T_k}^t d(\dot{\overline{y}}(s), F(s, x_k(T_k) + \overline{y}(s) - \overline{y}(T_k)) ds$$

with R not depending on k.

It is now possible to apply the result of Step 2 to each of the problems

$$(\mathbf{P}_k)\qquad \begin{array}{ll} \text{minimize} & d((T,x(t)),S)+\varepsilon_k\int_0^T\|\dot x(t)-\dot y_k(t)\|dt,\\ \text{s.t.} & \dot x\in F(t,x),\ \text{a.e.};\ x(0)=x_0. \end{array}$$

(The additional integral term, certainly, does not cause any additional problem.) It follows that for any k there are $\mu_k\in I\!R$ and $p_k(\cdot)\in W^{1,\infty}$ and $q_k(\cdot)\in L^\infty$ (both on $[0,T_k]$) such that $\|q_k(\cdot)\|_\infty\le 1$ and

(α) $(\mu,-p_k(T_k))\in\partial d(\cdot,S)(T_k,x(T_k)),\ 0\in\mu_k+D^*y_k(T_k)(-p_k(T_k))$;

(β) $\dot p_k(t)\in\operatorname{conv}\big\{w:(w,p_k(t)-\varepsilon_kq_k(t))\in N(\operatorname{Graph}F(t,\cdot),(x_k(t),\dot x_k(t)))$ $+\{0\}\times\epsilon_kB\big\}$ a.e. in $(0,T_k)$.

(γ) $\langle p_k(t),\dot x_k(t)\rangle=H(t,x_k(t),p_k(t)),$ a.e. on $[0,T_k]$.

Since $(T_k,x_k(\cdot))$ is not admissible in $(\mathbf{P})$, $(T_k,x_k(T_k))\notin S$ (as a $(T_kx_k(\cdot))\in X$). This means that $|\mu_k|+\|p_k(T_k)\|=1$.

On the other hand, since $x_k(\cdot)$ converges uniformly to $\overline{x}(\cdot)$ on every segment $[0,\tau],\ \tau<\overline{T}$ and $T_k\to\overline{T}$, it follows from $(\mathbf{A}_2)$ that $\|\dot p_k(\cdot)\|\le r(t)$ almost everywhere if k is sufficiently large. Standard arguments (e.g. Ioffe, 1967; Vinter, 2000) now prove the existence of a limiting pair $(\mu,p(\cdot))$ to which $(\mu_k,p_k(\cdot))$ converge, the latter weakly in $W^{1,1}$ on every $[0,\tau],\ \tau<\overline{T}$ and the limiting $(\mu,p(\cdot))$ satisfies the relations (a), (c) and (d) and the first part of (b).

To prove that the second relation in (b) also holds, we first note that, thanks to $(\mathbf{A}_3)$, $y_k(t)$ satisfies the Lipschitz condition in a neighborhood of $\overline{T}$ with constant, not depending on k. In this case, again applying the scalarization formula of Ioffe (1984, Proposition 8), we get

$$D^*y_k(t)(p)=\partial\langle p,y_k(\cdot)\rangle(T).$$

Fix a $\tau<\overline{T}$. Then $T_k>\tau$ for sufficiently large k. For $t\le T_k$ we have by (γ), setting $\alpha_k(\tau)=\langle p_k(T_k),y_k(\tau)\rangle$:

$$\begin{aligned}\langle p_k(T_k),y_k(t)\rangle&=\alpha_k(\tau)+\int_0^t\langle p_k(T_k),\dot x_k(s)\rangle ds\\ &=\alpha_k(\tau)+\int_0^t\langle p_k(T_k)-p_k(s),\dot x_k(s)\rangle ds+\int_0^t H(s,x_k(s),p_k(s))ds.\end{aligned}$$

Since $\dot x_k(t)$ are uniformly bounded in a neighborhood of $\overline{T}$ and $p_k(\cdot)$ are uniformly continuous and converge to $p(\cdot)$, this equality implies that for any $\varepsilon>0$ there is a τ such that the derivative of $\langle p_k(T_k),y_k(t)\rangle$ at any $t\in(\tau,T_k)$ (if it exists) differs from $H(t,\overline{x}(t),p(\overline{T}))=\langle p(\overline{T},\overline{y}(t)-\overline{y}(\tau))\rangle$ by less than ε.

On the other hand as follows from Theorem 1 of Ioffe (2005) (and its proof) and $(\mathbf{A}_3)$, there is an N not depending on k for $t>T_k$

$$\|\dot y_k(t)-\dot{\overline{y}}(t)\|\le N\int_{T_k}^t d(\dot{\overline{y}}(s),F(s,\overline{y}(s)+x_k(T_k)-\overline{y}(T_k)))ds.$$

(Here we set $\overline{y}(t) = \overline{x}(t)$ for $t < \overline{T}$.)

This shows that for any $\varepsilon > 0$, we can choose a positive δ such that the derivatives of $y_k(t)$ and $\overline{y}(t)$ differ by less than ε if k is sufficiently large also if $T_k < t < \overline{T} + \delta$. It is now an easy matter to conclude, that the coderivatives $D^*y_k(T_k)(-p_k(T_k))$ Hausdorff converge to $D^*\overline{y}(\overline{T})(-p(\overline{T}))$.

References

Aubin, J.-P. and Cellina, A. (1984) *Differential Inclusions.* Springer.

Clarke, F. and Vinter, R.B. (1989) Application of optimal multiprocesses, *SIAM J. Control Optimization* **27, 1048-1071.**

Filippov, A.F. (1967) Classical solutions of differential inclusions with multivalued right-hand sides. *SIAM J. Control Optimization* **5, 609-621.**

Ioffe, A.D. (1984) Approximate subdifferentials and applications 1: The finite dimensional theory. *Trans. Amer. Math. Soc.* **281**, 389-416.

Ioffe, A.D. (1997) Euler-Lagrange and Hamiltonian formalisms in dynamic optimization. *Trans. Amer. Math. Soc.* **349**, 2871-2900.

Ioffe, A.D. (2005) Existence and relaxation theorems for unbounded differential inclusions. *J. Convex Analysis.* To appear.

Ioffe, A.D. and Rockafellar, R.T. (1996) The Euler and Weierstrass conditions for nonsmooth variational problems. *Cal. Variations and Partial. Differential Eq.* **4**, 59-87.

Rockafellar, R.T. and Wets, R.J.B. (1998) *Variational Analysis.* Springer.

Vinter, R.B. (2000) *Optimal Control.* Birkhäuser.