Maximum principles for optimal control problems with differential inclusions

Alexander D. Ioffe
Department of Mathematics, Technion

International Conference:

Challenges and advances in modern variational analysis

dedicated to Jean-Paul Penot's 80th birthday

Limoge, March 2023

The problem:

$$(\mathbf{OC}) \qquad \begin{aligned} \text{minimize} \quad & \varphi(x(\cdot)), \\ \text{s.t.} \quad & \dot{x} \in F(t,x), \\ & (x(0), x(T)) \in S, \end{aligned}$$

Here:

φ is a function on the space of $\mathbf{R}^n$-valued continuous functions

F is a set-valued mapping into $\mathbf{R}^n$

S is a closed subset of $\mathbf{R}^{2n}$.

We are interested in necessary conditions for a given absolutely continuous $\overline{x}(\cdot)$ to be a local minimum in the problem in one or another sense.

Specifically, we shall consider local minima in the $W^{1,1}$ topology and in the topology of uniform convergence.

Standard optimal control problem with state constraints is a particular case of the stated problem. Namely, the problem

$$\begin{aligned} \text{minimize} \quad & \ell(x(0), x(T)), \\ \text{s.t.} \quad & \dot{x} \in F(t, x), \\ & (x(0), x(T)) \in S, \\ & g(t, x(t)) \leq 0, \ \forall\, t \end{aligned}$$

is reduced to (**OC**) if we set

$$\varphi(x(\cdot)) = \max\{\ell(x(0), x(T)) - \ell(\overline{x}(0), \overline{x}(T)), \max_t g(t, x(t))\}$$

Subdifferentials and normal cones in $\mathbf{R}^n$

Let f be an lsc extended-real-valued function on $\mathbf{R}^n$ finite at x.

Proximal subdifferential: $y \in \partial_p f(x)$ if there are $\varepsilon > 0$ and $k > 0$ such that

$$f(x+h) - f(x) \geq \langle y, h\rangle - k|h|^2, \quad \text{whenever } |h| < \varepsilon.$$

Limiting subdifferential: $y \in \partial f(x)$ if there are $x_k \to x$ and $y_k \to y$ such that $y_k \in \partial_p f(x_k)$.

Limiting normal cone: given $S \subset \mathbf{R}^n$ and $x \in S$; then $N(S,x) := \partial i_S(x)$, where i_S is the indicator of S (the function equal to zero on S and $+\infty$ outside of S).

Clarke's normal cone: $N_c(S,x) = \operatorname{cl}\operatorname{conv} N(S,x)$.

Clarke's generalized gradient:

$$\partial_c f(x) = \{y : \ (y,-1) \in N_c(\operatorname{epi} f, (x, f(x)))\}$$

NB: $\partial f(x)$ is a bounded set iff f is Lipschitz near x in which case $\partial_c f(x) = \operatorname{conv} \partial f(x)$.

Subdifferentials and normal cones in Banach spaces

Let X be a separable Banach space and f a function on X which is Lipschitz in a neighborhood of x

Dini–Hadamard subdifferential of f at x:

$$\partial^- f(x) = \{x^* \in X^* : \ \langle x^*, h\rangle \le d^- f(x,h), \ \forall\, h \in X\},$$

where $d^- f(x;h) = \liminf_{t \searrow 0} t^{-1}(f(x+th) - f(x))$

Limiting subdifferential: $x^* \in \partial f(x)$ if there are $x_k \to x$ in the norm topology and $x_k^* \to x^*$ in the weak*-topology such that $x_k^* \in \partial^- f(x_k)$.

Clarke's generalized gradient: $\partial_c f(x) = \operatorname{cl}\operatorname{conv} \partial f(x)$.

Limiting and Clarke normal cones: given an $S \subset X$ and $x \in S$,

$$N(S,x) = \bigcup_{\lambda \ge 0} \lambda \partial d(x,S), \quad N_c(S,x) = \bigcup_{\lambda \ge 0} \lambda \partial_c d(x,S)$$

where $d(u,S)$ is the distance from u to S.

General assumptions:

(H_1) φ is is Lipschitz in a C-neighborhood of $\overline{x}(\cdot)$

(H_2) $S \subset \mathbf{R}^n \times \mathbf{R}^n$ is a closed set;

(H_3) F is a measurable mapping with respect to the σ-algebra generated by products of Lebesgue measurable subsets of $[0, T]$ and Borel subsets of $\mathbf{R}^n$,

(H_4) Graph $F(t, \cdot)$ is a closed set in $\mathbf{R}^{2n}$ for almost every t.

Statements of main theorems: $W^{1,1}$-minimum

Let $U(t)$ be the set of $u \in F(t, \overline{x}(t))$ such that the inequality

$$|d(y, F(t,x)) - d(y, F(t,x'))| \le k|x - x'| \qquad (*)$$

holds for all $x, x' \in B(\overline{x}(t), \varepsilon),\ y \in B(u, \delta)$ with some positive $k,\ \varepsilon,\ \delta$ (depending on u and t).

NB: This simply means that $U(t)$ is the set of $u \in F(t, \overline{x}(t))$ such that $F(t, \cdot)$ has the Aubin (pseudo-Lipschitz) property at $(\overline{x}(t), u)$.

To state the theorem we also need the following additional assumption:

(H_5) there are measurable $\delta(t) > 0$, $\xi \in (0,1)$, $\varepsilon > 0$ and summable $k(t) > 0$ such that for almost every t the inequality $(*)$ holds with $u = \dot{\overline{x}}(t)$, $k = k(t)$ and $\delta = \delta(t)$, along with $F(t,x) \cap B(\dot{\overline{x}}(t), \xi\delta(t)) \neq \emptyset$ for almost every t.

Theorem 1. *Assume* (H_1)*-*(H_5)*. If* $\overline{x}(\cdot) \in W^{1,1}$ *is a* $W^{1,1}$*-local minimum in* (**OC**)*, then there are* $\lambda \geq 0$*, an* $\mathbf{R}^n$*-valued function* $p(t)$ *of bounded variation and a nonnegative measure* $\nu \in \lambda \partial \varphi(\overline{x}(\cdot))$ *such that* $\lambda + |p(t)| \neq 0$ *and the following relations* (i)*–*(iv) *are satisfied with some* $\mathbf{R}^n$*-valued summable* $q(t)$*:*

(i) $\quad p(t) + \int_t^T q(\tau)d\tau + \int_t^T \nu(d\tau) = \text{const};$

(ii) $\quad q(t) \in \text{conv}\,\{q : \ (q, p(t)) \in N(\text{Graph } F(t,\cdot), (\overline{x}(t), \dot{\overline{x}}(t)))\}$

(iii) $(p(0), -p(T)) \in (\nu\{0\}, \nu\{T\}) + N(S, (x(0), x(T)));$

(iv) $\quad \langle p(t), \dot{\overline{x}}(t)\rangle \geq \langle p(t), u\rangle, \ \forall\, u \in U(t)$ a.e. on $[0, T]$.

Theorem 1. *Assume* (H_1)*-*(H_5)*. If* $\overline{x}(\cdot) \in W^{1,1}$ *is a* $W^{1,1}$*-local minimum in* (**OC**)*, then there are* $\lambda \geq 0$*, an* $\mathbf{R}^n$*-valued function* $p(t)$ *of bounded variation and a nonnegative measure* $\nu \in \lambda \partial \varphi(\overline{x}(\cdot))$ *such that* $\lambda + |p(t)| \neq 0$ *and the following relations* (*i*)*–*(*iv*) *are satisfied with some* $\mathbf{R}^n$*-valued summable* $q(t)$*:*

(*i*) $p(t) + \int_t^T q(\tau)d\tau + \int_t^T \nu(d\tau) = \text{const};$

(*ii*) $q(t) \in \text{conv}\,\{q : \ (q, p(t)) \in N(\text{Graph } F(t,\cdot), (\overline{x}(t), \dot{\overline{x}}(t)))\}$

(*iii*) $(p(0), -p(T)) \in (\nu\{0\}, \nu\{T\}) + N(S, (x(0), x(T)));$

(*iv*) $\langle p(t), \dot{\overline{x}}(t)\rangle \geq \langle p(t), u\rangle, \ \forall\, u \in U(t)$ a.e. on $[0, T]$.

Remarks. 1. If $\varphi(x(\cdot)) = \ell(x(0), x(T))$, then (i) reduces to $\dot{p}(t) = q(t)$ and (iii) can be transformed to

$$(p(0), -p(T)) \in \lambda \partial \ell(x(0), x(T)) + N(S, (x(0), x(T)))$$

2. Conditions (i)-(iii) are necessary for the weak minimum (with respect to the norm $W^{1,\infty}$-topology).

Comments. 1. The problem was originally studied by Clarke in 1975 with $\varphi(x(\cdot)) = \ell(x(0), x(T))$, F with convex and bounded values and fully Lipschitz in x, that is such that $F(t,x) \subset r(t)B$, and $F(t,x) \subset F(t,x') + c(t)B$, for all $x, x' \in B(\overline{x}(t), \varepsilon)$ (with some summable $k(t)$, $c(t)$ and $\varepsilon > 0$). Clearly, in this case $U(t) = F(t, \overline{x}(t))$. The adjoint inclusion that appeared in Clarke's paper had the form

$$(\dot{p}(t), p(t)) \in N_c(\text{Graph } F(t,\cdot), (\overline{x}(t), \dot{\overline{x}}(t)))$$

and the transversalty condition also involved Clarke's normal cone

2. The "partially convexified" adjoint inclusion (ii) and transversality condition with limiting normal were introduced by Smirnov in 1991 (for the same type of F as in Clarke's 1975 paper) and by Loewen and Rockafellar in 1994 for a class of F with unbounded, but still convex, values, satisfying a certain "integrably sub-Lipschitz" condition again implying that $U(t) = F(t, \overline{x}(t))$.

3. Convexity assumption on the values of F was somewhat weakened by Mordukhovich in 1995 (under otherwise very strong assumptions on F) and fully removed in 1997 in papers by Ioffe and Vinter–Zheng.

4. Finally, Clarke in 2005 extended the theorem to F satisfying a certain "pseudo-Lipschitz condition of radius $R(t)$" in which $U(t)$ appeared to be equal to $F(t, \overline{x}(t)) \cap B(\dot{\overline{x}}(t), R(t))$.

Necessary conditions for a strong minimum

Assume that there are a summable positive-valued $k(\cdot)$, and positive β and ε such that

(H_6) the function $x \to d(y, F(t,x))$ is $(k(t) + \beta|y|)$-Lipschitz on $B(\overline{x}(t), \varepsilon)$ for almost all t;

(H_7) for all $N > 0$ and all $x \in B(\overline{x}(t), \varepsilon)$

$$(\operatorname{conv} F(t,x)) \cap B(\dot{\overline{x}}(t), N) \subset \operatorname{conv}\,(F(t,x) \cap B(\dot{\overline{x}}(t), k(t) + \beta N)).$$

Let further $Y(t,p) = \{y \in \operatorname{conv} F(t, \overline{x}(t)) : \ \langle p, y\rangle = \langle p, \dot{\overline{x}}(t)\rangle\}$

and $P(t) = N(\operatorname{conv} F(t, \overline{x}(t)), \dot{\overline{x}}(t))$.

(H_8) for almost every t the inclusion $Y(t,p) \subset k(t)B$ holds whenever $p \in P(t)$.
Set $H(t,x,p) = \sup\{\langle p, y\rangle : \ y \in F(t,x)\}$ and let $x(\cdot)$ be a feasible trajectory in the problem. We say that $x(\cdot)$ is an *H-normal trajectory* if there is no $p(\cdot) \neq 0$ such that

$$(p(0), -p(T)) \in N(S, (x(0), x(T))); \quad (-\dot{p}(t), \dot{x}(t)) \in \partial_c H(t, x(t), p(t))$$

Theorem 2. *Assume* (H_1)*-*(H_4) *and* (H_6)*-* (H_8)*. Let* $\overline{x}(\cdot)$ *be a local minimum in* (**OC**) *in the topology of uniform convergence. Then there are* $\lambda \geq 0$, $\mathbf{R}^n$*-valued function* $p(\cdot)$ *of bounded variation a nonnegative measure* $\nu \in \lambda \partial \varphi(\overline{x}(\cdot))$ *such that* $\lambda + |p(t)| \neq 0$ *and the relations (i)–(iii) below are satisfied with some* $\mathbf{R}^n$*-valued summable* $q(t)$*:*

$$(i) \quad p(t) + \int_t^T (q(s)ds + \nu(ds)) = \text{const},$$

$$(ii) \quad (p(0), -p(T) - \nu(\{T\}) \in N(S, (x(0), x(T))),$$

$$(iii) \ (-q(t), \dot{\overline{x}}(t)) \in \partial_c H(t, \cdot, \cdot)(\overline{x}(t), p(t)) \quad \text{a.e.}$$

Moreover, if $\overline{x}(\cdot)$ *is* H*-normal, then* $\lambda > 0$ *and (iii) can be replaced by the "partially convexified" Hamiltonian inclusion*

$$-q(t) \in \text{conv}\,\{w : \ (w, \dot{\overline{x}}(t)) \in \partial H(t, \cdot, \cdot)(\overline{x}(t), p(t)\}.$$

Comments. Developments of the Hamiltonian theory were more or less parallel to the studies associated with the Euler-Lagrange approach. A version of Theorem 2 for fully Lipschitz F with convex and bounded values was obtained by Clarke in 1976 and boundedness and Lipschitz conditions were weakened in the mentioned 1994 work of Loewen and Rockafellar. The partially convexified Hamiltonian condition for problems with convex-valued F was established in papers by Rockafellar (1996) and Ioffe (1997). But the convexity assumption on the values of F was dropped only in a 2014 paper by Vinter, also for bounded-valued F.

So Theorem 2 seems to be the first result containing Hamiltonian adjoint inclusions for problems with non-convex and unbounded F.

About the proofs.

The proofs of both theorems are based on reduction of the problem to one or a sequence of generalized problems of Bolza:

$$\text{minimize} \quad \psi(x(\cdot)) + \int_0^T L(t, x(t), \dot{x}(t))dt$$

with, generally, extended real valued ψ and integrand L.

Necessary conditions for minima in such problems under very general assumptions will shortly appear in a paper to be published soon in Serdica Math. Journal (in the issue dedicated to the memory of Asen Dontchev).

Here I shall only be able to briefly describe the constructions of the Bolza functionals in the proofs of both theorems. They are actually very different.

Theorem 1. Recall that $U(t)$ is the set of $u \in F(t, \overline{x}(t))$ such that (with some positive $k,\ \varepsilon,\ \delta$) the inequality

$$|d(y, F(t,x)) - d(y, F(t,x'))| \le k|x - x'| \qquad (*)$$

holds for $x, x' \in B(\overline{x}(t), \varepsilon), y \in B(u, \delta)$.

We shall consider all possible triples $\sigma = (k, \varepsilon, \delta)$. Fix such a σ and let $V_\sigma(t)$ be the collection of $u \in F(t, \overline{x}(t))$ such that $d(y, F(t, \cdot))$ is k-Lipschitz on $B(\overline{x}(t), \varepsilon)$ if $|y - u| < \delta$.

Let further $\Delta_\sigma = \{t :\ k(t) \le k\}$ with $k(t)$ from (H_5). We define the integrand $L_\sigma(t, x, y)$ equal to $d(y, F(t,x))$ (with slight modification) either on the δ-neighborhood of $U_\sigma(t)$ if $t \in \Delta_\sigma$ or on the $\delta(t)$-neighborhood of $\overline{x}(t)$ otherwise (with $\delta(t)$ from (H_5)) and consider Bolza functionals

$$J_{m\sigma}(x(\cdot)) = \max\{\varphi(x(\cdot)) + m^{-2}, d((x(0), x(T)), S)\} + \int_0^T L_\sigma(t, x(t), \dot{x}(t))dt,$$

Clearly $J_{m\sigma}(\overline{x}(\cdot)) = m^{-2}$, so applying Ekeland's principle, we shall get a slightly modified Bolza functional that attains minimum at a certain $x_{m\delta}(\cdot) \to \overline{x}(\cdot)$ as $m \to \infty$.

Theorem 2. Here the reduction is based on the following observation:

Optimality alternative Let Y be a metric space, $M \subset Y$ and $\overline{y} \in M$. Let further $f(y)$ be a function defined and Lipschitz in a neighborhood of $\overline{y}$ and attaining at $\overline{y}$ a local minimum on M. Let finally, $\psi(y)$ be a nonnegative lower semicontinuous function equal to zero at $\overline{y}$. Then the following alternative holds:

– either there is a $\lambda > 0$ such that $\lambda f + \psi$ has an unconditional local minimum at $\overline{y}$;

– or for any sequence of positive numbers $(\eta_m) \to 0$ there are $z_m \notin M$ converging to $\overline{x}$ such that $\varphi(z_m) < \eta_m d(z_m, M)$.

In particular, if $\varphi(y) = d(y, M)$, then $f(y) + Kd(y, M) \geq f(\overline{y})$ for all y of a neighborhood of $\overline{y}$, provided K is greater than the Lipschitz constant of f.

To apply the alternative to the problem, it is convenient to assume that $\overline{x}(t) \equiv 0$, $k(t) \equiv 1$ and to work with $x(\cdot) \in W^{1,2}$. Let X be the collection of trajectories of the inclusion $\dot{x} \in F(t, x)$.

Take a sufficiently large $r > 0$ and let X_r stand for the collection of elements of X such that $|x(t)| \leq \varepsilon$ for all t and $|\dot{x}(t)| \leq r$ almost everywhere. Let finally M be the subset of X_r containing those $x(\cdot)$ which satisfy $(x(0), x(T)) \in S$. We consider X_r with the topology of uniform convergence and, applying the optimality alternative, conclude that

(a) either there is a $\lambda > 0$ such that zero is an unconditional strong local minimum of $\lambda\psi(x(\cdot)) + d((x(0), x(T)), S)$ on X_r;

(b) or there is a sequence on $(z_m(\cdot)) \subset X_r$ uniformly converging to zero and such that $d((z_m(0), z_m(T)), S) < m^{-2} d_C(z_m(\cdot), M)$.

In the first case the functional is Lipschitz and we can find a $N > 0$ such that

$$J(x(\cdot)) = \lambda\psi(x(\cdot)) + d((x(0), x(T)), S) + N \int_0^T L(t, x(t), \dot{x}(t))dt \geq J(0) =$$

if $\|x(\cdot)\|_C < \varepsilon_0$ and $\|\dot{x}(t)\| \leq r$ a.e..

So J is the desired Bolza functional in this case.

The second case is more complicated. Let $\psi_m(x(\cdot))$ be the function equal to

$$d((x(0), x(T)), S) + \int_0^T ((|\dot{x}(t)| - r)^+)^2 dt + m^{-1}\|x(\cdot) - z_m(\cdot)\|_C.$$

if $x(\cdot) \in X$ and to $+\infty$ otherwise. Then

$$\psi_m(z_m(\cdot)) \leq m^{-2} d(z_m(\cdot), M).$$

We can apply Stegall's variational principle to ψ_m, and find an $a_m \in \mathbf{R}^n$ and a $w_m(\cdot) \in L^2$ such that $|a_m| < m^{-2}$, $\|w_m(\cdot)\|_2 < m^{-2} d(z_m(\cdot), M)$ and the function

$$J_m(x(\cdot)) = \psi_m(x(\cdot)) + \langle a_m, x(0) - z_m(0)\rangle + \int_0^T \langle w_m(t), \dot{x}(t) - \dot{z}_m(t)\rangle dt$$

attains its minimum on $W^{1,2}$ at some $x_m(\cdot) \in X$.

It remains to note that J_m is a Bolza functional because setting

$$L_m(t,x,y) = \begin{cases} ((\|y\| - r)^+)^2 + \langle w_m(t), y\rangle, & \text{if } y \in F(t,x); \\ +\infty, & \text{otherwise,} \end{cases}$$

and

$$\varphi_m(x(\cdot)) = d(x(0), x(T), S) + \langle a_m, x(0) - z_m(0)\rangle + m^{-1}\|x(\cdot) - z_m(\cdot)\|_C,$$

we get

$$J_m(x(\cdot)) = \varphi_m(x(\cdot)) + \int_0^T L_m(t, x(t), \dot{x}(t))dt$$

THANK YOU