OPTIMAL CONTROL OF DIFFERENTIAL INCLUSIONS: NEW DEVELOPMENTS AND OPEN PROBLEMS [1]

Alexander Ioffe[2]

The paper is dedicated to Jack Warga on the occasion of his 80th birthday

Abstract

The note is concerned with necessary optimality conditions in optimal control problems involving differential inclusions. We sketch proofs of two theorems giving answers to longstanding questions and discuss some related questions which still remain unanswered.

1 Introduction

We shall discuss two longstanding questions in the theory of necessary optimality conditions in optimal control problems for systems governed by differential inclusions:

$$\begin{array}{ll} \textit{minimize} & l(x(0),x(1)), \\ \textit{s.t.} & \dot{x} \in F(t,x), \\ & (x(0),x(1)) \in S, \end{array} \tag{1}$$

where $F : [0,1] \times I\!\!R^n \to I\!\!R^n$ is a set-valued mapping satisfying some standard requirements to be specified later.

The first question concerns Hamiltonian necessary conditions for systems with non-convex-valued F, the second is about possible weakening of Lipschitz-type requirements on F without affecting the power of the necessary condition available. This is a certain departure from my original intention to discuss mainly the second of the mentioned problems, partly in connection with (1) augmented with mixed state-velocity constraints. But very recently Clarke [2] has proved his 25 years old conjecture (see §3) and it was tempting to look at his proof through the the prism of the methods and techniques developed in [4]. This turned out to be quite useful for, combined with some new ideas of Clarke's proof, these methods allowed to considerably improve Clarke's result and extend the scope of validity of the Hamiltonian condition (in the fully convexified Clarke's form). This result is discussed in §4.

The second result presented in the note contains a partial answer to a question stated in [4] and very much stimulated by earlier improvements of Pontriagin's maximum principle by Lojasievicz and Sussmann which thereby become a part of the subdifferential-based theory of necessary conditions. The proof of the theorem sketched in §6 is also an application of the technology tested in [4] which is based on so called "optimality alternative" formulated in §4.

In sections 5 and 7 we discuss questions that remain open while §§2 and 3 are introductory and contain notation, terminology, basic assumptions and a brief summary of published results.

2 Notation and assumptions

When dealing with vectors in $I\!\!R^n$ we denote by $(x|y)$ the usual inner product, by $\|x\|$ the Euclidean norm, by B the closed unit ball; for any metric space we denote by $B(x,\varepsilon)$ the closed ball of radius ε around x and by $d(x,Q)$ the distance from x to Q. The distance to the empty set is set to be equal to infinity.

We define the *Hamiltonian* of F by

$$\mathcal{H}(t,x,p) = \sup\{(p|u) : \ u \in F(t,x)\}.$$

By F_t we shall denote the set-valued mapping from $I\!\!R^n$ into itself defined by $F_t(x) = F(t,x)$, for any t, Gr F_t stands for the graph of F_t: Gr $F_t = \{(x,y) : \ y \in F_t(x)\}$. Likewise, $\mathcal{H}_t$ is the function of $I\!\!R^n \times I\!\!R^n$ defined by $\mathcal{H}_t(x,p) = \mathcal{H}(t,x,p)$. By ∂_p, ∂ and ∂_c we denote respectively the *proximal subdifferential*, the *limiting subdifferential* and the *generalized gradient of Clarke*. Namely, if f is an extended-real-valued function on $I\!\!R^n$ which is finite at x, then $p \in \partial_p f(x)$ if there is an $N > 0$ such that the inequality $f(x+h) - f(x) \geq (p|h) - N\|h\|^2$ holds for all h of a neighborhood of zero. The limiting subdifferential is defined as the upper Painlévé-Kuratowski limit of proximal subdifferentials, that is

[1] The research was supported by the USA-Israel Binational Science Foundation under the grant 2000157 and by the Fund for Promotion of Research at the Technion.

[2] Department of Mathematics, Technion, Haifa 32000, Israel e-mail: ioffe@math.technion.ac.il

if $|f(x)| < \infty$, then $p \in \partial f(x)$ if there is a sequence of pairs (x_n, p_n) converging to (x, p) such that $f(x_n) \to f(x)$ and $p_n \in \partial_p f(x_n)$. Finally, if f satisfies the Lipschitz condition near x, then $\partial_c f(x)$ coincides with the convex hull of $\partial f(x)$.

If $S \subset I\!R^n$, then by $I_S(x)$ we denote the *indicator* of S, that is the function equal to zero on S and infinity outside of S and by conv S the closed convex hull of S. By $N(S, x)$ we denote respectively the *normal cone to S at x*, associated with the limiting subdifferential, that is $N(x) = \partial I_S(x)$. We do not need here normal cones associated with other types of subdifferentials.

Throughout the paper, if nothing else is specified, we fix an absolutely continuous trajectory $x_*(t)$ which is admissible in (1) and assume the following about the data of (1):

($\mathbf{A}_1$) l satisfies the Lipschitz condition in a neighborhood of $(x_*(0), x_*(1))$;

($\mathbf{A}_2$) S is a closed subset of $I\!R^n \times I\!R^n$;

($\mathbf{A}_3$) for any t the graph of F_t is closed and the function $t \to d(y, F(t,x))$ is measurable for every $x,\ y$;

($\mathbf{A}_4$) there are $\varepsilon > 0,\ \beta > 0$ and a positive summable $k(t)$ such that for any $N > 0$

$$F(t,x) \bigcap B(\dot{x}_*(t), N) \subset \\ F(t,x') + (k(t) + \beta N)\|x - x'\| B,$$

if $\|x - x_*(t)\| < \varepsilon,\ \|x' - x_*(t)\| < \varepsilon$.

The first three conditions are technical and very general.They will not change. The fourth requirement is a sort of a 'reference" pseudo-Lipschitz condition. It will be a subject of change and discussion. Thanks to ($\mathbf{A}_4$), we can in principle assume that $x_*(t) \equiv 0$.

3 Hamiltonian necessary conditions: a brief overview

The following theorem was announced at the CDC94 [3]: *if x_* is a local minimum in (1) with respect to the (norm) $W^{1,1}$-topology, then there are a $\lambda \geq 0$ and a $p(\cdot) \in W^{1,1}$ such that the following four conditions are satisfied:*

(a) *nontriviality* ($\|\cdot\|_\infty$ stands for the L_∞-norm):

$$\lambda + \|p(\cdot)\|_\infty = 1;$$

(b) the *transversality condition*:

$(p(0), -p(1) \in \lambda \partial l(x_*(0), x_*(1)) + N((x_*(0), x_*(1)), S)$;

(c) the *maximum principle*:

$$(p(t)|\dot{x}_*(t)) = \mathcal{H}(t, x_*(t), p(t)) \text{ a.e.};$$

(d) the *adjoint inclusion*:

$$\dot{p}(t) \in \text{conv } \{w :\ (w, p(t)) \in N(\text{Gr } F_t(x_*(t), \dot{x}_*(t))\}.$$

(e) *If moreover, F is convex-valued than $p(\cdot)$ also satisfies the differential inclusion*

$$-\dot{p}(t) \in \text{conv } \{w :\ (w, \dot{x}_*(t)) \in \partial \mathcal{H}_t(x_*(t), p(t)). \quad (2)$$

The differential inclusion in (d) and (e) are known as the *Euler-Lagrange* and the *Hamiltonian inclusion*:

The theorem was a result of sustained efforts of a number of reseachers. We refer to [4, 6, 11] for proofs, historical comments and further references.

The necessity of the Hamiltonian inclusion (2) is a far-reaching extension of an earlier Clarke's result who proved in [1], under an assumption stronger than ($\mathbf{A}_4$), namely

($\mathbf{A}_5$) there are an $\varepsilon > 0$ and a summable $k(t)$ such that $y \in F(t,x) \Rightarrow \|y\| \leq k(t)$ and $F(t,x) \subset F(t,x') + k(t)\|x - x'\| B$ if x and x' are within ε of $x_*(t)$,

that the inclusion

$$(-\dot{p}(t), \dot{x}_*(t)) \in \partial_c \mathcal{H}_t(x_*(t), p(t)) \quad (3)$$

is a necessary condition (along with (a) – (c)) for a convex-valued F.

We observe that (2) is a better (more selective) condition than (3) even if ($\mathbf{A}_5$) holds . An example in which (3) is satisfied and (3) is not can be found in [4].

Clarke conjectured that, under ($\mathbf{A}_5$), (3) remains a necessary condition even without the convexity assumption on F. A similar problem for (2) was causiously formulated in [4] as a question.

So far the best published result was established in [4]: (2) is a necessary condition, along with (a)–(c) if, in addition to ($\mathbf{A}_1$)-($\mathbf{A}_4$), we assume that x_* is a *normal* solution, that is, if the lambdas of all pairs $(\lambda, p(\cdot))$ satisfying (a)-(d) are bounded away from zero. This condition is generic and it is not "more a posteriori" in comparison with ($\mathbf{A}_4$) as it may look at the first glance (since it can be reformulated in other terms not involving multipliers). However verification of the condition is not simple and abnormality cannot be a priori excluded anyway.

4 New developments

In [2] Clarke proves his conjecture. In the proof he followed the route already used by him many years ago but with one fundamental technical innovation: the use of Stegall's variational principle instead of the variational principle of Ekeland. It follows that (3) is likely a necessary condition for a strong minimum while all previously known first order conditions in optimal control relate to weaker types of minima (see next section for more details).

In other words, the result actually proved in [2] can be stated as follows: *Assume* $(\mathbf{A}_1)$*-*$(\mathbf{A}_3)$ *and* $(\mathbf{A}_5)$*. If* x_* *is a strong local minimum in (1), then there are* $(\lambda, p(\cdot))$ *satisfting (a)-(c) and (3).*

Below we sketch the arguments showing how the scheme developed in [4] mainly with regard to the $W^{1,1}$-minimum can be modified to cover the case of a strong minimum. In particular, this will allow us to keep $(\mathbf{A}_4)$ augmented by the following structural assumption on F:

$(\mathbf{A}_6)$ For almost every t there are a $\delta > 0$ and an $R > 0$ such that

$$\begin{aligned}\text{conv } F(t,x) \bigcap B(\dot{x}_*(t), \delta) = \\ \text{conv } (F(t,x) \bigcap B(\dot{x}_*(t), R)).\end{aligned}$$

The assumption has a simple meaning: the closed convex hulls of $F(t,x)$ are defined by bounded portions of the sets. Clearly $(\mathbf{A}_4)$, together with $(\mathbf{A}_6)$, still remains much weaker than $(\mathbf{A}_5)$.

We claim that the following theorem holds.

Theorem 1. *Assume* $(\mathbf{A}_1)$*–*$(\mathbf{A}_4)$ *and* $(\mathbf{A}_6)$*. If* x_* *is a local minimum in (1) with respect to the topology of uniform convergence in* $W^{1,1}$*, then there are* $\lambda \geq 0$ *and absolutely continuous* $p(\cdot)$ *satisfying (a)-(c) and (3).*

Sketch of the proof. We denote by Φ the collection of all trajectories $x(t)$ of F satisfying the boundary condition in (1). Set $G(t,x,y) = d(y, F(t,x))$, and for any natural N let $G_N(t,x,y) = G(t,x,y) + I_{NB}(y)$. Thanks to $(\mathbf{A}_4)$, the function $G_N(t,x,\cdot)$ is epi-Lipschitz w.r.t. x for almost every t. It is also clear that any collection of trajectories $x(\cdot)$ on $[0,1]$ with $\int G_N(t,x(t),\dot{x}(t))dt < \infty$ taking initial (or terminal) values in a bounded set is L_∞-compact and the restrictions to such set of the L_∞- and L_1-norms induce topologically equivalent metrics.

Suppose x_* is a local minimum in (1) with respect to the uniform convergence of trajectories. We may always assume the set S bounded. As was mentioned, we shall basically follow the scheme developed in [4] in the proof. The starting point is the following alternative which is a generalization of the analogous fact stated in [3], Proposition 1 (see also [4], Proposition 4).

Optimality alternative. *Let* X *be a metric space (not necessarily complete), let* $C \subset X$*, let* $l(x)$ *be a function on* X *satisfying the Lipschitz condition on* X*, and let* $\varphi(x)$ *be a non-negative extended-real-valued function equal to zero at* x_**. Suppose* l *attains its minimum on* C *at* x_**. Then*

– either (regular case) there is an $r > 0$ *such that* dist $(x, C) \leq r\varphi(x)$ *for all* x *of a neighborhood of* x_* *in which case* x_* *is an unconditional local minimum of* $l(x) + R\varphi(x)$ *for a sufficiently large* R*,*

– or (singular case) there is a sequence $(x_k) \to x_*$ *such that* $\alpha_k = d(x_k, C) > k\varphi(x_k)$ *for every* $k = 1, 2, \ldots$*. If moreover* X *is a complete metric space,* C *is a closed set and* φ *is l.s.c. then there is a sequence* $(z_k) \to x_*$*,* $z_k \notin C$ *such that for any* k *the function* $\varphi(x) + k^{-1}d(x, z_k)$ *attains a unique global minimum at* z_k*.*

To prove Theorem 1 we apply the optimality alternative with $X = \{x(\cdot) \in W^{1,1} : \|x(t) - x_*(t)\| \leq \delta, \ \forall\, t\}$ (for a sufficiently small δ) endowed with the L_1-norm, $C = \Phi \cap X$ and

$$\varphi(x(\cdot)) = d((x(0), x(1)), S) + \int_0^1 G_N(t, x(t), \dot{x}(t))dt.$$

The analysis of the regular case is very much the same as in [4] and we omit it. In the singular case we have a sequence $(x_k(\cdot))$ of trajectories uniformly converging to x_* and such that

$$\begin{aligned}\alpha_k = \text{dist } (x_k(\cdot), \Phi) > k[d((x_k(0), x_k(1)), S) \\ + \int_0^1 G_N(t, x_k(t), \dot{x}_k(t))dt]\end{aligned} \tag{4}$$

Consider the functional

$$\begin{aligned}&J_k(x(\cdot)) = d((x(0), x(1)), S) \\ &+ \textstyle\int_0^1 G_N(t, x(t), \dot{x}(t))dt + \frac{1}{\sqrt{k}} \displaystyle\int_0^1 \|x(t) - x_k(t)\|dt.\end{aligned}$$

This is a nonnegative function and

$$0 < J_k(x_k(\cdot)) \leq \frac{\alpha_k}{k} \tag{5}$$

The functional is lower semicontinuous in the $W^{1,2}$-norm and its domain is bounded, hence we can apply Stegall's variational principle: for any $\varepsilon > 0$ there are $\mu \in \mathbb{R}^n$ and $u(\cdot) \in \mathrm{L}_2$ such that $\|\mu\| < \varepsilon$, $\int_0^1 \|u(t)\|^2 dt < \varepsilon^2$ such that

$$\mathcal{J}_k(x(\cdot)) = J_k(x(\cdot)) + (\mu|x(0)) + \int_0^1 (u(t)|\dot{x}(t))dt$$

attains absolute minimum at a certain point. We take $\varepsilon < \alpha_k/4kN$, and let $z_k(\cdot)$ be the point at which $\mathcal{J}_k$ attains its minimum. We claim that $z_k \notin \Phi$.

Indeed, otherwise (5) implies, as $\mathcal{J}_k(z_k) \leq \mathcal{J}_k(x_k)$, that

$$\tfrac{1}{\sqrt{k}} \int_0^1 \|x_k(t) - zk(t)\| dt \leq$$
$$\frac{\alpha_k}{k} + (\mu | z_k(0) - x_k(0)) + \int_0^1 (u(t)|\dot{x}_k(t) - \dot{z}_k(t)) dt$$

The $I\!R^n$-norms of derivatives of z_k and x_k are not greater than N almost everywhere, and we can harmlessly assume that the norms of $z_k(0)$ and $x_k(0)$ are also smaller than N. Therefore by the choice of ε

$$\frac{1}{\sqrt{k}} \int_0^1 \|z_k(t) - x_k(t)\| dt \leq \frac{2\alpha_k}{k}. \tag{6}$$

On the other hand, if $z_k \in \Phi$, then $\|z_k - x_k\|_1 \geq \alpha_k$ by (4) and we arrive at a contradiction whenever $k > 4$.

It also follows from (6) that z_k converge uniformly to x_* (as so do x_k). This is basically the end of the "new part" of the proof. The condition $z_k \notin \Phi$ is the key fact that will allow to prove the nontriviality condition.

From this point and almost up to the end the proof follows the same route as in [4]: we first convexify the integral part of $\mathcal{J}_k$ with respect to derivatives (using standard relaxation theorems, rather than the more advanced Theorem 4 of [5], thanks to the fact that we deal with the uniform convergence and not with the $W^{1,1}$-convergence) and get that z_k minimizes also

$$\overline{\mathcal{J}}_k(x(\cdot) = \overline{J}_k(x(\cdot) + (\mu | x(0)) + \int_0^1 (u(t)|\dot{x}(t)) dt,$$

where

$$\overline{J}_k(x(\cdot)) = d((x(0), x(1)), S) + \int_0^1 \overline{G}_N(t, x(t), \dot{x}(t)) dt$$
$$+ \tfrac{1}{\sqrt{k}} \int_0^1 \|x(t) - z_k(t)\| dt + (\mu | x(0)) + \int_0^1 (u(t)|\dot{x}(t)) dt$$

and $\overline{G}_N(t, x, y)$ is the convexification (second Fenchel conjugate) of G_N with respect to the third argument.

The next step is a variational analysis of minima in the convexified problems. Again it is very much the same as in [4, 5]. (The only modification needed is connected with the interior point requirement in Lemma 1 of [4]. But an attentive look at the proof of Theorem 5 of [5] on which the lemma is based shows that the needed property is that $z_k(1) - z_k(0)$ belong to the interior of the attainable sets in the corresponding problems which is clearly true.) The analysis leads to the conclusion that for any k there are $q_k(\cdot) \in W^{1,1}$ and $\lambda_k \geq 0$ which satisfy together with z_k the nontriviality and transversality conditions, the Weirstrass condition: for all y

$$\begin{aligned} \overline{G}_N(t, z_k(t), y) - \overline{G}_n(t, z_k(t), \dot{z}_k(0)) \\ \geq (q_k(t)| y - \dot{z}_k(t)) \end{aligned} \tag{7}$$

and the adjoint inclusion

$$\dot{q}_k(t) \in \text{conv } \{w : \ (w, q_k(t)) \in \partial \overline{G}_N(t, z_k(t), \dot{z}_k(t))\}.$$

As $\overline{G}_N$ is convex with respect to the last argument, the partial dualization theorem of [4, 8] implies that

$$-\dot{q}_k(t) \in \text{conv } \{w : \ (w, \dot{z}_k(t)) \in \partial \mathcal{H}_N(t, z_k(t), q_k(t))\},$$

where $\mathcal{H}_N(t, x, p)$ is the Fenchel conjugate of $\overline{G}_N$ with respect to the last argument. As $\mathcal{H}_N$ is Lipschitz, the latter implies

$$(-\dot{q}_k(t), \dot{z}_k(t)) \in \partial_c \mathcal{H}_N(t, z_k(t), q_k(t)) \text{ a.e.}$$

From this point to the final conclusion the proof follows the standard argument based on Mazur theorem, first when $k \to \infty$ and then as $N \to \infty$ with $(\mathbf{A}_6)$ applied at the very last stage to show that $\mathcal{H}(t, x, p) = \mathcal{H}_N(t, x, p)$ near the points of the interest so that the subdifferentials of the functions coincide at the points.

5 Open problems

The classical Pontriagin's maximum principle and the Euler-Lagrange maximum principle for problems with differential inclusions are necessary condition for the minimum of the following kind: for any $N > 0$ there is an $\varepsilon > 0$ such that x_* is not worse than any admissible trajectory with the norm of the derivative everywhere not greater than N (or corresponding to a uniformly bounded control) and which is close to x_* in the $W^{1,1}$-norm. This type of a minimum was specified by Miljutin (see [7]) as *Pontriagin minimum* (although the *Miljutin minimum* could probably be a better name).

On the other hand, as follows from the proof of Theorem 1, as well as from Clarke's proof of his theorem, (3) is a necessary condition for a local minimum of a different kind which is formulated similarly but with the $W^{1,1}$-norm replaced by the L_∞-norm (the *bounded strong minimum* in terminology of [7]).

Problem 1. *What is the relationship between the two types of minima ?*

It follows from Theorem 4 of [5] that the minima are equivalent for variational problems with finite Lagrangians under fairly general assumptions. Therefore they are equivalent for regular problems of optimal control which reduce to variational problems with finite Lagrangian through the optimality alternative. Miljutin and Osmolovskii also showed that the minima are equivalent if the solution satisfies the strong maximum principle [7]. Thus the difference of the concepts, if it exists, can only be observed in abnormal situations and in the absence of the strong maximum principle.

Problem 2. *Is the "partially convexified" Hamiltonian inclusion (2) a necessary condition for the $W^{1,1}$-minimum also in the abnormal case (for non-convex-valued F).*

This problem is certainly connected with the first although seems to be independent. The other remaining part of the Hamiltonian inclusion puzzle comes from comparison of conditions of Theorem 1 and the original Clarke's result. It seems that the condition ($\mathbf{A}_6$) cannot be completely removed for this may affect necessary Lipschitz properties of the convexified map. However the following question can be asked:

Problem 3. *Is the fully convexified Hamiltonian inclusion (3) a necessary optimality condition if ($\mathbf{A}_4$) is replaced by a stronger Lipschitz assumption:*

($\mathbf{A}_7$) there are $\varepsilon > 0$ and $k(\cdot) \in L_1$ such that

$$F_t(x) \subset F_t(x') + k(t)\|x - x'\|B$$

whenever x and x' are within ε of $x_*(t)$ (and ($\mathbf{A}_6$) is removed)?

In other words: whether the boundedness assumption on F can be excluded from Clarke's theorem?

6 Weakening the Lipschitz requirements: new developments

The second problem is connected with the Lipschitz-type requirement ($\mathbf{A}_4$). The question was stated in [4] as follows: do the conditions (a)-(e) in the beginning of Section 3 remain necessary if ($\mathbf{A}_4$) is replaced by the following weaker condition:

($\mathbf{A}_8$) there are $\varepsilon > 0$, $\delta > 0$ and $k(\cdot) \in L_1$ such that

$$\emptyset \neq F_t(x) \bigcap B(\dot{x}_*(t), \varepsilon) \subset F_t(x') + k(t)\|x - x'\|B$$

if x and x' are within δ of $x_*(t)$.

The question was motivated by recent studies by Sussmann [9, 10] who proved a version of maximum principle (other then being discussed here) under even weaker continuity requirements on F. A positive answer to the question for the case of convex-valued F was given in [4] Here is a partial answer to the question in the absence of the convexity assumption on F.

Theorem 2. *Assume ($\mathbf{A}_1$)-($\mathbf{A}_3$)and ($\mathbf{A}_8$). Assume further that F admits Lipschitz approximations in the following sense:*

($\mathbf{A}_9$) For any $N > 0$ there is a set-valued mapping F_N and a $\gamma > 0$ such that

$$F_N(t, x) \bigcap B(\dot{x}_*(t), \gamma) = F(t, x) \bigcap B(\dot{x}_*(t), \gamma),$$

F_N satisfies ($\mathbf{A}_8$) if $\|x - x_*(t)\| \leq \varepsilon$ and $\|y - \dot{x}_*(t)\| \leq N$ and

$$\int_0^1 \sup_{\substack{\|y-\dot{x}_*(t)\| \leq N \\ \|x-x_*(t)\| \leq \varepsilon}} |L_N(t, x, y) - G_N(t, x, y)| dt \to 0 \quad \text{as } N \to \infty,$$

where $L_N(t, x, y) = d(y, F_N(t, x)) + I_{NB}(y - \dot{x}_*(t))$.

If x_* is a local solution of the problem (in the norm $W^{1,1}$-topology), then there are $\lambda \geq 0$ and $p(t)$ such that conditions (a)-(d) of Section 3 are satisfied.

Sketch of the proof. We again apply the optimality alternative, this time with $X = W^{1,1}$ (with the norm topology), $C = \Phi$ and

$$\varphi(x(\cdot)) = d(x(0), x(1)), S) + \int_0^1 L_N(t, x(t), \dot{x}(t))dt.$$

The subsequent analysis basically follows the lines of the proof of the main Theorem 1 of [4] and we conclude as a result that for any N there are $(\lambda_N, p_N(\cdot))$ satisfying (a), (b) together with the adjoint inclusion

$$\dot{p}_N(t) \in \text{conv } \{w : \ (w, p_N(t)) \in \partial G_t(x_*(t), \dot{x}_*(t))$$

and the Weierstrass condition

$$L_N(t, x_*(t), y) - L_N(t, x_*(t), \dot{x}_*(t) \geq (p_N(t)|y - \dot{x}_*(t)).$$

In the adjoint inclusion we can write G instead of L_N because the functions coincide in a neighborhood of $(x(t), \dot{x}_*(t))$ and therefore the subdifferentials coincide at the reference points. From that point the limiting process is implemented without problems and the result follows.

7 Discussion and Open problems

The approximation condition ($\mathbf{A}_9$) is satisfied (after some reconstruction) if F has a dense collection of Carathéodory selections with moduli of continuity on bounded sets bounded by summable functions of t. This is in particular the case if F is the union of finitely many Carathéodory mappings. In the latter case, applying the argument similar to that used in the proof of the Pontriagin maximum principle in [4], we can easily get from Theorem 2 the Lojasievicz improvement of the maximum principle for the standard problem of optimal control (which says that only the vector field corresponding to the optimal control needs to be smooth with respect to the state variable) and Sussman's nonsmooth modification of the improvement (Theorems 8.2 and 8.4 in [9]). Thus Theorem 2 is a certain step in bridging the "mainstream" subdifferential theory of nonsmooth optimal control and the new theory being developed by Sussmann.

But there seems to be still a long way to go to explore the possibility to develop a unified theory. The theories differ both in their "philosophies" and the techniques they use which makes the results very different and difficult to compare. The subdifferential theory makes an extreme effort to preserve the constructive and intrinsic nature of the maximum principle. This means that every object that appears in statements is defined either through explicit formula or is "computable" by a certain, again explicitly produced, algorithm and no information other than explicitly given in the statement of the problem is incorporated.

In contrast, Sussmann's theory is basically an existence theory. It includes powerful and sophisticated existence results for some "good" selections of set-valued maps and, in certain publications, differentiation mechanisms in the spirit of derivate containers of Warga (see e.g. [9, 10, 12]). The selection theorems do not offer any instrument for a search of selections, so the success of their application remains to a large extent a matter of luck and skillfulness. However if a suitable selection is found it may work very efficiently. In the simple example of the control system $\dot{x} = u + x^2 \sin(1/|x|)$ and the functional $\int e^t u dt$ given in [9] non-optimality of the trivial solution is easily detected by Sussman's theory but not by any result of the subdifferential theory. On the other hand, one can imagine a mapping (even convex-valued) having better Lipschitzean characteristics than any of its selections. The subdifferential theory may work better in this case.

The explanation why Sussmann's conditions work in the above example while the subdifferential necessary conditions fail lies in the fact that the first are based on differentiation of flows rather than right-hand sides of equations or inclusions. This idea, probably common in the geometric control theory, is absolutely new in nonsmooth theory of optimal control. Incorporating it into the subdifferential theory may be an important step towards possible unification.

References

[1] F. Clarke, Extremal arcs and extended Hamiltonian systems, *Trans. Amer. Math. Soc.* **231** (1977), 349-367.

[2] F. Clarke, *Necessary Conditions in Dynamic Optimization*, manuscript 111 pages, July 2002.

[3] A. Ioffe, Nonsmoothness and nonconvexity in calculus of variations and optimal control, in *Proceedings of the 33th IEEE Conference on Decision and Control*, Springer 1994, pp. 3986-3991.

[4] A. Ioffe, Euler-Lagrange and Hamiltonian formalisms in dynamic optimization, *Trans. Amer. math. Soc.* **349** (1997), 2871-2900.

[5] A. Ioffe and R.T. Rockafellar, The Euler and Weierstrass conditions for nonsmooth variational problems, *Calculus of variations and PDEs*, **4** (1996), 58-87.

[6] P.D. Loewen and R.T. Rockafellar, New necessary condition for generalized problem of Bolza, *SIAM J. Control Optimization* **34** (1996), 1496-1511.

[7] A.A. Miljutin and N.P. Osmolovskii, *Calculus of Variations and Optimal Control*, Transl. Math. Monographs, vol. **180** AMS 1998.

[8] R.T. Rockafellar, Equivalent subgradient versions of the Hamiltonian and Euler-Lagrange equations in variational analysis, *SIAM J. Control Optimization* **34** (1996), 1300-1314.

[9] H.J. Sussmann, Geometry and optimal control, in *Mathematical Control Theory*, J. Baillieul and J.C. Willems, eds. Springer, N.Y. 1998, pp. 140-198.

[10] H.J. Sussmann, Needle variations and almost lower semicontinuous differential inclusions, to appear in *Set-Valued Analysis.*

[11] R. Vinter, *Optimal Control*, Birkhäuser 2000.

[12] J. Warga, Controllability, extremality and abnormality in nonsmooth optimal control, *J. Optim. Theory Appl.* **41** (1983), 239-260.