

EULER-LAGRANGE AND HAMILTONIAN FORMALISMS IN DYNAMIC OPTIMIZATION

ALEXANDER IOFFE

ABSTRACT. We consider dynamic optimization problems for systems governed by differential inclusions. The main focus is on the structure of and interrelations between necessary optimality conditions stated in terms of Euler–Lagrange and Hamiltonian formalisms. The principal new results are: an extension of the recently discovered form of the Euler–Weierstrass condition to nonconvex valued differential inclusions, and a new Hamiltonian condition for convex valued inclusions. In both cases additional attention was given to weakening Lipschitz type requirements on the set-valued mapping. The central role of the Euler type condition is emphasized by showing that both the new Hamiltonian condition and the most general form of the Pontriagin maximum principle for equality constrained control systems are consequences of the Euler–Weierstrass condition. An example is given demonstrating that the new Hamiltonian condition is strictly stronger than the previously known one.

1. INTRODUCTION

The model of dynamic optimization to be basically studied in this paper will includes a *differential inclusion*

$$(1.1) \quad \dot{x} \in F(t, x)$$

to represent the dynamics. In many respects, this is the most natural abstraction of a controlled dynamical system, for it directly associates the set of possible velocities $\dot{x}$ with any given state x and moment t . It is also a very universal model, for it embraces closed loop feedback control systems and others that cannot be incorporated into standard models based on differential equations with control parameters explicitly entering the right-hand sides (see e.g. [9, 14] for more details). It is finally a model in which the inherent nonsmoothness of dynamic optimization reveals itself eloquently through the fact that the *Hamiltonian* of the system

$$H(t, x, p) = \sup\{p \cdot y : y \in F(t, x)\}$$

is typically a nondifferentiable function under natural assumptions (which even assumes the value infinity if the set of velocities is unbounded).

Problems involving differential inclusions attracted attention at a very early stage of development of optimal control theory (perhaps mainly in the Russian school in

Received by the editors January 23, 1995 and, in revised form, January 17, 1996.

1991 *Mathematics Subject Classification.* Primary 49K24, 49K15; Secondary 34A60, 34H05.

Key words and phrases. Optimal control, calculus of variations, differential inclusion, Lagrangian, Hamiltonian, maximum principle, nonsmooth analysis, approximate subdifferential.

The research was supported by the USA–Israel BSF grant 90–00455, by the Fund of Promotion of Science at the Technion grant 100-954 and in later stages, by the NSF grant DMS 9404128.

the beginning—e.g. [3, 12, 2]), but only progress in nonsmooth analysis provided necessary equipment to tackle the problem and a push for intensive studies (e.g. [5, 9, 13, 14, 28, 29, 30, 31, 33, 39, 41, 42, 49, 53] and many more). As a result, a variety of necessary optimality conditions were obtained that either followed the patterns of the classical Euler–Lagrange or Hamiltonian formalisms or were based on the Pontriagin maximum principle for “parametrized” problems of optimal control. In each case two fundamental relations associated with the differential inclusion appear in the statement of the necessary condition. One of them, by which the formalisms differ, usually called the *adjoint inclusion*, is a differential inclusion defining the *adjoint trajectory* or *adjoint arc* $p(t)$, which is the basic Lagrange multiplier associated with the optimal trajectory $x_*(t)$. The other, common for all realizations of both formalisms, called the *maximum principle*, states that

$$(1.2) \quad p(t) \cdot \dot{x}_*(t) = H(t, x_*(t), p(t)), \quad \text{a.e.},$$

or, equivalently, that the linear function $y \mapsto p(t)y$ attains a maximum on $F(t, x_*(t))$ at $\dot{x}_*(t)$. No unique form like this exists for adjoint inclusions in either of the formalisms, and it would not be wrong to say that the story of developments in optimal control of differential inclusions is, to a great extent, a story of the search for the right way to write the adjoint inclusions.

The following is a very brief (and undoubtedly incomplete) account of the story. We shall freely use basic notation and concepts of nonsmooth analysis, such as subdifferential and normal cone, and refer to the next section for definitions and information. To the best of our knowledge, a necessary condition for optimal control problems involving differential inclusions was considered for the first time by Boltianski in [3]. It contained a Hamiltonian type adjoint inclusion which in the subdifferential form can be written as follows:

$$(1.3) \quad -\dot{p} \in \partial_x H(t, x, p),$$

with ∂_x being the partial subdifferential with respect to x . In fact, under the assumptions of [3] the Hamiltonian function was even continuously differentiable in x . The main assumption was the existence of some smooth “local sections” of the inclusion. It was shown in [19] that Boltianski’s result actually follows from the standard Pontriagin maximum principle. There were a number of subsequent results (e.g. [53]) where a condition like (1.3) was obtained each time under the additional assumption of the existence of families of some nice (say, smooth) selections of $F(t, \cdot)$ that typically cannot be verified. A recent example given by Zhu [53] shows that without such assumptions (1.3) may fail to be a necessary condition at all (even for convex valued F).

Analogy with the subdifferential analysis carried out by Rockafellar [44] for variational problems with convex Lagrangians suggests, however, another form for the Hamiltonian condition, namely

$$(1.4) \quad (-\dot{p}, \dot{x}) \in \partial H(t, x, p).$$

In the context of differential inclusions such a condition, with ∂ being the generalized gradient, was first introduced by Clarke in his 1976 paper [5], which marked a breakthrough of nonsmooth analysis into dynamic optimization, under the assumption that F has closed bounded values and satisfies the Lipschitz condition in x (in the sense that the Hausdorff distance between the values corresponding to two different x is majorized by the norm of the difference of the arguments times

a constant). Based on Clarke's result, Loewen and Rockafellar [30, 31] extended (1.4), by means of a sophisticated technique, to inclusions with unbounded values satisfying weaker Lipschitz-like conditions.

Clarke's proof was valid under the additional assumption that F is also *convex-valued* (see also [9]). This assumption turned out to be central also in [30, 31] and many other studies (containing "intrinsic" results in the sense that they do not use any assumption about existence of additional objects such as selections or approximation—see e.g. [13, 14]) unless a strong regularity assumption (typically, at least one free end point) is imposed. The question of whether or not the adjoint inclusion of one or another type and the maximum principle are jointly necessary for optimality without the convexity and regularity assumptions remained among the main unsolved problem in the theory.

The Euler-Lagrange adjoint inclusion was first obtained by Clarke (in the somewhat different context of a generalized Bolza problem) in the form

$$(1.5) \quad (\dot{p}, p) \in N(\text{Graph}F(t, \cdot), (x, \dot{x}))$$

(see also [42]). The next step was made by Mordukhovich [37] (see [38, 39] for the proof), who used his method of discrete approximations to show that (for F with bounded convex values satisfying the Lipschitz condition in x and continuous in t) the adjoint inclusion can be written in the form

$$(1.6) \quad (\dot{p}, \dot{x}) \in \text{conv}\{(u, v) : p \cdot v = H(t, x, p), (u, p) \in N(\text{Graph}F(t, \cdot), (x, v))\}.$$

Essential for understanding the difference is that (1.5) contains Clarke's normal cone (which is a convex set), while (1.6) contains a (typically smaller) nonconvex "approximate" normal cone earlier introduced by Mordukhovich [36] so that the set of possible $(\dot{p}, p)$ defined by (1.6) can even be nonconvex, contrary to the then common belief in the necessity of "full convexification" in the adjoint inclusion. Finally, Smirnov [49] (under the same conditions) and Loewen and Rockafellar [31] (also for convex valued inclusions but with other assumptions substantially weakened) established the adjoint inclusion in the form

$$(1.7) \quad \dot{p} \in \text{conv}\{w : (w, p) \in N(\text{Graph}F(t, \cdot), (x, \dot{x}))\}.$$

As in (1.6), the approximate subdifferential enters into (1.7), so that (1.6) and (1.7) coincide when the values of F are strictly convex (see e.g. [39]), but otherwise (1.7) is strictly stronger than either of (1.5), (1.6).

Again, it is to be observed that in all three cases the convexity assumption on F was really central in the absence of a strong regularity condition (as, say, a free end point or the "calmness" condition [9], which may be difficult to verify). Very recently Kaśkosz and Łojasiewicz [29] obtained the Euler-Lagrange inclusion in Clarke's form (1.5) as a necessary condition without assuming F convex valued (just bounded-valued satisfying the Lipschitz condition), and then Mordukhovich [41] succeeded in obtaining (1.7) for the same setting. Both times, however, the necessity of the maximum principle (1.2) jointly with the adjoint inclusion was not established.

The question about the relationship between the Hamiltonian and the Euler-Lagrange formalisms has been actively discussed lately. In the classical calculus of variations the Euler equation and the Hamiltonian system are equivalent, provided

the integrand is strongly convex with respect to the velocity argument. As a partial generalization of this fact, Loewen and Rockafellar showed in [31] that for an optimal control problem involving a convex valued inclusion (1.1) the existence of an adjoint arc satisfying simultaneously (1.2), (1.4) and (1.7) can be stated as a necessary condition for a minimum, provided a certain Lipschitz-like assumption on F is satisfied. It was also shown by Rockafellar [47] that (1.4) implies (1.6) if the Hamiltonian is finite valued and satisfies the Lipschitz condition, and the converse follows from another partial dualization result announced by Ioffe [19], which is actually an easy corollary of Theorem 3 of this paper. Thus, the Hamiltonian inclusion (1.4) with Clarke's generalized gradients and the Euler-Lagrange inclusion (1.6) with approximate subdifferentials turn out to be equivalent, at least for inclusions with convex bounded values satisfying the Lipschitz condition. The situation with nonconvex valued inclusions is probably more complicated. Examples given by Kaśkosz and Łojasiewicz [29] suggest that no equivalence of the above sort exists without the convexity assumption.

Another question is the relationship between Pontriagin's maximum principle for a control system

$$(1.8) \quad \dot{x} = f(t, x, u), \quad u \in U,$$

and the known necessary conditions of either type for the corresponding differential inclusions (1.1) with $F(t, x) = f(t, x, U)$.

Kaśkosz and Łojasiewicz [28] interpreted (1.8) as a family of equations

$$(1.9) \quad \dot{x} = f(t, x), \quad f \in \mathcal{F},$$

where $\mathcal{F}$ are Carathéodory functions satisfying the Lipschitz condition with respect to x and having the "decomposition" property: if f_1 and f_2 belong to $\mathcal{F}$, then for any measurable Δ , the function f equal to f_1 for $t \in \Delta$ and to f_2 otherwise is also in $\mathcal{F}$, and obtained for it a necessary condition for optimality in the form of Pontriagin's maximum principle (actually a particular case of the one established in [22]; see [33] for an alternative and shorter proof of the result of Kaśkosz and Łojasiewicz).

Applying this result to the collection of all selections of F (convex valued and satisfying the Lipschitz condition), Kaśkosz and Łojasiewicz obtained in [28] another necessary condition for inclusion constrained problems, and Zhu [53] extended their result to nonconvex inclusions (also bounded and Lipschitz) by elaborating on a controllability theorem of Warga [52]. An obvious drawback of these conditions is the absence of any analytic mechanism for obtaining selections (even in the case of a convex valued inclusion). However, in [28] Kaśkosz and Łojasiewicz furnished an example showing that their condition can be stronger (better sorting away non-optimal arcs) than Clarke's Hamiltonian condition (1.2), (1.4). On the other hand, it is not a difficult matter to construct an opposite example of a family of vector fields for which the Kaśkosz-Łojasiewicz maximum principle is unable to recognize a nonoptimal trajectory and, say, Clarke's Hamiltonian condition can easily do this.

This paper is concerned with the three questions discussed above: (a) structures of the Euler-Lagrange and Hamiltonian necessary conditions; (b) the interrelation between them, and (c) the relation between them and Pontriagin's maximum principle. The two principal results of the paper are Theorems 1 and 2, stated in the next section. Theorem 1 establishes necessity of the Euler-Lagrange formalism for

problems with nonconvex valued F including the adjoint equation in the state-of-the-art form (1.7), the maximum principle and the transversality condition. The second theorem adds to this list of necessary conditions a Hamiltonian condition of a new type

$$(1.10) \quad \dot{p} \in \text{conv}\{w : (w, \dot{x}) \in \partial H(t, x, p)\}$$

for the case of a convex valued F . Moreover, it is shown that this inclusion follows from (1.7) no matter whether there is a minimum in the problem or not. An example is given showing that as a necessary condition, (1.10) is strictly stronger than (1.4).

The proof of Theorem 1 is a combination of regularity arguments with a subsequent application of the necessary condition for a nonconvex Bolza problem recently found by Ioffe and Rockafellar in [25] (quoted below as Theorem 3). A nice alternative proof of the condition was offered by Vinter and Zheng in a concurrent paper [51]; they also use it to get a version of Theorem 1 under stronger Lipschitz type requirements on F .

Theorem 2 is stated under different, generally much weaker, Lipschitz type requirements, so the proof of necessity of the Euler inclusion, although following the same scheme, is more technical. The second part of the theorem in which (1.10) is proved to follow from (1.7) is an easy consequence of a general formula for the subdifferential of a partially conjugate function (i.e. a function of two variables obtained as a result of the Fenchel conjugacy operation applied to one of them) established in Theorem 4 of this paper.

In this part, our study has a solid overlap with simultaneous research of Rockafellar [48] in which actual equivalence of (1.7) and (1.10) was proved under the additional assumption that F depends continuously on x . Approximation arguments have been used both in [48] and here, but the techniques were rather different: Moreau–Yosida approximation in [48] and a sort of elaboration on Clarke’s approximation techniques ([9], Section 3.2).

Rockafellar’s equivalence theorem was the basis for parallel discovery by Loewen and Rockafellar in [32] of the necessity of the Hamiltonian condition (1.7). The assumptions applied in [32] and here in Theorem 2 are different in several points (see the next section for a comparison). More substantial is the difference in the “genealogies” of the Euler condition (1.10) here and in [32]: in both cases (1.10) follows from the Euler–Lagrange inclusion (1.7), but the latter appears in [32] as a derivative of a Hamiltonian type condition whereas here it emerges as the basic fact from which all others follow.

We show further in Theorem 4 that Pontriagin’s maximum principle is a direct consequence of the Euler formalism provided by Theorem 1. Observe in this connection that the proof of the main result of [25] given by Vinter and Zheng in [51] is based on the maximum principle and the variational principle of Ekeland. This leads to the surprising conclusion that Pontriagin’s maximum principle for systems (1.8), on the one hand, and the necessary optimality condition (1.2), (1.7) for inclusion constrained problems are equivalent as mathematical theorems.

An interesting problem arises in connection with this and the very recent studies of Łojasiewicz [35] (see also Sussmann [50]) containing extensions of the above mentioned Kaśkosz–Łojasiewicz version of the maximum principle in which, in particular, the Lipschitz property is required only from the optimal vector field. These results suggest that further attempts to weaken Lipschitz-type assumptions on F

should be made to understand, in particular, whether a suitable formulation of the Euler condition is valid which would imply this result. A few unsolved problems along these lines are stated at the end of Section 2. In Theorem 6 of the last section we prove a generalization of the necessary condition of [25] for the generalized Bolza problem [25] which implies the positive answer in “regular cases”.

The problem which remains unsolved, as it was for years, is whether a Hamiltonian formalism of either kind can be derived for problems not satisfying a convexity assumption. As a corollary of the first two theorems, we get in Theorem 7 a Hamiltonian condition in the form (1.10), (1.2) but under an additional normality assumption which is an easily verifiable equivalent of a mild regularity requirement, that also becomes available thanks to the development of the Euler formalism for a generalized Bolza problem in the absence of convexity assumptions in [25] and here. But the problem in general remains open.

The plan of the paper is the following. In the next section we state the problem and the main theorems. In Section 3 we describe the basic concepts of nonsmooth analysis involved in the statements and proofs. In Section 4 we prove Theorem 1. In Section 5 we establish a formula for the subdifferential of a partially conjugate function (Theorem 4) which is instrumental in the proof of Theorem 2 in Section 6. Section 7 contains the statements and the proofs of Theorems 5, 6 and 7 and an example showing that the new Hamiltonian condition (1.10) is stronger than (1.4).

ACKNOWLEDGEMENT

I am thankful to Richard Vinter and the referee for helpful remarks, and to Boris Mordukhovich for useful conversations.

2. STATEMENTS OF THE MAIN THEOREMS

The following notation will be used throughout the paper:

- $|x|$ for the Euclidean norm of x in $\mathbb{R}^n$;
- $x \cdot y$ for the inner product of x and y ;
- $\rho(x, C)$ for the (Euclidean) distance from x to C ($C \subset \mathbb{R}^n$);
- B for the unit ball in $\mathbb{R}^n$: $B = \{x \in \mathbb{R}^n : |x| \leq 1\}$;
- $B(x, \alpha)$ for the ball of radius α about x ;
- $W^{1,1}$ for the space of absolutely continuous functions $x(t)$ on $[0, 1]$ with the norm $\|x(\cdot)\|_{1,1} = |x(0)| + \int_0^1 |\dot{x}(t)| dt$.

The paper is mainly devoted to the problem **(P)** stated as follows:

$$(\mathbf{P}) \quad \left. \begin{array}{l} \text{minimize } l(x(0), x(1)), \\ \text{s.t. } \dot{x}(t) \in F(t, x(t)), \\ (x(0), x(1)) \in S. \end{array} \right\}$$

We shall fix an admissible element $x_*(\cdot) \in W^{1,1}$ of **(P)** (i.e. a solution to the differential inclusion (1.1) satisfying the boundary conditions $(x_*(0), x_*(1)) \in S$) and assume throughout the following hypotheses on the components of the problem:

- (H₁)**: $l(\cdot, \cdot)$ is Lipschitz in a neighborhood of $(x_*(0), x_*(1)) \in \mathbb{R}^{2n}$;
- (H₂)**: $S \subset \mathbb{R}^{2n}$ is a closed set;
- (H₃)**: the graph of $F(t, \cdot)$ is closed for almost every t , and F is measurable with respect to the σ -algebra $\mathcal{B} \oplus \mathcal{L}$ generated by products of Lebesgue measurable subsets of $[0, 1]$ and Borel subsets of $\mathbb{R}^n$.

There are also three “optional” hypotheses characterizing the behavior of F as a function of x :

(**H**₄): there are an $\varepsilon > 0$ and a summable nonnegative function $k(t)$ such that

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

for all x, x' with $|x - x_*(t)|, |x' - x_*(t)| \leq \varepsilon$ a.e.;

(**H**₅): there are a summable function $k(t)$ and constants $\varepsilon > 0, \beta > 0$ such that

$$(2.1) \quad F(t, x) \cap B(\dot{x}_*(t), N) \subset F(t, x') + (k(t) + \beta N)|x - x'|B \quad \text{a.e.}$$

for all x, x' with $|x - x_*(t)|, |x' - x_*(t)| \leq \varepsilon$, and all N ;

(**H**₆): there are an $\varepsilon > 0$ and a summable function $k(t)$ such that

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

for all x, x' with $|x - x_*(t)|, |x' - x_*(t)| \leq \varepsilon$ a.e.

We observe that (**H**₆) characterizes the standard Lipschitz condition with respect to the Hausdorff distance between sets, while (**H**₄) and (**H**₅) correspond respectively to the pseudo Lipschitz condition introduced by Aubin in [1] and the “integrable sub-Lipschitz” condition of Loewen–Rockafellar [31].

We observe further that the problem (**P**) covers a seemingly more general class of problems in which the functional to be minimized has the form

$$l(x(0), x(1)) + \int_0^1 L(t, x(t), \dot{x}(t))dt,$$

provided $L(t, x, y)$ satisfies the standard measurability requirements, is everywhere finite and l.s.c in (x, y) for almost every t , and satisfies a suitable type of optional Lipschitz condition for F (or $\rho(y, F(t, x))$) with respect to x . To see this it is sufficient to introduce the set valued mapping $\Phi(x) = \{(y, \alpha) : y \in F(t, x), \alpha \geq L(t, x, y)\}$ and consider the problem (**P**) with the inclusion replaced by $(\dot{x}, \dot{\mu}) \in \Phi(t, x)$ and with the additional terminal condition $\mu(0) = 0$.

Nowhere in the paper do we consider state constraints. The main reason for that is the desire to avoid technicalities, often tedious and heavy, that typically accompany any work with the state constraints. However we do not expect any substantial difficulties in incorporating such constraints in the frameworks of any of the three main theorems, and we hope to study them elsewhere in a more technique-oriented paper.

We are ready now to state the main results of the paper. In all of the statements ∂ and N stand for the *limiting* subdifferential and normal cone defined in the next section.

We say that $(\lambda, p(\cdot)) \in \mathbb{R} \times W^{1,1}$ is a *nontrivial pair* (of multipliers) if $|\lambda| + \|p(\cdot)\|_{1,1} > 0$.

Theorem 1. *We posit (**H**₁) – (**H**₃) and (**H**₅). If $x_*(\cdot)$ is a local minimum in (**P**) in the $W^{1,1}$ -norm topology, then there is a nontrivial pair of multipliers $(\lambda, p(\cdot))$ satisfying the Euler inclusion (1.7), the maximum principle (1.2) and the transversality condition*

$$(2.2) \quad (-p(0), p(1)) \in \lambda \partial l(x(0), x(1)) + N(S, (x(0), x(1)))$$

together with $x_(\cdot)$.*

As was mentioned in the introduction, Vinter and Zheng [51] simultaneously proved a similar result. They actually assume that F satisfies the more restrictive assumption $(\mathbf{H}_6)$ rather than $(\mathbf{H}_5)$, but it seems that their method (which does not directly address the regularity alternative and uses somewhat lengthier calculations and arguments instead) can also be used to prove Theorem 1.

Theorem 2. *Let F be a convex valued multifunction satisfying $(\mathbf{H}_1) - (\mathbf{H}_4)$. Assume further that*

- *either there is a constant K such that $\rho(\dot{x}_*(t), F(t, x)) \leq K|x - x_*(t)|$ if x is within ε of $x_*(\cdot)$;*
- *or there is a Carathéodory selection $y(t, x)$ of F satisfying the Lipschitz condition in x with constant $k(t)$ (say, the same as in $(\mathbf{H}_4)$) and such that $y(t, x_*(t)) = \dot{x}_*(t)$.*

If under these conditions $x_(\cdot)$ is a local minimum in $(\mathbf{P})$, then there is a nontrivial pair $(\lambda, p(\cdot))$ of multipliers satisfying simultaneously the Euler inclusion (1.7), the Hamiltonian inclusion (1.8), the maximum principle (1.2) and the transversality condition (2.2).*

Moreover, for a convex valued F , under $(\mathbf{H}_1) - (\mathbf{H}_4)$ alone, the Euler inclusion (1.7) implies the Hamiltonian inclusion (1.8) and the maximum principle (1.2), regardless of whether there is a local minimum in $(\mathbf{P})$ at $x_(\cdot)$.*

We obtain the result of Loewen and Rockafellar [32] if we assume in addition that the set valued mapping $F(t, \cdot)$ is continuous a.e. (in the sense that $x_k \rightarrow x$, $y \in F(t, x)$ implies the existence of $y_k \in F(t, x_k)$ converging to y) and that $k(t)$ of $(\mathbf{H}_4)$ is essentially bounded, in which case the first of the two additional optional requirements is satisfied.

Observe that the assumptions characterizing Lipschitzian behavior of F in Theorem 2, although substantially weaker than $(\mathbf{H}_5)$ in certain respects, do not cover it completely. Needless to say, of course, Theorem 1 can be applied to convex valued inclusions as well. A natural question is whether either of the theorems is valid under the “intersection” of the assumptions. This question also closely relates to the latest Łojasiewicz necessary condition mentioned in the introduction. Below we formulate some unsolved problems relating to this and other similar questions.

Problems. 1. *Is the conclusion of Theorem 1 valid for non-convex valued inclusions if we replace $(\mathbf{H}_5)$ by $(\mathbf{H}_4)$ (possibly augmented by the assumption that for almost every t the set valued mapping $F(t, \cdot)$ is continuous in the ε -neighborhood of $x_*(t)$ in the following sense: if $x_m \rightarrow x$ and $y \in F(t, x)$, then there is a sequence of y_m converging to y and such that $y_m \in F(t, x_m)$) and/or by the assumption that there is a selection of F as in the statement of Theorem 2)?*

2. *Is the conclusion of Theorem 1 valid if we replace $(\mathbf{H}_5)$ by the following “sub-Lipschitzness” assumption (cf. [46]): for any N there is a summable $k_N(t)$ such that*

$$\rho(y, F(t, x)) - \rho(y, F(t, x')) \leq k_N(t)|x - x'|$$

for all x, x', y with $|x - x_(t)|, |x' - x_*(t)| \leq \varepsilon$, $|y - \dot{x}_*(t)| \leq N$?*

3. *Is the conclusion of Theorem 2 (for convex valued F) valid under $(\mathbf{H}_4)$ alone, without additional assumptions (e.g. existence of a selection of F)?*

The second problem can be considered as a step on the way to the first. But it also has independent significance. Consider for instance the problem obtained by adding to $(\mathbf{P})$ an isoperimetric constraint, say, suppose we are interested in necessary conditions for a local minimum in $(\mathbf{P})$ but in $W^{1,p}$ ($1 \leq p < \infty$) rather

than $W^{1,1}$ (an “intermediate minimum” in the terminology of Mordukhovich [41]). In case of convex valued inclusion the corresponding necessary condition follows easily from Theorem 2. But for nonconvex inclusions Theorem 1 does not allow us to get the result for p other than 1. On the other hand, it would be an easy task if Problem 2 has a positive answer.

3. THE APPARATUS: SUBDIFFERENTIALS AND NORMAL CONES

Given a function f on $\mathbb{R}^n$, we write $u \rightarrow_f x$ to mean that $u \rightarrow x$ and $f(u) \rightarrow f(x)$. We denote by $\text{dom } f = \{x : |f(x)| < \infty\}$ the *domain* of f and by $\text{epi } f = \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : \alpha \geq f(x)\}$ the *epigraph* of f .

With any f and any $x \in \text{dom } f$ we can associate the following three *subdifferentials*:

– the *proximal* subdifferential $\partial_p f(x)$ of f at x , which is the collection of y such that

$$y \cdot h \leq f(x+h) - f(x) + O(|h|^2), \quad \forall h;$$

– the *Fréchet* subdifferential $\partial_F f(x)$ of f at x , which is the collection of y such that

$$y \cdot h \leq f(x+h) - f(x) + o(|h|), \quad \forall h;$$

– and the *Dini* subdifferential $\partial^- f(x)$ of f at x , which is the collection of y such that

$$y \cdot h \leq d^- f(x; h), \quad \forall h,$$

where $d^- f(x; \cdot)$ is the *Dini* directional derivative of f at x :

$$d^- f(x; h) = \liminf_{t \rightarrow 0, h' \rightarrow h} t^{-1}(f(x + th') - f(x)).$$

Each of the subdifferentials is a closed and convex set, and f is necessarily lower semicontinuous at x if one of them is nonempty. The subdifferentials obviously satisfy the inclusions

$$\partial_p f(x) \subset \partial_F f(x) \subset \partial^- f(x),$$

and it is not difficult to find examples when a smaller subdifferential is empty while a bigger one is not. A truly remarkable fact is that, if we consider the subdifferentials as setvalued maps, then the closures of their graphs under the f -convergence of arguments coincide, namely

$$\limsup_{u \rightarrow_f x} \partial_p f(u) = \limsup_{u \rightarrow_f x} \partial_F f(u) = \limsup_{u \rightarrow_f x} \partial^- f(u),$$

provided f is lower semicontinuous (see [21]). (Here $u \rightarrow_f x$ means that $u \rightarrow x$ and $f(u) \rightarrow f(x)$.)

We denote the common limit as $\partial f(x)$. If f is l.s.c. at x , we set $\partial f(x) = \partial \tilde{f}(x)$, where $\tilde{f}$ is the lower closure of f (i.e. the function whose epigraph is the closure of $\text{epi } f$). Finally, if f is not l.s.c. at x , we set $\partial f(x) = \emptyset$. Various names have been used for this object in the literature: *generalized derivative* [36, 38], *approximate subdifferential* [18, 20, 27] or *limiting proximal subdifferential* [10]. As a rule we shall call them just *limiting* in this article.

All four subdifferentials admit alternative geometric definitions: the interplay of analytic and geometric characterizations proved to be a powerful and convenient instrument in analysis of nonsmoothnesses. Given a set $S \subset \mathbb{R}^n$ and an $x \in S$, we define the *proximal normal cone* $N_p(S, x)$ to S at x as the collection of all y such

that x is the closest in S to $x + \lambda y$ for sufficiently small λ , and the *Fréchet* normal cone $N_F(S, x)$ to S at x as the collection of all y such that

$$y \cdot (u - x) \leq o(|u - x|), \quad \forall u \in S.$$

We also consider the *contingent* cone to S at x :

$$T(S, x) = \limsup_{\lambda \rightarrow 0} \lambda^{-1}(S - x)$$

and denote by $T^\circ(S, x)$ its polar, i.e. the collection of Y such that $y \cdot h \leq 0$ for all $h \in T(S, x)$. Then

$$N_p(S, x) \subset N_F(S, x) \subset T^\circ(S, x)$$

and, if S is closed,

$$\limsup_{u \rightarrow_S x} N_p(S, u) = \limsup_{u \rightarrow_S x} N_F(S, u) = \limsup_{u \rightarrow_S} T^\circ(S, u)$$

(where $u \rightarrow_S$ means " $u \rightarrow x$ & $u \in S$ "). As in case of subdifferentials, we denote the common limit as $N(S, x)$ and set $N(S, x) = N(\text{cl } S, x)$ if $x \in S$ ($\text{cl } S$ being the closure of S).

The two way relationship between the analytic concept of a subdifferential and the geometric concept of a normal cone is described by the two formulas below, in which $(\hat{\partial}, \hat{N})$ can be any of the four pairs (∂_p, N_p) , (∂_F, N_F) , (∂^-, T°) or (∂, N) :

$$(3.1) \quad \hat{N}(S, x) = \hat{\partial} \delta_S(x),$$

where $\delta_S(x)$ is the *indicator* of S (i.e. function equal to zero if x is in S and $+\infty$ otherwise) and

$$(3.2) \quad \hat{\partial} f(x) = \{y : (y, -1) \in \hat{N}(\text{epi } f, (x, f(x)))\}.$$

For the limiting subdifferential and normal cone we have an additional pair of characteristics when $x \in S$ and S is closed:

$$(3.3) \quad N(S, x) = \bigcup_{\lambda \geq 0} \lambda \partial \tilde{\rho}(x, S),$$

where $\tilde{\rho}$ is the distance function associated with some norm in $\mathbb{R}^N$, no matter which one, and

$$(3.4) \quad \partial \tilde{\rho}(x, S) = N(S, x) \cap \tilde{B},$$

where $\tilde{B}$ is the unit ball in the same norm. (For the other three subdifferentials these relations are trivial consequences of convexity.)

It turns out further that the convex closed hull of $N(S, x)$ always coincides with the normal cone in the sense of Clarke to S at x (we denote the latter by $\bar{N}$):

$$\bar{N}(S, x) = \text{conv cl } N(S, x),$$

and that the Clarke generalized gradient can be defined by

$$\bar{\partial} f(x) = \text{conv cl } (\partial f(x) + \partial^\infty f(x)),$$

where $\partial^\infty f(x) = \{y : (y, 0) \in N(\text{epi } f, (x, f(x)))\}$ is the *singular* subdifferential of f at x .

Systematic study of limiting subdifferentials started in the late 70s when Mordukhovich [36, 37] introduced them using the geometric approach and obtained the first optimality conditions involving such subdifferentials, and then an alternative analytic approach was used by Ioffe [18, 21] to establish the basic calculus rules.

Many remarkable analytic properties of limiting subdifferentials were gradually discovered [11, 21, 24, 27, 38, 40], etc., which made them one of the basic instruments of nonsmooth analysis, especially as far as applications to optimization are concerned. We refer to the above mentioned works for more details, and here only quote the *scalarization formula* established in [21], which will be of special use in the last section of this paper: given a mapping $S : \mathbb{R}^n \mapsto \mathbb{R}^m$ satisfying the Lipschitz condition near $x \in \mathbb{R}^n$, then (we set $y = S(x)$)

$$(3.5) \quad \partial(p \cdot S)(x) = \{u : (u, -p) \in N(\text{Graph } S, (x, y))\}.$$

4. PROOF OF THEOREM 1

4.1. **Several remarks about the distance function.** The function

$$G(t, x, y) = \rho(y, F(t, x))$$

will play a prominent role in the proof. This subsection contains several facts concerning the (x, y) -dependence of the function, so we suppress the time dependence till the end of the section.

Proposition 1 (e.g. [15]). *Let $F(x)$ be a set-valued mapping from $\mathbb{R}^m$ into $\mathbb{R}^n$ which is pseudo-Lipschitz near $(\bar{x}, \bar{y}) \in \text{Graph } F$ with a constant K . Consider the norm $|(u, v)|_K = K|u| + |v|$ in $\mathbb{R}^m \times \mathbb{R}^n$, and let ρ_K be the corresponding distance function. Then*

$$(4.1) \quad \rho_K((x, y), \text{Graph } F) = \rho(y, F(x)) = G(x, y)$$

for all (x, y) of a neighborhood of $(\bar{x}, \bar{y})$. Therefore

$$N(\text{Graph } F, (\bar{x}, \bar{y})) = \bigcup_{\lambda \geq 0} \lambda \partial \rho(\bar{y}, F(\bar{x})) = \bigcup_{\lambda \geq 0} \lambda \partial G(\bar{x}, \bar{y}).$$

Proof. We have for (x, y) sufficiently close to $(\bar{x}, \bar{y})$

$$K|x - u| + |y - v| \geq K|x - u| + \rho(y, F(u)) \geq \rho(y, F(x))$$

whenever $v \in F(u)$. This shows that $\rho_K(\text{Graph } F, (x, y)) \geq \rho(y, F(x))$ for such x and y . The reverse inequality is trivial, so the first statement is true. On the other hand, as was described in the previous section,

$$N(\text{Graph } F, (\bar{x}, \bar{y})) = \bigcup_{\lambda \geq 0} \lambda \partial \rho_K((\bar{x}, \bar{y}), \text{Graph } F)$$

and comparison with (4.1) gives the second equality.

We further observe that

$$(4.2) \quad (u, p) \in \partial G(x, y) \Rightarrow |p| \leq 1 \quad \text{and} \quad |p| = 1 \quad \text{if} \quad G(x, y) > 0.$$

Indeed, the same relations are trivially valid for, say, the Dini subdifferential, and are preserved when passing to the limit.

Proposition 2. *Set $H(x, p) = \max\{p \cdot y : y \in F(x)\}$. Then the relation $p \cdot y = H(x, p)$ is equivalent to the following condition: there is an $\alpha \in (0, 1]$ such that*

$$G(x, u) - G(x, y) - \alpha p \cdot (u - y) \geq 0, \quad \forall u \in \mathbb{R}^n.$$

Proof. Given a $Q \subset \mathbb{R}^n$ and $f(y) = \rho(y, Q)$, the Young–Fenchel transform of f is

$$f^*(v) = \sup_y (v \cdot y - f(y)) = \begin{cases} \sup_{y \in Q} v \cdot y & \text{if } |v| \leq 1; \\ \infty & \text{if } |v| > 1. \end{cases}$$

Applying this to $G(x, \cdot)$, we see that for $|p| \leq 1$ the equality $p \cdot y = H(x, p)$ is the same as

$$p \cdot y - G(x, y) = \max_u (p \cdot u - G(x, u)),$$

which implies the desired inequality with $0 < \alpha \leq \min\{1, |p(t)|^{-1}\}$ if $p \neq 0$; if $p = 0$, the result is trivial.

Proposition 3. *Assume that F has the following property: $y_* \in F(x_*)$ and there are constants $\varepsilon > 0$, $k_1 > 0$ and a nonnegative nondecreasing function $r(t)$ on $[0, \infty)$ (possibly extended-real-valued) such that*

$$F(x) \cap B(y_*, N) \subset F(x') + r(N)|x - x'|B, \quad \forall N, \quad \text{and} \quad \rho(y_*, F(x)) \leq k_1|x - x_*|,$$

provided $|x - x_|, |x' - x_*| \leq \varepsilon$. Then for any y and any $\delta \in (0, \varepsilon)$ we have*

$$\rho(y, F(x')) - \rho(y, F(x)) \leq r(2|y - y_*| + k_1\delta)|x - x'|$$

if $|x - x_|, |x' - x_*| \leq \delta$.*

Proof. Take x, x' within δ of x_* and a $y \in \mathbb{R}^n$. Let $u \in F(x)$ be the nearest to y in $F(x)$, that is, $|u - y| = \rho(y, F(x))$. We have (as $\rho(y_*, F(x)) \leq k_1\|x - x_*\| \leq k_1\delta$)

$$|u - y_*| \leq |y - y_*| + \rho(y, F(x)) \leq 2|y - y_*| + k_1\delta = N_1,$$

so that $u \in F(x') + r(N_1)|x - x'|B$. The latter implies that

$$\rho(y, F(x)) = |u - y| \geq \rho(y, F(x')) - r(N_1)|x - x'|,$$

which is the desired inequality.

4.2. Regularity. The following proposition plays a central role in the proof of necessity in Theorems 1 and 6. Let $\varphi(t, x, y)$ be a function on $[0, 1] \times \mathbb{R}^n \times \mathbb{R}^n$ which is non-negative, lower semicontinuous in (x, y) (possibly extended-real-valued) and satisfies measurability requirements sufficient to ensure that $\varphi(t, x(t), y(t))$ is measurable if $x(t)$ is continuous and $y(t)$ is measurable, and such that $\varphi(t, x_*(t), \dot{x}_*(t)) = 0$ a.e. We shall call such a function a *test function* for $x_*(\cdot)$.

Proposition 4. *We posit $(\mathbf{H}_1) - (\mathbf{H}_3)$. Let $\varphi(t, x, y)$ be a test function for $x_*(\cdot)$. If $x_*(\cdot)$ is a local minimum in $(\mathbf{P})$, then the following alternative is always valid:*

- either there is an $N > 0$ such that the function

$$M(x(\cdot)) = l(x(0), x(1)) + N \left(\int_0^1 \varphi(t, x(t), \dot{x}(t)) dt + \rho((x(0), x(1)), S) \right)$$

has an unconditional local minimum in $W^{1,1}$ at $x_(t)$;*

- or there is a sequence $\{x_k(\cdot)\}$ converging to $x_*(\cdot)$ in $W^{1,1}$ and such that for any $k = 1, 2, \dots$ the function

$$M_k(x(\cdot)) = \int_0^1 \varphi(t, x(t), \dot{x}(t)) dt + \rho((x(0), x(1)), S) + k^{-1}\|x(\cdot) - x_k(\cdot)\|_{1,1}$$

attains an unconditional local minimum in $W^{1,1}$ at $x_k(\cdot)$ and either $(x_k(0), x_k(1)) \notin S$ or $\dot{x}_k(t) \notin F(t, x_k(t))$ on a set of positive measure.

Proof. Let Φ denote the set of feasible elements in $(\mathbf{P})$, that is, of such $x(\cdot) \in W^{1,1}$ which satisfy $\dot{x}(t) \in F(t, x(t))$ a.e. and $(x(0), x(1)) \in S$. Assume first that there are $\delta > 0$, $C > 0$ such that the distance (in $W^{1,1}$) from $x(\cdot)$ to Φ is not greater than $C(\int_0^1 \varphi(t, x(t), \dot{x}(t))dt + \rho((x(0), x(1)), S))$ whenever $\|x(\cdot) - x_*(\cdot)\|_{1,1} \leq \delta$. If so, then for any $x(\cdot)$ sufficiently close to $x_*(\cdot)$ there is a $u(\cdot) \in \Phi$ such that

$$\|x(\cdot) - u(\cdot)\|_{1,1} \leq C(\int_0^1 \varphi(t, x(t), \dot{x}(t))dt + \rho((x(0), x(1)), S)).$$

On the other hand, as $u(\cdot) \in \Phi$, we have

$$l(x_*(0), x_*(1)) \leq l(u(0), u(1)) \leq l(x(0), x(1)) + 2r\|x(\cdot) - u(\cdot)\|_{1,1},$$

where r is the Lipschitz constant of l . Combining these two inequalities, we see, taking $N = 2Cr$, that $M(x(\cdot)) \geq l(u(0), u(1)) \geq M(x_*(\cdot))$, so that the first of the two claimed possibilities takes place.

If no δ and C with such properties exist, then there is a sequence of $u_k(\cdot) \in W^{1,1}$ converging to $x_*(\cdot)$ and such that

$$0 < a_k = \text{dist}(u_k(\cdot), \Phi) \leq 2k(\int_0^1 \varphi(t, u_k(t), \dot{u}_k(t))dt + \rho((u_k(0), u_k(1)), S)).$$

The function on the right is nonnegative and $a_k \rightarrow 0$. By the variational principle of Ekeland (applied to the function in the parentheses) there are $x_k(\cdot) \in W^{1,1}$ such that $\|x_k(\cdot) - u_k(\cdot)\| \leq a_k/2$ and $M_k(\cdot)$ attains a local minimum at $x_k(\cdot)$ as claimed. It remains to observe that $x_k(\cdot) \notin \Phi$, which means that either $(x_k(0), x_k(1)) \notin S$ or $\dot{x}_k(t) \notin F(t, x_k(t))$ on a set of positive measure.

Of course, this proposition is a variation on the regularity theme (cf. [15]), but it offers greater flexibility in choosing test functions. Clarke seems to have been the first to use such constructions, as early as in 1976 (see [9, 10]), without emphasis on the difference between the regular and irregular cases and with φ equal to $G(t, x, y)$; we find it extremely helpful to also use other test functions in proofs. Following tradition, we say that the problem is φ -regular at $x_*(\cdot)$ if the possibility takes place. Otherwise we say that the problem is *irregular* (with respect to φ).

4.3. Proof of the theorem. The proof is a result of joint application of Proposition 4 and the following basic fact about generalized Bolza problems with everywhere finite integrands established in [25]

Consider the following problem:

$$\text{minimize } J(x(\cdot)) = l(x(0), x(1)) + \int_0^1 L(t, x(t), \dot{x}(t))$$

with L being everywhere finite.

Theorem 3. Suppose that $x_*(\cdot)$ gives a local minimum to J in $W^{1,1}$. Assume that

- (a) $l(x, y)$ is l.s.c. in a neighborhood of $(x_*(0), x_*(1))$;
- (b) L is $\mathcal{B} \oplus \mathcal{L}$ -measurable and lower semicontinuous in (x, y) ;
- (c) for any N there are an $\varepsilon > 0$ and a summable function $k_N(t)$ such that

$$L(t, x, y) - L(t, x', y) \leq k_N(t)|x - x'| \quad \text{and} \quad L(t, x, y) \geq -k(t)$$

if $|x - x_*(t)| < \varepsilon$, $|x' - x_*(t)| < \varepsilon$, $|y - \dot{x}_*(t)| \leq N$.

Then there is a $p(\cdot) \in W^{1,1}$ such that

$$\begin{aligned} \dot{p}(t) &\in \text{conv}\{w : (w, p(t)) \in \partial L(t, x_*(t), \dot{x}_*(t))\}; \\ L(t, x_*(t), u) - L(t, x_*(t), \dot{x}_*(t)) - p(t) \cdot (u - \dot{x}_*(t)) &\geq 0, \quad \forall u; \\ (p(0), -p(1)) &\in \partial l(x(0), x(1)). \end{aligned}$$

We apply Proposition 4 with $G(t, x, y)$ as a test function. Let the problem be G -regular at $x_*(\cdot)$. Then there is a $\lambda > 0$ such that $x_*(\cdot)$ is a local minimum in $W^{1,1}$ of the function

$$(4.3) \quad J(x(\cdot)) = \lambda l(x(0), x(1)) + \rho((x(0), x(1)), S) + \int_0^1 G(t, x(t), \dot{x}(t)) dt.$$

In this case by Theorem 3 there is a $p(\cdot)$ such that

$$\begin{aligned} \dot{p}(t) &\in \text{conv}\{w : (w, p(t)) \in \partial G(t, x_*(t), \dot{x}_*(t))\}; \\ (4.4) \quad G(t, x_*(t), y) - G(t, x_*(t), \dot{x}_*(t)) - p(t) \cdot (y - \dot{x}_*(t)) &\geq 0, \quad \forall y; \\ (p(0), -p(1)) &\in \lambda \partial l(x_*(0), x_*(1)) + \partial \rho((x_*(0), x_*(1)), S). \end{aligned}$$

If the problem is not G -regular, then by Proposition 4 there is a sequence $\{x_k(\cdot)\}$ norm converging to $x_*(\cdot)$ in $W^{1,1}$ and such that each function

$$M_k(x(\cdot)) = \rho((x(0), x(1)), S) + \int_0^1 G_k(t, x(t), \dot{x}(t)) dt + \varepsilon k^{-1} |x_k(0)|$$

attains at $x_k(\cdot)$ a local minimum in $W^{1,1}$, and for any k either $(x_k(0), x_k(1))$ does not belong to S or $\dot{x}_k(t)$ does not belong to $F(t, x_k(t))$ on a set of positive measure. Here

$$G_k(t, x, y) = G(t, x, y) + k^{-1} |y - \dot{x}_k(t)|.$$

(As always, we may assume that the $\dot{x}_k(t)$ converge to $\dot{x}_*(t)$ almost everywhere.) Each of these problems also satisfies all the assumptions of Theorem 3 near its solution $x_k(\cdot)$. Indeed, only the assumption (c) needs verification; but it easily follows from $(\mathbf{H}_5)$ and Proposition 3 applied with $r(N) = k(t) + \beta N$.

This means that there are functions $p_k(\cdot)$ such that

$$\begin{aligned} \dot{p}_k(t) &\in \text{conv}\{w : (w, p_k(t)) \in \partial G(t, x_k(t), \dot{x}_k(t))\}; \\ G(t, x_k(t), y) - G(t, x_k(t), \dot{x}_k(t)) - p_k(t) \cdot (y - \dot{x}_k(t)) &\geq 0, \quad \forall y; \\ (p_k(0), -p_k(1)) &\in \partial \rho((x_k(0), x_k(1)), S) + k^{-1} B. \end{aligned}$$

and either $\max\{|p_k(0)|, |p_k(1)|\} = 1$ if $(x_k(0), x_k(1)) \notin S$ (if we take the sum norm in $\mathbb{R}^n \times \mathbb{R}^n$ to define $\rho((x(0), x(1)), S)$) or $|p_k(t)| = 1$ for all t of a set of a positive measure on which $\dot{x}_k(t) \notin F(t, x_k(t))$ (by (4.2)). As we observed above, one of these two possibilities must occur.

On the other hand, the sequence of $(p_k(0), p_k(1))$ is bounded, and also $|\dot{p}_k(t)| \leq k(t) + \beta |\dot{x}_k(t)|$ by $(\mathbf{H}_5)$. This means that $p_k(\cdot)$ form a weakly precompact sequence, so by Mazur's theorem there is a sequence of convex combinations $q_k(\cdot)$ of $p_k(\cdot)$ converging to a certain $p(\cdot)$ in the $W^{1,1}$ -norm, and, assuming that $\dot{q}_k(t) \rightarrow \dot{p}(t)$ almost everywhere, we conclude that $p(\cdot)$ satisfies (4.4) with $\lambda = 0$. The latter is obvious as far as the last two relations of (4.4) are concerned. To see that the limit function also satisfies the first, we first observe that

$$(w, p) \in \partial G(t, x_k(t), \dot{x}_k(t)) \Rightarrow |w| \leq k(t) + \beta |\dot{x}_k(t)|.$$

Together with the upper semicontinuity property of limiting subdifferentials and uniform convergence of $p_k(\cdot)$ this gives

$$\begin{aligned} & \text{conv}\{w : (w, p(t)) \in \partial G(t, x_k(t), \dot{x}_k(t))\} \\ &= \bigcap_{m=1}^{\infty} \text{conv}\{w : (w, p) \in \partial G(t, x_k(t), \dot{x}_k(t)), |p - p(t)| \leq 1/m\}, \end{aligned}$$

from which the first inclusion of (4.4) follows as $\dot{q}_k(t)$ converge to $\dot{p}(t)$ almost everywhere.

Finally, the limit function $p(\cdot)$ cannot be identically equal to zero, for $p_k(\cdot)$ converges to it uniformly and each of them is equal to one at least at one point.

Thus, in either case there is a nontrivial pair $(\lambda, p(\cdot))$ satisfying (4.4). The first of these relations implies the Euler inclusion (1.7) by Proposition 1 and (3.3), the second implies the maximum principle (1.2) by Proposition 2, and the third coincides with the transversality condition (2.1).

This completes the proof the theorem.

5. PARTIAL DUALIZATION OF SUBGRADIENT INCLUSIONS

5.1. Statement of the dualization theorem. In this section we first give a description of subgradients of the partial conjugate to a partially convex function which is the basis for derivation of the Hamiltonian necessary condition in the next section. We shall consider a function $f(x, y)$ on $\mathbb{R}^m \times \mathbb{R}^n$ and denote by $\bar{f}(x, p)$ the Fenchel conjugate of f as a function of y only:

$$\bar{f}(x, p) = \sup_y (p \cdot y - f(x, y)).$$

Let us agree to say following [32] that f is *Aubin continuous* near (x, y) if the set valued mapping $x \mapsto \text{epi } f(x, \cdot)$ is pseudo-Lipschitz near $(x, y, f(x, y))$.

Theorem 4. *Let $f(x, y)$ be a lower semicontinuous function which is convex in y , finite at $(\bar{x}, \bar{y})$ and Aubin continuous in x near $(\bar{x}, \bar{y})$. Then*

$$(\bar{u}, \bar{p}) \in \partial f(\bar{x}, \bar{y}) \Rightarrow -\bar{u} \in \text{conv}\{w : (w, \bar{y}) \in \partial f(\bar{x}, \bar{p})\} \Rightarrow (-\bar{u}, \bar{y}) \in \partial_c \bar{f}(\bar{x}, \bar{p}).$$

Remark. In [48] Rockafellar proves that the sets

$$\text{conv}\{w : (w, \bar{p}) \in \partial f(\bar{x}, \bar{y})\}$$

and

$$-\text{conv}\{w : (w, \bar{y}) \in \partial \bar{f}(\bar{x}, \bar{p})\}$$

coincide, provided that f is Aubin continuous near $(\bar{x}, \bar{y})$ and $\text{epi } f(x, \cdot)$ depends continuously on x near $\bar{x}$. Theorem 4, thus, states only a “half” of that (the first set is contained in the second), but without any requirement on the global behavior (epi-continuity).

5.2. Mollifying a convex function. Let f be a convex lower semicontinuous function on $\mathbb{R}^n$ which is finite at least at one point. Define f_ε by

$$\text{epi } f_\varepsilon = \text{epi } f + \varepsilon B,$$

where B is the unit ball in $\mathbb{R}^n \times \mathbb{R}$ associated with the norm

$$|(x, \alpha)| = \sqrt{|x|^2 + \alpha^2}.$$

Clearly,

$$(5.1) \quad f_\varepsilon(x) = \min_{|u| \leq \varepsilon} (f(x - u) - \sqrt{\varepsilon^2 - |u|^2})$$

and the minimum is attained at some u with $|u| < \varepsilon$ if x belongs to the interior of $\text{dom } f$ (which is an immediate consequence of the standard summation rule of the convex subgradient calculus).

Proposition 5. *Given $p \in \mathbb{R}^n$, set $l = \sqrt{1 + |p|^2}$. Then*

- (a) $p \in \partial f(x) \Rightarrow p \in \partial f_\varepsilon(x + \frac{\varepsilon}{l}p) \quad \& \quad f_\varepsilon(x + \frac{\varepsilon}{l}p) = f(x) - \frac{\varepsilon}{l};$
- (b) $p \in \partial f_\varepsilon(x) \Rightarrow p \in \partial f(x - \frac{\varepsilon}{l}p);$
- (c) $f_\varepsilon \in C^1$ in the interior of its domain, that is, $p \in \partial f_\varepsilon(x) \Leftrightarrow p = \nabla f_\varepsilon(x);$
- (d) $f(x) = \max_{|u| \leq \varepsilon} (f_\varepsilon(x + u) + \sqrt{\varepsilon^2 - |u|^2})$ and for each x at which f is differentiable the maximum is attained at a unique point u with $|u| < \varepsilon$.

Proof. Geometrically, the first two statements are clear as the epigraph of f_ε is the epigraph of f plus the unit ball (Euclidean). Analytic arguments follow.

(a) Denote by $\bar{u}$ the point at which $\min(-p \cdot u - \sqrt{\varepsilon^2 - |u|^2})$ is attained: $\bar{u} = \frac{\varepsilon}{l}p$. Set $y = x + \bar{u}$. Then $f(y - u) \geq f(y - \bar{u}) + p \cdot (\bar{u} - u)$ for any u , and, consequently,

$$\begin{aligned} f(y - u) - \sqrt{\varepsilon^2 - |u|^2} \\ \geq f(y - \bar{u}) + p \cdot (\bar{u} - u) - \sqrt{\varepsilon^2 - |u|^2} \geq f(y - \bar{u}) - \sqrt{\varepsilon^2 - |\bar{u}|^2}, \end{aligned}$$

which means that

$$f_\varepsilon(y) = f(x) - \sqrt{\varepsilon^2 - |\bar{u}|^2} = f(x) - \frac{\varepsilon}{l}.$$

Likewise

$$\begin{aligned} f_\varepsilon(y + v) &= \inf_u (f(y + v - u) - \sqrt{\varepsilon^2 - |u|^2}) \\ &\geq \inf_u (f(x) + p \cdot (\bar{u} + v - u) - \sqrt{\varepsilon^2 - |u|^2}) \\ &\geq f_\varepsilon(y) + p \cdot v + \inf_u (p \cdot (\bar{u} - u) + \sqrt{\varepsilon^2 - |\bar{u}|^2} - \sqrt{\varepsilon^2 - |u|^2}) \\ &= p \cdot v + f_\varepsilon(y), \end{aligned}$$

that is, $p \in \partial f_\varepsilon(y)$.

(b) If $p \in \partial f_\varepsilon(x)$ and the minimum in (5.1) is attained at a certain $\bar{u}$, then for all u, v

$$f(x + v - u) - \sqrt{\varepsilon^2 - |u|^2} - (f(x - \bar{u}) - \sqrt{\varepsilon^2 - |\bar{u}|^2}) \geq p \cdot v.$$

In particular

$$f(x + v - \bar{u}) - f(x - \bar{u}) \geq p \cdot v$$

which means that $p \in \partial f(x - \bar{u})$, so that applying (a), we obtain the desired conclusion.

(c) This is a consequence of the fact that the conjugate of f_ε , namely $f_\varepsilon^*(v) = f^*(v) + \varepsilon \sqrt{1 + |v|^2}$, is a strictly convex function.

(d) Replacing $x - u$ by y in (5.1), we see that

$$f(y) \geq f_\varepsilon(y + u) + \sqrt{\varepsilon^2 - |u|^2}, \quad \forall u.$$

On the other hand, if $p \in \partial f(y)$, then by (a)

$$f(y) = f_\varepsilon(y + \frac{\varepsilon}{l}p) - \sqrt{\varepsilon^2 - \frac{\varepsilon^2}{l^2}|p|^2},$$

which implies the formula. The uniqueness of a minimum again follows from the fact that we maximize a strictly concave function (which is $\sqrt{\varepsilon^2 - |u|^2}$).

5.3. Two useful properties of limiting subdifferentials.

Proposition 6. *Let f be an (extended-real-valued) l.s.c. function which is finite at $\bar{x}$. Let $\{f_k\}$ be a sequence of l.s.c. functions such that $\rho((x, \alpha), \text{epi } f_k) \rightarrow \rho((x, \alpha), \text{epi } f)$ for all (x, α) of a neighborhood of $(\bar{x}, f(\bar{x}))$. Then*

$$\partial f(\bar{x}) \subset \limsup_{(k, x, f_k(x)) \rightarrow (\infty, \bar{x}, f(\bar{x}))} \partial_p f_k(x).$$

Proof. Let x be sufficiently close to $\bar{x}$ and $u \in \partial_p f(x)$. This means that there is a $\lambda > 0$ such that $(x, f(x))$ is the unique nearest point in $\text{epi } f$ to $(x_\lambda, \alpha_\lambda)$, $x_\lambda = x + \lambda u$, $\alpha_\lambda = f(x) - \lambda$. Let (x_k, α_k) be any point in $\text{epi } f_k$ which is nearest to $(x_\lambda, \alpha_\lambda)$. Then $(x_k, \alpha_k) \rightarrow (x, f(x))$ (as $\rho((x_\lambda, \alpha_\lambda), \text{epi } f_k) \rightarrow \rho((x_\lambda, \alpha_\lambda), \text{epi } f)$ and $\text{epi } f$ contains no point nearest to $(x_\lambda, \alpha_\lambda)$ other than $(x, f(x))$) and $(h_k, \beta_k) = (x_\lambda - x_k, \alpha_\lambda - \alpha_k)$ is a proximal normal to $\text{epi } f_k$ at (x_k, α_k) . We have $(h_k, \beta_k) \rightarrow \lambda(u, -1)$, so that $(h_k, \alpha_k) = \lambda_k(u_k, -1)$, where $\lambda_k \rightarrow \lambda$, $u_k \rightarrow u$. As $\lambda > 0$, the vector $(u_k, -1)$ is a proximal normal to $\text{epi } f_k$ (at least for large k) at (x_k, α_k) , which can only be if $\alpha_k = f(x_k)$ and $u_k \in \partial_p f_k(x_k)$. It remains to recall that $\partial f(\bar{x})$ is the upper limit of $\partial_p f(x)$ when $x \rightarrow \bar{x}$ and $f(x) \rightarrow f(\bar{x})$.

Proposition 7. *Let $f(x, y)$ be an l.s.c. function which is convex in the second argument. If $(\bar{u}, \bar{p}) \in \hat{\partial} f(\bar{x}, \bar{y})$, where $\hat{\partial}$ stands for either proximal or Fréchet or Dini subdifferential, then $\bar{p}$ belongs to the subdifferential of $f(\bar{x}, \cdot)$ at $\bar{y}$ in the sense of convex analysis. The same is true for $\hat{\partial}$ being the limiting subdifferential if, in addition, f is continuous near $(\bar{x}, \bar{y})$. In each case f being continuous at $(\bar{x}, \bar{y})$ implies that $\bar{f}$ is lower semicontinuous at $(\bar{x}, \bar{p})$.*

Proof. The first statement is immediate from the definitions in Section 2. Let us check that the same conclusion holds if $(\bar{u}, \bar{p}) \in \partial f(\bar{x}, \bar{y})$ and f is continuous near $(\bar{x}, \bar{y})$. Indeed, in this case we have a sequence of (x_k, y_k, u_k, p_k) converging to $(\bar{x}, \bar{y}, \bar{u}, \bar{p})$ and such that $f(x_k, y_k) \rightarrow f(\bar{x}, \bar{y})$ and, say, $(u_k, p_k) \in \partial^- f(x_k, y_k)$. Therefore $f(x_k, y) - f(x_k, y_k) \geq p_k \cdot (y - y_k)$ for all y . Passing to the limit as $k \rightarrow \infty$ we get by continuity that for y sufficiently close to $\bar{y}$ the inequality $f(\bar{x}, y) - f(\bar{x}, \bar{y}) \geq p \cdot (y - \bar{y})$ holds. And this is precisely $p \in \partial_y f(x, y)$. To check the last statement we notice that if f is continuous at $(\bar{x}, \bar{y})$ and $(x_k, p_k) \rightarrow (\bar{x}, \bar{p})$, then

$$\liminf \bar{f}(x_k, p_k) \geq \lim(p_k \cdot \bar{y} - f(x_k, \bar{y})) = \bar{p} \cdot \bar{y} - f(\bar{x}, \bar{y}) = \bar{f}(\bar{x}, \bar{p}).$$

5.4. Proof of Theorem 4: the case of a Lipschitz function. We shall first prove the theorem under the additional assumption that $f(x, y)$ satisfies the Lipschitz condition with constant K near $(\bar{x}, \bar{y})$ (recall that it is convex in y by assumption). Consider the function $f_\varepsilon(x, y)$ which is the ε -mollification of f with respect to y only:

$$f_\varepsilon(x, y) = \min_{|u| \leq \varepsilon} (f(x, y - u) - \sqrt{\varepsilon^2 - |u|^2}),$$

and denote by $\bar{f}$ and $\bar{f}_\varepsilon$ the conjugates of f and f_ε with respect to y :

$$\bar{f}_\varepsilon(x, p) = \sup_y (p \cdot y - f_\varepsilon(x, y)).$$

We notice that if f satisfies the Lipschitz condition in the ball of radius r around $(\bar{x}, \bar{y})$, then f_ε satisfies the Lipschitz condition with the same constant at least in the

ball of radius $r - \varepsilon$ (if, say, we endow $\mathbb{R}^m \times \mathbb{R}^n$ with the norm $|(x, y)| = \sqrt{|x|^2 + |y|^2}$).

Fix an $\varepsilon > 0$ and an (x, y) within $r - \varepsilon$ of $(\bar{x}, \bar{y})$ such that f_ε is differentiable at the point. Let $(u, p) = \nabla f_\varepsilon(x, y)$. Then, of course, $p = \nabla_y f_\varepsilon(x, y)$, which means that

$$(5.2) \quad \bar{f}_\varepsilon(x, p) = p \cdot y - f_\varepsilon(x, y),$$

in particular, $\bar{f}_\varepsilon$ is lower semicontinuous at (x, y) . It follows that

$$\begin{aligned} d^- \bar{f}_\varepsilon((x, p); (h, q)) &= \liminf_{\substack{(h', q') \rightarrow (h, q) \\ t \searrow 0}} t^{-1} (\bar{f}_\varepsilon(x + th', p + tq') - \bar{f}_\varepsilon(x, y)) \\ &\geq \liminf_{\substack{(h', q') \rightarrow (h, q) \\ t \searrow 0}} t^{-1} ((p + tq') \cdot y - f_\varepsilon(x + th', y) - p \cdot y + f_\varepsilon(x, y)) \\ &= q \cdot y - u \cdot h, \end{aligned}$$

which means that

$$(5.3) \quad (-u, y) \in \partial^- \bar{f}_\varepsilon(x, p).$$

Denote by $S_\varepsilon(x, y)$ the collection of limits of sequences of $(u_k, p_k) = \nabla f_\varepsilon(x_k, y_k)$ corresponding to $x_k \rightarrow x$, $y_k \rightarrow y$. It follows from Proposition 5(c) and the continuity of gradients of convex function ([43], Theorem 25.7) that $p_k \rightarrow p(y) = \nabla_y f_\varepsilon(x, y)$. In other words, the elements of $S_\varepsilon(x, y)$ have the form $(u, p(y))$, with the same second component for all of them.

For any sequence of $(x_k, y_k) \rightarrow (x, y)$ with $(u_k, p_k) = \nabla_y f_\varepsilon(x_k, y_k) \rightarrow (u, p(y))$ we have by (5.2)

$$\lim \bar{f}_\varepsilon(x_k, p_k) = \lim (p_k \cdot y_k - f_\varepsilon(x_k, y_k)) = p(y) \cdot y - f_\varepsilon(x, y) = \bar{f}_\varepsilon(x, p(y)).$$

Together with (5.3) applied to (x_k, y_k, u_k, p_k) , this allows us to conclude that

$$(u, p) \in S_\varepsilon(x, y) \Rightarrow (-u, y) \in \partial \bar{f}_\varepsilon(x, p).$$

Moreover, as $|u_k| \leq K$, we can be sure that $|u| \leq K$.

By definition $\partial_c f_\varepsilon(x, y) = \text{conv } S_\varepsilon(x, y)$, so that

$$(5.4) \quad (u, p) \in \partial_c f_\varepsilon(x, y) \Rightarrow p = p(y) \text{ and } -u \in \text{conv } \{w : (w, p) \in \partial \bar{f}_\varepsilon(x, p), |w| \leq K\}.$$

We notice further that (by the standard conjugacy formula for an infimal convolution [43])

$$\bar{f}_\varepsilon(x, p) = \bar{f}(x, p) + \varepsilon \sqrt{1 + |p|^2},$$

and by the standard rules of subdifferential calculus (e.g [20, 38])

$$\partial \bar{f}_\varepsilon(x, p) \subset \partial \bar{f}(x, p) + \{0\} \times \varepsilon B,$$

or in other words,

$$(5.5) \quad (w, y) \in \partial \bar{f}_\varepsilon(x, p) \Rightarrow \exists v, |v| \leq \varepsilon \text{ s.t. } (w, y - v) \in \partial \bar{f}(x, p).$$

By Proposition 6, $(\bar{u}, \bar{p}) \in \partial f(\bar{x}, \bar{y})$ implies the existence of a sequence

$$\{(x_k, y_k, u_k, p_k, \varepsilon_k)\}$$

converging to $(\bar{x}, \bar{y}, \bar{u}, \bar{p}, 0)$ and such that

$$(u_k, p_k) \in \partial f_{\varepsilon_k}(x_k, y_k) \subset \partial_c f_{\varepsilon_k}(x_k, y_k).$$

By (5.4), (5.5), this means the existence of α_{ik} , w_{ik} , v_{ik} , $k = 1, \dots, m+1$, such that

$$\alpha_{ik} \geq 0, \quad \sum_{i=1}^{m+1} \alpha_{ik} = 1; \quad \sum_{i=1}^{m+1} \alpha_{ik} w_{ik} = -u_k |w_{ik}| \leq K; \quad |v_{ik}| \leq \varepsilon_k$$

and

$$(w_{ik}, y_k - v_{ik}) \in \partial \bar{f}(x_k, p_k).$$

We observe also that $\bar{f}(x_k, p_k) \rightarrow \bar{f}(\bar{x}, \bar{p})$. Indeed, $\liminf \bar{f}(x_k, p_k) \geq \bar{f}(\bar{x}, \bar{p})$ as $\bar{f}$ is l.s.c. at $(\bar{x}, \bar{p})$ by Proposition 7. On the other hand,

$$\bar{f}(x_k, p_k) \leq \bar{f}_{\varepsilon_k}(x_k, p_k) = p_k \cdot y_k - f_{\varepsilon_k}(x_k, y_k) \rightarrow \bar{p} \cdot \bar{y} - f(\bar{x}, \bar{y}) = \bar{f}(\bar{x}, \bar{p}).$$

Assuming (by taking subsequences, if necessary) that α_{ik} and w_{ik} converge to some α_i and w_i respectively, we get from here that

$$-\bar{u} = \sum \alpha_i w_i; \quad |w_i| \leq K; \quad (w_i, \bar{y}) \in \partial \bar{f}(\bar{x}, \bar{p}),$$

and, consequently,

$$-\bar{u} \in \text{conv} \{w : (w, \bar{y}) \in \partial \bar{f}(\bar{x}, \bar{p})\}.$$

This completes the proof of the theorem in the case of a Lipschitz function.

5.5. Completion of the proof. Consider the function

$$\varphi(x, y, \alpha) = \rho((y, \alpha), \text{epi } f(x, \cdot)).$$

It is convex in (y, α) and satisfies the Lipschitz condition in the neighborhood of $(\bar{x}, \bar{y}, \bar{\alpha})$, $\bar{\alpha} = f(\bar{x}, \bar{y})$. Applying the result of the previous subsection, we conclude that

$$(5.6) \quad (\bar{v}, \bar{q}, \bar{\beta}) \in \partial \varphi(\bar{x}, \bar{y}, \bar{\alpha}) \Rightarrow -\bar{v} \in \text{conv} \{v : (v, \bar{y}, \bar{\alpha}) \in \partial \bar{\varphi}(\bar{x}, \bar{q}, \bar{\beta})\},$$

where $\bar{\varphi}$ is the conjugate to φ with respect to (y, α) .

An easy calculation shows that $|(q, \beta)| \leq 1$ and $\beta \leq 0$ whenever $(y, \alpha) \in \text{dom } \varphi(x, \cdot)$, and if, actually, $\beta < 0$, then

$$(5.7) \quad \bar{\varphi}(x, q, \beta) = |\beta| \bar{f}(x, q/|\beta|).$$

Going further with still simple but rather tedious calculation, we find from (5.7) that

$$(5.8) \quad (v, y, \alpha) \in \partial \bar{\varphi}(x, q, \beta) \Rightarrow \left(\frac{v}{|\beta|}, y\right) \in \partial \bar{f}(x, \frac{q}{|\beta|})$$

if $\beta < 0$ and $|(q, \beta)| \leq 1$.

On the other hand, $(\bar{u}, \bar{p}) \in \partial f(\bar{x}, \bar{y})$ is equivalent to $(\bar{u}, \bar{p}, -1) \in N(\text{epi } f, (\bar{x}, \bar{y}, \bar{\alpha}))$ and (as was explained in Section 3)

$$N(\text{epi } f, (x, y, \alpha)) = \bigcup_{\lambda \geq 0} \lambda \partial \rho_K((x, y, \alpha), \text{epi } f),$$

where ρ_K is the distance function associated with the norm $|(x, y, \alpha)| = K|x| + |(y, \alpha)|$ in $\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}$ and K is the Lipschitz constant of φ in a neighborhood of $(\bar{x}, \bar{y}, \bar{\alpha})$. By Proposition 1

$$(5.9) \quad \rho_K((x, y, \alpha), \text{epi } f) = \rho((y, \alpha), \text{epi } f(x, \cdot)) = \varphi(x, y, \alpha)$$

in a neighborhood of $(\bar{x}, \bar{y}, \bar{\alpha})$.

The final circumstance we have to take into account is that by (3.4) for any set S and any $x \in S$ we have the inclusion $\lambda \partial \rho(x, S) \subset \partial \rho(x, S)$ whenever $0 \leq \lambda \leq 1$, so that we always can find a positive η such that

$$(5.10) \quad \eta(\bar{u}, \bar{p}, -1) \in \partial \varphi(\bar{x}, \bar{y}, \bar{\alpha}) \quad \text{and} \quad \eta|(\bar{p}, -1)| < 1.$$

Together with (5.6), (5.8) this gives

$$\begin{aligned} -\eta \bar{u} &\in \text{conv}\{v : (v, \bar{y}, \bar{\alpha}) \in \partial \bar{\varphi}(\bar{x}, \eta \bar{p}, -\eta)\} \\ &= \text{conv}\{v : (\frac{v}{\eta}, \bar{y}) \in \partial \bar{f}(\bar{x}, \bar{p})\}, \end{aligned}$$

from which we get the desired inclusion by setting $v = \eta w$. This completes the proof of the theorem.

6. CONVEX VALUED INCLUSIONS. PROOF OF THEOREM 2

In this section we prove Theorem 2 for convex valued F . The statement of Theorem 2 actually consists of two independent parts. The first says that the Euler inclusion (1.7) and the transversality condition (2.2) are necessary for optimality in $(\mathbf{P})$ under certain assumptions; the second states that, in a more general situation, the Hamiltonian condition (1.10) and the maximum principle (1.2) follow from the Euler inclusion, no matter whether the corresponding trajectory is optimal or not. The proofs of the statements are also separate and independent. We start with the second, which is an easy consequence of Theorem 4.

6.1. Hamiltonian condition and the maximum principle. Suppose now that F satisfies $(\mathbf{H}_1) - (\mathbf{H}_4)$ and the Euler inclusion is satisfied for $x_*(\cdot)$ and some $p(\cdot)$. If $(w, p(t)) \in N(\text{Graph } F, (x_*(t), \dot{x}_*(t)))$, then by Proposition 1 there is a $\lambda > 0$ such that $(w, p(t)) \in \lambda \partial G(t, x_*(t), \dot{x}_*(t))$. As G is continuous (even Lipschitz continuous) near $(x_*(t), \dot{x}_*(t))$, Proposition 6 guarantees that $p(t)$ belongs to the subdifferential of $G(t, x_*(t), \cdot)$ in the sense of convex analysis, which is a property equivalent to the Weierstrass condition, and it remains to refer to Proposition 2 to get the maximum principle.

Furthermore, $|p(t)| \leq \lambda$ by (3.4). We now recall (see the proof of Proposition 2) that the Fenchel conjugate of $\lambda G(t, x, \cdot)$ is $H(t, x, p)$ if $|p| \leq \lambda$, so applying Theorem 5, we conclude that $(w, \dot{x}_*(t)) \in \partial H(t, x_*(t), p(t))$.

6.2. Necessary condition for generalized Bolza problems with constraints on derivatives. To prove the first part of Theorem 2 we shall use test functions obtained by restricting $G(t, x, y)$ to bounded sets of y 's. Theorem 3 cannot be applied in this case in the way it was used in the proof of Theorem 1. Instead, we need its modification, whose proof is actually similar and can be easily recovered from the proof of Theorem 1 in [25].

Let $R(t)$ be a set valued mapping from $[0, 1]$ into $\mathbb{R}^n$. Consider the problem of minimizing the functional

$$J(x(\cdot)) = l(x(0), x(1)) + \int_0^1 L(t, x(t), \dot{x}(t)) dt$$

over all $x(\cdot) \in W^{1,1}$ such that $\dot{x}(t) \in R(t)$ almost everywhere. Let $x_*(\cdot)$ be a solution to the problem. We assume the following:

(A₁): $l(x, y)$ satisfies the Lipschitz condition near $(x(0), x(1))$;

(A₂): $R(t)$ is a measurable set valued mapping with closed convex values having nonempty interiors;

(A₃): $L(t, x, y)$ is extended-real-valued and $\mathcal{B} \otimes \mathcal{L}$ -measurable, and there are an $\varepsilon > 0$ and a summable function $k(t)$ on $[0, 1]$ such that $L(t, \cdot, \cdot)$ satisfies the Lipschitz condition with constant $k(t)$ in a neighborhood of every point (x, y) with $|x - x_*(t)| < \varepsilon, y \in R(t)$.

Lemma 1. *If under the assumptions (A₁) – (A₃), $x_*(\cdot)$ is a local minimum of $J(\cdot)$ in $W^{1,1}$ and the measure of the set $\{t : \dot{x}_*(t) \in \text{int } R(t)\}$ is positive, then there is a $p(\cdot) \in W^{1,1}$ such that*

$$(6.1) \quad \begin{aligned} \dot{p}(t) &\in \text{conv}\{w : (w, p(t)) \in \partial L(t, x_*(t), \dot{x}_*(t)) + \{0\} \times N(R(t), \dot{x}_*(t))\}; \\ L(t, x_*(t), u) - L(t, x_*(t), \dot{x}_*(t)) - p(t) \cdot (u - \dot{x}_*(t)) &\geq 0, \quad \forall u \in R(t); \\ (p(0), -p(1)) &\in \partial l(x(0), x(1)). \end{aligned}$$

6.3. Euler inclusion and the transversality condition. We are ready to prove the first part of the theorem. Actually, we shall prove a slightly more general fact.

Claim. *Suppose that there are a measurable set $\Delta \subset [0, 1]$, a summable function $k(t)$ on $[0, 1]$ and constants $\varepsilon > 0$, $K > 0$ such that for x within ε of $x_*(t)$, F satisfies (H₄) and the distance from $\dot{x}_*(t)$ to $F(t, x)$ is majorized by $K|x - x_*(t)|$ if $t \in \Delta$, and F satisfies the Lipschitz condition (H₆) with the same $k(t)$, for $t \in [0, 1] \setminus \Delta$. If under these assumptions $x_*(\cdot)$ is a local minimum in (P), then the Euler inclusion (1.7) and the transversality condition (2.2) are satisfied by a nontrivial pair $(\lambda, p(\cdot))$.*

Assume that we have already proved the claim. Then to prove that the Euler inclusion and the transversality condition hold under the assumptions of the first part of Theorem 2 when the distance from $\dot{x}_*(t)$ to $F(t, x)$ is majorized by $K|x - x_*(t)|$, we need to just set $\Delta = [0, 1]$.

When there is a Carathéodory selection $y(t, x)$ of F satisfying the Lipschitz condition with the constant $k(t)$ (we may assume this constant coincides with that of (H₄)) and such that $y(t, x_*(t)) = \dot{x}_*(t)$, for any N we set $\Delta_N = \{t : k(t) \leq N\}$ and consider the set valued mapping

$$F_N(t, x) = \begin{cases} F(t, x), & \text{if } t \in \Delta_N, \\ \{y(t, x)\}, & \text{if } t \notin \Delta_N. \end{cases}$$

Then, of course, $x_*(\cdot)$ is a local minimum in the problem (P_N) obtained if we replace F by F_N . Clearly, this problem satisfies the conditions specified in the claim. Applying the claim to this latter problem, we conclude that there is a nontrivial pair $(\lambda, p(\cdot))$, say, such that

$$(6.2) \quad \lambda + \max_{t \in [0, 1]} |p(t)| = 1,$$

satisfying the Euler inclusion on Δ_N and, outside of Δ_N , the inclusion

$$\dot{p}(t) \in \partial(p(t) \cdot y(t, \cdot))(x_*(t))$$

(following from the scalarization formula (3.5)). It follows that for $t \notin \Delta_N$

$$(6.3) \quad |\dot{p}(t)| \leq k(t)|p(t)|.$$

The same is true for all other t as well, since in view of Proposition 1, $(\mathbf{H}_4)$ implies that

$$(6.4) \quad (w, p) \in N(\text{Graph} F, (x_*(t), \dot{x}_*(t))) \Rightarrow |w| \leq k(t)|p(t)|$$

Denote by Λ_N the collection of pairs $(\lambda, p(\cdot))$ satisfying the Euler inclusion (1.7) on Δ_N , the transversality condition (2.2) along with the normalization condition (6.2) and the inequality (6.3) for almost all t . As we have seen, $\Lambda_N \neq \emptyset$ for all N . On the other hand, (6.2) and (6.3) imply that every Λ_N is weak compact. The same argument as at the end of the proof of Theorem 1 based on the application of Mazur's theorem along with (6.4) and upper semicontinuity of the mapping $x \mapsto N(S, x)$ (see e.g. [20]) shows that every Λ_N is actually weak closed, hence weak compact. It follows that there is a pair $(\lambda, p(\cdot))$ belonging to all Λ_N . Clearly, this pair satisfies (1.7) almost everywhere, and (2.2), as claimed by the theorem.

It remains to prove the claim. Set

$$\varphi(t, x, y) = \begin{cases} G(t, x, y), & \text{if } t \notin \Delta \text{ or } t \in \Delta \text{ and } |y - \dot{x}_*(t)| \leq \varepsilon, \\ \infty, & \text{otherwise.} \end{cases}$$

In other words, this function is the restriction of $G(t, x, \cdot)$ to

$$R(t) = \begin{cases} \mathbb{R}^n, & \text{if } t \notin \Delta, \\ B(\dot{x}_*(t), \varepsilon), & \text{if } t \in \Delta. \end{cases}$$

Clearly, we can use φ as a test function. Applying Proposition 4 in the same way as in the proof of Theorem 1, we conclude that either there is a $\lambda > 0$ such that $x_*(\cdot)$ is an unconditional local minimum of

$$J(x(\cdot)) = \lambda l(x(0), x(1)) + \rho(S, (x(0), x(1))) + \int_0^1 \varphi(t, x(t), \dot{x}(t)) dt$$

in $W^{1,1}$, or there is a sequence $\{x_k(\cdot)\} \subset W^{1,1}$ converging to $x_*(\cdot)$ in $W^{1,1}$ and such that for every k the functional

$$M_k(x(\cdot)) = \rho(S, (x(0), x(1))) + k^{-1}|x(0) - x_k(0)| + \int_0^1 \varphi_k(t, x(t), \dot{x}(t)) dt$$

attains a (finite) local minimum in $W^{1,1}$ at $x_k(\cdot)$. Here $\varphi_k(t, x, y) = \varphi(t, x, y) + k^{-1}|y - \dot{x}_k(t)|$.

Let us check that both J and M_k satisfy the condition of our lemma. Again only the condition $(\mathbf{A}_3)$ needs to be verified, namely that φ satisfies the required Lipschitz condition. According to the assumptions specified in the claim, we can write

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

for x' within ε of $x_*(t)$, and

$$r(t, N) = \begin{cases} K, & \text{if } t \in \Delta, N \leq \varepsilon, \\ \infty, & \text{if } t \in \Delta, N > \varepsilon, \\ k(t), & \text{if } t \notin \Delta. \end{cases}$$

By Proposition 3

$$\rho(y, F(x')) - \rho(y, F(x)) \leq s(t)|x' - x|$$

for, say, x' , x within $\delta = \varepsilon/3K$ of $x_*(t)$ and y within $\varepsilon/3$ of $\dot{x}_*(t)$ and $s(t)$ equal to K for $y \in \Delta$ and to $k(t)$ outside of Δ . This means that $(\mathbf{A}_3)$ is satisfied for J , and also for M_k with sufficiently large k as the $x_k(\cdot)$ converge uniformly to $x_*(\cdot)$.

As $\dot{x}_*(t) \in \text{int } R(t)$ for every t , application of the lemma in the first (regular) case immediately leads to proof of the claim. So we consider the second (irregular) case in greater detail.

By the lemma, for every k there is a $p_k(\cdot) \in W^{1,1}$ such that

$$(6.5) \quad \begin{aligned} \dot{p}_k(t) &\in \text{conv}\{w : (w, p_k(t)) \in \partial G(t, x_k(t), \dot{x}_k(t)) \\ &\quad + \{0\} \times (N(R(t), \dot{x}_*(t))) + B(0, k^{-1})\}; \\ (p_k(0), -p_k(1)) &\in \partial \rho(S, (x_k(0), x_k(1))) + B(0, k^{-1} \times \{0\}). \end{aligned}$$

It follows that there is a $q_k(\cdot)$ such that for almost all t

$$(6.6) \quad \begin{aligned} \dot{p}_k(t) &\in \text{conv}\{w : (w, q_k(t)) \in \partial G(t, x_k(t), \dot{x}_k(t))\}; \\ p_k(t) - q_k(t) &\in N(R(t), \dot{x}_k(t)) + B(0, k^{-1}). \end{aligned}$$

It follows from (6.4) that $|\dot{p}_k(t)| \leq k(t)$ a.e., which means that the sequence of $p_k(\cdot)$ is weak precompact in $W^{1,1}$. We may assume without loss of generality that it converges weakly in $W^{1,1}$, and hence uniformly, to some $p(\cdot)$. Clearly, this function satisfies the transversality condition, and the already used argument involving the Mazur theorem (as we can, as always, assume that $\dot{x}_k(t)$ converge to $\dot{x}_*(t)$ almost everywhere) readily leads to the conclusion that it also satisfies the Euler inclusion together with $x_*(\cdot)$. It remains to show that $p(\cdot)$ is not identically zero. As p_k are all continuous and converge uniformly, all we need to verify is that there is a positive constant c such that $p_k(t) \geq c$ at least at one point t for each k .

Proposition 4 states also that for each k either $(x_k(0), x_k(1))$ does not belong to S , or $\dot{x}_k(t) \notin F(t, x_k(t))$ on a set of positive measure. In the first case we have $|(p_k(0), p_k(1))| \geq 1 - k^{-1}$, so if this is the case for infinitely many indices k , then $p_k(\cdot)$ goes to one at least at one of the end points.

Now assume that $(x_k(0), x_k(1)) \in S$ for all k . In this case for any k there is at least one τ such that (6.6) holds and $\dot{x}_k(t) \notin F(t, x_k(t))$ for $t = \tau$. The latter implies that $|q_k(\tau)| = 1$. If for such τ we have $\dot{x}_k(\tau) \in \text{int } R(\tau)$, then the normal cone to $R(\tau)$ at $\dot{x}_k(\tau)$ reduces to zero and we have $|p_k(\tau)| \geq 1 - k^{-1} \geq 1/2$.

Otherwise $\tau \in \Delta$ and $|\dot{x}_k(\tau) - \dot{x}_*(\tau)| = \varepsilon$. Choose a $\delta > 0$ so small that for any y with $|y| = \varepsilon$ and any line segment meeting the ball of radius δ around zero we have $(y - v) \cdot y \geq -(\varepsilon/2)|y - v|$, where v is the point in the segment which is closest to y . Let k_0 be so big that $|x_k(t) - x_*(t)| < \delta$ for all t . This means that there is a $v \in F(\tau, x_k(\tau))$ with $|v - \dot{x}_k(\tau)| < \delta$. Let z be the closest element to $\dot{x}_*(\tau)$ of $F(\tau, x_k(\tau))$. As $G(\tau, \cdot, \cdot)$ is continuous, it follows from Proposition 7 that $q_k(\tau)$ belongs to the subdifferential (in the sense of convex analysis) of $G(\tau, x_k(\tau), \cdot)$ at $\dot{x}_k(\tau)$. This means that $q_k(\tau)$ is the unit vector proportional to $\dot{x}_k(\tau) - z$. On the other hand, as F is convex valued, the line segment joining v and z belongs to $F(\tau, x_k(\tau))$. This means that

$$q_k(\tau) \cdot (\dot{x}_k(\tau) - \dot{x}_*(\tau)) \geq -(1/2)|\dot{x}_k(\tau) - \dot{x}_*(\tau)| = -(1/2)\varepsilon.$$

Finally, the normal cone to $R(t)$ at $\dot{x}_k(\tau)$ is the ray generated by $\dot{x}_k(\tau) - \dot{x}_*(\tau)$. Therefore for any element a of the cone we have $q_k(\tau) \cdot a \geq -(1/2)|a|$, and a simple calculation gives $|q_k(\tau) + a| \geq 2/3$. It follows that $|p_k(\tau)| \geq (2/3) - k^{-1} \geq 1/2$ also in this case. This proves the claim and the theorem.

7. APPLICATIONS AND EXAMPLES

7.1. The Pontriagin maximum principle. In this subsection we show that the Pontriagin maximum principle for open loop optimal control problems is a corollary of the Euler formalism of Theorem 1. We shall prove this for the maximum principle in the form of Kařkocz and Łojasiewicz [28] (which is equivalent to an earlier result established in [22]). In fact we shall prove even a stronger result, in which the transversality condition uses smaller limiting subdifferentials instead of the generalized gradient of Clarke in [21, 28]. As a consequence, we get that the necessary optimality condition for differential inclusions stated in [28] (with the corresponding refinement of the transversality condition) also follows from Theorem 1.

Consider a family $\mathcal{F}$ of (generally non-autonomous) vector fields $f(t, x)$ defined in the tube $t \in [0, 1]$, $x \in \Gamma(t) = \{x : |x - x_*(t)| < \varepsilon\}$. We assume that the family satisfies the following decomposability condition:

(D): If $f_1 \in \mathcal{F}$ and $f_2 \in \mathcal{F}$ and Δ is a measurable subset of $[0, 1]$, then the field f coinciding with f_1 for $t \in \Delta$ and with f_2 outside of Δ also belongs to $\mathcal{F}$.

We assume further

(H₇): Any $f \in \mathcal{F}$ is a Carathéodory function, and there is a summable $k(t)$ (depending on f) for which the inequality $|f(t, x) - f(t, x')| \leq k(t)|x - x'|$ holds whenever $x, x' \in \Gamma(t)$.

Consider the following problem:

$$(\mathbf{P}_1) \quad \left. \begin{array}{l} \text{minimize } l(x(0), x(1)), \\ \text{s.t. } \dot{x} = f(t, x), \quad f \in \mathcal{F}, \\ (x(0), x(1)) \in S. \end{array} \right\}$$

Theorem 5 ([28]). *We posit (H₁), (H₂), (H₇) and (D). If $x_*(\cdot)$ is a solution of (P₁) and f_* is the corresponding vector field, then there are a $\lambda \geq 0$ and a $p(\cdot) \in W^{1,1}$ such that either $\lambda > 0$ or $p(\cdot)$ is not identically zero and the relations*

$$(7.1) \quad -\dot{p}(t) \in \bar{\partial}_x(p(t) \cdot f_*(t, x_*(t))) \quad \text{a.e.}$$

and for any other $f \in \mathcal{F}$

$$(7.2) \quad p(t) \cdot f_*(t, x_*(t)) \geq p(t) \cdot f(t, x_*(t)) \quad \text{a.e.}$$

hold along with the transversality condition (2.2).

Proof. Take a finite collection of fields $f_1, \dots, f_k$ and a $\delta > 0$ and define the set valued mapping $F(t, x) = F(\delta, f_1, \dots, f_k, t, x)$ as the union of $f_*(t, x)$ and those $f_i(t, x)$ for which $|f_i(t, x_*(t)) - f_*(t, x_*(t))| > \delta$. It is an easy matter to get from (H₇) that both (H₃) and (H₅) are satisfied for F so defined. Consider the problem (P) with this F . It follows from (D) that $x_*(\cdot)$ is a solution of the problem, so Theorem 1 can be applied and there are a λ_0 and $p(\cdot)$, not equal to zero simultaneously and satisfying (1.2), (1.7) and (2.2) with the given F . It follows from the definition of F that $\dot{x}_*(t)$ is an isolated point of every $F(t, x_*(t))$. Therefore

$$N(\text{Graph} F(t, \cdot), (x_*(t), \dot{x}_*(t))) = N(\text{Graph} f_*(t, \cdot), (x_*(t), f_*(t, x_*(t)))).$$

As $f_*(t, \cdot)$ satisfies the Lipschitz condition near $x_*(t)$, the scalarization formula (3.5) shows that

$$(w, p) \in N(\text{Graph} f_*(t, \cdot), (x_*(t), f_*(t, x_*(t))))$$

if and only if

$$-w \in \partial(p \cdot f_*(t, x_*(t))),$$

which together with (1.7) (through the fact that Clarke's generalized gradient of a Lipschitz function is the convex hull of its limiting subdifferential) immediately leads to (7.1). On the other hand, the maximum principle (1.2) implies in our case that

$$(7.3) \quad p(t) \cdot f_*(t, x_*(t)) \geq p(t) \cdot f_i(t, x_*(t)), \quad \text{if } |f_i(t, x_*(t)) - f_*(t, x_*(t))| > \delta.$$

Denote by $\Lambda(\delta, f_1, \dots, f_k)$ the collection of pairs $(\lambda_0, p(\cdot))$ satisfying (7.1), (7.3) and (2.2) along with the normalization condition

$$\lambda_0 \geq 0, \quad \lambda_0 + \max_{0 \leq t \leq 1} |p(t)| = 1.$$

Then $\Lambda(\delta, f_1, \dots, f_k)$ is a nested family of weak compact subsets of $\mathbb{R} \times W_1^1$ (for $|\dot{p}(t)| \leq k_*(t) \in L_1$ for all elements of $\Lambda(\delta, f_1, \dots, f_k)$ by $(\mathbf{H}_7)$) and the normalization is weak continuous on bounded sets. Therefore there is an element $(\lambda_0, p(\cdot))$ common to all $\Lambda(\delta, f_1, \dots, f_k)$. This completes the proof of the theorem.

7.2. A strengthening of the Euler formalism for the generalized Bolza problem. Recently Łojasiewicz extended Theorem 5 to the case when only the optimal field f_* satisfies the Lipschitz hypothesis $(\mathbf{H}_6)$ while the others are just naturally bounded Carathéodory functions (see [35, 50]). If the problem formulated at the end of Section 2 is valid, then the same arguments as in the proof of Theorem 5 will also lead to the proof of this new extension.

Here we give a simple proof of a strengthened version of Theorem 3 which implies the positive answer to the problem under a regularity assumption.

Consider a function $f(x)$ on a Banach space X , and let $Q \in X$ be a set on which $f(x)$ is bounded from above. For any $m = 1, 2, \dots$ we set

$$f_m(x) = \sup_{u \in Q} (f(u) - m\|u - x\|).$$

(In other words, F_m is the sup-convolution of f and $m\|\cdot\|$.)

Proposition 8 (cf. [16]). *The following statements are true:*

- (1) $f_m(x) \geq f(x), \quad \forall x \in Q;$
- (2) $f_m(x) = f(x)$ on Q if f satisfies on Q a Lipschitz condition with constant $\leq m$;
- (3) f_m satisfies the Lipschitz condition with constant m ;
- (4) $f_m(x)$ converges decreasingly to $f(x)$ at every $x \in Q$ at which f is u.s.c.

Proof. (1) is obvious as f_m cannot increase with m . If f satisfies on Q a Lipschitz condition with constant $k \leq m$, then $f(x) \geq f(u) - m\|u - x\|$ for $u, x \in Q$, which gives (2); (3) follows from the simple calculation

$$\begin{aligned} f_m(x) - f_m(x') &= \sup_u \inf_v (f(u) - f(v) + m(\|v - x'\| - \|u - x\|)) \\ &\leq \sup_u \inf_v (f(u) - f(v) + m\|u - v\|) + m\|x - x'\| \\ &\leq m\|x - x'\|. \end{aligned}$$

Finally, if f is u.s.c. at $x \in Q$, then, given an $\varepsilon > 0$, we can choose a $\delta > 0$ such that $f(u) \leq f(x) + \varepsilon$ if $u \in Q$ is within δ of x . Then for m so big that $f(u) \leq f(x) + m\delta$ for all $u \in Q$ we have $f_m(x) \leq f(x) + \varepsilon$.

Now we consider the same functional as in 4.3:

$$J(x(\cdot)) = l(x(0), x(1)) + \int_0^1 L(t, x(t), \dot{x}(t)) dt.$$

Our purpose is to prove the following result.

Theorem 6. *Let $x_*(\cdot)$ be a local minimum of $J(\cdot)$ in $W^{1,1}$. Assume that*

- (i) $l(u, v)$ is lower semi-continuous (extended-real-valued);
- (ii) $L(t, x, y)$ is $\mathcal{B} \otimes \mathcal{L}$ -measurable;
- (iii) for almost every t the function $L(t, \cdot, \cdot)$ is everywhere finite, lower semicontinuous in (x, y) , continuous in x for any fixed y , and for any N there is a summable function $c(t)$ such that $L(t, x_*(t), y) \geq c(t)$ a.e. if $|y - \dot{x}_*(t)| \leq N$;
- (iv) there are an $\varepsilon > 0$ and a summable function $k(t)$ on $[0, 1]$ such that

$$|L(t, x, y) - L(t, x', y)| \leq k(t)|x - x'|, \text{ if } |x, x' - x_*(t)| < \varepsilon, |y - \dot{x}_*(t)| < \varepsilon.$$

Then there is an adjoint arc $p(t)$ such that the following three conditions are satisfied: the Euler inclusion:

$$(7.4) \quad \dot{p}(t) \in \text{conv}\{w : (w, p(t)) \in \partial L(t, x_*(t), \dot{x}_*(t))\};$$

the Weierstrass condition:

$$(7.5) \quad L(t, x_*(t), u) - L(t, x_*(t), \dot{x}_*(t)) - p(t) \cdot (u - \dot{x}_*(t)) \geq 0, \quad \forall u;$$

and the transversality condition

$$(7.6) \quad (p(0), -p(1)) \in \partial l(x(0), x(1)).$$

The basic difference with Theorem 3 is that here we impose a weaker requirement on Lipschitzian (in x) behavior of L : namely we assume that the set valued mapping $\text{epi } L(t, x, \cdot)$ is pseudo-Lipschitz at points of the solution, whereas in [25] it is assumed sub-Lipschitz in the sense of Rockafellar [46].

Proof. Set

$$L_m(t, x, y) = \sup_{|u - \dot{x}_*(t)| \leq \varepsilon} (L(t, x, u) - mk(t)|u - y|)$$

and define $J_m(\cdot)$ exactly as J but with L_m instead of L under the integral.

By Proposition 7, $L_m(t, x, y) \geq L(t, x, y)$ and, by virtue of (iv), $L_m(t, x, y) = L(t, x, y)$ if (x, y) is within ε of $(x_*(t), \dot{x}_*(t))$. It follows that

$$J_m(x(\cdot)) \geq J(x(\cdot)) \geq J(x_*(\cdot)) = J_m(x_*(\cdot))$$

for all $x(\cdot)$ in a neighborhood of $x_*(\cdot)$. Furthermore, $L_m(t, \cdot, y)$ satisfies the Lipschitz condition with constant $mk(t)$, so the above quoted theorem of [25] can be applied to J_m . It follows that for any m the set Λ_m of adjoint curves $p(\cdot)$ satisfying the Euler condition (7.4). (as L_m and L coincide in a neighborhood of $(x_*(t), \dot{x}_*(t))$), the Weierstrass condition obtained from (7.5) on replacing L by L_m , and the transversality condition (7.6). Now the standard compactness argument based on (iv) (and not once used here in this paper) implies the existence of an adjoint arc common to all Λ_m . It remains to note that, as $L(t, \cdot, y)$ is continuous, $L_m(t, x, \cdot)$ converges to $L(t, x, \cdot)$ everywhere for every x almost everywhere in t , which implies that the Weierstrass condition (7.5) also holds for any adjoint arc common to all Λ_m . This completes the proof of the theorem.

If we now assume that the problem **(P)** is regular (that is, G -regular with $G(t, x, y) = \rho(y, F(t, x))$, as in the proof of Theorem 1), then applying the “regular part” of the proof of Theorem 1 with Theorem 3 replaced by Theorem 6 and taking into account that the continuity requirement in Problem 1 implies continuity of G with respect to x , we arrive at the positive answer to Problem 1 under the regularity assumption.

7.3. A Hamiltonian condition for a nonconvex inclusion. As was mentioned in the introduction, the question of whether or not a Hamiltonian type necessary optimality condition exists for nonconvex differential inclusions remains among the oldest unsolved problems in the theory of necessary conditions in optimal control.

An obvious first step in the search for such a necessary condition in an unrelaxed problem is to verify whether the solution of the problem remains a local minimum of its *relaxation* (which is the problem obtained by replacing $F(t, x)$ by its convex closure $\bar{F}(x)$). If the answer is yes, then any necessary condition in the second problem is also a necessary condition in the first and, in case of the Hamiltonian condition, this answers the question completely, provided $\bar{F}$ satisfies appropriate analytical requirements, because the problem and its relaxation have the same Hamiltonian.

Let us say that the problem **(P)** is *normal* at $x_*(\cdot)$ if for any $(\lambda, p(\cdot))$ satisfying the three relations of the Euler formalism (the Euler condition (1.7), the maximum principle (1.4) and the transversality condition (2.2)) we have $\lambda > 0$.

Proposition 9. *Suppose that the assumptions of Theorem 1 are satisfied with $(\mathbf{H}_5)$ replaced by the standard Lipschitz condition $(\mathbf{H}_6)$. Suppose further that $x_*(\cdot)$ is a solution of **(P)** and the problem is normal at $x_*(\cdot)$. Then $x_*(\cdot)$ is also a local minimum in the relaxed problem with $\bar{F}$ replacing F in (1.1).*

Proof. We first observe that under the normality assumption the problem is G -regular with $G(t, s, y) = \rho(y, F(t, x))$. Indeed, if this were not true, then arguing as in the proof of Theorem 1, we would find a sequence of $x_k(\cdot)$ converging to $x_*(\cdot)$ in $W^{1,1}$ such that the function

$$M_k(x(\cdot)) = \rho((x(0), x(1))) + \int_0^1 G(t, x(t), \dot{x}(t)) dt + k^{-1} \|x(\cdot) - x_k(\cdot)\|_{1,1}$$

attains a local minimum at $x_k(\cdot)$ and either $(x(0), x(1)) \notin S$ or $\dot{x}_k(t) \notin F(t, x_k(t))$ on a set of positive measure. By the assumption

$$|G(t, x, y) - G(t, x', y)| \leq k(t)|x - x'|, \quad \text{if } |x, x' - x_*(t)| < \varepsilon,$$

so applying Theorem 3 we conclude as in the proof of Theorem 1 that there is a nonzero $p(\cdot)$ such that (1.7), (1.2) and (2.2) are satisfied for $p(\cdot)$, $x_*(\cdot)$ and $\lambda = 0$, in contradiction to the normality assumption.

As **(P)** is G -regular, there is a constant N such that $x_*(\cdot)$ is an unconditional local minimum of the functional

$$J(x(\cdot)) = l(x(0), x(1)) + N(\rho(x(0), x(1))) + \int_0^1 G(t, x(t), \dot{x}(t)) dt.$$

Let $\bar{G}(t, x, y)$ be the convexification of G with respect to the last variable. The relaxation theorem of [25] shows that $x_*(\cdot)$ is also a local minimum of the functional $\bar{J}(\cdot)$ obtained by replacing G by $\bar{G}$ under the integral. It is easy to see, on the other

hand, that $\bar{G}(t, x, y) = \rho(y, \bar{F}(t, x))$ and therefore $x_*(\cdot)$ is a local minimum in the relaxed problem obtained from $(\mathbf{P})$ by replacing F by $\bar{F}$.

We observe further that $\bar{F}$ satisfies the Lipschitz condition of the proposition with the same ε and $k(t)$ (provided it is valid for F), which means that we can apply Theorem 2 for the relaxed problem $(\mathbf{P})$ and conclude with the following statement.

Theorem 7. *We posit $(\mathbf{H}_1) - (\mathbf{H}_3)$ and $(\mathbf{H}_6)$. Let $x_*(\cdot)$ be a solution of $(\mathbf{P})$, and let the problem be normal at $x_*(\cdot)$. Then there is a nontrivial pair $(\lambda, p(\cdot))$ satisfying the Hamiltonian condition (1.10), the maximum principle (1.2) and the transversality condition (2.2).*

7.4. Example: the new Hamiltonian condition (1.10) is strictly stronger than (1.4). Namely, we construct below a set valued mapping $F(x)$ such that

$$(7.7) \quad \text{conv}\{w : (w, \bar{y}) \in \partial H(\bar{x}, \bar{p})\} \neq \{w : (w, \bar{y}) \in \bar{\partial} H(\bar{x}, \bar{p})\}$$

(where $\bar{\partial}$ stands for the generalized gradient of Clarke) so that the set on the left can be strictly smaller.

Fix a closed set S and set

$$F(\bar{x}) = S, \quad F(x) = C(x)S,$$

where $C(x)$ for any x is a linear isomorphism of $\mathbb{R}^n$ depending continuously on x . Then

$$H(x, p) = \sup\{p \cdot y : y \in C(x)S\} = \sup\{p \cdot C(x)v : v \in S\} = \varphi(C^*(x)p),$$

where φ is the support function of S .

If S is bounded, then φ is continuous, so that if $C^*(x)$ satisfies the Lipschitz condition with respect to x , by the scalarization formula (3.5)

$$\partial H(x, p) = \bigcup_{v \in \partial \varphi(0)} \partial \psi_v(\bar{x}, \bar{p}),$$

where $\psi_v(x, p) = v \cdot C^*(x)p = C(x)v \cdot p$. As this function is linear in p , we have

$$\partial \psi_v(\bar{x}, \bar{p}) = (\partial_x \psi_v(\bar{x}, \bar{p}), C(\bar{x})v).$$

Therefore $(w, 0) \in \partial \psi_v(\bar{x}, \bar{p})$ implies $v = 0$ and, consequently, $w = 0$.

Thus to construct an example satisfying (7.7), we need to find a set S and a continuous family $C(x)$ of linear isomorphisms of $\mathbb{R}^n$ such that $0 \in S$ and $\bar{\partial} H(\bar{x}, \bar{p}) = \text{conv } \partial H(\bar{x}, \bar{p})$ contains a point $(w, 0)$ with $w \neq 0$.

For example, let $n = 2$ and $S = \{(y_1, y_2) : |y_1| \leq 1, y_2 = 0\}$, so that $\varphi(q) = q_1$. Let further

$$C^*(x) = \begin{pmatrix} 1 & |x_1| \\ 1 & 1 \end{pmatrix}; \quad \bar{p} = (0, 1), \quad \bar{x} = (0, 0).$$

Then $H(x, p) = |p_1 + |x_1|p_2|$; moreover, $\bar{\partial} H(\bar{x}, \bar{p})$ is the convex hull of four points $(\pm 1, 0, \pm 1, 0)$, which is $\{(w, 0, y, 0) : \max\{|w|, |y|\} \leq 1\}$. This means that $\{w : (w, 0) \in \bar{\partial} H(\bar{x}, \bar{p})\} = [-1, 1]$.

REFERENCES

- [1] J.-P. Aubin, *Lipschitz behavior of solutions to convex minimization problems*, Math. Oper. Res. **9** (1984), 87–111. MR **86f**:90119
- [2] V. I. Blagodatskikh, *Maximum principle for differential inclusions*, Trudy Mat. Inst. Steklov **166** (1984), 23–43 (Russian); English translation in Proc. Steklov Inst. Math. **166** (1986). MR **85m**:49050
- [3] V. G. Boltianski, *Local section method in the theory of optimal processes*, Differentsial'nye Uravneniya **4** (1968), 2166 – 2183; English transl. in Differential Equations **4** (1968). MR **39**:3862
- [4] F.H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc. **205** (1975), 247–262. MR **51**:3373
- [5] F. H. Clarke, *Necessary conditions for a general control problems*, in Calculus of Variations and Control Theory, D. Russel, editor, Academic Press 1976, pp 259–278. MR **58**:30664
- [6] F. H. Clarke, *The generalized problem of Bolza*, SIAM J. Control Optimization **14** (1976), 469–478. MR **54**:1047
- [7] F. H. Clarke, *Extremal arcs and extended Hamiltonian systems*, Trans. Amer. Math. Soc. **231** (1977), 349–367. MR **56**:1163
- [8] F. H. Clarke, *Optimal control and the true Hamiltonian*, SIAM Review **21** (1979), 157–166. MR **80e**:49001
- [9] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983; reprinted by CRM, Université de Montréal, 1989. MR **85m**:49002; MR **90g**:49001
- [10] F. H. Clarke, *Methods of Dynamic and Nonsmooth Optimization*, SIAM Publications, Philadelphia, 1989. MR **91j**:49001
- [11] R. Correa, A. Jofré and L. Thibault, *Characterization of lower semicontinuous convex functions*, Proc. Amer. Math. Soc. **116** (1992), 67–72. MR **92k**:49027
- [12] R. P. Fedorenko, *Maximum principle for differential inclusions (necessity)*, Zh. Vychisl. Mat. i Mat. Fiz. **11** (1971), 885–893; English transl. in USSR Comput. Math. and Math. Phys. **11** (1971). MR **45**:9762
- [13] H. Frankowska, *The maximum principle for an optimal solution to a differential inclusion with end points constraints*, SIAM J. Control Optimization **25** (1987), 145–157. MR **88b**:49025
- [14] H. Frankowska, *Contingent cones to reachable sets of control systems*, SIAM J. Control Optimization **27** (1989), 170–198. MR **89m**:49034
- [15] B. Ginsburg and A. Ioffe, *The maximum principle in optimal control of systems governed by semilinear equations*, Proceedings of the IMA Workshop on Nonsmooth Analysis and Geometric Methods in Deterministic Optimal Control, B. Mordukhovich and H. Sussman, eds, IMA Vol. Math. Appl., 78, Springer, 1996, pp. 81–110. CMP 97:03
- [16] J.-B. Hiriart-Urruty, *Extensions of Lipschitz functions*, J. Math. Anal. Appl. **77** (1980), 539–544. MR **83i**:58013
- [17] A. Ioffe *Necessary and sufficient conditions for a local minimum* SIAM J. Control Optimization **17** (1979), 245–250. MR **82j**:49005
- [18] A. Ioffe *Sous-différentielles approchées des fonctions numériques*, C.R. Acad. Sci. Paris Ser. I Math. **292** (1981), 675–678. MR **82e**:49026
- [19] A. Ioffe, *Single-valued representation of set-valued mappings 2. Application to differential inclusions*, SIAM J. Control Optimization **21** (1983), 641 – 651. MR **85b**:49017
- [20] A. Ioffe, *Calculus of Dini subdifferentials and contingent derivatives of set-valued maps*, Nonlinear Anal. Theory Meth. Appl. **8** (1984), 517–539. MR **85k**:46049
- [21] A. Ioffe, *Approximate subdifferentials and applications*, Trans. American Math. Soc. **284** (1984), 389–416. MR **84m**:49029
- [22] A. Ioffe, *Necessary conditions in nonsmooth optimization*, Math. Oper. Res. **9** (1984), 159–189. MR **85h**:49043
- [23] A. Ioffe, *Proximal analysis and approximate subdifferentials*, J. London Math. Soc. **41** (1990), 175–192. MR **91i**:46045
- [24] A. Ioffe, *Nonsmooth subdifferentials: their calculus and applications*, Proceedings of the 1st Congress of Nonlinear Analysts, de Gruyter, Berlin, 1996, pp. 2299–2310. CMP 96:12
- [25] A. Ioffe and R. T. Rockafellar, *The Euler and Weierstrass conditions for nonsmooth variational problems*, Calc. Var. Partial Differential Equations **4** (1996), 59–87. MR **96k**:49031

- [26] A. Ioffe and V. Tikhomirov, *Theory of Extremal Problems*, Nauka, Moscow, 1974; English translation, North-Holland, 1979. MR **53**:14251; MR **80d**:49001
- [27] A. Jourani and L. Thibault, *Verifiable conditions for openness and regularity of multivalued mappings in Banach spaces*, Trans. Amer. Math. Soc. **347** (1995), 1255–1268. MR **95h**:49021
- [28] B. Kaśkosz and S. Lojasiewicz Jr., *A maximum principle for generalized control systems*, Nonlinear Analysis. Theory, Methods & Appl. **9** (1985), 109–130. MR **86f**:49062
- [29] B. Kaśkosz and S. Lojasiewicz Jr., *Lagrange-type extremal trajectories in differential inclusions*, Systes and Control Letters **19** (1992), 241–247. MR **93i**:49030
- [30] P. Loewen and R. T. Rockafellar, *The adjoint arc in nonsmooth optimization*, Trans. Amer. Math. Soc. **325** (1991), 39–72. MR **91h**:49019
- [31] P. Loewen and R. T. Rockafellar, *Optimal control of unbounded differential inclusions*, SIAM J. Control Opt. **32** (1994), 442–470. MR **95h**:49043
- [32] P. Loewen and R. T. Rockafellar, *New necessary conditions for generalized problem of Bolza*, SIAM J. Control Optimization **34** (1996), 1496–1511. CMP 96:17
- [33] P. D. Loewen and R. B. Vinter, *Pontriagin-type necessary conditions for differential inclusion problem*, System and Control Letters **9** (1987), 263–265.
- [34] S. Lojasiewicz, *Lipschitz selectors of orientor fields*
- [35] S. Lojasiewicz, *Local controllability of parametrized differential equations*, in preparation.
- [36] B. Mordukhovich, *The maximum principle in the problem of time-optimal control with nonsmooth constraints*, J. Appl. Math. Mech. **40** (1976), 960–969. MR **58**:7284
- [37] B. Mordukhovich, *Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems*, Soviet Math. Doklady **22** (1980), 526–530. MR **82b**:90104
- [38] B. Mordukhovich, *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow, 1988; English transl., Wiley, New York (to appear). MR **89m**:49001
- [39] B. Mordukhovich, *On variational analysis of differential inclusions*, in A. Ioffe, M. Marcus and S. Reich (eds.), *Optimization and Nonlinear Analysis*, Pitman Research Notes in Math. 244, Longman Sci. Tech., Harlow, 1992, pp. 199–213. MR **93d**:49004
- [40] B. Mordukhovich, *Complete characterization of openness, metric regularity and Lipschitzian properties of multifunctions*, Trans. Amer. Math. Soc. **340** (1993), 1–35. MR **94a**:49011
- [41] B. Mordukhovich, *Discrete approximations and refined Euler-Lagrange conditions for nonconvex differential inclusions*, SIAM J. Control Opt. **33** (1995), 882–915. MR **96d**:49028
- [42] E.S. Polovinkin and G.V. Smirnov, *An approach to differentiation of many-valued mappings and necessary conditions for optimization of solutions of differential inclusions*, Differential Equations **22** (1986), 660–668. MR **87i**:49047
- [43] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, 1970. MR **43**:445
- [44] R. T. Rockafellar, *Generalized Hamiltonian equations for convex problems of Lagrange*, Pacific J. Math. **33** (1970), 411–428. MR **43**:2593
- [45] R. T. Rockafellar, *Existence and duality theorems for convex problems of Bolza*, Trans. Amer. Math. Soc. **159** (1971), 1–40. MR **43**:7995
- [46] R. T. Rockafellar, *Lipschitzian properties of multifunctions*, Nonlinear Analysis, Theory, Methods, Appl. **9** (1985), 867–885. MR **87a**:90149
- [47] R. T. Rockafellar, *Dualization of subgradient conditions for optimality*, Nonlinear Anal. Theory Meth. Appl. **20** (1993), 627–646. MR **94c**:49021
- [48] R. T. Rockafellar, *Equivalent subgradient versions of Hamiltonian and Euler-Lagrange equations in variational analysis*, SIAM J. Control Optim. **34** (1996), 1300–1314. CMP 96:14
- [49] G. V. Smirnov, *Discrete approximations and optimal solutions to differential inclusions*, Kibernetika (Kiev) **1991**, no. 1, 76–79; English transl., Cybernetics **27** (1991), no. 1, 101–107. MR **92h**:49019
- [50] H. Sussmann, *A strong version of the Lojasiewicz maximum principle*, in Optimal Control of Differential Equations, N. H. Pavel, Ed., M. Dekker, N.Y. 1994, pp. 293–309. MR **95g**:49035
- [51] R. Vinter and H. Zheng, *The extended Euler-Lagrange condition for nonconvex variational problems*, SIAM J. Control Optimization, to appear.
- [52] J. Warga, *Controllability, extremality and abnormality in nonsmooth optimal control*, J. Opt. Theory Appl. **41** (1983), 239–260. MR **85a**:49045
- [53] Q. Zhu, *Necessary optimality conditions for nonconvex differentiable inclusions with endpoint constraints*, J. Differential Equations **124** (1996), 186–204. CMP 96:06