Can. J. Math., Vol. XXXII, No. 2, 1980, pp. 494–509

THE ERDMANN CONDITION AND HAMILTONIAN INCLUSIONS IN OPTIMAL CONTROL AND THE CALCULUS OF VARIATIONS

FRANK H. CLARKE

1. Introduction. Consider the basic problem in the calculus of variations, that of minimizing

$$(1.1) \quad \int_0^1 L(t, x(t), \dot{x}(t))dt$$

over a class of functions x satisfying certain boundary conditions at 0 and 1. One of the classical first order necessary conditions for optimality is the second Erdmann condition, which asserts, in the case in which L is independent of t, that

$$(1.2) \quad L(x(t), \dot{x}(t)) - \dot{x}(t) \cdot L_{\dot{x}}(x(t), \dot{x}(t)) = \text{constant}$$

along any local solution x. This formula is the customary basis for solving many of the classical problems, such as the brachistochrone. When it is possible to define via the Legendre transform a Hamiltonian $H(t, x, p)$ corresponding to L, the second Erdmann condition, again in the autonomous case, is the assertion that

$$(1.3) \quad H(x(t), p(t)) = \text{constant},$$

a relation which always evokes classical Hamiltonian mechanics and conservation laws.

In [**3**] the author considered a variational problem involving (1.1) in which L was assumed measurable in t and locally Lipschitz in $(x, \dot{x})$. In lieu of absent derivatives, the necessary conditions are couched in terms of generalized gradients [**2**], which reduce in particular to the customary derivative or subdifferential if the function is C^1 or convex. In [**4**], we analyzed a general control problem defined in terms of a differential inclusion

$$(1.4) \quad \dot{x}(t) \in E(t, x(t)),$$

where the multifunction E is Lipschitz in x and measurable in t. It was

Received August 2, 1978 and in revised form February 6, 1979. This work was supported in part by NRC Grant No. A9082.

https://doi.org/10.4153/CJM-1980-039-9 Published online by Cambridge University Press

shown that a necessary condition for optimality was the existence of a solution of the Hamiltonian inclusion

$$(1.5) \quad (-\dot{p}(t), \dot{x}(t)) \in \partial H(t, x(t), p(t)),$$

where H is defined by

$$(1.6) \quad H(t, x, p) = \max \{p \cdot v \colon v \in E(t, x)\},$$

and where ∂H refers to the generalized gradient of H in (x, p) (note the analogy with a Hamiltonian system: $-\dot{p} = H_x, \dot{x} = H_p$). The properties and uses of (1.4) (1.5) are discussed in [**4**] and [**2**]; one of the motivations for studying (1.4) is to enable consideration of cases in which the control set depends on the state, a situation which renders the maximum principle inapplicable.

It is natural to ask whether, if the behaviour of L and E on t is assumed more regular, the classical Erdmann conditions find their counterpart in this new setting. Our main purpose here is to answer this in the affirmative, in § 2 for the optimal control problem, and in § 6 for the variational setting. In the latter case, when L is independent of t, we prove in particular that the useful formula (1.2) is valid even if L has only a derivative in $\dot{x}$. The reader who is familiar with the Pontryagin maximum principle of optimal control theory may recall that, in the autonomous version of that result, the constancy of the pseudo-Hamiltonian is a consequence of the other necessary conditions. In the present setting, however, the Erdmann condition is an independent conclusion. This is analogous to the classical setting, where (1.2) is not a consequence of the Euler equation unless L is assumed C^2 and $\dot{x}$ smooth. We now describe briefly the other contents of this article.

In § 3 we discuss "calmness", a constraint qualification assuring normality of the necessary conditions in § 2. It is shown next that the case in which the endpoint constraints in a control problem are given by inequalities is technically reducible to the free-endpoint case, a fact which seems to have been unrecognized. We go on to give a precise meaning to the statement "most inequality constrained problems are normal". We conclude the section by showing that the multicriterion (Pareto optimum) problem can be reduced to the single criterion case, thus avoiding the need for special treatment. This technique, like the removal of inequality constraints mentioned before, makes essential use of the capability of treating nondifferentiable functions. In § 4 we present a controllability result for trajectories on the boundary of the attainable set, while in § 5 we develop a version of the results of § 2 which is convenient when an integral cost functional is part of the problem. As an example of its application, we derive via a new variational principle involving a differential inclusion a theorem which asserts the existence of periodic solutions of certain Hamiltonian equations (full details appear elsewhere [**12**]).

The results of this paper have little overlap with the existing literature. Hamiltonian inclusions were studied in the concave-convex case by R. T. Rockafellar [**17**] [**18**]. J. Warga [**20**] and H. Halkin [**15**] have derived necessary conditions for control problems with nondifferentiable data; since these results are obtained within the Pontryagin framework, they are comparable to [**6**] rather than to the present work. Finally, R. P. Fedorenko [**13**] and V. G. Boltjanskii [**1**] have studied the differential inclusion problem under strong regularity hypotheses.

2. The Erdmann condition in optimal control.

2.1. *Statement of the problem.* Let X be a given open subset of $\mathbf{R}^n$. We shall suppose that the multifunction E appearing in (1.4) has the following properties:

(E_1) for each t in $[0, 1]$ and s in X, $E(t, s)$ is a non-empty compact set.

(E_2) there is a constant K with the following property: Given any points (t_1, s_1), (t_2, s_2) in $[0, 1] \times X$ and v_1 belonging to $E(t_1, s_1)$, there is some v_2 in $E(t_2, s_2)$ such that

$$|v_1 - v_2| \leqq K|(t_1 - t_2, s_1 - s_2)|.$$

(That is, E is Lipschitz in (t, s) in the Hausdorff metric.)

We define the *Hamiltonian* function $H:[0, 1] \times X \times \mathbf{R}^n \rightarrow \mathbf{R}$ as follows:

$$H(t, s, p) = \max \{p \cdot v : v \in E(t, s)\}.$$

Our hypotheses imply that H is Lipschitz in (t, s, p).

Given two closed subsets C_0, C_1 of $\mathbf{R}^n$ and a locally Lipschitz function $\phi : C_1 \rightarrow \mathbf{R}$, the problem we consider is that of minimizing $\phi(x(1))$ over the arcs x satisfying

$$x(t) \in X$$
$$\dot{x}(t) \in E(t, x(t)) \text{ a.e.}$$
$$x(0) \in C_0, x(1) \in C_1.$$

(An *arc* is an absolutely continuous function mapping $[0, 1]$ to $\mathbf{R}^n$.) An arc $\hat{x}$ is said to be a *local solution* if, for some $\delta > 0$, it solves this problem relative to the arcs x satisfying

$$|x(t) - \hat{x}(t)| < \delta.$$

(The word "local" will always have this meaning in later sections when we consider other problems.)

2.2. *Statement of the theorem.*

THEOREM 2.1. *If the arc x is a local solution, then there exist an arc p, and a scalar λ equal to either 0 or 1, such that $|p(t)| + \lambda$ is nonvanishing and*

$$(2.1)\quad (\dot{H}(t), -\dot{p}(t), \dot{x}(t)) \in \partial H(t, x(t), p(t)) \quad a.e.,$$

$$(2.2)\quad p(0) \in N(C_0; x(0)), -p(1) \in \lambda\partial\phi(x(1)) + N(C_1; x(1)),$$

where $\dot{H}(t)$ denotes the derivative of the function

$$H(t) = H(t, x(t), p(t)).$$

Remark. In (2.1), ∂H refers to the generalized gradient with respect to the variables (t, x, p). Since the function $H(t)$ is Lipschitz, as is easily seen, it is differentiable for almost all t, and so $\dot{H}(t)$ exists a.e. The notation $N(C; \cdot)$ for a closed set C refers to the generalized normal cone [**2**], which reduces to the customary entities if C is either a convex set or a smooth manifold. The case $\lambda = 1$, which has the desirable feature that ϕ is actually implicated in (2.2), is called "normal"; we shall discuss this in § 3. The inclusion (2.1) implies $H(t) = p(t) \cdot \dot{x}(t)$.

COROLLARY 2.1. *Suppose that E (and consequently, H) does not depend on t. Then p satisfies*

$$(2.3)\quad (-\dot{p}, \dot{x}) \in \partial H(x(t), p(t)),$$

$$(2.4)\quad H(x(t), p(t)) = constant.$$

Remark. We thus recover in the autonomous case an analogue of the traditional Hamiltonian system evolving on a level surface of H. An important distinction, however, is the fact that in the present context, (2.4) is not merely a consequence of (2.3), as in the classical situation in which H, $\dot{x}$, $\dot{p}$ are smooth.

2.3. *Reduction to the autonomous case.* Our purpose here is to show that it suffices to prove Corollary 2.1, since the theorem then follows from it. To see this, suppose that $x = \hat{x}$ satisfies the hypotheses of the theorem. We shall shift the problem to $\mathbf{R}^{n+1}$ from $\mathbf{R}^n$ by adding a zero$^{\text{th}}$ coordinate; variables in $\mathbf{R}^{n+1}$ will be indicated by an overbar, so that $\bar{x} = (x_0, x)$ lies in $\mathbf{R}^{n+1}$, with x_0 in $\mathbf{R}$, x in $\mathbf{R}^n$. We define

$$\bar{E}(\bar{x}) = \{1\} \times E(x_0, x),$$
$$\bar{C}_0 = \{0\} \times C_0,\ \bar{C}_1 = R \times C_1,\ \bar{X} = R \times X,$$
$$\bar{\phi}(\bar{x}) = \phi(x),$$
$$\bar{H}(x, p) = p_0 + H(x_0, x, p).$$

Then the arc $\bar{x} = (t, x)$ minimizes (locally) $\bar{\phi}(\bar{y}(1))$ subject to

$$\bar{y}(t) \in \bar{X},$$
$$\dot{\bar{y}} \in \bar{E}(\bar{y}) \text{ a.e.},$$
$$\bar{y}(0) \in \bar{C}_0,\ \bar{y}(1) \in \bar{C}_1.$$

https://doi.org/10.4153/CJM-1980-039-9 Published online by Cambridge University Press

We apply the corollary to deduce the existence of p and λ such that $|p(t)| + \lambda$ is never 0 and

(2.5) $(-\dot{p}_0, -\dot{p}, 1, \dot{x}) \in \partial\bar{H}(\bar{x}, \bar{p})$

(2.6) $p(0) \in N(C_0; \dot{x}(0)), \ -p(1) \in \lambda\partial\phi(x(1)) + N(C_1; x(1)),$

$$p_0(1) = 0,$$

(2.7) $\bar{H}(\bar{x}, \bar{p}) = \text{constant} = c.$

Consequently (2.2) holds, and furthermore these conditions imply

$$(-\dot{p}_0, -\dot{p}, \dot{x}) \in \partial H(t, x, p)$$
$$\bar{H}(\bar{x}, \bar{p}) = \bar{p}\cdot\dot{\bar{x}} = p_0 + p\cdot\dot{x} = p_0 + H(t, x, p) = c.$$

It follows that $\dot{p}_0 = -\dot{H}(t)$, and so (2.1) ensues. It remains to prove that $|p(t)| + \lambda$ is nonvanishing. Suppose to the contrary that $\lambda = 0$, and that $p(t)$ is somewhere 0. Relation (2.1) implies a differential inequality $|\dot{p}| \leqq K|p|$, so that p is identically zero. But then $|(p_0(t), p(t))| + \lambda$ vanishes at $t = 1$, a contradiction.

2.4. *Proof of Corollary* 2.1. As in § 2.3, $\bar{x} = (x_0, x)$ denotes points in $\mathbf{R}^{n+1}$. Let α in $(0, 1)$ be given. The following observation is fundamental.

LEMMA 2.1. *The arc* $(0, x)$ *is a local solution to the problem of minimizing* $\phi(y(1))$ *over the* $(n + 1)$*-dimensional arcs* (y_0, y) *satisfying*

(2.8) $\dot{y}(\tau) \in (1 + \dot{y}_0(\tau))E(y(\tau))$ a.e.,

(2.9) $|\dot{y}_0(\tau)| \leqq \alpha,$

(2.10) $y(0) \in C_0, \ y(1) \in C_1, \ y_0(0) = y_0(1) = 0.$

Proof. (By contradiction). Suppose (y_0, y) satisfies (2.8)–(2.10) and $\phi(y(1)) < \phi(x(1))$. Let us define an arc $z(t)$ as follows. For each t in $[0, 1]$, determine τ in $[0, 1]$ from the equation

$$\tau + y_0(\tau) = t.$$

This is possible because the function $\tau \rightarrow \tau + y_0(\tau)$ is monotonic, with value 0 at 0 and value 1 at 1. We set

$$z(t) = y(\tau).$$

Since the implicit function $\tau(t)$ is Lipschitz by [**5**, Theorem 1], it follows that z is an arc, and since $z(0) = y(0)$, $z(1) = y(1)$, it follows also that z satisfies the boundary conditions. If we verify that $\dot{z}(t)$ is given by $\dot{y}(\tau)/(1 + \dot{y}_0(\tau))$, then by (2.8) we have

$$\dot{z}(t) \in E(z(t)) \text{ a.e.}$$

Thus z is feasible for the problem that x solves, and if (y_0, y) is close to $(0, x)$, then z is close to x. Yet $\phi(z(1)) < \phi(x(1))$. This is the required contradiction, so the lemma is proved.

We now proceed by applying the necessary conditions of [**4**, Theorem 2] to the problem (2.8–2.10). These imply the existence of an arc (p_0, p) and a scalar $s \geqq 0$ such that $|(p_0, p)| + s$ is non-vanishing, $\|(p_0, p)\| + s = 1$ ($\|\cdot\|$ denotes L^∞ norm), and

$$(2.11) \quad (-\dot{p}_0, -\dot{p}, 0, \dot{x}) \in \partial H^*(0, x, p_0, p) \text{ a.e.},$$

$$(2.12) \quad p(0) \in N(C_0; x(0)), \; -p(1) \in s\partial\phi(x(1)) + N(C_1; x(1)),$$

where

$$(2.13) \quad \begin{aligned} H^*(y_0, y, p_0, p) &= \max\{p_0v_0 + p\cdot v: |v_0| \leqq \alpha, v \in (1 + v_0)E(y)\} \\ &= \max\{p_0v_0 + (1 + v_0)H(y, p): |v_0| \leqq \alpha\} \\ &= H(y, p) + \alpha|p_0 + H(y, p)|. \end{aligned}$$

It follows from [**10**, Art. 15] and from convex analysis that (2.11) implies

$$H^*(0, x, p_0, p) = p_0 0 + p\cdot\dot{x} \leqq H(x, p),$$

so we deduce (in view of (2.13))

$$(2.14) \quad H(t) = -p_0,$$

where $H(t) = H(x(t), p(t))$.

Since H^* does not actually depend on y_0, (2.11) implies that $\dot{p}_0 = 0$ a.e., and so p_0 is constant. The expression (2.13) for H^*, together with (2.11) and [**10**, Art. 8] yields:

$$(2.15) \quad (-\dot{p}, \dot{x}) \in \partial H(x, p) + \alpha K B^{2n} \text{ a.e.},$$

where B^j is the open unit ball in R^j. Finally, we note

$$(2.16) \quad \|H(t), p(t))\| + s = 1.$$

We now proceed to do the above for a sequence of values of α decreasing to 0. In so doing, we generate a sequence of arcs p (uniformly bounded) and scalars p_0, s (bounded) such that (2.12), (2.14), (2.15) and (2.16) hold (with the K in (2.15) independent of α). As shown in [**9**, Lemma 8] and [**6**, Lemma 5] we may choose a subsequence of (p, p_0, s) converging (in an appropriate sense) to a triple (p, p_0, λ) for which (2.2), (2.3) and (2.4) hold. These limits also satisfy (2.16), which rules out the simultaneous vanishing of λ and $p(t)$ (since $H(\cdot, 0) = 0$). If λ is positive, the conclusions still hold if p is replaced by p/λ and λ by 1 (see [**9**, Lemma 3], in the statement of which q, in its last appearance, should be multiplied by ϵ), so that there is no loss of generality in taking $\lambda = 0$ or 1.

3. Constraint behaviour, calmness and normality. Consider now the problem $P(s)$ of minimizing $\phi(x(1))$ subject to $x(t) \in X$, $x(0) \in C_0$, $\dot{x} \in E(t, x)$ a.e. and

$$x(1) \in C_1 + s.$$

https://doi.org/10.4153/CJM-1980-039-9 Published online by Cambridge University Press

Let the infimum for this problem be $\Phi(s) \in [-\infty, \infty]$ (we take the infimum over the empty set to be $+\infty$).

Definition. The problem $P(0)$ is calm if $\Phi(0)$ is finite and

$$\liminf_{s\to 0} [\Phi(s) - \Phi(0)]/|s| > -\infty.$$

3.1. *Calmness implies normality.*

THEOREM 3.1. *If to Theorem* 2.1 *we add the hypothesis that the problem is calm and* C_0 *is compact then the conclusions hold with* $\lambda = 1$.

Proof. We claim that, for some integer j, x solves the problem of minimizing ($d(C_1; s)$ is the Euclidean distance from s to C_1)

$$\phi(y(1) + jd(C_1; y(1)) \tag{3.1}$$

over the trajectories y for E satisfying $y(0) \in C_0$, $y(t) \in X$. If this were false, there would be an admissible arc y_j for each j such that

$$\phi(y_j(1)) + jd(C_1; y_j(1)) < \phi(x(1)) = \Phi(0).$$

It follows that $d(C_1; y_j(1)) \to 0$. Let s_j be a point of least magnitude such that

$$y_j(1) \in C_1 + s_j.$$

Then $d(C_1; y_j(1)) = |s_j| > 0$, and the last inequality implies

$$\Phi(s_j) - \Phi(0) < -j|s_j|.$$

But this would contradict the calmness assumption.

We proceed to apply Theorem 2.1 to the problem in which the roles of ϕ and C_1 in the theorem are played by (3.1) and $\mathbf{R}^n$. In the theorem, it always follows that $\lambda = 1$ when $C_1 = \mathbf{R}^n$, for otherwise (2.2) would imply that $p(1)$ vanishes along with λ. The result now follows when we recall [**2**] that the set

$$j\partial d(C_1; x(1))$$

is contained in $N(C_1; x(1))$.

Remark. The calmness of a given problem is automatic when the solution x is such that $x(1)$ lies in the interior of C_1. It can also be guaranteed by certain convexity hypotheses (see [**7**, Proposition 5]) or other special structure.

3.2. *Inequality endpoint constraints.* The endpoint constraints in an optimal control problem are sometimes given in the form

$$g_i(x(1)) \leqq 0 \quad (i = 1, 2, \ldots K), \tag{3.2}$$

where the g_i: $\mathbf{R}^n \to \mathbf{R}$ are given functions. Let us consider the problem

https://doi.org/10.4153/CJM-1980-039-9 Published online by Cambridge University Press

of minimizing $\phi(x(1))$ over the trajectories x for E satisfying $x(0) \in C_0$, $x(t) \in X$, and (3.2). We posit the hypotheses of § 2, and we assume as well that each g_i is locally Lipschitz.

The following result shows that (in the context of non-differentiable optimization) the inequality-constraint case is technically equivalent to the free-endpoint case.

LEMMA 3.1. *If x solves the above problem, then x minimizes $\tilde{\phi}(y(1))$ over all trajectories y for E satisfying $y(t) \in X$, $y(0) \in C$, where the locally Lipschitz function $\tilde{\phi}$: $\mathbf{R}^n \to \mathbf{R}$ is defined by*

$$\tilde{\phi}(s) = \max \{\phi(s) - \phi(x(1)), g_1(s), g_2(s), \ldots, g_K(s)\}.$$

Proof. Note that $\tilde{\phi}(x(1)) = 0$. If y is such that

$$\tilde{\phi}(y(1)) < 0,$$

then y satisfies (3.2), and

$$\phi(y(1)) < \phi(x(1)),$$

a contradiction.

COROLLARY 3.1. *If x solves the above problem, then there exist an arc p and nonnegative numbers $\lambda_0, \lambda_1, \ldots \lambda_K$ not all zero such that the inclusion* (2.1) *holds, as well as*

$$-p(1) \in \lambda_0 \partial\phi(x(1)) + \lambda_1 \partial g_1(x(1)) + \ldots + \lambda_K \partial g_K(x(1)), \tag{3.3}$$

$$\lambda_i g_i(x(1)) = 0 \quad (i = 1, 2, \ldots K), \tag{3.4}$$

$$p(0) \in N(C_0; x(0)). \tag{3.5}$$

Proof. We apply Theorem 2.1 to the problem described in the lemma; since $C_1 = \mathbf{R}^n$, it follows that $\lambda = 1$, so that we have

$$-p(1) \in \partial\tilde{\phi}(x(1)). \tag{3.6}$$

From a general characterization of generalized gradients of pointwise maxima [**8**, Proposition 9], it follows that (3.6) implies (3.3)–(3.4).

Remark. Note that if 0 belongs to $\sum \lambda_i \partial g_i(x(1))$ for suitable λ_i, then we generally satisfy all the necessary conditions by taking $\lambda_0 = 0$, $p \equiv 0$. Consequently, if one attempts to represent a constraint such as $x(1) = \zeta$ by inequalities

$$g_i(x(1)) \leqq 0,$$

the fact that 0 necessarily belongs to $\sum \lambda_i \partial g_i(x(1))$ for certain λ_i not all 0 implies that no non-trivial necessary conditions will result. This further delineates the fact that assuming inequality constraints (2.2) rather than the abstract constraint $x(1) \in C_1$ (as many authors do) is a considerable simplification.

3.3. *Most inequality constrained problems are normal.* Consider again the problem of § 3.2, but with the inequality endpoint constraints (3.2) perturbed to:

$$(3.7) \qquad g_i(x(1)) \leqq \alpha_i \ (i = 1, 2, \ldots K),$$

where $\alpha = (\alpha_1, \alpha_2, \ldots \alpha_K)$ lies in $\mathbf{R}^K$. We denote by $\Phi(\alpha)$ the infimum in the problem when the endpoint constraints are given by (3.7) (cf. § 3.1)

THEOREM 3.2. *Suppose that in a neighbourhood N of* 0, *Φ is finite. Then for almost every α in N, the problem with constraints* (3.7) *is normal, in the following sense: given any solution to the problem, the necessary conditions of Corollary* 3.1 *hold with* $\lambda = 1$.

Proof. The very definition of Φ implies that it is decreasing in each coordinate. Such a function is known to be differentiable a.e. [**19**], so that for almost every α, we must have

$$\liminf_{s \to 0} [\Phi(\alpha + s) - \Phi(\alpha)] > -\infty.$$

A simple argument paralleling that in the proof of Theorem 3.1 shows that any solution x to the problem must (for j large) minimize $\tilde{\phi}_j(x(1))$ over the admissible trajectories, where the locally Lipschitz function $\tilde{\phi}_j$ is defined by

$$\tilde{\phi}_j(x) = \phi(x) + j \max \{g_i(x) - \alpha_i, 0 \colon 1 \leqq i \leqq K\}.$$

It suffices then to apply Theorem 2.1 to this problem with ($\lambda = 1$ automatic) to see that the conclusions of Corollary 3.1 (with $\lambda = 1$) hold (for g_i replaced by $g_i - \alpha_i$, of course).

Remark. Other authors use the word "normal" to mean that the necessary conditions must hold with $\lambda = 1$, for any possible p. Our use of this term signifies rather the existence of some p for which the necessary conditions hold with $\lambda = 1$. A result similar to the above, in the context of mathematical programming, is given in [**8**, Theorem 3].

3.4. *Pareto optima.* Considerable attention has been paid of late to optimization problems incorporating multiple criteria. Our purpose here is to demonstrate that necessary conditions for such problems can be obtained as a simple corollary to the usual case; once again our capability to treat nondifferentiable functions is vital. Consider K locally Lipschitz functions $g_1, g_2, \ldots g_K$ mapping $\mathbf{R}^n$ to $\mathbf{R}$, in relation to the trajectories y of a given differential inclusion $\dot{y} \in E(t, y)$, $y(0) \in C_0$, $y(1) \in C_1$, $y(t) \in X$. Such a trajectory x is said to be (locally) *Pareto optimal* if there is no other trajectory y such that $g_i(y(1)) < g_i(x(1))$, $i = 1, 2, \ldots K$ (this is the least restrictive sense in which x can be optimal). We posit the hypotheses of § 2.

THEOREM 3.3. *Let x be locally Pareto optimal. Then there exist an arc p and nonnegative numbers $\lambda_1, \ldots \lambda_K$ such that $|p(t)| + \Sigma\lambda_i$ is nonvanishing, the inclusion* (2.1) *holds, and*

$$-p(1) \in \sum \lambda_i \partial g_i(x(1)) + N(C_1; x(1)), \tag{3.8}$$

$$p(0) \in N(C_0; x(0)). \tag{3.9}$$

Proof. The key is the easily verified assertion that x must minimize (locally) the function

$$\tilde{\phi}(y(1)) = \max_i \{g_i(y(1)) - g_i(x(1))\}$$

over the admissible trajectories y for E. We now apply Theorem 3.1, and the characterization of $\partial\tilde{\phi}$ [**8**, Proposition 9] provides the conclusion.

4. Controllability. Let C be a closed subset of $\mathbf{R}^n$, X an open subset, and consider the differential inclusion

$$\dot{x} \in E(t, x) \text{ a.e.}, \ 0 \leqq t \leqq 1, \ x(0) \in C, \ x(t) \in X, \tag{4.1}$$

where E satisfies the hypotheses (E_1) (E_2) of § 2.1. We define the *attainable set* $A(C)$ as the set of all points $x(1)$ where $x(\cdot)$ is a solution of (4.1).

We now give necessary conditions that an arc x be such that $g(x(1))$ lies on the boundary of $g(A(C))$, where $g: \mathbf{R}^n \to \mathbf{R}^m$ is a given locally Lipschitz function.

THEOREM 4.1. *Let the trajectory x satisfy* (4.1), *and suppose that $g(x(1))$ lies in $\partial g(A(C))$. Then there exist an arc p and a unit vector v such that*

$$(\dot{H}(t), -\dot{p}(t), \dot{x}(t)) \in \partial H(t, x(t), p(t)) \text{ a.e.}, \tag{4.2}$$

$$p(0) \in N(C; x(0)), \ p(1) \in \partial g(x(1))v. \tag{4.3}$$

Remark. The notation ∂g refers to the generalized Jacobian of g [**4**]. The proof is sufficiently like that of Theorem 2.1 to be left to the reader; reduce to the autonomous case (§ 2.3), apply the transformation of Lemma 2.1, and then invoke [**4**, Theorem 1].

5. Explicit integral cost functionals; example. Our purpose here is to deduce from the results of § 2 necessary conditions for the following problem: to minimize

$$\phi(x(1)) + \int_0^1 L(t, x, \dot{x})dt \tag{5.1}$$

over the arcs x satisfying $x(t) \in X$ and

$$\dot{x} \in E(t, x) \text{ a.e.},$$
$$x(0) \in C_0, \ x(1) \in C_1.$$

We suppose that $L: [0, 1] \times \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ is locally Lipschitz; the hypotheses of § 2 remain in force for the other data.

It is not hard to reformulate the above problem so as to have it assume the shape of the problem considered in § 2; it is a matter of redefining E so as to "absorb" L (see [**4**, Example]). But cost functionals are often given in the form (5.1), so conditions specifically adapted to this case are convenient; it is the case also that the theorem below says something rather different than what is obtained by merely reformulating. When L exhibits discontinuous behaviour in $\dot{x}$ (which is the case in the example cited above), the reformulation must, however, be resorted to. In the following, we assume either that the problem is calm or that E has no dependence on x.

THEOREM 5.1. *If x is a local solution to the above problem, then there is an arc p, and a scalar λ equal to* 0 *or* 1, *such that $|p(t)| + \lambda$ is nonvanishing and*

$$(5.1) \qquad (\dot{H}^*(t), -\dot{p}(t), \dot{x}(t)) \in \partial H^*(t, x(t), p(t)) \text{ a.e.},$$

$$(5.2) \qquad p(0) \in N(C_0; x(0)), \ -p(1) \in \lambda\partial\phi(x(1)) + N(C_1; x(1)),$$

where H^* is defined by

$$H^*(t, x, p) = \max \{p \cdot v - \lambda L(t, x, v) : v \in E(t, x)\},$$

and where

$$H^*(t) = H^*(t, x(t), p(t)).$$

Remark. As in § 2.2, the conclusions in the autonomous case become

$$(5.3) \qquad (-\dot{p}, \dot{x}) \in \partial H^*(x, p) \text{ a.e.},$$

$$(5.4) \qquad H^*(t) = \text{constant}.$$

It may be shown as in § 2.3 that the general case can be reduced to this one, for which we now prove the theorem. Using once again the notation $\bar{x} = (x_0, x)$ for points in $\mathbf{R}^{n+1}$, we define

$$\bar{E}(\bar{x}) = \{(L(x, v), v) : v \in E(x)\}$$
$$\bar{C}_0 = \{0\} \times C_0, \ \bar{C}_1 = \mathbf{R} \times C_1, \ \bar{X} = \mathbf{R} \times X$$
$$\bar{\phi}(\bar{x}) = x_0 + \phi(x).$$

It is not difficult to see that the arc $\bar{x}(t) = [\int_0^t L(x, \dot{x})d\tau, x(t)]$ minimizes $\bar{\phi}(\bar{y}(1))$ subject to

$$\bar{y}(t) \in \bar{X},$$
$$\dot{\bar{y}} \in \bar{E}(\bar{y}),$$
$$\bar{y}(0) \in \bar{C}_0, \ \bar{y}(1) \in \bar{C}_1.$$

We apply Corollary 2.1 to deduce the existence of $\bar{p} = (p_0, p)$ and λ such that $|\bar{p}(t)| + \lambda$ does not vanish, and

$$
\begin{aligned}
(5.5)\qquad & (-\dot{p}_0, -\dot{p}, L(x, \dot{x}), \dot{x}) \in \partial H(x_0, x, p_0, p) \text{ a.e.}, \\
& p(0) \in N(C_0; x(0)),\ -p(1) \in \lambda \partial \phi(x(1)) + N(C_1; x(1)), \\
& -p_0(1) = \lambda, \\
& H(x_0, x, p_0, p) = p_0 L(x, \dot{x}) + p \cdot \dot{x} = \text{constant},
\end{aligned}
$$

where

$$H(x_0, x, p_0, p) = \max \{p \cdot v + p_0 L(x, v) : v \in E(x)\}.$$

These relations imply that p_0 is identically $-\lambda$ (and so $|p(t)| + \lambda$ is nonvanishing), and consequently

$$H^*(t) = H(t) = \text{constant},$$

as required. It suffices therefore to prove that (5.5) implies (5.3), which we now proceed to do.

Suppose first that $\lambda = 1$, as we may if the problem is calm; then we have

$$(-\dot{p}, L(t), \dot{x}) \in \partial H(x, -1, p),$$

where $L(t) = L(x(t), \dot{x}(t))$. This was shown in [**9**, p. 364] to imply that $(-\dot{p}, \dot{x})$ belongs to the generalized gradient of the function $(x, p) \to H(x, -1, p)$, as required. The case $\lambda = 0$ must be treated differently, and hinges upon the fact that $\partial H^*(x, p)$ consists of the convex hull of all vectors of the form $[0, v]$, where v in $E(t)$ yields the maximum defining H^* [**2**, Theorem 2.1]. On the other hand, it follows that $\partial H(x, 0, p)$ is contained in the convex hull of all points of the form $[0, L(x, \dot{x}), v]$, see [**6**, Lemma 6]. Hence (5.3) holds.

Consider now the problem treated in Theorem 5.1 when $E(t, \cdot)$ and $L(t, \cdot, \cdot)$ are Lipschitz for each t but merely measurable in t. The same proof, with [**4**, Theorem 1] being invoked instead of Corollary 2.1, will lead to the following result, which simply eliminates from (5.1) the reference to derivatives of H^* in t: (we continue to assume either calmness or no dependence of E on x).

COROLLARY 5.1. *Let x solve locally the problem described at the beginning of this section, where we assume instead that E and L are measurable in t, and that there exists a function $k(t) \in L^1(0, 1)$ such that $E(t, \cdot)$ is Lipschitz on X (with constant $k(t)$), and such that*

$$|L(t, s_1, v_1) - L(t, s_2, v_2)| \leqq k(t) |(s_1 - s_2, v_1 - v_2)|$$

whenever s_1, s_2 *belong to* X, v_1 *and* v_2 *belong to* $E(t, s_1)$, $E(t, s_2)$ *respectively. Then there exist an arc* p, *and a scalar* λ *equal to* 0 *or* 1, *such that* $|p(t)| + \lambda$ *is nonvanishing*, (5.2) *holds, and*

$$(5.6) \qquad (-\dot{p}(t), \dot{x}(t)) \in \partial H^*(t, x(t), p(t)) \text{ a.e.},$$

where ∂H^* *refers to the generalized gradient in* (x, p).

Example. Most readers will be familiar with the importance in physics of Hamiltonian systems of differential equations:

$$(5.7) \qquad -\dot{p}(t) = H_x(x(t), p(t)), \dot{x}(t) = H_p(x(t), p(t))$$

$$(5.8) \qquad H(x(t), p(t)) = \text{constant} = c.$$

A long-standing problem concerning such systems is that of finding conditions on H guaranteeing the existence of periodic solutions for a given c (i.e., $x(\cdot)$, $p(\cdot)$ satisfying (5.7) (5.8) such that $x(0) = x(T)$, $p(0) = p(T)$ for some positive T). Significant progress on this question has recently been made by P. H. Rabinowitz [**16**] and A. Weinstein [**21**]. In [**12**] we present an approach which for the first time deals with this problem by means of a direct variational principle, one which involves a differential inclusion. As shown in [**12**], the existence problem for a general class of Hamiltonians can be reduced to the case in which H is the support function of a compact convex subset S of $\mathbf{R}^{2n}$ containing 0 in its interior:

$$H(x, p) = \max \{(x, p) \cdot (v, w) : (v, w) \in S\}.$$

The crux of the approach is the following result, a consequence of Theorem 5.1 (details and related matters may be found in [**12**]):

THEOREM. *If* H *is as above, then there exists, for some* $c > 0$, *a periodic nonvanishing solution* $(\hat{x}, \hat{p})$ *of* (5.7) (5.8).

Proof. Consider the problem of minimizing

$$\int_0^1 -p(t) \cdot \dot{x}(t) dt$$

subject to $x(0) = x(1) = 0$, $p(0) = p(1) = 0$, and

$$(-\dot{p}, \dot{x}) \in S \text{ a.e.}$$

This problem is calm, and admits a solution (x, p). The function H^* of Theorem 5.1 is easily calculated:

$$\begin{aligned} H^*(x, p, y, q) &= \max \{(y, q) \cdot (v, w) + p \cdot v : (-w, v) \in S\} \\ &= H(-q, y + p). \end{aligned}$$

(Note that the role of x is played here by (x, p), and that of p be (y, q).) Then there exist (y, q) and a constant c such that (5.3) (5.4) hold, and we have merely to set

$$\hat{x} = -q, \hat{p} = y + p$$

to obtain the desired conclusion.

6. The Erdmann condition in the calculus of variations. Let $L: \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ and $\phi: \mathbf{R}^n \to \mathbf{R}$ be locally Lipschitz. We consider the problem of minimizing

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

over the arcs x satisfying $x(0) \in C_0$, $x(1) \in C_1$ and having essentially bounded derivatives. We assume that C_0 and C_1 are closed. Below, we denote the partial generalized gradients of $L(s, v)$ by $\partial_s L$, $\partial_v L$ respectively.

THEOREM 6.1. *If x is a local solution, then there exist a constant c, and an arc p such that $(L(t)$ stands for $L(x(t), \dot{x}(t)))$*

(6.1) $\quad \dot{p}(t) \in \partial_s L(t), p(t) \in \partial_v L(t),$

(6.2) $\quad p(0) \in N(C_0; x(0)), -p(1) \in \partial\phi(x(1)) + N(C_1; x(1)),$

(6.3) $\quad L(t) - p(t) \cdot x(t) = c$ *a.e.*,

(6.4) $\quad L(x(t), \dot{x}(t) + v) - L(x(t), \dot{x}(t)) \geqq p(t) \cdot v$ *for all v, a.e.*

Remark. The above extends all the classical first-order necessary conditions of the calculus of variations. The Euler-Lagrange equation

$$\frac{d}{dt} \nabla_v L = \nabla_s L$$

is represented via (6.1). The first Erdmann condition (that $\nabla_v L(t)$ is continuous) corresponds to the continuity of $p(t)$, and the second to (6.3). The Weierstrass and Legendre conditions appear in (6.4), while (6.2) generalizes the natural boundary (or transversality) conditions. The theorem may be used (as in § 2.3) to derive an extended form of the above when L has Lipschitz dependence on t. We shall only sketch the proof of the theorem, since it uses techniques not unlike those employed in § 2. The first step is to establish (cf. Lemma 2.1)

LEMMA 6.1. *The arc $(0, x)$ minimizes*

$$\phi(y(1)) + \int_0^1 L(y, \dot{y}/(1 + \dot{y}_0))(1 + \dot{y}_0)dt$$

over all arcs (y_0, y) *having* $\dot{y}$ *essentially bounded*, $|\dot{y}_0| \leqq 1/2$, $y_0(0) = y_0(1) = 0$, $y(0) \in C_0$, $y(1) \in C_1$.

Next, we apply the necessary conditions of [**6**] to the above problem. An interpretation of these leads directly to the desired conclusions.

COROLLARY 6.1. *Suppose in the above that* L *is differentiable in* v. *Then*

$$L(x(t), \dot{x}(t)) - \dot{x}(t)\nabla_v L(x(t), \dot{x}(t)) = \textit{constant}.$$

Proof. It follows from (6.4) that $p(t) = \nabla_v L(t)$ a.e.; note (6.3).

REFERENCES

1. V. G. Boltjanksii, *The maximum principle for problems of optimal steering*, Differencial'nye Uravnenija *9* (1973), 1363–1370.

2. F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Soc. *205* (1975), 247–262.

3. ——— *The Euler-Lagrange differential inclusion*, J. Differential Equations *19* (1975), 80–90.

4. ——— *Necessary conditions for a general control problem*, in *Calculus of variations and control theory*, edited by D. Russell (Mathematics Research Center, Pub. No. *36*, University of Wisconsin, September 1975), Academic Press (1976).

5. ——— *On the inverse function theorem*, Pacific J. Math. *64* (1976), 98–102.

6. ——— *The maximum principle under minimal hypotheses*, SIAM J. Control and Optimization *14* (1976), 1078–1091.

7. ——— *The generalized problem of Bolza*, SIAM Journal of Control and Optimization *14* (1976), 682–699.

8. ——— *A New approach to Lagrange multipliers*, Math. Operations Research (1976), 165–174.

9. ——— *Extremal arcs and extended Hamiltonian systems*, Trans. Amer. Math. Soc., *231* (1977), 349–367.

10. ——— *Generalized gradients of Lipschitz functionals*, Tech. Report, Math. Res. Center, #1687 (1976). To appear in Advances in Math.

11. ——— *Nonsmooth analysis and optimization*, Proceedings of the International Congress of Mathematicians, Helsinki (1978).

12. ——— *Periodic solutions of Hamiltonian inclusions* (1978), Journal of Differential Equations, to appear.

13. R. P. Fedorenko, *A maximum principle for differential inclusions*, Z. Vycisl. Mat. i Mat. Fiz *10* (1970), 1385–1393. USSR Comp. Math. and Math. Phys. *10* (1970), 57–68.

14. A. F. Filippov, *Differential equations with discontinuous right-hand side*, Amer. Math. Soc. Trans. (2) *42* (1964), 199–231.

15. H. Halkin, *Optimization without differentiability*, to appear in the Proceedings of the Conference on Optimal Control Theory, Canberra, August (1977), Springer-Verlag.

16. P. H. Rabinowitz, *Periodic solutions of Hamiltonian systems*, Comm. Pure and Applied Math. *31* (1978), 157–184.

17. R. T. Rockafellar, *Conjugate convex functions in optimal control and the calculus of variations*, J. Math. Anal. Appl. *34* (1970), 174–222.

18. ——— *Generalized Hamiltonian equations for convex problems of Lagrange*, Pacific J. Math. *33* (1970), 411–427.

19. A. J. Ward, *On the differential structure of real functions*, Proc. London Math. Soc. (2) *39* (1935), 339–362.

20. J. Warga, *Necessary conditions without differentiability assumptions in optimal control*, J. Differential Equations *18* (1975), 41–62.

21. A. Weinstein, *Periodic orbits for convex Hamiltonian systems*, Annals of Math., to appear.

University of British Columbia,
Vancouver, British Columbia