JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **153**, 37–51 (1990)

Coextremals and the Value Function for Control Problems with Data Measurable in Time

R. B. VINTER AND P. R. WOLENSKI

Department of Electrical Engineering, Imperial College, Exhibition Road, London SW7 2BZ, England

Submitted by Frank H. Clarke

Received June 14, 1989

Suppose $x^*(\cdot)$ is a solution to an optimal control problem formulated in terms of a differential inclusion. Known first-order necessary conditions of optimality assert existence of a coextremal, or adjoint function, $p(\cdot)$, which together with $x^*(\cdot)$ satisfies the Hamiltonian inclusion and associated transversality condition. In this paper we interpret extremals in terms of generalized gradients of the value function V by demonstrating that $p(\cdot)$ can in addition be chosen to satisfy

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

The hypothesis imposed are more or less the weakest under which the Hamiltonian inclusion condition is known to apply and permit, in particular, measurable time dependence of the data. The proof of the results relies on recent developments in Hamilton Jacobi theory applicable in such circumstances. An analogous result is proved for problems where the dynamics are modelled by a differential equation with control term.

1. INTRODUCTION

Two major elements in optimal control theory are first-order necessary conditions (exemplified by the Pontryagin maximum principle or Clarke's Hamiltonian inclusion) and dynamic programming. We briefly review these in relation to a free right endpoint problem, labelled (P), where the dynamics are modelled by a differential inclusion:

$$\begin{aligned}
&\text{Minimize } g(x(1)) \\
&\text{over arcs } x \in W^{1,1}(0, 1; \mathbb{R}^n) \\
&\text{which satisfy} \\
&\dot{x}(t) \in F(t, x(t)), \qquad \text{a.e. } [0, 1] \\
&x(t_0) = x_0.
\end{aligned}$$

0022-247X/90 $3.00

Suppose $x^*(\cdot)$ is a minimizer. On one hand, Clarke [2] has shown that there exists an absolutely continuous function $p(\cdot)\colon [0, 1] \to \mathbb{R}^n$, named a coextremal, such that

$$(-\dot{p}(t), \dot{x}^*(t)) \in \partial H(t, x^*(t), p(t)), \qquad \text{a.e. } [0, 1] \tag{1.1}$$

and

$$-p(1) \in \partial g(x^*(1)). \tag{1.2}$$

Here H is the Hamiltonian function

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

∂g denotes the (Clarke) generalized gradient, and ∂H is the generalized gradient of $H(t, x, p)$ in the x, p variables.

On the other hand, dynamic programming concerns properties of the value function V. The value function $V(\tau, \xi)$, evaluated at $(\tau, \xi) \in [0, 1] \times \mathbb{R}^n$, is the infimum cost of a modified version of problem (P) in which $[\tau, 1]$ replaces the underlying time interval $[0, 1]$ and ξ replaces x_0 as the initial state.

Control theorists have long been familiar with heuristic arguments which suggest a relationship between the two theories. Expressed in terms of the differential inclusion problem (P), it asserts

$$(h(t), -p(t)) = \nabla V(t, x^*(t)), \tag{1.3}$$

where $h(t)$ is the Hamiltonian function evaluated along $x^*(t)$ and its coextremal $p(t)$:

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

As is well known, the assumption that V is continuously differentiable is not justified outside a narrow class of problems (see [6]). Thus (1.3) does not make sense, strictly speaking, in a framework of any generality. However, one may reasonably suppose that V will be locally Lipschitz continuous. It is natural then to seek in place of (1.3) a relationship involving the generalized gradient ∂V of V and the coextremal

$$(h(t), -p(t)) \in \partial V(t, x^*(t)), \qquad \text{a.e. } [0, 1]. \tag{1.4}$$

If we supplement the hypotheses under which the necessary conditions (1.1) and (1.2) have been proved by the extra hypotheses ($\overline{\text{H1}}$) and ($\overline{\text{H2}}$), then it is easy to establish the validity of (1.4).

($\overline{\text{H1}}$) F is locally Lipschitz continuous in both the t and the x variables, and

($\overline{\text{H2}}$) for each $t \in [0, 1]$, there is at most one absolutely continuous function p_t: $[t, 1] \to \mathbb{R}^n$ satisfying

$$(-\dot{p}_t(s), \dot{x}^*(s)) \in \partial H(s, x^*(s), p_t(s)), \qquad \text{a.e. } [t, 1] \tag{1.5}$$

$$-p_t(1) \in \partial g(x^*(1)). \tag{1.6}$$

We must just note that, for each $t \in (0, 1)$, the time $\tau = t$ and the restriction of $x^*(\cdot)$ to $[t, 1]$ provide a solution to the problem (labelled (P_t)):

Minimize $g(x(1)) + V(\tau, x(\tau))$

over times $\tau \in (0, 1)$ and arcs $x(\cdot) \in W^{1,1}(\tau, 1; \mathbb{R}^n)$

which satisfy

$\dot{x}(s) \in F(s, x(s))$, a.e. $[\tau, 1]$.

(This fact is easily deduced from the definition of the value function.) Note that, in passing from (P) to (P_t), besides replacing the time interval $[0, 1]$ by $[\tau, 1]$ in which τ is a choice variable, we have substituted an initial cost term $V(\tau, x(\tau))$ for the initial state condition $x(0) = x_0$. Under hypotheses which include ($\overline{\text{H1}}$), known "free time" necessary conditions for problems such as (P_t) are available [2] which give information about $x^*(\cdot)$. They tell us that, for each $t \in (0, 1)$ there exists an absolutely continuous function p_t: $[t, 1] \to \mathbb{R}^n$ such that (1.5) and (1.6) are satisfied and

$$(h(t), -p_t(t)) \in \partial V(t, x^*(t)). \tag{1.7}$$

Now take $p(\cdot)$ to be the coextremal in (1.1) and (1.2). Assuming ($\overline{\text{H2}}$) we have that, for each $t \in (0, 1)$, the restriction of $p(\cdot)$ to $[t, 1]$ coincides with p_t; the desired inclusion (1.4) now follows from (1.7). Versions of this argument were given in a specialized setting by Barbu [1] and are implicit in the material in [2, Chap. 3].

Our objective in this paper is to prove the inclusion (1.4) under more or less the weakest hypotheses invoked in the proof of the necessary conditions (1.1), (1.2) and for which V is known to be Lipschitz continuous. They are

(H1) F is compact valued, $F(\cdot, x)$ is measurable for each $x \in \mathbb{R}^n$ and g is locally Lipschitz continuous;

(H2) there exists an integrable function λ such that

$$F(t, x) \subset F(t, y) + \lambda(t)\,|y - x|\, B, \qquad \text{for all} \quad x, y \in \mathbb{R}^n \quad \text{and} \quad t \in [0, 1]$$

(B denotes the open unit ball); and

(H3) there exists a constant K such that

$$F(t, x) \subset K(1+|x|)\, B, \qquad \text{for all} \quad t \in [0, 1], \qquad x \in \mathbb{R}^n.$$

Note in particular that no conditions akin to ($\overline{\text{H2}}$) are imposed; this is important, since we cannot expect, outside special cases, that the costate differential inclusion will have a unique solution.

Here the simple arguments outlined above justifying the inclusion (1.4) break down. For one thing, free time optimality conditions for problem (P_t), of the type required, are not available for problem (P_t) when F is a *measurable* multifunction in t (although new optimality conditions in [3] give some information about minimizers for (P_t)). For another, there may be many solutions $(x^*(\cdot), p(\cdot))$ to the Hamiltonian inclusion (1.1) and transversality condition (1.2), and we do not know that $p(t)$ can be matched up to $p_t(t)$.

Clearly then more subtle arguments are required. Our approach is based on new results of Vinter and Wolenski [8] which establish that, under hypotheses somewhat weaker than (H1)–(H3), V is a "lower Dini" solution of the Hamiltonian Jacobi equation. This fact is used to construct a family of auxiliary problems involving V, each having solution $x^*(\cdot)$, whose coextremals satisfy the desired inclusion (1.4) in the limit.

The relationship we establish between the value function and coextremals in this paper improves upon those obtainable by the methods of [7], which require F to be continuous with respect to the time variable, and upon those of [4], which merely establish "$-p(t) \in \partial_x V$" and supply no information about $h(t)$.

2. The Main Result

THEOREM 2.1. *Let $x^*(\cdot)$ be a solution to problem* (P) *and suppose that hypotheses* (H1)–(H3) *are in force. Then there exists an absolutely continuous function $p(\cdot)$: $[0, 1] \to \mathbb{R}^n$ such that*

$$(-\dot{p}(t), \dot{x}^*(t)) \in \partial H(t, x^*(t), p(t)), \qquad \text{a.e. } [0, 1] \tag{2.1}$$

$$-p(1) \in \partial g(x^*(1)) \tag{2.2}$$

$$-p(0) \in \partial_x V(0, x_0)$$

and

$$(h(t), -p(t)) \in \partial V(t, x^*(t)), \qquad \text{a.e. } [0, 1],$$

where $h(t)$ is the function

$$h(t) := \max\{p(t) \cdot e : e \in F(t, x^*(t))\}$$

and ∂H denotes the generalized gradient of $H(t, x, p)$ in the (x, p) variables.

Existence of a function $p(\cdot)$ satisfying (2.1) and (2.2) constitutes familiar optimality conditions, of "Hamiltonian inclusion" type. These are derivable from [2, Thm. 3.2.6] in view of the fact that (P) is a free right endpoint problem. Since g is continuous and state trajectories for the convexified differential inclusion $\dot{x} \in \text{co}\, F$ can be uniformly approximated by state trajectories for $\dot{x} \in F$, this means that $x^*(\cdot)$ remains a solution when $F(t, x)$ in problem (P) is replaced by $\text{co}\, F(t, x)$ (see, for example, [2, p. 117]); the right side of the differential inclusion convex is thereby rendered convex, as is required for satisfaction of the hypotheses in [2, Thm. 3.2.6]. It also means that the transversality condition (2.2) is expressible in normalized form; i.e., the cost multiplier λ in [2, Thm. 3.2.6] can be set to 1.

The novel component of this theorem is of course the assertion that $p(\cdot)$ can be chosen to satisfy the stated relationship involving the value function, in addition to (2.1) and (2.2).

3. Properties of the Value Function

Given an open set $A \subset \mathbb{R}^n$, a function $\psi: A \to \mathbb{R}$, and points a in A and u in $\mathbb{R}^n$, we define the lower Dini derivative of ψ at a in the direction u, written $d^-\psi(a; u)$, as

$$d^-\psi(a; u) := \liminf_{h \downarrow 0} \frac{1}{h} [\psi(a + hu) - \psi(a)]. \tag{3.1}$$

We make use of properties of the value function, summarized in the following proposition:

Proposition 3.1. *Let V be the value function, as defined in Section* 1, *and assume hypotheses* (H1)–(H3). *Then V is a (finite valued) locally Lipschitz continuous function, and there exists a set $J \subset [0, 1]$ of zero measure such that, for all $t \notin J$ and $\xi \in \mathbb{R}^n$, we have*

$$\min_{v \in F(t, \xi)} d^- V((t, \xi); (1, v)) \geqslant 0. \tag{3.2}$$

The demonstration that V is locally Lipschitz continuous under the stated hypotheses is easily accomplished, along the lines of the proof of [5, Thm. 4.2]. That V satisfies the Hamilton Jacobi inequality (3.2) is implied by the somewhat stronger assertions of [8, Thm. 2.3].

4. Proof of Theorem 2.1

Proof of the theorem hinges on consideration of an appropriate auxiliary differential inclusion. It is constructed from data for problem (P) and generalized gradients of the value function.

For any $\varepsilon \in (0, 1)$ define the set valued function $G_\varepsilon(\cdot)$ on $[0, 1]$ and the function $\sigma_\varepsilon: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ as

$$G_\varepsilon(t) := \{(\alpha, \beta) \in \partial V(s, y): 0 \leqslant s \leqslant 1, (s, y) \in (t, x^*(t)) + \varepsilon \bar{B}\} \tag{4.1}$$

and

$$\sigma_\varepsilon(t, v, w) := \sup\{(\alpha, \beta) \cdot (w, -(1+w)v): (\alpha, \beta) \in G_\varepsilon(t)\}. \tag{4.2}$$

Here $x^*(\cdot)$ is the solution to (P) under consideration.

Observe the following boundedness and continuity properties of the function σ_ε, which are straightforward consequences of the fact that the graph of ∂V, viewed as a multifunction of its base point, has closed graph (see [2, Prop. 2.1.5]).

LEMMA 4.1. *For $\varepsilon > 0$, the function $\sigma_\varepsilon(t, v, w)$ is upper semicontinuous, continuous in (v, w) for fixed t, measurable in t for fixed (v, w) and bounded on bounded subsets of* $[0, 1] \times \mathbb{R}^n \times \mathbb{R}$.

The auxiliary differential inclusion is

$$(\dot{y}, \dot{x}) \in F_\varepsilon(t, (y, w)), \tag{4.3}$$

where

$$F_\varepsilon(t, (y, x)) := \{(\sigma_\varepsilon(t, v, w), (e+v)(1+w)): e \in F(t, x), v \in \varepsilon \bar{B}, w \in \varepsilon \bar{B}\}.$$

Associated with (4.3) is the cost function

$$J_\varepsilon(y(\cdot), x(\cdot)) := g(x(1)) - V(0, x(0)) + y(1) - y(0).$$

We now show that $(x^*(\cdot), y(\cdot) \equiv 0)$ supplies a solution to the problem of minimizing J_ε over arcs satisfying the differential inclusion (4.3).

LEMMA 4.2. *For any $\varepsilon \in (0, 1)$ let (y, x) be an absolutely continuous function which satisfies*

$$|x(t) - x^*(t)| < \varepsilon \qquad \textit{for all} \quad t \in [0, 1] \tag{4.4}$$

and the differential inclusion (4.3). *Then*

$$J_\varepsilon(0, x^*(\cdot)) \leqslant J_\varepsilon(y(\cdot), x(\cdot)).$$

Proof. We make use of the following "chain rule" which in turn follows from elementary properties of locally Lipschitz continuous functions applied at points t where the absolutely continuous function $s \to \phi(s, x(s))$ (introduced below) is differentiable, and which are Lebesgue points of

$\dot{x}$: "suppose that $\phi: \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ is a locally Lipschitz continuous function and $x: [0, 1] \to \mathbb{R}^n$ is absolutely continuous. Then

$$\frac{d}{dt}\phi(t, x(t)) = d^-\phi(t, x(t); (1, \dot{x}(t)) \qquad \text{a.e.}"$$

Here, as usual, $d^-\phi$ denotes the lower Dini derivative (3.1).

Let $(y(\cdot), x(\cdot))$ be as in Lemma 4.2. In view of the hypotheses on F and the assertions of Lemma 4.1, we may appeal to a measurable selection theorem [2, p. 111] to obtain measurable functions $e(t)$, $v(t)$ and $w(t)$, and a set $S_1 \subset [0, 1]$ of zero measure such that

$$\dot{y}(t) = \sigma_\varepsilon(t, v(t), w(t)),$$

$$\dot{x}(t) = (e(t) + v(t))(1 + w(t)), \tag{4.5}$$

$$e(t) \in F(t, x(t)), \tag{4.6}$$

$$\|v(t)\| \leqslant \varepsilon \qquad \text{and} \qquad |w(t)| \leqslant \varepsilon,$$

for all $t \in [0, 1]\backslash S_1$.

Let S_2 be the null set on which the Hamilton Jacobi equation (3.2) is not satisfied and set $S = S_1 \cup S_2$. Then for $t \in [0, 1)\backslash S$ we have (evaluating the functions x, e, and w at t)

$$0 \leqslant \liminf_{h \downarrow 0} [V(t+h, x+he) - V(t, x)] h^{-1}(1+w)$$

by (3.1) and (3.2) and since $1 + w \geqslant 0$,

$$= \liminf_{h \downarrow 0} [V(t+h(1+w), x+he(1+w)) - V(t, x)] h^{-1}$$

by positive homogeneity

$$\leqslant \liminf_{h \downarrow 0} [V(t+h, x+h(e+v)(1+w)) - V(t, x)] h^{-1}$$

$$+ \limsup_{h \downarrow 0} [V(t'+hw, x'-hv(1+w)) - V(t', x')] h^{-1},$$

where $t' = t + h$ and $x' = x + h(e+v)(1+w)$

$$\leqslant d^- V(t, x; 1, (e+v)(1+w))$$

$$+ \limsup [V(t'+hw, x'-hv(1+w)) - V(t', x')] h^{-1},$$

where now the lim sup is taken over $h \downarrow 0$, $t' \to t$, $x' \to x$

$$= d^- V(t, x; 1, (e+v)(1+w)) + D^0 V(t, x; w, -v(1+w)).$$

Here D^0V denotes the Clarke generalized directional derivative [2]. It follows now from (4.1), (4.4), and the characterization of D^0V in terms of subgradients provided by [2, p. 10] that

$$D^0V(t, x; w, -v(1+w)) \leqslant \sup\{(\alpha, \beta)\cdot(w, -v(1+w): (\alpha, \beta) \in G_\varepsilon(t)\}$$
$$= \sigma_\varepsilon(t, v, w)$$

by (4.2). Putting these estimates together, and noting that S is a null set, we obtain the inequality

$$d^- V(t, x(t); 1, (e(t)+v(t))(1+w)) + \sigma_\varepsilon(t, v(t), w(t)) \geqslant 0 \qquad \text{a.e..} \qquad (4.7)$$

We calculate

$$J_\varepsilon(y, x) = g(x(1)) - V(0, x(0)) + \int_0^1 \sigma_\varepsilon(t, v(t), w(t))\, dt$$

$$= V(1, x(1)) - V(0, x(0)) + \int_0^1 \sigma_\varepsilon\, dt$$

$$= \int_0^1 \frac{d}{dt} V(t, x(t)\, dt + \int_0^1 \sigma_\varepsilon\, dt$$

since $t \to V(t, x(t))$ is absolutely continuous,

$$= \int_0^1 \{[d^- V(t, x(t); 1, (e(t)+v(t))(1+w(t))] + \sigma_\varepsilon(t, v(t), w(t))\}\, dt$$

by the chain rule referred to above, and

$$\geqslant 0$$

by (4.7). This establishes that

$$J_\varepsilon(y(\cdot), x(\cdot)) \geqslant 0$$

for the arbitrary arc$(y(\cdot), x(\cdot))$ satisfying the auxiliary differential inclusion. Proof of the lemma is concluded by noting that $V(1, x^*(1)) = V(0, x^*(0))$ by the properties of the value function, and $\sigma_\varepsilon(\cdot, v \equiv 0, w \equiv 0) \equiv 0$, whence

$$J_\varepsilon(0, x^*(\cdot)) = 0.$$

Lemma 4.2 tells us that, for each $\varepsilon \in (0, 1)$, $(y(\cdot) \equiv 0, x^*(\cdot))$ is a local solution to the problem

$$\text{Minimize } J_\varepsilon(y(\cdot), x(\cdot))$$
$$\text{subject to}$$
$$(\dot y(t), \dot x(t)) \in F_\varepsilon(t, (y(t), x(t))), \qquad \text{a.e.}$$

(The solution is "local" in the sense that it is minimizing with respect to all arcs$(y(\cdot), x(\cdot))$ satisfying $\|x(\cdot)-x^*(\cdot)\|_C<\varepsilon$.) Now $(y(\cdot)\equiv 0, x^*(\cdot))$ remains a local solution when we replace F_ε by its convex hull. This is because J_ε is continuous with respect to the supremum norm and state trajectories for $\dot{z}\in \text{co}\, F_\varepsilon$ can be uniformly approximated by those for $\dot{z}\in F_\varepsilon$.

Following convexification, the hypotheses under which we may apply the necessary conditions [2, Thm. 3.2.6] at the minimizer $(y(\cdot)\equiv 0, x^*(\cdot))$ are now satisfied. Note that we are justified in setting the cost multiplier $\lambda=1$, since the problem has a free right endpoint. This tells us that, for each $t\in(0,1)$, there exist absolutely continuous functions $p'(\cdot)$ and $q'(\cdot)$ such that

$$(-\dot{q}'(t), -\dot{p}'(t), \dot{y}(t)\equiv 0, \dot{x}^*(t))\in \partial H_\varepsilon(t, (y(t)\equiv 0, x^*(t)), (q'(t), p'(t)) \quad \text{a.e.} \tag{4.8}$$

$$p'(0)\in -\partial_x V(0, x^*(0)), \qquad -p'(1)\in\partial g(x^*(1)), \qquad q'(1)=q'(0)=-1. \tag{4.9}$$

Here $H_\varepsilon\colon \mathbb{R}\times\mathbb{R}^{n+1}\times\mathbb{R}^{n+1}\to\mathbb{R}$ is the Hamiltonian function of the auxiliary differential inclusion, that is,

$$H_\varepsilon(t, (y, x), (q, p)) := \max\{q\sigma_\varepsilon(t, v, w) + p\cdot(e+v)(1+w)\}$$

in which the maximum is taken over $e\in F(t, x)$, $v\in\varepsilon\bar{B}$, and $w\in\varepsilon\bar{B}$. In (4.8), the generalized gradient is taken jointly with respect to the (y, x), (q, p) variables.

Bearing in mind that $(1+w)>0$ for all $w\in\varepsilon\bar{B}$, we can decompose H_ε:

$$H_\varepsilon(t, (y, x), (q, p)) = H(t, x, p) + f_\varepsilon(t, x, q, p). \tag{4.10}$$

Here, the function f_ε is defined to be

$$f_\varepsilon(t, x, q, p) := \max_{v, w}\{H(t, x, p)w + (1+w)\, p\cdot v + q\sigma_\varepsilon(t, v, w)\}.$$

It is a simple matter to show that there exists a constant K_1 with the property

$$|\mathscr{P}_{x,p}\,\partial f_\varepsilon(t, x^*(t), q, p)| \leqslant K_1\varepsilon(|p|+1) \tag{4.11}$$

for all $t\in[0, 1]$, $p\in\mathbb{R}^n$, $q\in\mathbb{R}^n$. Here ∂f_ε denotes the generalized gradient of $f_\varepsilon(t, x, q, p)$ with respect to the x, q, p variables and $\mathscr{P}_{x,p}$ denotes projection onto the x, p variables.

It follows now from (4.8)–(4.11) and basic properties of generalized gradients that

$$(-\dot{p}'(t), \dot{x}^*(t))\in\partial H(t, x^*(t), p'(t)) + K_1\varepsilon(|p'(t)|+1)\bar{B} \tag{4.12}$$

and

$$q'(\cdot) \equiv -1.$$

It is known [2, Thm. 3.2.4] that the Hamiltonian inclusion (4.8) implies that, almost everywhere, values of the function $(\dot{y}(\cdot) \equiv 0, \dot{x}^*(\cdot))$ achieve the maximum in the definition of the Hamiltonian function, i.e.,

$$H_\varepsilon(t, (y(t)=0, x^*(t)), (q'(t), p'(t)) = p'(t) \cdot \dot{x}^*(t) \qquad \text{a.e.}$$

Recalling the definition of $\sigma_\varepsilon(t, v, w)$ (see (4.2)), we deduce from this equation and our knowledge that $q'(\cdot) \equiv -1$ that

$$\inf_{(\alpha, \beta) \in G_\varepsilon(t)} \{wH(t, x^*(t), p'(t)) + (1+w)\, p'(t) \cdot v - \alpha w + \beta \cdot v(1+w)\} \leqslant 0$$

for all $w \in \varepsilon B$, $v \in \varepsilon B$.

Since $(1+w) > 0$ for $w \in \varepsilon B$, we preserve the inequality if, inside the brackets $\{\cdots\}$, we divide by $(1+w)$. Making the substitutions $w' = w/(1+w)$ and $v' = -v$, we obtain

$$H(t, x^*(t), p'(t))\, w' - p'(t) \cdot v' \leqslant \sup\{\alpha w' + \beta v' : (\alpha, \beta) \in G_\varepsilon(t)\}$$

for all points (w', v') in some ball in $\mathbb{R}^n$ about the origin. A simple application of the separation theorem now gives

$$(H(t, x^*(t), p'(t)), -p'(t)) \in \overline{\text{co}}\, G_\varepsilon(t). \tag{4.13}$$

Up to this point we have treated $\varepsilon \in (0, 1)$ as arbitrary. Now let $\{\varepsilon_i\}$ be a sequence in $(0, 1)$, converging to zero. We follow through the preceding steps with ε_i replacing ε, for $i = 1, 2, \ldots$. Write $p_i(\cdot)$ in place of the coextremal $p(\cdot)$ corresponding to $\varepsilon = \varepsilon_i$. The $p_i(\cdot)$'s are uniformly bounded and equicontinuous. So by replacing the original sequence with a subsequence if necessary, we can arrange that the $p_i(\cdot)$'s converge uniformly to some absolutely continuous function $p(\cdot)$.

Along the sequence, (4.13) is satisfied with ε_i taking the place of ε and $p_i(\cdot)$ that of $p(\cdot)$. Passage to the limit, $i \to \infty$, gives

$$(H(t), -p(t)) \in \bigcap_{\delta > 0} \overline{\text{co}} \cup \{\partial V(s, y) : (s, y) \in (t, x^*(t)) + \delta \bar{B}, 0 \leqslant s \leqslant 1\}$$

for all $t \in T$, where $T \subset [0, 1]$ is a set of full measure. $H(t, x^*(t), p(t))$ has been written briefly $H(t)$. We wish to show that

$$(H(t), -p(t)) \in \partial V(t, x^*(t)), \qquad \text{for all} \quad s \in T. \tag{4.14}$$

If this were not true we could strictly separate the point $(H(t), -p(t))$ and the closed convex set $\partial V(t, x^*(t))$. In other words there exist $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^n$, and $\gamma > 0$ such that

$$\alpha H(t) - p(t) \cdot \beta - \gamma > \max\{\alpha\tau + \xi \cdot \beta : (\tau, \xi) \in \partial V(t, x^*(t))\}$$
$$= D^0 V((t, x^*(t)); (\alpha, \beta)). \tag{4.15}$$

We have used here the fact that the generalized directional derivative $D^0 V$ is the support function of the generalized gradient.

Appealing to the upper semicontinuity of $D^0 V$ in its arguments [2, Prop. 2.1.5], we can assert the existence of some $\delta_1 > 0$, such that

$$\alpha H(t) - p(t) \cdot \beta - \gamma/2 > D^0 V((s, y); (\alpha, \beta))$$

for all points $(s, y) \in (t, x^*(t)) + \delta_1 \bar{B}$ with $0 \leqslant s \leqslant 1$. It follows that

$$\alpha H(t) - p(t) \cdot \beta - \gamma/2 > \sup\{\alpha\tau + \xi \cdot \beta : (\tau, \xi) \in W(t)\},$$

where

$$W(t) := \bigcup \{\partial V(s, y) : (s, y) \in (t, x^*(t)) + \delta_1 \bar{B}, 0 \leqslant s \leqslant 1\}.$$

We conclude that

$$(H(t), -p(t)) \notin \overline{\text{co}}\, W(t)$$

in contradiction of (4.15). Equation (4.14) is proved. Finally we note that, in the limit, (4.9) and (4.12) yield

$$-p(0) \in \partial_x V(0, x^*(0)), \; -p(1) \in \partial g(x^*(1))$$

and

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

Proof of Theorem 2.1 is complete.

5. Differential Equations with Control

So far, we have adopted a differential inclusion formulation of the optimal control problem. Our methods carry over to relate value functions and coextremals when the control system dynamics are described by differential equations.

Take functions $L:[0,1]\times\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$, $h:\mathbb{R}^n\to\mathbb{R}$, $f:[0,1]\times\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^n$, a multifunction $t\to U_t$ with values subsets of $\mathbb{R}^m$, and a point $x_0\in\mathbb{R}^n$. Now consider the following minimization problem, labelled (C),

Minimize $\int_0^1 L(t, x(t), u(t))\,dt + h(x(1))$

over $x(\cdot)\in W^{1,1}(0, 1; \mathbb{R}^n)$ and measurable functions $u(\cdot)$: $[0, 1]\to\mathbb{R}^n$,

subject to

$\dot{x}(t)\in f(t, x(t), u(t))$, a.e. $[0, 1]$

$u(t)\in U_t$, a.e. $[0, 1]$

and

$x(0)=x_0$.

Define the function $\bar{f}(\cdot,\cdot,\cdot)$: $[0, 1]\times\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^{n+1}$

$$\bar{f}(t, x, u) := \mathrm{col}\{L(t, x, u), f(t, x, u)\}.$$

We invoke the following hypotheses:

(C1) $\bar{f}(t, x, u)$ is measurable in t for fixed $(x, u)\in\mathbb{R}^n\times\mathbb{R}^m$ and continuous in u for fixed $(t, x)\in[0, 1]\times\mathbb{R}^n$, and h is Lipschitz continuous;

(C2) there exists $\alpha(\cdot)\in L^1$ such that

$$|\bar{f}(t, x, u)-\bar{f}(t, y, u)|\leqslant\alpha(t)\,|x-y| \quad \text{for all } x, y\in\mathbb{R}^n, u\in U_t \text{ a.e. } t\in[0, 1];$$

(C3) there exists a constant K such that

$$|\bar{f}(t, x, u)|\leqslant K(1+|x|) \qquad \text{for all} \quad x\in\mathbb{R}^n, u\in U_t \qquad \text{a.e.} \quad t\in[0, 1];$$

and

(C4) U_t is compact for all $t\in[0, 1]$ and graph $\{t\to U_t\}$ is Borel measurable.

The value function $W(\cdot,\cdot)$: $[0, 1]\times\mathbb{R}^n\to\mathbb{R}$ we associate with problem (C) is the following: for $(\tau, \xi)\in[0, 1]\times\mathbb{R}^n$, $W(\tau, \xi)$ is the minimum cost for a related problem in which $[\tau, 1]$ replaces the underlying time interval and ξ replaces the initial condition x_0. Under the hypotheses, $W(\cdot,\cdot)$ is locally Lipschitz continuous.

Define the function $\tilde{H}(\cdot,\cdot,\cdot,\cdot)$: $[0, 1]\times\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ to be

$$\tilde{H}(t, x, u, p) := p\cdot f(t, x, u) - L(t, x, u).$$

THEOREM 5.1. *Let* $(x^*(\cdot), u^*(\cdot))$ *solve* (C). *Assume hypotheses* (C1)–(C4) *are satisfied. Then there exists an absolutely continuous function* $p(\cdot)\colon [0,1] \to \mathbb{R}^n$ *such that*

$$-\dot{p}(t) \in \partial_x \tilde{H}(t, x^*(t), p(t), u^*(t)) \qquad a.e.$$

$$\tilde{H}(t, x^*(t), p(t), u^*(t)) = \max_{v \in U_t} \tilde{H}(t, x^*(t), p(t), v) \qquad a.e.$$

and

$$-p(1) \in \partial h(x(1)).$$

Furthermore $p(\cdot)$ *can be chosen so that*

$$-p(0) \in \partial_x W(0, x_0)$$

and

$$(\tilde{h}(t), -p(t)) \in \partial W(t, x^*(t)) \qquad a.e.$$

Here $\tilde{h}(\cdot)$ *is the function*

$$\tilde{h}(t) := p(t) \cdot f(t, x^*(t), u^*(t)) - L(t, x^*(t), u^*(t)).$$

Once again we have asserted that functions $p(\cdot)$ exist which are coextremals, i.e., they meet the requirements of the appropriate necessary conditions (in this case the Pontryagin maximum principle), and which are, at the same time, related to gradients of the value function in the desired manner.

Proof. We associate with (C) the following differential inclusion problem, labelled ($\tilde{\mathrm{P}}$):

Minimize $y(1) + h(x(1))$

over arcs $(y(\cdot), x(\cdot)) \in W^{1,1}(0, 1; \mathbb{R}^{1+n})$

subject to

$(\dot{y}(t), \dot{x}(t)) \in \{(L(t, x, w), f(t, x, w))\colon u \in U_t\}$

and

$(y(0), x(0)) = (0, x_0)$.

The arc $x^*(\cdot)$ solves problem ($\tilde{\mathrm{P}}$), as follows from application of an appropriate measurable selection theorem [2, p. 111].

Let $\tilde{V}(t, (y, x))$ be the value function for ($\tilde{\mathrm{P}}$), in the sense of Section 1. It is easy to see that $\tilde{V}$ is related to the value function W referred to in Theorem 5.1 according to

$$\tilde{V}(t, (y, x)) = W(t, x) + y, \qquad \text{for all} \quad (t, y, x) \in [0, 1] \times \mathbb{R} \times \mathbb{R}^n. \tag{5.1}$$

In conseqence of hypotheses (C1)–(C4), the data for problem ($\tilde{\mathrm{P}}$) satisfy hypotheses (H1)–(H3). The conclusions of Lemma 4.2 are available to us and, along with some calculations involving (5.1) and application of a measurable selection theorem, yield the following information: for arbitrary $\varepsilon > 0$, $(x^*(\cdot), (u^*(\cdot), v(\cdot) \equiv 0, w(\cdot) \equiv 0))$ solves the following optimal control problem (Q_ε):

Minimize $\int_0^1 L(t, x(t), u(t))\,[1 + w(t)]\,dt + h(x(1)) + \int_0^1 \sigma_\varepsilon(t, v(t), w(t))\,dt$ over arcs $x(\cdot) \in W^{1,1}(0, 1; \mathbb{R}^n)$ and measurable functions $u(\cdot)\colon [0, 1] \to \mathbb{R}^m$, $v(\cdot)\colon [0, 1] \to \mathbb{R}^n$, and $w(\cdot)\colon [0, 1] \to \mathbb{R}$, subject to

$$\dot{x}(t) = [f(t, x(t), u(t)) + v(t)](1 + w(t)) \qquad \text{a.e.}$$

$$(u(t), v(t), w(t)) \in U_t \times (\varepsilon \bar{B}) \times (\varepsilon \bar{B}) \qquad \text{a.e.}$$

and

$$\|x(\cdot) - x^*(\cdot)\|_C \leqslant \varepsilon_1.$$

Here $\varepsilon_1 > 0$ is a number whose magnitude is determined by ε and the function $\alpha(\cdot)$ in hypothesis (C2). The function $\sigma_\varepsilon(\cdot, \cdot, \cdot)$ is as defined by (4.1) and (4.2).

From this point we proceed along lines very similar to the analysis in Section 4, except that we apply a nonsmooth version of the Pontryagin maximum principle to problem (Q_ε) [2, Thm.] with reference to the minimizer $(x^*(\cdot), (u^*(\cdot), v(\cdot) \equiv 0, w(\cdot) \equiv 0))$, in place of the previously used Hamiltonian inclusion conditions. Detailed examination of the maximum principle conditions and passage to the limit, $\varepsilon \downarrow 0$, gives the assertions of Theorem 5.1. The analysis, which is in fact simpler than that required for the differential inclusion problem, was previously carried out in [7, pp. 103–105] under extra hypotheses on the data, extra hypotheses, however, which have no role in the arguments involved at this stage of the proof.

References

1. V. Barbu, Optimal control of variational inequalities, *in* "Research Notes in Mathematics," p. 100, Pitman, London, 1984.
2. F. H. Clarke, "Optimization and Nonsmooth Analysis," Wiley, New York, 1983.
3. F. H. Clarke, P. D. Loewen, and R. B. Vinter, Differential inclusions with free time, *Ann. Inst. H. Poincaré* **5** (1988), 573–593.
4. F. H. Clarke and R. B. Vinter, The relationship between the maximum principle and dynamic programming, *SIAM J. Control Optim.* **25** (1987), 1291–1311.

5. W. H. FLEMING AND R. W. RISHEL, "Deterministic and Stochastic Optimal Control," Springer-Verlag, New York, 1975.
6. L. S. PONTRYAGIN *et al.*, "The Mathematical Theory of Optimal Processes," Interscience, New York, 1962.
7. R. B. VINTER, New results on the relationship between dynamic programming and the maximum principle, *Math. Control. Signals Systems* **1** (1988), 97–105.
8. R. B. VINTER AND P. WOLENSKI, Hamilton Jacobi theory for optimal control problems with data measurable in time, *SIAM J. Control Optim.*, to appear.