The maximum principle in optimal control, then and now*

Francis Clarke †

October 22, 2004

Abstract

We discuss the evolution of the Pontryagin maximum principle, focusing primarily on the hypotheses required for its validity. We proceed to describe briefly a unifying result giving rise to both classical and new versions, a recent theorem of the author giving necessary conditions for optimal control problems formulated in terms of differential inclusions. We conclude with a new application of this result for the case in which mixed constraints on the state and control are imposed in terms of equalities, inequalities, and unilateral set constraints. In order to lighten the exposition, the discussion is limited to differentiable data, thereby avoiding mention of generalized gradients or normal cones, except in the technical section on differential inclusions.

1 Three maximum principles

1.1 The classical maximum principle

Our purpose in this introductory section is to discuss the celebrated maximum principle of Pontryagin and some of its variants. We begin in a setting which is essentially the one in which it was first developed, in the fixed-time case. The dynamics on the (prescribed) interval $[0, T]$ are given by

$$\dot{x}(t) = f(t, x(t), u(t)) \text{ a.e.},$$

where the measurable function u, the control, takes its values in a given set U. We also require that the state x satisfy certain boundary conditions: $(x(0), x(T)) \in S$. Subject to these constraints, the problem is to choose (x, u) so as to minimize a cost $\ell(x(0), x(T))$, and the issue

*Presented at the Banach Center Workshop in honor of Czesław Olech, 2004.
†Professeur à l'Institut universitaire de France. Clarke@igd.univ-lyon1.fr

is to give a set of necessary conditions that a solution (x_*, u_*) must satisfy. It is enough that the solution be local in some sense with respect to the x variable; we shall be more precise on this point later.

We assume that $D_x f$ exists and, together with f, is continuous in (t, x, u); we also assume that S admits (outward) 'normal vectors' in some sense (for example, we may take S to be a C^1 manifold with or without boundary). The classical maximum principle then says:

Theorem 1.1
Let the optimal control u_* be essentially bounded. Then there is an arc p and a scalar λ_0 equal to 0 or 1 which satisfy the adjoint equation

$$-\dot{p}(t) = D_x \left\{\langle p(t), f(t, \cdot, u_*(t))\rangle\right\}(x_*(t)) \text{ a.e.,}$$

the maximum condition

$$\langle p(t), f(t, x_*(t), u)\rangle \leq \langle p(t), f(t, x_*(t), u_*(t))\rangle \quad \forall\, u \in U \text{ a.e.,}$$

the transversality condition:

$$(p(0), -p(T)) - \lambda_0 \nabla \ell(x_*(0), x_*(T)) \text{ is normal to } S \text{ at the point } (x_*(0), x_*(T)),$$

and nontriviality:

The maximum principle has been and remains an important tool in the many areas in which optimal control plays a role. It is natural to seek to extend its applicability, for example by reducing the hypotheses under which it can be derived, or extending the type of problem to which it applies. There have been very many advances of this type in the last fifty years; we refer for example to [1][9][10][11] [12][13][14] for references, but of course the list could be much longer.

Some of the advances referred to are of the type that weaken the regularity requirements on the underlying data, sometimes to the point of not positing differentiability in x. In such cases, the derivatives that appear in the adjoint equation and the transversality condition are replaced by some generalized derivative.

One approach along these lines, initiated by Clarke in the early 1970's, may be called the 'nonsmooth analysis' approach; other contributors to its development include Ioffe, Loewen, Mordukhovic, Rockafellar, and Vinter. The recent monograph [1] gives the state of the art with respect to this school of thought, and discusses various aspects of the theory at more length than is possible here.

Because generalized derivatives and normal cones appear in the statements of results obtained through the nonsmooth analysis approach, they may be somewhat difficult to absorb for those unfamiliar with such notions. And the presence of these constructs may sometimes obscure the fact that the underlying methodology makes new advances even in situations in which smoothness is present. For that reason, we hereby ban all generalized derivatives and normal cones from the discussion, except for Section 2, where the very nature of the (differential inclusion) problem leaves us no alternative, and in the proof in Section 3.

We proceed in the remainder of this introductory section to give two illustrative extensions of the classical maximum principle. In Section 2 we briefly discuss new necessary conditions for differential inclusions (taken from [1]) which serve as a general template by means of which all other results in this paper can be derived. In the final section we illustrate the versatility of this result by using it to deduce the maximum principle in the case of state-dependent control constraints.

1.2 A solution-specific maximum principle

Consider again the optimal control problem described above, but let us replace the classical assumptions on the data by the following:

Hypotheses for Theorem 1.2.

1. f is Lebesgue $\times$ Borel measurable in t and (x,u);
2. f is continuously differentiable with respect to x;
3. For some $\varepsilon > 0$, for t a.e. and $u \in U$, we have

$$k(t,u) := \sup_{|x - x_*(t)| < \varepsilon} \|D_x f(t,x,u)\| < +\infty;$$

4. $k(t, u_*(t))$ is summable.

Our second maximum principle is

Theorem 1.2
Under these hypotheses, any solution (x_*, u_*) satisfies the conclusions of Theorem 1.1.

It is easy to see that the hypotheses of Theorem 1.2 (Clarke 1976 [3]) are weaker than those of the classical setting (Theorem 1.1). (And they apply with D_x replaced by a generalized gradient, but we have agreed not to mention that.) On the other hand, it may be felt that there is a somewhat *ad hoc* nature to the hypotheses, since the solution (x_*, u_*) is mentioned in their formulation. But we remark that this was already a feature (albeit a discrete one) of the classical result: the first sentence in the statement of Theorem 1.1 could be rephrased as requiring the existence of a constant M such that $|u_*(t)| \leq M$ a.e.

In fact, the conclusions of Theorem 1.1 are false if the assumption that $u_* \in L^\infty$ is not made, as shown by Clarke and Vinter [6]. It turns out that *some* assumption of an *ad hoc* nature must be made after the fact, or else an *a priori* structural assumption about the data of the problem. (For example, in Theorem 1.1, one could add to the hypotheses on the data that U be bounded.) The formulation of structural assumptions (on Lagrangians) guaranteeing regularity (and hence validity of the necessary conditions) in the calculus of variations is a well-known enterprise since the work of Tonelli. It appears to be rare in an optimal control setting. Our next maximum principle illustrates recent progress in this direction, by means of a growth assumption on f.

1.3 A structural hypothesis

We continue to consider the same optimal control problem, but we modify the hypotheses as follows. We retain (1)(2) of Hypotheses 1.2, but replace (3)(4) by the following growth condition (that makes no explicit reference to the solution): for any bounded set X of x-values, there exist a constant c and a summable function $d(t)$ such that

$$||D_x f(t,x,u)|| \leq c|f(t,x,u)| + d(t) \quad \forall\, x \in X, u \in U, t \text{ a.e.}$$

The following is a special case of Theorem 5.2.1 in [1]. Note that we do not require the optimal control to be bounded.

Theorem 1.3
Under these hypotheses, any solution (x_*, u_*) satisfies the conclusions of Theorem 1.1.

We have emphasized here the distinction between 'solution-specific' necessary conditions (Theorem 1.2) and the type that depend instead on a structural hypothesis (Theorem 1.3). But the results of [1] have a number of other features (aside from nondifferentiability) that extend the classical maximum principle in a variety of directions , namely:

- It is possible to consider the problem in a novel 'stratified' way by means of a 'radius function' $R(t)$. This means that both the hypotheses and the optimality need only hold for states x satisfying $|\dot{x}(t) - \dot{x}_*(t)| \leq R(t)$. ($R$ need be neither small nor large; the theorems above correspond to the choice $R \equiv +\infty$.) Then the maximum condition is asserted only relative to the given radius (the other necessary conditions are unchanged).
- The theorem holds when x is merely a local minimum in the following weak sense: relative to those x for which

$$|x(0) - x_*(0)| + ||\dot{x} - \dot{x}_*||_{L^1} < \varepsilon$$

(that is, a local $W^{1,1}$ minimum).

- It is possible to allow an integral term in the cost functional at a new level of generality. In fact, the hypotheses that are required on the cost integrand are so minimal that one obtains new results even for the (smooth) basic problem in the calculus of variations.
- The results apply to so-called 'generalized control systems', in which the control dynamics are given by a parametrized family of vector fields.

We stress, however, that all these situations are treated as special cases of a single basic differential inclusion problem, together with its three necessary conditions for optimality. (This is also the case for the problem with mixed state/control constraints that we treat in Section 3.) These optimality conditions are recognizable as analogues of the Euler, Weierstrass and transversality conditions. In the next section we describe them briefly. But we must introduce some basic concepts in nonsmooth analysis.

2 A maximum principle for inclusions

Normal cones and subdifferentials

Although the proof of the theorem below requires nonsmooth analysis in infinite-dimensional (Hilbert) spaces as well as some proximal calculus, the statement itself requires only certain constructs on $\mathbb{R}^n$, to which we limit ourselves here. We refer to [5] for more on nonsmooth analysis.

Given a nonempty closed subset S of $\mathbb{R}^n$ and a point x in S, we say that $\zeta \in \mathbb{R}^n$ is a *proximal normal* (vector) to S at x if there exists $\sigma \geq 0$ such that

$$\langle \zeta, x' - x \rangle \leq \sigma \|x' - x\|^2 \quad \forall\, x' \in S.$$

(This is the *proximal normal inequality.*) The set (convex cone) of such ζ, which always contains 0, is denoted $N_S^P(x)$ and referred to as the proximal normal cone.

Given a lower semicontinuous function $f : X \to \mathbb{R} \cup \{+\infty\}$ and a point x in the effective domain of f, that is, the set

$$\operatorname{dom} f := \{x' \in \mathbb{R}^n : f(x') < +\infty\},$$

we say that ζ is a *proximal subgradient* of f at x if there exists $\sigma \geq 0$ such that

$$f(x') - f(x) + \sigma \|x' - x\|^2 \geq \langle \zeta, x' - x \rangle$$

for all x' in a neighborhood of x. The set of such ζ, which may be empty, is denoted $\partial_P f(x)$ and referred to as the proximal subdifferential.

The *limiting normal cone* $N_S^L(x)$ to S at x is obtained by applying a sequential closure operation to N_S^P:

$$N_S^L(x) := \left\{\lim \zeta_i : \zeta_i \in N_S^P(x_i), x_i \to x, x_i \in S\right\}.$$

A similar procedure defines the *limiting subdifferential*:

$$\partial_L f(x) := \{\lim \zeta_i : \zeta_i \in \partial_P f(x_i), x_i \to x, f(x_i) \to f(x)\}.$$

It can be shown that ζ belongs to $\partial_P f(x)$ iff the vector $(\zeta, -1)$ belongs to $N^P_{\mathrm{epi}\, f}(x, f(x))$, where $\mathrm{epi}\, f$, the *epigraph* of f, is the set

$$\mathrm{epi}\, f := \{(x, r) \in X \times \mathbb{R} : f(x) \leq r\}.$$

A similar characterization holds for $\partial_L f(x)$.

The differential inclusion problem

We are given a multifunction F mapping $[0, T] \times \mathbb{R}^n$ to the subsets of $\mathbb{R}^n$. It is assumed that F is $\mathcal{L} \times \mathcal{B}$ measurable and that for each t the graph $G(t)$ of the multifunction $F(t, \cdot)$ is closed. A trajectory x of F refers to an absolutely continuous function satisfying

$$\dot{x}(t) \in F(t, x(t)) \text{ a.e.}$$

We consider the problem $\mathcal{P}$ of minimizing $\ell(x(0), x(T))$ over the trajectories x of F satisfying the boundary constraints $(x(0), x(T)) \in S$. It is assumed that ℓ is locally Lipschitz and that S is closed.

Let R be a measurable function on $[0, T]$ with values in $(0, +\infty]$ (the radius function). We are given an arc x_* feasible for $\mathcal{P}$ which is a *local* $W^{1,1}$ *minimum of radius* R in the following sense: for some $\varepsilon > 0$, for every other feasible arc x satisfying

$$|\dot{x}(t) - \dot{x}_*(t)| \leq R(t) \text{ a.e.}, \int_0^T |\dot{x}(t) - \dot{x}_*(t)|\, dt \leq \varepsilon, \; \|x - x_*\|_\infty \leq \varepsilon,$$

one has $\ell(x(0), x(T)) \geq \ell(x_*(0), x_*(T))$.

The bounded slope condition

We denote by $\mathcal{T}_*^\varepsilon$ the tube of radius ε around x_*:

$$\mathcal{T}_*^\varepsilon := \{(t, x) : 0 \leq t \leq T, |x - x_*(t)| < \varepsilon\}.$$

We say that F satisfies the *bounded slope condition* near x_* if, for some $\varepsilon > 0$, there exist a constant c and a summable function d such that for almost every t, for all $(t, x) \in \mathcal{T}_*^\varepsilon$ and $(x, v) \in G(t)$, for all $(\alpha, \beta) \in N^P_{G(t)}(x, v)$, one has $|\alpha| \leq (c\,|v| + d(t))\,|\beta|$.

The necessary conditions

Theorem 2.1
If F satisfies the bounded slope condition near x_*, then there exist an arc p and a number λ_0 in $\{0,1\}$ satisfying the nontriviality condition

$$\lambda_0 + \|p\|_\infty \neq 0.$$

and the transversality condition:

$$(p(0), -p(T)) \in \partial_L \lambda_0 \ell(x_*(0), x_*(T)) + N_S^L(x_*(0), x_*(T)),$$

and such that p satisfies the Euler (or adjoint) inclusion:

$$\dot{p}(t) \in \mathrm{co}\left\{\omega : (\omega, p(t)) \in N^L_{G(t)}(x_*(t), \dot{x}_*(t))\right\} \text{ a.e. } t \in [0,T]$$

as well as the Weierstrass condition of radius R:

$$\langle p(t), v\rangle \leq \langle p(t), \dot{x}_*(t)\rangle \ \forall v \in F(t, x_*(t)) \cap \overline{B}(\dot{x}_*(t), R(t)), \ \text{ a.e.} t \in [0,T].$$

If the above holds for a sequence of radius functions R_i (with ε, c and d possibly depending on i) for which

$$\liminf_{i\to\infty} R_i(t) = +\infty \ \text{ a.e.}$$

then the conclusions hold for an arc p which satisfies the global Weierstrass condition:

$$\langle p(t), v\rangle \leq \langle p(t), \dot{x}_*(t)\rangle \ \forall v \in F(t, x_*(t)), \ \text{ a.e. } t \in [0,T].$$

We remark that the above is the 'global structural version' of a family of such results proven in [1]. Note that through the use of a sequence R_i as above, we are able to assert the global necessary conditions (that is, of infinite radius) even when x_* is only known to be a minimum for every *finite* radius. The fact that this latter property does not in itself imply that x_* is a minimum of infinite radius is related to the well-known Lavrentiev phenomenon.

The role of the bounded slope condition is to guarantee a 'pseudo-Lipschitz' property (introduced in this connection in [2]) near any given solution, a property which can also be posited on a 'solution- specific' basis.

Theorem 2.1 represents a highly evolved version of the original maximum principle. (It is perhaps the adjoint equation that has mutated the most.) The striking thing about the theorem, not evident at a glance, is how the hypotheses turn out to be sufficiently weak and general, and the conclusions sufficiently strong so as to apply to a variety of situations and yield state of the art necessary conditions for them. A number of examples are developed in [1]; the next section gives another.

3 A problem with mixed constraints

Optimal control problems in which the admissible control values depend explicitly on the current value of the state are fundamentally different from those discussed in Section 1. They have traditionally been considered an important test problem in evaluating the scope of a given theory of necessary conditions. In this section we show how a general problem of this type can be treated directly from Theorem 2.1 of the previous section.

We consider on the interval $[0, T]$ the dynamics

$$\dot{x}(t) = f(t, x(t), u(t), v(t)) \text{ a.e.},$$

where now the control variable has two components u and v which are treated somewhat differently. We impose the mixed state/control constraints

$$g(t, x(t), u(t), v(t)) \leq 0, \; h(t, x(t), u(t), v(t)) = 0, \; v(t) \in V(t) \text{ a.e.},$$

and the boundary conditions

$$(x(0), x(T)) \in S,$$

where the sets S and $V(t)$ (t a.e.) are closed. All the variables above belong to certain Euclidean spaces, and the functions f, g, h are vector-valued. Problems of this type seem to have been considered first by Dubovickiĭ and Milyutin [8]; see also Dmitruk [7].

We denote by $\mathcal{C}(t, x)$ the set of admissible control values corresponding to (t, x): those points (u, v) satisfying the conditions

$$g(t, x, u, v) \leq 0, \; h(t, x, u, v) = 0, \; v \in V(t).$$

We are given an admissible triple (x_*, u_*, v_*) for the problem and a positive ε. We denote by $\mathcal{T}_*^\varepsilon$ the tube of radius ε around x_*:

$$\mathcal{T}_*^\varepsilon := \{(t, x) : 0 \leq t \leq T, |x - x_*(t)| < \varepsilon\}.$$

We suppose that the set

$$\mathcal{C} := \{(t, x, u, v) : (t, x) \in \mathcal{T}_*^\varepsilon, (u, v) \in \mathcal{C}(t, x)\}$$

is bounded, and that on some neighborhood of $\mathcal{C}$ the functions f, g, and h have derivatives in x and u which, together with the functions themselves, are continuous in (t, x, u, v). We also posit the $\mathcal{L} \times \mathcal{B}$ (Lebesgue$\times$Borel) measurability (see [4]) of the multifunction $V(\cdot)$.

The following *constraint qualification* is assumed to hold along x_*:

For any $t \in [0, T]$ and $(u, v) \in \mathcal{C}(t, x_*(t))$, the only pair (γ, λ) satisfying the relations

$$D_u\{\langle \gamma, g\rangle + \langle \lambda, h\rangle\}(t, x_*(t), u, v) = 0$$
$$\langle \gamma, g(t, x_*(t), u, v)\rangle = 0,\ \gamma \geq 0$$

is $\gamma = 0$ and $\lambda = 0$.

The problem consists of minimizing $\ell(x(0), x(T))$ over the admissible state/control functions (x, u, v), where ℓ is continuously differentiable.

Theorem 3.1
Let (x_*, u_*, v_*) be a solution of the problem relative to the tube $\mathcal{T}_*^{\varepsilon}$. Then there exists an arc p, bounded measurable functions $\gamma \geq 0$ a.e. and λ, and a scalar λ_0 equal to 0 or 1 which satisfy the complementary slackness condition:

$$\langle \gamma(t), g(t, x_*(t), u_*(t), v_*(t))\rangle = 0 \ \text{ a.e.},$$

the maximum condition:

$$\langle p(t), f(t, x_*(t), u', v')\rangle \leq$$
$$\langle p(t), f(t, x_*(t), u_*(t), v_*(t))\rangle \quad \forall\, (u', v') \in \mathcal{C}(t, x_*(t)) \ \text{ a.e.},$$

the adjoint equation:

$$-\dot{p}(t) = D_x \left\{\langle p, f\rangle + \langle \gamma, g\rangle + \langle \lambda, h\rangle\right\} (t, x_*(t), u_*(t), v_*(t)) \ \text{ a.e.},$$

the transversality condition:

$$(p(0), -p(T)) - \lambda_0 \nabla \ell(x_*(0), x_*(T)) \in N_S^L(x_*(0), x_*(T)),$$

and nontriviality:

$$\lambda_0 + ||p||_\infty \neq 0.$$

Proof.

We require the following fact.

Lemma 1. There exists $M > 0$ and $\varepsilon' > 0$ with the following property: given any $(t, x) \in \mathcal{T}_*^{\varepsilon'}$ and $(u, v) \in \mathcal{C}(t, x)$, together with any pair (γ, λ) satisfying

$$\gamma \geq 0, \quad \langle \gamma, g(t, x, u, v)\rangle = 0,$$

then one has

$$|(\gamma, \lambda)| \leq M\, |D_u\{\langle \gamma, g\rangle + \langle \lambda, h\rangle\}(t, x, u, v)|\,.$$

We omit the simple proof, based for example on assuming the contrary and obtaining a contradiction of the constraint qualification. Without loss of generality, we assume that the ε' of the Lemma equals the original ε.

We now consider the optimal control of a system governed by the differential inclusion

$$(\dot{x}(t), \dot{y}(t)) \in F(t, x(t), y(t)) \text{ a.e.},$$

where y is a scalar variable and $F(t,x,y)$ is given by

$$\{(f(t,x,u,v), \theta|u - u_*(t)|^2 + \theta|v - v_*(t)|^2) : (u,v) \in \mathcal{C}(t,x)\}.$$

Here θ is a fixed parameter in $(0,1)$; in the last step of the proof it will tend to zero. Its presence here serves to single out the particular controls u_* and v_* that generate x_* (in general they are not unique).

We observe that the arc $(x_*(t), 0)$ solves (locally uniquely) the *auxiliary problem* of minimizing $\ell(x(0, x(T)) + y(T)$ over the trajectories (x,y) of F which satisfy the boundary conditions

$$(x(0), x(T)) \in S, \; y(0) = 0.$$

The aim is to apply Theorem 2.1 to this context, so in this proof we must allow normal cones to enter.

It is easy to see that F satisfies the basic hypotheses of the preceding section (graph closedness and measurability); we proceed now to verify that a strong form of the bounded slope condition holds. That is, we demonstrate the existence of a constant c such that for any $(t,x) \in \mathcal{T}_*^\varepsilon$, whenever $(\alpha, \mu, \beta, \nu)$ is a proximal normal to $G(t)$ at a point

$$(x, y, f(t,x,u,v), r), \text{ with } r = \theta\left[|u - u_*(t)|^2 + |v - v_*(t)|^2\right],$$

the following inequality holds:

$$|(\alpha, \mu)| \leq c|(\beta, \nu)|. \tag{$*$}$$

Let K be a bound for the set $\mathcal{C}$, and also for all derivatives in (x,u) evaluated on that set. The definition of proximal normal tells us that as a function of (x', u', v'), the quantity

$$\begin{aligned} &\langle \alpha, x' \rangle + \mu y' + \langle \beta, f(t, x', u', v') \rangle \\ &\qquad + \nu\theta\left[|u' - u_*(t)|^2 + |v' - v_*(t)|^2\right] - Q(x', u', v') \end{aligned}$$

is maximized relative to the constraints

$$g(t, x', u', v') \leq 0, \; h(t, x', u', v') = 0, \; v' \in V(t)$$

at the point (x, u, v), where Q is a quadratic-like term whose derivative vanishes at that point. Note first that this gives $\mu = 0$ (which is to be expected since F has no dependence on y).

We fix $v' = v$ and apply the Lagrange multiplier rule (see for example [5]) to the reduced (x', u') optimization problem. We derive the existence of multipliers $\gamma \geq 0$ and λ such that $\langle \gamma, g(t, x, u, v)\rangle = 0$ and

$$\alpha + \langle \beta, D_x f(t, x, u, v)\rangle \\ = -\langle \gamma, D_x g(t, x, u, v)\rangle - \langle \lambda, D_x h(t, x, u, v)\rangle$$

$$\langle \beta, D_u f(t, x, u, v)\rangle + 2\nu\theta(u - u_*(t)) \\ = -\langle \gamma, D_u g(t, x, u, v)\rangle - \langle \lambda, D_u h(t, x, u, v)\rangle$$

(The constraint qualification, as given by Lemma 1, implies that these necessary conditions have the given 'normal' form.)

The second of these equations, combined with Lemma 1, implies

$$|(\gamma, \lambda)| \leq M |\langle \beta, D_u f(t, x, u, v)\rangle + 2\nu\theta(u - u_*(t))|.$$

This in turn, when combined with the first equation, yields $(*)$ for an appropriate value of c depending only on M and K.

The characterization of proximal normal vectors just derived leads to one for certain limiting normals:

Lemma 2. Let $(\alpha(t), \mu(t), \beta(t), \nu(t))$ be measurable and belong to $N^L_{G(t)}(x_*(t), 0, \dot{x}_*(t), 0)$ a.e. Then $\mu = 0$ a.e. and there exist measurable functions $\gamma(t) \geq 0$ and $\lambda(t)$ with

$$|(\gamma(t), \lambda(t))| \leq 4KM(|\nu(t)| + |\beta(t)|) \text{ a.e.}$$

such that one has

$$-\alpha(t) = D_x \{\langle \beta, f\rangle + \langle \gamma, g\rangle + \langle \lambda, h\rangle\} (t, x_*(t), u_*(t), v_*(t)) \text{ a.e.}$$

To see this, we recall that by definition, for fixed t a.e., $(\alpha, \mu, \beta, \nu)$ is the limit of points $(\alpha_i, \mu_i, \beta_i, \nu_i)$ which are proximal normals to $G(t)$ at points $(x_i, y_i, f(t, x_i, u_i, v_i), r_i)$ converging to $(x_*(t), 0, \dot{x}_*(t), 0)$. Then the μ_i are zero, as we have seen, and since $r_i \to 0$, it follows that (u_i, v_i) converges to $(u_*(t), v_*(t))$. (This is where $\theta > 0$ is invoked.) Then passing to the limit in the proximal normal characterization obtained previously implies that $-\alpha(t)$ has a representation of the given type. Standard measurable selection theorems (see for example [5]) show that the functions γ and λ can be defined measurably.

We turn now to the application of Theorem 2.1 (with $R \equiv +\infty$), which implies the existence of an arc (p, q) and a nonnegative scalar

λ_0 such that $\lambda_0 + ||(p,q)||_\infty \neq 0$, and such that the Euler inclusion, the Weierstrass condition, and the transversality condition are satisfied (for the auxiliary problem). We proceed now to interpret those three conclusions in the present context.

First, the transversality condition gives

$$(p(0), -p(T)) - \lambda_0 \nabla \ell(x_*(0), x_*(T)) \in N_S^L(x_*(0), x_*(T)),\ q(T) = -\lambda_0.$$

The Euler inclusion implies that q is constant (see Lemma 2), so that in fact q is identically $-\lambda_0$. It follows that $\lambda_0 + ||p||_\infty \neq 0$; by scaling we can take $\lambda_0 + ||p||_\infty = 1$.

The following result builds upon Lemma 2 and allows us to interpret the Euler inclusion.

Lemma 3. There exist measurable functions $\gamma(t) \geq 0$ and $\lambda(t)$ with

$$|(\gamma(t), \lambda(t))| \leq 4KM(\lambda_0 + |p(t)|) \text{ a.e.}$$

such that one has

$$-\dot{p}(t) = D_x \{\langle p, f\rangle + \langle \gamma, g\rangle + \langle \lambda, h\rangle\} (t, x_*(t), u_*(t), v_*(t)) \text{ a.e.}$$

To see this, it suffices to observe that almost everywhere, the Euler inclusion gives $\dot{p}(t)$ as a convex combination of points α_j such that $(\alpha_j, 0, p(t), -\lambda_0)$ is a limiting normal to $G(t)$ at $(x_*(t), 0, \dot{x}_*(t), 0)$. But then the result follows from Lemma 2.

Finally, we observe that the Weierstrass condition is equivalent to the assertion that almost everywhere the maximum over $C(t, x_*(t))$ of the following function of (u,v):

$$\langle p(t), f(t, x_*(t), u, v)\rangle - \lambda_0 \theta \left[|u - u_*(t)|^2 + |v - v_*(t)|^2\right]$$

is attained at $(u_*(t), v_*(t))$.

Note that all the required assertions of the theorem would now hold if θ were zero. The final step in the proof is a sequencing argument to arrive at this.

We consider the above for a sequence of θ_i decreasing to 0, letting the associated data be denoted $p_i, \gamma_i, \lambda_i, \lambda_0^i$. A routine use of Gronwall's Lemma (based upon the bound for $|\dot{p}_i(t)|$ furnished by Lemma 3) shows that the p_i are equi- Lipschitz. We may suppose that the sequence converges uniformly to a limit p, and that λ_0^i converges to λ_0; then we have $||p||_\infty + \lambda_0 = 1$. (We can normalize at the end to obtain λ_0 equal to 0 or 1.) The limiting transversality and Weierstrass conditions are the ones affirmed in the statement of the theorem, so only the adjoint equation needs verifying.

The functions (γ_i, λ_i) are bounded in L^∞ (Lemma 3), and can be assumed to converge weak* to a limit (γ, λ). Since this convergence preserves linear equalities and inequalities, we see that the limiting data have all the required properties, and the proof is complete.

Remarks

The proof. Note that the proof is essentially a direct appeal to Theorem 2.1. For the most part it consists of interpreting the normal cones that are involved. The device which makes unique the optimal control is a familiar one in dealing with inclusions, as is the sequential compactness argument for the adjoint inclusion. Other familiar devices (see [4] and Chapter 5 of [1]) would be involved in extending the result to cover nonsmooth data, or in reducing the continuity or boundedness requirements, or in considering variable time problems. We remark that a more general type of constraint qualification could also be used above, one in which normal cones to the graph of V intervene. The extension to unilateral (or pure) state constraints, ruled out here by the constraint qualification, can also be envisaged (but then the adjoint variable is of bounded variation, see [4]), as can a 'hybrid' form similar to the one described in Chapter 5 of [1], in which a cost integral of very general nature is admitted.

Transversality. We have violated our ban against normal cones in the statement of the Theorem, as regards transversality. To compensate, let us remark that if S is described by a system of equalities and inequalities in terms of smooth functions, then the limiting normal vectors can be written as linear combinations of the derivatives of those functions (see [5]), thus eliminating the explicit reference to a normal cone. Classically, transversality conditions have often been expressed in this 'Lagrange multiplier form'.

Type of minimum. We have assumed in the Theorem that x_* is a strong local minimum for the problem, but the proof is unchanged if it is supposed instead that we have a local $W^{1,1}$ minimum as defined earlier, since that is adequate in order to invoke the differential inclusion results. In a similar vein, when the case of unbounded controls is considered, the notion of 'Pontryagin minimum' can be treated by introducing a radius function.

Acknowledgement

The author thanks Andrei Dmitruk for several useful discussions, and the Banach Center for its warm hospitality and the use of its facilities.

References

[1] F. Clarke. Necessary conditions in dynamic optimization. *Memoirs of the Amer. Math. Soc.*, 173(816), 2005.

[2] F. H. Clarke. The generalized problem of Bolza. *SIAM J. Control Optim.*, 14:682–699, 1976.

[3] F. H. Clarke. The maximum principle under minimal hypotheses. *SIAM J. Control Optim.*, 14:1078–1091, 1976.

[4] F. H. Clarke. *Optimization and Nonsmooth Analysis.* Wiley-Interscience, New York, 1983. Republished as vol.5 of *Classics in Applied Mathematics*, SIAM, 1990.

[5] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth Analysis and Control Theory.* Graduate Texts in Mathematics, vol. 178. Springer-Verlag, New York, 1998.

[6] F. H. Clarke and R. B. Vinter. On the conditions under which the euler equation or the maximum principle hold. *Applied Math. and Optimization*, 12:73–79, 1984.

[7] A. V. Dmitruk. Maximum principle for a general optimal control problem with state and regular mixed constraints. *Comp. Math. and Modeling*, 4:364–377, 1993.

[8] A. Ya. Dubovitskiĭ and A. A. Milyutin. Theory of the principle of the maximum. (Russian). In *Methods of the Theory of Extremal Problems in Economics.* Nauka, 1981.

[9] M. R. Hestenes. *Calculus of Variations and Optimal Control Theory.* Wiley, New York, 1966.

[10] A. D. Ioffe and V. Tikhomirov. *Theory of Extremal Problems.* Nauka, Moscow, 1974. English translation, North-Holland (1979).

[11] A. A. Milyutin and N. P. Osmolovskii. *Calculus of Variations and Optimal Control.* American Math. Soc., Providence, 1998.

[12] L. S. Pontryagin, R. V. Boltyanskii, R. V. Gamkrelidze, and E. F. Mischenko. *The Mathematical Theory of Optimal Processes.* Wiley-Interscience, New York, 1962.

[13] R. B. Vinter. *Optimal Control.* Birkhäuser, Boston, 2000.

[14] J. Warga. *Optimal Control of Differential and Functional Equations.* Academic Press, New York, 1972.