SIAM J. Control Optim.
Vol. 32, No. 2, pp. 442–470, 1994

OPTIMAL CONTROL OF UNBOUNDED DIFFERENTIAL INCLUSIONS *

PHILIP D. LOEWEN† AND R. T. ROCKAFELLAR‡

Abstract. We consider a Mayer problem of optimal control, whose dynamic constraint is given by a convex-valued differential inclusion. Both state and endpoint constraints are involved. We prove necessary conditions incorporating the Hamiltonian inclusion, the Euler-Lagrange inclusion, and the Weierstrass-Pontryagin maximum condition. Our results weaken the hypotheses and strengthen the conclusions of earlier works. Their main focus is to allow the admissible velocity sets to be unbounded, provided they satisfy a certain continuity hypothesis. They also sharpen the assertion of the Euler-Lagrange inclusion by replacing Clarke's subgradient of the essential Lagrangian with a subset formed by partial convexification of limiting subgradients. In cases where the velocity sets are compact, the traditional Lipschitz condition implies the continuity hypothesis mentioned above, the assumption of "integrable boundedness" is shown to be superfluous, and our refinement of the Euler-Lagrange inclusion remains a strict improvement on previous forms of this condition.

Key words. Optimal control, differential inclusion, Hamiltonian inclusion, maximum principle, nonsmooth analysis

AMS(MOS) subject classifications. 49K24

This paper describes necessary conditions for optimality in the following Mayer problem of optimal control: choose an arc (i.e., an absolutely continuous function) $x\colon [a,b] \to \mathbb{R}^n$ in order to

$$(P) \qquad \left[\begin{array}{l} \text{minimize } \ell(x(a),x(b)) \\ \text{subject to } \dot{x}(t) \in F(t,x(t)) \text{ a.e. } t \in [a,b], \\ \qquad (x(a),x(b)) \in S, \\ \qquad x(t) \in X(t)\ \forall t \in [a,b]. \end{array}\right.$$

Experts will recognize the endpoint constraint $(x(a),x(b)) \in S$ and the state constraint $x(t) \in X(t)\ \forall t \in [a,b]$ as aspects of the model which are indispensable for applications, but which account for considerable complexity in the statement and derivation of necessary conditions. Clarke [2, Chap. III] gives an excellent introduction to this problem and describes several applications. Our main result can be viewed as a generalization of Clarke's necessary conditions in [2, Thm. 3.5.2]; however, the calculus described by Ioffe [6] and Rockafellar [29], and the careful Hamiltonian analysis of Loewen and Rockafellar [13] are important steps along the way from the cited result to the work at hand. The first two

*Received by the editors August 15, 1991; accepted for publication (in revised form) September 99, 1992. This work was supported by Canada's Natural Science and Engineering Research Council under grant 5-89441, and by the National Science Foundation under grant DMS-8819586.

†Department of Mathematics, University of British Columbia, Vancouver, CANADA, V6T 1Z2.

‡Department of Mathematics University of Washington Seattle, Washington USA 98195.

sections of [13] describe our reasons for choosing the formulation (P), and the relationship between this version of the problem and others current in the literature.

The results presented here improve upon those in [2] and [13] in three important ways. First, the problem is more general than any considered before, since we do not require the sets $F(t,x)$ of admissible velocities to be bounded. (We insist throughout, however, that these sets be convex.) Second, our necessary conditions are more precise than any previously published, since they involve sharper forms of the transversality condition and the Euler-Lagrange inclusion than those in [2] and [13]. Finally, our method of proof allows a simpler approach to the main result of [13], which is recovered as a corollary. We expect all of these improvements to serve in future developments of the theory.

Several sets of necessary conditions for optimal control problems without boundedness assumptions already exist in the literature. For example, Clarke proves necessary conditions analogous to the Euler-Lagrange equation for such a differential inclusion problem in [1] (see equation (5.1) below). Although his result does not require the velocity sets to be bounded, it does involve a Lipschitz hypothesis on the state dependence of F—an unacceptably strong condition when the velocity sets are actually unbounded. Polovinkin and Smirnov [19, 20] prove a form of the Euler-Lagrange inclusion which is sharper than Clarke's, using a truncation argument to weaken the Lipschitz hypothesis considerably. Their results also dispense with the convexity condition on the values of the multifunction $F(t,x)$. Kaskosz and Łojasiewicz [10] consider a Mayer problem whose dynamic constraint is a controlled differential equation in which both the control sets and the resulting velocity sets are allowed to be unbounded. However, their adjoint inclusions involve Carathéodory selections of the resulting multifunction F, and are not directly comparable to those of our main theorem. (A simple connection in the bounded case is indicated by Loewen and Vinter [14].) Also, Lipschitz conditions enter [10] at several points, making direct comparison with our main result difficult. The current paper breaks new ground in presenting Hamiltonian necessary conditions for optimality in problem (P) without assuming either that the velocity sets are bounded, or that they display full Lipschitz dependence on the state. Like our previous paper [13], it asserts the Hamiltonian and Eulerian forms of the necessary conditions simultaneously.

Two simple themes underlie our approach: truncation and strict convexity. Let us explain these ideas before pursuing the details. Suppose $\overline{x}$ solves problem (P). In the case where the optimal solution $\overline{x}$ is Lipschitzian, i.e., $\dot{\overline{x}} \in L^\infty([a,b],\mathbb{R}^n)$, we observe that for any $R>0$, the arc $\overline{x}$ also solves the version of problem (P) in which the given multifunction F is truncated to produce the bounded multifunction $\widetilde{F}(t,x) := F(t,x) \cap \big(\dot{\overline{x}}(t) + R\,\mathrm{cl}\,\mathbb{B}\big)$. Therefore $\overline{x}$ must fulfill the known necessary conditions for bounded differential inclusions, provided that $\widetilde{F}$ satisfies a suitable Lipschitz condition. Identifying hypotheses on F which ensure this is one of this paper's main contributions. Then, of course, there is the question of relating the necessary conditions derived using $\widetilde{F}$ to those one might expect for F. This is not trivial either: Section 3 contains the detailed arguments. Finally, when $\overline{x}$ is absolutely continuous but not Lipschitzian, we must allow the truncation radius R to vary with time. Our presentation treats this case in parallel with the Lipschitz case. By coordinating the hypotheses on the multifunction F with the regularity of the solution, we derive the same necessary conditions in both instances. If F is "integrably sub-Lipschitzian in the large" (see Definition 2.3(b)) at every point $(t, \overline{x}(t))$ of $\operatorname{gph} \overline{x}$, the necessary conditions are satisfied without any regularity hypothesis on $\overline{x}$; when $\overline{x}$ is known to be Lipschitzian, we require only that F be "sub-Lipschitzian" (see Definition 2.3(a)) at every point $(t, \overline{x}(t))$ of $\operatorname{gph} \overline{x}$.

Strict convexity has a unifying effect on the necessary conditions of nonsmooth optimal control, as noted in our previous work [13]. Continuing to assume that $\overline{x}$ solves (P), we note that $\overline{x}$ remains optimal for the problem $(\mathcal{P})$ in which the objective function $\ell(x(a), x(b))$ is augmented by an integral term to become

$$\ell(x(a), x(b)) + \int_a^b \left[\sqrt{1 + \left|\dot{x}(t) - \dot{\overline{x}}(t)\right|^2} - 1 \right] dt.$$

Hence the Hamiltonian necessary conditions for optimality in $(\mathcal{P})$ must apply to $\overline{x}$. In the bounded case, the analysis of [13] shows that the Hamiltonian inclusion for $\overline{x}$ in the convexified problem $(\mathcal{P})$ implies the Hamiltonian inclusion, the sharpened Euler-Lagrange inclusion, and the Weierstrass-Pontryagin maximum condition we ultimately intend to assert for the original problem (P). (This analysis hinges upon the strict convexity of the integrand above as a function of the velocity variable $\dot{x}$.) To make the results of [13] applicable here, we first truncate the problem as described in the previous paragraph, and then introduce strict convexity.

The small right-hand side in our transversality inclusion will surprise no one working in the field. Similar transversality conditions appeared first in the work of Mordukhovich [15], who has applied similar ideas to a range of problems in recent years—see [17]. The new condition is obtained by replacing Clarke's normal cone and subgradient set with their (possibly nonconvex) subsets consisting of limits of proximal normals and proximal subgradients. Clarke actually uses limiting proximal normals to prove his transversality conditions in [2, Thm. 3.5.2], and his proof requires only the slightest modifications to obtain the transversality conditions used here. (This is noted explicitly in [4, Thm. 4.1, footnote].) The sharpened transversality condition also figures in Rowland and Vinter's recent work [31] on necessary conditions for controlled differential equations with free time. We take pains to incorporate it here in order that Theorem 4.3 below can legitimately claim to have the weakest hypotheses and the strongest conclusions of any set of necessary conditions for the optimal control of differential inclusions on a fixed time interval.

The refinement of the Euler-Lagrange inclusion used here is also obtained by using the cone of limiting proximal normals in place of its convex hull (Clarke's normal cone) on the right-hand side in [2] and [13]. Some convexification is still required, but it now pertains only to the components involving derivatives of the adjoint function instead of to all components at once. A related inclusion has recently been given under considerably stronger hypotheses by Mordukhovich [18]: our inclusion implies Mordukhovich's, and can be strictly better in certain cases. The key to our refined formulation is the introduction of strict convexity through a suitable integral cost term, as outlined above. A description of Mordukhovich's condition and a detailed comparison with ours appears in Section 5.

The paper's first section describes the starting point for this work—the well known Hamiltonian necessary conditions of Clarke [2] as formulated by Loewen and Rockafellar [13]. It outlines the minor modifications to existing arguments required to sharpen the transversality inclusions as described above. The next two sections concern truncation: Section 2 introduces the truncated multifunction $\widetilde{F}$ and describes hypotheses under which it satisfies a suitable Lipschitz condition, while Section 3 elucidates the relationship between the subgradients of the two Hamiltonians corresponding to the original and truncated multifunctions. Section 4 draws its antecedents together to produce a set of Hamiltonian necessary conditions for unbounded differential inclusion problems (Theorem 4.1). It then brings in strict convexity as outlined above. The methods of Sections 2 and 3 (together with Loewen and Rockafellar [13]) then allow the simultaneous derivation of the Hamiltonian inclusion, the refined Euler-Lagrange inclusion, and the Weierstrass-Pontryagin maximum condition. This effort culminates in Theorem 4.3, the main result of this paper. The concluding Section 5 offers a comparison between Theorem 4.3 and other published work, and gives some examples which clarify the distinctions between the various adjoint inclusions appearing here and elsewhere in the literature.

Readers interested in a quick overview of the work should observe that the notation for generalized derivatives and normals introduced in Section 1 differs from that in such standard works as Clarke [2]. Clarke subgradients and normals are indicated by the "barred" symbols $\overline{\partial} f(x)$ and $\overline{N}_C(x)$, while proximal subgradients and normals wear a double hat: $\hat{\hat{\partial}} f(x)$ and $\hat{\hat{N}}_C(x)$. The unadorned notation $\partial f(x)$ and $N_C(x)$ is reserved for sets of limiting proximal subgradients and limiting proximal normals.

1. Hypotheses and Preliminary Results. In this section we establish the technical foundation on which our later results will rest. We state the hypotheses under which we will later analyse the given problem (P), and review the constraint qualification we must impose when the state constraint is active along the optimal arc. We also review the necessary conditions for bounded differential inclusions due to Clarke [2, Thm. 3.5.2], and observe that they remain valid with a somewhat sharper transversality condition. Since our formulation of the state constraint differs from Clarke's, we will use the form of his result appearing in our previous work [13, Thm. 2.8]. (The relationship between these two modes of presentation is clearly spelled out in [13]: while it is almost true to say that a simple change of variable makes them equivalent, the extra analysis appearing in Lemma 2.4 of [13] makes the nontriviality assertion of [13, Thm. 2.8] stronger than Clarke's.) We sharpen the transversality condition in the known result by replacing its right-hand side with a smaller set. Instead of the Clarke subgradient and normal cone, we use the limiting subgradient and the limiting normal cone. These are the fundamental objects in the theory of proximal analysis, which is described in Rockafellar [27], [29], and Clarke [2, Sect. 2.5], for example; see also the book by Mordukhovich [17].

Proximal Analysis. Consider a closed set $C \subseteq \mathbb{R}^m$ containing some point c. A vector $\zeta \in \mathbb{R}^m$ is called a *proximal normal to C at c*, written $\zeta \in \hat{\hat{N}}_C(c)$, if there is some $M > 0$ so large that

$$\langle \zeta , c' - c \rangle \le M |c' - c|^2 \quad \text{for all } c' \in C. \tag{1.1}$$

Theorem 1.2 below refers to the cone of *limiting normals to C at c*, namely,

$$N_C(c) := \big\{ \zeta \in \mathbb{R}^m : \zeta = \lim_{k\to\infty} \zeta_k \text{ for some sequences } \zeta_k \in \hat{\hat{N}}_C(c_k) \text{ and } c_k \xrightarrow[C]{} c \big\}. \tag{1.2}$$

(Here $c_k \xrightarrow[C]{} c$ means that $c_k \to c$ and $c_k \in C$ for all k.) The important properties of the limiting normal cone (easily deduced, for example, from [2, Section 2.5]) are
(a) If $c \in \operatorname{bdry} C$, then $N_C(c)$ contains nonzero elements;
(b) The multifunction $c' \rightrightarrows N_C(c')$ has closed graph; and
(c) Clarke's normal cone $\overline{N}_C(c)$ is given by

$$\overline{N}_C(c) = \operatorname{cl} \operatorname{co} N_C(c). \tag{1.3}$$

When the object of study is not a set but a locally Lipschitzian function $f: \mathbb{R}^m \to \mathbb{R}$, we apply the previous notions to the set $C := \operatorname{epi} f = \{(x, r) \in \mathbb{R}^m \times \mathbb{R} : r \geq f(x)\}$. This leads to the following definition: Given a point x, a vector ζ is called a *proximal subgradient of f at x,* written $\zeta \in \hat{\hat{\partial}} f(x)$, if there is some $M > 0$ so large that, on some neighbourhood U of x, one has

$$f(x') \geq f(x) + \langle \zeta, x' - x \rangle - M|x' - x|^2 \qquad \forall x' \in U. \tag{1.4}$$

The set of *limiting subgradients of f at x* is defined by

$$\partial f(x) = \Big\{ \zeta \in \mathbb{R}^m : \zeta = \lim_{k\to\infty} \zeta_k \text{ for some sequences } \zeta_k \in \hat{\hat{\partial}} f(x_k),\ x_k \to x \Big\}. \tag{1.5}$$

For locally Lipschitzian functions f, the set $\partial f(x)$ is nonempty and compact-valued everywhere, and the multifunction $x' \rightrightarrows \partial f(x')$ has closed graph. Moreover, Clarke's generalized gradient $\overline{\partial} f(x)$ may be obtained from the set of limiting subgradients as follows:

$$\overline{\partial} f(x) = \operatorname{co} \partial f(x). \tag{1.6}$$

(A relationship somewhat more complicated than (1.6) gives $\overline{\partial} f(x)$ in the case where f is assumed only to be lower semicontinuous and extended real valued.) Boris Mordukhovich has used the limiting normal cone in the formulation of necessary conditions since 1976 [15], [16], [17]. In collaboration with his student A. Y. Kruger [11], he has extended certain aspects of the theory to infinite-dimensional spaces. More recently, Ioffe [6] has studied the limiting normal cone and limiting subgradient set described here under the names "approximate normal cone" and "approximate subdifferential", and given a more comprehensive extension to the infinite-dimensional case [7], [8], [9].

Hypotheses. Throughout the paper we confine our attention to a relatively open subset Ω of $[a, b] \times \mathbb{R}^n$ having nonempty sections

$$\emptyset \neq \Omega_t = \{x \in \mathbb{R}^n : (t, x) \in \Omega\} \qquad \forall t \in [a, b].$$

In order to treat a local solution $\overline{x}$, we assume that $F(t, x)$ is empty-valued for $(t, x) \notin \Omega$. This makes the requirement that $x(t) \in \Omega_t$ for all t implicit for admissibility in problem (P). (Notice that for any continuous function $x: [a, b] \to \mathbb{R}^n$ whose graph lies in Ω, a simple compactness argument implies the existence of some $\varepsilon > 0$ so small that $x(t) + \varepsilon\mathbb{B} \subseteq \Omega_t$

for all $t \in [a, b]$. Here, and throughout the paper, $\mathbb{B}$ denotes the open unit ball in $\mathbb{R}^n$.) Furthermore, we assume

(H1) The endpoint cost functional ℓ is locally Lipschitz on the closed set $S_0 := (\mathrm{cl}\,\Omega_0) \times (\mathrm{cl}\,\Omega_1)$, and the localized endpoint constraint set $S \cap S_0$ is closed;

(H2) The sets $F(t, x)$ are nonempty, closed, and convex for each (t, x) in Ω;

(H3) The multifunction F is measurable with respect to the σ-field $\mathcal{L} \times \mathcal{B}$ generated by products of Lebesgue subsets of $[a, b]$ with Borel subsets of $\mathbb{R}^n$;

(H4) The state constraint multifunction X has closed values $X(t)$ and is lower semicontinuous, in the sense that for every point $(t_0, x_0) \in \Omega \cap (\mathrm{gph}\, X)$, and for every sequence $t_k \to t_0$ in $[a, b]$, there exists a sequence $x_k \to x_0$ satisfying $x_k \in X(t_k)$ for all k.

Jump Directions. It is well known that the action of state constraints on an optimal trajectory manifests itself in the necessary conditions by producing discontinuities in the corresponding adjoint arc. Roughly speaking, the adjoint vector is allowed to jump in an outward normal direction to the constraint set at an instant when the constraint is active. In the general setting proposed here, the possible jump directions lie in the closed convex cone defined as follows for each (t, x) in $\Omega \cap (\mathrm{gph}\, X)$:

$$\overline{N}_X(t,x) = \mathrm{cl\,co}\Big\{\nu \in \mathbb{R}^n \ : \nu = \lim_{k\to\infty} \nu_k \text{ for some sequences } \nu_k \in \hat{\hat{N}}_{X(t_k)}(x_k)\,,\ (t_k, x_k) \xrightarrow[\mathrm{gph}\, X]{} (t,x)\Big\}. \tag{1.7}$$

A discussion of this cone and its relation to other formulations of the state constraint is given in Section 2 of our previous paper [13]. (In that work the same cone was denoted by $N(t, x)$; the change of notation here is meant to emphasize the fact that this cone is related to the multifunction X and that, like Clarke's normal cone, it has closed convex values.)

Hypothesis (H4) and definition (1.7) together imply that for any continuous function $x\colon [a, b] \to \mathbb{R}^n$ satisfying $x(t) \in X(t) \cap \Omega_t$ for all t, the convex cone valued multifunction $t \rightrightarrows \overline{N}_X(t, x(t))$ is Borel measurable. In this case, to call an $\mathbb{R}^n$-valued measure μ "$\overline{N}_X(t, x(t))$-valued" means that μ is absolutely continuous with respect to some nonnegative measure μ_0 on $[a, b]$, and that some measurable selection $\nu(t) \in \overline{N}_X(t, x(t))$ satisfies $d\mu(t) \equiv \nu(t) d\mu_0(t)$. (See Rockafellar [22, Section 5].)

Necessary conditions for optimality in which the adjoint function is merely of bounded variation, with jump directions described in terms of cone-valued measures, were first given for convex problems of Bolza by Rockafellar [23], [24], [26].

The Constraint Qualification. Our necessary conditions require that the cone $\overline{N}_X(t, \overline{x}(t))$ be pointed everywhere on the graph of the optimal arc $\overline{x}$. This constraint qualification is also essential in Clarke's formulation (see [2, Remark 3.2.7(iii)]), as explained by Loewen and Rockafellar [13, Section 2]. Let us call the state constraint "active" (relative to $\overline{x}$) at any time t when $(t, \overline{x}(t))$ lies on the boundary of $\mathrm{gph}\, X$, and "inactive" when $(t, \overline{x}(t))$ lies in the interior of $\mathrm{gph}\, X$. It follows easily from (1.7) that $\overline{N}_X(t, \overline{x}(t))$ collapses to the trivial cone $\{0\}$ if and only if the state constraint is inactive at time t. In particular, if the state constraint is inactive for all $t \in [a, b]$—perhaps because $X(t) \equiv \mathbb{R}^n$—then the constraint qualification mentioned above holds automatically. (Notice that there can be times when the state constraint is active even though $\overline{x}(t) \in \int X(t)$ for all t. An example is provided by the arc $\overline{x}(t) = 2t$ and the multifunction $X(t) = \{y : |y| \geq t\}$: the state constraint is active at $t = 0$ even though $\overline{x}(t) \in \int X(t)$ for all t.) Another common case in which the constraint qualification holds automatically arises when the state constraint sets $X(t)$ are convex and have nonempty interior, for then the cone $\overline{N}_X(t, x)$ coincides with the usual normal cone $N_{X(t)}(x)$ of convex analysis; and the latter cone is pointed if and only if $\int X(t) \neq \emptyset$.

The state constraint we impose can be given a simple geometric interpretation, based on Rockafellar [25, Thm. 3]. That result states that if a closed subset Ξ of $\mathbb{R}^n$ contains a point ξ at which the Clarke normal cone $\overline{N}_\Xi(\xi)$ is pointed, then there is a neighbourhood of ξ in which Ξ is indistinguishable from the isometric linear image of the epigraph of some Lipschitz function on $\mathbb{R}^{n-1}$. (The set Ξ is then called *epi-Lipschitzian at* ξ.) The set $\overline{N}_X(t, x)$ defined by (1.7) is generally larger than Clarke's normal cone $\overline{N}_{X(t)}(x)$ by (1.2)–(1.3), since it contains information not only about the shape of the set $X(t)$, but also about its behaviour as t varies. This leads to the following result, which makes precise the sense in which we can regard our constraint qualification as a requirement of "uniform epi-Lipschitzian behaviour" of the multifunction X.

1.1. PROPOSITION. *Let (t, x) be a point in $\Omega \cap \operatorname{gph} X$ at which the cone $\overline{N}_X(t, x)$ is pointed. Then there exist a neighbourhood U of (t, x) in $[a, b] \times \mathbb{R}^n$, a linear isometry A on $\mathbb{R}^n$, and a constant L with the following properties. For any $(s, y) \in U \cap \operatorname{gph} X$, there is a Lipschitz function $\phi_{(s,y)}$ of rank L having the property that*

$$A\left(\operatorname{epi} \phi_{(s,y)}\right) = X(s) \text{ near } y.$$

(In detail, this conclusion means that there is a neighbourhood V of y such that $A\left(\operatorname{epi} \phi_{(s,y)}\right) \cap V = X(s) \cap V$.)

Proof. This result follows from a careful quantitative analysis of the cited theorem of Rockafellar. Details are available in [12]; here we merely indicate the main steps in the proof.

For any $\varepsilon \in (0, 1)$ and any unit vector $v \in \mathbb{R}^n$, define the closed, pointed convex cone

$$K_\varepsilon(v) := \{\zeta \in \mathbb{R}^n : \langle \zeta, v \rangle \geq \varepsilon|\zeta|\}.$$

Deduce from the hypothesis that there exist some $\varepsilon \in (0, 1)$ and some v of unit length, together with a neighbourhood U_0 of (t, x) relative to Ω, such that

$$(*) \qquad \overline{N}_{X(s)}(y) \subseteq K_\varepsilon(v) \quad \forall (s, y) \in U_0.$$

Let A be any linear isometry of $\mathbb{R}^n$ into $\mathbb{R}^{n-1} \times \mathbb{R}$ such that $Av = (0, -1)$. (One certainly exists.) Then taking polars in $(*)$ gives

$$\overline{T}_{A(X(s))}(Ay) \supseteq K_{\varepsilon'}(0, 1) \quad \forall (s, y) \in U_0,$$

where $\varepsilon' = \sqrt{1 - \varepsilon^2}$. For each (s, y) in U_0, Rockafellar's proof of [25, Thm. 3], provides a Lipschitzian function $\phi_{(s,y)}$ on $\mathbb{R}^{n-1}$ whose epigraph coincides with $A(X(s))$ throughout some neighbourhood of Ay. The Lipschitz rank of $\phi_{(s,y)}$ can be estimated using the bound on the size of Clarke's generalized gradient of $\phi_{(s,y)}$ implicit in the identification with $A(X(s))$ and $\operatorname{epi} \phi_{(s,y)}$. (In fact the estimate gives $L = \frac{\varepsilon}{\sqrt{1-\varepsilon^2}}$.) The conclusion of the proposition now follows, but we have interchanged the names of A and A^{-1} for clarity. □

Notice that we have little control over the time-dependence of the functions $\phi_{(s,y)}$ in Proposition 1.1. For example, the multifunction

$$X(t) := \{(x,y) : y \geq 0\} \text{ if } t < 1/2, \qquad X(t) := \{(x,y) : y \geq 1\} \text{ if } t \geq 1/2$$

satisfies all our hypotheses but has a discontinuity at $t = 1/2$. Since X is convex-valued, $\overline{N}_X(\frac{1}{2},(0,1)) = N_{X(\frac{1}{2})}(0,1) = \{0\} \times (-\infty, 0]$. This cone is clearly pointed. When $s < \frac{1}{2}$ and y is near 1, the set $X(s)$ near y looks like the epigraph of $\phi_{(s,y)} \equiv 0$; but when $s \geq \frac{1}{2}$ and y is near 1, the set $X(s)$ near y looks like the epigraph of the function $\phi_{(s,y)} \equiv 1$.

Hamiltonian Necessary Conditions. We are now in a position to state the necessary conditions for bounded differential inclusions on which our main results are based. These involve the *Hamiltonian* associated with the multifunction F, defined by $H(t,x,p) := \sup\{\langle p\,,\,v\rangle : v \in F(t,x)\}$. Theorem 1.2, below, is essentially a transcription of [13], Thm. 2.8, except that the transversality inclusion in part (b) involves limiting subgradients and normals instead of the Clarke subgradients and normals used in [13]. (Clarke subgradients are still required in the Hamiltonian inclusion.) This distinction is immaterial in the smooth and convex cases, for which Clarke's notions are indistinguishable from the corresponding limiting constructions. In general, however, it is possible that the right-hand side of (b) is a proper subset of its counterpart in [13]. A detailed proof of Theorem 1.2 would be both long and repetitive, since many of the steps in the argument are now (or should be) well known. For this reason we simply outline a derivation of the result based on small adjustments to proofs in the literature. Theorem 4.1, below, will significantly weaken the boundedness and Lipschitz continuity hypotheses (i) and (ii) in the following statement.

1.2. THEOREM. *Assume (H1)–(H4). Suppose the arc $\overline{x}$ solves problem (P), and that the constraint qualification below is satisfied:*

$$(CQ) \qquad \text{the cone } \overline{N}_X(t,\overline{x}(t)) \text{ is pointed for all } t \text{ in } [a,b].$$

Suppose further that there exist integrable functions ϕ and k such that

$$\text{(i)} \quad F(t,x) \subseteq \phi(t)\,\mathrm{cl}\,\mathbb{B} \quad \forall (t,x) \in \Omega,$$
$$\text{(ii)} \quad F(t,y) \subseteq F(t,x) + k(t)|y-x|\,\mathrm{cl}\,\mathbb{B} \quad \forall t \in [a,b],\ x,y \in \Omega_t.$$

Then there exist a scalar $\lambda \in \{0,1\}$ and a function $p \in BV([a,b];\mathbb{R}^n)$, not both zero, together with an integrable selection $\nu(t) \in \overline{N}_X(t,\overline{x}(t))$ for all $t \in [a,b]$, such that

(a) $(-\dot{p}(t) + \nu(t), \dot{\overline{x}}(t)) \in \overline{\partial} H(t,\overline{x}(t),p(t))$ for almost all $t \in [a,b]$,

(b) $(p(a), -p(b)) \in \lambda\partial\ell(\overline{x}(a),\overline{x}(b)) + N_S(\overline{x}(a),\overline{x}(b))$;

(c) The singular part of the measure dp is $\overline{N}_X(t,\overline{x}(t))$-valued, and in particular is supported on the set

$$\{t : \overline{N}_X(t,\overline{x}(t)) \neq \{0\}\} = \{t \in [a,b] : (t,\overline{x}(t)) \in \mathrm{bdry\,gph}\,X\}.$$

Proof (Outline). Our first step is to reconsider Clarke's necessary conditions for optimality in [2, Thm. 3.5.2], noting that they apply to a slightly different problem than ours. We claim that these remain valid when the transversality condition at the final time [2,

p. 143, eq. (2)] is written as $-E \in N_{C_1}(x(b))$ instead of $-E \in \overline{N}_{C_1}(x(b))$. (That is, with the limiting normal cone in place of Clarke's normal cone.) To justify this, we must review the proof of [2, Thm. 3.4.3]. In the notation used there, choosing $x' = x$ in line (4) of Lemma 2 shows immediately that $-v$ is a proximal normal ("perpendicular" in [2]) to the set C_1 at the point c: hence the third displayed conclusion of Lemma 2 can be replaced by

$$\lambda\beta\zeta + p(b) + \int_{[a,b]} \gamma(s)\,\mu(ds) \in -\widehat{\widehat{N}}_{C_1}(x(b) - u)\,.$$

In the limiting analysis of Step 4, this relation becomes

$$\lambda\beta_0\zeta + \widetilde{p}(b) + \int_{[a,b]} \gamma(s)\,\widetilde{\mu}(ds) = \lambda v_0 \in -N_{C_1}(x(b))\,.$$

The proof of [2, Thm. 3.4.3] concludes as before, and the trick used to make the given solution unique employed in the proof of [2, Thm. 3.5.2] respects the refined formulation of the transversality condition.

Our second step is to extend the necessary conditions described above to cope with an endpoint cost functional and endpoint constraint set involving both $x(a)$ and $x(b)$ jointly. To do so, simply notice that if an arc $\overline{x}$ solves (P) then the extended arc $(\overline{r}, \overline{x})$ with $\overline{r}(t) \equiv \overline{x}(a)$ solves the following problem

$$\left[\begin{array}{l} \text{minimize } \ell(r(b), x(b)) \\ \text{subject to } \dot{r}(t) = 0,\ \dot{x}(t) \in F(t, x(t)) \text{ a.e. } t \in [a,b], \\ \qquad (r(a), x(a)) \in D,\ (r(b), x(b)) \in S, \\ \qquad (r(t), x(t)) \in \mathbb{R}^n \times X(t)\ \forall t \in [a,b], \end{array}\right.$$

where $D = \{(z,z) : z \in \mathbb{R}^n\}$ is the diagonal of $\mathbb{R}^n \times \mathbb{R}^n$. This is a situation to which our refinement of [2, Thm. 3.5.2] can be applied, and the resulting transversality condition is

$$(\dagger) \qquad (p(a), -p(b)) \in \lambda\overline{\partial}\ell(\overline{x}(a), \overline{x}(b)) + N_S(\overline{x}(a), \overline{x}(b))\,.$$

This differs from the desired conclusion (b) only in its use of the Clarke subgradient of the endpoint cost function ℓ.

Passing from the parametric form of the state constraint $g(t, x(t)) \le 0$ to the intrinsic form $x(t) \in X(t)$ as described in Section 2 of [13] does not affect the transversality inclusion. The same methods, therefore, imply that [13, Thm. 2.8] remains valid with the transversality inclusion replaced by (†)—and in particular that conclusion (b) holds if ℓ is smooth.

We now turn to the statement of Theorem 1.2. Suppose $\overline{x}$ solves problem (P). Then the extended arc $(\overline{x}, \overline{z})$ in which $\overline{z}(t) \equiv \ell(\overline{x}(a), \overline{x}(b))$ must solve the following problem

$$\left[\begin{array}{l} \text{minimize } z(b) \\ \text{subject to } \dot{x}(t) \in F(t, x(t)),\ \dot{z}(t) = 0 \text{ a.e. } t \in [a,b], \\ \qquad (x(a), x(b), z(b)) \in \operatorname{epi}(\ell + \Psi_S),\ z(a) \in \mathbb{R}, \\ \qquad (x(t), z(t)) \in X(t) \times \mathbb{R}\ \forall t \in [a,b]. \end{array}\right.$$

(Here Ψ_S denotes the *indicator function* of the set S, defined by setting $\Psi_S(x) = 0$ if x lies in S, and $\Psi_S(x) = +\infty$ otherwise.) Applying the intermediate form of [13, Thm. 2.8] described above leads to the conclusion that for some $\lambda \in \{0, 1\}$ and $p \in BV([a, b]; \mathbb{R}^n)$, not both zero, and some selection $\nu(t) \in \overline{N}_X(t, \overline{x}(t))$, we have the desired conclusions (a) and (c) of the current theorem, along with the transversality condition

$$(*) \qquad (p(a), -p(b), -\lambda) \in N_{\mathrm{epi}(\ell+\Psi_S)} \left(\overline{x}(a), \overline{x}(b), \ell(\overline{x}(a), \overline{x}(b))\right).$$

If $\lambda > 0$, this assertion is equivalent to

$$(p(a), -p(b)) \in \lambda\partial(\ell + \Psi_S)(\overline{x}(a), \overline{x}(b)) :$$

thanks to the calculus rules for limiting subgradients due to Ioffe [6, Thm. 4], we deduce that

$$(\ddagger) \qquad (p(a), -p(b)) \in \lambda\partial\ell(\overline{x}(a), \overline{x}(b)) + N_S\left(\overline{x}(a), \overline{x}(b)\right).$$

In particular, when $\lambda = 1$ we obtain the desired conclusion (b). When $\lambda = 0$, the proximal subgradient formula (Rockafellar [27, proof of Thm. 1]) asserts that the vector $(p(a), -p(b), 0)$ appearing on the left-hand side of $(*)$ can be expressed as a limit of some sequence $(p_k(a), -p_k(b), -\lambda_k)$ with $\lambda_k > 0$, along which $(*)$ holds relative to a sequence of base points $(x_k(a), x_k(b), \ell(x_k(a), x_k(b)))$ converging to $(\overline{x}(a), \overline{x}(b), \ell(\overline{x}(a), \overline{x}(b)))$. These sequences therefore satisfy an analogue of $(\ddagger)$ in which a subscript k appears throughout; taking the limit as $k \to \infty$ we obtain $(\ddagger)$ with $\lambda = 0$. Thus conclusion (b) is also valid in this case. □

Remark. The transversality condition $(*)$ from which (b) is derived in the foregoing proof could conceivably be sharper than (b) in some cases, since it involves the subgradients of the essential endpoint cost functional $\ell + \Psi_S$ instead of the sum of subgradients of its two terms.

2. Localization of Unbounded Multifunctions. Suppose the arc $\overline{x}$ solves problem (P). Then $\overline{x}$ must also solve any problem with the same objective function as (P) but fewer admissible arcs: such a problem can be described by replacing the given velocity sets $F(t, x)$ by their bounded subsets

$$(2.1) \qquad \widetilde{F}(t, x) := F(t, x) \cap \left(\dot{\overline{x}}(t) + R(t)\,\mathrm{cl}\,\mathbb{B}\right)$$

for some real-valued function $R(t)$. Known necessary conditions for differential inclusion problems with compact right-hand sides, like Theorem 1.2 above, then provide some information about $\overline{x}$. Our goal is to translate this information into necessary conditions which refer only to the data of the original problem (P). This translation is not completely straightforward; neither is it obvious which hypotheses on F and which choice of $R(t)$ will make the application of Theorem 1.2 to the reduced problem both legitimate and informative.

This section deals with the hypotheses: it is rather obvious that for any nonnegative integrable function $R(t)$, the truncated multifunction $\widetilde{F}$ defined above has compact convex values satisfying Theorem 1.2(i). But how can we ensure nonemptiness and Lipschitz continuity? We approach these by way of the following lemma, whose uncluttered notation is intended to clarify the essential geometry.

2.1. LEMMA. *Let $\overline{v} \in \mathbb{R}^n$, and let F_1 and F_2 be two subsets of $\mathbb{R}^n$ such that for some $\delta > 0$ and $0 < r < R$,*

(i) $F_2 \cap (\overline{v} + R \operatorname{cl} \mathbb{B}) \subseteq F_1 + \delta \operatorname{cl} \mathbb{B}$,

(ii) $F_1 \cap (\overline{v} + r \operatorname{cl} \mathbb{B}) \neq \emptyset$,

(iii) F_1 *is convex.*

Then $F_2 \cap (\overline{v} + R \operatorname{cl} \mathbb{B}) \subseteq F_1 \cap (\overline{v} + R \operatorname{cl} \mathbb{B}) + \left(\dfrac{2R\delta}{R - r}\right) \operatorname{cl} \mathbb{B}$.

Proof. Without loss of generality, take $\overline{v} = 0$. Choose any $v_2 \in F_2 \cap R \operatorname{cl} \mathbb{B}$. According to (i), there exists $v_1 \in F_1$ such that $|v_2 - v_1| \leq \delta$. We also have some $v_0 \in F_1$ such that $|v_0| \leq r$, thanks to (ii). And hypothesis (iii) ensures that

$$v_t := (1 - t)v_0 + tv_1 \in F_1 \qquad \forall t \in [a, b].$$

We estimate

$$\begin{aligned} |v_t| &\leq (1 - t)|v_0| + t|v_1| \\ &\leq (1 - t)r + t\,(\delta + |v_2|) \\ &\leq r + t(\delta + R - r). \end{aligned}$$

This implies that $v_t \in F_1 \cap R \operatorname{cl} \mathbb{B}$ whenever $r + t(\delta + R - r) \leq R$, in particular whenever

$$0 \leq t \leq \widehat{t} := \frac{R - r}{\delta + R - r}.$$

Let $\widehat{v} = v_{\widehat{t}}$. Then $\widehat{v} \in F_1 \cap (\overline{v} + R \operatorname{cl} \mathbb{B})$, and

$$\begin{aligned} |v_2 - \widehat{v}| &\leq |v_2 - v_1| + |v_1 - \widehat{v}| \\ &\leq \delta + (1 - \widehat{t})|v_0 - v_1| \\ &\leq \delta + (1 - \widehat{t})\,(|v_0| + |v_1|) \\ &\leq \delta + (1 - \widehat{t})(r + R + \delta) \\ &= 2\delta \left[1 + \frac{r}{\delta + R - r}\right]. \end{aligned}$$

The right-hand side increases if we discard the δ appearing in the denominator: this yields $|v_2 - \widehat{v}| \leq 2\delta R/(R - r)$. Since $v_2 \in F_2$ is arbitrary and $\widehat{v} \in F_1 \cap (\overline{v} + R \operatorname{cl} \mathbb{B})$, the desired inclusion follows. □

Remark. Clarke's Lemma 3 on p. 172 of [2] is proven by a very similar argument, but starts with a stronger hypothesis.

Using Lemma 2.1, we now provide a set of sufficient conditions for our localization technique to produce a multifunction satisfying the hypotheses of Theorem 1.2.

2.2. PROPOSITION. *Let Ω and F be given as in the formulation of problem (P); assume (H2)–(H3). Let $\overline{x}$ be an F-trajectory. Suppose there exists $\overline{\varepsilon} > 0$ together with* *nonnegative integrable functions* m *and* R *such that* $m/R \in L^\infty[a,b]$ *and for almost every* $t \in [a,b]$ *one has*

$$F(t,y) \cap \left(\dot{\overline{x}}(t) + R(t)\,\mathrm{cl}\,\mathbb{B}\right) \subseteq F(t,x) + m(t)|y-x|\,\mathrm{cl}\,\mathbb{B} \qquad \forall x,y \in \overline{x}(t) + \overline{\varepsilon}\mathbb{B}. \tag{2.2}$$

Then there is a relatively open subset $\widetilde{\Omega}$ *of* $[a,b]\times\mathbb{R}^n$ *containing the graph of* $\overline{x}$ *on which the truncated multifunction* $\widetilde{F}(t,x)$ *of (2.1) satisfies not only (H2)–(H3), but also hypotheses (i)–(ii) of Theorem 1.2.*

Proof. Notice that the requirement that $\overline{x}(t)+\overline{\varepsilon}\mathbb{B} \subseteq \Omega_t$ for all t is implicit in hypothesis (2.2), since the choice $y = \overline{x}(t)$ forces $F(t,x) \neq \emptyset$ for all $x \in \overline{x}(t) + \overline{\varepsilon}\mathbb{B}$. Therefore any choice of $\varepsilon \in (0,\overline{\varepsilon}]$ will ensure that $\mathrm{gph}\,\overline{x} \subseteq \widetilde{\Omega} \subseteq \Omega$ for the set

$$\widetilde{\Omega} := \{(t,x)\,:\, t\in[a,b], |x-\overline{x}(t)| < \varepsilon\}\,.$$

We therefore fix $\varepsilon \in (0,\overline{\varepsilon}]$, taking care to arrange that

$$\varepsilon m(t)/R(t) \le 1/2 \qquad \text{a.e. } t\in[a,b]. \tag{$*$}$$

(This is possible because m/R is essentially bounded by hypothesis.)

Let us fix a time $t \in [a,b]$ at which $(*)$ and (2.2) hold, $\dot{\overline{x}}(t)$ exists, and $\dot{\overline{x}}(t) \in F(t,\overline{x}(t))$. (Such t-values form a subset of $[a,b]$ with full measure.) The sets $\widetilde{F}(t,x)$ are evidently compact and convex valued for each $x \in \overline{x}(t)+\varepsilon\mathbb{B}$. To see that they are nonempty, choose $y = \overline{x}(t)$ in (2.2): then

$$\dot{\overline{x}}(t) \in F(t,\overline{x}(t)) \cap \left(\dot{\overline{x}}(t) + R(t)\,\mathrm{cl}\,\mathbb{B}\right) \subseteq F(t,x) + m(t)|\overline{x}(t)-x|\,\mathrm{cl}\,\mathbb{B} \qquad \forall x \in \overline{x}(t)+\varepsilon\mathbb{B}.$$

This inclusion implies that

$$F(t,x) \cap \left(\dot{\overline{x}}(t) + \varepsilon m(t)\,\mathrm{cl}\,\mathbb{B}\right) \neq \emptyset \qquad \forall x \in \overline{x}(t)+\varepsilon\mathbb{B}, \tag{$\dagger$}$$

and $\widetilde{F}(t,x)$ contains the left-hand side because of $(*)$. Thus (H2) holds relative to $\widetilde{\Omega}$; the measurability property required by (H3) is evident. As for condition (i) of Theorem 1.2, the choice $\phi(t) := \left|\dot{\overline{x}}(t)\right| + R(t)$ will clearly serve. Only condition (ii) remains to check. With t fixed as above, choose any $x,y \in \overline{x}(t)+\varepsilon\mathbb{B}$ and let $F_1 = F(t,x)$, $F_2 = F(t,y)$. Then the hypotheses of Lemma 2.1 hold with $\overline{v} = \dot{\overline{x}}(t)$, $R = R(t)$, $\delta = m(t)|y-x|$ from (2.2), and $r = \varepsilon m(t) \le R(t)/2$ from $(\dagger)$. The conclusion is that $\widetilde{F}(t,y) \subseteq \widetilde{F}(t,x) + k(t)|y-x|\,\mathrm{cl}\,\mathbb{B}$, where

$$k(t) = \frac{2R(t)m(t)}{R(t)-\varepsilon m(t)} = \frac{2m(t)}{1-\varepsilon m(t)/R(t)} \le 4m(t).$$

Hypothesis (ii) of Theorem 1.2 requires that the function k be integrable; this is ensured by the integrability of m. □

The central hypothesis (2.2) of Proposition 2.2 is a quantitative version of Aubin's *pseudo-Lipschitzian continuity* for the multifunctions $F(t,\cdot)$ at the points $(\overline{x}(t), \dot{\overline{x}}(t))$ along the trajectory $\overline{x}$ (see Rockafellar [28]). Although the conditions of Proposition 2.2 are sufficient for the development of our theory, they require that the arc $\overline{x}$ be known in advance, and offer few suggestions about effective choices of the functions R and m. Before continuing the development, we pause to describe hypotheses on the multifunction F which can be used to verify (2.2) along any admissible arc. These involve the following concepts.

2.3. Definition. Let $\Gamma\colon\Omega \rightrightarrows \mathbb{R}^m$ be a multifunction with closed values, and suppose Γ is $\mathcal{L}\times\mathcal{B}$ measurable on Ω. Consider a point $(\overline{t},\overline{x})$ in Ω.

(a) The multifunction Γ is called *sub-Lipschitzian at* $(\overline{t},\overline{x})$ if for every constant $\rho \geq 0$, there exist constants $\varepsilon > 0$ and $\alpha \geq 0$ such that

$$\Gamma(t,y)\cap\rho\,\mathrm{cl}\,\mathbb{B} \subseteq \Gamma(t,x) + \alpha|y-x|\,\mathrm{cl}\,\mathbb{B} \tag{2.3}$$

for all $t \in (\overline{t}-\varepsilon,\overline{t}+\varepsilon)\cap[a,b]$ and all x,y in $\overline{x}+\varepsilon\mathbb{B}$.

(b) The multifunction Γ is called *integrably sub-Lipschitzian in the large at* $(\overline{t},\overline{x})$ if there exist constants $\varepsilon > 0$ and $\beta \geq 0$, together with a nonnegative function α integrable on $(\overline{t}-\varepsilon,\overline{t}+\varepsilon)$, such that

$$\Gamma(t,y)\cap\rho\,\mathrm{cl}\,\mathbb{B} \subseteq \Gamma(t,x) + (\alpha(t)+\beta\rho)|y-x|\,\mathrm{cl}\,\mathbb{B} \tag{2.4}$$

for all $t \in (\overline{t}-\varepsilon,\overline{t}+\varepsilon)\cap[a,b]$, all x,y in $\overline{x}+\varepsilon\mathbb{B}$, and all $\rho \geq 0$.

Definition 2.3(a) is very similar to the notion of sub-Lipschitzian behaviour introduced by Rockafellar [28]—the only difference being that here we consider multifunctions with explicit time-dependence, and require a certain uniformity of the parameters ε and α with respect to t. Rockafellar [28] offers a detailed discussion of (autonomous) sub-Lipschitzian multifunctions and the relationship between this property and the pseudo-Lipschitz continuity introduced by Aubin; he also describes several classes of sub-Lipschitzian multifunctions.

Definition 2.3(b) introduces a new type of sub-Lipschitzian assumption even in the autonomous case. It looks like a stricter hypothesis than that of Definition 2.3(a), because it places certain restrictions on the growth of the right-hand side with ρ. If (b) holds for a constant function $\alpha(t)\equiv\alpha$, then certainly (a) follows; it is not obvious that (b) always implies (a), however, since (b) allows α to depend on t, whereas α must be constant in (a). Each of these hypotheses has a role as a sufficient condition for the applicability of Proposition 2.2.

2.4. Proposition. *Let Ω and F be given as in the formulation of problem (P); assume (H2)–(H3). Let $\overline{x}$ be an F-trajectory. Under either of the two hypotheses below, all the conditions of Proposition 2.2 are met. In particular, there is a relatively open subset $\widetilde{\Omega}$ of Ω containing the graph of $\overline{x}$ on which the truncated multifunction $\widetilde{F}$ defined by (2.1) satisfies all the hypotheses of Theorem 1.2.*

(a) The arc $\overline{x}$ is Lipschitzian, and the multifunction F is sub-Lipschitzian at every point $(t,\overline{x}(t))$ in $\mathrm{gph}\,\overline{x}$.

(b) The multifunction F is integrably sub-Lipschitzian in the large at every point $(t,\overline{x}(t))$ in $\mathrm{gph}\,\overline{x}$.

2.5. Remarks. 1. Notice that the two parts of Proposition 2.4 correspond exactly to the two parts of Definition 2.3. Part (b) imposes apparently stricter conditions on F and applies to any arc $\overline{x}$, while part (a) imposes apparently weaker requirements on F but pertains only to Lipschitzian arcs $\overline{x}$.

2. The proof of part (a) below allows for an arbitrarily small positive constant value of R in the localization of (2.1): this may eventually link our results with the necessary conditions for "weak local minima" in the calculus of variations.

3. The conditions of the Proposition make explicit reference to the arc $\overline{x}$, but they would obviously follow from corresponding hypotheses regarding sub-Lipschitzian behaviour of F throughout the set Ω.

4. Proposition 2.4 remains valid when Definition 2.3 is weakened by replacing the phrase "$\rho \geq 0$" with "$\rho \geq 0$ sufficiently large" in parts (a) and (b).

Proof (Proposition 2.4). (b) Both hypotheses in the statement of the Proposition must first be extended to the whole interval $[a, b]$ by a compactness argument. We illustrate this just once, taking the more delicate case, situation (b). Applying Definition 2.3(b) to a point $(s, \overline{x}(s))$ in $\operatorname{gph} \overline{x}$ yields constants $\varepsilon_s > 0$, $\beta_s \geq 0$, and a nonnegative function $\alpha_s(t)$ integrable on $(s - \varepsilon_s, s + \varepsilon_s)$ such that (2.4) holds for any $\rho \geq 0$ and any triple (t, x, y) chosen from the set

$$G_s = \{(t, x, y) \,:\, |t - s| < \varepsilon_s,\ x \in \overline{x}(s) + \varepsilon_s \mathbb{B},\ y \in \overline{x}(s) + \varepsilon_s \mathbb{B}\}\,.$$

Now each set G_s, $s \in [a, b]$, is open, and the family of these sets covers the compact set $\{(t, \overline{x}(t), \overline{x}(t)) \,:\, t \in [a, b]\}$. Therefore we may extract a finite subcover indexed by $s_1, \ldots, s_N$, and define

$$G := \bigcup_{j=1}^{N} G_j.$$

For simplicity, we have written G_{s_j} as G_j: we define ε_j, β_j, and $\alpha_j(t)$ similarly. Observe that there exists $\overline{\varepsilon} > 0$ so small that for every $t \in [a, b]$,

$$\{t\} \times (\overline{x}(t) + \overline{\varepsilon}\mathbb{B}) \times (\overline{x}(t) + \overline{\varepsilon}\mathbb{B}) \subseteq G.$$

(If this were not true, then there would be a sequence of points outside G converging to some point $(\overline{t}, \overline{x}(\overline{t}), \overline{x}(\overline{t}))$ in the interior of G, a contradiction.) Next, choose $\beta = \max\{\beta_1, \ldots, \beta_N\}$ and define

$$\alpha(t) := \max_{j=1,\ldots,N} \{\alpha_j(t) \,:\, t \in (s_j - \varepsilon_j, s_j + \varepsilon_j)\}\,.$$

Clearly β is finite and $\alpha(t)$ is integrable. Moreover, for any pair of points (t, x) and (t, y) chosen from the set $\widehat{\Omega} = \{(t, x) \,:\, t \in [a, b],\ x \in \overline{x}(t) + \overline{\varepsilon}\mathbb{B}\}$, we have $(t, x, y) \in G$, so $(t, x, y) \in G_j$ for some $j = 1, \ldots, N$. Thus (2.4) holds for (t, x, y) with parameters β_j and $\alpha_j(t)$, and therefore it holds with the larger parameters β and $\alpha(t)$. This shows that the set $\widehat{\Omega}$ is a relatively open subset of Ω containing the graph of $\overline{x}$ throughout which (2.4) holds uniformly with parameters $\overline{\varepsilon}$, β, and $\alpha(t)$.

Now consider the function $R(t) = 1 + \alpha(t) + \left|\dot{\overline{x}}(t)\right|$. For any t in $[a, b]$, choosing $\rho = \left|\dot{\overline{x}}(t)\right| + R(t)$ in the extension of (2.4) just proved leads to the following estimate, valid for all x, y in $\overline{x}(t) + \overline{\varepsilon}\mathbb{B}$:

$$F(t,y) \cap \big(\dot{\overline{x}}(t) + R(t)\,\mathrm{cl}\,\mathbb{B}\big) \subseteq F(t,y) \cap \big(\big|\dot{\overline{x}}(t)\big| + R(t)\big)\,\mathrm{cl}\,\mathbb{B}$$
$$\subseteq F(t,x) + \big(\alpha(t) + \beta\,\big[\big|\dot{\overline{x}}(t)\big| + R(t)\big]\big)\,|y-x|\,\mathrm{cl}\,\mathbb{B}.$$

This confirms (2.2), where the function

$$m(t) = \alpha(t) + \beta\,\big[\big|\dot{\overline{x}}(t)\big| + R(t)\big] = (1+\beta)\alpha(t) + 2\beta\big|\dot{\overline{x}}(t)\big| + \beta$$

is clearly integrable, as is $R(t)$, while $m(t)/R(t) \leq 2 + 2\beta$ almost everywhere. All the hypotheses of Proposition 2.2 are in place; part (b) of the desired result follows.

(a) Under hypothesis (a), we fix any $R > 0$ (perhaps quite small) and let $\rho = R + \big\|\dot{\overline{x}}\big\|_{\infty}$. Then a compactness argument very similar to the one described in detail above leads to a pair of constants $\overline{\varepsilon} > 0$ and $\alpha \geq 0$ for which (2.3) holds for any pair of points (t,x) and (t,y) in $\widehat{\Omega} := \{(t,x) \,:\, t \in [a,b], |x - \overline{x}(t)| < \overline{\varepsilon}\}$. In particular, since $\rho \geq R + \big|\dot{\overline{x}}(t)\big|$, we have

$$F(t,y) \cap \big(\dot{\overline{x}}(t) + R\,\mathrm{cl}\,\mathbb{B}\big) \subseteq F(t,y) \cap \rho\,\mathrm{cl}\,\mathbb{B}$$
$$\subseteq F(t,x) + \alpha|y-x|\,\mathrm{cl}\,\mathbb{B} :$$

thus (2.2) holds with the constants $m = \alpha$ and R identified here. □

3. Hamiltonian Calculus. We now take up the second question raised at the beginning of Section 2. Given a multifunction F satisfying our standing hypotheses, and an F-trajectory $\overline{x}$, suppose it is possible to choose a function $R(t)$ for which the localization (2.1) produces a multifunction $\widetilde{F}$ with suitable boundedness and Lipschitz properties. What is the relationship between the Hamiltonian of $\widetilde{F}$ and that of the given multifunction F? More specifically, how are their subgradients linked? We answer these questions using simplified notation which suppresses the time-dependence of F, since we are concerned only with partial subgradients computed at fixed times.

Throughout this section, we consider a multifunction F defined on some neighbourhood $\overline{x} + \overline{\varepsilon}\mathbb{B}$ of a given point $\overline{x}$, and taking on nonempty closed convex subsets of $\mathbb{R}^n$ as values. We assume that $F(x)$ depends continuously on x in the set $\overline{x} + \overline{\varepsilon}\mathbb{B}$, in the sense that the inner and outer limits of the sets $F(x')$ share the common value $F(x)$ as $x' \to x$. Given a point $\overline{v}$ in $F(\overline{x})$, we consider the localized multifunction $\widetilde{F}(x) := F(x) \cap (\overline{v} + R\,\mathrm{cl}\,\mathbb{B})$ for some fixed $R > 0$. Like its predecessor F, the multifunction $\widetilde{F}$ has closed convex values on the set $\overline{x} + \overline{\varepsilon}\mathbb{B}$. Concerning $\widetilde{F}$, we assume that

$$\widetilde{F} \text{ is Lipschitz of rank } k \text{ on } \overline{x} + \overline{\varepsilon}\mathbb{B} \tag{3.1}$$

(in particular $\widetilde{F}$ is nonempty-valued there), and consider a vector $\overline{p}$ with the property that

$$\langle \overline{p}\,,\, \overline{v}\rangle \geq \langle \overline{p}\,,\, v\rangle \qquad \text{for all } v \in \widetilde{F}(\overline{x}). \tag{3.2}$$

Our concern is to relate the subgradients at $(\overline{x}, \overline{p})$ of the two Hamiltonians corresponding to F and $\widetilde{F}$, namely

$$\begin{aligned} H(x,p) &:= \sup\{\langle p\,,\, v\rangle \,:\, v \in F(x)\},\\ \widetilde{H}(x,p) &:= \sup\{\langle p\,,\, v\rangle \,:\, v \in \widetilde{F}(x)\}. \end{aligned} \tag{3.3}$$

In particular, we plan to prove that under the hypotheses described above,

$$\overline{\partial}\widetilde{H}(\overline{x},\overline{p}) \subseteq \overline{\partial}H(\overline{x},\overline{p}). \tag{3.4}$$

Recall the Legendre-Fenchel transform, which associates to any function $f\colon \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ its *conjugate*

$$f^*(p) := \sup\left\{\langle p\,,\,v\rangle - f(v) \;:\; v\in\mathbb{R}^n\right\}.$$

When f is a proper convex function, its conjugate is too, and the duality sponsored by this transformation is the cornerstone of many fundamental results in convex analysis. In our current setting, we recognize the Hamiltonians H and $\widetilde{H}$ as the conjugates of certain indicator functions: using the notation $\Psi_C(v) := 0$ if $v\in C$ and $\Psi_C(v) := +\infty$ if $v\notin C$, we have

$$\begin{aligned} H(x,p) &= \left(\Psi_{F(x)}\right)^*(p),\\ \widetilde{H}(x,p) &= \left(\Psi_{\widetilde{F}(x)}\right)^*(p) = \left(\Psi_{F(x)} + \Psi_{\overline{v}+R\,\mathrm{cl}\,\mathbb{B}}\right)^*(p). \end{aligned}$$

One important consequence of this observation is that the possibly extended-valued function H is lower semicontinuous on $(\overline{x}+\overline{\varepsilon}\mathbb{B})\times\mathbb{R}^n$. Indeed, the continuity of F assumed above implies that $\operatorname{epi}\Psi_{F(x)}$ varies continuously with x on $\overline{x}+\overline{\varepsilon}\mathbb{B}$. According to Wijsman's theorem [32, Thm. 6.2], epi-continuity is preserved under the Legendre-Fenchel transform. In particular, $\operatorname{epi} H(x,\cdot)$ also varies continuously with x on $\overline{x}+\overline{\varepsilon}\mathbb{B}$. The very definition of epi-continuity now implies that H is lower semicontinuous near $(\overline{x},\overline{p})$.

Infimal Convolution. Addition of proper convex functions corresponds to infimal convolution of their conjugates under the Legendre-Fenchel transform: according to *Convex Analysis* [21, Thm. 16.4], we have the following identity for all x near $\overline{x}$ and all $p\in\mathbb{R}^n$:

$$\begin{aligned} \widetilde{H}(x,p) &= \left(\Psi^*_{F(x)} \,\square\, \Psi^*_{\overline{v}+R\,\mathrm{cl}\,\mathbb{B}}\right)(p)\\ &= \inf\left\{\Psi^*_{F(x)}(p-z) + \Psi^*_{\overline{v}+R\,\mathrm{cl}\,\mathbb{B}}(z) \;:\; z\in\mathbb{R}^n\right\}\\ &= \inf\left\{H(x,p-z) + \langle\overline{v}\,,\,z\rangle + R|z| \;:\; z\in\mathbb{R}^n\right\}. \end{aligned} \tag{3.5}$$

(The hypotheses of [21, Thm. 16.4] require that for each x near $\overline{x}$, the convex sets $\operatorname{ri}(\operatorname{dom} H(x,\cdot))$ and $\operatorname{ri}(\operatorname{dom}\widetilde{H}(x,\cdot))$ have a point in common. But since $\widetilde{F}$ is bounded, the latter set is the whole space $\mathbb{R}^n$; the former set is nonempty, so this hypothesis holds.)

Subgradient Analysis. Equation (3.5) expresses $\widetilde{H}$ as the value function associated with a minimization problem depending upon the parameters x and p. Proximal analysis is a powerful technique for estimating the subgradients of such functions: the situation we now face is covered by Rockafellar [29, Thm. 8.3]. If we write $f(z,x,p) := H(x,p-z) + \langle\overline{v}\,,\,z\rangle + R|z|$, that result affirms that

$$\begin{aligned} \overline{\partial}\widetilde{H}(\overline{x},\overline{p}) \subseteq \mathrm{cl}\,\mathrm{co}\Bigg[&\bigcup_{z\in\Sigma(\overline{x},\overline{p})} \left\{(\pi,v) \;:\; (0,\pi,v)\in\overline{\partial}f(z,\overline{x},\overline{p})\right\}\\ &+ \bigcup_{z\in\Sigma(\overline{x},\overline{p})} \left\{(\pi,v) \;:\; (0,\pi,v)\in\overline{\partial}^{\infty}f(z,\overline{x},\overline{p})\right\}\Bigg], \end{aligned} \tag{3.6}$$

where $\Sigma(\overline{x},\overline{p})$ denotes the set of all points $z \in \mathbb{R}^n$ at which the infimum in (3.5) is attained. Our assumption (3.2) and the convexity of the set $F(\overline{x})$ together ensure that one such point is $z = 0$, where $\widetilde{H}(\overline{x},\overline{p}) = \langle \overline{p}\,,\, \overline{v}\rangle = H(\overline{x},\overline{p})$. But since $\overline{v} \in F(\overline{x})$, we have $H(\overline{x},\overline{p}-z) \geq \langle \overline{p}-z\,,\, \overline{v}\rangle$ for any $z \in \mathbb{R}^n$, whence

$$H(\overline{x},\overline{p}-z) + \langle \overline{v}\,,\, z\rangle + R|z| \geq H(\overline{x},\overline{p}) + R|z|.$$

The right-hand side here strictly exceeds the minimum value $H(\overline{x},\overline{p})$ for all z except $z = 0$: therefore $\Sigma(\overline{x},\overline{p}) = \{0\}$, and inclusion (3.6) simplifies to

$$\overline{\partial}\widetilde{H}(\overline{x},\overline{p}) \subseteq \operatorname{cl}\operatorname{co}\Big[\{(\pi,v)\,:\,(0,\pi,v)\in\overline{\partial}f(0,\overline{x},\overline{p})\} + \Big\{(\pi,v)\,:\,(0,\pi,v)\in\overline{\partial}^{\infty}f(0,\overline{x},\overline{p})\Big\}\Big]. \tag{3.7}$$

Before completing our analysis of (3.7), we must verify that the derivation of (3.6) from [29, Thm. 8.3] is justified. This requires that we check three hypotheses. First, the function $\widetilde{H}$ must be finite at $(\overline{x},\overline{p})$: this requires only that $\widetilde{F}(\overline{x})$ be nonempty, which we have assumed from the start. Second, a certain constraint qualification must hold at $(z,\overline{x},\overline{p})$ for every $z \in \Sigma(\overline{x},\overline{p})$: this turns out to be trivial, since the constraint structure of our problem is so much simpler than that involved in the general situation of the cited theorem. Third, there must exist constants $\varepsilon > 0$ and $\overline{\alpha} > \widetilde{H}(\overline{x},\overline{p})$ such that the following set is bounded:

$$S := \{(z,x,p)\,:\, H(x,p-z) + \langle \overline{v}\,,\, z\rangle + R|z| \leq \overline{\alpha},\ |(x,p)-(\overline{x},\overline{p})| \leq \varepsilon\}\,.$$

To prove this, we apply the Lipschitz hypothesis (3.1), which implies (since $\overline{v} \in \widetilde{F}(\overline{x})$) that for any x in $\overline{x} + \varepsilon\mathbb{B}$,

$$(\overline{v} + k|x-\overline{x}|\operatorname{cl}\mathbb{B}) \cap \widetilde{F}(x) \neq \emptyset.$$

It follows that for any such x, and for any $p, z \in \mathbb{R}^n$,

$$H(x,p-z) \geq \widetilde{H}(x,p-z) \geq \langle p-z\,,\, \overline{v}\rangle - k|x-\overline{x}||p-z|.$$

Therefore if we choose $\varepsilon > 0$ small enough that $R - \varepsilon k > \varepsilon/2$, any triple (z,x,p) satisfying the defining inequalities in S will obey

$$\begin{aligned}\overline{\alpha} &\geq H(x,p-z) + \langle \overline{v}\,,\, z\rangle + R|z| \\ &\geq R|z| + \langle p\,,\, \overline{v}\rangle - k|x-\overline{x}||p-z| \\ &\geq (R - k|x-\overline{x}|)\,|z| - (|\overline{v}| + k|x-\overline{x}|)\,|p| \\ &\geq (\varepsilon/2)|z| - (|\overline{v}| + k\varepsilon)(|\overline{p}| + \varepsilon).\end{aligned}$$

This clearly imposes an upper bound on $|z|$, and it follows that the set S is bounded. Our verification of the hypotheses of [29, Thm. 8.3] is complete, and we can apply its conclusions (3.6) and (3.7) with confidence.

To complete our derivation of inclusion (3.4), it remains only to compute the subgradient sets appearing in (3.7). Recall that $f(z,x,p) = H(x,p-z) + \langle \overline{v}\,,\ z\rangle + R|z|$: thus f is the sum of a lower semicontinuous function and a continuous convex function. According to Thm. 8.1 (the sum rule) and Cor. 7.1.2 (the chain rule) of Rockafellar [29], we have

$$\begin{aligned}\overline{\partial} f(0,\overline{x},\overline{p}) &\subseteq \left\{(-v,\pi,v) \,:\, (\pi,v)\in\overline{\partial} H(\overline{x},\overline{p})\right\} + (\overline{v} + R\,\mathrm{cl}\,\mathbb{B})\times\{(0,0)\}\\ \overline{\partial}^{\infty} f(0,\overline{x},\overline{p}) &\subseteq \left\{(-v,\pi,v) \,:\, (\pi,v)\in\overline{\partial}^{\infty} H(\overline{x},\overline{p})\right\}.\end{aligned}$$

Thus (3.7) yields

(3.8)

$$\begin{aligned}\overline{\partial}\widetilde{H}(\overline{x},\overline{p}) &\subseteq \mathrm{cl}\,\mathrm{co}\left(\left[\overline{\partial} H(\overline{x},\overline{p})\cap(\mathbb{R}^n\times(\overline{v}+R\,\mathrm{cl}\,\mathbb{B}))\right] + \left[\overline{\partial}^{\infty} H(\overline{x},\overline{p})\cap(\mathbb{R}^n\times\{0\})\right]\right)\\ &\subseteq \mathrm{cl}\,\mathrm{co}\left(\overline{\partial} H(\overline{x},\overline{p}) + \overline{\partial}^{\infty} H(\overline{x},\overline{p})\right).\end{aligned}$$

Since the set on the left side is nonempty, the set on the right must also be nonvoid. This forces $\overline{\partial} H(\overline{x},\overline{p}) \neq \emptyset$, a situation in which $\overline{\partial}^{\infty} H(\overline{x},\overline{p})$ is known to equal the recession cone of $\overline{\partial} H(\overline{x},\overline{p})$. In particular, $\overline{\partial} H(\overline{x},\overline{p}) + \overline{\partial}^{\infty} H(\overline{x},\overline{p})$ is a subset of the closed convex set $\overline{\partial} H(\overline{x},\overline{p})$. Therefore (3.8) implies (3.4), and the objective of this section is accomplished.

4. General Necessary Conditions.

We now combine the efforts of the first three sections to prove necessary conditions for optimality in (P) without the boundedness and Lipschitz continuity assumptions used previously. Our first result, Theorem 4.1, extends the Hamiltonian necessary conditions of Theorem 1.2 to the unbounded case. Although this is a significant advance in itself, it is superseded by Theorem 4.3 below, in which the same hypotheses are used to produce an adjoint function satisfying the Hamiltonian inclusion, a refined Euler-Lagrange inclusion, and the Weierstrass-Pontryagin maximum condition simultaneously. Our purpose in proving Theorem 4.1 first is to clarify the roles of Sections 2 and 3 in eliminating boundedness assumptions. This provides a convenient point to reflect on what has been achieved, and to gather strength for the next step.

Hypotheses (H1)–(H4) mentioned in the statement below are listed in Section 1; the notions required in assumptions (i) and (ii) are described in Definition 2.3. As observed in Section 2, assumptions (i) and (ii) can be replaced by stronger hypotheses requiring appropriate sub-Lipschitzian behaviour at every point of Ω if the arc $\overline{x}$ is not known in advance.

4.1. Theorem (Hamiltonian Necessary Conditions). *Assume (H1)–(H4). Suppose that the arc $\overline{x}$ solves problem (P), and that the constraint qualification below is satisfied:*

(CQ) *the cone $\overline{N}_X(t,\overline{x}(t))$ is pointed for all t in $[a,b]$.*

Suppose further that one of the following two conditions holds:

(i) The arc $\overline{x}$ is Lipschitzian, and the multifunction F is sub-Lipschitzian at every point $(t,\overline{x}(t))$ of $\mathrm{gph}\,\overline{x}$; or

(ii) The multifunction F is integrably sub-Lipschitzian in the large at every point $(t,\overline{x}(t))$ of $\mathrm{gph}\,\overline{x}$.

Then there exist a scalar $\lambda \in \{0,1\}$ *and a function* $p \in BV([a,b];\mathbb{R}^n)$*, not both zero, such that one has*
(a) the Hamiltonian inclusion

$$(-\dot{p}(t), \dot{\overline{x}}(t)) \in \overline{\partial} H(t, \overline{x}(t), p(t)) - \overline{N}_X(t, \overline{x}(t)) \times \{0\} \text{ a.e. } t \in [a,b],$$

(b) the transversality inclusion

$$(p(a), -p(b)) \in \lambda \partial \ell(\overline{x}(a), \overline{x}(b)) + N_S\left(\overline{x}(a), \overline{x}(b)\right), \text{ and}$$

(c) the singular part of the measure dp is $\overline{N}_X(t,\overline{x}(t))$*-valued, and in particular is supported on the set*

$$\left\{t \,:\, \overline{N}_X(t,\overline{x}(t)) \neq \{0\}\,\right\} = \{t \in [a,b] \,:\, (t, \overline{x}(t)) \in \operatorname{bdry} \operatorname{gph} X\}.$$

Remark. The interpretation of inclusion (a) in Theorem 4.1 is the same as that given in Theorem 1.2. That is, (a) asserts that for some integrable selection $\nu(t) \in \overline{N}_X(t,\overline{x}(t))$ for all $t \in [a,b]$, one has

$$(-\dot{p}(t) + \nu(t), \dot{\overline{x}}(t)) \in \overline{\partial} H(t, \overline{x}(t), p(t)) \text{ a.e. } t \in [a,b].$$

Proof. Under either hypothesis (i) or (ii), Proposition 2.4 describes a choice of $R(t)$ for which the truncated multifunction $\widetilde{F}(t,x) := F(t,x) \cap \left(\dot{\overline{x}}(t) + R(t) \operatorname{cl} \mathbb{B}\right)$ satisfies both assumptions (i) and (ii) of Theorem 1.2. Of course the arc $\overline{x}$ is a trajectory for $\widetilde{F}$, and consequently solves the problem $(\widetilde{P})$ defined by replacing F with $\widetilde{F}$ in (P). Apply Theorem 1.2 to $\overline{x}$ in $(\widetilde{P})$: this produces a constant λ and an adjoint function p of bounded variation, not both zero, together with a selection $\nu(t)$ of $\overline{N}_X(t,\overline{x}(t))$, satisfying all the conclusions of Theorem 1.2. Let us denote these by $(\widetilde{\text{a}})$–$(\widetilde{\text{c}})$, since they involve the multifunction $\widetilde{F}$ and its associated Hamiltonian $\widetilde{H}$. We will show that these three conditions imply the desired conclusions (a)–(c) for the same λ, p, and ν. Indeed, conditions $(\widetilde{\text{b}})$ and $(\widetilde{\text{c}})$ are the same as the desired assertions (b) and (c), while $(\widetilde{\text{a}})$ implies (a). To justify the latter assertion, fix $t \in (a,b)$ and consider the multifunctions $F(t,\cdot)$ and $\widetilde{F}(t,\cdot)$. Our assumption of either (i) or (ii) implies that the given multifunction $F(t,\cdot)$ is continuous in the weak sense required in Section 3, and that the truncated multifunction $\widetilde{F}(t,\cdot)$ satisfies the Lipschitz condition (3.1) (see Proposition 2.4). Hypothesis (3.2) for $\overline{p} = p(t)$ is a well-known consequence of $(\widetilde{\text{a}})$—see Clarke [2, Prop. 3.2.4(d)]. The conclusion is that for almost every time $t \in [a,b]$, $\overline{\partial}\widetilde{H}(t,\overline{x}(t),p(t)) \subseteq \overline{\partial} H(t,\overline{x}(t),p(t))$: hence (a) follows from $(\widetilde{\text{a}})$, as required. □

Strict Convexity. The crucial observation which allowed us to unify the adjoint inclusions of Hamilton, Euler-Lagrange, and Weierstrass-Pontryagin in [13] was that the Hamiltonian inclusion actually implies the other two inclusions when $\dot{\overline{x}}(t)$ is almost always an extreme point of the (convex) velocity set $F(t,\overline{x}(t))$. We now use the same observation to extend Theorem 4.1. Let us continue under the hypotheses of that result.

Consider the function

$$L(t,v) := \sqrt{1+\left|v-\dot{\overline{x}}(t)\right|^2} - 1.$$

Notice that L is nonnegative, smooth, and strictly convex, with $L(t,\dot{\overline{x}}(t)) \equiv 0$ and $L_v(t,\dot{\overline{x}}(t)) \equiv 0$. Observe also that for each fixed t, the function $L(t,\cdot)$ is globally Lipschitzian of rank 1 on $\mathbb{R}^n$. These properties are important in our analysis of the following auxiliary problem, whose state (x,y) evolves in $\mathbb{R}^n \times \mathbb{R}$:

$$(\mathcal{P}) \qquad \left[\begin{array}{l} \text{minimize } \ell(x(a),x(b)) + y(b) \\ \text{subject to } (\dot{x}(t),\dot{y}(t)) \in [F(t,x(t)) \times \mathbb{R}] \cap \operatorname{epi} L(t,\cdot) \text{ a.e. } t \in [a,b], \\ \qquad\qquad (x(a),x(b)) \in S,\ y(a) = 0, \\ \qquad\qquad (x(t),y(t)) \in X(t) \times \mathbb{R}\ \forall t \in [a,b], \end{array}\right.$$

It is clear that any absolutely continuous function $(x(t),y(t))$ admissible for the auxiliary problem $(\mathcal{P})$ has a first component admissible for the original problem (P), while the second component obeys $y(b) \geq 0$. Therefore the objective value in $(\mathcal{P})$ is always at least as large as the objective value in (P). But the arc $(\overline{x},\overline{y})$ for which $\overline{y}(t) \equiv 0$ is admissible for $(\mathcal{P})$, and it has an objective value equal to the minimum value in (P): therefore it must be optimal in $(\mathcal{P})$.

The dynamic constraint in $(\mathcal{P})$ involves the unbounded multifunction $\mathcal{F}\colon \Omega\times\mathbb{R} \rightrightarrows \mathbb{R}^n \times \mathbb{R}$ defined by

$$\mathcal{F}(t,x,y) := [F(t,x) \times \mathbb{R}] \cap \operatorname{epi} L(t,\cdot).$$

We now show that $\mathcal{F}$ inherits the sub-Lipschitzian property of F along $\overline{x}$, and consequently admits a truncation displaying the boundedness and Lipschitz continuity properties required for the application of Theorem 1.2.

4.2. LEMMA. *Suppose that hypothesis (i) or (ii) of Theorem 4.1 holds for the multifunction F relative to the arc $\overline{x}$. Then the same hypothesis holds for $\mathcal{F}$ relative to $\overline{x}$. Moreover, there exists a nonnegative function R such that both truncated multifunctions below satisfy hypotheses (i) and (ii) of Theorem 1.2 on some relatively open subset of $[a,b] \times \mathbb{R}^n$ containing the graph of $\overline{x}$:*

$$\widetilde{F}(t,x) := F(t,x) \cap \left[\dot{\overline{x}}(t) + R(t)\operatorname{cl}\mathbb{B}\right],$$
$$\widetilde{\mathcal{F}}(t,x) := \mathcal{F}(t,x,y) \cap \left[(\dot{\overline{x}}(t),0) + R(t)(\operatorname{cl}\mathbb{B} \times [-1,1])\right].$$

Proof. (ii) Suppose F satisfies hypothesis 4.1(ii) relative to $\overline{x}$. Let any $\overline{t} \in [a,b]$ be given. Then by hypothesis, there must be constants $\varepsilon > 0$ and $\beta \geq 0$, together with a nonnegative function α integrable on $(\overline{t}-\varepsilon,\overline{t}+\varepsilon) \cap [a,b]$ such that

$$(*) \qquad F(t,x') \cap \rho\operatorname{cl}\mathbb{B} \subseteq F(t,x) + (\alpha(t) + \beta\rho)\left|x'-x\right|\operatorname{cl}\mathbb{B}$$

for all $t \in (\bar{t}-\varepsilon, \bar{t}+\varepsilon) \cap [a,b]$, all x, x' in $\overline{x}(\bar{t}) + \varepsilon \mathbb{B}$, and all $\rho \geq 0$. To prove a similar statement involving $\mathcal{F}$, let any $\rho \geq 0$ and $t \in (\bar{t}-\varepsilon, \bar{t}+\varepsilon) \cap [a,b]$ be given, together with any two points $x, x' \in \overline{x}(\bar{t}) + \rho \operatorname{cl} \mathbb{B}$. Then for any point (v', r') in $\mathcal{F}(t, x') \cap \rho(\operatorname{cl} \mathbb{B} \times [-1,1])$, we have $v' \in F(t, x') \cap \rho \operatorname{cl} \mathbb{B}$, so $(*)$ provides a point $v \in F(t,x)$ such that $|v' - v| \leq (\alpha(t) + \beta\rho)|x' - x|$. Now $L(t, \cdot)$ is Lipschitz of rank 1, and $r' \geq L(t, v')$. Hence there must be a point $r \geq L(t, v)$ for which $|r' - r| \leq |v' - v|$. Thus (v, r) is a point in $\mathcal{F}(t, x)$ for which

$$\begin{aligned} |(v', r') - (v, r)| &\leq |v' - v| + |r' - r| \\ &\leq 2|v' - v| \\ &\leq 2(\alpha(t) + \beta\rho)|x' - x|. \end{aligned}$$

Since (v', r') is arbitrary, this argument proves that

$$(**) \qquad \mathcal{F}(t, x') \cap \rho\,(\operatorname{cl} \mathbb{B} \times [-1,1]) \subseteq \mathcal{F}(t,x) + 2\,(\alpha(t) + \beta\rho)\,|x' - x|\,(\operatorname{cl} \mathbb{B} \times [-1,1])\,.$$

Hypothesis 4.1(ii) for $\mathcal{F}$ follows.

Now if we multiply both α and β in $(*)$ by 2, we find that both multifunctions F and $\mathcal{F}$ are integrably sub-Lipschitzian in the large at $(\bar{t}, \overline{x}(\bar{t}))$ with the same choices of ε, $2\alpha(t)$, and 2β in the definition. Reviewing the proof of Proposition 2.4, we deduce that the function $R(t) := 1 + 2\alpha(t) + |\dot{\overline{x}}(t)|$ provides a truncation radius for which each multifunction $\widetilde{F}$, $\widetilde{\mathcal{F}}$ satisfies the hypotheses (i) and (ii) of Theorem 1.2 on some neighbourhood of $\operatorname{gph} \overline{x}$. We restrict attention to the intersection of these two neighbourhoods to obtain the desired conclusion.

(i) If F satisfies hypothesis 4.1(i) relative to $\overline{x}$, then an argument similar to that just given shows that for every point $\bar{t} \in [a,b]$ and every $\rho \geq 0$, there exist constants $\varepsilon > 0$ and $\alpha \geq 0$ such that

$$(\dagger) \qquad F(t,y) \cap \rho \operatorname{cl} \mathbb{B} \subseteq F(t,x) + 2\alpha|y - x| \operatorname{cl} \mathbb{B}$$

$$(\ddagger) \qquad \mathcal{F}(t,y) \cap \rho\,(\operatorname{cl} \mathbb{B} \times [-1,1]) \subseteq \mathcal{F}(t,x) + 2\alpha|y - x|\,(\operatorname{cl} \mathbb{B} \times [-1,1])$$

for all $t \in (\bar{t}-\varepsilon, \bar{t}+\varepsilon) \cap [a,b]$ and all x, y in $\overline{x}(\bar{t}) + \varepsilon \mathbb{B}$. Just as above, the proof of Proposition 2.4 shows that any constant value of $R > 0$ will provide a truncation radius suitable for both multifunctions F and $\mathcal{F}$ at once. □

Now, just as in the proof of Theorem 4.1, the arc $(\overline{x}, 0)$ which solves $(\mathcal{P})$ remains optimal for the problem $(\widetilde{\mathcal{P}})$ obtained from $(\mathcal{P})$ by changing $\mathcal{F}$ to $\widetilde{\mathcal{F}}$. We apply Theorem 1.2 to deduce that there exist a scalar $\lambda \geq 0$ and a function $(p,q)\colon [a,b] \to \mathbb{R}^n \times \mathbb{R}$ of bounded variation, not both zero, together with a selection $\nu(t) \in \overline{N}_X(t, \overline{x}(t))$ for all $t \in [a,b]$ such that

(a) $(-\dot{p}(t) + \nu(t), \dot{\overline{x}}(t), 0) \in \overline{\partial}\widetilde{\mathcal{H}}(t, \overline{x}(t), p(t), q(t))$ a.e. $t \in [a,b]$,
$\dot{q}(t) = 0$ a.e. $t \in [a,b]$,

(b) $(p(a), -p(b)) \in \lambda \partial \ell(\overline{x}(a), \overline{x}(b)) + N_S\,(\overline{x}(a), \overline{x}(b))$; $q(b) = -\lambda$;

(c) The singular part of the measure (dp, dq) is $\overline{N}_X(t, \overline{x}(t)) \times \{0\}$-valued, and in particular is supported on the set

$$\left\{t \,:\, \overline{N}_X(t, \overline{x}(t)) \neq \{0\}\right\} = \{t \in [a,b] \,:\, (t, \overline{x}(t)) \in \operatorname{bdry} \operatorname{gph} X\}\,.$$

Here we have used the fact that $\widetilde{\mathcal{F}}$ is independent of y to simplify the Hamiltonian inclusion; we need only deal with the reduced Hamiltonian given by

$$\widetilde{\mathcal{H}}(t,x,p,q) := \sup\left\{\langle p\,,\,v\rangle + qr \,:\, (v,r)\in\widetilde{\mathcal{F}}(t,x)\right\}.$$

Conditions (a)–(c) together imply that the adjoint function's $q(t)$ component is actually constant, with the value $-\lambda$. Thus conclusions (b) and (c) reduce to the expected transversality and support conditions associated with the adjoint function p, while conclusion (a) may be written as follows:

$$(-\dot{p}(t)+\nu(t),\dot{\overline{x}}(t),0)\in\overline{\partial}\widetilde{\mathcal{H}}(t,\overline{x}(t),p(t),-\lambda)\text{ a.e. }t\in[a,b]. \tag{4.1}$$

In the remainder of this section we use inclusion (4.1) to show that the function p satisfies the Hamiltonian inclusion, a refined Euler-Lagrange inclusion, and the Weierstrass-Pontryagin maximum condition for the original problem (P).

The Maximum Condition. We have seen in Lemma 4.2 that $\widetilde{\mathcal{F}}$ is a multifunction satisfying hypotheses (i) and (ii) of Theorem 1.2. Under these assumptions, Clarke [2, Prop. 3.2.4(d)] shows that inclusion (4.1) implies

$$(p(t),-\lambda)\in N_{\widetilde{\mathcal{F}}(t,\overline{x}(t))}\left(\dot{\overline{x}}(t),0\right)\quad\text{a.e. }t\in[a,b]. \tag{4.2}$$

But for each $t\in[a,b]$, the compact set $\widetilde{\mathcal{F}}(t,\overline{x}(t))$ coincides with the unbounded set $\mathcal{F}(t,\overline{x}(t))$ on a neighbourhood of $(\dot{\overline{x}}(t),0)$: hence these two sets have the same normal cone at this point. Using the calculus of convex normal cones (Rockafellar [21], or [29, Cor. 8.1.1]), we deduce that for almost every $t\in[a,b]$,

$$\begin{aligned}(p(t),-\lambda)&\in N_{\mathcal{F}(t,\overline{x}(t))}\left(\dot{\overline{x}}(t),0\right)\\&=N_{F(t,\overline{x}(t))\times\mathbb{R}\cap\operatorname{epi}L(t,\cdot)}\left(\dot{\overline{x}}(t),0\right)\\&\subseteq N_{F(t,\overline{x}(t))\times\mathbb{R}}\left(\dot{\overline{x}}(t),0\right)+N_{\operatorname{epi}L(t,\cdot)}\left(\dot{\overline{x}}(t),0\right)\\&=N_{F(t,\overline{x}(t))}\left(\dot{\overline{x}}(t)\right)\times\{0\}+\{0\}\times(-\infty,0].\end{aligned}$$

(The last step uses the fact that $N_{\operatorname{epi}L(t,\cdot)}\left(\dot{\overline{x}}(t),0\right)$ is the convex cone generated by $\partial L(t,\dot{\overline{x}}(t))\times\{-1\}=\{(0,-1)\}$.) The first component of this inclusion gives the desired maximum condition for p, namely,

$$p(t)\in N_{F(t,\overline{x}(t))}\left(\dot{\overline{x}}(t)\right)\text{ a.e. }t\in[a,b]. \tag{4.3}$$

The Hamiltonian Inclusion. Observe that the function $\widetilde{\mathcal{H}}$ can be written as follows:

$$\widetilde{\mathcal{H}}(t,x,p,q)=\begin{cases}\sup\{\langle p\,,\,v\rangle+qL(t,v)\,:\,v\in\widetilde{F}(t,x)\}, & \text{if } q<0,\\ \sup\{\langle p\,,\,v\rangle\,:\,v\in\widetilde{F}(t,x)\}+qR(t), & \text{if } q\ge 0.\end{cases}$$

This is precisely the sort of function studied in Section 4 of our previous work [13], where we examined its relationship to the function below:

$$\widetilde{\mathcal{H}}_\lambda(t,x,p):=\sup\left\{\langle p\,,\,v\rangle-\lambda L(t,v)\,:\,v\in\widetilde{F}(t,x)\right\}.$$

Lemma 4.2 ensures that for each fixed $t \in [a,b]$, the multifunction $\widetilde{F}(t,\cdot)$ in this expression obeys the standing assumptions (A1)–(A3) of [13, Section 4]. This observation allows us to apply [13, Thm. 4.4] to inclusion (4.1), and thereby derive

$$(-\dot{p}(t)+\nu(t), \dot{\overline{x}}(t)) \in \overline{\partial}\widetilde{\mathcal{H}}_\lambda(t, \overline{x}(t), p(t)) \text{ a.e. } t \in [a,b]. \tag{4.4}$$

In the case where $\lambda = 0$, $\widetilde{\mathcal{H}}_\lambda$ coincides with $\widetilde{H}$, so inclusion (4.4) is equivalent to

$$(-\dot{p}(t)+\nu(t), \dot{\overline{x}}(t)) \in \overline{\partial}\widetilde{H}(t, \overline{x}(t), p(t)) \text{ a.e. } t \in [a,b] \tag{4.5}$$

In the case where $\lambda = 1$, inclusion (4.4) implies (4.5), thanks to [13, Cor. 4.3(b)]. To justify this, fix any time t where (4.4) holds and apply Clarke [2, Prop. 2.5.3] to deduce that $\dot{\overline{x}}(t) \in \partial_p\widetilde{\mathcal{H}}_1(t, \overline{x}(t), p(t))$. Conversely, elementary convex analysis shows that any vector v lying in $\partial_p\widetilde{\mathcal{H}}_1(t, \overline{x}(t), p(t))$ must maximize the function $v' \mapsto \langle p\,,\, v'\rangle - L(t,v')$ over the set $\widetilde{F}(t, \overline{x}(t))$. Since this function is strictly concave by construction, only one maximizer can exist, namely $\dot{\overline{x}}(t)$. Consequently $\overline{\partial}_p\widetilde{\mathcal{H}}_1(t, \overline{x}(t), p(t)) = \{\dot{\overline{x}}(t)\}$. The union appearing in [13, Cor. 4.3(b)] therefore involves the choices $v = \dot{\overline{x}}(t)$ and $z \in \overline{\partial}_v L(t, \dot{\overline{x}}(t)) = \{0\}$: this implies that the right-hand side of (4.4) is a subset of the right-hand side of (4.5).

With (4.5) in hand, we note that the Hamiltonian analysis of Section 3 allows us to replace $\widetilde{H}$ with H in (4.5) just as we did in the proof of Theorem 4.1. The result is the desired Hamiltonian inclusion for p:

$$(-\dot{p}(t)+\nu(t), \dot{\overline{x}}(t)) \in \overline{\partial}H(t, \overline{x}(t), p(t)) \text{ a.e. } t \in [a,b]. \tag{4.6}$$

The Euler-Lagrange Inclusion. Once again we rely upon the technical results of Loewen and Rockafellar [13], as extended by Rockafellar [30]. To streamline the discussion, we fix a time $t \in [a,b]$ at which (4.1) holds, and suppress t in the notation below. Thus our starting point is the inclusion

$$(-\dot{p}+\nu, \dot{\overline{x}}, 0) \in \overline{\partial}\widetilde{\mathcal{H}}(\overline{x}, p, -\lambda). \tag{4.7}$$

Rockafellar [30, Thm. 3.1] provides a far-reaching analogue of [13, Lemma 4.5] in which limiting normals and subgradients replace Clarke normals and subgradients. Applying this result to our problem (with $f = \Psi_{\widetilde{\mathcal{F}}}$), we find that

$$\begin{aligned} -\partial(-\widetilde{\mathcal{H}})(\overline{x}, p, -\lambda) \subseteq \{(-u, v, \ell)\, :(u, p, -\lambda) &\in N_{\operatorname{gph}\widetilde{\mathcal{F}}}(\overline{x}, v, \ell)\,, \\ (p, -\lambda) &\in N_{\widetilde{\mathcal{F}}(x)}(v, \ell)\,\}. \end{aligned} \tag{4.8}$$

Since the function $\widetilde{\mathcal{H}}$ is Lipschitzian, we have $\overline{\partial}\widetilde{\mathcal{H}} = -\overline{\partial}(-\widetilde{\mathcal{H}}) = \operatorname{co} -\partial(-\widetilde{\mathcal{H}})$. Thus it must be possible to express the point $(-\dot{p}+\nu, \dot{\overline{x}}, 0)$ as a convex combination of elements from the right side of (4.8). That is, there must be some $N \in N$ and some constants $\alpha_i \geq 0$ with $\sum \alpha_i = 1$ such that

$$(-\dot{p}+\nu, \dot{\overline{x}}, 0) = \sum_{i=1}^{N} \alpha_i(-u_i, v_i, \ell_i), \tag{4.9a}$$

where, for each i,

$$(u_i, p, -\lambda) \in N_{\operatorname{gph} \widetilde{\mathcal{F}}}(\overline{x}, v_i, \ell_i), \qquad (p, -\lambda) \in N_{\widetilde{\mathcal{F}}(\overline{x})}(v_i, \ell_i). \tag{4.9b}$$

We have already shown that $(p, -\lambda) \in N_{\widetilde{\mathcal{F}}(\overline{x})}\left(\dot{\overline{x}}, 0\right)$ (see (4.2)). We now add the observation that $(\dot{\overline{x}}, 0)$ is an extreme point of the set $\widetilde{\mathcal{F}}(\overline{x})$. This is obvious, since $\widetilde{\mathcal{F}}(\overline{x})$ is the intersection of the compact convex sets $\operatorname{epi} L$ and $\widetilde{F}(\overline{x}) \times \mathbb{R}$, and $(\dot{\overline{x}}, 0)$ is an extreme point of the first of these by the strict convexity of L. It follows that $(\dot{\overline{x}}, 0)$ is the only point (v, ℓ) for which $(p, -\lambda) \in N_{\widetilde{\mathcal{F}}(\overline{x})}(v, \ell)$. This forces $(v_i, \ell_i) = (\dot{\overline{x}}, 0)$ in (4.9), and thus implies

$$\dot{p} - \nu \in \operatorname{co}\left\{u \,:\, (u, p, -\lambda) \in N_{\operatorname{gph} \widetilde{\mathcal{F}}}\left(\overline{x}, \dot{\overline{x}}, 0\right)\right\}. \tag{4.10}$$

To simplify this assertion, temporarily think of L as a function of both x and v ($L(x, v) \equiv L(v)$) in order to write $\operatorname{gph} \widetilde{\mathcal{F}} = \operatorname{epi}\left(L + \Psi_{\operatorname{gph} \widetilde{F}}\right)$. This allows us to transcribe the inclusion characterizing the right-hand side of (4.10) as

$$(u, p, -\lambda) \in N_{\operatorname{epi}(L + \Psi_{\operatorname{gph} \widetilde{F}})}\left(\overline{x}, \dot{\overline{x}}, 0\right).$$

The same arguments used in the last paragraph of the proof of Theorem 1.2, together with the observation that $\partial L(\overline{x}, \dot{\overline{x}}) = \{(0, 0)\}$, show that this inclusion implies

$$(u, p) \in N_{\operatorname{gph} \widetilde{F}}\left(\overline{x}, \dot{\overline{x}}\right) = N_{\operatorname{gph} F}\left(\overline{x}, \dot{\overline{x}}\right).$$

The equality here holds because the sets $\operatorname{gph} \widetilde{F}$ and $\operatorname{gph} F$ coincide on a neighbourhood of the point $(\overline{x}, \dot{\overline{x}})$, so their limiting normal cones at this point are identical. Using this statement in (4.10) leads to the following inclusion, in which we revert to fully explicit notation:

$$\dot{p}(t) - \nu(t) \in \operatorname{co}\left\{u \,:\, (u, p(t)) \in N_{\operatorname{gph} F(t, \cdot)}\left(\overline{x}(t), \dot{\overline{x}}(t)\right)\right\}. \tag{4.11}$$

This inclusion holds for all t outside a null subset of $[a, b]$: it is the form of the Euler-Lagrange inclusion we wish to record.

The Main Result. We now summarize the results of the derivation above. Conclusions (a)–(c) in the following formal statement have already been established as lines (4.3), (4.6), and (4.11) above. We remind the reader that our notation differs slightly from that of Clarke [2], as indicated in the last paragraph of the introduction.

4.3. THEOREM (GENERAL NECESSARY CONDITIONS). *Assume (H1)–(H4). Suppose that the arc $\overline{x}$ solves problem (P), and that the constraint qualification below is satisfied:*

(CQ) *the cone $N_X(t, \overline{x}(t))$ is pointed for all t in $[a, b]$.*

Suppose further that the one of the following two conditions holds:

(i) The arc $\overline{x}$ *is Lipschitzian, and the multifunction* F *is sub-Lipschitzian at every point* $(t,\overline{x}(t))$ *of* $\operatorname{gph}\overline{x}$*; or*
(ii) The multifunction F *is integrably sub-Lipschitzian in the large at every point* $(t,\overline{x}(t))$ *of* $\operatorname{gph}\overline{x}$*.*

Then there exist a scalar $\lambda \in \{0,1\}$ *and a function* $p \in BV([a,b];\mathbb{R}^n)$*, not both zero, such that for almost all* $t \in [a,b]$*, one has*
(a) the Hamiltonian inclusion

$$(-\dot{p}(t),\dot{\overline{x}}(t)) \in \overline{\partial} H(t,\overline{x}(t),p(t)) - N_X(t,\overline{x}(t)) \times \{0\}\,,$$

(b) the Euler-Lagrange inclusion

$$\dot{p}(t) \in \operatorname{co}\left\{u \,:\, (u,p(t)) \in N_{\operatorname{gph} F(t,\cdot)}\left(\overline{x}(t),\dot{\overline{x}}(t)\right)\right\} + \overline{N}_X(t,\overline{x}(t)),$$

(c) the Weierstrass-Pontryagin maximum condition

$$\left\langle p(t)\,,\,\dot{\overline{x}}(t)\right\rangle = \max\left\{\langle p(t)\,,\,v\rangle \,:\, v \in F(t,\overline{x}(t))\right\}.$$

The adjoint function p *also satisfies*
(d) the transversality inclusion

$$(p(a),-p(b)) \in \lambda\partial\ell(\overline{x}(a),\overline{x}(b)) + N_S\left(\overline{x}(a),\overline{x}(b)\right), \text{ and}$$

(e) the singular part of the measure dp *is* $N_X(t,\overline{x}(t))$*-valued, and in particular is supported on the set*

$$\{t \,:\, N_X(t,\overline{x}(t)) \neq \{0\}\} = \{t \in [a,b] \,:\, (t,\overline{x}(t)) \in \operatorname{bdry}\operatorname{gph} X\}.$$

4.4. Remarks. 1. If the state constraint is inactive along the optimal arc $\overline{x}$ (i.e., $\overline{x}(t) \in \int X(t)$ for all $t \in [a,b]$) and the endpoint constraint set S has the form $C \times \mathbb{R}^n$ or $\mathbb{R}^n \times D$ for some closed sets C, D, then one may take $\lambda = 1$ in Theorem 4.3. This is not completely obvious from the theorem's statement, but it does follow from the proof given above. To see this, note that the scalar λ and the function p described in the conclusions of the theorem actually arise as the dual variables in the auxiliary problem $(\widetilde{\mathcal{P}})$. Problem $(\widetilde{\mathcal{P}})$ has a Hamiltonian $\widetilde{\mathcal{H}}$ for which the mapping $x' \mapsto \mathcal{H}(t,x',p)$ is Lipschitz of rank $k(t)|p|$ for some integrable function k. Under the extra assumptions above we have $\nu(t) \equiv 0$, so the adjoint function p must be absolutely continuous and satisfy the differential inequality $|\dot{p}(t)| \le k(t)|p(t)|$ almost everywhere. Now if the endpoint conditions described above are satisfied, then assuming $\lambda = 0$ leads to either $p(a) = 0$ or else $p(b) = 0$: in either case, Gronwall's lemma implies $p(t) \equiv 0$, a contradiction.

2. A separated form of the Hamiltonian inclusion can be asserted concurrently with (a)–(c) above. To derive it, note that since $\widetilde{H}$ is locally Lipschitz, inclusion (4.5) implies

$$(-\dot{p}(t)+\nu(t),\dot{\overline{x}}(t)) \in \operatorname{co}\left[\partial_x\widetilde{H}(t,\overline{x}(t),p(t)) \times \partial_p\widetilde{H}(t,\overline{x}(t),p(t))\right] \text{ a.e. } t \in [a,b].$$

The second component simply reiterates (4.3), while the first asserts that

$$-\dot{p}(t) + \nu(t) \in \operatorname{co} \partial_x \widetilde{H}(t, \overline{x}(t), p(t)) \text{ a.e. } t \in [a, b].$$

Arguments similar to those in Section 3 allow us to replace $\widetilde{H}$ with H in this statement: the result is the separated Hamiltonian inclusion

$$\begin{aligned} -\dot{p}(t) &\in \operatorname{co} \partial_x H(t, \overline{x}(t), p(t)) - \overline{N}_X(t, \overline{x}(t)), \\ \dot{\overline{x}}(t) &\in \partial_p H(t, \overline{x}(t), p(t)) \text{ a.e. } t \in [a, b]. \end{aligned} \tag{4.12}$$

5. Epilogue. Theorem 4.3 is the most general set of necessary conditions available for differential inclusion control problems. To substantiate this claim, we review the literature and discuss several pertinent examples in this section.

The Bounded Case. Notice first that Theorem 4.3 is a strict extension of our best previous result for bounded differential inclusions, the case $L \equiv 0$ of [13, Thm. 1.1]. To prove this, it suffices to show that the boundedness and Lipschitz continuity hypotheses of [13], which coincide with conditions (i)–(ii) of the current Theorem 1.2, imply the hypotheses of Theorem 4.3. Indeed, suppose that the multifunction F satisfies condition (ii) of Theorem 1.2. Then the choices $\beta = 0$ and $\alpha = k$ in Definition 2.3(b) show that F is integrably sub-Lipschitzian in the large at every point (t, x) in Ω. Hence hypothesis (ii) of Theorem 4.3 is satisfied; the conclusions either reproduce those of [13, Thm. 1.1] or else are strictly stronger. Notice in particular that hypothesis (i) is completely superfluous both in [13] and in Theorem 1.2.

The two conclusions of Theorem 4.3 which differ from their counterparts in [13, Thm. 1.1] are the Euler-Lagrange inclusion (b) and the transversality inclusion (d). The right-hand side of (d) is always a subset of its cognate phrased in terms of Clarke subgradients and normals, although the two right-hand sides coincide whenever the function ℓ and the set S are Clarke regular at the point $(\overline{x}(a), \overline{x}(b))$. Likewise, the Euler-Lagrange inclusion (b) readily implies (but may not be equivalent to) the more familiar form involving Clarke's normal cone:

$$(\dot{p}(t), p(t)) \in \operatorname{co} N_{\operatorname{gph} F}\left(\overline{x}(t), \dot{\overline{x}}(t)\right) - \overline{N}_X(t, \overline{x}(t)) \times \{0\} \text{ a.e.} \tag{5.1}$$

The formulation in (b) has the advantage of applying the convex hull only to variables associated with the derivatives of the adjoint function p. The routine use of weak convergence of the derivatives both in existence theory and in the derivation of necessary conditions makes it hard to imagine making do with less convexity than this.

The possibility of refining the Euler-Lagrange inclusion in our main theorem was suggested by a recent preprint of Boris Mordukhovich, a pioneer in the systematic reduction of convexity hypotheses in nonsmooth analysis. His manuscript [18] introduces a version of the Euler-Lagrange inclusion whose counterpart in our problem would read as follows:

$$\begin{aligned} (\dot{p}(t) - \nu(t), \dot{\overline{x}}(t)) \in \operatorname{co}\{(u, v) \ : (u, p(t)) &\in N_{\operatorname{gph} F(t,\cdot)}\left(\overline{x}(t), v\right), \\ p(t) &\in N_{F(t, \overline{x}(t))}\left(v\right) \}. \end{aligned} \tag{5.2}$$

(Here, as in Section 4, $\nu(t)$ is a selection of $\overline{N}_X(t,\overline{x}(t))$: Mordukhovich's work does not allow for state constraints, so his version of (5.2) involves an absolutely continuous function p and $\nu \equiv 0$.) This is clearly a consequence of inclusion (4.11). The two are equivalent if, for almost every t, the maximum value of $\langle p(t)\,,\, v\rangle$ over $v \in F(t,\overline{x}(t))$ is attained at the unique point $v = \dot{\overline{x}}(t)$. Without this hypothesis, however, the right-hand side of (5.2) may be a proper superset of the right-hand side of (4.11): this is demonstrated by Example 5.2 below. Thus the necessary conditions of Mordukhovich [18] are strictly superseded by those given here. Indeed, Rockafellar's dualization result [30, Thm. 3.1] used to prove the Euler-Lagrange inclusion (4.11) implies that under Mordukhovich's hypotheses in [18], inclusion (5.2) actually follows from the Hamiltonian inclusion in Theorem 1.2.

Although several technical results from our previous work [13] were used to prove Theorem 4.3, this paper's development starts from [13], Thm. 2.8. Since we recover [13, Thm. 1.1] as a corollary (at least in the case $L \equiv 0$), this paper provides an much simpler alternative to the formidable sequential arguments of [13, Section 3]. This makes sense, because the sequences of adjoint functions required there arose directly out of a less sophisticated truncation procedure than the one introduced in the current work.

The simultaneous assertion of the adjoint inclusion in both Hamiltonian and Eulerian forms is a significant feature Theorem 4.3 shares with the main result of [13]. The relationship between these two inclusions in their various forms is still not completely understood. For example, we now show that the Euler-Lagrange inclusion in Clarke's form (5.1) bears no simple relationship to the Hamiltonian inclusion.

5.1. EXAMPLE. *There exist a compact convex valued, Lipschitzian multifunction $F\colon \mathbb{R} \rightrightarrows \mathbb{R}$ and a pair of arcs x, p on $[a,b]$ such that for all $t \in [a,b]$ the Clarke form of the Euler-Lagrange inclusion holds, i.e.,*

$$(\dot{p}(t), p(t)) \in \overline{N}_{\mathrm{gph}\, F}\,(x(t), \dot{x}(t))\,, \tag{5.3a}$$

but the following two inclusions fail:

$$(\dot{p}(t), \dot{x}(t)) \in \mathrm{co}\left\{(u,v)\,:\,(u,p(t)) \in N_{\mathrm{gph}\, F}\,(x(t),v)\,,\ p(t) \in N_{F(x(t))}\,(v)\,\right\}, \tag{5.3b}$$

$$(-\dot{p}(t), \dot{x}(t)) \in \overline{\partial} H(x(t), p(t)). \tag{5.3c}$$

Proof. Let $F(x) := [-|x|, |x|]$. This multifunction is compact convex valued and Lipschitz continuous; its graph is the plane set obtained by filling in the vertical space between the lines $y = x$ and $y = -x$. The limiting normal cone to $\mathrm{gph}\, F$ at the point $(0,0)$ consists of the two lines $y = \pm x$ in the plane; the corresponding Clarke normal cone is therefore the whole space $\mathbb{R}^2$. Thus for the arc $x(t) \equiv 0$, the right-hand side in Clarke's form of the Euler-Lagrange inclusion (5.3a) is simply $\mathbb{R}^2$, so any arc p will serve. On the other hand, Mordukhovich's form of the Euler-Lagrange inclusion (5.3b) makes a nontrivial restriction on the choice of p. The inclusion $p(t) \in N_{F(0)}\,(v)$ forces $v = 0$, so that (5.3b) becomes

$$(\dot{p}(t), 0) \in \mathrm{co}\,\{(u,0)\,:\,|u| = |p(t)|\} = [-|p(t)|, |p(t)|] \times \{0\}\,.$$

Any arc p which obeys $|\dot{p}(t)| > |p(t)|$ will confirm (5.3a) but violate (5.3b). For example, $p(t) = e^{2t}$ will serve. By the result of Rockafellar cited above, inclusion (5.3c) implies (5.3b): hence the same choice of p must also violate (5.3c). Of course, this can be confirmed directly by noting that the Hamiltonian corresponding to F is

$$H(x,p) = \sup\{pv \,:\, |v| \le |x|\} = |px|,$$

and that for $p \neq 0$, one has $\overline{\partial} H(0,p) = [-|p|, |p|] \times \{0\}$. □

Notice that in Example 5.1, the Mordukhovich form of the Euler-Lagrange inclusion is equivalent to the refined form used in Theorem 4.3 because $F(0)$ is a one-point set. Thus Example 5.1 shows that the Euler-Lagrange inclusion in Clarke's form (5.1) does not imply the Hamiltonian inclusion, and that (5.1) can be strictly weaker than our refined Euler-Lagrange inclusion (4.11). However, it does not rule out the possibility that (4.11) implies the Hamiltonian inclusion.

Our next example shows that the Hamiltonian inclusion does not imply either Euler-Lagrange inclusion (4.11) or (5.1) "pointwise"; recall, however, that the Hamiltonian inclusion does imply the Euler-Lagrange inclusion (5.2) in Mordukhovich's form.

5.2. Example. *There exist a compact convex valued, Lipschitzian multifunction $F\colon \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ and a collection of points x, v, p, u in $\mathbb{R}^2$ such that*

$$(-u, v) \in \overline{\partial} H(x,p) \qquad \textit{but} \qquad (u,p) \notin \overline{N}_{\operatorname{gph} F}(x,v)\,.$$

In particular,

$$u \notin \operatorname{co}\{u' \,:\, (u',p) \in N_{\operatorname{gph} F}(x,v)\}\,.$$

Proof. Define $F\colon \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ as follows:

$$F(x_1,x_2) := \{(t, t|x_1| + r) \,:\, t \in [-1,1], r \in [a,b]\}\,.$$

For each $x = (x_1,x_2)$ in $\mathbb{R}^2$, the set $F(x)$ is a solid parallelogram in the plane. The corresponding Hamiltonian is

$$H(x_1,x_2,p_1,p_2) = \big|p_1 + p_2|x_1|\big| + \max\{p_2, 0\}\,.$$

We consider the points $x = (0,0)$, $v = (0,0)$, and $p = (0,-1)$. With these choices, $F(x)$ is the plane rectangle $[-1,1] \times [a,b]$, and p is an outward normal vector to this set at the boundary point v. The crucial feature of this example is that the hyperplane $x_2 = 0$ which supports the set $F(x)$ at v intersects the set $F(x)$ in more than one point. (In other words, the maximum of $\langle p\,,\, v\rangle$ over v in $F(x)$ is attained at infinitely many points.) Clarke's subgradient of H at the point $(x,p) = (0,0,0,-1)$ can be calculated using [2, Thm. 2.5.1]: it is the two-dimensional square $[-1,1]\times\{0\}\times[-1,1]\times\{0\}$ in $\mathbb{R}^4$. One point in this square is $(1,0,0,0)$, which suggests the choice $u = (-1,0)$. We claim that $(u,p) = (-1,0,0,-1)$ lies outside $\overline{N}_{\operatorname{gph} F}(x,v)$. To prove this, notice that up to a permutation of the coordinates, $\operatorname{gph} F = E \times \mathbb{R}$ for the set $E := \{(x_1, t, |x_1|t + r) \,:\, x_1 \in \mathbb{R},\ t \in [-1,1],\ r \in [a,b]\}$. Near the point $(0,0,0)$, E coincides with the epigraph of the function $g\colon \mathbb{R}^2 \to \mathbb{R}$ defined by $g(y,t) := t|y|$. This function g is Lipschitzian, and it is easy to show that $\overline{\partial} g(0,0) = \{(0,0)\}$. Therefore

$$\overline{N}_E(0,0,0) = \overline{N}_{\operatorname{epi} g}(0,0,g(0,0)) = \bigcup_{\lambda \ge 0} \lambda\left[\overline{\partial} g(0,0) \times \{-1\}\right] = \{(0,0)\} \times (-\infty, 0].$$

We deduce that $\overline{N}_{\operatorname{gph} F}(0,0,0,0) = \{(0,0,0)\} \times (-\infty,0]$. In particular, $\overline{N}_{\operatorname{gph} F}(x,v)$ does not contain the point $(u,p) = (-1,0,0,-1)$, even though $(-u,v) \in \overline{\partial} H(x,p)$. □

It follows from Example 5.2 that the Mordukhovich form of the Euler-Lagrange inclusion (5.2) may fail to imply either the Clarke form (5.1) or the sharper form (4.11). To see this, recall that the Hamiltonian inclusion in Example 5.2 implies the Mordukhovich inclusion (5.2) by the result of Rockafellar [30] cited above: hence this is an example in which (5.2) holds, but both (5.1) and (4.11) fail.

The Unbounded Case. Since Theorem 4.3 incorporates a form of the Euler-Lagrange inclusion at least as sharp as (5.1), it subsumes the main result of Clarke [1]. That result requires the multifunction F to display integrably Lipschitz dependence on the state, a hypothesis strictly stronger than assumption (ii) of Theorem 4.3.

Conditions (i) and (ii) of Theorem 4.3 are not directly comparable to the basic hypothesis of Polovinkin and Smirnov [19]. Their truncation scheme involves a constant truncation radius in place of the positive-valued function $R(t)$ in (2.1), and their work involves explicit assumptions about the behaviour of the truncated multifunction $\widetilde{F}$ along the nominal arc $\overline{x}$. Our Section 2 has the advantage of introducing hypotheses only on the pointwise behaviour of the given multifunction F near the nominal arc. Indeed, our entire Section 2 can be viewed as a set of verifiable sufficient conditions for a weakened form of Polovinkin and Smirnov's "Condition 1" [19, p. 662] to hold. (See especially Proposition 2.2.)

Polovinkin and Smirnov's conclusions [19, 20] pertain to differential inclusions whose right-hand side may take on nonconvex values, whereas the convexity of the sets $F(t,x)$ is crucial to our approach. But their work offers only a version of the Euler-Lagrange inclusion, whereas ours incorporates a Hamiltonian inclusion as well. Even in the case of bounded differential inclusions, no one knows whether the Hamiltonian inclusion is a correct necessary condition in the absence of this convexity hypothesis.

A detailed comparison of our Euler-Lagrange inclusion with that in [19, (17)] is beyond the scope of this discussion. However, two comments are in order. First, the approach in [19, 20] is completely different from ours. It is based on "linearizing" the given differential inclusion about the nominal arc, and examining the manner in which solutions of the linearized system provide approximations for the resulting reachable set. (A similar approach is taken by Frankowska [5], and has recently been extended to second-order approximations by Zheng [33].) Second, we note that in Example 5.1 the inclusion [19, (17)] is equivalent to (5.3a). (In general, [19, (17)] is a sharper condition than (5.3a).) As such it may generate adjoint arcs which do not satisfy either the Hamiltonian inclusion (4.6) or the refined Euler-Lagrange inclusion (4.11). So we have at least one example in which our results outperform those of Polovinkin and Smirnov.

Let us note that Theorem 4.3 cannot be obtained simply by reformulating problem (P) as an instance of the Generalized Problem of Bolza. For simplicity, we discuss only the case without state constraints by setting $X(t) \equiv \mathbb{R}^n$. Then the definition $L(t,x,v) := \Psi_{F(t,x)}(v)$ puts problem (P) into the following form:

$$(P_B) \qquad \left[\begin{array}{c} \text{minimize } \ell(x(a),x(b)) + \displaystyle\int_0^1 L(t,x(t),\dot{x}(t))\,dt \\ \text{subject to } (x(a),x(b)) \in S. \end{array}\right.$$

The Hamiltonian for this problem is the same as the one we have already associated with F. In particular, since the sets $F(t, \overline{x}(t))$ is not necessarily bounded, the convex functions $p \mapsto H(t, \overline{x}(t), p)$ are not necessarily finite-valued everywhere. This places the current instance of (P_B) beyond the scope of the necessary conditions in Clarke [2, Chap. 4], since the strong Lipschitz condition used there tacitly requires the finiteness of H. (See [2, Remark 4.2.1].) Likewise, the possibility that H could take the value $+\infty$ makes it impossible to verify the basic growth condition assumed in Clarke [3]. Thus Theorem 4.3 not only generalizes the necessary conditions formulated explicitly in terms of differential inclusions, but also lies beyond the reach of the best results on the Generalized Problem of Bolza. Indeed, there is good reason to expect that Theorem 4.3 may lead to strict improvements of the necessary conditions for the Bolza problem. The authors are now pursuing this prospect.

REFERENCES

[1] Clarke, F. H., *Optimal solutions to differential inclusions,* J. Optim. Th. Applic. 19 (1976), 469–478.

[2] Clarke, F. H., *Optimization and Nonsmooth Analysis,* New York: John Wiley, 1983.

[3] Clarke, F. H., *Hamiltonian analysis of the generalized problem of Bolza,* Trans. Amer. Math. Soc. 301 (1987), 385–400.

[4] Clarke, F. H., *Methods of Dynamic and Nonsmooth Optimization,* CBMS-NSF Regional Conference Series, vol. 57. Philadelphia: SIAM, 1989.

[5] Frankowska, H., *Contingent cones to reachable sets of control systems,* SIAM J. Control Optim. 27 (1989), 170–198.

[6] Ioffe, A. D., *Approximate subdifferentials and applications 1: The finite dimensional theory,* Trans. Amer. Math. Soc. 281 (1984), 389–416.

[7] Ioffe, A. D., *Approximate subdifferentials and applications 2,* Mathematika 33 (1986), 111–128.

[8] Ioffe, A. D., *Approximate subdifferentials and applications 3: The metric theory,* Mathematika 36 (1989), 1–38.

[9] Ioffe, A. D., *Proximal analysis and approximate subdifferentials,* J. London Math. Soc. 41 (1990), 175–192.

[10] Kaskosz, B., and S. Łojasiewicz Jr., *Boundary trajectories of systems with unbounded controls,* J. Optim. Th. Applic. (1991), to appear.

[11] Kruger, A. Y., and B. S. Mordukhovich, *Extremal points and the Euler equation in a nonsmooth optimization problem,* Dokl. Acad. Nauk BSSR 24 (1980), 684–687. In Russian.

[12] Loewen, P. D., *Quantitative analysis of epi-Lipschitzian sets and functions,* preprint, UBC, 1991.

[13] Loewen, P. D., and R. T. Rockafellar, *The adjoint arc in nonsmooth optimization,* Trans. Amer. Math. Soc. 325 (1991), 39–72.

[14] Loewen, P. D., and R. B. Vinter, *Pontryagin-type necessary conditions for differential inclusion problems,* Systems and Control Letters 9 (1987), 263–265.

[15] Mordukhovich, B. S., *Maximum principle in the problem of time optimal response with nonsmooth constraints,* J. Appl. Math. Mech. 40 (1976), 960–969.

[16] Mordukhovich, B. S., *Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems,* Soviet Math. Dokl. 22 (1980), 526–530.

[17] Mordukhovich, B. S., *Approximation Methods in Problems of Optimization and Control,* Moscow: Nauka, 1988. (In Russian; English translation to appear.)

[18] Mordukhovich, B. S., *On variational analysis of differential inclusions,* preprint, Wayne State Univ., 1991.

[19] Polovinkin, E. S., and G. V. Smirnov, *An approach to the differentiation of many-valued mappings, and necessary conditions for optimization of solutions of differential inclusions,* Differential Equations 22 (1986), 660–668. Russian original in Differen. Uravn. 22 (1986), 944–954.

[20] Polovinkin, E. S., and G. V. Smirnov, *Time-optimum problem for differential inclusions,* Differential equations 22 (1986), 940–952. Russian original in Differen. Uravn. 22 (1986), 1351–1365.

[21] Rockafellar, R. T., *Convex Analysis.* Princeton, New Jersey: Princeton University Press, 1970.

[22] Rockafellar, R. T., *Integrals which are convex functionals II,* Pacific J. Math. 39 (1971), 429–469.

[23] Rockafellar, R. T., *State constraints in convex problems of Bolza,* SIAM J. Control 10 (1972), 691–715.

[24] Rockafellar, R. T. *Dual problems of Lagrange for arcs, of bounded variation,* in *Calculus of Variations and Control Theory,* D. L. Russell (ed.), Academic Press, 1976, 155–192.

[25] Rockafellar, R. T., *Clarke's tangent cones and the boundaries of closed sets in* $\mathbb{R}^n$*,* Nonlinear Analysis, Theory, Methods & Applications 3 (1979), 145–154.

[26] Rockafellar, R. T., *Optimality conditions for convex control problems with nonnegative states and the possibility of jumps.* In *Game Theory and Mathematical Economics,* O. Moeschlin (ed.), North-Holland, 1981, 339–349.

[27] Rockafellar, R. T., *Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization,* Math. of Oper. Res. 6 (1981), 424–436.

[28] Rockafellar, R. T., *Lipschitzian properties of multifunctions,* Nonlinear Analysis, Theory, Methods & Applications 9 (1985), 867–885.

[29] Rockafellar, R. T., *Extensions of subgradient calculus with applications to optimization,* Nonlinear Analysis, Theory, Methods & Applications 9 (1985), 665–698.

[30] Rockafellar, R. T., *Dualization of subgradient conditions for optimality,* preprint, Univ. of Washington, 1991.

[31] Rowland, J. D. L., and R. B. Vinter, *Dynamic optimization problems with free time and active state constraints,* preprint, Imperial College, 1990.

[32] Wijsman, R. A., *Convergence of sequences of convex sets, cones and functions II,* Trans. Amer. Math. Soc. 123 (1966), 32–45.

[33] Zheng, H., *Second-order necessary conditions for differential inclusion problems,* preprint, UBC, 1991.