SIAM J. CONTROL AND OPTIMIZATION—PREPRINT
Vol. 1, No. 1, pp. 000–000, April 1995

NEW NECESSARY CONDITIONS FOR THE GENERALIZED PROBLEM OF BOLZA *

P. D. LOEWEN† AND R. T. ROCKAFELLAR‡

Abstract. Problems of optimal control are considered in the neoclassical Bolza format, which centers on states and velocities and relies on nonsmooth analysis. Subgradient versions of the Euler-Lagrange equation and the Hamiltonian equation are shown to be necessary for the optimality of a trajectory, moreover in a newly sharpened form that makes these conditions equivalent to each other. At the same time, the assumptions on the Lagrangian integrand are weakened substantially over what has been required previously in obtaining such conditions.

Key words. Optimal control, calculus of variations, nonsmooth analysis, problem of Bolza, Euler-Lagrange condition, Hamiltonian condition, transversality condition

AMS subject classifications. 49K15, 49K05, 49K24

1. Introduction.

Among the classical problems in the calculus of variations, that of Bolza marked a high point of complication, involving all the kinds of side conditions then viewed as important. With deceptive simplicity, the *generalized* problem of Bolza can be stated in one line:

$$(\mathcal{P})\qquad \text{minimize}\ \ \Lambda[x] \;:=\; l(x(a),x(b)) + \int_a^b L(t,x(t),\dot{x}(t))\,dt,$$

where the minimization takes place over all absolutely continuous functions ("arcs") $x\colon[a,b]\to\mathbb{R}^n$. Its generality rests on allowing l and L to be extended-real-valued, hence not necessarily differentiable or even continuous.

The tactic of admitting such a broad range of choices for l and L, first adopted in Rockafellar [21], enables $(\mathcal{P})$ to encompass a vast array of dynamic optimization problems, including those governed by controlled differential equations, differential inclusions, and incorporating endpoint constraints of every conceivable form. For example, $(\mathcal{P})$ subsumes the problem

$$(\mathcal{P}_1)\qquad \begin{aligned}&\text{minimize}\ \ \Lambda_1[x] \;:=\; l_1(x(a),x(b)) + \int_a^b L_1(t,x(t),\dot{x}(t))\,dt\\ &\text{subject to}\ \ (x(a),x(b))\in S\ \ \text{and}\ \ \dot{x}(t)\in F(t,x(t))\ \text{a.e.}\ t\in[a,b]\end{aligned}$$

for a set $S\subset\mathbb{R}^n\times\mathbb{R}^n$ and a multifunction $F:[a,b]\times\mathbb{R}^n\rightrightarrows\mathbb{R}^n$. Indeed, it suffices to take $l=l_1+\Psi_S$ and $L=L_1+\Psi_{\mathrm{gph}\,F}$, where Ψ_S and $\Psi_{\mathrm{gph}\,F}$ are the indicators of S and the graph of F (having the value 0 on these sets but ∞ outside). In the classical problem of Bolza, S and the graph of F were specified by side conditions of the kind $l_i(x(a),x(b))=0$ and $L_j(t,x(t),\dot{x}(t))=0$, with i and j in given finite index sets, all functions being assumed smooth, cf. Bliss [2, p. 189]; eventually the equations

*Received by the editors October 24, 1994; accepted by the editors March 30, 1995.

†Dept. of Mathematics, Univ. of British Columbia, Vancouver, British Columbia, Canada (loew@math.ubc.ca). Research supported by Canada's Natural Science and Engineering Research Council.

‡Dept. of Mathematics, Univ. of Washington, Seattle, WA, 98195 (rtr@math.uwashington.edu). Research supported by the National Science Foundation under grant DMS-9200303.

were supplemented by inequalities, and "isoperimetric" constraints were listed too, cf. Hestenes [8, p. 348]. Isoperimetric constraints fit into $(\mathcal{P}_1)$ by the trick of adding more state variables and modifying S and F accordingly. (In these classical formulations the interval $[a, b]$ was permitted to vary, and this could be built into $(\mathcal{P}_1)$ and $(\mathcal{P})$ as well, but we focus on the fixed-interval case here, reserving the variable-interval extension for elsewhere.)

On the other hand, problems in optimal control of the wide form below can also be fitted into the pattern of $(\mathcal{P})$:

$$(\mathcal{P}_C)\qquad \begin{aligned}&\text{minimize } \Phi[x,u] := \phi(x(a), x(b)) + \int_a^b f(t, x(t), u(t))\, dt\\ &\text{subject to } \dot{x}(t) \in F(t, x(t), u(t)),\ \ u(t) \in U(t, x(t)) \text{ a.e. } t \in [a, b],\\ &\qquad \text{and } (x(a), x(b)) \in S.\end{aligned}$$

To arrange this, simply take $l = \phi + \Psi_S$ as before and

$$L(t, x, v) = \inf_u \{ f(t, x, u) \,:\, u \in U(t, x),\ v \in F(t, x, u)\},$$

interpreting the right side as ∞ when there is no u in $U(t, x)$ for which $F(t, x, u)$ contains v. Notice that the dynamics here involve a controlled differential inclusion and that the set of admissible controls displays explicit state-dependence—two features beyond the scope of the classical theory. It is more difficult to force $(\mathcal{P}_C)$ into the framework of $(\mathcal{P}_1)$, which underscores the importance of $(\mathcal{P})$ as the model of choice when a full spectrum of control applications is envisioned. For more on this approach to optimal control, see [24] and [29].

Our aim is to establish necessary conditions for optimality in $(\mathcal{P})$ that retain both the form and the power of their classical precursors, the equations of Euler-Lagrange and Hamilton, despite the nonsmooth, extended-real-valued setting. This program for the generalized problem of Bolza is not new: it began with Rockafellar's work in the case where both functions l and $L(t, \cdot, \cdot)$ are convex [21, 22, 23, 25], and it was greatly advanced beyond such full convexity by Clarke [3, 4, 5, 6] and others. Most recently there have been contributions by Loewen and Rockafellar [13, 14], Mordukhovich [19] and Ioffe and Rockafellar [10].

The current work has two especially distinguishing features. First, it provides a sharpened version of the Hamiltonian optimality condition that is *equivalent* to the sharpened form of the Euler-Lagrange condition we introduced in [14]. Second, it assumes significantly less than before about the Lagrangian L; it does not demand that L have the form $L_1 + \Psi_{\operatorname{gph} F}$ in which Lipschitz properties are expected of L_1 and F, as for instance in [13]. It does ask for the convexity of L in the velocity argument, in contrast to the recent papers [19] and [10], but those works are more restrictive in other respects and anyway concern the Euler-Lagrange condition only.

The convexity of L in the velocity argument is essential for the equivalence between the Euler-Lagrange condition and the Hamiltonian condition, whatever their versions. Indeed, aside from the classical case of a smooth function L, or the fully convex case where L is convex in the state and velocity arguments together and some other special cases covered by [30], results asserting the simultaneous necessity of both conditions were elusive. The best that could be claimed, in [14], was the existence of at least one adjoint arc for which both conditions in a certain form were satisfied. (Other adjoint arcs might fulfill just one of the two.)

The sharpened Euler-Lagrange condition that we use in relating an extremal arc $\overline{x}$ to an adjoint arc p asserts that

$$\dot{p}(t) \in \operatorname{co}\left\{v \,:\, (v, p(t)) \in \partial L(t, \overline{x}(t), \dot{\overline{x}}(t))\right\} \text{ a.e. } t \in [a,b]. \tag{1.1}$$

Here ∂ refers to the possibly nonconvex *limiting subgradient* set (see Loewen [12] for notation and terminology), known also under various other names: limiting proximal subgradient set in Clarke [6], approximate subdifferential in Ioffe [9], subdifferential in Mordukhovich [19], subgradient set in the general sense in Rockafellar [31]. (The subgradients are those of $L(t,\cdot,\cdot)$ with t fixed.) Under the hypotheses of this paper (see Section 2), the inclusion (1.1) implies that for almost all t the vector $\dot{\overline{x}}(t)$ maximizes the function $v \mapsto \langle p(t)\,,\, v\rangle - L(t, \overline{x}(t), v)$.

The sharpened Hamiltonian condition that we establish for the first time as necessary for optimality, by virtue of its equivalence to (1.1), is

$$\dot{p}(t) \in \operatorname{co}\left\{w \,:\, (-w, \dot{\overline{x}}(t)) \in \partial H(t, \overline{x}(t), p(t))\right\} \text{ a.e. } t \in [a,b]. \tag{1.2}$$

The Hamiltonian H is, as usual, the Legendre-Fenchel transform of the Lagrangian L in its velocity variable:

$$H(t,x,p) := \sup\left\{\langle p\,,\, v\rangle - L(t,x,v) \,:\, v \in \mathbb{R}^n\right\}.$$

Clearly, (1.2) is a strict improvement on the form $(-\dot{p}(t), \dot{\overline{x}}(t)) \in \operatorname{co}\partial H(t, \overline{x}(t), p(t))$ taken as standard until now, since it convexifies only in the first argument. It implies, in particular, that for almost all t, the vector $p(t)$ maximizes the function $q \mapsto \langle q\,,\, \dot{\overline{x}}(t)\rangle - H(t, \overline{x}(t), q)$.

The weakened assumptions on L that suffice for these developments are set out in hypotheses (H4) and (H5) of Section 2. The first of these is a very mild "epi-continuity" assumption. Geometrically it amounts to insisting that, for each fixed t, the set $\operatorname{epi} L(t,x,\cdot)$ should vary continuously with x. The second is a growth condition on subgradients, reducing when $L(t,x,v)$ is smooth to a local inequality of the form $|\nabla_x L| \le \kappa\big(1 + |\nabla_v L|\big)$. It implies, through a result of Mordukhovich [18], the Aubin ("pseudo-Lipschitz") continuity of the multifunction $x \rightrightarrows \operatorname{epi} L(t,x,\cdot)$ near the optimal arc. Our need for Aubin continuity on a tube of uniform size around the minimizing trajectory makes it necessary to formulate a quantitative generalization of Mordukhovich's result in Section 4.

We give special attention in Section 7 to the Lipschitz-plus-indicator case where $L = L_1 + \Psi_{\operatorname{gph} F}$, showing for that version of the problem that the present results yield a full suite of (sharpened) Lagrangian and Hamiltonian necessary conditions for optimality in both normal and abnormal forms, beyond what we had previously obtained in [13] and [14]. This recalls the work of Smirnov [32], who proposed the version of (1.1) for $L = \Psi_{\operatorname{gph} F}$ as a necessary condition in 1991, but whose requirements that F be bounded and autonomous are significantly relaxed here. (Smirnov's result and proof are linked to prior work of Mordukhovich [15, 16, 17], who has recently given conditions [19] under which the necessity of (1.1) can be established in the absence of convexity hypotheses.) The main thrust of our effort, however, goes the other way: we demonstrate how to transform $(\mathcal{P})$ in its full generality into an instance of the differential inclusion problem in [14], and with some new machinery we then apply the results in that paper in combination with the Lagrangian-Hamiltonian equivalence theorem in [31].

State constraints requiring $x(t)$ to belong to a set $X(t) \subset \mathbb{R}^n$ can in principle be incorporated into problem $(\mathcal{P})$ by adding an indicator term in the specification of L, but for technical reasons it is better, at least in the theory as it now stands, to keep them explicit. The treatment of such constraints is taken up in Section 6.

2. The Main Result. Our main result is Theorem 2.1, a set of necessary conditions for an arc $\overline{x}$ to provide a local minimum in problem $(\mathcal{P})$. So let $\overline{x}$ be given, and fix some $\varepsilon > 0$ in order to define a suitable neighbourhood of $\overline{x}$:

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

We impose five conditions on $\overline{x}$ and the functions l and L relative to the set Ω; these are described below as (H1)–(H5). For simplicity in dealing with subgradients of $L(t,x,v)$ and $H(t,x,p)$ we use the notation ∂L and ∂H instead of the more cumbersome (but precise) $\partial_{(x,v)} L$ and $\partial_{(x,p)} H$. In general, as already mentioned, we write $\partial f(z)$ for the set of limiting subgradients associated with a lower semicontinuous function f at the point z; the singular counterpart to this set is $\partial^\infty f(z)$. See Loewen [12] for details.

THEOREM 2.1. *Assume* (H1)–(H5). *Suppose that for every arc x with graph in Ω, one has $\Lambda[x] \ge \Lambda[\overline{x}]$. Then either the normal conditions or the degenerate conditions written below are valid.*

[Normal Conditions]: *For some arc p on $[a,b]$,*

(a) $\dot p(t) \in \operatorname{co}\big\{v \,:\, (v, p(t)) \in \partial L(t, \overline{x}(t), \dot{\overline{x}}(t))\big\}$ a.e. $t \in [a,b]$,

(b) $(p(a), -p(b)) \in \partial l(\overline{x}(a), \overline{x}(b))$.

[Degenerate Conditions]: *For some nonzero arc p on $[a,b]$,*

(a^∞) $\dot p(t) \in \operatorname{co}\big\{v \,:\, (v, p(t)) \in \partial^\infty L(t, \overline{x}(t), \dot{\overline{x}}(t))\big\}$ a.e. $t \in [a,b]$,

(b^∞) $(p(a), -p(b)) \in \partial^\infty l(\overline{x}(a), \overline{x}(b))$.

(In particular, if the only arc p on $[a,b]$ satisfying conditions (a^∞)–(b^∞) *is the zero arc, then the normal conditions are satisfied.) In the normal conditions, assertion* (a) *is equivalent to*

(a') $\dot p(t) \in \operatorname{co}\big\{w \,:\, (-w, \dot{\overline{x}}(t)) \in \partial H(t, \overline{x}(t), p(t))\big\}$ a.e. $t \in [a,b]$.

Also, conditions (a) *and* (a') *imply that for almost all t in $[a,b]$,*

(c) $p(t) \in \partial_v L(t, \overline{x}(t), \dot{\overline{x}}(t)) = \operatorname*{argmax}_{q \in \mathbb{R}^n} \big\{\big\langle q\,,\, \dot{\overline{x}}(t)\big\rangle - H(t, \overline{x}(t), q)\big\}$, *and*

$\dot{\overline{x}}(t) \in \partial_p H(t, \overline{x}(t), p(t)) = \operatorname*{argmax}_{v \in \mathbb{R}^n} \{\langle p(t)\,,\, v\rangle - L(t, \overline{x}(t), v)\}$.

Hypotheses. The terms in the Bolza functional Λ are required to have the following properties, expressed in terms of the constant $\varepsilon > 0$ in the definition of Ω and two positive-valued integrable functions δ and κ on $[a,b]$.

(H1) The endpoint cost function $l(x_a, x_b)\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is lower semicontinuous on $\Omega_a \times \Omega_b$;

(H2) The integrand $L(t,x,v)\colon \Omega \times \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ 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 $\mathbb{R}^n \times \mathbb{R}^n$;

(H3) For each fixed pair (t,x) in Ω, the function $v \mapsto L(t,x,v)$ is convex;

(H4) For almost every t in $[a,b]$, the function $(x,v) \mapsto L(t,x,v)$ is lower semicontinuous on $\Omega_t \times \mathbb{R}^n$ and has the following epi-continuity property: for any point $(\widehat{x}, \widehat{v})$ where $|\widehat{x} - \overline{x}(t)| < \varepsilon$ and $L(t, \widehat{x}, \widehat{v})$ is finite, and any sequence $x_k \to \widehat{x}$ in Ω_t, there exists a sequence $v_k \to \widehat{v}$ along which $L(t, x_k, v_k) \to L(t, \widehat{x}, \widehat{v})$.

(H5) The ratio $\kappa(t)/\delta(t)$ is essentially bounded. For almost all t in $[a,b]$, one has

$$|w| \le \kappa(t)\,(1+|p|) \;\text{ for all }\; (w,p) \in \partial L(t,x,v),$$

whenever $|x-\overline{x}(t)| < \varepsilon$, $\big|(v, L(t,x,v)) - (\dot{\overline{x}}(t), L(t,\overline{x}(t),\dot{\overline{x}}(t)))\big| < \delta(t)$.

As the absence of measures in the statement of Theorem 2.1 signals, state constraints cannot be implicit in the instance of $(\mathcal{P})$ under consideration at this stage. To see how such restrictions are ruled out, notice that (H4) makes the set

$$G_t = \big\{ x \in \Omega_t \,:\, L(t,x,v) < \infty \text{ for some } v \text{ with } \big|v - \dot{\overline{x}}(t)\big| < \delta(t) \big\}$$

be *open* for almost all t. Indeed, (H4) says that no point $\widehat{x}$ in G_t can be a boundary point. Of course, $\overline{x}(t)$ lies in G_t. Thus our paradigm allows for L to impose certain velocity constraints through the use of infinite penalties, but does not allow unilateral state constraints to be covered in the same way. State constraints in the explicit form "$x(t) \in X(t)\ \forall t$" can nonetheless be handled by our methods, as will be explained in Section 6.

Hypothesis (H5) can be viewed as a combined growth condition and Lipschitz condition. As a growth condition, it resembles the "condition of Morrey type" underlying Clarke and Vinter's Prop. 3.2 in [7], a result which establishes the validity of the Euler-Lagrange equation in the calculus of variations without the *a priori* assumption that the minimizing arc is Lipschitzian. Indeed, in the special case where $L(t,\cdot,\cdot)$ is continuously differentiable on Ω_t, (H5) reduces to

$$|\nabla_x L(t,x,v)| \le \kappa(t)\big(1 + |\nabla_v L(t,x,v)|\big) \qquad \forall (x,v) \in \Omega_t.$$

As a Lipschitz condition, (H5) is a subgradient characterization of the Aubin continuity of the epigraphical multifunction associated with L, as we shall see in Section 4. A Hamiltonian formulation of this assumption is derived in Section 5, where its relationship to Clarke's "strong Lipschitz condition" [4] is easiest to discern.

Notice that the seemingly weaker form of (H5) obtained by substituting the proximal subgradient set $\hat{\hat{\partial}} L$ for ∂L is actually equivalent to the form stated here, because ∂L is defined by taking limits of proximal subgradients.

The Method of Proof. There is a well known equivalence between Bolza problems and Mayer problems, mediated by the technique of state augmentation. Indeed, consider the domain $\widetilde{\Omega} = \Omega \times \mathbb{R}$ in one more state dimension, the extra state variable being denoted by y, and define the epigraphical multifunction $E : \widetilde{\Omega} \rightrightarrows \mathbb{R}^{n+1}$ by

$$E(t,x,y) := \operatorname{epi} L(t,x,\cdot).$$

(The set $E(t,x,y)$ does not actually depend on y.) If the arc $\overline{x}$ figuring in our hypotheses solves $(\mathcal{P})$, then the arc $(\overline{x},\overline{y})$, with

$$\overline{y}(t) := \int_a^t L(r,\overline{x}(r),\dot{\overline{x}}(r))\,dr,$$

solves the following differential inclusion problem:

$$(\mathcal{P}') \qquad \begin{array}{l} \text{minimize } k\big(x(a),y(a),x(b),y(b)\big) := l(x(a),x(b)) + y(b) + \Psi_{\{0\}}(y(a)) \\ \text{subject to } (\dot{x}(t),\dot{y}(t)) \in E(t,x(t),y(t)) \text{ a.e. } t \in [a,b]. \end{array}$$

The right side in the dynamic constraint here is unbounded. Necessary conditions for optimality in problems of this sort were the subject of a previous paper [14]. Our procedure in the current paper is basically to check the hypotheses in [14], and then to translate the conclusions of that work into the context of $(\mathcal{P})$. Checking the hypotheses takes a certain amount of work, since the transition from the subgradient hypothesis (H5) to the Lipschitz conditions required by [14] is not completely straightforward (see Section 4). Likewise, an additional state-augmentation argument is necessary to reduce the case of a general lower semicontinuous endpoint cost l to the Lipschitz-plus-indicator form treated in [14] (see the proof of Theorem 3.1). Finally, it is insufficient simply to transcribe the conclusions of [14]: the sharpened Hamiltonian inclusion featured here relies on a careful analysis of the relationship between the Hamiltonian and Eulerian forms of the necessary conditions, as carried out by Rockafellar [31].

3. Proof of the Main Result. To prove Theorem 2.1, we shall apply an intermediate result for unbounded differential inclusions which is readily derived from [14, Theorem 4.3]. The reformulated problem $(\mathcal{P}')$ under consideration fits the general pattern:

$$(\overline{\mathcal{P}}) \qquad \begin{aligned} &\text{minimize } k(z(a), z(b)) \\ &\text{subject to } \dot{z}(t) \in E(t, z(t)) \text{ a.e. } t \in [a, b]. \end{aligned}$$

The hypotheses of [14] for this kind of problem refer to a distinguished arc $\overline{z}$, and, for some fixed $\eta > 0$, its "graphical neighbourhood"

$$\begin{aligned} U &= \{(t, z) \,:\, t \in [a, b],\ |z - \overline{z}(t)| < \eta\}, \\ U_t &= \{z \,:\, |z - \overline{z}(t)| < \eta\}, \quad a \le t \le b. \end{aligned}$$

They read as follows:

(h1) The endpoint cost function $k\colon U_a \times U_b \to \mathbb{R} \cup \{\infty\}$ is lower semicontinuous;

(h2) The sets $E(t, z)$ are nonempty, closed, and convex for each (t, z) in U, and empty for each (t, z) outside U;

(h3) The multifunction E 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 $\mathbb{R}^m$;

(h4) There are integrable functions δ and K on $[a, b]$, with K/δ essentially bounded, such that almost every t in $[a, b]$ obeys

$$E(t, y) \cap (\dot{\overline{z}}(t) + \delta(t)\mathbb{B}) \subseteq E(t, z) + K(t)|y - z|\mathbb{B} \quad \forall y, z \in U_t.$$

Here and elsewhere in this paper, $\mathbb{B}$ denotes the closed unit ball in the Euclidean space of appropriate dimension.

THEOREM 3.1 [14]. *If hypotheses* (h1)–(h4) *hold and* $\overline{z}$ *solves problem* $(\overline{\mathcal{P}})$*, then there exists an arc* q *on* $[a, b]$ *such that*

(a) $\dot{q}(t) \in \operatorname{co}\left\{w \,:\, (w, q(t)) \in N_{\operatorname{gph} E(t,\cdot)}\left(\overline{z}(t), \dot{\overline{z}}(t)\right)\right\}$ a.e. $t \in [a, b]$; *and*

(b) *one of the following transversality conditions holds:*

(i) $(q(a), -q(b)) \in \partial k(\overline{z}(a), \overline{z}(b))$*, or*

(ii) $(q(a), -q(b)) \in \partial^\infty k(\overline{z}(a), \overline{z}(b))$*, with* q *not identically zero.*

Proof. A simple trick reduces the general lower semicontinuous endpoint cost function k to one in the Lipschitz-plus-indicator form analyzed in [14]. Indeed, it suffices to define the constant arc $\overline{r} = k(\overline{z}(a), \overline{z}(b))$, and then to observe that the pair $(\overline{z}, \overline{r})$ solves the problem:

$$\begin{aligned} &\text{minimize } \ r(b) \\ &\text{subject to } \ (z(a), z(b), r(a)) \in \operatorname{epi} k, \quad r(b) \in \mathbb{R}, \\ &\qquad\qquad (\dot{z}(t), \dot{r}(t)) \in E(t, z(t)) \times \{0\} \ \text{ a.e. } t \in [a,b]. \end{aligned}$$

The stated result follows from conditions (b) and (d) of [14, Theorem 4.3] by elementary subgradient calculus.

Note that although the statement of [14, Theorem 4.3] involves stronger Lipschitz conditions on the multifunction E, specifically tailored to the modulus of integrability of the function $\dot{\overline{z}}$, these are present only to facilitate a statement free of explicit references to the quantity $\dot{\overline{z}}(t)$. In fact, the conditions of [14, Proposition 2.2] are sufficient for the conclusions of [14, Theorem 4.3], and it is these we have applied here—taking $R = \delta$ and $m = K$. □

Leaving aside the verification of hypotheses (h1)–(h4) for now, let us derive the conclusions of Theorem 2.1 from those of Theorem 3.1. To apply the latter result, we take $m = n + 1$, with $z = (x, y)$ as a pattern and $\overline{z} = (\overline{x}, \overline{y})$ as the optimal arc. Of course, $E(t, z) = E(t, x, y)$ and $k(x_a, y_a, x_b, y_b) = l(x_a, x_b) + y_b + \Psi_{\{0\}}(y_a)$. Observe that

$$\begin{aligned} \partial k(\overline{z}(a), \overline{z}(b)) &= \{(\alpha, \zeta, \beta, 1) \, : \, (\alpha, \beta) \in \partial l(\overline{x}(a), \overline{x}(b)), \ \zeta \in \mathbb{R}\}, \\ \partial^\infty k(\overline{z}(a), \overline{z}(b)) &= \{(\alpha, \zeta, \beta, 0) \, : \, (\alpha, \beta) \in \partial^\infty l(\overline{x}(a), \overline{x}(b)), \ \zeta \in \mathbb{R}\}. \end{aligned}$$

In terms of these data, Theorem 3.1 provides an adjoint arc $(p, q)\colon [a, b] \to \mathbb{R}^n \times \mathbb{R}$ satisfying two conditions. First, the transversality condition 3.1(b) implies that either $q(b) = -1$ and $(p(a), -p(b)) \in \partial l(\overline{x}(a), \overline{x}(b))$, or else $q(b) = 0$ and $(p(a), -p(b)) \in \partial^\infty l(\overline{x}(a), \overline{x}(b))$ with (p, q) not identically zero. Second, the Euler-Lagrange condition 3.1(a) asserts that for almost all t in $[a, b]$,

$$(3.1) \quad (\dot{p}(t), \dot{q}(t)) \in \operatorname{co} \left\{ (v, w) \, : \, (v, w, p(t), q(t)) \in N_{\operatorname{gph} E(t,\cdot)} \left(\overline{x}(t), \overline{y}(t), \dot{\overline{x}}(t), \dot{\overline{y}}(t) \right) \right\}.$$

Now for fixed t, one has

$$\operatorname{gph} E(t, \cdot) = \{(x, y, v, r) \, : \, (x, v, r) \in \operatorname{epi} L(t, \cdot, \cdot), \ y \in \mathbb{R}\}.$$

In terms of $\overline{L}(t) = L(t, \overline{x}(t), \dot{\overline{x}}(t))$, this implies

$$\begin{aligned} &N_{\operatorname{gph} E(t,\cdot)} \left(\overline{x}(t), \overline{y}(t), \dot{\overline{x}}(t), \dot{\overline{y}}(t) \right) \\ &\qquad = \left\{ (v, 0, \pi, -\rho) \, : \, (v, \pi, -\rho) \in N_{\operatorname{epi} L(t,\cdot,\cdot)} \left(\overline{x}(t), \dot{\overline{x}}(t), \overline{L}(t) \right) \right\}. \end{aligned}$$

Using this relation in (3.1), we get

$$(3.2) \qquad (\dot{p}(t), \dot{q}(t)) \in \operatorname{co} \left\{ v \, : \, (v, p(t), q(t)) \in N_{\operatorname{epi} L(t,\cdot,\cdot)} \left(\overline{x}(t), \dot{\overline{x}}(t), \overline{L}(t) \right) \right\} \times \{0\}.$$

In particular, the second component of this inclusion implies that $\dot{q}(t) = 0$ almost everywhere. Thus q is a constant function, whose value is either 0 or -1. In the case where $q = -1$, one has the normal conditions of Theorem 2.1:

(a) $\dot{p}(t) \in \operatorname{co} \left\{ v \, : \, (v, p(t)) \in \partial L(t, \overline{x}(t), \dot{\overline{x}}(t)) \right\}$ a.e. $t \in [a, b]$, and

(b) $(p(a), -p(b)) \in \partial l(\overline{x}(a), \overline{x}(b))$.

In the case where $q = 0$, the degenerate conditions of Theorem 2.1 follow instead:

(a^∞) $\dot{p}(t) \in \operatorname{co} \left\{ v \, : \, (v, p(t)) \in \partial^\infty L(t, \overline{x}(t), \dot{\overline{x}}(t)) \right\}$ a.e. $t \in [a, b]$, and

(b^∞) $(p(a), -p(b)) \in \partial^\infty l(\overline{x}(a), \overline{x}(b))$, and p is not the zero arc.

In [31, Theorem 1.1], Rockafellar proves that the inclusions (a) and (a′) in Theorem 2.1 are equivalent for each fixed t with the properties described in (H4), provided that every such t also satisfies the condition

$$(w, 0) \in \partial^\infty L(t, \overline{x}(t), \dot{\overline{x}}(t)) \implies w = 0. \tag{3.3}$$

Note that (3.3) holds for almost all t, by (H5). Indeed, any point $(w, 0)$ in the cone $\partial^\infty L(t, \overline{x}(t), \dot{\overline{x}}(t))$ must have the form $(w, 0) = \lim_{\nu\to\infty} r_\nu(w_\nu, p_\nu)$ for sequences $r_\nu \to 0^+$ and $(w_\nu, p_\nu) \in \hat{\hat{\partial}} L(t, x_\nu, v_\nu)$ along which $(x_\nu, v_\nu) \xrightarrow[L]{} (\overline{x}(t), \dot{\overline{x}}(t))$. Since $\hat{\hat{\partial}} L \subseteq \partial L$ always, (H5) implies that for all ν sufficiently large,

$$|r_\nu w_\nu| \le \kappa\left(|r_\nu p_\nu| + r_\nu\right).$$

In the limit as $\nu \to \infty$, it follows that $|w| \le 0$, so (3.3) holds. Under the same hypotheses, Rockafellar [31] shows that the equivalent conditions (a) and (a′) both imply the argmax conditions in (c).

To complete the proof of Theorem 2.1, we must demonstrate that Theorem 3.1 is truly applicable—by checking hypotheses (h1)–(h4). Conditions (h1)–(h3) hold for any $\eta \in (0, \varepsilon]$, as obvious consequences of the corresponding hypotheses (H1)–(H3) on l and L. The real issue is hypothesis (h4), which calls for the Aubin continuity of the multifunction E (see Aubin [1]) with respect to a certain restricted tube around $\overline{z}$. This condition follows from hypothesis (H5) and Theorem 4.3 in the next section.

To see this, fix a time t in $[a, b]$ at which the conditions in (H4)–(H5) hold. Since t will be fixed throughout this argument, and since E does not actually depend on y, we suppress both the t- and y-dependence of E and L, writing simply $E(x) = \operatorname{epi} L(x, \cdot)$ for $|x - \overline{x}| < \varepsilon$. (We also write $\dot{\overline{x}}$ instead of $\dot{\overline{x}}(t)$ and use the shorthand $\overline{L} = L(\overline{x}, \dot{\overline{x}})$.) As noted above, $\operatorname{gph} E = \operatorname{epi} L$; thus (H4) implies that $\operatorname{gph} E$ is closed, and that condition (i) of Theorem 4.3 holds. Condition (ii), on the other hand, requires that

$$|w| \le R|(p, -\lambda)| \text{ for all } (w, p, -\lambda) \in N_{\operatorname{epi} L}(x, v, r), \quad \text{whenever } |x - \overline{x}| < \varepsilon \text{ and } \left|(v, r) - (\dot{\overline{x}}, \overline{L})\right| < \delta.$$

Now the "proximal subgradient formula" [26, 11] asserts that every nonzero vector $(w, p, -\lambda)$ in $N_{\operatorname{epi} L}(x, v, r)$ can be realized as the limit of a sequence of proximal normals $(w_\nu, p_\nu, -\lambda_\nu)$ in $\hat{\hat{N}}_{\operatorname{epi} L}(x_\nu, v_\nu, r_\nu)$ for which $\lambda_\nu > 0$ and the corresponding base points obey $(x_\nu, v_\nu, r_\nu) \xrightarrow[\operatorname{epi} L]{} (x, v, r)$. For every term in such a sequence, one has $(w_\nu/\lambda_\nu, p_\nu/\lambda_\nu) \in \hat{\hat{\partial}} L(x_\nu, v_\nu)$, so (H5) gives

$$|w_\nu/\lambda_\nu| \le \kappa\left(1 + |p_\nu/\lambda_\nu|\right).$$

Multiplying through by $\lambda_\nu > 0$ and letting $\nu \to \infty$, we obtain

$$|w| \le \kappa\left(|p| + |\lambda|\right) \le 2\kappa|(p, -\lambda)|.$$

This argument applies to every triple $(w, p, -\lambda)$ in $N_{\operatorname{epi} L}(x, v, r)$, so condition (ii) of Theorem 4.3 holds with $R = 2\kappa$. The conclusion of Theorem 4.3 establishes (h4), with $K(t) = \sqrt{1 + 4\kappa(t)^2}$, any constant $0 < \eta \le \varepsilon_0(t) := \min\{\varepsilon, \delta(t)/(9K(t))\}$, and $\delta(t)$ equal to one-sixth the value of the $\delta(t)$ provided in (H5). (The function ε_0 is bounded away from zero because $\kappa/\delta \in L^\infty$ by (H5).) This completes the proof of Theorem 2.1.

4. On Uniform Aubin Continuity. This section and the next furnish technical support for the proof and interpretation of Theorem 2.1 above. Both are intended, though, to stand alone as having independent interest: they involve functions and constants whose names are deliberately suggestive, but are logically distinct from those identified in Sections 1–3 and 6–7.

DEFINITION 4.1. *Let* $\Gamma\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *be a multifunction; let* $(\overline{x},\overline{\gamma})$ *be a point in* $\operatorname{gph}\Gamma$. *To say that* Γ *is Aubin continuous at* $(\overline{x},\overline{\gamma})$ *with parameters* $\varepsilon > 0$, $\delta > 0$, *and* $K > 0$ *means that one has*

$$\Gamma(x) \cap (\overline{\gamma} + \delta\mathbb{B}) \subseteq \Gamma(y) + K|y-x|\mathbb{B} \qquad \forall x, y \in \overline{x} + \varepsilon\mathbb{B}. \tag{4.1}$$

The modulus of Aubin continuity for a given multifunction Γ *at a point* $(\overline{x},\overline{\gamma})$ *in* $\operatorname{gph}\Gamma$ *is the number* $\kappa_\Gamma(\overline{x},\overline{\gamma})$, *defined as the infimum of all* $K > 0$ *satisfying (4.1) for some* $\varepsilon > 0$ *and* $\delta > 0$. *Mordukhovich [18, Theorem 5.7] has shown that if* $\operatorname{gph}\Gamma$ *is closed, then*

$$\kappa_\Gamma(\overline{z},\overline{\gamma}) = \sup\left\{|\alpha| \,:\, (\alpha,\beta) \in N_{\operatorname{gph}\Gamma}(\overline{x},\overline{\gamma})\,,\ |\beta| \le 1\right\}.$$

For the purposes of this paper, Aubin continuity with some fixed $\delta > 0$ *is required at every point in some* ε*-neighbourhood of a given arc, and knowing only that* κ_Γ *is finite at every point along the arc is not sufficient. We need quantitative estimates of the constants* ε, δ, *and* K *in terms of a neighbourhood of* $(\overline{x},\overline{\gamma})$ *in which the (generalized) slope of vectors normal to the graph of* Γ *is bounded.*

Our approach to this problem is patterned on that introduced in Rockafellar [28, Remark 3.14]: we prove that Γ *has the desired Aubin continuity properties by showing that the function* d_Γ *defined in the next Lemma satisfies a corresponding Lipschitz continuity condition. This in turn is accomplished by using Rockafellar's 1985 results [27] for estimating the subgradients of marginal functions. (Although the facts we appropriate from these earlier papers were phrased in terms of Clarke subgradients, they apply equally well to the limiting subgradients we are working with here.)*

LEMMA 4.2. *Given a multifunction* Γ *with closed graph* G, *consider* $d_\Gamma(x,v) := d_{\Gamma(x)}(v)$. *If* d_Γ *is Lipschitz of rank* K *on the set* $(\overline{x}+\varepsilon\mathbb{B})\times(\overline{\gamma}+\delta\mathbb{B})$ *for some constants* $\varepsilon > 0$, $\delta > 0$, *then condition* (4.1) *holds, with the same constants.*

Proof. This is elementary—see Rockafellar [28, Theorem 2.3, (b)⇒(a)]. ∎

THEOREM 4.3. *Let* $\Gamma\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *be a multifunction; write* $G = \operatorname{gph}\Gamma$, *and assume that* G *is a closed set. Let* $(\overline{x},\overline{\gamma})$ *in* G *be a point for which some constants* $\varepsilon > 0$, $\delta > 0$, *and* $R > 0$ *satisfy two conditions*:

(i) *For any point* (x,γ) *in* G *with* $|x-\overline{x}| < \varepsilon$ *and* $|\gamma-\overline{\gamma}| < \delta$, *and any sequence* $x_k \to x$, *there is a sequence* $\gamma_k \in \Gamma(x_k)$ *such that* $\gamma_k \to \gamma$.

(ii) $|\alpha| \le R|\beta|$ *for all* $(\alpha,\beta) \in N_{\operatorname{gph}\Gamma}(x,\gamma)$, *whenever* $|x-\overline{x}| < \varepsilon$ *and* $|\gamma-\overline{\gamma}| < \delta$.

Then Γ *is Aubin continuous at* $(\overline{x},\overline{\gamma})$ *with parameters* $K = \sqrt{1+R^2}$, $\delta_0 = \delta/6$, *and* $\varepsilon_0 = \min\{\varepsilon, \delta/(9K)\}$, *i.e.*,

$$\Gamma(y) \cap (\overline{x} + \delta_0\mathbb{B}) \subseteq \Gamma(x) + K|y-x|\mathbb{B} \qquad \forall x, y \in \overline{x} + \varepsilon_0\mathbb{B}. \tag{4.2}$$

Proof. According to Lemma 4.2, we only have to show that d_Γ is Lipschitz continuous of rank K on the set $(\overline{x}+\varepsilon_0\mathbb{B})\times(\overline{\gamma}+\delta_0\mathbb{B})$. And to accomplish this, it suffices to show that $\partial d_\Gamma(x,v) \subseteq K\mathbb{B}$ for all (x,v) in this set. We therefore set out to estimate ∂d_Γ, relying on Rockafellar [27].

A convenient characterization of d_Γ, valid for all x and v without restriction, is

$$\begin{aligned} d_\Gamma(x,v) &= \min\{|v-\gamma| \,:\, (x,\gamma)\in G\} \\ &= \min\{f(x,v,\gamma) \,:\, (x,v,\gamma)\in S\}, \end{aligned} \tag{4.3}$$

where $f(x,v,\gamma) = |x-\gamma|$ and $S = \{(x,v,\gamma) : (x,\gamma)\in G,\ v\in \mathbb{R}^m\}$. In the latter form, d_Γ is revealed as the marginal function associated with an optimization problem depending on parameters (x,v). Such functions have been studied extensively, in particular by Rockafellar [27], whose results we shall employ here. Let us denote by $\Sigma(x,v)$ the set of minimizing vectors γ in (4.3) above. Theorem 8.3 of [27] implies that any point (x,v) has a neighbourhood in which d_Γ is Lipschitzian and satisfies the following estimate for limiting subgradients:

$$\partial d_\Gamma(x,v) \;\subseteq\; \{(\xi,\eta) : (\xi,\eta,0)\in \partial f(x,v,\gamma) + N_S\,(x,v,\gamma)\;\; \exists \gamma\in\Sigma(x,v)\}\,.$$

(The hypotheses of Rockafellar's result are easy to verify, because the function f here is Lipschitzian, so $\partial^\infty f \equiv \{0\}$, and because f grows rapidly enough to make the inf-compactness condition obvious.) We note that whenever $\gamma\in\Sigma(x,v)$,

$$\begin{aligned} \partial f(x,v,\gamma) &\subseteq \{(0,u,-u) : |u|\le 1\}\,, \\ N_S\,(x,v,\gamma) &= \{(\alpha,0,\beta) : (\alpha,\beta)\in N_G\,(x,\gamma)\,\}\,. \end{aligned}$$

(A sharper version of the first inclusion is possible, but this one is adequate for our purposes.) It follows that any point (ξ,η) in $\partial d_\Gamma(x,v)$ obeys

$$(\xi,\eta,0) = (0,u,-u) + (\alpha,0,\beta) \text{ for some } \gamma\in\Sigma(x,v),\ (\alpha,\beta)\in N_G\,(x,\gamma)\,,\ u\in \mathbb{B}.$$

Thus

$$\partial d_\Gamma(x,v) \subseteq \{(\alpha,\beta)\in N_G\,(x,\gamma) : \gamma\in\Sigma(x,v),\ |\beta|\le 1\}\,.$$

For those points (x,v) where all the pairs (x,γ) with $\gamma\in\Sigma(x,v)$ lie in the set specified by hypothesis (ii), that condition implies that every pair (α,β) on the right side has $|\alpha|\le R$ and $|\beta|\le 1$, so $|(\alpha,\beta)|^2 \le (R^2+1) = K^2$. Thus

$$\{x\}\times\Sigma(x,v) \subseteq (\overline{x}+\varepsilon\mathbb{B})\times(\overline{\gamma}+\delta\mathbb{B}) \implies \partial d_\Gamma(x,v)\subseteq K\mathbb{B}. \tag{4.4}$$

This reveals the key to the result: the location of the set $\Sigma(x,v)$.

Claim. *Fix $(\widehat{x},\widehat{\gamma})$ in G with $|\widehat{x}-\overline{x}|<\varepsilon$, $|\widehat{\gamma}-\overline{\gamma}|<\delta$. Then for some $\mu>0$, one has $\Sigma(x,v)\subseteq \overline{\gamma}+\delta\mathbb{B}$ whenever $|x-\widehat{x}|<\mu$, $|v-\widehat{\gamma}| < (\delta-|\widehat{\gamma}-\overline{\gamma}|)/3$.*

To prove this, suppose not: then there are sequences $x_k\to\widehat{x}$ and v_k for which

$$|v_k-\widehat{\gamma}| < \frac{\delta-|\widehat{\gamma}-\overline{\gamma}|}{3} \tag{4.5}$$

and yet $\Sigma(x_k,v_k)$ contains some point outside $\overline{\gamma}+\delta\mathbb{B}$. Call this point π_k. Then $|\pi_k-\overline{\gamma}|>\delta$ and consequently

$$\begin{aligned} d_{\Gamma(x_k)}\,(v_k) &= |(\pi_k-\overline{\gamma}) - (v_k-\overline{\gamma})| \\ &\ge |\pi_k-\overline{\gamma}| - |v_k-\overline{\gamma}| \\ &> \delta - |v_k-\widehat{\gamma}| - |\widehat{\gamma}-\overline{\gamma}|. \end{aligned} \tag{4.6}$$

But the semicontinuity property (i) provides a sequence $\gamma_k \in \Gamma(x_k)$ such that $\gamma_k \to \widehat{\gamma}$. For this sequence,

$$d_{\Gamma(x_k)}(v_k) \leq |v_k - \gamma_k| \leq |v_k - \widehat{\gamma}| + |\widehat{\gamma} - \gamma_k|. \tag{4.7}$$

Concatenating inequalities (4.6) and (4.7) and applying condition (4.5) yields

$$\begin{aligned} 2|v_k - \widehat{\gamma}| &\geq \delta - |\widehat{\gamma} - \overline{\gamma}| - |\widehat{\gamma} - \gamma_k| \\ &> 3|v_k - \widehat{\gamma}| - |\widehat{\gamma} - \gamma_k|, \end{aligned}$$

whence $|v_k - \widehat{\gamma}| < |\gamma_k - \widehat{\gamma}|$. In particular, $v_k \to \widehat{\gamma}$. Taking the limit in the previous inequality then gives $0 \geq \delta - |\widehat{\gamma} - \overline{\gamma}|$. This is a contradiction, since the right side here is positive by construction. The claim holds.

We apply the claim first to the point $(\widehat{x}, \widehat{\gamma}) = (\overline{x}, \overline{\gamma})$. In view of (4.4), this shows that ∂d_Γ is bounded by K throughout the interior of some set $(\overline{x} + \mu\mathbb{B}) \times (\overline{\gamma} + (\delta/3)\mathbb{B})$, but provides no information about the size of $\mu > 0$. To balance the need for quantitative information in both directions, consider

$$\widehat{\mu} = \sup\left\{\mu \in (0, \varepsilon) \,:\, \partial d_\Gamma(x, v) \subseteq K\mathbb{B} \;\forall x \in \overline{x} + \mu\mathbb{B},\; v \in \overline{\gamma} + \delta_0\mathbb{B}\right\},$$

where δ_0 is defined in the theorem statement. Notice that for every (x, v) where $|x - \overline{x}| < \widehat{\mu}$ and $|v - \overline{\gamma}| \leq \delta_0$, one has $\partial d_\Gamma(x, v) \subseteq K\mathbb{B}$. Thus d_Γ is Lipschitz continuous of rank K on the set just described. In particular, Γ is Aubin continuous there, and consequently every y with $|y - \overline{x}| < \widehat{\mu}$ obeys

$$\begin{aligned} &\overline{\gamma} \in \Gamma(\overline{x}) \cap (\overline{\gamma} + \delta_0\mathbb{B}) \subseteq \Gamma(y) + K|y - \overline{x}|\mathbb{B}, \text{ i.e.,} \\ &\Gamma(y) \cap (\overline{\gamma} + K|y - \overline{x}|\mathbb{B}) \neq \emptyset. \end{aligned} \tag{4.8}$$

The closed-graph property of Γ makes it elementary to extend (4.8) to all y in the closed set $\overline{x} + \widehat{\mu}\mathbb{B}$.

Let us prove that $\widehat{\mu} \geq \varepsilon_0$. Suppose this statement is false, i.e., $\widehat{\mu} < \varepsilon_0$: then every sufficiently large integer k admits a corresponding point (x_k, v_k) with $|x_k - \overline{x}| < \widehat{\mu} + 1/k < \varepsilon$, $|v_k - \overline{\gamma}| \leq \delta_0$, but $\partial d_\Gamma(x, v) \not\subseteq K\mathbb{B}$. By passing to a subsequence if necessary we can assume that (x_k, v_k) converges to some point $(\widehat{x}, \widehat{v})$ satisfying $|\widehat{x} - \overline{x}| \leq \widehat{\mu} < \varepsilon_0$ and $|\widehat{v} - \overline{\gamma}| \leq \delta_0 = \delta/6$. From the strong form of (4.8) described in the previous paragraph there exists some point $\widehat{\gamma}$ in $\Gamma(\widehat{x})$ such that $|\widehat{\gamma} - \overline{\gamma}| \leq K|\widehat{x} - \overline{x}|$. Our claim applies to the point $(\widehat{x}, \widehat{\gamma})$: it says that the estimate $\partial d_\Gamma \subseteq K\mathbb{B}$ holds throughout some set of the form

$$(\widehat{x} + \mu\mathbb{B}) \times \left(\widehat{\gamma} + \frac{\delta - |\widehat{\gamma} - \overline{\gamma}|}{3}\mathbb{B}\right),$$

where $\mu > 0$ and (by choice of ε_0)

$$\frac{\delta - |\widehat{\gamma} - \overline{\gamma}|}{3} \geq \frac{\delta - K|\widehat{x} - \overline{x}|}{3} \geq \frac{\delta - K\left(\frac{\delta}{9K}\right)}{3} = \frac{8\delta}{27}.$$

But the point $(\widehat{x}, \widehat{v})$ satisfies

$$\begin{aligned} |\widehat{v} - \widehat{\gamma}| \leq |\widehat{v} - \overline{\gamma}| + |\overline{\gamma} - \widehat{\gamma}| &\leq \frac{\delta}{6} + K|\widehat{x} - \overline{x}| \\ &\leq \frac{\delta}{6} + K\left(\frac{\delta}{9K}\right) = \frac{5\delta}{18}. \end{aligned}$$

Now $\frac{8\delta}{27} > \frac{5\delta}{18}$, so these two estimates show that $(\widehat{x}, \widehat{v})$ lies in the *interior* of a set in which ∂d_Γ is bounded by K. This contradicts the stated properties of the sequence (x_k, v_k), and completes our justification that $\widehat{\mu} \geq \varepsilon_0$.

These arguments show that d_Γ is Lipschitz with rank K on a set containing $(\overline{x} + \varepsilon_0 \mathbb{B}) \times (\overline{\gamma} + \delta_0 \mathbb{B})$. The desired result now follows from Lemma 4.2. □

5. Aubin Continuity in Lagrangian and Hamiltonian Terms. Like Section 4, this section is logically independent of the others in the paper, although the similarity of the notation is deliberate.

Given $\overline{x}$ in $\mathbb{R}^n$ and $\varepsilon > 0$, write $\Omega = \overline{x} + \varepsilon\mathbb{B}$ and consider a Lagrangian $L\colon \Omega \times \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. Assume that for every $x \in \Omega$, the function $v \mapsto L(x, v)$ is closed, proper, and convex. Use L to define the multifunction $E(x) := \operatorname{epi} L(x, \cdot)$. Notice that for every $x \in \Omega$, this multifunction has nonempty closed convex values. In this section we characterize the Aubin continuity of E near $(\overline{x}, \overline{v})$ in terms of conditions on L and its associated Hamiltonian $H(x,p) := \sup\{\langle p\,,\, v\rangle - L(x,v) \,:\, v \in \mathbb{R}^n\}$.

THEOREM 5.1. *Fix any point $(\overline{v}, \overline{L})$ in $E(\overline{x})$, along with scalars $\varepsilon > 0$, $K \geq 0$. Write $\mathbb{B}' = \mathbb{B} \times [-1, 1]$ for the unit ball in in $\mathbb{R}^n \times \mathbb{R}$. Then for any $x, y \in \overline{x} + \varepsilon\mathbb{B}$, conditions* (a)–(c) *below are equivalent:*

(a) $E(x) \cap ((\overline{v}, \overline{L}) + \delta\mathbb{B}') \subseteq E(y) + K|y - x|\mathbb{B}'$;

(b) *For any $u \in \overline{v} + \delta\mathbb{B}$ obeying $L(x,u) \leq \overline{L} + \delta$, there exists $v \in \mathbb{R}^n$ satisfying*

  (i) $|v - u| \leq K|y - x|$, *and*

  (ii) $L(y, v) \leq \max\{\overline{L} - \delta, L(x,u)\} + K|y - x|$;

(c) *For any $p \in \mathbb{R}^n$,*

$$\inf_{\substack{p' \in \mathbb{R}^n \\ \theta > 0}} \left\{\theta H(x, p'/\theta) + \delta|p' - p| + \delta|\theta - 1| + \langle p - p'\,,\, \overline{v}\rangle + (\theta - 1)\overline{L}\right\} \leq H(x,p) + K\,(1 + |p|)\,|y - x|. \tag{5.1}$$

Proof. (a⇒b) Suppose (a) holds. If $u \in \mathbb{R}^n$ satisfies $L(x, u) \leq \overline{L} + \delta$, then the point $\left(u, \max\{\overline{L} - \delta, L(x,u)\}\right)$ lies in the left side shown in condition (a): thus

$$\left(u, \max\{\overline{L} - \delta, L(x,u)\}\right) \in E(y) + K|y - x|\mathbb{B}'.$$

In particular, there has to be a point (h, r) with $\max\{|h|, |r|\} \leq K|y - x|$ such that

$$\left(u, \max\{\overline{L} - \delta, L(x,u)\}\right) + (h, r) \in E(y).$$

The special shape of the epigraph set $E(y)$ allows us to replace r by the larger value $K|y - x|$ on the left side: in this case, defining $v = u + h$ gives $|v - u| \leq K|y - x|$ and

$$\left(v, \max\{\overline{L} - \delta, L(x,u)\} + K|y - x|\right) \in E(y),$$
$$\text{i.e., } L(y, v) \leq \max\{\overline{L} - \delta, L(x,u)\} + K|y - x|.$$

(b⇒a) Suppose (b) holds. Let (u, r) be a vector on the left side in (a). Then $L(x,u) \leq r$ and $\overline{L} - \delta \leq r \leq \overline{L} + \delta$, so (b) provides a vector v such that

(i′) $|v - u| \leq K|y - x|$,

(ii′) $L(y, v) \leq \max\{\overline{L} - \delta, L(x,u)\} + K|y - x| \leq r + K|y - x|$.

Thus $r \geq L(y, v) - K|y - x|$: the special shape of the epigraph set $E(y)$ ensures that $(u, r) \in E(y) + K|y - x|\mathbb{B}$, as required.

(c⇔a) The right side in (a) is a nonempty, closed convex set, since it arises as the sum of a closed convex set and a compact convex set. A separation theorem customized for epigraphs implies that an equivalent formulation of (a) is

$$\sigma_{\mathrm{LS}}(p,-1) \le \sigma_{\mathrm{RS}}(p,-1) \quad \forall p \in \mathbb{R}^n. \tag{5.2}$$

We calculate

$$\begin{aligned} \sigma_{\mathrm{RS}}(p,-1) &= \sigma_{E(y)}(p,-1) + K|y-x|\sigma_{\mathbb{B}'}(p,-1) \\ &= H(t,y,p) + K\|(p,-1)\|_*|y-x|, \end{aligned} \tag{5.3}$$

where $\|(v,r)\|_* = |v| + |r|$ is the norm on $\mathbb{R}^n \times \mathbb{R}$ dual to the one defining $\mathbb{B}'$ there.

Basic convex analysis (Rockafellar [20], Chap. 16) affirms that for any nonempty convex sets C and D, with D compact, one has

$$\sigma_{C\cap D} = \mathrm{cl}(\sigma_C \,\square\, \sigma_D) = \sigma_C \,\square\, \sigma_D.$$

(The second equation here holds because the convex function $\sigma_C \,\square\, \sigma_D$ is finite, hence continuous, on the whole space.) It follows that

$$\begin{aligned} &\sigma_{\mathrm{LS}}(p,-1) \\ &\quad = \left(\sigma_{E(x)} \,\square\, \sigma_{(\overline{v},\overline{L})+\delta\mathbb{B}'}\right)(p,-1) \\ &\quad = \inf_{(p',q')} \left\{\sigma_{E(x)}(p',q') + \delta\|(p,-1)-(p',q')\|_* + \langle p-p'\,,\,\overline{v}\rangle + (-1-q')\overline{L}\right\}. \end{aligned}$$

Now $\sigma_{E(x)}(p',q') = \infty$ whenever $q' > 0$, so the latter infimum can be restricted to those points where $q' \le 0$. Furthermore, at any point (p',q') where $q' = 0$, the special features of epigraph sets imply that the quantity $\sigma_{E(x)}(p',0) + \delta\|(p,-1)-(p',0)\|_*$ can be realized as a limit of some sequence $\sigma_{E(x)}(p'_k,q'_k) + \delta\|(p,-1)-(p'_k,q'_k)\|_*$ with $q'_k < 0$. Thus the infimum can be restricted to points where $q' < 0$. So we write $\theta = -q' > 0$, and use the observation that

$$\sigma_{E(x)}(p',-\theta) = \theta\sigma_{E(x)}(p'/\theta,-1) = \theta H(x,p'/\theta)$$

to obtain

$$\sigma_{\mathrm{LS}}(p,-1) = \inf_{\substack{p'\in\mathbb{R}^n \\ \theta>0}} \left\{\theta H(x,p'/\theta) + \delta\|(p'-p,\theta-1)\|_* + \langle p-p'\,,\,\overline{v}\rangle + (\theta-1)\overline{L}\right\}. \tag{5.4}$$

Equations (5.3) and (5.4) reveal that condition (5.2) is equivalent to (c), whereas (5.2) is equivalent to (a) by construction. □

Clarke's Strong Lipschitz Condition. In treating the generalized problem of Bolza in [4, Chap. 4], Clarke imposes a Hamiltonian requirement called the "strong Lipschitz condition," which asks that for all x and y in some large enough ball,

$$H(y,p) \le H(x,p) + K\,(1+|p|)\,|y-x| \qquad \forall p \in \mathbb{R}^n. \tag{5.5}$$

Our next corollary shows that Aubin continuity of the sort utilized here is a less demanding hypothesis.

5.2. COROLLARY. *If H satisfies the strong Lipschitz condition* (5.5)*, then H satisfies each of the equivalent conditions* (a)–(c) *in Theorem* 5.1 *for every $\delta > 0$.*

Proof. Choose $p' = p$, $\theta = 1$ in (5.1) to see that RS(5.1) $\le$ RS(5.5). Thus (5.5) implies condition (c) of Theorem 5.1. □

To see that (5.1) can be strictly weaker than (5.5), consider the example of $L(x, v) = \frac{1}{2}[v^2 + x^2 v^2]$. It is easy to compute that $H(x, p) = \frac{1}{2}p^2/(1 + x^2)$. For any $\varepsilon > 0$, then, there is a constant $\sigma_\varepsilon > 0$ such that $|H_x(x, p)| \ge \sigma_\varepsilon |p|^2$ for some x in $[-\varepsilon, \varepsilon]$. In particular, the strong Lipschitz condition (5.5) fails. However, for any fixed x, y in $[-\varepsilon, \varepsilon]$, the choices $\theta = 1$ and $p' = p\sqrt{1 + y^2}/\sqrt{1 + x^2}$ in (5.1) give

$$\mathrm{LS}(5.1) \le H(x, p) + \delta\left| p\frac{\sqrt{1 + y^2}}{\sqrt{1 + x^2}} - p\right| \le H(x, p) + (\delta\sqrt{1 + x^2})|p||y - x|.$$

Thus inequality (5.1) holds for any $\delta > 0$, with $K = \delta\sqrt{1 + \varepsilon^2}$.

In his later treatment of the generalized problem of Bolza in [5], Clarke replaces his "strong Lipschitz condition" with a "weak Lipschitz condition." Although the latter condition is difficult to compare to our Aubin continuity assumption, it does hold for the simple example introduced above.

6. Problems with Explicit State Constraints. In deriving Theorem 3.1 from Loewen and Rockafellar [14], we have transcribed only the conclusions that pertain in the absence of explicit state constraints. Such constraints are handled in [14], however, and a proof perfectly analogous to the one given in Section 3 allows us to incorporate them into the main result of this paper as well. In this section we summarize the new ideas required in this broader context, and develop the associated enlargement of Theorem 2.1. Fuller explanations of the new ingredients and ideas appear in [13], [14].

Consider the following extension of problem $(\mathcal{P})$ in which state constraints now explicitly enter:

$$(\mathcal{P}^*) \qquad \begin{aligned} &\text{minimize } \ \Lambda[x] := l(x(a), x(b)) + \int_a^b L(t, x(t), \dot{x}(t))\, dt \\ &\text{subject to } \ x(t) \in X(t) \ \ \forall t \in [a, b]. \end{aligned}$$

We retain hypotheses (H1)–(H5) of Section 2 and impose the following conditions on the state constraint multifunction X:

(H6) Each set $X(t)$ is closed, and the multifunction $t \rightrightarrows X(t)$ is lower semicontinuous, which means that for every point $(t_0, x_0) \in \Omega \cap (\operatorname{gph} X)$, and for every sequence $t_k \to t_0$ in $[a, b]$, there exists a sequence $x_k \to x_0$ with $x_k \in X(t_k)$ for all k.

For each (t, x) in $\Omega \cap (\operatorname{gph} X)$ let

$$\overline{N}_X(t, x) = \operatorname{cl}\operatorname{co}\Big\{\nu \in \mathbb{R}^n : \nu = \lim_{k\to\infty} \nu_k \text{ for some sequences } \nu_k \in \hat{\hat{N}}_{X(t_k)}(x_k),\ (t_k, x_k) \xrightarrow[\operatorname{gph} X]{} (t, x)\Big\}. \tag{6.1}$$

This closed convex cone specifies the directions in which the adjoint function p is allowed to jump when the state constraint is active. Recall that a vector-valued measure dp is called "$\overline{N}_X(t, \overline{x}(t))$-valued" when dp can be written as $\nu(t)\, d\mu(t)$ for some nonnegative measure μ on $[a,b]$ with $dp \ll \mu$ and some measurable selection $\nu(t) \in \overline{N}_X(t,\overline{x}(t))$ μ-a.e.

With these additional ingredients, our main result takes the form stated below. This version differs from the original one, Theorem 2.1, primarily in that its adjoint function p is only of bounded variation, not absolutely continuous as it must be when $X \equiv \mathbb{R}^n$. In particular, the endpoints $p(a)$ and $p(b)$ may differ from the one-sided limits $p(a+)$ and $p(b-)$ in cases where the measure dp has an atom at one or both ends of the interval $[a,b]$.

THEOREM 6.1. *Assume* (H1)–(H6). *Suppose that the arc $\overline{x}$ solves problem $(\mathcal{P}^*)$, and that the constraint qualification below is satisfied:*

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

Then either the normal conditions or the degenerate conditions below are satisfied by some function $p \in BV([a,b];\mathbb{R}^n)$ for which the singular part of the measure dp is $\overline{N}_X(t,\overline{x}(t))$-valued, and hence in particular is supported on the set

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

[Normal Conditions]:

(a) $\dot{p}(t) \in \operatorname{co}\big\{w \,:\, (w,p(t)) \in \partial L(t,\overline{x}(t),\dot{\overline{x}}(t))\big\} + \overline{N}_X(t,\overline{x}(t))$ a.e. $t \in [a,b]$,

(b) $(p(a),-p(b)) \in \partial l(\overline{x}(a),\overline{x}(b))$.

[Singular Conditions]:

(a$^\infty$) $\dot{p}(t) \in \operatorname{co}\big\{w \,:\, (w,p(t)) \in \partial^\infty L(t,\overline{x}(t),\dot{\overline{x}}(t))\big\} + \overline{N}_X(t,\overline{x}(t))$ a.e. $t \in [a,b]$,

(b$^\infty$) $(p(a),-p(b)) \in \partial^\infty l(\overline{x}(a),\overline{x}(b))$.

(*In particular, if the only such function p satisfying conditions* (a$^\infty$)–(b$^\infty$) *is identically zero, then the normal conditions are satisfied.*) *In the normal conditions, assertion* (a) *is equivalent to*

(a′) $\dot{p}(t) \in \operatorname{co}\big\{w \,:\, (-w,\dot{\overline{x}}(t)) \in \partial H(t,\overline{x}(t),p(t))\big\} + \overline{N}_X(t,\overline{x}(t))$ a.e. $t \in [a,b]$.

Also, conditions (a) *and* (a′) *imply that for almost all t in $[a,b]$,*

(c) $p(t) \in \partial_v L(t,\overline{x}(t),\dot{\overline{x}}(t)) = \operatorname*{argmax}_{q\in\mathbb{R}^n}\big\{\langle q\,,\,\dot{\overline{x}}(t)\rangle - H(t,\overline{x}(t),q)\big\}$, *and*

$\dot{\overline{x}}(t) \in \partial_p H(t,\overline{x}(t),p(t)) = \operatorname*{argmax}_{v\in\mathbb{R}^n}\{\langle p(t)\,,\,v\rangle - L(t,\overline{x}(t),v)\}$.

7. Application: The Lipschitz-plus-Indicator Case. Many practical problems permit a clear distinction to be drawn between the constraints and the costs. They can thus be expressed in the form of $(\mathcal{P}_1)$ in Section 1. To this model we can now add the possibility of explicit state constraints. We focus then on the problem:

$$(\mathcal{P}_1^*)\qquad \begin{aligned} &\text{minimize } \Lambda_1[x] := l_1(x(a),x(b)) + \int_a^b L_1(t,x(t),\dot{x}(t))\,dt\\ &\text{subject to } \ (x(a),x(b)) \in S \ \text{ and } \ \dot{x}(t) \in F(t,x(t)) \text{ a.e. } t \in [a,b]\\ &\text{along with } \ x(t) \in X(t) \ \ \forall t \in [a,b], \end{aligned}$$

in which the endpoint cost function l_1 and, for each t, the running cost function $L_1(t,\cdot,\cdot)$ are assumed to be locally Lipschitz continuous. To display this problem as an instance of the general problem's state-constrained version $(\mathcal{P}^*)$ treated in Section 6, it suffices, as we have noted above, to take $l = l_1 + \Psi_S$ and $L = L_1 + \Psi_{\operatorname{gph} F}$, where Ψ_S and $\Psi_{\operatorname{gph} F}$ are the indicators of S and the graph of the multifunction F.

Suitable hypotheses on l_1 and S, as well as L_1 and F, are as follows. Again, they refer to the constant $\varepsilon > 0$ appearing in the definition of Ω and to two positive-valued integrable functions δ and κ.

(H1+) The endpoint cost function $l_1\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitzian on $\Omega_a \times \Omega_b$; the target set $S \subseteq \mathbb{R}^n \times \mathbb{R}^n$ is closed.

(H2+) The integrand $L_1\colon \Omega \times \mathbb{R}^n \to \mathbb{R}$ and the multifunction $F\colon \Omega \rightrightarrows \mathbb{R}^n$ are $\mathcal{L}\times\mathcal{B}$-measurable.

(H3+) For each fixed pair (t,x) in Ω, the function $v \mapsto L_1(t,x,v)$ is convex on $\mathbb{R}^n$, while the set $F(t,x)$ is convex.

(H4+) For almost every t in $[a,b]$, the function $(x,v) \mapsto L_1(t,x,v)$ is finite-valued and lower semicontinuous on $\Omega_t \times \mathbb{R}^n$, while the multifunction $x \mapsto\!\!\!\rightrightarrows F(t,x)$ has closed graph. Furthermore, one has the following epi-continuity property: for any point $(\widehat{x},\widehat{v})$ in $\operatorname{gph} F$ where $|\widehat{x}-\overline{x}(t)| < \varepsilon$ and $L(t,\widehat{x},\widehat{v})$ is finite, and any sequence $x_k \to \widehat{x}$ in Ω_t, there exists a sequence $v_k \to \widehat{v}$ satisfying both $v_k \in F(t,x_k)$ and $L_1(t,x_k,v_k) \to L_1(t,\widehat{x},\widehat{v})$.

(H5+) The ratio $\kappa(t)/\delta(t)$ is essentially bounded. For almost all t in $[a,b]$, the function $(x,v) \mapsto L_1(t,x,v)$ is Lipschitz of rank $\kappa(t)$ on the set $(\overline{x}(t)+\varepsilon\mathbb{B}) \times (\dot{\overline{x}}(t) + \delta(t)\mathbb{B})$, while the multifunction F satisfies

$$|w| \le \kappa(t)\,(1+|p|) \ \text{ for all } (w,p) \in N_{\operatorname{gph} F(t,\cdot)}\,(x,v)\,,$$

whenever $|x-\overline{x}(t)| < \varepsilon$, $\left|v-\dot{\overline{x}}(t)\right| < \delta(t)$.

This case is especially interesting because the Lipschitz continuity of l_1 and L_1 ensures that the singular subgradients of l and L coincide with the usual subgradients of the reduced functions $l_0 = \Psi_S$ and $L_0(t,x,v) = \Psi_{\operatorname{gph} F}(t,x,v)$. This makes it possible to expand the degenerate conditions of Theorem 6.1, which now take the form

(a^∞) $\dot{p}(t) \in \operatorname{co}\left\{v \,:\, (v,p(t)) \in \partial L_0(t,\overline{x}(t),\dot{\overline{x}}(t))\right\} + \overline{N}_X(t,\overline{x}(t))$ a.e.,

(b^∞) $(p(a),-p(b)) \in \partial l_0(\overline{x}(a),\overline{x}(b)) = N_S\,(\overline{x}(a),\overline{x}(b))$.

Rockafellar's equivalence results in [31] certainly apply to L_0 as well as to L, and consequently condition (a^∞) has the equivalent Hamiltonian form

$$\dot{p}(t) \in \operatorname{co}\left\{w \,:\, (-w,\dot{\overline{x}}(t)) \in \partial H_0(t,\overline{x}(t),p(t))\right\} \ \text{ a.e. } t \in [a,b],$$

where, of course, $H_0(t,x,p) := \sup\{\langle p\,,\,v\rangle \,:\, v \in F(t,x)\}$ is the Hamiltonian corresponding to L_0. Either of the equivalent forms of (a^∞) implies a corresponding argmax condition analogous to (c) in Theorem 6.1.

To summarize these developments, define the Lagrangian and Hamiltonian of index λ, for any $\lambda > 0$, by

$$\begin{aligned} L_\lambda(t,x,v) &:= \lambda L_1(t,x,v) + \Psi_{\operatorname{gph} F}(t,x,v),\\ H_\lambda(t,x,p) &:= \sup\left\{\langle p\,,\,v\rangle - \lambda L_1(t,x,v) \,:\, v \in F(t,x)\right\}. \end{aligned}$$

Then the following result holds.

THEOREM 7.1. *Assume* (H1+)–(H5+) *and* (H6)*. Suppose that the arc* $\overline{x}$ *solves problem* $(\mathcal{P}_1^*)$*, and that the constraint qualification* (CQ) *of Theorem* 6.1 *is satisfied. Then there exist* $p \in BV([a,b];\mathbb{R}^n)$ *and a constant* $\lambda \in \{0,1\}$*, not both zero, such that for almost all* t *in* $[a,b]$*,*

(a) $\dot{p}(t) \in \operatorname{co}\left\{w \,:\, (-w,\dot{\overline{x}}(t)) \in \partial H_\lambda(t,\overline{x}(t),p(t))\right\} + \overline{N}_X(t,\overline{x}(t))$,

(b) $\dot{p}(t) \in \operatorname{co}\left\{w \,:\, (w,p(t)) \in \partial L_\lambda(t,\overline{x}(t),\dot{\overline{x}}(t))\right\} + \overline{N}_X(t,\overline{x}(t))$,

(c) $p(t) \in \partial_v L_\lambda(t, \overline{x}(t), \dot{\overline{x}}(t)) = \operatorname*{argmax}_{q \in \mathbb{R}^n} \big\{ \langle q\, ,\, \dot{\overline{x}}(t) \rangle - H_\lambda(t, \overline{x}(t), q) \big\},$

$\dot{\overline{x}}(t) \in \partial_p H_\lambda(t, \overline{x}(t), p(t)) = \operatorname*{argmax}_{v \in F(t, \overline{x}(t))} \{ \langle p(t)\, ,\, v \rangle - \lambda L_1(t, \overline{x}(t), v) \}.$

Furthermore,

(d) $(p(a), -p(b)) \in \partial\left(\lambda l_1 + \Psi_S\right)(\overline{x}(a), \overline{x}(b)) \subseteq \lambda \partial l_1(\overline{x}(a), \overline{x}(b)) + N_S\left(\overline{x}(a), \overline{x}(b)\right),$

(e) *the singular part of the measure dp is $\overline{N}_X(t, \overline{x}(t))$-valued, and thus is supported on the set* $\big\{t \,:\, \overline{N}_X(t, \overline{x}(t)) \neq \{0\}\big\} = \{t \,:\, (t, \overline{x}(t)) \in \operatorname{bdy} \operatorname{gph} X\}.$

Notice that when $L_1 \equiv 0$, problem $(\mathcal{P}_1^*)$ reproduces the unbounded differential inclusion control problem of Loewen and Rockafellar [14]. The conclusions of Theorem 7.1 then correspond exactly to those of [14, Theorem 4.3], but with three major improvements: they allow for a nonzero integrand L_1, offer the alternative formulation of the Aubin continuity hypothesis in (H5+), and present a sharper Hamiltonian inclusion in (a).

REFERENCES

[1] J.-P. Aubin, *Lipschitz behavior of solutions to convex minimization problems*, Math. Oper. Res., 9(1984), pp. 87–111.

[2] G. A. Bliss, *Lectures on the Calculus of Variations*, University of Chicago Press, Chicago, 1946.

[3] F. H. Clarke, *The generalized problem of Bolza*, SIAM J. Control Optim., 14(1976), pp. 682–699.

[4] ———, *Optimization and Nonsmooth Analysis,* New York: Wiley-Interscience, 1983. (Now reissued by SIAM, Classics in Applied Mathematics, 1991.)

[5] ———, *Hamiltonian analysis of the generalized problem of Bolza*, Trans. Amer. Math. Soc., 301(1987), pp. 385–400.

[6] ———, *Methods of Dynamic and Nonsmooth Optimization,* CBMS-NSF Regional Conference Series, vol. 57. SIAM, Philadelphia, 1989.

[7] F. H. Clarke and R. B. Vinter, *Regularity properties of solutions to the basic problem in the calculus of variations*, Trans. Amer. Math. Soc., 289(1985), pp. 73–98.

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

[9] A. D. Ioffe, *Approximate subdifferentials and applications I: the finite dimensional theory*, Trans. Amer. Math. Soc., 281(1984), pp. 389–416.

[10] A. D. Ioffe and R. T. Rockafellar, *The Euler and Weierstrass conditions for nonsmooth variational problems*, preprint, 1994.

[11] P. D. Loewen, *The proximal subgradient formula in Banach space*, Canad. Math. Bull., 31(1988), pp. 353–361.

[12] ———, *Optimal Control via Nonsmooth Analysis,* CRM-AMS Lecture Notes in Mathematics vol. 2, American Math. Society, Providence, RI, 1993.

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

[14] ———, *Optimal control of unbounded differential inclusions*, SIAM J. Control Optim., 32(1994), pp. 442–470.

[15] B. S. Mordukhovich, *Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems*, Dokl. Acad. Nauk SSSR 254(1980), pp. 1072–1076. (Russian; English translation in Soviet Math. Dokl., 22(1980), pp. 526–530.)

[16] ———, *Optimization and approximation of differential inclusions*, Kibernetika 6(1988), pp. 83–89. (Russian; English translation in Cybernetics, 24(1988), pp. 781–788.)

[17] ———, *Approximation Methods in Problems of Optimization and Control,* Nauka, Moscow, 1988 (Chapter 3).

[18] ———, *Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions*, Trans. Amer. Math. Soc., 340(1993), pp. 1–35.

[19] ———, *Discrete approximations and refined Euler-Lagrange conditions for nonconvex differential inclusions*, SIAM J. Control Optim., to appear.

[20] R. T. Rockafellar, *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[21] ———, *Conjugate convex functions in optimal control and the calculus of variations*, J. Math. Analysis Appl., 23(1970), pp. 174–222.

[22] ———, *Existence and duality theorems for convex problems of Bolza*, Trans. Amer. Math. Soc., 159(1971), pp. 1–40.

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

[24] ———, *Existence theorems for general control problems of Bolza and Lagrange*, Adv. in Math., 15(1975), pp. 312–333.

[25] ———, *Dual problems of Lagrange for arcs of bounded variation*, in: Calculus of Variations and Control Theory, D. L. Russell (ed.), Academic Press, New York, pp. 155–192, 1976.

[26] ———, *Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization*, Math. of Oper. Res., 6(1981), pp. 424–436.

[27] ———, *Extensions of subgradient calculus with applications to optimization*, Nonlinear Analysis, Theory, Meth. & Appl., 9(1985), pp. 665–698.

[28] ———, *Lipschitzian properties of multifunctions*, Nonlinear Analysis, Theory, Meth. & Appl., 9(1985), pp. 867–885.

[29] ———, *Hamiltonian trajectories and duality in the optimal control of linear systems with convex costs*, SIAM J. Control Optim., 27(1989), pp. 1007–1025.

[30] ———, *Dualization of subgradient conditions for optimality*, Nonlinear Analysis, Theory, Meth. & Appl., 20(1993), pp. 627–646.

[31] ———, *Equivalent subgradient versions of Hamiltonian and Euler-Lagrange equations in variational analysis*, SIAM J. Control Optim., submitted.

[32] G. V. Smirnov, *Discrete approximations and optimal colutions to differential inclusions*, Kibernetika, 10(1991), pp. 76–79. (Russian; English translation in Cybernetics, 27(1991), pp. 101–107.)