TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 325, Number 1, May 1991

THE ADJOINT ARC IN NONSMOOTH OPTIMIZATION

PHILIP D. LOEWEN AND R. T. ROCKAFELLAR

ABSTRACT. We extend the theory of necessary conditions for nonsmooth problems of Bolza in three ways: first, we incorporate state constraints of the intrinsic type $x(t) \in X(t)$ for all t; second, we make no assumption of calmness or normality; and third, we show that a single adjoint function of bounded variation simultaneously satisfies the Hamiltonian inclusion, the Euler-Lagrange inclusion, and the Weierstrass-Pontryagin maximum condition, along with the usual transversality relations.

1. INTRODUCTION

In this paper we refine the necessary conditions for optimality in the following nonsmooth problem of dynamic optimization:

$$
\text{(P)} \qquad \begin{aligned}
&\text{minimize} && \Lambda[x] := l(x(a), x(b)) + \int_a^b L(t, x(t), \dot{x}(t))\, dt \\
&\text{subject to} && x \in \mathrm{AC}\big([a, b]; \mathbb{R}^n\big), \\
& && \dot{x}(t) \in F(t, x(t)) \quad \text{a.e. } t \in [a, b], \\
& && x(t) \in X(t) \quad \forall t \in [a, b], \\
& && (x(a), x(b)) \in S.
\end{aligned}
$$

This problem involves the constrained minimization of a continuous Bolza functional Λ over the space of absolutely continuous functions x carrying the given interval $[a, b]$ into $\mathbb{R}^n$. The constraints are typical of those in optimal control: they are the dynamic constraint $\dot{x}(t) \in F(t, x(t))$ a.e. $t \in [a, b]$; the state constraint $x(t) \in X(t)$ $\forall t \in [a, b]$; and the endpoint constraint $(x(a), x(b)) \in S$. The range of problems that fit into this general framework is very broad (see Clarke [5] for examples).

Our main result, stated below as Theorem 1.1, provides necessary conditions for optimality in (P) which represent a threefold advance in the state of the art. First, they allow a cost functional Λ involving an integral term without imposing any constraint qualification such as normality or calmness. In the case

Received by the editors September 11, 1989.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 49B10, 49B34.
The first author's research was supported by the Natural Science and Engineering Research Council (NSERC) of Canada through grant 5-89441; the second author's research was supported by the National Science Foundation (NSF) through grant DMS-8819586.

©1991 American Mathematical Society
0002-9947/91 $1.00 + $.25 per page

where $L \equiv 0$, our problem is closely related to the extensively studied *differential inclusion problem,* for which necessary conditions without calmness are already available ([2]; see §2). The traditional approach to (P) has been to reduce it to the case $L \equiv 0$ through the introduction of an auxiliary state variable: we take this approach too, but give a sufficiently careful analysis of the resulting differential inclusion problem to exploit this reduction without sacrificing the generality of the original problem. It is not surprising that this is possible, since the functional Λ and the multifunction F are Lipschitzian; however, we believe that this useful observation has not appeared in print before.

Second, our formulation of problem (P) features an intrinsic representation of all the problem's constraints. We regard the state $x \in \mathbb{R}^n$ and its evolution $x(\cdot)$ as the fundamental quantities of interest, and therefore write down the constraints on $x(\cdot)$ without any reference to such parametrizations as $\Phi(x(a), x(b)) \leq 0$, say, for the endpoint constraints. (A satisfactory treatment of problems whose velocity constraint is described intrinsically by a differential inclusion is one of the principal accomplishments of nonsmooth analysis.) In particular, we represent the problem's state constraints in the inclusion form $x(t) \in X(t)$ for all $t \in [a, b]$, where $X(t)$ is a closed set whose time dependence may be rather general. This representation removes some arbitrariness from the statement of the problem, and gives rise to necessary conditions whose significance is more transparent. Our results make explicit the common intuitive understanding that the influence of state constraints is reflected in the possibility of jumps in the adjoint function. Thus we consider adjoint functions ("arcs") p of *bounded variation,* whose jumps (or, more precisely, departures from absolute continuity) occur at instants when the state constraint acts upon the optimal solution, when they drive the adjoint function in the direction of an "outward normal" to the state constraint set $X(t)$. (If $x(t) \in \operatorname{int} X(t)$ for all t in some open interval I, then the only outward normal vector is 0, and it follows that p is absolutely continuous on I.) We note that two functions p_1, p_2 of bounded variation are indistinguishable if $p_1(t-) = p_2(t-)$ and $p_1(t+) = p_2(t+)$ for all $t \in (a, b)$, while $p_1(a) = p_2(a)$ and $p_1(b) = p_2(b)$. Indistinguishability is an equivalence relation on the set of functions with bounded variation on $[a, b]$, and the corresponding set of equivalence classes forms the vector space denoted by $\mathrm{BV}([a, b]; \mathbb{R}^n)$. The "outward normal jump condition" mentioned above asserts, in the case where the sets $X(t)$ are convex, that

$$\text{the singular part of } dp \text{ is } N_{X(t)}(x(t))\text{-valued.}$$

(Here N denotes the usual normal cone of convex analysis.) This implies, for example, that any simple discontinuity in the adjoint function p must be such that the jump vector $p(t+) - p(t-)$ is an outward normal to the state constraint set $X(t)$ at the position $x(t)$.

The third, and most significant, advance reported here is a unification of three types of necessary conditions, whose forms are inspired by the classical conditions of Euler, Hamilton, and Weierstrass. Whereas the Hamiltonian system

and the Euler equation are equivalent in the classical calculus of variations and in some smooth control problems (under suitable hypotheses), this equivalence does not persist in the nonsmooth case. And although necessary conditions for nonsmooth problems have been given in all three of the forms described here, ours is the first work to produce a single adjoint arc which plays all three roles simultaneously.

We now state our main result, although some of the terms in its statement, particularly $N(t, x)$, will not be defined until §2. The maximized Hamiltonian H_λ is defined as follows, for all $\lambda \geq 0$:

$$H_\lambda(t, x, p) := \max\{\langle p, v\rangle - \lambda L(t, x, v) : v \in F(t, x)\}.$$

The hypotheses mentioned in the statement below are given in the early stages of §§2 and 3; the proof occupies §§3 and 4.

1.1. **Theorem.** *Assume* (H1)–(H7). *Let* $\bar{x}$ *solve problem* (P). *Assume the constraint qualification*

(CQ) $\qquad N(t, \bar{x}(t))$ *is pointed, for each* $t \in [a, b]$.

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

(a) $(-\dot{p}(t) + \nu(t), \dot{\bar{x}}(t)) \in \partial H_\lambda(t, \bar{x}(t), p(t))$;
(b) $(\dot{p}(t) - \nu(t), p(t)) \in \lambda \partial L(t, \bar{x}(t), \dot{\bar{x}}(t)) + N_{\mathrm{Gr}\, F(t, \cdot)}(\bar{x}(t), \dot{\bar{x}}(t))$;
(c) $p(t) \in \lambda \partial_v L(t, \bar{x}(t), \dot{\bar{x}}(t)) + N_{F(t, \bar{x}(t))}(\dot{\bar{x}}(t))$.

Moreover,

(d) $(p(a), -p(b)) \in \lambda \partial l(\bar{x}(a), \bar{x}(b)) + N_S(\bar{x}(a), \bar{x}(b))$;
(e) *the singular part of* dp *is* $N(t, \bar{x}(t))$*-valued, and in particular is supported on the set*

$$\{t : N(t, \bar{x}(t)) \neq \{0\}\} = \{t \in [a, b] : (t, \bar{x}(t)) \in \mathrm{bdry}\, \mathrm{Gr}\, X(\cdot)\}.$$

Note that inclusion (a) is the usual Hamiltonian inclusion; inclusion (b) is the Euler inclusion; and inclusion (c) is the appropriate analogue of the classical Weierstrass condition and the maximum principle of Pontryagin. (The generalized gradients appearing in (a) and (b) refer to the last two components only.) Taken independently, inclusions (a) and (b) are equivalent to the more suggestive inclusions

$$(-\dot{p}(t), \dot{\bar{x}}(t)) \in \partial H_\lambda(t, \bar{x}(t), p(t)) - N(t, \bar{x}(t)) \times \{0\},$$

$$(\dot{p}(t), p(t)) \in \lambda \partial L(t, \bar{x}(t), \dot{\bar{x}}(t)) + N_{\mathrm{Gr}\, F(t, \cdot)}(\bar{x}(t), \dot{\bar{x}}(t)) + N(t, \bar{x}(t)) \times \{0\}.$$

(A proof of this fact may be based on the results in [16].) However, the statement of Theorem 1.1 asserts that the same selection $\nu(t)$ from the multifunction $t \rightrightarrows N(t, \bar{x}(t))$ satisfies both of these inclusions at once.

The Generalized Problem of Bolza. It is instructive to compare our problem (P) with the more comprehensive Generalized Problem of Bolza, which is to

$$(\mathrm{P}_1) \qquad \text{minimize } \Lambda_1[x] := l_1(x(a), x(b)) + \int_a^b L_1(t, x(t), \dot{x}(t))\, dt$$

over all $x \in \mathrm{AC}([a, b]; \mathbb{R}^n)$. The key feature of (P_1) is that the functions l_1 and L_1 are allowed to take values in $\mathbb{R} \cup \{+\infty\}$, and that the objective value $+\infty$ may be used to incorporate constraints directly into the objective functional. For example, our problem (P) may be put into this form by setting (for $\lambda \geq 0$)

$$(1.1) \qquad \begin{aligned} l_\lambda(x, y) &:= \lambda l(x, y) + \Psi_S(x, y), \\ L_\lambda(t, x, v) &:= \lambda L(t, x, v) + \Psi_{\mathrm{Gr}\, F(t, \cdot)}(x, v) + \Psi_{X(t)}(x). \end{aligned}$$

When $X(t) \equiv \mathbb{R}^n$, this reduction leads to an instance of (P_1) which satisfies the usual hypotheses imposed on the Generalized Problem of Bolza, as described in [2, Chapter IV]. A transition in the opposite direction is also possible, of course: when confronted with (P_1), we need only define

$$\begin{aligned} F(t, x) &:= \operatorname{dom} L_1(t, x, \cdot), \\ X(t) &:= \bigcup \{\operatorname{dom} L_1(t, \cdot, v) : v \in \mathbb{R}^n\}, \\ S &:= \operatorname{dom} l_1(\cdot, \cdot), \end{aligned}$$

to deduce from (1.1) (with $\lambda = 1$) the definitions of the finite-valued functions l and L needed to recast (P_1) in the form of (P). However, this latter reduction does not allow us to treat the Generalized Problem of Bolza completely, since our analysis of (P) relies upon the assumption that the finite functions l and L, along with the multifunction F, are Lipschitzian.

It is possible to remove our structural assumptions and give a result analogous to Theorem 1.1 for the Generalized Problem of Bolza, provided that state constraints are absent and the problem is calm. These conditions are among the standing hypotheses in [2, Chapter IV]: if we place ourselves in that context, then our results may be combined with Clarke's arguments to yield Theorem 1.2 below. In its statement, we write $H_\lambda(t, x, p) := \sup\{\langle p, v\rangle - L_\lambda(t, x, v) : v \in \mathbb{R}^n\}$: note that this definition is consistent with our earlier notation in the context of problem (P), the link being furnished by (1.1).

1.2. **Theorem.** *Let the data of* (P_1) *satisfy the hypotheses of* [2, *Chapter* IV]. *Suppose that the arc* $\bar{x}$ *solves* (P_1), *and that* (P_1) *is calm at* $\bar{x}$. *Then there exists an absolutely continuous function* $p: [a, b] \to \mathbb{R}^n$ *such that for almost all* $t \in [a, b]$,

(a) $(-\dot{p}(t), \dot{\bar{x}}(t)) \in \partial H_1(t, \bar{x}(t), p(t))$;
(b) $(\dot{p}(t), p(t)) \in \partial L_1(t, \bar{x}(t), \dot{\bar{x}}(t))$;
(c) $p(t) \in \partial_v L_1(t, \bar{x}(t), \dot{\bar{x}}(t))$.

Moreover,

(d) $(p(a), -p(b)) \in \partial l_1(\bar{x}(a), \bar{x}(b))$.

Returning to our original problem, let us suppose that an arc $\overline{x}$ solves (P). If state constraints are absent ($X(t) \equiv \mathbb{R}^n$) and (P) is calm at $\overline{x}$, then the reformulation of (P) as (P_1) sponsored by (1.1) allows us to apply Theorem 1.2 instead of Theorem 1.1. This may lead to necessary conditions which are sharper than those of Theorem 1.1. In conclusion (b), for example, Theorem 1.1 states

$$(\dot{p}, p) \in \partial L(t, \overline{x}, \dot{\overline{x}}) + \partial \Psi_{\mathrm{Gr}\, F(t,\cdot)}(\overline{x}, \dot{\overline{x}}), \tag{1.2}$$

while Theorem 1.2 asserts

$$(\dot{p}, p) \in \partial [L(t, \cdot, \cdot) + \Psi_{\mathrm{Gr}\, F(t,\cdot)}](\overline{x}, \dot{\overline{x}}). \tag{1.3}$$

In general, the right side of (1.3) is a subset of that in (1.2), although equality prevails in certain important cases. It is natural to ask to what extent this sharpening of Theorem 1.1 remains valid when state constraints are retained and the calmness condition is abandoned. For example, when can conclusions (a)–(d) of Theorem 1.1 be replaced by the generally sharper conditions (a′)–(d′) below?

(a′) $(-\dot{p}(t) + \nu(t), \dot{\overline{x}}(t)) \in \partial H_\lambda(t, \overline{x}(t), p(t))$;
(b′) $(\dot{p}(t) - \nu(t), p(t)) \in \partial L_\lambda(t, \overline{x}(t), \dot{\overline{x}}(t))$;
(c′) $p \in \partial_v L_\lambda(t, \overline{x}(t), \dot{\overline{x}}(t))$;
(d′) $(p(a), -p(b)) \in \partial l_\lambda(\overline{x}(a), \overline{x}(b))$.

And how must these conditions be modified if l, L, and F are not assumed to be Lipschitz?

Comments on the literature. Part of the motivation for the research reported here comes from the extensive and relatively complete theory of necessary conditions available for the "convex case" of the Generalized Problem of Bolza (P_1). This theory, put forward by Rockafellar in [12, 13, 14, 19], requires that the extended-valued mappings l_1 and $L_1(t, \cdot, \cdot)$ be convex. The desire to give the problem's constraints an intrinsic representation and the need to consider adjoint arcs of bounded variation are clearly evident in Rockafellar's work; indeed, each of the necessary conditions in Theorem 1.1 has a precursor in [14] or [12].

In the current paper, we assume only that the map $L(t, x, \cdot)$ and the sets $F(t, x)$ are convex: this reduced hypothesis is the current standard, and its widespread acceptance is firmly grounded upon its near necessity for existence theory. However, the full power of Clarke's nonsmooth analysis is required to treat the resulting version of (P), and Clarke's methods and results (as set forth in [2], for example) are fundamental to our approach. In particular, our development starts by quoting Clarke's Hamiltonian necessary conditions for the case $L \equiv 0$ of problem (P). The Euler-Lagrange inclusion may also be found in Clarke's early work [3], but a satisfactory bridge between the two types of necessary conditions has remained unfinished until now. (We note that Raïssi studied their relationship in her thesis [9], where she established their equivalence under certain additional hypotheses.)

A recent advance in another direction has been Clarke's proof of Hamiltonian necessary conditions for the Generalized Problem of Bolza in the absence of a calmness hypothesis [6]. The results presented here also dispense with calmness, largely by assuming enough Lipschitz continuity to sidestep some of the thorny technical issues at the heart of [6]. We hope that the explicit consideration of a Bolza functional in problem (P) will not only be attractive to the users of our necessary conditions, but will also—in conjunction with [6]—provide some useful insight into the questions we have raised above.

2. The differential inclusion problem

In this section we review and extend the known necessary conditions for problem (P) in the case $L \equiv 0$: this special case is denoted by (P_D). Throughout this paper, we use only the Euclidean norm and let B denote the open unit ball in $\mathbb{R}^n$.

Basic hypotheses. The time interval $[a, b]$ is fixed, along with a relatively open subset Ω of $[a, b] \times \mathbb{R}^n$ having nonempty sections,

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

The role of the set Ω is to allow the consideration of local minimizers in (P_D): in the problem we study below, the requirement that $x(t) \in \Omega_t$ for all $t \in [a, b]$ is an implicit constraint. Throughout this section, we make the following hypotheses:

(H1) The endpoint functional l is Lipschitz on the closed set $S_0 := \overline{\Omega_a} \times \overline{\Omega_b}$, and the local endpoint constraint set $S \cap S_0$ is closed.

(H2) The multifunction F is $\mathscr{L} \times \mathscr{B}$ measurable on Ω.

(H3) The sets $F(t, x)$ are nonempty, convex, and compact for each (t, x) in Ω; and there exists a nonnegative function $\phi_F \in L^1[a, b]$ for which

$$F(t, x) \subseteq \phi_F(t)\overline{B} \qquad \forall (t, x) \in \Omega.$$

(H4) There exists a nonnegative function $k_F(\cdot) \in L^1[a, b]$ for which one has the Lipschitz condition

$$F(t, y) \subseteq F(t, x) + k_F(t)|y - x|\overline{B} \qquad \forall t \in [a, b], \ \forall x, y \in \Omega_t.$$

(H5) The state constraint multifunction X has closed values $X(t)$ and is lower semicontinuous on Ω, in the sense that for every point $(t_0, x_0) \in \operatorname{Gr} X \cap \Omega$ and for every sequence $t_i \to t_0$, there exists a sequence $x_i \to x_0$ satisfying $x_i \in X(t_i)$ for all i.

State constraint representation. Hypotheses (H1)–(H4) are standard, as is the assumption that the multifunction X has closed values. Starting with only these hypotheses, let us show that the additional assumption of lower semicontinuity in (H5) does not restrict the class of problems to which our results apply. Let $\overline{x}(\cdot)$ be any arc admissible for (P_D). Then $\operatorname{Gr}\overline{x}$ is a compact subset of Ω, so there exists $\delta > 0$ such that $\overline{x}(t) + \delta\overline{B} \subseteq \Omega_t$ for all $t \in [a, b]$. Define

$T := \{(t, x) : t \in [a, b],\ |x - \overline{x}(t)| < \delta\}$. It is clear that replacing Ω by T in (P_D) has no effect on the structure of the problem near $\overline{x}$. Having made this change, we consider the multifunction

$$\begin{aligned} Y(t) := \{x(t) : & x \in \mathrm{AC}([a, b]; \mathbb{R}^n), \\ & |\dot{x}(\tau)| \le k_F(\tau) \text{ a.e. } \tau \in [a, b], \\ & x(\tau) \in X(\tau) \cap (\overline{x}(\tau) + \delta \overline{B})\ \forall \tau \in [a, b]\}. \end{aligned}$$

Obviously $Y(t) \subseteq X(t)\ \forall t \in [a, b]$. But it is also true that any arc x admissible for (P_D) must obey $x(t) \in Y(t)$ for all $t \in [a, b]$. Therefore we may write Y in place of X without altering the set of arcs admissible for (P_D). Note, however, that Y has closed values and is lower semicontinuous. In making hypothesis (H5), we are essentially assuming that the state constraint sets $X(t)$ have been trimmed to contain only those points which are "reachable" in a very weak sense. This can reasonably be regarded as an aspect of proper formulation of the problem.

Instead of writing state constraints in the intrinsic form $x(t) \in X(t)$, many authors impose a pointwise inequality of the form

$$g(t, x(t)) \le 0 \quad \forall t \in [a, b] \tag{2.1}$$

for some function $g: \Omega \to \mathbb{R}$. While the formulation (2.1) may be technically convenient, especially when g is smooth, it has the serious drawback of being nonintrinsic. In other words, the same constraint can be represented by many different functions g, and an unsuitable choice may unduly complicate a given problem. Moreover, the state constraint regions defined by

$$X(t) := \{x \in \mathbb{R}^n : g(t, x) \le 0\} \tag{2.2}$$

may form a multifunction X which fails to be lower semicontinuous. (Consider, for example, the analytic function $g(t, x) = \sin^2 t \sin^2 x$.) We prefer the inclusion formulation both because it is intrinsic, and because the hypothesis of lower semicontinuity turns out to be useful in itself. On the other hand, once the state constraint sets $X(t)$ have been identified (either directly or via (2.2)), the definition $g(t, x) := \operatorname{dist}(X(t), x)$ can be used to harness theoretical results designed for problems whose state constraints are given in inequality form (2.1). Indeed, we may regard this choice of g as a canonical representation of the state constraint in the formulation (2.1). Hypothesis (H5) ensures that the resulting function g is Lipschitz in x and (jointly) upper semicontinuous in (t, x).

As we saw in the Introduction, jumps in the adjoint function $p(\cdot)$ must be vectors normal to the state constraint region $X(t)$. In the case where X is convex-valued, "normality" refers to the usual normal cone of convex analysis. In the general case, where the shapes of X are relatively unrestricted, an extended notion of normality is required. We describe a suitable candidate here, first directly, and then in relation to known concepts. Our description relies upon the notion of "graphical closure": recall that given a multifunction

$\Gamma\colon \Omega \rightrightarrows \mathbb{R}^n$, the *graphical closure of* Γ is the multifunction $\overline{\Gamma}$ whose graph equals $\overline{\operatorname{Gr}\Gamma}$. More explicitly,

$$\overline{\Gamma}(\omega) = \left\{ \lim_{i\to\infty} \gamma_i \,:\, \gamma_i \in \Gamma(\omega_i),\ \ \omega_i \to \omega \right\} \quad \forall \omega \in \Omega. \tag{2.3}$$

Consider the proximal normal multifunction $\pi\colon \operatorname{Gr} X \rightrightarrows \mathbb{R}^n$ defined by

$$\pi(t, x) := \{\zeta \in \mathbb{R}^n \,:\, \exists \sigma > 0 \text{ s.t. } \langle \zeta\,,\ y - x\rangle \le \sigma |y - x|^2\ \forall y \in X(t)\}. \tag{2.4}$$

This gives the cone of proximal normals to the set $X(t)$ at the point x. The normal cone we propose to use is $N(t, x) := \operatorname{cl}\operatorname{co}\overline{\pi}(t, x)$, i.e., the closed convex cone generated by the graphical closure of π.

Readers familiar with the proximal normal formula [2, 8] will recall that the (Clarke) normal cone to the set $X(t)$ at the point x, denoted $N_{X(t)}(x)$, equals, for each fixed t, the closed convex cone generated by the graphical closure of the multifunction $x \rightrightarrows \pi(t, x)$. Therefore the inclusion $N(t, x) \supseteq N_{X(t)}(x)$ always holds, and it may be strict, since the cone $N(t, x)$ contains information about the t-dependence of $X(\cdot)$ which is absent from $N_{X(t)}(x)$. On the other hand, since $\pi(t, x) \subseteq N_{X(t)}(x)$ for all $(t, x) \in \operatorname{Gr} X$, the graphical closure of the multifunction $(t, x) \rightrightarrows N_{X(t)}(x)$ generates a closed convex cone which always contains $N(t, x)$. The latter inclusion may also be strict. Roughly speaking, this is true because the operations of taking the graphical closure and computing the closed convex hull do not commute. An example which illustrates both strict inclusions is furnished by the state constraint set

$$X(t) := \{(x, y) \in \mathbb{R}^2 \,:\, y \ge tx \text{ or } y \le 0 \text{ or } x \le 0\}.$$

Clearly $X(0) = \mathbb{R}^2$, so $N_{X(0)}(0, 0) = \{0\}$. However, $N(0, (0, 0)) = \{0\} \times \mathbb{R}$, while the graphical closure of $N_{X(\cdot)}(\cdot, \cdot)$ at the point $(0, (0, 0))$ is the right half-plane. This multifunction $X(\cdot)$ is clearly not convex-valued. For convex-valued multifunctions X, the distinctions displayed by this example disappear (see Proposition 2.3(d)).

Another way to understand the multifunction $N(t, x)$ is to choose $g(t, x) := \operatorname{dist}(X(t), x)$ as recommended above, and then to note that

$$N(t, x) = \operatorname{cl}\left[\{0\} \cup \bigcup_{r>0} r\partial_x^{>} g(t, x)\right] \quad \forall (t, x) \in \operatorname{Gr} X \cap \Omega, \tag{2.5}$$

where $\partial_x^{>} g$ is the multifunction introduced by Clarke [2, Theorem 3.5.2] in this same situation:

(2.6)
$$\partial_x^{>} g(t, x) := \operatorname{cl}\operatorname{co}\{\lim \zeta_i \,:\, \zeta_i \in \partial_x g(t_i, y_i),\ \ (t_i, y_i) \to (t, x),\ g(t_i, y_i) > 0\ \forall i\}.$$

Our justification of (2.5) appears in Proposition 2.3 below, whose proof relies upon the following observation.

2.1. **Lemma.** *Let* $g: \Omega \to \mathbb{R}$ *be a function which, for some constant* $K \geq 0$, *satisfies*

$$|g(t, y) - g(t, x)| \leq K|y - x| \quad \forall t \in [a, b], \ \forall x, y \in \Omega_t.$$

Then for every $(t, x) \in \Omega$, *one has*

$$\partial_x^> g(t, x) = \mathrm{co}\{\lim \nabla_x g(t_i, y_i) : (t_i, y_i) \to (t, x), \ g(t_i, y_i) > 0 \ \forall i\}. \tag{2.7}$$

Proof. Fix $(t, x) \in \Omega$ and let $R(t, x)$ denote the set appearing on the right side of (2.7). Observe that the inclusion $\partial_x^> g(t, x) \supseteq R(t, x)$ follows immediately from [2, Proposition 2.2.2].

To prove the reverse inclusion, note that $R(t, x)$ is a compact convex set. Thus it suffices to show that any ζ of the form below lies in $R(t, x)$:

$$\zeta = \lim_{i \to \infty} \zeta_i, \qquad \zeta_i \in \partial_x g(t_i, y_i), \ (t_i, y_i) \to (t, x), \ g(t_i, y_i) > 0. \tag{2.8}$$

To do this, fix i and apply [2, Theorem 2.5.1]. Then ζ_i may be represented as a convex combination of limiting gradients, thus:

$$\begin{aligned} \zeta_i = \sum_{k=1}^{n+1} \lambda_i^k \zeta_i^k, \ \text{where } \zeta_i^k &= \lim_{j \to \infty} \nabla_x g(t_i, y_{ij}^k), \\ & y_{ij}^k \to y_i \text{ as } j \to \infty \text{ (for each } i, k), \\ & \lambda_i^k \geq 0 \ (\forall i, k), \ \text{and} \ \sum_{k=1}^{n+1} \lambda_i^k = 1 \ (\forall i). \end{aligned} \tag{2.9}$$

Observe that since i is fixed, and $g(t_i, y_i) > 0$, we may use the Lipschitz continuity of $g(t_i, \cdot)$ to assume $g(t_i, y_{ij}^k) > 0$ for all j, k.

The representation (2.8) remains valid along any subsequence of the indices i. Thus we may use representation (2.9) to restrict our attention to a subsequence along which

- for each $k = 1, \dots, n+1$, the sequence $\{\lambda_i^k\}_{i=1}^\infty$ converges to some limiting value denoted by λ_∞^k; and
- for each $k = 1, \dots, n+1$, the sequence $\{\zeta_i^k\}_{i=1}^\infty$ converges to some limiting value denoted by ζ_∞^k.

Then for fixed i, k, we use representation (2.9) to generate an index $j = j(i, k)$ so large that

- $|y_{ij(i,k)}^k - y_i| < 1/i$ for all $k = 1, \dots, n+1$, for all i;
- $|\zeta_i^k - \nabla_x g(t_i, y_{ij(i,k)}^k)| < 1/i$ for all $k = 1, \dots, n+1$, for all i.

Then (2.8) and (2.9) give the representation

$$\zeta = \lim_{i \to \infty} \zeta_i = \sum_{k=1}^{n+1} \lambda_\infty^k \zeta_\infty^k,$$

where $\lambda_\infty^k \geq 0$, $\sum \lambda_\infty^k = 1$, and

$$\zeta_\infty^k = \lim_{i\to\infty} \nabla_x g(t_i, y_{ij(i,k)}^k)$$

for a sequence $\{(t_i, y_{ij(i,k)}^k)\}_{i=1}^\infty$ which converges to (t, x), while satisfying $g(t_i, y_{ij(i,k)}^k) > 0$ for all i. Thus $\zeta \in R(t, x)$, and the proof is complete. □

2.2. **Corollary.** *Let X satisfy* (H5), *and let $g(t, x) := \operatorname{dist}(X(t), x)$. Then g is (jointly) upper semicontinuous, and for every $(t, x) \in \operatorname{Gr} X \cap \Omega$, one has*

$$\partial_x^> g(t, x) = \operatorname{co}\{\lim \zeta_i : |\zeta_i| = 1,\ \zeta_i \in \pi(t_i, x_i),\ (t_i, x_i) \to (t, x) \text{ in } \operatorname{Gr} X\}. \tag{2.10}$$

Proof. The upper semicontinuity of g can be verified easily using its definition and the lower semicontinuity of X.

We will show that the right side of (2.10) equals the right side of (2.7). In fact, we will show that the sets of sequences whose convex hulls make up these two sets are identical. In each direction, we begin with an individual entry in such a sequence without its subscript. Let $(t, y) \in \Omega$ be a point where $g(t, y) > 0$ and $\nabla_x g(t, y)$ exists. According to [2, Propositions 2.5.4 and 2.5.5], the vector $\zeta = \nabla_x g(t, y)$ has unit length, and belongs to $\pi(t, x)$ for some point $x \in X(t)$ obeying $|x - y| = g(t, y)$.

Now let $(t, x) \in \operatorname{Gr} X \cap \Omega$ be fixed, and let ζ_i be any sequence as described on the right side of (2.7). Then, as we have shown, ζ_i is a sequence of unit vectors in $\pi(t_i, x_i)$, where $|x_i - y_i| = g(t_i, y_i)$. Now $(t_i, y_i) \to (t, x)$, so $\limsup g(t_i, y_i) \leq g(t, x) = 0$. Hence $(t_i, x_i) \to (t, x)$ also, and the sequence ζ_i is admissible on the right side of (2.10).

Conversely, if ζ is a unit vector in $\pi(t, x)$ for some $x \in X(t)$, then for all $\lambda > 0$ sufficiently small, the point $y = x + \lambda\zeta$ obeys both $g(t, y) = \lambda > 0$ and $\nabla_x g(t, y) = \zeta$. (The proof of this fact is an entertaining exercise in the geometry of Euclidean balls.) In particular, if ζ_i is any sequence as described on the right side of (2.10), then there exists a sequence λ_i decreasing to zero such that the choices $y_i = x_i + \lambda_i \zeta_i$ make the sequence ζ_i admissible on the right side of (2.7). □

2.3. **Proposition.** *Assume* (H1)–(H5). *Let $g(t, x) := \operatorname{dist}(X(t), x)$. Then for any $(t, x) \in \operatorname{Gr} X \cap \Omega$, one has*

(a) $N(t, x) = \operatorname{cl}[\{0\} \cup \bigcup_{r>0} r\,\partial_x^> g(t, x)]$;

(b) $0 \notin \partial_x^> g(t, x)$ *if and only if $N(t, x)$ is pointed, in which case one has*

$$N(t, x) = \{0\} \cup \bigcup_{r>0} r\,\partial_x^> g(t, x).$$

(c) *If the graph of the multifunction $(t', x') \rightrightarrows N_{X(t')}(x')$ is closed, then $N_{X(t)}(x) = N(t, x)$.*

(d) *If the multifunction* X *is convex-valued, then* $N(t, x) = N_{X(t)}(x)$, *and this cone is pointed if and only if* $\operatorname{int} X(t) \neq \varnothing$.

Proof. (a) Corollary 2.2 shows that $\partial_x^> g(t, x) \subseteq \operatorname{co} \overline{\pi}(t, x)$, from which the forward inclusion ($\supseteq$) follows immediately.

To prove the reverse inclusion, let ζ_i be any convergent sequence satisfying $\zeta_i \in \pi(t_i, x_i)$ for a sequence $(t_i, x_i) \to (t, x)$ in $\operatorname{Gr} X$. Let $\zeta = \lim \zeta_i$. If $\zeta = 0$, then certainly ζ belongs to the right-hand side. Otherwise, one has $|\zeta_i| \to |\zeta| > 0$, so $\zeta_i/|\zeta_i| \to \zeta/|\zeta|$. By Corollary 2.2, $\zeta/|\zeta| \in \partial_x^> g(t, x)$, so ζ belongs to the right-hand side of (a) in this case also. Thus the right-hand side contains $\overline{\pi}(t, x)$; since the right side is closed and convex, it therefore contains $N(t, x) = \operatorname{cl} \operatorname{co} \overline{\pi}(t, x)$ also.

(b) By Corollary 2.2, $\partial_x^> g(t, x)$ is the convex hull of a set of unit vectors. Hence each of its extreme points is a unit vector. If $0 \in \partial_x^> g(t, x)$ it follows that 0 is not an extreme point, so it must be the midpoint of some line segment in $\partial_x^> g(t, x)$. The line generated by this segment is a subspace of $N(t, x)$ by part (a), so $N(t, x)$ is not pointed.

Conversely, suppose $0 \notin \partial_x^> g(t, x)$. Then there exist a vector $\mu \neq 0$ and a constant $\varepsilon > 0$ for which one has

$$\langle \mu , \zeta \rangle \geq \varepsilon > 0 \quad \forall \zeta \in \partial_x^> g(t, x).$$

Since $\partial_x^> g(t, x)$ is a subset of $\overline{B}$, it follows that

$$\left\langle \mu , \frac{\zeta}{|\zeta|} \right\rangle \geq \varepsilon > 0 \quad \forall \zeta \in \partial_x^> g(t, x).$$

Now the set of all $\zeta \in \mathbb{R}^n$ satisfying the inequality above is a closed and pointed convex cone. This cone contains $N(t, x)$ by (a), so $N(t, x)$ must also be pointed.

Finally, suppose $N(t, x)$ is pointed. Since $N(t, x) \supseteq \overline{\pi}(t, x)$, it follows that $\overline{\pi}(t, x)$ is also pointed, while $\overline{\pi}(t, x)$ is obviously closed. According to [15, p. 57], $\operatorname{co} \overline{\pi}(t, x)$ is then closed and pointed. Thus $N(t, x) = \operatorname{co} \overline{\pi}(t, x)$: the identity in part (b) follows from (2.10).

(c) Let $(t', x') \rightrightarrows \overline{N}_{X(t')}(x')$ denote the graphical closure of the multifunction $(t', x') \rightrightarrows N_{X(t')}(x')$. As we have already observed, the following inclusions always hold:

$$\overline{N}_{X(t)}(x) \supseteq N(t, x) \supseteq N_{X(t)}(x).$$

Whenever the leftmost set in this statement equals the rightmost set, equality prevails throughout.

(d) In view of (c), it suffices to show that the multifunction $(t', x') \rightrightarrows N_{X(t')}(x')$ has closed graph. To do so, suppose $\zeta_i \in N_{X(t_i)}(x_i)$ for some convergent sequence ζ_i, where one has $(t_i, x_i) \to (t, x)$ in $\operatorname{Gr} X \cap \Omega$. Let $\zeta = \lim \zeta_i$. By definition of the convex normal cone,

$$\langle \zeta_i , y - x_i \rangle \leq 0 \quad \forall y \in X(t_i).$$

Now let any point $y \in X(t)$ be given. Since the multifunction $X(\cdot)$ is lower semicontinuous, there exists a sequence $y_i \to y$ satisfying $y_i \in X(t_i)$ $\forall i$. Using this sequence in the previous inequality, we have

$$\langle \zeta_i , y_i - x_i \rangle \leq 0 \quad \forall i.$$

In the limit as $i \to \infty$, this inequality yields $\langle \zeta , y - x \rangle \leq 0$: since $y \in X(t)$ is arbitrary, it follows that $\zeta \in N_{X(t)}(x)$, as required. □

Having characterized the multifunction $(t, x) \rightrightarrows N(t, x)$ in various ways, let us now observe that it is (jointly) Borel measurable. This is true because $N(t, x) = \operatorname{cl}\operatorname{co}\overline{\pi}(t, x)$: the multifunction $\overline{\pi}$ is Borel measurable because it has closed graph, and hence $\operatorname{cl}\operatorname{co}\overline{\pi}$ is Borel measurable by [16, Proposition 1H]. Likewise, for any continuous function $x: [a, b] \to \Omega$ satisfying $x(t) \in X(t)$ for all t, the multifunction $N(t) := N(t, x(t))$ is Borel measurable. To see this, note that $\overline{\pi}(t) := \overline{\pi}(t, x(t))$ is the composition of Borel measurable multifunctions, hence Borel measurable, and apply [16, Proposition 1H] again. Given a Borel measurable multifunction $N: [a, b] \rightrightarrows \mathbb{R}^n$ whose values are closed convex cones, and a measure μ on $[a, b]$ taking values in $\mathbb{R}^n$, the statement that *μ is $N(t)$-valued* means that there is a nonnegative measure μ_0 on $[a, b]$ and a measurable selection $\nu(t) \in N(t)$ μ_0-a.e. such that $d\mu(t) \equiv \nu(t)\, d\mu_0(t)$. (See [17, § 5].)

The Hamiltonian. The Hamiltonian associated with the multifunction F is defined by

$$H(t, x, p) := \sup\{\langle p, v \rangle : v \in F(t, x)\}.$$

(Compare the definition immediately preceding Theorem 1.1.) Its basic properties are described in [2, Proposition 3.2.4]; the following is an extension of part (e) of that result.

2.4. **Lemma.** *Under* (H1)–(H5),

$$\partial H(t, x, 0) = \{0\} \times F(t, x).$$

Proof. We argue for fixed $t \in [a, b]$, which will be suppressed throughout the notation below. The identity asserted by the lemma involves compact convex sets, which are equal if and only if their support functions are identical. It therefore suffices to prove that

$$H^0(x, 0; v, w) = H(x, w) \qquad \forall (v, w) \in \mathbb{R}^n \times \mathbb{R}^n. \tag{2.11}$$

To prove the inequality $\geq$ in (2.11), consider the definition of the left-hand side:

$$H^0(x, 0; v, w) = \limsup_{\substack{(y, q) \to (x, 0) \\ t \to 0^+}} \frac{H((y, q) + t(v, w)) - H(y, q)}{t}.$$

The choices $y = x - tv$ and $q = 0$ can only reduce the value of this expression, whence

$$H^0(x, 0; v, w) \geq \limsup_{t \to 0^+} \frac{H(x, tw)}{t} = H(x, w).$$

The reverse inequality ($\leq$) is more challenging. Choose sequences $(y_i, q_i) \to (x, 0)$ and $t_i \to 0^+$ such that

$$H^0(x, 0; v, w) = \lim_{i\to\infty} \frac{H((y_i, q_i) + t_i(v, w)) - H(y_i, q_i)}{t_i}.$$

Apply the Mean Value Theorem [2, Theorem 2.3.7] to obtain the following estimate, for some $\lambda_i \in [0, 1]$:

$$\begin{aligned} &H(y_i + t_i v, q_i + t_i w) - H(y_i, q_i) \\ &\quad = H(y_i, q_i + t_i w) - H(y_i, q_i) + H(y_i + t_i v, q_i + t_i w) - H(y_i, q_i + t_i w) \\ &\quad \in H(y_i, q_i + t_i w) - H(y_i, q_i) + t_i \partial_x H(y_i + \lambda_i t_i v, q_i + t_i w) v. \end{aligned}$$

Now $\partial_x H(x, p) \subseteq |p| k_F \overline{B}$ in general notation [2, Proposition 3.2.4(b)], so the previous inclusion implies

$$\begin{aligned} &\frac{H(y_i + t_i v, q_i + t_i w) - H(y_i, q_i)}{t_i} \\ &\qquad \leq \frac{H(y_i, q_i + t_i w) - H(y_i, q_i)}{t_i} + k_F |q_i + t_i w| |v|. \end{aligned}$$

Since the rightmost term converges to 0 as $i \to \infty$, it suffices to prove

$$\limsup_{i\to\infty} \frac{H(y_i, q_i + t_i w) - H(y_i, q_i)}{t_i} \leq H(x, w). \tag{2.12}$$

To do so, choose a subsequence of indices (without relabelling) along which the left side of (2.12) actually converges to its upper limit, and then apply the Mean Value Theorem again. There results a sequence

$$\zeta_i \in \partial_p H(y_i, q_i + \lambda_i t_i w), \qquad \lambda_i \in [0, 1],$$

for which the difference quotient on the left side of (2.12) equals $\langle \zeta_i, w \rangle$. Now $\zeta_i \in F(y_i)$ for each i [2, Proposition 3.2.4(d)], so the sequence $\{\zeta_i\}$ is contained in the compact set $k_F \overline{B}$ and admits a convergent subsequence. Along this subsequence, ζ_i converges to a point ζ in $F(x)$ because F has closed graph. In summary,

$$\text{LHS}(2.12) = \lim_{i\to\infty} \langle \zeta_i, w \rangle = \langle \zeta, w \rangle$$

for some $\zeta \in F(x)$. This implies (2.12) and completes the proof. □

The Hamiltonian inclusion. We are now ready to review the standard Hamiltonian necessary conditions for (P_D). Our first result is a refinement of [2, Theorem 3.5.2]. It pertains to the localized version of (P_D) in which the requirement that $x(t) \in \Omega_t$ for all $t \in [a, b]$ is added to the constraints.

2.5. **Theorem.** *Assume* (H1)–(H5). *Let the arc* $x(\cdot)$ *solve* (P_D), *and assume the constraint qualification*

(CQ) $\qquad N(t, x(t))$ *is pointed for each* $t \in [a, b]$.

Define $g(t, x) := \operatorname{dist}(X(t), x)$. *Then there exist a scalar* $\lambda \in \{0, 1\}$, *an arc* $q \in \mathrm{AC}([a, b]; \mathbb{R}^n)$, *and a finite nonnegative Borel measure* μ *on* $[a, b]$, *together with a measurable selection* $\gamma(t) \in \partial_x^> g(t, x(t))$ *for all* $t \in [a, b]$, *such that*

(a) $(-\dot{q}(t), \dot{x}(t)) \in \partial H(t, x(t), q(t) + \int_{[a,t)} \gamma(s)\, d\mu(s))$ *a.e.* $t \in [a, b]$;
(b) $(q(a), -q(b) - \int_{[a,b]} \gamma(s)\, d\mu(s)) \in \lambda \partial l(x(a), x(b)) + N_S(x(a), x(b))$;
(c) $\operatorname{Supp}(\mu) \subseteq \{t : N(t, x(t)) \neq \{0\}\} = \{t : (t, x(t)) \in \operatorname{bdry} \operatorname{Gr} X\}$;
(d) $\lambda + \|q(t) + \int_{[a,t)} \gamma(s)\, d\mu(s) + \gamma(b)\mu\{b\}\delta_{\{b\}}(t)\|_\infty > 0$.

Proof. The quantities λ, q, μ, and γ are all provided by [2, Theorem 3.5.2]. That result gives the conclusions (a) and (b) recorded above, and asserts the following in place of (c) and (d):

$$\operatorname{Supp}(\mu) \subseteq \{t : \partial_x^> g(t, x(t)) \neq \varnothing\}, \tag{2.13}$$

$$\lambda + \|q\|_\infty + \mu[a, b] > 0. \tag{2.14}$$

To recover (c), we prove that the set on the right side of (2.13) is the same as the set on the right side of (c). Indeed, Proposition 2.3(a) implies that

$$\partial_x^> g(t, x) \neq \varnothing \Leftrightarrow N(t, x) \neq \{0\},$$

so it suffices to show that

$$N(t, x) \neq \{0\} \Leftrightarrow (t, x) \in \operatorname{bdry} \operatorname{Gr} X. \tag{2.15}$$

To do so, note first that a point (t, x) lies in $\operatorname{int} \operatorname{Gr} X$ if and only if any sequence (t_i, x_i) converging to (t, x) in $\operatorname{Gr} X$ obeys $x_i \in \operatorname{int} X(t_i)$ for all i sufficiently large. So if $(t, x) \in \operatorname{int} \operatorname{Gr} X$, then $\overline{\pi}(t, x) = \{0\}$ and it follows that $N(t, x) = \{0\}$. Conversely, if there exists a sequence (t_i, x_i) in $\operatorname{Gr} X$ for which $x_i \in \operatorname{bdry} X(t_i)$ for all i while $(t_i, x_i) \to (t, x)$, then $\overline{\pi}(t_i, x_i) \neq \{0\}$ for all i (cf. [2, Theorem 2.5.6, Corollary 1]), so $\overline{\pi}(t, x) \neq \{0\}$. Hence $N(t, x) = \{0\}$ implies $(t, x) \in \operatorname{int} \operatorname{Gr} X$. We have shown

$$N(t, x) = \{0\} \Leftrightarrow (t, x) \in \operatorname{int} \operatorname{Gr} X,$$

which is clearly equivalent to (2.15).

Now our conclusion (d) is strictly stronger than Clarke's nontriviality condition (2.14), and hence requires an assumption which is absent from [2, Theorem 3.5.2]. This is the constraint qualification (CQ), which we put to work as follows. If $\lambda = 1$ then (d) obviously holds, so suppose $\lambda = 0$. If (d) is false, i.e.,

$$\left\| q(t) + \int_{[a,t)} \gamma(s)\, d\mu(s) + \gamma(b)\mu\{b\}\delta_{\{b\}}(t) \right\|_\infty = 0, \tag{2.16}$$

then we must have $q(t) + \int_{[a,t)} \gamma(s)\, d\mu(s) + \gamma(b)\mu\{b\}\delta_{\{b\}}(t) = 0$ for all $t \in [a, b]$. Applying Lemma 2.4 to the Hamiltonian inclusion (a), we deduce that $\dot{q}(t) = 0$ a.e., so $q(t) = q(a) = 0$ for all $t \in [a, b]$. Consequently

$$\int_{[a,t)} \gamma(s)\, d\mu(s) + \gamma(b)\mu\{b\}\delta_{\{b\}}(t) = 0 \quad \forall t \in [a, b].$$

Subtraction yields the two identities

$$\gamma(b)\mu\{b\} = 0, \qquad \int_{[\alpha,\beta)} \gamma(s)\,d\mu(s) = 0 \quad \forall[\alpha,\beta] \subseteq [a,b]. \tag{2.17}$$

We will use (CQ) to deduce that μ is the zero measure: This will contradict (2.14), thereby vindicating condition (d) and completing the proof.

Let $C = \operatorname{Supp}(\mu)$, and fix $t \in C$. By (c), $N(t, x(t)) \neq \{0\}$, while $N(t, x(t))$ is pointed by (CQ). Hence the compact convex set $G(t) := \partial_x^> g(t, x(t))$ does not contain 0 by Proposition 2.3(b), while $G(t) \neq \varnothing$ by (2.13). Therefore there exist a unit vector $p \in \mathbb{R}^n$ and a constant $\varepsilon > 0$ such that $\langle p, \gamma\rangle \geq 2\varepsilon > 0$ for all $\gamma \in G(t)$. Now the multifunction G defined here has closed graph, so there exists $\delta > 0$ so small that

$$t' \in C, \qquad |t' - t| < \delta \Rightarrow \varnothing \neq G(t') \subseteq G(t) + \varepsilon\overline{B}.$$

In particular, for the given selection $\gamma(t') \in G(t')$ μ-a.e. $t' \in C \cap [t-\delta, t+\delta]$, one has

$$\langle p\gamma(t')\rangle \geq \varepsilon > 0 \quad \mu\text{-a.e. } t' \in C \cap [t-\delta, t+\delta].$$

Together with (c) and (2.17), this implies

$$0 = \langle p, 0\rangle = \left\langle p, \int_{C\cap[t-\delta,t+\delta)} \gamma(s)\,d\mu(s)\right\rangle \geq \int_{C\cap[t-\delta,t+\delta)} \varepsilon\,d\mu(s).$$

Consequently $\mu(C \cap [t-\delta, t+\delta)) = 0$. This implies $t \notin \operatorname{Supp}(\mu) = C$. Since $t \in C$ is arbitrary, it follows that $C = \varnothing$, i.e., that μ is the zero measure. □

On the constraint qualification. In [2, p. 122], Clarke notes that whenever the constraint qualification (CQ) (or its cognate in his notation, namely $0 \notin \partial_x^> g(t, x(t))$ for all t; see Proposition 2.3(b)) is violated, then all the conclusions of [2, Theorem 3.5.2] may hold without providing any useful information. Indeed, if one has $0 \in \partial_x^> g(t, x(t))$ for some $t \in [a, b]$, then the choices $\mu = \delta_{\{t\}}$, $\gamma \equiv 0$, $p \equiv 0$, and $\lambda = 0$ satisfy conclusions (a)–(c) with $\|\mu\| = 1$, but give no insight into the behavior of $x(\cdot)$. Thus condition (CQ) is *necessary* for the conclusions of [2, Theorem 3.5.2] to be useful. The point of Theorem 2.5 is that (CQ) is also *sufficient* for this purpose, since it gives the improved nontriviality condition (d). (Note that (d) disallows the unenlightening choices above, since they produce a function $q(t) + \int_{[a,t)} \gamma(s)\,d\mu(s) + \gamma(b)\mu\{b\}\delta_{\{b\}}(t)$ which is identically zero.)

Having identified (CQ) as a necessary and sufficient condition for the conclusions of Theorem 2.5 to be useful, we now observe that this condition also has a desirable technical consequence: namely, that the set of measures which are $N(t, x(t))$-valued is weak*-closed. This is the assertion of Proposition 2.7 below, which we will apply in §3. To prove it, we will need the following lemma.

2.6. **Lemma.** *Let a multifunction $\gamma: [a, b] \rightrightarrows \mathbb{R}^n$ be given. If the values of γ are pointed cones and γ has closed graph, then the same properties are enjoyed by the multifunction $\Gamma(t) := \operatorname{co}\gamma(t)$.*

Proof. For each $t \in [a, b]$, the cone $\gamma(t)$ is closed and pointed. It follows from the arguments of [15, p. 57], that each cone $\Gamma(t) = \operatorname{co}\gamma(t)$ is likewise closed and pointed.

Now fix $t \in (a, b)$. (The cases $t = a$ and $t = b$ are treated similarly.) Since $\Gamma(t)$ is closed and pointed, its polar cone $\Gamma(t)^0$ has nonempty interior. Let $-\overline{v}$ be a unit vector in $\operatorname{int}\Gamma(t)^0$, and let $\varepsilon > 0$ be chosen so that

$$-\overline{v} + 2\varepsilon B \subseteq \Gamma(t)^0.$$

Observe that this implies

$$\langle w, \overline{v}\rangle \geq 2\varepsilon|w| \qquad \forall w \in \Gamma(t).$$

Consider the cone $K_\varepsilon(\overline{v}) = \{v\colon \langle v, \overline{v}\rangle \geq \varepsilon|v|\}$. Since γ has closed graph, there exists $\delta > 0$ such that

$$\gamma(t') \subseteq K_\varepsilon(\overline{v}) \qquad \forall t' \in (t-\delta, t+\delta).$$

(Suppose not: then there exist sequences $t_i \to t$ and $v_i \in \gamma(t_i)\backslash K_\varepsilon(\overline{v})$. Let $\widehat{v}_i = v_i/|v_i|$ and limit attention to a subsequence along which $\widehat{v}_i \to \widehat{v}$. Since γ has closed graph, $\widehat{v} \in \gamma(t) \subseteq \Gamma(t)$. But $\langle \widehat{v}, \overline{v}\rangle = \lim\langle \widehat{v}_i, \overline{v}\rangle \leq \varepsilon = \varepsilon|\widehat{v}|$, a contradiction.) In particular,

$$\langle w, \overline{v}\rangle \geq \varepsilon|w| \quad \forall t' \in (t-\delta, t+\delta), \ \forall w \in \Gamma(t'). \tag{2.18}$$

Now we show that $\operatorname{Gr}\Gamma$ is closed at t. Let any sequence $t_i \to t$ be given, together with a convergent sequence $v_i \in \Gamma(t_i)$: write $v = \lim v_i$. We must show that $v \in \Gamma(t)$, an exercise which is trivial unless $v \neq 0$. For each i, Carathéodory's theorem allows us to write

$$v_i = \sum_{k=1}^{n+1} v_k^i, \quad \text{where } v_k^i \in \gamma(t_i) \text{ for } k = 1, \dots, n+1.$$

Since the sequence v_i converges, it is bounded. That is, for some $R > 0$ we have $R \geq |v_i|$ for all i. We may therefore use (2.18) to estimate (for each i)

$$\begin{aligned} R \geq |v_i| &= \left|\sum_{k=1}^{n+1} v_k^i\right| = \left|\sum_{k=1}^{n+1} v_k^i\right||\overline{v}| \\ &\geq \left\langle \sum_{k=1}^{n+1} v_k^i, \overline{v}\right\rangle \geq \varepsilon \sum_{k=1}^{n+1} |v_k^i| \geq \varepsilon \max_k |v_k^i|. \end{aligned}$$

In particular, each of the $n+1$ sequences $\{v_k^i\}_{i=1}^\infty$ is bounded by R/ε and the successive extraction of subsequences allows us to assume that $v_k^i \to v_k^\infty$ for each $k = 1, \dots, n+1$. Since γ has closed graph, $v_k^\infty \in \gamma(t)$ for each k. Hence the vector

$$v = \lim_{i\to\infty} v_i = \lim_{i\to\infty} \sum_{k=1}^{n+1} v_k^i = \sum_{k=1}^{n+1} v_k^\infty$$

belongs to $\Gamma(t)$, as required. □

2.7. **Proposition.** *Let* $\Gamma\colon [a, b] \rightrightarrows \mathbb{R}^n$ *be a multifunction whose values are closed pointed convex cones, and suppose* $\operatorname{Gr}\Gamma$ *is closed. Then the set of measures* $G = \{\mu : \mu$ *is* $\Gamma(t)$*-valued*$\}$ *is weak***-closed.*

Proof. Let $K = \{u \in C([a, b]; \mathbb{R}^n)\colon u(t) \in \Gamma(t)^0 \ \forall t \in [a, b]\}$. It suffices to prove that G equals the convex cone

$$K^0 = \bigcap_{u \in K} \{\mu\colon \langle \mu , u\rangle \le 0\},$$

since the latter is obviously weak*-closed. To do so, note first that the multifunction $t \rightrightarrows \Gamma(t)^0$ is lower semicontinuous, by Lemma 2.6 and [1, Proposition 18, p. 177]. According to [17, Theorem 6] (with $Q(t) = \Gamma(t)^0$), the set K introduced above is nonempty, closed, and convex in $C[a, b]$, and its support function on the space of measures is given by

$$\Psi_K^*(\mu) = \int_a^b \Psi_{\Gamma(t)^0}^* \left(\frac{d\mu}{d\theta}(t)\right) d\theta(t) \quad \forall \mu,$$

where θ is any nonnegative measure such that $\mu \ll \theta$. But $\Psi_{\Gamma(t)^0}^* \equiv \Psi_{\Gamma(t)}$, so this representation may be rewritten as

$$\Psi_K^*(\mu) = \int_a^b \Psi_{\Gamma(t)} \left(\frac{d\mu}{d\theta}(t)\right) d\theta(t) \quad \forall \mu.$$

Consequently a measure μ is $\Gamma(t)$-valued if and only if $\Psi_K^*(\mu) = 0$. To complete the proof, note that $K^0 = N_K(0) = \partial\Psi_K(0)$, while Young's identity gives

$$\begin{aligned} \mu \in \partial\Psi_K(0) &\Leftrightarrow \Psi_K^*(\mu) + \Psi_K(0) = \langle 0 , \mu\rangle \\ &\Leftrightarrow \Psi_K^*(\mu) = 0. \quad \square \end{aligned}$$

Now the formulation of Hamiltonian necessary conditions in terms of an adjoint function of bounded variation represents a simple change of variables in Theorem 2.5. The recipe for this change appears in the proof of the next result, in which the improved nontriviality condition (d) of Theorem 2.5 is indispensable in establishing the crucial phrase "not both zero".

2.8. **Theorem.** *Assume* (H1)–(H5), *and suppose an arc* $x(\cdot)$ *solves* (P_D) *and satisfies the constraint qualification*

(CQ) $\qquad N(t, x(t))$ *is pointed for each* $t \in [a, b]$.

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

(a) $(-\dot{p}(t) + \nu(t), \dot{x}(t)) \in \partial H(t, x(t), p(t))$ *a.e.* $t \in [a, b]$;
(b) $(p(a), -p(b)) \in \lambda\partial l(x(a), x(b)) + N_S(x(a), x(b))$;
(c) *the singular part of* dp *is* $N(t, x(t))$*-valued, and in particular is supported on the set*

$$\{t\colon N(t, x(t)) \neq \{0\}\} = \{t\colon (t, x(t)) \in \operatorname{bdry}\operatorname{Gr} X\}.$$

Proof. Apply Theorem 2.5, then adjust γ on a set of μ-measure zero to arrange

$$(t, x(t)) \in \operatorname{bdry} \operatorname{Gr} X \Rightarrow \gamma(t) \in \partial_x^{>} g(t, x(t)).$$

Then introduce a function $p \in \mathrm{BV}([a, b]; \mathbb{R}^n)$ by defining

$$p(t) = q(t) + \int_{[a,t)} \gamma(s)\, \mu(ds) + \gamma(b)\mu\{b\}\delta_{\{b\}}(t).$$

Note that $p(a) = q(a)$ since $[a, a) = \varnothing$, while $p(b) = q(b) + \int_{[a,b]} \gamma(s)\, d\mu(s)$. Hence (b) follows from Theorem 2.5(b).

Next, decompose $d\mu(t) = m(t)\, dt + d\mu_s(t)$ for some nonnegative function $m \in L^1[a, b]$ and some nonnegative Borel measure μ_s totally singular with respect to Lebesgue measure. Clearly $m(t) = 0$ for all t where $(t, x(t)) \in \operatorname{int} \operatorname{Gr} X$, so the definition

$$\nu(t) := \begin{cases} \gamma(t)m(t), & \text{if } (t, x(t)) \in \operatorname{bdry} \operatorname{Gr} X, \\ 0, & \text{otherwise,} \end{cases}$$

gives both $\nu(t) \in N(t, x(t))$ for all $t \in [a, b]$ (by Proposition 2.3(b) and Theorem 2.5(c)) and $\dot{p}(t) = \dot{q}(t) + \gamma(t)m(t)$ a.e. Hence (a) holds.

Finally, the singular part of dp is the vector measure $\gamma(t)\, d\mu_s(t)$: it is $N(t, x(t))$-valued by definition. □

3. Proof of Theorem 1.1

We now pass from the case $L \equiv 0$ of the differential inclusion problem to the general case of problem (P). This requires a precise description of our hypotheses concerning the integrand L. Thus we augment hypotheses (H1)–(H5) of §2 with the following hypotheses:

(H6) There exists $\varepsilon > 0$ such that the Lagrangian $L: \operatorname{Gr}(F + \varepsilon B) \to \mathbb{R}$ is $\mathscr{L} \times \mathscr{B}$-measurable; for each fixed $(t, x) \in \Omega$, the function $v \to L(t, x, v)$ is convex on the set $F(t, x) + \varepsilon B$.

(H7) There exists a nonnegative function $k_L(\cdot) \in L^1[a, b]$ such that

$$|L(t, x', v') - L(t, x, v)| \le k_L(t)|(x', v') - (x, v)| \qquad \forall t \in [a, b],\ (x', v'), (x, v) \in \operatorname{Gr}(F(t, \cdot) + \varepsilon B);$$

also, the product $k_L(t)k_F(t)$ is integrable.

A function of fundamental importance is the Hamiltonian introduced in §1: for each (t, x) in Ω and p in $\mathbb{R}^n$, we define

$$H_\lambda(t, x, p) := \sup\{\langle p, v\rangle - \lambda L(t, x, v) : v \in F(t, x)\}.$$

As we now show, our standing hypotheses on F and L imply that H_1 obeys the "strong Lipschitz condition".

3.1. **Lemma.** *Assume* (H1)–(H7). *Then there is a function* $k_H(\cdot) \in L^1[a, b]$ *for which*

$$|H_1(t, y, p) - H_1(t, x, p)| \le \tfrac{1}{2} k_H(t)(1 + |p|)|y - x| \qquad \forall t \in [a, b],\ x, y \in \Omega_t,\ p \in \mathbb{R}^n.$$

Moreover, for each $t \in [a, b]$, *the multifunction* E *defined by*

$$E(t, x) := \{(v, w) : v \in F(t, x),\ w \ge L(t, x, v)\}$$

is Lipschitz of rank $k_H(t)$ *on* Ω_t.

Proof. Taking $\lambda = 1$ in Lemma 4.1(c) below, we find that

$$|H_1(t, y, p) - H_1(t, x, p)| \le (k_F(t)|p| + k_L(t) + k_F(t)k_L(t))|y - x|.$$

Consequently (3.1) will hold for the function $k_H = 2\max\{k_F, k_L(1 + k_F)\}$, which is integrable by (H7).

The desired Lipschitz continuity of $E(t, \cdot)$ is a consequence of the strong Lipschitz condition (3.1): this is proven as the first step of [2, Lemma 2, pp. 170–171]. □

An auxiliary problem. Now suppose an arc $\overline{x}(\cdot)$ solves problem (P). Then $\overline{x}$ remains admissible, and hence optimal, if the set Ω is assumed to obey $\Omega_t = \overline{x}(t) + \delta B$ for each $t \in [a, b]$, where $\delta \in (0, 1/6)$ is fixed. Similarly, $\overline{x}(\cdot)$ must also solve any problem of the following form, in which the running cost is augmented by a nonnegative function M which satisfies $M(t, \dot{\overline{x}}(t)) = 0$ a.e. $t \in [a, b]$:

$$\begin{aligned} &\text{minimize} && l(x(a), x(b)) + \int_a^b [L(t, x(t), \dot{x}(t)) + M(t, \dot{x}(t))]\, dt \\ &\text{subject to} && x \in \mathrm{AC}([a, b]; \mathbb{R}^n), \\ &&& \dot{x}(t) \in F(t, x(t)) \quad \text{a.e. } t \in [a, b], \\ &&& x(t) \in X(t) \quad \forall t \in [a, b], \\ &&& (x(a), x(b)) \in S. \end{aligned} \tag{3.2}$$

For our purposes the choice $M(t, v) := \theta(v - \dot{\overline{x}}(t))$ is convenient, where

$$\theta(v) := [1 + |v|^2]^{1/2} - 1.$$

Since $\theta(0) = 0$ and $\nabla\theta(0) = 0$, the addition of M to the running cost makes a change which is undetectable to first order along the arc $\overline{x}$, while rendering the new Lagrangian $\widetilde{L} := L + M$ a strictly convex function of its third argument. This property is crucial to the simultaneous derivation of the Eulerian and Hamiltonian inclusions.

Now if we use the solution $\overline{x}$ of (3.2) to define $\overline{y}(t) = \int_a^t L(r, \overline{x}(r), \dot{\overline{x}}(r))\, dr$, then the arc $(\overline{x}, \overline{y})$ evidently solves the following problem, in which no integral

term appears:

$$(\mathscr{P})\qquad \begin{array}{ll} \text{minimize} & l(x(a), x(b)) + y(b) \\ \text{subject to} & x \in \mathrm{AC}([a, b]; \mathbb{R}^n), \\ & (\dot{x}(t), \dot{y}(t)) \in \widetilde{E}(t, x(t)) \quad \text{a.e. } t \in [a, b], \\ & (x(t), y(t)) \in X(t) \times \mathbb{R} \quad \forall t \in [a, b], \\ & (x(a), y(a); x(b), y(b)) \in \mathscr{S}. \end{array}$$

Here the dynamics are governed by the multifunction

$$\widetilde{E}(t, x) := \{(v, w) : v \in F(t, x),\ \widetilde{L}(t, x, v) \le w\},$$

and the endpoint constraint set is

$$\mathscr{S} := \{(\xi_0, \eta_0; \xi_1, \eta_1) : (\xi_0, \xi_1) \in S,\ \eta_0 = 0\}.$$

Roughly speaking, the results announced by Theorem 1.1 follow from an application of Theorem 2.8 to problem $(\mathscr{P})$, followed by a careful reduction of the Hamiltonian inclusion involving $\widetilde{E}$ to the desired conclusions involving F and L. Let us outline the first step in this program before confronting the fact that $(\mathscr{P})$ fails to satisfy (H1)–(H5). If we could apply Theorem 2.8 directly to $(\mathscr{P})$, then we would find a scalar $\lambda \in \{0, 1\}$ and a function $(p, q) \in \mathrm{BV}([a, b]; \mathbb{R}^n \times \mathbb{R})$, not both zero, together with an integrable selection $(\nu(t), \rho(t)) \in \mathscr{N}(t, \overline{x}(t), \overline{y}(t))$ $\forall t \in [a, b]$, such that

$$\text{(a)}\qquad \begin{aligned} &(-\dot{p}(t) + \nu(t), -\dot{q}(t) + \rho(t), \dot{\overline{x}}(t), L(t, \overline{x}(t), \dot{\overline{x}}(t))) \\ &\quad \in \partial\mathscr{H}(t, \overline{x}(t), \overline{y}(t), p(t), q(t)); \end{aligned}$$

$$\text{(b)}\qquad \begin{aligned} &(p(a), q(a); -p(b), -q(b)) \\ &\quad \in \lambda\{(\alpha_0, 0; \alpha_1, 1) : (\alpha_0, \alpha_1) \in \partial l(\overline{x}(a), \overline{x}(b))\} \\ &\qquad + N_{\mathscr{S}}(\overline{x}(a), 0; \overline{x}(b), \overline{y}(b)); \end{aligned}$$

(c) the singular part of (dp, dq) is $\mathscr{N}(t, \overline{x}(t), \overline{y}(t))$-valued.

In these conditions the Hamiltonian is

$$\mathscr{H}(t, x, y, p, q) := \sup\{\langle p, v\rangle + qw : (v, w) \in \widetilde{E}(t, x)\}.$$

Since it is independent of y, the y-component of its generalized gradient is always zero, and the remaining information is contained in the generalized gradient of the function

$$\widetilde{H}(t, x, p, q) := \mathscr{H}(t, x, y, p, q).$$

The tilde notation in $\widetilde{H}$ is inspired by the tilde on $\widetilde{L} = L + M$ via

$$\widetilde{E}(t, x) := \operatorname{epi}[\widetilde{L}(t, x, \cdot) + \Psi_{F(t,x)}(\cdot)].$$

In conditions (a)–(c), the endpoint constraint set $\mathscr{S} = \{(\xi_0, \eta_0; \xi_1, \eta_1) : (\xi_0, \xi_1) \in S,\ \eta_0 = 0\}$ has such a simple structure that its normal cone is closely related to that for the original set S: thus (b) implies that $-q(b) = \lambda$

and that the usual transversality condition ((b′) below) holds for p. Similarly, the cone $\mathscr{N}(t, x, y)$ is constructed from the state constraint multifunction $\mathscr{X}(t) = X(t) \times \mathbb{R}$ just as in §2: evidently $\mathscr{N}(t, x, y) = N(t, x) \times \{0\}$. Thus (c) implies that the singular part of dq is zero, while (a) implies that $\dot{q} = 0$ a.e. Hence $q \equiv -\lambda$ follows from (b), and we are left with

(a′) $(-\dot{p}(t) + \nu(t), \dot{\overline{x}}(t), L(t, \overline{x}(t), \dot{\overline{x}}(t))) \in \partial \widetilde{H}(t, \overline{x}(t), p(t), -\lambda)$ a.e.;
(b′) $(p(a), -p(b)) \in \lambda \partial l(\overline{x}(a), \overline{x}(b)) + N_S(\overline{x}(a), \overline{x}(b))$;
(c′) the singular part of dp is $N(t, \overline{x}(t))$-valued.

Conclusions (b′) and (c′) are just what we hope to get, while the extended Hamiltonian inclusion (a′) contains the germs of the ordinary Hamiltonian inclusion, the Euler inclusion, and the Weierstrass-Pontryagin condition.

A technical necessity. Now as we mentioned above, the data of problem $(\mathscr{P})$ fail to satisfy the hypotheses of Theorem 2.8, so the arguments of the previous paragraph are not completely rigorous. The main obstacle is that the multifunction $\widetilde{E}$ is unbounded, and hence fails to satisfy both the compactness condition and the integrable bound required by (H3). We surmount this difficulty by considering, instead of problem $(\mathscr{P})$, a sequence of problems $(\mathscr{P}_m)$ defined as follows: For each $m \in \mathbb{N}$, problem $(\mathscr{P}_m)$ is identical to $(\mathscr{P})$, except that the multifunction $\widetilde{E}$ is replaced by

$$\widetilde{E}_m(t, x) := \widetilde{E}(t, x) \cap \left[\left(\dot{\overline{x}}(t) + mk_H(t)\overline{B}\right) \times \left(\dot{\overline{y}}(t) + \left[-mk_H(t), mk_H(t)\right]\right)\right].$$

It is clear that $(\overline{x}, \overline{y})$ remains admissible, and hence optimal, for each of the problems $(\mathscr{P}_m)$. We now verify that each problem $(\mathscr{P}_m)$ satisfies the hypotheses of Theorem 2.8. To confirm (H3), simply note that $\widetilde{E}_m(t, x) \subseteq \phi_m(t)\overline{B}$ $\forall (t, x) \in \Omega$, where

$$\phi_m(t) := |\dot{\overline{x}}(t)| + |\overline{L}(t)| + 2mk_H(t).$$

Conditions (H1) and (H5) hold for $(\mathscr{P}_m)$ because they are independent of the dynamics. The measurability condition (H2) obviously holds. The sets $\widetilde{E}_m(t, x)$ are clearly compact and convex for all $(t, x) \in \Omega$: it remains only to show that they are nonempty and satisfy the Lipschitz condition of (H4). To do so, we proceed directly, aided by [2, pp. 170–172]. Fix $t \in [a, b]$, and let $x_1 \in \Omega_t$ be given, together with $(v_1, w_1) \in \widetilde{E}(t, x_1)$ and $x_2 \in \Omega_t$. By Lemma 3.1, the multifunction $x \rightrightarrows E(t, x)$ is Lipschitz of rank $k_H(t)$ on Ω_t. Since $(v_1, w_1 - M(t, v_1)) \in E(t, x_1)$, it follows that there is some point $(v_2, w_2) \in E(t, x_2)$ satisfying

$$|(v_2, w_2) - (v_1, w_1 - M(t, v_1))| \le k_H(t)|x_2 - x_1|. \tag{3.3}$$

For the corresponding point $(v_2, w_2 + M(t, v_2))$ in $\widetilde{E}(t, x_2)$, we have

$$\begin{aligned} &|(v_2, w_2 + M(t, v_2)) - (v_1, w_1)| \\ &\quad \le |(v_2, w_2) - (v_1, w_1 - M(t, v_1))| + |M(t, v_2) - M(t, v_1)| \\ &\quad \le k_H(t)|x_2 - x_1| + |v_2 - v_1|. \end{aligned} \tag{3.4}$$

The last term in this estimate arises because the function θ is globally Lipschitz of rank 1. Combining (3.3) and (3.4), we have

$$|(v_2, w_2 + M(t, v_2)) - (v_1, w_1)| \le 2k_H(t)|x_2 - x_1|.$$

Hence the multifunction $x \rightrightarrows \widetilde{E}(t, x)$ is Lipschitz of rank $2k_H(t)$ on Ω_t.

Now recall that $\Omega_t = \overline{x}(t) + \delta B$ for some fixed $\delta \in (0, 1/6)$. It follows that for any $(t, x) \in \Omega$, one has

$$\begin{aligned}(\dot{\overline{x}}(t), \dot{\overline{y}}(t)) \in \widetilde{E}(t, \overline{x}(t)) \subseteq \widetilde{E}(t, x) + 2k_H(t)|x - \overline{x}(t)|\overline{B} \\ \Rightarrow \widetilde{E}(t, x) \cap ((\dot{\overline{x}}(t), \dot{\overline{y}}(t)) + 2\delta k_H(t)\overline{B}) \neq \varnothing.\end{aligned} \tag{3.5}$$

Hence $\widetilde{E}_m(t, x) \neq \varnothing$ for all $x \in \Omega_t$ provided $m > 2\delta$; and this holds for all $m \ge 1$ by our choice of δ. Moreover, upon choosing $r_m = mk_H(t)/3$ and $M' = 2k_H(t)$, we have $r_m > 2\delta k_H(t)$, so (3.5) remains valid with r_m in place of $2\delta k_H(t)$. According to [2, Lemma 3, p. 172], the multifunction $\widetilde{E}_m(t, \cdot, \cdot) = \widetilde{E}(t, \cdot, \cdot) \cap ((\dot{\overline{x}}(t), \dot{\overline{y}}(t)) + 3r_m\overline{B})$ is nonempty, compact-convex-valued, and Lipschitz of rank $5M' = 10k_H(t)$ on Ω_t. Thus (H4) holds for the integrable function $k_{F_m}(t) := 10k_H(t)$. It is significant that this function is independent of m.

We have justified the application of Theorem 2.8 to problem $(\mathscr{P}_m)$. Exactly the same reasoning used above leads to the following conclusions: For each $m \in \mathbb{N}$, there exist a constant $\lambda_m \in \{0, 1\}$ and a function $p_m \in \mathrm{BV}([a, b]; \mathbb{R}^n)$, not both zero, together with a measurable selection $\nu_m(t) \in N(t, \overline{x}(t))$ for all $t \in [a, b]$, such that

(a_m) $(-\dot{p}_m(t) + \nu_m(t), \dot{\overline{x}}(t), L(t, \overline{x}(t), \dot{\overline{x}}(t))) \in \partial\widetilde{H}_m(t, \overline{x}(t), p_m(t), -\lambda_m)$ a.e.;

(b_m) $(p_m(a), -p_m(b)) \in \lambda_m \partial l(\overline{x}(a), \overline{x}(b)) + N_S(\overline{x}(a), \overline{x}(b))$;

(c_m) the singular part of dp_m is $N(t, \overline{x}(t))$-valued.

The key step. Conditions (a_m)–(c_m) involve the Hamiltonian function

$$\begin{aligned}\widetilde{H}_m(t, x, p, q) := \sup\{\langle p, v\rangle + qw : v \in F_m(t, x), \\ \widetilde{L}(t, x, v) \le w \le L(t, \overline{x}(t), \dot{\overline{x}}(t)) + mk_H(t)\}\end{aligned} \tag{3.6}$$

corresponding to the Lagrangian $\widetilde{L}$ through the multifunction $\widetilde{E}_m$. The strict convexity of $\widetilde{L}(t, x, v)$ as a function of v is the key to the fact that (a_m) simultaneously implies the Euler inclusion, the Hamiltonian inclusion, and the Weierstrass-Pontryagin condition for the original Lagrangian L. The reasons for this are given in detail in §4: here we simply apply them. Throughout this process, we consider a fixed time $t \in [a, b]$ and index $m \in \mathbb{N}$ for which $\dot{\overline{y}}(t) = L(t, \overline{x}(t), \dot{\overline{x}}(t))$ and inclusion (a_m) holds, and suppress t in the notation. We also drop the subscript m from the multipliers λ_m, p_m, and ν_m. Thus our starting point, the inclusion (a_m), is written more concisely as

$$(-\dot{p} + \nu, \dot{\overline{x}}, \dot{\overline{y}}) \in \partial\widetilde{H}_m(\overline{x}, p, -\lambda), \tag{3.7}$$

with

$$\widetilde{H}_m(x, p, q) = \sup\{\langle p, v\rangle + qw : v \in F_m(x),\ \widetilde{L}(x, v) \le w \le R\}. \tag{3.8}$$

(Here $R = \overline{L} + mk_H$.) In parallel to (3.8), we introduce the notation

$$\widetilde{H}_{m,\lambda}(x, p) := \sup\{\langle p, v\rangle - \lambda\widetilde{L}(x, v) : v \in F_m(x)\}. \tag{3.9}$$

The functions H_m and $H_{m,\lambda}$ are defined exactly as in (3.8) and (3.9), except that L replaces $\widetilde{L}$.

3.2. **Lemma.** *Inclusion* (3.7) *implies that*

$$|\dot{p} - \nu| \le 10k_H|(p, -\lambda)|.$$

Proof. By our previous discussion of (H4) and [2, Proposition 3.2.4(d)], the Lipschitz rank of $x \mapsto \widetilde{H}_m(x, p, q)$ is $10k_H|(p, q)|$. The result follows. □

The crux of this section is the following result, whose proof relies upon the material in §4 below.

3.3. **Proposition.** *Inclusion* (3.7) *implies the following inclusions*:

$$(-\dot{p} + \nu, \dot{\overline{x}}) \in \partial H_{m,\lambda}(\overline{x}, p), \tag{3.10}$$

$$(\dot{p} - \nu, p) \in \lambda\partial L(\overline{x}, \dot{\overline{x}}) + N_{\mathrm{Gr}\,F}(\overline{x}, \dot{\overline{x}}), \tag{3.11}$$

$$p \in \lambda\partial_v L(\overline{x}, \dot{\overline{x}}) + N_{F(\overline{x})}(\dot{\overline{x}}). \tag{3.12}$$

Proof. All three inclusions follow from the results of §4. We apply those results to the domain $U = \Omega_t$, the multifunction F_m, the Lagrangian L (or sometimes $\widetilde{L}$), and the perturbing function M. Notice that the technical hypotheses (A1)–(A3) of §4 all hold for these choices: Ω_t is a bounded open set on which F_m is compact and convex-valued by (H3), while F_m is nonempty on Ω_t because $\widetilde{E}_m$ is nonempty there. Moreover, F_m is Lipschitz of rank k_m on Ω_t; this property is inherited from $\widetilde{E}_m$. The possible choices L, M, and $\widetilde{L} = L + M$ all satisfy (A2) (compare (H6)–(H7)). Finally, assumption (A3) holds because $\widetilde{E}_m(t, x) \ne \varnothing\ \ \forall x \in \Omega_t$. Thus all the results of §4 are available to us.

Consider the multifunction $\widetilde{E}_m$ introduced above, now written concisely as

$$\widetilde{E}_m(x) = \{(v, w) : v \in F_m(x),\ \widetilde{L}(x, v) \le w \le R\}.$$

Now $(\dot{\overline{x}}, \dot{\overline{y}})$ lies in $\widetilde{E}_m(\overline{x})$ by assumption: we observe that in fact,

$$(\dot{\overline{x}}, \dot{\overline{y}}) \in \operatorname{ext}\widetilde{E}_m(\overline{x}). \tag{3.13}$$

To prove this, suppose two points $(v_i, w_i) \in \widetilde{E}_m(\overline{x})$, $i = 0, 1$, obey

$$(\dot{\overline{x}}, \dot{\overline{y}}) = \tfrac{1}{2}\left[(v_0, w_0) + (v_1, w_1)\right].$$

Then $w_i \ge \widetilde{L}(\overline{x}, v_i)$ implies that while $\dot{\overline{x}} = \frac{1}{2}(v_0 + v_1)$,

$$\dot{\overline{y}} = \widetilde{L}(\overline{x}, \dot{\overline{x}}) = \tfrac{1}{2}(w_0 + w_1) \ge \tfrac{1}{2}\widetilde{L}(\overline{x}, v_0) + \tfrac{1}{2}\widetilde{L}(\overline{x}, v_1).$$

This is compatible with the strict convexity of $\widetilde{L}(\overline{x}, \cdot)$ only if $v_0 = v_1 = \dot{\overline{x}}$ and $w_0 = w_1 = \dot{\overline{y}} = \widetilde{L}(\overline{x}, \dot{\overline{x}})$. Thus (3.13) holds. We may therefore apply Theorem 4.7 (with $E = \widetilde{E}_m$, $\overline{q} = -\lambda$, and $-\pi = -\dot{p} + \nu$) to deduce that

$$(\dot{p} - \nu, p) \in \lambda \partial \widetilde{L}(\overline{x}, \dot{\overline{x}}) + N_{\operatorname{Gr} F_m}(\overline{x}, \dot{\overline{x}}),$$
$$p \in \lambda \partial_v \widetilde{L}(\overline{x}, \dot{\overline{x}}) + N_{F_m(\overline{x})}(\dot{\overline{x}}).$$

But the sets $F_m(\overline{x})$ and $F(\overline{x})$ are indistinguishable on a neighborhood of $\dot{\overline{x}}$, and similarly $\operatorname{Gr} F_m$ is identical to $\operatorname{Gr} F$ on some open set containing $(\overline{x}, \dot{\overline{x}})$. Therefore the previous inclusions are unchanged if the subscript m is deleted from F. Since $\widetilde{L}(x, v) = L(x, v) + M(v)$ for a smooth function M whose derivative is 0 at the point $(\overline{x}, \dot{\overline{x}})$, it follows that these inclusions are identical to (3.11) and (3.12).

To recover (3.10), we apply Theorem 4.4. It follows that

$$(-\dot{p} + \nu, \dot{\overline{x}}) \in \partial \widetilde{H}_{m,\lambda}(\overline{x}, p). \tag{3.14}$$

Now if $\lambda = 0$, then $\widetilde{H}_{m,\lambda} \equiv H_{m,\lambda}$, so (3.10) follows directly. However, if $\lambda = 1$, then we may use Corollary 4.3(b). The estimate given there involves the set $\partial_p \widetilde{H}_{m,1}(\overline{x}, p)$: let us show that

$$\partial_p \widetilde{H}_{m,1}(\overline{x}, p) = \{\dot{\overline{x}}\}. \tag{3.15}$$

The inclusion $\supseteq$ follows directly from (3.14) and [2, Proposition 2.5.3]. The reverse inclusion follows from Young's identity, which implies that

$$v \in \partial_p \widetilde{H}_{m,1}(\overline{x}, p) \Leftrightarrow v \in \arg\max_{v' \in F_m(\overline{x})} \{\langle p, v' \rangle - \widetilde{L}(\overline{x}, v')\}.$$

Now the function $v' \to \langle p, v' \rangle - \widetilde{L}(\overline{x}, v')$ is strictly convex, so there can be at most one vector v satisfying the latter relation. This vector must be $\dot{\overline{x}}$, so (3.15) holds. Since $\partial M(\dot{\overline{x}}) = \{0\}$, the union in Corollary 4.3(b) actually involves only one pair (v, z): thus that inclusion reduces to

$$\partial \widetilde{H}_{m,1}(\overline{x}, p) \subseteq \operatorname{co}\{(-\pi, v) : (-\pi, v) \in \partial H_{m,1}(\overline{x}, p)\} = \partial H_{m,1}(\overline{x}, p).$$

The desired result (3.10) follows from this estimate and (3.14). □

Limiting conclusions. We now revert to fully explicit notation. Recall that for each $(t, x) \in \Omega$, the set $F(t, x)$ is compact. Consequently $\bigcup_{m=2}^{\infty} A_m = [a, b]$, where the measurable sets A_m are defined by

$$A_m = \{t \in [a, b] : F_{m-1}(t, \overline{x}(t)) = F(t, \overline{x}(t))\}.$$

But for $t \in A_m$, one has $F_m(t, x') = F(t, x')$ for all x' sufficiently near $\overline{x}(t)$, since F_m is Lipschitz in x'. Hence $H_{m,\lambda}(t, x', p) = H_\lambda(t, x', p)$ for all x' near $\overline{x}(t)$ and $p \in \mathbb{R}^n$. This gives $\partial H_{m,\lambda}(t, \overline{x}(t), p) = \partial H_\lambda(t, \overline{x}(t), p)$ for all $t \in A_m$, $p \in \mathbb{R}^n$. We use this observation to simplify the inclusion (3.10).

Taken together, inclusions (3.10)–(3.12) and statements (b_m)–(c_m) provide sequential versions of all the desired conclusions of Theorem 1.1. It remains to pass to the limit as $m \to \infty$. To do so, it is convenient to treat the measures represented by each p_m more explicitly, using the change of variables below:

$$q_m(t) := p_m(a) + \int_a^t \left[\dot{p}_m(r) - \nu_m(r)\right] dr,$$
$$d\mu_m(t) := dp_m(t) - \dot{q}_m(t)\, dt.$$

Note that the functions q_m are absolutely continuous and satisfy $q_m(a) = p_m(a)$, while the $\mathbb{R}^n$-valued measures μ_m are $N(t, \overline{x}(t))$-valued. It follows that for almost every $t \in [a, b]$,

$$(3.16)\quad |\dot{q}_m(t)| \le 10k_H(t)\left(|q_m(t) + \mu_m[a, t)| + \lambda_m\right),$$
$$(3.17)\quad (-\dot{q}_m(t), \dot{\overline{x}}(t)) \in \partial H_{\lambda_m}(t, \overline{x}(t), q_m(t) + \mu_m[a, t)) \quad (t \in A_m),$$
$$(3.18)\quad (\dot{q}_m(t), q_m(t) + \mu_m[a, t)) \in \lambda_m \partial L(t, \overline{x}(t), \dot{\overline{x}}(t)) + N_{\mathrm{Gr}\, F(t,\cdot)}(\overline{x}(t), \dot{\overline{x}}(t)),$$
$$(3.19)\quad q_m(t) + \mu_m[a, t) \in \lambda_m \partial_v L(t, \overline{x}(t), \dot{\overline{x}}(t)) + N_{F(t, \overline{x}(t))}(\dot{\overline{x}}(t)).$$

We also have

$$(3.20)\quad (q_m(a), -q_m(b) - \mu_m[a, b]) \in \lambda_m \partial l(\overline{x}(a), \overline{x}(b)) + N_S(\overline{x}(a), \overline{x}(b)),$$
$$(3.21)\quad \mathrm{Supp}(\mu_m) \subseteq \{t \in [a, b] : (t, \overline{x}(t)) \in \mathrm{bdry}\, \mathrm{Gr}\, X\},$$
$$(3.22)\quad \lambda_m + |q_m(a)| + \mathrm{Var}(q_m(t) + \mu_m[a, t) + \mu_m\{b\}\delta_{\{b\}}(t); [a, b]) = 1.$$

The last assertion is equivalent to $\lambda_m + |p_m(a)| + \mathrm{Var}(p_m; [a, b]) = 1$; it is arranged by scaling. Inequality (3.16) follows from Lemma 3.2, while (3.17)–(3.19) are the results of Proposition 3.3. Conditions (3.20) and (3.21) are transcriptions of (b_m) and (c_m) above.

By successive extraction of subsequences, which we do not relabel, we may assert the following:

1. The numbers λ_m converge to some $\lambda \in [0, 1]$ (by the Heine-Borel Theorem).

2. The arcs q_m converge uniformly to some arc $q \in \mathrm{AC}([a, b]; \mathbb{R}^n)$, with $\dot{q}_m \xrightarrow{w} \dot{q}$ weakly in $L^1([a, b]; \mathbb{R}^n)$ (since $|\dot{q}_m(t)| \le 10k_H(t)$ a.e. by (3.16) and (3.22), the Dunford-Pettis criterion implies that $\{\dot{q}_m\}$ has a weakly convergent subsequence in L^1).

3. The measures μ_m converge weak* to some finite $\mathbb{R}^n$-valued Borel measure μ on $[a, b]$ (since (3.22), together with assertion 2, implies that each component sequence of $\{\mu_m\}$ consists of real-valued measures with uniformly bounded variation; we apply Helly's selection theorem to each component).

It follows from Proposition 2.7 that $\mu = \mathrm{w}^*\text{-}\lim \mu_m$ is itself $N(t, \overline{x}(t))$-valued. Consequently the limiting forms of (3.19)–(3.22) follow immediately,

since only pointwise convergence is involved: we have

$$q(t) + \mu[a, t) \in \lambda \partial_v L(t, \overline{x}(t), \dot{\overline{x}}(t)) + N_{F(t, \overline{x}(t))}(\dot{\overline{x}}(t)) \text{ a.e.}, \tag{3.23}$$

$$(q(a), -q(b) - \mu[a, b]) \in \lambda \partial l(\overline{x}(a), \overline{x}(b)) + N_S(\overline{x}(a), \overline{x}(b)), \tag{3.24}$$

$$\operatorname{Supp}(\mu) \subseteq \{t \in [a, b] : (t, \overline{x}(t)) \in \operatorname{bdry} \operatorname{Gr} X\}, \tag{3.25}$$

$$\lambda + |q(a)| + \operatorname{Var}(q(t) + \mu[a, t) + \mu\{b\}\delta_{\{b\}}(t); [a, b]) = 1. \tag{3.26}$$

Next we turn to the two inclusions involving weak convergence, starting with (3.18). If $\lambda = \lim \lambda_m = 0$, then consider the set of functions

$$\mathscr{S}_0 = \{(r(\cdot), s(\cdot)) \in L^1([a, b]; \mathbb{R}^n \times \mathbb{R}^n) : \\ (r(t), s(t)) \in N_{\operatorname{Gr} F(t, \cdot)}(\overline{x}(t), \dot{\overline{x}}(t)) \text{ a.e.}\}.$$

This set is convex, and strongly closed in $L^1([a, b]; \mathbb{R}^{2n})$. Hence it is weakly closed. The sequence of functions on the left side of (3.18) converges weakly, and its terms approach the set $\mathscr{S}_0$ in norm since $\|\lambda_m \partial L(t, \overline{x}(t), \dot{\overline{x}}(t))\|_1 \to 0$. It follows that the weak limit of this sequence lies in $\mathscr{S}_0$.

On the other hand, if $\lambda_m \to \lambda > 0$, then the same argument may be applied, once $\mathscr{S}_0$ has been replaced by

$$\mathscr{S}_\lambda := \{(r, s) : (r(t), s(t)) \in \lambda \partial L(t, \overline{x}(t), \dot{\overline{x}}(t)) + N_{\operatorname{Gr} F(t, \cdot)}(\overline{x}(t), \dot{\overline{x}}(t)) \text{ a.e.}\}.$$

In short, we have the following inclusion for almost all $t \in [a, b]$:

$$(\dot{q}(t), q(t) + \mu[a, t)) \in \lambda \partial L(t, \overline{x}(t), \dot{\overline{x}}(t)) + N_{\operatorname{Gr} F(t, \cdot)}(\overline{x}(t), \dot{\overline{x}}(t)). \tag{3.27}$$

As for (3.17), if $\lambda_m = 0 \ \forall m$ then the multifunction on the right side of (3.17) is always ∂H_0, and the limiting validity of this relationship follows from [2, Theorem 3.1.7].

If $\lambda_m \to 0+$, then for each m there is a measurable selection $z_m(t) \in \overline{B}$ such that, by Corollary 4.3(a) below,

$$(-\dot{q}_m(t), \dot{\overline{x}}(t)) \in \partial H_0(t, \overline{x}(t), q_m(t) + \mu_m[a, t) + \lambda_m z_m(t) k_L(t)) \\ + \lambda_m k_L(t) \overline{B} \times \{0\} \quad \text{a.e. } t \in A_m.$$

Again, the result follows from [2, Proposition 3.1.7].

Finally, if $\lambda_m \to \lambda > 0$, then $q_m/\lambda_m \to q/\lambda$ uniformly, with weak convergence of the derivatives in L^1. The limiting validity of (3.17) is established by dividing through (3.17) by λ_m to obtain the following inclusion for almost all t in A_m:

$$\left(-\left(\dot{q}_m(t)/\lambda_m\right), \dot{\overline{x}}(t)\right) \in \partial H_1\left(t, \overline{x}(t), \left(q_m(t)/\lambda_m\right) + \mu[a, t)/\lambda_m\right),$$

taking the limit with the aid of [2, Theorem 3.1.7], and then multiplying both sides by λ to get the final result:

$$(-\dot{q}(t), \dot{\overline{x}}(t)) \in \partial H_\lambda(t, \overline{x}(t), q(t) + \mu[a, t)) \quad \text{a.e. } t \in [a, b]. \tag{3.28}$$

This notation is consistent with the results of all three cases.

Now conclusions (3.23)–(3.28) correspond closely to the conclusions of Theorem 2.8. The desired assertions of Theorem 1.1 follow immediately upon reversing the change of variables made above: the proof of Theorem 2.8 provides a model for this reduction. This completes the proof of Theorem 1.1.

4. Hamiltonian calculus

Necessary conditions for problem (P) typically involve functions of the following form:

$$H_\lambda(x, p) := \sup\{\langle p, v\rangle - \lambda L(x, v) : v \in F(x)\}. \tag{4.1}$$

In this section we study the generalized gradients of these functions for $\lambda \geq 0$ from three different viewpoints. We also consider the generalized gradients of the closely related function

$$H(x, p, q) := \begin{cases} H_{-q}(x, p), & \text{if } q < 0, \\ H_0(x, p) + qR, & \text{if } q \geq 0, \end{cases} \tag{4.2}$$

at points (x, p, q), where $q \leq 0$. Our presentation offers a completely self-contained description of the technical results underlying the main theorems of this paper. All standing assumptions and notational conventions in §§1–3 are abandoned (except for the useful motivation they provide), and we start with a clean slate.

Assumptions. Let $U \subseteq \mathbb{R}^n$ be a bounded open set. Throughout this section, we assume the following.

(A1) The multifunction $F: U \rightrightarrows \mathbb{R}^n$ has nonempty, compact, convex values, and obeys the following Lipschitz condition for some constant $k_F \geq 0$:

$$F(y) \subseteq F(x) + k_F|y - x|\overline{B} \quad \forall x, y \in U.$$

(A2) For some constant $\varepsilon > 0$, the function $L: \mathrm{Gr}(F + \varepsilon B) \to \mathbb{R}$ is Lipschitz of rank k_L; for each fixed $x \in U$, the map $v \to L(x, v)$ is convex on $F(x) + \varepsilon B$. (Here, of course, $\mathrm{Gr}(F + \varepsilon B) = \{(x, v) : x \in U,\ v \in F(x) + \varepsilon B\}$.)

(A3) The constant $R \geq 0$ appearing in definition (4.2) is large enough that the sets $F(x) \cap \{v : L(x, v) \leq R\}$ are nonempty for all x in U.

The most elementary properties of the functions H_λ are recorded below.

4.1. **Lemma.** *Fix $(\overline{x}, \overline{p}) \in U \times \mathbb{R}^n$ and $\lambda \geq 0$. Then*

(a) *the function $p \to H_\lambda(\overline{x}, p)$ is finite-valued and convex on $\mathbb{R}^n$;*

(b) *Young's identity holds: that is, for any $\overline{v} \in \mathbb{R}^n$,*

$$\begin{aligned} \overline{v} \in \partial_p H_\lambda(\overline{x}, \overline{p}) &\Leftrightarrow H_\lambda(\overline{x}, \overline{p}) = \langle \overline{p}, \overline{v}\rangle - \lambda L(\overline{x}, \overline{v}) \\ &\Leftrightarrow \overline{p} \in \lambda \partial_v L(\overline{x}, \overline{v}) + N_{F(\overline{x})}(\overline{v}); \end{aligned} \tag{4.3}$$

(c) *the function $x \to H_\lambda(x, \overline{p})$ is Lipschitz of rank $\lambda k_L + k_F(\lambda k_L + |\overline{p}|)$ near $\overline{x}$.*

Proof. (a) Obvious.

(b) Well known.

(c) Let $v_\lambda(x, p)$ be any selection from $\arg\max H_\lambda(x, p)$. Given $x, y \in U$ choose $v \in F(y)$ such that $|v - v_\lambda(x, \overline{p})| \le k_F|y - x|$. Then

$$\begin{aligned} H_\lambda(y, \overline{p}) &\ge \langle \overline{p}, v\rangle - \lambda L(y, v) \\ &\ge \langle \overline{p}, v_\lambda(x, \overline{p})\rangle + \langle \overline{p}, v - v_\lambda(x, \overline{p})\rangle \\ &\qquad - \lambda L(x, v_\lambda(x, \overline{p})) - \lambda k_L|(x, v_\lambda(x, \overline{p})) - (y, v)| \\ &\ge H_\lambda(x, \overline{p}) - |\overline{p}|k_F|y - x| - \lambda k_L(|y - x| + k_F|y - x|). \end{aligned}$$

Since x and y are interchangeable in this computation, the result follows. □

(a) Additional Lagrangian terms. Suppose that in addition to the multifunction F and the Lagrangian L, a second function $M\colon \mathrm{Gr}(F + \varepsilon B) \to \mathbb{R}$ is given. Assume that M satisfies (A2) with Lipschitz constant k_M. Let $\widetilde{L}(x, v) := L(x, v) + M(x, v)$, and define $\widetilde{H}_\lambda(x, p) := \sup\{\langle p, v\rangle - \lambda\widetilde{L}(x, v)\colon v \in F(x)\}$ just as in (4.1). Our first order of business is to establish a relationship between $\partial H_1(x, p)$ and $\partial\widetilde{H}_1(x, p)$ valid in $U \times \mathbb{R}^n$.

4.2. **Theorem.** *For each* $(\overline{x}, \overline{p}) \in U \times \mathbb{R}^n$, *the following estimate holds*:

$$\partial\widetilde{H}_1(\overline{x}, \overline{p}) \subseteq \mathrm{co}\bigcup\{(-\pi, v) + (\gamma, 0)\colon (-\pi, v) \in \partial H_1(\overline{x}, \overline{p} - z),\ (\gamma, v) \in \partial K_1(\overline{x}, z)\}, \tag{4.4}$$

where the union is taken over all $v \in \partial_p\widetilde{H}_1(\overline{x}, \overline{p})$ *and* $z \in \partial_v M(\overline{x}, v)$, *and where* $K_1(x, p) := \sup\{\langle p, v\rangle - M(x, v)\colon v \in F(\overline{x}) + \frac{1}{2}\varepsilon\overline{B}\}$. *In particular, one has the following estimate, useful when* k_M *is small*:

$$\partial\widetilde{H}_1(\overline{x}, \overline{p}) \subseteq \mathrm{co}\bigcup_{|z|\le k_M}(\partial H_1(\overline{x}, \overline{p} - z) + k_M\overline{B} \times \{0\}). \tag{4.5}$$

Proof. Let $G = \mathrm{Gr}\,F$, and $S = \mathbb{R}^n \times (F(\overline{x}) + \frac{1}{2}\varepsilon\overline{B})$. Since $\overline{x}$ has a neighborhood throughout which $F(x)$ is a subset of $F(\overline{x}) + \frac{1}{2}\varepsilon\overline{B}$ by (A1), the following identity holds for all (x, p) in some neighborhood of $(\overline{x}, \overline{p})$:

$$\begin{aligned} \widetilde{H}_1(x, p) &= \sup_{v\in\mathbb{R}^n}\{\langle p, v\rangle - L(x, v) - \Psi_G(x, v) - M(x, v) - \Psi_S(x, v)\} \\ &= \min_{z\in\mathbb{R}^n}\{H_1(x, p - z) + K_1(x, z)\}. \end{aligned} \tag{4.6}$$

The second identity in (4.6) reflects the duality between addition and infimal convolution of convex functions (see [11, Theorem 16.4]). This computation reveals $\widetilde{H}_1$ as the value function of a perturbed minimization problem. The generalized gradient of such a value function can be evaluated using proximal analysis, and a result of this sort appears in [18, Theorem 3.1]. Let $h(z; x, p)$ denote the function in braces in the last line of (4.6). Then h is locally Lipschitz in all variables by Lemma 4.1, and satisfies the inf-compactness condition [18, (3.2)]. Lemma 4.1 also implies that $\widetilde{H}_1$ is locally Lipschitz. In the notation

of [18], it follows that $\hat{\partial}^{\infty}\widetilde{H}_1(\overline{x}, \overline{p}) = \{0\}$, so we may conclude via [18, (2.5)] that

$$\partial\widetilde{H}_1(\overline{x}, \overline{p}) \subseteq \operatorname{co}\{(\alpha, v) : (0; \alpha, v) \in \partial h(z; \overline{x}, \overline{p}),\ z \in \Xi(\overline{x}, \overline{p})\}. \tag{4.7}$$

Here $\Xi(\overline{x}, \overline{p})$ denotes the set of points $z \in \mathbb{R}^n$ at which the minimum defining $\widetilde{H}_1(\overline{x}, \overline{p})$ is attained. To obtain the estimate (4.4), we first enlarge the right-hand side of (4.7) by replacing $\Xi(\overline{x}, \overline{p})$ by $\mathbb{R}^n$. Then we estimate

$$\begin{aligned}\partial h(z; \overline{x}, \overline{p}) \subseteq \{(-\beta; -\pi, \beta) &: (-\pi, \beta) \in \partial H_1(\overline{x}, \overline{p} - z)\} \\ &+ \{(\delta; \gamma, 0) : (\gamma, \delta) \in \partial K_1(\overline{x}, z)\}.\end{aligned}$$

It follows that if $(0; \alpha, v)$ belongs to $\partial h(z; \overline{x}, \overline{p})$, then there exist points $(-\pi, \beta) \in \partial H_1(\overline{x}, \overline{p} - z)$ and $(\gamma, \delta) \in \partial K_1(\overline{x}, z)$ for which $v = \beta = \delta$ and $\alpha = -\pi + \gamma$. Thus (4.7) implies

$$\begin{aligned}\partial\widetilde{H}_1(\overline{x}, \overline{p}) \subseteq \operatorname{co}\{(-\pi, v) + (\gamma, 0) &: (-\pi, v) \in \partial H_1(\overline{x}, \overline{p} - z), \\ &(\gamma, v) \in \partial K_1(\overline{x}, z)\}.\end{aligned} \tag{4.8}$$

To obtain the restrictions on v and z which appear in (4.4), we observe that

$$\begin{aligned}(-\pi, v) \in \partial H_1(\overline{x}, \overline{p} - z) &\Rightarrow v \in \partial_p H_1(\overline{x}, \overline{p} - z) \\ &\Rightarrow \overline{p} - z \in \partial_v[L + \Psi_G](\overline{x}, v),\end{aligned} \tag{4.9a}$$

$$\begin{aligned}(\gamma, v) \in \partial K_1(\overline{x}, z) &\Rightarrow v \in \partial_p K_1(\overline{x}, z) \\ &\Rightarrow z \in \partial_v[M + \Psi_S](\overline{x}, v).\end{aligned} \tag{4.9b}$$

Now (4.9a) implies that $v \in F(\overline{x}) \subseteq \operatorname{int}(F(\overline{x}) + \frac{1}{2}\varepsilon\overline{B})$, so (4.9b) implies $z \in \partial_v M(\overline{x}, v)$. This is the second restriction made explicit in (4.4). Moreover, adding the two inclusions in (4.9) yields

$$\overline{p} \in \partial_v L(\overline{x}, v) + \partial_v M(\overline{x}, v) + N_{F(\overline{x})}(v) = \partial_v[L + M + \Psi_G](\overline{x}, v).$$

This implies $v \in \partial_p\widetilde{H}_1(\overline{x}, \overline{p})$, and hence establishes the first restriction appearing in (4.4). The proof of (4.4) is complete.

Now as we have noted above, the vectors v participating in the right side of (4.4) all lie in $F(\overline{x})$. Therefore the inclusion $z \in \partial_v M(\overline{x}, v)$ implies $|z| \leq k_M$. Moreover, the inclusion $(\gamma, v) \in \partial K_1(\overline{x}, z)$ implies that $|\gamma|$ is bounded above by the Lipschitz rank of the function $x \to K_1(x, z)$ near $(\overline{x}, z)$. This rank is estimated in Lemma 4.1(c): replacing k_F by $k_S = 0$, k_L by k_M, and λ by 1 in that result, we deduce that $|\gamma| \leq k_M$. The estimate (4.5) follows immediately. □

4.3. **Corollary.** (a) *For all $\lambda \geq 0$ and any $(x, p) \in U \times \mathbb{R}^n$, one has*

$$\partial H_\lambda(x, p) \subseteq \operatorname{co} \bigcup_{|z| \leq \lambda k_L} \left(\partial H_0(x, p - z) + \lambda k_L \overline{B} \times \{0\}\right). \tag{4.10}$$

(b) *If the function* $M = M(v)$ *appearing in Theorem* 4.2 *is independent of* x, *then one has*

$$\partial \widetilde{H}_1(\overline{x}, \overline{p}) \subseteq \operatorname{co} \bigcup_{\substack{v \in \partial_p \widetilde{H}_1(\overline{x}, \overline{p}) \\ z \in \partial M(v)}} \{(-\pi, v) : (-\pi, v) \in \partial H_1(\overline{x}, \overline{p} - z)\}.$$

(*Note that equality holds if* $M \equiv 0$.)

Proof. (a) The inclusion (4.10) is an obvious identity for the case $\lambda = 0$. We therefore fix $\lambda > 0$ and apply Theorem 4.2 with L replaced by the zero function and M replaced by λM. These two choices obviously satisfy (A2), while the Lipschitz rank of λM is λk_M. Moreover, they imply $H_1 \equiv H_0$, so we may write the estimate (4.5) in the form

$$\partial \widetilde{H}_1(\overline{x}, \overline{p}) \subseteq \operatorname{co} \bigcup_{|z| \le \lambda k_M} \left(\partial H_0(\overline{x}, \overline{p} - z) + \lambda k_M \overline{B} \times \{0\}\right). \tag{4.11}$$

This is precisely the desired assertion (4.10), except that the names of L and M have been interchanged.

(b) If M is independent of x, then so is $K_1(x, p) = M^*(p)$. Hence all vectors (γ, v) appearing on the right side of (4.4) have $\gamma = 0$. Also, one has $(0, v) \in \partial K_1(\overline{x}, z)$ if and only if $z \in \partial M(v)$: this condition needs to be imposed only once. □

(b) The q-dependence in (4.2). We now consider the generalized gradient of the function H defined by (4.2). Corollary 4.3(a) proves useful in this analysis.

4.4. **Theorem.** *For any* $(x, p) \in U \times \mathbb{R}^n$ *and any* $q \le 0$, *one has*

$$(-\pi, v, w) \in \partial H(x, p, q) \;\Rightarrow\; (-\pi, v) \in \partial H_{-q}(x, p). \tag{4.12}$$

Furthermore, for any $(x, p, q) \in U \times \mathbb{R}^n \times (-\infty, 0]$ *and* $\lambda > 0$,

$$(\alpha/\lambda, \beta) \in \partial H_{-q}(x, p/\lambda) \;\Leftrightarrow\; (\alpha, \beta) \in \partial H_{-\lambda q}(x, p). \tag{4.13}$$

Proof. We first prove (4.13). Observe that for any $(x, p, q) \in U \times \mathbb{R}^n \times (-\infty, 0]$ and $\lambda > 0$,

$$H_{-\lambda q}(x, p) = \lambda H_{-q}(x, p/\lambda). \tag{4.14}$$

This identity implies that $H_{-\lambda q}$ is differentiable at the point (x, p) if and only if H_{-q} is differentiable at the point $(x, p/\lambda)$, and that in this case

$$\nabla H_{-\lambda q}(x, p) = (\alpha, \beta) \;\Leftrightarrow\; (\alpha/\lambda, \beta) = \nabla H_{-q}(x, p/\lambda).$$

Therefore the equivalence (4.13) follows from [2, Theorem 2.5.1].

To prove (4.12), we consider two cases. Suppose first that $q < 0$. Throughout the open set $U \times \mathbb{R}^n \times (-\infty, 0)$, we have

$$H(x, p, q) = -qH_1(x, p/(-q))$$

by definition. Thus H is a product and composition of locally Lipschitz functions, whose generalized gradient may be estimated using the product and chain rules of nonsmooth calculus [2, Propositions. 2.3.13 and 2.6.6]:

$$\begin{aligned}\partial H(x, p, q) &\subseteq \{(-q\alpha, \beta, \beta p/(-q) - H_1(x, p/(-q))) : \\ &\qquad\qquad (\alpha, \beta) \in \partial H_1(x, p/(-q))\} \\ &\subseteq \{(-q\alpha, \beta) : (\alpha, \beta) \in \partial H_1(x, p/(-q))\} \times \mathbb{R}.\end{aligned}$$

According to (4.13), the set on the right side coincides with $\partial H_{-q}(x, p) \times \mathbb{R}$, and (4.12) follows for $q < 0$.

Now consider the case $q = 0$. Since H is locally Lipschitz, we have the general relation (using [2, Theorem 2.5.1] with $S = \mathbb{R}^{2n} \times \{0\}$ in conjunction with [2, Proposition 2.2.2] and [2, Proposition 2.1.5])

$$\begin{aligned}\partial H(x, p, q) = \operatorname{co}\{\lim(-\pi_i, v_i, w_i) &: \\ &(-\pi_i, v_i, w_i) \in \partial H(x_i, p_i, q_i), \\ &(x_i, p_i, q_i) \to (x, p, q), q_i \neq 0 \ \forall i\}.\end{aligned} \tag{4.15}$$

Suppose, therefore, that a sequence $(x_i, p_i, q_i) \to (x, p, 0)$ is given with $q_i \neq 0$ $\forall i$. By passing to a subsequence, we may assume either that $q_i > 0$ for all i, or that $q_i < 0$ for all i. To treat the first possibility, note that in the half-space where $q > 0$, one has $H(x, p, q) = H_0(x, p) + qR$ and thus

$$\partial H(x, p, q) = \partial H_0(x, p) \times \{R\}.$$

On the other hand, if the sequence (x_i, p_i, q_i) has $q_i < 0$ for all i, then the arguments of the previous paragraph, together with Corollary 4.3(a), yield the estimates

$$\begin{aligned}\partial H(x_i, p_i, q_i) &\subseteq \partial H_{-q_i}(x_i, z_i) \times \mathbb{R} \\ &\subseteq \operatorname{co} \bigcup_{|z| \le |q_i| k_L} (\partial H_0(x, p - z) + |q_i| k_L \overline{B} \times \{0\}) \times \mathbb{R}.\end{aligned}$$

The upper semicontinuity of the multifunction ∂H_0 implies that the limit of the sequence $(-\pi_i, v_i, w_i)$ must lie in $\partial H_0(x, p) \times \mathbb{R}$. Therefore any convergent sequence as described in (4.15) must have its limit point in $\partial H_0(x, p) \times \mathbb{R}$. This establishes (4.12) in the case $q = 0$. □

(c) An estimate involving the Lagrangian. Note that definition (4.2) is equivalent to

$$H(x, p, q) = \sup\{\langle (p, q), (v, w)\rangle : (v, w) \in E(x)\},$$

where $E(x) := \{(v, w) : v \in F(x),\ L(x, v) \le w \le R\}$. Our next result details the relationship between the multifunctions ∂H and $\operatorname{Gr} E$.

4.5. **Lemma.** *For each* $(x, p) \in U \times \mathbb{R}^n$ *and* $q \le 0$, *one has*

$$\begin{aligned}\partial H(x, p, q) \subseteq \operatorname{co} \bigcup_{(v, w)} \{(-\pi, v, w) &: (\pi, p, q) \in N_{\operatorname{Gr} E}(x, v, w), \\ &(p, q) \in N_{E(x)}(v, w)\}.\end{aligned} \tag{4.16}$$

Proof. Observe that definition (4.2) can be rewritten as

$$(4.17)\qquad (-H)(x, p, q) = \min_{\substack{v\in\mathbb{R}^n\\ w\in\mathbb{R}}} \{\langle -(p, q), (v, w)\rangle + \Psi_{\operatorname{Gr} E}(x, v, w)\},$$

where E is the multifunction introduced above. The negative signs in (4.17) appear because *sub*differential analysis is best suited to marginal functions arising from *min*imization. The key to such analysis is the proximal subgradient (or, equivalently, the proximal normal), and a suitable estimate of the generalized gradient of $-H$ may be extracted from [18, Theorem 3.1]. Indeed, let $h(v, w; x, p, q)$ denote the function whose minimum is computed on the right side of (4.17), and let $\Xi(x, p, q)$ denote the set of minimizing points (v, w) in $\mathbb{R}^n \times \mathbb{R}$. Then h is lower semicontinuous everywhere, the value $-H(x, p, q)$ is everywhere finite, and the inf-compactness condition [18, (3.2)] is easy to verify. In fact, $-H$ is locally Lipschitzian, so

$$\partial(-H)(x, p, q) \subseteq \operatorname{co} \bigcup_{(v,w)\in\,\Xi(x,p,q)} \{(\pi, -\phi, -\psi) : (0, 0; \pi, -\phi, -\psi) \in \hat{\partial} h(v, w; x, p, q)\},$$

whence

$$\partial H(x, p, q) \subseteq \operatorname{co} \bigcup_{(v,w)\in\,\Xi(x,p,q)} \{(-\pi, \phi, \psi) : (0, 0; \pi, -\phi, -\psi) \in \partial h(v, w; x, p, q)\}.$$

Now elementary subdifferential calculus implies that

$$(v, w) \in \Xi(x, p, q) \Leftrightarrow (p, q) \in N_{E(x)}(v, w)$$

and that

$$(0, 0; \pi, -\phi, -\psi) \in \partial h(v, w, x, p, q) \Rightarrow \begin{cases} (\pi, p, q) \in N_{\operatorname{Gr} E}(x, v, w), \\ (\phi, \psi) = (v, w). \end{cases}$$

Therefore (4.16) follows. □

In applications of the estimate (4.16), the following result is useful.

4.6. **Proposition.** *For any* $(x, p) \in U \times \mathbb{R}^n$ *and* $q \le 0$, *one has*

$$(4.18)\qquad \begin{aligned} &(p, q) \in N_{E(x)}(v, L(x, v)) \;\Rightarrow\; p \in -q\partial_v L(x, v) + N_{F(x)}(v), \\ &(\pi, p, q) \in N_{\operatorname{Gr} E}(x, v, L(x, v)) \\ &\qquad\qquad \Rightarrow\; (\pi, p) \in -q\partial L(x, v) + N_{\operatorname{Gr} F}(x, v). \end{aligned}$$

Proof. Fix any point $(x, v, L(x, v))$, where $x \in U$ and $v \in F(x)$. Then there is a neighborhood of this point throughout which the following identities hold locally:

$$\operatorname{Gr} E = \operatorname{epi}(L + \Psi_{\operatorname{Gr} F}), \quad E(x) = \operatorname{epi}(L(x, \cdot) + \Psi_{F(x)}(\cdot)).$$

The desired inclusions follow immediately from this observation and the general estimate

$$N_{\mathrm{epi}(l+\Psi_S)}(\bar{s}, l(\bar{s})) \subseteq \{(\zeta, -\varepsilon) : \zeta \in \varepsilon\partial l(\bar{s}) + N_S(\bar{s})\},$$

which holds for any closed set $S \subseteq \mathbb{R}^n$, locally Lipschitz function $l : S \to \mathbb{R}$, and point $\bar{s} \in S$. A proof of this estimate may be found in [7, Lemma 4.1]. □

4.7. **Theorem.** *Let $(\bar{x}, \bar{p}) \in U \times \mathbb{R}^n$ and $\bar{q} \leq 0$. Let $(\bar{v}, L(\bar{x}, \bar{v}))$ be an extreme point of the compact convex set $E(\bar{x})$. Then any point $(-\pi, \bar{v}, L(\bar{x}, \bar{v}))$ in $\partial H(\bar{x}, \bar{p}, \bar{q})$ satisfies*

$$\bar{p} \in -\bar{q}\partial_v L(\bar{x}, \bar{v}) + N_{F(\bar{x})}(\bar{v}), \tag{4.20}$$

$$(\pi, \bar{p}) \in -\bar{q}\partial L(\bar{x}, \bar{v}) + N_{\mathrm{Gr}\,F}(\bar{x}, \bar{v}). \tag{4.21}$$

Proof. Applying inclusion (4.16) at the point $(\bar{x}, \bar{p}, \bar{q})$, we find that the vector $(-\pi, \bar{v}, L(\bar{x}, \bar{v}))$ belonging to the left-hand side must admit the decomposition

$$(-\pi, \bar{v}, L(\bar{x}, \bar{v})) = \sum_{i=1}^{N} \alpha_i(-\pi_i, v_i, w_i), \tag{4.22}$$

for some $N \in \mathbb{N}$ and $\alpha_i \geq 0$ with $\sum \alpha_i = 1$, where

$$(\pi_i, \bar{p}, \bar{q}) \in N_{\mathrm{Gr}\,E}(\bar{x}, v_i, w_i), \tag{4.23}$$

$$(p, \bar{q}) \in N_{E(\bar{x})}(v_i, w_i). \tag{4.24}$$

Inclusion (4.23) implies that each (v_i, w_i) belongs to $E(\bar{x})$. By assumption, $(\bar{v}, L(\bar{x}, \bar{v}))$ is an extreme point of this set, so the last two components of (4.22) force $(v_i, w_i) = (\bar{v}, L(\bar{x}, \bar{v}))$ for all i. Consequently the right sides of (4.23) and (4.24) are actually convex cones independent of i, and (4.22) implies

$$(-\pi, \bar{p}, \bar{q}) \in N_{\mathrm{Gr}\,E}(\bar{x}, \bar{v}, L(\bar{x}, \bar{v})), \tag{4.25}$$

$$(\bar{p}, \bar{q}) \in N_{E(\bar{x})}(\bar{v}, L(\bar{x}, \bar{v})). \tag{4.26}$$

The desired results follow from (4.25)–(4.26) via Proposition 4.6. □

Remarks on $\partial H_0(\bar{x}, \bar{p})$. As noted in the Introduction, Hamiltonian necessary conditions for the Generalized Problem of Bolza in the absence of a calmness hypothesis have been given by Clarke in [6]. In the normal case, the right-hand side of Clarke's Hamiltonian inclusion is simply ∂H_1, as it is in ours. But in the abnormal case, the right-hand side of Clarke's inclusion is $\Gamma(t, \bar{x}(t), p(t))$, where (dropping the t-dependence for clarity)

$$\Gamma(x, p) := \mathrm{cl\,co}\{\lim(\varepsilon_i\phi_i, v_i) : (\phi_i, v_i) \in \partial H_1(x_i, p_i/\varepsilon_i), \\ (x_i, p_i) \to (x, p), \varepsilon_i \to 0^+\}.$$

(This follows from [6, Proposition 4.1].) Upon substituting $-\pi_i = \varepsilon_i\phi_i$ and using identity (4.13), we find that

$$\Gamma(x, p) = \mathrm{cl\,co}\{\lim(-\pi_i, v_i) : (-\pi_i, v_i) \in \partial H_{\varepsilon_i}(x_i, p_i), \\ (x_i, p_i) \to (x, p), \varepsilon_i \to 0^+\}.$$

Note that every sequence $(-\pi_i, v_i)$ appearing on the right-hand side of this relation is bounded by Corollary 4.3(a). Indeed, any such sequence obeys

$$(-\pi_i, v_i) \in \operatorname{co} \bigcup_{|z| \le \varepsilon_i k_L} \left(\partial H_0(x_i, p_i - z) + \varepsilon_i k_L \overline{B} \times \{0\}\right) \quad \forall i.$$

Consequently $\lim(-\pi_i, v_i) \in \partial H_0(x, p)$ and we deduce that

$$\Gamma(x, p) \subseteq \partial H_0(x, p).$$

Hence Clarke's form of the abnormal Hamiltonian inclusion implies ours.

References

1. J.-P. Aubin and I. Ekeland, *Applied nonlinear analysis*, Wiley-Interscience, New York, 1984.
2. Frank H. Clarke, *Optimization and nonsmooth analysis*, Wiley-Interscience, New York, 1983.
3. ——, *Optimal solutions to differential inclusions*, J. Optim. Theory Appl. **19** (1976), 469–478.
4. ——, *Extremal arcs and extended Hamiltonian systems*, Trans. Amer. Math. Soc. **231** (1977), 349–367.
5. ——, *The generalized problem of Bolza*, SIAM J. Control Optim. **14** (1976), 682–699.
6. ——, *Hamiltonian analysis of the generalized problem of Bolza*, Trans. Amer. Math. Soc. **301** (1987), 385–400.
7. P. D. Loewen, F. H. Clarke, and R. B. Vinter, *Differential inclusions with free time*, Ann. Inst. H. Poincaré Anal. Non Linéaire **5** (1988), 573–593.
8. P. D. Loewen, *The proximal normal formula in Hilbert space*, Nonlinear Anal. **11** (1987), 979–995.
9. Nadia Raissi, *Analyse proximale en optimisation*, Ph.D. thesis, Université de Montréal, 1987.
10. R. T. Rockafellar, *Existence theorems for general control problems of Bolza and Lagrange*, Adv. in Math. **15** (1975), 312–333.
11. ——, *Convex analysis*, Princeton Univ. Press, Princeton, N. J., 1970.
12. ——, *State constraints in convex control problems of Bolza*, SIAM J. Control Optim. **10** (1972), 691–715.
13. ——, *Dual problems of Lagrange for arcs of bounded variation*, Calculus of Variations and Control Theory (D. L. Russell, ed.), Academic Press, New York, 1976, pp. 155–192.
14. ——, *Optimality conditions for convex control problems with nonnegative states and the possibility of jumps*, Game Theory and Mathematical Economics (O. Moeschlin and D. Pallaschke, eds.), North-Holland, Amsterdam, 1981, pp. 339–349.
15. ——, *Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming*, Math. Programming Stud. **17** (1982), 28–66.
16. ——, *Integral functionals, normal integrands, and measurable selections*, Nonlinear Operators and the Calculus of Variations (Bruxelles 1975) (J. P. Gossez et al., eds.), Lecture Notes in Math., vol. 543, Springer-Verlag, 1976, pp. 157–207.
17. ——, *Integrals which are convex functionals*. II, Pacific J. Math. **39** (1971), 429–469.
18. ——, *Extensions of subgradient calculus with applications to optimization*, Nonlinear Anal. **9** (1985), 665–698.
19. ——, *Hamiltonian trajectories and duality in the optimal control of linear systems with convex costs*, SIAM J. Control Optim. **27** (1989), 1007–1025.

Department of Mathematics, University of British Columbia, Vancouver, British Columbia V6T 1Y4, Canada

Department of Mathematics, University of Washington, Seattle, Washington 98195