Reprinted from:
CALCULUS OF VARIATIONS AND CONTROL THEORY

ACADEMIC PRESS, INC.
New York San Francisco London

Dual Problems of Lagrange for Arcs of Bounded Variation

R. Tyrrell Rockafellar

Introduction.

In Lagrange problems in optimal control and the calculus of variations, an integral functional of state and velocity is minimized over a class of arcs in R^n satisfying an endpoint condition and other constraints. For problems with joint convexity properties in state and velocity, a theory of duality is available in the context of arcs which are absolutely continuous. However, for inherent reasons, a fully satisfactory treatment of state constraints is not possible without an extension of the basic foundations so as to admit arcs which are merely of bounded variation. Such an extension in symmetric form is carried out here for the first time. Results are obtained on the characterization of optimal arcs in terms of a generalized Hamiltonian "equation", as well as on their existence and the possibility of identifying or approximating them by absolutely continuous arcs.

1. Lagrange Problems with Abstract Constraints.

Let $[t_0, t_1]$ be a fixed, bounded interval. An extended-real-valued function h on $[t_0, t_1] \times R^m$ is said to be a Lebesgue-normal-integrand (or respectively, a Borel-normal integrand) if the epigraph

$$\text{epi } h(t,\cdot) = \{(z,\alpha) \in R^m \times R \mid \alpha \geq h(t,z)\}$$

is closed and depends Lebesgue (resp. Borel) measurably on t, in the sense that for each closed $K \subset R^m \times R$ the set

$$\{t \in [t_0, t_1] \mid K \cap \text{epi } h(t,\cdot) \neq \emptyset\}$$

is Lebesgue (resp. Borel) measurable. This concept has been introduced in the study of integral functionals. General results covering all facts cited below may be found in [1], but we shall be concerned mainly with the case treated extensively in [2], [3], [4], [5] where the integrand is convex and proper, i.e. $h(t,\cdot)$ is for each t a convex function which is not identically $+\infty$ and which nowhere has the value $-\infty$.

Normality implies that $h(t,z(t))$ is Lebesgue (resp. Borel) measurable in t when $z(t)$ is. The closedness of epi $h(t,\cdot)$ is equivalent to the lower semicontinuity of $h(t,\cdot)$. In fact, h is Lebesgue normal if and only if the latter property holds for every $t \in [t_0,t_1]$ and h is measurable with respect to the σ-algebra in $[t_0,t_1] \times R^m$ generated by products of Lebesgue sets in $[t_0,t_1]$ and Borel sets in R^m. However, the same is not true with "Lebesgue" replaced by "Borel"; the sufficiency fails.

Let $\mathcal{A}$ be the space of all absolutely continuous functions on the fixed interval $[t_0,t_1]$. By a Lagrange functional on $\mathcal{A}$, we shall mean an extended-real-valued functional of the form

$$J_L(x) = \int_{t_0}^{t_1} L(t,x(t),\dot{x}(t))dt , \tag{1.1}$$

where the function

$$L : [t_0,t_1] \times R^n \times R^n \to R \cup \{+\infty\} \tag{1.2}$$

is a Lebesgue-normal integrand called the Lagrangian. Normality ensures that $L(\cdot,x(\cdot),\dot{x}(\cdot))$ is Lebesgue measurable (defined almost everywhere) for each $x \in \mathcal{A}$. We adopt the convention that $J_L(x) = +\infty$ if neither the positive nor the negative part of $L(\cdot,x(\cdot),\dot{x}(\cdot))$ is integrable over $[t_0,t_1]$. Then J_L is well-defined on all of $\mathcal{A}$, although both $+\infty$ and $-\infty$ are generally possible as values.

For each choice of x_0 and x_1 in R^n, there is an associated problem of Lagrange:

$$\text{minimize } J_L(x) \text{ subject to } x(t_0) = x_0, x(t_1) = x_1 . \tag{1.3}$$

Here the minimization is formally over all $x \in \mathcal{A}$, but because of the way $+\infty$ is admitted certain constraints are abstractly represented. Unless J_L is identically $+\infty$ on $\mathcal{A}$, the minimization effectively concerns only the arcs x such that

(1.4) $\dot{x}(t) \in E(t, x(t))$ a.e. (control constraint),

(1.5) $x(t) \in X(t)$ a.e. (state constraint),

where

(1.6) $E(t,x) = \{v \in R^n \mid L(t,x,v) < +\infty\}$,

(1.7) $X(t) = \{x \in R^n \mid E(t,x) \neq \emptyset\}$.

The abstract model (1.3) thus serves for a wide variety of problems. It appears very economical for existence theorems in optimal control (see [6]). In that context it is natural (and virtually essential) to require also that $L(t,x,v)$ be convex in v for fixed (t,x). Fenchel's notion of conjugate convex functions [7] then provides a one-to-one correspondence between such Lagragians L and certain functions

(1.8) $H : [t_0, t_1] \times R^n \times R^n \to R \cup \{\pm\infty\}$

called Hamiltonians, namely

(1.9) $H(t,x,p) = \sup_{v \in R^n} \{p \cdot v - L(t,x,v)\}$,

(1.10) $L(t,x,v) = \sup_{p \in R^n} \{p \cdot v - H(t,x,p)\}$.

Existence theorems in the context of $x \in \mathcal{A}$ typically entail the assumption that $H(t,x,p) < +\infty$ for all (t,x,p). (This is a growth condition on the functions $L(t,x,\cdot)$.) Note by way of contrast that

(1.11) $H(t,x,p) > -\infty \iff x \in X(t)$.

In the present paper we shall be occupied with extending the duality theory developed in [8], [9] and [10]. This requires $L(t,x,v)$ to be convex not just in v, but in x and v jointly. We therefore assume henceforth that the Lebesgue-normal integrand L on $[t_0, t_1] \times R^{2n}$ is convex and proper (as defined earlier). The convexity of $L(t,x,v)$ in (x,v) is equivalent to the concavity of $H(t,x,p)$ in x for each p. (Trivially, $H(t,x,p)$ is always convex in p.) It implies the convexity of the set $X(t)$ and the functional J_L.

A generalized Hamiltonian "equation" can then be formulated using concepts of convex analysis. For each $t \in [t_0, t_1]$ and $(x,p) \in R^n \times R^n$, let $\partial H(t,x,p)$ denote the set of all subgradients of the concave-convex function $H(t,\cdot,\cdot)$ at (x,p) [7, p. 374], i.e. the closed convex set consisting of all $(y,q) \in R^n \times R^n$ such that q is a subgradient of the convex function $H(t,x,\cdot)$ at p, while y is a subgradient of the convex function $-H(t,\cdot,p)$ at x. The generalized equation is

(1.12) $$(-\dot{p}(t), \dot{x}(t)) \in \partial H(t,x(t),p(t)) \quad \text{a.e.}$$

It was shown in [9] that if $H(t,x,p) > -\infty$ everywhere (no state constraints), solutions $x \in \mathcal{A}$ to Lagrange problems (1.3) can typically be characterized in terms of (1.12) being satisfied for some $p \in \mathcal{A}$. Moreover, such an arc p solves a parallel Lagrange problem corresponding to a certain dual Lagrangian,

$$M : [t_0, t_1] \times R^n \times R^n \to R \cup \{+\infty\}.$$

The latter function, which is likewise a Lebesgue-normal integrand, convex and proper, is related to L by

(1.13) $$M(t,p,s) = \sup_{x,v} \{p \cdot v + s \cdot x - L(t,x,v)\},$$

(1.14) $$L(t,x,v) = \sup_{p,s} \{p \cdot v + s \cdot x - M(t,p,s)\}.$$

The dual class of Lagrange problems

(1.15) $$\text{minimize } J_M(p) \text{ subject to } p(t_0) = p_0, \; p(t_1) = p_1,$$

also involves abstract constraints, in particular

(1.16) $$p(t) \in P(t) \quad \text{a.e.} \quad \text{(dual state constraint)},$$

where

(1.17) $$P(t) = \{p \in R^n \mid \exists\, s \in R^n \text{ with } M(t,p,s) < +\infty\}.$$

It can be shown from (1.13) and (1.9) that

(1.18) $\quad P(t) \subset \{p \in R^n \mid H(t,x,p) < +\infty\} \subset \mathrm{cl}\, P(t)$ for each $x \in X(t)$

(see the remarks in §4 just prior to Theorem 2) and therefore

(1.19) $\quad P(t) = R^n \iff H(t,x,p) < +\infty$ for all (x,p) .

The absence of effective state constraints for L does not imply the absence of such constraints for M , and therefore the Hamiltonian condition (1.12) does not lead to a fully satisfactory theory, from the point of view of duality, unless $H(t,x,p)$ is actually finite everywhere. On the other hand, the introduction of state constraints apparently requires far-reaching extensions, not only in the formulation of the "equation", but also in the scope of the problem.

Experience in the theory of optimal control leads us to expect that, if the arc x is really affected by state constraints, this should be reflected in optimality conditions by the possibility of the dual arc p having certain jumps. Thus both the Hamiltonian "equation" and the dual class of Lagrange problems must be broadened to allow for p to be not absolutely continuous, but merely of bounded variation.

We have demonstrated the feasibility of such a generalization in [10], but in the context of a particular structure for L which does not carry over to M . It is desirable to broaden this to a more symmetric framework. Clearly, this means by duality that arcs x which are only of bounded variation must be admitted into the original class of Lagrange problems.

This then is our goal: to extend the convex functionals J_L and J_M from $\mathcal{A}$ to the larger space $\mathcal{B}$ of arcs of bounded variation in such a way that optimality in the Lagrange problems is characterized by an extended Hamiltonian condition. It happens that this can be accomplished without, as might be feared, losing contact with the original problems: the extended problems will be seen usually to differ only in allowing "ideal" solutions when solutions in the sense of absolutely continuous arcs might not otherwise exist. One reaches in this way the conclusion that solutions in the latter sense, when available, are optimal in a wider context than has been realized.

The results obtained here largely encompass the ones for state-constrainted Lagrange problems in [10], as well as a necessary condition derived by Halkin [11], although we shall not go into the details at this time. We plan to show later how to apply them with slight further development, to problems of Bolza to get generalizations of various other theorems of [9] and [10].

2. Arcs of Bounded Variation.

Every function $x: [t_0, t_1] \to R^n$ of bounded variation gives rise to an R^n-valued regular Borel measure dx on $[t_0, t_1]$. The atoms for dx occur only at discontinuities of x, of which there are at most countably many. At any discontinuity, the right and left limits of x exist; if these are equal, there is actually no atom, and the discontinuity is said to be removable. This cannot be the case if the discontinuity is at $t = t_0$ or $t = t_1$.

Removable discontinuities have no useful role in our development, and we therefore regard as equivalent any two R^n-valued functions x_1 and x_2 of bounded variation on $[t_0, t_1]$ such that all the discontinuities of $x_1 - x_2$ are of this type. An arc of bounded variation in R^n over $[t_0, t_1]$ is defined to be such an equivalence class, and the space of all these arcs is denoted by $\mathcal{B}$. Nevertheless, where there is no harm in it, we often speak of an element x of $\mathcal{B}$ as a function, being careful only to associate with x objects that would not be affected if x were replaced by an equivalent function. We thus regard the space $\mathcal{A}$ of absolutely continuous arcs as a subspace of $\mathcal{B}$.

Observe that we can unambiguously associate with each arc $x \in \mathcal{B}$ the endpoints $x(t_0)$ and $x(t_1)$, as well as the value $x(t)$ at any $t \in (t_0, t_1)$ where dx does not have an atom. Furthermore, there are uniquely determined functions x_+ and x_- from $[t_0, t_1]$ to R^n, right and left continuous respectively, such that $x_+(t) = x_-(t) = x(t)$ at all the nonatomic points just mentioned, while

$$x_-(t_0) = x(t_0) \quad \text{and} \quad x_+(t_1) = x(t_1) . \tag{2.1}$$

It is sometimes convenient to use the notation

$$x(t+) = x_+(t) \text{ and } x(t-) = x_-(t) \text{ for } t_0 \le t \le t_1 . \tag{2.2}$$

The quantity

$$\Delta x(t) = x(t+) - x(t-) = x_+(t) - x_-(t) \tag{2.3}$$

is called the jump of the arc x at t, and if it is nonzero there is an atom of dx at t with this value.

There is no ambiguity in associating with each arc $x \in \mathcal{B}$ the derivative function $\dot{x} = dx/dt$, since this is only defined in an almost everywhere sense anyway. (A function of bounded variations is differentiable almost everywhere.) Strictly speaking, $\dot{x}$ is an element of the Lebesgue space $\mathcal{L}^1_n = \mathcal{L}^1([t_0, t_1], R^n)$. We denote by $\dot{x}\,dt$ the absolutely continuous part of the measure dx. The singular part of dx, which may of course consist of more than just the atoms described above, can be represented as $(dx/d\theta)d\theta$, where $d\theta$ is some nonnegative singular measure (a regular Borel measure), and $dx/d\theta$ is the Radon-Nikodym derivative of dx with respect to $d\theta$.

The following formula for "integration by parts" will be needed.

Proposition 1. For any $x \in \mathcal{B}$ and $p \in \mathcal{B}$, one has

$$\begin{aligned} x(t_1)\cdot p(t_1) - x(t_0)\cdot p(t_0) &= \int_{t_0}^{t_1} x_+ dp + \int_{t_0}^{t_1} p_- dx \\ &= \int_{t_0}^{t_1} x_- dp + \int_{t_0}^{t_1} p_+ dx . \end{aligned} \tag{2.4}$$

Proof. Let $\tilde{x} : (-\infty, +\infty) \to R^n$ be any function whose restriction to $[t_0, t_1]$ belongs to the equivalence class constituting the arc x and which has the constant values $x(t_0)$ on $(-\infty, t_0)$ and $x(t_1)$ on $(t_1, +\infty)$. Similarly $\tilde{p} : (-\infty, +\infty) \to R^n$ corresponding to p. Then $\tilde{x}$ and $\tilde{p}$ are of locally bounded variation, and over any bounded open interval (a, b) we may apply an integration-by-parts formula furnished by Asplund and Bungart [12, Prop. 8.5.5 on p. 374]:

$$\int_{(a,b)} \tilde{x}\, d\tilde{p} + \int_{(a,b)} \tilde{p}\, d\tilde{x} = \tilde{x}(b-) \cdot \tilde{p}(b-) - \tilde{x}(a+) \cdot \tilde{p}(a+)$$

$$+ \sum_{a<t<b} [\tilde{x}(t) - \tilde{x}(t-)] \cdot [\tilde{p}(t) - \tilde{p}(t-)] + \sum_{a<t<b} [\tilde{x}(t+) - \tilde{x}(t)] \cdot [\tilde{p}(t+) - \tilde{p}(t)].$$

Taking (a,b) to contain $[t_0, t_1]$ and

$$\tilde{x}(t) = x_+(t) \text{ and } \tilde{p}(t) = p_-(t) \text{ for } t_0 \le t \le t_1 ,$$

we have $\tilde{x}(t+) - \tilde{x}(t) = 0$ and $\tilde{p}(t) - \tilde{p}(t-) = 0$ for all t, so that the two sums reduce to zero and the formula turns into the first equation in Proposition 1. The second equation follows by symmetry.

3. Extended Lagrange Functionals.

The question of how to extend the Lagrange functional J_L from arcs in $\mathcal{A}$ to arcs in $\mathcal{B}$ has a natural answer in the light of recent developments in the theory of convex integral functionals [5]. Since $L(t,x,v)$ is a lower semicontinuous, proper convex function of (x,v) for each t, the quantity

$$(3.1) \qquad r_L(t,z) = \lim_{\lambda \to +\infty} [L(t,x_0,v_0 + \lambda z) - L(t,x_0,v_0)]/\lambda$$

is well-defined and independent of (x_0,v_0), as long as $L(t,x_0,v_0) < +\infty$, and as a function of z it is lower semicontinuous, positively homogeneous, convex and proper [7, §8]. In fact, $r_L(t,\cdot)$ is the so-called recession function of the convex function $L(t,x,\cdot)$ for every $x \in X(t)$. Since the conjugate of $L(t,x,\cdot)$ is $H(t,x,\cdot)$ by (1.9) it follows from [7, Theorem 13.3] and (1.18) that

$$(3.2) \qquad r_L(t,z) = \sup\{z \cdot p \mid p \in P(t)\}.$$

We extend the functional J_L from $\mathcal{A}$ to $\mathcal{B}$ by the formula

$$(3.3) \qquad J_L(x) \triangleq \int_{t_0}^{t_1} L(t,x(t),\dot{x}(t))dt + \int_{t_0}^{t_1} r_L(t,\xi(t))d\theta(t) \text{ for } x \in \mathcal{B},$$

where $\xi\, d\theta$ is any representation of the singular measure $dx - \dot{x}dt$

(with $d\theta$ nonnegative and ξ Borel measurable). Here our earlier conventions about $\pm\infty$ remain in force for both integrals, and $J_L(x)$ is interpreted as $+\infty$ if the integrals are oppositely infinite; the requisite normality of the integrand r_L is discussed below. The expression is independent of the particular representation $\xi d\theta$, because $r_L(t,\cdot)$ is positively homogeneous: if $\xi' d\theta'$ were another representation, we would have

$$\xi \frac{d\theta}{d\tau} = \xi' \frac{d\theta'}{d\tau} \quad \text{a.e. } d\tau \text{ with } d\tau = d\theta + d\theta' ,$$

and consequently

$$\int_{t_0}^{t_1} r_L(t,\xi(t)) \frac{d\theta}{d\tau}(t) d\tau(t) = \int_{t_0}^{t_1} r_L(t,\xi'(t)) \frac{d\theta'}{d\tau}(t) d\tau(t) .$$

Note that if the singular part of dx is purely atomic, one has

$$J_L(x) = \int_{t_0}^{t_1} L(t,x(t),\dot{x}(t))dt + \sum_{t_0 \le t \le t_1} r_L(t,\Delta x(t)) . \tag{3.4}$$

Of course, if $x \in \mathcal{A}$ then (3.3) reduces to the earlier definition (1.1), because then $\xi(t) = 0$ and $r_L(t,0) = 0$.

For (3.3) to make sense, we need $r_L(t,\xi(t))$ to be Borel measurable in t whenever $\xi(t)$ is, so that the integral can be taken with respect to a general Borel measure $d\theta$. An appropriate condition ensuring this would be the Borel-normality of r_L. This can be translated into a condition on $P(t)$.

<u>Proposition 2.</u> <u>The integrand</u> r_L <u>on</u> $[t_0,t_1] \times R^n$ <u>is Borel-normal if and only if the convex set</u> $\mathrm{cl}\, P(t)$ <u>depends Borel measurably on</u> T.

<u>Proof.</u> The second property is equivalent to the Borel-normality of the integrand

$$h(t,p) = \begin{cases} 0 & \text{if } p \in \mathrm{cl}\, P(t) \\ +\infty & \text{if } p \notin \mathrm{cl}\, P(t) \end{cases} \tag{3.5}$$

[2, Theorem 3]. But $h(t,\cdot)$ and $r_L(t,\cdot)$ are conjugate to each other by (3.2), and normality is known to be preserved under conjugacy [4], [2].

Of course, J_M is extended from $\mathcal{A}$ to $\mathcal{B}$ in the same manner. In analogy to (3.2) and Proposition 2, the corresponding function r_M satisfies

(3.6) $$r_M(t,w) = \sup\{w\cdot x \mid x \in X(t)\},$$

and it is a Borel-normal integrand if and only if $\operatorname{cl} X(t)$ depends Borel measurably on t.

For our purposes, the Borel measurability of the multifunctions

(3.7) $$t \to \operatorname{cl} X(t) \text{ and } t \to \operatorname{cl} P(t), \quad t_0 \leq t \leq t_1,$$

while ensuring that J_L and J_M are well-defined, is not a strong enough property for the results of real interest. We shall need to assume:

(S_1) <u>the multifunctions</u> (3.7) <u>are upper semicontinuous.</u>

The upper semicontinuity of $t \to \operatorname{cl} X(t)$ means that the graph $\{(t,x) \mid x \in \operatorname{cl} X(t)\}$ is closed in $[t_0,t_1] \times R^n$, or equivalently that the set

(3.8) $$\{t \in [t_0,t_1] \mid K \cap \operatorname{cl} X(t) \neq \emptyset\}$$

is closed for every compact $K \subset R^n$. Our main duality theorems will actually necessitate the <u>continuity</u> of the multifunctions. Continuity equals upper semicontinuity plus lower semicontinuity. <u>Lower semicontinuity</u> means that the set (3.8) is open relative to $[t_0,t_1]$ for every open $K \subset R^n$.

<u>Theorem 1.</u> <u>Assume</u> (S_1). <u>Then the extended-real-valued functionals</u> J_L <u>and</u> J_M <u>are well-defined on</u> $\mathcal{B}$ <u>and convex.</u> <u>Moreover, the inequality</u>

(3.9) $$J_L(x) + J_M(p) \geq x(t_1)\cdot p(t_1) - x(t_0)\cdot p(t_0)$$

(<u>with the convention</u> $+\infty - \infty = +\infty$ <u>if necessary</u>) <u>holds for all arcs</u> $x \in \mathcal{B}$ <u>and</u> $p \in \mathcal{B}$ <u>such that</u>

(3.10) $$x(t_0) \in \operatorname{cl} X(t_0) \text{ or } p(t_0) \in \operatorname{cl} P(t_0),$$

<u>and</u>

(3.11) $$x(t_1) \in \operatorname{cl} X(t_1) \text{ or } p(t_1) \in \operatorname{cl} P(t_1).$$

Proof. The upper semicontinuity of the multifunctions (3.7) implies their Borel measurability so r_L and r_M are Borel-normal by Proposition 2, and hence J_L and J_M are well-defined. We next show convexity. Let x^1 and x^2 be any two elements of $\mathcal{B}$ and let

$$x = (1-\lambda)x^1 + \lambda x^2 \text{ for } \lambda \in (0,1) .$$

Represent the singular measures $dx^i - \dot{x}^i dt$ in the form $\xi^i d\theta$, as is indeed possible with a common $d\theta$. Then $dx - \dot{x}dt = \xi d\theta$, where $\xi = (1-\lambda)\xi^1 + \lambda\xi^2$, and we get

$$J_L(x) \leq (1-\lambda) J_L(x^1) + \lambda J_L(x^2)$$

from the definition (3.3) and the convexity of $L(t,\cdot,\cdot)$ and $r_L(t,\cdot)$, keeping in mind the convention that $+\infty - \infty = +\infty$.

Of course, the same convention makes (3.9) trivial unless both $J_L(x) < +\infty$ and $J_M(p) < +\infty$, in which case the state constraints (1.5) and (1.15) must in particular be satisfied. The assumed upper semicontinuity of the state constraint multifunctions then implies

(3.12) $\quad x_+(t) \in \text{cl } X(t)$ and $p_+(t) \in \text{cl } P(t)$ for $t_0 \leq t < t_1$,

(3.13) $\quad x_-(t) \in \text{cl } X(t)$ and $p_-(t) \in \text{cl } P(t)$ for $t_0 < t \leq t_1$,

so that by (3.2) and (3.6) we have

(3.14) $$\text{for all } t \in [t_0,t_1),\ r_L(t,\xi(t)) \geq \xi(t)\cdot p_+(t) \text{ and } r_M(t,\pi(t)) \geq \pi(t)\cdot x_+(t),$$

(3.15) $$\text{for all } t \in (t_0,t_1],\ r_L(t,\xi(t)) \geq \xi(t)\cdot p_-(t), \text{ and } r_M(t,\pi(t)) \geq \pi(t)\cdot x_+(t),$$

where $\xi d\theta$ and $\pi d\theta$ represent the singular measures $dx - \dot{x}dt$ and $dp - \dot{p}dt$. The first equation in Proposition 1 can be expressed as

(3.16) $$x\cdot p\Big|_{t_0}^{t_1} = \int_{t_0}^{t_1}(x\cdot\dot{p} + \dot{x}\cdot p)dt + \int_{(t_0,t_1)}(x_+\cdot\pi + p_-\cdot\xi)d\theta + x_+(t_0)\cdot\Delta p(t_0) + p(t_0)\cdot\Delta x(t_0) + x(t_1)\cdot\Delta p(t_1) + p_-(t_1)\cdot\Delta x(t_1) .$$

We then have

$$(3.17)\quad J_L(x) + J_M(p) - x\cdot p\Big|_{t_0}^{t_1} = \int_{t_0}^{t_1}(L(t,x,\dot{x}) + M(t,p,\dot{p}) - x\cdot\dot{p} - \dot{x}\cdot p)dt$$

$$+\int_{(t_0,t_1)}(r_L(t,\xi) - \xi\cdot p_- + r_M(t,\pi) - \pi\cdot x_+)d\theta + c_0(x,p) + c_1(x,p),$$

where

$$(3.18)\qquad c_0(x,p) = r_L(t_0,\Delta x(t_0)) - \Delta x(t_0)\cdot p(t_0) + r_M(t_0,\Delta p(t_0)) - \Delta p(t_0)\cdot x_+(t_0)\ ,$$

$$(3.19)\qquad c_1(x,p) = r_L(t_1,\Delta x(t_1)) - \Delta x(t_1)\cdot p_-(t_1) + r_M(t_1,\Delta p(t_1)) - \Delta p(t_1)\cdot x(t_1)\ .$$

The two integrals on the right side of (3.17) are nonnegative by virtue of (3.14), (3.15) and the inequality

$$(3.20)\qquad L(t,x(t),\dot{x}(t)) + M(t,p(t),\dot{p}(t)) \geq x(t)\cdot\dot{p}(t) + \dot{x}(t)\cdot p(t)\ ,$$

which is a consequence of the definition (1.13) of M. If $x(t_0)\in \mathrm{cl}\, X(t_0)$ and $p(t_1)\in \mathrm{cl}\, P(t_1)$, we also have from (3.2) and (3.6) the nonnegativity of $c_0(x,p)$ and $c_1(x,p)$ in (3.17), so that (3.9) is true as claimed. The other cases where (3.10) and (3.11) are satisfied yield the same conclusion by consideration of the following expressions equivalent to (3.18) and (3.19):

$$(3.21)\qquad c_0(x,p) = r_L(t_0,\Delta x(t_0)) - \Delta x(t_0)\cdot p_+(t_0) + r_M(t_0,\Delta p(t_0)) - \Delta p(t_0)\cdot x(t_0)\ ,$$

$$(3.22)\qquad c_1(x,p) = r_L(t_1,\Delta x(t_1)) - \Delta x(t_1)\cdot p(t_1) + r_M(t_1,\Delta p(t_1)) - \Delta p(t_1)\cdot x_-(t_1)\ .$$

4. Extended Hamiltonian Condition.

A vector w is said to be normal to a convex set $C \subset R^n$ at the point C if for all $z'\in C$ one has $w\cdot(z'-z) \leq 0$ [7, p. 215]. If $z\in \mathrm{int}\, C$, this holds only for $w=0$.

We shall say that the arcs $x \in \mathcal{B}$ and $p \in \mathcal{B}$ satisfy the extended Hamiltonian condition if (1.12) holds,

(4.1) $x_+(t)$ and $x_-(t)$ belong to $\mathrm{cl}\, X(t)$ for all t ,

(4.2) $p_+(t)$ and $p_-(t)$ belong to $\mathrm{cl}\, P(t)$ for all t ,

and for any representation

(4.3) $$dx - \dot{x}\,dt = \xi d\theta \quad \text{and} \quad dp - \dot{p}dt = \pi d\theta$$

(where ξ and π are Borel measurable and $d\theta$ is nonnegative) it is true that

(4.4) $\pi(t)$ is normal to $\mathrm{cl}\, X(t)$ at $x_+(t)$ and $x_-(t)$ almost everywhere relative to $d\theta$,

(4.5) $\xi(t)$ is normal to $\mathrm{cl}\, P(t)$ at $p_+(t)$ and $p_-(t)$ almost everywhere relative to $d\theta$.

It is easily verified that the validity of (4.4) and (4.5) does not depend on the particular representation (4.3).

The following jump condition is implied by (4.4) (and equivalent to it if the singular part of dp is purely atomic):

(4.6) $\Delta p(t)$ is normal to $\mathrm{cl}\, X(t)$ at $x_+(t)$ and $x_-(t)$ for all $t \in [t_0, t_1]$.

On the other hand, if $x(t) \in \mathrm{int}\, X(t)$ for all t in some subinterval (a,b), we have by (4.4) that $\pi d\theta$ is the zero measure on (a,b), and therefore p is absolutely continuous over (a,b). Similar observations can be made about (4.5).

It is important that, in this way, conclusions about whether x or p must actually be absolutely continuous in certain cases can be drawn from the Hamiltonian conditions itself. For example, if there are no effective state constraints associated with L (i.e. $X(t) \equiv R^n$), the condition implies $p \in \mathcal{A}$, while dually if $P(t) \equiv R^n$ (a property amounting to a growth condition on L) then $x \in \mathcal{A}$.

Proposition 3. Assume (S_1). Then any pair of arcs $x \in \mathcal{A}$ and $p \in \mathcal{A}$ satisfying the Hamiltonian "equation" (1.12) also satisfies the extended Hamiltonian condition.

Proof. The subgradient set $\partial H(t, x(t), p(t))$ is empty unless $x(t) \in X(t)$ and $p(t) \in P(t)$ [7, Theorem 37.4]. Therefore (1.12) implies (1.5) and (1.16). From the upper semicontinuity in (3.7) and the continuity of x and p, we then have (4.1) and (4.2). Since (4.4) and (4.5) are trivial for arcs in $\mathcal{A}$, we conclude that x and p satisfy the extended Hamiltonian condition.

In order to describe the relationship between the extended Hamiltonian condition and optimality in problems of Lagrange, we introduce the functions

(4.7) $$f_L^{\mathcal{B}}(x_0, x_1) = \text{inf in problem (1.3) for } x \in \mathcal{B},$$

(4.8) $$f_M^{\mathcal{B}}(p_0, p_1) = \text{inf in problem (1.15) for } p \in \mathcal{B},$$

(4.9) $$f_L(x_0, x_1) = \begin{cases} f_L^{\mathcal{B}}(x_0, x_1) & \text{if } x_0 \in \operatorname{cl} X(t_0),\ x_1 \in \operatorname{cl} X(t_1), \\ +\infty & \text{otherwise} \end{cases}$$

(4.10) $$f_M(p_0, p_1) = \begin{cases} f_M^{\mathcal{B}}(p_0, p_1) & \text{if } p_0 \in \operatorname{cl} P(t_0),\ p_1 \in \operatorname{cl} P(t_1), \\ +\infty & \text{otherwise}. \end{cases}$$

Condition (S_1) is always assumed in this context.

The reason for introducing f_L in addition to $f_L^{\mathcal{B}}$ is that sometimes we shall want to restrict attention to arcs $x \in \mathcal{B}$ satisfying the terminal conditions $x_0 \in \operatorname{cl} X(t_0)$ and $x_1 \in \operatorname{cl} X(t_1)$, which do not follow from $J_L(x) < +\infty$ due to the possibility of jumps at t_0 and t_1. The next result sets forth some basic relationships between these optimal value functions.

Proposition 4. Suppose (S_1) holds. Then $f_L, f_L^{\mathcal{B}}, f_M$ and $f_M^{\mathcal{B}}$ are well-defined, extended-real-valued functions on $R^n \times R^n$ satisfying

$$(4.11)\qquad f_L(x_0,x_1) + f_M^{\mathcal{B}}(p_0,p_1) \geq x_1 \cdot p_1 - x_0 \cdot p_0 ,$$

$$(4.12)\qquad f_L^{\mathcal{B}}(x_0,x_1) + f_M(p_0,p_1) \geq x_1 \cdot p_1 - x_0 \cdot p_0 ,$$

$$(4.13)\qquad f_L^{\mathcal{B}}(x_0,x_1) = \inf_{x_0', x_1'} \{f_L(x_0',x_1') + r_L(t_0,x_0'-x_0) + r_L(t_1,x_1-x_1')\} ,$$

$$(4.14)\qquad f_M^{\mathcal{B}}(p_0,p_1) = \inf_{p_0', p_1'} \{f_M(p_0',p_1') + r_M(t_0,p_0'-p_0) + r_M(t_1,p_1-p_1')\}$$

(where the convention $\infty - \infty = +\infty$ is understood).

Proof. All but the last two formulas is apparent from Theorem 1. To demonstrate (4.13), we consider arbitrary pairs of arcs x and x' in $\mathcal{B}$ which agree on (t_0,t_1) but have possibly different endpoint pairs, denoted by (x_0,x_1) and (x_0',x_1'). Let $\tilde{x}$ be the arc which agrees with x and x' on (t_0,t_1) but is right-continuous at t_0 and left-continuous at t_1. Then by the definition (3.3) of J_L:

$$J_L(x) = J_L(\tilde{x}) + r_L(t_0,\tilde{x}(t_0) - x_0) + r_L(t_1,x_1 - \tilde{x}(t_1)) ,$$

$$J_L(x') = J_L(\tilde{x}) + r_L(t_0,\tilde{x}(t_0) - x_0') + r_L(t_1,x_1' - \tilde{x}(t_1)) .$$

But since $r_L(t_0,\cdot)$ and $r_L(t_1,\cdot)$ are convex and positively homogeneous, we have

$$r_L(t_0,\tilde{x}(t_0) - x_0') + r_L(x_0' - x_0) \geq r_L(t_0,\tilde{x}(t_0) - x_0) ,$$

$$r_L(t_1,x_1 - x_1') + r_L(t_1,x_1' - \tilde{x}(t_1)) \geq r_L(t_1,x_1 - \tilde{x}(t_1)) .$$

Therefore

$$(4.15)\qquad J_L(x) \leq J_L(x') + r_L(t_0,x_0' - x_0) + r_L(t_1, x_1 - x_1') ,$$

and, for this reason, at least $\leq$ must hold in (4.13). The truth of $=$ is seen from the fact that if $x' = \tilde{x}$ in the above and $J_L(x) < +\infty$, then equality holds in (4.15), and also $x_0 \in \mathrm{cl}\, X(t_0)$ and $x_1 \in \mathrm{cl}\, X(t_1)$ by the upper semicontinuity of the multifunction $t \to \mathrm{cl}\, X(t)$ and the automatic state constraint (1.5). The verification of (4.14) is parallel.

Corollary. Assuming (S_1) holds, one has $f_L^{\mathcal{B}} = f_L$ if

$$r_L(t_0,w) = r_L(t_1,w) = +\infty \text{ for all } w \neq 0 ,$$

or equivalently, if $P(t_0) = P(t_1) = R^n$.

Proposition 4 furnishes in particular the inequality

$$f_L(x_0,x_1) + f_M(p_0,p_1) \geq x_1 \cdot p_1 - x_0 \cdot p_0 . \tag{4.16}$$

For reasons clear from the next theorem, we shall say that (x_0,x_1) and (p_0,p_1) are endpoint pairs in duality for L and M if (4.16) is realized as an equation. (Note that then $f_L(x_0,x_1)$ and $f_M(p_0,p_1)$ must both be finite.) An arc $x \in \mathcal{B}$ will be called optimal for L if

$$J_L(x) = f_L(x(t_0),x(t_1)) \tag{4.17}$$

and extremal for L if there is an arc $p \in \mathcal{B}$ (called a coextremal for L corresponding to x) such that the extended Hamiltonian condition is satisfied.

The latter definitions are also applicable to the dual Lagrangian M, but extremality in this case refers to the Hamiltonian $\tilde{H}$ associated with M in the same way that H is associated with L :

$$\tilde{H}(t,p,x) = \sup_{s \in R^n} \{x \cdot s - M(t,p,s)\} , \tag{4.18}$$

$$M(t,p,s) = \sup_{x \in R^n} \{x \cdot s - \tilde{H}(t,p,x)\} . \tag{4.19}$$

These formulas, in conjunction with the previous four relating H, L and M, yield the fact that $\tilde{H}(t,p,\cdot)$ is for each t and p the biconjugate of the convex function $-H(t,\cdot,p)$, while $H(t,x,\cdot)$ is for each t and x the biconjugate of the convex function $-\tilde{H}(t,\cdot,x)$. This means that $-\tilde{H}(t,p,x)$ and $H(t,x,p)$ are equivalent "closed saddle-functions" of x and p in the sense studied in [7, §34]. Among the consequences of this equivalence and the fact (obvious from (1.17) and (4.18)) that

$$\tilde{H}(t,x,p) = -\infty \text{ if and only if } p \notin P(t) \tag{4.20}$$

are the already mentioned relation (1.18) and its dual

(4.21) $\qquad X(t) \subset \{x \in R^n \mid \tilde{H}(t,p,x) < +\infty\} \subset \mathrm{cl}\, X(t)$ for each $p \in P(t)$.

Another consequence [7, Corollary 34.2.1] is that

(4.22) $\qquad -\tilde{H}(t,p,x) = H(t,x,p)$ if $x \in \mathrm{int}\, X(t)$ or $p \in \mathrm{int}\, P(t)$.

If $P(t) = R^n$ for every t (i.e. $r_L(t,w)$ is always $+\infty$ for $w \neq 0$), one has $-\tilde{H}(t,p,x) = H(t,x,p)$ for all t,x,p. The same is true if $X(t) = R^n$ for every t (i.e. there are no state constraints).

Theorem 2. Assume (S_1). Let $x \in \mathcal{B}$ and $p \in \mathcal{B}$ be any pair of arcs such that $J_L(x)$ and $J_M(p)$ are not oppositely infinite. Then the following assertions are equivalent (and imply $J_L(x)$ and $J_M(p)$ are both actually finite):

(a) x is extremal for L with coextremal p,

(b) p is extremal for M with coextremal x,

(c) x is optimal for L, p is optimal for M, and the endpoint pairs $(x(t_0), x(t_1))$ and $(p(t_0), p(t_1))$ are in duality in the above sense.

Proof. The cited equivalence of $H(t,x,p)$ and $-\tilde{H}(t,p,x)$ as saddle-functions of (x,p) for each t implies that these functions have the same subgradients everywhere [7, Corollary 37.4.1]. In fact, the Hamiltonian "equation" (1.12) for H and the corresponding one for $\tilde{H}$ can both be expressed equivalently in the symmetric form:

(4.23) $\qquad L(t,x(t),\dot{x}(t)) + M(t,p(t),\dot{p}(t)) = x(t)\cdot\dot{p}(t) + \dot{x}(t)\cdot p(t) \qquad$ a.e.

[8]. In particular, (a) and (b) are equivalent.

In light of (3.6), the normality condition (4.4) is equivalent (assuming (4.1)) to

(4.24) $\qquad r_M(t,\pi(t)) = \pi(t)\cdot x_+(t) = \pi(t)\cdot x_-(t)$

almost everywhere relative to $d\theta$,

which at t_0 and t_1 entails

(4.25) $$r_M(t_0, \Delta p(t_0)) = \Delta p(t_0) \cdot x_+(t_0) = \Delta p(t_0) \cdot x(t_0) ,$$

(4.26) $$r_M(t_1, \Delta p(t_1)) = \Delta p(t_1) \cdot x(t_1) = \Delta p(t_1) \cdot x_-(t_1) .$$

Similarly from (3.2), the other normality condition (4.5) is equivalent (assuming (4.2)) to

(4.27) $$r_L(t, \xi(t)) = \xi(t) \cdot p_+(t) = \xi(t) \cdot p_-(t)$$
almost everywhere relative to $d\theta$,

which at t_0 and t_1 entails

(4.28) $$r_L(t_0, \Delta x(t_0)) = \Delta x(t_0) \cdot p_+(t_0) = \Delta x(t_0) \cdot p(t_0) ,$$

(4.29) $$r_L(t_1, \Delta x(t_1)) = \Delta x(t_1) \cdot p(t_1) = \Delta x(t_1) \cdot p_-(t_1) .$$

Now we refer back to the proof of Theorem 1, where relations like the preceding were crucial, but appeared in the form of inequalities rather than equations. Inspection of the argument given there shows that the extended Hamiltonian condition, as expressed by these equations, is in fact equivalent to having all of the alternatives in (3.10) and (3.11) satisfied and

(4.30) $$J_L(x) + J_M(p) = x(t_1) \cdot p(t_1) - x(t_0) \cdot p(t_0) .$$

In view of the general inequality (4.16) and the definitions of f_L and f_M , this means exactly that

(4.31) $$J_L(x) = f_L(x(t_0), x(t_1)) \text{ and } J_M(p) = f_M(p(t_0), p(t_1)) ,$$

(4.32) $$f_L(x(t_0), x(t_1)) + f_M(p(t_0), p(t_1)) = x(t_1) \cdot p(t_1) - x(t_0) \cdot p(t_0),$$

which is (c) .

(A central role in this proof is played by the identity (3.17), which however might fail in certain cases where terms involving both $+\infty$ and $-\infty$ might be present. In the proof of Theorem 1, we were able to reduce consideration to $J_L(x) < +\infty$ and $J_M(p) < +\infty$ in which event neither term in (3.3) or the corresponding formula for $J_M(p)$ can be $+\infty$, and (3.17) is valid. The same holds here in arguing from (c) to (a) and (b), since (c) can be expressed by (4.31) and (4.32), and these equations imply the finiteness of $J_L(x)$ and $J_M(p)$. In arguing in the other direction, we have at our disposal (4.24) and (4.27), which imply the finiteness of the "singular" integrals in the definition of $J_L(x)$ and $J_M(p)$. Furthermore we have (4.23), where the right side is summable. Thus the integrals

$$\int_{t_0}^{t_1} L(t, x(t), \dot{x}(t))dt \quad \text{and} \quad \int_{t_0}^{t_1} M(t, p(t), \dot{p}(t))dt$$

cannot be both $+\infty$ or both $-\infty$. Since the theorem assumes $J_L(x)$ and $J_M(p)$ are not oppositely infinite, these integrals must both be finite, so that (3.17) is again usable.)

Corollary 1. Assume (S_1). Let $x \in \mathcal{A}$ and $p \in \mathcal{A}$ be arcs satisfying the Hamiltonian "equation" (1.12) and such that $J_L(x)$ and $J_M(p)$ are not oppositely infinite. Then x is actually optimal for L in the present sense of arcs in $\mathcal{B}$, not just relative to other arcs in $\mathcal{A}$, and similarly p for M.

Proof. This follows via Proposition 3.

Corollary 2. Assume (S_1). In order that an arc $x \in \mathcal{B}$ be a solution to the Lagrange problem (1.3) for a given endpoint pair (x_0, x_1), it is sufficient that x be an extremal arc for L satisfying $x(t_0) = x_0$ and $x(t_1) = x_1$.

This condition is also necessary, if there exists an endpoint pair (p_0, p_1) which is in duality with (x_0, x_1) and such that the corresponding Lagrange problem (1.15) has a solution.

A deeper result on the necessity in Corollary 2 will be stated as Theorem 4 at the end of this paper.

5. Analysis of Some Further Regularity Conditions.

For the purpose of the results to be obtained in §6, some stronger assumptions on the Hamiltonian H and the behavior of the convex sets $X(t)$ and $P(t)$ will be required:

(S_2) $\quad \operatorname{int} X(t) \neq \emptyset$ and $\operatorname{int} P(t) \neq \emptyset$ for all t,

(S_3) for arbitrary t, $x \in \text{int}\, X(t)$ and $p \in \text{int}\, P(t)$, there is a relatively open subinterval of $[t_0, t_1]$ containing t over which $H(\cdot, x, p)$ is summable.

Despite appearances, (S_3) does not give special weight to the Lagrangian L and therefore call for the introduction, at some stage, of a corresponding assumption on the Hamiltonian $\tilde{H}$ associated with M. It is shown below, at the beginning of the proof of Proposition 5, that (S_3) in the presence of (S_1) and (S_2) is equivalent to a stronger property (S_3'). Moreover, (S_3') involves only the values of H over int X(t) and int P(t), where H and $-\tilde{H}$ agree according to (4.22), and thus can be viewed equally as a property of H or one of $\tilde{H}$. In assuming (S_1), (S_2) and (S_3), we therefore do not lose symmetry between L and M and so are justified in invoking the "principle of duality" as a tool in our proofs.

Proposition 5. Suppose (S_1), (S_2) and (S_3) hold. Then the multifunctions (3.7) are continuous. Furthermore, for any continuous function $\bar{x} : [t_0, t_1] \to R^n$ with $\bar{x}(t) \in \text{int}\, X(t)$ for all t (and such functions do exist), there are summable functions

$$\beta : [t_0, t_1] \to R, \quad b : [t_0, t_1] \to R^n, \quad B : [t_0, t_1] \to R^{n \times n}$$

and an $\varepsilon > 0$ such that

(5.1) $$|x - \bar{x}(t)| < \varepsilon \implies L(t, x, B(t)x + b(t)) \le \beta(t) .$$

The same property holds with respect to M for any continuous function $p : [t_0, t_1] \to R^n$ with $p(t) \in \text{int}\, P(t)$ for all t.

Proof. Our first step is to show that (S_1), (S_2) and (S_3) imply:

(S_3') For any elements $t \in [t_0, t_1]$, $x \in \text{int}\, X(t)$ and $p \in \text{int}\, P(t)$, there is a relatively open subinterval I of $[t_0, t_1]$, along with open neighborhoods U of x and V of P, such that

$$|H(t, x', p')| \le \rho(t) \text{ for all } t' \in I \;\; x' \in U \text{ and } p' \in V ,$$

where $\rho(t)$ is finite and summable in t.

Fixing t, x, p as described, we start by selecting $\delta > 0$ sufficiently small that

$$p \pm 2\delta e_j \in \operatorname{int} P(t) \quad \text{for } j = 1, \ldots, n ,$$

where

$$e_1 = (1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1) .$$

By (S_3), there is a relatively open interval I_0 containing t over which the functions $H(\cdot, x, p \pm 2\delta e_j)$ are all summable and hence in particular finite almost everywhere. Then $p \pm 2\delta e_j \in \operatorname{cl} P(t')$ for almost all $t' \in I_0$ by (1.18) and therefore actually for all $t' \in I_0$ by our upper semicontinuity assumption (S_1). It follows that $p \in \operatorname{int} P(t')$ and $p \pm \delta e_j \in \operatorname{int} P(t')$ for all $t' \in I_0$, so that the functions $H(\cdot, x, p)$ and $H(\cdot, x, p \pm \delta e_j)$ are all finite throughout I_0 and, again by virtue of (S_3), all summable with respect to some relatively open interval I_1 such that $t \in I_1 \subset I_0$. Let

$$V_1 = \operatorname{co} \{ p \pm \delta e_j, \ldots, p \pm \delta e_n \} . \tag{5.2}$$

Then V_1 is a neighborhood of p. For each $p' \in V_1$, we have by the convexity of $H(t', x, \cdot)$ that

$$H(t', x, p') \leq \max_{j=1}^{n} H(t', x, p \pm \delta e_j) \triangleq \rho_0(t') ,$$

and $\rho_0(t')$ is finite and summable in $t' \in I_1$. At the same time we have $p'' = p - (p' - p) \in V_1$ by the symmetry in (5.2), implying $H(t', x, p'') \leq \rho_0(t')$ on I_1. Since $p = \frac{1}{2}(p' + p'')$, it follows that

$$H(t', x, p) \leq \tfrac{1}{2}[H(t', x, p') + H(t', x, p'')] ,$$

so that

$$2H(t', x, p) - \rho_0(t') \leq H(t', x, p') \leq \rho_0(t')$$

for all $t' \in I_1$. Let

$$\rho_1(t') = \max \{ |\rho_0(t')| , |2H(t', x, p) - \rho_0(t')| \} .$$

Then

$$|H(t', x, p')| \leq \rho_1(t') \text{ for all } t' \in I_1 , \; p' \in V_1 , \tag{5.3}$$

where ρ_1 is finite and summable on I_1 and V_1 is a neighborhood of p.

The construction now proceeds in the second argument of H. Let $\delta' > 0$ be small enough that

$$x \pm \delta' e_j \in \text{int}\, X(t) \quad \text{for } j = 1,\dots,n .$$

The derivation of (5.7) is valid for any of the points $x \pm \delta' e_j$ in place of x, and we therefore have

$$(5.4) \qquad |H(t', x \pm \delta' e_j, p')| \le \rho_2(t'), \; j = 1,\dots,n, \quad \text{for all } t' \in I \text{ and } p' \in V,$$

where $V \subset V_1$ is some open neighborhood of p, $I \subset I_1$ is some relatively open interval containing t, and ρ_2 is finite and summable on I. Let

$$U = \text{int co}\,\{x \pm \delta' e_1, \dots, x \pm \delta' e_n\}.$$

Then U is an open neighborhood of x containing along with each of its points x' the point $x'' = x - (x' - x)$. The concavity of $H(t', \cdot, p')$ yields from (5.4) for each $x' \in U$, $p' \in V$, $t' \in I$, that

$$H(t', x', p') \ge \min_{j=1}^{n} H(t', x \pm \delta' e_j, p') \ge -\rho_2(t').$$

and simultaneously $H(t', x'', p') \ge -\rho_2(t')$. Since $x = \frac{1}{2}(x' + x'')$, we obtain

$$H(t', x, p') \ge \tfrac{1}{2}[H(t', x', p') + H(t', x'', p')]$$

and consequently

$$2H(t', x, p) - \rho_2(t') \ge H(t', x', p') \ge -\rho_2(t').$$

Thus for

$$\rho(t') = \max\{\rho_2(t'), |2H(t', x, p) - \rho_2(t')|\}$$

we have ρ finite and summable on I and

$$|H(t', x', p')| \le \rho(t') \text{ for all } t' \in I, \; x' \in U, \; p' \in V$$

as desired. This finishes the verification of (S_3').

An immediate consequence of (S_3') and (S_2) is that the interior of $\{(t,x) \mid x \in \text{cl}\, X(t)\}$ relative to $[t_0, t_1] \times R^n$ is $\{(t,x) \mid x \in \text{int}\, X(t)\}$. This implies (because $X(t)$ is convex) that the multifunction $t \to \text{cl}\, X(t)$ is lower semicontinuous [5, p. 458], hence by (S_1) continuous. Similarly, the multifunction $t \to \operatorname{cl} P(t)$ is continuous.

It follows from Michael's selection theorem [13] that there exist continuous functions $\bar{x} : [t_0, t_1] \to R^n$ satisfying $\bar{x}(t) \in \operatorname{int} X(t)$ for all t; similarly $\bar{p}(t) \in \operatorname{int} P(t)$. For any such $\bar{x}$ and $\bar{p}$ there exist by a compactness argument using (S_3') some $\bar{\varepsilon} > 0$ and summable function $\bar{\rho} : [t_0, t_1] \to [0, +\infty)$ such that

(5.5) $\quad |H(t, \bar{x}(t) + w, \bar{p}(t) + z)| \leq \bar{\rho}(t)$ whenever $|w| \leq \bar{\varepsilon}$, $|z| \leq \bar{\varepsilon}$.

Consider for each w with $|w| \leq \bar{\varepsilon}$ the functions

$$h_w(t, p) = H(t, \bar{x}(t) + w, p)$$

$$g_w(t, v) = L(t, \bar{x}(t) + w, v) .$$

In view of (1.9) and (1.10), $h_w(t, \cdot)$ and $g_w(t, \cdot)$ are convex functions conjugate to each other. Moreover $h_w(t, \cdot)$ is not identically infinite because of (5.5), so $h_w(t, \cdot)$ and $g_w(t, \cdot)$ are both "proper". Since L is Lebesgue-normal , it follows that g_w is a Lebesgue-normal proper convex integrand on $[t_0, t_1] \times R^n$ [2, Corollary 4.5], and therefore so is h_w [4, Lemma 5]. Property (5.9) then implies by [5, Theorem 2] the existence of at least one summable function $v : [t_0, t_1] \to R^n$ such that

$$+\infty > \int_{t_0}^{t_1} f_w(t, v(t))dt = \int_{t_0}^{t_1} L(t, \bar{x}(t) + w, v(t))dt .$$

We apply this now to a set of affinely independent points $w_0, w_1, \ldots, w_n$, such that

$$0 \in \operatorname{int} \operatorname{co} \{w_0, w_1, \ldots, w_n\} \text{ and } |w_i| \leq \bar{\varepsilon} \quad \text{for all } i ,$$

obtaining summable functions

$$v_i : [t_0, t_1] \to R^n \quad \text{and} \quad \beta_i : [t_0, t_1] \to R$$

such that

(5.6) $\qquad L(t, \bar{x}(t) + w_i, v_i(t)) \leq \beta_i(t) \quad$ for all t .

Let $\varepsilon > 0$ be such that

(5.7) $\qquad |w| \leq \varepsilon \implies w \in \operatorname{co} \{w_0, w_1, \ldots, w_n\}$.

For each t let $b(t) \in R^n$ and $B(t) \in R^{n \times n}$ be the unique vector and matrix with the property that

$$(5.8) \quad x - \bar{x}(t) = \sum_{i=0}^{n} \lambda_i w_i, \ \sum_{i=0}^{n} \lambda_i = 1 \implies B(t)x + b(t) = \sum_{i=0}^{m} \lambda_i v_i(t) .$$

The components of $b(t)$ and $B(t)$ are then summable in t. If $|x - \bar{x}(t)| < \varepsilon$, the coefficients λ_i in (5.8) are positive because of (5.7), and the convexity of $L(t,\cdot,\cdot)$ yields

$$L(t,x,B(t)x + b(t)) = L(t,\sum_{i=0}^{n} \lambda_i(\bar{x}(t) + w_i), \sum_{i=0}^{n} \lambda_i v_i(t))$$

$$\leq \sum_{i=0}^{n} \lambda_i L(t,\bar{x}(t) + w_i, v_i(t)) \leq \max_{i=0}^{n} \beta_i(t)$$

by (5.6). Thus (5.1) holds for the summable function $\beta = \max_{i=0}^{n} \beta_i$.

As for the corresponding result for the dual Lagrangian M, this follows by symmetry in view of the remarks made prior to the statement of Proposition 5. This completes the proof.

Proposition 6. Suppose (S_1), (S_2) and (S_3) hold. Then J_L and J_M nowhere have the value $-\infty$ on $\mathcal{B}$, and the assumption in Theorem 2 is thus satisfied for any pair of arcs $x \in \mathcal{B}$ and $p \in \mathcal{B}$.

Proof. According to Proposition 5, we can find a continuous function $\bar{x} : [t_0,t_1] \to R^n$ and summable functions $\bar{v} : [t_0,t_1] \to R^n$ and $\beta : [t_0,t_1] \to R$ such that $\bar{x}(t) \in X(t)$ and

$$L(t,\bar{x}(t),\bar{v}(t)) \leq \beta(t) \text{ for all } t .$$

We then have

$$M(t,p,s) \geq s \cdot \bar{x}(t) + p \cdot \bar{v}(t) - \beta(t)$$
$$\text{for all } t \in [t_0,t_1], \ p \in R^n, \ s \in R^n$$

by (1.13) and

$$r_M(t,w) \geq w \cdot \bar{x}(t) \text{ for all } t \in [t_0,t_1], \ w \in R^n$$

by (3.6). This implies for arbitrary $p \in \mathcal{B}$ that

$$J_M(p) \geq \int_{t_0}^{t_1} \bar{x}(t)dp(t) + \int_{t_0}^{t_1}[\bar{v}(t) \cdot p(t) - \beta(t)]dt > -\infty >$$

The corresponding property of J_L is true by symmetry.

Proposition 7. Suppose (S_1) and (S_2) hold. Then (S_3) is equivalent to the following. For any $t \in [t_0, t_1]$ and any $x \in \operatorname{int} X(t)$, there is a relatively open subinterval I containing t, along with summable functions $v : I \to R^n$ and $\beta : I \to R$ such that

$$L(t', x, v(t')) \leq \beta(t') \text{ for all } t' \in I. \tag{5.9}$$

Similarly, for any $t \in [t_0, t_1]$ and any $p \in \operatorname{int} P(t)$, there is a relatively open subinterval I containing t, along with summable functions $s : I \to R^n$ and $\gamma : I \to R$ such that

$$M(t', p, s(t')) \leq \gamma(t') \text{ for all } t' \in I. \tag{5.10}$$

Proof. The conditions are sufficient for (S_3), because (5.9) and (5.10) imply via (1.9) and (1.13) that

$$p \cdot v(t') - \beta(t') \leq H(t', x, p) \leq \gamma(t') - x \cdot s(t')$$

for all $t' \in I$, where the two outer expressions are summable. On the other hand, the necessity is shown by Proposition 5: since the multifunctions (3.7) are continuous, we can find for any t and $x \in \operatorname{int} X(t)$ a continuous function $\bar{x} : [t_0, t_1] \to R^n$ with $\bar{x}(t') \in \operatorname{cl} X(t')$ for all $t' \in [t_0, t_1]$ and $\bar{x}(t) = x$ (Michael's selection theorem [13]).

6. Optimal Value Functions and Duality.

In this section we explore the necessity of the extended Hamiltonian condition for optimality in Lagrange problems through the framework of Corollary 2 above. Thus we investigate the extent to which the fundamental inequality (4.16) for the convex "optimal value" functions f_L and f_M can hold as an equation, as well as the existence of solutions to the Lagrange problems over $\mathcal{B}$ in the definitions of these functions.

A byproduct of our technique is a simple condition under which the infimum in problem (1.3) or (1.15) is the same, whether taken over $\mathcal{A}$ or $\mathcal{B}$. This is obtained by a study of the relationship between f_L, f_M, $f_L^{\mathcal{B}}$, $f_M^{\mathcal{B}}$, and the optimal value functions

$$(6.1)\qquad f_L^{\mathcal{A}}(x_0,x_1) = \inf \text{ in } (1.3) \text{ for } x \in \mathcal{A},$$

$$(6.2)\qquad f_M^{\mathcal{A}}(p_0,p_1) = \inf \text{ in } (1.15) \text{ for } p \in \mathcal{A}.$$

These functions are likewise convex on $R^n \times R^n$, and under (S_1) they satisfy

$$(6.3)\qquad f_L^{\mathcal{A}} \geq f_L \geq f_L^{\mathcal{B}} \quad \text{and} \quad f_M^{\mathcal{A}} \geq f_M \geq f_M^{\mathcal{B}}.$$

(This is true because the state constraints (1.5) and (1.16) are satisfied whenever $J_L(x) < +\infty$ and $J_M(p) < +\infty$.)

The next theorem and its dual are our deepest results. In stating them, we recall the notion of the <u>effective domain</u> of an extended-real-valued convex function f, which is the convex set consisting of all the elements where the value of f is not $+\infty$. Of direct interest here are the effective domains of the various optimal value functions, which can also be expressed in the following manner as <u>sets of attainable endpoint pairs</u> relative to the Lagrangians L and M:

$$(6.4)\qquad \operatorname{dom} f_L^{\mathcal{B}} = \{(x_0,x_1) \in R^n \times R^n \mid \exists x \in \mathcal{B} \text{ with } J_L(x) < +\infty,\ x(t_0) = x_0,\ x(t_1) = x_1\},$$

$$(6.5)\qquad \operatorname{dom} f_L = \{(x_0,x_1) \in \operatorname{cl} X(t_0) \times \operatorname{cl} X(t_1) \mid \exists x \in \mathcal{B} \text{ with } J_L(x) < +\infty,\ x(t_0) = x_0,\ x(t_1) = x_1\},$$

$$(6.6)\qquad \operatorname{dom} f_L^{\mathcal{A}} = \{(x_0,x_1) \in \operatorname{cl} X(t_0) \times \operatorname{cl} X(t_1) \mid \exists x \in \mathcal{A} \text{ with } J_L(x) < +\infty,\ x(t_0) = x_0,\ x(t_1) = x_1\},$$

and analogously for $f_M^{\mathcal{B}}$, f_M and $f_M^{\mathcal{A}}$. We recall also the notation for a convex set C that

$$\operatorname{ri} C = \text{interior of } C \text{ relative to its affine hull.}$$

<u>Theorem 3. Suppose</u> (S_1), (S_2) <u>and</u> (S_3) <u>hold, and there is an arc</u> $\bar{x} \in \mathcal{A}$ <u>with</u> $J_L(\bar{x}) < +\infty$ <u>and</u> $\bar{x}(t) \in \operatorname{int} X(t)$ <u>for all</u> t. <u>Then</u>

(a) $f_L(x_0,x_1) = f_L^{\mathcal{A}}(x_0,x_1)$ <u>for all</u> $(x_0,x_1) \in \operatorname{ri} \operatorname{dom} f_L$, <u>and in particular</u>

$$(6.7)\qquad \operatorname{ri} \operatorname{dom} f_L = \operatorname{ri} \operatorname{dom} f_L^{\mathcal{A}} \subset \operatorname{dom} f_L^{\mathcal{A}} \subset \operatorname{dom} f_L.$$

(b) <u>For every</u> $(p_0, p_1) \in R^n \times R^n$, <u>there exists an arc</u> $p \in \mathcal{B}$ <u>furnishing the minimum in the Lagrange problem</u> (1.15).

(c) f_M <u>and</u> $f_M^{\mathcal{B}}$ <u>are lower semicontinuous.</u> <u>For every</u> $(p_0, p_1) \in R^n \times R^n$, <u>the minimum in</u> (4.14) <u>is attained and</u>

$$f_M^{\mathcal{B}}(p_0, p_1) = \sup_{x_0, x_1} \{x_1 \cdot p_1 - x_0 \cdot p_0 - f_L(x_0, x_1)\} , \tag{6.8}$$

$$f_M(p_0, p_1) = \sup_{x_0, x_1} \{x_1 \cdot p_1 - x_0 \cdot p_0 - f_L^{\mathcal{B}}(x_0, x_1)\} . \tag{6.9}$$

<u>Theorem 3'.</u> <u>Suppose</u> (S_1), (S_2) <u>and</u> (S_3) <u>hold, and there is an arc</u> $\bar{p} \in \mathcal{A}$ <u>with</u> $J_M(\bar{p}) < +\infty$ <u>and</u> $\bar{p}(t) \in \operatorname{int} P(t)$ <u>for all</u> t. <u>Then</u>

(a) $f_M(p_0, p_1) = f_M^{\mathcal{A}}(p_0, p_1)$ <u>for all</u> $(p_0, p_1) \in \operatorname{ri} \operatorname{dom} f_L$, <u>and in particular</u>

$$\operatorname{ri} \operatorname{dom} f_M = \operatorname{ri} \operatorname{dom} f_M^{\mathcal{A}} \subset \operatorname{dom} f_M^{\mathcal{A}} \subset \operatorname{dom} f_M . \tag{6.10}$$

(b) <u>For every</u> $(x_0, x_1) \in R^n \times R^n$, <u>there exists an arc</u> $x \in \mathcal{B}$ <u>furnishing the minimum in the Lagrange problem</u> (1.3).

(c) f_L <u>and</u> $f_L^{\mathcal{B}}$ <u>are lower semicontinuous.</u> <u>For every</u> $(x_0, x_1) \in R^n \times R^n$, <u>the minimum in</u> (4.13) <u>is attained and</u>

$$f_L^{\mathcal{B}}(x_0, x_1) = \sup_{p_0, p_1} \{x_1 \cdot p_1 - x_0 \cdot p_0 - f_M(p_0, p_1)\} , \tag{6.11}$$

$$f_L(x_0, x_1) = \sup_{p_0, p_1} \{x_1 \cdot p_1 - x_0 \cdot p_0 - f_M^{\mathcal{B}}(p_0, p_1)\} . \tag{6.12}$$

<u>Proofs.</u> Theorem 3 and 3' are equivalent, in view of the symmetry of our assumptions (cf. the remarks preceding Proposition 5), and therefore it will be enough to prove Theorem 3. In fact, almost everything can be reduced to establishing a single relation: that for any (p_0, p_1) there exists $p \in \mathcal{B}$ such that

$$J_M(p) \le \sup_{x_0, x_1} \{x_1 \cdot p_1 - x_0 \cdot p_0 - f_L^{\mathcal{A}}(x_0, x_1)\} . \tag{6.13}$$

This immediately yields part (b) of Theorem 3 and formula (6.8) in (c), because of the basic inequalities (4.11) and (6.3). It also tells us that the right sides of (6.8) and (6.13) are equal. Thus the convex functions

f_L and f_L^a have the same conjugate.

The latter implies part (a) of Theorem 3, at least if $f_{L'}$ does not have the value $-\infty$ anywhere, and even in that case if it can be verified by some other means that $\operatorname{dom} f_L$ and $\operatorname{dom} f_L^a$ have the same closure. However, this can indeed be accomplished by applying result already described to $L_0 = \max\{L, 0\}$, which (we claim) also satisfies the hypotheses, and using the obvious fact that f_{L_0} and $f_{L_0}^a$ are nonnegative with

$$\operatorname{dom} f_{L_0} = \operatorname{dom} f_L \quad \text{and} \quad \operatorname{dom} f_{L_0}^a = \operatorname{dom} f_L^a .$$

As for L_0 satisfying our hypotheses, it is at least clear that L_0 is another Lebesgue-normal integrand which is convex and proper and has $X_0(t) = X(t)$ for all t. (We mark all objects associated with L_0 instead of L by a subscript 0.) By duality, $M_0(t,\cdot,\cdot)$ is the "closed convex hull" of $M(t,\cdot,\cdot)$ and the indicator function of the origin $(0,0)$ [7, Theorem 16.5], i.e. the lower semicontinuous hull of the function

$$(p,s) \to \inf_{0<\lambda\leq 1} \lambda M(t, \lambda^{-1}p, \lambda^{-1}s) .$$

Therefore

$$\operatorname{cl} P_0(t) = \operatorname{cl} \operatorname{co} \{P(t), 0\} ,$$

and the properties of $X(t)$ and $P(t)$ in (S_1), (S_2) and Proposition 5 carry over to $X_0(t)$ and $P_0(t)$. The fact that (S_3) carries over is seen easily from the equivalent formulation of (S_3) in Proposition 7 and the above description of M_0.

The only things that will not follow from establishing (6.13), but need a supplementary argument, are the assertions in (c) of Theorem 3 besides (6.8). The lower semicontinuity of f_M^β is a consequence of (6.8) itself, and that of f_M is then trivial from the definition (4.9). Formula (6.9) is easy to derive from (6.8), (4.13) and the fact that (3.2) implies

$$\sup_{w \in R^n} \{w \cdot z - r_L(t,z)\} = \begin{cases} 0 & \text{if} \quad w \in \operatorname{cl} P(t) , \\ +\infty & \text{if} \quad w \notin \operatorname{cl} P(t) \end{cases}$$

[7, Theorem 13.1]. Finally, the attainment of the minimum in (4.14) can be deduced in the following manner. Consider the convex functions

$$g_1(x_0,x_1) = f_L^{\mathcal{B}}(-x_0,x_1)$$

$$g_2(x_0,x_1) = \begin{cases} 0 & \text{if } -x_0 \in \operatorname{cl} X(t_0),\ x_1 \in \operatorname{cl} X(t_1), \\ +\infty & \text{otherwise.} \end{cases}$$

The conjugates are $g_1^* = f_M$ by (6.9), $(g_1+g_2)^* = f_M^{\mathcal{B}}$ by (6.8), and

$$g_2^*(p_0,p_1) = r_M(t_0,-p_0) + r_M(t_1,p_1)$$

by (3.6). Furthermore, the hypothesis of Theorem 3 concerning $\bar{x}$ implies that

$$\operatorname{dom} g_1 \cap \operatorname{int} \operatorname{dom} g_2 \neq \emptyset .$$

The theorem about the conjugate of a sum of convex functions [7, Theorem 16.4] tells us then that

$$(g_1 + g_2)^*(p_0,p_1) = \min_{p'_0,p'_1} \{g_1^*(p'_0,p'_1) + g_2^*(p_0-p'_0,p_1-p'_1)\} ,$$

and this is identical to our assertion about (4.14).

The task now is to verify for an arbitrary pair (p_0,p_1) that (6.13) holds for some $p \in \mathcal{B}$. We can assume the right side of (6.13) is not $+\infty$, since otherwise the inequality holds trivially for all $p \in \mathcal{B}$. Let C be the Banach space of all continuous functions $y:[t_0,t_1] \to R^n$, and for each $y \in C$ let

$$(6.14) \qquad \varphi_L(y) = \inf_{x \in \mathcal{A}} \{x(t_0)\cdot p_0 - x(t_1)\cdot p_1 + \int_{t_0}^{t_1} L(t,x(t)+y(t),\dot{x}(t))dt\}.$$

Then φ_L is an extended-real-valued convex function on C such that

$$(6.15) \qquad \varphi_L(0) = \inf_{x_0,x_1} \{x_0\cdot p_0 - x_1\cdot p_1 + f_L(x_0,x_1)\} > -\infty .$$

We demonstrate that φ_L is bounded above on neighborhood of the origin in C. Let $\bar{x}$ have the properties assumed in Theorem 3 and take B, b,β and ε as in Proposition 5. Define

$$F : [t_0,t_1]\times R^n \to R^n$$

as follows:

$$(6.16)\qquad F(t,x) = \dot{\bar{x}}(t) \text{ if } x = \bar{x}(t) ,$$

whereas if $x \neq \bar{x}(t)$ and thus can be written in the form

$$x = (1 - \lambda(t))\bar{x}(t) + \lambda(t)\xi(t)$$

with

$$\lambda(t) = \varepsilon^{-1}|x - \bar{x}(t)| > 0 ,$$

$$\xi(t) = \bar{x}(t) + \lambda(t)^{-1}(x - \bar{x}(t)) ,$$

then

$$F(t,x) = (1 - \lambda(t))\dot{\bar{x}}(t) + \lambda(t)[B(t)\xi(t) + b(t)] .$$

In direct terms, but more opaquely,

$$(6.17)\qquad F(t,x) = B(t)(x-\bar{x}(t)) + \dot{\bar{x}}(t) + \varepsilon^{-1}|x-\bar{x}(t)|[B(t)\bar{x}(t) + b(t)-\dot{\bar{x}}(t)] .$$

For any t such that $|x - \bar{x}(t)| < \varepsilon$, we have (5.1) holding and also $\lambda(t) < 1$, so that by convexity of $L(t,\cdot,\cdot)$:

$$\begin{aligned} L(t,x,F(t,x)) &\leq (1-\lambda(t))L(t,\bar{x}(t),\dot{\bar{x}}(t)) \\ &\qquad + \lambda(t)\, L(t,\xi(t),B(t)\xi(t) + b(t)) \\ &\leq \max\{L(t,\bar{x}(t),\dot{\bar{x}}(t)), \beta(t)\} . \end{aligned}$$

Thus, denoting the last expression by $\alpha(t)$, we have $\alpha(t)$ summable in $t \in [t_0,t_1]$ and

$$(6.18)\qquad |x - \bar{x}(t)| < \varepsilon \Longrightarrow L(t,x,F(t,x)) \leq \alpha(t) .$$

Observe from (6.17) that $F(t,x)$ is Lipschitz continuous in $x \in R^n$ and summable in t. The differential equation

$$(6.19)\qquad \dot{x}(t) = F(t,x(t) + y(t)) \quad \text{a.e.,} \quad x(t_0) = x_0 ,$$

thus has a unique solution $x \in \mathcal{A}$ for any $x_0 \in R^n$ and $y \in \mathcal{C}$. Denote by x^y the solution to (6.20) corresponding to $x_0 = \bar{x}(t_0)$. If $\|(x^y + y) - \bar{x}\| < \varepsilon$ in the norm of $\mathcal{C}$, we have

$$L(t,x^y(t) + y(t), \dot{x}^y(t)) \leq \alpha(t) \text{ for all } t$$

by (6.18) and therefore

(6.20) $$\varphi_L(y) \leq \int_{t_0}^{t_1} \alpha(t)dt < +\infty .$$

The mapping $y \to x^y$ is continuous (as a transformation from $\mathcal{C}$ into $\mathcal{C}$), and for $y = 0$ we have $x^y = \bar{x}$ by (6.16). Hence there exists $\delta > 0$ such that

$$\|y\| < \delta \implies \|(x^y + y) - \bar{x}\| < \varepsilon .$$

Then (6.20) holds for all $y \in \mathcal{C}$ satisfying $\|y\| < \delta$, and φ_L is bounded above on a neighborhood of the origin in $\mathcal{C}$, as claimed.

The latter fact, with φ_L convex and $\varphi_L(0) > -\infty$ (by (6.15)), implies the existence of a subgradient of φ_L at 0, i.e. a continuous linear functional ψ on $\mathcal{C}$ such that

$$\varphi_L(y) \geq \varphi_L(0) + \psi(y) \quad \text{for all } y \in \mathcal{C} ,$$

or equivalently

(6.21) $$\sup_{y \in \mathcal{C}} \{\psi(y) - \varphi_L(y)\} = -\varphi_L(0) .$$

We can represent

(6.22) $$\psi(y) = \int_{t_0}^{t_1} y(t)dp(t), \quad \text{where } p \in \mathcal{B},\ p(t_0) = p_0 .$$

Combining (6.21) with (6.15), we see that

(6.23) $$\sup_{y \in \mathcal{C}} \{ \int_{t_0}^{t_1} y(t)dp(t) - \varphi_L(y)\} < +\infty ,$$

and that to establish (6.13) it will suffice to show this implies $p(t_1) = p_1$ and the supremum equals $J_M(p)$.

It is evident from the definition (6.14) of φ_L that the supremum in (6.23) is the same as the supremum over all $x \in \mathcal{A}$ and $y \in \mathcal{C}$ of the expression

$$x(t_1) \cdot p_1 - x(t_0) \cdot p(t_0) + \int_{t_0}^{t_1} y(t)dp(t) - \int_{t_0}^{t_1} L(t, x(t)+y(t), \dot{x}(t))dt ,$$

or equivalently for $z = x+y$, the supremum over all $x \in \mathcal{A}$ and $z \in \mathcal{C}$ of

$$x(t_1)\cdot p_1 - x(t_0)\cdot p(t_0) + \int_{t_0}^{t_1}[z(t)-x(t)]dp(t) - \int_{t_0}^{t_1} L(t,z(t),\dot{x}(t))dt$$

$$= x(t_1)\cdot(p_1 - p(t_1)) + \int_{t_0}^{t_1} z(t)dp(t) + \int_{t_0}^{t_1} p(t)\cdot\dot{x}(t)dt - \int_{t_0}^{t_1} L(t,z(t),\dot{x}(t))dt.$$

Thus it is the supremum of

$$x_1\cdot(p_1 - p(t_1)) + \int_{t_0}^{t_1} z(t)dp(t) + \int_{t_0}^{t_1} p(t)\cdot v(t)dt - \int_{t_0}^{t_1} L(t,z(t),v(t))dt$$

over all $x_1 \in R^n$, $z \in \mathcal{C}$ and $v \in \mathcal{L}^1 = \mathcal{L}^1([t_0,t_1], R^n)$. This would be $+\infty$ if $p_1 - p(t_1) \neq 0$, so we may conclude from (6.23) that $p(t_1) = p_1$, and the problem is reduced to proving that the expression

(6.24) $$\sup_{z\in\mathcal{C}} \{\int_{t_0}^{t_1} z(t)dp(t) + \sup_{v\in\mathcal{L}^1}\{\int_{t_0}^{t_1} p(t)\cdot v(t)dt - \int_{t_0}^{t_1} L(t,z(t),v(t))dt\}\},$$

equals $J_M(p)$, under the assumption it is not $+\infty$.

The theory of integral functionals and normal integrands will be applied in two steps to calculate (6.24). Let

(6.25) $$\mathcal{Z} = \{z \in \mathcal{C} \mid \exists\, v \in \mathcal{L}^1 \text{ with} \int_{t_0}^{t_1} L(t,z(t),v(t))dt < +\infty\}.$$

(Recall our convention that the integral is $+\infty$ if and only if $L(\cdot,z(.),v(\cdot))$ is not majorized by any summable function.) Note that $\mathcal{Z}$ is convex and

(6.26) $$z \in \mathcal{Z} \Longrightarrow \begin{cases} z(t) \in X(t) & \text{for almost } t, \\ z(t) \in \operatorname{cl} X(t) & \text{for all } t. \end{cases}$$

Proposition 5 implies then that

(6.27) $$\operatorname{int}\mathcal{Z} = \{z \in \mathcal{C} \mid z(t) \in \operatorname{int} X(t) \text{ for all } t\},$$

(6.28) $$\operatorname{cl}\mathcal{Z} = \{z \in \mathcal{C} \mid z(t) \in \operatorname{cl} X(t) \text{ for all } t\}.$$

If $z \notin \mathcal{Z}$, then the inner supremum in (6.24) is $-\infty$. Thus $\mathcal{C}$ can be replaced by $\mathcal{Z}$ in (6.24).

For each $z \in \mathcal{Z}$, the inner supremum in (6.24) is not $-\infty$ in view of (6.25), nor is it $+\infty$, since the overall supremum in (6.24) is assumed not to be $+\infty$. Thus it is finite. Fix any $z \in \mathcal{Z}$ and let

$$T = \{t \in [t_0, t_1] \mid z(t) \in X(t)\},$$
$$\ell(t,v) = L(t,z(t),v) \text{ for } t \in T, \ v \in R^n .$$

The complement of T in $[t_0, t_1]$ is a null set by (6.26) while $\ell(t,\cdot)$ is for each $t \in T$ a proper convex function on R^n by the definition of $X(t)$. The Lebesgue normality of L on $[t_0, t_1] \times R^n \times R^n$ implies the Lebesgue normality of ℓ as an integrand on $T \times R^n$ [2, Corollary 4.5], and the inner supremum in (6.24) can be written as

(6.29) $$\sup_{v \in \mathfrak{L}^1} \{\int_T p(t) \cdot v(t)dt - \int_T \ell(t,v(t))dt\},$$

where p can be regarded as an element of $\mathfrak{L}^\infty$. According to a fundamental theorem on conjugate integral functionals, (6.29) equals

(6.30) $$\int_T [\sup_{v \in R^n} \{p(t) \cdot v - \ell(t,v)\}]dt .$$

But the supremum in (6.30) is $H(t,z(t),p(t))$ by (1.9). Putting everything together, we now see that

(6.31) $$\sup_{v \in \mathfrak{L}^1} \{\int_{t_0}^{t_1} p(t) \cdot v(t)dt - \int_{t_0}^{t_1} L(t,z(t),v(t))dt\} = \int_{t_0}^{t_1} H(t,z(t),p(t))dt ,$$

where $H(\cdot,z(\cdot),p(\cdot))$ is summable over $[t_0, t_1]$. Moreover, this is true for arbitrary $z \in \mathscr{Z}$. The summability implies

(6.32) $p(t) \in P(t)$ for almost every t.

(Consider $z \in \operatorname{int} \mathscr{Z}$ as in (6.27) and apply (4.20) and (4.22).) We conclude further that the functional

(6.33) $$\Phi(z) = \begin{cases} -\int_{t_0}^{t_1} H(t,z(t),p(t))dt & \text{if } z \in \mathscr{Z} \\ +\infty & \text{if } z \notin \mathscr{Z} \end{cases}$$

is finite on $\mathscr{Z}$, and that (6.24) can be expressed as

(6.34) $$\sup_{z \in \mathcal{C}} \{\int_{t_0}^{t_1} z(t)dp(t) - \Phi(z)\}.$$

Observe that Φ is actually a convex functional, due to the convexity of $\mathcal{Z}$ and the concavity of $H(t,\cdot,p(t))$ for each t. Furthermore, for any $\bar{x} \in \operatorname{int} \mathcal{Z}$ (cf. (6.27)) we can apply Proposition 5 to get a bound

$$L(t,z(t),B(t)z(t) + b(t)) \leq \beta(t)$$

valid when $\|z - \bar{x}\| < \varepsilon$, and then

$$\int_{t_0}^{t_1} H(t,z(t),p(t))dt \geq \int_{t_0}^{t_1} (p(t) \cdot [B(t)z(t) + b(t)] - \beta(t))dt$$

when $\|z - \bar{z}\| < \varepsilon$, implying that Φ is uniformly bounded above on a neighborhood of $\bar{x}$. Therefore, in view of convexity, Φ is continuous on $\operatorname{int} \mathcal{Z}$.

At this point we employ again the Hamiltonian $\tilde{H}$ corresponding to M (cf. (4.18) and (4.19)) and define in terms of it the integrand

$$h(t,z) = \begin{cases} \tilde{H}(t,p(t),z) & \text{if } t \in S \\ 0 & \text{if } t \notin S \text{ but } z \in \operatorname{cl} X(t), \\ +\infty & \text{if } t \notin S \text{ and } z \notin \operatorname{cl} X(t), \end{cases} \tag{6.35}$$

where

$$S = \{t \in [t_0,t_1] \mid p(t) \in P(t)\}. \tag{6.36}$$

It should be borne in mind that the complement of S in $[t_0,t_1]$ is of measure zero (cf. (6.32)). Also,

$$h(t,z) = -H(t,p(t),z) \text{ if } t \in S \text{ and } z \in \operatorname{int} X(t), \tag{6.37}$$

by (4.22), while

$$h(t,z) = +\infty \quad \text{if} \quad z \notin \operatorname{cl} X(t) \tag{6.38}$$

by (4.21). Thus for $z \in \operatorname{int} \mathcal{Z}$ we have

$$h(t,z(t)) = -H(t,p(t),z(t)) \quad \text{a.e.}$$

(a summable function, but for $z \notin \operatorname{cl} \mathcal{Z}$ there exists a subinterval (a,b) of $[t_0,t_1]$ where $z(t) \notin \operatorname{cl} X(t)$ (due to the continuous dependence of $\operatorname{cl} X(t)$ on t; Proposition 5), and hence $h(t,z(t))$ is not majorized by any summable function of t. This shows that

(6.39) $$\Phi(z) = \int_{t_0}^{t_1} h(t, z(t))dt \text{ if } z \in \operatorname{int} \mathcal{J} \text{ or } z \notin \operatorname{cl} \mathcal{J} .$$

We know from its definition (6.33) that $h(t,\cdot)$ is for each $t \in [t_0, t_1]$ a lower semicontinuous proper convex function on R^n. The conjugate integrand

(6.40) $$m(t,s) = \sup_{z \in R^n} \{s \cdot z - h(t,z)\}$$

thus likewise has $m(t,\cdot)$ lower semicontinuous, convex and proper for every $t \in [t_0, t_1]$, and indeed

(6.41) $$m(t,s) = \begin{cases} M(t,p(t),s) & \text{if } t \in S \\ r_M(t,s) & \text{if } t \notin S \end{cases}$$

by (4.19) and (3.6). The normality of the integrands M and r_M yields through this the Lebesgue-normality of m [2, Corollary 4.5]. Since h and m are conjugate to each other, it follows that h too is Lebesgue-normal [4, Lemma 5]. The integral functional

(6.42) $$I_h(z) = \int_{t_0}^{t_1} h(t, z(t))dt \quad \text{for } z \in \mathcal{C}$$

is therefore well defined and convex. By virtue of (6.39), it agrees with Φ on int $\mathcal{J}$ (where Φ is finite and continuous) as well as outside of cl $\mathcal{J}$ (where Φ is identically $+\infty$). Hence the supremum (6.34), which we want to prove equal to $J_M(p)$, is the same as

(6.43) $$\sup_{z \in \mathcal{C}} \left\{ \int_{t_0}^{t_1} z(t)dp(t) - \int_{t_0}^{t_1} h(t,z(t))dt \right\} .$$

We are going to apply to (6.43) a theorem of [5] on the conjugates of integral functionals on spaces of continuous functions. This requires noting some further properties of h that result from its definition and the facts already established. Firstly, we have

(6.44) $$X(t) \subset \operatorname{dom} h(t,\cdot) \subset \operatorname{cl} X(t) \text{ for all } t$$

according to (6.35) and (4.21). Thus cl dom $h(t,\cdot)$ depends continuously on t,

(6.45) $$\operatorname{int} \operatorname{dom} h(t,\cdot) = \operatorname{int} X(t) \neq \emptyset ,$$

and by (3.6)

(6.46) $$\sup\{w \cdot z \mid z \in \operatorname{dom} h(t,\cdot)\} = r_M(t,w) .$$

Secondly, if $x \in R^n$ belongs to the interior (6.45) for all t in some subinterval $[a,b]$, then $h(\cdot,x)$ is summable over $[a,b]$. This follows because of the existence of a function $z \in \operatorname{int} \mathscr{Z}$ such that $z(t) \equiv x$ for $t \in [a,b]$ (cf. the result on extensions of continuous selections stated as Proposition 3 of [14]); the integral (6.42) has been shown to be finite for $z \in \operatorname{int} \mathscr{Z}$.

All the conditions needed in invoking [5, Theorem 5] for the functional (6.42) are met, and the consequence is that (6.43) can be identified via (6.40) and (6.46) with

(6.47) $$\int_{t_0}^{t_1} m(t,\dot p(t))dt + \int_{t_0}^{t_1} r_M(t,\pi(t))d\theta(t) ,$$

where $dp - \dot p dt = \pi d\theta$. Since in (6.41) we have $t \in S$ for almost every $t \in [t_0,t_1]$, this expression coincides with $J_M(p)$, and Theorem 3 has been proved.

<u>Theorem 4.</u> <u>Assume the hypothesis of Theorem</u> 3, <u>and let</u> (x_0,x_1) <u>be any endpoint pair which is strongly attainable for</u> L <u>in the sense of belonging to the set</u>

(6.48) $$(\operatorname{ri} \operatorname{dom} f_L^{\mathscr{B}}) \cap (\operatorname{cl} X(t_0) \times \operatorname{cl} X(t_1)) \supset \operatorname{ri} \operatorname{dom} f_L .$$

<u>Then an arc</u> $x \in \mathscr{B}$ <u>with</u> $x(t_0) = x_0$ <u>and</u> $x(t_1) = x_1$ <u>solves the Lagrange problem</u> (1.3) <u>for</u> L <u>over</u> $\mathscr{B}$ <u>if and only if it is extremal for</u> L, <u>i.e. satisfies the extended Hamiltonian condition for some</u> $p \in \mathscr{B}$.

<u>Proof.</u> The inclusion (6.48) is valid because the hypothesis of Theorem 3 concerning $\bar x$ implies

$$\operatorname{dom} f_L^{\mathscr{B}} \cap (\operatorname{int} X(t_0) \times \operatorname{int} X(t_1)) \neq \emptyset$$

and consequently by [7, Theorem 6.5] that

$$(\operatorname{ri} \operatorname{dom} f_L^{\mathscr{B}}) \cap (\operatorname{int} X(t_0) \times \operatorname{int} X(t_0))$$
$$= \operatorname{ri}[(\operatorname{dom} f_L^{\mathscr{B}}) \cap (\operatorname{cl} X(t_0) \times \operatorname{cl} X(t_1))] = \operatorname{ri} \operatorname{dom} f_L .$$

For the rest, we observe from Corollary 2 (at the end of §4) and T 3(b) that only the existence of an endpoint pair (p_0, p_1) in duality (x_0, x_1) needs to be established. The convex function f_L^{β} is subdifferentiable throughout ri dom f_L^{β} [7, Theorem 23.4] and therefore has at (x_0, x_1) a subgradient which can be expressed as $(-p_0, p_1)$:

$$f_L^{\beta}(x_0', x_1') \geq f_L^{\beta}(x_0, x_1) - p_0 \cdot (x_0' - x_1) + p_1 \cdot (x_1' - x_1) \text{ for all } x_0', x_1',$$

or equivalently

$$x_1 \cdot p_1 - x_0 \cdot p_0 - f_L^{\beta}(x_0, x_1) = \sup_{x_0', x_1'} \{x_1' \cdot p_1 - x_0' \cdot p_0 - f_L^{\beta}(x_0', x_1')\}. \tag{6.49}$$

But $f_L(x_0, x_1) = f_L^{\beta}(x_0, x_1) < +\infty$ by our assumption on (x_0, x_1), so (6.49) can be written in accordance with Theorem 3(c) as

$$-\infty < x_1 \cdot p_1 - x_0 \cdot p_0 - f_L(x_0, x_1) = f_M(p_0, p_1). \tag{6.50}$$

If $f_L(x_0, x_1) = -\infty$, there can be no solution to the Lagrange problem (1.3), inasmuch as J_L nowhere has the value $-\infty$ (Proposition 6). Then the necessity of Hamiltonian condition is vacuously true. On the other hand, if $f_L(x_0, x_1) > -\infty$, then $f_L(x_0, x_1)$ must be finite, and (6.50) asserts that (x_0, x_1) and (p_0, p_1) are in duality. Theorem 4 is thereby proved.

References

1. R. T. Rockafellar, Integral functionals, normal integrands and measurable selections, in Nonlinear Operators and the Calculus of Variations, edited by L. Waelbroeck, Springer Verlag series Lecture Notes in Math., 1976.
2. R. T. Rockafellar, Measurable dependence of convex sets and functions on parameters, J. Math. Anal. Appl. 28 (1969), 4-25.
3. R. T. Rockafellar, Convex integral functionals and duality, in Contributions to Nonlinear Functional Analysis, Academic Press, 1971, 215-236.
4. R. T. Rockafellar, Integrals which are convex functionals, Pacific J. Math. 24 (1968), 525-539.

R. T. Rockafellar, Integrals which are convex functionals, II, Pacific J. Math. 39 (1971), 439-469.

R. T. Rockafellar, Existence theorems for general control problems of Bolza and Lagrange, Advances in Math. 15 (1975), 312-333.

R. T. Rockafellar, Convex Analysis, Princeton Univ. Press, 1970.

8. R. T. Rockafellar, Conjugate convex functions in optimal control and the calculus of variations, J. Math. Anal. Appl. 32 (1970), 174-222.

9. R. T. Rockafellar, Existence and duality theorems for convex problems of Bolza, Trans. Amer. Math. Soc. 159 (1971), 1-40.

10. R. T. Rockafellar, State constraints in convex control problems of Bolza, SIAM J. Control 10 (1972), 691-715.

11. H. Halkin, Extremal properties of biconvex contingent equations, in Ordinary Differential Equations, edited by L. Weiss, Academic Press, 1972.

12. E. Asplund and L. Bungart, A First Course in Integration, Holt, Rinehart and Winston, 1966.

13. E. Michael, Continuous Selections, I, Annals of Math. 63 (1956), 361-382.

14. R. T. Rockafellar and R. J. B. Wets, Continuous versus measurable dependence in N-stage stochastic programming, J. Math. Analysis Appl. 48 (1974), 836-858.

Department of Mathematics
University of Washington
Seattle, Washington 98195

This research was supported in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under AFOSR grant number 72-2269.