Reprinted from JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS Vol. 32, No. 1, October 1970

Conjugate Convex Functions in Optimal Control and the Calculus of Variations

R. T. ROCKAFELLAR*

University of Washington, Seattle, Washington 98105

Submitted by R. Bellman

1. INTRODUCTION

This paper will be concerned with generalized problems of Bolza of the form: Minimize

$$\Phi_{l,L}(x) = l(x(0), x(T)) + \int_0^T L(t, x(t), \dot{x}(t))\, dt \tag{1.1}$$

subject to the constraints

$$(x(0), x(T)) \in C, \tag{1.2}$$

$$(x(t), \dot{x}(t)) \in D(t) \text{ for almost every } t, \tag{1.3}$$

where $x(t)$ is an absolutely continuous function from the real interval $[0, T]$ (T fixed and positive) to R^n with derivative $\dot{x}(t)$ (almost everywhere), C and $D(t)$ are subsets of $R^n \times R^n$, l is a real-valued function on C and $L(t, \cdot, \cdot)$ is for each $t \in [0, T]$ a real-valued function on $D(t)$. Here C and $D(t)$ could be defined, for example, by systems of equations or inequalities. Not only classical problems, but many problems of optimal control can be expressed in this form, as will be seen below.

Our treatment of such problems of Bolza differs from previous treatments in several respects. On the one hand, we impose convexity, not only in $\dot{x}$, but in x and $\dot{x}$ jointly. Thus we consider only the case where the sets C and $D(t)$ are convex, l is a convex function on C, and $L(t, \cdot, \cdot)$ is a convex function on $D(t)$. This, of course, excludes many important problems from consideration, although it still allows a substantial class of applications.

On the other hand, we make unusually weak assumptions concerning the regularity of l and L. No differentiability is assumed whatsoever. In deriving necessary and sufficient conditions for a given arc $x(t)$ to be optimal, we rely entirely on the "subdifferentiability" properties of l and L which automatically follow from convexity. Furthermore, only lower semicontinuity, rather

* This research was supported in part by the Air Force Office of Scientific Research under Grant AF-AFOSR-1202-67A.

than continuity, in $x(0)$, $x(T)$, x and $\dot{x}$ is imposed on the functions l and $L(t, \cdot, \cdot)$. The pair $(D(t), L(t, \cdot, \cdot))$ is only required to depend measurably on t in a certain general sense.

Convexity theory is, of course, the tool which makes it possible to survive under such weak regularity assumptions. The concepts and special results of convex analysis can be substituted in many instances for those of classical differential analysis, as has long been known in the calculus of variations and optimal control theory. For example, the notion of a tangent hyperplane to a smooth manifold can be replaced by that of a supporting hyperplane to a convex set. To some extent, the convexity assumptions in this paper are motivated by the desire to explore what happens if this substitution of convex analysis for differential analysis, already widely carried out in the literature, is brought to a logical extreme. From such an exploration, even if its domain is restrictive in certain respects, one may hope to learn something about the "limits of the possible". Knowledge of what is, or is not, true in the "purely convex" case could help shape conjectures in more general cases. And, needless to say, there is always the hope that the methods in the "purely convex" case, which are quite different from the usual ones, may lead to new insights.

The main justification for our convexity assumptions, however, is that they lead to a theory of duality which would otherwise not be possible. By means of the theory of conjugate convex functions, we shall show that each (mildly regular) convex problem of Bolza of the type described above has associated with it a dual problem, which is likewise a convex problem of Bolza. The dual of the dual problem is the original problem again. Extremal arcs x of the original problem and extremal arcs p of the dual problem are related to each other by several conditions, involving subgradients of convex functions, which generalize the classical Euler-Lagrange equations, Hamiltonian equations and transversality conditions. These subgradient conditions, in the case of certain optimal control problems formulated as convex problems of Bolza, also generalize the well known maximum principle of control theory.

The duality theory developed here may be viewed as an extension of the one originally proposed in the calculus of variations by Friedrichs [10] (see also Courant-Hilbert [7, p. 234 ff.]). This earlier theory likewise required, in effect (in view of the necessary condition of Legendre and its dual), joint convexity in x and $\dot{x}$, but it was based on the classical Legendre transformation, rather than the much more general conjugacy correspondence of Fenchel [8]. Thus it actually required *strict* convexity, as well as differentiability, in x and $\dot{x}$, and it was unable, except in very special cases, to handle constraints of the type (1.2) or (1.3). Since the Fenchel conjugate of a convex function is essentially equivalent to the Legendre conjugate when the function is strictly convex and differentiable [21, Section 26], the earlier theory is essentially contained in the one in this paper.

The relationship between our dual problem and the dual or reciprocal problems of Pearson [18, 19], Mond-Hanson [14] and Kreindler [12] (which depend on differentiability for their definition) is less apparent. Basically, however, it is the same as that already known in the theory of convex programs between the dual in the sense of Wolfe [30] and the dual in the sense of conjugate convex functions; see [21, p. 320-22, and p. 430]. Our dual problem is an outgrowth of the abstract duality theory developed in [21], [24], and [25], and also given, in a somewhat different form (where conjugate functions do not appear explicitly), by Wets-VanSlyke [29].

The duality between "continuous" infimal convolution and "continuous" addition of convex functions, as discussed by Ioffe and Tikhomirov in [11], [33], and [34], may be regarded as a special case of the duality in Example 7 below (for convex functions on R^n). The duality in the continuous linear programming problems of Bellman [2, p. 197 ff.] and Tyndall [28] could also be regarded as a special case of our duality, although not in as simple a manner.

To our knowledge, the general theory of conjugate convex functions has not previously been applied to the calculus of variations (or optimal control) in a broad and systematic way, although Ioffe and Tikhomirov have recently used this theory as a vehicle for expressing certain convexification results in [33, Part II, Section 2]. However, nonclassical conjugate functions have appeared in special types of problems treated by Moreau [16, 17] and Ioffe-Tikhomirov [11, 33, 34], and they have been used by Young [31] in defining "Hamiltonians in the large". They have also been mentioned by Zachrisson in an informal note [32] which anticipates several of the ideas exploited in this paper, such as generalized Hamiltonian equations in terms of subgradients.

(*Note added in proof*: some related ideas have also been pursued by Tsvetanov [35].)

Most of the background material in convex analysis relevant to this paper can be found in the book [21], the 1967 lecture notes of Moreau [15], and the survey of Ioffe and Tikhomirov [33]. The principal exception is the special theory of measurability developed by the author in [22] and [23] with the present application in mind. This theory makes it possible, by taking advantage of convexity, to avoid certain assumptions of continuity and to allow the constraint (1.3) to depend on t in a very general way. It contains, in particular, results on measurable selections which take on the role played elsewhere by Filippov's implicit functions lemma [9].

The plan of the paper is fairly apparent from the section titles:

1. Introduction
2. Basic assumptions

3. Convex problems of Bolza
4. Some examples in optimal control
5. The dual Bolza functional
6. Examples of dual problems
7. Conjugates of Bolza functionals
8. Duality of infima
9. Subdifferential conditions for a minimum
10. Extremal arcs and the maximum principle.

Sections 2 through 6 are concerned principally with the proper technical formulation of a convex problem of Bolza and its dual. Section 7 builds machinery. The main duality results (Theorems 4 and 5 and their corollaries) are harvested in Sections 8 and 9. Various applications, such as to optimal control theory, are treated in general examples in sections 4, 6, 8, and 10. Many of these examples are accompanied by lengthy proofs, and they thus contain much of the substance of this paper.

Due to the length of the exposition, we have had to omit a number of results of a more difficult nature which are needed to balance out the theory, such as theorems about the existence and regularity properties of optimal arcs and extremals. These will be published separately [26, 27].

2. Basic Assumptions

Let $L_n^p = L_n^p[0, T]$ denote the usual Banach space of (equivalence classes of) Lebesgue measurable functions from $[0, T]$ to R^n under the norm

$$\| v \|_p = \left(\int_0^T | v(t) |^p \, dt \right)^{1/p} \qquad \text{if} \qquad 1 \leqslant p < +\infty,$$
$$\| v \|_\infty = \operatorname*{ess.\,sup}_{0 \leqslant t \leqslant T} | v(t) | ,$$

where $| \cdot |$ denotes the Euclidean norm in R^n. Let B_n^p be the linear space $R^n \oplus L_n^p$ under the norm

$$\begin{aligned} \|(c, v)\|_p &= (| c |^p + \| v \|_p^p)^{1/p} \qquad \text{if} \qquad 1 \leqslant p < +\infty, \\ \|(c, v)\|_\infty &= \max\{| c | , \| v \|_\infty\}. \end{aligned} \tag{2.1}$$

Obviously B_n^p is a Banach space. If $1 \leqslant p < +\infty$, the dual of B_n^p can be identified with B_n^q, where $(1/p) + (1/q) = 1$, under the pairing

$$\langle (c, v), (d, w) \rangle = \langle c, d \rangle + \int_0^T \langle v(t), w(t) \rangle \, dt.$$

We denote by A_n^p the space of all absolutely continuous functions from $[0, T]$ to R^n whose derivative (defined almost everywhere) belongs to L_n^p. The norm on A_n^p is taken to be

$$||| x |||_p = \|(x(0), \dot{x})\|_p . \tag{2.2}$$

The mapping $x \to (x(0), \dot{x})$ is thus a linear isometry of A_n^p onto B_n^p, so that A_n^p is a Banach space whose dual, in the case where $1 \leqslant p < +\infty$, can be identified with B_n^q, $(1/p) + (1/q) = 1$, under the pairing

$$\langle x, (d, w)\rangle = \langle x(0), d\rangle + \int_0^T \langle \dot{x}(t), w(t)\rangle\, dt. \tag{2.3}$$

Note that if $p \geqslant p'$ one has $A_n^p \subset A_n^{p'}$, and convergence in $||| \cdot |||_p$ implies convergence in $||| \cdot |||_{p'}$.

It will be convenient to reformulate a problem of Bolza as a problem of minimizing a certain functional Φ over the space A_n^1, where no constraints appear explicitly, but Φ is extended-real-valued. The idea is simply to incorporate the constraints (1.2) and (1.3) into the functional $\Phi_{l,L}$ by defining (or redefining, as the case may be)

$$l(c_0, c_T) = +\infty \qquad \text{if} \qquad (c_0, c_T) \notin C, \tag{2.4}$$

$$L(t, x, v) = +\infty \qquad \text{if} \qquad (x, v) \notin D(t). \tag{2.5}$$

Heuristically, (2.4) and (2.5) may be interpreted as imposing an infinite penalty when the given constraints are violated.

Assume for a moment that the regularity properties of $L(t, u, v)$ and $D(t)$ are such that, under this extended definition of L, the integral in (1.1) is *well-defined* in the following sense: For each x in A_n^1, the (extended-real-valued) integrand is a measurable function of t which majorizes at least one summable function of t. Then $\Phi_{l,L}(x)$ will be well-defined and equal to either a real number or $+\infty$. In fact, one will have $\Phi_{l,L}(x) = +\infty$ whenever x fails to satisfy either of the constraints (1.2) or (1.3), so that minimizing $\Phi_{l,L}$ over all of A_n^1 will be equivalent to minimizing $\Phi_{l,L}$ subject to (1.2) and (1.3).

From this discussion, it is clear that problems of Bolza can be described simply by specifying two extended-real-valued functions l and L, which are everywhere defined on $R^n \times R^n$ and $[0, T] \times R^n \times R^n$, respectively. This is the approach we shall take. The sets C and $D(t)$ are defined in terms of l and L by

$$C = \{(c_0, c_T) \in R^n \times R^n \mid l(c_0, c_T) < +\infty\}, \tag{2.6}$$

$$D(t) = \{(x, v) \in R^n \times R^n \mid L(t, x, v) < +\infty\}. \tag{2.7}$$

However, in many contexts, other than examples, these sets need not be treated explicitly.

It should be observed that, in suppressing $D(t)$ and passing to an extended-real-valued L, one is nevertheless faced with a technical question of what conditions on L are appropriate to insure that the integrand in (1.1) is always measurable. Familiar conditions such as those of Caratheodory (continuity in (x, v) and measurability in t) are not applicable, since $L(t, \cdot, \cdot)$ may jump abruptly to $+\infty$ at the boundary of $D(t)$. Furthermore, for the study of duality, one needs conditions which are self-dual with respect to taking conjugates of convex functions. The conditions given below will meet this criterion.

We now state the basic assumptions which will be in effect throughout this paper. The first is:

(A) *Each of the functions l and $L(t, \cdot, \cdot)$ is a lower semicontinuous convex function (everywhere defined) on $R^n \times R^n$ with values in $R^1 \cup \{+\infty\}$, not identically $+\infty$.*

The lower semicontinuity assumption in (A), of course, requires all level sets of the form

$$\{(c_0, c_T) \in R^n \times R^n \mid l(c_0, c_T) \leqslant \mu\},$$
$$\{(x, v) \in R^n \times R^n \mid L(t, x, v) \leqslant \mu\},$$

to be closed, but it does not actually require the sets C and $D(t)$ to be closed. On the other hand, (A) does imply that C and $D(t)$ are convex and non-empty.

The remaining assumptions concern L only, and *under* (A) *they are all automatically satisfied when L is independent of t.* The main purpose of these assumptions is to guarantee in a suitable way that the integral

$$\int_0^T L(t, x(t), \dot{x}(t))\, dt$$

is meaningful for every $x \in A_n^1$.

(B) *L is Lebesgue normal in the sense of* [22].

(C) *L majorizes at least one function r on $[0, T] \times R^n \times R^n$ of the form*

$$r(t, x, v) = \langle x, s(t)\rangle + \langle v, p(t)\rangle - \alpha(t)$$

with $s \in L_n^1$, $p \in L_n^\infty$, $\alpha \in L_1^1$.

(D) *There exists at least one pair of functions $x \in L_n^\infty$ and $v \in L_n^1$ such that the function $L(\cdot, x(\cdot), v(\cdot))$ (which is extended-real-valued on $[0, T]$) is majorized by a function $\beta \in L_1^1$.*

(Here $\langle \cdot, \cdot \rangle$ denotes the usual inner product in R^n.)

The *Lebesgue normality* condition (B) is satisfied by definition if and only if [in addition to (A) being satisfied] there exists a *countable* collection $\{(x_i, v_i) \mid i \in I\}$, where x_i and v_i are Lebesgue measurable functions from $[0, T]$ to R^n, such that $L(t, x_i(t), v_i(t))$ is a Lebesgue measurable function of t for each $i \in I$, and the set

$$D(t) \cap \{(x_i(t), v_i(t)) \mid i \in I\}$$

is dense in $D(t)$ for each $t \in [0, T]$. It is easily seen that this holds in particular whenever $L(t, x, v)$ is independent of t, or whenever $L(t, x, v)$ is Lebesgue measurable in t for each (x, v) and $D(t)$ has a nonempty interior for each t [22, p. 528].

We have shown in [23, Corollary 5.1] that [in the presence of (A)] condition (B) is satisfied if and only if L is $\mathscr{L} \times \mathscr{B} \times \mathscr{B}$ measurable, i.e., measurable with respect to the σ-field of subsets of $[0, T] \times R^n \times R^n$ generated by products of Lebesgue sets in $[0, T]$ and Borel sets in R^n. [The latter certainly is true if L is Borel measurable, and in particular if L is lower semicontinuous in (t, x, v).] Also, according to [23, Theorem 4], (B) is satisfied if and only if the multifunction

$$E : t \to E(t) = \{(x, v, \mu) \mid (x, v) \in D(t), L(t, x, v) \leqslant \mu < +\infty\} \qquad (2.8)$$

is *Lebesgue measurable* from $[0, T]$ to $R^n \times R^n \times R^1$, in the sense that

$$E^{-1}(S) = \{t \mid E(t) \cap S \neq \emptyset\}$$

is a Lebesgue measurable subset of $[0, T]$ for every closed subset S of $R^n \times R^n \times R^1$.

The fact that (C) automatically holds [assuming (A)] when L is independent of t follows from the fact that a lower semicontinuous convex function from $R^n \times R^n$ to $R^1 \cup \{+\infty\}$ necessarily majorizes at least one affine function [21, Theorem 12.1]. Thus when L is independent of t, the functions s, p, and α in (C), as well as x and v in (D), may be taken to be constant functions.

We shall note in Section 4 that (C) and (D) are "dual" to each other, while (A) and (B) are "self-dual".

3. Convex Problems of Bolza

The conditions described in the preceding section lead to a well-defined problem of Bolza. In the first place, (A) and (B) guarantee according to [22, p. 531] that $L(t, x(t), v(t))$ is a Lebesgue measurable function of t whenever $x(t)$ and $v(t)$ are Lebesgue measurable in t.

Furthermore, if $x \in L_n^\infty$ and $v \in L_n^1$, we have

$$\begin{aligned} L(t, x(t), v(t)) &\geq r(t, x(t), v(t)) \\ &= \langle x(t), s(t)\rangle + \langle v(t), p(t)\rangle - \mu(t) \end{aligned} \tag{3.1}$$

by (C), where the latter function of t is summable on $[0, T]$. It follows that the integral

$$I_L(x, v) = \int_0^T L(t, x(t), v(t))\, dt \tag{3.2}$$

is well-defined (unambiguously either a real number or $+\infty$) for every $(x, v) \in L_n^\infty \oplus L_n^1$.

PROPOSITION 1. *Under* (A), (B) *and* (C), *the integral I_L is a well-defined convex function from $L_n^\infty \oplus L_n^1$ to $R^1 \cup \{+\infty\}$. Moreover, when $L_n^\infty \oplus L_n^1$ is regarded as a topological vector space in the product of the norm topologies, I_L is lower semicontinuous, not only with respect to this normable topology, but also with respect to the corresponding weak topology.*

Proof. It has already been seen that I_L is well-defined. The convexity of I_L is an immediate consequence of the convexity of $L(t, \cdot, \cdot)$ for every t. To prove the lower semicontinuity of I_L in the product of the norm topologies, consider any function r as in (C), and set

$$L'(t, x, v) = L(t, x, v) - r(t, x, v) \geq 0. \tag{3.3}$$

Then

$$\begin{aligned} I_L(x, v) &= \int_0^T L'(t, x(t), v(t))\, dt + \int_0^T r(t, x(t), v(t))\, dt \\ &= I_{L'}(x, v) + I_r(x, v), \end{aligned} \tag{3.4}$$

where I_r is a continuous linear functional on $L_n^\infty \oplus L_n^1$, so it suffices to prove the lower semicontinuity of $I_{L'}$. Since L' is nonnegative, the latter is easily deduced from Fatou's lemma and the fact that every convergent sequence (x_i, v_i) in $L_n^\infty \oplus L_n^1$ has a subsequence in which the functions v_i, as well as the functions x_i, converge pointwise almost everywhere.

A convex functional I on a topological vector space B which is lower semi-continuous with respect to the given topology on B is necessarily lower semicontinuous also with respect to the corresponding weak topology on B. (This is immediate from the fact that I is lower semicontinuous with respect to some topology if and only if all the level sets of the form $\{y \in B \mid I(y) \leq \mu\}$, $\mu \in R^1$, are closed with respect to that topology. When I is convex, these level sets are convex subsets of B, and, as is well-known, a *convex* subset of a locally convex space is weakly closed if and only if it is closed in the given topology.) This establishes Proposition 1.

Condition (D), which was not needed in Proposition 1, is of course merely the condition that $I_L(x, v) < +\infty$ for at least one $(x, v) \in L_n^\infty \oplus L_n^1$. Certainly no harm is done when this condition is added to the others, and it will be needed later for the sake of duality.

Given any $x \in A_n^p$, $1 \leqslant p \leqslant \infty$, we have $x \in L_n^\infty$ and $\dot{x} \in L_n^1$, so that the integral

$$I_L(x, \dot{x}) = \int_0^T L(t, x(t), \dot{x}(t))\, dt \tag{3.5}$$

is well-defined by the above. In fact, this integral is a convex function of x, inasmuch as $x = (1 - \lambda) x_1 + \lambda x_2$ implies $\dot{x} = (1 - \lambda) \dot{x}_1 + \lambda \dot{x}_2$. It is also lower semicontinuous in the norm topology of A_n^p by Proposition 1, since convergence of a sequence $\{x_i\}$ in the norm of A_n^p entails convergence of $\{\dot{x}_i\}$ in L_n^1 and convergence of $\{x_i\}$ in L_n^∞. Moreover, strong lower semicontinuity, together with convexity, implies weak lower semicontinuity, as observed at the end of the proof of Proposition 1.

On the other hand, since l is assumed to be a lower semicontinuous convex function on $R^n \times R^n$, it is apparent that the term

$$J_l(x) = l(x(0), x(T)) \tag{3.6}$$

is a lower semicontinuous convex function of $x \in A_n^p$.

Adding $J_l(x)$ and $I_L(x, \dot{x})$, we obtain:

THEOREM 1. *Under* (A), (B), *and* (C), *the function*

$$\Phi_{l,L}(x) = l(x(0), x(T)) + \int_0^T L(t, x(t), \dot{x}(t))\, dt$$

is, for any p $(1 \leqslant p \leqslant +\infty)$, *a well-defined convex function from* A_n^p *to* $R^1 \cup \{+\infty\}$. *Moreover*, $\Phi_{l,L}$ *is lower semicontinuous on* A_n^p, *not only with respect to the norm topology, but also with respect to the weak topology.*

We shall call $\Phi_{l,L}$ the *Bolza functional* corresponding to l and L, where l is the *boundary function* and L is the *Lagrangian function*. A problem of minimizing a function of the form $\Phi_{l,L}$ over A_n^1 [under conditions (A), (B), (C), and (D)] will be called a *convex problem of Bolza*.

By a *feasible arc* in a convex problem of Bolza, we shall mean an $x \in A_n^1$ such that $\Phi_{l,L}(x) < +\infty$. Clearly, a feasible arc must satisfy conditions (1.2) and (1.3) [where C and $D(t)$ are given by (2.6) and (2.7)], although these conditions are not always sufficient for feasibility. It follows from the convexity of $\Phi_{l,L}$ that the set of all feasible arcs in a given convex problem of Bolza is a convex subset of A_n^1, not necessarily closed and possibly empty.

Minimizing $\Phi_{l,L}$ over all of A_n^1 is equivalent to minimizing $\Phi_{l,L}$ over this convex subset.

A feasible arc at which the minimum of $\Phi_{l,L}$ over A_n^1 is achieved will be called an *optimal arc*. (We do not speak of optimal arcs when $\Phi_{l,L}$ is identically $+\infty$, even though in that case the minimum of $\Phi_{l,L}$ is achieved at every point of A_n^1.) Theorem 1 implies that the set of all optimal arcs in a given convex problem of Bolza is a (possibly empty) convex subset of A_n^1, which is weakly closed, as well as strongly closed.

Since $\Phi_{l,L}$ is convex, a local minimum of $\Phi_{l,L}$ is a global minimum, and no difficulties arise because of a possibility of arcs yielding more complicated kinds of extrema or stationary points. Besides optimal arcs, we shall define in Section 9, in terms of subgradients of the convex functions l and $L(t, \cdot, \cdot)$, a class of so-called *extremal* arcs of $\Phi_{l,L}$, but it will turn out that every such arc is optimal. (A major task is to establish conditions under which an optimal arc is necessarily an extremal arc.)

4. Some Examples in Optimal Control

As mentioned in the introduction, various problems in optimal control can be formulated as convex problems of Bolza. We shall now demonstrate this in several examples. These examples are chosen mainly to illustrate how the basic assumptions (A), (B), (C), and (D) can be verified in some important cases, and they do not pretend to give the most general problems to which the theory is applicable.

Example 1. Consider an optimal control problem of the following type: Minimize

$$\int_0^T K(t, z(t), u(t))\, dt \tag{4.1}$$

in $z \in A_r^1$ and $u \in L_s^1$ (with T fixed), subject to the constraints

$$\dot{z}(t) = A(t)\, z(t) + B(t)\, u(t) \text{ for almost every } t, \tag{4.2}$$

$$u(t) \in U(t) \text{ for almost every } t, \tag{4.3}$$

$$z(0) \in Z_0 \qquad \text{and} \qquad z(T) \in Z_T, \tag{4.4}$$

where, for each $t \in [0, T]$, $K(t, \cdot, \cdot)$ is a real-valued (finite and everywhere defined) convex function on $R^n \times R^s$, $A(t)$ and $B(t)$ are real matrices of dimensions $r \times r$ and $r \times s$, respectively, $U(t)$ is a nonempty closed convex subset of R^s, and Z_0 and Z_T are nonempty closed convex subsets of R^r. (In particular, Z_0 or Z_T may consist of a single point or be all of R^r.)

To formulate this as a problem of Bolza, it is convenient to regard $u(t)$ as the derivative of a function $y(t)$ in A_s^1 and then set $x(t) = (z(t), y(t))$ in R^n, where $n = r + s$. Then l and L are defined by

$$l(x(0), x(T)) = \begin{cases} 0 & \text{if} \quad z(0) \in Z_0 \quad \text{and} \quad z(T) \in Z_T, \\ +\infty & \text{if} \quad z(0) \notin Z_0 \quad \text{or} \quad z(T) \notin Z_T, \end{cases} \tag{4.5}$$

$$L(t, x, v) = \begin{cases} K(t, z, u) & \text{if} \quad u \in U(t) \quad \text{and} \quad w = A(t)\, z + B(t)\, u, \\ +\infty & \text{if} \quad u \notin U(t) \quad \text{or if} \quad w \neq A(t)\, z + B(t)\, u, \end{cases} \tag{4.6}$$

where $(z, y) = x$ and $(w, u) = v$ in $R^r \times R^s = R^n$. The given optimal control problem is equivalent to minimizing

$$l(x(0), x(T)) + \int_0^T L(t, x(t), \dot{x}(t))\, dt$$

over all $x \in A_n^1$, provided that the latter problem is well-defined, as is always the case when conditions (A), (B), and (C) are satisfied.

It is elementary here that (A) is satisfied. (Recall that the functions $K(t, \cdot, \cdot)$, being finite and convex, are necessarily continuous throught $R^r \times R^s$.) To get (B), we assume further that $K(t, z, u)$ is a Lebesgue measurable function of t for each $(z, u) \in R^r \times R^s$, that the components of $A(t)$ and $B(t)$ are Lebesgue measurable functions of t, and that the multifunction $U : t \to U(t)$ is Lebesgue measurable (in the sense that the set $\{t \mid U(t) \cap S \neq \emptyset\}$ is a Lebesgue measurable subset of $[0, T]$ for every closed subset S of R^s; some criteria for this are compiled in [5] and [23]).

LEMMA. *Condition* (B) *is satisfied under the preceding assumptions.*

Proof. This will be deduced from results in [23]. Let D_1 and D_2 be the multifunctions from $[0, T]$ to $R^r \times R^s \times R^r \times R^s$ defined by

$$D_1(t) = \{(z, y, w, u) \mid w - A(t)\, z - B(t)\, u = 0\},$$
$$D_2(t) = \{(z, y, w, u) \mid u \in U(t)\}.$$

It is clear that D_2 is Lebesgue measurable, and the Lebesgue measurability of D_1 is assured, for instance, by [23, Corollary 3.6]. Then the multifunction

$$D : t \to D_1(t) \cap D_2(t)$$

is Lebesgue measurable by [23, Corollary 1.3]. Now let

$$L_1(t, x, v) = K(t, z, u),$$
$$L_1'(t, x, v) = \begin{cases} 0 & \text{if} \quad (z, y, w, u) \in D(t), \\ +\infty & \text{if} \quad (z, y, w, u) \notin D(t). \end{cases}$$

Since L_1 is finite, measurable in t, and convex in (z, y, w, u), L_1 is Lebesgue normal by [22, p. 529]. On the other hand, the Lebesgue measurability of the multifunction D implies the Lebesgue normality of L_1' by [23, Theorem 3]. We have $L = L_1 + L_1'$, and therefore L is Lebesgue normal by [23, Corollary 4.2]. Thus (B) is satisfied as claimed.

Assuming that the components of $A(t)$ belong to L_1^1, while those of $B(t)$ belong to L_1^∞, it can easily be seen that condition (C) holds if and only if there exist functions $a \in L_r^1$, $b \in L_s^\infty$ and $\mu \in L_1^1$ such that

$$K(t, z, u) \geqslant \langle z, a(t)\rangle + \langle u, b(t)\rangle - \mu(t)$$

for all $t \in [0, 1]$, $z \in R^r$, and $u \in U(t)$. There are various ways of insuring the existence of such functions, but in particular it follows from [22, Theorem 4] that this condition is satisfied (regardless of the nature of $U(t)$), if $K(t, z, u)$ is an essentially bounded function of t for every $(z, u) \in R^r \times R^s$.

Condition (D) merely requires here the existence of functions $z \in L_r^\infty$ and $u \in L_s^1$ such that $u(t) \in U(t)$ for almost every t, and $K(t, z(t), u(t))$ is summable in t. Again, this is satisfied in particular, according to [22, Theorem 4], if $K(t, z, u)$ is an essentially bounded function of t for every $(z, u) \in R^r \times R^s$, and if the function

$$d(t) = \min\{|u| \mid u \in U(t)\}$$

is essentially bounded above in t. (The boundedness of $d(t)$ implies the existence of a function $u \in L_s^\infty$ such that $u(t) \in U(t)$ for every $t \in [0, T]$; see [5] or [23]. One may take this choice of $u(t)$, together with $z(t) \equiv 0$.)

Example 2. Consider an optimal control problem of the following type: minimize

$$k_0(z(0), z(T)) + \int_0^T K_0(t, z(t), \dot{z}(t), u(t))\, dt \tag{4.7}$$

in $z \in A_r^1$ and $u \in L_s^1$, subject to the constraints

$$k_i(z(0), z(T)) \leqslant 0 \quad \text{for} \quad i = 1, \ldots, m_1, \tag{4.8}$$

$$K_j(t, z(t), \dot{z}(t), u(t)) \leqslant 0 \quad \text{for} \quad j = 1, \ldots, m_2 \text{ and almost every } t, \tag{4.9}$$

where the functions k_i are all finite and convex on $R^r \times R^r$, and for each $t \in [0, T]$ the functions $K_j(t, \cdot, \cdot, \cdot)$ are all finite and convex on $R^r \times R^r \times R^s$. (Here system (4.9) may involve constraints on the state $z(t)$ alone, or constraints on the control $u(t)$ alone.)

This may be formulated as a convex problem of Bolza in much the same manner as the preceding example. Setting $x = (z, y)$ and $v = (w, u)$, we define

$$l(x(0), x(T)) = \begin{cases} k_0(z(0), z(T)) & \text{if (4.8) is satisfied,} \\ +\infty & \text{otherwise,} \end{cases} \tag{4.10}$$

$$L(t, x, v) = \begin{cases} K_0(t, z, q, u) & \text{if} \quad K_j(t, z, q, u) \leqslant 0, \quad j = 1,..., m_2 , \\ +\infty & \text{otherwise,} \end{cases} \tag{4.11}$$

Condition (A) will be satisfied, assuming that the constraints (4.9) are consistent in $R^r \times R^r$ and that the constraints

$$K_j(t, z, w, u) \leqslant 0, \qquad j = 1,..., m_2 , \tag{4.12}$$

are consistent in $R^r \times R^r \times R^s$ for each fixed $t \in [0, T]$. We shall assume in addition to this that $K_j(t, z, w, u)$ is an essentially bounded Lebesgue measurable function of t for each index j $(0 \leqslant j \leqslant m_2)$ and each (z, w, u) in $R^r \times R^r \times R^s$.

LEMMA. *Conditions* (B) *and* (C) *are satisfied under the preceding assumptions.*

Proof. With $x = (z, y)$ and $v = (w, u)$ as above, let

$$L_j(t, x, v) = K_j(t, z, w, u), \qquad j = 0, 1,..., m_2 , \tag{4.13}$$

and for each $t \in [0, T]$ let $D(t)$ be the (nonempty, closed, convex) set of all (x, v) in $R^n \times R^n$ satisfying (4.12), i.e.,

$$D(t) = \{(x, v) \mid L_j(t, x, v) \leqslant 0, j = 1,..., m_2\}.$$

The functions L_j, being finite, convex in (x, v), and Lebesgue measurable in t, are all Lebesgue normal [22, p. 529]. This implies by [23, Corollary 4.4] that the multifunction $D : t \to D(t)$ is Lebesgue measurable, and hence by [23, Theorem 3] that the function

$$L_0'(t, x, v) = \begin{cases} 0 & \text{if} \quad (x, v) \in D(t), \\ +\infty & \text{if} \quad (x, v) \notin D(t), \end{cases}$$

is Lebesgue normal. Inasmuch as $L = L_0 + L_0'$, L is Lebesgue normal by [23, Corollary 4.2], and (B) is established.

The fact that $L_0(t, x, v)$ is essentially bounded in t for each $(x, v) \in R^n \times R^n$ implies by [22, Theorem 4] that $L_0(t, x(t), v(t))$ is summable in t for every $x \in L_n^\infty$ and $v \in L_n^\infty$, and furthermore that L_0 satisfies (C) (where s can actually be chosen in L_n^∞). Therefore L satisfies (C).

In fact L also satisfies (D), under the preceding assumptions, provided that it is possible to select a point $(x(t), v(t))$ from $D(t)$ for each $t \in [0, T]$ in such a way that $x \in L_n^\infty$ and $v \in L_n^\infty$. The latter will be true in particular by a selection theorem of Kuratowski and Ryll-Nardzewski [13] (quoted as [23, Corollary 1.1]) if there exists a bounded subset S of $R^n \times R^n$ such that $D(t) \cap S \neq \phi$ for every t.

It may be noted that Example 2 contains Example 1 as the special case where $m_1 = 1$, $m_2 = 2r + 1$, and

$$
\begin{aligned}
&k_0(z(0), z(t)) \equiv 0,\\
&k_1(z(0), z(T)) = \min\{| z(0) - a_0 |^2 + | z(T) - a_T |^2 \mid a_0 \in Z_0, a_T \in Z_T\},\\
&K_0(t, z, w, u) = K(t, z, u)\\
&K_j(t, z, w, u) = w^j - \sum_{k=1}^{r} a_{jk}(t)\, z^k - \sum_{l=1}^{s} b_{jl}(t)\, u^l, \qquad j = 1,..., r, \qquad (4.14)\\
&K_{r+j}(t, z, w, u) = -K_j(t, z, w, u), \qquad j = 1,..., r,\\
&K_{2r+1}(t, z, w, u) = \min\{| u - u' |^2 \mid u' \in U(t)\}.
\end{aligned}
$$

Although the constraints (4.9) do not in general represent an ordinary differential equation, they can always be expressed, of course, as a so-called contingent differential equation:

$$\dot{z}(t) \in F(t, z(t), u(t)) \text{ for almost every } t, \tag{4.15}$$

where $F(t, z, u)$ denotes for each $t \in [0, T]$, $z \in R^r$ and $u \in R^s$ the (closed, convex, possibly empty) set of all $w \in R^r$ such that (4.12) is satisfied.

EXAMPLE 3. The two preceding examples illustrate what seems to be the most convenient method of formulating an optimal control problem as a problem of Bolza. However, there is another method, technically harder to work with, which makes clearer in some ways the relationship between the present approach and other approaches to optimal control problems in the literature. To avoid a lengthy technical discussion, we shall treat this method only in a very special case, although it is really of much greater generality.

Consider a fixed endpoint problem of the following type: minimize

$$\int_0^T K(t, x(t), u(t))\, dt \tag{4.16}$$

in $x \in A_n^1$ and $u \in L_m^\infty$, subject to the constraints

$$\dot{x}(t) = f(t, x(t), u(t)) \text{ for almost every } t, \tag{4.17}$$

$$u(t) \in U \text{ for almost every } t, \tag{4.18}$$

$$x(0) = c_0 \quad \text{and} \quad x(T) = c_T, \tag{4.19}$$

where U denotes a nonempty compact subset of R^m, and K and f are continuous functions from $[0, T] \times R^n \times R^m$ to R^1 and R^n, respectively. Define L to be the function from $[0, T] \times R^n \times R^n$ to $R^1 \cup \{+\infty\}$ such that $L(t, x, v)$ is the minimum of $K(t, x, u)$ over all vectors $u \in U$ such that $f(t, x, u) = v$. (If there are no such vectors u, the minimum is $+\infty$ by convention.) Then L is lower semicontinuous (in all variables). Moreover, the given control problem is equivalent to minimizing the integral

$$\int_0^T L(t, x(t), \dot{x}(t))\, dt \tag{4.20}$$

over all $x \in A_n^1$ satisfying (4.19). (The integral is well-defined, although it may be $+\infty$, because, under our assumptions, $L(t, x(t), \dot{x}(t))$ is measurable and essentially bounded below as a function of $t \in [0, T]$ for each $x \in A_n^1$. It can be seen via Filippov's implicit functions lemma [9] that this integral is finite for a given x if and only if there exists a $u \in L_m^\infty$ satisfying (4.17) and (4.18), such that (4.20) and (4.16) are equal.)

The theory in this paper is applicable to the reformulated problem if $L(t, x, v)$ turns out to be convex in (x, v) for each t, as is always the case in particular when $K(t, x, u)$ is convex in (x, u), f is affine in (x, u), and U is convex. (One may verify that conditions (A), (B), (C), and (D) are then satisfied, where $l(x(0), x(T))$ is taken to be 0 if (4.19) holds and $+\infty$ otherwise.) The convexity of $L(t, x, v)$ in (x, v) means that, for each t, the epigraph

$$\{(x, v, \mu) \in R^n \times R^n \times R^1 \mid \mu \geqslant L(t, x, v)\}.$$

is convex. This may be compared with the weaker condition, developed by Cesari [6, p. 390] as a generalization of a similar condition of Filippov [9], that for each (t, x) the epigraph

$$\{(v, \mu) \in R^n \times R^1 \mid \mu \geqslant L(t, x, v)\}$$

is convex, or in other words that $L(t, x, v)$ is merely convex as a function of v, rather than as a function of (x, v).

5. The Dual Bolza Functional

Every convex problem of Bolza leads to a certain dual problem of the same type. This dual problem will be defined below in terms of the conjugates of the convex functions l and $L(t, \cdot, \cdot)$. The theory of conjugate convex functions will be used further in Sections 7 and 8 in establishing relationships between the dual problem and the original one.

We begin by reviewing some basic facts about conjugates (for a fuller exposition, see [3], [15], [21], [33]). A context more general than that needed simply for the definition of the dual problem is chosen for the purposes of Section 7.

Let X and Y be arbitrary real vector spaces paired by a bilinear form $\langle\cdot,\cdot\rangle$, and let X and Y be supplied with arbitrary locally convex topologies compatible with this pairing. (Thus it is assumed that $x \to \langle x, y\rangle$ is always a continuous linear function on X, and that every continuous linear function on X can be expressed in this form; at the same time, $y \to \langle x, y\rangle$ is always a continuous linear function on Y, and every continuous linear function on Y can be expressed in this form. In the case where $X = Y = R^n$, we take $\langle\cdot,\cdot\rangle$ to be the ordinary inner product.)

Let f be any extended-real-valued convex function on X. We allow f possibly to take on $-\infty$, as well as $+\infty$, in which case the convexity of f is interpreted to mean that the epigraph of f, i.e., the set

$$\operatorname{epi} f = \{(x, \mu) \mid x \in X, \mu \in R^1, \mu \geqslant f(x)\},$$

is convex in $X \oplus R^1$. We say that f is *proper*, if f does not take on $-\infty$, and f is not identically $+\infty$. The extended-real-valued function f^* on Y defined by

$$f^*(y) = \sup\{\langle x, y\rangle - f(x) \mid x \in X\} \tag{5.1}$$

is called the *conjugate* of f (with respect to the given pairing). The conjugate of f^*, i.e. the function f^{**} on X defined by

$$f^{**}(x) = \sup\{\langle x, y\rangle - f^*(y) \mid y \in Y\}, \tag{5.2}$$

is called the *biconjugate* of f.

The functions f^* and f^{**} are always convex and lower semicontinuous, and if they are not both proper then one must be identically $+\infty$ and the other identically $-\infty$. If f is proper and lower semicontinuous, one has $f^{**} = f$; thus the operation $f \to f^*$ yields a one-to-one symmetric correspondence between the lower semicontinuous proper convex functions on X and those on Y.

More generally, if f is not lower semicontinuous, let $\bar{f}$ denote the lower semicontinuous hull of f, i.e.

$$\bar{f}(x) = \liminf_{x' \to x} f(x'). \tag{5.3}$$

Then $\bar{f}$ is the greatest lower semicontinuous convex function majorized by f. If $\bar{f}(x) > -\infty$ for every x, one has $f^{**} = \bar{f}$. On the other hand, if $\bar{f}$ takes on $-\infty$ somewhere, there exists a convex set C (namely the closure of

$\{x \in X \mid f(x) < +\infty\})$ such that $\bar{f}(x) = -\infty$ for every $x \in C$ and $\bar{f}(x) = +\infty$ for every $x \notin C$. In the latter case f^{**} is identically $-\infty$, so that f^{**} and $\bar{f}$ agree on C but disagree outside of C.

We proceed now with the definition of the dual Bolza functional. Given the functions l and L, we denote by l^* the conjugate of l on $R^n \times R^n$, and we denote by L^* the function on $[0, T] \times R^n \times R^n$ such that, for each t, $L^*(t, \cdot, \cdot)$ is the conjugate of $L(t, \cdot, \cdot)$. (Here the ordinary inner product gives the pairings.) We then set

$$m(d_0, d_T) = l^*(d_0, -d_T), \tag{5.4}$$

$$M(t, p, s) = L^*(t, s, p). \tag{5.5}$$

Thus by definition

$$m(d_0, d_T) = \sup\{\langle c_0, d_0\rangle - \langle c_T, d_T\rangle - l(c_0, c_T) \mid c_0 \in R^n, c_T \in R^n\}, \tag{5.6}$$

$$M(t, p, s) = \sup\{\langle x, s\rangle + \langle v, p\rangle - L(t, x, v) \mid x \in R^n, v \in R^n\}. \tag{5.7}$$

The function m will be called the *boundary function dual to* l, and M will be called the *Lagrangian function dual to* L. We shall call the functional

$$\Phi_{m,M}(p) = m(p(0), p(T)) + \int_0^T M(t, p(t), \dot{p}(t))\, dt \tag{5.8}$$

the *Bolza functional dual to* $\Phi_{l,L}$.

THEOREM 2. *The conditions* (A), (B), (C), *and* (D) *on* l *and* L *imply that* m *and* M *likewise satisfy* (A), (B), (C), *and* (D). *Thus Theorem 1 is applicable to* $\Phi_{m,M}$, *as well as to* $\Phi_{l,L}$.

Moreover, l *is in turn the boundary function dual to* m, *and* L *is the Lagrangian function dual to* M, *so that the Bolza functional dual to* $\Phi_{m,M}$ *is just* $\Phi_{l,L}$ *again.*

Proof. It is immediate from the facts cited above that m and M again satisfy (A), and that

$$\begin{aligned} l(c_0, c_T) &= \sup\{\langle c_0, c_0'\rangle + \langle c_T, c_T'\rangle - l^*(c_0', c_T') \mid c_0' \in R^n, c_T' \in R^n\} \\ &= \sup\{\langle c_0, d_0\rangle - \langle c_T, d_T\rangle - m(d_0, d_T) \mid d_0 \in R^n, d_T \in R^n\} \\ &= m^*(c_0, -c_T), \end{aligned} \tag{5.9}$$

$$\begin{aligned} L(t, x, v) &= \sup\{\langle x, s\rangle + \langle v, p\rangle - L^*(t, s, p) \mid s \in R^n, p \in R^n\} \\ &= \sup\{\langle x, s\rangle + \langle v, p\rangle - M(t, p, s) \mid s \in R^n, p \in R^n\} \\ &= M^*(t, v, x). \end{aligned} \tag{5.10}$$

The latter formulas say that l and L are the functions dual to m and M as claimed. We have already proved elsewhere [22, Lemma 5] that Lebesgue normality is preserved when conjugates are taken. Thus M again satisfies (B). As for conditions (C) and (D), we observe from (5.7) and (5.10) that the functions $p \in L_n^\infty$, $s \in L_n^1$, and $\alpha \in L_1^1$ satisfy (C) for L, if and only if they satisfy

$$M(t, p(t), s(t)) \leqslant \alpha(t) \text{ for almost every } t.$$

Similarly, the function $x \in L_n^\infty$, $v \in L_n^1$, and $\beta \in L_1^1$ satisfy (D) for L, if and only if they satisfy

$$M(t, p, s) \geqslant \langle x(t), s\rangle + \langle v(t), p\rangle - \beta(t)$$

for every $t \in [0, T]$, $p \in R^n$ and $s \in R^n$. Thus (C) for L implies (D) for M, and (D) for L implies (C) for M. The proof of Theorem 2 is now complete.

The problem of minimizing $\Phi_{m,M}$ over A_n^1 will be called the *convex problem of Bolza dual* to the problem of minimizing $\Phi_{l,L}$ and A_n^1, and vice versa. The close connection between these two problems will be seen in Sections 8 and 9.

6. Examples of Dual Problems

In passing from $\Phi_{l,L}$ to $\Phi_{m,M}$, it is necessary to determine the conjugates of certain convex functions on R^{2n}, and this can be easy or difficult, depending on the nature of the functions in question. Many examples of conjugate convex functions are given in [21] and elsewhere in the literature. Often, as in the problems described in Section 4, l and $L(t, \cdot, \cdot)$ arise by various operations from other convex functions, as well as convex sets, and in such situations formulas like those in [21, Section 16] may be helpful in calculating conjugates.

The examples which follow indicate the calculation of the dual problem in some typical cases. They also bring out the fact that the dual problem can sometimes be finite-dimensional in character, and therefore more elementary in principle than the original problem.

Example 4. Let l and L be as in Example 1 in Section 4. We shall determine m and M. Setting $p = (q, h)$, as well as $x = (z, y)$, we have by (4.5)

$$\begin{aligned} m(p(0), p(T)) &= m(q(0), h(0), q(T), h(T)) \\ &= \sup\{\langle z(0), q(0)\rangle + \langle y(0), h(0)\rangle - \langle z(T), q(T)\rangle \\ &\qquad - \langle y(T), h(T)\rangle - l(z(0), y(0), z(T), y(T))\} \qquad (6.1) \\ &= \sup_{z(0)\in Z_0} \langle z(0), q(0)\rangle + \sup_{y(0)\in R^s} \langle y(0), h(0)\rangle \\ &\qquad + \sup_{z(T)\in Z_T} \langle z(T), -q(T)\rangle + \sup_{y(T)\in R^s} \langle y(T), -h(T)\rangle, \end{aligned}$$

or, in other words,

$$m(p(0), p(T)) = \begin{cases} f_0(q(0)) + f_T(-q(T)) & \text{if} \quad h(0) = 0 = h(T), \\ +\infty & \text{if} \quad h(0) \neq 0 \quad \text{or} \quad h(T) \neq 0, \end{cases} \tag{6.2}$$

where f_0 and f_T are the *support functions* of the convex sets Z_0 and Z_T, respectively (cf. [21, Section 13]). On the other hand, we have by (4.6)

$$\begin{aligned} M(t, p, s) &= m(t, q, h, r, k) \\ &= \sup_{z,y,w,u} \{\langle z, r\rangle + \langle y, k\rangle + \langle w, q\rangle + \langle u, h\rangle - L(t, z, y, w, u)\} \\ &= \sup_{z,y,u}\{\langle z, r\rangle + \langle y, k\rangle + \langle A(t)\, z + B(t)\, u, q\rangle + \langle u, h\rangle \\ &\quad - K_0(t, z, u)\}, \\ &= \sup_{y}\langle y, k\rangle + \sup_{z,u}\{\langle z, r + A^*(t)\, q\rangle + \langle u, h + B^*(t)\, q\rangle \\ &\quad - K_0(t, z, u)\}, \end{aligned} \tag{6.3}$$

where $A^*(t)$ and $B^*(t)$ denote the transposes of $A(t)$ and $B(t)$, respectively, and

$$K_0(t, z, u) = K(t, z, u) + \Psi(t, u), \tag{6.4}$$

$$\Psi(t, u) = \begin{cases} 0 & \text{if} \quad u \in U(t), \\ +\infty & \text{if} \quad u \notin U(t). \end{cases} \tag{6.5}$$

Let K^* and $K_0{}^*$ be the functions on $[0, T] \times R^r \times R^s$ such that, for each t, $K^*(t, \cdot, \cdot)$ is the conjugate of $K(t, \cdot, \cdot)$, and $K_0{}^*(t, \cdot, \cdot)$ is the conjugate of $K_0(t, \cdot, \cdot)$. Let Ψ^* be the function on $[0, T] \times R^s$ such that, for each t, $\Psi^*(t, \cdot)$ is the conjugate of $\Psi(t, \cdot)$, i.e., the support function of $U(t)$:

$$\Psi^*(t, h) = \sup\{\langle u, h\rangle \mid u \in U(t)\}. \tag{6.6}$$

Then from (6.3) we have

$$M(t, p, s) = \begin{cases} K_0{}^*(t, r + A^*(t)\, q, h + B^*(t)\, q) & \text{if} \quad k = 0, \\ +\infty & \text{if} \quad k \neq 0. \end{cases} \tag{6.7}$$

Moreover, according to [21, Theorem 16.4], we have

$$K_0{}^*(t, r', h') = -\min_{\bar{h} \in R^s}\{K^*(t, r', h' - \bar{h}) + \Psi^*(t, \bar{h})\}. \tag{6.8}$$

The convex problem of Bolza dual to the one in Example 1 in Section 4 is thus the problem of minimizing $\Phi_{m,M}$ over $A_n{}^1$, where m is given by (6.2) and M is given by (6.7). In other words, it is the problem of minimizing

$$f_0(q(0)) + f_T(-q(T)) + \int_0^T K_0^*(t, \dot{q}(t) + A^*(t)\, q(t), B^*(t)\, q(t))\, dt \qquad (6.9)$$

over all $q \in A_r^1$, where K_0^* is given by (6.8).

EXAMPLE 4′. To be more specific, let us suppose in Example 4 that Z_0 consists of a single point a, Z_T is a certain subspace of R^r with orthogonal complement $Z_T^\perp$, and for a given $\rho (1 \leqslant \rho \leqslant +\infty)$

$$U(t) = \{u \mid \| u \|_\rho \leqslant 1\}, \qquad (6.10)$$

$$K(t, z, u) = \alpha \| z \|_\rho + \beta \| u \|_\rho, \qquad (6.11)$$

where α and β are nonnegative constants and

$$\| u \|_\rho = [(u^1)^\rho + \cdots + (u^s)^\rho]^{1/\rho} \quad \text{if} \quad 1 \leqslant \rho < +\infty,$$
$$\| u \|_\infty = \max\{| u^1 |, \ldots, | u^s |\}.$$

(Note that K is not differentiable everywhere with respect to z if $\alpha > 0$, so that here we have a type of control problem not covered by the standard theory.) Then, as is easily verified,

$$f_0(q(0)) = \langle a, q(0) \rangle,$$
$$f_T(q(T)) = \begin{cases} 0 & \text{if} \quad q(t) \in Z_T^\perp, \\ +\infty & \text{if} \quad q(T) \notin Z_T^\perp. \end{cases}$$

Furthermore, we have

$$\Psi^*(t, h) = \| h \|_\sigma,$$
$$K^*(t, r, h) = \begin{cases} 0 & \text{if} \quad \| r \|_\sigma \leqslant \alpha \quad \text{and} \quad \| h \|_\sigma \leqslant \beta, \\ +\infty & \text{if} \quad \| r \|_\sigma > \alpha \quad \text{or} \quad \| h \|_\sigma > \beta, \end{cases}$$

where $(1/\rho) + (1/\sigma) = 1$, and therefore by (6.8)

$$K_0^*(t, r', h') = \begin{cases} \max\{\| h' \|_\sigma - \beta, 0\} & \text{if} \quad \| r' \|_\sigma \leqslant \alpha, \\ +\infty & \text{if} \quad \| r' \|_\sigma > \alpha. \end{cases}$$

Hence in this case the dual convex problem of Bolza consists of minimizing

$$\langle a, q(0) \rangle + \int_0^T \max\{\| B^*(t)\, q(t) \|_\sigma - \beta, 0\}\, dt \qquad (6.12)$$

in $q \in A_r^1$, subject to the constraints

$$q(T) \in Z_T^{\perp}, \tag{6.13}$$

$$\| \dot{q}(t) + A^*(t)\, q(t) \|_\sigma \leqslant \alpha \text{ for almost every } t. \tag{6.14}$$

Note that, if $\alpha = 0$, we have

$$\dot{q}(t) = -A^*(t)\, q(t) \tag{6.15}$$

by (6.14), so that (assuming, say, that the matrix components in (6.15) are summable functions of t) $q(t)$ is determined for all t by $q(0)$, and the dual problem is essentially finite-dimensional. In fact, the dual problem consists of minimizing a certain finite (everywhere defined, nondifferentiable) convex function of $q(0) \in R^s$ subject to a finite system of linear equations [representing the constraint (6.13)].

Similarly, in the more general dual problem where one minimizes (6.9), it can be seen that the constraint (6.15), and hence the finite-dimensionality of the problem, will be implicit whenever $K(t, z, u)$ is actually independent of z.

EXAMPLE 5. Suppose the Lagrangian function L can be expressed in the form

$$L(t, x, v) = f(t, x) + g(t, v - E(t)\, x), \tag{6.16}$$

where $E(t)$ denotes an $n \times m$ matrix, and $f(t, \cdot)$ and $g(t, \cdot)$ are lower semi-continuous convex functions from R^n to $R^1 \cup \{+\infty\}$, not identically $+\infty$. (This is true in Example 1, for instance, if

$$K(t, z, u) = K_1(t, z) + K_2(t, u),$$

where $K_1(t, \cdot)$ and $K_2(t, \cdot)$ are convex functions.) An L of this form satisfies (A), and it also satisfies (B), as can be shown by the arguments similar to those in Example 1 of Section 4, if the components of $E(t)$ are Lebesgue measurable functions of t, and f and g are Lebesgue normal in the sense of [22]. The dual Lagrangian M may then be computed directly from (5.7) as

$$M(t, p, s) = g^*(t, p) + f^*(t, s + E^*(t)\, p), \tag{6.17}$$

where $E^*(t)$ is the transpose of $E(t)$, and $f^*(t, \cdot)$ and $g^*(t, \cdot)$ are the conjugates of $f(t, \cdot)$ and $g(t, \cdot)$, respectively. Thus, in the given problem of Bolza one minimizes

$$l(x(0), x(T)) + \int_0^T f(t, x(t))\, dt + \int_0^T g(t, u(t))\, dt \tag{6.18}$$

subject to

$$\dot{x}(t) = E(t)\,x(t) + u(t), \tag{6.19}$$

while in the dual problem one minimizes

$$m(p(0), p(T)) + \int_0^T g^*(t, p(t))\,dt + \int_0^T f^*(t, w(t))\,dt \tag{6.20}$$

subject to

$$\dot{p}(t) = -E^*(t)\,p(t) + w(t). \tag{6.21}$$

If $f(t, x) \equiv 0$, then

$$f^*(t, w) = \begin{cases} 0 & \text{if} \quad w = 0, \\ +\infty & \text{if} \quad w \neq 0, \end{cases}$$

so that in the dual problem one actually minimizes

$$m(p(0), p(T)) + \int_0^T g^*(t, p(t))\,dt \tag{6.22}$$

over all solutions p to the differential equation

$$\dot{p}(t) = -E^*(t)\,p(t) \text{ for almost every } t. \tag{6.23}$$

7. Conjugates of Bolza Functionals

The dual of A_n^1 can be identified with the Banach space B_n^∞ under the pairing (2.3), as already pointed out in Section 2, and the convex functions $\Phi_{l,L}$ and $\Phi_{m,M}$ on A_n^1 have certain conjugates on B_n^∞ with respect to this pairing. The study of these conjugates will reveal the connection between the problems of Bolza corresponding to $\Phi_{l,L}$ and $\Phi_{m,M}$. It will be seen, in fact, that the conjugate of $\Phi_{m,M}$ describes the behavior, under perturbations, of the infimum in the problem of minimizing $\Phi_{l,L}$, while the conjugate of $\Phi_{l,L}$ describes the behavior, under perturbations, of the infimum in the problem of minimizing $\Phi_{m,M}$.

Given any $(a, y) \in B_n^\infty$, we denote by $\Phi_{l,L}^{a,y}$ the Bolza functional obtained by replacing l by l^a and L by L^y, where

$$l^a(c_0, c_T) = l(c_0 + a, c_T), \tag{7.1}$$

$$L^y(t, x, v) = L(t, x + y(t), v). \tag{7.2}$$

Thus by definition

$$\Phi_{l,L}^{a,y}(x) = l(x(0) + a, x(T)) + \int_0^T L(t, x(t) + y(t), \dot{x}(t))\,dt. \tag{7.3}$$

It is easy to see that our assumptions (A), (B), (C), and (D) on l and L imply that l^a and L^y satisfy these same assumptions, so that $\Phi_{l,L}^{a,y}$ is well-defined on A_n^1 by Theorem 1.

We define the functional $\varphi_{l,L}$ on B_n^∞ by

$$\varphi_{l,L}(a, y) = \inf\{\Phi_{l,L}^{a,y}(x) \mid x \in A_n^1\}. \tag{7.4}$$

Of course $\Phi_{l,L}^{a,y} = \Phi_{l,L}$ for $(a, y) = (0, 0)$, and consequently

$$\varphi_{l,L}(0, 0) = \inf\{\Phi_{l,L}(x) \mid x \in A_n^1\}. \tag{7.5}$$

We shall be interested in the lower semicontinuity of $\varphi_{l,L}$ at $(a, y) = (0, 0)$ with respect to the *weak* topology* on B_n^∞, in other words, the weak topology on B_n^∞ induced by A_n^1 under the pairing (2.3).

We define the functional $\Phi_{m,M}^{b,q}$ on A_n^1 for $(b, q) \in B_n^\infty$ similarly by

$$\Phi_{m,M}^{b,q}(p) = m(p(0) + b, p(T)) + \int_0^T M(t, p(t) + q(t), \dot{p}(t))\, dt, \tag{7.6}$$

and we define $\varphi_{m,M}$ on B_n^∞ by

$$\varphi_{m,M}(b, q) = \inf\{\Phi_{m,M}^{b,q}(p) \mid p \in A_n^1\}, \tag{7.7}$$

where in particular

$$\varphi_{m,M}(0, 0) = \inf\{\Phi_{m,M}(p) \mid p \in A_n^1\}. \tag{7.8}$$

THEOREM 3. *The functions $\varphi_{l,L}$ and $\varphi_{m,M}$ are convex on B_n^∞, and their conjugates on A_n^1 [with respect to the pairing (2.3)] are given by*

$$\varphi_{l,L}^* = \Phi_{m,M} \quad \text{and} \quad \varphi_{m,M}^* = \Phi_{l,L}. \tag{7.9}$$

The conjugate of $\Phi_{l,L}$ on B_n^∞ is in turn given by the formula

$$\Phi_{l,L}^*(b, q) = \liminf_{(b',q') \to (b,q)} \varphi_{m,M}(b', q'), \tag{7.10}$$

except in the case where there are no feasible arcs for $\Phi_{l,L}$ at all, and where at the same time $\varphi_{m,M}$ is identically $+\infty$ on some weak neighborhood of (b, q). Similarly, the conjugate of $\Phi_{m,M}$ on B_n^∞ is given by*

$$\Phi_{m,M}^*(a, y) = \liminf_{(a',y') \to (a,y)} \varphi_{l,L}(a', y'), \tag{7.11}$$

except in the case where there are no feasible arcs for $\Phi_{m,M}$ at all, and where at the same time $\varphi_{l,L}$ is identically $+\infty$ on some weak neighborhood of (a, y). Here the limits are to be taken over all nets converging in the weak* topology to the indicated points.*

Proof. The key to the proof is a fact which we have already established elsewhere [22]: the convex functions I_L and I_M on $L_n^\infty \oplus L_n^1$ [see (3.2) and Proposition 1] are conjugate to each other with respect to the pairing

$$\langle (x, v), (p, s) \rangle = \int_0^T \langle x(t), s(t) \rangle \, dt + \int_0^T \langle v(t), p(t) \rangle \, dt \qquad (7.12)$$

between $L_n^\infty \oplus L_n^1$ and itself. In other words,

$$I_M(p, s) = \sup_{x \in L_n^\infty} \sup_{v \in L_n^1} \{\langle (x, v), (p, s) \rangle - I_L(x, v)\} \qquad (7.13)$$

for every $p \in L_n^\infty$ and $s \in L_n^1$, and dually. This is a special case of Theorem 2 of [22], in view of conditions (A), (B), (C), and (D) on L and the definition (5.5) of M. Calculating the conjugate function $\varphi_{l,L}^*$ on A_n^1 directly from the definitions, we have, using this fact,

$$\begin{aligned}
\varphi_{l,L}^*(p) &= \sup\{\langle (a, y), p \rangle - \varphi_{l,L}(a, y) \mid (a, y) \in B_n^1\} \\
&= \sup_{a \in R^n} \sup_{y \in L_n^\infty} \{\langle (a, y), p \rangle - \inf_{x \in A_n^1} \Phi_{l,L}^{a,y}(x)\} \\
&= \sup_{a \in R^n} \sup_{y \in L_n^\infty} \sup_{x \in A_n^1} \Big\{ \langle a, p(0) \rangle + \int_0^T \langle y(t), \dot{p}(t) \rangle \, dt \\
&\qquad - l(x(0) + a, x(T)) - \int_0^T L(t, x(t) + y(t), \dot{x}(t)) \, dt \Big\} \\
&= \sup_{a' \in R^n} \sup_{y' \in L_n^\infty} \sup_{x \in A_n^1} \Big\{ \langle a' - x(0), p(0) \rangle + \int_0^T \langle y'(t) - x(t), \dot{p}(t) \rangle \, dt \\
&\qquad - l(a', x(T)) - \int_0^T L(t, y'(t), \dot{x}(t)) \, dt \Big\} \\
&= \sup_{a' \in R^n} \sup_{y' \in L_n^\infty} \sup_{x \in A_n^1} \Big\{ \langle a', p(0) \rangle + \int_0^T \langle y'(t), \dot{p}(t) \rangle \, dt - \langle x(T), p(T) \rangle \\
&\qquad + \int_0^T \langle \dot{x}(t), p(t) \rangle \, dt - l(a', x(T)) - \int_0^T L(t, y'(t), \dot{x}(t)) \, dt \Big\} \\
&= \sup_{a' \in R^n} \sup_{y' \in L_n^\infty} \sup_{x \in A_n^1} \{\langle a', p(0) \rangle - \langle x(T), p(T) \rangle - l(a', x(T)) \\
&\qquad + \langle (y', \dot{x}), (p, \dot{p}) \rangle - I_L(y', \dot{x})\} \\
&= \sup_{a' \in R^n} \sup_{c \in R^n} \{\langle a', p(0) \rangle - \langle c, p(T) \rangle - l(a', c)\} \\
&\qquad + \sup_{y' \in L_n^\infty} \sup_{v \in L_n^1} \{\langle (y', v), (p, \dot{p}) \rangle - I_L(y', v)\} \\
&= m(p(0), p(T)) + I_M(p, \dot{p}) = \Phi_{m,M}(p).
\end{aligned}$$

Thus $\varphi^*_{l,L} = \Phi_{m,M}$, as claimed. Consequently, as explained at the beginning of Section 5, the conjugate function $\Phi^*_{m,M}$ on $B_n{}^\infty$ is given by

$$\Phi^*_{m,M}(a, y) = \varphi^{**}_{l,M}(a, y) = \liminf_{(a',y')\to(a,y)} \varphi_{l,L}(a', y'), \tag{7.14}$$

except in the case where $\varphi^*_{l,L} = \Phi_{m,M}$ is identically $+\infty$ on $A_n{}^1$ and $\varphi_{l,L}$ is identically $+\infty$ in some neighborhood of (a, y). Here the neighborhood of (a, y) and the limit in (7.14) may be taken with respect to any locally convex topology on $B_n{}^\infty$ such that $A_n{}^1$ can be identified with the space of all continuous linear functionals on $B_n{}^\infty$ under the pairing (2.3). The weak* topology, in particular, meets this requirement.

This proves (7.11) and the first half of (7.9). Formula (7.10) and the second half of (7.9) follow by symmetry, and Theorem 3 is thereby established.

A slight generalization of the above argument enables one to determine also the conjugates on $B_n{}^\infty$ of the restrictions of $\Phi_{l,L}$ and $\Phi_{m,M}$ to the spaces $A_n{}^p$, $1 < p \leqslant +\infty$. (In what follows, the letters p and q will be used to denote the traditional Lebesgue exponents, as well as functions, but no confusion should arise if the reader bears this in mind.)

THEOREM 3′. *Let* $1 < q_0 \leqslant +\infty$ *and* $(1/p_0) + (1/q_0) = 1$, *and suppose that condition* (C) *can be satisfied with* $s \in L_n^{q_0}$. *Then the conjugate of* $\varphi_{l,L}$ *on* $A_n^{q_0}$, *with respect to the pairing* (2.3) *between* $A_n^{q_0}$ *and* $B_n{}^\infty$, *is the restriction of* $\Phi_{m,M}$ *to* $A_n^{q_0}$. *The conjugate on* $B_n{}^\infty$ *of the restriction of* $\Phi_{m,M}$ *to* $A_n^{q_0}$ *is in turn given by* (7.11), *except when* $\Phi_{m,M}$ *has no feasible arc in* $A_n^{q_0}$ *and* $\varphi_{l,L}$ *is identically* $+\infty$ *on some neighborhood of* (a, y). *Here the neighborhood of* (a, y) *is taken with respect to the* $B_n^{p_0}$ *norm* (2.1) *on* $B_n{}^\infty$, *and the limit in* (7.11) *is taken over all sequences in* $B_n{}^\infty$ *converging to* (a, y) *in this norm.*

Moreover, the same facts hold here, and in Theorem 3, *if the infimum in the definition of* $\Phi_{l,L}$ *is taken over* $A_n^{p_1}$ *instead of* $A_n{}^1$, *where* $1 < p_1 \leqslant +\infty$, *provided that condition* (D) *on L can be satisfied with* $v \in L_n^{p_1}$ *(as is true in particular when* $\Phi_{l,L}$ *has a feasible arc in* $A_n^{p_1}$*).*

The roles of (l, L) *and* (m, M) *can be reversed in the preceding to obtain a dual result.*

Proof. Let p_1 be such that (D) can be satisfied with $v \in L_n^{p_1}$, where $1 \leqslant p_1 \leqslant +\infty$, and consider the pairing (7.12) between the spaces $X = L_n{}^\infty \oplus L_n^{p_1}$ and $Y = L_n{}^\infty \oplus L_n^{q_0}$. The convex functions I_L on X and I_M on Y are well-defined and conjugate to each other with respect to this pairing; this is another special case of the result cited in the proof of Theorem 3, namely Theorem 2 of [22]. Using this fact, we may calculate $\varphi^*_{l,L}$ just as before and see that $\varphi^*_{l,L} = \Phi_{m,M}$ (restricted to $A_n^{q_0}$). It follows, then, that the conjugate of the latter is given by formula (7.11), except in the case noted, but the weak * topology must be replaced by a locally convex topology on B_n^∞ with respect to which the continuous linear functionals on B_n^∞ can be identified with the elements of $A_n^{q_0}$ under the pairing (2.3). Since B_n^∞ is dense in $B_n^{p_0}$ and $1 \leqslant p_0 < +\infty$, the $B_n^{p_0}$ topology on B_n^∞ satisfies this condition.

Remark. It can also be shown by the same arguments that, when condition (C) on L can be satisfied with $s \in L_n^{q_0}$ (as is true in particular when $\Phi_{m,M}$ has a feasible arc in $A_n^{q_0}$), then $\varphi_{l,L}$ is actually a well-defined convex function on all of $B_n^{p_0}$, and the conjugacy formulas in Theorem 3′ hold with respect to the pairing (2.3) between $A_n^{q_0}$ and $B_n^{p_0}$, rather than between $A_n^{q_0}$ and B_n^∞.

The significance of Theorems 3 and 3′ for the study of convex problems of Bolza lies in the extensive duality between properties of the nest of (convex) level sets $\{x \mid \Phi_{l,L}(x) \leqslant \mu\}$ of $\Phi_{l,L}$ and the behavior of the conjugate function $\Phi_{l,L}^*$ at the origin. For instance, boundedness or compactness properties of the level sets of $\Phi_{l,L}$ correspond to continuity properties of $\Phi_{l,L}^*$ at 0, while the manner in which the minimum of $\Phi_{l,L}$ is attained can be analyzed in terms of the differentiability of $\Phi_{l,L}^*$ at 0 in various senses. Many facts of this type have been established in the general theory of conjugate convex functions by E. Asplund, J. J. Moreau and the author; see [1], [15], and [21, Section 27]. In the present paper we shall not exploit this duality to the fullest, but the reader can get some idea of what is possible by examining the theory of finite-dimensional extremum problems in [21, Part 6].

8. Duality of Infima

We shall now derive some facts relating the infimum in a given convex problem of Bolza and the infimum in the corresponding dual problem.

A basic fact is near the surface. Consider any x and p in A_n^1. By the definition of m and M (see formulas (5.6) and (5.7)), we have

$$l(x(0), x(T)) + m(p(0), p(T)) \geqslant \langle x(0), p(0)\rangle - \langle x(T), p(T)\rangle, \tag{8.1}$$

$$L(t, x(t), \dot{x}(t)) + M(t, p(t), \dot{p}(t)) \geqslant \langle x(t), \dot{p}(t)\rangle + \langle \dot{x}(t), p(t)\rangle. \tag{8.2}$$

Integrating (8.2), we get

$$\begin{aligned}\int_0^T L(t, x(t), \dot{x}(t))\, dt + \int_0^T M(t, p(t), \dot{p}(t))\, dt &\geqslant \int_0^T [\langle x(t), \dot{p}(t)\rangle + \langle \dot{x}(t), p(t)\rangle]\, dt \\ &= \int_0^T \frac{d}{dt}\langle x(t), p(t)\rangle\, dt \qquad (8.3) \\ &= \langle x(T), p(T)\rangle - \langle x(0), p(0)\rangle.\end{aligned}$$

The latter inequality, when added to (8.1), yields the following result.

PROPOSITION 2. *For every* $x \in A_n^1$ *and* $p \in A_n^1$, *one has*

$$\Phi_{l,L}(x) + \Phi_{m,M}(p) \geqslant 0. \tag{8.4}$$

Thus every $p \in A_n^1$ *furnishes a lower bound* $-\Phi_{m,M}(p)$ *to* $\Phi_{l,L}$ *on* A_n^1, *while every* $x \in A_n^1$ *furnishes a lower bound* $-\Phi_{l,L}(x)$ *to* $\Phi_{m,M}$ *on* A_n^1, *and one has*

$$\inf\{\Phi_{l,L}(x) \mid x \in A_n^1\} \geqslant -\inf\{\Phi_{m,M}(p) \mid p \in A_n^1\}. \tag{8.5}$$

An obvious question to ask is whether, under some kind of general conditions, equality holds in (8.5). An answer can be provided, via Theorems 3 and 3′, in terms of lower semicontinuity properties of the infima in (8.5) with respect to certain "perturbations" of the functionals $\Phi_{l,L}$ and $\Phi_{m,M}$, namely the "perturbations" which replace these functionals by $\Phi_{l,L}^{a,y}$ and $\Phi_{m,M}^{b,q}$, as defined in (7.3) and (7.6), for various pairs (a, y) and (b, q) near the origin of B_n^∞.

We assume here, of course, as everywhere else, that (A), (B), (C), and (D) are satisfied.

THEOREM 4. *Let* $1 \leqslant p_i \leqslant +\infty$ *and* $(1/p_i) + (1/q_i) = 1$ *for* $i = 0, 1$. *Suppose that conditions* (C) *and* (D) *can be satisfied with* $s \in L_v^{q_0}$ *and* $v \in L_n^{p_1}$, *respectively. If either* $\Phi_{l,L}$ *has a feasible arc in* $A_n^{p_1}$ *or* $\Phi_{m,M}$ *has a feasible arc in* $A_n^{q_0}$, *then*

$$\inf\{\Phi_{l,L}(x) \mid x \in A_n^{p_1}\} = -\liminf_{(b,q)\to(0,0)}(\inf\{\Phi_{m,M}^{b,q}(p) \mid p \in A_n^{q_0}\}), \tag{8.6}$$

$$\liminf_{(a,y)\to(0,0)}(\inf\{\Phi_{l,L}^{a,y}(x) \mid x \in A_n^{p_1}\}) = -\inf\{\Phi_{m,M}(p) \mid p \in A_n^{q_0}\}. \tag{8.7}$$

Here, if $p_1 = 1$ *and* $q_1 = +\infty$, *the limit in* (8.6) *is to be taken over all nets in* B_n^∞ *converging to* $(0, 0)$ *in the weak* topology, while, if* $p_1 > 1$ *and* $q_1 < +\infty$, *it is to be taken over all sequences in* B_n^∞ *converging to* $(0, 0)$ *in the* $B_n^{q_1}$ *norm. Similarly, if* $p_0 = +\infty$ *and* $q_0 = 1$, *the limit in* (8.7) *is to be taken over all nets in* B_n^∞ *converging to* $(0, 0)$ *in the weak* topology, while, if* $p_0 < +\infty$ *and* $q_0 > 1$, *it is to be taken over all sequences converging to* $(0, 0)$ *in the* $B_n^{p_0}$ *norm.*

Proof. Assume first that $p_1 = 1$ and $q_0 = 1$, and consider formulas (7.10) and (7.11) in Theorem 3 in the case where $(a, y) = (0, 0)$ and $(b, q) = (0, 0)$. By definition of the conjugates, we have

$$\Phi_{l,L}^*(0) = -\inf\{\Phi_{l,L}(x) \mid x \in A_n^1\}, \tag{8.8}$$

$$\Phi_{m,M}^*(0) = -\inf\{\Phi_{m,M}(p) \mid p \in A_n^1\}. \tag{8.9}$$

If a feasible arc exists for $\Phi_{l,L}$ in A_n^1, then (7.10) is valid, and hence (8.6). Furthermore, in this case $\varphi_{l,L}(0) < +\infty$, so that the lim inf in (7.11) is not

$+\infty$, and (7.11) is valid, yielding (8.7). Similarly, both (8.6) and (8.7) are valid if a feasible arc exists for $\Phi_{m,M}$ in A_n^1.

The case of general exponents p_i and q_i follows in exactly the same way from Theorem 3′.

COROLLARY. *If a feasible arc exists for either $\Phi_{l,L}$ or $\Phi_{m,M}$ in A_n^1, then the following assertions are equivalent, where the limits are taken over all nets in B_n^∞ converging in the weak* topology:*

(a) $\inf_{x \in A_n^1} \Phi_{l,L}(x) = - \inf_{p \in A_n^1} \Phi_{m,M}(p)$,

(b) $\inf_{x \in A_n^1} \Phi_{l,L}(x) = \liminf_{(a,y)\to(0,0)} \left(\inf_{x \in A_n^1} \Phi_{l,L}^{a,y}(x)\right)$,

(c) $\inf_{p \in A_n^1} \Phi_{m,M}(p) = \liminf_{(b,q)\to(0,0)} \left(\inf_{p \in A_n^1} \Phi_{m,M}^{b,q}(p)\right)$.

It is possible to develop "reasonable" conditions of some generality on l and L guaranteeing that the three equivalent properties in this corollary are present. However, this is a lengthy undertaking in itself, and we therefore relegate it to a separate paper [27]. Here we shall only give examples to show that the properties do hold in some cases and do not hold in other cases.

EXAMPLE 6. Assume there exist $n \times n$ matrices $E(t)$ and a real number ρ such that

$$(c_0, c_T) \in C \quad \text{implies} \quad |c_0| \leqslant \rho,$$
$$(x, v) \in D(t) \quad \text{implies} \quad |v - E(t)\,x| \leqslant \rho,$$

where C and $D(t)$ are the convex sets defined in (2.6) and (2.7), respectively, and the components of $E(t)$ are measurable, essentially bounded functions of $t \in [0, T]$. (This is satisfied in particular, of course, if C is bounded and the union of the sets $D(t)$, $0 \leqslant t \leqslant T$, is bounded, in which case one can take $E(t) = 0$.) Under this assumption, the level sets

$$\{x \in A_n^1 \mid \Phi_{l,L}(x) \leqslant \mu\}, \qquad \mu \in R^1,$$

are weakly compact in A_n^1, in view of Theorem 1, and therefore the Bolza functional $\Phi_{l,L}$ attains its minimum somewhere on A_n^1. In other words, the convex problem of Bolza corresponding to l and L has an optimal arc, provided that it has at least one feasible arc. We now show that, under the same assumption, properties (a), (b), and (c) of the Corollary to Theorem 4 are present. To do this, it is enough to prove the following result.

LEMMA. *Under the above assumption, the function $\varphi_{m,M}$, if not identically $-\infty$ on B_n^∞, is everywhere finite and lower semicontinuous in the weak* topology. In fact, $\varphi_{m,M}$ is everywhere continuous on B_n^∞ with respect to the B_n^1 norm.*

Proof. Let $s \in L_n^1$, $p \in L_n^\infty$, and $\alpha \in L_1^1$ be functions with the property in condition (C) of Section 2, and let $d_0 \in R^n$ and $d_T \in R^n$ be points such that $\alpha_0 = m(d_0, d_T) < +\infty$. Let $\bar{p}$ be the arc in A_n^1 such that $\bar{p}(T) = d_T$ and $\dot{\bar{p}} = s$, and let $\bar{q} = p - \bar{p}$ and $\bar{b} = d_0 - \bar{p}(0)$. Then $l \geqslant l_0$ and $L \geqslant L_0$, where

$$l_0(c_0, c_T) = \begin{cases} \langle c_0, \bar{p}(0) + \bar{b}\rangle - \langle c_T, \bar{p}(T)\rangle - \alpha_0 & \text{if} \quad | c_0 | \leqslant \rho, \\ +\infty & \text{if} \quad | c_0 | > \rho, \end{cases}$$

$$L_0(t, x, v) = \begin{cases} \langle x, \dot{\bar{p}}(t)\rangle + \langle v, \bar{p}(t) + \bar{q}(t)\rangle - \alpha(t) & \text{if} \quad | v - E(t)\, x | \leqslant \rho, \\ +\infty & \text{if} \quad | v - E(t)\, x | > \rho. \end{cases}$$

Therefore $m \leqslant m_0$ and $M \leqslant M_0$, where m_0 and M_0 are the functions dual to l_0 and L_0 in the sense of formulas (5.6) and (5.7). For each $q \in L_n^\infty$, let p_q denote the unique arc in A_n^1 such that $p_q(T) = \bar{p}(T)$ and

$$\dot{p}_q(t) - \dot{\bar{p}}(t) = -E^*(t)\,[p_q(t) - \bar{p}(t)] + E^*(t)\,[q(t) - \bar{q}(t)]$$

for almost every t. [Here $E^*(t)$ is the transpose of $E(t)$.] Note that $p_q = \bar{p}$ for $q = \bar{q}$, and that the definition of p_q implies (by the essential boundedness of the components of $E(t)$ as functions of t) the existence of a constant σ, independent of q, such that

$$| p_q(t) - \bar{p}(t) | \leqslant \sigma \, \| q - \bar{q} \|_1 , \qquad 0 \leqslant t \leqslant T.$$

For any $(b, q) \in B_n^\infty$, we have (by direct calculation)

$$m_0(p_q(0) + b, p_q(T)) = \alpha_0 + \rho \, | p_q(0) - \bar{p}(0) + b - \bar{b} | ,$$
$$M_0(t, p_q(t) + q(t), \dot{p}_q(t)) = \alpha(t) + \rho \, | p_q(t) - \bar{p}(t) + q(t) - \bar{q}(t) | .$$

Since $m \leqslant m_0$ and $M \leqslant M_0$, it follows that

$$\begin{aligned} \Phi_{m,M}^{b,q}(p_q) &= m(p_q(0) + b, p_q(T)) + \int_0^T M(t, p_q(t) + q(t), \dot{p}_q(t))\, dt \\ &\leqslant \bar{\alpha} + \rho \Big[| p_q(0) - \bar{p}(0) | + | b - \bar{b} | \\ &\qquad + \int_0^T | p_q(t) - \bar{p}(t) |\, dt + \int_0^T | q(t) - \bar{q}(t) |\, dt \Big] \\ &\leqslant \bar{\alpha} + \rho \, | b - \bar{b} | + \rho[1 + \sigma(1 + T)] \, \| q - \bar{q} \|_1 , \end{aligned}$$

where

$$\bar{\alpha} = \alpha_0 + \int_0^T \alpha(t)\, dt.$$

On the other hand, we have

$$\varphi_{m,M}(b, q) \leqslant \Phi_{m,M}^{b,q}(p_q)$$

by the definition of $\varphi_{m,M}$ in Section 7. Thus there exists a constant $\bar{\rho}$ such that

$$\varphi_{m,M}(b, q) \leqslant \bar{\alpha} + \bar{\rho}\, \|(b, q) - (\bar{b}, \bar{q})\|_1$$

for every $(b, q) \in B_n^\infty$. This implies that $\varphi_{m,M}$ is bounded above in a $\|\cdot\|_1$-neighborhood of every point of B_n^∞. Since $\varphi_{m,M}$ is convex (Theorem 3), we may conclude that $\varphi_{m,M}$ is either identically $-\infty$ or finite everywhere, and that $\varphi_{m,M}$ is continuous everywhere on B_n^∞ with respect to the norm $\|\cdot\|_1$ (see [4, Chap. 2, p. 92]). Therefore, by the observation in the last paragraph of the proof of Proposition 1, $\varphi_{m,M}$ is lower semicontinuous with respect to the weak topology induced on B_n^∞ by the $\|\cdot\|_1$-continuous linear functionals on B_n^∞, in other words, the weak topology induced on B_n^∞ by A_n^∞ under the pairing (2.3). Then, *a fortiori*, $\varphi_{m,M}$ is lower semicontinuous with respect to stronger topology induced in the same way by A_n^1, which is the weak* topology on B_n^∞.

EXAMPLE 7. Consider the special case of Example 5 in which L [satisfying (A), (B), (C), and (D)) can be expressed in the form

$$L(t, x, v) = g(t, v - E(t)\, x), \tag{8.10}$$

where $E(t)$ is an $n \times n$ matrix whose components are summable functions of t, and g is a function from $[0, T] \times R^n$ to $R^1 \cup \{+\infty\}$. Let $E^*(t)$ denote the transpose of $E(t)$, and $g^*(t, \cdot)$ the conjugate of $g(t, \cdot)$, for each t.

LEMMA. *Under these assumptions, the three properties in the Corollary to Theorem 4 are present (and moreover an optimal arc exists for $\Phi_{m,M}$ in A_n^1), whenever the following conditions are satisfied:*

(a) *There exists at least one feasible arc for $\Phi_{m,M}$, in other words (in view of the formula for $\Phi_{m,M}$ determined in Example 5) at least one solution $p \in A_n^1$ to the differential equation*

$$\dot{p}(t) = -E^*(t)\, p(t) \tag{8.11}$$

such that

$$m(p(0), p(T)) + \int_0^T g^*(t, p(t))\, dt < +\infty. \tag{8.12}$$

(b) *If $\bar{p}$ is a solution to* (8.11) *such that the expression*

$$m(p(0) + \lambda\bar{p}(0), p(T) + \lambda\bar{p}(T)) + \int_0^T g^*(t, p(t) + \lambda\bar{p}(t))\, dt \qquad (8.13)$$

is a nonincreasing function of $\lambda \in R^1$ for every arc p satisfying (8.11) *and* (8.12), *then $-\bar{p}$ has this same property.*

Proof. Condition (b) means, in the sense of [21, Section 8], that the convex function $\Phi_{m,M}$ has no directions of recession other than directions of constancy. Since $\Phi_{m,M}$ is lower semicontinuous by Proposition 1, and also finite-dimensional, that is, the set of all p such that $\Phi_{m,M}(p) < +\infty$ forms a finite-dimensional subset of $A_n{}^1$ (because every such p satisfies (8.11)), it follows from this that $\Phi_{m,M}$ attains its minimum (finitely) on $A_n{}^1$ [21, Theorem 27.1b]. Thus an optimal arc exists for $\Phi_{m,M}$.

The idea is now to pass from the finite-dimensionality of $\Phi_{m,M}$ to a dual property, the finite-codimensionality of $\varphi_{l,L}$, and thereby reduce the limit in (b) of the Corollary to Theorem 4 to a more elementary finite-dimensional case.

Let N be the finite-dimensional subspace of $A_n{}^1$ consisting of all p satisfying (8.11), and let $N^\perp$ be the annihilator of N in $B_n{}^\infty$ with respect to the pairing (2.3). The quotient space $B_n{}^\infty/N^\perp$ is finite-dimensional, and it is paired with N in a natural way. Suppose it can be shown that $\varphi_{l,L}(a, y)$ depends only on the canonical image of (a, y) in $B_n{}^\infty/N^\perp$, so that $\varphi_{l,L}$ can be regarded as a convex function on $B_n{}^\infty/N^\perp$ whose conjugate on N is the restriction of $\Phi_{m,M}$ to N. Then, since $\Phi_{m,M}$ has no directions of recession other than directions of constancy, it will follow from [21, Theorems 7.4 and 27.1b] that $\varphi_{l,L}$, as a function on $B_n{}^\infty/N^\perp$, is lower semicontinuous at the origin (in the natural finite-dimensional topology). The latter property implies that $\varphi_{l,L}$, as a function on $B_n{}^\infty$, is lower semicontinuous at the origin in the weak* topology, and hence that condition (b) holds in the Corollary to Theorem 4.

To show that $\varphi_{l,L}(a, y)$ depends only on the canonical image of (a, y) in $B_n{}^\infty/N^\perp$, we observe first that [by direct calculation using the fundamental matrix of the differential equation (8.11)] $N^\perp$ consists of the pairs (c, w) such that $c \in R^n$, $w \in L_n{}^\infty$, and there exists a $z \in A_n{}^1$ with $z(0) = -c$, $z(T) = 0$, and

$$\dot{z}(t) = E(t)\,[z(t) + w(t)] \text{ for almost every } t.$$

Now, by definition, $\varphi_{l,L}(a + c, y + w)$ is the infimum of

$$l(x(0) + a + c, x(T)) + \int_0^T g(t, \dot{x}(t) - E(t)\,[x(t) + y(t) + w(t)])\, dt \qquad (8.14)$$

over all $x \in A_n{}^1$. If $(c, w) \in N^\perp$, and z corresponds to (c, w) as just described, expression (8.14) can be rewritten as

$$l(u(0) + a, u(T)) + \int_0^T g(t, \dot{u}(t) - E(t)[u(t) + y(t)])\, dt, \qquad (8.15)$$

where $u = x - z \in A_n{}^1$. Thus the infimum of (8.14) over all $x \in A_n{}^1$ is the same as the infimum of (8.15) over all $u \in A_n{}^1$. In other words,

$$\varphi_{l,L}(a + c, y + w) = \varphi_{l,L}(a, y) \qquad \text{for any} \qquad (c, w) \in N^\perp,$$

and the proof is complete.

EXAMPLE 8. This is a counterexample showing that the three equivalent properties in the Corollary to Theorem 4 do not always hold, even when optimal arcs exist for both $\Phi_{l,L}$ and $\Phi_{m,M}$. It also shows that Theorem 4 and the Corollary would fail if the weak* convergence were replaced by convergence in the $B_n{}^\infty$ norm.

Let $n = 1$, and define L on $[0, T] \times R^1 \times R^1$ by

$$L(t, x, v) = \begin{cases} 0 & \text{if} \quad x \geqslant 0, \\ +\infty & \text{if} \quad x < 0. \end{cases} \qquad (8.16)$$

Define l on $R^1 \times R^1$ by

$$l(c_0, c_T) = c_0 . \qquad (8.17)$$

It is obvious that condition (A) is satisfied, and, since L is independent of t, conditions (B), (C), and (D) are satisfied too. We have

$$\Phi_{l,L}(x) = \begin{cases} x(0) & \text{if} \quad x(t) \geqslant 0 \quad \text{for all } t, \\ +\infty & \text{otherwise}, \end{cases} \qquad (8.18)$$

so that trivially

$$\inf\{\Phi_{l,L}(x) \mid x \in A_n{}^1\} = \varphi_{l,L}(0) = 0. \qquad (8.19)$$

On the other hand, for $(a, y) \in B_n{}^\infty$ we have

$$\Phi_{l,L}^{a,y}(x) = \begin{cases} x(0) & \text{if} \quad x(t) \geqslant -y(t) \quad \text{for almost every } t, \\ +\infty & \text{otherwise}. \end{cases}$$

Thus the inequality $\|(a, y)\|_\infty \leqslant \delta$ implies that $\varphi_{l,L}(a, y) \geqslant -\delta$. Therefore

$$\text{strong} \liminf_{(a,y)\to(0,0)} \varphi_{l,L}(a, y) = \varphi_{l,L}(0, 0) = 0, \qquad (8.20)$$

where the limit is taken over all sequences in B_1^∞ converging to (0, 0) with respect to the norm $\|\cdot\|_\infty$.

We shall show now, however, that the limit is actually $-\infty$ if taken over all sequences converging to (0, 0) with respect to the weak* topology on B_1^∞. Fixing any positive real number α, we define $(a_m, y_m) \in B_1^\infty$ and $x_m \in A_1^\infty$ for each positive integer m by $a_m = 0$,

$$y_m(t) = \begin{cases} -\alpha & \text{if} \quad 0 \leqslant t \leqslant \dfrac{1}{m}, \\ 0 & \text{if} \quad \dfrac{1}{m} < t \leqslant T, \end{cases}$$

$$x_m(t) = \begin{cases} \alpha(mt - 1) & \text{if} \quad 0 \leqslant t \leqslant \dfrac{1}{m}, \\ 0 & \text{if} \quad \dfrac{1}{m} \leqslant t \leqslant T. \end{cases}$$

Then for every m we have

$$\varphi_{l,L}(a_m, y_m) = \Phi_{l,L}(x_m) = -\alpha.$$

Furthermore, the function y_m converges to zero almost uniformly on $[0, T]$, and hence in particular (a_m, y_m) converges to (0, 0) in the weak* topology on B_n^∞. Therefore

$$\text{weak* } \liminf_{(a,y)\to(0,0)} \varphi_{l,L}(a, y) \leqslant -\alpha,$$

where the limit is taken over all sequences in B_n^∞ converging to (0, 0) in the weak* topology. Since α was an arbitrary positive number, we conclude that

$$\text{weak* } \liminf_{(a,y)\to(0,0)} \varphi_{l,L}(a, y) = -\infty. \tag{8.21}$$

Of course, (8.21) implies by Theorem 4 that, dually,

$$\begin{aligned} \inf\{\Phi_{m,M}(p) \mid p \in A_n^1\} &= \varphi_{m,M}(0, 0) = +\infty \\ &> \text{weak* } \liminf_{(b,q)\to(0,0)} \varphi_{m,M}(b, q) = 0. \end{aligned} \tag{8.22}$$

This can also be verified directly by computing m and M, if one so desires.

If in this example the definition of l is changed to

$$l(c_0, c_T) = e^{c_0},$$

the same arguments can be carried through, with the difference that, in this case, an optimal arc exists for $\Phi_{m,M}$ as well as for $\Phi_{l,L}$, the minimum of $\Phi_{l,L}$ on A_n^1 being 1, and the minimum of $\Phi_{m,M}$ on A_n^1 being 0.

9. Subdifferential Conditions for a Minimum

The familiar Euler–Lagrange equations and transversality conditions of the calculus of variations can be generalized to convex problems of Bolza, where there are no differentiability assumptions, by means of the theory of subdifferentiation of convex functions. Furthermore, a one-to-one correspondence between Lagrangian functions L and Hamiltonian functions H, extending the classical Legendre correspondence, can be defined in terms of conjugate convex functions. Under this correspondence, the Euler–Lagrange subdifferential condition is transformed into a Hamiltonian subdifferential condition, which is an ordinary differential equation with a multivalued right-hand side.

Given an extended-real-valued convex function f on R^n and a point $x \in R^n$, we denote by $\partial f(x)$ the set of all $x^* \in R^n$ such that

$$f(z) \geqslant f(x) + \langle z - x, x^* \rangle \qquad \text{for every} \qquad z \in R^n. \tag{9.1}$$

Such vectors x^* are called *subgradients* of f at x, and $\partial f(x)$ is called the *subdifferential* of f at x. It is immediate from (9.1) that $\partial f(x)$ is always a closed convex set (possibly empty). If f is actually (finite and) differentiable at x in the ordinary sense, then $\partial f(x)$ consists of a single element, namely the gradient vector

$$\nabla f(x) = \left(\frac{\partial f}{\partial x^1}(x), \ldots, \frac{\partial f}{\partial x^n}(x) \right).$$

Inequality (9.1) implies that

$$x^* \in \partial f(x) \qquad \text{if and only if} \qquad f^*(x^*) = \langle x, x^* \rangle - f(x), \tag{9.2}$$

where f^* is the conjugate of f. Thus, in the case where f is proper and lower semicontinuous (so that $f^{**} = f$), one has $x^* \in \partial f(x)$ if and only if $x \in \partial f^*(x^*)$.

The theory of subgradients of convex functions on R^n is presented at length in [21, Sections 23-26]. This theory includes formulas for calculating subgradients in various situations.

We shall denote by $\partial L(t, x, v)$ the subdifferential of the convex function $L(t, \cdot, \cdot)$ at (x, v). Thus $\partial L(t, x, v)$ will be a certain closed convex subset of $R^n \times R^n$ for each (t, x, v) in $[0, T] \times R^n \times R^n$. The subdifferential

$\partial l(c_0, c_T)$ of l at (c_0, c_T) will likewise be a closed convex subset of $R^n \times R^n$ for each (c_0, c_T) in $R^n \times R^n$.

We shall say that a given arc $x \in A_n^1$ satisfies the *Euler–Lagrange condition* for L if there exists an arc $p \in A_n^1$ such that

$$(\dot{p}(t), p(t)) \in \partial L(t, x(t), \dot{x}(t)) \text{ for almost every } t. \tag{9.3}$$

When $L(t, x, v)$ is actually differentiable with respect to

$$(x, v) = (x^1, \ldots, x^n, v^1, \ldots, v_n),$$

so that $\partial L(t, x, v)$ consists of the single vector

$$\left(\frac{\partial L}{\partial x^1}(t, x, v), \ldots, \frac{\partial L}{\partial x^n}(t, x, v), \frac{\partial L}{\partial v^1}(t, x, v), \ldots, \frac{\partial L}{\partial v^n}(t, x, v)\right),$$

(9.3) says simply that

$$p^i(t) = \frac{\partial L}{\partial v^i}(t, x(t), \dot{x}(t)), \qquad i = 1, \ldots, n,$$

$$\dot{p}^i(t) = \frac{\partial L}{\partial x^i}(t, x(t), \dot{x}(t)), \qquad i = 1, \ldots, n,$$

for almost every t. Then the Euler–Lagrange condition is the classical condition that x should satisfy the Euler–Lagrange differential equations:

$$\frac{d}{dt}\frac{\partial L}{\partial v^i}(t, x(t), \dot{x}(t)) = \frac{\partial L}{\partial x^i}(t, x(t), \dot{x}(t)), \qquad i = 1, \ldots, n. \tag{9.4}$$

In certain cases the Euler–Lagrange condition may also be construed as the maximum principle of optimal control theory, as will be seen in the next section.

We shall say that $x \in A_n^1$ is an *extremal arc* for the Bolza functional $\Phi_{l,L}$ if x satisfies the Euler–Lagrange condition for L and, in addition, the arc p in (9.3) (which is not always uniquely determined) can be chosen in such a way that

$$(p(0), -p(T)) \in \partial l(x(0), x(T)). \tag{9.5}$$

Such a $p \in A_n^1$ will be called a *coextremal of* $\Phi_{l,L}$ corresponding to x.

Condition (9.5) will be called the *transversality condition* for $\Phi_{l,L}$ in analogy with related conditions in the calculus of variations and optimal control. The reason for the terminology is found in the special case where, for certain nonempty closed convex sets C_0 and C_T in R^n one has

$$l(x(0), x(T)) = \begin{cases} 0 & \text{if} \quad x(0) \in C_0 \quad \text{and} \quad x(T) \in C_T, \\ +\infty & \text{if} \quad x(0) \notin C_0 \quad \text{or} \quad x(T) \notin C_T. \end{cases} \tag{9.6}$$

[Then in the given problem of Bolza one is minimizing $\int_0^T L(t, x(t), \dot{x}(t))\, dt$ subject to $x(0) \in C_0$ and $x(T) \in C_T$]. From (9.6) and the definition of ∂l, it is apparent that the transversality condition (9.5) is satisfied in this case if and only if $x(0) \in C_0$, $x(T) \in C_T$ and

$$\begin{aligned} \langle p(0), c_0 - x(0)\rangle &\leqslant 0 \qquad \text{for every} \qquad c_0 \in C_0,\\ \langle - p(T), c_T - x(T)\rangle &\leqslant 0 \qquad \text{for every} \qquad c_T \in C_T. \end{aligned}$$

In the language of convex analysis, this says that $p(0)$ is a *normal vector* to C_0 at $x(0)$, while $-p(T)$ is a normal vector to C_T at $x(T)$. When C_0 and C_T consist of single vectors c_0 and c_T, respectively, the transversality condition says simply that $x(0) = c_0$ and $x(T) = c_T$ ($p(0)$ and $p(T)$ being arbitrary).

Of course, an arc $p \in A_n^1$ will similarly be called an extremal of the dual Bolza functional $\Phi_{m,M}$ if there exists an $x \in A_n^1$ (a coextremal of $\Phi_{m,M}$ corresponding to p) such that

$$(\dot{x}(t), x(t)) \in \partial M(t, p(t), \dot{p}(t)) \text{ for almost every } t, \tag{9.7}$$

$$(x(0), -x(T)) \in \partial m(p(0), p(T)). \tag{9.8}$$

The relationship between extremal arcs and optimal arcs of the Bolza functionals $\Phi_{l,L}$ and $\Phi_{m,M}$ is exactly explained by the next result, which is a direct consequence of Theorem 4 and the theory of subdifferentiation.

THEOREM 5. *The following conditions on a pair of arcs $x \in A_n^1$ and $p \in A_n^1$ are equivalent:*

(a) *x is an extremal arc for $\Phi_{l,L}$ with co extremal p;*

(b) *p is an extremal arc for $\Phi_{m,M}$ with co extremal x;*

(c) *x is an optimal arc for $\Phi_{l,L}$, p is an optimal arc for $\Phi_{m,M}$, and the equivalent semicontinuity conditions* (b) *and* (c) *in the Corollary to Theorem 4 are satisfied;*

(d) $\Phi_{l,L}(x) = -\Phi_{m,M}(p)$.

Proof. In view of the conjugacy relations (5.5) and (5.10), conditions (9.3) and (9.7) are equivalent, and they are satisfied if and only if

$$L(t, x(t), \dot{x}(t)) + M(t, p(t), \dot{p}(t)) - \langle x(t), \dot{p}(t)\rangle - \langle \dot{x}(t), p(t)\rangle = 0 \tag{9.9}$$

for almost every t. Similarly, by (5.4) and (5.9) conditions (9.5) and (9.8) are equivalent, and they are satisfied if and only if

$$l(x(0), x(T)) + m(p(0), p(T)) - \langle x(0), p(0)\rangle + \langle x(T), p(T)\rangle = 0. \tag{9.10}$$

Of course, the left hand sides of (9.9) and (9.10) are always nonnegative, and the integral of the left hand side of (9.9) over $[0, T]$, plus the left hand side of (9.10) is

$$\Phi_{l,L}(x) + \Phi_{m,M}(p) \geqslant 0,$$

as observed in Proposition 2. Therefore (a), (b), and (d) are equivalent. The equivalence of (c) with (d) is obvious from Proposition 2 and the Corollary to Theorem 4.

COROLLARY 1. *Every extremal arc of $\Phi_{l,L}$ is an optimal arc. Conversely, if an extremal arc exists for $\Phi_{l,L}$, or if one of the semicontinuity conditions* (b) *or* (c) *in the Corollary of Theorem 4 is satisfied and the infimum of $\Phi_{m,M}$ over A_n^1 is attained, then every optimal arc for $\Phi_{l,L}$ is an extremal arc.*

COROLLARY 2. *If x is an extremal arc for $\Phi_{l,L}$, then the coextremals of $\Phi_{l,L}$ corresponding to x are the optimal arcs for $\Phi_{m,M}$.*

These corollaries provide, among other things, a dual method for solving a given convex problem of Bolza in cases (as in Examples 6 and 7) where it is known that the equivalent properties in the Corollary of Theorem 4 are present. In this method, one minimizes the dual Bolza functional $\Phi_{m,M}$ over A_n^1, rather than the given $\Phi_{l,L}$. Having determined any optimal arc p for $\Phi_{m,M}$, one gets all the optimal arcs x for $\Phi_{l,L}$ by finding all the arcs x which, together with p, satisfy (9.3) and (9.5). This dual method could be advantageous if minimizing $\Phi_{m,M}$ happened to be simpler than minimizing $\Phi_{l,L}$, perhaps because $\Phi_{m,M}$ was essentially finite-dimensional (as in Example 7), or because $\Phi_{m,M}$ was everywhere differentiable in some suitable sense, so that "steepest descent" algorithms could be used.

(Dual methods of solution of variational problems are, of course, nothing new, and they are well known in the case of linear control problems with fixed endpoints. However, dual methods have customarily been described in terms of supporting hyperplanes as in Wets-Van Slyke [29], whereas here we are able to give a more explicit form, provided that m and M can be calculated from l and L.)

A Hamiltonian form of the Euler–Lagrange condition will now be derived.

The *Hamiltonian function* corresponding to the Lagrangian function L will be defined as the extended-real-valued function H on $[0, T] \times R^n \times R^n$ obtained by taking the conjugate of the convex function $L(t, x, \cdot)$ for each (t, x). Thus

$$H(t, x, p) = \sup\{\langle v, p\rangle - L(t, x, v) \mid v \in R^n\}. \tag{9.11}$$

By the properties of the conjugacy correspondence, of course, L is completely determined by H [in view of assumption (A)], and one has

$$L(t, x, v) = \sup\{\langle v, p\rangle - H(t, x, p) \mid p \in R^n\}. \tag{9.12}$$

The correspondence between Lagrangian functions and Hamiltonian functions is thus one-to-one.

If $L(t, x, v)$ is differentiable in v, and if the gradient mapping

$$v \to \nabla_v L(t, x, v) = \left(\frac{\partial L}{\partial v^1}(t, x, v),\ldots, \frac{\partial L}{\partial v^n}(t, x, v)\right)$$

is one-to-one from R^n onto itself, (9.11) reduces essentially to the classical Legendre transformation: $H(t, x, p)$ is obtained by solving the equation

$$\nabla_v L(t, x, v) = p$$

for v in terms of t, x, and p and then substituting this in the expression

$$\langle v, p\rangle - L(t, x, v).$$

This is treated rigorously in [21, Section 26].

The convexity of $L(t, x, v)$ in (x, v) implies by [21, Theorem 33.1] that $H(t, x, p)$ *is concave as a function of* x *and convex as a function of* p, so that H, like L, is well suited for study by convexity methods. The properties of H will be discussed in more detail elsewhere [26]. (Note that the one-to-one correspondence between Lagrangians and Hamiltonians in (9.11) and (9.12) does not depend on all of our assumptions; it is well-defined as long as $L(t, x, \cdot)$ is, for each t and x, a lower semicontinuous convex function from R^n to $R^1 \cup \{+\infty\}$. In this general case, however, $H(t, x, p)$ would not be concave in x, although it would still be convex in p, and subdifferentiation with respect to x could not be employed as below.)

For each t and x, let us denote (somewhat imperfectly) by $\partial_p H(t, x, p)$ the set of all subgradients in R^n of the convex function $H(t, x, \cdot)$ at the point p. Similarly, for each t and p let us denote by $-\partial_x H(t, x, p)$ the set of all subgradients in R^n of the convex function $-H(t, \cdot, p)$ at the point x. In other words,

$$\partial_p H(t, x, p) = \{v \mid \forall p' \in R^n, H(t, x, p') \geqslant H(t, x, p) + \langle v, p' - p\rangle\}, \tag{9.13}$$

$$\partial_x H(t, x, p) = \{s \mid \forall x' \in R^n, H(t, x', p) \leqslant H(t, x, p) + \langle x' - x, s\rangle\}. \tag{9.14}$$

We shall say that given arcs $x \in A_n^1$ and $p \in A_n^1$ satisfy the *Hamiltonian condition* corresponding to H if

$$\begin{gathered} \dot{x}(t) \in \partial_p H(t, x(t), p(t)) \text{ for almost every } t \\ \text{and} \\ \dot{p}(t) \in -\partial_x H(t, x(t), p(t)) \text{ for almost every } t. \end{gathered} \tag{9.15}$$

If $H(t, \cdot, \cdot)$ happens to be differentiable at $(x(t), p(t))$ for every t, this reduces to the Hamiltonian equations:

$$\dot{x}(t) = \nabla_p H(t, x(t), p(t)) \text{ for almost every } t$$

and (9.16)

$$\dot{p}(t) = -\nabla_x H(t, x(t), p(t)) \text{ for almost every } t.$$

(Observe, incidentally, that H could be differentiable everywhere in x and p without L necessarily being differentiable, or even finite, everywhere in x and v.)

In the classical case, where the correspondence between Lagrangians and Hamiltonians is defined in terms of the Legendre transformation, x and p satisfy the Hamiltonian equations if and only if x satisfies the Euler–Lagrnage equations and

$$p(t) = \nabla_v L(t, x(t), v(t)).$$

In the present case, there is an analogous result, which is augmented by a game-theoretic characterization. The latter concerns the function J defined by

$$J(t, s, v) = -\sup\{\langle s, x\rangle - L(t, x, v) \mid x \in R^n\}. \tag{9.17}$$

We note that $J(t, \cdot, \cdot)$, like $H(t, \cdot, \cdot)$, is concave-convex on $R^n \times R^n$ for each t by [21, Theorem 33.1]. We shall say that arcs $x \in A_n^1$ and $p \in A_n^1$ satisfy the *minimax condition* corresponding to J if $(\dot{p}(t), \dot{x}(t))$ is for almost every $t \in [0, T]$ a saddle-point of the concave-convex function

$$K_t(s, v) = J(t, s, v) - \langle s, x(t)\rangle + \langle v, p(t)\rangle \tag{9.18}$$

on $R^n \times R^n$, in other words,

$$K_t(\dot{p}(t), v) \geqslant K_t(\dot{p}(t), \dot{x}(t)) \geqslant K_t(s, \dot{x}(t)) \tag{9.19}$$

for every $v \in R^n$ and $s \in R^n$.

THEOREM 6. *The following conditions on a pair of arcs $x \in A_n^1$ and $p \in A_n^1$ are equivalent:*

(a) *x and p satisfy the Hamiltonian condition corresponding to H;*

(b) *x satisfies, together with p, the Euler–Lagrange condition corresponding to L;*

(c) *p satisfies, together with x, the Euler–Lagrange condition corresponding to M;*

(d) *x and p satisfy the minimax condition corresponding to J.*

Proof. This is immediate from the subgradient relations established in [21, Theorem 37.5].

COROLLARY. *If $x \in A_n^1$ and $p \in A_n^1$ satisfy the Hamiltonian condition* (9.15), *then x minimizes $\int_0^T L\,dt$ over the class of all arcs in A_n^1 having the same endpoints as x, while p minimizes $\int_0^T M\,dt$ over the class of all arcs in A_n^1 having the same endpoints as p.*

Proof. Given x and p satisfying (9.15), let $c_0 = x(0)$ and $c_T = x(T)$, and l be the function on $R^n \times R^n$ which vanishes at (c_0, c_T) but has the value $+\infty$ everywhere else. Then x and p trivially satisfy the transversality condition for l. Since x and p also satisfy the Euler–Lagrange condition (9.3) by Theorem 6, x is an extremal arc for $\Phi_{l,L}$. Theorem 5 implies then that x is an optimal arc for $\Phi_{l,L}$. In other words, one has

$$\int_0^T L\,(t, x(t), \dot{x}(t))\,dt \leqslant \int_0^T L(t, z(t), \dot{z}(t))\,dt$$

for every arc $z \in A_n^1$ such that $z(0) = x(0)$ and $z(T) = x(T)$. The argument for p is parallel.

The advantage of the Hamiltonian form of the Euler–Lagrange condition is that it can be studied as a differential equation with a multivalued right-hand side. Thus results about the existence of solutions, and the dependence of such solutions on initial points, can be derived in certain broad cases, as we shall demonstrate in [26], from known generalizations of theorems about ordinary differential equations, such as the results of Castaing in [5, Section 9].

10. Extremal Arcs and the Maximum Principle

The Euler–Lagrange condition and transversality condition introduced in Section 9 can be made more explicit in particular cases by means of the rules given in [21, Section 23] for computing subgradients, together with various measurability results and selection theorems. We shall demonstrate this for some of the example problems already discussed. Example 12, especially, will clarify the relationship, in the case of "sufficiently differentiable" problems of optimal control reformulated as convex problems of Bolza, between extremal arcs in the sense of Section 9 and arcs and controls which are extremal in the sense of the maximum principle.

EXAMPLE 9. Suppose, as in Example 5, that L is of the form (6.16). The corresponding Hamiltonian function H may then be calculated as

$$\begin{aligned} H(t, x, p) &= \sup_{v \in R^n}\{\langle v, p\rangle - f(t, x) - g(t, v - E(t)\, x)\} \\ &= \sup_{w \in R^n}\{\langle w + E(t)\, x, p\rangle - f(t, x) - g(t, w)\} \\ &= \langle E(t)\, x, p\rangle - f(t, x) + g^*(t, p), \end{aligned} \tag{10.1}$$

where the last expression is to be interpreted as $-\infty$ if both $f(t, x) = +\infty$ and $g^*(t, p) = +\infty$. The generalized Hamiltonian "equations" (9.5) for this H, which are equivalent to the generalized Euler–Lagrange "equation" (9.3) by Theorem 6, are:

$$\begin{aligned} \dot{x}(t) &\in E(t)\, x(t) + \partial g^*(t, p(t)), \\ \dot{p}(t) &\in -E^*(t)\, p(t) + \partial f(t, x(t)) \end{aligned} \tag{10.2}$$

for almost every t, where $\partial f(t, x(t))$ is the set of all subgradients of $f(t, \cdot)$ at $x(t)$, and $\partial g^*(t, p(t))$ is the set of all subgradients of $g^*(t, \cdot)$ at $p(t)$. Of course, (10.2) can also be expressed in the form

$$\dot{x}(t) = E(t)\, x(t) + u(t), \tag{10.3a}$$

$$\dot{p}(t) = -E^*(t)\, p(t) + w(t), \tag{10.3b}$$

$$u(t) \in \partial g^*(t, p(t)), \tag{10.3c}$$

$$w(t) \in \partial f(t, x(t)), \tag{10.3d}$$

where $u \in L_n^1$ and $w \in L_n^1$.

Suppose now that f is identically 0, and that the conditions in Example 7 in Section 8 are satisfied, so that the dual problem of Bolza is essentially finite-dimensional and has at least one optimal arc, and the properties in the Corollary to Theorem 4 hold. Then, by Corollary 2 of Theorem 5, x is an optimal arc for $\Phi_{l,L}$ if and only if x is an extremal arc, i.e., satisfies the conditions (10.3a–d) and (9.5) for some $p \in A_n^1$. In this case $w(t)$ must be identically zero in (10.3d) and (10.3b), and the arc p is uniquely determined by $p(0)$. Once $p(0)$ and $x(0)$ have been specified, a function u can be obtained from (10.3c), and then (10.3a) can be solved for x. Here (10.3c) can be expressed equivalently as the condition that the maximum of the (extended-real-valued) concave function

$$\langle p(t), \cdot\rangle - g(t, \cdot)$$

over R^n be attained at the point $u(t)$; thus, if g is such that $\partial g^*(t, p(t))$ cannot be handled explicitly, it may still be possible to determine the function u by solving a certain optimization problem in R^n for each t.

Observe that, depending on the relationship between $E(t)$ and $g(t, \cdot)$, it could well happen that u is uniquely determined by (10.3) up to equivalence in L_n^1, since $\partial g^*(t, p(t))$ might reduce to a single element for almost every t when p is an arc satisfying the differential equation (10.3b) with $w(t) = 0$. (This is suggested by the fact that, on the interior of the set of points where $\partial g^*(t, \cdot)$ is nonempty, $\partial g^*(t, \cdot)$ reduces to a single element almost everywhere [21, Theorem 25.5].) In such cases, x is uniquely determined by $p(0)$ and $x(0)$. The only remaining problem (not necessarily easy) is then to choose $p(0)$ and $x(0)$ in such a way that the resulting $x(T)$ and $p(T)$ satisfy the transversality condition (9.5).

As pointed out following Corollary 2 of Theorem 5, the coextremal p needed in order to determine x can also be found by solving the dual problem of Bolza, which in this case (according to Example 5) consists of minimizing

$$m(p(0), p(T)) + \int_0^T g^*(t, p(t))\, dt$$

over the finite-dimensional subspace of A_n^1 consisting of all arcs p such that $\dot{p}(t) = -E^*(t)\, p(t)$ for almost every t. This may be regarded as a problem of minimizing a certain (extended-real-valued, not necessarily differentiable) convex function of a vector variable $d_0 \in R^n$, where $d_0 = p(0)$.

EXAMPLE 10. Consider the convex problem of Bolza in Example 1 in Section 4, where l and L are given by (4.5) and (4.6). Here we set

$$x(t) = (z(t), y(t)) \qquad \text{and} \qquad p(t) = (q(t), h(t)).$$

The transversality condition can be analyzed as in the example mentioned in Section 9: one has

$$(q(0), h(0), -q(T), -h(T)) \in \partial l(z(0), y(0), z(T), y(T))$$

if and only if $h(0) = 0 = h(T)$, $q(0)$ is a normal vector to Z_0 at $z(0)$, and $-q(T)$ is a normal vector to Z_T at $z(T)$.

Let us now write

$$L(t, x, v) = L_1(t, x, v) + L_2(t, x, v) + L_3(t, x, v), \tag{10.4}$$

where (for $x = (z, y)$ and $v = (w, u)$)

$$L_1(t, x, v) = K(t, z, u),$$

$$L_2(t, x, v) = \begin{cases} 0 & \text{if} \quad w = A(t)\, z + B(t)\, u, \\ +\infty & \text{if} \quad w \neq A(t)\, z + B(t)\, u, \end{cases}$$

$$L_3(t, x, v) = \begin{cases} 0 & \text{if} \quad u \in U(t), \\ +\infty & \text{if} \quad u \notin U(t). \end{cases}$$

For each fixed $t \in [0, T]$, the hypothesis of [21, Theorem 23.8] is easily seen to be satisfied for (10.4) as a sum of convex functions of (x, v), and hence

$$\partial L(t, x, v) = \partial L_1(t, x, v) + \partial L_2(t, x, v) + \partial L_3(t, x, v). \tag{10.5}$$

LEMMA. *An arc $x \in A_n^1$ satisfies the Euler–Lagrange condition here for L if and only if there exists a $p \in A_n^1$ such that one has*

$$(\dot{p}(t), p(t)) = (x_1^*(t), v_1^*(t)) + (x_2^*(t), v_2^*(t)) + (x_3^*(t), v_3^*(t)) \tag{10.6}$$

for Lebesgue measurable functions x_i^* and v_i^* satisfying

$$(x_i^*(t), v_i^*(t)) \in \partial L_i(t, x(t), \dot{x}(t)) \textit{ for almost every } t. \tag{10.7}$$

Proof. The functions L_i satisfy conditions (A) and (B) (see the discussion in Section 4), and therefore by [23, Corollary 4.6] the multifunctions

$$t \to \partial L_i(t, x(t), \dot{x}(t)) \subset R^n$$

are Lebesgue measurable for any $x \in A_n^1$. (At points t where the derivative $\dot{x}(t)$ does not exist, an arbitrary value may be assigned to $\dot{x}(t)$, so that these multifunctions are everywhere defined.) Fix any arc $p \in A_n^1$, let $Q_1(t)$ denote the set of all $(x_1^*, v_1^*, x_2^*, v_2^*, x_3^*, v_3^*)$ in R^{6n} such that

$$x_1^* + x_2^* + x_3^* = p(t) \qquad \text{and} \qquad v_1^* + v_2^* + v_3^* = \dot{p}(t),$$

and let

$$Q_2(t) = \partial L_1(t, x(t), \dot{x}(t)) \times \partial L_2(t, x(t), \dot{x}(t)) \times \partial L_3(t, x(t), \dot{x}(t)),$$
$$Q(t) = Q_1(t) \cap Q_2(t).$$

In view of (10.5), the Euler–Lagrange condition (9.3) is satisfied by x and p if and only if

$$Q(t) \neq \emptyset \text{ for almost every } t. \tag{10.8}$$

Now Q_1 and Q_2 are Lebesgue measurable as multifunctions from $[0, T]$ to R^{6n} by [23, Theorem 3] and [23, Corollary 1.2], and hence Q is likewise Lebesgue measurable by [23, Corollary 1.3]. This implies by a result of Kuratowski and Ryll-Nardzewski [13] (quoted as [23, Corollary 1.1]) that, under (10.8), there exists a Lebesgue measurable function from $[0, T]$ to R^{6n} whose value at t belongs to $Q(t)$ for almost every t. In other words, (9.3) holds if and only if there exist, as claimed, Lebesgue measurable functions x_i^* and v_i^* from $[0, T]$ to R^n satisfying (10.6) and (10.7). This proves the lemma.

We must now analyze conditions (10.7). For convenience, let us set

$$(x_i^*, v_i^*) = (z_i^*, y_i^*, w_i^*, u_i^*) \in R^r \times R^s \times R^r \times R^s.$$

Then obviously (10.7) holds for $i = 1$ if and only if, for almost every t, $y_1^*(t) = 0$, $w_1^*(t) = 0$, and

$$(z_1^*(t), u_1^*(t)) \in \partial K(t, z(t), u(t)) \tag{10.9}$$

(where $u = \dot{y}$). On the other hand, (10.7) holds for $i = 2$ if and only if for almost every t

$$\dot{z}(t) = A(t)\, z(t) + B(t)\, u(t), \tag{10.10}$$

$y_2^*(t) = 0$, $z_2^*(t) = -A^*(t)\, w_2^*(t)$, and $u_2^*(t) = -B^*(t)\, w_2^*(t)$ (where $A^*(t)$ and $B^*(t)$ denote the transposes of $A(t)$ and $B(t)$, respectively). Similarly, (10.7) holds for $i = 3$ if and only if, for almost every t, $z_3^*(t) = 0$, $y_3^* = 0$, $w_3^* = 0$ and

$$u_3^*(t) \text{ is a normal vector to } U(t) \quad \text{at} \quad u(t) \in U(t). \tag{10.11}$$

Thus (10.6) requires that, for almost every t,

$$\begin{aligned} \dot{q}(t) &= z_1^*(t) - A^*(t)\, w_2^*(t), \\ \dot{h}(t) &= 0, \\ q(t) &= w_2^*(t), \\ h(t) &= u_1^*(t) - B^*(t)\, w_2^*(t) + u_3^*(t), \end{aligned}$$

subject to (10.9), (10.10), and (10.11). Of course, the condition that $\dot{h}(t) = 0$, and the transversality condition that $h(0) = 0 = h(T)$, imply that $h(t)$ is identically 0.

Therefore the arc $x = (z, y)$ is an extremal arc in this example (i.e. the arc z and corresponding control $u = \dot{y}$ are an "extremal pair" for the given optimal control problem) if and only if there exist functions $q \in A_r^1$ and $u^* \in L_s^1$ such that, for almost every $t \in [0, T]$,

$$\dot{z}(t) = A(t)\, z(t) + B(t)\, u(t), \tag{10.12a}$$

$$(\dot{q}(t) + A^*(t)\, q(t), -u^*(t) + B^*(t)\, q(t)) \in \partial K(t, z(t), u(t)), \tag{10.12b}$$

$$u^*(t) \textit{ is a normal vector to } U(t) \quad \textit{at} \quad u(t) \in U(t), \tag{10.12c}$$

$$q(0) \textit{ is a normal vector to } Z_0 \quad \textit{at} \quad z(0) \in Z_0, \tag{10.12d}$$

$$-q(T) \textit{ is a normal vector to } Z_T \quad \textit{at} \quad z(T) \in Z_T. \tag{10.12e}$$

These conditions can be analyzed further when more information is given about K, $U(t)$, Z_0, and Z_T. Suppose, for instance, that these are defined as in Example 4′. Let B_ρ denote the unit ball for the norm $\|\cdot\|_\rho$, and for a vector w let $J_\rho(w)$ denote the set of points of B_ρ at which w is a normal vector. Define J_σ similarly for $\|\cdot\|_\sigma$. Then (10.12b) and (10.12c) say that

$$\dot{q}(t) + A^*(t)\, q(t) \in \alpha J_\sigma(z(t)), \tag{10.13}$$

$$u(t) \in J_\rho(B^*(t)\, q(t) - \beta J_\sigma(u(t))), \tag{10.14}$$

while (10.12d) and (10.12e) say that $z(0) = a$, $z(T) \in Z_T$, and $q(T) \in Z_T^{\perp}$ ($q(0)$ arbitrary). Of course, in the case where $\rho = \sigma = 2$, one has

$$J_2(w) = \begin{cases} |w|^{-1} w & \text{if} \quad w \neq 0, \\ \{w^* \mid |w^*| \leqslant 1\} & \text{if} \quad w = 0. \end{cases}$$

If $\rho = \infty$ and $\sigma = 1$, conditions (10.13) and (10.14) are more complicated, but they can still be written down explicitly.

The Hamiltonian function in this example is given by

$$\begin{aligned} H(t, x, p) &= H(t, z, y, q, h) \\ &= \langle A(t)\, z, q\rangle + \sup_{u \in U(t)} \{\langle u, h + B^*(t)\, q\rangle - K(t, z, u)\}. \end{aligned} \tag{10.15}$$

Example 11. We shall show that, in the case of Example 2 in Section 4, the Euler–Lagrange condition and transversality condition require the existence of certain *Lagrange multipliers* for the constraints (4.8) and (4.9). Here l and L are given by (4.10) and (4.11). We shall assume it is possible to choose a_0 and a_T in R^r in such a way that

$$k_i(a_0, a_T) \leqslant 0 \qquad \text{for} \qquad i = 1,\ldots, m_1,$$

with *strict* inequality for all i such that k_i is not affine. Similarly, we shall assume that, for almost every $t \in [0, T]$, it is possible to choose $z \in R^r$, $w \in R^r$, and $u \in R^s$ such that (4.12) holds with *strict* inequality for all j such that $K_j(t, \cdot, \cdot, \cdot)$ is not affine. Then ∂l and ∂L may be calculated as in [21, p. 283].

Setting $p = (q, h)$ and $x = (z, y)$ as before, we obtain the result that the transversality condition

$$(q(0), h(0), -q(T), -h(T)) \in \partial l(z(0), y(0), z(T), y(T))$$

is satisfied if and only if $h(0) = 0 = h(T)$, and there exist real numbers λ_i (Lagrange multipliers) such that

$$\begin{aligned} (q(0), -q(T)) \in \partial k_0(z(0), z(T)) &+ \lambda_1 \partial k_1(z(0), z(T)) \\ &+ \cdots + \lambda_{m_1} \partial k_{m_1}(z(0), z(T)), \end{aligned} \tag{10.16}$$

and

$$\begin{aligned} k_i(z(0), z(T)) \leqslant 0, &\qquad \lambda_i \geqslant 0 \\ \lambda_i k_i(z(0), z(T)) = 0, &\qquad i = 1,\ldots, m_1. \end{aligned} \tag{10.17}$$

Likewise, the Euler–Lagrange condition (9.3) is satisfied if and only if, for almost every t, one has $\dot{h}(t) = 0$, and it is possible to choose real numbers

$\mu_j(t)$, such that

$$(\dot{q}(t), q(t), h(t)) \in \partial K_0(t, z(t), \dot{z}(t), u(t)) + \mu_1(t)\, \partial K_1(t, z(t), \dot{z}(t), u(t)) \\ + \cdots + \mu_{m_2}(t)\, \partial K_{m_2}(t, z(t), \dot{z}(t), u(t)) \tag{10.18}$$

and

$$K_j(t, z(t), \dot{z}(t), u(t)) \leqslant 0, \qquad \mu_j(t) \geqslant 0 \\ \mu_j(t)\, K(t, z(t), \dot{z}(t), u(t)) = 0 \qquad \text{for} \qquad j = 1,..., m_2 . \tag{10.19}$$

Here the Lagrange multipliers $\mu_j(t)$, if they exist, can be chosen as Lebesgue measurable functions of t; this may be established by an argument similar to the one in the lemma in the preceding example.

Therefore $z \in A_r^1$ *and* $u \in L_s^1$ *are an "extremal pair" for this optimal control problem if and only if there exist real numbers* λ_i *and Lebesgue measurable functions* μ_j *from* $[0, T]$ *to* R^1 *such that conditions* (10.16), (10.17), (10.18), *and* (10.19) *hold (for almost every* t*), with* $h(t) \equiv 0$ *in* (10.18).

When the functions k_i and $K_j(t, \cdot, \cdot)$ are differentiable on $R^r \times R^s$, the subgradients in (10.16) and (10.18) can be replaced by gradients.

EXAMPLE 12. Relationships with the maximum principle [20] can be clarified by considering Example 3 in Section 4. In this fixed endpoint problem, the transversality condition is just that $x(0) = c_0$ and $x(T) = c_T$, with no restriction on $p(0)$ or $p(T)$. According to (9.11) and the definition of L, the Hamiltonian function is given by

$$H(t, x, p) = \max_{u \in U} \{\langle f(t, x, u), p \rangle - K(t, x, u)\}, \tag{10.20}$$

an expression familiar in the theory of optimal control. Thus [under the assumption that $L(t, x, v)$ is convex in (x, v)] formula (10.20) defines a function H which is concave in x, convex in p (and actually finite and continuous in all variables).

From Theorem 6 we know that $x \in A_n^1$ is an extremal arc if and only if $x(0) = c_0$, $x(T) = c_T$, and there exists a $p \in A_n^1$ such that x and p satisfy the generalized Hamiltonian differential equation (9.15) for this H. The existence of solutions to the latter for arbitrary initial points and sufficiently small t intervals can be deduced from the theory of contingent equations, as we shall show elsewhere [26].

The fact that the Hamiltonian condition here implies the maximum principle, assuming (as required in the formulation of the maximum principle) that K and f are differentiable with respect to x, can most easily be demonstrated by working directly with the equivalent Euler–Lagrange condition. The Euler–Lagrange condition (9.3) is satisfied if and only if one has

$$\langle x(t), \dot{p}(t)\rangle + \langle \dot{x}(t), p(t)\rangle - L(t, x(t), \dot{x}(t)) = M(t, \dot{p}(t), p(t))$$

for almost every t, where

$$L(t, x(t), \dot{x}(t)) = \min\{K(t, x(t), u) \mid u \in U, f(t, x(t), u) = \dot{x}(t)\} \tag{10.21}$$

(the minimum being $+\infty$ if there is no $u \in U$ such that $f(t, x(t), u) = \dot{x}(t)$) and

$$\begin{aligned} M(t, p(t), \dot{p}(t)) &= \sup_{x\in R^n} \sup_{v\in R^n}\{\langle x, \dot{p}(t)\rangle + \langle v, p(t)\rangle - L(t, x, v)\} \\ &= \sup_{x\in R^n} \max_{u\in U}\{\langle x, \dot{p}(t)\rangle + \langle f(t, x, u), p(t)\rangle - K(t, x, u)\}. \end{aligned} \tag{10.22}$$

This says that, for almost every t, there must exist a $u(t) \in U$ such that

$$\dot{x}(t) = f(t, x(t), u(t)) \tag{10.23}$$

and the "sup max" in (10.22) is attained at $(x(t), u(t))$. (It can be seen from Filippov's lemma [9] that, in this event, $u(t)$ can be chosen to be a Lebesgue measurable function of t.) In particular, one then has

$$\begin{aligned} &\langle f(t, x(t), u(t)), p(t)\rangle - K(t, x(t), u(t)) \\ &= H(t, x(t), p(t)) \\ &= \max_{u\in U}\{\langle f(t, x(t), u), p(t)\rangle - K(t, x(t), u)\}, \end{aligned} \tag{10.24}$$

$$\begin{aligned} &\langle x(t), \dot{p}(t)\rangle + \langle f(t, x(t), u(t)), p(t)\rangle - K(t, x(t), u(t)) \\ &= \max_{x\in R^n}\{\langle x, \dot{p}(t)\rangle + \langle f(t, x, u(t)), p(t)\rangle - K(t, x, u(t))\}. \end{aligned} \tag{10.25}$$

Assuming that K and f are differentiable with respect to x, (10.25) implies, of course, that

$$\dot{p}(t) = -\nabla_x f(t, x(t), u(t))\, p(t) + \nabla_x K(t, x(t), u(t)), \tag{10.26}$$

i.e., that p satisfies the familiar adjoint differential equation in optimal control theory. Thus, in this case, if $x \in A_n^1$ is an extremal arc, there must exist a control function $u \in L_n^\infty$ and an arc $p \in A_n^1$ such that (10.23), (10.24), and (10.26), the conditions of the maximum principle for a "normal" arc, are satisfied.

Conversely, suppose that K and f are not only differentiable in x, but also that $K(t, x, u)$ is convex in (x, u), $f(t, x, u)$ is affine in (x, u), i.e., of the form

$$f(t, x, u) = A(t)\, x + B(t)\, u + C(t),$$

and U is convex. (These are natural assumptions implying that $L(t, x, v)$ is indeed convex in (x, v), as mentioned in Section 4.) Then (10.26) is equivalent to (10.25), and it can be seen further that (10.24) and (10.25) imply the seemingly stronger condition that the maximum in (10.22) be attained at $(x(t), u(t))$. Under these assumptions, therefore, every "normal extremal" in the sense of the maximum principle is an extremal arc in the sense of this paper.

References

1. E. Asplund and R. T. Rockafellar, Gradients of convex functions, *Trans. Amer. Math. Soc.* **139** (1969), 443–467.
2. R. Bellman, "Dynamic Programming," Princeton Univ. Press, Princeton, N. J., 1957.
3. A. Brøndsted, Conjugate convex functions in topological vector spaces, *Mat.-Fys. Medd. Danske Vid. Selsk.* **34** (1964).
4. N. Bourbaki, "Espaces Vectoriels Topologiques," Hermann, Paris, 1953.
5. C. Castaing, Sur les multi-applications mesurables, Thèse, Faculté des Sciences, Caën, 1967. This has partly been published in *Rev. Française Informat. Recherche Opérationelle* **1** (1967), 3–34.
6. L. Cesari, Existence theorems for weak and usual optimal solutions in Lagrange problems with unilateral constraints, I, *Trans. Amer. Math. Soc.* **124** (1966), 369–412.
7. R. Courant and D. Hilbert, "Methods of Mathematical Physics," Vol. I, Interscience, New York, 1953.
8. W. Fenchel, On conjugate convex functions, *Canad. J. Math.* **1** (1949), 73–77.
9. A. F. Filippov, On certain questions in the theory of optimal control, *Vestnik Moskov. Univ. Ser. Math. Mech. Astronom.* **2** (1959), 25–32, and *SIAM J. Control* **1** (1962), 76–84.
10. K. O. Friedrichs, Ein Verfahren der Variationsrechnung das Minimum eines Integral als das Maximum eines anderen Ausdruckes darzustellen, *Nachr. Gesell. Wiss., Göttingen*, (1929), 13–20.
11. A. D. Ioffe and V. M. Tikhomirov, On duality in problems of the calculus of variations, *Dokl. Akad. Nauk SSSR* **180** (1968), 789–792, and *Sov. Math. Dokl.* **9** (1968), 685–688.
12. E. Kreindler, Reciprocal optimal control problems, *J. Math. Anal. Appl.* **14** (1966), 164–172.
13. K. Kuratowski and C. Ryll-Nardzewski, A general theorem on selectors, *Bull. Polish Acad. Sci.* XIII (1965), 273–411.
14. B. Mond and M. A. Hanson, Duality for variational problems, *J. Math. Anal. Appl.* **18** (1967), 355–364.
15. J.-J. Moreau, Fonctionelles convexes, mimeographed lecture notes, 107 pp., Séminaire sur les Équations aux Dérivées Partielles, Collège de France, 1967.

16. J.-J. MOREAU, Principes extremaux pour le problème de la naissance de la cavitation, *J. Mécanique* **5** (1966), 439–470.
17. J.-J. MOREAU, One-sided constraints in hydrodynamics, *in* "Nonlinear Programming," (J. Abadie, ed.), pp. 259–279, North-Holland, Amsterdam, 1967.
18. J. D. PEARSON, Reciprocity and duality in control programming problems, *J. Math. Anal. Appl.* **10** (1965), 388–408.
19. J. D. PEARSON, Duality and a decomposition technique, *SIAM J. Control* **4** (1966), 164–172.
20. L. S. PONTRYAGIN, V. G. BOLTYANSKII, R. V. GAMKRELIDZE, AND E. F. MISHCHENKO, "The Mathematical Theory of Optimal Processes," Wiley (Interscience), New York, 1962.
21. R. T. ROCKAFELLAR, "Convex Analysis," Princeton Univ. Press, Princeton, N. J., 1969.
22. R. T. ROCKAFELLAR, Integrals which are convex functionals, *Pacific J. Math.* **24** (1968), 525–539.
23. R. T. ROCKAFELLAR, Measurable dependence of convex sets and functions on parameters, *J. Math. Anal. Appl.* **28** (1969), 4–25.
24. R. T. ROCKAFELLAR, Extension of Fenchel's duality theorem for convex functions, *Duke Math. J.* **33** (1966), 81–90.
25. R. T. ROCKAFELLAR, Duality and stability in extremum problems involving convex functions, *Pacific J. Math.* **21** (1967), 167–187.
26. R. T. ROCKAFELLAR, Generalized Hamiltonian equations for convex problems of Lagrange, *Pacific J. Math.* **33** (May, 1970).
27. R. T. ROCKAFELLAR, Existence and duality theorems for convex problems of Bolza, (to be published).
28. W. F. TYNDALL, A duality theorem for a class of continuous linear programming problems, *SIAM J. Appl. Math.* **13** (1965), 644–666.
29. R. J.-B. WETS AND R. M. VAN SLYKE, A duality theory for abstract mathematical programs with applications to optimal control theory, *J. Math. Anal. Appl.* **22** (1968), 679–706.
30. P. WOLFE, A duality theorem for nonlinear programming, *Quart. Appl. Math.* **19** (1961), 239–244.
31. L. C. Young, "Lectures on the Calculus of Variations and Optimal Control Theory," Saunders, Philadelphia, 1969.
32. L. E. ZACHRISSON, Deparametrization of the Pontryagin maximum principle, *in* "Mathematical Theory of Control," (A. V. Balakrishnan and L. W. Neustadt, eds.), pp. 234–245. Academic Press, New York, 1967.
33. A. D. IOFFE AND V. M. TIKHOMIROV, Duality of convex functions and extremum problems, *Uspehi Mat. Nauk* XXIII (1968), 51–116 (Russian).
34. A. D. IOFFE AND V. M. TIKHOMIROV, On the minimization of integral functionals, *Functional Anal. Appl.* **3** (1969), 61–70 (Russian).
35. M. M. TSVETANOV, On duality in problems of the calculus of variations, *Comptes Rendus de l'Academie bulgare des Sciences* **21** (1968), 733–736 (Russian). (Note: Theorems 2, 3 and 4 of this paper are incorrect without further assumptions.)

PRINTED IN BELGIUM BY THE ST. CATHERINE PRESS, TEMPELHOF, 37, BRUGES, LTD.