NONLINEAR PROGRAMMING AND OPTIMAL CONTROL

by

Pravin P. Varaiya

ERL Technical Memorandum M-129

23 September 1965

ELECTRONICS RESEARCH LABORATORY

University of California, Berkeley

The research reported herein was supported by National Aeronautics and Space Administration under Grant NsG-354 (S-2).

NONLINEAR PROGRAMMING AND OPTIMAL CONTROL

Pravin Pratap Varaiya

ABSTRACT

Considerable effort has been devoted in recent years to three classes of optimization problems. These areas are nonlinear programming, optimal control, and solution of large-scale systems via decomposition. We have proposed a model problem which is a generalization of the usual nonlinear programming problem, and which subsumes these three classes. We derive necessary conditions for the solution of our problem. These conditions, under varying interpretations and assumptions, yield the Kuhn-Tucker theorem and the maximum principle. Furthermore, they enable us to devise decomposition techniques for a class of large convex programming problems.

More important than this unification, in our opinion, is the fact that we have been able to develop a point of view which is useful for most optimization problems. Specifically, we show that there exist sets which we call local cones and local polars (Chapter I), which play a determining role in maximization theories which consider only first variations, and that these theories give varying sets of assumptions under which these sets, or relations between them, can be obtained explicitly.

ACKNOWLEDGMENTS

I am extremely grateful to Professor L. A. Zadeh for his inspiring guidance throughout my stay in Berkeley, and for the great pleasure I have derived from his lectures. I also wish to thank Professor E. Polak for having read this manuscript and for his valuable suggestions regarding the presentation of this material. It is also a pleasure to acknowledge the great benefit I have received from the innumerable conversations with my colleague, Mr. L. Barbosa.

INTRODUCTION

HISTORICAL BACKGROUND

Many problems in engineering and industrial planning can be reduced to the maximization of functions under constraints. These problems can be mathematically formulated as follows:

$$\text{Maximize} \quad \left\{ f(x) \;\middle|\; g(x) \geq 0,\; x \geq 0 \right\} \tag{1}$$

where $x = (x_1, \ldots, x_n)$ is an n-dimensional variable, $g = (g_1, \ldots g_n)$ is an m-dimensional function of x and $g(x) \geq 0$, $x \geq 0$ (which means $g_i(x) \geq 0$, $x_i \geq 0$ for each i) represents the constraint on x. f is a real-valued function of x and is the performance index or profit function.

Methods for solving such problems in nonlinear programming almost invariably depend on some use of Lagrange multipliers. These methods are extensions of the classical theory of Lagrange multipliers, and use only the first variations of the functions involved. The first satisfactory theoretical solution of (1) was presented in the paper by Kuhn and Tucker [1]. They show that under certain qualifications on the function g (which insure the adequacy of the first variations), the solution of (1) is related to the determination of a saddle-point of the Lagrangian function:

$$\Phi(x, \lambda) = f(x) + \langle \lambda, g(x) \rangle \tag{2}$$

Some papers [2] have since appeared which deal with the situation where the variable x in (1) ranges over more general spaces. An essential weakening of the constraint qualification of Kuhn and Tucker, and a clarification of its role was given by Arrow et al [3]. Cases where x is subjected to more general constraints than in (1) have been investigated by Arrow et al [4].

It should be noted that the situation discussed above is a static situation. Time does not enter into the formulation of these problems. In contrast to this, control engineers are frequently faced with problems which are essentially dynamic in nature. These problems may be abstracted into the following general form. We are given a system which can be represented as a difference equation

$$x(n+1) - x(n) = f_n(x(n), u(n)) , \quad n \geq 0 \tag{3}$$

or as a differential equation

$$\frac{dx}{dt}(t) = f(x(t), u(t), t) , \quad \geq 0 \tag{4}$$

where x(n) and x(t) represent the state-vector of the system at time n (in the discrete case (3)) and at time t (in the continuous case (4)), respectively. u(n) and u(t) represent the control vectors. We are given certain constraints on the state and on the control and we are required to find a control sequence $(u(n), n \geq 0)$ or a control function $(u(t), t \geq 0)$, such that the constraints are met and some scalar-valued performance index is maximized. The main theoretical solution to the continuous-time problem (4) is the "maximum principle" of Pontryagin et al [5]. In his dissertation, Jordan [6] gives a maximum principle for the discrete case (3). His approach is essentially a translation of the methods of Pontryagin. The situation envisaged in (3) and (4) can be further complicated if we introduce randomness into the picture so that x and u are now random variables [7]. It should be remarked that these methods also limit themselves to first variations. In the formulation of the "maximum principles," an important part is played by a vector ψ which is the solution of the adjoint equations of (3) and (4). It is intuitively clear that this ψ vector is the ubiquitous Lagrange multiplier, and if so, we should be able to derive these results from suitable generalizations of the Kuhn-Tucker theorems, for example. So far, however, no such contributions have appeared in the literature.

The practical applicability of nonlinear programming to engineering and industrial problems has been limited to a certain extent by the size or "dimensionality" of the problem. In an attempt to meet this contingency, a considerable amount of effort has been directed to obtain a sort of "decomposition theory." The basic idea is the following. Many large problems can be "decomposed" into a number of autonomous smaller problems which are coupled either through constraints or through the profit function or both. Is it possible to reformulate this problem in such a way that the modified problem can be decomposed and solved by its parts as it were, so as to yield a solution to the original problem? In the linear case (i.e., where f and g of (1) are linear), one may use the Dantzig-Wolfe decomposition technique [8]. A dual approach to a more general class of problems has been presented by Rosen [9]. Lasdon [10], in his dissertation, has suggested a decomposition technique which can be applied to a different class of problems.

PURPOSE OF THIS PAPER

While the three classes of problems referred to above--namely, constrained maximization, deterministic and stochastic optimal control and decomposition techniques--appear to be more or less unrelated, we hope to show that they are different versions of the same constrained maximization problems. Our model is a slight generalization of the Kuhn-Tucker model (1). Namely, we wish to

$$\text{Maximize} \quad \{ f(x) \mid g(x) \in A \,,\; x \in A' \} \tag{5}$$

where A' is an arbitrary set and A is any convex set. We shall show that (5) is related to a saddle-value problem. We also hope to show that the solution to (5) rests upon a very elementary and well-known geometric fact that under certain conditions two disjoint convex sets can be separated by a closed hyperplane. In order to account for certain applications, we have found it useful to allow the variable x in (5) to be an element of a Banach space, rather than the more usual, but slightly less general, Euclidean space. We feel that the proofs are not appreciably complicated or prolonged by this generality.

Far more important, in our point of view, is the fact that for all these maximization problems there exist pairs of "dual" cones which we call local cones and local polars, which in a sense convey all the information about first-order variations. The various maximization theories (viz., Kuhn-Tucker theorem, the Maximum Principle) then, give various conditions under which these sets and the relationships that they satisfy, may be determined. We thus hope to show that through the introduction of the notions of a local cone and a local polar we have presented a common framework with which we can deal with maximization problems. In individual cases, furthermore, these sets may have a more intuitive structure. Thus, for example, in Chapter V, we show that the so-called "cone of attainability" (see Reference 5) is an approximation of the local cone.

The structure of this paper is as follows: In Chapter 0, we accumulate (without proof) some of the well-known results of the theory of linear topological spaces. Details and proofs of these statements can be found in Dunford and Schwartz [11]. In Chapter I, we introduce some terminology and discover sets (the local cone and the local polar) and relations between them which are essential to a maximization theory which limits itself to the first variation only. Theorem 1.1 demonstrates the relevance of these sets. In Chapter II, we hope to make transparent the necessity of some sort of constraint qualification. Chapter III gives an extension of the Kuhn-Tucker theorem. We tackle problem (5) and the related saddle-value problem in Chapter IV. A maximum principle for both the deterministic discrete case (under more general conditions than in [6]), as well as the stochastic case, is obtained in Chapter V. A section of Chapter V is devoted to an extremal problem in differential equations. This section is heavily dependent on the papers by Gamkrelidze [13] and Neustadt [14]. The connection between the problem that we consider and a class of continuous-time optimal control problems with state trajectory constraints is shown in Neustadt [14].

Finally, the relation of (5) with classes of decomposition techniques is presented in two papers [15, 16] which will soon be available. We therefore omit this material here.

CHAPTER 0

PRELIMINARIES

In this chapter, we collect (without proof) some of the well-known facts in the theory of linear topological spaces. Throughout our discussion, the field associated with a linear space will be the field of real numbers. For a detailed explanation and for the proofs of these statements, the reader may refer to Dunford and Schwartz [11]. We shall assume also that all the topologies which we encounter are Hausdorff.

1. Separation theorem in linear topological spaces.

(a) Let X be a linear topological space and K_1 and K_2 be disjoint convex sets in X with K_1 open. Then there exists a non-zero <u>continuous</u> linear functional f which separates them, i.e., there exists a number α such that

$$f(x_1) \geq \alpha \geq f(x_2) \qquad \forall x_1 \in K_1 , \forall x_2 \in K_2$$

(Remark: The existence of f is equivalent to the existence of a proper closed hyperplane $\{x \mid f(x) = \alpha\}$ which separates K_1 and K_2.)

(b) Strong separation theorem. Let X be a locally convex linear topological space and K_1 and K_2 be disjoint closed convex sets in X with K_1 compact. Then there exists a continuous linear functional f, real numbers α and c, $c > o$ such that

$$f(x_1) \geq \alpha > \alpha - c \geq f(x_2) \quad \forall x_1 \in K_1 , \forall x_2 \in K_2$$

2. Def. A Banach space is a normed linear space which is complete in the metric induced by the norm.

(a) A Banach space (with the topology induced by the norm) is a locally convex linear topological space.

Def. If X is a Banach space, the space of linear continuous functionals on X is a Banach space and will be denoted by $X^* = \{x^*\}$

(b) The Hahn-Banach theorem. Let X be a B-space and X_1 a subspace of X (i.e., a closed linear manifold of X). Let $x_1^* \in X_1^*$. Then there exists $x^* \in X^*$ such that

x^* extends x_1^* and

$$\|x^*\| = \|x_1^*\|$$

(c) The Open Mapping theorem. Let X and Y be B-spaces and f a linear continuous function from X onto Y. Then f is open, i.e., if U is an open set in X, f(U) is an open set in Y.

(d) Derivatives in Banach space. Let X and Y be B-spaces and f a map from X into Y. We say that f is differentiable at a point $x \in X$ iff,

$\exists$ a linear continuous function $f'(x)$ from X to Y such that

$$\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - f'(x)(h)\|}{\|h\|} = 0 , \forall h \in X$$

$f'(x)$ will be called the derivative (the Frechet-derivative) of f at x. $f'(x)$ is unique if it exists and the usual differential calculus applies [12].

(e) Weak topologies in Banach Spaces. Let X be a Banach space and X^* its dual. The usual topology on X is the one induced by the norm.

It is also called the norm-, or uniform-, topology. However, we can also induce another topology on X which is weaker than the norm-topology.

Def. The weak topology on X is the weakest topology on X which keeps all the elements $x^* \in X^*$ continuous.

Alternately, a net $\{x_\alpha\}$ in X converges weakly to an element $x \in X$ (i.e., in the weak topology of X) iff

$$x^*(x_\alpha) \rightarrow x^*(x) \quad \forall x^* \in X^*$$

Now let x be a fixed element in X. We can consider x as a function $\tilde{x}$ on X^* as follows:

$$\tilde{x}(x^*) \equiv x^*(x)$$

Thus the map $x \rightarrow \tilde{x}$ is an imbedding of X onto $\tilde{X} \subseteq X^{**}$. Dually we have,

Def. The weak* topology of X^* is the weakest topology on X^* which keeps all the elements $\tilde{x}$ of $\tilde{X}$ continuous.

Alternately, a net $\{x^*_\alpha\}$ in X^* converges weak*ly to an element $x^* \in X^*$ (i.e., in the weak* topology of X^*) iff

$$x^*_\alpha(x) \rightarrow x^*(x) \quad \forall x \in X$$

(f) The weak topology on X and the weak* topology on X^* are locally convex topologies.

Def. A B-space X is reflexive iff $\tilde{X} = X^{**}$.

Def. If X is a B-space and $x^* \in X^*$

$$\langle x^*, x\rangle \equiv x^*(x)$$

If X and Y are B-space and $g: X \rightarrow Y$ is a linear map,

$$\langle g, x\rangle \equiv g(x)$$

CHAPTER I

Here we introduce some terminology and define sets which we call local cones (LC) and local polars (LP) that are essential to a maximization theory that only takes into account first variations. We shall obtain some relations between these sets using elementary manipulations. One of the more important results will be an analog of the Bipolar theorem in the theory of dual spaces, and which as a special case yields Farkas' Lemma. The relevance of the local cones and local polar is given by Theorem 1.1 which gives necessary conditions for the maximization of a function when the variable ranges over a subset of the entire space.

Let X be a locally convex (real) linear topological space and X* its dual (X* is given the uniform topology).

Def. 1.1:[1] Let A be a non-empty set in X (X*). Let $\underline{x} \in A$ ($\underline{x}^* \in A$). By the closed cone[2] generated by A at $\underline{x}$ ($\underline{x}^*$), we mean the intersection of all closed cones containing the set $A-\underline{x} \equiv \{a-\underline{x} \mid a \in A\}$ ($A-\underline{x}^* \equiv \{a-\underline{x}^* \mid a \in A\}$). We denote this set by $C(A,\underline{x})$ ($C(A,\underline{x}^*)$).

Remark 1.1: If A is convex, $C(A,\underline{x})$ ($C(A,\underline{x}^*)$) is convex.

Def. 1.2: Let A be a non-empty set in X (X*). Let $\underline{x} \in A$ ($\underline{x}^* \in A$). By the polar of A at $\underline{x}$ ($\underline{x}^*$), we mean the set

---

[1] A set C is a cone if $(\forall \alpha \geq 0)(\forall x \in C)\ [\alpha x \in C]$

[2] Def. 1.1, 1.2, 1.3, 1.4 are not necessarily standard in the literature.

$$P(A,\underline{x}) = \{x^* | x^* \in X^*, \langle x^*, x\rangle \equiv x^*(x) \leq 0 \forall x \in C(A,\underline{x})\}$$

$$(P(A,\underline{x}^*) = \{x \mid x \in X, \langle x, x^*\rangle \equiv x^*(x) \leq 0 \forall x^* \in C(A,\underline{x}^*)\})$$

If A is a cone, $\underline{P(A) \equiv P(A, 0)}$

Remark 1.2: (a) $P(A,\underline{x})$ is a closed convex cone in X^*.

(b) $P(A,\underline{x}^*)$ is a closed convex cone in X.

Def. 1.3: Let A be a non-empty set in X and $\underline{x} \in A$. Let $\mathcal{N}(\underline{x})$ be the neighborhood system at $\underline{x}$ (i.e., the class of all neighborhoods of $\underline{x}$). By <u>the local closed cone generated by A at $\underline{x}$,</u> we mean the set

$$LC(A,\underline{x}) = \bigcap_{N \in \mathcal{N}(\underline{x})} C(A \cap N, \underline{x})$$

Remark 1.3: (a) If A is convex, $LC(A,\underline{x}) = C(A,\underline{x})$

(b) We are not interested in local cones in X^*.

Def. 1.4: Let A be a non-empty set in X and $\underline{x} \in A$. By <u>the local polar of A at $\underline{x}$</u> we mean the set

$$LP(A,\underline{x}) = \{x^* | x^* \in X^*, \langle x^*, x\rangle \leq 0 \quad \forall x \in LC(A,\underline{x})\}$$

Remark 1.4: (a) $LP(A,\underline{x})$ is a closed, convex cone in X^*.

(b) See Fig. 1.1 for an illustration of the objects defined above.

Def. 1.5: Let A be a cone. Then <u>Co(A)</u> is the intersection of all the closed convex cones containing A.

Fact 1.1 (Analog of the Bipolar Theorem). Let A be a cone in X. Then $(P(P(A)) = Co(A)$. In particular if A is closed and convex, $P(P(A)) = A$.

Proof: (a) $P(P(A)) \supseteq Co(A)$.

By Remark 1.2 (b) it is sufficient to show that $P(P(A)) \supseteq A$.

Let $x \in A$ be fixed and $x^* \in P(A)$.

Then $\langle x^*, x\rangle \leq 0 \ \forall x^* \in P(A)$. $\therefore x \in P(P(A))$.

(b) $P(P(A)) \subseteq Co(A)$.

Suppose $\underline{x} \in P(P(A))$ and $\underline{x} \notin Co(A)$.

Then by the strong separation theorem, $\exists\, x^* \in X^*$, $x^* \neq 0$, α real and $\epsilon > 0$, such that

$$\langle x^*, \underline{x}\rangle \geq \alpha > \alpha - \epsilon \geq \langle x^*, x\rangle \quad \forall x \in Co(A) \tag{1}$$

Since $Co(A)$ is a cone, $0 \in Co(A)$

$$\therefore\ \alpha - \epsilon \geq \langle x^*, 0\rangle = 0$$

$$\therefore\ \quad = > 0.$$

Again since $(Co(A)$ is a cone, $\therefore\ \lambda x \in Co(A)\ \forall \lambda \geq 0$, $\forall x \in Co(A)$.

$$\therefore\ \alpha > \langle x^*, x\rangle \quad \forall x \in Co(A)$$

$$\Rightarrow 0 \geq \langle x^*, x\rangle \quad \forall x \in Co(A)$$

$$\therefore\ \langle x^*, \underline{x}\rangle \geq \alpha > 0 \geq \langle x^*, x\rangle \quad \forall x \in Co(A)$$

$\therefore\ x^* \in P(Co(A))$ by definition.

Moreover $A \subseteq Co(A) \Rightarrow P(A) \supseteq P(Co(A))$.

$\therefore\ x^* \in P(A)$ and since $\langle x^*, \underline{x}\rangle > 0$, $\underline{x} \notin P(P(A))$.

Q.E.D.

Corollary 1.1 (Farkas'Lemma). Let $a_1, \ldots, a_n, b$ be vectors in a finite-dimensional Euclidean space. If,

$$\langle a_i, x\rangle \leq 0 \quad \forall_i \Rightarrow \langle b, x\rangle \leq 0, \text{ then}$$

$$\exists\, \lambda_1, \ldots, \lambda_n \geq 0 \text{ such that } b = \sum_{i=1}^{n} \lambda_i a_i$$

Fact 1.2 Let A be a cone in X. Then

$$P(A) = P(Co(A))$$

Proof: (a) $A \subseteq Co(A) \quad P(A) \supseteq P(Co(A))$.

(b) Let $x^* \in P(A)$ $\therefore\ \langle x^*, x\rangle \leq 0\ \forall x \in A$.

Let $\{x_1, \ldots, x_n\} \subseteq A$ and $\lambda_1, \ldots, \lambda_n$ be positive.

Then $\langle x^*, x_i\rangle \leq 0 \quad \forall_i \Rightarrow \langle x^*, \sum_{i=1}^{n} \lambda_i x_i\rangle \leq 0.$

By continuity we get,

$$\langle x^*, x \rangle \leq 0 \quad \forall x \in Co(A)$$

$$\therefore\ x^* \in P(Co(A)).$$

Q.E.D.

Corollary 1.2 $LP(A, \underline{x}) = P(Co(LC(A, x)))$.

Proof: $LP(A, \underline{x}) = P(LC(A, \underline{x}))$ by Def. 1.2 and Def. 1.4.

$= P(Co(LC(A, \underline{x})))$ by Fact 1.2.

Q.E.D.

Fact 1.3 Let A_1 and A_2 be closed convex cones in X. Then

$$P(A_1 \cap A_2) = \overline{P(A_1) + P(A_2)}$$

Proof: (a) $P(A_1 \cap A_2) \supseteq \overline{P(A_1) + P(A_2)}$

Let $x_1^* \in P(A_1)$, $x_2^* \in P(A_2)$ and $x \in A_1 \cap A_2$

$\therefore\ \langle x_1^*, x \rangle \leq 0$ and $\langle x_2^*, x \rangle \leq 0$ so that

$\langle x_1^* + x_2^*, x \rangle \leq 0$. $\therefore\ (x_1^* + x_2^*) \in P(A_1 \cap A_2)$

$P(A_1 \cap A_2) \supseteq P(A_1) + P(A_2)$.

Since $P(A_1 \cap A_2)$ is closed, the assertion follows.

(b) $P(A_1 \cap A_2) \subseteq \overline{P(A_1) + P(A_2)}$.

Let $\underline{x}^* \in P(A_1 \cap A_2)$ and $\underline{x}^* \notin \overline{P(A_1) + P(A_2)}$.

By the strong separation theorem, $\exists\, x \in X$, α real and $\varepsilon > 0$ such that

$$\langle \underline{x}^*, x \rangle \geq \alpha > \alpha - \varepsilon \geq \langle x^*, x \rangle \ \forall x^* \in \overline{P(A_1) + P(A_2)}$$

$$\therefore \langle x^*, x \rangle > 0 \geq \langle x^*, x \rangle \forall x^* \in \overline{P(A_1) + P(A_2)}$$

$$\therefore \langle x^*, x \rangle \leq 0 \ \forall x^* \in \overline{P(A_1) + P(A_2)}$$

$$\therefore\ x \in P(P(A_1)) \cap P(P(A_2)) = A_1 \cap A_2 \text{ by Fact 1.1}$$

But $\underline{x}^* \in P(A_1 \cap A_2)$ so that $\langle x^*, x \rangle \leq 0$.

Q.E.D.

Corollary 1.3 Let A_1 and A_2 be cones in X. Then

$$P(A_1 \cap A_2) = \overline{P(A_1) + P(A_2)} \tag{1}$$

$$\text{iff} \quad Co(A_1 \cap A_2) = Co(A_1) \cap Co(A_2) \tag{2}$$

Proof: (a) (2)$\Longrightarrow$(1).

By Fact 1.3,

$$P(Co(A_1 \cap A_2)) = \overline{P(Co(A_1)) + P(Co(A_2))}$$

By Fact 1.2,

$$P(Co(A_1 \cap A_2)) = P(A_1 \cap A_2) \text{ and}$$

$$P(Co(A_i)) = P(A_i) \; i = 1, 2.$$

(b) (1)$\Longrightarrow$(2).

$$(1) \Longrightarrow P(P(A_1 \cap A_2)) = \overline{P(P(A_1) + P(A_2))}$$

By Fact 1.1,

$$P(P(A_1 \cap A_2)) = Co(A_1 \cap A_2) \tag{3}$$

By Fact 1.3,

$$P(Co(A_1) \cap Co(A_2)) = \overline{P(Co(A_1)) + P(Co(A_2))} = \overline{P(A_1) + P(A_2)} \text{ by Fact 1.2.}$$

$$\therefore P(\overline{P(A_1) + P(A_2)}) = P(P(Co(A_1) \cap Co(A_2)))$$

$$Co(A_1) \cap Co(A_2) \text{ by Fact 1.1} \tag{4}$$

From (1), (3), and (4) we obtain (2).

Q.E.D.

Corollary 1.4 Let A_1 and A_2 be non-empty sets in X and $\underline{x} \in A_1 \cap A_2$.

Then, $LP(A_1 \cap A_2, \underline{x}) = \overline{LP(A_1, \underline{x}) + LP(A_2, \underline{x})}$

iff $Co(LC(A_1 \cap A_2, \underline{x})) = Co(LC(A_1, \underline{x})) \cap Co(LC(A_2, \underline{x}))$

Remark 1.4 (a) The previous corollary will be useful in obtaining results both in "decomposition techniques" as well as in "optimal control." Suppose we have a variable x which is constrained to lie in two sets A_1 and

A_2, i.e., $x \in A_1 \cap A_2$. Now, as will be demonstrated in Theorem 1.1, the important set for maximization is the set $LP(A_1 \cap A_2, x)$. If x were constrained to be in A_i only, the relevant set would be $LP(A_i, x)$, $i = 1, 2$. Under what conditions can we obtain the set $LP(A_1 \cap A_2, x)$ from the decomposed pieces $LP(A_i, x)$, $i = 1, 2$? This corollary gives a partial answer to this. It may be noted that $LP(A_1 \cap A_2, x)$ is not always equal to $\overline{LP(A_1, x) + LP(A_2, x)}$. A simple counter-example is the following:

Let $X = E^2 = \{(x_1, x_2)\}$

Let A_1 be the x_1 axis, i.e., $A_1 = \{(x_1, 0) \mid x_1 \text{ arbitrary}\}$

and $A_2 = \{(x_1, x_2) \mid x_1^2 + (x_2 + 1)^2 \leq 1\}$

Let $x = (0, 0)$

Then $LP(A_1 \cap A_2, x) = E^2$

but $LP(A_1, x) + LP(A_2, x) = x_2$ axis.

(b) It is also interesting and important to determine conditions under which $\overline{LP(A_1, x) + LP(A_2, x)} = LP(A_1, x) + LP(A_2, x)$. Stated differently, let A_1 and A_2 be closed convex cones. When is $\overline{A_1 + A_2} = A_1 + A_2$? Unfortunately we have been unable to find a satisfactory solution to this problem.

Local Maxima

Let X be a real B-space and f a real-valued differentiable function on X. Let A be a non-empty set in X.

Def. 1.6 f has <u>a local maximum at $\underline{x}$ in A</u> iff

(i) $\underline{x} \in A$

(ii) $\exists$ a neighborhood N of $\underline{x}$ such that

$$f(\underline{x}) = \sup\left\{f(x) \mid x \in N \cap A\right\}$$

Theorem 1.1 f has a local maximum at $\underline{x}$ in $A \Longrightarrow$

$$f'(\underline{x}) \in LP(A, \underline{x})$$

Proof: Let N be a neighborhood of $\underline{x}$ such that

$$f(\underline{x}) = \sup\left\{f(x) \mid x \in N \cap A\right\}$$

Let M be a sufficiently large positive integer such that

$S_0 = \left\{x \mid \| x-\underline{x} \| \leq \frac{1}{M}\right\} \subseteq N$ and define

$$S_n = \left\{x \mid \| x-\underline{x} \| \leq \frac{1}{M+n}\right\} \qquad n = 0, 1, \ldots$$

Then, $N \supseteq S_0 \supseteq S_1 \ \ldots$ and $\bigcap_{n=0}^{\infty} S_n = \{\underline{x}\}$.

Now, for each $x_n \in A \cap S_n$ we have

$$f(x_n) \leq f(\underline{x}) \qquad n = 0, 1, \ldots \tag{1}$$

and since f is differentiable

$$f(x_n) = f(\underline{x}) + \langle f'(\underline{x}), x_n - \underline{x} \rangle + o(\| x_n - x \|) \leq f(\underline{x})$$

$$\therefore \langle f'(\underline{x}), x_n - \underline{x} \rangle = f(x_n) - f(\underline{x}) + o(\|x_n - \underline{x}\|) \leq o(\|x_n - \underline{x}\|)$$

$$\therefore \langle f'(\underline{x}), \frac{x_n - \underline{x}}{\|x_n - \underline{x}\|} \rangle \leq \frac{o(\|x_n - \underline{x}\|)}{\|x_n - \underline{x}\|} \text{ for } x_n \in A \cap S_n \quad n = 0, 1, \ldots$$

$$\therefore \lim_{n \to \infty} \sup_{x_n \in A \cap S_n} \langle f'(\underline{x}), \frac{x_n - \underline{x}}{\|x_n - \underline{x}\|} \rangle \leq 0$$

$$\therefore \lim_{n \to \infty} \sup_{x_n - \underline{x} \in C(A \cap S_n, \underline{x})} \langle f'(\underline{x}), \frac{x_n - x}{\|x_n - x\|} \rangle \leq 0$$

$$\therefore \sup\left\{\langle f'(\underline{x}), \frac{x}{\|x\|} \rangle \mid x \in \bigcap_{n=0}^{\infty} C(A \cap S_n, \underline{x}) = LC(A, \underline{x})\right\} \leq 0$$

$$\therefore \langle f'(\underline{x}), x \rangle \leq 0 \quad \forall x \in LC(A, \underline{x})$$

$\therefore f'(\underline{x}) \in LP(A, \underline{x})$ by Def. 1.4

Q.E.D.

Corollary 1.1 Let A be convex and f a concave function.

Then f has a maximum at $\underline{x}$ in A (i.e., $f(\underline{x}) \geq f(x) \quad \forall x \in A$)

$$\Longleftrightarrow \; f'(\underline{x}) \in LP(A, \underline{x})$$

Proof: "$\Longrightarrow$" follows from Theorem 3.1.

"$\Longleftarrow$" $LP(A, \underline{x}) = P(LC(A, \underline{x})) = P(C(A, \underline{x}))$ since A is convex

$$\therefore \; f'(\underline{x}) \in P(C(A, \underline{x})) \Longrightarrow \langle f'(\underline{x}), x-\underline{x} \rangle \leq 0 \;\; \forall x \in A \tag{1}$$

Now since f is concave,

$$\begin{aligned} f(x) &\leq f(\underline{x}) + \langle f'(\underline{x}), x-\underline{x} \rangle \;\; \forall x \in A \\ &\leq f(\underline{x}) \;\; \forall x \in A \text{ by (1)} \end{aligned}$$

Q.E.D.

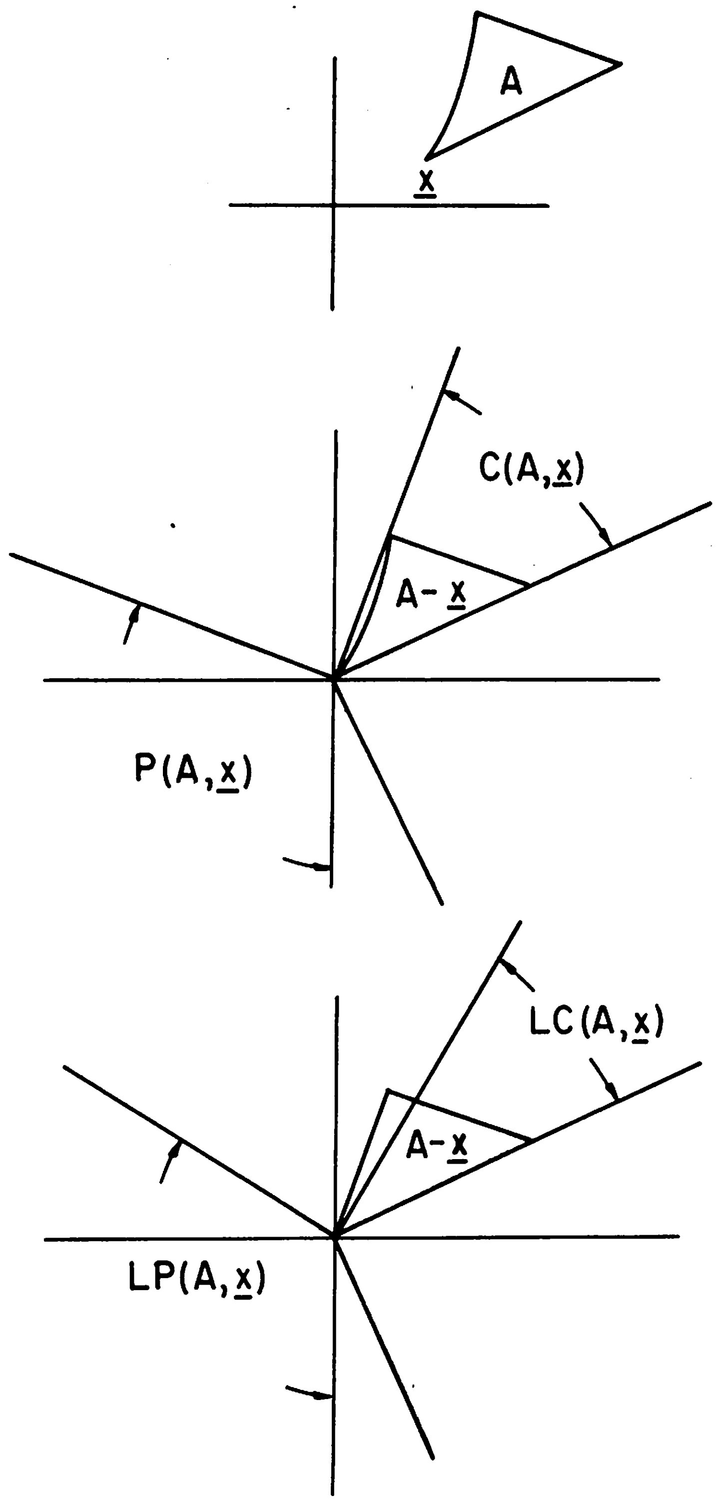

Fig. 1.1 Def 1.1 to 1.4

CHAPTER II

CONSTRAINT QUALIFICATION

In this chapter, we study the notion of constraint qualification. Two definitions are presented. One is that of Kuhn and Tucker [1]. The second is a weaker requirement, first suggested by Arrow et al [3]. We shall demonstrate the sufficiency of these requirements in terms of the sets introduced in Chapter I. Since we shall be dealing with derivatives of functions, we shall restrict ourselves to Banach spaces, because there is no adequate theory of differentiation in more general spaces. For a definition of the Fréchet derivative, the reader may refer to Chapter O. (For details, see [12]). It should be clear that the discussion of Chapter I is valid for B-spaces.

Let X and Y be real B-spaces; $g: X \to Y$ a differentiable map. Let A_Y be a non-empty subset of Y, and $A_X = \{x \mid g(x) \in A_Y\} = g^{-1}\{A_Y\}$. (Note: The definitions given below closely parallel Arrow et al [3].)

Let $\underline{x} \in A_X$.

Def. 2.1 We say that a vector $x \in X$ is an attainable direction at $\underline{x}$ if if there exists an arc $\{x(\theta) \mid 0 \leq \theta \leq 1\} \subseteq A_X$ such that

(1) $x(0) = \underline{x}$

(2) $x(\theta)$ is differentiable from the right at $\theta = 0$

and $\left.\frac{dx}{d\theta}\right|_{\theta=0} \equiv x'(0) = \Delta x$.

Let $AD(\underline{x}) = \{\Delta x \mid \Delta x \text{ is an attainable direction at } \underline{x}\}$

Clearly $AD(x)$ is a cone.

Let $A(\underline{x}) = Co(AD(\underline{x}))$ = closed, convex cone generated by $AD(\underline{x})$.

Def. 2.2 We say that a vector $\Delta x \in X$ is a <u>locally constrained direction at x</u> if $\langle g'(\underline{x}), \Delta x\rangle \equiv g'(\underline{x})(\Delta x) \in LC(A_Y, g(\underline{x}))$.

Let $L(\underline{x}) = \{\Delta x \mid \Delta x \text{ is a locally constrained direction at } \underline{x}\}$

Clearly $L(x)$ is a closed cone.

Fact 2.1 $AD(\underline{x}) \subseteq LC(A_X, \underline{x})$ hence

$$A(\underline{x}) \subseteq Co(LC(A_X, \underline{x}))$$

Proof: Let $\Delta x \in AD(\underline{x})$

$\therefore$ $\Delta x = x'(0)$ where $\{x(\theta) \mid 0 \leq \theta \leq 1\} \subseteq A_X$ and $x(0) = \underline{x}$

$$\therefore\ x(\theta) = x(0) + \theta x'(0) + o(\theta)$$

$$= \underline{x} + \theta \Delta x + o(\theta). \tag{1}$$

Let N be an arbitrary neighborhood of $\underline{x}$ and $\theta(N) > 0$ be sufficiently small so that

$x(\theta) \in A_X \cap N \quad \forall \theta \leq \theta(N)$

$\therefore$ $x(\theta) - \underline{x} \in A_X \cap N - \underline{x} \subseteq C(A_X \cap N, \underline{x})$

$\therefore$ $\Delta x + \frac{o(\theta)}{\theta} \in C(A_X \cap N, \underline{x})$ by (1)

$\therefore$ $\Delta x \in C(A_X \cap N, \underline{x})$ since it is closed.

As N was an arbitrary neighborhood of x we have

$\Delta x \in \bigcap_{N \in \mathcal{N}(\underline{x})} C(_X \cap N, \underline{x}) = LC(A_X, \underline{x})$ by Def. 1.3

Q.E.D.

Fact 2.2 $LC(A_X, \underline{x}) \subseteq L(\underline{x})$ hence,

$$Co(LC(A_X, \underline{x})) \subseteq Co(L(\underline{x})).$$

Proof: Let $\Delta x \in LC(A_X, \underline{x}) = \bigcap_{N \in \mathcal{N}(\underline{x})} C(A_X \cap N, \underline{x})$

$$= \bigcap_{n=1}^{\infty} C(A_X \cap S_n, \underline{x}) \text{ where}$$

$S_n = \left\{ x \mid \|\underline{x} - x\| \leq \frac{1}{n} \right\} \quad n = 1, 2, \ldots$

$\therefore \exists$ a sequence $\left\{ x_k^n \right\}_{k=1}^{\infty} \subseteq A_X \cap S_n$ and a sequence $\left\{ \lambda_k^n \right\}_{k=1}^{\infty}$

of positive numbers so that,

$$\lim_{k \to \infty} \lambda_k^n (x_k^n - \underline{x}) = \Delta x \qquad n = 1, 2, \ldots$$

By a diagonal argument, $\exists$ sequences $\left\{ x_{n_k}^n \right\}_{n=1}^{\infty}$ and $\left\{ \lambda_{n_k}^n \right\}_{n=1}^{\infty}$

so that,

$$\lim_{n \to \infty} \lambda_{n_k}^n (x_{n_k}^n - \underline{x}) = \Delta x \quad \text{and } x_{n_k}^n \in A_X \cap S_n \quad \therefore \lim_{n \to \infty} x_{n_k}^n = \underline{x}$$

Now, $g(x_{n_k}^n) = g(\underline{x}) + \langle g'(\underline{x}), x_{n_k}^n - \underline{x} \rangle + o(\|x_{n_k}^n - \underline{x}\|)$ (1)

Let N be an arbitrary neighborhood of $g(\underline{x})$ in Y. Let $n(N)$ be sufficiently large so that $g(x_{n_k}^n) \in A_Y \cap N$ for all $n \geq n(N)$.

$$\therefore (g(x_{n_k}^n) - g(\underline{x})) \in \left\{ A_Y \cap N - g(\underline{x}) \right\} \subseteq C(A_Y \cap N, g(\underline{x})) \ \forall n \geq n(N)$$

$$\therefore \frac{g(x_{n_k}^n) - g(\underline{x})}{\|x_{n_k}^n - \underline{x}\|} = \frac{\langle g'(\underline{x}), x_{n_k}^n - \underline{x} \rangle}{\|x_{n_k}^n - \underline{x}\|} + \frac{o(\|x_{n_k}^n - \underline{x}\|)}{\|x_{n_k}^n - \underline{x}\|} \in C(A_Y \cap N, g(\underline{x})) \tag{2}$$

$\forall n \geq n(N)$ by (1)

Also, $\Delta x = \lim_{n \to \infty} \lambda_{n_k}^n (x_{n_k}^n - \underline{x})$ so that if $\Delta x \neq 0$ we have,

$$\frac{\Delta x}{\|\Delta x\|} = \lim_{n \to \infty} \frac{\lambda_{n_k}^n (x_{n_k}^n - \underline{x})}{\|\lambda_{n_k}^n (x_{n_k}^n - \underline{x})\|} = \lim_{n \to \infty} \frac{x_{n_k}^n - \underline{x}}{\|x_{n_k}^n - \underline{x}\|}$$

In (2) letting $n \to \infty$ we get

$\langle g'(x), \Delta x \rangle \in C(A_Y \cap N, g(\underline{x}))$ for $\Delta x \neq 0$.

Since 0 always belongs to $C(A_Y \cap N, g(\underline{x}))$ we have for any neighborhood N of $g(\underline{x})$ that

$\langle g'(\underline{x}), \Delta x \rangle \in C(A_Y \cap N, g(\underline{x}))$

$$\therefore \langle g'(\underline{x}), \Delta x \rangle \in \bigcap_{N \in \mathcal{N}(g(\underline{x}))} C(A_Y \cap N, g(\underline{x})) = LC(A_Y, g(\underline{x}))$$

$\therefore \Delta x \in L(\underline{x})$. Q.E.D.

Combining Facts 2.1 and 2.2 we obtain

Lemma 2.1 (a) $AD(\underline{x}) \subseteq LC(A_X, \underline{x}) \subseteq L(\underline{x})$

(b) $A(\underline{x}) \subseteq Co(LC(A_X, \underline{x})) \subseteq Co(L(\underline{x}))$

Def. 2.3 We say that (g, A_X, A_Y) satisfies <u>the Kuhn-Tucker constraint qualification (KT)</u> if

$$AD(\underline{x}) \supseteq L(\underline{x}) \quad \forall \underline{x} \in A_X.$$

Def. 2.4 We say that (g, A_X, A_Y) satisfies <u>the weak constraint qualification (W)</u> if

$$A(\underline{x}) \supseteq L(\underline{x}) \quad \forall \underline{x} \in A_X$$

Remark 2.1 $KT \Longrightarrow W$ since $AD(\underline{x}) \subseteq A(\underline{x})$.

Corollary 2.1 (a) If (g, A_X, A_Y) satisfies KT, then

$$LC(A_X, \underline{x}) = L(\underline{x}) = \{\Delta x \mid \langle g'(\underline{x}), \Delta x \rangle \in LC(A_Y, g(\underline{x}))\}$$

(b) If (g, A_X, A_Y) satisfies W, and if A_Y is a convex set in Y then

$$Co(LC(A_X, \underline{x})) = L(\underline{x}) = \{\Delta x \mid \langle g'(\underline{x}), \Delta x \rangle \in LC(A_Y, g(\underline{x}))\}$$

Proof: (a) follows from Lemma 2.1 (a) and Def. 2.3.

(b) follows from Lemma 2.1 (b), Def. 2.4 and the fact that A_Y convex implies $A_Y - g(\underline{x})$ convex so that

$$LC(A_Y, g(\underline{x})) = Co(LC(A_Y, g(\underline{x}))) \qquad \text{Q.E.D.}$$

Remark 2.2 It was demonstrated in the last chapter that the sets important for our discussion are $LC(A_X, \underline{x})$ and $LP(A_X, \underline{x})$. Now usually the set A_X is given indirectly as $A_X = g^{-1}\{A_Y\}$ and cannot be explicitly determined. However, the set A_Y is given and $LC(A_Y, g(\underline{x}))$ can be easily computed. The constraint qualifications, presented above, enable us to determine the unknown sets $LC(A_X, \underline{x})$ from the sets $LC(A_Y, g(\underline{x}))$. In fact, as is shown in the next result, the set $LP(A_X, \underline{x})$ has an even simpler form if a constraint qualification is satisfied.

Theorem 2.1 Let A_Y be a convex set in Y and assume that (g, A_X, A_Y) satisfies W. Let $\underline{x} \in A_X$, then

$$LP(A_X, \underline{x}) = \overline{LP(A_Y, g(\underline{x})) \circ g'(\underline{x})} \quad \text{where}$$

$$LP(A_Y, g(\underline{x})) \circ g'(\underline{x}) = \left\{ y^* \circ g'(\underline{x}) \;\middle|\; y^* \in LP(A_Y, g(\underline{x})) \right\}$$

Proof: (a) $LP(A_X, \underline{x}) \subseteq LP(A_Y, g(\underline{x})) \circ g'(\underline{x}) = B$ say.

First notice that B is a closed convex cone in X^*. Suppose $\underline{x}^* \in LP(A_X, \underline{x})$ and $\underline{x}^* \notin B$. Then, by the strong separation theorem, $\exists \Delta x \in X$, α real and $\epsilon > 0$, such that

$$\langle \underline{x}^*, \Delta x\rangle = \alpha > \alpha - \epsilon \geq \langle x^*, \Delta x\rangle \quad \forall x^* \in B.$$

Since B is a cone, we have

$$\langle \underline{x}^*, \Delta x\rangle > 0 \geq \langle x^*, \Delta x\rangle \quad \forall x^* \in B. \qquad (1)$$

$$\therefore \Delta x \notin Co(LC(A_X, \underline{x})) \Longrightarrow \langle g'(\underline{x}), \Delta x\rangle \equiv \underline{\Delta y} \notin LC(A_Y, g(\underline{x}))$$

by Corollary 2.1 (b)

By hypothesis, A_Y is convex, so that $LC(A_Y, g(\underline{x})$ is a closed convex cone, not containing $\underline{\Delta y}$. Once again using the strong separation theorem, $\exists\, y^* \in Y^*$, β real and $\delta > 0$ so that

$$\langle y^*, \underline{\Delta y} \rangle = \beta > \beta - \delta \geq \langle y^*, \Delta y\rangle \quad \forall \Delta y \in LC(A_Y, g(\underline{x}))$$

Again,

$$\langle y^*, \underline{\Delta y}\rangle > 0 \geq \langle y^*, \Delta y\rangle \quad \forall \Delta y \in LC(A_Y, g(\underline{x}))$$

$$\therefore\ y^* \in LP(A_Y, g(\underline{x}))$$

$$\therefore\ y^* \circ g'(\underline{x}) \in LP(A_Y, g(\underline{x})) \circ g'(\underline{x}) \subseteq B.$$

But $\langle y^* \circ g'(\underline{x}), \Delta x \rangle = \langle y^*, \underline{\Delta y}\rangle > 0$ which contradicts (1)

(b) $LP(A_X, \underline{x}) \supseteq LP(A_Y, g(\underline{x})) \circ g'(\underline{x})$

First notice that it is sufficient to show that

$$LP(A_X, \underline{x}) \supseteq LP(A_Y, g(\underline{x})) \circ g'(\underline{x}).$$

Suppose $\exists\, y^* \in LP(A_Y, g(\underline{x}))$ such that $y^* \circ g'(\underline{x}) \notin LP(A_X, \underline{x})$.

Then, $\exists\, \underline{\Delta x} \in X$, α real and $\epsilon > 0$ such that

$$\langle \underline{y}^* \circ g'(\underline{x}), \underline{\Delta x}\rangle = \alpha > \alpha - \epsilon \geq \langle x^*, \underline{\Delta x}\rangle \quad \forall x^* \in LP(A_X, \underline{x})$$

$$\therefore \langle y^* \circ g'(\underline{x}), \underline{\Delta x}\rangle > 0 \geq \langle x^*, \underline{\Delta x}\rangle \quad \forall x^* \in LP(A_X, \underline{x})$$

$$\therefore\quad \underline{\Delta x} \in P(LP(A_X, \underline{x})) = Co(LC(A_X, \underline{x})) \tag{1}$$

Moreover $\langle y^* \circ g'(\underline{x}), \underline{\Delta x}\rangle = \langle \underline{y}^*, g'(x)(\underline{\Delta x})\rangle = \langle \underline{y}^*, \underline{\Delta y}\rangle = \alpha > 0$

$\therefore\ \underline{\Delta y} \notin Co(LC(A_Y, g(\underline{x}) = LC(A_Y, g(\underline{x})).$

But by (1) and Cor. 2.1(b) $\langle g'(\underline{x}), \underline{\Delta x}\rangle = \underline{\Delta y} \in LC(A_Y, g(\underline{x})).$

Corollary 2.2 If in the hypothesis of the above theorem, A_Y is a convex cone then

$$LP(A_X, \underline{x}) = \overline{LP(A_Y, g(\underline{x})) \circ g'(\underline{x})} \subseteq \overline{P(A_Y) \circ g'(\underline{x})}.$$

Proof: Since A_Y is a convex cone and $g(\underline{x}) \in A_Y$,

$\therefore \; A_Y \subseteq A_Y - g(\underline{x})$

$\therefore \; Co(A_Y) \subseteq LC(A_Y, \; g(\underline{x}))$

$\therefore \; P(A_Y) \supseteq LP(A_Y, \; g(\underline{x}))$. The rest follows by Theorem 2.1

Q.E.D.

Corollary 2.3 In the hypothesis of Theorem 2.1 let $A_Y = \{0\}$.

Then $\quad LC(A_X, \; \underline{x}) = \{\Delta x \mid \langle g'(\underline{x}), \; \Delta x \rangle = o\}$ and

$$LP(A_X, \; \underline{x}) = \overline{Y^* o \; g'(\underline{x})}.$$

Remark 2.2 It is necessary to determine conditions under which $Y^* o \; g'(\underline{x}) = Y^* o \; g'(\underline{x})$. If this were true, then in Corollary 2.3, any element $x^* \in LP(A_X, \; \underline{x})$ could be expressed as $x^* = y^* o \; g'(\underline{x})$ where $y^* \in Y^*$. A partial answer to this is given in the following assertions.

Fact 2.3 Let X and Y be real B-spaces and $f: X \longrightarrow Y$ be a linear continuous function. Define $\tilde{f}: Y^* \longrightarrow X^*$ by $\tilde{f}(y^*) \equiv y^* o \; f$.

Then

(a) f has closed range $\Rightarrow \tilde{f}$ has closed range i.e., $Y^* o \; f = \overline{Y^* o \; f}$.

(b) $\tilde{f}$ has closed range and X is reflexive $\Rightarrow f$ has closed range.

Proof: (a) Let $N \subseteq X$ be the null subspace of f, i.e., $N = \{x \mid f(x) = 0\}$ Let $X_1^* \subseteq X^*$ be the subspace of all elements $x^* \in X^*$ such that $x \in N \Rightarrow x^*(x) = 0$. Then by Corollary 2.3 (taking $g = f$) we have,

$$X_1^* = \overline{Y^* o \; f}.$$

Let $Y_1 \subseteq Y$ be the range of f. Then by hypothesis Y_1 is a (closed) subspace of Y. We give Y_1 its relative (induced) topology and regard it as a B-space, so that $f: X \rightarrow Y_1$ is a linear continuous onto map. By the Open Mapping Theorem f is an open map of X onto Y_1.

Let $x_1^* \in X_1^*$

Define a function $\tilde{y}_1$ on Y_1 as follows:

$$\tilde{y}_1(y_1) = x_1^*(x) \text{ where } x \in X \text{ such that } f(x) = y_1.$$

(i) $\tilde{y}_1$ is well-defined. $\because$ $f(x_1) = f(x_2) \Longleftrightarrow f(x_1 - x_2) = 0 \Longrightarrow$ $x_1^*(x_1 - x_2) = 0$ since $x_1^* \in X_1^* \Longleftrightarrow x_1^*(x_1) = x_1^*(x_2)$.

(ii) $\tilde{y}_1$ is linear

(iii) $\tilde{y}_1$ is continuous. Let $\{y_n\}_{n=1}^{\infty} \subseteq Y_1$ and $y_n \rightarrow 0$.

Since f is an open map, it can be easily shown that $\exists m < \infty$ and $\{x_n\}_{n=1}^{\infty} \subseteq X$ with $\|x_n\| \leq m\|y_n\|$ and $f(x_n) = y_n$.

$$\therefore |\tilde{y}_1(y_n)| = |x^*(x_n)| \leq \|x^*\| \; \|x_n\| \leq m \; \|x^*\| \; \|y_n\| \longrightarrow 0 \text{ as } n \rightarrow \infty$$

$\therefore$ $\tilde{y}_1$ is continuous at 0 and so $\tilde{y}_1 \in Y_1^*$.

(iv) By the Hahn-Banach Theorem $\exists y^* \in Y^*$ which extends $\tilde{y}_1$ so that

$$x_1^* = y^* \circ f. \quad \therefore x_1^* \in Y^* \circ f.$$

$\therefore$ (iv) $\Longrightarrow X_1^* = \overline{Y^* \circ f} = Y_1^* \circ f$.

(b) Let $X_1^* = \overline{Y^* \circ f} = Y^* \circ f$. Then $\tilde{f}: Y^* \longrightarrow X_1^*$ is onto.

Let $y \in \overline{f(X)}$

Define a function $\tilde{x}_1$ on X_1^* as follows:

$$\tilde{x}_1(x_1^*) = y^*(y) \text{ where } x_1^* = \tilde{f}(y^*) \qquad y^* \in Y^*.$$

By the same argument as before, we see that $\tilde{x}_1$ is a linear continuous function on X_1^*, i.e., $\tilde{x}_1 \in (X_1^*)^*$.

Now X is reflexive and X_1 is a subspace of X. $\therefore X_1$ is reflexive.

$\therefore \exists x_1 \in X_1$ such that $\tilde{x}_1(x_1^*) = \langle x_1^*, x_1 \rangle$.

$$y^*(y) = \tilde{x}_1(x_1^*) = \tilde{x}_1(\tilde{f}(y^*)) = \langle \tilde{f}(y^*), x_1 \rangle = \langle y^*, f(x_1) \rangle \quad \forall y^* \in Y^*$$

$$\therefore\ y = f(x_1). \quad \therefore\ y \in f(X).$$

Q.E.D.

Corollary 2.4 Suppose $(g, A_X, A_Y = \{0\})$ satisfies W. Let $\underline{x} \in A_X$. Then if $g'(\underline{x})$ has closed range,

$$LP(A_X, \underline{x}) = Y^* \circ g'(x) \overset{\text{def}}{=} \{ y^* \circ g'(\underline{x}) \mid y^* \in Y^* \}$$

CHAPTER III

AN EXTENSION OF THE KUHN-TUCKER THEOREM

This chapter may be considered to be a straightforward application of the results derived in Chapters I and II. Theorem 3.1 is a slight extension of the Kuhn-Tucker Theorem. The proof of this theorem demonstrates the need for some sort of constraint qualification.

Let X, Y be real B-spaces; g a differentiable map from $X \longrightarrow Y$, f a real-valued differentiable function on X. Let A_Y be a convex set in Y and $A_X = \{x \mid g(x) \in A_Y\} = g^{-1}\{A_Y\}$. We shall assume that (g, A_X, A_Y) satisfies the weak constraint qualification (W).

Theorem 3.1 (Extended Kuhn-Tucker Theorem)

(a) With the above hypothesis, if f has a local maximum at $\underline{x}$ in A_X, then,

$$f'(\underline{x}) \in \overline{LP(A_Y, g(\underline{x})) \circ g'(\underline{x})}$$

(b) In addition to the above hypothesis, suppose that A_X is convex and f concave. Then f has a maximum at $\underline{x}$ in A_X iff

$$f'(\underline{x}) \in \overline{LP(A_Y, g(\underline{x})) \circ g'(\underline{x})}$$

Proof: (a) By Theorem 1.1, f has a local maximum at $\underline{x}$ in A_X

$$\Longrightarrow f'(\underline{x}) \in LP(A_X, \underline{x}).$$

By Theorem 2.1, since (g, A_X, A_Y) satisfies W

$$LP(A_X, \underline{x}) = \overline{LP(A_Y, g(\underline{x})) \circ g'(\underline{x})}$$

(b) By Corollary 1.1, f has a maximum at $\underline{x}$ in A_X

$$\Longleftrightarrow f'(\underline{x}) \in LP(A_X, \underline{x})$$

$$\Longleftrightarrow f'(\underline{x}) \in \overline{LP(A_Y, g(\underline{x})) \circ g'(\underline{x})}$$ by Theorem 2.1.

Q.E.D.

Suppose now that X is a real B-space and $Y = E^m$. Let $A_Y = \{y = (y_1, \ldots\ldots, y_m) \mid y_i \geq 0 \quad \forall i\}$. Let $g(x) \equiv (g_1(x), \ldots\ldots, g_m(x))$, so that $A_X = \{x \mid g_i(x) \geq 0 \quad 1 \leq i \leq m\}$. Let $\underline{x} \in A_X$ and suppose that

$$\left.\begin{array}{l} g_1(\underline{x}) = \quad \ldots \quad = g_i(\underline{x}) = 0 \text{ and} \\ g_{i+1}(\underline{x}) > 0, \ldots, g_m(\underline{x}) > 0. \end{array}\right\} \qquad (1)$$

Then it should be clear that

$$LC(A_Y, g(\underline{x})) = \{(y_1, \ldots, y_m) \mid y_k \geq 0 \quad \forall k \leq i\}$$

so that $$LP(A_Y, g(\underline{x})) = \{(\lambda_1, \ldots, \lambda_m) \mid \lambda_k \leq 0,\ k \leq i;\ \lambda_k = 0,\ k > i\} \qquad (2)$$

Furthermore, $LP(A_y, g(\underline{x})) \circ g'(\underline{x})$ will be closed. Hence, if f has a local maximum at $\underline{x}$ in A_X, we will have

$$f'(\underline{x}) \in LP(A_Y, g(\underline{x})) \circ g'(\underline{x}) \text{ so that}$$

$$f'(\underline{x}) = \sum_{i=1}^{m} \lambda_i g_i'(\underline{x}) \text{ where } (\lambda_1, \ldots \lambda_m) \text{ satisfies (2).}$$

Also $\sum_{i=1}^{m} \lambda_i g_i(\underline{x}) = 0$ from (1) and (2).

We thus have the

Kuhn-Tucker Theorem: Let X be a real B-space and $g = (g_1, \ldots, g_m)$ be a differentiable mapping from $X \longrightarrow E^m$. Let f be a real-valued, differentiable function of x. Then

(a) A necessary condition that $\underline{x}$ solves the problem

$$\text{Maximize} \quad \left\{ f(x) \;\middle|\; g_i(x) \geq 0 \quad ; \quad i = 1, \ldots, m \right\} \text{ is}$$

(i) $g_i(\underline{x}) \geq 0$ and there exist numbers $\lambda_i \leq 0 \qquad 1 \leq i \leq m$

such that

$$\text{(ii)} \sum_{i=1}^{m} \lambda_i g_i(\underline{x}) = 0 \text{ and (iii) } f'(\underline{x}) = \sum_{i=1}^{m} \lambda_i g_i'(\underline{x}).$$

(b) If the functions $f, g_1, \ldots, g_m$ are also concave then, the conditions given above are sufficient.

Remark: We have demonstrated why we need some sort of a constraint qualification. By Theorem 1.1, we see that $f'(\underline{x}) \in LP(A_X, \underline{x})$. However, in order to relate $LP(A_X, \underline{x})$ with $LP(A_Y, g(\underline{x}))$ and $g'(\underline{x})$, we need a constraint qualification. This condition is sufficient but not necessary for $LP(A_X, \underline{x}) = \overline{LP(A_Y, g(\underline{x})) \circ g'(\underline{x})}$.

CHAPTER IV

A GENERALIZATION OF THE KUHN-TUCKER THEOREM AND THE RELATED SADDLE-VALUE PROBLEM

We recall the problem treated by Kuhn and Tucker [1].

$$\text{Maximize } \left\{ f(x) \;\middle|\; g(x) \geq 0, \quad x \geq 0 \right\} \tag{1}$$

where $x \in X(=E^n)$, $g: X \longrightarrow Y(=E^m)$ is a differentiable map and f is a real-valued, differentiable function of x. Equivalently,

$$\text{Maximize } \left\{ f(x) \;\middle|\; g(x) \in A_Y, \quad x \in A \right\}$$

where A_Y is the non-negative orthant in Y and A is the non-negative orthant in X. The related saddle-value problem is to find $\underline{x} \geq 0$, $\underline{y} \geq 0$ such that

$$\Phi(x, \underline{y}) \leq \Phi(\underline{x}, \underline{y}) \leq \Phi(\underline{x}, y) \qquad \forall x \geq 0, \quad \forall y \geq 0$$

where $\Phi(x, y) = f(x) + \langle y, g(x) \rangle$. We note that $x \geq 0 \Leftrightarrow x \in A$ and $y \geq 0 \Leftrightarrow y \in -P(A_Y)$.

We shall consider the following generalization of this problem,

$$\text{Maximize } \left\{ f(x) \;\middle|\; g(x) \in A_Y, \; x \in A \right\} \tag{2}$$

where A_Y is any convex set in Y and A is any set in X. This problem, however, does not have a natural corresponding saddle-value problem. If, however, we restrict A_Y to a closed convex cone, there is a related saddle-value problem. Namely find $\underline{x} \in A$, $\underline{y}^* \in -P(A_Y)$ such that

$$\Phi(x, \underline{y}^*) \leq \Phi(\underline{x}, \underline{y}^*) \leq \Phi(\underline{x}, y^*) \qquad \forall x \in A, \quad \forall y^* \in -P(A_Y)$$

and where $\Phi(x, y^*) = f(x) + \langle y^*, g(x) \rangle$.

In the first part of this chapter, we shall consider (2) with A_Y as any convex set. Then we shall specify the case where A_Y is a closed convex cone and put forward the corresponding saddle-value problem.

A. Let X and Y be real B-spaces; g, a differentiable map from X to Y and f, a real-valued, differentiable function of x. Let A_Y be a convex set in Y and $A_X = \{x \mid g(x) \in A_Y\}$. Assume that (g, A_X, A_Y) satisfies W. Let A be an arbitrary set in X.

Consider the following problem:

$$\text{Maximize } \{f(x) \mid g(x) \in A_Y,\ x \in A\} \tag{2}$$

Theorem 4.1 (Generalized Kuhn-Tucker Theorem)

(a) Suppose $\underline{x}$ solves (2). Then

$$f'(\underline{x}) \in LP(A \cap A_X, \underline{x}) \tag{3}$$

(b) If in addition $LP(A \cap A_X, \underline{x}) = LP(A, \underline{x}) + LP(A_X, \underline{x})$ (4)

then $\exists\, \underline{x}^* \in \overline{-LP(A_Y, f(\underline{x})) \circ g'(\underline{x})}$ such that

$$f'(\underline{x}) + \underline{x}^* \in LP(A, \underline{x}) \tag{5}$$

(c) If in addition $LP(A_Y, g(\underline{x})) \circ g'(\underline{x})$ is a closed set in X^*,

then $\exists\, \underline{y}^* \in -LP(A_Y, g(\underline{x}))$ such that

$$f'(\underline{x}) + \underline{y}^* \circ g'(\underline{x}) \in LP(A, \underline{x}) \tag{6}$$

(d) Conversely, suppose $\exists\, \underline{x} \in A \cap A_X$, $\exists\, \underline{y}^* \in -LP(A_Y, g(x))$ and suppose A is convex and $f(x) + \langle \underline{y}^*, g(x) \rangle$ is concave on A, then if

$\underline{y}^*$, $\underline{x}$ satisfy (6) and

$$\langle \underline{y}^*, g(\underline{x}) \rangle = 0 \tag{7}$$

then $\underline{x}$ solves (2).

(e) In any case, if $\underline{x}$ solves (2), $\exists\, \not{A} \geq 0$, $\exists\, \underline{y}^* \in -LP(A_Y, g(x))$ such that

$$\not{A}\, f'(x) + \underline{y}^* \circ g'(\underline{x}) \in LP(A, \underline{x}) \tag{8}$$

Proof: (a) By hypothesis, $\underline{x}$ solves (2), so that $f(\underline{x}) \geq f(x)$ $\forall x \in A \cap A_X$. $\therefore$ Theorem 1.1 $\Longrightarrow$ (3).

(b) By (3) and (4) we have,

$$f'(\underline{x}) \in LP(A_X, \underline{x}) = LP(A, \underline{x}) + LP(A_X, \underline{x})$$

$$= LP(A, \underline{x}) + \overline{LP(A_Y, g(x)) \circ g'(\underline{x})} \text{ by Theorem 2.1}$$

The last equality $\Longrightarrow$ (5).

(c) By (b) we have

$$f'(\underline{x}) \in \overline{LP(A_Y, g(x)) \circ g'(\underline{x})} + LP(A, \underline{x})$$

$$= LP(A_Y, g(x)) \circ g'(\underline{x}) + LP(A, \underline{x}) \text{ by hypothesis of (c).}$$

$\therefore \exists \underline{y}^* \in -LP(A_Y, g(\underline{x}))$ such that (6) holds.

(d) By the hypothesis of (d) $\exists \underline{x} \in A \cap A_X$ and $\exists \underline{y}^* \in -LP(A_Y, g(\underline{x}))$ such that

$$f'(\underline{x}) + \underline{y}^* \circ g'(x) \in LP(A, \underline{x}).$$

Moreover $f(x) + \langle \underline{y}^*, g(x) \rangle$ is concave on A so that by Corollary 1.1

$$f(x) + \langle \underline{y}^*, g(x) \rangle \leq f(\underline{x}) + \langle \underline{y}^*, g(x) \rangle \quad \forall x \in A.$$

Suppose in addition that $x \in A_X$, i.e., $g(x) \in A_Y$. Then since A_Y is convex, $(g(x) - g(\underline{x})) \in A_Y - g(\underline{x}) \subseteq LC(A_Y, g(\underline{x}))$.

Also $\underline{y}^* \in -LP(A_Y, g(x))$ so that

$$\langle \underline{y}^*, g(x) - g(\underline{x}) \rangle \geq 0. \quad \forall x \in A_X.$$

$\therefore \forall x \in A \cap A_X$ we have

$$f(x) \leq f(x) + \langle \underline{y}^*, g(x) \rangle \leq f(\underline{x}) + \langle \underline{y}^*, g(\underline{x}) \rangle = f(\underline{x})$$

since $\langle \underline{y}^*, g(\underline{x}) \rangle = 0$ by hypothesis. $\therefore$ $\underline{x}$ solves (2).

(e) This is obvious because 0 belongs to every cone.

Q.E.D.

For purposes of application to optimal control, we wish to strengthen part (e) of Theorem 4.1 for the following special situation.

Theorem 4.2. Let X and Y be real B-spaces and g a differentiable function from X to Y. Let f be a real-valued differentiable function of x. Let A be a non-empty set in X, and suppose that $\underline{x}$ solves (1).

$$\text{Maximize } \{ f(x) \mid g(x) = 0,\ x \in A \} \tag{1}$$

Then under assumptions A1 and A2, there is a $\mu \geq 0$ and a $\underline{y}^* \in Y^*$, not both zero such that

$$\mu f'(\underline{x}) + \underline{y}^* \circ g'(\underline{x}) \in LP(A, \underline{x}) \tag{2}$$

A1. Let $D = LC(A, \underline{x})$. We will assume that D is convex. Furthermore, if $D \neq \{0\}$, there is a subspace $Z \subseteq X$, such that D has a non-empty interior C relative to Z. Finally, if $z(\epsilon)$ for $\epsilon > 0$ is an arc in C, such that $\lim_{\epsilon \to 0} z(\epsilon) = \underline{x}$ and $z(\epsilon)$ is differentiable from the right at $\epsilon = 0$ with $z'(0) \in C$, then there is a sequence $\epsilon_n \to 0$ such that $z(\epsilon_n) \in A$.

A2. Let $G \equiv g'(\underline{x})$. We assume that if $\overline{G(D)} = Y$, then $G(D) = Y$. Let $N = \{ x \mid G(x) = 0 \}$. Then we shall assume that if $\overline{N + D} = X$, $N + D = X$. Also, if $LP(N) \cap LP(D) = \{0\}$ we will assume that $LP(N) + LP(D)$ is closed.

Remark 1: If X is finite-dimensional, then assumptions A1 and A2 are trivially satisfied when A is a finite union of disjoint closed convex sets.

Before we proceed to the proof of Theorem 4.2, we shall obtain some preliminary results which we shall need and which also have some independent interest.

Lemma 4.1. Let X and Y be B-spaces and G a continuous linear map from X to Y. Let D be a closed, convex cone in X such that $G(D) = Y$.

For $\rho > 0$, let $P_\rho = \{ \Delta x \mid \| \Delta x \| < \rho,\ \Delta x \in D \}$. Then there is a real number $m > 0$ such that

$$G(P_\rho) \supseteq S_{m\rho}$$

where $S_{m\rho}$ is the closed sphere in Y of center 0 and radius $m\rho$.

Proof: The proof of this lemma is a straightforward modification of the proof of the Open Mapping Theorem [11] and is, therefore, omitted.

Q.E.D.

Lemma 4.2. Let D be a closed convex cone in X, and g, a continuously differentiable function from X to Y such that $g(0) = 0$. Let $G \equiv g'(0)$ and suppose that $(\exists\, m>0)\ (\forall \rho>0)\ (G(P_\rho) \supseteq S_{m\rho})$. Let $z \in D$, $\|z\| = 1$ and $G(z) = 0$. Then there is a number $\epsilon_0 > 0$, and a function $o(\epsilon)$ such that for all $0 < \epsilon < \epsilon_0$, the set $g(\epsilon z + P_{o(\epsilon)})$ is a neighborhood of 0 in Y.

Proof: Let $v: X \to Y$ be the function defined by $v(x) = g(x) - G(x)$.

Then,
$$\|v(\epsilon z + x_1) - v(\epsilon z + x_2)\|$$
$$= \|g(\epsilon z + x_1) - g(\epsilon z + x_2) - G(x_1 - x_2)\|$$
$$= \|\langle g'(\epsilon z + x_1),\ x_1 - x_2\rangle + o_1(\|x_1 - x_2\|) - G(x_1 - x_2)\|$$

Therefore,
$$\frac{\|v(\epsilon z + x_1) - v(\epsilon z + x_2)\|}{\|x_1 - x_2\|} \le \|g'(\epsilon z + x_1) - G\| + \frac{o_1(\|x_1 - x_2\|)}{\|x_1 - x_2\|}$$

Also,
$$\|v(\epsilon z + x)\| = \|g(\epsilon z + x) - G(\epsilon z + x)\| = o_2(\|\epsilon z + x\|)$$

Pick a number $\epsilon_0 > 0$ such that if $0 < \epsilon < \epsilon_0$,
$$\frac{\|v(\epsilon z + x_1) - v(\epsilon z + x_2)\|}{\|x_1 - x_2\|} < \frac{m}{4} \quad \text{for } \|x_i\| < \epsilon \quad i = 1, 2$$

and
$$o_2(\|\epsilon z + x\|) = o(\epsilon) < \frac{m}{4}\,\epsilon \quad \text{for } \|x_i\| < \epsilon,\ i = 1, 2.$$

Fix $0 < \epsilon < \epsilon_0$ and let $y \in Y$ with $\|y\| < o(\epsilon)$.

Let $x_0 \in D$ such that $G(x_0) = y$ and $\|x_0\| < \frac{1}{m} \|y\| < \frac{1}{m} o(\epsilon)$.

Let $x_1 \in D$ such that $G(x_1 - x_0) = -v(\epsilon z + x_0)$ and $\|x_1 - x_0\| < \frac{1}{m}$

$$\|v(\epsilon z + x_0)\| < \frac{1}{m} o(\epsilon).$$

For $n \geq 1$, let $x_{n+1} \in D$ with $G(x_{n+1} - x_n) = -v(\epsilon z + x_n) + v(\epsilon z + x_{n-1})$ and $\| x_{n+1} - x_n\| < \frac{1}{m} o(\epsilon)$.

We first show that for $n \geq 0$, $\|x_n\| < \epsilon$ so that the above inequalities are valid. Firstly,

$$\|x_0\| < \frac{1}{m} o(\epsilon) < \frac{1}{4} \epsilon \quad \text{and} \quad \|x_1 - x_0\| < \frac{1}{m} o(\epsilon) < \frac{1}{4} \epsilon$$

$$\therefore \|x_1\| \leq \|x_0\| + \|x_1 - x_0\| < \frac{1}{2} \epsilon$$

By induction on n,

$$\|x_{n+1} - x_n\| < \left(\frac{o(\epsilon)}{m}\right)^n \|x_1 - x_0\| < \left(\frac{1}{4}\right)^n \|x_1 - x_0\|$$

$$\therefore \|x_{n+p} - x_p\| < \left(\frac{1}{4} \epsilon\right)^p \frac{1}{1 - \frac{1}{4}\epsilon} \|x_1 - x_0\| < \left(\frac{1}{4} \epsilon\right)^p \frac{2}{m} o(\epsilon)$$

In particular, $\| x_{n+1} - x_1\| < \frac{\epsilon}{2}$ so that $\|x_{n+1}\| < \frac{4}{m} o(\epsilon) < \epsilon$

Also x_n converges. Let $\lim_{n\to\infty} x_n = x$. Then $\|x\| < \frac{4}{m} o(\epsilon)$ and $x \in D$.

Now,

$$G(x_0) = y$$

$$G(x_1) - G(x_0) = -v(\epsilon z + x_0)$$

$$G(x_2) - G(x_1) = -v(\epsilon z + x_1) + v(\epsilon z + x_0)$$

$$\vdots$$

$$G(x_{n+1}) - G(x_n) = -v(\epsilon z + x_n) + v(\epsilon z + x_{n-1})$$

Adding both sides we get,

$$G(x_{n+1}) = y - v(\epsilon z + x_n), \; n \geq 0.$$

$$\therefore \quad y = G(x_{n+1}) - v(\epsilon z + x_n), \; n \geq 0.$$

Also $\quad \|y - g(\epsilon z + x_n)\| = \|y - G(x_n) - v(\epsilon z + x_n)\|$

$$= \|y - G(x_n) + G(x_{n+1}) - G(x_{n+1}) - v(\epsilon z + x_n)\|$$
$$= \|G(x_{n+1} - x_n)\| \leq \|G\| \, \|x_{n+1} - x_n\| \longrightarrow 0$$

$$\therefore \; g(\epsilon z + x_n) \longrightarrow y \quad \text{as} \quad n \longrightarrow \infty$$

But $x_n \to x$ so that $g(\epsilon z + x) = y$. Also $x \in D$ and $\|x\| < \frac{4}{m} o(\epsilon)$.

$$\therefore \; g(\epsilon z + P_{o(\epsilon)}) \supseteq S_{\frac{m}{4} o(\epsilon)}$$

Q.E.D.

Lemma 4.3 Let D and N be closed convex cones with $D + N = X$. For $\rho > 0$, let S_ρ be the closed sphere in X of center 0 and radius ρ Let $P_\rho = D \cap S_\rho$ and $N_\rho = N \cap S_\rho$. Then there is a number $m > 0$, such that $\quad P_\rho + N_\rho \supseteq S_{m\rho}$

Proof: $X = D + N = \bigcup_{n=1}^{\infty} (P_n + N_n)$. Now P_n is a closed, convex, bounded set and is, therefore, weakly compact. N_n is closed and convex and, hence, is weakly closed. Therefore, $P_n + N_n$ is a weakly closed, convex set, and, hence, it is strongly closed. The result follows by the Baire Category Theorem.

Q.E.D.

Lemma 4.4 Let D be a closed, convex cone in X such that there is a subspace $Z \subseteq X$ with $D \subseteq Z$ and such that D has a non-empty interior C relative to Z. Then, if N is any subspace of X such that $N + D = X$, we must have

$$D \cap N = \overline{C \cap N} \cup \{0\}.$$

Proof: Trivially, $D \cap N \supseteq \overline{C \cap N} \cup \{0\}$. To prove the converse, let $z \in D \cap N$ be any vector such that $\|z\| = 1$. Let $x \in X$ such that $z - x \in C$.

By lemma 3, there are $n \in N$ and $d \in D$ such that $x = n + d$ and $\|n\|, \|d\| < \frac{1}{m}\|x\|$. Then,

$$z-x = z-n-d$$

$$\therefore \quad z+d-x = z-n$$

Since $z-x \in C$ and $d \in D = \overline{C}$, therefore for all $\lambda > 0$ we must have, $z + \lambda(d-x) = z-\lambda n \in N \cap C$. Letting λ approach 0 we have, $z \in \overline{N \cap C}$.

Q.E.D.

Lemma 4.5. Let $g:X \longrightarrow Y$ be a continuously differentiable function. Let $A_X = \{x \mid g(x) = 0\}$. Let $\underline{x} \in A_X$ such that $G(X) = Y$ where $G \equiv g'(\underline{x})$. Then $(g, A_X, \{0\})$ satisfies the K.T. condition at $\underline{x}$.

Proof: Let $z \in X$ such that $G(z) = 0$. By Lemma 2, $\exists\, \epsilon_0 > 0$ such that for $0 < \epsilon < \epsilon_0$,

$$g(\underline{x} + \epsilon z + x(\epsilon)) = 0 \text{ and } \|x(\epsilon)\| < o(\epsilon).$$

Let $z(\epsilon) = \underline{x} + \epsilon z + x(\epsilon)$. Then $z(\epsilon) \in A_X$,

$$\lim_{\epsilon \to 0} z(\epsilon) = \underline{x} \text{ and } \lim_{\epsilon \to 0} \left\| \frac{z(\epsilon) - \underline{x} - \epsilon z}{\epsilon} \right\| = \lim_{\epsilon \to 0} \|x(\epsilon)\| = 0$$

so that $\left.\dfrac{dz(\epsilon)}{d\epsilon}\right|_{\epsilon = 0} = z.$

Q.E.D.

We now prove that under the assumptions A1 and A2, if $\underline{x}$ solves (1), then (2) is satisfied with $\mu \neq 0$ or $\underline{y}^* \neq 0$.

Case 1. Suppose $Q = \overline{g'(\underline{x})(D)} \neq Y$. Then Q is a proper closed convex cone in Y so that there is a $\underline{y}^* \in \underline{Y}^*$, $\underline{y}^* \neq 0$ such that

$$\langle \underline{y}^*, q \rangle \leq 0 \quad \forall q \in Q$$

$$\therefore \langle \underline{y}^*, g'(\underline{x})\, \Delta x \rangle \leq 0 \quad \forall \Delta x \in D$$

$$\therefore \langle \underline{y}^* \circ g'(\underline{x}), \Delta x \rangle \leq 0 \quad \forall \Delta x \in D$$

$$\therefore \quad \underline{y}^* \circ g'(\underline{x}) \in LP(D) = LP(A, \underline{x})$$

Hence (2) is satisfied with $\mu = 0$ and $\underline{y}^* \neq 0$.

Case 2a. Suppose $Q = \overline{g'(\underline{x})(D)} = Y$. Then, by assumption 2,

$$g'(\underline{x})(D) = Y.$$

For convenience let $G \equiv g'(\underline{x})$. Now, if $\underline{Y}^* \circ G \cap LP(D) \neq \{0\}$, then there is a $\underline{y}^* \neq 0$ with $\underline{y}^* \circ G \in LP(D)$, so that again (2) holds with $\mu = 0$, $\underline{y}^* \neq 0$.

Case 2b. $G(D) = Y$ and $\underline{Y}^* \circ G \cap LP(D) = \{0\}$. Let $N = \{\Delta x \mid G(\Delta x) = 0\}$. By Lemma 5, $(g, A_X, \{0\})$ satisfies K.T. so that $N = LC(A_X, \underline{x})$ and $\underline{Y}^* \circ G = LP(A_X, \underline{x})$. Now since $\underline{Y}^* \circ G \cap LP(D) = \{0\}$, we have by Fact 1.3,

$$\overline{N + D} = X.$$

By assumption 2 therefore, $N + D = X$. Hence by Lemma 4,

$$N \cap D = \overline{N \cap C} \cup \{0\}$$

We shall now show that,

$$LC(A_X \cap A, \underline{x}) = LC(A_X, \underline{x}) \cap LC(A, \underline{x}) = N \cap D$$

Trivially, $LC(A \cap A_X, \underline{x}) \subseteq N \cap D = \overline{N \cap C} \cup \{0\}$ where C is defined as in A1. Let $z \in N \cap C$, $\|z\| = 1$. By Lemma 2, for $0 < \epsilon < \epsilon_0$, $g(\underline{x} + \epsilon z + P_{o(\epsilon)})$ is a neighborhood of 0 in Y. Also, since $\epsilon z \in C$, we have $\epsilon z + P_{o(\epsilon)} \subseteq C$. Let $g(\underline{x} + \epsilon z + x_\epsilon) = 0$ where $\epsilon z + x_\epsilon \in C$ and $\|x_\epsilon\| < o(\epsilon)$. Let $x(\epsilon) = \underline{x} + \epsilon z + x_\epsilon$. Then $x(\epsilon) \longrightarrow \underline{x}$ as $\epsilon \longrightarrow 0$ and, $\frac{d\,x(\epsilon)}{d\,\epsilon} = z$. Therefore by A1, there is a sequence $\epsilon_n \longrightarrow 0$ such that $x(\epsilon_n) \in A$. Also $g(x(\epsilon_n)) = 0$ means, $x(\epsilon_n) \in A_X$. Therefore $x(\epsilon_n) \in A_X \cap A$. Since $\lim_{n \to \infty} \frac{1}{\epsilon_n}(x(\epsilon_n) - \underline{x}) = z$, we have $z \in LC(A_X \cap A)$.

$$\therefore \quad N \cap C \quad \{0\} \quad \subseteq \quad LC(A \cap A_X, \underline{x})$$

$$\therefore \quad \overline{N \cap C} \quad \{0\} \quad \subseteq \quad LC(A \cap A_X, \underline{x})$$

$$\therefore \quad N \cap D \quad = \quad LC(A \cap A_X, \underline{x})$$

Now by Theorem 1.1,

$$
\begin{aligned}
f'(\underline{x}) &\in LP(A_X \cap A,\ \underline{x}) \\
&= LP(LC(A_X \cap A,\ \underline{x})) \\
&= LP(N \cap D) \\
&= \overline{Y^* \circ g'(\underline{x}) + LP(A,\ \underline{x})} && \text{by Fact 1.3} \\
&= Y^* \circ g'(\underline{x}) + LP(A,\ \underline{x}) && \text{by A2}
\end{aligned}
$$

$\therefore\ \exists\ \underline{y}^* \in \underline{Y}^*$ such that,

$$f'(\underline{x}) + \underline{y}^* \circ g'(\underline{x}) \in LP(A,\ \underline{x})$$

Hence (2) is satisfied with $\mu \neq 0$.

Q.E.D.

B. Let X, Y be real B-spaces; g, a differentiable map from X into Y; f, a real-valued differentiable function of x. Let A_Y be a closed convex cone in Y and A an arbitrary subset in X. We assume that (g, A_X, A_Y) satisfies W. Consider the following three problems:

Problem 1. Saddle-Value Problem

Find $\underline{x} \in A$ and $\underline{y}^* \in -P(A_Y)$ such that

$$\Phi(x, \underline{y}^*) \leq \Phi(\underline{x}, \underline{y}) \leq \Phi(\underline{x}, y^*) \qquad \forall x \in A, \quad \forall y^* \in -P(A_Y)$$

where $\Phi(x, y^*) = f(x) + \langle y^*, g(x) \rangle$.

Problem 2.

Find $\underline{x}$ which solves

$$\text{Maximize } \left\{ f(x) \;\middle|\; g(x) \in A_Y,\ x \in A \right\}$$

Problem 3 ($\underline{y}^*$)

Find $\underline{x}$ which (for fixed $\underline{y}^*$) solves

$$\text{Maximize } \left\{ f(x) + \langle \underline{y}^*, g(x) \rangle \;\middle|\; x \in A \right\}.$$

Fact 4.1 a) If $(\underline{x}, \underline{y}^*)$ solves Problem 1, then

(i) $\dfrac{\partial \Phi}{\partial x}(\underline{x}, \underline{y}^*) \equiv f'(\underline{x}) + \underline{y}^* \circ g'(\underline{x}) \in LP(A,\ \underline{x})$

(ii) $g(\underline{x}) \in A_Y$ and (iii) $\langle \underline{y}^*, g(\underline{x}) \rangle = 0$.

b) If, moreover, A is convex and $\Phi(x, \underline{y}^*)$ is concave for $x \in A$, conditions (i), (ii), (iii) are sufficient for $(\underline{x}, \underline{Y}^*)$ to solve Problem 1.

c) If $(\underline{x}, \underline{y}^*)$ solves Problem 1, then $\underline{x}$ solves Problem 2.

Proof: a) $\Phi(x, \underline{y}^*) \leq \Phi(\underline{x}, \underline{y}^*) \quad \forall x \in A$.

$\therefore$ Theorem 1.1 $\Longrightarrow$ (i)

$$\Phi(\underline{x}, \underline{y}^*) \leq \Phi(\underline{x}, y^*) \quad \forall y^* \in -P(A_Y)$$

$$\therefore \quad \langle \underline{y}^*, g(x) \rangle \leq \langle y^*, g(x) \rangle \quad \forall y^* \in -P(A_Y) \tag{1}$$

Since $-P(A_Y)$ is a cone, $0 \in -P(A_Y)$

$\therefore \langle \underline{y}^*, g(\underline{x}) \rangle = \alpha \leq 0$.

Suppose $\alpha < 0$.

Then $\langle \underline{y}^*, g(\underline{x}) \rangle > \langle 2\underline{y}^*, g(x) \rangle$ and $2\underline{y}^* \in -P(A_Y)$

which contradicts (1).

$$\therefore \langle \underline{y}^*, g(x) \rangle = 0 \leq \langle y^*, g(x) \rangle \quad \forall y^* \in -P(A_Y) \tag{2}$$

This implies (iii) and $-g(\underline{x}) \in P(-P(A_Y)) = P(P(-A_Y)) = -A_Y$ by Fact 1.1

$\therefore \quad g(\underline{x}) \in A_Y$ giving (ii).

c) $g(\underline{x}) \in A_Y \Longrightarrow \underline{x}$ is a feasible solution to Problem 2.

Again, since $(\underline{x}, \underline{y}^*)$ solves Problem 1

$\therefore \Phi(x, \underline{y}^*) \leq \Phi(\underline{x}, \underline{y}^*) \quad \forall x \in A$.

$\therefore \quad f(x) + \langle \underline{y}^*, g(x) \rangle \leq f(\underline{x}) + \langle \underline{y}^*, g(\underline{x}) \rangle = f(\underline{x})$ by (iii).

Suppose $x \in A_X$, i.e., $g(x) \in A_Y$. Then since $\underline{y}^* \in -P(A_Y)$

we must have $\langle \underline{y}^*, g(x) \rangle \geq 0$.

$\therefore \forall x \in A \cap A_X$,

$$f(x) \leq f(x) + \langle \underline{y}^*, g(x) \rangle \leq f(\underline{x})$$

$\therefore \underline{x}$ solves Problem 2.

b) Suppose $\Phi(x, \underline{y}^*)$ is concave for $x \in A$ and A is convex. Then (i) and Corollary 1.1 $\Longrightarrow \Phi(x, \underline{y}^*) \leq \Phi(\underline{x}\ \underline{y}^*)$ $\forall x \in A$.
Now $\Phi(\underline{x}, y^*) = f(\underline{x}) + \langle \underline{y}^*, g(\underline{x}) \rangle = f(\underline{x})$ by (iii).
Also, for any $y^* \in -P(A_Y)$

$$\langle g(\underline{x}), y^* \rangle \geq 0 \text{ since } g(\underline{x}) \in A_Y \text{ by (ii).}$$

$\therefore \Phi(\underline{x}, \underline{y}^*) = f(\underline{x}) \leq f(\underline{x}) + \langle \underline{y}^*, g(\underline{x}) \rangle = \Phi(\underline{x}, y^*)$ $\forall y^* \in -P(A_Y)$

Q.E.D.

Fact 4.2 a) Suppose $\underline{x}$ solves Problem 2. Then

$$f'(\underline{x}) \in LP(A \cap A_X, \underline{x}) \tag{1}$$

b) If, in addition, $LP(A \cap A_X, \underline{x}) = LP(A, \underline{x}) + LP(A_X, \underline{x})$ (2)

Then $\exists\ \underline{x}^* \in \overline{-LP(A_Y, g(\underline{x})) \circ g'(\underline{x})} \subseteq -P(A_Y) \circ g'(\underline{x})$ such that

$$f'(\underline{x}) + \underline{x}^* \quad \in \quad LP(A, \underline{x}) \tag{3}$$

c) If, in addition, $\overline{-LP(A_Y, g(x)) \circ g'(\underline{x})} = -LP(A_Y, g(\underline{x})) \circ g'(\underline{x})$ (4)

then $\exists\ \underline{y}^* \in -LP(A_Y, g(\underline{x})) \subseteq -P(A_Y)$ such that

$$f'(\underline{x}) + \underline{y}^* \circ g'(\underline{x}) \in LP(A, \underline{x}) \tag{5}$$

d) If, in addition, A is convex, $f(x) + \langle \underline{y}^*, g(x) \rangle$ is concave on A and $\langle \underline{y}^*, g(\underline{x}) \rangle = 0$, then $(\underline{x}, \underline{y}^*)$ solves Problem 1.
e) In any case, if $\underline{x}$ solves Problem 2, there exists a $\mathcal{A} \geq 0$, $\exists$ $\underline{y}^* \in -LP(A_Y, g(\underline{x})) \subseteq -P(A_Y)$ such that

$$\mathcal{A} f'(\underline{x}) + \underline{y}^* \circ g'(\underline{x}) \in LP(A, \underline{x}) \tag{6}$$

Proof: a) By hypothesis $\underline{x}$ solves Problem 2
$\therefore f(\underline{x}) \geq f(x)$ $\forall x \in A \cap A_X$.

By Theorem 3.1

$$f'(\underline{x}) \in LP(A \cap A_X, \underline{x}) \text{ giving (1).}$$

b) By (2) and (1),

$$f'(x) \in LP(A \cap A_X, \underline{x}) = LP(A_X, \underline{x}) + LP(A, \underline{x})$$

$$= \overline{LP(A_Y, g(\underline{x})) \circ g'(\underline{x})} + LP(A, \underline{x}) \text{ by Theorem 2.1.}$$

$\therefore \exists\, \underline{x}^* \in \overline{LP(A_Y, g(\underline{x})) \circ g'(\underline{x})}$ such that (3) holds.

c) By b) we have

$$f'(\underline{x}) \in \overline{LP(A_Y, g(\underline{x})) \circ g'(\underline{x})} + LP(A, \underline{x})$$

$$= LP(A_Y, g(\underline{x})) \circ g'(\underline{x}) + LP(A, \underline{x}) \text{ by (4).}$$

$\therefore \exists\, \underline{y}^* \in -LP(A_Y, g(\underline{x}))$ such that (5) holds.

d) If (5) holds, A is convex and $f(x) + \langle \underline{y}^*, g(x) \rangle$ is concave on A, then by Cor. 1.1

$$f(x) + \langle \underline{y}^*, g(x) \rangle \leq f(\underline{x}) + \langle \underline{y}^*, g(\underline{x}) \rangle \quad \forall x \in A.$$

$$= f(\underline{x}). \quad \text{Since } \langle \underline{y}^*, g(\underline{x}) \rangle = 0 \text{ by hypothesis.}$$

Now $g(\underline{x}) \in A_Y$ so that if $y^* \in -P(A_Y)$ we have

$$\langle g(\underline{x}), y^* \rangle \geq 0.$$

Combining the above equations we get

$$f(x) + \langle \underline{y}^*, g(x) \rangle \leq f(\underline{x}) + \langle \underline{y}^*, g(\underline{x}) \rangle \leq f(\underline{x}) + \langle y^*, g(\underline{x}) \rangle$$
$$\forall x \in A, \ \forall y^* \in -P(A_Y)$$

$\therefore (\underline{x}, \underline{y}^*)$ solves Problem 1.

e) is obvious, since (6) is certainly satisfied for $\cancel{A} = 0$.

Fact 4.3: a) Suppose $\underline{x}$ solves Problem $3(\underline{y}^*)$. Then

$$f'(\underline{x}) + \underline{y}^* \circ g'(\underline{x}) \in LP(A, \underline{x}). \tag{1}$$

b) If, in addition, $\underline{y}^* \in -P(A_Y)$ then $(\underline{x}, \underline{y}^*)$ solves Problem 1

$$\Longleftrightarrow g(\underline{x}) \in A_Y \ , \tag{2}$$

$$\langle \underline{y}^*, g(\underline{x}) \rangle = 0 \tag{3}$$

Proof: a) Suppose $\underline{x}$ solves Problem $3(\underline{y}^*)$. i.e.,

$$f(x) + \langle \underline{y}^*, g(x) \rangle \leq f(\underline{x}) + \langle \underline{y}^*, g(\underline{x}) \rangle \quad \forall x \in A.$$

(1) follows by Theorem 1.1.

b) Suppose $(\underline{x}, y^*)$ solves Problem 1. Then

$$\Phi(x, \underline{y}^*) \le \Phi(\underline{x}, \underline{y}^*) \quad \forall x \in A.$$

$\therefore$ $\underline{x}$ solves Problem 3($\underline{y}^*$).

Conversely, if $\underline{x}$ solves Problem 3($\underline{y}^*$), then

$$\Phi(x, \underline{y}^*) \le \Phi(\underline{x}, \underline{y}^*) \quad \forall x \in A.$$

Now by hypothesis, $g(\underline{x}) \in A_Y$ so that

$$\langle y^*, g(\underline{x})\rangle \ge 0 \quad \forall y^* \in -P(A_Y).$$

$$\begin{aligned}\therefore\ \Phi(\underline{x}, \underline{y}^*) &= f(\underline{x}) + \langle \underline{y}^*, g(\underline{x})\rangle = f(\underline{x}) \quad \text{by hypothesis}\\ &\le f(\underline{x}) + \langle y^*, g(\underline{x})\rangle \quad \forall y^* \in -P(A_Y)\\ &= \Phi(\underline{x}, y^*).\end{aligned}$$

$\therefore$ $(\underline{x}, \underline{y}^*)$ solves Problem 1.

Q.E.D.

CHAPTER V

APPLICATIONS OF THE GENERALIZED KUHN-TUCKER THEOREM

Preliminaries: We shall need some notation for the direct product of B-spaces.

Def. 5.1 Let X_1 and X_2 be B-spaces with norms $\|\cdot\|_1$ and $\|\cdot\|_2$ respectively. Then the direct product $X_1 \otimes X_2 = \{(x_1, x_2) | x_1 \in X_1, x_2 \in X_2\}$ is a B-space under the norm

$$\|(x_1, x_2)\| = \max(\|x_1\|_1, \|x_2\|_2).$$

This norm is convenient for our purposes because of the following elementary fact.

Fact 5.1 The direct product $X_1{}^* \otimes X_2{}^*$ is a B-space under the norm $\|(x_1{}^*, x_2{}^*)\| = \|x_1{}^*\|_1 + \|x_2{}^*\|_2$. In fact, we have

$$X_1{}^* \otimes X_2{}^* = (X_1 \otimes X_2)^*$$

where $\langle (x_1{}^*. x_2{}^*), (x_1, x_2) \rangle \overset{\text{def}}{=} x_1{}^*(x_1) + x_2{}^*(x_2)$.

A. The Case of Discrete Optimal Control

Consider a system of difference equations.

$$x(k+1) = x(k) + f_k(x(k), u(k)) \qquad k = 0, 1, \ldots \tag{1}$$

where

$x \in X$ is the state vector

$u \in U$ is the control vector and

$f_k : X \otimes U \longrightarrow X$ is a differentiable map.

X, U are real B-spaces called the state space and control space, respectively.

The initial state $x(0)$ belongs to a subset $A(0) \subseteq X$. The target is a subset $A(N) \subseteq X$. The total gain incurred up to time $k > 0$ is given by $g_k(x(0), \ldots, x(k); u(0), \ldots u(k-1))$ where g_k is a real-valued differentiable map on $X^{k+1} \otimes U.^k$

Let $\Omega(k) \subseteq U$ for $k = 0, 1, \ldots$ be subsets. It is required to find

(i) a sequence of controls $\underline{u}(k) \in \Omega(k)$, $k = 0, 1, \ldots, N-1$

(ii) an initial state $\underline{x}(0) \in A(0)$ such that the sequence $(\underline{x}(0), \ldots, \underline{x}(N))$ satisfies (1) with $u(k) = \underline{u}(k)$ and $\underline{x}(N) \in A(N)$ and such that

$$g_N(\underline{x}(0), \ldots, \underline{x}(N); \underline{u}(0), \ldots, \underline{u}(N-1)) \text{ is maximized over all}$$

such sequences.

We can restate the problem in the following way

$$\operatorname{Max}\left\{ g_N(x(0), \ldots, x(N), u(0), \ldots, u(N-1)) \;\middle|\; \begin{array}{l} x(k) + f_k(x(k), u(k)) - x(k+1) \\ \qquad = 0 \quad k = 0, 1, \ldots N-1 \\ x(0) \in A(0),\ x(N) \in A(N) \text{ and} \\ u(k) \in \Omega(k) \quad k = 0, 1, \ldots, N-1 \end{array} \right.$$

or

$$\operatorname{Max}\left\{ g_N(x(0), \ldots, x(N), u(0), \ldots, u(N-1)) \;\middle|\; \begin{array}{l} h_k(x(k), u(k), x(k+1)) = 0 \\ \qquad k = 0, 1, \ldots, N-1 \\ x(0) \in A(0),\ x(N) \in A(N) \text{ and} \\ u(k) \in \Omega(k) \quad k = 0, 1, \ldots, N-1 \end{array} \right.$$

where the functions h_k are defined in the obvious way.

We now assume that the functions h_k and the constraint sets $\Omega(k)$ and $A(0)$ and $A(N)$ satisfy the assumptions of Theorem 4.2. Then by Theorem 4.2 we form the function

$$\Phi(\mu;\, x(0), \ldots, x(N);\, u(0), \ldots, u(N-1), \psi(1), \ldots, \psi(N)) =$$

$$\mu g_N + \sum_{k=0}^{N-1} \langle \psi(k+1), h_k \rangle \quad \text{where the } \psi(k) \in X^*.$$

Let $\{\underline{u}(0), \ldots, \underline{u}(N-1)\}$ be the optimal control and $\{\underline{x}(0), \ldots, \underline{x}(N)\}$ the optimal trajectory. Then by Theorem 4.2, the following conditions are satisfied.

There exist functions $\underline{\psi}(1), \ldots, \underline{\psi}(N)$ in X^* and a $\underline{\mu} \geq 0$ not all zero such that

$$\frac{\partial \Phi}{\partial x(0)} \in LP(A(0), \underline{x}(0))$$

$$\frac{\partial \Phi}{\partial x(k)} = 0 \qquad 0 < k \leq N-1$$

$$\frac{\partial \Phi}{\partial x(N)} \in LP(A(N), \underline{x}(N))$$

$$\frac{\partial \Phi}{\partial u(k)} \in LP(\Omega(k), \underline{u}(k)) \qquad 0 \leq k \leq N-1$$

where the partial derivatives are evaluated at

$$\mu = \underline{\mu};\; x(k) = \underline{x}(k);\; u(k) = \underline{u}(k);\; \psi(k) = \underline{\psi}(k) \quad \forall\, k.$$

Expanding these relations we get

$$\frac{\partial \Phi}{\partial x(0)} = \underline{\mu}\, \frac{\partial g_N}{\partial x(0)} + \underline{\psi}(1) + \underline{\psi}(1) \circ \left(\frac{\partial f_0}{\partial x(0)}\right) \in LP(A(0), \underline{x}(0)) \tag{2}$$

$$\frac{\partial \Phi}{\partial x(k)} = \underline{\mu}\, \frac{\partial g_N}{\partial x(0)} + \underline{\psi}(k+1) + \underline{\psi}(k+1) \circ \left(\frac{\partial f_k}{\partial x(k)}\right) - \underline{\psi}(k) = 0 \quad 0 < k < N \tag{3}$$

$$\frac{\partial \Phi}{\partial x(N)} = \underline{\mu}\, \frac{\partial g_N}{\partial x(N)} - \underline{\psi}(N) \in LP(A(N), \underline{x}(N)) \tag{4}$$

$$\frac{\partial \Phi}{\partial u(k)} = \underline{\mu}\, \frac{\partial g_N}{\partial u(k)} + \underline{\psi}(k+1) \circ \left(\frac{\partial f_k}{\partial u(k)}\right) \in LP(\Omega(k), \underline{u}(k)) \quad 0 \leq k \leq N-1 \tag{5}$$

Define $H(u(0), \ldots, u(N-1)) = \mathcal{A} g_N(\underline{x}(0), \ldots, \underline{x}(N); u(0), \ldots, u(N-1))$

$$+ \sum_{k=0}^{N-1} \langle \psi(k+1), f_k(\underline{x}(k), u(k)) \rangle \tag{6}$$

Then the equations (5) can be expressed as

$$\frac{\partial H}{\partial u}(u) \in LP\,(\Omega(0) \otimes \Omega(1) \otimes \ldots \otimes \Omega(N-1)\,;\, \underline{u}) \tag{7}$$

where $u = (u(0), \ldots, u(N-1)) \in U^N$ (8)

and $\underline{u} = (\underline{u}(0), \ldots, \underline{u}(N-1))$

Equation (3) can be written in a more familiar form as

$$\underline{\psi}(k+1) = \underline{\psi}(k) - \underline{\psi}(k+1) \circ \left(\frac{\partial fk}{\partial x(k)}\right) - \mathcal{A} \frac{\partial g_N}{\partial x(k)} \qquad 0 < k < N$$

Equations (2) and (4) give us the so-called "transversality conditions" when $A(0)$ and $A(N)$ are replaced by prescribed sets (e.g., singleton, manifold, etc.). These results can be summarized in the following theorem.

Theorem 5.1 Suppose that $\{\underline{u}(0), \ldots, \underline{u}(N-1)\}$ is the optimal control and $\{\underline{x}(0), \ldots, \underline{x}(N)\}$ the optimal trajectory. Then there exist functions $\{\underline{\psi}(1), \ldots, \underline{\psi}(N)\} \subseteq X^*$ and $\mathcal{A}$ not both zero such that

$$\underline{\psi}(k+1) = \underline{\psi}(k) - \underline{\psi}(k+1) \circ \left(\frac{\partial f_k}{\partial x(k)}\right) - \mathcal{A} \frac{\partial g_N}{\partial x(k)} \qquad 0 < k < N$$

The transversality conditions are given by

$$\mathcal{A} \frac{\partial g_N}{\partial x(o)} + \underline{\psi}(1) + \underline{\psi}_1 \circ \left(\frac{\partial f_o}{\partial x(o)}\right) \in LP(A(0), \underline{x}(0))$$

and $\mathcal{A} \dfrac{\partial g_N}{\partial x(N)} - \underline{\psi}(N) \in LP(A(N), \underline{x}(N))$

Moreover, if $H(u)$ is defined as in (6) we must have

$$H'(\underline{u}) \in LP(\Omega(0) \otimes \ldots \otimes \Omega(N-1);\, U)$$

where u and $\underline{u}$ are defined in (8).

Remark The maximum principle for discrete optimal control as obtained by Jordan [6] is a special case of Theorem 5.1. There the x, u are finite dimensional vectors and the profit function depends only on one coordinate of the final state $x(N)$. If we substitute these additional restrictions, we obtain his result.

It should be pointed out that whereas we obtain Theorem 5.1 as an application of Theorem 4.2, Jordan uses a direct argument (which is essentially a translation of the argument in Ref. 5) to arrive at his results. His proof, therefore, has a very great intuitive appeal, unlike ours.

Our main concern for generalizing the state space to an arbitrary B-space is to treat the case of stochastic control where the x, u are random variables. The case in which the random variables can take more than a finite number of values cannot be treated by the result obtained by Jordan. With very slight modification, however, we can use our result for this problem. The next section deals with this case and it will be illustrated by a simple example.

B. The Case of Discrete Stochastic Optimal Control

Consider a system of stochastic difference equations

$$x(k+1) = x(k) + f_k(x(k), u(k)) \qquad k = 0, 1, \ldots, N-1 \tag{1}$$

where $x(k)$ is an n-dimensional random variable representing the state at time k; $u(k)$ is an r-dimensional random variable representing the control at time k. The sample space of these random variables is the probability triple (Ω, A, P) where Ω is the sample space, A is a specified σ-algebra of subsets of Ω and P is the probability measure on A.

We shall assume that the random variables $x(k)$ belong to some Banach space X of random variables over (Ω, A, P). For example, X may be the Hilbert space of all n-dimensional random variables which have finite second moments. Similarly, we shall assume that $u(k)$, for each k, belongs to some B-space U of random variables. For each k, the function $f_k : X \otimes U \longrightarrow X$ is assumed to be differentiable. We are given some constraints on the u(k)'s and on the initial and final states x(0) and x(N), and we are required to maximize some differentiable, real-valued function of the x(k)'s and u(k)'s.

It is clear that under this formalism this problem is a special case of part A. Therefore, instead of repeating the same arguments, we shall consider a simple example and work it out in some detail.

Example (Ref. 7) :

We are given a linear, stochastic difference equation

$$x(k+1) = ax(k) + u(k) + v(k) \qquad k = 0, \ldots, N-1 \tag{1}$$

where x(k), u(k), v(k) are scalar-valued random variables representing the state, control and noise at time k, and a is a fixed known constant. Let

$$y(k) = x(k) + w(k) \qquad k = 0, \ldots, N-1 \tag{2}$$

be the observation at time k of the state x(k) corrupted by the noise w(k). Let $\alpha(0)$ be a random variable representing the a priori knowledge about the initial state x(0). Let (Ω, A, P) be the sample space of all these random variables. We shall make the following additional assumptions:

1) The random variables $\alpha(0)$, $v(0), \ldots, v(N-1)$, $w(0), \ldots, w(N-1)$ are independent and all of these except possibly $\alpha(0)$ have zero mean.

2) All the random variables that we shall encounter are square integrable, i.e., they belong to $L_2(\Omega, A, P) = B$ say. Note that $B^* = B$.

Let U(k), for $k = 0, \ldots, N-1$, be the space of all functions $u(y(0), \ldots, y(k))$, of $y(0), \ldots, y(k)$, such that $E\, u^2 < \infty$. It is clear that U(k) is a subspace of B and that $U(0) \subseteq U(1) \subseteq \ldots \subseteq U(N-1) \subseteq B$. We are required to find $u(k) \in U(k)$, i.e., $u(k) = u(y(0), \ldots, y(k))$ so as to minimize $1/2E \sum_{k=1}^{N} x(k)^2$.

Formally we wish to

$$\text{Maximize } \left\{-1/2 \sum_{k=1}^{N} Ex(k)^2 \;\middle|\; \begin{array}{l} ax(k) - x(k+1) + u(k) + v(k) = 0; \\ \alpha(0) - x(0) = 0 \qquad k = 0, \dots, N-1 \\ u(k) \in U(k);\ k = 0, \dots, N-1 \end{array}\right. \tag{3}$$

We now apply Theorem 4.1. We first form the Lagrangian function

$$\Phi = -\frac{\mu}{2} \sum_{k=1}^{N} E\, x(k)^2 + \sum_{k=0}^{N-1} \langle \psi(k+1),\ ax(k) - x(k+1) + u(k) + v(k) \rangle$$

$$+ \langle \psi(0),\ \alpha(0) - x(0) \rangle .$$

where $\mu \geq 0$ and $\psi(k) \in B^* = B$ and $\langle g, h \rangle \equiv E(gh)$. Let $\{\underline{u}(0), \dots, \underline{u}(N-1)\}$ be the optimal control functions and $\{\underline{x}(0), \dots, \underline{x}(N)\}$ the corresponding state sequence. Then, by Theorem 4.1, $\exists \underline{\mu} \geq 0$, $\exists \{\underline{\psi}(0), \dots, \underline{\psi}(N) \subseteq B$ such that

$$\frac{\partial \Phi}{\partial x(k)} = 0 \qquad k = 0, \dots, N$$

and

$$\frac{\partial \Phi}{\partial u(k)} \in LP(U(k), u(k)) \qquad k = 0, \dots, N-1$$

where the derivations are evaluated at $\mu = \underline{\mu}$, $\psi(k) = \underline{\psi}(k)$, $x(k) = \underline{x}(k)$, $u(k) = \underline{u}(k)$.

Expanding the first of these relations we get

$$a\,\underline{\psi}(1) - \underline{\psi}(0) = 0 \tag{4}$$

$$a\,\underline{\psi}(k+1) - \underline{\psi}(k) = \underline{\mu}\,\underline{x}(k) \qquad k = 1, \dots, N-1 \tag{5}$$

and

$$-\underline{\psi}(N) = \underline{\mu}\,\underline{x}(N) \tag{6}$$

while the second relation gives us

$$\underline{\psi}(k+1) \in LP(U(k), \underline{u}(k)) \qquad k = 0, \dots, N-1 \tag{7}$$

Now, $U(k)$ is a subspace of B, and $\underline{u}(k) \in U(k)$ so that $U(k) - \underline{u}(k) = U(k)$.

Therefore, $\underline{\psi}(k+1) \in LP(U(k), \underline{u}(k))$

iff $\langle \underline{\psi}(k+1), u(k) \rangle \leq 0 \quad \forall\, u(k) \in U(k)$

iff $\langle \underline{\psi}(k+1), u(k) \rangle = 0 \quad \forall\, u(k) \in U(k)$

iff $\langle \psi(k+1), u(y(0), \ldots, y(k)) \rangle \equiv E(\psi(k+1)\, u(y(0), \ldots, y(k)) = 0$

for all square integrable functions u of $(y(0), \ldots, y(k))$.

It is easy to see that this requirement is satisfied iff

$$E(\psi(k+1) \mid y(0), \ldots, y(k)) = 0 \qquad k = 0, \ldots, N-1 \tag{8}$$

Now equations (4) - (6) and (8) can be satisfied for $\mu > 0$.

$\therefore$ Taking $\mu = 1$ we have

$$a\,\underline{\psi}(1) - \underline{\psi}(0) = 0 \tag{9}$$

$$a\,\underline{\psi}(k+1) - \underline{\psi}(k) = \underline{x}(k) \qquad k = 1, \ldots, N-1 \tag{10}$$

$$-\underline{\psi}(N) = \underline{x}(N) \tag{11}$$

and
$$E(\underline{\psi}(k+1) \mid y(0), \ldots, y(k)) = 0 \qquad k = 0, \ldots, N-1 \tag{12}$$

From (11) we get

$$-\underline{\psi}(N) = a\underline{x}(N-1) + \underline{u}(N-1) + v(N-1)$$

Using (12) we have

$$0 = aE(\underline{x}(N-1) \mid y(0), \ldots, y(n-1)) + E(u(N-1) \mid y(0), \ldots, y(N-1)) + E(v(N-1) \mid y(0), \ldots, y(N-1))$$

$$\therefore\ 0 = aE(x(N-1) \mid y(0), \ldots, y(N-1)) + \underline{u}(y(0), \ldots, y(N-1)) + E\, v(N-1)$$

so that

$$\underline{u}(N-1) = u(y(0), \ldots, y(N-1) = -aE(\underline{x}(N-1) \mid y(0), \ldots, y(N-1)) \tag{13}$$

From (10) we get

$$a\underline{\psi}(k+1) - \underline{\psi}(k) = \underline{x}(k) = a\underline{x}(k-1) + \underline{u}(k-1) + v(k-1)$$

Taking conditional expectations with respect to $y(0), \ldots, y(k-1)$

and using the fact that

$E(\underline{\psi}(k) \mid y(0), \ldots, y(k-1)) = 0$ by (12) and

$$E(\underline{\psi}(k+1) \mid y(0), \ldots, y(k-1)) = E(E(\underline{\psi}(k+1) \mid y(0), \ldots, y(k)) \mid y(0), \ldots, y(k-1)) = 0$$

we have

$$0 = aE(\underline{x}(k-1) \mid y(0), \ldots, y(k-1)) + E(\underline{u}(k-1) \mid y(0), \ldots, y(k-1)) + E(v(k-1) \mid y(0), \ldots, y(k-1)).$$

$$\therefore\ \underline{u}(k-1) = \underline{u}(y(0), \ldots, y(k-1)) = -aE(\underline{x}(k-1) \mid y(0), \ldots, y(k-1)).$$

$$k = 1, \ldots, N-1 \qquad (14)$$

Combining (13) and (14) we have

$$\underline{u}(k-1) = -aE(\underline{x}(k) \mid y(0), \ldots, y(k-1)) \quad k = 0, \ldots, N-1 \qquad (15)$$

C. A Maximization Problem in Differential Equations.

The problem considered in this section and the methods employed for its solution are based to a very large extent on the papers by Gamkrelidze [13] and Neustadt [14].

Let $\mathcal{F}$ be a linear space whose elements $f(x,t)$ are n-dimensional vector-valued functions for x in R_n and t in $I = [t_0, t_1]$. We assume that the functions f in $\mathcal{F}$ satisfy the following conditions. 1. Each f is measurable in t over I for every fixed x, and is of class C^1 with respect to x in R_n. 2. For every f in $\mathcal{F}$, and compact set X in R_n, there exists a function $m(t)$, integrable over I and possibly dependent on f and X such that

$$\left| f(x,t) \right| \le m(t), \left| \frac{\partial f}{\partial x}(x,t) \right| \le m(t) \qquad x \text{ in } X, \ t \text{ in } I.$$

where the vertical bars denote the usual Euclidean norm in R_n.

Let P^r denote the set of all vectors $\alpha = (\alpha_1, \ldots, \alpha_r)$ where $\alpha_i \geq 0$ and $\sum_{i=1}^{r} \alpha_i = 1$. Let $F \subseteq \mathcal{F}$. Then the convex hull $[F]$ of F is given by

$$[F] = \left\{ \sum_{i=1}^{r} \alpha_{i} f_{i} \;\middle|\; f_i \in F,\ (\alpha_1, \ldots, \alpha_r) \in P^r,\ r > 0 \right\}$$

Def. The set $F \subseteq \mathcal{F}$ will be called <u>quasi-convex</u> if for every compact set

X in R_n, every finite collection $f_1, \ldots, f_r$ of elements in F, every $\epsilon > 0$, there are functions $f_\alpha \in F$, defined for every $\alpha \in P^r$ (and dependent on X, the f_i and ϵ), such that the functions $g(x, t; \alpha) = \sum_{i=1}^{r} \alpha_i f_i(x, t) - f_\alpha(x, t)$ satisfy the following conditions:

1. $\left| g(x, t; \alpha) \right| \leq \overline{m}(t)$, $\left| \frac{\partial g}{\partial x}(x, t; \alpha) \right| \leq \overline{m}(t)$
$\forall \cdot x \in X,\ t \in I$ and $\alpha \in P^r$

where $\overline{m}(t)$ is some function integrable over I and possibly dependent on X and the f_i (but not on ϵ);

2. $\left| \int_{\tau_1}^{\tau_2} g(x, t; \alpha)\, dt \right| < \epsilon$, $\tau_1 \in I$, $\tau_2 \in I$, $\forall x \in X$

3. for every sequence $\{\alpha^i\}$ with $\alpha^i \in P^r$, which converges to some $\overline{\alpha} \in P^r$, $g(x, t; \alpha^i)$ converges in measure (as a function of t on I) to $g(x, t; \overline{\alpha})$, for every $x \in X$.

Suppose we are given such a quasi-convex set F. Let f in F, and let x(t), t in I be any absolutely continuous solution of the differential equation

$$\dot{x}(t) = f(x(t), t) \qquad t \text{ in } I. \tag{1}$$

We shall regard such an x as an element of the Banach space B of continuous functions from the compact interval I into R_n. Now let A be the subset of B consisting of those elements x in B which are solutions of (1) for some f in F. Let h be a real-valued differentiable function of x in B and let $q: B \longrightarrow R_m$ be a differentiable mapping. We wish to solve the following problem.

$$\text{Maximize } \left\{ h(x) \,\middle|\, q(x) = 0, \ x \text{ in } A \right\} \tag{2}$$

Remark. The notion of quasi-convexity was first introduced by Gamkrelidze in [13] where he shows that it "encompasses almost all the extremal problems in solving the minimization of integral type functionals which arise in the classical calculas of variations and in the theory of optimal control . . ." Thus, for instance, suppose that we are given a fixed set Ω in E^r and let $\overline{\Omega}$ be the set of all measurable functions on some interval I which are essentially bounded. Then the set

$$F = \left\{ f(x,t) \;\middle|\; f(x,t) = h(x, u(t), t) \quad , \quad u(\cdot) \in \overline{\Omega} \right\} \text{- (a)}$$ is quasi-convex if h is of class C^1 with respect to x and measurable in (u, t) for every fixed x. (See [13].) The problem considered by Gamkrelidze can be phrased as follows. We are given a quasi-convex family F of functions defined on a bounded open interval I. Let A denote as before the set of functions $x(\cdot)$ which satisfy the differential equation

$$\dot{x}(t) = f(x(t), t), \qquad t \in I$$

for some element f in F. Then the problem is to find an element $\underline{x}(\cdot)$ in A, and a pair t_0, t_1 in I with $t_0 \leq t_1$ such that the $(2n + 2)$-ple $(x(t_0), x(t_1), t_0, t_1)$ is an extremal of the set $Q \cap N$ in E^{2n+2} where

$$Q = \left\{ (x(\tau_0), x(\tau_1), \tau_0, \tau_1) \;\middle|\; x(\cdot) \in A; \; \tau_0, \tau_1 \in I \text{ and } \tau_0 \leq \tau_1 \right\}$$

and N in some differentiably manifold of E^{2n+2} which represents the constraints on the initial and final values of the trajectory.

The problem that we consider is closely related to the one discussed by Neustadt in [14]. First of all, the initial and terminal moments t_0 and t_1 are fixed. He then supposes that the quasi-convex family F is given via a set of admissible controls as in (a) above. The function q in our eq (2) above may then be construed to represent a finite number of constraints on the entire trajectory (rather than just on the end-points as in [13]. The results presented by Neustadt are very similar to ours. Unfortunately we do not have a proof of his results so that we cannot compare our method with his.

We now return to the solution of the problem stated in (2). Let $\underline{x}$ be a solution of (2) and suppose that

$$\dot{\underline{x}}(t) = \underline{f}(\underline{x},(t),t), \ t \text{ in } I$$

for some $\underline{f} \in F$. We first obtain an estimate for the set $LC(A, \underline{x})$. Consider the linear variational equation of (2),

$$\Delta \dot{x}(t) = \frac{\partial f}{\partial x}(\underline{x}(t), \ t)\,\Delta x(t) + \Delta f(\underline{x}(t), \ t) \tag{3}$$

for t in $I = [t_0, t_1]$, Δf any arbitrary element of $[F] - \underline{f}$ and $x(t_0) = \xi$ any arbitrary vector in R_n. Let $\Phi(t)$ be a non-singular matrix solution of the homogeneous matrix differential equation

$$\dot{\Phi}(t) = \frac{\partial f}{\partial x}(\underline{x}(t), \ t)\,\Phi(t)$$

with $\Phi(t_0) = I$, the identity matrix. Then,

$$\Delta x(t) = \Phi(t)\left\{\xi + \int_{t_0}^{t} \Phi^{-1}(\tau)\,\Delta f(\underline{x}(\tau), \ \tau)\,d\tau\right\} \tag{4}$$

Let $K \subseteq B$ be the collection of all such solutions $\Delta x(t)$ of (3) for some $\xi \in R_n$ and some $\Delta f \in [F]-\underline{f}$. Clearly K is convex, $0 \in K$. Our first observation is the following Lemma.

Lemma: Since F is quasi-convex, $LC(A, \underline{x}) \supseteq K$.

Proof: Let $\Delta x(t) \in K$, $\xi \in R_n$, $\Delta f \in [F] - f$ such that $\Delta x(t_0) =$ and

$$\Delta \dot{x}(t) = \frac{\partial f}{\partial x}(\underline{x}(t), \ t)\ \ \Delta x(t) + \ \Delta f(\underline{x}(t), \ t \in I$$

Let $\epsilon > 0$. Since F is quasi-convex, there exists a function $g_\epsilon(x,t)$ in class C^1 with respect to x, and dependent on Δf and ϵ such that

$$\underline{f} + \epsilon\,\Delta\,f + g_\epsilon \qquad \in \quad F$$

$$|g_\epsilon(x,t)| < \overline{m}(t), \qquad \left|\frac{\partial g_\epsilon}{\partial x}(x,t)\right| < \overline{m}(t) \qquad , t \in I, \ x \in X$$

$$\left|\int_{\tau_1}^{\tau_2} g_\epsilon\,(x_\epsilon(\tau) \ , \ \tau)\,d\tau\right| < \epsilon^2$$

for $t_0 \leq \tau_1 \leq \tau_2 \leq t_1$, for every solution $x_\epsilon(t)$ of (5) below sufficiently near $\underline{x}(t)$ and for a compact set X in R_n which contains the trajectory $\underline{x}(t)$ in its interior. Here $x_\epsilon(t)$ is the solution of

$$\dot{x}_\epsilon = \underline{f}(x_\epsilon, t) + \epsilon \Delta f(x_\epsilon, t) + g_\epsilon(x_\epsilon, t) \tag{5}$$

and $\quad x_\epsilon(t_0) = \underline{x}(t_0) + \epsilon \xi$

It can be shown then that

$$x_\epsilon(t) = \underline{x}(t) + \epsilon \Delta x(t) + o(\epsilon) \qquad t \text{ in } I \tag{6}$$

where $\frac{o(\epsilon)}{\epsilon} \longrightarrow 0$ as $\epsilon \longrightarrow 0$ uniformly for t in I.

Clearly x_ϵ belongs to A. Also $x_\epsilon(t) \longrightarrow \underline{x}(t)$ as $\epsilon \longrightarrow 0$ and

$$\lim_{\epsilon \to 0} \left| \frac{x_t - \underline{x}}{\epsilon} \right| = \Delta x.$$

Hence $\Delta x \in LC(A, \underline{x})$.

Q.E.D.

Remark: It should be emphasized that the function $o(\epsilon)$ in (6) is not a continuous function of ϵ since the function g_ϵ is not chosen continuously.

We can now state the main result of this section. Let Q denote the derivative $q'(\underline{x})$ of q at the optimal $\underline{x}$.

Theorem 5.2: If $\underline{x}$ is a solution of (2), then there is a $\not{\mu} \geq 0$ and a vector $\underline{\lambda} \in R_m$ not both zero such that

$$\not{\mu}\, h'(\underline{x}) + \underline{\lambda} \circ Q \in LP(K, 0) \tag{7}$$

where the set K is defined as above.

Proof: Let C be the cone generated by K. C is convex because K is convex. We proceed as in Theorem 4.2.

Case 1. Suppose $Q(C) \neq R_m$. Then there is a $\underline{\lambda}$ in R_m, such that

$\langle \underline{\lambda}, Q(c) \rangle \leq 0$ for all c in C.

$\langle \underline{\lambda} \circ Q, c \rangle \leq 0$ for all c in C.

$\therefore$ $\underline{\lambda} \circ Q$ is in $LP(K, 0)$ and (7) is satisfied with $\not\mu = 0$, $\underline{\lambda} \neq 0$.

Case 2a. Suppose $Q(C) = R_m$. We know $LP(C) = LP(K)$. If $R_m \circ Q \cap LP(C) \neq \{0\}$, then there is a $\underline{\lambda} \in R_m$, $\underline{\lambda} \neq 0$ such that again (7) holds with $\not\mu = 0$, $\underline{\lambda} \neq 0$.

Case 2b. Suppose $Q(C) = R_m$, $R_m \circ Q \cap LP(C) = \{0\}$. Since $Q(C) = R_m$, it can be easily shown that $(q, A_q, \{0\})$ satisfies K.T. at $\underline{x}$. Here A_q is the set of all x in B such that $q(x) = 0$. Then, $LC(A_q, \underline{x}) = N$ where N is the null space of $Q \equiv q'(\underline{x})$, and $LP(A_q, \underline{x}) = R_m \circ Q$. We shall now prove,

$$LC(A_q \cap A, \underline{x}) \supseteq \overline{C} \cap N. \tag{8}$$

Since $Q(C) = R_m$ is finite-dimensional, it is easy to show that $\overline{C} \cap N = \overline{C \cap N}$. Using this fact and that C is generated by K to show (8) it suffices to prove (9).

$$LC(A_q \cap A, \underline{x}) \supseteq K \cap N \tag{9}$$

Let $\Delta x \in K \cap N$, i.e., $\Delta x \in K$ and $Q(\Delta x) = 0$. Using arguments which closely parallel those of Gamkrelidze [13], we can show, using condition 3 in the definition of quasi-convexity, that for sufficiently small ϵ, there is a vector x_ϵ in A such that $|x_\epsilon - \epsilon \Delta x| \leq o(\epsilon)$ and such that

$$q(x_\epsilon) = 0.$$

It follows then that $\Delta x \in LC(A \cap A_q, \underline{x})$.

Now by Theorem 3.1, since $\underline{x}$ solves (2), we must have,

$$h'(\underline{x}) \in LP(A_q \cap A, \underline{x})$$

$= LP(LC(A_q \cap A, \underline{x}))$

$\subseteq P(N \cap K)$ by (8) and (9)

Moreover, we know that $P(N) \cap P(K) = \{0\}$ and $P(N) = R_m \circ Q$ is a finite-dimensional subspace. Hence, $(P(N) + P(K))$ is closed.

$$\therefore \; h'(\underline{x}) \in P(N) + P(K)$$

$$= R_m \circ Q + P(K)$$

Therefore, there is a $\underline{\lambda}$ in R_m such that

$$h'(\underline{x}) + \underline{\lambda} \circ Q \in P(K)$$

and (7) is satisfied with $\not{A} = 1$.

Q.E.D.

REFERENCES

1. Kuhn H. W. and Tucker, A. W. "Nonlinear Programming," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1952, Vol. V, pp. 481-492.

2. Arrow, K., Hurwicz, L., and Uzawa, H. Studies in Linear and Non-Linear Programming. Stanford, California: Stanford University Press, 1958.

3. Arrow, K., Hurwicz, L., and Uzawa, H. "Constraint Qualifications in Maximization Problems," Naval Research Logistics Quarterly, 8 (1961), pp. 175-191.

4. Arrow, K. and Hurwicz, L. "Reduction of Constrained Maxima to Saddle-Point Problems," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1956.

5. Pontryagin, L. S. et al. The Mathematical Theory of Optimal Processes. Interscience Publishers, 1962.

6. Jordan, B. W. "Theory of a Class of Discrete Optimal Control Systems," Ph.D. dissertation, Dept. of Elec. Engg., University of California, Berkeley, 1964.

7. Sussman, R. "Optimal Control of Systems with Stochastic Disturbances," Electronics Research Laboratory Report No. 63-20, Berkeley, University of California.

8. Dantzig, G. B. and Wolfe, P. "Decomposition Principle for Linear Programs," Operations Research, Vol. 8 (1960), pp. 101-111.

9. Rosen, J. B. "Convex Partition Programming," Recent Advances in Mathematical Programming. Edited by R. Graves and P. Wolfe. McGraw Hill, 1963, pp. 159-176.

10. Lasdon, L. S. "A Multi-Level Technique for Optimization," Systems Research Centre, Case Institute of Technology, Cleveland, Ohio, SRC 50 - C - 64 - 19, April 1964.

11. Dunford, N. and Schwartz, J. T. Linear Operators Part I. New York: Interscience Publishers, Inc., 1964.

12. Dieudonné, J. Foundations of Modern Analysis. New York: Academic Press, 1960.

13. Gamkrelidze, R. V. "On some extremal problems in the theory of differential equations with applications to the theory of optimal control," J. SIAM, Ser. A: Control 3 (1965).

14. Neustadt, L. W. "Optimal Control Problems as extremal problems in a Banach space," Technical Report USCEE Report 133, Electronics Sciences Laboratory, University of Southern California, May 1965.

15. Varaiya, P. P. "Decomposition of Large-Scale Systems," presented at First International Conference on Programming and Control, Colorado Springs, 1965, to be published in J. SIAM, Ser. A: Control (1966).

16. Varaiya, P. P. "A decomposition technique for non-linear programming," I. B. M. Research Report RJ-345, July 1, 1965.

NONLINEAR PROGRAMMING AND OPTIMAL CONTROL

by

Pravin P. Varaiya

ERL Technical Memorandum M-129

23 September 1965

ELECTRONICS RESEARCH LABORATORY
University of California, Berkeley