c.1

NECESSARY CONDITIONS FOR A SOLUTION OF A NON-LINEAR PROGRAMMING PROBLEM

by

LINDA MAY LEE

B.Sc., University of British Columbia, 1969

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Mathematics

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September, 1973

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Mathematics

The University of British Columbia
Vancouver 8, Canada

Date Sept 1973

ABSTRACT

The conditions required for a solution of general non-linear programming problems of the form

$$\min\{f(x): \ x \in X \ , \ g(x) \leq 0 \ , \ h(x) = 0\} \ ;$$

where f is called the objective function, g the inequality constraint and h the equality constraint, are presented in this thesis. The following cases are studied:

(1) X, a finite dimensional space; f, a real valued function; and g and h finite dimensional vector functions.

(2) X, an infinite dimensional space; f, a real valued function; and g and h either finite or infinite dimensioanl vector functions.

An application of this type of problem to optimal control will be given and the recent developments in this area will be discussed.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I am indebted to Dr. Rodrigo Restrepo for suggesting the topic of the thesis and for his encouragement during the research of the topic. My special thanks goes to Dr. Ulrich Haussmann for allowing me a generous amount of his time and for his many constructive comments during the thesis preparation.

CHAPTER ONE: PRELIMINARY MATERIAL

1.0 Introduction

During the last decade, the problem of optimization has attracted a lot of attention, since such problems arise in many fields; for example, in automatic control theory, economics and even in biology. Although optimization problems are not new in mathematics, owing to the demands of economics and control theory and also owing to the appearance of the computer, an intensive and systematic investigation of such problems has only recently been started.

The first category of problems was studied as early as 1925. This field is called Calculus of Variations and deals with problems of maxima and minima where definite integrals involving one or more unknown functions are considered, subject to equality constraints. G. A. Bliss and O. Bolza did significant work with this type of problem.

The next type of problem to be classified were linear programming problems; problems where the objective function and constraints are all linear. The theory for this problem was widely developed by G. B. Dantzig and can be dated to 1948.

By 1951, the Kuhn-Tucker theory was developed. This gives the necessary conditions for an extremum in convex programming problems and, when in differential form, formulates the necessary conditions for non-convex programming problems in a finite dimensional space.

In the decade following, the theory of optimal control was developed. The basis of this theory is Pontryagin's Maximum Principle. This principle permitted the solution of various problems of mathematical and

applied nature and thus stimulated work in mathematical programming.

The embedding of optimal control theory into a general theory of necessary conditions was first carried out by A. A. Milyutin and A. Y. Dubovitski. They with H. Halkin and L. W. Neustadt have taken the present "modern" infinite dimensional approach.

This thesis presents the conditions required for a solution of general non-linear programming problems. The general problem is of the form

$$\min\{f(x): \ x \in X, \ g(x) \leq 0, \ h(x) = 0\} \ ;$$

f is called the objective function, g the inequality constraint and h the equality constraint. In Chapter Two, the following assumptions are made: the space X on which f, g, and h are defined is finite dimensional, f is a real-valued function and g and h are finite dimensional vector functions. In Chapter Three, X is assumed to be infinite dimensional and f a functional defined on X. From here the problem breaks down to two distinct problems depending on the dimension of the range of the constraint functions; that is, finite or infinite dimensional. Chapter Four gives a brief introduction to optimal control problems and to their solution using mathematical programming results . Also an example is given.
Finally, Chapter Five discusses the recent developments in non-linear programming problems. The rest of this first chapter deals with the tools required in the later chapters.

1.1 Differentiability Concepts.

Let X, Y be normed linear spaces and f be an operator defined

on a domain contained in X and range contained in Y.

1.1.1 Definition

The operator $f: X \to Y$ is Gateaux differentiable at $\bar{x}$ in X if there exists an operator $\delta f(\bar{x};e)$ which is linear[1] in e for all e in X and which satisfies

$$f(\bar{x} + \lambda e) - f(\bar{x}) = \lambda \delta f(\bar{x};e) + r(\bar{x},\lambda e) \quad \text{where}$$

$$\lim_{\lambda \to 0} \frac{\|r(\bar{x},\lambda e)\|}{\lambda} = 0 \quad \text{for all } e \text{ in } X .$$

1.1.2 Definition

The operator $f: X \to Y$ is Fréchet differentiable at $\bar{x}$ in X if there exists a continuous linear operator $f'(\bar{x}): X \to Y$ satisfying

$$f(x) - f(\bar{x}) = f'(\bar{x})(x - \bar{x}) + r(\bar{x};x - \bar{x}) \quad \text{for all } x \text{ in } X$$

where the function $r(x,z)$ is such that $\lim_{\|z\| \to 0} \frac{\|r(x,z)\|}{\|z\|} = 0$.

The Fréchet derivative of f is continuous at $\bar{x}$ if given $\varepsilon > 0$ there exists $\delta > 0$ such that $\|x - \bar{x}\| < \delta$ implies $\|f'(x) - f'(\bar{x})\| < \varepsilon$. If the derivative of f is continuous in some open sphere S, f is continuously Fréchet differentiable on S.

Remark:

If $f'(\bar{x})$ is a Fréchet differential then it is also a Gateaux differential i.e. $f'(\bar{x})e = \delta f(\bar{x},e)$ for all e in X. But the converse is not always true

---

1) In the usual definition of Gateau differentiability, the linearity is not assumed.

1.1.2.1 Some Elementary Properties of Fréchet Derivatives.

If follows from the definition that if f and g are Fréchet differentiable at $\bar{x}$ then $\alpha f + \beta g$ is Fréchet differentiable at $\bar{x}$ and

$$(\alpha f + \beta g)'(\bar{x}) = \alpha f'(\bar{x}) + \beta g'(\bar{x}) .$$

The chain rule and an inequality which replaces the Mean Value Theorem for the ordinary differentiable functions also hold for Fréchet differentiable function. The proofs will not be given here but may be found in Luenberger [10]. The inequality will be stated here as it will be referenced later.

Let f be Fréchet differentiable on an open set X_o in X. Let x be in X_o and suppose that $x + \alpha h$ is in X_o for all α, $0 \leq \alpha \leq 1$. Then

$$\|f(x + h) - f(x)\| \leq \|h\| \sup_{0 \leq \alpha \leq 1} \|f'(x + \alpha h)\| .$$

1.1.3 Definition.

A functional f defined on a normed linear space X is said to be quasi-differentiable at a point $\bar{x}$ if there exists a convex weak* closed set $F(\bar{x}) \subset X^*$ such that

$$\lim_{\lambda \to 0^+} \frac{f(\bar{x} + \lambda e) - f(\bar{x})}{\lambda} = \max_{f^* \in F(\bar{x})} f^*(e) \quad \text{for all } e \text{ in } X .$$

1.2 Inverse and Implicit Function Theorems

This section deals with the major results from analysis underlying the later theorems in optimization. The commonly known versions of the Inverse and Implicit Function Theorems for continuously differentiable functions in R^n will be stated. The reader is referred to Rudin [14] for proofs of these theorems. An Implicit Function Theorem generalized to functionals defined on linear spaces and an Inverse Function for Banach spaces will be presented. We follow the proofs in [10] and [13].

1.2.1 Inverse Function Theorem.

Let f be a continuously differentiable mapping of an open set E of R^n into R^n where $f'(\bar{x})$ is invertible for some $\bar{x}$ in E and where $\bar{y} = f(\bar{x})$. Then

(a) there exist open sets U and V in R^n where $\bar{x}$ is in U and $\bar{y}$ is in V and where f is one to one on U and $f(U) = V$;

(b) if g is the inverse of f (which exists by (a)) defined in V by $g(f(x)) = x$ for all x in U then g is a continuously differentiable function on V .

1.2.2 Implicit Function Theorem.

Let $(\bar{x},\bar{y})$ be a vector of an open set E contained in R^{n+m} and let f be a continuously differentiable n-dimensional vector function defined on E which satisfies the following conditions:

(1) $f(\bar{x},\bar{y}) = 0$,

(2) $\nabla_y f(\bar{x},\bar{y})$ is non-singular; that is, $\nabla_y f(\bar{x},\bar{y})K = 0$ implies $K = 0$.

Then there is a neighbourhood Z in R^n of $\bar{x}$, and an m-dimensional vector function e which is continuously differentiable on Z such that

(a) $\bar{y} = e(\bar{x})$,

(b) $f(x,e(x)) = 0$ for all x in Z.

1.2.3 Generalized Inverse Function Theorem.

Before the theorem can be stated, a definition is required.

1.2.3.1 Definition.

Let f be a continuously Fréchet differentiable mapping from an open set E in a Banach space X into a Banach space Y. If $\bar{x}$ in D is such that $f'(\bar{x})$ maps X onto Y, the point $\bar{x}$ is said to be a regular point of the transformation f.

1.2.3.2 Theorem:

Let $\bar{x}$ be a regular point of a continuously Fréchet differentiable transformation f mapping the Banach space X into a Banach space Y. Then there is a neighbourhood V of the point $\bar{y} = f(\bar{x})$ and a constant K such that the equation $f(x) = y$ has a solution for every y in V and the solution satisfies $\|x - \bar{x}\| \leq K\|y - \bar{y}\|$.

Proof:

Let L_o be the null space of $f'(\bar{x})$. Since L_o is closed, the quotient space X/L_o is a Banach space. If $[x]$ denotes the class of elements equivalent to x modulo L_o and if A is an operator on X/L_o defined by $A[x] = f'(\bar{x})x$ then A is well-defined since equivalent elements x yield identical elements y in Y. Also, by definition, A is linear,

continuous, one-to-one and onto and hence, by the Bounded Inverse Theorem [Appendix, Theorem 1], A has a continuous linear inverse.

Let y be an element of Y close to $\bar{y}$ and let $g_o = \theta$, the zero element in L_o. Now define the sequence of elements $\{L_n\}$ from X/L_o and a corresponding sequence $\{g_n\}$ with g_n, an element from L_n, recursively by

$$L_n - L_{n-1} = A^{-1}(y - f(\bar{x} + g_{n-1})) . \tag{1}$$

As $\|L_n - L_{n-1}\| = \inf_{g \in L_n} \|g - g_{n-1}\|$, select g_n from the coset L_n such that $\|g_n - g_{n-1}\| \leq 2\|L_n - L_{n-1}\|$.

Rewriting (1),

$$\begin{aligned} L_n &= A^{-1}(y - f(\bar{x} + g_{n-1})) + L_{n-1} , \\ &= A^{-1}(y - f(\bar{x} + g_{n-1}) + A[g_{n-1}]) \quad \text{by the definition of } L_{n-1} \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{and the properties of } A , \\ &= A^{-1}(y - f(\bar{x} + g_{n-1}) + f'(\bar{x})g_{n-1}) \quad \text{by the definition of } A \end{aligned}$$

and similarly, $L_{n-1} = A^{-1}(y - f(\bar{x} + g_{n-2}) + f'(\bar{x})g_{n-2})$. Thus

$$L_n - L_{n-1} = - A^{-1}(f(\bar{x} + g_{n-1}) - f(\bar{x} + g_{n-2}) - f'(\bar{x})(g_{n-1} - g_{n-2})) .$$

Define $g_t = tg_{n-1} + (1 - t)g_{n-2}$ and let $F(x) = - A^{-1}(f(x) - f'(\bar{x})x)$. Applying the generalized mean value inequality for Fréchet differentiable functions (Section 1.1.2.1), this implies

$$\|F(x + g_{n-1}) - F(x + g_{n-2})\| = \|g_{n-1} - g_{n-2}\| \sup_{0<t<1} \|F'(x + g_t)\| .$$

Hence

$$\|L_n - L_{n-1}\| \leq \|A^{-1}\| \, \|g_{n-1} - g_{n-2}\| \sup_{0<t<1} \|f'(\bar{x} + g_t) - f'(\bar{x})\| \ . \tag{2}$$

By the selection of g_1 and by the definition of L_1 ,

$$\|g_1\| = \|g_1 - g_o\| \leq 2\|L_1 - L_o\| = 2\|L_1\| = 2\|A^{-1}\| \, \|y - \bar{y}\| \ .$$

This implies for $\|y - \bar{y}\|$ small enough that $\|g_1\| < \frac{1}{2} r$ for some $r > 0$ and hence, in this particular case $\|g_t\| = \|tg_1 + (1 - t)g_o\| < r$ for $0 < t < 1$. By the continuity of f' at $\bar{x}$, for a given $\varepsilon > 0$ there exists an $r > 0$ such that $\|f'(x) - f'(\bar{x})\| < \varepsilon$ for $\|x - \bar{x}\| < r$. Therefore (2) becomes: $\|L_2 - L_1\| < \varepsilon\|A^{-1}\| \, \|g_1 - g_o\|$. By the selection of g_2 , the preceding inequality implies that $\|g_2 - g_1\| \leq 2\|L_2 - L_1\| \leq 2\varepsilon\|A^{-1}\| \, \|g_1 - g_o\|$. Hence for sufficiently small ε ,

$$\|g_2 - g_1\| \leq \frac{1}{2}\|g_1 - g_o\| \tag{3}$$

Similarly if $\|g_t\| = \|tg_{k-1} + (1 - t)g_{k-2}\| < r$, then

$$\|g_k - g_{k-1}\| \leq \frac{1}{2}\|g_{k-1} - g_{k-2}\| \ . \tag{4}$$

Moreover if (4) holds for all $k \leq n$, then

$$\begin{aligned} \|g_n\| &= \|g_1 + (g_2 - g_1) + \ldots + (g_n - g_{n-1})\| \\ &\leq (1 + \frac{1}{2} + \ldots + \frac{1}{2^{n-1}})\|g_1\| \\ &\leq 2\|g_1\| < r \ , \end{aligned}$$

so that $\|g_t\| = \|tg_n + (1 - t)g_{n-1}\| < r$. Thus by induction (4) holds for all k. Hence the sequence $\{g_n\}$ converges to an element g and correspondingly

the sequence $\{L_n\}$ converges to a coset L. Thus

$$f(\bar{x} + g) = y \quad \text{and} \quad \|g\| \leq 2\|g_1\| \leq 4\|A^{-1}\| \, \|y - \bar{y}\| .$$

Finally, by letting $K = 4\|A^{-1}\|$, the theorem is proved.

1.2.4 Generalized Implicit Function Theorem.

Let x be a m-dimensional vector, let $\lambda \in R$ and let $f_i(\lambda,x)$ for $i = 1, \ldots, m$ be continuous real valued functions which satisfy the following conditions:

(a) $f_i(0,0) = 0$ for $i = 1, \ldots, m$;

(b) $f_i(\lambda,x)$ is differentiable at $\lambda = 0$, $x = 0$;

(c) $\nabla_\lambda f_i(0,0) = 0$ for $i = 1, \ldots, m$

(d) $\nabla_x f(0,0)$ is non-singular

Then the system of equations $f_i(\lambda,x) = 0$ has a solution for sufficiently small λ and there exists a solution $x(\lambda)$ with the property that $\lim_{0} \frac{\|x(\lambda)\|}{\lambda} = 0$.

Proof:

Condition (b) is equivalent to

$$\|f(\lambda,x) - f(0,0) - \lambda\nabla_\lambda f(0,0) - x\nabla_x f(0,0)\| \leq \bar{r}(\sqrt{\lambda^2 + \|x\|^2}) \qquad (1)$$

where $\frac{\bar{r}(z)}{z} \to 0$ as $z \to 0$. Now applying conditions (a), (c), (d), the inequality (1) becomes

$$f(\lambda,x) = (\nabla_x f(0,0))(x) + r(\lambda,x) \qquad (2)$$

where $\|r(\lambda,x)\| \leq \bar{r}(\sqrt{\lambda^2 + \|x\|^2})$.

Consider the mapping $g(\lambda,x) = x - (\nabla_x f(0,0))^{-1} f(\lambda,x)$. Applying (2), this becomes $g(\lambda,x) = -(\nabla_x f(0,0))^{-1} r(\lambda,x)$ and $\|g(\lambda,x)\|$ $\leq \bar{r}(\sqrt{\lambda^2 + \|x\|^2})\|(\nabla_x f(0,0))^{-1}\|$. Without any loss of generality, the assumption that $\bar{r}(z)$ is a non-decreasing function can be made since if it were not, $\bar{r}(z)$ can be replaced by $w(z) = \sup_{0 \leq t \leq z} \bar{r}(t)$ where $w(z) \geq \bar{r}(z)$ and thus $\frac{w(z)}{z} \to 0$.

Now set $\tau(\lambda) = \inf\{\tau: K\bar{r}(\sqrt{\lambda^2 + \tau^2}) < \tau\}$ (3)

where $K = \|(\nabla_x f(0,0))^{-1}\|$. Since $K\bar{r}(\sqrt{\lambda^2 + \lambda^2}) < \lambda$ for sufficiently small λ , $\tau(\lambda) \leq \lambda$ for all such λ . By definition of infimum, for every such λ , there exists a $\tau^*(\lambda)$ such that $\tau(\lambda) \leq \tau^*(\lambda) \leq \tau(\lambda) + \lambda^2$ and $K\bar{r}(\sqrt{\lambda^2 + (\tau^*(\lambda))^2}) \leq \tau^*(\lambda)$. By definition of $\tau(\lambda)$, $\tau(\lambda) - \lambda^2 < K\bar{r}(\sqrt{\lambda^2 + (\tau(\lambda) - \lambda^2)^2})$. Since $\bar{r}(z)$ is a non-decreasing function and since $\sqrt{\lambda^2 + w^2} \leq \lambda + w$ for $\lambda > 0$, $w > 0$, this implies

$$K\bar{r}(\sqrt{\lambda^2 + (\tau(\lambda) - \lambda^2)^2}) \leq K\bar{r}(\lambda + |\tau(\lambda) - \lambda^2|) .$$ Thus,

$$\tau(\lambda) - \lambda^2 \leq K\bar{r}(\lambda + |\tau(\lambda) - \lambda^2|) .$$

Further since $\tau(\lambda) < \lambda$ for sufficiently small λ ,

$$K\bar{r}(\lambda + |\tau(\lambda) - \lambda^2|) \leq K\bar{r}(\lambda + \lambda) = K\bar{r}(2\lambda) .$$

Thus $\tau(\lambda) - \lambda^2 \leq K\bar{r}(2\lambda)$ or $\frac{\tau(\lambda)}{\lambda} \leq \frac{2K\bar{r}(2\lambda)}{2\lambda} + \lambda$ and since $\frac{r(z)}{z} \to 0$ as $z \to 0$, $\frac{\tau(\lambda)}{\lambda} \to 0$ as $\lambda \to 0$. Thus $\frac{\tau^*(\lambda)}{\lambda} \to 0$ as $\lambda \to 0$

If $\|x\| < \tau^*(\lambda)$ then

$$\|g(\lambda,x)\| \leq K\bar{r}(\sqrt{\lambda^2 + \|x\|^2}) \leq K\bar{r}(\sqrt{\lambda^2 + (\tau^*(\lambda))^2}) \leq \tau^*(\lambda) .$$

this implies that the continuous linear map $g(\lambda,x)$ maps the ball $\|x\| \leq \tau^*(x)$ into itself. Hence, by Brouwer's Fixed Point Theorem [Appendix, Theorem 2], $g(\lambda,x)$ has a fixed point; that is, there exists a point $x(\lambda)$ such that $x(\lambda) = g(\lambda,x(\lambda))$ and $\|x(\lambda)\| < \tau^*(\lambda)$. But the definition of $g(\lambda,x)$ implies $f(\lambda,x(\lambda)) = 0$; that is the set of non-linear equations under consideration has a solution. Also, since $\frac{\|x(\lambda)\|}{\lambda} \leq \frac{\tau^*(\lambda)}{\lambda}$, this implies $\frac{\|x(\lambda)\|}{\lambda} \to 0$ as $\lambda \to 0$.

CHAPTER TWO: OPTIMIZATION PROBLEMS IN FINITE DIMENSIONAL SPACES.

2.0 Introduction

This chapter presents the necessary and sufficients conditions for optimality when the objective function and constraints are finite dimensional. The mathematical programming problem to be studied here is:

$\mathcal{M}$: $\min f(x): x \in X_o, \; g(x) \leq 0, \; h(x) = 0$ where f is a functional, g is an m-dimensional vector function, h is a k-dimensional vector function, all defined on X_o, a subset of R^n.

2.1 Necessary Conditions.

In the necessary optimality conditions, the differentiability property of the functions play a crucial role since this is used to linearize the nonlinear programming problem.

Lemma 2.1.1

Let X_o be a convex set in R^n with a non-empty interior: int X_o, and let E be an open set in R^n. Let f be an ℓ-dimensional vector function and let h be a k-dimensional vector function, both defined on some open set containing X_o. Let $\bar{x}$ be an element from $X_o \cap E$, $f(\bar{x}) = 0$ and $h(\bar{x}) = 0$. Let f be differentiable at $\bar{x}$, let h be continuously differentiable in an open set containing $\bar{x}$ and let $\nabla h_j(\bar{x})$ for $j = 1, \ldots, k$ be linearly independent. If the equations $f(x) < 0$ and $h(x) = 0$ have no solution x in $X_o \cap E$ then the equations

$\nabla f(\bar{x})(x - \bar{x}) < 0$ and $\nabla h(\bar{x})(x - \bar{x}) = 0$ have no solution x in int X_o .

Proof:

Case $k > n$: This case is excluded because the assumption that $\nabla h_j(\bar{x})$, $j = 1, \ldots, k$, is linearly independent implies that $k \leq n$.

Case $k = n$: Since the linear independence of $\nabla h_j(\bar{x})$, $j = 1, \ldots, k$, is equivalent to the non-singularity of $\nabla h(\bar{x})$, $\nabla f(\bar{x})(x - \bar{x}) < 0$ cannot hold because $\nabla h(\bar{x})(x - \bar{x}) = 0$ implies that $(x - \bar{x}) = 0$. Thus neither the linear nor the nonlinear equations $\nabla f(\bar{x})(x - \bar{x}) < 0$ can be solved.

Case 3 $0 < k < n$:

The proof for this case follows from an indirect attack because the way the lemma is stated is the way it will be applied, not the way it is proved. Instead the proof shows that if the equations $\nabla f(\bar{x})(x - \bar{x}) < 0$ and $\nabla h(\bar{x})(x - \bar{x}) = 0$ have a solution x in int X_o then the equations $f(x) < 0$ and $h(x) = 0$ have a solution x in $X_o \cap E$.

Let $\hat{x}$ in int X_o be such that $\nabla f(\bar{x})(\hat{x} - \bar{x}) < 0$ and $\nabla h(\bar{x})(\hat{x} - \bar{x}) = 0$. For all x in R^n let (x_1, x_2) form a partition of x such that x_1 is in R^{n-k} and x_2 is in R^k . Then $\nabla h(\bar{x}) = (\nabla_{x_1} h(\bar{x})$, $\nabla_{x_2} h(\bar{x}))$. Since $\nabla h(\bar{x})$ is non-singular, this implies in particular that $\nabla_{x_2} h(\bar{x})$ is non-singular. Thus, since $h(\bar{x}) = 0$, and since $\nabla_{x_2} h(\bar{x})$ is non-singular and h is continuously differentiable in an open set containing $\bar{x}$, the Implicit Function Theorem [section 1.2.2] states that there exists an open set W in R^{n-k} containing $\bar{x}_1$ and a k-dimensional differentiable

vector function e on W such that

(a) $\bar{x}_2 = e(\bar{x}_1)$ and (b) $h(x_1, e(x_1)) = 0$ for all x_1 in W.

Since W is open and since $\bar{x}_1$ is in W, there exists $\delta_o > 0$ such that for all $\delta < \delta_o$, $(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1))$ is in W. Thus, since e is differentiable at $\bar{x}_1$ in W, this implies that

$$e(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1)) = e(\bar{x}_1) + \delta\nabla e(\bar{x}_1)(\hat{x} - \bar{x}_1) + o(\delta) \quad \text{for all} \quad \delta < \delta_o \; . \quad (1)$$

Since h is differentiable at $\bar{x}$ and since $h(x_1, e(x_1)) = 0$ for all x_1 in W, by chain rule, $\nabla_{x_1} h(\bar{x}) + \nabla_{x_2} h(\bar{x})\nabla e(\bar{x}_1) = 0$ and multiplying by $(\hat{x}_1 - \bar{x}_1)$ this becomes $\nabla_{x_1} h(\bar{x})(\hat{x}_1 - \bar{x}_1) + \nabla_{x_2} h(\bar{x})\nabla e(\bar{x}_1)(\hat{x}_1 - \bar{x}_1) = 0$. But the assumption $\nabla h(\bar{x})(\hat{x} - \bar{x}) = 0$ is equivalent to

$\nabla_{x_1} h(\bar{x})(\hat{x}_1 - \bar{x}_1) + \nabla_{x_2} h(\bar{x})(\hat{x}_2 - \bar{x}_2) = 0$. Thus $\nabla_{x_2} h(\bar{x})(\hat{x}_2 - \bar{x}_2)$ $= \nabla_{x_2} h(\bar{x})\nabla e(\bar{x}_1)(\hat{x}_1 - \bar{x}_1)$ and the non-singularity of $\nabla_{x_2} h(\bar{x})$ implies

$$(\hat{x}_2 - \bar{x}_2) = \nabla e(\bar{x}_1)(\hat{x}_1 - \bar{x}_1) \; . \quad (2)$$

Using (2) and the fact that $\bar{x}_2 = e(\bar{x}_1)$, equation (1) becomes

$$e(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1)) = \bar{x}_2 + \delta(\hat{x}_2 - \bar{x}_2) + o(\delta) \quad \text{for all} \quad \delta < \delta_o \; . \quad (3)$$

Because $\hat{x}$ is in $\text{int}\, X_o$ (by assumption) and by definition of $o(\delta)$, there exists $\delta_1 > 0$ such that for all $\delta < \delta_1$, $(\hat{x}_1, \hat{x}_2 + \frac{o(\delta)}{\delta})$ is in X_o. In particular δ_1 can be chosen such that $0 < \delta_1 << \delta_o$. Since $\bar{x}$ is in X_o and since X_o is convex, $(1 - \delta)(\bar{x}_1, \bar{x}_2) + \delta(\hat{x}_1, \hat{x}_2 + \frac{o(\delta)}{\delta})$ is in X_o for all $\delta < \delta_1$; that is, $(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1),\ \bar{x}_2 + \delta(\hat{x}_2 - \bar{x}_2) + o(\delta))$

is in X_o which by applying equation (3) is equivalent to $(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1), e(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1)))$ is in X_o for all $\delta < \delta_1$. Furthermore since $\bar{x}$ is in E and since E is open, there exists $\delta_2 > 0$ such that for all $\delta < \delta_2$, $(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1), e(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1)))$ is in E. Hence, letting $\delta_3 = \min\{\delta_1, \delta_2\}$,

$$(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1), e(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1))) \text{ is in } X_o \cap E \text{ for all } \delta < \delta_3 . \quad (4)$$

Since $h(x_1, e(x_1)) = 0$ for all x_1 in W, and since for all $\delta < \delta_o$ $(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1))$ is in W and since $\delta_3 \leq \delta_2 < \delta_o$, then for all $\delta < \delta_3$, $h(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1), e(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1))) = 0$.

Finally, since f is differentiable at $\bar{x}$, this implies the existence of δ_4 such that $\delta_4 < \delta_o$ and such that for all $\delta < \delta_4$

$$f(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1), e(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1)))$$

$$= f(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1), \bar{x}_2 + \delta(\hat{x}_2 - \bar{x}_2) + o(\delta)) , \text{ by equation (3)},$$

$$= f((\bar{x}_1, \bar{x}_2) + \delta(\hat{x}_1 - \bar{x}_1, \hat{x}_2 - \bar{x}_2 + \frac{o(\delta)}{\delta})) ,$$

$$= f(\bar{x}) + \delta[\nabla_{x_1} f(\bar{x})(\hat{x}_1 - \bar{x}_1) + \nabla_{x_2} f(\bar{x})(\hat{x}_2 - \bar{x}_2) + \nabla_{x_2} f(\bar{x}) \frac{o(\delta)}{\delta}] + o(\delta)$$

by the differentiablity of f at $\bar{x}$,

$$= \delta[\nabla f(\bar{x})(\hat{x} - \bar{x})] + \nabla_{x_2} f(\bar{x}) o(\delta) + o(\delta)$$

since $f(\bar{x}) = 0$.

But by assumption $\nabla f(\bar{x})(\hat{x} - \bar{x}) < 0$ which implies that there exists $\delta_5 > 0$ such that $\delta_5 < \delta_4$ and such that

$$f(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1), e(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1))) < 0 \text{ for all } \delta < \delta_5 .$$

Hence, letting $\hat{\delta} = \min\{\delta_3, \delta_5\}$, for all $\delta < \hat{\delta}$

$$f(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1), e(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1))) < 0$$

and $$h(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1) , e(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1))) = 0$$

where $(\bar{x}_1 + \delta(\hat{x}_1 - \bar{x}_1))$ is in $X_o \cap E$.

<u>Case 4 $k = 0$</u> :

Suppose there exists $\hat{x}$ in int X_o satisfying $\nabla f(\bar{x})(\hat{x} - \bar{x}) < 0$. Then it must be shown that there exists $\tilde{x}$ in X_o E satisfying $f(\tilde{x}) < 0$.

Since $\bar{x}$ is in $X_o \cap E$ (by hypothesis) and since X_o is convex and E is open, there exists $\delta_o > 0$ such that for all $\delta < \delta_o < 1$, $(\bar{x} + \delta(\hat{x} - \bar{x}))$ is in $X_o \cap E$. By the differentiability of f at $\bar{x}$, there exists $\delta_1 > 0$ such that $\delta_1 < \delta_o$ and such that for all $\delta < \delta_1$

$$f(\bar{x} + \delta(\hat{x} - \bar{x})) = f(\bar{x}) + \delta[\nabla f(\bar{x})(\hat{x} - \bar{x}) + \frac{o(\delta)}{\delta}] ,$$

$$= \delta[\nabla f(\bar{x})(\hat{x} - \bar{x}) + \frac{o(\delta)}{\delta}] \text{ since } f(\bar{x}) = 0 .$$

By the assumption $\nabla f(\bar{x})(\hat{x} - \bar{x}) < 0$ and by definition of $o(\delta)$, there exists $\delta_2 > 0$ such that $\delta_2 < \delta_1$ and such that for all $\delta < \delta_2$, $f(\bar{x} + \delta(\hat{x} - \bar{x})) < 0$.

<u>Lemma 2.1.2</u>

Let X_o, E, $\bar{x}$ and f be as given in Lemma 2.1.1. Let h be a k-dimensional vector function which is continuously differentiable in an open set containing $\bar{x}$. If $f(x) < 0$ and $h(x) = 0$ have no solution in $X_o \cap E$ then there exists $\bar{p}$ in R^ℓ, $\bar{q}$ in R^k with $\bar{p} \geq 0$, $(\bar{p}, \bar{q}) \neq 0$

satisfying

$$[\bar{p}\nabla f(\bar{x}) + \bar{q}\nabla h(\bar{x})](x - \bar{x}) \geq 0 \quad \text{for all } x \text{ in the closure of } X_o .$$

Proof:

Case 1: $\nabla h_j(\bar{x})$, $j = 1, \ldots, k$ linearly dependent.

This implies that there exists $\bar{q}$ in R^k where $\bar{q} \neq 0$ such that $\bar{q}\nabla h(\bar{x}) = 0$. Therefore for $\bar{p} = 0$, $[\bar{p}\nabla f(\bar{x}) + \bar{q}\nabla h(\bar{x})](x - \bar{x}) = 0$ for all x in the closure of X_o.

Case 2: $\nabla h_j(\bar{x})$, $j = 1, \ldots, k$, linearly independent.

Since X_o is convex, then int X_o is convex. Let $F(x) = \nabla f(\bar{x})(x - \bar{x})$ and $H(x) = \nabla h(\bar{x})(x - \bar{x})$. Define for each x in int X_o, the set $S(x) = \{(y,z): y \in R^\ell, z \in R^k, y > F(x), z = H(x)\}$ and let $S = \bigcup_{x \text{ int } X_o} S(x)$.

Observe that S is convex since if (y_1,z_1) and (y_2,z_2) are in S, then for $0 < \lambda < 1$,

$$\begin{aligned}(1 - \lambda)y_1 + \lambda y_2 &> (1 - \lambda)F(x_1) + \lambda F(x_2) , \\ &= (1 - \lambda)\nabla f(\bar{x})(x_1 - \bar{x}) + \lambda\nabla f(\bar{x})(x_2 - \bar{x}) , \\ &= \nabla f(\bar{x})[(1 - \lambda)(x_1 - \bar{x}) + \lambda(x_2 - \bar{x})] , \\ &= F((1 - \lambda)x_1 + \lambda x_2) ;\end{aligned}$$

and

$$\begin{aligned}(1 - \lambda)z_1 + \lambda z_1 &= (1 - \lambda)H(x_1) + \lambda H(x_2) , \\ &= (1 - \lambda)\nabla h(\bar{x})(x_1 - \bar{x}) + \lambda\nabla h(\bar{x})(x_2 - \bar{x}) ,\end{aligned}$$

$$= \nabla h(\bar{x})[(1-\lambda)(x_1 - \bar{x}) + \lambda(x_2 - \bar{x})] ,$$

$$= H[(1-\lambda)x_1 + \lambda x_2] .$$

Since $f(x) < 0$ and $h(x) = 0$ have no solution x in $X_o \cap E$, by Lemma 2.1.1, $F(x) < 0$ and $H(x) = 0$ have no solution x in $\text{int } X_o$. This implies that $(0,0)$ is not in S which is in $R^\ell \times R^k$. Apply the separation theorem [Appendix, Theorem 3] for the convex sets S and $\{(0,0)\}$. Then there exists $p \in R^\ell$ and $q \in R^k$ where $(p,q) \neq 0$ such that for (u,v) in S, $pu + qv \geq 0$. Note that $p \geq 0$ since each u_i can be made as large as desired.

Let $\varepsilon > 0$, $u = F(x) + e\varepsilon$ where e is the all ones vector and $v = H(x)$ where x is in $\text{int } X_o$. Then obviously (u,v) is in $S(x)$ and hence in S. Thus $pu + qv = pF(x) + pe\varepsilon + pH(x) \geq 0$ or equivalently,

$$pF(x) + qH(x) \geq -\varepsilon pe \quad \text{for all } x \text{ in } \text{int } X_o .$$

But since $\varepsilon > 0$ is chosen arbitrarily, one must have

$$pF(x) + qH(x) \geq 0 \quad \text{for all } x \text{ in } \text{int } X_o .$$

Hence $\inf_{x \in \text{int } X_o} (pF(x) + qH(x)) \geq 0$. (1)

Finally, since $[\bar{p}\nabla f(\bar{x}) - \bar{q}\nabla h(\bar{x})](x - \bar{x})$ is continuous in x and in fact linear, equation (1) holds for all x in the closure of X_o.

<u>Theorem 2.1.3</u>

Let X_o be a convex set in R^n with a non-empty interior: $\text{int } X_o$. Let $\bar{x}$ be a solution to the minimization problem M. Let f

and g be differentiable at $\bar{x}$ and h be continuously differentiable in an open set containing $\bar{x}$. Then there exists $\bar{r}_o$ in R, $\bar{r}$ in R^m and $\bar{s}$ in R^k such that

(a) $(\bar{r}_o \nabla f(\bar{x}) + \bar{r}\nabla g(\bar{x}) + \bar{s}\nabla h(\bar{x}))(x - \bar{x}) \geq 0$ for all x in the closure of X_o;

(b) $\bar{r}g(\bar{x}) = 0$;

(c) $(\bar{r}_o, \bar{r}) \geq 0$;

(d) $(\bar{r}_o, \bar{r}, \bar{s}) \neq 0$.

Proof:

Let $I = \{i: g_i(\bar{x}) = 0\}$ and $J = \{i: g_i(\bar{x}) < 0\}$ then $I \cup J = \{1, 2, \ldots, m\}$ and let m_I, m_J denote the number of elements in the set I and J respectively so that $m_I + m_J = m$. Since g is defined on some open set containing X_o and since g is continuous at $\bar{x}$, there exists $\delta_o > 0$ such that

$$E = \{x: g_J < 0, \|x - \bar{x}\| < \delta_o\} \text{ is an open set in } R^n .$$

If Lemma 2.1.2 can be applied, the theorem is proved. Let F be the mapping from R^n to $R \chi R^m$ where $F(x) = [f(x) - f(\bar{x}), g_I(x)]$. Note that $F(\bar{x}) = (f(\bar{x}) - f(\bar{x}), g_I(\bar{x})) = (0,0)$ and that F is differentiable at $\bar{x}$ since f and g are. Also, since $\bar{x}$ is in X_o and $\bar{x}$ is in E, $\bar{x}$ is in $X_o \cap E$. Now, since $\bar{x}$ is the solution to the minimization problem, then the equations $f(x) - f(\bar{x}) < 0$ and $h(x) = 0$ have no solution x in $X_o \cap E$, or equivalently $F(x) < 0$ and $H(x) = 0$ have no solution in $X_o \cap E$. Hence, by Lemma 2.1.2, there exists $(\bar{r}_o, \bar{r}_I)$ in $R \chi R^{m_I}$,

$\bar{s}$ in R^k with $(\bar{r}_o,\bar{r}_I) \geq 0$, $(\bar{r}_o,r_I,\bar{s}) \neq 0$ satisfying $((\bar{r}_o,\bar{r}_I)\nabla F(\bar{x}) + \bar{s}\nabla h(\bar{x}))(x - \bar{x}) \geq 0$ for all x in the closure of X_o, and, by definition of F, this implies that $(\bar{r}_o\nabla f(x) + \bar{r}_I\nabla g_I(\bar{x}) + \bar{s}\nabla h(\bar{x}))(x - \bar{x}) \geq 0$. By defining $\bar{r}_J = 0$ and $\bar{r} = (\bar{r}_I,\bar{r}_J)$ in R^m, $\bar{r}g(\bar{x}) = \bar{r}_I g_I(\bar{x}) + \bar{r}_J g_J(\bar{x}) = 0$ and $\bar{r}\nabla g(\bar{x}) = \bar{r}_I\nabla g_I(\bar{x}) + \bar{r}_J\nabla g_J(\bar{x}) = \bar{r}_I\nabla g_I(\bar{x})$. Hence the theorem is proved.

Remark:

If the convexity requirement on X_o is replaced by the requirement that X_o be open then a stronger necessary optimality condition than condition (a) in the previous theorem is obtained. This is known as the Fritz-John Stationary Point Necessary Optimality Theorem.

Theorem 2.1.4

Let X_o be an open set in R^n . Let $\bar{x}$ be a (global) solution of the minimization problem M or a local solution thereof; that is $f(\bar{x}) = \min\{f(x): \ x \in X_o \cap B_\delta(\bar{x}),\ g(x) \leq 0,\ h(x) = 0\}$ where $B_\delta(\bar{x})$ is an open ball around $\bar{x}$ with radius δ . Let f and g be differentiable at $\bar{x}$ and h be continuously differentiable in an open set containing $\bar{x}$. Then there exists $\bar{r}_o$ in R, $\bar{r}$ in R^m , $\bar{s}$ in R^k where $(\bar{r}_o,\bar{r},\bar{s}) \neq 0$ such that $\bar{r}_o\nabla f(\bar{x}) + \bar{r}\nabla g(\bar{x}) + \bar{s}\nabla h(\bar{x}) = 0$, $\bar{r}g(\bar{x}) = 0$ and $(\bar{r}_o,\bar{r}) \geq 0$.

Proof:

Let $\bar{x}$ be a global or local solution of the minimization problem. Then since X_o is open, there exists $B_\lambda(\bar{x})$, an open ball around $\bar{x}$ with radius λ such that $B_\lambda(\bar{x}) \subset B_\delta(\bar{x}) \subset X_o$ and such that

$$f(\bar{x}) = \min\{f(x): \ x \in B_\lambda(\bar{x}),\ g(x) \leq 0,\ h(x) = 0\} .$$

Since $B_\lambda(\bar{x})$ is convex and has a non-empty interior $(\bar{x} \in B_\lambda(\bar{x}))$, Theorem 2.1.3 can be applied thus giving $\bar{r}_o$ in R, $\bar{r}$ in R^m, $\bar{s}$ in R^k, $(\bar{r}_o, \bar{r}, \bar{s}) \neq 0$ where $(\bar{r}_o, \bar{r}) \geq 0$ such that

$$(\bar{r}_o \nabla f(\bar{x}) + \bar{r}\nabla g(\bar{x}) + \bar{s}\nabla h(\bar{x}))(x - \bar{x}) \geq 0 \quad \text{for all } x \text{ in } B_\lambda(\bar{x}) \qquad (1)$$

and $\bar{r}g(\bar{x}) = 0$.

Since for some $p > 0$, $(\bar{x} - p[\bar{r}_o \nabla f(\bar{x}) + \bar{r}\nabla g(\bar{x}) + \bar{s}\nabla h(\bar{x})])$ is in $B_\lambda(\bar{x})$, then, by inequality (1), this implies

$$\bar{r}_o \nabla f(\bar{x}) + \bar{r}\nabla g(\bar{x}) + \bar{s}\nabla h(\bar{x}) = 0$$

Remark:

In the Fritz-John necessary optimality criteria, there is no guarantee that $\bar{r}_o > 0$. If $\bar{r}_o = 0$ the necessary optimality criteria does not say much about the minimization problem since the function f, itself, disappears; thus any other function could play its role. It is possible to exclude such cases by introducing restrictions, known as constraint qualifications on the constraints $g(x) \leq 0$ and $h(x) = 0$. There are many constraint qualifications, some weaker than others but all giving the same result, namely $r_o > 0$. The one presented next gives a very elegant proof although its restrictions are not the weakest. Other constraint qualifications can be found in Mangasarian [11].

The modified Arrow-Hurwicz-Uzawa Constraint Qualification.

Let X_o be an open set in R^n, let g and h be m-dimensional and k-dimensional vector functions on X_o and let $X = \{x: x \in X_o, g(x) \leq 0, h(x) = 0\}$. The functions g and h are said to satisfy the modified

Arrow-Hurwicz-Uzawa constraint qualification at $\bar{x}$ in X if

(a) g is differentiable at $\bar{x}$

(b) h is continuously differentiable at $\bar{x}$

(c) $\nabla h_i(\bar{x})$ for $i = 1, \ldots, k$ are linearly independent

(d) there exists a solution z in R^n such that $\nabla g_I(\bar{x})z > 0$ and $\nabla h(\bar{x})z = 0$ where $I = \{i: g_i(\bar{x}) = 0\}$.

Mangasarian refers only to the restrictions on the constraints and not on the objective function. A different approach where the constraint qualifications involve both the objective function and the constraints has been developed by Geoffrion [6]. This approach significantly weakens the hypothesis demanded by Mangasarian but in this chapter Mangasarian's approach will be described.

The following theorem is known as the Kuhn-Tucker Stationary-point Necessary Optimality Theorem.

Theorem 2.1.5

Let X_o be an open set in R^n and let $\bar{x}$ be a global solution of the minimization problem M or a local solution thereof. Let f and g be differentiable at $\bar{x}$ and let h be continuously differentiable in an open set containing $\bar{x}$. Let g and h satisfy the modified Arrow-Hurwicz-Uzawa constraint qualification at $\bar{x}$. Then there exists $\bar{u}$ in R^m, $\bar{v}$ in R^k such that $\nabla f(\bar{x}) + \bar{u}\nabla g(\bar{x}) + \bar{v}\nabla h(\bar{x}) = 0$, $\bar{u} \geq 0$, and $ug(\bar{x}) = 0$.

Proof:

Since X_o, f, g, h and $\bar{x}$ satisfy the assumptions of Theorem 2.1.4, this implies that there exists $\bar{r}_o$ in R, $\bar{r}$ in R^m,

and $\bar{s}$ in R^k where $(\bar{r}_o,\bar{r},\bar{s}) \neq 0$, $(\bar{r}_o,\bar{r}) > 0$ such that $\bar{r}_o\nabla f(\bar{x}) + \bar{r}\nabla g(\bar{x}) + \bar{s}\nabla h(\bar{x}) = 0$ and $\bar{r}g(\bar{x}) = 0$. Note that if $\bar{r}_o > 0$ then by letting $\bar{u} = \bar{r}/\bar{r}_o$ and $\bar{v} = \bar{s}/\bar{r}_o$, the theorem is proved.

Suppose $r_o = 0$.

<u>Case 1: $\bar{r} = 0$</u> .

This implies that $\bar{s} \neq 0$ and $\bar{s}\nabla h(\bar{x}) = 0$. Since h satisfies the constraint qualification, $\nabla h_i(\bar{x})$ for $i = 1, \ldots, k$ are linearly independent therefore if $\bar{s}\nabla h(\bar{x}) = 0$ then $\bar{s} = 0$ which is a contradiction. Thus, if $\bar{r} = 0$ then $r_o > 0$.

<u>Case 2: $\bar{r} \neq 0$</u> .

This implies that $\bar{r} > 0$. Let $I = \{i: g_i(\bar{x}) = 0\}$ and $J = \{i: g_i(\bar{x}) < 0\}$. Then

$$\bar{r}g(\bar{x}) = \bar{r}_I g_I(\bar{x}) = 0 \tag{1}$$

and

$$\bar{r}\nabla g(\bar{x}) + \bar{s}\nabla h(\bar{x}) = \bar{r}_I\nabla g_I(\bar{x}) + \bar{r}_J\nabla g_J(\bar{x}) + \bar{s}\nabla h(\bar{x}) = 0 . \tag{2}$$

By definition of I and J , equation (1) implies that $\bar{r}_J = 0$ thus $\bar{r}_I > 0$. Substituting this into (2), this implies that

$$\bar{r}_I\nabla g_I(\bar{x}) + \bar{s}\nabla h(\bar{x}) = 0 . \tag{3}$$

Since f and g satisfy the modified Arrow-Hurwicz-Uzawa constraint qualification, then there exists z in R^n such that $\nabla g_I(\bar{x})z > 0$ and $\nabla h(\bar{x})z = 0$. Thus, since $\bar{r}_I > 0$ for all $\bar{s}$ in R^k , $\bar{r}_I\nabla g_I(\bar{x})z + \bar{s}\nabla h(\bar{x})z = (\bar{r}_I\nabla g_I(\bar{x}) + \bar{s}\nabla h(\bar{x}))z > 0$ contradicting equation (3).

Therefore, for $\bar{r} \neq 0$, $\bar{r}_o > 0$.

2.2 Sufficient Optimality Criteria.

Theorem 2.2.1

Let X_o be an open set in R^n and let f, g, h be, respectively, a numerical function, an m-dimensional vector function and a k-dimensional vector function, all defined on X_o . Let $\bar{x}$ be in X_o , let f and g be convex and differentiable at $\bar{x}$ and let h be a linear equality constraint. If there exists $\bar{u}$ in R^m and $\bar{v}$ in R^k such that $(\bar{x},\bar{u},\bar{v})$ satisfy the following conditions:

(a) $\nabla f(\bar{x}) + \bar{u}\nabla g(\bar{x}) + \bar{v}\nabla h(\bar{x}) = 0$,

(b) $\bar{u}g(\bar{x}) = 0$,

(c) $g(\bar{x}) \leq 0$,

(d) $h(\bar{x}) = 0$,

(e) $\bar{u} \geq 0$,

then $\bar{x}$ is a solution of the minimization problem M .

Proof:

Let $X = \{x: x \in X_o, g(x) \leq 0, h(x) = 0\}$. Since h is a linear equality constraint, $h(x) = Bx - d = 0$ for all x in X where B is a $k \chi n$ matrix and d is some constant vector in R^k . Then $Bx = d$ can be substituted for $h(x) = 0$ and $\nabla h(x) = B$.

Since f is convex and differentiable at $\bar{x}$,

$$f(x) - f(\bar{x}) \geq \nabla f(\bar{x})(x - \bar{x}) \quad \text{for all } x \text{ in } X . \tag{1}$$

Condition (a) implies that $\nabla f(\bar{x}) = -\bar{u}\nabla g(\bar{x}) - \bar{v}\nabla h(\bar{x})$ or equivalently,

$$\nabla f(\bar{x})(x - \bar{x}) = -\bar{u}\nabla g(\bar{x})(x - \bar{x}) - \bar{v}\nabla h(\bar{x})(x - \bar{x}) \ . \tag{2}$$

Since $h(\bar{x}) = 0$, $B\bar{x} = d$, and by the linearity of h, $\nabla h(\bar{x})(x - \bar{x}) = B(x - \bar{x}) = 0$ for all x in X. Also, since g is convex and differentiable at $\bar{x}$, $\nabla g(\bar{x})(x - \bar{x}) \leq g(x) - g(\bar{x})$. But because $\bar{u} \geq 0$, $g(x) \leq 0$ and $\bar{u}g(\bar{x}) = 0$, for all x in X

$$-\bar{u}\nabla g(\bar{x})(x - \bar{x}) \geq \bar{u}[-g(x) + g(\bar{x})] = -\bar{u}g(x) \geq 0 \ .$$

Therefore equation (2) becomes:

$$\nabla f(\bar{x})(x - \bar{x}) \geq -\bar{u}g(\bar{x}) \geq 0 \ .$$

Hence by inequality (1), $f(x) - f(\bar{x}) \geq 0$ for all x in X and since $g(\bar{x}) \leq 0$ and $h(\bar{x}) = 0$, $\bar{x}$ is in X. Thus $\bar{x}$ is the solution to M.

<u>Remark</u>:

The assumptions of this theorem, namely, the convexity of f and g and the linearity of the equality constraint, can be weakened somewhat since not all the properties of convex functions are needed. For example, f need only be pseudo-convex at $\bar{x}$; that is if f is differentiable at $\bar{x}$ in X and if x is in X where $\nabla f(\bar{x})(x - \bar{x}) \geq 0$, then $f(x) \geq f(\bar{x})$; g be quasi-convex at $\bar{x}$, that is for x in X where $g(x) \leq g(\bar{x})$ and where for, $0 < \lambda < 1$, $(1 - \lambda)\bar{x} + \lambda x$ is in X, then $g((1 - \lambda)\bar{x} + \lambda x) \leq g(\bar{x})$; and finally the equality constraint $h(x) = 0$ need not be linear so long as h is differentiable, quasi-convex and quasi-concave at $\bar{x}$. The proof is very similar to the previous proof and can be found in Mangasarian [11].

CHAPTER THREE: OPTIMIZATION PROBLEMS IN LINEAR SPACES

3.0 Introduction:

In this chapter, optimality criteria are developed for mathematical programming problems where the objective functional and the constraints are defined on linear spaces. The first section deals with the necessary and sufficient conditions for the problem with constraints which are mappings from a linear space into a normed space. The necessary criteria are approached in two ways: the global theory relying on convexity and the local theory using differentiability. Also, necessary conditions for the case where only equality constraints are present in the problem will be developed. Section 3.2 deals with the necessary criteria for a slightly modified problem; namely, where the constraints are actually functionals defined on a linear space. Finally, in Section 3.3, a comparison of the necessary conditions for the two problems is presented.

3.1 General Criteria for Optimization by Linear Space Methods.

This presentation follows Luenberger [10]

3.1.1 Global Necessary Conditions.

The basic problem to be considered in this section is:

L_g: $\min\{f(x): x \in X_o, g(x) \leq \theta\}$ where X_o is a convex subset of a linear space X, f is a convex functional on X_o and g is a convex mapping from X_o into a normed space Z which has a positive cone P.

Consider this problem in the space $R \chi Z$. If μ_o were the solution to this problem then, by convexity, there would exists a hyperplane which lies below $f(x)$ for all x in X_o where $g(x) \leq \theta$ and which goes through the point (μ_o,θ) . This separating hyperplane corresponds to the point $(1,z_o^*)$ in $R \chi Z^*$ where

$$\mu_o = \inf_{x \in X_o} \{f(x) + z_o^* g(x)\} .$$

<u>Theorem 3.1.1.1</u>

Let X be a linear space and X_o a convex subset of X . Let Z be a normed linear space with positive cone P having a non-empty interior. Let f be a real-valued convex functional on X_o and g a convex mapping from X to Z . Assume the existence of a point x_1 in X_o for which $g(x_1) < \theta$; that is , $g(x_1)$ is an interior point of $N = - P$. let $\mu_o = \inf\{f(x): x \in X_o, g(x) \leq \theta\}$ and assume μ_o is finite. Then there is an element z_o^* in Z^* , $z_o^* \geq \theta$ such that

$$\mu_o = \inf\{f(x) + z_o^* g(x): x \in X_o\} .$$

Furthermore, if the infimum is achieved by an $\bar{x}$ in X_o for which $g(\bar{x}) \leq \theta$ then $z_o^* g(\bar{x}) = 0$.

<u>Proof</u>:

In the space $W = R \chi Z$, define the following sets: let $A(x) = \{(r,z): r \geq f(x), z \geq g(x)\}$ for each x in X_o; let $A = \bigcup_{x \in X_o} A(x)$ and let $B = \{(r,z): r \leq \mu_o, z \leq \theta\}$. Obviously (μ_o,θ) is in B and,

since x_1 is in X_o, the point $(f(x_1), g(x_1))$ is in A.

Since f and g are convex, the sets A and B are convex and by definition of μ_o, A contains no interior points of B. Also, since $N = -P$ has an interior point, by the definition of B it contains an interior point. Thus by the separating hyperplane theorem, [Appendix, Theorem 4] there exists a non-zero element $w_o^* = (r_o, z_o^*)$ in W^* such that

$$r_o r_1 + z_o^* z_1 \geq r_o r_2 + z_o^* z_2 \quad \text{for all } (r_1, z_1) \text{ in } A \text{ and } (r_2, z_2) \text{ in } B \qquad (1)$$

Since $r_2 \leq \mu_o$ and $z_2 \leq \theta$ (by definition of B), equation (1) implies that $r_o \geq 0$ and $z_o^* \geq \theta$. But suppose $r_o = 0$ then since w_o^* is non-zero, $z_o^* > \theta$. Since (μ_o, θ) is in B, for $r_o = 0$ equation (1) implies that $z_o^* z_1 \geq 0$ for all (r_1, z_1) in A. In particular for $(f(x_1), g(x_1))$ in A, this implies that $z_o^* g(x_1) \geq 0$. But this contradicts the fact that $g(x_1) < 0$ and $z_o^* > 0$. Therefore $r_o > 0$ and without loss of generality, take $r_o = 1$.

Now applying $r_o = 1$ and the fact that (μ_o, θ) is an element of B and is arbitrarily close to A, equation 1 then imples that

$$\begin{aligned}
\mu_o &= \inf\{r + z_o^* z : \ (r,z) \in A\} ; \\
&= \inf\{f(x) + z_o^* g(x) : \ x \in X_o\} \text{ by defn. of } A ; \\
&\leq \inf\{f(x) : \ x \in X_o, \ g(x) \leq \theta\} \text{ since } z_o^* \geq \theta \text{ and} \\
&\qquad \text{considering only those } x \text{ in } X_o \text{ for} \\
&\qquad \text{which } g(x) \leq \theta ; \\
&= \mu_o \text{ by definition of } \mu_o .
\end{aligned}$$

Thus the first part of the theorem is proved.

Suppose there exists an $\bar{x}$ in X_o such that $g(\bar{x}) \leq \theta$ and $\mu_o = f(\bar{x})$. Since by above, $\mu_o \leq f(\bar{x}) + z_o^* g(\bar{x})$, this implies $\mu_o \leq f(\bar{x})$ because $z_o^* \geq \theta$, and $g(\bar{x}) \leq \theta$. Thus $z_o^* g(\bar{x}) = 0$.

Remarks:

(a) The theorem depends partially on the convexity of the set A . The way the theorem is stated, the convexity of A is guaranteed by the convexity of f and g . As in the finite dimensional study, this may be somewhat weakened as A may be convex without f and g being convex.

(b) The assumption the existence of an interior point of P, and the assumption that $g(x_1) < \theta$ for some x_1 in X_o guarantee the existence of a non-vertical separating hyperplane.

(c) Only convex constraints of the form $g(x) \leq \theta$ have been considered. Linear equality constraints, $h(x) = \theta$, although equivalent to convex inequalities $h(x) \leq \theta$ and $-h(x) \leq \theta$ cannot be treated in the same way as there never exists an x_1 which simultaneously renders $h(x_1) \leq \theta$ and $-h(x_1) \leq \theta$.

3.1.2 Local Necessary Conditions

The local theory of optimization parallels the theory for finite dimensions since generalizations of the concepts of differentials, gradients, and such to normed linear spaces are used. It also parallels the global case as the underlying principles are substantially the same: the separating hyperplane argument is again used.

The basic problem here is:

L_ℓ: $\min\{f(x): x \in X, g(x) \leq \theta\}$ where X is a normed linear space, f is a real-valued functional on X and g is a mapping from X into the normed space Z which has a positive cone P.

Again, as in the global case, an assumption must be included in the theorem to guarantee the existence of a non-vertical hyperplane. The analog to the interior point condition is the following definition of a regular point of an inequality.

Definition 3.1.2.1

Let X and Z be normed linear spaces. Let P be a positive cone in Z which has a non-empty interior. Let g be a mapping from X to Z which is Gateaux differentiable at $\bar{x}$ in X. The point $\bar{x}$ is said to be a regular point of the inequality $g(x) \leq \theta$ if $g(\bar{x}) \leq \theta$ and if there exists an e in X such that $g(\bar{x}) + \delta g(\bar{x};e) < \theta$.

Theorem 3.1.2.2

Let X be a normed linear space and Z be a normed linear space with a positive cone P having a non-empty interior. Let f be a Gateaux differentiable real-valued functional on X and g be a Gateaux differentiable mapping from X to Z. Suppose $\bar{x}$ is the solution to problem L_ℓ and $\bar{x}$ is a regular point of the inequaltiy $g(x) \leq \theta$, then there exists a z_o^* in Z^* such that

$$f'(\bar{x})e + z_o^* \delta g(\bar{x};e) = 0 \quad \text{for all } e \text{ in } X$$

and futhermore $z_o^* g(\bar{x}) = 0$.

Proof:

For each e in X define the set $A(e) = \{(r,z) \in R \times Z: \ r \geq f'(\bar{x})e, \ z \geq g(\bar{x}) + g'(\bar{x})e\}$ and let $A = \bigcup_{e \in X} A(e)$. Also define the set $B = \{(r,z): \ r \leq 0, \ z \leq \theta\}$. Obviously A and B are convex, $(0,\theta)$ is in $A(0)$, and B is a cone with vertex at $(0,\theta)$. Hence $(0,\theta)$ is in both A and B. Also, from the definition of B one may observe that B contains an interior point since $N = -P$ has an interior point.

In order to apply the Separating Hyperplane Theorem (Appendix, Theorem 4) it must be shown that A contains no interior points of B. Suppose A does. Then there exists an e in X such that $\delta f(\bar{x};e) < 0$ and $g(\bar{x}) + \delta g(\bar{x};e) < 0$. Consider the latter inequality. This implies that there exists an open sphere of some radius, say ρ, and center $g(\bar{x}) + \delta g(\bar{x};e)$ such that the sphere is contained in N. For $0 < \alpha < 1$, the point $\alpha(g(\bar{x}) + \delta g(\bar{x};e))$ is the center of an open sphere of radius $\alpha\rho$ contained in N. Thus the point

$$(1 - \alpha)g(\bar{x}) + \alpha(g(\bar{x}) + \delta g(\bar{x};e)) = g(\bar{x}) + \alpha\delta g(\bar{x};e) \quad \text{is in } N .$$

For fixed h since g is Gateaux differentiable, $|g(\bar{x} + \alpha h) - g(\bar{x}) - \alpha\delta g(\bar{x};e)| \leq o(\alpha)$, it follows that for sufficiently small α, $g(\bar{x} + \alpha e) < 0$. Similarly $\delta f(\bar{x};e) < 0$ implies that $f(\bar{x} + \alpha e) \leq f(\bar{x})$. This contradicts the optimality of $\bar{x}$. Hence A contains no interior points of B.

Thus, by the Separating Hyperplane Theorem, there is a closed

hyperplane H separating the sets A and B ; that is, there exists a non-zero element (r_o, z_o^*) in $R \chi Z^*$ such that

$$r_o r_2 + z_o^* z_2 \leq \xi \leq r_o r_1 + z_o^* z_1$$

for all (r_1, z_1) in A and (r_2, z_2) in B . Since $(0,\theta)$ is in both A and B , this implies that $\xi = 0$ and that

$$r_o r_1 + z_o^* z_1 \geq 0 \quad \text{for all} \quad (r_1, z_1) \quad \text{in} \quad A \tag{1}$$

and

$$r_o r_2 + z_o^* z_2 \leq 0 \quad \text{for all} \quad (r_2, z_2) \quad \text{in} \quad B \ . \tag{2}$$

In order for equation (2) to hold, the definition of B implies that $r_o \geq 0$ and $z_o^* \geq \theta$. Suppose $r_o = 0$. Then since (r_o, z_o^*) is a non-zero element in $R \chi Z^*$, $z_o^* > \theta$. Equation 1 implies that $z_o^* z_1 \geq 0$ for all (r_1, z_1) in A ; in particular, since by the definition of A , the point $(\delta f(\bar{x};e),\ g(\bar{x}) + \delta g(\bar{x};e))$ is in A for all e in X, this implies that $z_o^*(g(\bar{x}) + \delta g(\bar{x};e)) \geq 0$. But there exists an e in X such that $g(\bar{x}) + \delta g(\bar{x};e) < \theta$ since $\bar{x}$ is a regular point of the inequality $g(x) \leq \theta$. Thus a contradiction is reached since $z_o^* > 0$. Therefore $r_o > 0$ and without loss of generality, one can assume that $r_o = 1$.

Applying $r_o = 1$ and the fact that $(\delta f(\bar{x};e),\ g(\bar{x}) + \delta g(\bar{x};e))$ is in A , for all e in X, equation (1) becomes

$$\delta f(\bar{x};e) + z_o^*(g(\bar{x}) + \delta g(\bar{x};e)) \geq 0 \quad \text{for all} \quad e \quad \text{in} \quad X \ .$$

In particular, since θ is in X , this implies that $z_o^* g(\bar{x}) \geq 0$. But $z_o^* \geq \theta$ and $g(\bar{x}) \leq \theta$ imply that $z_o^* g(\bar{x}) \leq 0$. Thus $z_o^* g(\bar{x}) = 0$.

Finally, by the linearity of Gateaux differentials with respect to e ,

$$\delta f(\bar{x};e) + z_o^* \, \delta g(\bar{x};e) = 0 \ .$$

Remark

With the regularity condition on $\bar{x}$, this theorem cannot be extended to include equality constraints $h(\bar{x}) = \theta$ since there never exists an e in X such that $h(\bar{x}) + \delta h(\bar{x};e) < \theta$ and $-h(\bar{x}) - \delta h(\bar{x};e) < \theta$.

3.1.3 Necessary Conditions for Equality Constraints.

For problems with equality constraints only, a necessary optimality theorem by Luisternick as done in Luenberger [10] will be presented here. References will be made to the definition of a regular point of a transformation [Section 1.2.3] and to the Generalized Inverse Function Theorem [Section 1.2.4] .

The basic problem is

L_e: $\{\min f(x): \ x \in X, \ h(x) = \theta\}$ where X is a Banach space, f is a real-valued functional on X and h is a mapping from X into a Banach space Z .

Lemma 3.1.3.1

Let f achieve a local extremum subject to $h(x) = \theta$ at the point $\bar{x}$ and assume that f and h are continuously Fréchet differentiable in an open set containing $\bar{x}$ and that $\bar{x}$ is a regular point of the transformation h (see definition 1.2.3). Then $f'(\bar{x})e = 0$ for all e in X satisfying $h'(\bar{x})e = \theta$.

Proof:

Assume that the local extremum is a local minimum.

Suppose there exists an e in X such that $f'(\bar{x})e \neq 0$ and $h'(\bar{x})e = 0$. Define the mapping $T: X \to R \chi Z$ such that $T(x) = (f(x), h(x))$ then T is continuously Fréchet differentiable in an open set containing $\bar{x}$ and $T'(\bar{x}) = (f'(\bar{x}), h'(\bar{x}))$. Since $\bar{x}$ is a regular point of h, this implies that $h'(\bar{x})$ maps X onto Z; that is, for all z in Z there exists a y in X such that $h'(\bar{x})y = z$. By the linearity of Fréchet differentials, this implies that $h'(\bar{x})(y + \lambda e) = h'(\bar{x})y$ and $f'(\bar{x})(y + \lambda e) = f'(\bar{x})y + \lambda f'(\bar{x})e$ for all λ. For any t in R, λ can be chosen such that $f'(\bar{x})(y + \lambda e) = t$, and hence $T'(\bar{x})$ is an onto map from X to $R \chi Z$. Thus $\bar{x}$ is a regular point of the transformation T. By the Generalized Inverse Function Theorem [Section 1.2.4], for any $\varepsilon > 0$ there exists a vector x in X and $\delta > 0$ with $|x - \bar{x}| < \varepsilon$ such that $T(x) = (f(\bar{x}) - \delta, \theta)$, contradicting the assumption that $\bar{x}$ is a local minimum.

Theorem 3.1.3.2

Let f, h and $\bar{x}$ be as in the previous lemma. Then there exists an element z_o^* in Z^* such that $f'(\bar{x}) + z_o^* h'(\bar{x}) = \theta$.

Proof:

Lemma 3.1.3.1 implies that $f'(\bar{x})$ is orthogonal to the nullspace of the transformation $h'(\bar{x})$. By the definition of Fréchet differential, $h'(\bar{x})$ is a bounded linear operator. Since $h'(\bar{x})$ maps X onto Z, a Banach space, this implies that the range of the operator $h'(\bar{x})$ is closed. Thus,

by the property of bounded linear operators defined on Banach spaces (Appendix, Theorem 5), $f'(x)$ is an element in the range of $h'(\bar{x})^*$. This implies the existence of z^* in Z^* such that

$$f'(\bar{x}) = - h'(\bar{x})^* z^* ,$$

or an alternative notation

$$f'(\bar{x}) + z^* h'(\bar{x}) = \theta .$$

3.1.4 Sufficient Conditions

By the necessary conditions for optimality in problem L_g as seen in theorem 3.1.1.1, convexity and the existence of interior points guarantee the existence of a separating hyperplane. But these are too strong to impose for sufficiency since the appropriate hyperplane could exist in the absence of these conditions.

Theorem 3.1.4.1

Let f be a real-valued functional defined on a subset X_o of a linear space X. Let g be a mapping from X_o into the normed space Z having a non-empty positive closed cone P. Suppose there exists an element z_o^* in Z^*, $z_o^* \geq \theta$ and an element $\bar{x}$ in X_o such that

$$f(\bar{x}) + z^* g(\bar{x}) \leq f(\bar{x}) + z_o^* g(\bar{x}) \leq f(x) + z_o^* g(x)$$

for all x in X_o, $z^* \geq \theta$ in Z^*. Then $\bar{x}$ minimizes $f(x)$ subject to $g(x) \leq \theta$ with $\bar{x}$ in X_o.

Proof:

Since $f(\bar{x}) + z^* g(\bar{x}) \leq f(\bar{x}) + z_o^* g(\bar{x})$ for all $z^* \geq \theta$ in Z^*, this implies that $z^* g(\bar{x}) \leq z_o^* g(\bar{x})$. If $z_1^* \geq \theta$ then $(z_1^* + z_o^*) \geq \theta$ since $z_o^* \geq 0$ and thus $(z_1^* + z_o^*) g(\bar{x}) \leq z_o^* g(\bar{x})$ or equivalently $z_1^* g(\bar{x}) \leq 0$ for all $z_1^* \geq \theta$. Then, since $N = -P$ is a closed cone, this implies $g(\bar{x}) \leq \theta$ and thus, $z_o^* g(\bar{x}) \leq \theta$. Therefore, since $z^* g(\bar{x}) \leq z_o^* g(\bar{x})$ for all $z^* \geq \theta$ in Z^*, this implies that $z_o^* g(\bar{x}) = 0$.

Assume that x_1 is in X_o and that $g(x_1) \leq \theta$. Therefore, since $f(\bar{x}) + z_o^* g(\bar{x}) \leq f(x) + z_o^* g(x)$ for all x in X_o, this implies that $f(\bar{x}) + z_o^* g(\bar{x}) \leq f(x_1) + z_o^* g(x_1)$. Since $z_o^* g(\bar{x}) = 0$ by the previous part of this proof and also since $z_o^* \geq \theta$, $g(x_1) \leq \theta$, this implies that $f(\bar{x}) \leq f(x_1)$. Therefore $\bar{x}$ minimizes $f(x)$ subject to $g(x) \leq \theta$ and x in X_o.

3.2 Pshenichnyi's Approach

The necessary criteria for optimality derived in this section are for the following problem:

> P: $\min\{f(x): x \in X_o, g(x) \leq \theta, h(x) = \theta\}$ where X_o is some set in the linear space X and f, g_i for $i = 1, \ldots, m$ and h_j for $j = 1, \ldots, k$ are functionals defined on X.

This presentation follows Pshenichnyi [13]. The major results are:

1) Theorem 3.2.2 is the basic theorem of the section. It's assumptions on X, f, h are the weakest given for this type of problem.

The method of proof is very similar to the others in that a separating plane argument is used.

2) Theorem 3.2.4 restricts theorem 1 to the case where f and g are quasi-differentiable, h is Gateaux differentiable at $\bar{x}$ and X is a Banach Space.

3) Theorem 3.2.5 restricts theorem 1 to the case where f and g are convex and bounded, h is linear and X_o is a convex set in the Banach space X.

Pshenichnyi's first theorem is similar in statement to Mangasarian's Minimum Principle Necessary Optimality Theorem [Section 2.1.3] in the finite dimensional case. Recall that Mangasarian's required assumptions on X_o, f, g, h were:

1) X_o is a convex set in R^n with a non-empty interior.

2) f and g are differentiable at $\bar{x}$.

3) h is continuously differentiable in an open set containing $\bar{x}$.

Pshenichnyi's basic assumptions are:

1) X is a linear space, X_o is some set in X.

2) there exists a convex cone K such that if e is in K then for $\lambda >> 0$ sufficiently small, $\bar{x}(\lambda) = \bar{x} + e + \sum_{i=1}^{k} r_i(\lambda) e_i$ is in X_o where e_i for $i = 1, \ldots, k$ are any vectors in X and the functionals r_i satisfy $\lim_{\lambda \to 0} \frac{r_i(\lambda)}{\lambda} = 0$.

3) $\lim_{\lambda \to 0} \frac{f(\bar{x}(\lambda)) - f(\bar{x})}{\lambda} \leq F(e)$ and $\lim_{\lambda \to 0} \frac{g_i(\bar{x}(\lambda)) - g(\bar{x})}{\lambda} \leq G_i(e)$

for $i = 1, \ldots, m$ where F and G_i are convex functionals with respect to e.

4) $\lim_{\lambda \to 0} \frac{h_i(\bar{x}(\lambda)) - h_i(\bar{x})}{\lambda} = H_i(e)$ where H_i is a linear functional.

Observe that Pshenichnyi makes no convexity assumptions directly on X_o nor any differentiability assumptions directly on f, g and h. Thus his results are in terms of F, G, and H.

The lemma preceding theorem 3.1.2 proves that in the case where $H_1, \ldots, H_k$ are linearly independents, the separating plane argument can be applied in Theorem 3.2.2.

Lemma 3.2.1

Let $\bar{x}$ be the solution to the minimization problem P where X_o, X, f, g and h satisfy assumptions 1 through 4. Also, let $H_1, \ldots, H_k$ be linearly independent. Define $I = \{i: g_i(\bar{x}) = 0\}$, $J = \{i: g_i(\bar{x}) \not< 0\}$ and let m_I, m_J denote the number of elements in each set. Then the convex hull of the set

$$\tilde{K} = \bigcup_{e \in K} \{(r,s,t) \in R \chi R^{m_I} \chi R^k: \ r = F(e), \ s = G_I(e), \ t = H(e)\}$$

and the set

$$P = \{(r,s,t) \in R \chi R^{m_I} \chi R^k: \ r < 0, \ s < 0, \ t = 0\}$$

have an empty intersection.

Proof:

Suppose the intersection were not empty. Then it must be shown that the existence of an element in the intersection contradicts the

minimality of $f(\bar{x})$. Let $(\bar{r},\bar{s},\bar{t})$ be an element in $co\tilde{K} \cap P$. This implies that since $(\bar{r},\bar{s},\bar{t})$ is in $co\tilde{K}$ there exists (r^j,s^j,t^j) in $\tilde{K}$ such that

$$(\bar{r},\bar{s},\bar{t}) = \sum_j \lambda_j (r^j,s^j,t^j) \quad \text{where} \quad \lambda_j > 0 \quad \text{and} \quad \sum_j \lambda_j = 1 .$$

Furthermore, since (r^j,s^j,t^j) are in $\tilde{K}$ then for some e^j in K , $r^j = F(e^j)$, $s^j = G_I(e^j)$ and $t^j = H(e^j)$. Let $e^o = \sum_j \lambda_j e^j$ and observe that e^o is in K since K is convex and $\sum_j \lambda_j = 1$ where $\lambda_j > 0$. Therefore

$$\begin{aligned} F(e^o) &\leq \sum_j \lambda_j F(e^j) \quad \text{since } F \text{ is convex with respect to } e , \\ &= \sum_j \lambda_j r^j , \\ &= \bar{r} ; \end{aligned}$$

similarly $G_I(e^o) \leq \bar{s}$ and, since H is linear, $H(e^o) = \bar{t}$. Thus $F(e^o) < 0$, $G_I(e^o) < 0$ and $H(e^o) = 0$ since $(\bar{r},\bar{s},\bar{t})$ is also in P .

Consider the set of equations

$\psi_i(\lambda,r_1,r_2,\ldots,r_k) = h_i(\bar{x} + \lambda e^o + \sum_{j=1}^{k} r_j a^j) = 0$ for $i = 1, \ldots, k$ where a^j are in X and are chosen such that $H_i(a^j) = \delta_{ij}$; that is $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ij} = 1$ if $i = j$. Then by the Generalized Implicit Function Theorem [Section 1.2.5] the system of equations $\psi_i(\lambda,r_1,\ldots,r_k) = 0$ for $i = 1, \ldots, k$ has a solution $r_i(\lambda)$ for $i = 1, \ldots, k$ where $\lim_{\lambda\to 0} \frac{r_i(\lambda)}{\lambda} = 0$.

Consider the points $\bar{x}(\lambda) = \bar{x} + \lambda e^o + \sum_{j=1}^{k} r_j(\lambda) a^j$ where $\lambda > 0$.

then since e^o is in K, $\bar{x}(\lambda)$ is in X_o. Thus by definition of F

$$f(\bar{x}(\lambda)) \leq f(\bar{x}) + \lambda F(e^o) + o(\lambda) .$$

but $F(e^o) < 0$ implies that $f(\bar{x}(\lambda)) < f(\bar{x})$ for sufficiently small λ. Similarly by definition g_i, $g_i(\bar{x}(\lambda)) < g_i(\bar{x}) + \lambda_i G_i(e^o) + o(\lambda_i)$ for $i = 1, \ldots, m$, thus implying $g_I(\bar{x}(\lambda)) \leq g_I(\bar{x}) + \lambda_I G_I(e^o) + o(\lambda_I)$. But $g_I(\bar{x}) = 0$ and $G_I(e^o) < 0$. Thus for sufficiently small λ, $g_I(x(\lambda)) < 0$. Finally by construction of $\bar{x}(\lambda)$, $h(\bar{x}(\lambda)) = 0$ for sufficiently small λ. Therefore a contradiction to the assumption that $\bar{x}$ is the solution to P is obtained.

Theorem 3.2.2

Assume 1 through 4 again. If $\bar{x}$ is a solution to the minimization problem P then there exists $\bar{r}_o \in R$, $\bar{r} \in R^m$, $\bar{s} \in R^k$ where $(\bar{r}_o, \bar{r}, \bar{s}) \neq 0$ such that $\bar{r}_o F(e) + \bar{r}\, G(e) + \bar{s}H(e) \geq 0$ for all e in K where $(\bar{r}_o, \bar{r}) \geq 0$ and $\bar{r}g(\bar{x}) = 0$.

Proof:

Case 1: If $H_1, \ldots, H_k$ are linearly dependent then it is sufficient to set $(\bar{r}_o, \bar{r}) = 0$ and to choose $\bar{s}$ not equal to zero such that $\bar{s}H(e) = 0$.

Case 2: If $H_1, \ldots, H_k$ are linearly independent then by the previous lemma $co\tilde{K} \cap P = \phi$. Since $co\tilde{K}$ and P are finite dimensional convex sets, they can be separated; that is, there exists a vector $(\tilde{r}, \tilde{s}_I, \tilde{t})$ in $R \chi R^{m_I} \chi R^k$ such that $r_1\tilde{r} + s_1\tilde{s}_I + t_1\tilde{t} \geq 0 \geq r_2\tilde{r} + s_2\tilde{s}_I + t_2\tilde{t}$ for all (r_1, s_1, t_1) in $co\tilde{K}$, thus for all point s in $\tilde{K}$, and for all (r_2, s_2, t_2) in P. This implies that

$$\tilde{r}F(e) + \tilde{s}_I G_I(e) + \tilde{t}H(e) \geq 0 \quad \text{for all } e \text{ in } K$$

and also that $(\tilde{r},\tilde{s}_I) \geq 0$ since $r_2 < 0$ and $s_2 < 0$. By letting $\tilde{s}_J = 0$ and $\tilde{s} = (\tilde{s}_I,\tilde{s}_J)$, the theorem is proved since

$$\tilde{r}F(e) + \tilde{s}G(e) + \tilde{t}H(e) \geq 0 \quad \text{for all } e \text{ in } K \text{ with } (\tilde{r},\tilde{s}) \geq 0$$

and $\tilde{s}g(\bar{x}) = \tilde{s}_I g_I(\bar{x}) + \tilde{s}_J g_J(\bar{x}) = 0$ since $g_I(\bar{x}) = 0$ and $\tilde{s}_J = 0$.

Next, the particularization of Theorem 1 to the case where X is a Banach space and f and g_i for $i = 1, \ldots, m$ are quasi-differentiable will be presented. Recall that the class of all quasi-differentiable functions contains all Gateaux differentiable functionals and all convex functionals. It will be shown that if f and g_i also satisfy a Lipschitz condition then

$$\lim_{\lambda\to 0} \frac{f(\bar{x}(\lambda)) - f(\bar{x})}{\lambda} = \sup_{f^* \in F(\bar{x})} f^*(e)$$

and

$$\lim_{\lambda\to 0} \frac{g_i(\bar{x}(\lambda)) - g_i(\bar{x})}{\lambda} = \sup_{g_i^* \in G_i(\bar{x})} g_i^*(e)$$

i.e. the quasi-differentials. The results of Theorem 3.2.4 are comparable to Luenberger's local case (Theorem 3.1.2.2) which is in terms of Gateaux differentials.

<u>Lemma 3.2.3</u>

Let X be a Banach space. Assume the existence of K as in assumption 2. Let f, g_i for $i = 1, \ldots, m$ be functionals defined on X which are quasi-differentiable and let h_i be such that for

$i = 1, \ldots, k$, $\lim_{\lambda \to 0} \frac{h_i(\bar{x}(\lambda)) - h_i(\bar{x})}{\lambda} = h_i^*(e)$ for some h_i^* in X^* where $\bar{x}$ is the solution to the minimization problem P . If the sets $F(\bar{x})$ and $G_i(\bar{x})$ for $i = 1, \ldots, m$ are bounded then if

$$\bar{r}_o \sup_{f^* \in F(\bar{x})} f^*(e) + \sum_{i=1}^{m} \bar{r}_i \sup_{g_i^* \in G_i(\bar{x})} g_i^*(e) + \sum_{i=1}^{k} \bar{s}_i h_i^*(e) \geq 0$$

for all e in K , it is necessary and sufficient that there exists functionals $f^* \in F(\bar{x})$ and $g_i^* \in G_i(\bar{x})$ for $i = 1, \ldots, m$ such that

$$\bar{r}_o f^* + \sum_{i=1}^{m} \bar{r}_i g_i^* + \sum_{j=1}^{k} \bar{s}_j h_j^* \text{ is in } K^*$$

<u>Proof</u>:

The proof for sufficiency is obvious.

Let $N^* = \{x^* : x^* = \bar{r}_o f^* + \sum_{i=1}^{m} \bar{r}_i g_i^* + \sum_{j=1}^{k} \bar{s}_j h_j^*$ where $f^* \in F(\bar{x})$ and $g_i^* \in G_i(\bar{x})$ for $i = 1, \ldots, m\}$. Since $F(\bar{x})$ and $G_i(\bar{x})$ for $i = 1, \ldots, m$ are convex by definition, N^* is convex. Also since $F(\bar{x})$ and $G_i(\bar{x})$ for $i = 1, \ldots, m$ are, by definition, weak* closed and bounded, $F(\bar{x})$ and $G_i(\bar{x})$ for $i = 1, \ldots, m$ are weak* compact. These in turn imply that N^* is weak* closed and weak* compact.

Suppose there does not exist f^* in $F(\bar{x})$ and g_i^* in $G_i(\bar{x})$ for $i = 1, \ldots, m$ such that $r_o f^* + \sum_{i=1}^{m} r_i g_i^* + \sum_{j=1}^{k} s_j h_j^*$ is in K^* . This says that K^* and N^* have an empty intersection, or equivalently that $K^* - N^*$ does not contain the zero functional Since N^* is weak* closed and compact and since K^* is weak* closed, this implies [Appendix; Theorem 6] that $K^* - N^*$ is weak* closed. Also since N^* and K^* are

convex, $K^* - N^*$ is also convex. Thus $K^* - N^*$ is regularly convex [Appendix Theorem 7]; that is, for every functional x_o^* not in $K^* - N^*$ there exists e in X such that

$$z^*(e) < x_o^*(e) - \varepsilon \quad \text{for all } z^* \text{ in } K^* - N^* , \quad \varepsilon > 0 .$$

Since the zero functional is not in $K^* - N^*$, this implies that $z^*(e) < -\varepsilon < 0$. Thus $y^*(e) - x^*(e) > \varepsilon > 0$ for all y^* in K^* and x^* in N^*. This implies that $y^*(e)$ is bounded for all y^* in K^* but since K^* is a cone, $\inf_{y^* \in K^*} y^*(e) = 0$. Thus e is in $\bar{K}$ by properties of conjugate cones and $x^*(e) < -\varepsilon$ for all x^* in N^*.

Since $F(\bar{x})$ and $G_i(\bar{x})$ are bounded, there exists e_o in K where $|e - e_o|$ is so small that

$$|r_o| \; |x_o^*(e - e_o)| + \sum_{i=1}^{m} |r_i| \; |x_o^*(e - e_o)| + \sum_{j=1}^{k} |s_j| \; |h_j^*(e - e_o)| < \varepsilon/4$$

for all x_o^* in $F'(\bar{x})$ and x_i^* in $G_i(\bar{x})$ for $i = 1, \ldots, m$. Then by supremum property

$$0 \leq r_o \sup_{f^*} f^*(e_o) + \sum_{i=1}^{m} r_i \sup_{g_i^*} g_i^*(e_o) + \sum_{j=1}^{k} s_j \, h_i^*(e_o)$$

$$\leq r_o \, f^*(e_o) + \sum_{i=1}^{m} r_i \, g_i^*(e_o) + \sum_{j=1}^{k} s_j \, h_j^*(e_o) + \varepsilon/4$$

$$\leq r_o \, f^*(e) + \sum_{i=1}^{m} r_i \, g_i^*(e) + \sum_{j=1}^{k} s_j \, h_j^*(e) + \varepsilon/4 + \varepsilon/4$$

$$= x^*(e) + \varepsilon/2$$

Thus, $x^*(e) \geq -\varepsilon/2$ contradicting $x^*(e) < -\varepsilon$ for all x^* in N^* and so necessity is established.

Theorem 3.2.4

Let f, g_i for $i = 1, \ldots, m$ and h_j for $j = 1, \ldots, k$ be functionals on a Banach space X. Let X_o be some set in X. Assume the existence of K as in assumption 2. If f and g_i for $i = 1, \ldots, m$ satisfy a Lipschitz condition and are quasi-differentiable and if h_j for $j = 1, \ldots, k$ satisfy a Lipschitz condition and has a Gateaux differential h_j^* for $j = 1, \ldots, k$, then in order that $\bar{x}$ be a solution to the minimization problem P, it is necessary that there exists $\bar{r}_o$ in R, $\bar{r}$ in R^m and $\bar{s}$ in R^k where $(\bar{r}_o, \bar{r}, \bar{s}) \neq 0$ and functionals f^* in $F(\bar{x})$ and g_i^* in $G_i(\bar{x})$ for $i = 1, \ldots, m$ such that

$$\bar{r}_o f^* + \sum_{i=1}^{m} \bar{r}_i g_i^* + \sum_{i=1}^{k} \bar{s}_i h_i^* \in K^* \text{ with } (\bar{r}_o, \bar{r}) \geq 0$$

and futhermore $\bar{r}\, g(\bar{x}) = 0$.

Proof:

If it can be shown that the given assumptions on f, g_i for $i = 1, \ldots, m$ and h_j for $j = 1, \ldots, k$ imply that the conditions on f, g_i and h_j in theorem 3.2.2 are satisfied and that $F(\bar{x})$ and $G_i(\bar{x})$ for $i = 1, \ldots, m$ are bounded, then by Lemma 3.2.3, this theorem is proved.

Since f and g_i for $i = 1, \ldots, m$ satisfy the same conditions it is sufficient to consider just f. Since

$\frac{f(\bar{x}(\lambda)) - f(\bar{x})}{\lambda} = \frac{f(\bar{x}(\lambda)) - f(\bar{x} + \lambda e)}{\lambda} + \frac{f(\bar{x} + \lambda e) - f(\bar{x})}{\lambda}$, and since, by definition of $\bar{x}(\lambda)$, $\left|\frac{f(\bar{x}(\lambda)) - f(\bar{x} + \lambda e)}{\lambda}\right| \leq L \left|\frac{\sum_{i=1}^{k} r_i(\lambda) e_i}{\lambda}\right|$ because f satisfies a Lipschitz condition, then taking limits,

$\lim_{\lambda \to +0} \frac{f(\bar{x}(\lambda)) - f(\bar{x})}{\lambda} = \sup_{f^* \in F(\bar{x})} f^*(e)$ for all $x(\lambda)$ in X since $\lim_{\lambda \to +0} L \left|\frac{\sum_{i=1}^{k} r_i(\lambda) e_i}{\lambda}\right| = 0$ and by the quasi-differentiability of f , $\lim_{\lambda \to +0} \frac{f(\bar{x} + \lambda e) - f(\bar{x})}{\lambda} = \sup_{f^* \in F(\bar{x})} f^*(e)$. Thus by setting $F(e) = \sup_{f^* \in F(\bar{x})} f^*(e)$ and similarly $G_i(e) = \sup_{g_i^* \in G(\bar{x})} g_i^*(e)$ for $i = 1, \ldots, m$, this implies that F and G_i for $i = 1, \ldots, m$ are convex with respect to e and assumption 3 required of f and g_i for $i = 1, \ldots, m$ is more than met since equality has been established.

Finally, since h_j for $j = 1, \ldots, k$ has a Gateaux differential h_j^* at $\bar{x}$ for $j = 1, \ldots, k$ and since h_j for $j = 1, \ldots, k$ satisfy a Lipschitz condition, a similar argument as used for f and g_i for $i = 1, \ldots, m$ yields

$$\lim_{\lambda \to +0} \frac{h_j(\bar{x}(\lambda)) - h_j(\bar{x})}{\lambda} = h_j^*(e) \quad \text{for all } \bar{x}(\lambda) \text{ in } X .$$

Thus setting $H_j(e) = h_j^*(e)$, H_j, by definition of Gateaux differentials [Section 1.1.1] is the required linear functional for assumption 4 .

Suppose $F(\bar{x})$ is not bounded. This implies that for every $n > 0$ there exists e_n with $|e_n| = 1$ and a functional f_n^* in $F(\bar{x})$ such that $f_n^*(e_n) \geq n - \varepsilon$ for some $\varepsilon > 0$. Since f is a quasi-differentiable functional, this implies that

$$f(\bar{x} + \lambda e_n) - f(\bar{x}) = \lambda \sup_{f^* \in F(\bar{x})} f^*(e_n) + o(\lambda)$$

$$\geq \lambda f_n^*(e_n) + o(\lambda)$$

$$\geq \lambda (n - \varepsilon) + o(\lambda) .$$

Since f also satisfies a Lipschitz condition, $|f(\bar{x} + \lambda e_n) - f(\bar{x})| \leq L\lambda$, this now implies $L\lambda \geq \lambda(n - \varepsilon) + o(\lambda)$. Thus $L \geq n - \varepsilon$. But n can be chosen so large that $n - \varepsilon > L$ since L is fixed. Thus, a contradiction is reached. Finally, since the same argument holds for $G_i(\bar{x})$, $i = 1, \ldots, m$, the theorem is proved.

The final theorem presented in this section is a particularization of the problem P to the case where f and g_i for $i = 1, \ldots, m$ are convex bounded functionals, where h_j for $j = 1, \ldots, k$ are linear functionals, and X_o is a convex set in the Banach space X. This theorem is similar to Luenberger's global case [3.1.1].

<u>Theorem 3.2.5</u>

Let f and g_i for $i = 1, \ldots, m$ be convex bounded functionals on a Banach space X. Let h_j for $j = 1, \ldots, k$ be bounded linear functionals on X and let X_o be a convex set in X. If $\bar{x}$ is the solution to the minimization problem P, it is necessary that there

exists $\bar{r}_o$ in R, $\bar{r}$ in R^m and $\bar{s}$ in R^k with $(\bar{r}_o,\bar{r},\bar{s}) \neq 0$ such that

$$\bar{r}_o f(x) + \bar{r}g(x) + \bar{s}h(x) \geq \bar{r}_o f(\bar{x}) + \bar{r}g(\bar{x}) + \bar{s}h(\bar{x})$$

for all x in X_o where $(\bar{r}_o,\bar{r}) \geq 0$ and moreover $\bar{r}g(\bar{x}) = 0$.

<u>Proof</u>:

Let $K = \{e: \ e = \lambda(x - \bar{x}) \ \text{for } x \text{ in } X_o, \ x \neq \bar{x}, \ \lambda\lambda > 0\}$. Then K is a cone and if e is in K then $\bar{x}(\lambda) = \bar{x} + \lambda e$ is in X_o for small λ since X_o is convex. Since f is a convex functional

$\lim\limits_{\lambda \to +0} \dfrac{f(\bar{x} + \lambda e) - f(\bar{x})}{\lambda} = \dfrac{\partial f(\bar{x})}{\partial e} \leq f(\bar{x} + e) - f(\bar{x})$. Let $F(e) = f(\bar{x} + e) - f(\bar{x})$. Then F is a convex functional with respect to e. Similarly define $G_i(e)$ for $i = 1, \ldots, m$. Since h_i is a linear functional,

$\lim\limits_{\lambda \to +0} \dfrac{h_i(x + \lambda e) - h_i(\bar{x})}{\lambda} = \dfrac{\partial h_i(\bar{x})}{\partial e} = h_i(\bar{x} + e) - h_i(\bar{x}) = H_i(e)$. Thus by theorem 1, there exists $\bar{r}_o$ in R, $\bar{s}$ in R^k where $(\bar{r}_o,\bar{r},\bar{s}) \neq 0$ such that

$$\bar{r}_o F(e) + \bar{r}G(e) + \bar{s}H(e) \geq 0 \quad \text{for all } e \text{ in } K$$

where

$$(\bar{r}_o,\bar{r}) \geq 0 \quad \text{and} \quad \bar{r}_g(\bar{x}) = 0 \ ;$$

or equivalently $\bar{r}_o(f(\bar{x} + e) - f(\bar{x})) + \bar{r}(g(\bar{x} + e) - g(\bar{x})) + \bar{s}(h(\bar{x} + e) - h(\bar{x})) \geq 0$ and by setting $e = x - x_o$ for some x in X_o, this implies

$$\bar{r}_o f(x) - \bar{r}g(x) + \bar{s}h(x) \geq \bar{r}_o f(\bar{x}) + \bar{r}\, g(\bar{x}) + \bar{s}h(\bar{x}) \quad \text{for all } x \text{ in } X_o .$$

3.3 Comparison of Pshenichnyi's Approach to Luenberger's

In Luenberger's minimization problems there can be an infinite number of constraints since the inequality constraint g is a mapping into Z, a normed linear space of any dimension, finite or infinite and similarly the equality constraint h is a mapping into Z, a Banach space of any dimension. Thus Pshenichnyi's problems are actually a particularization of Luenberger's to a finite number of constraints. In this case Pshenichnyi's results are better than Luenberger's since Pshenichnyi's assumptions are weaker.

In this section, a comparison of assumptions and results of Luenberger's problem L restricted to the case where the number of constraints is finite, to Pshenichnyi's problem P will be presented. Since, in Luenberger's presentation, global and local cases with inequality constraints and the case with equality constraints only are all handled separately, the comparison with Pshenichnyi's assumptions will also be handled separately but first observe that the problems L_g, L_ℓ and L_e can be deduced from P by setting X_o to be the whole space X. The choosing of X_o from the linear space X is a way of further qualifying a mathematical programming problem. For an example of where this is used refer to Kushner's paper: Necessary Condition for Continuous Parameter Stochastic Optimal Problem [8].

3.3.1 Inequality Constraints - Local Case.

First it will be shown that Theorem 3.2.2 can be applied to derive Theorem 3.1.2.2. The set X_o and the cone K can both be defined

as X . The assumption of Gateaux differentiability of f and g implies that assumption (3) (Theorem 3.2.2) holds with F and G being the Gateaux differentials. Hence from theorem 3.2.2, there exists $\bar{r}_o \in R$, $\bar{r} \in R^m$ where $(\bar{r}_o,\bar{r}) \neq 0$ such that $\bar{r}_o F(e) + \bar{r}G(e) \geq 0$ for all e in X where $(\bar{r}_o,\bar{r}) \geq 0$ and such that $\bar{r}g(\bar{x}) = 0$.

Now, if the condition that $\bar{x}$ be a regular point of the inequality constraint is assumed in Theorem 3.2.2, this would imply the existence of e in X such that $g(\bar{x}) + \delta g(\bar{x};e) < \theta$ and $g(\bar{x}) < \theta$. But Theorem 3.2.2 says that $\bar{r}g(\bar{x}) = 0$, $(\bar{r}_o,\bar{r}) \geq 0$, $(\bar{r}_o,\bar{r}) \neq 0$ and $\bar{r}_o\delta f(\bar{x};e) + \bar{r}\delta g(\bar{x};e) \geq 0$ for all e in X . If $\bar{r}_o = 0$ this then implies

$$\bar{r}\delta g(\bar{x};e) \geq 0 \text{ for all } e \text{ in } X . \tag{1}$$

If e is as given in the regularity condition, then $\bar{r}(g(\bar{x}) + \delta g(\bar{x};e)) < 0$ since $\bar{r} > 0$. But $\bar{r}g(\bar{x}) = 0$ hence $\bar{r}\delta g(\bar{x};e) < 0$ contradicting (1) . Thus $\bar{r} \neq 0$ and, without loss of generality, r_o can be taken as 1 . Hence with the regular point assumption added, theorem 3.2.2 says $F(e) + \bar{r}G(e) \geq 0$ for all e in X . This, in turn, implies that $\delta f(\bar{x};e) + \bar{r}\delta g(\bar{x};e) \geq 0$ and so, by the linearity of the Gateaux differentials in e , $\delta f(\bar{x};e) + \bar{r}\delta g(\bar{x};e) = 0$, the result of theorem 3.1.2.2.

3.3.2 Inequality Constraints - Global Case

As stated earlier, Theorem 3.2.5 is very similar to Theorem 3.1.1.1. In fact, both require convexity of X_o , f and g . Thus the results of Theorem 3.2.5 follow. If the interior point assumption is added to Theorem 3.2.5, then the assumptions of 3.2.5 are equivalent to 3.1.1.1 Recall that the interior point condition implies the existence of x_1 in

X such that $g(x_1) < \theta$. With this condition added, the results of Theorem 3.2.5, namely $(\bar{r}_o,\bar{r}) \neq 0$, $(\bar{r}_o,\bar{r}) \geq 0$, $\bar{r}g(\bar{x}) = 0$ and $\bar{r}_o f(x) + \bar{r}g(x) \geq \bar{r}_o f(\bar{x}) + \bar{r}g(\bar{x})$ for all x in X , would imply $r_o > 0$ since if $r_o = 0$ then $\bar{r}g(x_1) < 0$. Hence, with $r_o > 0$, without loss of generality $r_o = 1$ and the results of Theorem 3.1.1.1 appear.

3.3.3 Equality Constraints

The assumptions of Theorem 3.2.2 will be derived from those of theorem 3.1.3.2. The set X_o and the cone K can both be defined as X . Since in 3.1.3.2, the equality constraint h and the objective functional f are continuously Fréchet differentiable, assumption 3 and assumption 4 in 3.2.2 follow immediately with F and H equal to the Fréchet differentials. Thus theorem 3.2.2 can be applied and the following results hold:

$$(\bar{r}_o,\bar{s}) = 0 \ , \quad \bar{r}_o G(e) + \bar{s}H(e) \geq 0 \quad \text{for all } e \text{ in } X \text{ and } \bar{r}_o \geq 0 \ .$$

Now, let the condition that $\bar{x}$ be a regular point of the transformation h be added to the assumptions of theorem 3.2.2. This implies that $h'(\bar{x})$ maps onto R^k . If $\bar{r}_o = 0$ then $\bar{s}h'(\bar{x})e \geq 0$ for all e in X , and thus, by the linearity in e of the Fréchet differentials, $\bar{s}h'(\bar{x})e = 0$. Since $\bar{s} \neq 0$, this implies that some of the components of $h'(\bar{x})e$ must be 0 for all e in X , hence the rank of $h'(\bar{x})$ is less than k but this contradicts the fact that $h'(\bar{x})$ maps onto R^k . Therefore, without loss of generality, $r_o = 1$ and $f'(\bar{x})e + \bar{s}h'(\bar{x})e \geq 0$ for all e in X so that by linearity in e of the Fréchet differentials, $f'(\bar{x}) + \bar{s}h(\bar{x}) = 0$.

CHAPTER FOUR: APPLICATION TO OPTIMAL CONTROL

4.0 Introduction

In an optimal control problem, the dynamics are described by a system of differential equations of the form

$$\frac{dx}{dt} = f(x(t), u(t)) \tag{1}$$

where $x(t)$ is an n-dimensional "state" vector, $u(t)$ is an n-dimensional control vector and f is a mapping of $R^n \chi R^m \to R^n$. This system when supplied with an initial state $x(t_o)$ and a control input function u, produces a vector-valued function x. Let the interval $[t_o, t_1]$ represent the interval on which x and u are defined. Also, in addition to this dynamic system, the classical optimal control problem has an objective functional of the form

$$J = \int_{t_0}^{t_1} \ell(x,u)dt$$

and a finite number of terminal constraints.

$$g_i(x(t_i)) = c_i \quad \text{for} \quad i = 1, \ldots r$$

or equivalently $$g(x(t_1)) = c \ . \tag{2}$$

Thus, an optimal control problem consists of finding the pair of functions (x,u) minimizing J while satisfying the system (1) and the terminal constraints (2).

By considering the problem as one formulated in $R^n \chi R^m$ and by treating the differential equation (1) and the terminal constraint (2) as

connecting u and x, the optimal control would reduce to a mathematical programming problem. Some assumptions must be made first. Let the vector-valued function f have continuous partial derivatives with respect to x and u. Let $u \in U \subset C^m[t_o,t_1]$ where U is the class of admissible control functions and $C^m[t_o,t_1]$ the set of continuous m-dimensional functions on $[t_o,t_1]$. Also assume for any given $u \in U$ and for an initial condition $x(t_o)$, equation (1) defines a unique continuous function $x(t)$ where $t > t_o$. If x is the function resulting from the application of a given control u, then x is said to be the trajectory of the system produced by u. Let X denote the class of all admissible trajectories. Finally assume that ℓ and g have continuous partial derivatives with respect to their arguments.

Now, let $X = C^n[t_o,t_1]$, $U = C^m[t_o,t_1]$ and define

$A(x,u) = x(t) - x(t_o) - \int_{t_o}^{t_1} f(x(s), u(s))ds = \Theta$. Observe that this is is simply the integrated form of the differential equation (1). Then A is a mapping from $C^n[t_o,t_1] \chi C^m[t_o,t_1]$ into $C^n[t_o,t_1]$. Thus the Fréchet differential of A exists, is continuous and is given by the formula

$$A'(x,u)(h,v) = h(t) - \int_{t_o}^{t_1} \nabla_x f(x,u)h(x)ds - \int_{t_o}^{t_1} \nabla_u f(x,u)v(s)ds . \qquad (3)$$

for all (h,v) in $X \chi U$. Also, since g has continuous partial derivatives, the terminal constraint g is a mapping from $C^n[t_o,t_1]$ into R^r with Fréchet differential:

$$g'(x)h = \nabla_x g(x)h(t_1) \quad \text{for all } h \text{ in } X . \qquad (4)$$

Since the transformation A and g define the constraints of the optimal control problem and since these constraints are actually ones of equality, the question of regularity of these constraints is equivalent to asking if, at the optimal trajectory, that is, at $(\bar{x},\bar{u})$, the Fréchet differentials $A'(\bar{x},\bar{u})$ and $g'(\bar{x})$, taken as a pair, map onto $X \chi R^r$ (see definition 3.1.2.1). To establish this, two assumptions are needed: (i) $\nabla_x g(x)$ has rank r and (ii) the system (3) is controllable, that is, for any e in R^n there exists $\tilde{v}$ in U such that

$$\hat{h}(t) - \int_{t_o}^{t_1} \nabla_x f(x,u)\tilde{h}(s)ds - \int_{t_o}^{t_1} \nabla_u f(x,u)\tilde{v}(s)ds = 0 , \tag{5}$$

and

$$\tilde{h}(t_1) = \tilde{e} \tag{6}$$

have a solution $\tilde{h}$ in X. Using (i), it is clear that the pair (A',g') is onto if for any e in R^n and any y in X there is an (h,v) in $X \chi U$ such that

$$h(t) - \int_{t_o}^{t_1} \nabla_x f(x,u)h(s)ds - \int_{t_o}^{t_1} \nabla_u f(x,u)v(s)ds = y(t) \tag{7}$$

and

$$h(t_1) = e \tag{8}$$

If $v = 0$ then by the fundamental existence theorem for linear Volterra integral equations [Appendix, Theorem 6], equation (7) has a solution $\bar{h}$. Now, let $\tilde{h}$ be the solution of (5) and (6) with $\tilde{e} = e - h(t_1)$. Then it is clear that $h = \tilde{h} + \bar{h}$ satisfies (7) and (8). Hence the constraints are regular.

4.1 Basic Necessary Conditions For Optimality.

The basic necessary conditions satisfied by the solution to the optimal control problem will be given here.

Theorem 4.1.1

Let $(\bar{x},\bar{u})$ minimize $J = \int_{t_o}^{t_1} \ell(x,u)dt$ subject to $\frac{dx}{du} = f(x,u)$, $x(t_o)$ fixed and $g(x(t_1)) = C$. Also, assume that the regularity conditions are satisfied at $(\bar{x},\bar{u})$. Then there is an n-dimensional vector-valued function $\lambda(t)$ and an r-dimensional vector μ such that for all t in $[t_o,t_1]$

$$\text{(a)} \quad -\frac{d\lambda}{dt} = [\nabla_x f(\bar{x}(t), \bar{u}(t))]^T\lambda(t) + [\nabla_x \ell(\bar{x}(t), \bar{u}(t))]^T \tag{9}$$

$$\text{(b)} \quad \lambda(t_1) = [\nabla_x g(x(t_1))]^T\mu \tag{10}$$

$$\text{(c)} \quad [\lambda(t)]^T\nabla_u f(\bar{x}(t), \bar{u}(t)) + \nabla_u \ell(x(t), \bar{u}(t)) = 0 . \tag{11}$$

Proof:

Since $\frac{dx}{dt} = f(x,u)$ can be rewritten in the form $A(x,u) = \Theta$ as seen in the introduction and since $g(x(t_1)) - C = \Theta$, the minimization problem to be considered has only equality constraints. Since A, g and J are continuously Fréchet differentiable since the regularity conditions are satisfied, and since $(\bar{x},\bar{u})$ is the solution to the minimization problem, by theorem 3.1.3.2, there exists $\lambda \in NBV[t_o,t_1]$, the normalized space of functions of bounded variation (see Luenberger [10], Dual of $C[t_o,t_1]$), and $\mu \in R^r$ such that

$$\int_{t_o}^{t_1} \nabla_x \ell(\bar{x},\bar{u})h(t)dt + \int_{t_o}^{t_1} d[\lambda(t)]^T[h(t) - \int_{t_o}^{t} \nabla_x f(\bar{x},\bar{u})h(s)ds]$$

$$+ \mu^T \nabla_x g(\bar{x}(t_1))h(t_1) = 0 \tag{12}$$

and

$$\int_{t_o}^{t_1} \nabla_u \ell(\bar{x},\bar{u})v(t)dt + \int_{t_o}^{t_1} d[\lambda(t)]^T \int_{t_o}^{t} \nabla_u f(\bar{x},\bar{u})v(s)ds = 0 \tag{13}$$

for all (h,v) in $X \chi U$. Without loss of generality, set $\lambda(t_1) = \theta$.

Equation (12) is equivalent to

$$\int_{t_o}^{t_1} \nabla_x \ell(\bar{x},\bar{u})h(t)dt + \int_{t_o}^{t_1} h(t)d\lambda(t) - \int_{t_o}^{t} d\lambda(t) \int_{t_o}^{t} \nabla_x f(\bar{x},\bar{u})h(s)ds$$

$$+ \mu^T \nabla_x g(\bar{x}(t_1))h(t_1) = 0$$

Then integrating by parts the third term of the above equation and substituting $\lambda(t_1) = \theta$, equation (12) can be rewritten as

$$\int_{t_o}^{t_1} \nabla_x \ell(\bar{x},\bar{u})h(t)dt + \int_{t_o}^{t_1} h(t)d\lambda(t) + \int_{t_o}^{t_1} [\lambda(t)]^T \nabla_x f(\bar{x},\bar{u})h(t)dt$$

$$+ \mu^T \nabla_x g(\bar{x}(t_1))h(t_1) = 0 \quad \text{for all } h \text{ in } X . \tag{14}$$

Thus, it is clear that λ can have no jumps in (t_o,t_1) since otherwise a suitable h in X can be constructed to make the second term of (13) large compared with the other terms. However there must be a jump at t_1 of magnitued $- \nabla g_x(\bar{x}(t_1))\mu$. Since (13) holds for all continuous h , it holds, in particular for all continuously differentiable h with $h(t_1) = h(t_o) = 0$. Therefore, integrating by parts the second term,

equation (12) becomes

$$\int_{t_o}^{t_1} \{ \nabla_x \ell(\bar{x},\bar{u})h(t) - [\lambda(t)]^T \frac{dh}{dt} + [\lambda(t)]^T \nabla_x f(\bar{x},\bar{u})h(t)\}dt = 0$$

or equivalently,

$$\int_{t_o}^{t_1} \{(\nabla_x \ell(\bar{x},\bar{u}) + [\ (t)]^T \nabla_x f(\bar{x},\bar{u}))h(t) - [\lambda(t)]^T \frac{dh}{dt}\}dt = 0 . \qquad (15)$$

Let $B(t) = \int_{t_o}^{t} \{\nabla_x \ell(\bar{x},\bar{u}) + [\lambda(s)]^T \nabla_x f(\bar{x},\bar{u})\}ds$. Integrating by parts,

$\int_{t_o}^{t_1} [\nabla_x \ell(\bar{x},\bar{u}) + [\lambda(t)]^T \nabla_x f(\bar{x},\bar{u})]h(t)dt = - \int_{t_o}^{t_1} B(t) \frac{dh}{dt} dt$ since

$h(t_o) = h(t_o) = h(t_1) = 0$. Thus equation (15) becomes

$$- \int_{t_o}^{t_1} \{B(t) + [\lambda(t)]^T\} \frac{dh}{dt} dt = 0 .$$

This implies that $- \lambda(t) = [B(t)]^T + c$ [Appendix: Theorem 8] for some constant c and hence by the definition of B , equation (9) holds.

Equation (13), after integrating the second term by parts, becomes

$$\int_{t_o}^{t_1} \nabla_u \ell(\bar{x},\bar{u})v(t)dt + \int_{t_o}^{t_1} [\lambda(t)]^T \nabla_u f(\bar{x},\bar{u})v(t)dt = 0$$

or equivalently,

$$\int_{t_o}^{t_1} [\nabla_u \ell(\bar{x},\bar{u}) + [\lambda(t)]^T \nabla_u f(\bar{x},\bar{u})]v(t)dt = 0 \quad \text{for all } v \text{ in } U . \qquad (16)$$

Let $D(t) = \nabla_u \ell(\bar{x},\bar{u}) + [\lambda(t)]^T \nabla_u f(\bar{x},\bar{u})$. Then equation (16) becomes

$$\int_{t_o}^{t_1} D(t)v(t)dt = 0 \qquad (17)$$

If $D(t) \neq 0$ at a point, say $D(t) > 0$, then by continuity, $D(t) > 0$ in a neighbourhood of that point. Let v be any continuous function which is zero outside the neighbourhood but which is greater than zero somewhere inside. Then $\int_{t_o}^{t_1} D(t)v(t)dt > 0$, contradicting (17). Thus $D(t)$ is equal to zero and equation (11) holds.

Finally, by changing the boundary condition on $\lambda(t_1)$ from $\lambda(t_1) = \theta$ to $\lambda(t_1) = [\nabla_x g(x(t_1))]^T \mu$ to account for the jump, λ will be continuous throughout $[t_o, t_1]$ and the theorem is proved.

4.2 An Example Problem in Optimal Control.

Consider the problem of finding the m-dimensional control function u that minimizes the quadratic objective functional

$$J = \tfrac{1}{2} \int_{t_o}^{t_1} ([x(t)]^T Q\, x(t) + [u(t)]^T R\, u(t))dt$$

subject to the linear dynamic constraint

$$\frac{dx}{dt} = F\, x(t) + B\, u(t), \quad x(t_o) \text{ fixed}, \tag{18}$$

where Q is an $n \chi n$ symmetric positive-semidefinite matrix, R is an $m \chi m$ symmetric positive-definite matrix, F is an $n \chi n$ matrix and B is an $n \chi m$ matrix.

Applying the necessary conditions of Theorem 4.1.1,

$$-\frac{d\lambda}{dt} = F^T \lambda(t) = Q\, x(t), \quad \lambda(t_1) = \theta \tag{19}$$

$$[\lambda(t)]^T B + [u(t)]^T R = 0 .$$

Since R is positive definite, this implies that $u(t) = - R^{-1} B^T \lambda(t)$ and thus the dynamic linear constraint (18) becomes

$$\frac{dx}{dt} = F\, x(t) - B\, R^{-1} B^T \lambda(t), \quad x(t_o) \text{ fixed} . \tag{20}$$

Observe that equations (19) and (20) form a linear system of differential equations whose solution satisfies the relation:

$\lambda(t) = P(t)x(t)$ where $P(t)$ is the $n \chi n$ matrix solution

of the Riccati differential equation

$$\frac{dP}{dt} + PF + F^T P - P\, B\, R^{-1}\, B^T P + Q = U, \quad P(t_1) = 0 .$$

CHAPTER FIVE: DEVELOPMENTS.

In the last few years, a lot of work has been done in the development of necessary optimality criteria for programming problems with both equality and inequality constraints. The general form of this problem is

$$G: \quad \min\{f(x): \ x \in X, \ g(x) \leq 0, \ h(x) = 0\} \quad \text{where}$$

$$f: \ X \to R, \ g: \ X \to Y \ \text{ and } \ h: \ X \to Z \ .$$

The following cases have been presented:

1) X, Y, Z finite dimensional

2) X, Y linear spaces and no equality constraints

3) X, Z Banach spaces and no inequality constraints

4) X, a linear space and Y, Z finite dimensional spaces.

Differentiability assumptions were required for optimality results in the local case of (2) as well as cases (1) and (3). In case (4), Pshenichnyi required the existence of convex approximations of f and g and a linear approximation of h as well as a convex approximation of X.

B. D. Craven [4] does an extension to case (3). In his problem f is not a functional but rather a mapping into a Banach space. His main result will be stated here for comparison to Theorem 3.1.3.2.

Craven's Main Result.

Let X, Y and Z be Banach spaces and let U be an open subset in X. Let $f: U \to Y$ be Fréchet differentiable and $h: U \to Z$ be continuously Fréchet differentiable. Assume (by restricting Y and Z) that $f(U)$ is dense in Y and $h(U)$ is dense in Z. Let $E = \{x: \ x \in U, \ h(x) = 0, \ h'(x) \text{ is an onto map}\}$. Then $f(x)$ is stationary

subject to the constraints $h(x) = 0$ at $x = \bar{x} \in E$ if and only if there exists a continuous linear map $M: Z \to Y$ such that $f'(\bar{x}) = Mh'(\bar{x})$.

H. Halkin and L. Neustadt [7] and L. Neustadt [12] present the case where the number of inequality constraints is infinite and the number of equality constraints is finite; that is, in problem $\mathcal{G}$, the spaces X, Y are infinite dimensional and Z is finite dimensional. The assumptions on f, g, h and X were similar to Pshenichnyi's; that is convex approximations were used and thus, the results were in terms of these convex approximations. M. Altman [1] developed the necessary criteria for the reverse problem; that is where the number of equality constraints is infinite and the number of inequality constraints is finite.

In Bazaraa and Goode's paper [3], the necessary conditions for the case where there are infinitely many equality and inequality constraints, are developed. These results are probably the most general so far available. Before the main results can be stated a few definitions are required:

(a) The problem to be considered is:

$\mathcal{B}$: $\min\{f(x): x \in S, g(x) \in c\ell\, C, h(x) = 0\}$ where $f: X \to R$, $g: X \to Y$ and $h: X \to Z$; X, Y, Z are normed linear spaces; S is a subset of X, $c\ell\, C$ is the closure of C where C is a convex cone in Y.

(b) The cone of interior directions for an arbitrary set S and $\bar{x}$ in $c\ell\, S$ is defined as:

$$D(S,\bar{x}) = \{x: \text{ there is a ball } B \text{ about the origin and } \delta > 0 \text{ such that } y \in x + B \text{ and } \lambda \in (0,\delta) \text{ imply that } \bar{x} + \lambda y \in S\}.$$

(c) The cone of tangents for an arbitrary set S and $\bar{x}$ in $c\ell\, S$ is defined as:

$$T(S,\bar{x}) = \{x: \text{ given any } \delta > 0, \text{ there exists a } \lambda \in (o,\delta) \text{ and } z \in B_\delta \text{ such that } \bar{x} + \lambda x + \lambda z \in S \text{ where } B_\delta \text{ is a ball around the origin with radius } \delta\} .$$

(d) The functions f and g are differentiable at $\bar{x}$ with derivatives F and G in the following sense:

$$1/\lambda\ (f(\bar{x} + \lambda z) - f(\bar{x})) \to F(y) \quad \text{for } \lambda \to 0^+,\ z \to y$$

and

$$1/\lambda\ (g(\bar{x} + \lambda z) - g(\bar{x})) \to G(y) \quad \text{for } \lambda \to 0^+,\ z \to y .$$

(e) G is $c\ell\, C$-convex if $G(\lambda x + (1 - \lambda)y) - \lambda G(x) - (1 - \lambda)G(y) \in c\ell\, C$ for each x, y and $\lambda \in (0,1)$.

(f) The level set L is defined as $L = \{x: h(x) = 0\}$.

(Note: if $\bar{x}$ is the solution of $\mathcal{B}$ then, obviously, $\bar{x} \in L$).

Remarks

(1) The cone of interior directions is similar to the cone K used by Pshenichnyi.

(2) If f and g are differentiable at $\bar{x}$ in the above sense then f and g are obviously Gateaux differentiable at $\bar{x}$ when $z = y$. But Gateaux differentiable at $\bar{x}$ need not imply the above definition since z may not converge to y.

Main Result:

Theorem:

Suppose that $\bar{x}$ solves problem $\mathcal{B}$ and suppose that $D(S,\bar{x})$ is convex. Let M be a nonempty convex subset of $T(L,\bar{x})$. Then, there exists a non-zero (u,v,w) in $R \chi Y^* \chi X^*$ such that

(i) $u \geq 0$, $v \in C^*$, $w \in M^*$

(ii) $vg(\bar{x}) = 0$

(iii) $(uF + vG + w)x \geq 0$ for all x in $D(S,\bar{x})$.

Extensions:

(1) If $\bar{x}$ is in the interior of S then $uF + vG + w = 0$.

(2) If S is convex and has a non-empty interior then condition (iii) is equivalent to $(uF + vG + w)(x) \geq (uF + vG + w)(\bar{x})$ for all x in S.

(3) If h is Fréchet differentiable at $\bar{x}$ with derivative H, it can be shown that $T(L,\bar{x}) \subset N(H)$. If $T(L,\bar{x}) = N(H)$ then the set M can be chosen such that $M = N(H)$ and thus $w \in (N(H))^{\perp}$. The assumption $N(H) \subset T(L,\bar{x})$ can be viewed as a regularity assumption. Here $N(H)$ is the null space of H. This condition is implied when either

(a) h is affine

or (b) X, Z are Banach spaces, h is continuously Fréchet differentiable, at $\bar{x}$, and the range of H is equal to Z. This is equivalent to Luenberger's regular point definition (Section 1.2.3)

(4) If either Z is finite dimensional or X and Z are Banach spaces and range of H is closed, then w can be written as a composition of H and a continuous linear functional on Y.

From this theorem and points (3) and (4) the following two theorems are immediate.

Theorem:

Let X, Y, Z be normed linear spaces, where Z is finite dimensional. Suppose that f and g are differentiable in the sense of definition (d) with derivatives F and G and let H be a continuous linear transformation from X to Z with $N(H) \subset T(L,\bar{x})$, eg. by letting h be affine. If $\bar{x}$ is the solution to $\mathcal{B}$ then there exists a non-zero $(u,v,w) \in R \chi Y^* \chi Z^*$ such that

(i) $u \geq 0$, $v \in C^*$

(ii) $vg(\bar{x}) = 0$

(iii) $(uF + vG + wH)(x) \geq 0$ for all x in $D(S,\bar{x})$.

Theorem:

Let X and Z be Banach spaces and Y be a normed linear space. Suppose that f and g are differentiable at $\bar{x}$ as in definition (d) with derivatives F and G. Further suppose that h is continuously Fréchet differentiable at $\bar{x}$ with derivative H . If $\bar{x}$ solves $\mathcal{B}$ then there exists a non-zero (u,v,w) in $R \chi Y^* \chi Z^*$ such that

(i) $u \geq 0$, $v \in C^*$

(ii) $vg(\bar{x}) = 0$

(iii) $(uF + vG + wH)(x) \geq 0$ for all x in $D(S,\bar{x})$.

Remark:

If, in the above theorem, S is convex, f is convex, g is

$c\ell$ C-convex and g is affine then condition (iii) is equivalent to

$$uf(x) + vG(x) + wH(x) \geq uf(\bar{x}) + vg(\bar{x}) + wh(\bar{x}) = uf(\bar{x}) .$$

Hence if $\bar{x}$ solves $\mathcal{B}$ then there exists $(u,v,w) \neq 0$ such that conditions (i), (ii) and the above inequality hold.

APPENDIX

1. Bounded Inverse Theorem [2]

Let X be a Banach space, D a subspace of X and Y, a normed linear space. Suppose $A: D \to Y$ is a closed linear transformation and suppose the range of A, $A(D)$ is of category II. Then:

(a) A is onto; that is $A(D) = Y$

(b) there exists an $m > 0$ such that for any $y \in Y$ there is some $x \in D$ such that $Ax = y$ and $\|x\| \leq m\|y\|$.

(c) If A^{-1} exists, it is a bounded linear transformation.

2. Brouwer's Fixed Point Theorem [5]

Any continuous map f of a closed ball in R^n into itself has at least one fixed point; that is, a point x such that $f(x) = x$.

3. Separating Plane Theorem [11]

Let X and Y be two non-empty disjoint convex sets in R^n. Then there exits a plane $\{x: x \in R^n, cx = \alpha\}$, $c \neq 0$ which separates them; that is, $cx \leq \alpha \leq cy$ for all x in X, y in Y.

4. Separating Hyperplane Theorem [10]

Let K_1 and K_2 be convex sets in the normed linear space X such that K_1 has interior points and K_2 has no interior points of K_1. Then there is a closed hyperplane H separating K_1 and K_2; that is, there is an x^* in X^* such that $\sup_{x \in K_1} x^*(x) \leq \inf_{x \in K_1} x^*(x)$.

5. Property of Bounded Linear Operators in Banach Spaces [10]

Let X and Y be normed spaces and let f be an element in the normed space of all bounded linear operators from X into Y. Let the

range of f, denoted by $R(f)$, be closed. Then $R(f^*) = [N(f)]^\perp$ where $N(f)$ is the null space of f.

6. Fundamental Existence Theorem of Linear Volterra Integral equations [9]

If

(a) $y(x) = h(x) + \lambda \int_a^x K(x,t)y(t)dt$ where a is constant;

(b) $K(x,t)$ is real and continuous in the rectangle $a \leq x \leq b$, $|K(x,t)| \leq M$ in R, $K(x,t) \not\equiv 0$;

(c) $h(x) \not\equiv 0$ is real and continuous in I: $a \leq x \leq b$;

then the equation (a) has one and only one continuous solution $y(x)$.

7. Regularly Convex Sets [11]

Definition: Let B^* be the conjugate space to a Banach space B. A set X^* in B^* is said to be regularly convex if, for every functional x_o^* not in X^*, there exists an element x_o in B such that

$$x^*(x_o) < x_o^*(x_o) - \varepsilon \text{ for all } x^* \text{ in } X^* \text{ and some } \varepsilon \geq 0 .$$

Theorem: A set X^* is regularly convex if and only if it is convex and weak* closed.

8. Property of Euler-Lagrange Equations [10]

If $\alpha(t)$ is continuous in $[t_1,t_2]$ and $\int_{t_1}^{t_2} \alpha(t)h(t)dt = 0$ for every h in the normed linear space consisting of all functions on the interval $[t_1,t_2]$ which are continuous and have continuous derivatives with $h(t_1) = h(t_2) = 0$ then $\alpha(t) \equiv c$ in $[t_1,t_2]$ where c is a constant.

BIBLIOGRAPHY

1. M. Altman, "A General Maximum Principle for Optimization Problems", Studia Mathematica, Vol. 31, No. 4, 1968.

2. G. Bachman and L. Narici, Functional Analysis, Academic Press.

3. M. S. Bazaraa and J. J. Goode, "Necessary Optimality Criteria in Mathematical Programming in Normed Linear Spaces", Journal of Optimization Theory and Applications, March 1973, Vol. 11, No. 3.

4. B. D. Craven," A Generalization of Lagrange Multipliers", Bulletin of Australian Mathematical Society, Vol. 3, 1970.

5. N. Dunford and J. T. Schwarz, Linear Operators: Part I General Theory, Interscience, New York 1958.

6. A. M. Geoffrion, "Duality in Non-Linear Programming: A Simplified Applications-oriented Development", Perspectives in Optimization, Addison-Wesley.

7. H. Halkin and L. W. Neustadt, "General Necessary Conditions for Optimization Problems", Proceedings of the National Academy of Science, Vol. 56, No. 4, 1966.

8. Kushner, "Necessary Conditions for Continuous Parameter Stochastic Processes", SIAM Journal of Control, Aug. '72 Vol. 10, No. 3.

9. W. V. Lovitt, Linear Integral Equations, McGraw-Hill, 1924 .

10. D. G. Luenberger, Optimization by Vector Space Methods, John Wiley and Sons, 1969.

11. O. L. Mangasarian, Non-Linear Programming, McGraw-Hill 1969.

12. L. W. Neustadt, "A General Theory of Extremals", Journal of Computer and System Sciences. Vol. 3, No. 1, 1969.

13. B. D. Pshenichnyi, Necessary Conditions for an Extremum, Marcel Dekker Inc., New York 1971.

14. W. Rudin, Principles of Mathematical Analysis, McGraw-Hill, 1964.