OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

A LINEAR APPROXIMATION APPROACH TO

DUALITY IN NONLINEAR PROGRAMMING

by

T.L. Magnanti*

OR 016-73
April 1973

OPERATIONS RESEARCH CENTER

Massachusetts Institute of Technology

Cambridge Massachusetts 02139

* Supported in part by the US Army Research Office (Durham) under Contract DAHC04-70-C-0058

ABSTRACT

Linear approximation and linear programming duality theory are used as unifying tools to develop saddlepoint, Fenchel and local duality theory. Among results presented is a new and elementary proof of the necessity and sufficiency of the stability condition for saddlepoint duality, an equivalence between the saddlepoint and Fenchel theories, and nasc for an optimal solution of an optimization problem to be a Kuhn-Tucker point. Several of the classic "constraint qualifications" are discussed with respect to this last condition. In addition, generalized versions of Fenchel and Rockafeller duals are introduced. Finally, a shortened proof is given of a result of Mangasarian and Fromowitz that under fairly general conditions an optimal point is also a Fritz John point.

Introduction

The notion of duality plays an essential role in both the theoretical and applied aspects of nonlinear programming. Results in duality theory can be broadly classified as belonging to one of the following three closely related areas.

(1) optimality conditions which are local in nature

(2) saddlepoint (min-max) theory which we also refer to as global Lagrange duality, and

(3) Fenchel duality.

Our purpose here is to use a unifying approach of linear approximation coupled with linear programming duality theory to develop several results in each of these three areas. This is to be distinguished from the usual practice of utilizing such tools as separating hyperplanes, conjugate functions and/or subgradients to develop the nonlinear results and then treating linear programming as a special case. The two approaches are, of course, intimately related (e.g., linear programming duality implies separating hyperplane theorems [22]). A feature of our approach is the elementary proofs that it provides for many well known results as well as several extensions to be discussed below. Only basic results from real analysis and linear programming duality theory are required as a prerequisite.

Given the scope of our coverage, we believe that this paper might function not only to present new results and proofs, but also to survey the field. Due to limitations in space, we have omitted applications of the theory and have given only brief accounts of underlying geometry. For coverage of this material see Geoffrion [13], Luenberger [21] and Variaya [32]. Geoffrion's article in particular might be useful as co-reading to complement our discussion.

Of course it has been impossible to give a complete coverage of duality. For example, converse duality and second order conditions have not been included. Nor have special dual forms such as symmetric duals [6], quadratic programming [4], [7], [8], or the dual problems of Wolfe [33], Stoer [30] or Mangasarian and Ponstein [24]. As a guiding principle, we have incorporated new results, material for which we have a new proof, or results which we believe are required for the unity of the presentation. Standard texts such as Luenberger [21], Mangasarian [22], Rockafeller [24], Stoer and Witzgal [31] and Zangwill [34] discuss further topics as well as point to other primary sources in the literature. See also Fisher and Shapiro [10], Shapiro [27], Gould [14], and Kortanek and Evans [19] for recent extensions to integer programming, generalized Kuhn-Tucker vectors and asymptotic duality.

Summary of Results

Section I introduces notation, presents a linear version of duality which forms the basis for following linear approximations, and gives some basic results such as weak duality for the general duality problem.

Section II.A treats the global Lagrange dual. Primary topics that are covered include the stability condition of Gale [12] and Rockafeller [25], the classic Slater Condition [29], and the generalized Slater Condition. Our approach is to relate these results to boundedness of optimal dual variables to certain linear programming approximations. In each case, non-convex extensions are given. Even for the convex case, we believe that the proofs provided by this approach may be of interest.

Fenchel duality is considered in section II.B. Here we establish an extended version of Fenchel's original result. Instead of using linear approximation directly as above, we treat this material by displaying an equivalence between the Fenchel and global Lagrange theories so that previous results can be applied. The connection between the two theories is well known, e.g., both can be derived from Rockafeller's general perturbation theory [25] (see also [21]), but apparently the equivalence that we exhibit is new. The final topic in this section is Rockafeller's dual [25].

In section III, optimality conditions are presented which generalize the classical Lagrange theorem and extensions to systems with inequality constraints given by Fritz John [16] and Kuhn and Tucker [20]. We begin by giving necessary and sufficient conditions for an optimal solution to satisfy the Kuhn-Tucker conditions. These results are based upon familiar linearization procedures (see [3] and [17]) and global Lagrange duality. To our knowledge, though, nasc of this nature have not been treated before. (See [11] for a related approach when $C=R^n$). Many well known "constraint qualifications" are then discussed with respect to these nasc.

For the Fritz John condition, we present a simplified proof of a result due to Mangasarian and Fromowitz [22], [23].

The final section is devoted to extensions and further relationships with other work.

I. Preliminaries

A. Problem statement and notation

We shall consider duality correspondences for the mathematical programming problem (called the primal): Determine v,

$$v = \inf_{x \in C} f(x) \quad \text{subject to } g(x) \leq 0 \tag{P}$$

where C is a subset of R^n, f is a real valued function defined on C and $g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}$ is a column vector of real valued functions each defined on R^n. By $g(x) \leq 0$, we mean each component of g(x) is non-positive.

$\hat{x}$ is called feasible for P if $\hat{x} \varepsilon C$ and $g(\hat{x}) \leq 0$. The feasible point $\hat{x}$ is called optimal if $f(y) \geq f(\hat{x})$ for any y feasible for P. Note that there need not be any optimal points for P, and yet v is well defined. Finally, if no point is feasible for P, we set $v = +\infty$.

The problem below, which is called the perturbed primal, plays a central role in the theory:

$$v(\theta) = \inf_{x \varepsilon C} f(x)$$

$$\text{s.t.} \quad g(x) \leq \theta e^m$$

where e^m is a column vector of m ones, θ is a real valued parameter, and $v(\theta)$ is the value of the perturbed problem as a function of the perturbation θ. Observe that $v(0)$ is the value of the original primal problem.

Note: The results to be obtained do not depend upon our choice of e^m for the right hand side above. Any m-vector with positive components would suffice.

To avoid repitition of needless transposes lower case Greek letters will be used for row vectors and lower case Latin letters for column vectors. The latter will also be used for scalars, e.g., θ above; context should indicate the distinction. Following this convention, we let e^m and ε^m denote respectively a column and row vector of m ones. In addition, we reserve subscripting for components of vectors and use superscripting to denote distinct vectors.

If π is a row vector and x a column vector of the same dimension, then $\pi x = \Sigma\pi_i x_i$, i.e., inner product. As above, vector equalities and inequalities are assumed to hold componentwise, and 0 is used either as a scalar or zero row or column vector of appropriate dimension.

For any subset C of R^n, int(C) denotes the interior of C and ri(C) its relative interior, i.e., interior with respect to the smallest linear variety containing C.

Finally, $||x||$ denotes the Euclidian norm of x, i.e., $||x|| = (\Sigma x_j^2)^{1/2}$.

B. A Linear Case

The following linear version of P provides the basis for our analysis in Part II. Let C be the convex hull of the fixed points $x^1,\ldots,x^k$ of R^n, let A be an m by n matrix, b an m-dimensional column vector, and γ an n-dimensional row vector. The problem is:

$$v(\theta) = \min_{x\varepsilon C} \gamma x \qquad \text{or} \qquad v(\theta) = \min \sum_1^k (\gamma x^j) w_j$$

$$\text{s.t.} \quad Ax \leq b + \theta e^m \qquad\qquad \text{s.t.} \quad \sum_1^k (Ax^j - b) w_j \leq \theta e^m$$

$$\varepsilon^k w = 1$$

$$w \geq 0$$

which is a linear program (LP) in the variables w. We assume that this problem has a feasible solution for $\theta = 0$. By LP duality theory [5], [27].

$$v(\theta) = \max\{\sigma - \theta\pi e^m\}$$

$$\text{s.t.} \qquad \sigma \leq \gamma x^j + \pi(Ax^j - b) \qquad j = 1,\ldots,k$$

$$\pi \geq 0 \quad .$$

Given any π, it is optimal to select $\sigma = \min\limits_{x^j} [\gamma x^j + \pi(Ax^j - b)]$. Thus

$$v(\theta) = \max_{\pi\geq 0} \min_{1\leq j\leq k} [\gamma x^j + \pi(Ax^j - b - \theta e^m] \quad . \tag{1}$$

In fact, since $\gamma x + \pi(Ax - b - \theta e^m)$ is linear in x for any fixed π,

$$v(\theta) = \max_{\pi\geq 0} \min_{x\varepsilon C} [\gamma x + \pi(Ax - b - \theta e^m)] \quad .$$

Remark 1: From paramteric linear programming theory [5], there is a $\theta_0 > 0$ and $\hat{\pi} \geq 0$ dual optimal for $\theta = 0$ such that $v(0) - v(\theta) = \theta\hat{\pi}e^m$ for all $0 \leq \theta \leq \theta_0$.

C. The Saddlepoint Problem and Stability

An elementary argument shows that for the LP problem above (with $f(x) = \gamma x$, $g(x) = Ax-b$)

$$v(\theta) = \inf_{x\varepsilon C} \sup_{\pi\geq 0} [f(x) + \pi(g(x) - \theta e^m)] \tag{2}$$

and we have shown that $v(\theta) = D(\theta)$ where

$$D(\theta) \equiv \sup_{\pi\geq 0} \inf_{x\varepsilon C} [f(x) + \pi(g(x) - \theta e^m)] \quad . \tag{3}$$

Indeed, (2) is true for arbitrary f, g and C. We call the sup inf in (3) with $\theta = 0$ either the saddlepoint or (global) Lagrange dual to P, $L(\pi,x) \equiv f(x) + \pi\, g(x)$ being called the Lagrangian function. For notational convenience, let $D(\theta,\pi) \equiv \inf_{x\varepsilon C} [f(x) + \pi(g(x) - \theta e^m)]$ so that (3) becomes

$$D(\theta) = \sup_{\pi\geq 0} D(\theta,\pi) \ .$$

Observe that $D(\theta,\pi)$ is concave as a function of π.

Note that the primal has no feasible solution if and only if for $\theta = 0$ the sup in (2) is $+\infty$ for each $x\varepsilon C$, i.e., $v(0) = +\infty$. This conforms with our previous convention.

The next two well known results are immediate consequences of these definitions.

Lemma 1 (Weak Duality): $v(0) \geq D(0)$.

Proof: If $v(0) = +\infty$, there us nothing to prove. If $\hat{x}$ is primal feasible, then $f(\hat{x}) \geq f(\hat{x}) + \pi g(\hat{x}) \geq D(0,\pi)$ for any $\pi \geq 0$. But then $v(0) \geq D(0,\pi)$ for all $\pi \geq 0$, thus $v(0) \geq D(0)$. ///

Under certain conditions to be discussed later, $v(0) = D(0)$ and we say that duality holds. Otherwise $v(0) > D(0)$ and we say that there is a duality gap.

Lemma 2: Suppose that $v(0) = D(0) < +\infty$ and that there is a $\hat{\pi} \geq 0$ such that $D(0) = D(0,\hat{\pi})$. Let $M = \hat{\pi}e^m$. Then $v(0) - v(\theta) \leq \theta M$ for all $\theta \geq 0$.

Proof: From the hypothesis, Lemma 1, and the definition of $D(\theta)$,

$$v(0) - v(\theta) \leq D(0) - D(\theta) \leq \inf_{x\in C} [f(x) + \hat{\pi}g(x)] - \inf_{x\in C} [f(x) + \hat{\pi}(g(x)-\theta e^m)]$$

$$= \theta\hat{\pi}e^m \quad .$$

///

As we shall see in the next section, under certain circumstances, the condition that there is a finite number M such that $v(0) - v(\theta) \leq \theta M$ for all $\theta \geq 0$ is not only necessary but sufficient for Lagrange duality to hold. If this condition (often stated in other forms in the literature) holds for the perturbed problem, then we say that the primal is stable.

As an example of a simple problem that is not stable, let $C = \{x:x\geq 0\} \subseteq R$, take $f(x) = -\sqrt{x}$ and $g_1(x) = x$. In this case, $v(0) - v(\theta) = \sqrt{\theta} \not\leq \theta M$ for any M and $v(0) \neq D(0,\pi)$ for any $\pi \in R^m$, $\pi \geq 0$. This example as well as several others that illustrate many of the concepts to be considered below are presented in [13].

II. Global Theory

A. Lagrange (saddlepoint) Duality

An elementary fact from real analysis states that if C is a nonempty subset of R^n then C is separable with respect to the norm $||\cdot||$, i.e., there are countable points $x^1, x^2, \ldots$ that are dense in C (each $x^j \in C$ and given any $y \in C$ and $\varepsilon > 0$ there is an x^j such that $||y - x^j|| < \varepsilon$).

Let $f^k = (f(x^1), f(x^2), \ldots, f(x^k))$ and let $g^k = (g(x^1), \ldots, g(x^k))$ denote respectively a k-dimensional row vector and an m by k matrix determined by the first k x^j's. Consider the following LP approximation to the perturbed problem (in the variables w_j):

$$
\begin{aligned}
P^k(\theta): \quad v^k(\theta) = \inf \quad & f^k w \\
\text{s.t.} \quad & g^k w \leq \theta e^m \\
& \varepsilon^k w = 1 \\
& w \geq 0 \quad .
\end{aligned}
$$

This LP makes piecewise linear approximations to f and the g_j over the convex hull of the points $x^1, \ldots, x^k$. We assume that x^1 is feasible for P; thus, $w = (1,0,\ldots,0)$ is feasible for P^k.

We call the original problem P <u>regular</u> if there exists $x^1, \ldots, x^k, \ldots$ contained in C with the properties:

(R1) For $k=1,2,\ldots$, $v^k(\theta)$ overestimates $v(\theta)$ locally about $\theta \geq 0$, i.e., there is a $\bar{\theta}_k > 0$ such that $v^k(\theta) \geq v(\theta)$ for $0 \leq \theta \leq \bar{\theta}_k$.

(R2) Given any $y \varepsilon C$ and any $\pi \geq 0$, there is a subsequence $\{k_j\}$ of $\{1,2,\ldots\}$ such that $x^{k_j} \to y$ and $\underline{\lim}\, [f(x^{k_j}) + \pi g(x^{k_j})] \leq f(y) + \pi g(y)$, where $\underline{\lim}$ denotes limit inferior [26].

P is called <u>y-regular</u> if it is regular with $x^1 = y$.

<u>Remark 2</u>: (i) If C is a convex set, f is convex on C and each g_j are convex on R^n, then $v^k(\theta) \geq v(\theta)$ for all $\theta \geq 0$ and any choice of $x^j \varepsilon C$. Also (R2) is a standard result of convex functions (see the appendix for a proof). Thus, the problem is y-regular for any $y \varepsilon C$ and thus a fortiori regular.

(ii) If each g_j is continuous, then R2 is equivalent to $\underline{\lim}\, f(x^{k_j}) \leq f(y)$. If, in addition, f is (upper semi-) continuous on C, then R2 holds.

Our main result in this section is that if x^* solves P and if P is x^*-regular, then either stability of P or boundedness of the dual variables in the LP approximations is a necessary and sufficient condition for duality to hold. The geometry of these results is illustrated in Figure 1. From LP theory, $v^k(\theta)$ is convex and piecewise linear. Selecting π^k optimal for the dual to $P^k(0)$ satisfying remark 1, $-\pi^k e^m$ is the slope of the initial segment of $v^k(\theta)$. Under R1 and R2, these slopes approach the dashed line segment supporting $v(\theta)$ at $\theta = 0$ and consequently their boundedness is essentially equivalent to stability, i.e., $M < +\infty$. Furthermore, the vector $\hat{\pi}$ with $\hat{\pi}e^m = M$ solves the dual problem.

We begin by establishing a sufficient condition without the full regularity assumption.

<u>Theorem 1</u>: Suppose that $v^k(0) \geq v(0)$ for $k=1,2,\ldots$ and assume R2. Let π^k be optimal variables for the dual to $P^k(0)$. If $\pi^{k_j} \to \hat{\pi}\varepsilon R^m$ for some subsequence $\{k_j\}$ of $\{1,2,\ldots\}$, then $v(0) = D(0) = D(0,\hat{\pi})$.

<u>Proof</u>: From section 1.B, $v^k(0) = \min_{1\leq j\leq k} \{f(x^j) + \pi^k g(x^j)\}$.

Since $v^k(0) \geq v(0)$ this implies

$$f(x^j) + \pi^k g(x^j) \geq v(0) \quad \text{for all } j \leq k .$$

Consequently, $f(x^j) + \hat{\pi}g(x^j) \geq v(0)$ for all x^j.

Finally, let $y\varepsilon C$. Then by R2 there is a subsequence $\{k_j'\}$ of $\{1,2,\ldots\}$ with $x^{k_j'} \to y$ and $f(y) + \hat{\pi}g(y) \geq \underline{\lim}\, [f(x^{k_j'}) + \hat{\pi}g(x^{k_j})] \geq v(0)$. Thus, $D(0,\hat{\pi}) = \inf_{x\varepsilon C} [f(x) + \hat{\pi}g(x)] \geq v(0)$. But $D(0,\hat{\pi}) \leq D$ and by weak duality $D(0,\hat{\pi}) \leq v(0)$. ///

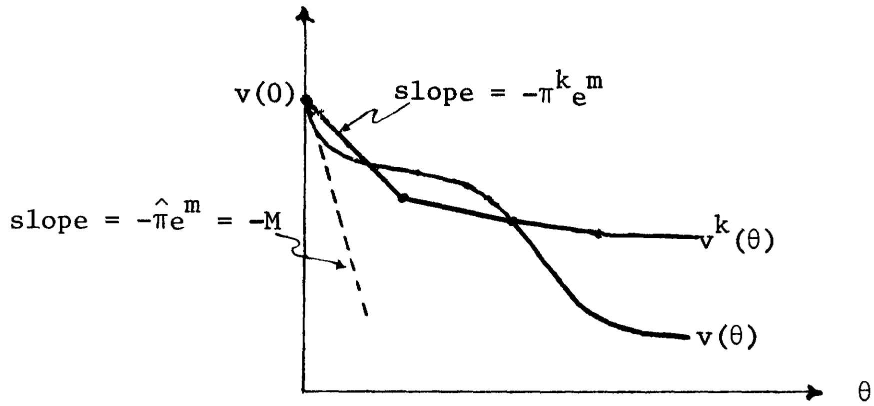

Figure 1

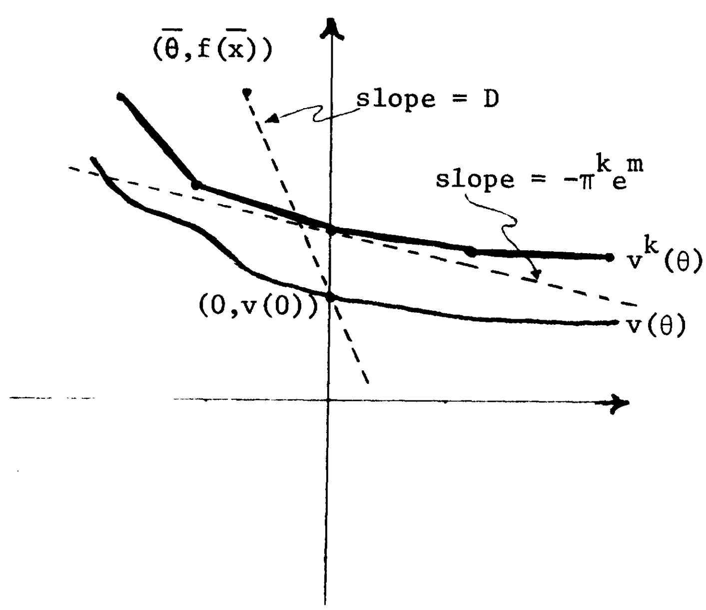

Figure 2

<u>Corollary 1.1</u>: Suppose that $v^k(0) \geq v(0)$ for $k=1,2,\ldots$ and assume R2. If the sequence π^k of optimal dual variables for $P^k(0)$ are bounded, then $v(0) = D(0) = D(0,\hat{\pi})$ for some $\hat{\pi}\varepsilon R^m$, $\hat{\pi} \geq 0$.

<u>Proof</u>: Immediate.

<u>Corollary 1.2</u>: (Slater Condition) Suppose that there is an $\bar{x}\varepsilon C$ satisfying $g(\bar{x}) < 0$. Assume R2 and that $v^k(0) \geq v(0)$ with $x^1 = \bar{x}$. Then there is a $\hat{\pi} \geq 0$ such that $v(0) = D(0,\hat{\pi})$.

<u>Proof</u>: From the LP dual to $P^k(0)$, for $k \geq 1$, $f(\bar{x}) + \pi^k g(\bar{x}) \geq v(0)$ or $\pi^k[-g(\bar{x})] \leq f(\bar{x}) - v(0)$. Since each component of $-g(\bar{x})$ is positive and $\pi^k \geq 0$, this implies that the π^k are bounded. Apply Corollary 1.1. ///

The geometry of this result is illustrated in Figure 2. From LP duality theory $-\pi^k e^m$ is the slope of a supporting line to $v^k(\theta)$ at $\theta = 0$. Letting $\bar{\theta} = \max_j [g_j(\bar{x})]$, $v^k(\bar{\theta}) \leq f(\bar{x})$ and consequently if $v^k(0) \geq v(0)$, $-\pi^k e^m$ is no smaller than the slope D of the line joining $(\bar{\theta},f(\bar{x}))$ and $(0,v(0))$.

<u>Theorem 2</u>: Assume that x* solves P and that P is x*-regular. Let π^k be optimal dual variables for $P^k(0)$. Then $v(0) = D(0,\hat{\pi})$ for some $\hat{\pi} \geq 0$ if and only if the π^k are bounded.

<u>Proof</u>: In light of Corollary 1.1, we must only prove necessity. We simply note that there exists a $\theta_0^k > 0$ such that, for all $0 \leq \theta \leq \theta_0^k$

$$\theta\pi^k e^m = v(0) - v^k(\theta) \leq v(0) - v(\theta) \leq \theta\hat{\pi}e^m.$$

The equality is a result of Remark 1 and the fact that $v(0) = v^k(0)$ if P is x*-regular and x* solves P; the first inequality is a result of $v^k(0) \geq v(0)$ and the final inequality a consequence of Lemma 2. Thus, $\pi^1,\pi^2,\ldots$ are bounded from above by $\hat{\pi}e^m$ and from below by 0. ///

<u>Theorem 3</u>: Let x* solve P and suppose that P is x*-regular. Then there is a $\hat{\pi} \geq 0$ satisfying $v(0) = D(0,\hat{\pi})$ if and only if P is stable.

Proof: Lemma 2 provides the necessity. For sufficiency, simply note that by x*-regularity (each $v^k(0) = v(0)$) and Remark 1, $\theta\pi^k e^m = v(0) - v^k(\theta) \leq v(0) - v(\theta) \leq \theta M$ for some $\theta > 0$. Thus the π^k are bounded and Theorem 2 applies. ///

Actually, x*-regularity is not required above when C, f and g are convex. A proof of this fact is given in [13] and [25]. We state the result here for use in coming sections.

Theorem 3A: Suppose that in P, C is convex, f is a convex function on C, and each g_j is a convex function on R^n. Then there is a $\hat{\pi}\epsilon R^m$, $\hat{\pi} \geq 0$ satisfying $v(0) = D(0,\hat{\pi})$ if and only if P is stable.

Next, consider the optimization problem

$$P_1: \quad \inf_{x\epsilon C} f(x) \quad \text{s.t. } g(x) \leq 0,\ h(x) = 0 \qquad \text{or} \qquad \inf_{x\epsilon C} f(x) \quad \text{s.t. } g(x) \leq 0,\ h(x) \leq 0,\ -h(x) \leq 0 \quad \text{where } h(x) = \begin{pmatrix} h_1(x) \\ \vdots \\ h_r(x) \end{pmatrix}$$

We say that the problem is stable if the form to the right is stable. Writing the dual to the second form and simplifying, we find

$$D(0) = \sup_{\substack{\pi\geq 0 \\ \alpha\epsilon R^r}} \inf_{x\epsilon C} [f(x) + \pi g(x) + \alpha h(x)] \equiv \sup_{\substack{\pi\geq 0 \\ \alpha\epsilon R^r}} D(0,\pi,\alpha)$$

i.e., the dual variables α corresponding to $h(x)$ are unconstrained in sign. Let P_1^k denote the LP approximation to P_1, i.e., let $h^k = (h(x^1),\ldots,h(x^k))$ and add $h^k w = 0$ to $P^k(0)$.

Arguing as in the proof of Theorem 1, we may prove:

Theorem 1A: Suppose that $v^k(0) \geq v(0)$ for k=1,2,... and assume R2. Let (π^k,α^k) be optimal dual variables for P_1^k. If the sequences $\{\pi^k\}$ and $\{\alpha^k\}$ are bounded, then there is a $\hat{\pi} \geq 0$, $\alpha \varepsilon R^r$ satisfying $v(0) = D(0) = D(0,\hat{\pi},\hat{\alpha})$.

A consequence of this theorem is a generalized version of Corollary 1.2. We say that P satisfies the generalized Slater condition if:

(i) there is an $\bar{x} \varepsilon C$ satisfying $g(\bar{x}) < 0$, $h(\bar{x}) = 0$

(ii) for j=1,...,r, there exist $y^j \varepsilon C$, satisfying $h(y^j) = \theta_j u^j$
$\bar{y}^j \varepsilon C$, satisfying $h(\bar{y}^j) = -\bar{\theta}_j u^j$

where u^j is the j^{th} unit vector, and the θ_j and $\bar{\theta}_j$ are positive constants.

Corollary 1A.1: Assume R2 and that P satisfies the generalized Slater condition. Also let $v^k(0) \geq v(0)$ for k=1,2,... and assume that the points $\bar{x}, y^j, \bar{y}^j$ in the definition of the generalized Slater condition appear in $x^1, x^2, \ldots$. Then there is a $\hat{\pi} \geq 0$, $\hat{\alpha} \varepsilon R^r$ with $v(0) = D(0,\hat{\pi},\hat{\alpha})$.

Proof: Let P^d contain the points $\bar{x}$, y^j, $\bar{y}^j$, j=1,...,r. Then for $k \geq d \geq i$

$$\pi^k[-g(\bar{x})] \leq f(\bar{x}) - v(0) \qquad \text{(i)}$$

$$\alpha^k h(x^i) \geq v(0) - \pi^k g(x^i) - f(x^i) \qquad \text{(ii)}$$

From (i) the π^k are bounded. Consequently, as x^i varies over y^j, $\bar{y}^j$, (ii) implies that the α^k are bounded. Now apply Theorem 1A. ///

Suppose that h(x) is composed of affine functions, $h(x) = Ax + b$, and let C be a convex set. Then we may state a condition that easily implies the generalized Slater condition. Here for the first time we formally use a separating hyperplane theorem for convex sets. This is somewhat illusory, however, since it is well known that LP duality theory (or one of its equivalents such as Gordon's Lemma) is equivalent to this separation theorem [22]. In fact, Theorem 1 itself supplies a simple proof of a separating hyperplane theorem [13].

Lemma 3: Let $h:C \to R^r$ be affine and let $C \subseteq R^n$ be convex. Suppose that there is no $\alpha \neq 0$ such that $\alpha h(x) \geq 0$ for all $x \varepsilon C$. Then there exist y^j, $\overline{y}^j \varepsilon C$ $j=1,\ldots,r$ satisfying condition (ii) above.

Proof: Suppose there does not exist a $y^j \varepsilon C$ with $h(y^j) = \theta_j u^j$ for any $\theta^j > 0$. By linearity and the convexity of C, $h(C) = \{h(x):x\varepsilon C\}$ is a convex set disjoint from $\overline{C} = \{\theta u^j:\theta > 0\}$. By the separating hyperplane theorem, this implies that there is an $\alpha \neq 0$, $\beta \varepsilon R$ such that $\alpha h(x) \geq \beta$ for all $x\varepsilon C$, $\alpha y \leq \beta$ for all $y \varepsilon \overline{C}$. This last relation implies that $\beta \geq 0$.

A similar argument applies if no $\overline{y}^j \varepsilon C$ gives $h(\overline{y}^j) = -\overline{\theta}_j u^j$ for any $\overline{\theta}_j > 0$. ///

Remark 3: The results presented in this section (besides Remark 2) do not depend upon properties of R^n. In fact, they are applicable to optimization problems with a finite number of constraints (equality or inequality) in any linear topological space as long as C is separable (i.e., there are points $x^1, x^2, \ldots$ of C such that for any $y\varepsilon C$ and any neighborhood $N(y)$ of y, there is an $x^j \varepsilon N(y)$). In particular, the results are valid in separable metric spaces.

B. Fenchel Duality

Suppose that f is a convex function and g a concave function defined respectively on the convex subsets C_1 and C_2 of R^n. The Fenchel duality theorem [9] states that under appropriate conditions $v \equiv \inf_{\text{s.t. } x\varepsilon C_1 \cap C_2} [f(x) - g(x)]$ satisfies

$$v = D \equiv \max_{\pi} \; [g^*(\pi) - f^*(\pi)]$$

$$\text{s.t.} \quad \pi \varepsilon C_1^* \cap C_2^*$$

where $f^*(\pi) = -\inf\limits_{x \varepsilon C_1} [f(x) - \pi x]$ and $g^*(\pi) = \inf\limits_{x \varepsilon C_2} [\pi x - g(x)]$

$$C_1^* = \{\pi : f^*(\pi) < +\infty\}, \quad C_2^* = \{\pi : g^*(\pi) > -\infty\}.$$

f^* and g^* are usually referred to respectively as the conjugate convex and concave functions of f and g.

In this section, an extended version of this result is established. Our approach is to state the Fenchel problem in the form of P from the preceding section and then to apply the Lagrange theory. We also establish an equivalence between the two theories by showing that every problem P may be treated as a Fenchel problem. For simplicity, attention will be restricted to the convex case, though non-convex generalization may be given along the lines of the preceding section.

Let C_j $j=1,\ldots,r$ be convex sets in R^n and for $j=1,\ldots,r$ let f_j be a convex function defined C_j. Consider the optimization problem

$$v \equiv \inf \sum_{j=1}^{r} f_j(x) \qquad \text{(F)}$$

$$\text{s.t.} \quad x \varepsilon \bigcap_{j=1}^{r} C_j$$

Define $C \equiv C_1 x C_2 x \ldots x C_r x R^n$, let $x = (x^1,\ldots,x^r,\bar{x}) \varepsilon C$ and define $f: C \to R$ by $f(x) = \sum_{j=1}^{r} f_j(x^j)$. Then C is convex, f is convex on C and F is equivalent to the Lagrange problem:

$$v = \inf_{x \varepsilon C} f(x) \qquad (\overline{P})$$

$$\text{s.t.} \quad x^j = \bar{x} \quad j=1,\ldots,r \;.$$

The Lagrange dual to $\overline{P}$ is:

$$D \equiv \sup_{\pi^j} \inf_{x \varepsilon C} \left[f(x) + \sum_{j=1}^{r} \pi^j (x^j - \bar{x}) \right] \quad .$$

For the inf to be finite, it is necessary that $\sum_{1}^{r} \pi^j = 0$, and thus after rearrangement the dual becomes

$$D = \sup_{\Sigma\pi^j=0} \sum_{j=1}^{r} \inf_{x\varepsilon C_j} [f_j(x) + \pi^j x] \equiv \max_{\Sigma\pi^j=0} D(\pi^1,\ldots,\pi^r).$$

Note that for r=2, this has the form of the Fenchel dual from the first paragraph of this section. From previous results, we immediately have:

Theorem 4: $v = D = D(\hat{\pi}^1,\ldots,\hat{\pi}^r)$ for some $\hat{\pi}^j \varepsilon R^n$, $j=1,\ldots,r$ with $\sum_{1}^{r} \pi^j = 0$ if and only if $\overline{P}$ is stable.

Below we establish another sufficient condition for this duality to hold. Often it is easier to establish this condition than stability of $\overline{P}$ directly.

Lemma 4: Nonsingular affine transformations of R^n do not affect the duality relationship for F.

Proof: Let $L(x) = Ax - y$ for A a nonsingular n by n matrix and $y\varepsilon R^n$ fixed. Define $L(C_j) \equiv \{z\varepsilon R^n : z=Ax-y \text{ for some } x\varepsilon C_j\}$. Then substituting $A^{-1}(z+y)$ for x in F gives the equivalent problem:

$$\inf \sum_{1}^{r} f_j[A^{-1}(z+y)]$$
$$\text{s.t. } z \ \varepsilon \bigcap_{1}^{r} L(C_j).$$

Note that $f_j[A^{-1}(z+y)]$ is convex as a function of z and that each $L(C_j)$ is convex. The dual to the transformed problem is:

$$\sup_{\Sigma\hat{\pi}^j=0} \sum_{1}^{r} \inf_{z\varepsilon L(C_j)} [f_j[A^{-1}(z+y)] + \hat{\pi}^j z] = \sup_{\Sigma\hat{\pi}^j=0} \sum_{j=1}^{r} \inf_{x\varepsilon C_j} [f_j(x) + \hat{\pi}^j (Ax-y)].$$

But $\sum_{1}^{r}\hat{\pi}^j y = (\sum_{1}^{r}\hat{\pi}^j)y = 0$; also defining $\pi_j = \hat{\pi}_j A$, $\Sigma\pi_j = (\Sigma\hat{\pi}_j)A = 0$. Consequently, this dual is equivalent to the Fenchel dual of F. ///

Remark 4: Suppose that $\bigcap_{j=1}^{r} \text{int } (C_j) \neq \emptyset$. By the preceding lemma, we may assume that $0 \ \varepsilon \bigcap_{j=1}^{r} \text{int } (C_j)$. But then for $j=1,\ldots,r$ and $k=1,\ldots,n$, let $x^{jk} = (0,\ldots,0,x^j=\theta u^k,0,\ldots,0)$ where u^k is the k^{th} unit vector in R^n and θ is chosen small enough so that each x^{jk} and $-x^{jk}$ is contained in C. Substituting in $\overline{P}$, the x^{jk}'s and $-x^{jk}$'s provide the requirements for the generalized Slater condition and consequently, duality holds. The following result generalizes this observation.

Theorem 5: Suppose that $\bigcap_{j=1}^{r} ri(C_j) \neq \phi$ in F. Then $v = D = D(\pi^1,\ldots,\pi^r)$ for some $\pi^j \varepsilon R^n$, $j=1,\ldots,r$ with $\sum_1^r \pi^j = 0$.

Proof: For simplicity we assume r=2. The general case is treated analogously. By Lemma 4, we may assume that $0\varepsilon ri(C_1)\cap ri(C_2)$. Let S be the smallest linear subspace containing $C_1 \cap C_2$ and for j=1,2 let S_j be the smallest linear subspace containing C_j. By a change of basis and lemma 4, we may assume that the unit vectors $u^1,\ldots,u^{d_o}$ $d_o \leq n$ are a basis for S and that these vectors together with $u^{d_o+1},\ldots,u^{d_1}$ are a basis for S_1 and together with $u^{d_1+1},\ldots,u^{d_2}$ are a basis for S_2.

Note that for $x^1 \varepsilon C_1$, the components $x_k^1 = 0$ for $k > d_1$ and for $x^2 \varepsilon C_2$, $x_k^2 = 0$ for $k > d_2$ or $d_o < k \leq d_1$. Thus F may be rewritten as

$$v = \inf_{x\varepsilon C}[f_1(x^1) + f_2(x^2)]$$

$$\text{s.t.} \quad \left.\begin{array}{l} x_j^1 = \overline{x}_j \\ x_j^2 = \overline{x}_j \end{array}\right\} \; j=1,\ldots d_o$$

$$x_j^1 = 0 \quad j=d_o+1,\ldots,d_1$$

$$x_j^2 = 0 \quad j=d_1+1,\ldots,d_2 \;.$$

The notation here corresponds to that in $\overline{P}$. But now since $0\varepsilon ri(C_1)\cap ri(C_2)$, for each x_k^i appearing above there is an $x\varepsilon C$ with $x_k^i = \pm\theta$ for some $\theta > 0$ and with each x_j^1, x_j^2, $\overline{x}_j = 0$ $j\neq k$. That is, the problem above satisfies the generalized Slater conditions and duality holds. Let π_j^i be optimal dual variables for the problem. For $j=p+1,\ldots,d_1$ define $\pi_j^2 = -\pi_j^1$, noting that $\pi_j^2 x_j^2 = 0$ since $x_j^2 = 0$ for $x^2 \varepsilon C_2$. Similarly, for $j=d_1+1,\ldots,d_2$ define $\pi_j^1 = -\pi_j^2$ and note that $\pi_j^1 x_j^1 = 0$ for all $x^1 \varepsilon C_2$. In addition, define $\pi_j^1 = \pi_j^2 = 0$ for $j = d_2+1,\ldots,n$. Combined with the dual to the problem above, these variables give the desired result $v = D(\pi^1,\pi^2)$. ///

For the general case, $u^1,\ldots,u^{do}$ above is taken as a basis for the smallest linear subspace containing $\bigcap_{j=1}^{r} C_j$, is extended to a basis for the intersection of every (r-1) sets, which in turn is extended to a basis for the intersection of every (r-2) sets, etc..

Consider next the extension of Fenchel's problem introduced by Rockafeller:

$$\begin{aligned} v \equiv \inf & \{f_1(x) + f_2(Ax)\} \\ \text{s.t.} \quad & x\varepsilon C_1 \\ & Ax\varepsilon C_2 \end{aligned} \qquad (\overline{F})$$

where A is a given m by n matrix. C_2 is a convex subset of R^m, $f_2: R^m \to R$ is a convex function and C_1 and f_1 are as above. As a Lagrange problem, this becomes

$$\inf_{(x^1,x^2,\bar{x})\varepsilon C_1 x C_2 x R^n} [f_1(x^1) + f_2(x^2)]$$

$$x^1 = \bar{x}$$

$$x^2 = A\bar{x}$$

Its dual is: $\sup_{\pi^1,\pi^2} \{ \sum_{j=1}^{2} \inf_{x^j \varepsilon C_j} [f_j(x^j) + \pi^j x^j] + \inf_{\bar{x}\varepsilon R^n}[-(\pi^1+\pi^2 A)\bar{x}]\}$

For the last inf to be finite, $\pi^1 = -\pi^2 A$, thus the dual becomes:

$$\sup_{\pi} \{\inf_{x\varepsilon C_1} [f_1(x) - \pi Ax] + \inf_{x\varepsilon C_2} [f_2(x) + \pi x]\} = D(-\pi A,\pi) .$$

By arguing as in the previous theorem, our approach provides an alternate proof to Rockafeller's theorem [25]:

<u>Theorem 6</u>: If there is $x\varepsilon ri(C_1)$ with $Ax\varepsilon ri(C_2)$ in $\overline{F}$, then $v = D(-\pi A,\pi)$ for some $\pi\varepsilon R^m$.

For the proof, first translate C_1 so that $0 \varepsilon ri(C_1)$ with $A0 = 0 \varepsilon ri(C_2)$. Then let $\bar{S}_2$ be the smallest linear subspace of R^m containing C_2, let $S_2 = \{x: Ax \varepsilon \bar{S}_2\}$ and proceed as in the proof of theorem 5.

Of course, we also have the obvious result that the conclusion here holds if and only if the Lagrange version of $\bar{F}$ is stable.

Above, we have shown that Fenchel duality is a consequence of Lagrange duality. Now we show the converse, thus establishing that the Fenchel and Lagrange saddlepoint theories are actually equivalent. Consider the Lagrange problem P with C a convex set and f and each g_j convex functions. Let $C_1 = \{\bar{y} = \binom{y_0}{y} \varepsilon R^{m+1}: y_0 \geq f(x)$ and $y \geq g(x)$ for some $x \varepsilon C\}$ and let $C_2 = \{\bar{y} = \binom{y_0}{y} \varepsilon R^{m+1}: y \leq 0, y_0 \varepsilon R\}$. C_1 and C_2 are convex sets and the Lagrange primal is equivalent to:

$$\inf y_0$$

$$\text{s.t. } \bar{y} \varepsilon C_1 \cap C_2 .$$

Identifying $f_1(\bar{y}) = y_0$ and $f_2(\bar{y}) \equiv 0$, the Fenchel dual to this problem is:

$$\max_{\pi_0, \pi} \{\inf_{\bar{y} \varepsilon C_1} [y_0 + \pi_0 y_0 + \pi y] + \inf_{\bar{y} \varepsilon C_2} [\pi_0 y_0 - \pi y]\}$$

Since the second inf is $-\infty$, if $\pi_0 \neq 0$ or some $\pi_j < 0$ $j=1,\ldots,m$, this dual can be rewritten as:

$$\max_{\pi \geq 0} \inf_{\bar{y} \varepsilon C_1} [y_0 + \pi y] = \max_{\pi \geq 0} \inf_{x \varepsilon C} [f(x) + \pi g(x)]$$

which is the Lagrange dual. (Note that $\text{int}(C_1) \cap \text{int}(C_2) \neq \phi$ iff P satisfies the Slater condition.)

III. Local Duality

Throughout this section, we assume that $\bar{x}$ solves the optimization problem:

$$
\begin{aligned}
v \equiv \min_{x \in C} \; & f(x) \\
\text{s.t.} \;\; & g(x) \leq 0 \\
& h(x) = 0
\end{aligned}
\qquad (P')
$$

where $C \subseteq R^n, f: C \to R,\ g: R^n \to R^m,\ h: R^n \to R^r$; also f and each g_j and h_i are assumed to be differentiable at $\bar{x}$.

We say that $\hat{x}$ is a Fritz John point of P' if there exists $\pi \varepsilon R^m$, $\alpha \varepsilon R^r$ and $\pi_0 \varepsilon R$ not all zero such that

$$
\begin{aligned}
[\pi_0 \nabla f(\hat{x}) + \pi \nabla g(\hat{x}) + \alpha \nabla h(\hat{x})](x - \hat{x}) & \geq 0 \text{ for all } x \varepsilon C \\
\pi g(\hat{x}) & = 0 \\
\pi & \geq 0
\end{aligned}
\qquad (4)
$$

where $\nabla f(\hat{x}) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)_{x=\hat{x}}$ is a row vector, the gradient of f at $\hat{x}$,

$\nabla g(\hat{x}) = \begin{pmatrix} \nabla g_1(\hat{x}) \\ \vdots \\ \nabla g_m(\hat{x}) \end{pmatrix}$ an m by n matrix and similarly $\nabla h(\hat{x})$ is an r by n matrix.

(π_0, π, α) is called a Fritz John vector at $\hat{x}$ and (4) are called the Fritz John conditions.

$\hat{x}$ is called a Kuhn-Tucker point if π_0 above is 1. In this case, (π, α) is called a Kuhn-Tucker vector at $\hat{x}$ and (4) the Kuhn-Tucker conditions.

Remark 5: If $\hat{x}$ is interior to C, then (4) can be replaced by

$$
\begin{aligned}
[\pi_0 \nabla f(\hat{x}) + \pi \nabla g(\hat{x}) + \alpha \nabla h(\hat{x})] & = 0 \\
\pi g(\hat{x}) & = 0. \\
\pi & \geq 0
\end{aligned}
$$

Our purpose here is

(i) to give necessary and sufficient conditions for a Kuhn-Tucker vector to exist at the point $\bar{x}$ and to discuss several of the well-known "constraint qualifications" from the literature from this viewpoint.

(ii) To supply a shortened proof that a Fritz John vector exists at $\bar{x}$ whenever C is convex and has a non-empty interior and h has continuous first partial derivatives in a neighborhood of $\bar{x}$.

We use linear approximation and either LP duality or the saddlepoint theory as our principal tools. In this case, however, approximations are based upon first order Taylor expansions, not upon inner linearization of C as used previously.

The following elementary fact from advanced calculus will be used often. If r is differentiable at $\bar{x}\varepsilon R^n$ and $y\varepsilon R^n$, then the directional derivative in the direction y satisfies

$$\lim_{\alpha\to 0} \frac{r(\bar{x}+\alpha y) - r(\bar{x})}{\alpha} = \nabla r(\bar{x})y \ . \tag{*}$$

Consequently, if $\nabla r(\bar{x})y < 0$, $r(\bar{x}+\alpha y) < r(\bar{x})$ for $\alpha > 0$ sufficiently small.

Also, if $r:R \to R^n$, we denote $\left.\frac{dr(\theta)}{d\theta}\right|_{\theta=\bar{\theta}}$ by $r'(\bar{\theta})$.

A. Kuhn-Tucker Conditions

Let $A = \{i: 1\leq i\leq m, \ g_i(\bar{x})=0\}$, let $g_A(\bar{x})$ denote the vector $(g_j(\bar{x}):j\varepsilon A)$ and $\nabla g_A(\bar{x})$ the matrix of gradients determined by $g_A(\bar{x})$, i.e., $\nabla g_A(\bar{x}) = (\nabla g_j(\bar{x}):j\varepsilon A)$. Consider making a linear approximation to P' locally about $\bar{x}$. Since the $g_j(\cdot)$ are continuous at $\bar{x}$, $g_j(x) < 0$ for all $j\notin A$ and x in a neighborhood of $\bar{x}$. Thus we ignore these constraints and write the approximation as:

$$\begin{aligned} \hat{v} \quad & \min_{x\varepsilon C} \{f(\bar{x}) + \nabla f(\bar{x})(x-\bar{x})\} \\ & \text{s.t.} \quad g_A(\bar{x}) + \nabla g_A(\bar{x})(x-\bar{x}) = \nabla g_A(\bar{x})(x-\bar{x}) \leq 0 \\ & \qquad\quad h(\bar{x}) + \nabla h(\bar{x})(x-\bar{x}) = \nabla h(\bar{x})(x-\bar{x}) = 0 \ . \end{aligned}$$

We say that P' is <u>regular at $\bar{x}$</u> (not to be confused with the global regularity of section II) if $\hat{v} = v = f(\bar{x})$. Since $x=\bar{x}$ is feasible for the linear approximation, $\hat{v} \leq v$ and regularity is equivalent to $v^L = 0$ where

$$
\begin{aligned}
v^L \equiv \inf_{y\varepsilon C-\bar{x}} \quad & \nabla f(\bar{x})y \\
\text{s.t.} \quad & \nabla g_A(\bar{x})y \leq 0 \\
& \nabla h(\bar{x})y = 0
\end{aligned}
\tag{L}
$$

i.e., y=0 solves L. The dual to L is $\sup_{\substack{\pi_A \geq 0 \\ \alpha\varepsilon R^r}} \inf_{y\varepsilon C-\bar{x}} [\nabla f(\bar{x})+ \pi_A\nabla g_A(\bar{x})+ \alpha\nabla h(\bar{x})]y.$

An example of a problem that is not regular at $\bar{x}$ is min x_1 subject to $x_2 \leq 0$ and $x_1^2 - x_2 \leq 0$ which has $(x_1,x_2) = (0,0)$ as its only feasible solution. Note also, that (0,0) is not a Kuhn-Tucker point.

The basic result for Kuhn-Tucker vectors which generalizes this observation is:

<u>Lemma 5</u>: There is a Kuhn-Tucker vector at $\bar{x}$ if and only if P' is regular at $\bar{x}$ and saddlepoint duality holds for L.

<u>Proof</u>: (Sufficiency) If $v^L = 0$ and saddlepoint duality holds, then there is a $\pi_A \geq 0$ and $\alpha\varepsilon R^r$ with

$$[\nabla f(\bar{x})+ \pi_A\nabla g_A(\bar{x}) + \alpha\nabla h(\bar{x})]y \geq 0 \quad \text{for all } y\varepsilon C-\bar{x}. \tag{5}$$

Letting $\pi_j = 0$ for $j\notin A$, then $\pi g(\bar{x}) = 0$ and (5) is equivalent to

$$[\nabla f(\bar{x}) + \pi\nabla g(\bar{x}) + \alpha\nabla h(\bar{x})](x-\bar{x}) \geq 0 \text{ for all } x\varepsilon C.$$

(Necessity) If (π,α) is a Kuhn-Tucker vector, then $\pi_j = 0$ for $j\notin A$, so that the gradient condition is just (5). But then the inf of (5) over $y\varepsilon C-\bar{x}$ is non-negative, so that weak duality implies that $v^L \geq 0$. Since $v^L \leq 0$, $v^L = 0$ thus L is regular at $\bar{x}$ and saddlepoint duality holds. ///

<u>Corollary 5.1</u>: Suppose that C is convex. Then there is a Kuhn-Tucker vector at $\bar{x}$ if and only if P' is regular at $\bar{x}$ and L is stable.

<u>Proof</u>. By Theorem 3 saddlepoint duality holds for L in this case if and only if L is stable. ///

If $\bar{x}$ is an interior point of C, then regularity at $\bar{x}$ is equivalent to $v^{L'} = 0$ where

$$v^{L'} = \inf_{y \in R^n} \nabla f(\bar{x})y$$
$$\text{s.t. } \nabla g_A(\bar{x})y \leq 0 \qquad \text{(L')}$$
$$\nabla h(\bar{x})y = 0 \quad .$$

Clearly $v^{L'} \leq v^L$. But if there is a y feasible for L' with $\nabla f(\bar{x})y < 0$, then for some $\theta > 0$, $\theta y \varepsilon C-\bar{x}$ is feasible in L with $\nabla f(\bar{x})(\theta y) < 0$. Thus $v^{L'} \geq 0$ iff $v^L \geq 0$. Since L' is a linear program, duality is assured. But the Linear programming dual gives precisely the Kuhn-Tucker conditions as stated in remark 5. In fact, this same argument may be used for the somewhat weaker hypothesis that C satisfies:

for any $y \varepsilon R^n$ there is a $\theta > 0$ such that $x + \theta y \varepsilon C$. If this condition holds, we say that C encloses $\bar{x}$.

Consequently, for this case, Lemma 5 becomes

Lemma 6: Suppose that C encloses $\bar{x}$. Then there is a Kuhn-Tucker vector at $\bar{x}$ if and only if P' is regular at $\bar{x}$.

In particular, if $\bar{x}$ is an interior point of C, then there is a Kuhn-Tucker vector at $\bar{x}$ if and only if P' is regular at $\bar{x}$.

There are several properties of problem P' that imply the existence of Kuhn-Tucker vectors. Among the most well known are:

(i) Kuhn-Tucker constraint qualification [20]: ($\bar{x} \varepsilon \text{ int}(C)$).

Let $y \varepsilon R^n$, $\nabla g_A(\bar{x})y \leq 0$, $\nabla h(\bar{x})y = 0$.

Qualification: there is a function $e:[0,1] \to R^n$ such that

(a) $e(0) = \bar{x}$

(b) $e(\theta)$ is feasible for P' for all $\theta \varepsilon [0,1]$

(c) e is differentiable at $\theta = 0$ and $e'(0) = \lambda y$ for some real number $\lambda > 0$.

(ii) Tangent cone qualification [32]: ($\bar{x} \in \text{int}(C), h(x) \equiv 0$ (i.e. no equality constraints))

Let $C(\bar{x}) = \{y \in R^n$: there are x^k $k=1,2,\ldots$ feasible for P' and

$\varepsilon_k > 0$, $\varepsilon_k \in R$ such that $x^k \to \bar{x}$, $\frac{1}{\varepsilon_k}(x^k - \bar{x}) = y\}$

be the tangent cone to P' at $\bar{x}$ and let $D(\bar{x}) = \{y \in R^m: \nabla g_A(\bar{x})y \leq 0\}$.

Qualification: $C(\bar{x}) = D(\bar{x})$.

(iii) Arrow-Hurwicz-Uzawa constraint qualification [2]: ($\bar{x} \in \text{int}(C)$, $h(x) \equiv 0$)

Let $Q = \{j \in A$: for all $x \in C$, $\nabla g_j(\bar{x})(x-\bar{x}) \leq 0$ implies that $g_j(x) \leq g_j(\bar{x}) = 0\}$

and $R = A - Q$.

Qualification: There is a $z \in R^n$ satisfying $\nabla g_Q(\bar{x})z \leq 0$

$$\nabla g_R(\bar{x})z < 0.$$

(Note that linear equations may be incorporated into $g_Q(\bar{x})$.)

(iv) Slater condition (C convex, f and each g_j convex, and $h(x) \equiv 0$)

Condition: there is an $x \in C$ with $g(x) < 0$.

(v) Karlin's constraint qualification [18]: (C convex, each g_j convex, $h(x) \equiv 0$)

Qualification: there is no $\pi \in R^m$, $\pi \geq 0$, $\pi \neq 0$ such that $\pi g(x) \geq 0$ for all $x \in C$.

(vi) Nonsingularity conditions:

Qualification ($h(x) \equiv 0$, $C = R^n$): $\nabla g_j(\bar{x})$ $j=1,\ldots,m$ are linearly independent.

Lagrange's condition ($g(x) \equiv 0$, $C = R^n$): $\nabla h_j(\bar{x})$ $j=1,\ldots,r$ are linearly independent.

Theorem 7: Each of (i)-(vi) implies the existence of a Kuhn-Tucker vector at $\bar{x}$.

Proof: By lemma 6, for (i)-(iii), it suffices to show that P' is regular at $\bar{x}$. That is for (i) assume that y solves $\nabla g_A(\bar{x})y \leq 0$, $\nabla h(\bar{x})y = 0$ and for (ii) and (iii) assume that $\nabla g_A(\bar{x})y \leq 0$. We must show that $\nabla f(\bar{x})y \geq 0$.

(i) By the chain rule, $\left.\frac{df[e(\theta)]}{d\theta}\right|_{\theta=0} = \nabla f[e(0)]\ e'(0) = \lambda\nabla f(\bar{x})y$ for some $\lambda > 0$. But then if $\nabla f(\bar{x})y < 0$, $f[e(\theta)] < f[e(0)] = f(\bar{x})$ for $\theta\varepsilon[0,1]$ small enough. Since $e(\theta)$ is feasible for P', this contradicts $\bar{x}$ optimal.

(ii) By definition of derivatives, given $\varepsilon > 0$ there is a K_ε such that for $k \geq K_\varepsilon$, $\left|\left|\frac{1}{\varepsilon_k}[f(x^k)-f(\bar{x})] - \nabla f(\bar{x})(\frac{x^k-\bar{x}}{\varepsilon_k})\right|\right| \leq \varepsilon \left|\left| \frac{x^k-x}{\varepsilon_k} \right|\right|$

By the qualification, $\frac{x^k-x}{\varepsilon_k} \to y$ and this inequality implies that if $\nabla f(\bar{x})y < 0$ then $f(x^k) < f(\bar{x})$ for k large enough. But since x^k is feasible for P' this contradicts $\bar{x}$ optimal.

(iii) The qualification implies that

$$\nabla g_Q(\bar{x})[y+ \theta z] \leq 0$$
$$\nabla g_R(\bar{x})[y+ \theta z] < 0$$

for any $\theta > 0$. By definition of Q and (*) this implies that $g_A[\bar{x}+\alpha(y+\theta z)] \leq g_A(\bar{x}) \leq 0$ for all $0 < \alpha < \bar{\alpha}$ some $\bar{\alpha}$. Also since $\bar{x}\ \varepsilon\ \text{int}(C)$, $\bar{x} + \alpha(y+ \theta z)\varepsilon C$ if $\bar{\alpha}$ is small enough. Finally, if $\nabla f(\bar{x})y < 0$ then $\nabla f(\bar{x})[y+ \theta z] < 0$ for $0 < \theta \leq \bar{\theta}$ for some $\bar{\theta}$. But this again implies that if $\bar{\alpha}$ is small enough, $f(\bar{x} + \alpha[y+ \theta z]) < f(\bar{x})$, i.e., $\bar{x} + \alpha[y+ \theta z]$ is feasible for P' and contradicts $\bar{x}$ optimal.

(iv) By Corollary 1.2, there is a $\hat{\pi} \geq 0$ such that $f(x) + \hat{\pi}g(x) \geq v = f(\bar{x})$ for all $x\varepsilon C$. Since $g(\bar{x}) \leq 0$, $f(\bar{x}) + \hat{\pi}g(\bar{x}) \leq f(\bar{x})$ thus equality holds and $\hat{\pi}g(\bar{x}) = 0$. Also $\bar{x}$ solves $\min_{x\varepsilon C} [f(x) + \hat{\pi}g(x)]$.

But this implies that $[\nabla f(\bar{x}) + \hat{\pi}\nabla g(\bar{x})](x-\bar{x}) \geq 0$ for all $x\varepsilon C$, i.e., $\hat{\pi}$ is a Kuhn-Tucker vector at $\bar{x}$.

(v) The problem $v = \inf_{\substack{x\varepsilon C\\ \sigma\varepsilon R}} \{\sigma\colon g(x) \leq \sigma e^m\}$ satisfies the Slater condition.

Thus by saddlepoint duality $v = \inf_{x\varepsilon C} [\pi g(x)]$ for some $\pi \geq 0$, $\pi e^m = 1$

the last condition being implied by $\sigma\varepsilon R$ in the dual. By the qualification, though, $\pi g(x) < 0$ for some $x\varepsilon C$. Thus $v < 0$, i.e., P' satisfies the Slater condition and (iv) applies.

(vi) If the $\nabla g_j(\bar{x})$ are linearly independent, then $\pi\nabla g(\bar{x}) = 0$, $\pi \geq 0$, $\pi e^m = 1$ has no solution. LP duality (take objective $\equiv 0$ with the constraints above) implies that sup $\{\sigma\colon \nabla g(\bar{x})y + \sigma e^m \leq 0\}$ is $+\infty$, i.e., there is a y with $\nabla g(\bar{x})y < 0$. Then apply (iii).

The proof for Lagrange's condition utilizes the implicit function theorem and is similar, though easier, than the proofs in the next section. It is omitted here. ///

For discussions of yet further sufficient conditions for the existence of Kuhn-Tucker vectors, see [22]. As a final observation in this section, we show that the Arrow-Hurwicz-Uzawa constraint qualification implies the Kuhn-Tucker constraint qualification. Abadie [1] has previously given a partial proof of this result.

<u>Lemma 7</u>: The Arrow-Hurwicz-Uzawa constraint qualification implies the Kuhn-Tucker constraint qualification.

<u>Proof</u>: Let $y\varepsilon C-\bar{x}$ satisfy $\nabla g_A(\bar{x})y \leq 0$, and let z satisfy A-H-U. For any $\theta > 0$, $\nabla g_R(\bar{x})[z+\theta y] < 0$, thus by (*) there is an $\bar{\alpha} > 0$ such that

$$g_j(\bar{x} + \alpha[z+\theta y]) < g_j(\bar{x}) \leq 0 \text{ for all } j\varepsilon R, \text{ for all } 0 \leq \alpha \leq \bar{\alpha}.$$

Thus for $j\varepsilon R$, $g_j(\bar{x} + \alpha[\theta y + \theta^2 z]) \leq 0$ for all $\alpha\varepsilon[0,\bar{\alpha}]$, for all $\theta\varepsilon[0,1]$. Also $\nabla g_Q(\bar{x})[\alpha\theta y + \alpha\theta^2 z] \leq 0$ so $g_j(\bar{x} + \alpha[\theta y + \theta^2 z]) \leq 0$, for all $j\varepsilon Q$. Finally, since $\bar{x} \varepsilon \operatorname{int}(C)$ there is an $\bar{\bar{\alpha}} > 0$ such that in $\bar{x} + \alpha[\theta y + \theta^2 z] \varepsilon C$ for $0 \leq \alpha \leq \bar{\bar{\alpha}}$.

Let $e_y(\theta) = \bar{x} + \alpha[\theta y + \theta^2 z]$ with $\alpha \leq \min(\bar{\alpha}\ \bar{\bar{\alpha}})$ and small enough that $g_j[e_y(\theta)] \leq 0$ for all $j \notin A$ and $\theta\varepsilon[0,1]$. Then $e_y(0) = \bar{x}$, $e'(0) = \alpha y$ and from above, $e_y\colon [0,1] \to C \cap \{x\colon g(x) \leq 0\}$. ///

<u>Remark 6:</u> Conditions (i), (ii) and (iii) can be modified by assuming that C is convex, and not necessarily that $\bar{x}\varepsilon\mathrm{int}(C)$. As above the conditions imply regularity at $\bar{x}$, thus by corollary 5.1 under any of these conditions there is a Kuhn-Tucker vector at $\bar{x}$ if L is stable. Additionally, lemma 7 will hold for the modified Arrow-Hurwicz-Uzawa and Kuhn-Tucker constraint qualifications. The proof is modified by noting:

$$\bar{x} + \alpha[\theta x + \theta^2 z] = [1 - \alpha\theta - \alpha\theta^2]\bar{x} + (\alpha\theta)(\bar{x}+y) + (\alpha\theta^2)(\bar{x}+z)\varepsilon C$$

if $0 \leq \alpha \leq \bar{\bar{\alpha}} = 1/2$ (so that $1 - \alpha\theta - \alpha\theta^2 \geq 0$).

B. Fritz John Conditions

If x is partitioned as $x = \begin{pmatrix} y \\ z \end{pmatrix}$, let $\nabla_y h(\bar{x}) \equiv [\frac{\partial h(x)}{\partial y_1}, \ldots, \frac{\partial h(x)}{\partial y_k}]\Big|_{x=\bar{x}}$ and similarly define $\nabla_z h(\bar{x})$.

Remark 7: Let $h(\bar{x}) = 0$, $\bar{x} = \begin{pmatrix} \bar{y} \\ \bar{x} \end{pmatrix}$. Suppose h has continuous first partial derivatives in a neighborhood of $\bar{x}$ and that $\nabla_y h(\bar{x})$ is non-singular. Then by the implicit function theorem there is an open set $\Gamma \subseteq R^{n-r}$ with $\bar{z} \in \Gamma$ and a function $I:\Gamma \to R^r$ such that with $y = I(z)$ and $\bar{y} = I(\bar{z})$, $h(y,z) = 0$ for $z \in \Gamma$. Furthermore, I has continuous first partial derivatives on Γ and satisfies:

FACT: Assume $\nabla h(\bar{x})v = 0$. Let $\nabla h(\bar{x})v = \nabla_y h(\bar{x})v_y + \nabla_z h(\bar{x})v_z$, where the partitioning $v = \begin{pmatrix} v_y \\ v_z \end{pmatrix}$ of v conforms with that of x. Also, for all θ such that $z + \theta v_z \in \Gamma$ let $\phi(\theta) = I(\bar{z} + \theta v_z)$. Then $\phi'(0) = v_y$.

Proof: From above, $h[\phi(\theta), \bar{z}+\theta v_z] = 0$ for all θ such that $\bar{z}+\theta v_z \in \Gamma$ so $\frac{d}{d\theta} h[\phi(\theta), \bar{z}+\theta v_z]\Big|_{\theta=0} = 0$. Thus by the chain rule $\nabla_y h(\bar{x})\phi'(0) + \nabla_z h(\bar{x})v_z = 0$ (**). But since $\nabla_y h(\bar{x})$ is non-singular the solution for $\phi'(0)$ in (**) is unique, thus by hypothesis equal to v_y. ///

Armed with this fact about implicit functions, our proofs are quite easy. The next result and approach taken here is a shortened version of Mangasarian and Fromowitz's development [22], [23]:

Lemma 7: (Linearization Lemma) Let $C \subseteq R^n$ be convex; let h and g satisfy the hypothesis of P' and assume that h has continuous first partial derivatives in a neighborhood of $\bar{x}$.

$$\nabla h(\bar{x})v = 0$$
$$\text{If} \quad \nabla g(\bar{x})v < 0$$
$$\bar{x} + v \in \text{int}(C)$$

has a solution and the vectors $\nabla h_j(\bar{x})$ $j=1,\ldots,r$ are linearly independent, then there is an $x' \in \text{int}(C)$ arbitrarily close to $\bar{x}$ with $h(x') = 0$ and $g(x') < g(\bar{x})$.

Proof: $\nabla h_j(\bar{x})$ linearly independent implies that there is a partitioning of x, $\bar{x} = (\bar{y}, \bar{z})$ with $\nabla_y h(\bar{x})$ non-singular. Using the notation in Remark 6 above, define $x(\theta) = [\phi(\theta) - \phi(0), \theta v_z]$. By the fact $\frac{x(\theta)}{\theta} \to v$ as $\theta \to 0$ and $h(x(\theta) + \bar{x}) = 0$ in a neighborhood of $\theta = 0$. By the definition of derivative, given $\varepsilon > 0$, for each j

$$||\{g_j(x(\theta) + \bar{x}) - g_j(\bar{x})\} - \nabla g_j(\bar{x})x(\theta)|| \leq \varepsilon \; ||\, x(\theta)||$$

for $|\theta| \leq$ some constant $K_{\varepsilon,j}$. Thus

$$||\tfrac{1}{\theta}\{g_j(x(\theta) + \bar{x}) - g_j(\bar{x})\} - \nabla g_j(\bar{x})\frac{x(\theta)}{\theta}|| \leq \varepsilon \; || \; \frac{x(\theta)}{\theta}||$$

and so since $\frac{x(\theta)}{\theta} \to v$, $\nabla g(\bar{x})[\frac{x(\theta)}{\theta}] \to \nabla g(\bar{x})v < 0$ as $\theta \to 0$, and

$$g_j(x(\theta) + \bar{x}) < g_j(\bar{x}) \text{ for } \theta > 0 \text{ small enough.}$$

Finally, since $\bar{x} + \frac{x(\theta)}{\theta} \underset{\theta\to 0}{\to} \bar{x} + v \;\varepsilon$ int (C), $\bar{x} + \frac{x(\theta)}{\theta} \;\varepsilon$ int (C) for θ small enough. But if $x + \frac{x(\theta)}{\theta} \;\varepsilon$ int (C) and $0 < \theta < 1$

$$\bar{x} + x(\theta) = \theta[\bar{x} + \frac{x(\theta)}{\theta}] + (1-\theta)\bar{x} \;\varepsilon \text{ int (C) since C is convex.}$$

Putting everything together: $\bar{x} + x(\theta) \;\varepsilon$ int (C)

$$h(\bar{x} + x(\theta)) = 0, \; g(\bar{x} + x(\theta)) < g(\bar{x})$$

for $\theta>0$ small enough. Let $x' = \bar{x} + x(\theta)$. ///

Remark 7: An analogous result holds when $\nabla h(\bar{x}) \equiv 0$, i.e., there are no equality constraints above.

Theorem 8: Consider problem P', suppose that C is convex and has a non-empty interior, and that h has continuous first partial derivatives in a neighborhood of $\bar{x}$. Then there is a Fritz John vector at $\bar{x}$.

Proof: If $\nabla h_j(\bar{x})$ are linearly dependent, then there is an $\alpha \neq 0$ such that $\alpha\nabla h(\bar{x}) = 0$. Letting $\pi_0 = 0$ and $\pi = 0$, (π_0,π,α) is a Fritz John vector. If the $\nabla h_j(\bar{x})$ are linearly independent (or $h(x) \equiv 0$), then since $\bar{x}$ is optimal, the previous lemma (for θ small enough in Lemma 7 $g_j(\bar{x} + x(\theta)) < 0$ for $j \notin A$, thus these constraints may be ignored) implies that $\hat{\sigma} = 0$ where

$$\hat{\sigma} = \min -\sigma$$

$$\text{s.t.} \quad \nabla h(\bar{x})v = 0$$

$$\nabla g_A(\bar{x})v + \sigma e^m \leq 0$$

$$\nabla f(\bar{x})v + \sigma \leq 0$$

$$\bar{x} + v \in \text{int } (C)$$

has no solution, with $\hat{\sigma} < 0$.

Note that $(\sigma,v) = (0,0)$ is feasible here, thus $\hat{\sigma} \leq 0$.

By the hypothesis and Lemma 3 of section I, this optimization problem satisfies the generalized Slater Condition. Thus by Corollary 1A.1 there is a vector (π_0,π_A,α), $(\pi_0,\pi_A) \geq 0$ such that

$$\inf_{\substack{\sigma \in R \\ v \in \text{int}(C)-\bar{x}}} \{(\pi e^m + \pi_0 - 1)\sigma + [\pi_0 \nabla f(\bar{x}) + \pi \nabla g_A(\bar{x}) + \alpha \nabla h(\bar{x})]v\} = 0$$

But then $\pi e^m + \pi_0 = 1$ and

$$[\pi_0 \nabla f(\bar{x}) + \pi_A \nabla g_A(\bar{x}) + \alpha \nabla h(\bar{x})]v \geq 0 \quad \text{for all } v+\bar{x} \in \text{int } (C).$$

Letting $\pi_j = 0$ for $j \notin A$ and noting that this relationship is continuous in v implies

$$[\pi_0 \nabla f(\bar{x}) + \pi \nabla g(\bar{x}) + \alpha \nabla h(\bar{x})](x-\bar{x}) \geq 0 \quad \text{for all } x \in \text{closure of } C. \quad ///$$

IV. Concluding Remarks

In this section, we give some brief remarks relating our results with previous work and indicate some possible extensions.

Our approach to the saddlepoint theory is related to Dantzig's generalized programming approach to solve P [5, chapter 24]. There he generates the x^j not as a dense subset of C, but rather by solving a Lagrangian subproblem.

His results furnish Theorem 1A when C is convex and compact, f is convex on C, each g_j is convex on R^n, and the h_i are affine. Note that whereas he develops a computational procedure for solving P, we do not.

Previous proofs of Theorem 3 [13], [25] have been given when v(·) is a convex function. We have not considered this case, but conjecture that convexity of v(·) implies that P is regular. We also believe that theorem 3A can be obtained from our results by approximating P by problems that have an optimal solution and applying theorem 3 to these approximations.

In [15], Gould and Tolle give an alternate development of necessary and sufficient conditions for optimality conditions. They consider: when is there a Kuhn-Tucker vector at $\bar{x}$ for every function f such that P has a local minimum at $\bar{x}$?

As to extensions, we note that our approach will provide what seems to be an almost unlimited variety of dual forms and suggests a number of additional theorems. For example, for j=1,...,r let $C_j \subseteq R^n$ and $f_j : C_j \to R$, let $g_j : R^n \to R^m$, and let $\bar{C} \subseteq R^n$. Consider the composite problem:

$$\inf_{x \in \bar{C}} \sum_{j=1}^{r} f_j(x)$$

$$\text{s.t.} \quad x \in \bigcap_{j=1}^{r} C_j$$

$$g(x) \leq 0 \quad .$$

By replacing C with $C_1 x \ldots x C_r x \bar{C}$ and introducing $g(\bar{x}) \leq 0$ into $\bar{P}$ from section II.B, we easily see that the problem has the dual

$$\sup_{\substack{\pi^j \in R^n \\ \pi \in R^m}} \{\inf_{x \in \bar{C}} [\pi g(x) + (\sum_{1}^{r} \pi^j)x] + \sum_{j=1}^{r} \inf_{x \in C_j} [f_j(x) + \pi^j x]\} \; .$$

By using $g(x^j) \leq 0$ for some j in $\bar{P}$, we obtain a different dual. As another example, consider the following extended version of Rockafeller's dual:

$$\inf \sum_{j=1}^{r} [f_j(A^j x)] \qquad A^j \text{ is an } m_j \text{ by } n \text{ matrix}$$

$$\text{s.t.} \quad A^j x \in C_j \qquad j=1,\ldots,r \qquad C_j \subseteq R^{m_j}$$

Arguing as before, its dual is

$$\sup_{\sum_1^r \pi^j A^j = 0} \sum_{j=1}^{r} \inf_{x \in C_j} [f_j(x) + \pi^j x] \quad .$$

Finally, we note that the techniques of III.B can be applied to give Kuhn-Tucker conditions for problems with equality constraints. For details, see [22].

Acknowledgement: I am grateful to Gabriel Bitran for his careful reading of a preliminary manuscript of this paper and for his valuable suggestions to improve the exposition.

APPENDIX

If C is convex and $f:C \to R$ is a convex function, then it is well known [22] that f is continuous on ri(C). In particular, if $C = R^n$ then f is continuous. Consequently, each g_j in P is continuous and to prove (R2) from section II.A as asserted in Remark 2, it suffices to establish the condition only for f. This also is well known but often misquoted; for completeness we provide a proof here of a somewhat stronger result.

<u>Lemma</u>: Let $x^1, x^2, \ldots$ be a dense set of points from the convex set C and let $f:C \to R$ be a convex function on C. Given any $y \in C$ there is a subsequence $\{k_j\}$ of $\{1,2,\ldots\}$ such that $x^{k_j} \to y$ and $f(y) \geq \overline{\lim}\, f(x^{k_j})$. ($\overline{\lim}$ denotes limit superior [26]).

<u>Proof</u>: If C is a single point, there is nothing to prove. Otherwise, let z be contained in the relative interior of C and for $j \in \{1,2,\ldots\}$ let $z^j = (\frac{1}{j})z + (\frac{j-1}{j})y$. Then $z^j \to y$ and $f(y) \geq \overline{\lim}\, f(z^j)$ since by convexity $f(z^j) \leq (\frac{1}{j})f(z) + (\frac{j-1}{j})f(y)$. Also $z^j \in ri(C)$. Since f is continuous on ri(C) there is an x^{k_j} with $||x^{k_j} - z^j|| < (\frac{1}{j})$ and $|f(x^{k_j}) - f(z^j)| < (\frac{1}{j})$. Then $x^{k_j} \to y$ and $\overline{\lim}\, f(x^{k_j}) \leq f(y)$. ///

Note that this lemma does not say that if $x^{k_j} \to y$ then $\overline{\lim}\, f(x^{k_j})$ or even $\underline{\lim}\, f(x^{k_j}) \leq f(y)$. As a counter-example to this assertion let

$$C = \{x \in R^2 : x_1^2 + x_2^2 \leq 1\} \text{ and for } x \in C \text{ let}$$

$$f(x) = \begin{cases} 0 \text{ for } x_1^2 + x_2^2 < 1 \\ 1/2 \text{ for } x = (1,0) \\ 1 \text{ otherwise.} \end{cases}$$

Letting $x^{k_j} \in \{x : x_1^2 + x_2^2 = 1\} - \{(1,0)\}$, $x^{k_j} \to (1,0)$ gives $\lim f(x^{k_j}) = 1$ with $f((1,0)) = 1/2$.

REFERENCES

1. J. Abadie, "On the Kuhn-Tucker Theorem," in Nonlinear Programming, J. Abadie (ed.), North Holland Publishing Company, Amsterdam, 1967.

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

3. M. Cannon, C. Cullum and E. Polak, "Constrained Minimization in Finite Dimensional Spaces," SIAM J. on Control 4, 528-547 (1966).

4. R. Cottle, "Symmetric Dual Quadratic Programs," Quart. Appl. Math., 21, 237 (1963).

5. G. Dantzig, Linear Programming and Extensions, Princeton University Press, Princeton, New Jersey, (1963).

6. G. Dantzig, E. Eisenberg and R. Cottle, "Symmetric Dual Nonlinear Programs," Pacific J. Math., 15, 809-812 (1965).

7. J. Dennis, Mathematical Programming and Electrical Networks, Technology Press, Cambridge, Mass. (1959).

8. W. Dorn, "Duality in Quadratic Programming," Quart. Appl. Math., 18, 155-162 (1960).

9. W. Fenchel, "Convex Cones, Sets and Functions," mimeographed lecture notes, Princeton University (1951).

10. M. Fisher and J. Shapiro, "Constructive Duality in Integer Programming," Tech. Report OR 008-72, Operations Research Center, M.I.T., Cambridge, Mass. (April 1972).

11. A. Fiacco and G. McCormick, Sequential Unconstrained Minimization Techniques for Nonlinear Programming, John Wiley and Sons, New York(1968)

12. D. Gale, "A Geometric Duality Theorem with Economic Applications," Review of Economic Studies, XXXIV, 97, 19-24 (1967).

13. A. Geoffrion, "Duality in Nonlinear Programming: A Simplified Applications-Oriented Development," SIAM Review 13, 1-37, 1971.

14. F. Gould, "Extensions of Lagrange Multipliers in Nonlinear Programming," SIAM J. on Appl. Math., 17, 1280-1297 (1969).

15. F. Gould and J. Tolle, "A Necessary and Sufficient Qualification for Constrained Optimization," SIAM J. on Appl. Math., 20, 164-172 (1971).

16. F. John, "Extremum Problems with Inequalities as Subsidiary Conditions, Studies and Essays," Courant Anniversary Volume, Interscience, New York (1948).

17. H. Halkin, "A Maximum Principle of the Pontryagin Type For Systems Described by Nonlinear Difference Equations," SIAM J. on Control, 4, 90-111 (1966).

18. S. Karlin, "Mathematical Methods and Theory in Games, Programming and Economics," Vol. I, Addison-Wesley, Reading, Mass. (1959).

19. K. Kortanek and J. Evans, "Asymptotic Lagrange Regularity for Pseudo-concave Programming with Weak Constraint Qualification," Opns. Res., 16, 849-857 (1968).

20. H. Kuhn and A. Tucker, "Nonlinear Programming" in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 481-492 (1951).

21. D. Luenberger, Optimization by Vector Space Methods, John Wiley & Sons, New York (1969).

22. O. Mangasarian, Nonlinear Programming, McGraw-Hill, New York (1969).

23. O. Mangasarian and S. Fromowitz, "The Fritz John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints," J. Math. Anal. and Appl., 17, 37-47 (1967).

24. O. Mangasarian and J. Ponstein, "Minimax and Duality in Nonlinear Programming," J. Math. Anal. and Appl., 11, 504-518 (1965).

25. R.T. Rockafeller, Convex Analysis, Princeton University Press, Princeton, New Jersey, (1970).

26. W. Rudin, Principles of Mathematical Analysis, McGraw-Hill, New York (1953).

27. J. Shapiro, "Generalized Lagrange Multipliers in Integer Programming," Opns. Res., 19, 68-76 (1971).

28. M. Simonnard, Linear Programming, Prentice-Hall, Englewood Cliffs, N.J. (1966).

29. M. Slater, "Lagrange Multipliers Revisited: A Contribution to Nonlinear Programming," Cowles Commission Discussion Paper, Math. 403 (1950).

30. J. Stoer, "Duality in Nonlinear Programming and the Minimax Theorem," Numer. Math., 5, 371-379 (1963).

31. J. Stoer and C. Witzgal, Convexity and Optimization in Finite Dimensions, Springer-Verlag, Berlin (1970).

32. P. Varaiya, Notes on Optimization, Van Nostrand Reinhold, New York (1972).

33. P. Wolfe, "A Duality Theorem for Non-Linear Programming," Quart. Appl. Math., 14, 239-244 (1961).

34. E. Zangwill, Nonlinear Programming: A Unified Approach, Prentice-Hall, Englewood Cliffs, N.J. (1969).