PRIMAL AND DUAL OPTIMALITY CRITERIA IN CONVEX PROGRAMMING

by

A. Ben-Tal and A. Charnes

Technical Report No.66
December 1975

*) Director, CENTER FOR CYBERNETIC STUDIES, The University of Texas at Austin, Austin, Texas 78712, USA.

This research was partly supported by Project No. NR 047-021, ONR Contract N00014-75-C-0569 with the Center for Cybernetic Studies, The University of Texas. Reproduction in whole or in part is permitted for any purpose of the United States Government.
Abstract

This paper considers the problem of minimizing a convex differentiable function subject to convex differentiable constraints. Necessary and sufficient conditions (not requiring any constraints qualification) for a point to be an optimal solution are given in terms of a parametric linear program. Dual characterization theorems are then derived, which generalizes the classical results of Kuhn-Tucker and Fritz John.
1. INTRODUCTION

In this paper we consider the problem (P) of minimizing a convex differentiable function \( f^0: \mathbb{R}^n \to \mathbb{R} \) subject to finite number of constraints,

\[
\begin{align*}
    f^k(x) &\leq 0, \quad k \in \{1, 2, \ldots, p\}
\end{align*}
\]

where each of the functions \( f^k \) is convex and differentiable.

Necessary and sufficient conditions for a feasible point to be an optimal solution of (P) were given recently [3]. These conditions (termed in the sequel BBZ), unlike the well-known Kuhn-Tucker (KT) conditions [6], do not depend on the customary assumption of constraints qualification.

The BBZ conditions are stated in terms of consistency, or inconsistency, of a family of systems involving linear inequalities and cone relations. In this paper we derive a different form of the BBZ conditions. The new characterization (Theorem 1) is given in terms of a parametric linear program. Here by a "linear program" we mean a problem of the form

\[
\text{(LP)} \quad \text{Min} \ \{c^t x: Ax = b, \ x \in S\}
\]

where \( A \) is a matrix and \( S \) is a convex cone.

The main advantage of the new formulation is the possibility of dualizing the parametric linear program, thus obtaining "dual characterization theorems" which generalize the KT-conditions, on the one hand (Theorem 3) and the
Fritz John conditions [5] on the other hand (Theorem 4). Those results are obtained in Section 3. It is shown also that the KT conditions are obtained as a limit case of the parametric formulation when the parameter decreases to zero (Corollary 1.1).

Unlike the primal characterization, the dual characterization is not necessarily valid for any convex programming problem, but its applicability is rather large and perhaps sufficient for most practical situations. In particular this characterization (Theorem 2) is valid if the constraint functions are analytical or strictly convex in their actual variables. For the latter case the characterization (given in Corollary 2.1) is especially simple and perhaps more open to interpretations.

Duality theorems for linear programs such as (LP) were originated by Duffin in [4] and were refined later by Ben-Israel, Charnes and Kortanek [2].

2. PRIMAL CHARACTERIZATION THEOREM

Consider the programming problem

\[
\begin{align*}
(P) & \quad \text{Min } f^0(x) \\
    \text{s.t.} & \quad f^k(x) \leq 0, \ k \in P \setminus \{1, 2, \ldots, p\}.
\end{align*}
\]

where \( f^k \) is a function: \( \mathbb{R}^n \to \mathbb{R} \), \( k \in \{0\} \cup P \).

For a feasible solution \( x^* \), i.e. \( f^k(x^*) \leq 0, \ k \in P \), we denote the set of
binding constraints by \( P^* \triangleq \{k: f^k(x^*) = 0, k \in P\} \).

Let also
\[
p^* \triangleq \text{cardinality of } P^*.
\]

For \( k \in P \) we define \( D_k(x^*) \), the cone of directions of constancy of \( f^k \) at \( x^* \), as follows
\[
D_k(x^*) \triangleq \{d: \exists \alpha > 0 \exists f^k(x^* + \alpha d) = f^k(x^*), \forall \alpha \in [0, \alpha]\}.
\]

Whenever it is clear from the context, we will write \( D_k \) for \( D_k(x^*) \).

It was shown in [3] that, for \( f^k \) differentiable and convex, \( D_k \) is a convex cone contained in the subspace \( \{d: d^t \nabla f^k(x^*) = 0\} \).

The following theorem gives necessary and sufficient condition for optimality in terms of the "parametric linear program" \((G, \theta)\) below.

**Theorem 1.** Let problem \((P)\) have differentiable convex functions \( \{f^k: k \in \{0\} \cup P\} \), and let \( x^* \) be a feasible solution of \((P)\). Then \( x^* \) is optimal if, and only if, there exists a positive scalar \( \theta^* \) such that the program \((G, \theta)\) below has optimal value zero for every \( \theta \in (0, \theta^*) \).

\[
\begin{align*}
(G, \theta) & \quad \text{Min } d^t f^0(x^*) \\
\text{s.t.} & \\
(1, \theta) & \quad d^t \nabla f^k(x^*) + \theta \sum_{i=1}^{n} |d_i - \delta_i^k| \leq 0, k \in P^* \\
(2) & \quad \delta^k \triangleq (\delta_1^k, \delta_2^k, \ldots, \delta_n^k)^t \in D_k(x^*), k \in P^*.
\end{align*}
\]
Proof. We use the following result (for details see the proof of Theorem 1 in [3]):

\[
\begin{align*}
\begin{cases}
    x^* \text{ is not optimal if and only if there exists a} \\
    \text{subset } \tilde{\Omega} \subset P^* \text{ and a vector } d \in \mathbb{R}^n \text{ such that} \\
    d^T \nabla f^0(x^*) < 0 \\
    d^T \nabla f^k(x^*) < 0, \ k \in \tilde{\Omega} \\
    d \in D_k(x^*), \quad k \in P^*/\tilde{\Omega}.
\end{cases}
\end{align*}
\]

We denote below by \(d(\theta), \{\delta^k(\theta): k \in P^*\}\) an optimal solution of \((G, \theta)\).

Let \(x^*\) be non-optimal, then, for some \(\tilde{\Omega} \subset P^*\), there exist \(\tilde{d} \in \mathbb{R}^n\) satisfying (3) - (5). Define

\[
\tilde{d} \Delta \begin{cases}
    \min \left\{ \frac{-d^T \nabla f^k(x^*)}{\sum_{i=1}^{n} |d_i|} \right\} & \text{if } \tilde{\Omega} \neq \emptyset \\
    1 & \text{if } \tilde{\Omega} = \emptyset
\end{cases}
\]

Then

\(\tilde{d} > 0\).

Let

\[
\delta^k = \begin{cases}
    \tilde{d} & \text{if } \tilde{d}^T \nabla f^k(x^*) = 0 \\
    0 & \text{if } \tilde{d}^T \nabla f^k(x^*) < 0.
\end{cases}
\]
Then \( \tilde{d} \) and \( \{ \tilde{\delta}^k: k \in P^* \} \) satisfy (1,\( \tilde{\theta} \)) and (2). Since \( \tilde{d} \) also satisfies (3) it follows that
\[
\tilde{d}(\tilde{\theta})^T \nabla f^0(x^*) < 0.
\]

For \( \theta \in (0,\tilde{\theta}] \) the vectors \( d(\tilde{\theta}), \{ \delta^k(\tilde{\theta}), k \in P^* \} \) still form a feasible solution for \((G,\theta)\) and hence
\[
d(\theta)^T \nabla f^0(x^*) < 0; \forall \theta \in (0,\tilde{\theta}]
\]

contradicting the existence of \( \theta^* > 0 \) for which
\[
d(\theta)^T \nabla f^0(x^*) = 0; \forall \theta \in (0,\theta^*].
\]

**Only if**

Suppose that for some \( \theta > 0 \), the problem \((G,\theta)\) has negative optimal value corresponding to an optimal solution \( \tilde{d}, \{ \tilde{\delta}^k: k \in P^* \} \) i.e.
\[
\tilde{d}^T \nabla f^0(x^*) < 0
\]
(7)
\[
\tilde{d}^T \nabla f^k(x^*) + \theta \sum_{i=1}^{n} |\tilde{d}_i| - |\tilde{\delta}^k_i| \leq 0, \quad k \in P^*
\]
(8)
\[
\tilde{\delta}^k \in D_k
\]

Let
\[
\tilde{\Omega} = \{ k: \tilde{d}^T \nabla f^k(x^*) < 0 \}.
\]

By the nature of the constraint (7) it then follows that \( k \in P^* / \tilde{\Omega} \Rightarrow \nabla f^k(x^*) = 0 \Rightarrow \tilde{d} = \tilde{\delta}^k \Rightarrow \tilde{\delta} \in D_k \), by (8).
Thus, \( \tilde{d} \) and \( \tilde{\Omega} \) satisfy (3) - (5), contradicting the optimality of \( x^* \).

For \( \theta \geq 0 \) let \( d(\theta), \{\delta^k(\theta), k \in P^*\} \) denote an optimal solution of \( (G, \theta) \) and let

\[
\mathbf{(9)} \quad z^*(\theta) \triangleq d(\theta)^t \nabla f^O(x^*)
\]

be the corresponding optimal value of \( (G, \theta) \). In fact \( z^*(\theta) \) is equal either to zero or to \( (-\infty) \), but, adding the constraints

\[
|d_i| \leq 1 \quad i = 1, \ldots, n
\]

to \( (G, \theta) \) Theorem 1 remains true, and \( z^*(\theta) \) is bounded below by

\[
- \sum_{i=1}^{n} \left| \frac{\partial}{\partial x^*_i} f^O(x^*) \right|.
\]

In terms of the "optimal value function" \( z^*(\theta) \), the relation of the optimality conditions, given in Theorem 1 to the classical Kuhn-Tucker conditions are expressed in the following

**Corollary 1.1** The consistency of the Kuhn-Tucker system

\[
(KT) \quad \left\{ \begin{array}{l}
\nabla f^O(x^*) + \sum_{k \in P^*} y_k \nabla f^k(x^*) = 0 \\

y_k \geq 0, \quad k \in P^*
\end{array} \right.
\]

is a necessary condition for \( x^* \) to be an optimal solution of \( (P) \) if, and only if,

\[
\lim_{\theta \to 0^+} \inf z^*(\theta) = z^*(0).
\]
Proof. First note that by Motzkin's Theorem of the alternative the consistency of (KT) is equivalent to the inconsistency of the system

\[ a^t \vee f^0(x^*) < 0 \]
\[ a^t \vee f^k(x^*) \leq 0, \ k \in P^* \]

or equivalently, to zero being the optimal value of the problem (G,0) (note that for this program the constraints (2) are redundant) i.e. \( z^*(0) = 0 \).

On the other hand, Theorem 1 just states that

\[ \lim_{\theta \to 0^+} \inf z^*(\theta) = 0. \]

Hence the conclusion of the corollary follows.

\[ \square \]

3. DUAL CHARACTERIZATIONS

We start this section with some preliminary results on linear programs of the type (LP). In fact, for our purposes, it suffices to study only "homogenized programs"

\[ (L) \inf \{c^t x : Ax = 0, \ x \in S\} \]

and their corresponding "linear inequality" system

\[ (D) \ y^t A + c \in S^+, \ y \in R^m \]
where $S^+$ is the polar cone of the convex cone $S$, i.e.

$$S^+ = \{ z : x \in S \Rightarrow z^T x \geq 0 \}$$

we let $m$ denote the optimal value of (L). The system (D) is said to be subconsistent if there are sequences $y^n \in \mathbb{R}^m$, $s^n \in S^+$ such that

$$\lim_{n \to \infty} ((y^n)^T A - s^n) = -c.$$ 

**Lemma 1** Let $S$ be a closed convex cone. Then

(i) $m = 0 \iff (D)$ is subconsistent.

Moreover, if $S$ is a polyhedral convex cone, then

(ii) $m = 0 \iff (D)$ is consistent.

If $S$ is a convex cone, not necessarily closed, then

(iii) (D) subconsistent $\Rightarrow m = 0$.

**Proof** Part (i) follows from ([4], Corollary 2), and part (ii) from e.g. ([2], Theorem 4.6). To prove (iii) note that (D) is also dual to the problem

$$\inf \{ c^T x : A x = 0, x \in \text{cl}S \}.$$ 

This is simply a consequence of the fact $S^+ = (\text{cl}S)^+$. If we let $\bar{m}$ be the optimal value of the program (10) then clearly

$$\bar{m} \leq m.$$
Now, part (1) implies:

(D) subconsistent $\iff \bar{m} = 0$.

But (11) then implies $m \geq 0$. This lower bound is actually attained by $x = 0$, hence $m = 0$.

Lemma 2 Let $S$ be a convex cone (not necessarily closed). Consider the system

$$(D_o) \quad \{ y^T A + y_o c \in S^+, y \in \mathbb{R}^n, y_o \geq 0, (y, y_o) \neq 0 \}.$$ 

Then

$m = 0 \Rightarrow (D_o)$ consistent.

Proof This is a special case of Theorem 1 in [1]. For $t \geq 0$ let $I(t) \triangleq \{ x \in \mathbb{R}^n : |x_i| \leq t \}$ denote the $n$-dimensional cube with side $[-t, t]$.

The dual characterization theorem follows.

Theorem 2 Let problem (P) have differentiable convex functions $\{ f^k : k \in \{0\} \cup P \}$. Let $x^*$ be a feasible solution of (P) and suppose further that the cones $\{ D_k(x^*) : k \in P^* \}$ are closed (polyhedral). Then $x^*$ is optimal if, and only if, there exists a positive scalar $\theta^*$ such that the system $(DG, \theta)$ below is subconsistent (consistent) for every $\theta \in (0, \theta^*)$.

\[
(DG, \theta) \quad \begin{cases}
\nabla f^0(x^*) + \sum_{k \in P^*} y_k \nabla f^k(x^*) \in \sum_{k \in P^*} [D_k^+ \cap I(y_k)] \\
y_k \geq 0, \quad k \in P^*
\end{cases}
\]
Proof First we transform the constraint \((1,\theta)\) to the following pair of constraints

\[
d^t \nabla f^k(x^*) + \theta \sum_{i=1}^{n} |\gamma_i^k| \leq 0
\]

\[-d_i + \delta_i^k + \gamma_i^k = 0\]

Then, substituting \(|\gamma_i^k| = u_i^k + v_i^k\), \(\gamma_i^k = u_i^k - v_i^k\), \((u_i^k \geq 0, v_i^k \geq 0)\) we write the program \((G,\theta)\) as follows

\[(G,\theta) \quad \text{Min } d^t \nabla f^\theta(x^*)\]

\[
\begin{align*}
s.t.
\quad & d^t \nabla f^k(x^*) + \theta \sum_{i=1}^{n} u_i^k + \theta \sum_{i=1}^{n} v_i^k + w_k = 0 \\
\quad & -d_i + \delta_i^k + u_i^k - v_i^k = 0, \ i = 1, \ldots, n \\
\quad & \delta_i^k \in D_k \\
\quad & \nu_k^k \geq 0, \ u_k^k \geq 0, \ w = \{w_j: j \in \mathcal{P}^*\} \geq 0
\end{align*}
\]

This program is of the form \((L)\) with

\[
\begin{align*}
\quad x &= (d, \delta, u, v, w)^t \in \mathbb{R}^n \times \mathbb{R}^{n \times \mathcal{P}^*} \times \mathbb{R}^{n \times \mathcal{P}^*} \times \mathbb{R}^{n \times \mathcal{P}^*} \times \mathbb{R}^p^* \\
\quad c &= (\nabla f^\theta(x^*), 0, 0, 0, 0)^t \\
\quad S &= \mathbb{R}^n \times (\times_{k \in \mathcal{P}^*} D_k) \times \mathbb{R}^{n \times \mathcal{P}^*} \times \mathbb{R}^{n \times \mathcal{P}^*} \times \mathbb{R}^p^* \\
\quad \text{where} \\
\quad \mathbb{R}^m_+ &\quad \text{the nonnegative orthant of } \mathbb{R}^m.
\end{align*}
\]
By Theorem 1, \( x^* \) is optimal iff \( d \theta^* > 0 \)

\[(12) \quad \min (G, \theta) = 0, \forall \theta \in (0, \theta^*]. \]

Since the cones \( \{D_k : k \in P^*\} \) are closed (polyhedral) we can use part (i) (part (ii)) of Lemma 1 and conclude that (12) is equivalent to the subconsistency (consistency) of the system (D) which corresponds to \((G, \theta)\). Note first that

\[
S^+ = \{0\} \times ( \times _{k \in P^*} D_k^+ ) \times R^{nxp^*} \times R^{nxp^*} \times R^n
\]

Let \( f_j \Delta \frac{\partial}{\partial x_j} f^k(x^*) \). Then (0) is the following:

\[
\begin{align*}
\sum_{k \in P^*} y_k f^k_j - \sum_{k \in P^*} \mu^k_j + f_j^0 &= 0 \\
\theta y_k + \mu^k_j &\geq 0 \quad \text{j} = 1, \ldots, n \\
\theta y_k - \mu^k_j &\geq 0 \quad \text{j} = 1, \ldots, n \\
\mu^k_j &\in D_k^+ \\
y_k &\geq 0 \quad k \in P^*
\end{align*}
\]

This system is equivalent to the system \((DG, \theta)\) in the theorem.

The applicability of Theorem 2 is rather wide, because for a large family of convex functions the cone of directions of constancy is indeed polyhedral, in fact a subspace.
Recall that a function \( f: \mathbb{R}^n \to \mathbb{R} \) is \textit{faithfully convex} if it can be represented as
\[
(13) \quad f(x) = \varphi(Ax + b) + a^T x + \beta
\]
where \( A \) is an \( m \times n \) matrix, \( b \in \mathbb{R}^m \), \( a \in \mathbb{R}^n \), \( \beta \in \mathbb{R} \) and \( \varphi: \mathbb{R}^m \to \mathbb{R} \) is strictly convex.

The family \( F \) of the faithfully convex functions contains the strictly convex functions, quadratic functions, and in fact all analytic convex functions. See [7] and [8]. The constancy cone corresponding to \( f \) represented as in (13) is simply the null space of the \((m+1) \times n\) matrix \[
\begin{bmatrix}
A \\
a^T
\end{bmatrix}.
\]

An important subfamily of \( F \) is the functions whose restrictions are strictly convex. Let \( f^k: \mathbb{R}^n \to \mathbb{R} \) and \( [k] \) (read "block \( k \)"") denote the indices of variables \( \{x_j\} \) on which \( f^k \) actually depends, i.e.
\[
[k] = \{ j : \exists x_i = \xi_i, \text{ } i \neq j \text{ such that the function } f^k(\xi_1, \xi_2, \ldots, \xi_{j-1}, *, \xi_{j+1}, \ldots, \xi_n) \text{ is not a constant} \}.
\]
For any \( x \in \mathbb{R}^n \) the subvector \( x_{[k]} \) is obtained by deleting from \( x \) the components \( \{x_j : j \notin [k]\} \). The \textit{restriction} \( f^{[k]} \) is the function:
\[
f^{[k]}(x_{[k]} + \xi) \mid_{\text{card}[k]} \rightarrow \mathbb{R}
\]
obtained by restricting \( f^k \) to \( x_{[k]} \).

For a feasible point \( x^* \) of problem (P), with corresponding binding constraints \( P^* \), we denote by \( \langle i \rangle^* \) (read "cage \( i \) star") the indices of the functions \( \{ f^k : k \in P^* \} \) which actually depend on the variable \( x_i \), i.e.
Let \( \mathbf{x} = (x_1, x_2, \ldots, x_n)^t \), we denote by \( \vert \mathbf{x} \vert \) the vector
\[
\vert \mathbf{x} \vert = (\vert x_1 \vert, \vert x_2 \vert, \ldots, \vert x_n \vert)^t.
\]

For problems \((P)\) with strictly convex restrictions Theorem 2 reduces to a rather simple form.

**Corollary 2.1** Let problem \((P)\) have differentiable convex functions \( \{f^k: k \in \{0\} \cup P\} \) and suppose further that at the feasible point \( x^* \) the restrictions \( f^{[k]}: k \in P^* \) are strictly convex. Then \( x^* \) is optimal if, and only if, there exists \( \theta^* > 0 \) such that the system

\[
\begin{bmatrix}
\nabla f^0(x^*) + \sum_{k \in P^*} y_k \nabla f^k(x^*)
\end{bmatrix} \leq \theta
\]

\[
\begin{bmatrix}
\sum_{j \in \langle i \rangle} y_j \\
\vdots \\
\sum_{j \in \langle n \rangle} y_j
\end{bmatrix}
\]

is consistent for every \( \theta \in (0, \theta^*] \).

**Proof** For a function \( f^k \) with strictly convex restriction
\[
D_k = \{d \in \mathbb{R}^n: d_{[k]} = 0\}
\]
and hence
\[
D_k^+ = \{w \in \mathbb{R}^n: w_i = 0, i \notin [k]\}.
\]
Therefore \( \mu^k \in D_k^+ \cap I(y_k) \) if and only if \( \mu_i^k = 0 \) if \( i \notin [k] \) and \( |\mu_i^k| \leq y_k \) if \( i \in [k] \).

Consequently

\[
\mu \in \sum_{k \in P^*} D_k^+ \cap I(y_k) \text{ if and only if } |\mu_i| \leq \sum_{k \in [i]} y_k ,
\]

and hence the system \((D_G, \theta)\) reduces to \((S, \theta)\).

To illustrate Corollary 2.1, consider

**Example 1**

\[
\begin{align*}
\min f^0(x) &= e^x + e^{-x^2} + x^3 \\
\text{s.t.} & \\
\quad f^1(x) &= e^x \\
\quad f^2(x) &= e^{-x^2} \\
\quad f^3(x) &= (x_1 - 1)^2 + x_2^2 \\
\quad f^4(x) &= x_1^2 + x_2^2 + e^{-x_3} \\
\quad x &= (x_1, x_2, x_3)^t
\end{align*}
\]

Here \([1] = \{1\}, [2] = \{2\}, [3] = \{1, 2\}, [4] = \{1, 2, 3\} \). The feasible set is

\[
\left\{ \begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix} : x_3 \geq 0 \right\}
\]

and \( x^* = (0, 0, 0)^t \) is the optimal solution. For this point

\( P^* = \{1, 2, 3, 4\}, <1>^* = \{1, 3, 4\}, <2>^* = \{2, 3, 4\} \) and \( <3>^* = \{3\} \).
The system \( (S, \varepsilon) \) is
\[
\begin{bmatrix}
\gamma_1 + \gamma_3 + \gamma_4 \\
\gamma_2 + \gamma_3 + \gamma_4 \\
\gamma_4
\end{bmatrix}
\leq
\begin{bmatrix}
1 & 0 & 0 & -2 & 0 \\
-1 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & -1
\end{bmatrix}
\begin{bmatrix}
\gamma_1 \\
\gamma_2 \\
\gamma_3 \\
\gamma_4
\end{bmatrix}
\leq
\begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

\[
\theta
\begin{bmatrix}
\gamma_1 + \gamma_3 + \gamma_4 \\
\gamma_2 + \gamma_3 + \gamma_4 \\
\gamma_4
\end{bmatrix}
, \quad \gamma_1 \geq 0, \quad i=1, \ldots, 4
\]

For every \( 0 < \theta \leq \theta^* = \frac{1}{2} \) a solution for this system is
\[
\gamma_1 = \frac{2(1-2\theta)}{\theta(1-\theta)^2}; \quad \gamma_2 = 0; \quad \gamma_3 = \frac{1-2\theta}{\theta(1-\theta)}; \quad \gamma_4 = \frac{1}{1-\theta}
\]

For \( \theta = 0 \), \((S, 0)\) is inconsistent, i.e. Kuhn-Tucker conditions are not satisfied at the optimal point \( x^* = (0,0,0)^t \).

Unlike Theorem 1, which is valid regardless of the closedness of \( D_k \), Theorem 2 may fail without this assumption. To illustrate this consider the following

Example 2

Let
\[
\ell^2(x_1, x_2) = \begin{cases} 
(x_1^2 + x_2^2 - 2)^2 & \text{if } x_1^2 + x_2^2 - 2 \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]
and consider the convex program

\[
\text{Min } x_1 \\
\text{s.t. } f^1 = e^{2-x_1}-x_2 - 1 \leq 0 \\
\phantom{f^1 = e^{2-x_1}-x_2 - 1} f^2 \leq 0
\]

\(x^* = (1,1)^t\) is the only feasible solution and hence optimal.

At this point \(P^* = \{1,2\}\),

\[
D_1 = \{d \in \mathbb{R}^2: d_1 + d_2 = 0\}, \\
D_2 = \{d \in \mathbb{R}^2: d_1 + d_2 < 0\} \cup \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\forall f^0 = [1], \forall f^1 = [-1], \forall f^2 = [0]
\]

problem \((G,\theta)\) is here

\[
\text{Min } d_1 \\
\text{s.t. } \\
(15) \quad -d_1 - d_2 + \theta |d_1 - \delta^1_1| + \theta |d_2 - \delta^2_1| \leq 0 \\
(16) \quad \theta |d_1 - \delta^2_1| + \theta |d_2 - \delta^2_2| \leq 0 \\
(17) \quad \delta^1_1 + \delta^2_1 = 0 \\
(18) \quad \delta^2_1 + \delta^2_2 < 0 \quad \text{or} \quad \delta^2_1 = \delta^2_2 = 0.
\]

The constraint (16) implies \(d_1 = \delta^2_1\), \(d_2 = \delta^2_2\) hence from (18) either \(d_1 + d_2 < 0\) or \(d_1 = d_2 = 0\). The first possibility contradicts (15) hence
\[ d_1 = d_2 = 0 \iff \min(G, \theta) = 0, \ \forall \theta > 0. \]

This shows the optimality of \( x^* \).

To apply Theorem 2, first note that

\[ D_1^+ = \{ \lambda_1 [^1] : \lambda_1 \in \mathbb{R} \}, \quad D_2^+ = \{ \lambda_2 [^1] : \lambda_2 \leq 0 \}. \]

Hence \((DG, \theta)\) is here

\[
\begin{bmatrix} 1 \\ 0 \end{bmatrix} + y_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \lambda_1 + \lambda_2 \\ \lambda_1 + \lambda_2 \end{bmatrix}
\]

\[ |\lambda_1| \leq \theta y_1 \]

\[ -\theta y_2 \leq \lambda_2 \leq 0 \leq \theta y_2 \]

Clearly (19) cannot hold and \((DG, \theta)\) is therefore not consistent and not even subconsistent. Hence for this example Theorem 2 fails to indicate the optimality of \( x^* = (1,1)^+ \).

Although the conditions in Theorem 2 are not necessary for optimality if the cones \( \{ D_k : k \in P^* \} \) are not closed, these conditions are always sufficient.

**Theorem 3.** Under the assumptions of Theorem 1, the existence of \( \theta^* > 0 \) for which \((DG, \theta)\) is subconsistent (and in particular consistent) for every \( \theta \in (0, \theta^* ) \), is a sufficient condition for \( x^* \) to be an optimal solution of \((P)\).

**Proof.** The result follows from Theorem 1, by dualizing \((G, \theta)\) as in the proof.
of Theorem 2 using part (iii) of Lemma 1.

A general necessary criterion for optimality is the following.

**Theorem 4**  Under the assumptions of Theorem 1, the existence of $\theta^* > 0$ for which the system $(DG^0, \theta)$ (given below) is consistent for every $\theta \in (0, \theta^*)$ is a necessary condition for $x^*$ to be an optimal solution of $(P)$.

\[
(DG^0, \theta) \begin{cases}
y_0 \nabla f^0(x^*) + \sum_{k \in P^*} y_k \nabla f^k(x^*) \in \theta \sum_{k \in P^*} D^+_k \cap I(y_k) \\
y_0 \geq 0, \ y_k \geq 0, \ k \in P^*.
end{cases}
\]

Proof. For $\theta \in (0, \theta^*)$ the result follows from Theorem 1 by dualizing $(G, \theta)$ as in the proof of Theorem 2, and using Lemma 2. For $\theta = 0$ $(DG^0, \theta)$ reduces to the well known Fritz John system, whose consistency is a necessary condition for the optimality of $x^*$.

Remark. Note that the zero vector is always contained in

\[
\sum_{k \in P^*} D^+_k \cap I(y_k)
\]

for $y_k \geq 0, \ k \in P^*$. Therefore, the sufficiency criterion stated in Theorem 3 contains the classical Kuhn-Tucker sufficiency criterion (i.e. the consistency of the system $(DG, 0)$) as a special case.
To illustrate Theorem 4 we present

**Example 3** Consider again the problem given in Example 1. For the optimal point \( x^* = (1,1)^t \) the system \((DG^0, \theta)\) is

\[
\begin{align*}
\gamma_0[1] + \gamma_1[-1] &= \lambda_1 + \lambda_2 \\
|\lambda_1| &\leq \theta y \\
-\theta y_2 &\leq \lambda_2 \leq 0 \\
y_0 \geq 0, y_1 \geq 0, y_2 \geq 0 \quad (y_0, y_1, y_2)^t \neq 0.
\end{align*}
\]

For every \( 0 \leq \theta \leq \theta^* = 1 \), a solution of this system is

\[
\begin{align*}
\gamma_0 &= 0, \quad y_1 = \theta, \quad y_2 = 1-\theta \\
\lambda_1 &= -\theta^2, \quad \lambda_2 = -\theta(1-\theta).
\end{align*}
\]
REFERENCES

[1] Bazaraa, M.S., "A Theorem of the Alternative with Application to


[4] Duffin, R.J.: "Infinite Programs", in Linear Inequalities and
Related Systems (H.W.Kuhn and A.W. Tucker, Eds.), Annals of
Math. Studies No. 38, Princeton University Press, Princeton,

Conditions", in Studies and Essays Courant Anniversary Volume
(K.O. Friedrichs, I.E. Neugebauer and J.J. Stoker, Eds.),

Second Berkeley Symp. on Math. Statistics and Probability,

[7] Rockafellar, R.T.; "Some Convex Programs whose Duals are Linearly
Constrained", in Nonlinear Programming (J.B.Rosen, O.L. Mangasarian