OPTIMALITY CONDITIONS FOR CONVEX SEMI-INFINITE PROGRAMMING PROBLEMS

by

A. Ben-Tal*) , L. Kerzner**) and S. Zlobec**

Technical Report No. 65
December 1975

*) Department of Computer Science, Technion-Israel Institute of Technology, Haifa, Israel and Center of Cybernetic Studies, The University of Texas at Austin, Austin, Texas 78712, USA.

**) Department of Mathematics, McGill University, Montreal, Quebec, Canada H3C 3GL.

This research was partly supported by Project No. NR047-021, ONR Contract N00014-75-C0569 with the Center for Cybernetic Studies, The University of Texas and by the National Research Council of Canada.
ABSTRACT

This paper gives characterizations of optimal solutions for convex semi-infinite programming problems. These characterizations are free of a constraint qualification assumption. Thus they overcome the deficiencies of the semi-infinite versions of the Fritz John and the Kuhn-Tucker theories, which give only necessary or sufficient conditions for optimality, but not both. An application to the problem of best linear Chebyshev approximation with constraints is demonstrated.
1. INTRODUCTION

A mathematical programming problem with infinitely many constraints is termed a "semi-infinite programming problem". Such problems occur in many situations including production scheduling [10], air pollution problems [6],[7], approximation theory [5], statistics and probability [9]. For a rather extensive bibliography on semi-infinite programming the reader is referred to [8].

The purpose of this paper is to give necessary and sufficient conditions of optimality for convex semi-infinite programming problems. It is well known that the semi-infinite versions of both the Fritz John and the Kuhn-Tucker theories fail to characterize optimality (even in the linear case) unless a certain hypothesis, known as a "constraint qualification", is imposed on the problem, e.g. [4],[12]. This paper gives a characterization of optimality without assuming a constraint qualification.

Characterization theorems without a constraint qualification for ordinary (i.e. with a finite number of constraints) mathematical programming problems have been obtained in [1]. It should be noted that the analysis of the semi-infinite case is significantly different; the special feature being here the topological properties of all constraint functions including the particular role played by the non-binding constraints.
The optimality conditions are given in Section 2 for the general differentiable convex semi-infinite programming problems. For a particular class of such programs, namely the programs with "uniformly decreasing" constraint functions, the optimality conditions can be strengthened, as shown in Section 3. A connection with the semi-infinite analogs of the Fritz John and Kuhn-Tucker theories is presented in Section 4. An application to the problem of best linear Chebyshev approximation with constraints is demonstrated in Section 5. A linear semi-infinite problem taken from [4], for which the Kuhn-Tucker theory fails, is solved in this section using our theory.

2. CHARACTERIZATION THEOREMS

Consider the convex semi-infinite programming problem

\[(P)\]
\[
\begin{align*}
\text{Min} & \quad f^0(x) \\
\text{s.t.} & \quad f^k(x,t) \leq 0 \quad \text{for all } t \in T_k, \quad k \in \mathbb{N} \setminus \{1, \ldots, p\} \\
& \quad x \in \mathbb{R}^n
\end{align*}
\]

where

- $f^0$ is convex and differentiable,
- $f^k(x,t)$ is convex and differentiable in $x$
- for every $t \in T_k$ and continuous in $t$ for every $x$,
- $T_k$ is a compact subset of $\mathbb{R}^\ell$ ($\ell \geq 1$).
The feasible set of problem (P) is

\[ F = \{ x \in \mathbb{R}^n : f_k(x, t) \leq 0 \text{ for all } t \in T_k, \ k \in P \}. \]

Note that \( F \) is a convex set being the intersection of convex sets.

For \( x^* \in F \),

\[ T_k^* \triangleq \{ t \in T_k : f_k(x^*, t) = 0 \}, \]

\[ P^* \triangleq \{ k \in P : T_k^* = \emptyset \}. \]

A vector \( d \in \mathbb{R}^n \) is called a feasible direction at \( x^* \) if \( x^* + d \in F \).

For a given function \( f^k(\cdot, t) \), \( k \in \{ 0 \} \cup P \) and for a fixed \( t \in T_k \), we define

\[ D_k(x^*, t) \triangleq \{ d \in \mathbb{R}^n : \exists \bar{a} > 0 \ \exists f^k(x + ad, t) = f^k(x^*, t) \text{ for all } 0 \leq a \leq \bar{a} \}. \]

This set is called the cone of directions of constancy in [1], where it has been shown that, for a differentiable convex function \( f^k(\cdot, t) \), it is a convex cone contained in the subspace

\[ \{ d : d^t v f^k(x^*, t) = 0 \}. \]

Furthermore, if \( f^k(\cdot, t) \) is an analytic convex function, then \( D_k(x^*, t) \) is a subspace (not depending on \( x^* \)), see [1, Example 4]. In the sequel
the derivative of $f$ with respect to $x$, i.e. $f_x(x,t)$, is denoted by $V_f(x,t)$.

Optimality conditions will now be given for problem (P).

**THEOREM 1.** Let $x^*$ be a feasible solution of problem (P). Then $x^*$ is an optimal solution of (P) if, and only if, for every $\alpha^* > 0$ the system

(A) \[ d^t V^0(x^*) < 0, \]

(B) \[ d^t V^k(x^* + \alpha^* d, t) \leq 0 \text{ for all } t \in T_k^* \]

\[ \frac{d^t V^k(x^* + \alpha^* d, t)}{f^k(x^*, t)} \geq \frac{1}{\alpha^*} \]

(C) \[ \text{for all } t \in T_k \setminus T_k^* \]

is inconsistent.

**Proof.** We will show that $x^*$ is nonoptimal if, and only if, there exists $\alpha^* > 0$ such that the system (A), (B), (C) is consistent. A feasible $x^*$ is nonoptimal if, and only if, there exist $d^* > 0$ and $d \in \mathbb{R}^n$, $d \neq 0$, such that
(1) \[ f^0(x^* + \bar{d}) < f^0(x^*) \]

(2) \[ f^k(x^* + \bar{d}, t) \leq 0 \quad \text{for every} \quad t \in T_k, \quad k \in P. \]

By the convexity of \( f^0 \) and the gradient inequality, the existence of \( \bar{\alpha} > 0 \) satisfying (1) is equivalent to

\[ d^t \nabla f^0(x^*) < 0. \]

By the continuity of \( f^k(\cdot, t), k \in P \), the constraints with \( k \in P \setminus P^* \) can be omitted from discussion. We consider (2), for some given \( k \in P^* \), and discuss separately the two cases: \( t \in T_k^* \) and \( t \in T_k \setminus T_k^* \). Thus (2) can be written

(2-a) \[ f^k(x^* + \bar{d}, t) \leq 0 \quad \text{for every} \quad t \in T_k^* \]

(2-b) \[ f^k(x^* + \bar{d}, t) \leq 0 \quad \text{for every} \quad t \in T_k \setminus T_k^* . \]

Consider first (2-a) for some fixed \( k \in P^* \) and \( t \in T_k^* \). By the mean value theorem

(3) \[ f^k(x^* + \bar{d}, t) = f^k(x^*, t) + \bar{d}^t \nabla f^k(x^* + \alpha d, t) \]

for some

\[ 0 < \alpha_k < \bar{\alpha}. \]

Since \( t \in T_k^* \) and \( \bar{\alpha} > 0 \), (2-a) implies
(4) \[ d^t \nu_k^k(x^* + \alpha_k d, t) \leq 0. \]

Denote

(5) \[ \hat{\alpha} = \min \{\alpha_k\}. \]

Clearly, \( \hat{\alpha} \) always exists (since \( P \) is finite) and it is positive. By the convexity of \( f^k(\cdot, t) \), (5) and (4),

(6) \[ d^t \nu_k^k(x^* + \hat{\alpha} d, t) \leq d^t \nu_k^k(x^* + \alpha_k d, t) \leq 0. \]

On the other hand, the existence of \( \alpha^* > 0 \) such that, for some \( t \in T^*_k \) and all \( k \in P^* \),

\[ d^t \nu_k^k(x^* + \alpha^* d, t) \leq 0 \]

implies (2-a) with \( 0 < \tilde{\alpha} \leq \alpha^* \).

It is left to show that the existence of \( \tilde{\alpha} > 0 \), such that (2-b) holds, is equivalent to the existence of \( \tilde{\alpha} > 0 \), such that (C) holds. Suppose that (2-b) holds for some \( \tilde{\alpha} > 0 \). Then

\[ f^k(x^* + \tilde{\alpha} d, t) = f^k(x^*, t) + \tilde{\alpha} d^t \nu_k^k(x^* + \tilde{\alpha} d, t) \leq 0, \]
for some

(7) \[ 0 < \tilde{\alpha}_k < \bar{\alpha}, \]

by the mean value theorem. Hence

\[
\frac{d^t \nabla f^k(x + \tilde{\alpha}_k d, t)}{f^k(x, t)} \geq \frac{1}{\bar{\alpha}}, \text{ since } t \in T_k \setminus T_k^k
\]

(8) \[ z = \frac{1}{\bar{\alpha}}, \text{ by (7)}. \]

Denote

(9) \[ \tilde{\alpha} = \min_{k \in P^k} (\tilde{\alpha}_k) > 0. \]

Using the monotonicity of the gradient of the convex function \( f^k(\cdot, t) \), one obtains here

(10) \[
\frac{d^t \nabla f^k(x + \tilde{\alpha}_k d, t)}{f^k(x, t)} \geq \frac{d^t \nabla f^k(x + \tilde{\alpha} d, t)}{f^k(x, t)} \text{ for every } 0 \leq \alpha \leq \tilde{\alpha}_k.
\]

This gives

\[
\frac{d^t \nabla f^k(x + \tilde{\alpha}_k d, t)}{f^k(x, t)} \geq \frac{1}{\tilde{\alpha}_k}, \text{ by (10) and (8)}
\]

\[ z = \frac{1}{\tilde{\alpha}_k}, \text{ by (9)} \]
which is (C) with $\alpha^* = \bar{\alpha}$.

Suppose now that (C) is true for some $\alpha^* > 0$. Using again the monotonicity of the gradient of the convex function $f^k(\cdot, t)$, and the fact that $f^k(x^*, t) < 0$ for $t \in T^k \setminus T^k_*$, one easily obtains

$$f^k(x^*, t) + \alpha^* d^t Vf^k(x^* + \alpha d, t) \leq 0, \quad \text{for every } 0 < \alpha < \alpha^*. \tag{11}$$

But

$$f^k(x^* + \alpha^* d, t) = f^k(x^*, t) + \alpha^* d^t Vf^k(x^* + \alpha d, t),$$

for some particular $0 < \alpha < \alpha^*$,

by the mean value theorem

$$\leq 0, \quad \text{by (11)}$$

which is (2-b) with $\bar{\alpha} = \alpha^*$.

Summarizing the above results one derives the following conclusion: If $x^*$ is not optimal then there exists $\alpha^* = \min(\alpha, \bar{\alpha}) > 0$ such that the system (A), (B) and (C) is consistent. If there exists $\alpha^* > 0$ such that the system (A), (B) and (C) is consistent, then there exist $\alpha_0 > 0$ and $\bar{\alpha} > 0$ such that

$$\begin{align*}
\{ & f^0(x^* + \alpha_0 d) < f^0(x^*) \\
& f^k(x^* + \alpha d, t) \leq 0 \quad \text{for every } t \in T^k_* \\
& f^k(x^* + \bar{\alpha} d, t) \leq 0 \quad \text{for every } t \in T^k \setminus T^k_* \\
& k \in P^* \end{align*}$$

\tag{12}$$
If one denotes
\[ \hat{\alpha} = \min \{ \alpha_0, \bar{\alpha} \} > 0 \]
then, again by the convexity of \( f^k(\cdot, t) \), \( k \in \{0\} \cup P \), (12) can be written
\[
\begin{align*}
& f^0(x^* + \hat{\alpha} d) < f^0(x^*) \\
& f^k(x^* + \hat{\alpha} d, t) \leq 0 \quad \text{for every} \quad t \in T_k, \quad k \in P^* .
\end{align*}
\]
implying that \( x^* \) is not optimal.

\[ \square \]

**Remark 1.** Since \( Vf^k(x, \cdot) \) is continuous for every \( x \) in some neighbourhood of \( x^* \) (this follows from e.g. [14, Theorem 25.7]), condition (C) in Theorem 1 needs checking only at the points \( t \in T_k \) which are in
\[
N_k \triangleq U \bigcup_{t^* \in T_k^*} N(t^*),
\]
where \( N(t^*) \) is a fixed open neighbourhood of \( t^* \). For the points \( t \in T_k \setminus N_k \) one can always find \( \alpha^* \) which satisfies (C). This follows from the fact that for every \( \tilde{\alpha} \),
\[
(13) \quad \frac{d^t Vf^k(x^* + \tilde{\alpha} d, t)}{f^k(x^*, t)} \geq -M
\]
for some positive constant \( M \), by the compactness of \( T_k \setminus N_k \). Choose \( M \) in (13) large enough, so that
\[
(14) \quad \alpha^* \triangleq \frac{1}{M} \leq \tilde{\alpha} .
\]
Now
\[
\frac{d^k_v f(x^* + \alpha d, t)}{f^k(x^*, t)} \geq \frac{d^k_v f(x^* + \bar{\alpha d}, t)}{f^k(x^*, t)}, \text{ by (10) and (14)}
\]
\[
\geq \frac{1}{\alpha^*}, \text{ by (13) and (14)}.
\]

**Example 1.** Consider

\[
\min f^0(x) = -x
\]

s.t.
\[
f(x, t) = t^2[(x-t)^2 - t^2] \leq 0, \text{ for all } t \in T = [0,1].
\]

The feasible set consists here of the single vector \(x^* = 0\), which is therefore the optimal solution. At \(x^* = 0, T^* = T\), the condition (C) is redundant, while conditions (A) and (B) are

\[
-d < 0
\]

\[
2t^2(\alpha d - t) \leq 0 \text{ for all } t \in [0,1].
\]

Since these conditions imply

\[
\alpha^* \leq \frac{t}{d} \text{ for all } t \in (0,1),
\]

\(\alpha^*\) cannot be positive here. Therefore, for every \(\alpha^* > 0\) the system (A), (B), (C) is inconsistent, implying that \(x^*\) is optimal.

An example demonstrating the importance of condition (C) is given in Section 5.
In order to state our next result, which is a characterization of optimality for a subclass of convex functions, i.e. strictly convex functions in their "actual variables", we adopt some notions from [1].

For every \( k \in P \) and \( t \in T_k \), denote by \([k](t)\) (read "block k"), the following index subset of \( P \): \( j \in [k](t) \) if, and only if \( \gamma^k : \mathbb{R} \rightarrow \mathbb{R} \), defined by

\[
\gamma^k(i) \triangleq (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n)
\]

is not a constant function for some fixed \( x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n \). Thus, for a given \( t \in T_k \), \([k](t)\) is the set of indices of those variables on which \( f^k(\cdot, t) \) actually depends. These "actual variables" determine the vector \( x_{[k]}(t) \), obtained from \( x = (x_1, \ldots, x_n)^t \) by deleting the variables \( \{x_j : j \notin [k](t)\} \), without changing the order of the remaining ones. Similarly, we denote by \( f_{[k]}(t) : R^{\text{card}[k]} \rightarrow \mathbb{R} \) the restriction of \( f^k \) to \( R^{\text{card}[k]} \).

**Definition 1.** A function \( f^k : \mathbb{R}^n \times T_k \rightarrow \mathbb{R} \) is strictly convex in its actual variables if for every \( t \in T_k \) its restriction \( f_{[k]}(t) (\cdot, t) \) is strictly convex.

The above concept will be illustrated by an example.

**Example 2.** Consider

\[
f^1(x, t) = x_1^2 + tx_2^2, \quad t \in T = [0, 1],
\]
Note that function \( f^1(\cdot, t) \) is not strictly convex for every \( t \in T \). Here
\[
[1](t) = \begin{cases} 
1 & \text{if } t = 0 \\
1,2 & \text{if } t \in (0,1],
\end{cases}
\]
\[
\mathbf{x}[1](t) = \begin{cases} 
(x_1) & \text{if } t = 0 \\
(x_1, x_2) & \text{if } t \in (0,1]
\end{cases}
\]
and
\[
f[1](t) = \begin{cases} 
x_1^2 & \text{if } t = 0 \\
x_1^2 + tx_2^2 & \text{if } t \in (0,1]
\end{cases}
\]
clearly a strictly convex function in its actual variables for every \( t \in T \).
Hence \( f^1 \) is a strictly convex function in its actual variables.

**Corollary 1.** Let \( x^* \) be a feasible solution of problem (P), where \( f^k(\cdot, t), \ k \in p^* \) are strictly convex in their actual variables. Then \( x^* \) is an optimal solution of (P) if, and only if, for every \( \alpha^* > 0 \) and every subset \( \Omega_k \subset T_k^* \) the system

(A) \[ d^t Vf^0(x^*) < 0 \]
(B,\( \Omega \)) \[ d^t Vf^k(x^* + \alpha^* d, t) < 0 \text{ for all } t \in T_k^* \backslash \Omega_k \]
(C) \[ \frac{d^t Vf^k(x^* + \alpha^* d, t)}{f^k(x^*, t)} \geq - \frac{1}{\alpha^*} \text{ for all } t \in T_k \backslash T_k^* \]
(D,\( \Omega \)) \[ d_{[k]}(t) = 0 \text{ for all } t \in \Omega_k, \]
\( k \in p^* \)

is inconsistent.
Proof. We know, by Theorem 1, that \( x^* \) is nonoptimal if, and only if, there exists \( \alpha^* > 0 \) such that the system (A), (B), (C) is consistent.

In order to prove Corollary 1, it is enough to show that (B) is consistent if, and only if, for some subsets \( \Omega_k \subset T_k \), \( k \in \mathcal{P}^* \), the system (B,\( \Omega \)), (D,\( \Omega \)) is consistent. Suppose that (B) holds. For every \( k \in \mathcal{P}^* \) define

\[
\hat{\Omega}_k \triangleq \{ t \in T_k : d^t \forall f^k(x^*+\alpha d,t) = 0 \quad \text{for all} \quad 0 < \alpha \leq \alpha^* \}.
\]

Hence, by the mean value theorem, for every \( t \in \hat{\Omega}_k \)

\[
f^k(x^*+\alpha d,t) = f^k(x^*,t) \quad \text{for all} \quad 0 < \alpha \leq \alpha^*.
\]

Since \( f^k(\cdot,t) \) is strictly convex in its actual variables, this is equivalent to

\[
\hat{d}[k](t) = 0 \quad \text{for all} \quad t \in \hat{\Omega}_k.
\]

If \( t \in T_k \setminus \hat{\Omega}_k \), then obviously \( d^t \forall f^k(x^*+\alpha d,t) < 0 \) for some \( 0 < \alpha < \alpha^* \), by (B). Thus (B,\( \Omega \)), (D,\( \Omega \)) holds for \( \Omega_k = \hat{\Omega}_k \). (Note that some or all \( \hat{\Omega}_k \)'s may be empty.) The reverse statement follows from the observation that \( \hat{d}[k](t) = 0 \) implies \( d^t \forall f^k(x^*+\alpha d,t) = 0 \).

\[
\blacksquare
\]

If a function \( f^k(\cdot,t) \) is strictly convex (in all variables \( x_1, \ldots, x_n \)) for every \( t \in T_k \), \( k \in \mathcal{P}^* \), then \( D_k(x^*,t) = \{0\} \). This implies
that the system (A), (B,Ω), (C), (D,Ω) is inconsistent for every nonempty Ωₖ, k ∈ ℘. Thus condition (D,Ω) is redundant. In fact condition (C) is also redundant, which follows by the following lemma.

**Lemma 1.** Let \( f(x,t) \) be convex and differentiable in \( x \in \mathbb{R}^n \) for every \( t \) in a compact set \( T \subset \mathbb{R}^l \) and continuous in \( t \) for every \( x \). If for some \( d \in \mathbb{R}^n \),

\[
\text{(15)} \quad d^T \nabla f(x^*,t) < 0 \quad \text{for all} \quad t \in T^* = \{ t : f(x^*,t) = 0 \},
\]

then there exists \( \alpha > 0 \) such that

\[
\text{(16)} \quad f(x^* + \alpha d, t) \leq 0 \quad \text{for all} \quad t \in T \setminus T^*.
\]

*Proof.* It is to show that the hypothesis (15) and the negation of the conclusion (16), which is

"For every \( \alpha > 0 \) there is \( t = t(\alpha) \in T \setminus T^* \) such that \( f(x^* + \alpha d, t(\alpha)) > 0 \),"

are not simultaneously satisfied. If this were true one would have the following situation:

For every \( \alpha_n \) of the sequence \( \alpha_n = 2^{-n} \), there is a \( t_n = t_n(\alpha_n) \in T \setminus T^* \) such that

\[
\text{(17)} \quad f(x^* + \alpha_n d, t_n(\alpha_n)) > 0, \quad n = 0, 1, 2, \ldots
\]
Since $T$ is compact, $\{t_n\}$ has an accumulation point $\hat{t} \in T$, i.e. there is a convergent subsequence $\{t_{n_i}\}$ with $\hat{t}$ as its limit point. We discuss separately two possibilities and arrive at contradictions in each case.

**Case I.** $\hat{t} \in T^*$. Since $f(x^*,\hat{t}) = 0$ and $d^tVf(x^*,\hat{t}) < 0$, by (15), there exists $\alpha > 0$ such that

$$f(x^* + \alpha d, \hat{t}) < 0.$$  

(18)

For all large values of index $i$, $\alpha_{n_i} < \alpha$ and

$$f(x^*, t_{n_i}) < 0,$$  

(19)

since $t_{n_i} \in T \setminus T^*$. This implies

$$f(x^* + \alpha d, t_{n_i}) > 0.$$  

(20)

(If (20) were not true, one would have, for some particular $n_i$,

$$f(x^* + \alpha d, t_{n_i}) \leq 0.$$  

(21)

Now $\alpha_{n_i} < \alpha$, (19), (21) and the convexity of $f$ imply

$$f(x^* + \alpha_{n_i} d, t_{n_i}) \leq 0.$$  

(22)
which contradicts (17). But (18) and (20) contradict the continuity of $f(x^*+\alpha d,\cdot)$.

**Case II.** $\hat{t} \in T\setminus T^*$. Since $f(x^*,\hat{t}) < 0$, there exists $\alpha > 0$ such that (18) holds, by the continuity of $f(\cdot,\hat{t})$. The rest of the proof is the same as in Case I.

A characterization of optimality for strictly convex constraints follows.

**Corollary 2.** Let $x^*$ be a feasible solution of problem (P), where $f^k(\cdot,t)$ are strictly convex for every $t \in T_k$, $k \in P^*$. Then $x^*$ is an optimal solution of (P) if, and only if, for every $\alpha^* > 0$ the system

\[
(A) \quad d^t v^0(x^*) < 0
\]

\[
(B_1) \quad d^t v^k(x^*,t) < 0 \quad \text{for all } t \in T_k^*, \quad k \in P^*
\]

is inconsistent.

**Proof.** If $x^*$ is not optimal, then the system (A), (B_1), (C) is consistent, by Corollary 1. This implies that the less restrictive system (A), (B_1) is consistent. Suppose that the system (A), (B_1) is consistent. Then for every $k \in P^*$ there is $\alpha_k > 0$ such that
\[ f^k(x^* + a_k^*, t) \leq 0 \text{ for all } t \in T_k \setminus T_k^* \]

by Lemma 1. Let
\[ a^* = \min\{a_k : k \in P^*\} \]

By the convexity of \( f^k \), it follows that
\[ f^k(x^* + a_k^*, t) \leq 0 \text{ for all } t \in T_k \setminus T_k^* \text{ and } k \in P^*. \]

This is equivalent to (C) of Theorem 1 (see (2-b)). Therefore the system (A), (B), (C) is consistent. This implies that the system (A), (B), (C) is consistent. (The reader may verify it by the technique used in the proof of Lemma 2.) Hence \( x^* \) is optimal, by Corollary 1.

\[ \square \]

When the constraint functions (but not necessarily the objective function) are linear, i.e. when (P) is of the form

(L)
\[
\text{Min } f^o(x)
\]
\[
g^k_0(t) + \sum_{i=1}^{n} x_i g^k_i(t) \leq 0, \text{ for all } t \in T_k, \ k \in P
\]
then Theorem 1 can be considerably simplified.

**Corollary 3.** Let \( x^* \) be a feasible solution of problem (L). Then \( x^* \) is optimal if, and only if, the system
(A) \[ d^T v f^o(x^*) < 0 \]

(B2) \[ \frac{1}{\lambda} \sum_{i=1}^{n} d_i g_i^k(t) \leq 0, \text{ for all } t \in \mathbb{T}_k \]

(C1) \[ \frac{1}{\lambda} \sum_{i=1}^{n} d_i g_i^k(t) \geq -1, \text{ for all } t \in \mathbb{T}_k \backslash \mathbb{T}_k^* \]

is inconsistent.

Proof. If \( f_k(\cdot, t) \) is linear, then for every \( t \in \mathbb{T}_k \)

\[ D_k(x, t) = \{ d \in \mathbb{R}^n : d^T v f_k(x, t) = 0 \}. \]

Thus (B) reduces to (B2). The left hand side of (C) reduces to the left hand side of (C1), which does not depend on \( \alpha^* \). Moreover, \( \alpha^* \) on the right hand side of (C) can be taken \( \alpha^* = 1 \), because whenever \( \mathbb{J} \) satisfies (A) and (B2), so does \( d = \frac{1}{\alpha^*} \mathbb{J} \).

In many practical situations the sets \( \mathbb{T}_k, k \in \mathbb{P} \) are compact intervals and the sets \( \mathbb{T}^*_k, k \in \mathbb{P}^* \) are finite. (This is always the case when \( f(x^*, \cdot) \) are analytic functions not identically zero.) For such cases condition (B) can be replaced by a finite number of linear inequalities.
Corollary 4. Let \( x^* \) be a feasible solution of problem (P).

Suppose that all the sets \( T_k^*, \ k \in P^* \) are finite. Then a feasible solution \( x^* \) of problem (P) is optimal if, and only if, for every \( \alpha^* > 0 \) and for every subset \( \Omega_k^* \) of \( T_k^* \) the system

\[
\begin{align*}
(A) & \quad d^T \nabla f^0(x^*) < 0 \\
(B_3) & \quad d^T \nabla f^k(x^*, t) < 0, \ t \in \Omega_k \\
& \quad d \in D_k(x^*, t), \ t \in T_k^* \setminus \Omega_k \\
(C) & \quad \frac{d^T \nabla f^k(x^* + \alpha^* d, t)}{f^k(x^*, t)} \geq \frac{1}{\alpha^*} \\
& \quad \text{for all } t \in T_k \setminus T_k^*, \ k \in P^* \\
\end{align*}
\]

is inconsistent.

An important special case of Corollary 4 is when the sets \( T_k \) themselves are finite. Then problem (P) can be reduced to a mathematical program of the form

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

This is obtained by setting \( T_k = \{k_1, k_2, \ldots, k_{\text{card } T_k}\} \) and identifying \( \{f^k(x, k_i); k_i \in T_k; k = 1, 2, \ldots, p\} \) with \( \{f^k(x); k \in P \setminus \{1, 2, \ldots, \text{card } T_k\}\} \). Here \( P^* = \{k \in P; f^k(x^*) = 0\} \).

Also \( \{D_k(x^*, k_i); k_i \in T_k; k = 1, 2, \ldots, p\} \) is denoted by \( \{D_k(x^*); k \in P\} \).
The major difference between the semi-infinite problem (P) and the mathematical problem (MP) is that for the latter the condition (C) is redundant; Theorem 1 then reduces to the following result obtained in [1, Theorem 1].

Corollary 5. Consider problem (MP), where \( \{ f^k : k \in \{0\} \cup \mathbb{P}\} \) are differentiable convex functions: \( \mathbb{R}^n \rightarrow \mathbb{R} \). A feasible solution \( x^* \) of (MP) is optimal if, and only if, for every subset \( \Omega \) of \( \mathbb{P}^* \) the system

\[
\begin{align*}
&d^t v f^0(x^*) < 0 \\
&d^t v f^k(x^*) < 0, \quad k \in \Omega \\
&d \in D_k(x^*), \quad k \in \mathbb{P}^* \setminus \Omega
\end{align*}
\]

is inconsistent.

Proof. Here condition (C) becomes

\[
\frac{d^t v f^k(x^* + \alpha^* d)}{f^k(x^*)} \leq -\frac{1}{\alpha^*}, \quad k \in \mathbb{P} \setminus \mathbb{P}^*
\]

for some \( \alpha^* > 0 \). Since here the set \( \mathbb{P} \setminus \mathbb{P}^* \) is finite, and hence compact, the redundancy of condition (C) is shown as in Remark 1.

The following result gives a characterization of a unique optimal solution of problem (P).

**Theorem 2.** Let \( x^* \) be a feasible solution of problem (P). Then \( x^* \) is the unique optimal solution of problem (P) if, and only if, for every \( \alpha^* > 0 \) there is no \( d \) satisfying conditions \( (B) \), \( (C) \) and
(A₁) \[ d^T \nabla f^0(x^*) < 0 \text{ or } d \in D_0(x^*). \]

**Proof.** Suppose that the system \((A₁), (B), (C)\) is inconsistent. Then so is the system \((A), (B), (C)\). Hence, by Theorem 1, \(x^*\) is an optimal solution. Suppose that \(x^*\) is not the unique optimal solution. Then there exist \(\bar{\alpha} > 0\) and \(\bar{d} \neq 0\) such that \(\bar{x} = x^* + \bar{\alpha}\bar{d}\) is feasible, which implies that \(\bar{d}\) satisfies \((B), (C)\) and \(f^0(\bar{x}) = f^0(x^* + \bar{\alpha}\bar{d})\).

Since the set of all optimal solutions of a convex program is convex, the latter implies \(f^0(x^*) = f^0(x^* + \alpha d)\) for all \(0 \leq \alpha \leq \bar{\alpha}\), i.e. \(d \in D_0(x^*)\). Thus \(d\) satisfies \((A₁), (B)\) and \((C)\), which is impossible. Therefore \(x^*\) is the unique optimum. The necessity follows by a similar argument.

3. **PROGRAMS WITH UNIFORMLY DECREASING CONSTRAINTS**

The applicability of Theorem 1 is, in general, obscured by the appearance of parameter \(\alpha^*\) in conditions \((B)\) and \((C)\). The purpose of this section is to point out some of the topological difficulties which arise in the removing of \(\alpha^*\) from condition \((B)\). A class of convex functions for which the optimality conditions can be stated without reference to \(\alpha^*\) in condition \((B)\) will be called the uniformly decreasing functions.

In what follows we assume that \(f: \mathbb{R}^n \times T \rightarrow \mathbb{R}\) is convex and differentiable in \(x \in \mathbb{R}^n\) for every \(t\) of a compact set \(T\) in \(\mathbb{R}^m\). Further, \(\nabla f(x^*, t)\) denotes \(\nabla f_x(x^*, t)\).
Definition 2. Let \( f: \mathbb{R}^n \times T \rightarrow \mathbb{R} \) and \( x^* \in \mathbb{R}^n \) be such that \( T^* \neq \emptyset \).
Then for a given \( d \in \mathbb{R}^n \), \( d \neq 0 \), the function \( f \) is uniformly decreasing at \( x^* \) in the direction \( d \), if (i) the set
\[
S(x^*,d) = \{ t \in T^*: \mathcal{D}^v f(x^*,t) < 0 \}
\]
is compact and if (ii) there exists \( \bar{\alpha} > 0 \) such that \( f(x^* + \bar{\alpha}d, t) = 0 \) for all \( t \in T^* \) for which \( d \in D(x^*, t) \).

It is not easy to recognize whether a general convex function \( f \) is uniformly decreasing.

Example 3. Consider the following functions from \( \mathbb{R} \times \mathbb{R} \) into \( \mathbb{R} \):
\[
\begin{align*}
  f^1(x,t) &= t^2[(x-t)^2 - t^2], \quad t \in T \text{ (used in Example 1)} \\
  f^2(x,t) &= x^2 - tx, \quad t \in T \\
  f^3(x,t) &= -tx, \quad t \in T.
\end{align*}
\]
These functions are all convex, \( f^2 \) is actually strictly convex and \( f^3 \) is linear in \( x \) for every \( t \in T \). If \( T = [0,1] \), then neither function is uniformly decreasing at \( x^* = 0 \) in the direction \( d = 1 \). However, if \( T = [1,2] \) then all three functions are uniformly decreasing at \( x^* = 0 \) in the same direction \( d = 1 \).
As suggested by the above example, a convex function \( f \) is uniformly decreasing at \( x^* \) in the direction \( d \neq 0 \), whenever \( \nabla f(x^*,\cdot) \) is continuous and the set
\[
E(x^*,d) = \{ t \in T^*: d^T \nabla f(x^*,t) = 0 \}
\]
is empty. Its complement
\[
S(x^*,d) = T^* \setminus E(x^*,d) = T^*
\]
is then compact. Clearly all analytic functions not identically zero are uniformly decreasing. However, a characterization of optimality for problem \( (P) \) with such constraint functions is already given by Corollary 3.

An important uniformity property of convex functions with compact \( S(x^*,d) \) follows:

**Lemma 2.** Let \( f(x,t) \) be convex and differentiable in \( x \), for every \( t \) in a compact set \( T \subseteq \mathbb{R}^n \), and continuous in \( t \), for every \( x \in \mathbb{R}^n \). Suppose further that for some \( x^* \) and \( d \neq 0 \) in \( \mathbb{R}^n \), the set \( S(x^*,d) \) is nonempty and compact. Then there exists \( \bar{a} > 0 \) such that

\[
(22) \quad f(x^* + \bar{a} d, t) < 0, \quad 0 < \alpha \leq \bar{a}
\]
for all \( t \in S(x^*,d) \).
Proof. Suppose that such $\bar{\alpha} > 0$ does not exist. Then there exists a sequence $\{t_i\} \subset S(x^*, d)$ and a sequence $\{\alpha_i\}$, $\alpha_i = \alpha_i(t_i) > 0$ such that

$$f(x^* + \alpha_i d, t_i) = 0, \quad f(x^* + \alpha d, t_i) < 0, \quad 0 < \alpha < \alpha_i$$

and

$$f(x^* + \alpha d, t_i) > 0, \quad \alpha > \alpha_i$$

with $\inf \{\alpha_i\} = 0$. Since $S(x^*, d)$ is compact, $\{t_i\}$ contains a convergent subsequence $\{t_{i_j}\}$. Let $\hat{t} \in S(x^*, d)$ be the limit point of $\{t_{i_j}\}$. Now

$$d^{\perp} v f(x^*, \hat{t}) < 0$$

implies that there exists $\hat{\alpha} > 0$ such that

$$f(x^* + \alpha d, \hat{t}) < 0, \quad 0 < \alpha < \hat{\alpha}.$$ 

In particular

$$f(x^* + \hat{\alpha} d, \hat{t}) < 0.$$ 

For any $\epsilon > 0$ there exists $j_0 = j_0(\epsilon)$ such that

$$|t_{i_j} - \hat{t}| < \epsilon \quad \text{and} \quad \alpha_{i_j} < \hat{\alpha} \quad \text{for all} \quad j > j_0.$$ 

Now (23) and (25) imply

$$f(x^* + \hat{\alpha} d, t_{i_j}) > 0 \quad \text{for all} \quad j > j_0.$$ 

But the inequalities (24) and (26) contradict the continuity of $f(x^* + \hat{\alpha} d, \cdot)$. 

\[ \Box \]
Example 4. Consider again

\[ f^2(x,t) = x^2 - tx, \quad t \in T = [1,2]. \]

This function is uniformly decreasing at \( x^* = 0 \) in the direction \( d = 1 \).

The inequality (22) holds for every \( 0 < \alpha < 1 \), in particular \( \alpha = \frac{1}{2} \). If

the above interval \( T \) is replaced by \( T = [0,1] \), then \( f^2 \) is not uniformly

decreasing at \( x^* = 0 \) with \( d = 1 \). An \( \alpha > 0 \) satisfying (22) here does

not exist.

A characterization of optimality for programs \((P)\), with uniformly decreasing constraint functions, follows.

THEOREM 3. Let \( x^* \) be a feasible solution of problem \((P)\). Suppose that \( \{f_k(x,t); k \in P^* \} \) are uniformly decreasing at \( x^* \) in every feasible direction \( d \). Then \( x^* \) is an optimal solution of \((P)\) if, and only if,

for every \( \alpha^* > 0 \) the system

\[
\begin{align*}
(A) \quad & d^T \nabla f^0(x^*) < 0, \\
(B^*_k) \quad & \left\{ \begin{array}{l}
\quad d^T \nabla f^k(x^*,t) < 0 \quad \text{or} \quad d \in D_k(x^*,t) \\
\quad \text{for all} \quad t \in T^*_k, \\
\end{array} \right. \\
(C) \quad & \left\{ \begin{array}{l}
\quad \frac{d^T \nabla f^k(x^* + \alpha^* d,t)}{f^k(x^*,t)} \geq \frac{1}{\alpha} \\
\quad \text{for all} \quad t \in T^*_k \setminus T^*_k, \\
\end{array} \right. \\
\end{align*}
\]

is inconsistent.
Proof. Parts (A) and (C) are proved as in the case of Theorem 1. It is left to show that the existence of $\bar{\alpha} > 0$ satisfying (2-a) is equivalent to the consistency of $(B_4)$. It is clear that (?-a) implies $(B_4)$. In order to show that $(B_4)$ implies (2-a) we use the assumption that the functions $\{f^k(x, t); k \in P^*\}$ are uniformly decreasing at $x^*$ in the direction $d$. When $(B_4)$ holds, then for every $k \in P^*$ there exist $\alpha_k > 0$ and $\alpha^0_k > 0$ such that

$$f^k(x^* + \alpha d, t) < 0, \quad 0 < \alpha \leq \alpha_k$$

for all $t \in S_k \setminus \{t \in T_k^* : d^T v f^k(x^*, t) < 0\},$ (27)

by Lemma 2, and

$$f^k(x^* + \alpha d, t) = 0, \quad 0 < \alpha \leq \alpha^0_k$$

for all $t \in T_k^* \setminus S_k$, (28)

since $d \neq 0$. The latter follows by part (ii) of Definition 2 and the convexity of $f^k$. Let

$$\bar{\alpha} \triangleq \min_{k \in P^*} \{\alpha_k, \alpha^0_k\} > 0.$$ (29)

Clearly (27) and (28) can be written as the single statement (2-a) with $\bar{\alpha}$ chosen as in (29).
The following example shows that the assumption that \( \{f^k(x,t) : k \in \mathbb{P}^* \} \) be uniformly decreasing at \( x^* \) cannot be omitted in Theorem 3.

**Example 5.** Consider again the program

\[
\text{Min } f^0(x) = -x \\
\text{s.t.} \\
f(x,t) = t^2[(x-t)^2 - t^2] \leq 0, \text{ for all } t \in T = [0,1].
\]

The constraint function \( f(x,t) \) is not uniformly decreasing. The optimal solution of this program is \( x^* = 0 \), as established in Example 1 using Theorem 1. Here \( T^* = T \), condition (C) is redundant, while conditions (A) and (B_4) are

\[ -d < 0 \]

\[
\left\{ \begin{array}{l}
-2dt^2 < 0 \text{ or } d \in D(x^*,t) \\
\text{for all } t \in [0,1].
\end{array} \right.
\]

Since \( D(x^*,0) = \mathbb{R} \) (the real line), these conditions are clearly satisfied, e.g. for \( d = 1 \). (If \( t = 0 \), then \( d \in D(x^*,0) \); if \( t > 0 \), then \( -2dt^2 < 0 \).) Thus the system (A), (B_4), (C) in Theorem 3 is consistent, although \( x^* = 0 \) is optimal.
4. The Fritz John and Kuhn-Tucker Theories for Semi-Infinite Programming

In this section we show how, under suitable "regularity conditions", Theorem 1 reduces to the analogs of the Fritz John and Kuhn-Tucker theories in semi-infinite programming. In the sequel we use the following concept from the duality theory of semi-infinite programming, e.g. [3].

Definition 3. Let I be an arbitrary index set, \( \{p^i : i \in I \} \) a collection of vectors in \( \mathbb{R}^m \) and \( \{c_i : i \in I \} \) a collection of scalars. The linear inequality system

\[
u^t p^i \leq c_i, \text{ for all } i \in I
\]

is canonically closed if the set of coefficients \( \{((p^i)^t, c_i) : i \in I \} \) is compact in \( \mathbb{R}^{m+1} \) and there exists a point \( u^0 \) such that

\[
(u^0)^t p^i < c_i, \text{ for all } i \in I
\]
i.e. the system has interior points.

We will say that problem (P) is canonically closed at \( x^* \) if the system

\[
(B_3). \quad d^v f^k(x^*, t) \leq 0 \quad \text{for all } t \in T_k^*, k \in P^*
\]
is canonically closed.

Remark 2. Problem (P) can have all constraint functions uniformly decreasing without being itself canonically closed. The opposite can also occur.
Lemma 3 below is a specialized version of Theorem 3 from [3], adjusted to our need. It is related to the following pair of the semi-infinite linear programs:

\[
\begin{align*}
(I) & \quad \text{Min } u^t p^0 \\
\text{s.t.} & \quad u^t p^i \geq c_i, \text{ all } i \in I \\
& \quad u \in \mathbb{R}^m
\end{align*}
\]

\[
\begin{align*}
(II) & \quad \text{Max } \sum_{i \in I} c_i \lambda_i \\
\text{s.t.} & \quad \sum_{i \in I} p^i \lambda_i = p^0 \\
& \quad \lambda \in S, \lambda \geq 0,
\end{align*}
\]

where \( S \) is the vector space of all vectors \( \{\lambda_i : i \in I\} \) with only finitely many nonzero entries.

**Lemma 3.** If the feasible set of problem (I) is nonempty and canonically closed, then problem (II) is consistent and the optimal values of the objective functions of problems (I) and (II) coincide. Furthermore, if either problem is consistent, so is the other.

The concept of a canonically closed system is used in the following theorem.

**Theorem 4.** Let \( x^* \) be a feasible solution of problem (P). Suppose that problem (P) is canonically closed at \( x^* \). Then \( x^* \) is an optimal solution of (P) if, and only if, the system

\[
(A) \quad d^t v f^0(x^*) < 0
\]
is consistent. Or dually, \( x^* \) is optimal if, and only if, there exist nonnegative scalars

\[
\lambda^0, \{\lambda^k_t : t \in T^*_k, k \in \mathcal{P}^*\}
\]

not all zero, and only finitely many of which are positive, such that

\[
\lambda^0 D^0(x^*) + \sum_{k \in \mathcal{P}^*} \sum_{t \in T^*_k} \lambda^k_t V^k(x^*, t) = 0.
\]

**Proof.** The proof consists of three parts: (i) condition (B) reduces to (B), (ii) condition (C) is redundant and (iii) the inconsistency of the system (A), (B) is equivalent to the consistency of (30).

(i). Suppose that some \( \bar{d} \) satisfies (A) and (B), i.e. that for some subsets \( \Omega_k^* \) of \( T^*_k, k \in \mathcal{P}^* \),

\[
\bar{d}^T V^0(x^*) < 0\
\bar{d}^T V^k(x^*, t) < 0 \quad \text{for all } t \in \Omega^*_k\
\bar{d} \in D_k(x^*, t) \quad \text{for all } t \in T^*_k \setminus \Omega^*_k, k \in \mathcal{P}^*.
\]

From a property of the cone \( D_k \) (see the beginning of Section 2), (33) implies

\[
\bar{d}^T V^k(x^*, t) = 0 \quad \text{for all } t \in T^*_k \setminus \Omega^*_k.
\]

By the canonical closure assumption there exists \( \hat{d} \) such that

\[
\hat{d}^T V^k(x^*, t) < 0 \quad \text{for all } t \in T^*_k, k \in \mathcal{P}^*.
\]
Now it follows from (31), (32), (33) and (35) that, for a sufficiently small \( \alpha \), the vector \( d = \tilde{d} + \alpha \tilde{d} \) satisfies (A) and (B). On the other hand, if the system (A), (B) is consistent, so is the system (A), (B). This follows by the following argument: Since (B) holds and \( T_k^* \) is compact, \( k \in P^* \), one can apply Lemma 2 to conclude that there exists \( \alpha_k > 0 \) such that

\[
f^k(x^{*} + \alpha d, t) < 0, \quad 0 < \alpha \leq \alpha_k
\]

for all \( t \in T_k^* \). By the mean value theorem, there exists \( \alpha_k^* \), \( 0 < \alpha_k^* < \alpha_k \) such that

\[
f(x^{*}, t) + \alpha_k^* d \cdot \nabla f^k(x^{*} + \alpha_k^* d, t) < 0
\]

for all \( t \in T_k^* \). Hence

\[
d \cdot \nabla f^k(x^{*} + \alpha_k^* d, t) < 0
\]

for all \( t \in T_k^* \). Therefore (B) holds with \( \alpha^* = \min \{ \alpha_k^* \}_{k \in P^*} \).

(ii) Condition (B) implies, by Lemma 1, condition (2-b), which is equivalent to (C). Thus whenever (B) holds, condition (C) is redundant.

(iii) The inconsistency of the system (A), (B) is equivalent to \( \mu^* = 0 \) being the optimal value of the semi-infinite linear program.
The dual of (I) is

(II)

\[ \text{Max } 0 \]

s.t.

\[ \lambda^0 v^0(x^*_{x^*}) + \sum_{k \in P} \sum_{t \in T_k} \lambda^k v^k(x^*, t) = 0 \]

\[ \sum_{k \in P} \sum_{t \in T_k} \lambda^k = 1, \]

\[ \lambda^k_t \geq 0, \text{ only finitely many are positive.} \]

The feasible set of problem (I) is clearly nonempty and canonically closed \((d=0, \mu=1)\) satisfy the constraints of (I) with strict inequalities). Lemma 3 is now readily applicable to the pair (I),(II), which proves (iii).

Remark 3. The dual statement in Theorem 4 is the Fritz John optimality condition for semi-infinite programming.

Remark 4. In Theorem 4 one can set \(\lambda^0 = 1\). This is seen by multiplying the equation (30) by the vector \(\hat{d}\) which satisfies (35).
The resulting condition is a semi-infinite version of the Kuhn-Tucker condition, e.g. [12].

**Remark 5.** The canonical closedness assumption is a semi-infinite version of the Arrow-Hurwicz-Uzawa constraint qualification, e.g. [12]. It is closely related to the "Constraint Qualification II" of Gehner [4].

Under different assumptions and using a considerably different approach he gives a slightly stronger dual statement than the one in our Theorem 4.

**Remark 6.** The canonical closedness property is implied by Slater's condition:

\[ \exists \hat{x} \in R^n \ \exists f^k(\hat{x}, t) < 0 \text{ for all } t \in T_k, k \in P. \]

This is well-known in the finite case and the proof is similar in the semi-infinite case.

The following example illustrates a situation in which both the Fritz John and the Kuhn-Tucker conditions fail to characterize an optimal solution since the underlying problem is not canonically closed. In contrast, our results are applicable.

**Example 6.** Consider the semi-infinite convex problem

\[ \min f^0 = x_1 - x_2 \]

s.t.

\[ f^1 = x_1^2 + tx_2 - t^2 \leq 0 \text{ for all } t \in T_1 = [0,1] \]

\[ f^2 = -x_1 - tx_2 - t \leq 0 \text{ for all } t \in T_2 = [0,1]. \]
The feasible set is
\[ F = \{ (0, x_2) : -1 \leq x_2 \leq 0 \} \]
and the optimal solution is \( x^* = (0,0)^t \). For this point
\[ T_1^* = T_2^* = \{0\}, \quad \mathcal{P}^* = \{1,2\}. \]

The system \((B_3)\) is
\[
\begin{align*}
0 &\leq 0 \\
-x_1 &\leq 0,
\end{align*}
\]

obviously not canonically closed. The Kuhn-Tucker condition is
\[
\begin{pmatrix} 1 \\ -1 \end{pmatrix} + \lambda_1 \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\]
\[ \lambda_1 \geq 0, \quad \lambda_2 \geq 0 \]

which clearly fails.

We note that the constraint functions \( f_1 \) and \( f_2 \) are uniformly decreasing at \( x^* = 0 \) in every direction \( d \neq 0 \). (The sets \( T_1^* \) and \( T_2^* \) are singletons!)
Thus Theorem 3 is applicable. Conditions (A), (B_4) and (C) are here

\[ (A) \quad d_1 - d_2 < 0 \]
\[ (B_4) \quad \begin{cases} 0 < 0 \text{ or } d_1 = 0, \ d_2 \in \mathbb{R} \\ -d_1 \leq 0 \end{cases} \]
\[ \frac{a^* d_1^2 + td_2}{-t^2} \geq -\frac{1}{a^*} \text{ for all } t \in (0,1] \]
\[ (C) \quad \begin{cases} -d_1 - td_2 \leq 0 \\ \frac{-d_1 - td_2}{-t} \geq -\frac{1}{a^*} \text{ for all } t \in (0,1]. \end{cases} \]
This reduces to

\[ d_1 = 0, \quad d_2 > 0 \]

\[ \frac{d_2}{-t} \geq - \frac{1}{a^*} \quad \text{for all } t \in (0, 1] \]

(36)

\[ -d_2 \geq - \frac{1}{a^*} . \]

Since \( d_2 > 0 \), the inequality (36) cannot hold for any \( a^* > 0 \).

Hence, by Theorem 3, \( x^* = (0, 0)^t \) is optimal.

Consider now the point \( x^* = (0, -1)^t \). Here

\[ T_1^* = \{0\}, \quad T_2^* = [0, 1], \quad p^* = \{1, 2\}. \]

It is easy to verify that the Fritz John condition is satisfied in spite of the fact that \( x^* \) is not optimal. Conditions (A), (B) and (C) are here

(A) \[ d_1 - d_2 < 0 \]

(B) \[
\begin{align*}
0 & \leq 0 \\
-d_1 - td_2 & \leq 0 \quad \text{for all } t \in [0, 1]
\end{align*}
\]

(C) \[ \frac{\alpha^* d_2^2 + td_2}{-t - t^2} \geq - \frac{1}{a^*} \quad \text{for all } t \in (0, 1]. \]

For \( a^* = 1 \), these conditions are satisfied by \( d_1 = 0, \quad d_2 = 1 \).

Hence, by Theorem 1, the point \( x^* = (0, 1)^t \) is not optimal.

Although the Fritz John and Kuhn-Tucker theories fail to characterize optimality, they can be used to formulate, respectively, either the necessary or the sufficient conditions of optimality.
THEOREM 5. ("The Fritz John Necessity Theorem") Let $x^*$ be an optimal solution of problem $(P)$. Then the system

\begin{align*}
(A) & \quad d^T v^0(x^*) < 0 \\
(B_1) & \quad d^T v^k(x^*,t) < 0 \text{ for all } t \in T_k^*, \quad k \in \mathcal{P}^*
\end{align*}

is inconsistent. Dually, the system

\begin{align*}
(PJ) & \quad \lambda^0 v^0(x^*) + \sum_{k \in \mathcal{P}^*} \sum_{t \in T_k^*} \lambda^k v^k(x^*,t) = 0 \\
& \quad \lambda^0, \{\lambda^k_t; t \in T_k^*, k \in \mathcal{P}^*\} \text{ nonnegative scalars, not all zero and of which only finitely many are positive}
\end{align*}

is consistent.

Proof. Suppose that the system $(A), (B_1)$ is consistent. Then the condition $(C)$ is redundant (see the proof of Theorem 4, part (ii)) and one contradicts the optimality of $x^*$. The dual statement is obtained as in the proof of Theorem 4, part (iii), using the fact that $v^k(x, \cdot), \quad k \in \mathcal{P}^*$ is continuous for every $x$ in some neighborhood of $x^*$.

\qed
Remark 7. The dual statement in Theorem 4 is equivalent to the one obtained in [4, Theorem 2].

THEOREM 6. ("The Kuhn-Tucker Sufficiency Theorem") Let \( x^* \) be a feasible solution of problem (P). Then \( x^* \) is optimal if the system

\[
\begin{align*}
(A) & \quad d^t Vf^o(x^*) < 0 \\
(B_5) & \quad d^t Vf^k(x^*, t) \leq 0 \text{ for all } t \in T_k^*, k \in P^* 
\end{align*}
\]

is inconsistent or, dually, if the system

\[
\begin{align*}
(K-T) \quad \left\{ Vf^o(x^*) + \sum_{k \in P^*} \sum_{t \in T_k^*} \lambda^k Vf^k(x^*, t) = 0 \right\} \\
\quad \{ \lambda^k : t \in T_k^*, k \in P^* \} \text{ nonnegative scalars of which only finitely many are positive}
\end{align*}
\]

is consistent.

Proof. If the system (A), (B) is inconsistent, so is (A), (B).
(Recall that \( D_k(x^*, t) \subset \{ d : d^t Vf^k(x^*, t) = 0 \} \) Hence, in particular, the system (A), (B), (C) is inconsistent. Following the proof of Theorem 1, one concludes that \( x^* \) is optimal. The inconsistency of (A), (B) is equivalent to the consistency of (K-T), by e.g. [11, Corollary 5].

□
Remark 8. It is not needed in Theorems 5 and 6 that the constraint functions of problem (P) be uniformly decreasing.

Remark 9. The "asymptotic" form of the Kuhn-Tucker conditions (K-T) gives a weaker sufficient condition for optimality than the familiar (i.e. without the closure) condition

\[ \begin{align*}
\forall f^0(x^*) + \sum_{k \in P^*} \sum_{t \in T_k} \lambda_k^* \nabla f^k(x^*, t) &= 0 \\
(\lambda_k^*: t \in T_k^*, k \in P^*) \text{ nonnegative scalars}
\end{align*} \]

This is demonstrated by the next example.

Example 7. Consider the following problem in \( \mathbb{R}^2 \)

\[ \begin{align*}
\text{Min } f^0 &= -x_1 \\
\text{s.t. } f^1 &= \alpha(t)x_1 + \beta(t)x_2 \leq 0 \text{ for all } t \in [0, \frac{\pi}{2}]
\end{align*} \]

where

\[ \alpha(t) = t(\frac{\pi}{2} - t) \cos t, \quad \beta(t) = t(\frac{\pi}{2} - t) \sin t. \]
Note that

\[
\begin{pmatrix}
\alpha(t) \\
\beta(t)
\end{pmatrix} \in \mathbb{R}^2_+ = \{ x \in \mathbb{R}^2 : x \geq 0 \} \quad \text{for all } t \in [0, \frac{\pi}{2}].
\]

One may easily verify that the feasible set is precisely \( F = -\mathbb{R}^2_+ \) (the negative orthant). Let \( x^* = (0,0)^t \). The consistency of \((K-T)\) means that there is a sequence \( \{\lambda^{k,i}_t\} \) such that for each \( i = 1, 2, \ldots \), \( \lambda^{k,i}_t \leq 0 \), with only finitely many negative elements, and

\[
\lim_{i \to \infty} \sum_{k \in P^*} \sum_{t \in T^*_k} \lambda^{k,i}_t \nu f^k(x^*, t) = \nu f^0(x^*).
\]

Here

\[
P = P^* = \{1\}, \quad T_1 = T_1^* = [0, \frac{\pi}{2}], \quad \nu f^0(x^*) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \nu f^1(x^*, t) = \begin{pmatrix} \alpha(t) \\ \beta(t) \end{pmatrix}.
\]

Let

\[
\lambda^{1,i}_t = \begin{cases} 
- \frac{1}{t(\frac{\pi}{2} - t)} & \text{if } t = (\frac{1}{i})^i \\
0 & \text{otherwise}.
\end{cases}
\]

Then

\[
\lim_{i \to \infty} \sum_{k \in P^*} \sum_{t \in T^*_k} \lambda^{k,i}_t \nu f^k(x^*, t) = \lim_{i \to \infty} \lambda^{1,i}_t \nu f^1(x^*, t)
\]

\[
= \lim_{i \to \infty} \begin{pmatrix} - \cos(\frac{1}{i})^i \\ - \sin(\frac{1}{i})^i \end{pmatrix}
\]

\[
= \nu f^0(x^*).
\]
Therefore \( x^* = (0,0)^t \) is optimal, by Theorem 6. The nonasymptotic Kuhn-Tucker conditions (K-T) are here

\[
(37) \quad (-1) = \sum_{t \in T_1} \lambda^1_t \begin{pmatrix} \alpha(t) \\ -\beta(t) \end{pmatrix}.
\]

Either \( \lambda^1_t = 0 \), for all \( t \in T_1 \), or \( \lambda^1_t < 0 \) for some finite subset \( \Omega \) of \( T_1 \). The first case is absurd. If the second case held one would have \( \beta(t) = 0 \) for all \( t \in \Omega \) (since \( \beta(t) \geq 0 \) for all \( t \in T_1 \)). But this is equivalent to \( \alpha(t) = 0 \) for all \( t \in \Omega \), which again contradicts (37). Thus (37) does not hold, i.e. the conditions (K-T) here fail to establish optimality.

In situations the primal Kuhn-Tucker conditions (A), (B), may be easier to apply than (K-T). This will be illustrated on the following problem taken from [8, Example 2.4].

Example 8. Consider

\[
\text{Min } f^0 = 4x_1 + \frac{2}{3}(x_4 + x_6)
\]

s.t.

\[
f^1 = -x_1 - x_2 - t_2x_3 - t_2^2x_4 - t_1t_2x_5 - t_2^2x_6 + 3 - (t_1 - t_2)^2 - (t_1 + t_2)^2 \leq 0
\]

for all \( t \in T_1 = \left\{ \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} : -1 \leq t_i \leq 1, i=1,2 \right\} \).

We will show, using the Kuhn-Tucker theory, that \( x^* = (3,0,0,0,0,0)^t \) is an optimal solution. The optimality of \( x^* \) has been established in [8] by a different approach.

First note that here

\[
T_1^* = \left\{ \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} : t_1 - t_2 = 0 \text{ or } t_1 + t_2 = 0 \right\} \cap T_1.
\]
The system \((A),(B_2)\) becomes

\[
\begin{align*}
(A) & \quad 4d_1 + \frac{2}{3}d_4 + \frac{2}{3}d_6 < 0 \\
(B_2) & \quad d_1 + t_1d_2 + t_2d_3 + t_1^2d_4 + t_1t_2d_5 + t_2^2d_6 \leq 0
\end{align*}
\]

for all \(t \in T_1^*\).

Substitute in \((B_2)\) the following five points of \(T_1^*\):

\[
(0), (1), (-1), (1), (-1).
\]

This gives

\[
\begin{align*}
-d_1 & \quad \leq 0 \\
-d_1 - d_2 - d_3 & \quad - d_4 - d_5 - d_6 \leq 0 \\
-d_1 & \quad - d_2 + d_3 - d_4 + d_5 - d_6 \leq 0 \\
-d_1 & \quad + d_2 - d_3 - d_4 + d_5 - d_6 \leq 0 \\
-d_1 & \quad + d_2 + d_3 - d_4 - d_5 - d_6 \leq 0.
\end{align*}
\]

Multiply the first inequality by \(\frac{10}{3}\) and each of the remaining four inequalities by \(\frac{1}{6}\) then add all five inequalities. We get

\[
-4d_1 - \frac{2}{3}d_4 - \frac{2}{3}d_6 \leq 0
\]

which contradicts \((A)\). Thus the system \((A),(B_2)\) is inconsistent and \(x^* = (3,0,0,0,0,0)^t\) is optimal, by Theorem 6.
5. **AN APPLICATION TO CHEBYSHEV APPROXIMATION**

It is well-known that there is a close connection between convex programming and approximation theory, e.g. [5],[13]. In fact, many approximation problems can be formulated as convex semi-infinite programming problems in which case the results of this paper are readily applicable. In particular, the problem of linear Chebyshev approximation subject to side constraints

\[
\begin{align*}
\min & \quad \max_{t \in T} |f(t) - \sum_{i=1}^{n} x_i g_i(t)| \\
\text{s.t.} & \quad t(t) \leq \sum_{i=1}^{n} x_i g_i(t) \leq u(t) \quad \text{for all } t \in T
\end{align*}
\]

is equivalent to the linear semi-infinite programming problem

\[
\begin{align*}
\min & \quad x_{n+1} \\
\text{s.t.} & \quad -x_{n+1} \leq \sum_{i=1}^{n} x_i g_i(t) - f(t) \leq x_{n+1} \\
& \quad t(t) \leq \sum_{i=1}^{n} x_i g_i(t) \leq u(t) \quad \text{for all } t \in T.
\end{align*}
\]

Corollary 3 of this paper can be applied to (L) and it gives a characterization of the best approximation for the problem (MM). Uniqueness of the best approximation can be checked using Theorem 2. Rather than going into details we will illustrate this application by an example.
Example 9. The approximation problem stated in this example is taken from [4], see also [15]. It shows that situations when the Kuhn-Tucker theory for semi-infinite programming fails to establish optimum even in the case of linear constraints. However the optimality is established using the results of this paper.

The linear Chebyshev approximation problem is

\[
\begin{align*}
\text{Min} \left( \max_{t \in [0,1]} |t^4 - x_1 - x_2t| \right) \\
\text{s.t.} \\
-t \leq x_1 + x_2t \leq t^2, \text{ for all } t \in [0,1].
\end{align*}
\]

An equivalent linear semi-infinite programming problem is

\[
\begin{align*}
\text{Min} \ f^0 = x_3 \\
\text{s.t.} \\
f^1 = t^4 - x_1 - x_2t - x_3 \leq 0 \\
f^2 = -t^4 + x_1 + x_2t - x_3 \leq 0 \\
f^3 = -t^2 + x_1 + x_2t \leq 0 \\
f^4 = -t - x_1 - x_2t \leq 0
\end{align*}
\]

Is \( x^* = (0,0,1)^t \) optimal?
Here $T_1^* = \{1\}, \ T_2^* = \emptyset, \ T_3^* = \{0\}, \ T_4^* = \{0\}$ and $p^* = \{1,3,4\}$.

The system $(A),(B_5)$ is

$$(A) \quad d_3 < 0$$
$$d_1 \leq 0$$
$$d_1 \leq 0$$

and it is clearly consistent (set e.g. $d_1 = 0$, $d_2 = 1$, $d_3 = -1$).

Therefore Theorem 6 cannot be applied. (Since the system $(K-T)$ is inconsistent, $x^* = (0,0,1)^t$ is not a "Kuhn-Tucker point".)

But the system

$$(A) \quad d_3 < 0$$
$$d_1 \leq 0$$
$$d_1 \leq 0$$

$$\frac{-d_1 - d_2 t - d_3}{t^4 - 1} \geq -1, \text{ for all } t \in [0,1)$$

$$(C_1) \quad \frac{d_1 + d_2 t}{t} \geq -1, \text{ for all } t \in (0,1]$$
is inconsistent. (First, \( d_1 = 0 \), by the last two inequalities in \((B_2)\). Now \((A)\) and \((B_2)\) imply \( d_2 > 0 \). This contradicts \( d_2 \leq 0 \) obtained from the second inequality in \((C_1)\).) Therefore \( x^* = (0, 0, 1)^T \) is optimal, by Corollary 3.

**Acknowledgment.** The authors thank Professor G. Schmidt for providing some of the constraint functions used in Examples 1, 2 and 4.
References


