NECESSARY AND SUFFICIENT CONDITIONS
IN CONSTRAINED OPTIMIZATION

by

M.A. Hanson

Department of Statistics
Florida State University
Tallahassee, Florida 32306
USA

and

B. Mond

Department of Pure Mathematics
La Trobe University
Bundoora 3083
Melbourne, Australia

FSU Statistics Report M683

November, 1984

The Florida State University
Department of Statistics
Tallahassee, Florida 32306
1. INTRODUCTION

In inequality constrained optimization the Kuhn-Tucker conditions are sufficient for optimality if the functions involved are convex. Hanson [1] and Mond and Hanson [2], [3] have defined more general classes of functions, now known as invex functions, for which the Kuhn-Tucker conditions are sufficient for optimality. Although in the context of the Kuhn-Tucker Theory invex functions are sufficient, they are not necessary for optimality.

This raises the question of whether one can make an extended generalization of convexity which has this sufficiency property and which is also essential in the context of the Kuhn-Tucker theory.

The Kuhn-Tucker theory, being a generalization of the traditional Lagrangian theory of constrained optimization, requires a certain linear relationship among the objective and constraint functions at optimum. We show in this paper that additional linear relationships are required in general.

We give an answer to the preceding question by introducing two new classes of functions, both closely related to, but more general than, invex functions, which are not only sufficient but are also necessary for optimality in primal and dual problems respectively, under very general conditions.

We also examine converse duality, and the question of sufficient conditions for optimality of the dual problem, without reference to the feasibility of the primal problem.
Notation and Preliminaries

Let $f(x)$ be a differentiable scalar function and $g(x)$ an $m$-dimensional differentiable vector function defined on an open set $x \in \mathbb{R}^n$. Consider the problems:

**(Primal)**

Minimize $f(x)$ \hspace{1cm} (1.1)

$x \in X$

Subject to $g(x) \leq 0$ ; \hspace{1cm} (1.2)

**(Dual)**

Maximize $\psi(x,y)$ \hspace{1cm} (1.3)

$(x,y) \in Y$

Subject to $\nabla_x \psi(x,y) = 0$ \hspace{1cm} (1.4)

and $y \geq 0$ \hspace{1cm} (1.5)

where $\psi(x,y) = f(x) + [g(x)]^T y$

and $Y = \{(x,y) | x \in X, y \in \mathbb{R}^m, \nabla_x \psi(x,y) = 0, y \geq 0 \}$.

Let $D = \{x | (x,y) \in Y \}$ and $P = \{x | x \in X, g(x) \leq 0 \}$.

Associated with the two problems are primal constraint qualifications and dual constraint qualifications, respectively (see Mangasarian [4]).

Given a primal constraint qualification it is necessary, according to the Kuhn-Tucker Theorem, that for $x_0 \in P$ to be a local optimum in the primal problem there exists a vector $y_0 \in \mathbb{R}^m$ such that

$$\nabla_x f(x_0) + [\nabla_x g(x_0)]^T y_0 = 0,$$  \hspace{1cm} (1.6)

$$[g(x_0)]^T y_0 = 0,$$ \hspace{1cm} (1.7)

and $y_0 \geq 0$, \hspace{1cm} (1.8)

where $\nabla_x g(x_0)$ is the Jacobian of $g(x)$ at $x_0$. Since $g(x_0) \leq 0$ and $y_0 \geq 0$ it follows from (1.7) that if any component of $g(x_0)$ is nonzero then the
The corresponding component of \( y_0 \) is zero. Any component of \( g(x_0) \) which is zero is said to be an active constraint function at \( x_0 \).

For the dual problem there is the constraint qualification (Hanson [5]) that the dual constraints be such that there exist an open set \( W \in \mathbb{R}^m \) containing \( y_0 \) and an \( n \)-dimensional differentiable vector function \( e(y) : \mathbb{W} \to \mathbb{X} \) such that
\[
    x_0 = e(y_0) \quad \text{and} \quad \nabla_x \psi(x,y) \big|_{x = e(y)} = 0 \quad \text{for} \ y \in \mathbb{W},
\]
where \( (x_0, y_0) \) is optimal in the dual problem.

If \( (x_0, y_0) \) is optimal in the dual problem it follows from this dual constraint qualification that \( (x_0, y_0) \) satisfies the Kuhn-Tucker conditions (1.6) - (1.8) and \( x_0 \in P \).

If \( (x, y_0) \) is twice continuously differentiable at \( x_0 \) a special case of this constraint qualification (Huard [6]) is that the Hessian matrix \( \nabla_x^2 \psi(x_0, y_0) \) be nonsingular.

We shall say that \( f(x) \) and \( g(x) \) are Type I objective and constraint functions, respectively, with respect to \( \eta(x) \) at \( x_0 \) if there exists an \( n \)-dimensional vector function \( \eta(x) \) defined for all \( x \in P \) such that
\[
    f(x) - f(x_0) \geq [\nabla_x f(x_0)]^T \eta(x), \quad (1.9)
\]
and
\[
    -g(x_0) \geq [\nabla_x g(x_0)] \quad \eta(x); \quad (1.10)
\]
and \( f(x) \) and \( g(x) \) are Type II objective and constraint functions, respectively, with respect to \( \eta(x) \) at \( x_0 \) if there exists an \( n \)-dimensional vector function \( \eta(x) \) defined for all \( x \in D \) such that
\[
    f(x_0) - f(x) \geq [\nabla_x f(x)]^T \eta(x), \quad (1.11)
\]
and
\[
    -g(x) \geq [\nabla_x g(x)] \quad \eta(x). \quad (1.12)
\]

2. THE PRIMAL PROBLEM

**THEOREM 2.1.** For \( x_0 \in P \) to be optimal in the primal problem it is sufficient that \( f(x) \) and the active components of \( g(x) \) at \( x_0 \) are Type I functions with respect to a common \( \eta(x) \) at \( x_0 \) and conditions (1.6) - (1.8) hold at \( x_0 \) for some \( y_0 \).
Proof. Let \( x \in P \) be any vector satisfying the constraints of the primal problem. Then

\[
f(x) - f(x_0) \geq [\nabla_x f(x_0)]^T \eta(x), \text{ since } f \text{ is Type I,}
\]

\[
= -y_0^T \nabla x g(x_0) \eta(x), \text{ by (1.6),}
\]

\[
\geq y_0 \eta(x_0), \text{ by (1.5) since the components of } g(x_0)
\]

are Type I or not active,

\[
= 0, \text{ by (1.8).}
\]

So \( x_0 \) is optimal in the primal problem, which proves the theorem.

We note that if \( \eta(x) \) is identically zero then (1.9) and (1.10) become definitions of feasibility and optimality for the primal problem. It follows that if a primal constraint qualification holds, then for \( x_0 \in P \) to be optimal in the primal problem it is necessary as well as sufficient that conditions (1.6) - (1.8) hold at \( x_0 \) for some \( y_0 \) and that \( f(x) \) and \( g(x) \) are Type I functions with respect to a common \( \eta(x) \) at \( x_0 \) and that conditions (1.6) - (1.8) hold at \( x_0 \) for some \( y_0 \). However in the following Theorem we see that for optimality, under an additional condition there must exist an \( \eta(x) \) which is not identically zero for any feasible \( x \).

THEOREM 2.2. If \( x_0 \in P \) and the number of active constraints at \( x_0 \) is \( k \), where \( k < n - 1 \), then for \( x_0 \) to be optimal in the primal problem it is necessary that \( f(x) \) and \( g(x) \) are Type I functions with respect to a common \( \eta(x) \) at \( x_0 \), not identically zero for any \( x \).

Proof. Let \( x_0 \) be optimal in the primal problem, and let \( x \) be any feasible point in the primal problem.

Suppose that \( \eta(x) \equiv 0 \) is the only solution of the system

\[
f(x) - f(x_0) \geq [\nabla_x f(x_0)]^T \eta(x)
\]

\[
- g(x_0) \geq [\nabla_x g(x_0)] \eta(x);
\]
that is, 
\[ a_n \leq b \Rightarrow c^T n = 0 \text{ for all } c \in \mathbb{R}^n, \]
where
\[ A = \begin{bmatrix} \nabla_x f(x_0) \end{bmatrix}^T \quad \text{and} \quad b = \begin{bmatrix} f(x) - f(x_0) \\ g(x) - g(x_0) \end{bmatrix}. \]

So by the nonhomogeneous Farkas Theorem (see for example Thrall and Tornheim [7]) there exists \( z \in \mathbb{R}^{m+1} \) and that
\[ A^T z = c, \quad b^T z \leq 0, \quad z \geq 0; \]
that is, there exist \( z_1 \in \mathbb{R}^1 \) and \( z_2 \in \mathbb{R}^m \) such that
\[ \nabla_x f(x_0) z_1 + \nabla_x g(x_0) z_2 = c \]
\[ (f(x) - f(x_0) z_1 - g(x_0) z_2 \leq 0 \]
\[ z_1 \geq 0 \]
\[ z_2 \geq 0; \]
that is, since \( f(x) - f(x_0) \geq 0, -g(x_0) \geq 0, z_1 \geq 0, z_2 \geq 0, \)
\[ \nabla_x f(x_0) z_1 + \nabla_x g(x_0) z_2 = c \quad (2.1) \]
\[ (f(x) - f(x_0) z_1 = 0 \quad (2.2) \]
\[ -g(x_0) z_2 = 0 \quad (2.3) \]
\[ z_1 \geq 0 \quad (2.4) \]
\[ z_2 \geq 0. \quad (2.5) \]

Since \(-g(x_0) \geq 0\) and \(z_2 \geq 0\) it follows from (2.3) that if any component of the vector \( g(x_0) \) is nonzero then the corresponding component of \( z_2 \) is zero. By hypothesis, there are \( m-k \) such components. So the corresponding \( m-k \) columns of \( \nabla_x g(x_0) \) can be deleted from the \( nx(1+m) \) matrix \( A^T \equiv [\nabla_x f(x_0) [\nabla_x g(x_0)]^T] \),
without affecting the system (2.1)-(2.5), leaving a modified $n \times (1+k)$ matrix
where greatest possible rank is $1+k$ since $n > 1 + k$.

But $c$ has dimension $n$, and is arbitrary, and so can be chosen to violate
the over-determined modified system of linear equations (1).

So we have a contradiction to the supposition that $\eta(x) \equiv 0$ is the only
solution of the original system.

**Remark.** If $m < n - 1$ then Theorem 2.2 is applicable even if $x_0$ is not known,
provided it exists.

Theorems 2.1 and 2.2 can be combined to give a necessary and sufficient
Theorem for $x_0$ to be optimal in the primal problem:

**Theorem 2.3.** If $x_0 \in P$, conditions (1.6)-(1.8) hold, and $k < n - 1$, then for
$x_0$ to be optimal in the primal problem it is necessary and sufficient that
$f(x)$ and $g(x)$ are Type I functions with respect to a common $\eta(x)$ at $x_0$, not
identically zero for any $x$.

3. THE DUAL PROBLEM

**Theorem 3.1.** If $x_0 \in P$, or more generally if $y_0^T g(x_0) \leq 0$, then for $(x_0,y_0)$
to be optimal in the dual problem it is necessary that $f(x)$ and $g(x)$ are
Type II functions with respect to a common $\eta(x)$ at $x_0$.

**Proof.** Let $(x_0,y_0)$ be optimal in the dual problem. Then for any $(x,y)$
feasible in the dual problem we can write

$$
\nabla_x f(x) + [\nabla_x g(x)]^T y = 0, \; y \geq 0 \implies f(x_0) + [g(x_0)]^T y_0 \geq f(x) + [g(x)]^T y,
$$

that is

$$
Ay \leq b \implies c^T y \leq m, \quad (3.1)
$$

where

$$
A = \begin{bmatrix}
[\nabla_x g(x)]^T \\
[-\nabla_x g(x)]^T \\
-1
\end{bmatrix}, \quad b = \begin{bmatrix}
-\nabla_x f(x) \\
\nabla_x f(x) \\
0
\end{bmatrix}
$$

$$
c = g(x), \; m = -f(x) + f(x_0) + [g(x_0)]^T y_0.
$$
Since \((x_0, y_0) \in D\), the system \(Ax \leq b\) is not empty. So by the nonhomogeneous Farkas Theorem (3.1) holds if and only if there exists \(z \in \mathbb{R}^{2n+m}\) such that
\[
A^Tz = c, \quad b^Tz \leq m, \quad z \geq 0,
\]
that is, there exist \(z_1 \in \mathbb{R}^n, z_2 \in \mathbb{R}^n, z_3 \in \mathbb{R}^m\) such that
\[
\begin{bmatrix}
\nabla_x g(x) & -\nabla_x g(x) & -1
\end{bmatrix}
\begin{bmatrix}
z_1 \\
z_2 \\
z_3
\end{bmatrix} = g(x)
\]
\[
\begin{bmatrix}
-\nabla_x f(x)
\end{bmatrix}^T
\begin{bmatrix}
\nabla_x f(x)
\end{bmatrix}^T
\begin{bmatrix}
z_1 \\
z_2 \\
z_3
\end{bmatrix} = -f(x) + f(x_0) + [g(x_0)]^Ty_0
\]
and
\[
\begin{bmatrix}
z_1 \\
z_2 \\
z_3
\end{bmatrix} \geq 0.
\]
So
\[
[\nabla_x g(x)](z_1 - z_2) \geq g(x)
\]
\[
[\nabla_x f(x)]^T(z_1 - z_2) = -f(x) + f(x_0) + [g(x_0)]^Ty_0
\]
\[
\leq -f(x) + f(x_0), \text{ since } y_0^Tg(x_0) \leq 0.
\]
Putting \(z_1 - z_2 \equiv \eta(x)\) we have
\[
[\nabla_x g(x)]\eta(x) \leq -g(x)
\]
and \([\nabla_x f(x)]^T\eta(x) \leq -f(x) + f(x_0)\).
Thus \(f(x)\) and \(g(x)\) are Type II functions with respect to \(\eta(x)\).

**Theorem 3.2.** For \((x_0, y_0)\) to be optimal in the dual problem it is sufficient that \(f(x)\) and \(g(x)\) are Type II functions with respect to a common \(\eta(x)\) at \(x_0\) and conditions (1.6) - (1.8) hold at \((x_0, y_0)\).
Proof. By (1.6) and (1.8), \((x_0, y_0)\) satisfies the constraints of the dual problem. Let \((x, y) \in Y\) be any vector satisfying the constraints of the dual problem. Then

\[
\psi(x_0, y_0) - \psi(x, y) = f(x_0) + [g(x_0)]^T y_0 - f(x) - [g(x)]^T y
\]

\[
= f(x_0) - f(x) - [g(x)]^T y, \quad \text{by (1.7)},
\]

\[
\geq [\nabla_x f(x)]^T \eta(x) - [g(x)]^T y, \quad \text{since } f \text{ is Type II},
\]

\[
= -y^T [\nabla_x g(x)] \eta(x) - [g(x)]^T y, \quad \text{by (1.4)},
\]

\[
y^T (-g(x) - [\nabla_x g(x)] \eta(x)),
\]

\[
\geq 0, \quad \text{by (1.5) since } g \text{ is Type II}.
\]

So \((x_0, y_0)\) is optimal in the dual problem.

4. CONVERSE DUALITY

In the following theorem the concept of dual constraint qualification is used in the sense of Hanson or Huard (See the Introduction—Notation and Preliminaries). This is different from the usual terminology of "constraint qualification" which in this paper is called "primal constraint qualification".

THEOREM 4.1. If a dual constraint qualification holds at \((x_0, y_0)\), where \((x_0, y_0)\) is optimal in the dual problem, then a necessary and sufficient condition for \(x_0\) to be optimal in the primal problem is that \(f(x)\) and \(g(x)\) are Type I functions with respect to a common \(\eta(x)\) at \(x_0\). At optimum the two objective functions are equal.

Proof. The necessity follows as in Section 2. For the sufficiency we have that since \((x_0, y_0)\) is optimal in the dual problem and a dual constraint qualification holds at \((x_0, y_0)\) then \((x_0, y_0)\) satisfies the Kuhn-Tucker conditions (1.6)-(1.8) and \(x_0 \in P\). So the sufficiency follows from Theorem 2.1 since \(f(x)\) and \(g(x)\) are Type I functions with respect to \(\eta(x)\) at \(x_0\). By (1.7) the primal and dual objective functions are equal at their respective optima.
The results of Sections 2 and 3 can be combined into the following:

THEOREM 4.2. If $x_0 \in P$, $(x_0, y_0) \in Y$, $k < n - 1$, and $f(x_0) = \psi(x_0, y_0)$, then for $x_0$ and $(x_0, y_0)$ to be optimal in the two respective problems it is necessary and sufficient that $f(x)$ and $g(x)$ are both Type I functions with respect to some common $\eta_I(x)$ at $x_0$, not identically zero, and Type II functions with respect to some common $\eta_{II}(x)$ at $x_0$.

Proof. By Theorem 2.2, for $x_0$ to be optimal in the primal problem it is necessary that $f(x)$ and $g(x)$ are Type I functions at $x_0$. Since $(x_0, y_0) \in Y$ then (1.6) and (1.8) hold. Since the objective functions are equal equation (1.7) holds. Hence, since $x_0 \in P$, for $x_0$ to be optimal in the primal problem it is sufficient that $f(x)$ and $g(x)$ are Type I functions at $x_0$, by Theorem 2.1.

For $(x_0, y_0)$ to be optimal in the dual problem it is necessary by Theorem 3.1, since $x_0 \in P$, that $f(x)$ and $g(x)$ are Type II functions at $x_0$; and it is sufficient by Theorem 3.2, since (1.6) - (1.8) hold at $(x_0, y_0)$, that $f(x)$ and $g(x)$ are Type II functions at $x_0$. 
5. REFERENCES


