Necessary and Sufficient Conditions for Global Optimality in Constrained Optimization

by

M. A. Hanson

Department of Statistics, Florida State University, Tallahassee, Florida 32306

and

B. Mond

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

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

FSU Statistics Report M-600

November, 1981
1. INTRODUCTION

We consider the problem:

\[
\min_{x \in \mathcal{C}} f(x) \tag{1.1}
\]

subject to \( g(x) \leq 0 \), \( \tag{1.2} \)

where \( f(x) \) and the vector \( g(x) = (g_1(x), \ldots, g_m(x))^t \) are differentiable functions defined on a set \( \mathcal{C} \subset \mathbb{R}^n \). The Kuhn-Tucker theorem gives a set of necessary conditions for a point \( x_0 \) to be locally minimal in this problem; namely, under certain constraint qualifications, which will be discussed later in this section, it is necessary that there exist a vector \( y_0 \in \mathbb{R}^m \) such that

\[
\nabla f(x_0) + y_0^t g(x_0) = 0, \tag{1.3}
\]

\[
y_0^t g(x_0) = 0, \tag{1.4}
\]

\[
y_0 \geq 0. \tag{1.5}
\]

In general these conditions are not sufficient for \( x_0 \) to be minimal. Kuhn and Tucker showed that they are sufficient if the functions \( f(x) \) and \( g(x) \) are convex; and various other authors, for example Mangasarian [1], Hanson [2], Hanson and Mond [3], have defined wider classes of functions \( f(x) \) and \( g(x) \) for which the Kuhn-Tucker conditions are sufficient.

In the generalization of convexity in mathematical programming it is clear that since global results over some region are sought through use of a local concept, namely that of derivative, it is necessary to
establish some relationship between the difference operator $\Delta_X$, which we will define through the relationship $\Delta_X f(x_0) = f(x) - f(x_0)$, and the differential operator $\nabla_X$. (In the following the subscript $x$ will be implicitly assumed.)

The abovementioned definitions of convexity and its generalizations are designed precisely for this purpose in special cases; but, though they introduce sufficiency, they are not necessary definitions.

In this paper we now address the problem of finding such a necessary and sufficient definition by establishing a general relationship between $\Delta$ and $\nabla$ appropriate to the use of the Kuhn-Tucker theorem in global optimization. Specifically we find additional necessary conditions for optimality involving $\Delta$ and $\nabla$ which in conjunction with the Kuhn-Tucker conditions form a set of necessary and sufficient conditions for the solution of (1.1)-(1.2).

Let the index set of active constraints at $x_0$ be denoted by $K$, that is, $K = \{ i \mid g_i(x_0) = 0 \}$. Thus by (1.2), (1.4), and (1.5), $y_{0i} = 0$ if $i \notin K$ and (1.3)-(1.5) can be written

$$\nabla f(x_0) + \sum_{i \in K} \nabla y_{0i} g_i(x_0) = 0 , \quad (1.6)$$

$$y_{0i} g_i(x_0) = 0 , i \in K, \quad (1.7)$$

$$y_{0i} \geq 0 , i \in K. \quad (1.8)$$

Without loss of generality we suppose that the constraints are labelled so that $K = \{1, 2, \ldots, k\}$. In addition to the Kuhn-Tucker conditions
it is necessary to specify a constraint qualification. One of these is the modified Arrow-Hurwicz-Uzawa constraint qualification (see [4, p. 172]), which, for our purposes, is that there exists a vector $r \in \mathbb{R}^n$ such that

$$
\begin{bmatrix}
(vg_1(x_0))^t \\
(vg_2(x_0))^t \\
\vdots \\
(vg_k(x_0))^t
\end{bmatrix} \begin{bmatrix} r < 0. \end{bmatrix} \quad (1.9)
$$

We shall say that the vectors $a_1, a_2, \ldots, a_k$ are nonnegatively linearly independent if, for the scalars $t_1 \geq t_2 \geq 0, \ldots, t_k \geq 0$,

$$
t_1 a_1 + t_2 a_2 + \ldots + t_k a_k = 0 \text{ implies } t_1 = t_2 = \ldots = t_k = 0.
$$

In [5, p. 47] it is shown that a constraint qualification for convex programs is that the gradient vectors of the active constraints at $x_0$ be nonnegatively linearly independent.

We note that the proof in [5, p. 47] of this constraint qualification holds as well for nonconvex programs (see also [6]). Indeed, for our problem, nonnegative linear independence of the gradient vectors of the active constraints is equivalent to the modified Arrow-Hurwicz-Uzawa qualification as given in (1.9).

This follows from Gordon's theorem of the alternative (see [4, p. 34]):
Either the system (1.9) has a solution \( r \), or the system

\[
\begin{bmatrix}
  v_{g_1}(x_0) & v_{g_2}(x_0) & \ldots & v_{g_k}(x_0)
\end{bmatrix} s = 0,
\]

\( s \geq 0 \), not all \( s_i = 0 \),

has a solution \( s \in \mathbb{E}^k \), but not both.

The point \( x_0 \) will be said to be regular if the active constraints at \( x_0 \) are nonnegatively linearly independent, or, equivalently, satisfy the modified Arrow-Hurwicz-Uzawa qualification.

2. NECESSARY AND SUFFICIENT CONDITIONS FOR A GLOBAL MINIMUM

THEOREM 2.1. Let \( f(x), g_1(x), \ldots, g_m(x) \) be differentiable functions defined on \( C \subseteq \mathbb{E}^n \). For a regular point \( x_0 \) to be a global minimum in problem (1.1)-(1.2) it is necessary and sufficient that there exist a vector \( y_0 \in \mathbb{E}^m \), a vector function \( \eta(x, x_0) \in \mathbb{E}^n \) and a nonnegative scalar function \( \alpha(x, x_0) \), both defined for \( x \in C, x_0 \in C \), such that in addition to (1.3)-(1.5) the following conditions are satisfied for \( \{ x \mid g(x) \leq 0 \} \):

\[
f(x) - f(x_0) \geq \alpha(x, x_0) \eta^t(x, x_0) \forall f(x_0), \quad (2.1)
\]

where equality holds if \( \eta^t(x, x_0) \forall f(x_0) \neq 0 \),

and

\[
g_i(x) \geq \eta^t(x, x_0) \forall g_i(x_0), \quad i \in K. \quad (2.2)
\]

Proof. (Sufficiency) We have, for any \( x \in C \) satisfying \( g(x) \leq 0 \),

...
\[ f(x) - f(x_0) \geq \alpha(x, x_0) \, \eta^t(x, x_0) \, \forall f(x_0), \]  
\text{by (2.1)}
\[ = -\alpha(x, x_0) \, \eta^t(x, x_0) \, \sum_{i \in K} \nu_{o_i} g_i(x_0) , \text{by (1.6)} \]
\[ \geq -\alpha(x, x_0) \sum_{i \in K} \nu_{o_i} g_i(x) , \quad \text{by (2.2), since } \alpha(x, x_0) \geq 0 \]
\[ \geq 0 \quad , \text{by (1.2) and (1.8)}. \]

Hence \( x_0 \) is a global minimum.

(Necessity) Suppose \( x_0 \) is a global minimum. Then by the Kuhn-Tucker theorem it is necessary that there exist \( y_0 \) satisfying (1.5)-(1.5). It remains to be shown that, for this \( y_0 \), there exist \( \eta(x, x_0) \) and \( \alpha(x, x_0) \) such that (2.1) and (2.2) are satisfied.

Since \( x_0 \) is regular there exists a vector \( r \in \mathbb{R}^n \) such that

\[ [v_{g_i}(x_0)]^t r < 0 \quad , \quad i \in K. \]

Also
\[ g_i(x) \leq 0 \quad , \quad \text{by (1.2)}. \]

So there exists a positive function \( \lambda(x, x_0) \) sufficiently large that

\[ g_i(x) \geq [v_{g_i}(x_0)]^t r \lambda(x, x_0) , \quad i \in K. \]

Putting \( r \lambda(x, x_0) = \eta(x, x_0) \) we have

\[ g_i(x) \geq \eta^t(x, x_0) \, v_{g_i}(x_0) , \quad i \in K, \quad (2.3) \]

which is (2.2).

Now, by (1.6),

\[ \eta^t(x, x_0) \, \forall f(x_0) = -\eta^t(x, x_0) \, \sum_{i \in K} \nu_{o_i} g_i(x_0), \]
\[ \sum_{i \in K} y_{oi} g_i(x) \leq \sum_{i \in K} y_{oi} \alpha(x, x_o), \text{ by (2.3)}, \]
\[ \geq 0 \quad , \text{ by (1.2) and (1.8)}. \]  

(2.4)

Also \( f(x) - f(x_o) \geq 0 \) since \( x_o \) is a global minimum.

Hence \( \alpha(x, x_o) \geq 0 \) can be chosen so that

\[ f(x) - f(x_o) \geq \alpha(x, x_o) \eta^t(x, x_o) \forall f(x_o) \]

where equality holds if \( \eta^t(x, x_o) \forall f(x_o) = 0 \),

which is (2.1).

So the theorem is proved.

The conditions (2.1) and (2.2) can be written in the form:

\[ \Delta f(x_o) \geq \alpha(x, x_o) \eta^t(x, x_o) \forall f(x_o) \]

and

\[ \Delta g_i(x_o) \geq \eta^t(x, x_o) \forall g_i(x_o), \quad i \in K. \]

Note that linear independence is a special case of nonnegative linear independence. Clearly if \( \forall g_i(x_o), i \in K \), are linearly independent they are also nonnegatively linearly independent, and theorem 2.1 applies. It is possible, however, in the case of linear independence to obtain stronger necessary and sufficient conditions.

**Corollary 2.1.** If the gradient vectors \( \forall g_i(x_o), i \in K \), are linearly independent then theorem 2.1 holds where (2.2) is replaced by

\[ g_i(x) = \eta^t(x, x_o) \forall g_i(x_o), \quad i \in K. \]  

(2.5)

**Proof.**

Since the gradient vectors \( \forall g_i(x_o), i \in K \), are linearly independent they are nonnegatively linearly independent and \( x_o \) is regular. So theorem 2.1 holds, and the following system does not have a solution:

\[ [\forall g_1(x_o) \quad \forall g_2(x_o) \quad \ldots \quad \forall g_k(x_o)] s = 0, \text{ not all } s_i = 0. \]
So the following system does not have a solution:

\[
\begin{bmatrix}
\nabla g_1(x_o) & \nabla g_2(x_o) & \ldots & \nabla g_k(x_o)
\end{bmatrix} s = 0,
\]

\[
\begin{bmatrix}
g_1(x) & g_2(x) & \ldots & g_k(x)
\end{bmatrix} s = 1.
\]

Hence by Gale's theorem of the alternative (see [4, p. 34]) there exists a vector \(\eta(x, x_o)\) such that

\[
g_i(x) = \eta^t(x, x_o) \nabla g_i(x_o), \quad i \in K.
\]

3. NECESSARY CONDITIONS AND SUFFICIENT CONDITIONS

From (2.4) we see that if \(x_o\) is a global minimum, that is, \(f(x) - f(x_o) \geq 0\) then \(\eta^t(x, x_o) \nabla f(x_o) \geq 0\). So we have the necessary condition, analogous to the definition of quasiconcave function,

\[
f(x) - f(x_o) \geq 0 \implies \eta^t(x, x_o) \nabla f(x_o) \geq 0,
\]

and we have the following:

THEOREM 3.1. Let \(f(x), g_1(x), \ldots, g_m(x)\) be differentiable functions on \(C \subseteq \mathbb{R}^n\). For a regular point \(x_o\) to be a global minimum in problem (1.1) - (1.2), it is necessary that there exist a vector \(y_o \in \mathbb{R}^m\) and a vector function \(\eta(x, x_o) \in \mathbb{R}^n\) defined for \(x \in C\), \(x_o \in C\), such that in addition to (1.3) - (1.5) the following conditions are satisfied for \(\{x | g(x) \leq 0\}\):

\[
f(x) - f(x_o) \geq 0 \implies \eta^t(x, x_o) \nabla f(x_o) \geq 0, \quad (3.1)
\]

and

\[
g_i(x) \geq \eta^t(x, x_o) \nabla g_i(x_o), \quad i \in K. \quad (3.2)
\]
COROLLARY 3.1. If the gradient vectors $\nabla g_i(x_o)$, $i \in K$, are linearly independent then theorem 3.1 holds where (3.2) is replaced by

$$g_i(x) = \eta^t(x, x_o) \nabla g_i(x_o), \quad i \in K.$$ 

On the other hand, if we seek a corresponding result for sufficiency we obtain the following condition, analogous to the definition of pseudoconvex function,

$$\eta^t(x, x_o) \nabla f(x_o) \geq 0 \implies f(x) - f(x_o) \geq 0.$$ 

THEOREM 3.2. Let $f(x)$, $g_1(x)$, ..., $g_m(x)$ be differentiable functions on $C \subset E^n$. For a regular point $x_o$ to be a global minimum in problem (1.1)-(1.2) it is sufficient that there exist a vector $y_o \in E^m$ and a vector function $\eta(x, x_o) \in E^n$ defined for $x \in C$, $x_o \in C$, such that, in addition to (1.3) - (1.5) the following conditions are satisfied for $\{x | g(x) \leq 0\}$:

$$\eta^t(x, x_o) \nabla f(x_o) \geq 0 \implies f(x) - f(x_o) \geq 0, \quad (3.3)$$

$$g_i(x) \geq \eta^t(x, x_o) \nabla g_i(x_o), \quad i \in K. \quad (3.4)$$

Proof. Let $x \in C$ be any point satisfying $g(x) \leq 0$.

By (1.2) and (1.8),

$$0 \leq - \sum_{i \in K} y_{oi} g_i(x)$$

$$\leq - \eta^t(x, x_o) \sum_{i \in K} y_{oi} g_i(x_o), \text{ by (3.4)},$$

$$= \eta^t(x, x_o) \nabla f(x_o), \text{ by (1.6)}.$$
Hence by (3.3),

$$f(x) - f(x_0) \geq 0,$$

that is, $x_0$ is a global minimum.

**COROLLARY 3.2.** If the gradient vectors $\nabla g_i(x_0), i \in K$, are linearly independent then theorem 3.2 holds where (3.4) is replaced by

$$g_i(x) = n^*(x, x_0) \nabla g_i(x_0), i \in K.$$

4. **EXAMPLE**

The following example illustrates the existence of functions satisfying the requirements of corollary 2.1 and, a fortiori, theorems 2.1, 3.1, 3.2 and corollaries 3.1, 3.2.

Minimize $f(x) \equiv -2x_2^3 - 6x_1^2 + 3x_2^2 + 6x_1 + 6x_2 - 7$, subject to

$$g_1(x) \equiv -3x_1^4 + x_2^3 - 3x_1 - 3x_2 + 2 \leq 0,$$

$$g_2(x) \equiv 2x_1^4 + 2x_1^2 - x_2^2 + 1 \leq 0,$$

$$g_3(x) \equiv 2x_1x_2 - 6x_1 - 1 \leq 0,$$

where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. 
It can be seen that this problem has a Kuhn-Tucker point at \( x_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \). Let \( C \) be the constraint region \( \{x | g_1(x) \leq 0, g_2(x) \leq 0, g_3(x) \leq 0\} \). At \( x_0 \) the constraint \( g_3(x) \leq 0 \) is not active. We have

\[
\nabla f(x_0) = \begin{bmatrix}
-12x_1 + 6 \\
-6x_2^2 + 6x_2 + 6
\end{bmatrix}
= \begin{bmatrix} 6 \\ 6 \end{bmatrix},
\quad \text{x} = x_0
\]

\[
\nabla g_1(x_0) = \begin{bmatrix}
-12x_1^3 - 3 \\
3x_2^2 - 3
\end{bmatrix}
= \begin{bmatrix} -3 \\ 0 \end{bmatrix},
\quad \text{x} = x_0
\]

\[
\nabla g_2(x_0) = \begin{bmatrix}
8x_1^3 + 4x_1 \\
-2x_2
\end{bmatrix}
= \begin{bmatrix} 0 \\ -2 \end{bmatrix}.
\quad \text{x} = x_0
\]

So \( \nabla g_1(x_0) \) and \( \nabla g_2(x_0) \) are linearly independent vectors.

By corollary 2.1,

\[
g_1(x) = n^t(x, x_0) \nabla g_1(x_0)
\]

and

\[
g_2(x) = n^t(x, x_0) \nabla g_2(x_0).
\]

Hence

\[
n(x, x_0) = \begin{bmatrix}
(\nabla g_1(x_0))^t \\
(\nabla g_2(x_0))^t
\end{bmatrix}^{-1} \begin{bmatrix} g_1(x) \\ g_2(x) \end{bmatrix}
\]

\[
= \begin{bmatrix}
- \frac{1}{3} & 0 \\
0 & -\frac{1}{2}
\end{bmatrix}
\begin{bmatrix}
-3x_1^4 + x_2^3 - 3x_1 - 3x_2 + 2 \\
2x_1^4 + 2x_1^2 - x_2^2 + 1
\end{bmatrix}
\]
\[
\begin{bmatrix}
-x_1^4 & x_2^3 + x_1 + x_2 - \frac{2}{3} \\
-x_1^4 & x_1^2 + \frac{1}{2} x_2^2 - \frac{1}{2}
\end{bmatrix}
\]

So (2.5) is satisfied for this \( \eta(x, x_0) \).

By theorem 2.1,
\[
f(x) - f(x_0) \geq \alpha(x, x_0) \eta^t(x, x_0) \nabla f(x_0),
\]
that is,
\[
-2x_2^3 - 6x_1^2 + 3x_2^2 + 6x_1 + 6x_2 - 7 - 0
\]

\[
\geq \alpha(x, x_0) \begin{bmatrix} 6 & 6 \\ -6 & -6 \end{bmatrix} \begin{bmatrix}
x_1^4 & \frac{1}{3} x_2^3 + x_1 + x_2 - \frac{2}{3} \\
-x_1^4 & x_1^2 + \frac{1}{2} x_2^2 - \frac{1}{2}
\end{bmatrix}
\]

\[
= \alpha(x, x_0) (-2x_2^3 - 6x_1^2 + 3x_2^2 + 6x_1 + 6x_2 - 7).
\]

So (2.1) is satisfied for \( \alpha(x, x_0) = 1 \).

By corollary 2.1 \( x_0 \) is a global minimum.
REFERENCES


