† *Department of Statistics, University of North Carolina at Chapel Hill. This work was supported in part by a grant from the Office of Naval Research, Contract Number N00014-67-A-0321-003 (NR047-095).*

†† *Department of Mathematics, University of North Carolina at Chapel Hill.*

Optimality Conditions and Constraint Qualifications in Banach Space

by

F. J. Gould† and Jon W. Tolle††

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 707

August, 1970

Optimality Conditions and Constraint Qualifications in Banach Space

F. J. Gould[†] and Jon W. Tolle[††]

Abstract. In this paper necessary optimality conditions for nonlinear programs in Banach spaces and constraint qualifications for their applicability are considered. A new optimality condition is introduced and a constraint qualification ensuring the validity of this condition is given. When the domain space is a reflexive space, it is shown that the qualification is the weakest possible. If a certain convexity assumption is made, then this optimality condition is shown to reduce to the well-known extension of the Kuhn-Tucker conditions to Banach spaces. In this case the constraint qualification is weaker than those previously given.

Introduction. In this paper we will be concerned with a general constrained optimization problem in Banach space. Let X and Y be Banach spaces and let $g: X \to Y$ be a differentiable map. For subsets $A_x \subseteq X$ and $A_y \subseteq Y$ define the constraint set, S, by $S = A_x \cap g^{-1}(A_y)$. A function $f: X \to R$ will be called an objective function with a local constrained maximum at x_o if f is continuous on an open set containing S and is differentiable at $x_o \in S$ where x_o is a local solution to the constrained optimization problem

$$\text{(1.1)} \qquad \text{maximize } f(x), \text{ subject to } x \in S,$$

i. e. there exists an $\varepsilon > 0$ such that $f(x_o) \geq f(x)$ for all $x \in S \cap \{x: ||x - x_o|| < \varepsilon\}$. The set of all objective functions which have a local constrained maximum at x_o will be denoted by $F(x_o)$ and the set of all derivatives at x_o of elements in $F(x_o)$ will be denoted by $DF(x_o)$. Thus $DF(x_o)$ is a subset of X^*, the topological dual of X.

[†]Department of Statistics, University of North Carolina at Chapel Hill. This work was supported in part by a grant from the Office of Naval Research, contract number N00014-67-A-0321-003 (NR047-095).

[††]Department of Mathematics, University of North Carolina at Chapel Hill.

The optimization problem (1.1) is said to satisfy an optimality condition at x_o if there is a specified subset of X^*, say A^*, such that $Df(x_o) \in A^*$ for every objective function f with a local constrained maximum at x_o; i.e. $DF(x_o) \subseteq A^*$. The most familiar optimality condition for a problem of the type (1.1) is the Kuhn-Tucker condition [6] which was originally stated for the case when $A_x = X = R^n$, $Y = R^m$, and $A_y = \{y \in R^m: y_j \leq 0, j = 1,\ldots,m\}$. For this condition the set A^* is defined by

$$A^* = \{y \circ Dg(x_o): y_j \geq 0, j = 1,\ldots,m, \text{ and } y \circ g(x_o) = 0\}.$$

The Kuhn-Tucker condition has since been extended to apply to a more general problem which includes equality constraints and with A_x a proper subset of X. This extension usually necessitates a change in the set A^*; for details see Mangasarian and Fromovitz [7] or Gould and Tolle [3]. In addition considerable progress has been made in extending the Kuhn-Tucker condition to nonlinear programming problems in infinite-dimensional spaces. For results in this area as well as for lists of related references, the reader is directed to the papers of Guignard [4], Hurwicz [5], Ritter [8], Russell [9], and Varaiya [11]. In general, the approach of the above authors is to attempt to determine sets $C^* \subseteq X^*$ and $B^* \subseteq Y^*$ such that for each $f \in F(x_o)$,

$$Df(x_o) = y^* \circ Dg(x_o) + x^*$$

for some $y^* \in B^*$ and $x^* \in C^*$. In the cases where C^* is the trivial map these optimality conditions are of the Kuhn-Tucker type.

In order that an optimality condition be valid at x_o for the problem (1.1), certain restrictions, called constraint qualifications, must usually be placed on x_o, A_x, A_y, and g. In Euclidean space the development of successively weaker constraint qualifications for which the Kuhn-Tucker condition is valid has led to a qualification which is the weakest possible, i.e. both necessary and sufficient for the Kuhn-Tucker condition to hold for every objective function [3]. The constraint qualifications which have been used for similar optimality conditions in infinite-dimensional spaces have generally been essentially the same as those applied in the finite-dimensional cases. One advantage of the optimality condition with a nontrivial C^* is that often it requires a weaker constraint qualification than does a Kuhn-Tucker type of condition.

Herein we introduce a new optimality condition for the problem (1.1) and a constraint qualification which is sufficient to guarantee that the optimality condition is satisfied. In the case when X is reflexive, it will be shown that the problem (1.1) satisfies the optimality condition if and only if the constraint qualification is satisfied. Moreover, it will be proved that if A_y is convex the optimality condition reduces to earlier extensions of the Kuhn-Tucker condition and that the constraint qualification is weaker than those previously applied in this case.

These results generalize the previous work of the authors in Euclidean spaces [3] and extend the investigations recently carried out by Guignard [4], Ritter [8], and Varaiya [11].

2. *Definitions and terminology.* In this section we will introduce some terminology which will be necessary for the statement of our results.

DEFINITION 1. Let L be a linear topological space and let L^* be its topological dual given the weak* topology. For B a nonempty subset of L, the polar cone of B, B', is the subset of L^* defined by

$$B' = \{\ell^* \in L^*: \ell^*(\ell) \leq 0 \text{ for every } \ell \in B\}.$$

The following properties of polar cones will be important:

(i) If $B_1 \subseteq B_2$, then $B_2' \subseteq B_1'$.

(ii) $B' = $ (closed convex hull of B)'.

(iii) B' is a closed convex cone.

(iv) $B \subseteq (B')'$ with equality if and only if B is a closed convex cone.

DEFINITION 2. Let Z be a Banach space, $B \subseteq Z$, and $z_o \in B$. Then the <u>cone of tangents to</u> B <u>at</u> z_o, $T(B,z_o)$, is the set of all $z \in Z$ for which there exists a nonnegative real sequence, $\{\lambda_n\}$, and a sequence in B, $\{z_n\}$, such that

(i) $z_n \to z_o$,

(ii) $\lambda_n(z_n - z_o) \to z$.

DEFINITION 3. For Z, B, z_o as in definition 2, the <u>weak cone of tangents to</u> B <u>at</u> z_o, $T_w(B,z_o)$, is the set of all $z \in Z$ for which there exists a nonnegative real sequence, $\{\lambda_n\}$, and a sequence in B, $\{z_n\}$, such that

(i) $z_n \to z_o$

(ii) $\lambda_n(z_n - z_o) \to z$ weakly; i.e. $\lambda_n z^*(z_n - z_o) \to z^*(z)$ for every $z^* \in Z^*$.

We shall denote by $R(B,z_o)$ and $R_w(B,z_o)$ the closed convex hulls of $T(B,z_o)$ and $T_w(B,z_o)$ respectively. The properties of tangent cones which will be useful in this paper are listed below. Most of these properties are easily shown and no proofs are given. However, properties iv and v are less obvious and are dealt with more fully in section 6.

(i) $T(B,z_o)$, $T_w(B,z_o)$, $R(B,z_o)$, and $R_w(B,z_o)$ are nonempty cones.

(ii) $T(B,z_o) \subseteq T_w(B,z_o)$ and $R(B,z_o) \subseteq R_w(B,z_o)$ with equality holding in both cases if Z is finite-dimensional.

(iii) $T(B,z_o)$ is closed.

(iv) If B is convex, then $T(B,z_o)$, $T_w(B,z_o)$, $R(B,z_o)$, and $R_w(B,z_o)$ are all the same closed convex cone.

(v) $R'_w(B,z_o) \subseteq R'(B,z_o)$ with equality not holding in general.

In the following, we shall employ the same notation given in section 1 in stating problem (1.1). The dual spaces X^* and Y^* will be given their respective weak* topologies.

DEFINITION 4. Let $x_o \in S$. The pseudolinearizing cone at x_o, $K(x_o)$, and the weak pseudolinearizing cone at x_o, $K_w(x_o)$, are subsets of X defined by

$$K(x_o) = \{x \in X: Dg(x_o)(x) \in R(A_y, g(x_o))\}$$

and

$$K_w(x_o) = \{x \in X: Dg(x_o)(x) \in R_w(A_y, g(x_o))\}.$$

By using the properties of the tangent cones, it can be easily verified that $K(x_o)$ and $K_w(x_o)$ are closed convex cones in X and that $K(x_o) \subseteq K_w(x_o)$ with equality holding if Y is finite-dimensional or if A_y is convex.

DEFINITION 5. Let $x_o \in S$. The subsets, $B^*(x_o)$ and $B^*_w(x_o)$, of X^* are defined by

$$B^*(x_o) = \{x^* \in X^*: x^* = y^* \circ Dg(x_o) \text{ for some } y^* \in T'(A_y, g(x_o))\}$$

and

$$B^*_w(x_o) = \{x^* \in X^*: x^* = y^* \circ Dg(x_o) \text{ for some } y^* \in T'_w(A_y, g(x_o))\}.$$

Clearly $B^*_w(x_o) \subseteq B^*(x_o)$ with equality holding if A_y is convex or Y is finite-dimensional. The sets $B^*(x_o)$ and $B^*_w(x_o)$ are convex cones but need not be closed. In the classical finite-dimensional nonlinear programming problem, $B^*(x_o) = B^*_w(x_o)$ is a closed convex polyhedral cone.

For a discussion of the cases in which $B^*(x_o)$ is closed, the reader is referred to Varaiya [10].

3. Statement of main results. We shall be concerned with an optimality condition that establishes the set $DF(x_o)$ as a subset of $\overline{B_w^*(x_o)}$, the weak* closure of $B_w^*(x_o)$. Varaiya [11], in the case where A_y is convex, and Guignard [4], more generally, have shown that the following relations are valid:

$$\overline{B^*(x_o)} = K'(x_o) \subseteq T'(S,x_o), \tag{3.1}$$

$$DF(x_o) \subseteq T'(S,x_o), \tag{3.2}$$

where $\overline{B^*(x_o)}$ refers to the weak* closure of $B^*(x_o)$. Thus if the constraint qualification

$$T'(S,x_o) \subseteq K'(x_o) \tag{3.3}$$

holds, it follows that the optimality condition

$$DF(x_o) \subseteq \overline{B^*(x_o)} \tag{3.4}$$

is true. It should be noted that in the case when $B^*(x_o)$ is closed and (3.3) is satisfied, it follows from (3.4) that for any $f \in F(x_o)$ there is a $y^* \in T'(A_y,g(x_o))$ such that $Df(x_o) = y^* \circ Dg(x_o)$. This is a direct extension of the Kuhn-Tucker condition in Euclidean spaces.

Our first theorem gives a result similar to (3.1) and (3.2) utilizing the notion of the weak cone of tangents.

THEOREM 1. The following relations hold:

$$\overline{B_w^*(x_o)} = K_w'(x_o) \subseteq T_w'(S,x_o), \tag{3.5}$$

$$DF(x_o) \subseteq T_w'(S,x_o). \tag{3.6}$$

As a consequence of this theorem we have

COROLLARY 1. The optimality condition

$$DF(x_o) \subseteq \overline{B_w^*(x_o)} \tag{3.7}$$

holds if the weak constraint qualification

(3.8) $$T_w'(S,x_o) \subseteq K_w'(x_o)$$

is satisfied.

Since $\overline{B_w^*(x_o)} \subseteq \overline{B^*(x_o)}$, our specification of $DF(x_o)$ as a subset of $\overline{B_w^*(x_o)}$ is a sharper result than that given by (3.4) above. The corresponding constraint qualifications, (3.3) and (3.8), are not comparable in general.

In the important case in which A_y is convex, it follows from property iv of the tangent cones that $K_w'(x_o) = K'(x_o)$ and $B_w^*(x_o) = B^*(x_o)$. Thus we obtain

COROLLARY 2. *If* A_y *is convex and the constraint qualification*

(3.9) $$T_w'(S,x_o) \subseteq K'(x_o)$$

is satisfied, then the optimality condition (3.4) *holds*.

For this case the optimality conditions (3.4) and (3.7) are the same; however, the constraint qualifications (3.3) and (3.9) are not equivalent. From (3.5) and the fact that $T_w'(S,x_o) \subseteq T'(S,x_o)$ it is evident that the constraint qualification (3.9) is less stringent than (3.3). In section 6 an example is given which illustrates that $T_w'(S,x_o)$ and $T'(S,x_o)$ are, in general, not equivalent so that (3.9) and (3.3) are indeed different.

It is natural to ask if the constraint qualification (3.8) is the weakest qualification which will ensure the validity of the optimality condition (3.7). In the event that X is a reflexive Banach space, the following theorem allows us to answer this question in the affirmative.

THEOREM 2. *If* X *is reflexive, then*

(3.10) $$DF(x_o) = T_w'(S,x_o).$$

It now follows from (3.5) and (3.10) that for X reflexive, the weak constraint qualification (3.8) and the optimality condition (3.7) are equivalent.

COROLLARY 3. *If X is reflexive then the optimality condition (3.7) is satisfied if and only if the constraint qualification (3.8) is valid. In this case*

$$DF(x_o) = \overline{B_w^*(x_o)}.$$

For X reflexive and A_y convex we obtain the following extension of corollary 2:

COROLLARY 4. *If X is reflexive and A_y is convex, then the optimality condition (3.4) holds if and only if the constraint qualification (3.9) holds. Then*

$$DF(x_o) = \overline{B^*(x_o)}.$$

Finally we note that if X and Y are finite-dimensional and $A_y = \{y \in Y: y_j \leq 0\}$, then $\overline{B^*(x_o)} = B^*(x_o) = \{\Sigma\lambda_j \cdot Dg_j(x_o): \lambda_j \geq 0\}$. In this case we obtain from corollary 4 an earlier result of the authors [3].

COROLLARY 5. *Let X and Y be finite-dimensional and let $A_y = \{y \in Y: y_j \leq 0\}$. Then in order for*

$$Df(x_o) = \Sigma\, \lambda_j \cdot Dg_j(x_o),$$

$$\lambda_j \geq 0,$$

and

$$\lambda_j \cdot g_j(x_o) = 0$$

to hold for every objective function f with a local constrained maximum at x_o, it is necessary and sufficient that

$$T'(S,x_o) = K'(x_o).$$

4. *Proof of theorem 1.* We shall need the following well-known separation theorem: *Let Z be a locally convex linear topological space. Let K_1 and K_2 be disjoint closed convex subsets of Z with K_1 compact and K_2 a cone. Then there exist a $z^* \in Z^*$ and an $\alpha > 0$ such that*

$$z^*(z_1) \geq \alpha > 0 \geq z^*(z_2)$$

for all $z_1 \in K_1$ and $z_2 \in K_2$.

Theorem 1 will result from the following three lemmas. It should be pointed out that the methods of proof of these lemmas will be essentially the same as those used by Varaiya [11] and Guignard [4]. The only changes are those necessary to allow the use of the weak cone of tangents rather than the cone of tangents. The importance of this modification is contained in the statement of theorem 2.

LEMMA 4.1. $\overline{B_w^*(x_o)} = K_w'(x_o)$.

Proof. We first verify that $\overline{B_w^*(x_o)} \subseteq K_w'(x_o)$. Suppose $\phi^* \in B_w^*(x_o)$, then $\phi^* = \psi^* o Dg(x_o)$ where $\psi^* \in T_w'(A_y, g(x_o))$. By property iii of polar cones $\psi^* \in R_w'(A_y, g(x_o))$. Let $x \in K_w(x_o)$. Then $\phi^*(x) = \psi^* o Dg(x_o)(x) = \psi^*(y) \leq 0$ since $y = Dg(x_o)(x) \in R_w(A_y, g(x_o))$. Thus $\phi^* \in K_w'(x_o)$ and hence $B_w^*(x_o) \subseteq K_w'(x_o)$. The desired result follows from the fact that $K_w'(x_o)$ is closed.

To show that $K_w'(x_o) \subseteq \overline{B_w^*(x_o)}$, we assume that the inclusion does not hold. Then there exists an $x^* \in K_w'(x_o)$ with $x^* \notin \overline{B_w^*(x_o)}$. We now apply the separation theorem with $Z = X^*$, $K_1 = \{x^*\}$, and $K_2 = \overline{B_w^*(x_o)}$. Thus there is an $\hat{x} \in X$ such that

$$x^*(\hat{x}) > 0 \geq \phi^*(\hat{x}) \tag{4.1}$$

for all $\phi^* \in \overline{B_w^*(x_o)}$. Clearly $\hat{x} \notin K_w(x_o)$ so $\hat{y} = Dg(x_o)(\hat{x}) \notin R_w(A_y, g(x_o))$. Applying the separation theorem again with $Z = Y$, $K_1 = \{\hat{y}\}$, and $K_2 = R_w(A_y, g(x_o))$ we obtain a $\hat{y}^* \in \hat{Y}^*$ such that

$$\hat{y}^*(\hat{y}) > 0 \geq \hat{y}^*(y) \tag{4.2}$$

for all $y \in R_w(A_y, g(x_o))$. Thus $\hat{y}^* \in R_w'(A_y, g(x_o)) = T_w'(A_y, g(x_o))$ and hence $\hat{\phi}^* = y^* o Dg(x_o) \in \overline{B_w^*(x_o)}$. But $\hat{\phi}^*(\hat{x}) = \hat{y}^*(\hat{y}) > 0$ by (4.2) and therefore (4.1) is contradicted.

LEMMA 4.2. $K_w'(x_o) \subseteq T_w'(S,x_o)$.

Proof. We show $T_w(S,x_o) \subseteq K_w(x_o)$. The result then follows from property i of polar cones. Let $x \in T_w(S,x_o)$. Then there are sequences $\{x_n\} \in S$ and $\{\lambda_n\} \in R^+$ such that $||x_n - x_o|| \to 0$ and $\lambda_n(x_n - x_o) \to x$ weakly. Set $y_n = g(x_n)$. Then $y_n \in A_y$ and $||y_n - g(x_o)|| \to 0$ by the continuity of g. Since g is differentiable at x_o we have

$$(4.3) \qquad \lambda_n(y_n - g(x_o)) = \lambda_n \cdot Dg(x_o)(x_n - x_o) + \lambda_n \cdot \varepsilon(x_n - x_o)$$

where $\frac{\varepsilon(z)}{||z||} \to 0$ as $||z|| \to 0$.

From (4.3) we have that for $y^* \in Y^*$

$$(4.4) \qquad \lambda_n \cdot y^*(y_n - g(x_o)) = \lambda_n \cdot y^* \circ Dg(x_o)(x_n - x_o) + \lambda_n \cdot y^* \circ \varepsilon(x_n - x_o).$$

As $n \to \infty$, $||\lambda_n \cdot y^* \circ \varepsilon(x_n - x_o)|| = ||\lambda_n(x_n - x_o)|| \cdot \frac{|y^* \circ \varepsilon(x_n - x_o)|}{||x_n - x_o||}$

tends to zero since $\{\lambda_n(x_n - x_o)\}$ is a weakly convergent sequence and hence bounded. Therefore from (4.4)

$$\lim_{n\to\infty} y^*(\lambda_n(y_n - g(x_o))) = \lim_{n\to\infty} y^* \circ Dg(x_o)(\lambda_n(x_n - x_o)) = y^* \circ Dg(x_o)$$

since $y^* \circ Dg(x_o) \in X^*$.

Thus we have shown that for $\{y_n\} \in A_y$ and $\{\lambda_n\} \in R^+$, $||y_n - g(x_o)|| \to 0$, and $\lambda_n(y_n - g(x_o)) \to Dg(x_o)(x)$ weakly. Therefore $Dg(x_o)(x) \in T_w(A_y, g(x_o)) \subseteq R_w(A_y, g(x_o))$ and so $x \in K_w(x_o)$.

LEMMA 4.3. $DF(x_o) \subseteq T_w'(S,x_o)$.

Proof. Let $x \in T_w(S,x_o)$. Then there are sequences $\{x_n\} \in S$ and $\{\lambda_n\} \in R^+$ such that $||x_n - x_o|| \to 0$ and $\lambda_n(x_n - x_o) \to x$ weakly. Let $f \in F(x_o)$. Since f is differentiable at x_o and has a local constrained maximum at x_o, we have

$$(4.5) \quad \lambda_n(f(x_n) - f(x_o)) = \lambda_n \cdot Df(x_o)(x_n - x_o) + \lambda_n \cdot \varepsilon(x_n - x_o) \leq 0$$

for n sufficiently large. Using the same argument as given in the proof of the preceding lemma we have $||\lambda_n \cdot \varepsilon(x_n - x_o)|| \to 0$ as $n \to \infty$. Therefore since $Df(x_o) \in X^*$, we obtain the result

$$Df(x_o)(x) = \lim_{n\to\infty} Df(x_o)(\lambda_n(x_n - x_o)) = \lim_{n\to\infty} \lambda_n(f(x_n) - f(x_o)).$$

But then, from (4.5), we have $Df(x_o)(x) \leq 0$ which implies that $Df(x_o) \in T_w'(S,x_o)$.

5. Proof of theorem 2. By theorem 1, it is sufficient to show that if X is reflexive then for every $x^* \in T_w'(S,x_o)$ there is an $f \in F(x_o)$ such that $Df(x_o) = x^*$. For convenience we shall assume that x_o is the null vector of X, denoted by θ, and that the given $x^* \in T_w'(S,\theta)$ has norm $||x^*|| = \sup_{x\neq\theta} \frac{|x^*(x)|}{||x||} = 1$. This latter assumption causes no loss of generality since $T_w'(S,\theta)$ is a cone.

Let $\hat{x}_2 \in X$ be chosen such that $x^*(\hat{x}_2) = 1$ and let N be the one-dimensional subspace of X defined by $N = \{\alpha\hat{x}_2 : \alpha \in R\}$. Let M be the kernel of x^*. Now M and N are closed linear subspaces of X and each $x \in X$ has a unique decomposition, $x = x_1 + x_2$, where $x_2 = x^*(x)\cdot\hat{x}_2 \in N$ and $x_1 = x - x_2 \in M$. The projection map $p: X \to M$ defined by $p(x) = x_1$ is then a continuous linear operator.

For $\varepsilon > 0$ and k a positive integer, we define the following subsets of X:

$$S_\varepsilon = S \cap \{x \in X: ||x|| < \varepsilon\}$$

and

$$C_k = \{x \in X: x \neq \theta,\ \mathrm{Arccos}\left(\frac{x^*(x)}{||x||}\right) \leq \frac{\pi}{2} - \frac{\pi}{k+2}\} .$$

Clearly $C_k \subset C_{k+1}$ and $C_k \cup \{\theta\}$ is a cone.

LEMMA 5.1. <u>For every positive integer</u> k <u>there is an</u> $\varepsilon(k) > 0$ <u>such that</u> $S_{\varepsilon(k)} \cap C_k = \phi$.

In order to prove this lemma we employ the following result from [2]: <u>A subset of a reflexive Banach space is weakly sequentially compact if and only if it is bounded.</u>

<u>Proof of lemma 5.1</u>. Suppose otherwise. Then for every k there would exist a sequence $\{x_m^k\}_{m=1}^{\infty} \in C_k$ with $x_m^k \in S_{1/m}$ for each $m > 0$. If $x_m^k \in C_k$, then it follows from the definition of C_k that

$$(5.1) \qquad 1 \geq \frac{1}{||x_m^k||} \cdot x^*(x_m^k) \geq \cos\left(\frac{\pi}{2} - \frac{\pi}{k+2}\right) > 0.$$

The set $\left\{\dfrac{x_m^k}{||x_m^k||}\right\}_{m=1}^{\infty}$ is a bounded subset of the reflexive space X; thus it is weakly sequentially compact. Consequently there is a subsequence, $\left\{\dfrac{\hat{x}_n^k}{||\hat{x}_n^k||}\right\}_{n=1}^{\infty}$, which converges weakly to some $\hat{x} \in X$. For any $x_o^* \in X^*$

$$(5.2) \qquad x_o^*\left(\frac{1}{||\hat{x}_n^k||}(\hat{x}_n^k - \theta)\right) \to x_o^*(x) \quad \text{as} \quad n \to \infty.$$

Moreover since $\{\hat{x}_n^k\}$ is a subsequence of $\{x_m^k\}$ and $x_m^k \in S_{1/m}$, it follows that, as $n \to \infty$

$$(5.3) \qquad \hat{x}_n^k \to \theta.$$

Now (5.2) and (5.3) imply that $\hat{x} \in T_w(S,\theta)$. On the other hand, (5.1) and (5.2) imply that $x^*(\hat{x}) \geq \cos(\frac{\pi}{2} - \frac{\pi}{k+2}) > 0$ which yields the contradiction and thus completes the proof.

We now define

$$\epsilon_k = \sup\{\epsilon > 0: S_\epsilon \cap C_k = \phi\}$$

and set

$$\hat{\varepsilon}_k = \begin{cases} \min(\varepsilon_1, 1) & k = 1, \\ \min(\varepsilon_k, \frac{1}{2}\varepsilon_{k-1}) & k > 1, \end{cases}$$

to obtain a sequence, $\{\hat{\varepsilon}_k\}$, of finite positive numbers which strictly decrease to 0 as $k \to \infty$. Moreover, for all $\varepsilon < \hat{\varepsilon}_k$,

(5.4) $$S_\varepsilon \cap C_k = \phi.$$

Let L be the smallest positive integer such that

$$\sin \frac{\pi}{L+1} < \frac{1}{||\hat{x}_2||}$$

and define the sequence $\{\alpha_k\}_{k=L}^{\infty}$ by

$$\alpha_k = \frac{\sin \frac{\pi}{k+1}}{1 - ||\hat{x}_2|| \cdot \sin \frac{\pi}{k+1}} .$$

The sequence $\{\alpha_k\}_{k=L}^{\infty}$ is strictly decreasing and tends to 0 as $k \to \infty$. We now define the function $P: M \to R$ by

$$P(x_1) = \begin{cases} \alpha_L \hat{\varepsilon}_L & \text{for } ||x_1|| \geq \hat{\varepsilon}_L, \\ \alpha_{k+1}\hat{\varepsilon}_{k+1} + \dfrac{\alpha_k \hat{\varepsilon}_k - \alpha_{k+1}\hat{\varepsilon}_{k+1}}{\hat{\varepsilon}_k - \hat{\varepsilon}_{k+1}} \cdot (||x_1|| - \hat{\varepsilon}_{k+1}) & \\ & \text{for } k \geq L \text{ and } \hat{\varepsilon}_{k+1} \leq ||x_1|| < \hat{\varepsilon}_k, \\ 0 & \text{for } x_1 = \theta_1. \end{cases}$$

For $k \geq L$ and $\hat{\varepsilon}_{k+1} \leq ||x_1|| < \hat{\varepsilon}_k$, we may rewrite $P(x_1)$ as follows:

(5.5) $$P(x_1) = \frac{1}{\hat{\varepsilon}_k - \hat{\varepsilon}_{k+1}} \left\{ \alpha_k \hat{\varepsilon}_k (||x_1|| - \hat{\varepsilon}_{k+1}) + \alpha_{k+1}\hat{\varepsilon}_{k+1}(\hat{\varepsilon}_k - ||x_1||) \right\}.$$

Now, in (5.5), we first replace α_k with α_{k+1} and then replace α_{k+1} by α_k. Since the sequence $\{\alpha_k\}$ is decreasing, this yields the estimates

(5.6) $$\alpha_{k+1} \cdot ||x_1|| < P(x_1) < \alpha_k ||x_1||$$

for $\hat{\varepsilon}_{k+1} \leq ||x_1|| < \hat{\varepsilon}_k$, $k \geq L$. It now follows from (5.5) and (5.6) that P is continuous on M and differentiable at $x_1 = \theta_1$ with $DP(\theta_1)$ being the trivial map.

Let $f: X \to R$ be defined by

$$f(x) = x^*(x) - (Pop)(x).$$

Since x^* and p are continuous linear maps, it is apparent from the properties of P that f is continuous on X and differentiable at $x = \theta$ with $Df(\theta) = x^*$. Thus f is a functional on X having the desired derivative. It remains to show that f has a local constrained maximum at x_o, i.e. that $f \in F(x_o)$. The following lemma will yield this result.

LEMMA 5.2. <u>There exists an</u> $\varepsilon > 0$ <u>such that</u> $f(x) \leq 0$ <u>for all</u> $x \in S_\varepsilon$, $x \neq \theta$.

<u>Proof</u>. Let $x \in S_{\hat{\varepsilon}_{L+1}}$. Then by (5.4), $x \notin C_L$ and so

$$x^*\left(\frac{x}{||x||}\right) < \cos\left(\frac{\pi}{2} - \frac{\pi}{L+2}\right) = \sin\frac{\pi}{L+2}\ .$$

Thus

$$x^*(x) < \sin\frac{\pi}{L+2}\cdot||x||.$$

Writing $x = x_1 + x^*(x)\cdot\hat{x}_2$, we have

$$x^*(x) < \sin\frac{\pi}{L+2}(||x_1|| + |x^*(x)|\cdot||\hat{x}_2||). \tag{5.7}$$

If $x^*(x) \leq 0$, then $f(x) \leq 0$ since $P(x_1) = (Pop)(x) \geq 0$ for all $x \in X$. If $x^*(x) > 0$, then from (5.7) we obtain

$$x^*(x) < \frac{\sin\frac{\pi}{L+2}}{1 - ||\hat{x}_2||\sin\frac{\pi}{L+2}}\cdot||x_1|| = \alpha_{L+1}\cdot||x_1||.$$

It follows from (5.6) that $P(x_1) = (Pop)(x) \geq \alpha_{L+1}\cdot||x_1||$ if $||x_1|| = ||p(x)|| < \hat{\varepsilon}_L$. If we choose

$$\varepsilon = \min\left(\hat{\varepsilon}_{L+1}, \frac{\hat{\varepsilon}_L}{||p||}\right),$$

then for $x \in S_\varepsilon$, $x \neq \theta$, we have

$$||x_1|| \leq ||p||\cdot||x|| \leq \hat{\varepsilon}_L,$$

and hence

$$x^*(x) \leq a_{L+1}\cdot||x_1|| \leq P(x_1) = (P\circ p)(x).$$

Thus $f(x) \leq 0$ which completes the proof of the lemma.

Since $f(\theta) = 0$, it is clear that f has a local constrained maximum at x_o, thus completing the proof of theorem 2.

If X is a Banach space such that the function $x \to ||x||$ is differentiable except at $x = \theta$, e.g. a Hilbert space, then the function P, and hence the function f, can be modified so as to be differentiable everywhere. In this case the point x_o need not be specified to define the class of objective functions. The objective functions will be the smaller class of functions which are differentiable in an open set containing S. For a discussion of when the norm function is differentiable, see [1].

6. Supplementary results. In this section we will clarify and prove some of the statements concerning tangent cones made earlier. First we shall prove property iv of tangent cones.

PROPOSITION 6.1. *If* B *is a convex subset of a Banach space* Z *and* $z_o \in B$, *then* $T(B,z_o) = T_w(B,z_o)$.

To prove this proposition we shall employ the following result from [1]: *If* $\{x_n\}$ *is a sequence of elements in a Banach space* X *converging weakly to* $x \in X$, *then some sequence of convex combinations of the* $\{x_n\}$ *converges to* x.

Proof of proposition 6.1. From property ii of tangent cones it is evident that we need only show that $T_w(B,z_o) \subseteq T(B,z_o)$.

Let $\hat{z} \in T_w(B,z_o)$. Then there exist sequences $\{z_n\} \in B$ and $\{\lambda_n\} \in R^+$ such that $||z_n - z_o|| \to 0$ and $\lambda_n(z_n - z_o) \to \hat{z}$ weakly. For each $k > 0$ let N_k be chosen such that for all $n \geq N_k$,

$$(6.1) \qquad ||z_n - z_o|| < \frac{1}{k} .$$

Denote by $\{z_n^k\}$ and $\{\lambda_n^k\}$ the sequences $\{z_n\}$ and $\{\lambda_n\}$ with the first N_k terms removed and set

$$x_n^k = \lambda_n^k(z_n^k - z_o).$$

For each fixed k, $x_n^k \to \hat{z}$ weakly. Thus there is a sequence $\{\hat{x}_n^k\}$ such that each $\hat{x}_n^k$ is a convex combination of elements in $\{x_n^k\}$ and as $n \to \infty$,

$$(6.2) \qquad ||\hat{x}_n^k - \hat{z}|| \to 0.$$

Since the $\hat{x}_n^k$ are convex combinations of the elements $x_n^k = \lambda_n^k(z_n^k - z_o)$, it can be shown that each $\hat{x}_n^k$ can be written as

$$\hat{x}_n^k = \hat{\lambda}_n^k(\hat{z}_n^k - z_o)$$

where the $\hat{\lambda}_n^k$ and $\hat{z}_n^k$ are convex combinations of the elements in $\{\lambda_n^k\}$ and $\{z_n^k\}$ respectively. Thus the $\hat{z}_n^k$ are in B.

From (6.2), we have that for each k there is an n_k such that

$$||\hat{x}_{n_k}^k - \hat{z}|| < \frac{1}{k} .$$

Thus as $k \to \infty$, $\hat{x}_{n_k}^k = \hat{\lambda}_{n_k}(\hat{z}_{n_k}^k - z_o) \to \hat{z}$ in the norm topology. Moreover, since each $\hat{z}_{n_k}^k$ is a convex combination of elements in $\{z_n^k\}$, it follows from (6.1) that $||z_{n_k}^k - z_o|| \to 0$ as $k \to \infty$. It then follows from the definition of $T(B,z_o)$ that $\hat{z} \in T(B,z_o)$ which proves the proposition.

The following example shows that in general $T_w(B,z_o) \not\subseteq R(B,z_o)$ for B an arbitrary subset of a Banach space Z.

EXAMPLE 6.1. Let $Z = \ell^2$ and set $z_o = \theta$ and $B = \{x \in \ell^2: x = \frac{1}{k}(e_1 + e_k),\ k$ a positive integer$\} \cup \{\theta\}$. Here e_k

denotes the k^{th} unit vector. Taking $x_k = \frac{1}{k}(e_1 + e_k)$, we see that $||x_k - \theta|| \to 0$ as $k \to \infty$ and that $k(x_k - \theta) \to e_1$ weakly as $k \to \infty$. Thus $e_1 \in T_w(B,\theta)$. It is easily verified that $T(B,\theta) = R(B,\theta) = \{\theta\}$, so that $T_w(B,\theta) \not\subseteq R(B,\theta)$. This example may be modified to give an example for which $T_w(B,\theta) = Z$ and $R(B,\theta) = \{\theta\}$.

Examples for which $R(B,z_o) \not\subseteq T_w(B,z_o)$ may be easily constructed in finite-dimensional spaces where $T_w(B,z_o) = T(B,z_o)$.

That equality does not hold in general in property v of tangent cones follows from the above example, property iii of polar cones, and the following proposition.

PROPOSITION 6.2. $T_w'(B,z_o) = T'(B,z_o)$ *if and only if* $T_w(B,z_o) \subseteq R(B,z_o)$.

Proof. Suppose $T_w(B,z_o) \subseteq R(B,z_o)$. Then $R'(B,z_o) \subseteq T_w'(B,z_o) \subseteq T'(B,z_o)$. But $R'(B,z_o) = T'(B,z_o)$, so $T_w'(B,z_o) = T'(B,z_o)$.

Conversely, suppose $T_w(B,z_o) \not\subseteq R(B,z_o)$. Then there is a $\hat{z} \in T_w(B,z_o)$ such that $\hat{z} \notin R(B,z_o)$. Since $R(B,z_o)$ is a closed convex cone the separation theorem of section 4 can be applied. Thus there is a $z^* \in Z^*$ such that

$$z^*(\hat{z}) > 0 \geq z^*(z)$$

for all $z \in R(B,z_o)$. Therefore $z^* \in R'(B,z_o)$ and hence $z^* \in T'(B,z_o)$. But $z^* \notin T_w'(B,z_o)$ since $z^*(\hat{z}) > 0$. Thus $T_w'(B,z_o) \neq T'(B,z_o)$.

References

[1] M. M. Day, Normed Linear Spaces, Springer-Verlag, Berlin, 1958.

[2] N. Dunford, J. T. Schwartz, Linear Operators, Part I, Interscience Publishers, New York, 1964.

[3] F. J. Gould, J. W. Tolle, A necessary and sufficient qualification for constrained optimization, SIAM J. Appl. Math.,to appear.

[4] M. Guignard, Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space, SIAM J.Control, 7 (1969), pp. 232-241.

[5] L. Hurwicz, Programming in linear spaces, Studies in Linear and Nonlinear Programming, K. J. Arrow, L. Hurwicz, H. Uzawa, eds., Stanford University Press, Stanford, 1958.

[6] H. Kuhn, A. Tucker, Nonlinear programming, Proceedings of the Second Berkeley Symposium, on Mathematical Statistics and Probability, J. Neyman, ed., University of California Press, Berkeley, 1951, pp. 481-492.

[7] O. Mangasarian, S. Fromovitz, The Fritz John necessary optimality conditions in the presence of equality and inequality constraints, J. Math. Anal. Appl., 17 (1967), pp. 37-47.

[8] K. Ritter, Optimization theory in linear spaces, Math. Ann., 184 (1970), pp. 133-154.

[9] D. L. Russell, The Kuhn-Tucker conditions in Banach space with an application to control theory, J. Math. Anal. Appl., 15 (1966), pp. 200-212.

[10] P. P. Varaiya, Nonlinear programming and optimal control, ERL Tech. Memo. M-129, University of California, Berkeley, 1965.

[11] P. P. Varaiya, Nonlinear programming in Banach space, SIAM J. Appl. Math., 15 (1967), pp. 284-293.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Department of Statistics
University of North Carolina
Chapel Hill, N. C. 27514

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

3. REPORT TITLE

Optimality Conditions and Constraint Qualifications in Banach Space

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

F. J. Gould and Jon W. Tolle

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August, 1970	21	11

8a. CONTRACT OR GRANT NO.

N00014-67-A-0321-003 (NR047-095)

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

Institute of Statistics Mimeo Series
Number 707

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

The distribution of this report is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Operations Research Program (Code 434)
Office of Naval Research
Washington, D.C. 20360

13. ABSTRACT

In this paper necessary optimality conditions for non-linear programs in Banach spaces and constraint qualifications for their applicability are considered. A new optimality condition is introduced and a constraint qualification ensuring the validity of this condition is given. When the domain space is a reflexive space, it is shown that the qualification is the weakest possible. If a certain convexity assumption is made, then this optimality condition is shown to reduce to the well-known extension of the Kuhn-Tucker conditions to Banach spaces. In this case the constraint qualification is weaker than those previously given.

DD FORM 1 NOV 65 1473

Security Classification

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Kuhn-Tucker conditions						
Optimality conditions						
Constraint qualifications						
Nonlinear programming						
Banach spaces						

Security Classification