GENERAL NECESSARY CONDITIONS FOR OPTIMIZATION PROBLEMS*

BY HUBERT HALKIN AND LUCIEN W. NEUSTADT

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA (SAN DIEGO), AND DEPARTMENT OF
ELECTRICAL ENGINEERING, UNIVERSITY OF SOUTHERN CALIFORNIA, LOS ANGELES

Communicated by Solomon Lefschetz, August 3, 1966

Introduction.—In the derivation of necessary conditions for various optimization
problems one should distinguish two basic elements: the first element is common
to all problems and deals with optimization in itself, whereas the second element
is an exercise in ordinary differential equations, difference equations, or partial
differential equations, according to the particular nature of the problem under
consideration. The present paper is a contribution to the first of these elements:
we give a very general maximum principle for a mathematical programing problem
over an arbitrary set. Although nonlinear optimal control problems have been our
original motivation, we shall show, by an example, that the results of this paper in-
clude the standard Kuhn-Tucker necessary conditions and generalize them from
finite-dimensional spaces to arbitrary linear spaces. Moreover, this paper includes
and extends all the important necessary conditions in optimal control (with or
without restricted phase coordinates).

The reader will find in Dubovitskiy and Milyutin,\textsuperscript{2} Gamkrelidze,\textsuperscript{4, 5} Halkin,\textsuperscript{6, 7}
and Neustadt\textsuperscript{9, 10} a suitable background for the present paper.

1. Problem.—Given a set \( L \) and real-valued functions \( \varphi_i, i = -\mu, \ldots, 0, \ldots, m \)
defined on \( L \) (where \( m \) and \( \mu \) are given nonnegative integers), find an element \( z_0 \)
\( \in L \) which minimizes \( \varphi_0 \) on the set of all \( z \in L \) that satisfy the constraints \( \varphi_i(z) \leq 0 \)
for \( i = -\mu, \ldots, -1 \) and \( \varphi_i(z) = 0 \) for \( i = 1, \ldots, m \).

2. Assumption.—We shall make the following assumption: there exist convex
sets \( H_i (i = -\mu, \ldots, m) \) in a real linear space \( S \), real valued functions \( h_i (i = -\mu, \ldots, m) \)
defined on the corresponding \( H_i \), and a set \( M \subset S \) with the following properties:

\[
(i) \ 0 \in \bigcap_{i=-\mu}^{m} H_i \quad \text{and} \quad M \subset \bigcap_{i=-\mu}^{m} H_i;
\]

\[
(ii) \ M \text{ is convex;}
\]

\[
(iii) \text{the functionals } h_i \text{ are convex for } i \leq 0 \text{ and linear for } i > 0, \text{ and } h_i(0) = 0 \text{ for each } i; \quad \text{and}
\]

\[
(iv) \text{for every subset } A \text{ of } M \text{ which consists of } m + 1 \text{ elements in general position}
\text{[the elements } y_1, \ldots, y_{m+1} \text{ are in general position if the vectors } (y_j - y_{m+1}), \text{for}
\text{ } j = 1, \ldots, m \text{ are linearly independent], there exist a set } D \subset (0,1) \text{ and a mapping}
\text{ } \theta \text{ from } [A] \times D \text{ ([ } A \text{ ] is the convex hull of } A \text{) into } L \text{ such that}
\]

\[
(a) \ 0 \in D,
\]

\[
(b) \ \lim_{\delta \to 0} \frac{\varphi_i \ast \theta (y, \delta) - \varphi_i(z_0) - h_i(\delta y)}{\delta} \{ = 0 \text{ for each } i = 1, \ldots, m \} \leq 0 \text{ for each } i = -\mu, \ldots, 0
\]

\[
\text{uniformly over } [A], \quad \text{and}
\]

\[
(c) \text{ for every } \delta \in D \text{ and every } i = 1, \ldots, m \text{ the mappings } \varphi_i \ast \theta (\cdot, \delta) \text{ are}
\text{continuous on } [A].
\]

In (iv) (c) continuity on \( [A] \) is to be understood with respect to the ordinary
Euclidean, finite-dimensional topology on \( [A] \).

1066
3. **Maximum Principle.**—Let \( z_0 \) be a solution of the preceding problem satisfying the above hypotheses. Then there exist real numbers \( \alpha_{-\mu}, \ldots, \alpha_\mu \) such that

\[
\alpha \sum_{i=-\mu}^\mu \alpha_i h_i(y) \leq 0 \quad \text{for all } y \in M;
\]

\[
(\alpha) \quad \sum_{i=-\mu}^\mu |\alpha_i| > 0;
\]

\[
(\beta) \quad \sum_{i=-\mu}^\mu \alpha_i > 0;
\]

\[
(\gamma) \quad \alpha_i \leq 0 \quad \text{if } i \leq 0; \quad \text{and}
\]

\[
(\delta) \quad \alpha_i \varphi_i(z_0) = 0 \quad \text{if } i < 0.
\]

In addition, if for some subset \( J \) of \( \{-\mu, \ldots, 0\} \) the point \( 0 \) is an internal point (ref. 3, p. 410) of \( H_i \) for all \( i \in J \), and if \( 0 \in M \), then there exist linear functionals \( l_i(y) \in \mathbb{R}^m \) on \( S \) such that

\[
\sum_{i \in J} \alpha_i l_i(y) + \sum_{i \in \mathbb{R}_{-J}} \alpha_i h_i(y) \leq 0 \quad \text{for all } y \in M, \text{ and}
\]

\[
(\eta) \quad l_i(y) \leq h_i(y) \quad \text{for all } y \in H_i \text{ and every } i \in J.
\]

Moreover, if \( S \) is a topological linear space and, for some \( i \in J \), \( 0 \) is an interior point of \( H_i \) and \( h_i \) is continuous at \( 0 \), then \( l_i \) is continuous on \( S \).

4. **Proof of the Maximum Principle.**—Let us first suppose that \( \varphi_i(z_0) = 0 \) for each \( i < 0 \) and that \( m > 0 \).

Let \( h \) be the linear mapping from \( M \) into \( \mathbb{R}^m \) (Euclidean \( m \)-space) defined by the relation \( h(y) = (h_1(y), \ldots, h_m(y)) \). Let \( \mathbb{M} = \{ y : y \in M, h_i(y) < 0 \text{ for } i = -\mu, \ldots, 0 \} \) and let \( K = h(\mathbb{M}) \). It follows at once that \( \mathbb{M} \) and \( K \) are convex sets in \( S \) and \( \mathbb{R}^m \), respectively. Let us prove that \( 0 \) is not an interior point of \( K \). Indeed, suppose the contrary, so that there is a closed \( m \) simplex \( \mathbb{S} \subset K \) such that \( 0 \) is an interior point of \( S \). We shall construct a map \( \xi \) from \( S \) into \( L \) and a map \( \Phi \) from \( L \) into \( \mathbb{R}^m \) such that

\[
(i) \quad \Phi \cdot \xi \text{ is a continuous map from } S \text{ into } \mathbb{R}^m;
\]

\[
(ii) \quad x - \Phi \cdot \xi(x) \in S \text{ for all } x \in S;
\]

\[
(iii) \quad x \in S \text{ and } \Phi \cdot \xi(x) = 0 \text{ imply that}
\]

\[
\varphi_i \cdot \xi(x) < 0 \quad \text{if } i < 0
\]

\[
\varphi_0 \cdot \xi(x) < \varphi_0(z_0)
\]

\[
\varphi_i \cdot \xi(x) = 0 \quad \text{if } i > 0.
\]

It follows from Brouwer's fixed-point theorem and \( (i), (ii) \) that there exists an \( x_1 \in S \) such that \( x_1 \neq \Phi \cdot \xi(x_1) = x_1 \), i.e., \( \Phi \cdot \xi(x_1) = 0 \). Let \( z_1 = \xi(x_1) \). The point \( z_1 \in L \). Then, from \( (iii) \), we have \( \varphi_i(z_1) < 0 \) if \( i < 0 \) and \( \varphi_0(z_1) < \varphi_0(z_0) \) and \( \varphi_i(z_1) = 0 \) for \( i > 0 \), which contradicts the optimality of \( z_0 \).

Let \( h(y_j), j = 1, \ldots, m + 1 \), be the vertices of \( S \), where \( y_j \in \mathbb{M} \) for \( j = 1, \ldots, m + 1 \). It is easily seen that \( y_1, \ldots, y_{m+1} \) are in general position. For the set \( A = \{ y_1, \ldots, y_{m+1} \} \), let \( D \subset (0,1) \) and \( \theta : [A] \times D \rightarrow L \) be given as per part \( (iv) \) of the assumption. If \( x = \sum_{j=1}^{m+1} \beta_j h(y_j) \in S \), where \( \sum_{j=1}^{m+1} \beta_j = 1 \) and \( \beta_j \geq 0 \) for each \( j \), let \( \xi(x) = \theta \left( \sum_{j=1}^{m+1} \beta_j y_j, \delta^* \right) \), where \( \delta^* \) is some element of \( D \) which will be defined later. If \( z \in L \), let \( \Phi(z) = \frac{1}{\delta^*} (\varphi_1(z), \ldots, \varphi_m(z)) \). The number \( \delta^* \) is determined
Let \( \epsilon_i > 0 \) be such that \( \{x : x \in \mathbb{R}^m, |x| \leq \epsilon_i \} \subset S \), and let \( -2\epsilon_i = \max_{1 \leq j \leq m+1, -\mu \leq i \leq 0} h_i(y) \). By definition of \( M \), \( \epsilon_i > 0 \), and it follows from the convexity of the \( h_i \) that \( h_i(y) < -\epsilon_i \) for all \( y \in [4] \) and \( i \leq 0 \). Now let \( \delta^* \in D \) be such that, for every \( y \in [4] \),

\[
\frac{\varphi_i - \theta(y, \delta^*) - \varphi_i(z_0) - h_i(\delta^*y)}{\delta^*} < \epsilon_i \quad \text{for} \quad i = -\mu, \ldots, 0
\]

\[
\frac{|\varphi_i - \theta(y, \delta^*) - h_i(y)|}{\delta^*} < \frac{\epsilon_i}{m} \quad \text{for} \quad i = 1, \ldots, m.
\]

It is a trivial matter to verify that the maps \( \Phi \) and \( \xi \) satisfy the conditions (i), (ii), and (iii) stated above.

Hence, 0 is not an interior point of \( K \), and there is a hyperplane in \( \mathbb{R}^m \) separating 0 and \( K \), i.e., there are real numbers \( \alpha_i, i = 1, \ldots, m \), not all zero such that

\[
\sum_{i=1}^{m} \alpha_i h_i(y) \leq 0 \quad \text{for all} \quad y \in M.
\]  

(1)

Let \( B = \{(h_{-\mu}(y), \ldots, h_0(y), \sum_{i=1}^{m} \alpha_i h_i(y)) : y \in M \}, B_1 = B + \{(\xi_{-\mu}, \ldots, \xi_0, 0) : \xi_i \geq 0 \quad \text{for} \quad i \leq 0 \} \). It follows from the convexity of the \( h_i \) that \( B_1 \) is a convex subset of \( \mathbb{R}^{m+2} \), and we have, by virtue of (1), that \( B_1 \) does not meet the convex cone \( B_2 = \{(\xi_{-\mu}, \ldots, \xi_0, \xi_i) : \xi_i < 0 \quad \text{for} \quad i \leq 0, \xi_i > 0 \} \). Hence, there is a hyperplane through 0 \( \in \mathbb{R}^{m+2} \) separating \( B_1 \) from \( B_2 \), i.e., there are numbers \( \beta_i(i = -\mu, \ldots, 0, 1) \), not all zero, such that

\[
\sum_{i=-\mu}^{0} \beta_i h_i(y) + \beta_1 \sum_{i=1}^{m} \alpha_i h_i(y) \leq 0 \quad \text{for all} \quad y \in M
\]  

(2)

\[
\beta_i \leq 0 \quad \text{for} \quad i = -\mu, \ldots, 0.
\]  

(3)

Setting \( \alpha_i = \beta_i \) for \( i = -\mu, \ldots, 0 \) and \( \alpha_i = \beta_i \alpha_i \) for \( i > 0 \), the statements (\( \alpha \)), (\( \beta \)), (\( \gamma \)), (\( \delta \)) of the maximum principle follow at once from (2) and (3).

If \( m = 0 \), it is easy to show, on the basis of our hypotheses, that \( \{(h_{-\mu}(y), \ldots, h_0(y)) : y \in M \} \cap \{(\xi_{-\mu}, \ldots, \xi_0) : \xi_i < 0 \quad \text{for each} \quad i \} = \phi \). Consequently, (\( \alpha \)), (\( \beta \)), (\( \gamma \)), and (\( \delta \)) follow as before.

If \( \varphi_i(z_0) < 0 \) for some \( i < 0 \), we shall suppose, without loss of generality, that \( \varphi_i(z_0) = 0 \) for \( i = -\mu', \ldots, -1 \), and \( \varphi_i(z_0) < 0 \) for \( i = -\mu, \ldots, -\mu' - 1 \). We can repeat the preceding arguments with \( \mu \) replaced by \( \mu' \). The only change that need be made is to choose \( \delta^* \in D \) sufficiently small (when achieving the contradiction) so that \( \varphi_i + \xi(x) < 0 \) for \( x \in S \) and for \( i = -\mu, \ldots, -\mu' - 1 \) as well as for \( i = -\mu', \ldots, -1 \) (it is easily seen that this can be done on the basis of our assumption).

We then arrive at relations (\( \alpha \)), (\( \beta \)), and (\( \gamma \)), but with \( \mu \) replaced by \( \mu' \). Setting \( \alpha_i = 0 \) for \( i < -\mu' \), we obtain the desired relations (\( \alpha \)), (\( \beta \)), (\( \gamma \)), and (\( \delta \)).

Let us show that if \( 0 \in M \) and \( J \) as is indicated in the theorem statement, then there exist linear functionals \( l_i(i \in J) \) on \( S \) such that (\( \epsilon \)) and (\( \eta \)) are satisfied. Let \( J = \{i_1, \ldots, i_k\} \). We define the following two subsets of \( S \times \mathbb{R}^1 \):

\[
V_1 = \{(y, \lambda) : y \in H_{i_0}, \lambda > h_i(y)\}
\]

\[
V_2 = \{(y, \lambda) : y \in H_{i_0}, \lambda < h_i(y)\}
\]
\[ W_1 = \begin{cases} \{ (y, \lambda) : y \in M, \lambda < - \sum_{i=1}^{m} \alpha_{ti} \lambda_{i}(y) \} & \text{if } \alpha_{ti} \neq 0 \\ \{ (0,0) \} & \text{if } \alpha_{ti} = 0. \end{cases} \]

It is easily verified that \( V_1 \) and \( W_1 \) are convex and that \((0,1)\) is an internal point of \( V_1 \). Further, it follows from (a) that \( V_1 \cap W_1 = \emptyset \). Hence, (ref. 3, p. 412, Th. 12), there exist a real number \( \alpha \) and a nonzero linear functional \( l_i(y, \lambda) \) on \( \mathbb{R}^1 \) such that \( l_i(y', \lambda') \leq \alpha \leq l_i(y, \lambda) \) whenever \( (y, \lambda) \in V_1 \) and \( (y', \lambda') \in W_1 \). It is evident that \((\epsilon y, \epsilon \lambda) \in V_1 \) (respectively \( W_1 \)) whenever \((y, \lambda) \in V_1 \) (respectively \( W_1 \)) and \( 0 < \epsilon \leq 1 \). From this it follows that \( \alpha = 0 \). Further, \( l_i(y, \lambda) = l_i^*(y) + c_i \lambda \) where \( l_i^* \) is a linear functional on \( \mathbb{R} \) and \( c_i \) is a real number. If \( c_i = 0 \) then \( l_i^*(y) \geq 0 \) for all \( y \in H_{ti} \), and \( l_i^* \neq 0 \). But this is impossible since \( 0 \) is an internal point of \( H_{ti} \). Thus, since \((0,1) \in V_1, c_i > 0 \). If we set \( l_i(y) = -c_i^{-1}l_i^*(y) \), then it is straightforward to show that
\[
\alpha_{ti}l_i(y) + \sum_{i=-\mu}^{\mu} \alpha_{ti}h_i(y) \leq 0 \text{ for all } y \in M
\]
\[
l_i(y) \leq h_i(y) \text{ for all } y \in H_{ti}.
\]

We now define the sets \( V_2 \) and \( W_2 \) as follows:
\[
V_2 = \{ (y, \lambda) : y \in H_{ti}, \lambda > h_i(y) \}
\]
\[
W_2 = \begin{cases} \{ (y, \lambda) : y \in M, \lambda < - \sum_{i=-\mu}^{\mu} \frac{\alpha_{ti}}{\alpha_{ti}} h_i(y) - \frac{\alpha_{ti}}{\alpha_{ti}} l_i(y) \} & \text{if } \alpha_{ti} \neq 0 \\ \{ (0,0) \} & \text{if } \alpha_{ti} = 0. \end{cases}
\]

Just as above we can show that there exists a linear functional \( l_i \) defined on \( \mathbb{R} \) such that
\[
\sum_{j=1}^{2} \alpha_{ti}l_j(y) + \sum_{i=-\mu}^{\mu} \alpha_{ti}h_i(y) \leq 0 \text{ for all } y \in M
\]
\[
l_i(y) \leq h_i(y) \text{ for all } y \in H_{ti}.
\]

Successively repeating the above construction we arrive at relations (e) and (f).

If \( \mathbb{R} \) is a topological linear space, 0 is an interior point of \( H_{ti} \), and \( h_i \) is continuous at 0 (for some \( i \in J \)), it follows from (e) that \( l_i(y) \leq 1 \) in a neighborhood of 0. Therefore (ref. 3, p. 417, Lemma 7), \( l_i \) is continuous on \( \mathbb{R} \).

5. Remarks.—The Maximum Principle can be proved in essentially the same way if property (ii) of the given assumption is omitted and if the beginning of property (iv) is replaced by

(iv)* for every subset \( A \) of \( M \) which consists of at most \( 2m \) elements there exist . . .

In the proof given in §4, we use the fact that if the point 0 is interior to a set \( K \subset \mathbb{R}^m \), then 0 is interior to the convex hull of a subset of \( K \) consisting of \( m + 1 \) points. With the above modification of our assumption, this statement should be replaced by: if the point 0 is interior to the convex hull of a set \( K \subset \mathbb{R}^m \), then 0 is interior
to the convex hull of a subset of $K$ consisting of at most $2m$ points. (See ref. 1, for which we thank Victor Klee.)

6. **Two Applications.**—Let us suppose that the set $L$ in the problem statement is a convex set in a real linear space $S$, that the functionals $\varphi_i$, $i = 1, \ldots, m$, are continuous on any $m$ simplex contained in $L$, and that there exist linear functionals $h_i$, $i = 1, \ldots, m$, defined on $S$, such that, for $i = 1, \ldots, m$,

$$
\lim_{\delta \to 0^+} \frac{\varphi_i(z_0 + \delta y) - \varphi_i(z_0)}{\delta} = h_i(y) \quad \text{for every } y \in L - z_0,
$$

where $z_0 \in L$ is a solution of our problem and the convergence in (4) is uniform over any $m$ simplex in $L - z_0$. (If $S$ is a Banach space and if the $\varphi_i$ possess a Fréchet differential at $z_0$, then it is easily seen that this differentiability condition is satisfied.) In addition, we shall assume that either (I) the $\varphi_i$, for $i \leq 0$, are defined on all of $S$ and are convex on $S$, or (II) there exist convex functionals $h_i$, for $i = -\mu, \ldots, 0$, defined on $S$ such that (4) holds for each $i \leq 0$ with the convergence uniform over any $m$ simplex in $L - z_0$.

If we set $M = L - z_0$ and $H_i = S$ for each $i$, and let $h_i(y) = \varphi_i(y + z_0) - \varphi_i(z_0)$ for each $i \leq 0$ in case (I), then it is straightforward to verify that $S$, $M$, the $H_i$, and the $h_i$ satisfy the assumption described in §2—both in cases (I) and (II). Indeed it is only necessary to choose $D = (0,1)$ and $\theta(y, \delta) = z_0 + \delta y$ (independent of $A$) in (iv).

Consequently, the maximum principle [i.e., relations (a)-(q)] holds with $J = \{-\mu, \ldots, 0\}$. If $L$ is a subspace of $S$, $\leq$ in (e) can be replaced by $=$. If $L = S = \mathbb{R}^n$ and $m = 0$, we obtain the conventional Kuhn-Tucker conditions [for convex programing problems in case (I), and for general nonlinear programing problems in case (II)]. If $L = S$, $S$ is a Banach space, $m = 0$, and the $\varphi_i$ are continuous, our maximum principle yields the generalized Kuhn-Tucker conditions obtained by Hurwicz and Pshenichny.\(^{11}\)

To apply our theorem to nonlinear optimal control problems, it is necessary to consider sets $L$ (in suitable linear spaces), which are not necessarily convex, but which possess “first-order, convex approximations,” near $z_0$. For details, the reader is referred to references 9 and 10.

---


6. Two Applications.—Let us suppose that the set $L$ in the problem statement is a convex set in a real linear space $S$, that the functionals $\varphi_i$, $i = 1, \ldots, m$, are continuous on any $m$ simplex contained in $L$, and that there exist linear functionals $h_i$, $i = 1, \ldots, m$, defined on $S$, such that, for $i = 1, \ldots, m$,

$$
\lim_{\delta \to 0^+} \frac{\varphi_i(z_0 + \delta y) - \varphi_i(z_0)}{\delta} = h_i(y) \quad \text{for every } y \in L - z_0,
$$

where $z_0 \in L$ is a solution of our problem and the convergence in (4) is uniform over any $m$ simplex in $L - z_0$. (If $S$ is a Banach space and if the $\varphi_i$ possess a Fréchet differential at $z_0$, then it is easily seen that this differentiability condition is satisfied.) In addition, we shall assume that either (I) the $\varphi_i$, for $i \leq 0$, are defined on all of $S$ and are convex on $S$, or (II) there exist convex functionals $h_i$, for $i = -\mu, \ldots, 0$, defined on $S$ such that (4) holds for each $i \leq 0$ with the convergence uniform over any $m$ simplex in $L - z_0$.

If we set $M = L - z_0$ and $H_i = S$ for each $i$, and let $h_i(y) = \varphi_i(y + z_0) - \varphi_i(z_0)$ for each $i \leq 0$ in case (I), then it is straightforward to verify that $S$, $M$, the $H_i$, and the $h_i$ satisfy the assumption described in §2—both in cases (I) and (II). Indeed it is only necessary to choose $D = (0,1)$ and $\theta(y, \delta) = z_0 + \delta y$ (independent of $A$) in (iv).

Consequently, the maximum principle [i.e., relations (a)-(q)] holds with $J = \{-\mu, \ldots, 0\}$. If $L$ is a subspace of $S$, $\leq$ in (e) can be replaced by $=$. If $L = S = \mathbb{R}^n$ and $m = 0$, we obtain the conventional Kuhn-Tucker conditions [for convex programing problems in case (I), and for general nonlinear programing problems in case (II)]. If $L = S$, $S$ is a Banach space, $m = 0$, and the $\varphi_i$ are continuous, our maximum principle yields the generalized Kuhn-Tucker conditions obtained by Hurwicz and Pshenichny.\(^{11}\)

To apply our theorem to nonlinear optimal control problems, it is necessary to consider sets $L$ (in suitable linear spaces), which are not necessarily convex, but which possess “first-order, convex approximations,” near $z_0$. For details, the reader is referred to references 9 and 10.

---

This research was supported by the U.S. Air Force Office of Scientific Research, Office of Aerospace Research, under AFOSR grant 1039-66 (Hubert Halkin, Department of Mathematics, University of California, San Diego) and under grant AF-AFOSR 1039-66 (Lucien W. Neustadt, Department of Electrical Engineering, University of Southern California, Los Angeles).


Neustadt, L. W., "An abstract variational theory with applications to a broad class of optimization problems II: Applications," to appear. Available as report of the Electronic Sciences Laboratory, University of Southern California, Los Angeles.