REVUE FRANÇAISE D'AUTOMATIQUE, INFORMATIQUE, RECHERCHE OPÉRATIONNELLE. RECHERCHE OPÉRATIONNELLE

T. L. MAGNANTI

Non-linear programming and the maximum principle for discrete time optimal control problems

Revue française d'automatique, informatique, recherche opérationnelle. Recherche opérationnelle, tome 9, n° V3 (1975), p. 75-91

<http://www.numdam.org/item?id=RO_1975__9_3_75_0>

© AFCET, 1975, tous droits réservés.

L'accès aux archives de la revue « Revue française d'automatique, informatique, recherche opérationnelle. Recherche opérationnelle » implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme Numérisation de documents anciens mathématiques

http://www.numdam.org/

R.A.I.R.O.
(9e année, octobre 1975, V-3, p. 75 à 91)

NON-LINEAR PROGRAMMING AND THE MAXIMUM PRINCIPLE FOR DISCRETE TIME OPTIMAL CONTROL PROBLEMS (*)

by T. L. MAGNANTI (1)

Abstract. — *Results in non-linear programming are used to prove a generalized version of the maximum principle for fixed-time discrete optimal control problems. Proofs are based upon the implicit function theorem and a theorem of the alternative for systems of linear inequalities over a convex set: they do not, as in the past, require Brouwer's fixed-point theorem.*

INTRODUCTION

It is becoming apparent that optimal control theory and non-linear programming are highly related. This point of view is illustrated by the recent text of Canon, Cullum and Polak [1]. It seems yet to be fully exploited, however.

Our purpose here is to utilize results from non-linear programming to establish a generalized version of the maximum principle for fixed-time discrete optimal control problems. These problems have been treated in various degrees of generality by a number of researchers ([1], [3], [4], [5], [10], [11], [12]). We shall consider an extension of a version originally discussed by Halkin [3] and later generalized by Holtzman [4] and Canon, Cullum and Polak [1]. Thus, we will not require convexity of the set of admissible controls (though a certain convexity condition will be imposed) nor do we require differentiability with respect to the control variables. We also admit inequality constraints on the state variables as well as equality constraints on the initial and terminal state vectors.

(*) Reçu septembre 1974.
Supported in part by the U.S. Army Research Office (Durham) under contract DAHC04-70-C-0058.

(1) Sloan School of Management M.I.T.

For motivation, we might note that linear optimal problems trivially fit out framework. In addition, Halkin [3] has shown that discrete approximations to continuous problems satisfy the hypothesis that we shall impose.

The non-linear programming approach that we adopt should be contrasted with previous approaches. We do not require intricate arguments based upon Brouwer's fixed point theorem, nor do we rely upon canonical approximations [1]. Instead, we utilize two basic results from non-linear programming stated as Lemma's 1 and 2 below. The first result is a direct consequence of the implicit function theorem and a well-known lemma (*see* [7]) due to Motzkin. It was established by Mangasarian and Fromowitz ([7], [8]). A short proof is given in [5]. Though the result is stated as a "maximum principle" in [7], it has not been used previously in the present optimal control context, but rather as a direct extension of the classical Fritz John theorem of non-linear programming applied with differentiability requirements that we do not impose.

The second result is essentially a special case of a theorem of Fan, Glicksburg and Hoffman ([2], [7], p. 63). The version that we give can be established easily *via* an elementary separating hyperplane argument.

Before stating these results, let us set some notation. R denotes the rea numbers, R^n n-dimensional real space (with the usual topology). Subscripting denotes distinct vectors and superscripting vector components. This same convention will be applied to functions in the sense that if $g : R^n \to R^m$, then g^i denotes the i-th coordinate function, $g^i : R^n \to R$.

$\nabla g(\bar{x})$ denotes the matrix $(h_{ij}) = (\partial g^i(x)/\partial x^j)_{x=\bar{x}}$. Similarly, if $f : R^n \to R$ is a function, $\nabla f(\bar{x})$ is the gradient vector evaluated at $x = \bar{x}$, i. e., $(\partial f(x)/\partial x^1, \ldots, \partial f(x)/\partial x^n)_{x=\bar{x}}$.

Finally, if α, $v \in R^n$ and A is a real valued matrix, then αv denotes inner product as does αA and $A v$, i. e., $(A v)^i = \sum_{j=1}^{n} A_{ij} v^j$. Also, vector equalities and inequalities hold componentwise.

Lemma 1 (first linearization lemma): Let $C \subseteq R^n$ be a convex set with a non-empty interior and let $f : R^n \to R$, $\varphi : R^n \to R^{m_1}$, $\xi : R^n \to R^{m_2}$ be given functions. Let $\bar{x}$ solve

$$\max f(x),$$

subject to

$$\varphi(x) = 0,$$

$$\xi(x) \leqq 0,$$

$$x \in C.$$

Assume that $\varphi(x)$ is continuously differentiable in a neighborhood of $\bar{x}$ and that $f(x)$ and $\xi(x)$ are differentiable at $\bar{x}$. Then, if the vectors

$$\{\nabla \varphi^j(\bar{x}) : j = 1, \ldots, m_1\}$$

are linearly independent, the system

$$\begin{aligned} \nabla f(\bar{x})(\Delta x) &> 0, \\ \nabla \varphi(\bar{x})(\Delta x) &= 0, \\ \nabla \xi^i(\bar{x})(\Delta x) < 0 \quad &\text{for} \quad i \in \{1 \leqq j \leqq m_2 : \xi^j(\bar{x}) = 0\}, \\ \bar{x} + \Delta x &\in \{\text{interior of } C\}, \end{aligned}$$

has no solution in the variables $(\Delta x)^k$, $k = 1, \ldots, n$.

Lemma 2: Let $C \subseteq R^n$ be a convex set and let A and B be m_1 by n and m_2 by n real valued matrices. If the system

$$\begin{aligned} Ax &= 0, \\ Bx &< 0, \\ x &\in C, \end{aligned}$$

has no solution, then there exist $\sigma \in R^{m_1}$, $\alpha \in R^{m_2}$, $\alpha \leqq 0$, $(\sigma, \alpha) \neq 0$ such that

$$\sigma(Ax) + \alpha(Bx) \leqq 0 \quad \text{for all } x \in C. \tag{$\bigstar$}$$

For convenience, let us record the following special case of this result to be used later.

Remark 1: Let $\bar{C} \subseteq R^m$ be a convex set; let A_1, A_2, B_1 and B_2 be respectively m_1 by N, m_1 by M, m_3 by N, and m_4 by M real valued matrices. Then, by Lemma 2 if the system

$$\begin{aligned} A_1 v + A_2 w + \rho &= 0, \\ B_1 v &< 0, \\ B_2 w &< 0, \\ \rho &\in \bar{C}, \end{aligned}$$

has no solution, then there exist $\sigma \in R^{m_1}$, $\psi \in R^{m_3}$. $\mu \in R^{m_4}$, $\psi \leqq 0$, $\mu \leqq 0$, $(\sigma, \psi, \mu) \neq 0$ satisfying

$$\begin{aligned} \sigma A_1 + \psi B_1 &= 0, \\ \sigma A_2 + \mu B_2 &= 0, \\ \sigma \rho \leqq 0 \quad &\text{for all } \rho \in C. \end{aligned}$$

If there are no B_1 and/or B_2 constraints, we may eliminate ψ and/or μ above. To obtain the equality constraints above, simply take positive and negative unit vectors for v and w in (★).

1. PROBLEM STATEMENT

Let $\Omega_0, \Omega_1, \ldots, \Omega_{T-1}$ be given subsets of R^r. Assume that each of the following functions is given:

$$\begin{aligned} f_t &: R^n \times \Omega_t \to R^n \qquad (t = 0, 1, \ldots, T-1), \\ h &: R^n \to R^l, \\ g_0 &: R^n \to R, \\ g &: R^n \to R^m, \\ q_t &: R^n \to R^{m_t} \qquad (t = 0, 1, \ldots, T). \end{aligned}$$

Problem: Determine state vectors $x_0, x_1, \ldots, x_T \in R^n$ and control vectors $u_0, u_1, \ldots, u_{T-1} \in R^r$ to

$$\max g_0(x_T),$$

subject to

$$\left.\begin{aligned} h(x_0) &= 0, \\ x_{t+1} - x_t = f_t(x_t, u_t) & \qquad (t = 0, 1, \ldots, T-1), \\ g(x_T) &= 0, \\ q_t(x_t) \leqq 0 & \qquad (t = 0, 1, \ldots, T), \\ u_t \in \Omega_t & \qquad (t = 0, \ldots, T-1). \end{aligned}\right\} \quad \text{(P)}$$

Suppose that $\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_T$ and $\bar{u}_0, \bar{u}_1, \ldots, \bar{u}_{T-1}$ solve (P). Some typical hypotheses for the problem are:

H1) For every $u \in \Omega_t$ the vector valued function $f_t(x, u)$ is continuously differentiable (1) with respect to x in a neighborhood of $x = \bar{x}_t$ $(t = 0, 1, \ldots, T-1)$.

H2) For every $x \in R^n$, the set $\{f_t(x, u) : u \in \Omega_t\}$ is a convex subset of R^n $(t = 0, 1, \ldots, T-1)$.

H3) The vector valued functions $h(x)$ and $g(x)$ are each continuously differentiable in neighborhoods of $x = \bar{x}_0$ and $x = \bar{x}_T$ respectively. $g_0(x)$ and the vector valued functions $q_t(x)$ $(t = 0, 1, \ldots, T)$ are differentiable at respectively $x = \bar{x}_T$, $x = \bar{x}_0, \ldots, x = \bar{x}_T$.

---

(1) A referee has suggested that the results of this paper remain valid if we replace all continuous differentiability assumptions with weaker strong differentiability assumptions which are discussed, for example, in [9].

H4) The vectors $\nabla h^j(\bar{x}_0)$ $(j = 1, \ldots, l)$ are linearly independent as are the vectors $\nabla g^j(\bar{x}_T)$ $(j = 1, \ldots, m)$.

H5) For each $t = 0, 1, \ldots, T$ with $m_t \neq 0$, no nonzero vector $\lambda_t \in R^{m_t}$, $\lambda_t \leqq 0$ satisfies

$$\lambda_t \nabla q_t(\bar{x}_t) = 0,$$
$$\lambda_t q_t(\bar{x}_t) = 0.$$

(Note that if for each fixed t, the vectors $\{ \nabla q_t^i(x_t) : q_t^i(x_t) = 0 \}$ are linearly independent, then condition H5 is valid.)

The *discrete maximum principle* for (P) states that there exists (adjoint) vectors $p_0, p_1, \ldots, p_T \in R^n$, (multiplier) vectors $\lambda_t \in R^{m_t}$, $\lambda_t \leqq 0$ $(t = 0, 1, \ldots, T-1)$, vectors $\alpha \in R^l$, $\beta \in R^m$ and a scalar $\beta_0 \geqq 0$ satisfying :

0) Not all the quantities $p_0, p_1, \ldots, p_T$, α, β and β_0 are zero.

1) Hamiltonian maximization: for all $u \in \Omega_t$,

$$p_{t+1} f_t(x_t, u_t) \geqq p_{t+1} f_t(\bar{x}_t, u) \qquad (t = 0, 1, \ldots, T-1).$$

2) Adjoint equations [1]:

$$p_t - p_{t+1} = p_{t+1} \nabla_x f_t(x_t, \bar{u}_t) + \lambda_t \nabla q_t(\bar{x}_t) \qquad (t = 0, 1, \ldots, T-1).$$

3) Transversality conditions :

$$p_0 = \alpha \nabla h(x_0),$$
$$p_T = \beta \nabla g(\bar{x}_T) + \beta_0 \nabla g_0(x_T) + \lambda_T \nabla q_T(x_T).$$

4) Complementary slackness:

$$\lambda_t q_t(\bar{x}_t) = 0 \qquad (t = 0, 1, \ldots, T).$$

Our purpose will be to prove an extended version of the fact that H1-H5 imply this discrete maximum principle. We will use the first linearization Lemma and Remark 1 as basic tools.

Towards this end, it will be convenient to consider a generalized version of (P). Let U be a given subset of R^L and suppose that each of the following functions are given:

$$F : R^N \to R,$$
$$H : R^N \times R^M \times U \to R^r,$$
$$G : R^N \to R^m,$$
$$Q : R^N \to R^{s_1},$$
$$\bar{Q} : R^M \to R^{s_2}.$$

[1] $\nabla_x f_t(x_t, u_t)$ denotes the matrix $\nabla d(x_t)$ where $d(x) = f_t(x, u_t)$. This same convention applies to any function of two or more vectors.

The new problem is:

Determine $y \in R^N$, $z \in R^N$ and $u \in R^L$ to

$$\max F(y)$$

subject to

$$\left.\begin{array}{r} G(y) = 0, \\ Q(y) \leqq 0, \\ \overline{Q}(z) \leqq 0, \\ H(y, z, u) = 0, \\ u \in U. \end{array}\right\} \qquad (\mathrm{P}')$$

Assume that y, $\bar{z}$, u solve (P'). We shall impose the following analogs of hypothesis H1-H5.

$\overline{\mathrm{H1}}$) For every $u \in U$ the vector valued function $H(y, z, u)$ is continuously differentiable with respect to (y, z) in a neighborhood of $(y, z) = (\bar{y}, \bar{z})$.

$\overline{\mathrm{H2}}$) For every $y \in R^N$, $z \in R^M$, the set $\{ H(y, z, u) : u \in U \}$ is a convex subset of R^r.

$\overline{\mathrm{H3}}$) $G(y)$ is continuously differentiable in a neighborhood of $y = y$. $F(y)$ and $Q(y)$ are differentiable at $y = \bar{y}$ and $\overline{Q}(z)$ is differentiable at $z = \bar{z}$.

$\overline{\mathrm{H4}}$) The vectors $\nabla G^j(y)$ $(j = 1, \ldots, m)$ are linearly independent.

$\overline{\mathrm{H5}}$) Let

$$A = \{ i \in \{ 1, \ldots, s_1 \} : Q^i(\bar{y}) = 0 \},$$
$$\overline{A} = \{ i \in \{ 1, \ldots, s_2 \} : \overline{Q}^i(\bar{z}) = 0 \}$$

denote indices of "active constraints" and let

$$\bar{s}_1 = |A|, \qquad s_2 = |\overline{A}|$$

($|\cdot|$ is cardinality).

Also, let $Q^A : R^N \to R^{s_1}$ denote the vector valued function defined by restricting the Q^i $(i = 1, \ldots, s_1)$ to A and similarly define $\overline{Q}^{\bar{A}}(z)$. Then $A = \varnothing$ or there is no non-zero $\mu^A \in R^{\bar{s}_1}$, $\mu^A \leqq 0$ with $\mu^A \nabla Q^A(\bar{y}) = 0$ and $\overline{A} = \varnothing$ or there is no non-zero $\bar{\mu}^{\bar{A}} \in R^{\bar{s}_2}$, $\mu^{\bar{A}} \leqq 0$ with $\bar{\mu}^A \nabla \overline{Q}^{\bar{A}}(\bar{z}) = 0$.

Note that problem (P) corresponds to the case where

$$z = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{T-1} \end{pmatrix}, \qquad u = \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_{T-1} \end{pmatrix}, \qquad y = x_T,$$

$$U = \Omega_0 \times \Omega_1 \times \ldots \times \Omega_{T-1}.$$

$$F(y) = g_0(x_T),$$

$$G(y) = g(x_T),$$

$$H(y, z, u) = \begin{bmatrix} h(x_0) & & & & \\ x_0 + f_0(x_0, u_0) & -x_1 & & & \\ & x_1 + f_1(x_1, u_1) & -x_2 & & \\ & & & \ddots & \\ & & & & -x_{T-2} \\ & & & & x_{T-1} + f_{T-1}(x_{T-1}, u_{T-1}) \quad -x_T \end{bmatrix},$$

$$\bar{Q}(z) = \begin{pmatrix} q_0(x_0) \\ q_1(x_1) \\ \vdots \\ q_{T-1}(x_{T-1}) \end{pmatrix} \quad \text{and} \quad Q(y) = q_T(x_T).$$

Note that

$$[\nabla_z H(y, z, u), \nabla_y H(y, z, u)] = \begin{bmatrix} \nabla h(x_0) & & & & \\ I + \nabla_x f_0(x_0, u_0) & -I & & & \\ & I + \nabla_x f_1(x_1, u_1) & -I & & \\ & & & \ddots & \\ & & & & I + \nabla_x f_{T-1}(x_{T-1}, u_{T-1}) \quad -I \end{bmatrix}.$$

With these associations $\overline{\text{H1}}$-$\overline{\text{H5}}$ are direct consequences of H1-H5.

Observe that the assumptions imposed upon problem (P′) do not require that y can be expressed as a function of z and u from $H(y, z, u) = 0$, although this is true for problem (P) with the associations given above.

II. GENERALIZED DISCRETE MAXIMUM PRINCIPLE

The proof of the generalized discrete maximum principle becomes particularly transparent when $U = \{ u_1, \ldots, u_k, u \}$. Let us begin by briefly introducing the argument in this case. The idea is to use the convexity of $\{ H(y, z, u) : u \in U \}$, i. e. assumption $\overline{\mathrm{H2}}$, to replace the condition $u \in U$ in (P′) so that Lemma 1 applies. Given any y and z and any $\theta_1, \theta_2, \ldots, \theta_k$ with

$$\sum_{j=1}^{k} \theta_j \leqq 1,$$

$$\theta_j \geqq 0 \qquad (j = 1, \ldots, k).$$

$\overline{\mathrm{H2}}$ implies the existence of a $u = u(\theta_1, \ldots, \theta_k, x, y) \in U$ such that

$$H(x, y, u) = \left(1 - \sum_{j=1}^{k} \theta_j\right) H(x, y, \bar{u}) + \sum_{j=1}^{k} \theta_j H(x, y, u_j).$$

Consequently, $\bar{x}$, $\bar{y}$ and $\theta_j = 0$ must solve problem (P′) when $H(x, y, u)$ is replaced by the right-hand-side in the last equality and conditions $\sum \theta_j \leqq 1$ and $\theta_j \geqq 0$ serve as the convex set C in Lemma 1. The maximum principle now can be established essentially by applying Lemmas 1 and 2 to the new problem. Details are presented below in the context of the general case.

When the set U is not finite, a similar argument can be used by determining "finite generators" for U. The following elementary fact concerning matrix valued functions is useful. Let U be a given set and suppose that for each $u \in U$, $\psi(u)$ is a real valued matrix. We will call ψ linearly independent over U if there is no $\pi \in R^{d_1}$, $\pi \neq 0$ such that $\pi \psi(u) = 0$ for all $u \in U$.

Remark 2: ψ is linearly independent over U if and only if there are $u_1, \ldots, u_D \in U$ such that the matrix

$$M = [\psi(u_1), \psi(u_2), \ldots, \psi(u_D)]$$

has full row rank.

Proof: If $\pi \in R^{d_1}$, $\pi \neq 0$ and $\pi \psi(u) = 0$ for all $u \in U$, then $\pi M = 0$ for any choice of $u_1, u_2, \ldots, u_D \in U$. Thus no M chosen as above can have full row rank.

Conversely, suppose that ψ is linearly independent over U and that M is given. Let rows $r_1, \ldots, r_k$, $k < d_1$ be linearly independent rows of M spanning its row space. Then any other row r_{k+1} of M can be expressed uniquely *via* $\pi M = 0$ with $\pi_{r_{k+1}} = 1$ and $\pi_i = 0$ for $i \neq r_1, r_2, \ldots, r_k$. By the linear independence of ψ, there is a u_{D+1} such that $\pi\psi(u_{D+1}) \neq 0$. Thus, rows

$r_1, \ldots, r_k, r_{k+1}$ of $[\psi(u_1), \ldots, \psi(u_D), \psi(u_{D+1})]$ are linearly independent. This matrix either has full row rank or we may continue adding $\psi(u_j)$ as above until the resulting matrix does. ■

Remark 3: Suppose that ψ above is given by $\psi(u) = \begin{pmatrix} A & \\ B & \varphi(u) \end{pmatrix}$ for fixed real matrices A and B where $\varphi(u)$ is a real valued matrix. Then ψ is linearly independent if and only if there are $u_1, \ldots, u_D \in U$ such that

$$M' = \begin{pmatrix} A & \\ B & \varphi(u_1), \ldots, \varphi(u_D) \end{pmatrix}$$

has full row rank.

Proof: The row ranks of $\begin{pmatrix} A & & A & & \ldots & A & \\ B & \varphi(u_1) & B & \varphi(u_2) & & B & \varphi(u_D) \end{pmatrix}$ and M' are the same. ■

Using these results, the first linearization Lemma and Remark 1, we now show:

Theorem 1 (generalized discrete maximum principle): Let $y = \bar{y}$, $x = \bar{x}$ and $u = \bar{u}$ solve control problem (P'); assume $\overline{\text{H1}}$-$\overline{\text{H3}}$ and $\overline{\text{H5}}$. Then there exist a real number $\beta_0 \geqq 0$ and vectors $\beta \in R^m$, $\sigma \in R^r$, $\mu \in R^{s_1}$, $\bar{\mu} \in R^{s_2}$, $\delta \in R^N$ with $\mu \leqq 0$, $\bar{\mu} \leqq 0$, $(\beta_0, \beta, \sigma) \neq 0$, satisfying

(*a*) $$-\delta = \beta_0 \nabla F(\bar{y}) + \beta \nabla G(\bar{y}) + \mu \nabla Q(\bar{y});$$

(*b*) $$\begin{cases} \delta = \sigma \nabla_y H(\bar{y}, \bar{z}, \bar{u}), \\ 0 = \sigma \nabla_z H(\bar{y}, \bar{z}, \bar{u}) + \bar{\mu} \nabla Q(\bar{z}); \end{cases}$$

(*c*) $$\mu Q(\bar{y}) = 0, \qquad \bar{\mu} \overline{Q}(\bar{z}) = 0;$$

(*d*) $$\sigma [H(\bar{y}, \bar{z}, \bar{u}) - H(\bar{y}, \bar{z}, u)] \geqq 0 \quad \text{for all } u \in U.$$

Proof: If the matrix valued function

$$\psi(u) = \begin{pmatrix} \nabla G(\bar{y}) & & \\ \nabla_y H(\bar{y}, \bar{z}, \bar{u}) & \nabla_z H(\bar{y}, \bar{z}, \bar{u}) & H(\bar{y}, \bar{z}, u) \end{pmatrix}$$

is not linearly independent over U, then there are $\beta \in R^m$, $\sigma \in R^r$, $(\beta, \sigma) \neq 0$ such that

$$\delta \equiv -\beta \nabla G(\bar{y}) = \sigma \nabla_y H(\bar{y}, \bar{z}, \bar{u})$$

$$\sigma \nabla_z H(\bar{y}, \bar{z}, \bar{u}) = 0$$

and $$\sigma H(\bar{y}, \bar{z}, u) = 0 \qquad \text{for all } u \in U.$$

Taking β_0, μ, $\bar{\mu}$ all zero and incorporating

$$H(\bar{y}, \bar{z}, \bar{u}) = 0 \quad \text{in} \quad \sigma H(\bar{y}, \bar{z}, u) = 0,$$

the conclusions (*a*)-(*d*) are satisfied.

On the other hand, if ψ is linearly independent over U, then by Remark 3, there are $u_1, \ldots, u_D \in U$ such that

$$M' = \begin{bmatrix} \nabla G(\bar{y}) & & & & & \\ \nabla_y H(\bar{y}, \bar{z}, \bar{u}) & \nabla_z H(\bar{y}, \bar{z}, \bar{u}) & H(\bar{y}, \bar{z}, u_1) & \ldots & H(\bar{y}, \bar{z}, u_D) & H(\bar{y}, \bar{z}, u) \end{bmatrix}$$

has full row rank. This result will provide the hypothesis for an application of Lemma 1.

Let $u \in U$ be fixed. By $\overline{\text{H2}}$, given any $\theta_j \geqq 0$ $(j = 1, \ldots, D+1)$, $\sum_{j=1}^{D+1} \theta_j \leqq 1$, $y \in R^N$, $z \in R^M$, there is a $u(\theta_1, \ldots, \theta_{D+1}, y, z) \in U$ such that

$$H(y, z, u(\theta_1, \ldots, \theta_{D+1}, y, z)) = \left(1 - \sum_{j=1}^{D+1} \theta_j\right) H(y, z, \bar{u}) + \sum_{j=1}^{D} \theta_j H(y, z, u_j) + \theta_{D+1} H(y, z, u).$$

Consequently, $y = \bar{y}$, $z = \bar{z}$, $\theta_j = \bar{\theta}_j = 0$ $(j = 1, \ldots, D+1)$ must solve the problem (in the variables y, z, $\theta_1, \ldots, \theta_{D+1}$):

$$\max F(y)$$

subject to

$$\left.\begin{aligned} & G(y) = 0, \\ & \underline{Q}(y) \leqq 0, \\ & \bar{Q}(z) \leqq 0, \\ & \left(1 - \sum_{j=1}^{D+1} \theta_j\right) H(y, z, \bar{u}) + \sum_{j=1}^{D} \theta_j H(y, z, u_j) + \theta_{D+1} H(y, z, u) = 0, \\ & (y, z, \theta_1, \ldots, \theta_{D+1}) \in C, \end{aligned}\right\} \quad (\bar{\text{P}})$$

where
$$C = \left\{(y, z, \theta_1, \ldots, \theta_{D+1}) : y \in R^N,\ z \in R^M, \sum_{j=1}^{D+1} \theta_j \leqq 1,\ \theta_j \geqq 0 (j = 1, \ldots, D+1)\right\}.$$

Utilizing $\overline{\text{H1}}$ and $\overline{\text{H3}}$, we see that this problem satisfies the hypothesis of the first linearization lemma with the association $x = (y, z, \theta_1, \ldots, \theta_{D+1})$ and the obvious associations for φ and ξ. Note that $\nabla \varphi(\bar{x}) = M'$ above which has full row rank. By that lemma, the following system has no solution:

$$\nabla F(\bar{y})\, v > 0, \tag{1}$$

$$\nabla G(\bar{y})\, v = 0, \tag{2}$$

$$\nabla Q^A(\bar{y})\, v < 0, \tag{3}$$

$$\nabla \bar{Q}^{\bar{A}}(\bar{z})\, w < 0, \tag{4}$$

$$\nabla_y H(\bar{y}, \bar{z}, \bar{u})\, v + \nabla_z H(\bar{y}, \bar{z}, \bar{u})\, w + \sum_{j=1}^{D} H(\bar{y}, \bar{z}, u_j)\,\hat{\theta}_j + H(\bar{y}, \bar{z}, u)\,\hat{\theta}_{D+1} = 0, \tag{5}$$

$$\sum_{j=1}^{D+1} \hat{\theta}_j < 1, \qquad \hat{\theta}_j > 0 \qquad (j = 1, \ldots, D+1). \tag{6}$$

Observe that we have used $H(\bar{y}, \bar{z}, \bar{u}) = 0$ here.

Note that by scaling, (1)-(6) has no solution if and only if (1)-(5) and

$$\hat{\theta}_j > 0 \qquad (j = 1, \ldots, D+1) \tag{6'}$$

has no solution. Also note that we may write (5) as

$$\nabla_y H(\bar{y}, \bar{z}, \bar{u})\, v + \nabla_z H(\bar{y}, \bar{z}, \bar{u})\, w + \sum_{j=1}^{D} H(\bar{y}, \bar{z}, u_j)\,\hat{\theta}_j + \rho = 0, \tag{5'}$$

where

$$\rho = H(\bar{y}, \bar{z}, u)\,\hat{\theta}_{D+1}.$$

Consequently (1)-(4), (5') and (6') have no solution for any

$$\rho \in S = \{ H(\bar{y}, \bar{z}, u)\,\hat{\theta}_{D+1} \text{ for some } u \in U,\ \hat{\theta}_{D+1} > 0 \}.$$

Observe that S is a convex set, since if ρ_1, $\rho_2 \in S$ then

$$\rho_1 = H(\bar{y}, \bar{z}, u)\,\hat{\theta}_{D+1}$$

and

$$\rho_2 = H(\bar{y}, \bar{z}, u')\,\hat{\theta}'_{D+1} \quad \text{for some } u,\ u' \in U,\ \hat{\theta}_{D+1} > 0,\ \hat{\theta}'_{D+1} > 0.$$

But then for any $\lambda \in [0, 1]$,

$$\begin{aligned}\lambda\rho_1 + (1-\lambda)\,\rho_2 = {} & [\lambda\hat{\theta}_{D+1} + (1-\lambda)\,\hat{\theta}'_{D+1}] \\ & \times \left\{ \frac{\lambda\,\hat{\theta}_{D+1}}{\lambda\,\hat{\theta}_{D+1} + (1-\lambda)\,\hat{\theta}'_{D+1}}\, H(\bar{y}, \bar{z}, u) \right. \\ & \left. + \frac{(1-\lambda)\,\hat{\theta}'_{D+1}}{\lambda\hat{\theta}_{D+1} + (1-\lambda)\,\hat{\theta}'_{D+1}}\, H(\bar{y}, \bar{z}, u') \right\}.\end{aligned}$$

By $\overline{\text{H2}}$, the bracketed term equals $H(\bar{y}, \bar{z}, \tilde{u})$ for some $\tilde{u} \in U$; thus $\lambda \rho_1 + (1 - \lambda) \rho_2 \in S$.

Consequently we may apply Remark 1 to system (1)-(4), (5') and (6') over S, i. e., there is a scalar $\beta_0 \geqq 0$ and vectors $\beta \in R^m$, $\sigma \in R^r$, $\mu^A \in R^{s_1}$, $\bar{\mu}^{A} \in R^{\bar{s}_2}$ with $\mu^A \leqq 0$, $\bar{\mu}^{\bar{A}} \leqq 0$ and $(\beta_0, \beta, \sigma, \mu^A, \bar{\mu}^{\bar{A}}) \neq 0$ satisfying

$$\beta_0 \nabla F(\bar{y}) + \beta \nabla G(\bar{y}) + \mu^A \nabla Q^A(\bar{y}) + \sigma \nabla_y H(\bar{y}, \bar{z}, \bar{u}) = 0,$$

$$\sigma \nabla_z H(\bar{y}, \bar{z}, \bar{u}) + \bar{\mu}^{\bar{A}} \nabla \bar{Q}^{\bar{A}}(\bar{z}) = 0,$$

$$\sigma H(\bar{y}, \bar{z}, u_j) \hat{\theta}_j \leqq 0 \quad \text{for all } \hat{\theta}_j > 0 \qquad (j = 1, \ldots, D),$$

$$\sigma H(\bar{y}, \bar{z}, u) \hat{\theta}_{D+1} \leqq 0 \quad \text{for all } \hat{\theta}_{D+1} > 0 \text{ and } u \in U.$$

Note that $(\beta_0, \beta, \sigma) \neq 0$, since otherwise either $\mu^A \neq 0$ or $\mu^{\bar{A}} \neq 0$ and hypothesis $\overline{\text{H5}}$ is violated. Also, the last inequality in this system implies that $\sigma H(\bar{y}, \bar{z}, u) \leqq 0$ for all $u \in U$. Incorporating $H(\bar{y}, \bar{z}, \bar{u}) = 0$ in this last statement, defining $\delta = \sigma \nabla_y H(\bar{y}, \bar{z}, \bar{u})$ and letting $\mu^i = 0$ for $i \notin A$, $\bar{\mu}^i = 0$ for $i \notin \bar{A}$, conclusions (*a*)-(*d*) of the theorem are satisfied. ■

Remark 4 : If the vectors $\nabla G^j(\bar{y})$ ($j = 1, \ldots, m$) are not linearly independent, then Theorem 1 holds by taking $\beta \nabla G(\bar{y}) = 0$, $\beta \neq 0$ and $(\beta_0, \sigma, \mu, \bar{\mu}, \delta) = 0$. Thus $\overline{\text{H4}}$ will rule out this trivial case.

Corollary 1.1 (discrete maximum principle): Let $u_t = \bar{u}_t$, ($t = 0, \ldots, T-1$) and $x_t = \bar{x}_t$ ($t = 0, \ldots, T$) solve control problem (P); assume H1 through H5. Then the discrete maximum principle holds.

Proof: We previously showed that (P) is a special case of (P') and that $\overline{\text{H1}}$-$\overline{\text{H5}}$ hold when we make this association. Let σ, μ, μ and δ from Theorem 1 be given by:

$$\sigma = (-\alpha, p_1, \ldots, p_{T-1}),$$
$$\mu = \lambda_T,$$
$$\bar{\mu} = (\lambda_0, \ldots, \lambda_{T-1}),$$
$$\delta = -p_T.$$

In these terms, consequences (*a*)-(*d*) of Theorem 1 are translated into the following terms for (P):

(*a*) $$p_T = \beta_0 \nabla g_0(\bar{x}_T) + \beta g(\bar{x}_T) + \lambda_T q_T(\bar{x}_T);$$

(*b*) $$\begin{cases} -p_T = -p_T, \\ -\alpha \nabla h(\bar{x}_0) + p_1 [I + \nabla_x f_0(\bar{x}_0, \bar{u}_0)] + \lambda_0 q_0(\bar{x}_0) = 0, \\ -p_t + p_{t+1} [I + \nabla_x f_t(\bar{x}_t, \bar{u}_t)] + \lambda_t q_t(\bar{x}_t) = 0 \qquad (t = 1, \ldots, T-1); \end{cases}$$

$$(c) \qquad \lambda_T q_T(\bar{x}_T) = 0, \qquad \sum_{t=0}^{T-1} \lambda_t q_t(\bar{x}_t) = 0;$$

$$(d) \qquad \sum_{t=0}^{T-1} p_{t+1} f_t(\bar{x}_t, \bar{u}_t) \geqq \sum_{t=0}^{T-1} p_{t+1} f_t(\bar{x}_t, u_t)$$

for all

$$(u_0, \ldots, u_{T-1}) \in \Omega_0 \times \Omega_1 \times \ldots \times \Omega_{T-1}.$$

Defining $p_0 = \alpha \nabla h(\bar{x}_0)$, (*a*) and (*b*) become the adjoint equations and transversality conditions. Since $q_t(\bar{x}_t) \leqq 0$ and $\lambda_t \leqq 0$, (*c*) implies the complementary conditions $\lambda_t q_t(\bar{x}_t) = 0$ $(t = 0, \ldots, T-1)$. Successively taking $u_j = \bar{u}_j$ $(j \neq t)$ for $t = 0, 1, \ldots, T-1$, (*d*) implies Hamiltonian maximization for each $t = 0, \ldots, T-1$. Finally, $p_0, p_1, \ldots, p_T, \alpha$, β and β_0 are not all zero from Theorem 1 and the above associations.

Remark 5: (i) As in Remark 4, H4 rules out the trivial case where $\beta \nabla g(\bar{y}) = 0$ and β is the only non-zero multiplier or adjoint variable. Similarly H4 rules out the analogous case where $\alpha \nabla h(\bar{y}) = 0$, $\alpha \neq 0$.

(ii) In the case of no inequality constraints on state variables, non-singularity of $[I + \nabla_x f_t(\bar{x}_t, \bar{u}_t)]$ $(t = 0, \ldots, T-1)$ implies that $p_t \neq 0$ $(t = 0, 1, \ldots, T-1)$. If in addition the vectors $\nabla g_0(x_T)$, $\nabla g^1(\bar{x}_T)$, $\ldots$, $\nabla g^m(\bar{x}_T)$ are linearly independent then $p_T \neq 0$.

III. AN EXTENSION

Holtzman [4] extended the discrete maximum principle by allowing g_0 to be a function of u as well as x_t and introducing the concept of directional convexity. We show here that an analogous assumption permits our proof of the generalized maximum principle to be extended easily. When only F and H below are functions of u, assumption $\overline{\text{H2}}'$ is equivalent to Holtzman's definition of directional convexity (*see* [1]).

Suppose that $y = \bar{y}$, $z = \bar{z}$, $u = \bar{u}$ solves the control problem:

$$\max F(y, u)$$

subject to

$$\left.\begin{aligned} G(y, u) &= 0, \\ Q(y, u) &\leqq 0, \\ \bar{Q}(z, u) &\leqq 0, \\ H(y, z, u) &= 0, \\ u &\in U. \end{aligned}\right\} \qquad (\text{P}'')$$

Let $\overline{\text{H1}}'$ be the same as $\overline{\text{H1}}$ and modify $\overline{\text{H2}}$-$\overline{\text{H5}}$ to:

$\overline{\text{H2}}'$) for every $y \in R^N$, $z \in R^M$, $u \in U, u' \in U, \lambda \in [0, 1]$ there is a $\tilde{u} \in U$ such that

$$\begin{aligned} F(y, \tilde{u}) &\geqq \lambda F(y, u) \quad +(1-\lambda) F(y, u'), \\ Q(y, \tilde{u}) &\leqq \lambda\, Q(y, u) \quad +(1-\lambda)\, Q(y, u'), \\ \overline{Q}(z, \tilde{u}) &\leqq \lambda\, \overline{Q}(z, u) \quad +(1-\lambda) \overline{Q}(z, u'), \\ G(y, \tilde{u}) &= \lambda\, G(y, u) \quad +(1-\lambda)\, G(y, u'), \\ H(y, z, \tilde{u}) &= \lambda H(y, z, u)+(1-\lambda) H(y, z, u'). \end{aligned}$$

$\overline{\text{H3}}'$) For every $u \in U$, $G(y, u)$ is continuously differentiable in a neighborhood of $y = \bar{y}$, $F(y, u)$ and $Q(y, u)$ are differentiable at $y = \bar{y}$, and $\overline{Q}(z, u)$ is differentiable at $z = \bar{z}$.

$\overline{\text{H4}}'$) The vectors $\nabla G^j(\bar{y}, \bar{u})$ $(j = 1, \ldots, m)$ are linearly independent.

$\overline{\text{H5}}'$) $\overline{\text{H5}}$ with $Q^i(\bar{y})$ and $\overline{Q}^i(\bar{z})$ replaced by $Q^i(\bar{y}, \bar{u})$ and $\overline{Q}^i(\bar{z}, \bar{u})$ in the definitions of A and $\overline{A}$.

With these modifications Theorem 1 becomes:

Theorem 1 A: Let $y = \bar{y}$, $x = \bar{x}$, $u = \bar{u}$ solve control problem (P″); assume $\overline{\text{H1}}'$-$\overline{\text{H3}}'$ and $\overline{\text{H5}}'$. Then there exist a real number $\beta_0 \geqq 0$ and vectors $\beta \in R^m$, $\sigma \in R^r$, $\mu \in R^{s_1}$, $\bar{\mu} \in R^{s_2}$, $\sigma \in R^r$ with $\mu \leqq 0$, $\bar{\mu} \leqq 0$, $(\beta_0, \beta, \sigma) \neq 0$ satisfying

$$(a') \qquad -\delta = \beta_0 \nabla_y F(\bar{y}, \bar{u}) + \beta \nabla_y G(\bar{y}, \bar{u}) + \mu \nabla_y Q(\bar{y}, \bar{u});$$

$$(b') \qquad \begin{cases} \delta = \sigma \nabla_y H(\bar{y}, \bar{z}, \bar{u}), \\ 0 = \sigma \nabla_z H(\bar{y}, \bar{z}, \bar{u}) + \bar{\mu} \nabla_z Q(\bar{z}, \bar{u}); \end{cases}$$

$$(c') \qquad \mu\, Q(\bar{y}, \bar{u}) = 0, \qquad \bar{\mu}\, \overline{Q}(\bar{y}, \bar{u}) = 0;$$

$$(d') \qquad [\mathscr{H}(\bar{u}) - \mathscr{H}(u)] \geqq 0 \quad \text{for all } u \in U$$

where

$$\mathscr{H}(u) \equiv \beta_0 F(\bar{y}, u) + \beta\, G(\bar{y}, u) + \mu\, Q(\bar{y}, u) + \bar{\mu}\, \overline{Q}(\bar{z}, u) + \sigma H(\bar{y}, \bar{z}, u).$$

Proof: The proof is analogous to that of Theorem 1 and will be omitted. We simply note that $\psi(u)$ is now defined as

$$\psi(u) = \begin{pmatrix} \nabla_y G(\bar{y}, u) & & G(\bar{y}, u) \\ \nabla_y H(\bar{y}, \bar{z}, \bar{u}) & \nabla_z H(\bar{y}, \bar{z}, \bar{u}) & H(\bar{y}, \bar{z}, u) \end{pmatrix}.$$

Also, in problem $(\overline{\mathrm{P}})$, functions F, G, Q^A and $\overline{Q}^{\overline{A}}$ are each replaced by analogs of the form for H in that problem. For example, $F(y)$ is replaced by

$$\left(1-\sum_{j=1}^{D+1}\theta_j\right)F(y, u)+\sum_{j=1}^{D}\theta_j F(y, u_j)+\theta_{D+1}F(y, u).$$

Remark 6: H4′ admits a statement analogous to Remark 4. In addition, the statement of Theorem 1 A can be specialized to an extended version of problem (P), giving an extension of the discrete maximum principle. Details are omitted.

IV. A SEPARATION PROPERTY

Let us consider the following linearized version of (P′) about $(\bar{y}, \bar{z}, \bar{u})$:

$$\max \nabla F(\bar{y})v$$

subject to

$$\nabla G(\bar{y})v = 0 \qquad [v = (y-\bar{y}),\ w = (x-\bar{x})],$$

$$\nabla Q^A(\bar{y})v < 0,$$

$$\nabla \overline{Q}^{\overline{A}}(\bar{z})w < 0,$$

$$\nabla_y H(\bar{y}, \bar{z}, \bar{u})v+\nabla_z H(\bar{y}, \bar{z}, \bar{u})w+H(\bar{y}, \bar{z}, u) = 0, \qquad u\in U.$$

This is a standard first order strict inequality approximation except for the linearized H equation. Here we have omitted $G(\bar{y}) = 0$, $Q^A(\bar{y}) = 0$, $\overline{Q}^{\overline{A}}(\bar{z}) = 0$, $H(\bar{y}, \bar{z}, \bar{u}) = 0$ and the constant term $F(\bar{y})$.

Let

$$S_1 = \{v\in R^n : \nabla F(\bar{y})v > 0,\ \nabla G(\bar{y})v = 0,\ \nabla Q^A(\bar{y})v < 0\}$$

and

$$S_2 = \{v\in R^N : \nabla \overline{Q}^{\overline{A}}(\bar{z})w < 0,$$
$$\nabla_y H(\bar{y}, \bar{z}, \bar{u})v+\nabla_z H(\bar{y}, \bar{z}, \bar{u})w+H(\bar{y}, \bar{z}, u) = 0$$
$$\text{for some } u\in U,\ w\in R^M\}.$$

S_1 and S_2 each reflect half of the above linearized problem.

Halkin's [3] approach to the discrete maximum principle was to first prove that there is a hyperplane separating S_1 and S_2. He considered the case with no state constraints Q or $\overline{Q}$ and with H having the associations for the control problem (P) that were given in section I. Since this separation property may be of independent interest, let us see how it results from the arguments of Theorem 1.

Lemma 3 (second linearization lemma): Let $y = y$, $x = x$ and $u = \bar{u}$ solve control problem (P'); assume $\overline{\text{H1}}$-$\overline{\text{H5}}$ and that the matrix $[\nabla_y H(\bar{y}, \bar{z}, \bar{u}), \nabla_z H(\bar{y}, \bar{z}, \bar{u})]$ has full row rank. Then S_1 and S_2 are separated by a hyperplane.

Proof: For the most part, the proof is the same as the proof of Theorem 1. Consequently, we adopt the notation used there and only sketch modifications to that proof.

If $\psi(u)$ is not linearly independent over u, then let $(\beta, \sigma) \neq 0$ and δ be defined as before, that is

$$\begin{aligned} \delta \equiv -\beta \nabla G(\bar{y}) &= \sigma \nabla_y H(\bar{y}, \bar{z}, \bar{u}) \\ \sigma \nabla_z H(\bar{y}, \bar{z}, \bar{u}) &= 0 \\ \sigma H(\bar{y}, \bar{z}, u) &= 0 \quad \text{for all } u \in U. \end{aligned}$$

Then consider the hyperplane $\{ x : \delta x = 0 \}$. Note that $\delta \neq 0$ since the rows of $\nabla G(\bar{y})$ are linearly independent as are those of $[\nabla_y H(\bar{y}, \bar{z}, \bar{u}), \nabla_z H(\bar{y}, \bar{z}, \bar{u})]$. For $v \in S_1$, $\delta v = -\beta \nabla G(\bar{y})\, v = 0$. For $v \in S_2$, there are $w \in R^m$ and $u \in U$ so that

$$\delta v = \sigma \nabla_y H(\bar{y}, \bar{z}, \bar{u}) v = \sigma \left[\nabla_y H(\bar{y}, \bar{z}, \bar{u}) v + \nabla_z H(\bar{y}, \bar{z}, \bar{u}) w + H(\bar{y}, \bar{z}, u)\right] = 0.$$

Thus $\{ x : \delta x = 0 \}$ separates S_1 and S_2.

If $\psi(u)$ is linearly independent over u, then let

$$\begin{aligned} S_2^* = \Big\{ v \in R^N :\ & \nabla \overline{Q^A}(z) w < 0, \\ & \nabla_y H(\bar{y}, \bar{z}, \bar{u}) v + \nabla_z H(\bar{y}, \bar{z}, \bar{u}) w \\ & \quad + \sum_{j=1}^{D} H(\bar{y}, \bar{z}, u_j) \hat{\theta}_j + H(\bar{y}, \bar{z}, u) \hat{\theta}_{D+1} = 0, \\ & \sum_{j=1}^{D+1} \hat{\theta}_j < 1,\ \hat{\theta}_j > 0,\ (j = 1, \ldots, D+1) \text{ for some } u \in U,\ w \in R^M \Big\}, \end{aligned}$$

where $u_1, \ldots, u_D$ are defined as in the proof of Theorem 1. By the argument used previously to show that the set S is convex, we can easily establish that S_2^* is convex. Since S_1 is convex and equations (1)-(6) in Theorem 1 have no solution for $u \in U$, S_1 and S_2^* are disjoint hence separated by a hyperplane. But $S_2 \subseteq$ closure S_2^* (let $\hat{\theta}_{D+1} \to 1$) so that S_1 and S_2 also are separated.

REFERENCES

[1] M. D. CANON, D. C. CULLUM and E. POLAK, *Theory of optimal control and mathematical programming*, McGraw-Hill, New York, 1970.

[2] K. FAN, I. GLICKSBURG and A. J. HOFFMAN, *Systems of inequalities involving convex functions*, A.M.S. Proc., 8, 1957, pp. 617-622.

[3] H. HALKIN, *A maximum principle of the Pontryagin type for systems described by non-linear difference equations*, S.I.A.M. J. Control, 4, 1966, pp. 90-111.

[4] J. M. HOLTZMAN, *On the maximum principle for non-linear discrete-time systems*, I.E.E.E. Trans. Automatic Control, 4, 1966, pp. 528-547.

[5] B. W. JORDAN and E. POLAK, *Theory of a class of discrete optimal control systems*, J. Electronics Control, 17, 1964, pp. 697-713.

[6] T. L. MAGNANTI, *A linear approximation approach to duality in non-linear programming*, Tech. Rep. OR 016-73, Oper. Res. Center, M.I.T., April 1973.

[7] O. L. MANGASARIAN, *Non-linear programming*, McGraw-Hill, New York, 1969.

[8] O. L. MANGASARIAN and S. FROMOVITZ, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints*, J. Math. Analysis and Applic., 17, 1967, pp. 37-47.

[9] J. M. ORTEGA and W. C. RHEINBOLDT, *Iterative solution of non-linear equations in several variables*, Academic Press, New York, 1970.

[10] A. I. PROPOI, *The maximum principle for discrete control systems*, Avtomatica i Telemachanica, 7, 1965, pp. 1177-1187.

[11] J. B. ROSEN, *Optimal control and convex programming*, I.B.M. Symp. Control Theory Applic., Yorktown Heights, New York, October 1964, pp. 223-237.

[12] R. M. VAN SLYKE and R. J. B. WETS, *A duality theory for abstract mathematical programs with applications to optimal control theory*, Math. Note No. 538, Math. Res. Lab., Boeing Scientific Research Laboratories, October 1967.