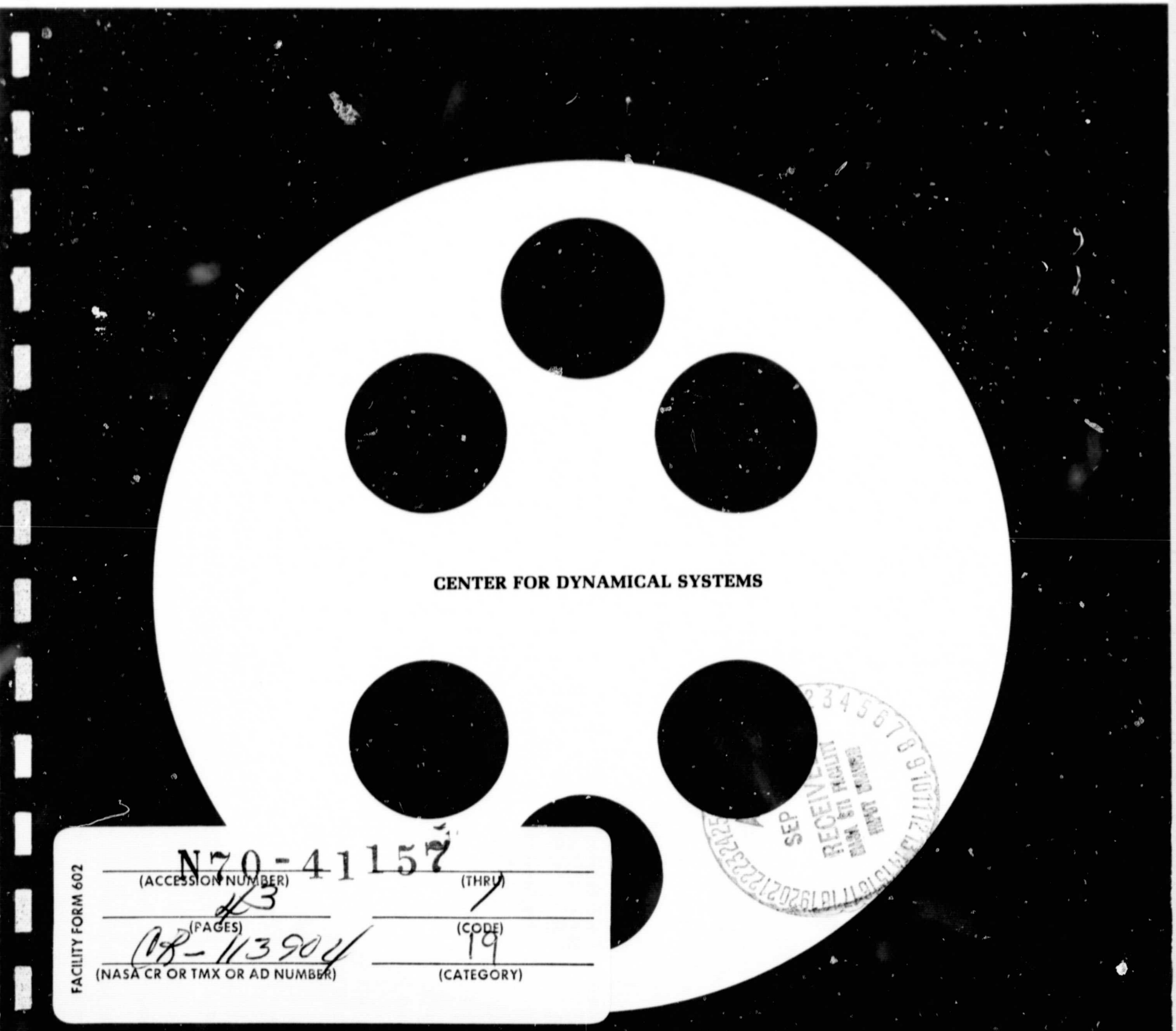
TECHNICAL REPORT 70-2

T70-01672

H. J. KUSHNER
AND
A. J. KLEINMAN

MATHEMATICAL PROGRAMMING AND THE CONTROL OF MARKOV CHAINS

MAY, 1970



CENTER FOR DYNAMICAL SYSTEMS

DIVISION OF APPLIED MATHEMATICS

BROWN UNIVERSITY

Mathematical Programming and the Control of Markov Chains

by

H. J. Kushner*

and

A. J. Kleinman**

*This research was supported in part by the National Science Foundation under Grant No. GK 2788, in part by the National Aeronautics and Space Administration under Grant No. NGL 40-002-015 and in part by the Air Force Office of Scientific Research under Grant No. AF-AFOSR 67-0693A.

**This research was supported by the National Science Foundation under Grant No. GK 2788.

Abstract

Linear programming versions of some control problems on Markov chains are derived, and are studied under conditions which occur in typical problems which arise by discretizing continuous time and state systems, or in discrete state systems. Control interpretations of the dual variables and simplex multipliers are given. The formulation allows the treatment of 'state space' like constraints which cannot be handled conveniently with dynamic programming. The relation between dynamic programming on Markov chains, and the deterministic discrete maximum principle is explored, and some insight is obtained into the problem of singular stochastic controls (with respect to a stochastic maximum principle).

1. Introduction

This paper is concerned with several problems occuring in the control of a Markov chain $\{X_n\}$ on the state space $(0,1,\ldots,N) = S$, with transition probabilities $p_{ij}(\alpha)$, where α, a control, takes values in a set U_i. State 0 is a desired target state and $p_{00}(\alpha) \equiv 1$; once in state 0, always in state 0. The terms $u = (u_1,\ldots,u_N)$, $u_i \in U_i$, denotes a control vector. I.e., if the control vector u is always used, and $X_n = i$, then the value of α in $p_{ij}(\alpha)$ is $u(X_n) = u_i$. Let τ denote the first time state 0 is attained, $k(i,\alpha)$ the cost paid when the state is i and control $u(X_n) = u_i = \alpha$ is used, and E_i^u the expectation operator given that $X_0 = i$, and the control vector u is used. Then the cost is

$$V(u;i) = E_i^u \sum_0^{\tau-1} k(X_n,u(X_n)).$$

Define $k(0,\alpha) \equiv 0$. Then

$$V(u;i) = E_i^u \sum_0^{\infty} k(X_n,u(X_n)). \tag{1}$$

Define the column vectors $V(u) = (V(u;1),\ldots,V(u;N))$ and $K(u) = (k(1,u_1),\ldots,k(N,u_N))$.

Note that, if the N step transition probability $p_{i0}^{(N)}(u) > 0$ for all i, then state 0 is attainable and $V(u)$ exists.

Define problem (P1): Let U_i contain a finite number of points (which, for convenience, we assume are $\alpha_1,\ldots,\alpha_q$), or let the $n+1$ dimensional

set $(p_{i1}(U_i),\ldots,p_{iN}(U_i),k(i,U_i))$ be a convex polyhedron with extreme points included in $\{(p_{i1}(\alpha_r),\ldots,p_{iN}(\alpha_r),k(i,\alpha_r)),\ r = 1,\ldots,q\}$. Assume (A1): $p_{iO}^{(N)}(u) > 0$ for all i and u, or (A2): $k(i,\alpha) > 0$ for all i,α and $p_{iO}^{(N)}(u) > 0$ for all i and some u. Find the control $u = (u_1,\ldots,u_N)$ which minimizes $V(u;i)$, $i = 1,\ldots,N$. Define $V_i = \min_u V(u;i)$.

The assumption on U_i can be weakened, although the form given allows a relatively simple notation. Indeed any compact U_i is suitable if the $p_{ij}(\cdot)$ and $k(\cdot)$ are continuous. The convex polyhedron assumption is satisfied for problems which are obtained by discretizing continuous time bang-bang problems. See the example.

In Section 2, a linear programming formulation of (P1) will be given. Linear programming (L.P.) versions of many types of dynamic programming problems are well known (see, e.g., [3] - [5], [9]). Indeed, a L. P. version of (P1) was given by Derman [6]. The variables in the L.P. form in [6] do not seem to have a simple physical interpretation. However, the form here seems more natural and has a more natural dual, namely the dynamic programming equations for (P1)

$$V_i \geqq \sum_{j=1}^{N} p_{ij}(\alpha_r)V_j + k(i,\alpha_r), \quad \text{all } i,r.$$

While experience indicates that the linear programming algorithm (Simplex method) is generally inferior, in computational efficiency, to the available dynamic programming iterative methods (for the type of problems discussed here), it is of interest since it is an alternative formulation which sheds further light on the Markov optimization problem and, in addition, the two important reasons:

(a) There may be additional constraints on the probabilities $P\{X_n = i\}$ (Section 2). The dynamic programming is not directly applicable, and the L.P. formulation yields useful insights into the optimization problem. Indeed, it is often desirable or necessary to add such constraints in Markov control problems. See Section 2 for example.

(b) The L.P. formulation gives us insight into a form of a stochastic maximum principle (Section 3), and the singularity problem of the stochastic maximum principle .

In Section 3, which treats a finite time Markov optimization problem, it is shown that the Holtzman form of the discrete maximum principle [7] is equivalent to dynamic programming, in the absense of 'state space' constraints on the variables $P\{X_n = i\}$, and that the control is often singular (in the sense that minimization of the relevant Hamiltonian yields no information on the form of the control) in the presence of such constraints, a situation which often occurs with deterministic systems with state space constraints.

2. Linear Programming and the Optimal Control Problem.

2.1. No 'state space' constraints. First a form of (P1) will be treated. Let $R(u) = \{p_{ij}(u_i),\ i,j = 1,\ldots,N\}$ denote the reduced transition matrix (state 0 omitted) corresponding to control vector $u = (u_1,\ldots,u_N)$. The following known results [2] will be used below.

Lemma 1. Assume (A1). Then state 0 is attained w.p.1. and $V(u)$ is the unique vector solution to the vector equation

$$C = R(u)C + K(u). \tag{2}$$

If $k(i,\alpha) > 0$ and (2) has a finite solution, then $p_{i0}^{(N)}(u) > 0$ and $p_{i0}^{(n)}(u) \to 1$ as $n \to \infty$ and $C = V(u)$. Under (A2), there is at least one such u.

Under (A1) or (A2) there is an optimal control, and the least cost vector V satisfies

$$V = \min_u [R(u)V + K(u)]. \tag{3}$$

Remark. The property $p_{i0}^{N}(u) > 0$ for all i assures that state 0 is ultimately attained with a corresponding finite average cost.

Lemma 2. (Howard's iteration in policy space procedure). Assume (A1) or (A2). Choose u^0 so that $V(u^0)$ exists. Assume u^n is given and $V(u^n)$ exists. Choose u^{n+1} as the minimizing vector u in

$$\min_u [R(u)V(u^n) + K(u)] \equiv R(u^{n+1})V(u^n) + K(u^{n+1}) \tag{4}$$

then $V(u^n) \downarrow V \equiv \min_u V(u)$.

Remark. The method in Lemma 2 is mentioned because of its relation to the simplex method (see below). For many problems, it seems to converge slower than the various backward iteration methods, e.g.

$$C^{n+1} = \min_u [R(u)C^n + K(u)].$$

See [1] for a discussion of a better iterative method.

2.1.1. Introduction of Randomized Controls. For purposes of the L.P.

formulation and its generalizations, its useful to rewrite (P1) in an equivalent form. We suppose that $U_i = (\alpha_1,\ldots,\alpha_q)$ and allow randomized controls. That is to say that, at each time, the actual control action which is used is randomly selected among the $\alpha_1,\ldots,\alpha_r$; the probability which governs the choice (or, equivalently, the control law) depends on the current state. Thus, the control u is replaced by a sequence γ of N_q elements.

$$\gamma = (\gamma_1,\ldots,\gamma_N), \quad \gamma_i \text{ is a } q \text{ vector}$$

$$\gamma_i = (\gamma_{i1},\ldots,\gamma_{iq}), \quad i = 1,\ldots,N$$

$$\sum_{j=1}^{q} \gamma_{ij} = 1, \quad \gamma_{ij} \geq 0, \quad \gamma_{ij} = P\{u(X_n) = j \mid X_n = i\}$$

If $\gamma_{ij} = 1$ then the control at state i is pure and $u(X_n) = \alpha_j$, when $X_n = i$. Under the control law γ, the transition probabilities take values

$$P\{X_1 = j \mid X_0 = i, \text{ law } \gamma \text{ used}\} = P_i^{\gamma}\{X_1 = j\} = p_{ij}(\gamma_i) = \sum_r p_{ij}(\alpha_r)\gamma_{ir}.$$

We now write $V(\gamma)$ and E_i^{γ} and P_i^{γ} instead of $V(u)$, E_i^u, P_i^u. It turns out, of course, that the L.P. formulation does give a non-random control. With this randomization, finiteness of $U_i = U$ is equivalent to the sets $S_i = (p_{i1}(U_i),\ldots,p_{iN}(U_i),\kappa(i,U_i))$ being convex polyhedrons.

Let M_{ij} denote the average number of times that α_j is actually used when state i is visited. Write $M_i = \sum_{j=1}^{q} M_{ij}$, and suppose that X_0 is random with $P\{X_0 = r\} = \mu_r$, where $\mu = (\mu_1,\ldots,\mu_N)$ is a column vector. Write

$$P_{\mu}^{\gamma}\{X_n = j\} = \sum_i p_{ij}^{(n)}(\gamma)\mu_i = P\{X_n = j \mid \text{control } \gamma \text{ used, initial distribution} = \mu\}.$$

2.1.2. The Constraints for L.P. By definition,

$$M_i = \sum_{n,r}^{\infty} P_r^{\gamma}\{X_n = i\}\mu_r \equiv \sum_{n=0}^{\infty} P_\mu^{\gamma}\{X_n = i\}$$

$$= \mu_i + \sum_{n=0}^{\infty} P_\mu^{\gamma}\{X_{n+1} = i\},$$

$$M_{ij} = \sum_{r,n} P_r^{\gamma}\{X_n = i,\ u(X_n) = \alpha_j\}\mu_r$$

$$= \sum_{n=0}^{\infty} P_\mu^{\gamma}\{X_n = i,\ u(X_n) = \alpha_j\}.$$

From the relation

$$P_\mu^{\gamma}\{X_{n+1} = i\} = \sum_{k,j} p_{ji}(\alpha_k) P_\mu^{\gamma}\{X_n = j,\ u(X_n) = \alpha_k\},$$

e obtain

$$\sum_k M_{ik} = M_i = \mu_i + \sum_{k,j} p_{ji}(\alpha_k) M_{jk}, \quad M_{ij} \geq 0, \quad i = 1,\ldots,N \tag{5}$$

or, equivalently,

$$\mu_i = \sum_{k,j} [-p_{ji}(\alpha_k) + \delta_{ij}] M_{jk}. \tag{6}$$

Define the transition matrix (again state 0 deleted) $R(\gamma) = \{p_{ji}(\gamma_j);\ j,i = 1,\ldots,N\}$. Now $\gamma_{jk} = M_{jk}/M_j$, and an alternate form to (5) is

$$(7) \qquad M_i = \mu_i + \sum_j \Big[\sum_k p_{ji}(\alpha_k)\gamma_{jk} \Big] M_j$$
$$= \mu_i + \sum_j \Gamma_{ji}(\gamma_j) M_j, \quad i = 1,\ldots,N.$$

In vector notation (where M is the column vector $(M_1,\ldots,M_N)$, and prime $'$ is transpose)

$$(8) \qquad M = \mu + R'(\gamma)M, \quad M_{ij} \geq 0.$$

We now address ourselves to the uniqueness of the solution of (7), (8). Unless an obtained solution of (8) is truly the vector of average occupancy times, the L.P. formulation may not give the correct solution. The matrix $R(\gamma)$ is said to be a contraction if its eigenvalues lie strictly inside the unit circle. This is equivalent to ([1]) the property $\sum_{j=1}^{N} p_{ij}^{(N)}(u) < 1$ for all i, which, in turn, is equivalent to $R^N(\gamma)$ being a contraction in the sense that $\max_i |C_i| < \max_i |D_i|$ in $C = R^N(\gamma)D$. These properties are equivalent to $R^n(u) \to 0$ as $n \to \infty$.

Lemma 2. Suppose $R(\gamma)$ is given. Assume either (i); (A1), or (ii); $\mu_i > 0$ for all i. Define the cost

$$(9) \qquad z = \sum_{i,j} M_{ij} k(i,\alpha_j) = \sum_i M_i k(i,\gamma_i)$$
$$k(i,\gamma_i) \equiv \sum_j \gamma_{ij} k(i,\alpha_j)$$

Then there is a unique non-negative solution (10) to (8),

(10) $$M = \sum_{n=0}^{\infty} (R'(\gamma))^n \mu, \quad R^0(\gamma) \equiv I,$$

and this solution is the vector of mean occupancy times. Furthermore (9) can be written as (11).

(11) $$z = \mu' \sum_{n=0}^{\infty} R^n(\gamma) K(\gamma), \; = \sum_i \mu_i V(\gamma; i)$$

where $V(\gamma; i)$ = cost for (P1) corresponding to randomized control γ and $K(\gamma)$ is the column vector

$$K(\gamma) = (k(1, \gamma_1), \ldots, k(N, \gamma_N)).$$

If $\mu_i > C$ for all i, then any control law $(\gamma_1, \ldots, \gamma_N)$, or equivalently, any $\{M_{ij}\}$ which minimizes (9) subject to (8), also solves (P1), and conversely. In particular $\min z = \sum_i \mu_i V_i$. The converse statement is true even if some of the $\mu_i = 0$.

Proof: Only the uniqueness of the solution to (8) will be shown, for the rest follows easily from this. Any solution of (8) is of the form

(12) $$M = \lim_n (R'(\gamma))^n M + \sum_{n=0}^{\infty} (R'(\gamma))^n \mu.$$

Thus, we need only show that $R(\gamma)$ or $R^N(\gamma)$ are contractions in the appropriate senses.

Assume (i). The eigenvalues of all $R(u)$ are interior to the unit circle for all pure controls u, and there are only a finite number of possibilities for u. Any $R(\gamma)$ has the form

$$R(\gamma) = \sum_i \lambda_i R(u^i), \quad \lambda_i \geq 0, \quad \sum_i \lambda_i = 1,$$

where u^i ranges over all possible pure control vectors with values in $U_1 \times \ldots \times U_N$. But, since $R(\gamma)$ is a non-negative matrix, the eigenvalue $e(R(\lambda))$ with largest absolute value is real and positive and

$$e(R(\lambda)) \leq \sum_i \lambda_i e(R(u^i)) < 1,$$

thus proving uniqueness under (i).

Assume (ii). If $\mu_i > 0$ for all i, and $R^n(\gamma)$ does not tend to the zero matrix, then (12) implies that some M_i is infinite, a contradiction. Q.E.D.

Remark. Lemma 2 can be strengthened under (A2). First we make the following observation. Let $\mu_i > 0$, $i = 1,\ldots,r$ with all other $\mu_i = 0$. Let $S_1(\gamma)$ denote the states $1,\ldots,r$, all those states connected to $1,\ldots,r$ and all transient states. Let $S_2(\gamma)$ denote the remaining states (a positive recurrent class). A modification of the proof under (A2) yields that $\sum_{n=0}^{\infty} p_{ij}^{(n)}(\gamma) < \infty$ for $i \in S_1(\gamma)$ and all j. Hence for $i \in S_1(\gamma)$, $p_{ij}^{(n)}(\gamma) \to 0$ and the form (12) implies that the component M_i of the solu-

tion to (8) is the mean occupancy time. For $i \in S_2$, (12) indicates that the component M_i of the solution to (8) can be larger than the mean occupancy time. This turns out to be unimportant under (A2).

We also note that, if $\mu_i > 0$ for all i, and $z < \infty$, and $k(i,\alpha_j) \geq \epsilon > 0$, then $R^N(\gamma)$ must be a contraction, for otherwise we would have $z = \infty$.

Lemma 3. Assume (A2). Let γ be optimal. Then the M_i solving (8) are the mean occupancy times, $M_i = 0$ for $i \in S_2(\gamma)$, and $p_{ij}(\gamma) = 0$ for $i \in S_1(\gamma)$, $j \in S_2(\gamma)$. Also $p_{ij}^{(n)}(\gamma) \to 0$ as $n \to \infty$, for $i \in S_1(\gamma)$. Thus (10) and (11) hold.

Proof: All states in $S_1(\gamma)$ are transient, and non-transient states (i.e., those in $S_2(\gamma)$) cannot be reached from states in $S_1(\gamma)$, for otherwise the representation

$$z = (\lim_n M' R^n(\gamma) K(\gamma) + \mu' \sum_0^\infty R^n(\gamma) K(\gamma),$$

and the positivity of $k(i,\gamma_i)$, imply that $z = +\infty$. Let $i \in S_2(\gamma)$. Then M_i is not effected by the values of M_j, $j \in S_1(\gamma)$, since $p_{ji}(\gamma) = 0$. Since $k(i,\gamma_i) > 0$ and γ is optimal, the form (9) implies that $M_i = 0$. Thus $M' R^n(\gamma) \to 0$ as $n \to \infty$, proving (10) and (11). Q.E.D.

Remark on the equality constraint (6). If $k(i,\alpha_j) \geq 0$, the equality constraint (6) (or (8)) can be replaced by

$$(6') \qquad M_i - \sum_{j,\ell} p_{ji}(\alpha_\ell) M_{j\ell} \geq \mu_i, \quad M_{ij} \geq 0$$

or

$$(6'') \qquad M_i \geq \mu_i + \sum_{j,\ell} p_{ji}(\alpha_\ell) M_{j\ell}.$$

We will give the proof for all $\mu_i > 0$ and show only that the minimum of (9) under (6') is not less than the minimum of (9) under (6) - which, in turn, implies that the optimal solution will give an equality in (6').

First observe that (6') implies that $M_i \geq \mu_i > 0$. Let $\{M_{ij}\}$ solve (6'). Define, again, $\gamma_{ij} = M_{ij}/M_i$, and let $R(\gamma)$ be the corresponding transition matrix. (6'') can be written as

$$M \geq \mu + R'(\gamma)M,$$

which implies that $(R'(\gamma))^N$ is a contraction. Thus there is a unique non-negative solution to

$$\tilde{M} = \mu + R'(\gamma)\tilde{M},$$

and

$$M-\tilde{M} \geqq R'(\gamma)(M-\tilde{M})$$

which (since $(R'(\gamma)) \to 0$) implies that

$$M \geqq \tilde{M}.$$

Then the set $\{\tilde{M}_{ij} = \tilde{M}_i \gamma_{ji}\}$ satisfies

$$\tilde{M}_{ij} \leqq M_{ij}.$$

Since $k(i,\alpha_j) \geqq 0$,

$$\tilde{z} = \sum_{i,j} k(i,\alpha_j)\tilde{M}_{ij} \leqq \sum_{i,j} k(i,\alpha_j)M_{ij}.$$

2.1.3. The Dual form for L.P. Write the system (7) in the vector form

$$\mu = \mathscr{A}\mathscr{M}, \quad z = \mathscr{B}\mathscr{M}$$

where $\mathscr{M} = (M_{11},M_{12},\ldots,M_{1q},M_{21},\ldots,M_{Nq})$ is the column vector of L.P. variables and $\mathscr{A}$ and $\mathscr{B}$ are the $N \times Nq$ matrix and Nq row vector, (13a) and (13b), resp.

$$
\begin{array}{l}
-p_{11}(\alpha_1)+1,\ldots,-p_{11}(\alpha_q)+1, \quad -p_{21}(\alpha_1),\ldots,-p_{21}(\alpha_q),\ldots,-p_{N1}(\alpha_1)\ldots,-p_{N1}(\alpha_q) \\
-p_{12}(\alpha_1) \ ,\ldots,-p_{12}(\alpha_q) \ , \quad -p_{22}(\alpha_1)+1,\ldots,-p_{22}(\alpha_q)+1,\ldots,-p_{N2}(\alpha_1),\ldots,-p_{N2}(\alpha_q) \\
\quad . \\
\quad . \qquad\qquad\qquad . \ . \ . \qquad\qquad . \ . \ . \\
\quad . \\
-p_{1N}(\alpha_1),\ldots,-p_{1N}(\alpha_q), \quad . \ . \ . \qquad\qquad\qquad\qquad , \ -p_{NN}(\alpha_1)+1,\ldots,-p_{NN}(\alpha_q)+1.
\end{array}
\tag{13a}
$$

$$
k(1,\alpha_1),\ldots,k(1,\alpha_q),\ k(2,\alpha_1),\ldots,k(2,\alpha_q),\ldots,k(N,\alpha_1),\ldots,k(N,\alpha_q) \tag{13b}
$$

Let $C = (C_1,\ldots,C_N)$ be the column vector of dual variables. By the usual rules, [8, p. 127], the dual form of the L.P. is

$$
\text{maximize} \ \sum_i \mu_i C_i \tag{14}
$$

with constraint

$$
\mathscr{A}'C \leq \mathscr{B}, \tag{15}
$$

and the C_i are unconstrained in sign.

Writing out (15) in detail and rearranging some terms gives the Nq inequalities

$$
C_i \leq \sum_{j=1}^{N} p_{ij}(\alpha_r)C_j + k(i,\alpha_r) \tag{16}
$$

((i,r^{th}) inequality) $i = 1,\ldots,N;\ r = 1,\ldots,q.$

It will be shown below that, if all $\mu_i > 0$, then for any optimal solution, $M_{ij} \neq 0$ for at most one j (depending on i), and $M_i \geq \mu_i > 0$. Denote this j by $r(i)$ and let u_i denote $\alpha_{r(i)}$. Let C denote the optimal dual vector. By the complementary slackness theorem of L.P., there is equality in (16) for the $(i,r(i))^{th}$ lines. Thus

$$(17) \qquad C_i = \min_{\alpha}\Big[\sum_{j=1}^{N} p_{ij}(\alpha)C_j + k(i,\alpha)\Big] = \sum_{j=1}^{N} p_{ij}(u_i)C_j + k(i,u_i)$$

where α ranges over $U = (\alpha_1,\ldots,\alpha_q)$, which are precisely the dynamic programming equations (3). Thus, for the optimal dual variable

$$C_i = V_i = \min_{u} V(u;i).$$

The L.P. dual requires a maximization (14), but, any vector C which actually satisfies (16) is not a true cost vector (for some control u), unless it is the optimal cost vector.

If not all $\mu_i > 0$, but $k(i,\alpha_j) \geq \epsilon > 0$, some of the optimal M_i will equal zero (see Lemma 3 and the remark preceeding it). Let $M_i > 0$ and $M_{is} > 0$. Then, by the complementary slackness theorem, for all α_t,

$$(17') \qquad C_i = \sum_{j=1}^{N} p_{ij}(\alpha_s)C_j + k(i,\alpha_s) \leq \sum_{j=1}^{N} p_{ij}(\alpha_t)C_j + k(i,\alpha_t).$$

Furthermore, by taking suitable linear combinations in (17') (γ = optimal

control law)

$$(17'') \qquad C_i = \sum_{j=1}^{N} p_{ij}(\gamma_i)C_j + k(i,\gamma_i) \leqq \sum_{j=1}^{N} p_{ij}(\alpha_t)C_j + k(i,\alpha_t)$$

Since $p_{ij}(\gamma_i) = 0$ for $i \in S_1(\gamma)$, $j \in S_2(\gamma)$, and $S_1(\gamma)$ are transient, we conclude that $C_i = \min_{\gamma} V(\gamma;i) = V_i$, and that there is a non-random optimal control (let $S_1(\gamma) = 1,\ldots,s$, and $M_{ir(i)} > 0$ for $i = 1,\ldots,s$; then $(\alpha_{r(1)},\ldots,\alpha_{r(s)})$ is an optimal control, and $S_2(\gamma)$ is never reached).

2.1.4. The Simplex Method and Iteration in Policy Space.

Theorem 1. Under (A1) or (A2), there is an optimal non-random control. I.e., there is an admissible set $\{M_{ij}\}$ which minimizes z, and for which $M_{ij} > 0$ for at most one j for each i. If $\mu_i > 0$, the basic solution at each iteration of the simplex method satisfies $M_{ij} > 0$ for only one j for each i.

Proof: All assertions have already been proved, except the last. There are at most N of the $\{M_{ij}\}$ which are non-zero at each iteration. Then $M_i \geqq \mu_i > 0$. If $M_{ir} > 0$, $M_{is} > 0$ for $s \neq r$, then $M_j = 0$ for some j, which contradicts $M_j \geqq \mu_j > 0$. Thus $M_{ij} > 0$ for one and only one j, for each i, at each iteration of the simplex method. Q.E.D.

Simplex Multipliers. Assume either (A1) or (A2) and also that, under (A2), the simplex routine is initiated with a pure control u or a random control γ for which $R^N(u)$ or $R^N(\gamma)$ is a contraction. Let

$A = [a_1,\ldots,a_M]$ be an $N \times M$ matrix with $N < M$ and columns a_i. Consider the L.P. problem of minimizing $c'x = z$ with constraint $Ax = b$, where $c' = (c_1,\ldots,c_M)$. Let $x_{i_1},\ldots,x_{i_N}$ be the basic solution at a given iteration. Then([8]), there are numbers (simplex multipliers) $\pi_1,\ldots,\pi_N$ so that

$$\pi' a_{i_n} - c_{i_n} = 0, \quad n = 1,\ldots,N \tag{18}$$

$$\pi' = (\pi_1,\ldots,\pi_N) = \text{row vector.}$$

Define q_i:

$$\pi' a_{i_n} - c_{i_n} = \sum_{r=1}^{N} \pi_r a_{ri_n} - c_{i_n} \equiv q_i. \tag{19}$$

Let $q_i = \max_j q_j$. Then the simplex algorithm chooses x_i as the new entry into the basis. If all $q_i \leqq 0$, the current basis is optimal.

Let $\mu_i > 0$ for all i, and let $\{M_{ij(i)},\ i = 1,\ldots,N\}$ be the basis at a given iteration. Let $v_i = \alpha_{j(i)}$, and $v = (v_1,\ldots,v_N)$. For our L.P. problem, the multiplier results is: there is a vector $(\pi_1,\ldots,\pi_N) = \pi$ so that

$$\pi_i - \sum_{j=1}^{N} p_{ij}(v_i)\,\pi_j - k(i,v_i) = 0, \quad i = 1,\ldots,N. \tag{20}$$

The new basis entry M_{ir} is chosen as follows: choose the i,r for which

$$\pi_i - \sum_{j=1}^{N} p_{ij}(\alpha_r)\pi_j - k(i,\alpha_r) = q_{ir}$$

is largest. At the optimal (optimal control = $u = (u_1,\ldots,u_r)$)

$$\pi_i = \sum_{j=1}^{N} p_{ij}(u_i)\pi_j + k(i,j) \leq \sum_{j=1}^{N} p_{ij}(\alpha_r)\pi_j + k(i,\alpha_r) \tag{21}$$

for all i and α_r.

By (20) and Lemma 1, $\pi_i = V(v;i)$, the cost corresponding to initial state i. Eqn. (21) is merely the principle of optimality once again. The method of selecting the new basis variable is clearly a special case of iteration in policy space (Lemma (2)), where only one control is changed at a time. This was first observed by DeGhellinck [9] for the average cost per unit time problem. This observation suggests that the L.P. algorithm is no better than algorithms which are available for the original dynamic programming problem.

2.1.5. Elaboration of the Dual Form (14), (16).

Assume either (A1) or (A2) in this Section. If either

(i) $U = U_i = (\alpha_1,\ldots,\alpha_q)$

or

(ii) $S_i \equiv (p_{i1}(U_i),\ldots,p_{iN}(U_i),\ k(i,U_i))$ is a convex polyhedron for each i,

then (P1) has an L.P. form, with dual form (14), (16). It has already been noted that (i) and (ii) are equivalent. Instead of finiteness of the U_i, suppose temporarily that (A3): $p_{ij}(\cdot)$ and $k(i,\cdot)$ are continuous and U_i is compact. If, in addition

(iii) S_i is convex,

then the generalized programming method (G.P.) of Wolfe [8] can be used to solve (P1), and the dual of the G.P. is precisely (14), (16), where α_r ranges over the U_i.

Under (A3) alone, we can convexify the S_i by allowing randomizations, and thus apply G.P. However, it is interesting to see, by a more direct argument, that the solution to (14) - (16) is also the solution to (P1).

Theorem 2. Assume (A3) and either (A1) or (A2). Then there is a solution to (P1). The optimal cost vector V solves (14), (16')

$$(16') \qquad C_i \leqq \sum_j p_{ij}(v_i)C_j + k(i,v_i), \quad v_i \in U_i$$

$$C \leqq R(v)C + K(v), \quad v = (v_1,\ldots,v_N).$$

Proof: 1°. The first statement is known to be true [2]. By the principle of optimality, the optimal control $u = (u_1,\ldots,u_N)$ and least cost satisfy

$$V = R(u)V + K(u) \leqq R(v)V + K(v), \text{ all } v_i \in U_i, \quad v = (v_1,\ldots,v_N).$$

Thus V satisfies (16').

2°. If vectors A,B satisfy (16'), then $\max(A,B)$ (take the max component by component) satisfies (16') by the following argument.

$$\begin{aligned} A_i &\leqq \sum_j p_{ij}(\alpha)A_j + k(i,\alpha) \\ B_i &\leqq \sum_j p_{ij}(\alpha)B_j + k(i,\alpha) \end{aligned} \qquad \text{all } \alpha,i$$

$$\max(A_i,B_i) \leqq \sum_j p_{ij}(\alpha)\max(A_j,B_j) + k(i,\alpha).$$

3°. Next, it is shown that all vectors W satisfying (16') also satisfy $W \leqq V$. This implies that the set of vectors satisfying (16') is a lattice with maximal element V, and proves the theorem. Let U satisfy (16') with $U_i > V_i$. Then $W = \max(U,V)$ satisfies (16'). Write $W_i = V_i+\epsilon_i$, $\epsilon_i > 0$ for $i = 1,\ldots,r$ and $\epsilon_s = 0$ for $s > r$. Then, using the fact $V = V(u)$,

$$\sum_j p_{ij}(u_i)W_j + k(i,u_i) = \sum_j p_{ij}(u_i)V_j + k(i,u_i) + \sum_j p_{ij}(u_i)\epsilon_j$$

$$= V_i + \sum_1^r p_{ij}(u_i)\epsilon_j \geqq W_i = V_i+\epsilon_i$$

and

$$\sum_1^r p_{ij}(u_i)\epsilon_j \geqq \epsilon_i, \quad i = 1,\ldots,r$$

which implies that $\sum_{j=1}^{r} p_{ij}^{(n)}(u) = 1$ for all n. This contradicts the fact that $R^n(u) \to 0$. Thus $W = V$. Q.E.D.

2.2. Additional constraints. In addition to (7) suppose that we require satisfaction of the inequality constraints

$$\sum_{i,r} e_{ir}^{s} M_{ir} \leqq \delta_s, \qquad s = 1,\ldots,\ell. \tag{22}$$

Let the dual variables be $C_1,\ldots,C_N, C_{N+1},\ldots,C_{N+\ell}$, where the C_i, $i \leqq N$, corresponds to the i^{th} equality in (7) and C_{N+i} corresponds to the i^{th} inequality in (22). Then, for the dual problem, the C_i, $i \leqq N$, are unconstrained in sign (see rules in [8], p. 125-7) and the C_{N+i}, $0 < i \leqq \ell$, are non-negative. The dual equations can be written as

$$C_i \leqq \sum_{j=1}^{N} p_{ij}(\alpha_s)C_j + [\sum_{r=1}^{\ell} e_{is}^{r} C_{N+r} + k(i,\alpha_s)] \tag{23}$$

$$s = 1,\ldots,q; \quad i = 1,\ldots,N,$$

and we maximize

$$\sum_{1}^{N} \mu_i C_i - \sum_{1}^{\ell} \delta_i C_{N+i} = z. \tag{24}$$

Suppose all C_{N+i} are given. Then (23), (24) is equivalent to the problem of computing the optimal control for the cost

$$\tilde{k}(i,\alpha_s) = \sum_{r=1}^{\ell} e_{is}^r C_{N+r} + k(i,\alpha_s).$$

In many control applications, the $e_{is}^r \geq 0$. See example below. Then the dual L.P. is, in a sense, equivalent to finding the optimal control for a cost rate $\tilde{k}(i,\alpha)$ which weighs (positively) the constraint. I.e., suppose $\ell = 1$ and $e_{ir}^1 = 1$. Then, we seek the control which minimizes $\sum_{i,j} M_{ij} k(i,\alpha_j)$ subject to the mean time to absorbtion being no greater than δ_1. Then

$$\tilde{k}(i,\alpha) = C_{N+1} + k(i,\alpha).$$

Thus, the equivalent cost $C = (C_1,\ldots,C_N)$ is

$$(25) \qquad C_i = [E_i^u C_{N+1} \cdot \text{time to absorbtion} + E_i^u \sum_{n=0}^{\infty} k(X_n, u(X_n))].$$

If $\delta_i \geq 0$, the form (24) suggests that we want to find the least weights C_{N+i}, for which the control which minimizes (25) also satisfies the constraints (22). Note that the optimal controls for at most ℓ states may possibly be randomized, since the basic solutions of the primal problem may have as many as $N+\ell$ of the $\{M_{ij}\}$ non zero.

2.3. Example. To see how 'state space' constraints of the form (22) may appear, we consider a simple Markov chain problem which is a discretization of a continuous time problem. Consider the system $\ddot{y} = u+\sigma\xi$, where ξ_t is white Gaussian noise and $|u| \leq 1$. In Itô equation form, the system is

$$dx_1 = x_2 dt$$

$$dx_2 = u dt + \sigma dz$$

where z_t is a Wiener process. Suppose that we wish to drive $x_t = (x_{1t}, x_{2t})$ to the target line T in Figure 1, in minimum average time. By the method in [1], an approximating Markov chain $\{X_n\}$ (whose state space is the collection of nodes in Fig. 1) can be obtained. Let h denote the distance between nodes in Fig. 1, with $h < \sigma^2$, and let e_i denote the unit vector in the i^{th} coordinate direction. Then for x on a node not on T, the transition probabilities of the Markov chain are

$$p_{x, x+e_1 h}(u) = h|x_2|/(\sigma^2 + h|x_2|) \quad \text{if} \quad x_2 \geq 0$$

$$= 0 \quad \text{if} \quad x_2 < 0$$

$$p_{x, x-e_1 h}(u) = 0 \quad \text{if} \quad x_2 \geq 0$$

$$= h|x_2|/(\sigma^2 + h|x_2|) \quad \text{if} \quad x_2 < 0$$

$$p_{x, x+e_2 h}(u) = (\sigma^2 + hu)/2(\sigma^2 + h|x_2|)$$

$$p_{x, x-e_2 h}(u) = (\sigma^2 - hu)/2(\sigma^2 + h|x_2|)$$

$$k(x,u) = k(x) = h^2/(\sigma^2 + h|x_2|)$$

In order to solve the minimum average time to (the nodes on) T problem for $\{X_n\}$, it is necessary to truncate the space. To do this we

fix an external boundary B as in Fig. 2, and assign transition probabilities on B to be consistent with the internal dynamics in some way. Several procedures are possible, and, for our purposes, the exact procedure is unimportant. Suppose only that, on the indicated segments of B, the process can move in the directions of the arrows with given probabilities. Of course, specification of an outerboundary may be part of the original problem statement.

Let us next consider some state space constraints. A reasonable constraint (considering that the model may not be adequate for large $|x|$ any way) is

$$\text{(i): Average time on boundary} = \sum_{i \in B} \sum_{j=1}^{q} M_{ij} \leq \delta.$$

(i) denotes the average time on the boundary for the $\{X_n\}$ process. If we wish an approximation to the original continuous time problem, with the additional constraint that the average time the original process is on the boundary B, we need to take into account the fact that a unit time for the $\{X_n\}$ process is not a unit time for the x_t process. The details must be omitted due to space limitations, and the reader is referred to [11]. It will suffice to say that the weighted average

$$\text{(ii):} \quad \sum_{i \in B} a_i \sum_{j=1}^{q} M_{ij} \leq \delta$$

is required in lieu of (i) where $a_i = (\sigma^2 + h|x_2|)^{-1}$, where x_2 is the second component of the vector x at node i.

In general, there may be a region Q which it is undesirable to enter, and we can introduce

$$\text{(iii)}: \sum_{i \in Q} a_i \sum_{j=1}^{q} M_{ij} \leqq \delta.$$

For another type of example, suppose that fuel has an associated cost. Note that the $p_{ij}(u)$ are <u>linear</u> in the control u. Let $\beta_i = P\{u(X_n) = +1 \mid X_n = i\}$. If $\beta_i = \frac{1}{2}$, then the average (or actual) control at state i is zero. Indeed, we can suppose that the actual applied control is $2\beta_i - 1$ since this gives the same transition probabilities as the random control. In general, the average cost of fuel at state i is $|2\beta_i - 1|$. Define M_i^+, M_i^- as the M_{ij}, where j corresponds to $u = +1$ and $u_i = -1$, resp. Then, the average fuel used is

$$\text{(iv)}: F = \sum_i |M_i^+ - M_i^-| a_i$$

and, we can optimize with constraint $F \leqq \delta$. The constraint (iv) can be put into a linear form by the introduction of suitable auxiliary variables as follows: Minimize $\sum_i (M_i^+ + M_i^-)a_i$ with the constraints (6) and $V^i - W^i = M_i^+ - M_i^-$ and $V^i \geqq 0$, $W^i \geqq 0$, and

$$\sum_i (V^i + W^i)a_i \leqq \delta.$$

(See [7], Sec. 5.3 for a similar substitution.)

Example Continued: Numerical Result. Let $h = .55$, $\sigma^2 = 2$. Then for the region of Fig. 2, we will have $N = 195$ states, including the 3 target states. The $k(x)$ on the outer boundary nodes are 1.5 of what their values would be were the node not on a boundary, and we let the $p_{ij}(u)$ be independent of u, for i on the upper and lower boundary. u_i (Eqn. (6)) equals one for the ◻ marked state in Fig. 2. Note that the immediate effects of the control u are on the vertical movement only. The control values (± 1) for the minimum average time problem are given in Fig. 3. Denote T^* = minimum average time = minimum average fuel. Figs. 4 and 5, plot the control values for $\delta = .9T^*$ and $.75T^*$, resp., and indicate the expected decrease of control effort on the counter clockwise side of the switching curve as δ decreases.

Note that the control value $u = 0$ is singular (see also the end subsection of the paper) in that either the right side of (23) is minimized (for this example) at $\alpha_s = \pm 1$, or else it does not vary as α varies in $[-1, +1]$; i.e., if the optimal control for state i is zero, it can never be determined by minimizing the right side of (23), as it could if there were no side constraints. The example also emphasizes the relationship between singularity and randomness of a control.

3. The L. P. Form of the Finite Time Problem.

Consider the dynamic programming problem (P3): minimize, for each $i = 1,\ldots,N$,

$$E_i^{\pi} \sum_0^n k(X_n, u(X_n))$$

where $\pi = (u^0, u^1, \ldots, u^{n-1})$ is a sequence of control vectors, u^i being used at time $n-i$. (P3) is equivalent to the following L.P. problem. Let $y_{ij}(m) = P\{X_m = i,\ u(X_m) = \alpha_j\}$. Minimize

$$z = \sum_{m=0}^{n} \sum_{j=1}^{q} \sum_{i=1}^{N} y_{ij}(m) k(i, \alpha_j) \tag{26}$$

with constraints (the Chapman-Kolmogorov equation)

$$y_i(0) \equiv \sum_j y_{ij}(0) = \mu_i, \quad i = 1, \ldots, N, \qquad y_{ij}(m) \geq 0$$

$$y_i(m+1) = \sum_j y_{ij}(m+1) = \sum_{\ell, k} y_{k\ell}(m) p_{ki}(\alpha_\ell), \quad m = 0, 1, \ldots, n-1; \tag{27}$$

where all $\mu_i > 0$ and $\sum \mu_i = 1$. We will write the L.P. eqns. for the more general problem (P4): minimize (26) with constraint (27), for any $\mu_i \geq 0$ and the inequality constraints

$$\sum_{i,j} a_{ij}(m) y_{ij}(m) \leq \delta_m, \quad m = 0, \ldots, n. \tag{28}$$

(28) includes only one constraint for each time m, but the general case is just as simple.

Define the row vectors with q components

$$\tilde{I} = (1,\ldots,1)$$

$$p_{ij} = (p_{ij}(\alpha_1),\ldots,p_{ij}(\alpha_q))$$

$$k(i) = (k(i,\alpha_1),\ldots,k(i,\alpha_q))$$

$$a_i(m) = (a_{i1}(m),\ldots,a_{iq}(m))$$

and the column vectors (with q and Nq components, resp.)

$$y_i(m) = (y_{i1}(m),\ldots,y_{iq}(m))$$

$$y(m) = (y_1(m),\ldots,y_N(m)).$$

Then the simplex tableau can be written in the form of Fig. 3.

Let the column vector $C(m) = (C_1(m),\ldots,C_N(m))$ be the dual vector to the m^{th} group of equations in Figure 3. The dual of (P4) is: maximize

$$C'(0)\mu - \sum_{0}^{n} \tilde{C}_i \delta_i, \quad \tilde{C}_i \geq 0 \tag{29}$$

with the constraints

$$\begin{aligned} &C_i(n) \leq \tilde{C}_n a_{i\ell}(n), \quad \ell = 1,\ldots,q \\ &C_i(m-1) \leq \sum_{j=1}^{N} p_{ij}(\alpha_\ell) C_j(m) + k(i,\alpha_\ell) + \tilde{C}_{m-1} a_{i\ell}(m-1) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{all } i,m,\ell. \end{aligned} \tag{30}$$

dual variables							
$C_1(0)$ $\vdots$ $C_N(0)$	$\mu_1 =$ $\vdots$ $\mu_N =$	$\tilde{I}$ $\cdot$ 0 $\cdot$ $\cdot$ 0 $\tilde{I}$	0	0	$\cdots$	0	$y_1(0)$ $\vdots$ $y_N(0)$
$C(1)$	$0 =$	$-p_{11}\cdots-p_{N1}$ $-p_{1N}\cdots-p_{NN}$	$\tilde{I}$ $\cdot$ 0 $\cdot$ $\cdot$ 0 $\tilde{I}$	0	$\cdots$	0	$y(1)$
	$\vdots$					0	$\vdots$
$C(n)$	$0 =$	0	$\cdots$	0	$-p_{11}\cdots-p_{N1}$ $-p_{1N}\cdots-p_{NN}$	$\tilde{I}$ $\cdot$ 0 $\cdot$ $\cdot$ 0 $\tilde{I}$	$y(N)$
$\tilde{C}_0$ $\tilde{C}_N$	$\delta_0 \geqq$ $\delta_N \geqq$	$a_1(0),\ldots,a_N(0)$	$a_1(1),\ldots,a_N(1)$			$a_1(n),\ldots,a_N(n)$	
		$k(1),\ldots,k(N),$	$k(1),\ldots,k(N),$		$\ldots,$	$k(n),\ldots,k(N)$	

Figure 3

In the absense of the inequality constraints (28), the system (30) is simply the dynamic programming equation. (30) can be put into a more convenient vector form as follows. Let $w = (\alpha_{i_1},\ldots,\alpha_{i_N})$ be an arbitrary control. Define the column vectors

$$K(w) = (k(1,\alpha_{i_1}),\ldots,k(N,\alpha_{i_N}))$$

$$a(w;m) = (a_{1i_1}(m),\ldots,a_{Ni_N}(m)).$$

Then (31) is equivalent to (30).

(31)
$$C(n) \leqq \tilde{C}_n a(w;n)$$
$$C(m-1) \leqq R(w)C(m) + K(w) + \tilde{C}_{m-1}a(w;m-1)$$

for all control vectors w.

4. A Maximum Principle for Markov Chains.

The linear programming formulation treats the control and state simultaneously, in that the M_{ij} or y_{ij} are the free variables. Next, by a direct application of the deterministic discrete time maximum principle, a form of stochastic maximum principle for the fixed finite time Markov problem will be derived, in which the control and state are treated analogously to their treatment in the deterministic problem.

Define $p_i^{(n)} = P\{X_n = i\}$. The probabilities $p_i^{(n)}$ will be the

dynamical variables. Again $U = (\alpha_1,\ldots,\alpha_q)$, and we suppose that the control variables are the probabilities

$$\beta_{ij}^n = P\{u(X_n) = \alpha_j | X_n = i\}.$$

Indeed, whether or not the solution is a pure control, it is (once more) only by allowing randomization that the discrete maximum principle will be applicable. Define the vectors $\beta_i^n = (\beta_{i1}^n,\ldots,\beta_{iq}^n)$ and $\beta^n = (\beta_1^n,\ldots,\beta_N^n)$. β^n takes the place of the $u = (u_1,\ldots,u_N)$ of the dynamic programming problem (and the γ of the L.P. problem). Let $R(\beta^m) = \{p_{ij}(\beta_i^m);$ $i,j = 1,\ldots,N\}$ denote the matrix of transition probabilities (with state 0 deleted) under the random rule β^m; i.e.,

$$p_{ij}(\beta_i^m) = \sum_\ell p_{ij}(\alpha_\ell)\beta_{i\ell}^m.$$

Define the (N) column vector $K(\beta^m) = (k(1,\beta_1^m),\ldots,k(N,\beta_N^m))$ where

$$k(i,\beta_i^m) = \sum_\ell k(i,\alpha_\ell)\beta_{i\ell}^m.$$

The problem to be treated is (P5), a slight extension of (P4). The dynamics are

$$(32) \qquad p^{(m+1)} = R'(\beta^m)p^{(m)} = [R'(\beta^m)p^{(m)} - p^{(m)}] + p^{(m)}$$
$$\equiv f(p^{(m)},\beta^m) + p^{(m)}, \quad m = 0,1,\ldots,n-1.$$

The cost is

$$z = \sum_{m=0}^{n} \sum_{i,j} k(i,\alpha_j) p_i^{(m)} \beta_{ij}^m = \sum_m K'(\beta^m) \cdot p^{(m)}, \tag{33}$$

with constraints

$$\begin{aligned} &G_0 p^{(0)} = \tilde{\delta}_0, \quad G_n p^{(n)} = \tilde{\delta}_n \\ &Q_i p^{(i)} \leq \delta_i, \quad i = 0,\ldots,n, \end{aligned} \tag{34}$$

where G_0, G_n and the Q_i are matrices of full rank, and $\tilde{\delta}_0$, $\tilde{\delta}_n$ and δ_i are suitable vectors. Define $p_0^{(m)}$ by $p_0^{(0)} = 0$ and

$$p_0^{(m+1)} = p_0^{(m)} + f_0(p^{(m)},\beta^m) \equiv p_0^{(m)} + K'(\beta^m) p^{(m)}$$

Then $p_0^{(n)} = z$.

Observe that the set

$$f(p^{(m)},\beta^m)$$

$$f_0(p^{(m)},\beta^m)$$

is convex in the control vector β^m. It is easy to see that the conditions of the discrete maximum principle hold for the set (33) - (34) (see [7, Chapter 4], and note that we change some signs here in order to bring the result in closer conformity with dynamic programming usage). A direct transcription of this discrete maximum principle yields

Theorem 3. Let $\hat{\beta}^0,\ldots,\hat{\beta}^{n-1}$ and $\hat{p}^{(0)},\ldots,\hat{p}^{(n)}$ be the optimal control and state, resp. Then there are costate vectors $\pi(0),\ldots,\pi(n)$, and vectors $\lambda_i \geq 0$ (all components are non-negative), and vectors $\tilde{\lambda}_0$, $\tilde{\lambda}_n$ and a scalar $\pi^0 \geq 0$. (Not all the π^0, $\pi(0),\ldots,\pi(n)$, $\tilde{\lambda}_0$, $\tilde{\lambda}_n$ are zero.) The $\pi(i)$ satisfy the adjoint equation

$$\text{(35a)} \qquad \pi(m) = \pi(m+1) + [R(\hat{\beta}^m)-I]\pi(m+1) + \pi^0 K(\hat{\beta}^m) + Q_m'\lambda_m$$

$$m = 0,\ldots,n-1$$

and the transversality conditions

$$\text{(35b)} \qquad \pi(0) = G_0'\tilde{\lambda}_0, \quad \pi(n) = G_n'\tilde{\lambda}_n + Q_n'\lambda_n$$

and

$$\text{(35c)} \qquad \lambda_m'(Q_m\hat{p}^{(m)} - \delta_m) = 0, \quad m = 0,\ldots,n.$$

Define the Hamiltonian

$$H(p^{(m)},\beta^{(m)},\pi,\pi^0,m) = \pi^0 K'(\beta^m)p^{(m)} + \pi' f(p^{(m)},\beta^m).$$

Then

$$\text{(36)} \qquad H(\hat{p}^{(m)},\hat{\beta}^m,\pi(m+1),\pi^0,m) \leq H(\hat{p}^{(m)},\beta^m,\pi(m+1),\pi^0,m)$$

for all β^i, or equivalently,

$$(37) \qquad (\hat{p}^{(m)})'[R(\hat{\beta}^m)\pi(m+1) + \pi^o K(\hat{\beta}^m)] \leq \hat{p}^{(m)}[R(\beta^m)\pi(m+1) + \pi^o K(\beta^m)].$$

In terms of components (37) is

$$(37a) \qquad \hat{p}_i^{(m)}[\sum_j p_{ij}(\hat{\beta}_i^m)\pi_j(m+1) + \pi^o k(i,\hat{\beta}_i^m)] \leq \hat{p}_i^{(m)}[\sum_j p_{ij}(\beta_i^m)\pi_j(m+1) + \pi^o k(i,\hat{\beta}_i^m)]$$

Remark. It can be shown that $\pi^o > 0$; thus we can set $\pi^o = 1$. Let $G_0 = I$. Then $\tilde{\delta}_0 = \mu$, a vector of given initial probabilities. Suppose that the other constraints of (34) are absent. Then, $\pi(i)$ is the optimal dynamic programming cost vector, with $n-i$ steps to go, and (35a) and (37) combine into

$$\pi(m) = R(\hat{\beta}^m)\pi(m+1) + K(\hat{\beta}^m) \leq$$

$$R(\beta^m)\pi(m+1) + K(\beta^m), \quad \pi(n) = 0, \; m = 0,1,\ldots,n-1,$$

which is precisely the dynamic programming equation (3).

Remark on Singular Controls. The set $\{p_{ij}(\beta_i),\ i = 1,\ldots,N,\ k(i,\beta^i)\}$ is a convex polyhedron, as β^i varies over its admissible values. Thus, the minimum of the r.h.s. of (37) lies on a vertex of the polyhedron - or, if the minimum falls on more than one vertiex, it also falls in the convex hull of the set of vertices on which the minimum occurs. Con-

sider the example (a typical discrete problem derived from a continuous time problem which is linear in the control). In that case there are at most two extreme points to the polyhedron and $p_{ij}(\beta^m)$ has the form $\beta_1^m p_{ij}(+1) + (1-\beta_1^m)p_{ij}(-1)$, and $k(i,\beta_i^m) = k(i)$ and we can write the r.h.s. of (37a) as

$$\hat{p}_i^{(m)}\Big[\sum_{j=1}^{N} p_{ij}(\beta_i^m)\pi_j(m+1) + k(i)\Big]$$

$$= \beta_i^m \cdot d_i^+(m+1) - \beta_i^m d_i^-(m+1) + e_{ij}(m+1)\ ,\ d_i^+ \geq 0,\ d_i^- \geq 0,$$

and the minimizing β_i^m satisfies

$$\beta_i^m = 1 \quad \text{if} \quad d_i^-(m+1) \geq d_i^+(m+1)$$

$$= 0 \quad \text{if} \quad d_i^+(m+1) > d_i^-(m+1)$$

$$= ? \quad \text{otherwise.}$$

However, we have seen in past sections that, in the presense of 'state variable' constraints (34) (except $G_0 p^{(0)} = \tilde{\delta}_0$), the control may be random for some times m and states i. Thus, with these state variable constraints, the control may well be singular; i.e., $d_i^+(m+1) = d_i^-(m+1)$, and the maximum principle yields no information directly, in analogy to the state variable constrained deterministic case. Existing works (e.g. [10], on continuous time stochastic maximum principles - dealing with terminal time 'average' constraints $g(EX_T) = 0$ have not adequately accounted for the possibility of randomization. It would also be worthwhile to study

methods for extracting information from the stochastic Hamiltonian formulation in the singular situation. One of the advantages of our study of the simple Markov chain problem, is that the singular - and randomization - problems are made quite apparent, a situation not easily seen from the continuous time formulations.

Bibliography

1. H. Kushner, A. Kleinman, "Numerical Methods for the Solution of the Degenerate Nonlinear Equations arising in Optimal Stochastic Control Theory", IEEE Trans. on Automatic Control, AC-13, No. 4, 1968.

2. H. Kushner, Introduction to Stochastic Control Theory, Holt, Rinehart and Winston, in press.

3. H. Wagner, Principles of Operations Research, Prentice Hall, 1969.

4. P. Wolfe, G. Dantzig, "Linear Programming in a Markov Chain", Oper. Res., 10, 1962, pp. 702-710.

5. A. Manne, "Linear Programming and Sequential Decisions", Man-Sci., April, 1960.

6. C. Derman, "On Sequential Decisions and Markov Chains", Man. Sci., 9, No. 1, 1962.

7. M. Canon, C. Cullum, Jr., E. Polak, Theory of Optimal Control and Mathematical Programming, McGraw-Hill, 1970.

8. G. Dantzig, Linear Programming and Extensions, Princeton Univ. Press, 1963.

9. G. de Ghellinck, "Les Problems de Decisions Sequentielles", Cahiers de Centre d'Etude de Recherche Operationelle, 2, No. 2, Brussels, 1960.

10. H. Kushner, "On the Stochastic Maximum Principle with 'Average' Constraints", J. Math. Anal. and Appl., 12, No. 1, 1965.

11. H. Kushner, "Probability Limit Theorems and the Convergence of Finite Difference Approximations of Partial Differential Equations", Brown Univ. C. D. S. Report, 69-4, to appear in J. Math. Anal. and Applications.

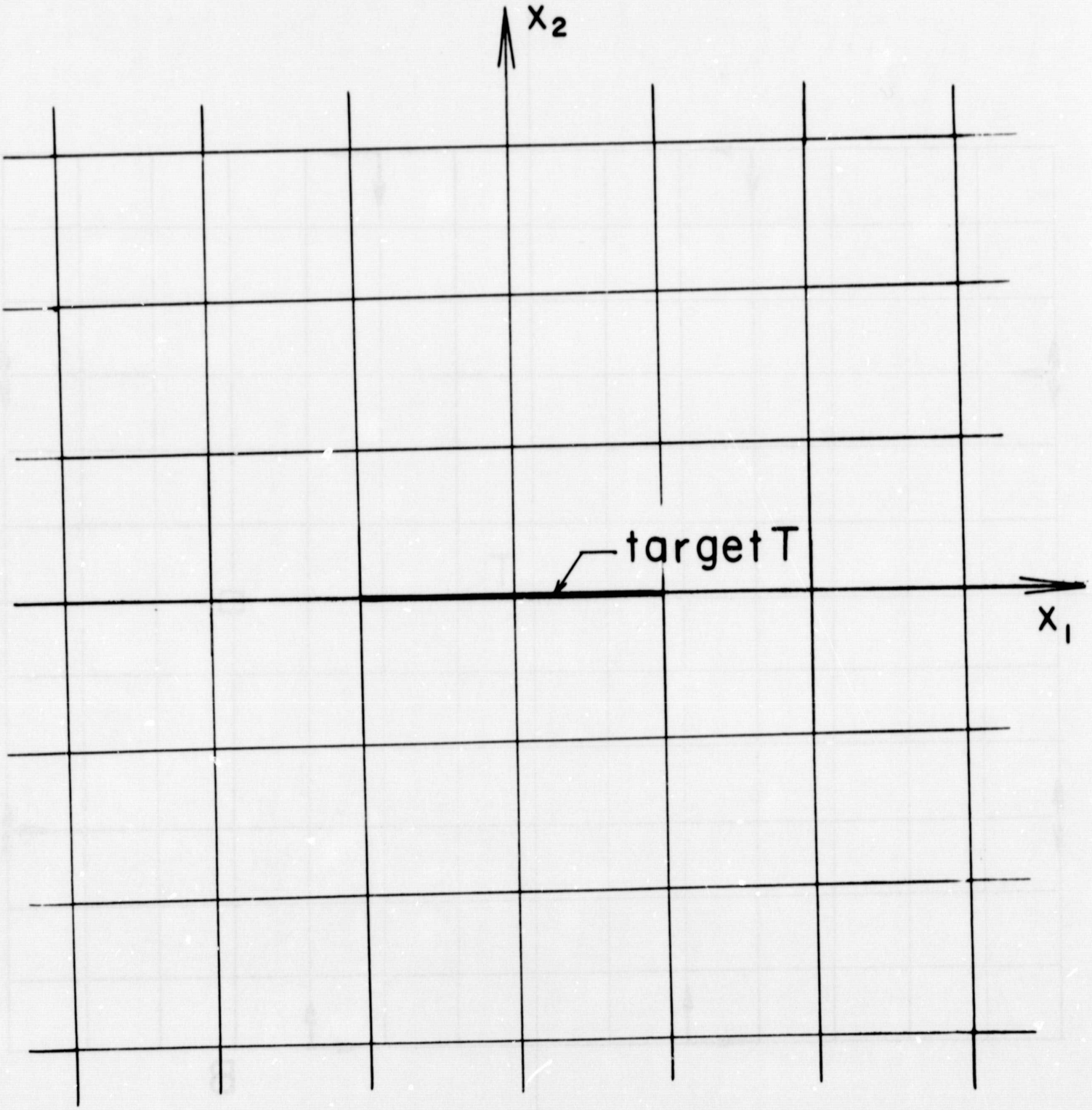


FIG. 1

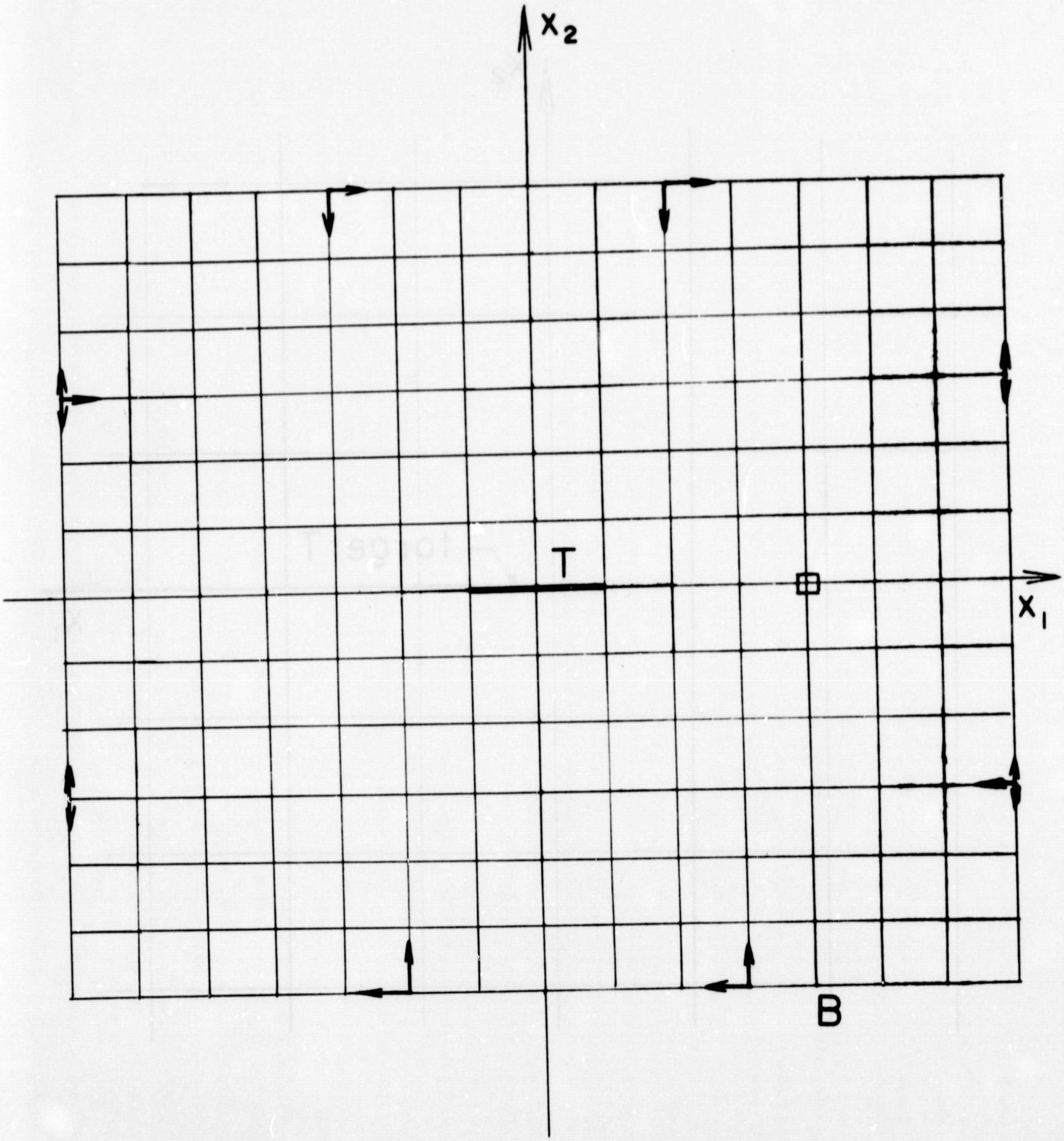


FIG. 2

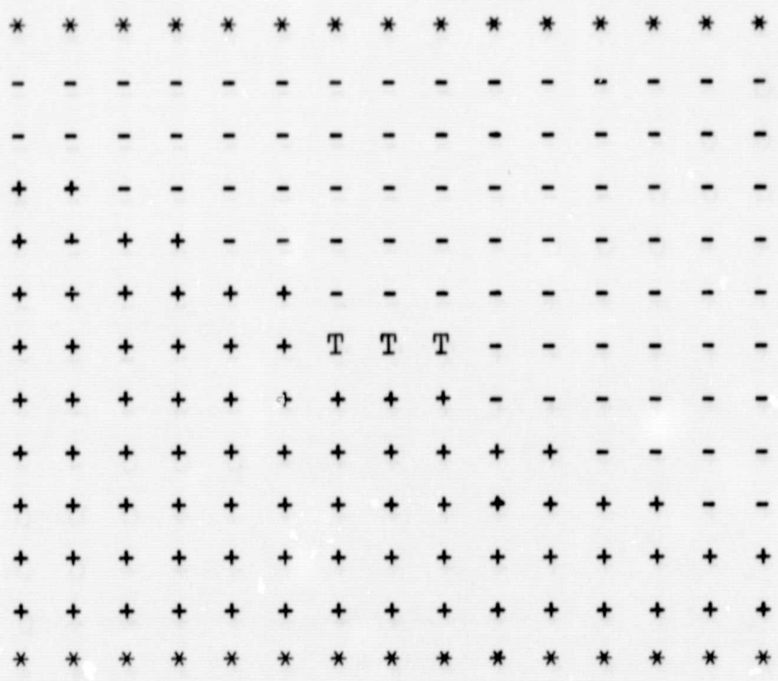


Figure 3. Optimal Control Values for the Minimum Average Time Problem.

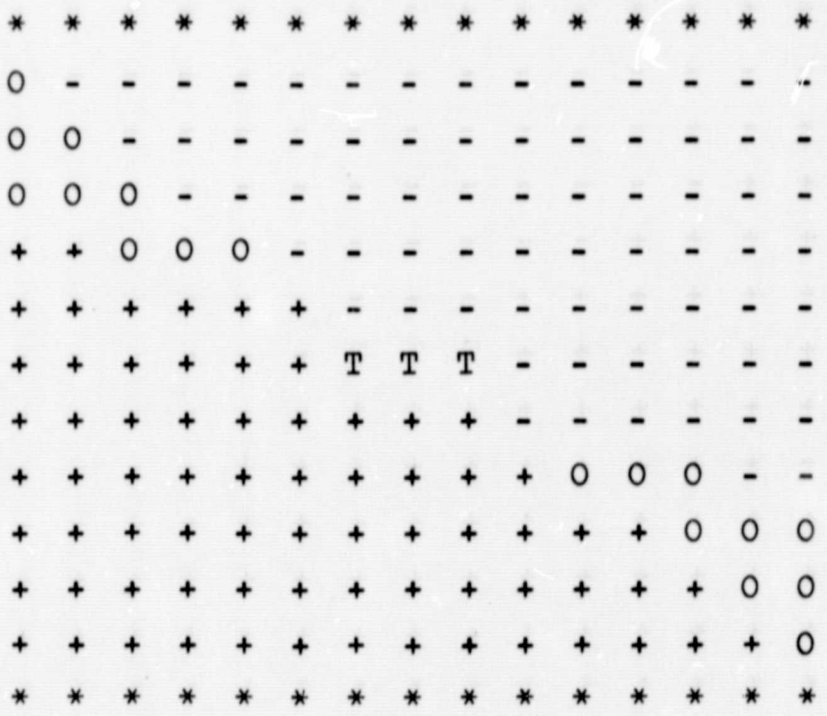


Figure 4. Optimal Control Values for Minimum Average Time Problem with Average Fuel $\leq .9T^*$

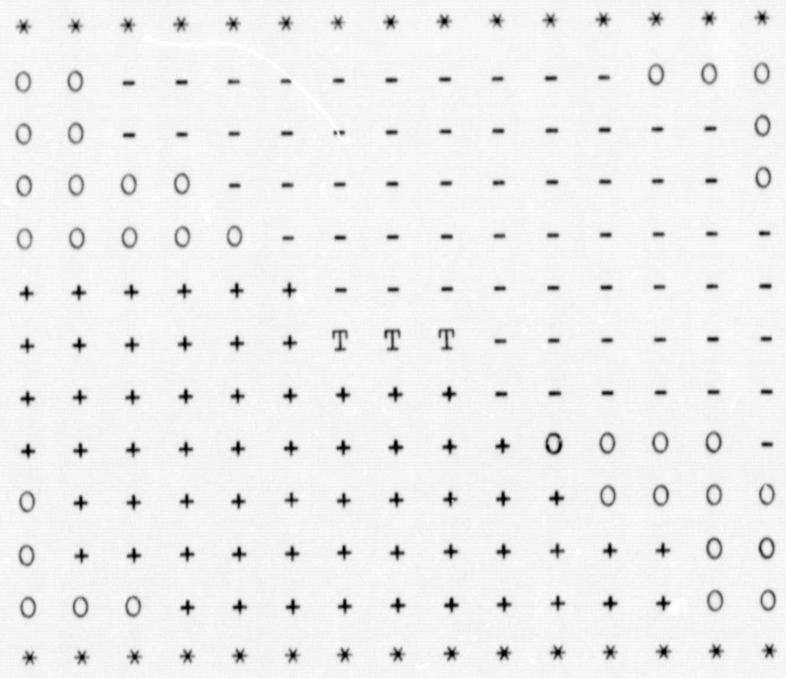


Figure 5. Optimal Control Values for Minimum Average Time with Average Fuel $\leq .75T^*$