On Sequential Decisions and Markov Chains
Author(s): Cyrus Derman
Source: *Management Science*, Vol. 9, No. 1 (Oct., 1962), pp. 16-24
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/2627184
Accessed: 11/02/2014 08:18

ON SEQUENTIAL DECISIONS AND MARKOV CHAINS*[1]

CYRUS DERMAN

Columbia University and Technion, Israel Institute of Technology

Several problems in the optimal control of dynamic systems are considered. When observed, a system is classifiable into one of a finite number of states and controlled by making one of a finite number of decisions. The sequence of observed states is a stochastic process dependent upon the sequence of decisions, in that the decisions determine the probability laws that operate on the system. Costs are associated with the sequence of states and decisions. It is shown that, for the problems considered, the optimal rules for controlling the system belong to a subclass of all possible rules and, within this subclass, the optimal rules can be derived by solving linear programming problems.

1. Introduction

We are concerned with a controlled dynamic stochastic system which, at times $t = 0, 1, \cdots$ is observed to be in one of $L + 1$ states $0, \cdots, L$. After each observation, the system is "controlled" by making one of K decisions $d_1, \cdots, d_K$. The decisions have the effect of determining the laws of chance which will operate on the system.

Let $X_0, X_1, \cdots$ denote the sequence of observed states and $\Delta_0, \Delta_1, \cdots$, the sequence of decisions. The class C of *all* decision procedures R consists of functions

$$D_k(X_0, \Delta_0, \cdots, X_t) = P(\Delta_t = d_k \mid X_0, \Delta_0, \cdots, X_t)$$

for $k = 1, \cdots, K; \quad t = 0, 1, \cdots$. For every $X_0, \Delta_0, \cdots, X_t, \quad t = 0, 1, \cdots$

$$D_k(X_0, \Delta_0, \cdots, X_t) \geqq 0, \qquad k = 1, \cdots, K$$

and

$$\sum_{k=1}^{K} D_k(X_0, \Delta_0, \cdots, X_t) = 1.$$

We shall assume throughout that

$$\begin{aligned} P(X_{t+1} = j \mid X_0, \Delta_0, \cdots, X_t = i) \\ = \sum_{k=1}^{K} q_{ij}(k)\, D_k(X_0, \Delta_0, \cdots, X_t = i) \end{aligned}$$

for $i, j, = 0, \cdots, L; t = 0, 1, \cdots$ where $\{q_{ij}(k)\}$ are such that

$$q_{ij}(k) \geqq 0, \quad i, j = 0, \cdots, L; k = 1, \cdots, K$$

and

$$\sum_{j=0}^{L} q_{ij}(k) = 1, \qquad i = 0, \cdots, L; k = 1, \cdots, K.$$

* Received December 1961.

[1] Work sponsored by the Office of Naval Research.

The $\{q_{ij}(k)\}$'s being stochastic matrices for each k are the chance laws brought into effect by the various decisions; they are functions only of the last observed state of the system.

Let C' be the class of decision procedures R such that

$$D_k(X_0, \Delta_0 \cdots, X_t = i) = D_{ik}, \qquad i = 0, \cdots, L; k = 1, \cdots, K,$$

independent of $X_0, \cdots, X_{t-1}, \Delta_{t-1}$ and t. If $R \varepsilon C'$, the sequence $\{X_t\}$, $t = 0, 1, \cdots$ is a Markov Chain with stationary transition probabilities $\{p_{ij}\}$ where

$$p_{ij} = \sum_{k=1}^{K} q_{ij}(k)\, D_{ik}, \qquad i, j = 0, \cdots, L.$$

Let C'' be the sub-class of C' such that $D_{ik} = 0$ or 1. Thus C'' contains only a finite number of procedures.

Let $w_{ik}(t) > 0$, $i = 0, \cdots, L; k = 1, \cdots, K; t = 0, 1 \cdots$, be finite values denoting the *expected cost* ascribed to time t given that the system is observed in state i at time t and that decision d_k is made. We *assume that* $w_{ik}(t) = w_{ik}$, *independent of* t.

Let W_t, $t = 0, 1 \cdots$ denote the *expected cost* ascribed to time t as a result of using a given decision procedure R. We are concerned with two problems.

Problem 1. Suppose $X_0 = i$ with probability 1. We want to find that procedure R_1 such that

$$Q_{R_1}(i) = \min_{R \varepsilon C} Q_R(i) \qquad i = 0, \cdots, L,$$

where

$$Q_R(i) = \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T} W_t,$$

the average expected cost per unit time whenever the limit exists.

Problem 2. Suppose $X_0 = i$ with probability 1, L is an absorbing state for all $R \varepsilon C$ (i.e. $X_t = L$ implies $X_{t'} = L$ if $t' > t$ for all $R \varepsilon C$), and $w_{Lk} = 0$, $k = 1, \cdots, K$. We want to find that procedure R_2 such that

$$S_{R_2}(i) = \min_{R \varepsilon C} S_R(i), \qquad i = 0, \cdots, L,$$

where

$$S_R(i) = \sum_{t=0}^{\infty} w_t.$$

In words, $S_R(i)$ is the expected total cost involved in taking the system from state i to state L using procedure R. Problem 2 bears some resemblance to the "Shortest-Route" problem (see [10]).

Our interest in these problems grew out of a "replacement" problem considered by Derman [2] and a paper by Eaton and Zadeh [4] on "pursuit strategies".

The method employed here is due to Manne [9][2] who, in the context of an inventory problem, showed the applicability of linear programming to Problem 1.[3] Howard [6] has provided another method for solving Problem 1.

Our results show (i) that both R_1 and R_2 are contained within the class C'' (hence, in C') so that the Procedures $C - C'$ can be ignored, and (ii) that under irreducibility assumptions, Problem 2 as well as Problem 1, can be formulated as a linear programming problem.

In the above cited references the procedures in C-C' have been ignored. Intuitively obvious as it may be that these rules can be ignored, a proof is required before a procedure optimal over C' can be considered as *the* optimal procedure. For a similar encounter with this type of question[4] see [3] and [5].

Following Manne's paper, Wagner [11] using the machinery of linear programming showed that $R_1 \varepsilon C''$. Our proof follows the functional equations approach.

Our lemma used for proving the applicability of linear programming to Problem 2 leads us to consider an extension of Problem 1 where the expression to be minimized is the ratio of average expected costs per unit time with respect to two different sets of costs. See Klein [8] for an example of this type of problem.

2. Reduction to the Class C″

In the following theorem we show that, in seeking an optimal procedure in C, it is sufficient to consider only the members of C''.

Theorem 1. There exists procedures R_1 and R_2 in C'' such that, for problem 1,

$$Q_{R_1}(i) = \min_{R \varepsilon C} Q_R(i), \qquad i = 0, \cdots, L, \tag{1}$$

and, for problem 2,

$$S_{R_2}(i) = \min_{R \varepsilon C} S_R(i) \text{ (possibly } \infty), \qquad i = 0, \cdots, L. \tag{2}$$

Proof: Since the proofs of both statements are almost identical, we can carry out most of the two proofs simultaneously. For any rule R and $\alpha (0 < \alpha < 1)$ let

$$V_R(i, \alpha) = \sum_{t=0}^{\infty} \alpha^t W_t, \qquad i = 0, \cdots, L$$

From Karlin [7] there exists, for each α, a procedure $R_\alpha \varepsilon C$ such that

$$V_{R_\alpha}(i, \alpha) = \min_{R \varepsilon C} V_R(i, \alpha).$$

[2] We are also indebted to R. Oliver for directing our attention to this method.

[3] Our attention has recently been called to an alternative linear programming approach to Problem 1 by Marshall Freimer [13]. He formulates the dual of the problem formulated here.

[4] See also, D. Blackwell, On The Functional Equation Dynamic Programming Journal of Mathematical Analysis and Applications, Vol. 2, No. 2, April 1961.

From a functional equation point of view, it is clear that we can consider R_α to be a member of C'. Let $V_i (i = 0, \cdots L)$ denote the value of $V_{R_\alpha}(i, \alpha)$ and $D_i = (D_{i1}, \cdots, D_{iK}), i = 0, \cdots, L$. Then it is well known that we can write

$$\begin{aligned} V_i &= \min_{D_i} \left\{ \sum_{k=1}^{K} D_{ik} w_{ik} + \alpha \sum_{j=0}^{L} p_{ij} V_j \right\} \\ &= \min_{D_i} \left\{ \sum_{k=1}^{K} D_{ik} w_{ik} + \alpha \sum_{j=0}^{L} \sum_{k=1}^{K} q_{ij}(k) D_{ik} V_j \right\} \\ &= \min_{D_i} \left\{ \sum_{k=1}^{K} D_{ik} \left(w_{ik} + \alpha \sum_{j=0}^{L} q_{ij}(k) V_j \right) \right\} \end{aligned}$$

from which it follows that the minimum on the right side is achieved by having $D_{ik} = 0$ or 1. That is, we can consider R_α to be a member of the finite class C''. Since C'' is a finite class it is possible to choose a sequence $\{\alpha_v\}$, $\lim_{v \to \infty} \alpha_v = 1$, such that $R_{\alpha_v} = R^*$, $v = 1, 2, \cdots$, where $R_{\alpha_v} = R^* \varepsilon C''$. By an Abelian theorem (Widder [12] p. 181) and an application of a Markov chain limit theorem (Chung [2] p. 32) we have

$$\lim_{v \to \infty} (1 - \alpha_v) V_{R^*}(i, \alpha_v) = Q_{R^*}(i).$$

Let R be any procedure in C. Since

$$V_R(i, \alpha) \geqq V_{R_{\alpha_v}}(i, \alpha_v) = V_{R^*}(i, \alpha_v), \qquad i = 0, \cdots, L,$$

we have (using Widder [17] p. 181)

$$\begin{aligned} Q_R(i) &= \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T} W_t \\ &\geqq \lim_{v \to \infty} (1 - \alpha_v) V_R(i, \alpha_v) \\ &\geqq \lim_{v \to \infty} (1 - \alpha_v) V_{R^*}(i, \alpha_v) \\ &= Q_{R^*}(i), \qquad i = 0, \cdots, L \end{aligned}$$

Setting $R_1 = R^*$ this proves (1). To prove (2) we have directly, since $w_{ik} > 0$ for all i and k,

$$\begin{aligned} S_R(i) &= \sum_{t=0}^{\infty} W_t \\ &= \lim_{v \to \infty} V_R(i, \alpha_v) \\ &\geqq \lim_{v \to \infty} S_{R_{\alpha_v}}(i, \alpha_v) \\ &= \lim_{v \to \infty} S_{R^*}(i, \alpha_v) \\ &= S_{R^*}(i), \qquad i = 0, \cdots, L. \end{aligned}$$

This holds whether or not $S_{R^*}(i)$ is finite. The proof is completed on setting $R_2 = R^*$.

3. Formulation of Problems 1 and 2 as Linear Programming Problems

The results of the preceding section enable us to restrict our attention to the procedures of class C'. Thus for all procedures considered, the sequence of states $\{X_t\}$, $t = 0, 1, \cdots$, will be a Markov Chain with stationary transition probabilities. The following well-known (see Chung [2], part 1) results from Markov Chain theory will be relevant.

Let $\{p_{ij}\}$ be the transition probabilities of a Markov Chain $\{X_t\}$, $t = 0, 1, \cdots$ with a finite set I of states all belonging to the same class; $f(j)$, $j \varepsilon I$, a function defined over the states; ${}_ip_{ij}^{(0)}$ denote the conditional probability that $X_t = j$ given that $X_0 = i$ and $X_s \neq i$ for $0 < s < t$ where we define ${}_ip_{ij}^{(o)} = 0$ for all i and j. Then

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} Ef(X_t) = \sum_{j \varepsilon I} \pi_j f(j) \tag{3}$$

and

$$\begin{aligned} \sum_{t=0}^{\infty} \sum_{j \varepsilon I} {}_ip_{ij}^{(t)} f(j) &= \sum_{j \varepsilon I} \sum_{t=0}^{\infty} {}_ip_{ij}^{(t)} f(j) \\ &= \frac{1}{\pi_i} \sum_{j \epsilon I} \pi_j f(j) \end{aligned} \tag{4}$$

where the $\{\pi_j\}$'s uniquely satisfy

$$\begin{aligned} \pi_j &\geqq 0 \\ \pi_j - \sum_{i \varepsilon I} \pi_i p_{ij} &= 0 \qquad j \varepsilon I \\ \sum_{j \varepsilon I} \pi_j &= 1 \end{aligned} \tag{5}$$

In fact, $\pi_j > 0$ for all $j \varepsilon I$.

In order to use (3), (4) and (5) we assume the following:

Assumption A. In problem 1, for every $R \varepsilon C'$ the states $0, \cdots, L$ belong to the same class.

Assumption B. In problem 2, for every $R \varepsilon C'$, the state L is accessible from states $0, \cdots, L - 1$ within a finite number of transitions with probability 1.

We shall show

Theorem 2. If Assumption A (B) holds, then problem 1 (2) can be formulated as a linear programming problem.

Proof: Setting $f(j) = \sum_{k=1}^{K} D_{jk} w_{jk}$, $j = 0, \cdots, L$, we have from assumption A and (3) that for every $R \varepsilon C'$

$$Q_R(i) = \sum_{j=0}^{L} \pi_j \sum_{k=1}^{K} D_{jk} w_{jk} \qquad i = 0, \cdots, L. \tag{6}$$

In problem 2, let us adjoin the state -1 to the states $0, \cdots, L$ and set, for every $R \varepsilon C'$

$$p_{-1,i} = \frac{1}{L+1} \qquad i = 0, \cdots L$$

$$p_{i,-1} = 0 \qquad i = 0, \cdots, L-1$$

$$p_{L,-1} = 1$$

$$w_{-1,k} = 0 \qquad k = 1, \cdots, K$$

Then under assumption B and (4) we have

$$\text{(7)} \qquad \begin{aligned} \frac{1}{L+1} \sum_{i=0}^{L} S_R(i) &= \sum_{t=0}^{\infty} \sum_{j=-1}^{L} {}_{-1}p_{-1,j}^{(t)} \sum_{k=1}^{K} D_{jk}\, w_{jk} \\ &= \frac{1}{\pi_{-1}} \sum_{j=-1}^{L} \pi_j \sum_{k=1}^{K} D_{jk}\, w_{jk} \end{aligned}$$

In both cases the π_j's satisfy (5) with $I = \{0, \cdots, L\}$ for problem 1 and $I = \{-1, \cdots, L\}$ for problem 2.

It is clear that any procedure which minimizes the right side of (7) will minimize $S_R(i)$ for $i = 0, \cdots, L$.

Let $x_{jk} = \pi_j D_{jk}, \quad j \varepsilon I; k = 1, \cdots K$. Then (6) becomes

$$\text{(8)} \qquad Q_R(i) = \sum_{j=0}^{L} \sum_{k=1}^{K} x_{jk} w_{jk}$$

and (7) becomes

$$\text{(9)} \qquad \frac{1}{L+1} \sum_{i=0}^{L} S_R(i) = \frac{\sum_{j=-1}^{L} \sum_{k=1}^{K} x_{jk}\, w_{jk}}{\sum_{k=1}^{K} x_{-1k}}$$

and (5) becomes

$$\text{(10)} \qquad \begin{aligned} x_{jk} &\geqq 0 && j \varepsilon I; k = 1, \cdots, K \\ \sum_{k=1}^{K} x_{jk} - \sum_{i \varepsilon I} \sum_{k=1}^{K} x_{ik} q_{ij}(k) &= 0 && j \varepsilon I \\ \sum_{j \varepsilon I} \sum_{k=1}^{K} x_{jk} &= 1 \end{aligned}$$

Because of the uniqueness of the solutions to (5), there corresponds to every $R \varepsilon C'$ a solution to (10) with $\sum_{n=1}^{K} x_{jk} > 0, \quad j \varepsilon I$. Moreover, every solution to (10) must satisfy $\sum_{k=1}^{K} x_{jk} > 0, \quad j \varepsilon I$; hence, setting

$$D_{jk} = \frac{x_{jk}}{\sum_{k=1}^{K} x_{jk}}, \qquad j \varepsilon I; k = 1, \cdots, K,$$

a solution to (10) corresponds to some $R \varepsilon C'$.

Thus, we have shown that solutions to problems 1 and 2 consist in minimizing (8) and (9) respectively, subject to the contraints (10). This proves the part of the theorem concerning problem 1. To complete the proof, we use the following lemma.

Lemma[5]: The non-linear function

$$g(x) = \frac{\sum_{i=1}^{n} c_i x_i}{\sum_{i=1}^{n} d_i x_i}$$

Can be minimized subject to

$$\text{(11)} \qquad \begin{aligned} x_i &\geqq 0 & i &= 1, \cdots, n \\ \sum_{j=1}^{n} a_{ji} x_i &= b_j & j &= 1, \cdots, m \end{aligned}$$

by solving a linear programming problem if

(i) $$\sum_{i=0}^{n} a_{ji} x_i = 0, \qquad j = 1, \cdots, m$$

and $x_i \geqq 0$, $i = 0, \cdots, n$ imply $x_i = 0$, $i = 1, \cdots, n$ and

(ii) every $x = (x_1, \cdots, x_n)$ satisfying (11) implies

$$\sum_{i=1}^{n} d_i x_i > 0.$$

Proof: The conditions (i) and (ii) imply that the transformation

$$z_i = \frac{x_i}{\sum_{i=1}^{n} d_i x_i}, \quad i = 1, \cdots, n, \qquad z_{n+1} = \frac{1}{\sum_{i=1}^{n} d_i x_i}$$

is one-to-one between (11) and the values $z = (z_1, \cdots, z_{n+1})$ satisfying

$$\text{(12)} \qquad \begin{aligned} z_i &\geqq 0, & i &= 1, \cdots, n+1 \\ \sum_{i=1}^{n} a_{ji} z_i - b_j z_{n+1} &= 0, & j &= 1, \cdots, m \\ \sum_{i=1}^{n} d_i z_i &= 1; \end{aligned}$$

since, by (ii) for every $x = (x_1, \cdots, x_n)$ satisfying (11) $\sum_{i=1}^{n} d_i x_i > 0$, and by (i) every $z = (z_1, \cdots, z_{n+1})$ satisfying (12) must satisfy $z_{n+1} > 0$. However, under the transformation

$$g(x) = h(z) = \sum_{i=1}^{n} c_i z_i$$

[5] We are indebted to A. F. Veinott for helpful suggestions.

is a linear function. Thus the linear programming problem of minimizing $h(z)$ subject to (12) yields, on setting $x_i = z_i/z_{n+1}$, $i = 1, \cdots, n$, the minimum of $g(x)$ subject to (11). This proves the lemma.

One can verify that the lemma applies to (9) and (10), thus completing the proof of the theorem.

4. Ratios of Costs

The lemma of the preceding section is suggestive. Suppose $w'_{ik} > 0$ and $w''_{ik} > 0$, $i = 0, \cdots, L; k = 1, \cdots, K$ are two sets of expected costs. Denote by W'_t and W''_t, $t = 0, 1, \cdots$ the respective expected costs ascribed to time t for a fixed procedure R. Consider the cost criterion

$$\psi_R(i) = \limsup_{T\to\infty} \frac{\sum_{t=0}^{T} W'_t}{\sum_{t=0}^{T} W''_t}, \qquad i = 0, \cdots, L$$

when $X_0 = i$ with probability 1. If $R \varepsilon C'$, and assumption A holds then

$$\psi_R(i) = \frac{\sum_{i=0}^{L}\sum_{k=1}^{K} \pi_i D_{ik} w'_{ik}}{\sum_{i=0}^{L}\sum_{k=1}^{K} \pi_i D_{ik} w''_{ik}}, \qquad i = 0, \cdots, L$$

where π_i, $i = 0, \cdots, L$ satisfy (5). Hence ψ_R can be minimized over all $R \varepsilon C'$ according to the development of the preceding section. The question again arises as to whether this provides an overall optimal procedure, i.e. optimal over all procedures C. To this end we prove

Theorem 3. Under assumption A, there exists a procedure $R_3 \varepsilon C''$ such that

$$\psi_{R_3}(i) = \min_{R\varepsilon C} \psi_R(i)$$

Proof: Suppose $X_0 = i$ with probability 1. Let $M = \inf_{R\varepsilon C} \psi_R(i)$. Then there is a sequence $R(i)$, $R(2)$, $\cdots$ of procedures in C such that $\lim_{v\to\infty} \psi_{R(v)}(i) = M$. Let $\psi_{R(v)}(i) = m_v$, $v = 1, \cdots$. Consider, Problem 1 with $w_{ik} = w'_{ik} - m_v w''_{ik}$ (the hypothesis that $w_{ik} > 0$ is unnecessary for Problem 1). Using rule $R(v)$ we have $Q_{R(v)}(i) \leqq 0$. However from theorem 1, there exists a rule $R^*(v) \varepsilon C''$ such that $Q_{R^*(v)}(i) \leqq 0$, which implies $\psi_{R^*(v)}(i) \leqq m_v$. Since C'' is a finite class there is a subsequence $v_1, v_2, \cdots$ such that $R^* = R^*(v_i)$, $i = 1, 2, \cdots$. Thus $\psi_{R^*}(i) \leqq \lim_{i\to\infty} m_{v_i} = M$. Setting $R_3 = R^*$ the theorem is proved.

References

1. Chung, Kai Lai, *Markov Chains with Stationary Transition Probabilities*, Springer, Berlin, 1960.
2. Derman, C., "On Optimal Replacement Rules When Changes of State are Markovian," to appear in a *Project Rand Report on Optimal Decision Processes*.
3. Derman, C., M. V. Johns, and G. J. Lieberman, "Continuous Sampling Procedures Without Control," *Annals of Mathematical Statistics*, December, 1959.

4. Eaton, J. H. and L. A. Zadeh, "Optimal Pursuit Strategies in Discrete-State Probabilistic Systems," to appear *ASME Transactions* (*Journal of Basic Engineering*).
5. Elfing, G., "The AOQL of Multi-Level Continuous Sampling Plans," *Technical Report No. 48, Applied Mathematics and Statistics Laboratories*, Stanford University, Stanford, California, December, 1960.
6. Howard, Ronald A., *Dynamic Programming and Markov Processes*, John Wiley, New York, 1960.
7. Karlin, S., "Structure of Dynamic Programming," *Naval Research Logistics Quarterly*, Vol. 2, 1955.
8. Klein, M., "Inspection—Maintenance—Replacement Schedules Under Markovian Deterioration," *Management Science*, Vol. 9, No. 1, (October 1962).
9. Manne, Alan S., "Linear Programming and Sequential Decisions," *Management Science*, Vol. 6, No. 3, 1960).
10. Pollack, Maurice and Walter Wiebenson, "Solutions of the Shortest-Route Problem—A Review," *Operations Research*, Vol. 8, March-April 1960, pp. 224–230.
11. Wagner, Harvey M., "On the Optimality of Pure Strategies," *Management Science*, Vol. 6, No. 3, (April 1960).
12. Widder, D. V., *The Laplace Transform*, Princeton University Press, Princeton, N. J., 1946.
13. Freimer, M., "On Solving a Markovian Decision Problem by Linear Programming," (Unpublished Paper, Institute for Defense Analysis, Cambridge, Mass., 12 Dec. 1961, rev. 16 March 1961).