

STOCHASTIC FINITE-STATE SYSTEMS IN CONTROL THEORY

L. A. ZADEH (*)

1 - INTRODUCTION -

Much of the current research in the theory of control tends to be concerned with the determination of optimal policies under specified performance criteria. This mode of formulation of control problems was initiated principally by Bellman [1] in his work on dynamic programming, and it has come into increasingly wide use since the enunciation of the "maximum principle" by Pontryagin, Boltyanskii and Gamkrelidze [2].

In general, the determination of an optimal policy presents difficult analytical problems when the system to be optimized has a continuum of states. Indeed, in the case of the minimum-time problem, no explicit expression for the optimal policy is available even in the simple case of a linear, continuous-time time-invariant system. However, under favorable conditions, iterative and steepest descent techniques such as those developed by L. Neustadt [3], A. Bryson [4], H. Kelley [5], Y. C. Ho [6], J. H. Eaton [7] and others are quite effective when used in conjunction with large scale digital computers.

For a variety of reasons some of which have to do with the nature of current applications of optimal control in space research and others are largely a matter of tradition, much of the literature of control theory has been and still is concerned with continuous-state continuous-time systems whose behavior is characterized by a set of ordinary differential equations. Within the past few years, however, there appeared a trend toward the formulation of some control problems in the context of finite-state systems, that is, systems whose state space consists of a finite number of points rather than a continuum. Such systems are usually more readily amenable to analysis and simulation by digital computers than continuous-state systems, and are particularly well-suited for the application of the techniques of dynamic programming [8]. By contrast, the techniques of the calculus of variations are generally not effective in dealing with finite-state systems on account of the lack of requisite continuity and differentiability properties of the performance functionals.

In what follows, we shall focus our attention on a few basic control problems in the context of stochastic finite-state systems, that is, finite-state systems in which the transitions from one state to another are governed probabilistically rather than deterministically by the input sequence. Such systems - regarded as information channels - were introduced by Shannon in his classical paper on the mathematical theory of communication [9]. However, within the past decade the concept of a stochastic system has found applications in a number of other fields, among them learning theory, econometrics, game theory and, more recently, control theory. Before proceeding to discuss some of its applications in control theory, we shall define what is meant by a stochastic finite-state system and examine some of their elementary properties.

2 - PRELIMINARIES CONCERNING STOCHASTIC FINITE-STATE SYSTEMS -

Informally, by a stochastic finite-state system A we mean a system having the following characteristics : (a) The input and output of A have the form of sequences of symbols drawn from a finite alphabet ; more specifically, the input and output sequences are denoted by $u_0 u_1 u_2 \dots u_t \dots$ and $y_0 y_1 y_2 \dots y_t \dots$, respectively, with u_t and y_t ranging over a fixed alphabet $\{a_1, \dots, a_m\}$ independent

(*) Professor of Electrical Engineering, University of California, Berkeley, California ; presently on leave at the Research Laboratory of Electronics, MIT, Cambridge, Mass. This work was supported in part by the National Science Foundation.

of t ($t = 0, 1, 2, \dots$). For convenience, we shall assume that the alphabet consists of just two symbols, 0 and 1. (b) The state of A at time t , s_t , ranges over a finite set (state space) $1, \dots, N$, which is the same for all t . (c) The input at time t , the output at time t , the state at time t and the state at time $t + 1$ are related by an equation of the form :

$$(y_t, s_{t+1}) = f(s_t, u_t, r_t) \quad , \quad t = 0, 1, 2, \dots \quad (1)$$

where (y_t, s_{t+1}) denotes a pair of values of y_t and s_{t+1} ; f is a non-random function of its arguments (for simplicity we assume that f is independent of t , which implies that we restrict ourselves to time-invariant systems) ; and the r_t are identically distributed random variables. (d) The r_t and f in (1) are such that the state s_t has the separation (Markoff) property expressed by the relation :

$$p(s_{t+1}, y_t | s_t, u_t, \text{past states, past outputs, past and future inputs}) = p(s_{t+1}, y_t | s_t, u_t) \quad (2)$$

for all t , all inputs, outputs and states, where the symbol $p(\alpha|\beta)$ is an abbreviation for the probability that a random variable α will assume a value α^* , given a value of the random variable β ; "past states" stands for the states s_{t-1}, s_{t-2}, \dots ; "past outputs" stands for y_{t-1}, y_{t-2}, \dots ; and "past and future inputs" stands for u_{t-1}, u_{t-2}, \dots , and u_{t+1}, u_{t+2}, \dots .

In plain words, (2) implies that, given the present state, s_t , and the present input, u_t , the knowledge of the past states, outputs and inputs as well as future inputs has no influence on the present output, y_t , and the next state, s_{t+1} .

It is easy to show that a sufficient condition for (2) to hold is that r_0, r_1, \dots be independent random variables. Specifically, we note that the left-hand member of (2) is given by :

$$\Pr \{ r_t \in R | \text{past states, past outputs, past and future inputs} \} \quad (3)$$

where R is the set of values of r_t satisfying the relation $(s_{t+1}, y_t) = f(s_t, u_t, r_t)$ for fixed s_{t+1}, y_t, s_t and u_t . Clearly, in view of the assumed independence of the r_t , $t = 0, 1, 2, \dots$, (3) is numerically equal to $\Pr \{ r_t \in R \}$, which in turn is equal to right-hand member of (2). Q. E. D.

Henceforth we shall assume that the r_t are independent, identically distributed random variables. Then (1) implies that A is completely characterized by the conditional probability distribution function :

$$p(s_{t+1}, y_t | s_t, u_t) \quad (4)$$

It will be convenient for our purposes to write this function as $p_{i,u}(j,y)$, with the understanding that it represents the conditional probability that the state at time $t + 1$ is j and the output at time t is y , given that the state at time t is i and the input at time t is u , $i, j = 1, \dots, N$, $t = 0, 1, \dots$. Thus, A is characterized by the totality of values which p assumes as i and j vary over the set $\{1, \dots, N\}$ and u and y vary over the alphabet $\{a_1, \dots, a_m\}$.

Following J. Carlyle [10], let $P(y|u)$ denote an $N \times N$ matrix whose i, j th element is $p_{i,u}(j,y)$. Then, A is characterized by the set of matrices $\{P(y|u) ; u, y = a_1, \dots, a_m\}$. In terms of these matrices, it is easy to give an expression for the conditional probability of A being in state j at the end of an output sequence $y_0 \dots y_t$ when A is initially (at $t = 0$) in state i and an input sequence $u_0 \dots u_t$ is applied to A . Specifically, let this probability be denoted by $p_{i,u_0 \dots u_t}(j, y_0 \dots y_t)$ and let $P(y_0 \dots y_t | u_0 \dots u_t)$ or $P(\bar{y} | \bar{u})$ for short denote the $N \times N$ matrix whose i, j th element is $p_{i,u_0 \dots u_t}(j, y_0 \dots y_t)$. Then we have the identity :

$$P(y_0 \dots y_t | u_0 \dots u_t) = P(y_0 | u_0) \dots P(y_t | u_t) \quad (5)$$

whose right-hand member is the matrix product of $P(y_0 | u_0), \dots, P(y_t | u_t)$.

The manner in which this identity can be established in the general case can readily be seen by considering the special case of $t = 1$. Here the i, j th element of the product $P(y_0 | u_0) P(y_1 | u_1)$ is given by the sum :

$$\sum_{k=1}^N p_{i,u_0}(k, y_0) p_{k,u_1}(j, y_1) \quad (6)$$

(*) For simplicity, we use the same symbol to denote both a random variable and its value.

Now by virtue of the separation property (2), we can write :

$$p_{i,u_0}(k, y_0) = p_{i,u_0 u_1}(k, y_0)$$

$$p_{k,u_1}(j, y_1) = p_{k,y_0, i, u_0 u_1}(j, y_1)$$

and hence :

$$p_{i,u_0}(k, y_0) p_{k,u_1}(j, y_1) = p_{i,u_0 u_1}(k, y_0) p_{k,y_0, i, u_0 u_1}(j, y_1)$$

$$= p_{i,u_0 u_1}(k, j, y_0, y_1) \quad (7)$$

which upon summing on k reduces to :

$$p_{i,u_0 u_1}(j, y_0, y_1) \quad (8)$$

Since this is by definition the i, j^{th} element of $P(y_0, y_1 | u_0, u_1)$, we have established the desired identity, namely :

$$P(y_0, y_1 | u_0, u_1) = P(y_0 | u_1) P(y_1 | u_1) \quad (9)$$

From the probability distribution $p_{i,u}(j, y)$ we can obtain the marginal distributions $p_{i,u}(j)$ and $p_{i,u}(y)$ by summing on y and j , respectively. These distributions define the following matrices :

$$P(u) \triangleq \sum_y P(y|u) \quad (10)$$

$$q(y|u) \triangleq P(y|u) \mathbf{1} \quad (11)$$

where \triangleq stands for "is defined to be" and $\mathbf{1}$ is the column N -vector $(1, \dots, 1)$. Stated explicitly, $P(u)$ is an $N \times N$ matrix whose i, j^{th} element is the conditional probability that the state of A at time $t+1$ is j given that $u_t = u$ and $s_t = i$. As for $q(y|u)$, it is an N -vector whose i^{th} component is the conditional probability that the output of A at time t is y , given $s_t = i$ and $u_t = u$.

Example - A simple example of a stochastic system having two states is shown in figure 1. The figure in question represents the state diagram of A , with the circles marked 1 and 2 denoting the

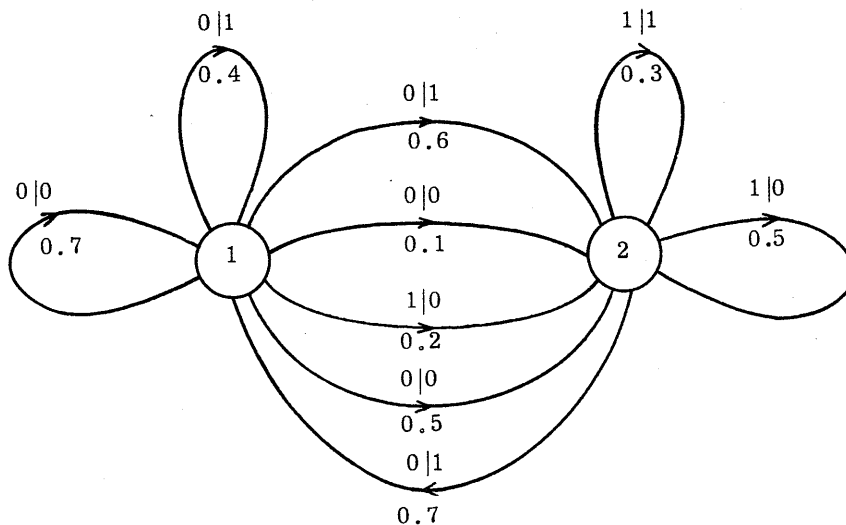


Figure 1 - An example of a stochastic finite-state system with two states and binary alphabet.

two states of A, and the lines connecting the states denoting those transitions that have positive probability. E.g., the line starting on 1 and terminating on 2 and labeled 1|0, 0.2 signifies that $p_{1,0}(2,1) = 0.2$; that is, the conditional probability that A starting in 1 and subjected to the input 0 would go into state 2 and produce the output 1 is 0.2.

The expressions for $P(y|u)$, $P(u)$ and $q(y|u)$ read :

$P(0 0)$		$P(1 0)$		$P(0 1)$		$P(1 1)$	
S_{t+1}		S_{t+1}		S_{t+1}		S_{t+1}	
S_t		S_t		S_t		S_t	

$$P_{\lambda}(y|u) = \langle \lambda, q(y|u) \rangle \quad (13)$$

$$\lambda_{t+1} = \lambda_t P(u_t) \quad (\lambda_t \text{ is a row vector}) \quad (14)$$

The first of these relations plays a key role in defining the notion of equivalence. Specifically, consider two stochastic systems A^1 and A^2 having identical input and output alphabets, and let A^1 and A^2 be characterized by the matrices $P^1(y|u)$ and $P^2(y|u)$ respectively. Furthermore, let λ^1 and λ^2 be distribution states associated with A^1 and A^2 (the number of states, N^1 , in the state space of A^1 need not be the same as the number of states, N^2 , in the state space of A^2). Then λ^1 and λ^2 are said to be equivalent distribution states, written as $\lambda^1 \simeq \lambda^2$, if and only if :

$$P_{\lambda^1}(\bar{y}|\bar{u}) = P_{\lambda^2}(\bar{y}|\bar{u}) \quad (15)$$

for all sequences \bar{u} and \bar{y} (of equal length). Clearly, in the special case where A^1 and A^2 are identical systems, (15) defines the equivalence between two distinct distribution states of a single system.

If instead of requiring that (15) hold for all input sequences \bar{u} and all output sequences \bar{y} , we require only that (15) hold for all input and output sequences of length k , then λ^1 and λ^2 are k-equivalent distribution states. More specifically, on denoting a sequence \bar{v} of length k by the symbol \bar{v}^k , and denoting k -equivalence by $\overset{k}{\simeq}$, we can write compactly :

$$\lambda^1 \overset{k}{\simeq} \lambda^2 \iff P_{\lambda^1}(\bar{y}^k|\bar{u}^k) = P_{\lambda^2}(\bar{y}^k|\bar{u}^k) \quad \text{for all } \bar{u}^k, \bar{y}^k. \quad (16)$$

It is evident that if λ^1 and λ^2 are k -equivalent, then they are $(k-1)$ -equivalent, $(k-2)$ -equivalent, \dots , 1 -equivalent, for if (15) holds for all sequences of length k , then it holds a fortiori for all sequences of shorter length.

In terms of equivalent distribution states, the equivalence between two stochastic systems A^1 and A^2 is defined as follows : A^1 and A^2 are equivalent, written as $A^1 \equiv A^2$, if and only if to every distribution state in Σ^1 (see 12) there is an equivalent distribution state in Σ^2 , and vice-versa. The task of verifying that A^1 and A^2 are equivalent is greatly simplified by observing that if $\lambda_1, \dots, \lambda_{n-1}$ is a basis for Σ^1 (i. e., a set of linearly independent vectors in Σ^1 such that every distribution state in Σ^1 can be expressed as a linear combination of $\lambda_1, \dots, \lambda_{n-1}$) then every λ^1 in Σ^1 has an equivalent distribution state in Σ^2 if each of the basis vectors has an equivalent distribution state in Σ^2 . In particular, if to every state of A^1 (i. e., to every distribution state which assign probability 1 to a state of A^1 and zero probabilities to other states of A^1) corresponds an equivalent distribution state in A^2 , then to every distribution state in Σ^1 corresponds an equivalent distribution state of Σ^2 .

Returning to the discussion of k -equivalent distribution states, let q^1, \dots, q^r denote the distinct N -vectors $q(\bar{y}^k|\bar{u}^k)$ corresponding to all possible \bar{y}^k and \bar{u}^k which can be associated with a stochastic system A characterized by $P(y|u)$, and let L_k be a subspace of \mathbb{R}^N (space of N -tuples of real numbers) which is spanned by q^1, \dots, q^r . Furthermore, let λ and μ be two distribution states in Σ . Then by (13) and (16) :

$$\lambda \overset{k}{\simeq} \mu \iff \langle (\lambda - \mu), q^\rho \rangle = 0, \rho = 1, \dots, r. \quad (17)$$

which implies that λ and μ are k -equivalent if and only if the vector $\lambda - \mu$ lies in the orthogonal complement of L_k . Thus, if Q_μ is the set of all distribution states equivalent to μ :

$$Q_\mu = \{ \lambda : \langle (\lambda - \mu), q^\rho \rangle = 0, \rho = 1, \dots, r \}$$

then Σ can be reduced mod Q_μ .

Example - For the system considered at the end of Section 2, $\lambda = (\lambda_1, \lambda_2)$ and $\mu = (\mu_1, \mu_2)$ are 1-equivalent if and only if :

$$0.8 (\lambda_1 - \mu_1) + 0.5 (\lambda_2 - \mu_2) = 0$$

$$0.2 (\lambda_1 - \mu_1) + 0.2 (\lambda_2 - \mu_2) = 0$$

$$(\lambda_1 - \mu_1) + 0.7 (\lambda_2 - \mu_2) = 0$$

$$(\lambda_2 - \mu_2) = 0$$

Clearly, in this case there do not exist 1-equivalent distribution states. (We exclude of course the trivial case $\lambda = \mu$).

Let A be a stochastic system with N states. A basic property of equivalent distribution states of A which was established by J. Carlyle may be stated compactly in the form of an implication :

$$\lambda \stackrel{N-1}{\approx} \mu \iff \lambda \approx \mu \quad (19)$$

which means that if two distribution states are (N-1)-equivalent, then they are equivalent. This result is an extension to stochastic systems of an analogous property of deterministic systems which was first stated by E.F. Moore [11]. It can be proved in a manner similar to that employed by Moore, but a simpler way is to demonstrate first that if $L_{k+1} = L_k$ then $L_{k+2} = L_{k+1}$ and hence k-equivalence implies equivalence. We shall forego the details, which can be found in [10].

Following this brief introduction to stochastic finite-state systems, we turn our attention to some of their applications in control theory. More specifically, we shall focus our discussion on so-called optimum pursuit problems, in which the pursuer's aim is to intercept a moving target in the shortest possible time.

4 - OPTIMAL PURSUIT PROBLEMS -

Consider a situation in which the motions of both the target and its pursuer can be described as stochastic transitions from one state to another in a finite-state system. Let ρ_t and τ_t denote, respectively, the states of the pursuer and the target at time t, with both ρ_t and τ_t ranging over $1, \dots, N$. We assume that the motion of the pursuer is characterized by a conditional probability distribution function $p_{i,u}^{\rho}(j)$, $i, j = 1, \dots, N$, where $i = \rho_t$, $j = \rho_{t+1}$ and u is the input to the pursuer at time t, with u ranging over a finite set, say 0,1. As for the target, we have to distinguish between two cases. (a) The target does not try to evade the pursuer (non-competitive pursuit problem), in which case we assume that the motion of the target is characterized by a conditional probability distribution function $p_i^{\tau}(j)$, where $i = \tau_t$ and $j = \tau_{t+1}$. (b) The target tries to evade the pursuer (competitive control), in which case we assume that its motion is characterized by a conditional probability distribution function $p_{i,v}^{\tau}(j)$, where v is the input to the target at time t, with v like u ranging over a finite set, say 0,1.

By a policy for the pursuer we mean an N-vector $\pi^{\rho} = (\pi_1^{\rho}, \dots, \pi_N^{\rho})$ whose i^{th} component is the input which the pursuer applies in state i. Similarly, in Case (b), by a policy for the target we mean an N-vector $\pi^{\tau} = (\pi_1^{\tau}, \dots, \pi_N^{\tau})$ whose i^{th} component is the input which the target applies in state i. As an illustration, if u and v range over the set $\{0,1\}$, then π^{ρ} and π^{τ} are vectors whose components are 0 or 1, e.g., $\pi^{\rho} = (0,1,0,1,1,0)$ for $N = 6$. Note that both π^{ρ} and π^{τ} are assumed to be stationary, that is, independent of time. Furthermore, we restrict ourselves to pure strategies.

For fixed π^{ρ} and π^{τ} , the time it takes the pursuer to intercept the target is a random variable. The expected value of this random variable is denoted by $C(\pi^{\rho})$ in Case (a) and $C(\pi^{\rho}, \pi^{\tau})$ in Case (b). In Case (a), the objective of the pursuer is to intercept the target in the shortest possible time (or, equivalently, to reach the target in the smallest number of steps). In Case (b), the objective of the target is to maximise $C(\pi^{\rho}, \pi^{\tau})$ and, if possible, make it infinite.

The problem of interception in the non-competitive case can readily be reduced to the more conventional problem of reaching a specified fixed set of states in the shortest possible time. Specifically, consider a composite stochastic system whose states are pairs of the form (ρ_t, τ_t) , with ρ_t and τ_t ranging over the set $\{1, \dots, N\}$ and the transition probability from a state say (i,k) at time t to a state (j,l) at time $t+1$ expressed by the product :

$$p_{i,u}^{\rho}(j) p_k^{\tau}(l) \quad (20)$$

where u is the input applied by the pursuer. Note that in expressing this probability as a product we are tacitly assuming that the random variables which influence the motion of the pursuer are independent of those which influence the motion of the target.

The interception of the target by the pursuer takes place when the composite system enters a state in the set of N states $\{(1,1), \dots, (N,N)\}$. Let this set of states of the composite system be denoted by Γ . Then, in terms of the composite system the problem of interception reduces to that of reaching a specified set of states Γ from a given initial state (i,k) .

Based on this observation, we can focus our attention on the problem of taking a stochastic system A from a specified initial state to a prescribed set of terminal states Γ . In Case (a), the objective is to do this in the shortest possible time. In Case (b) on the other hand, there is a conflict between the objective of the pursuer, which is to reach Γ in the shortest possible time, and that of the target, which is to reach Γ in the longest possible time.

A method for determining an optimal policy for the pursuer in Case (a) was described in a paper by J. H. Eaton and the author [12]. The essence of this method can be summarized as follows. Let A be a stochastic system with N states $1, \dots, N$ of which the states $n+1, \dots, N$ are designated to belong to Γ . The policy, then, is an n-vector $\pi = (\pi_1, \dots, \pi_n)$, since the process stops once A reaches a state in Γ . Furthermore, since A is characterized by the conditional probability distribution function $p_{i,j}(\pi)$, once π is specified the transitions of A from one state to another form a Markoff process characterized by the transition probability matrix $P(\pi)$ whose i, j^{th} element is given by $p_{i, \pi_i}(j)$.

Let $X(\pi) = (X_1, \dots, X_n)$ denote a vector, called the expected time vector, whose i^{th} component is the expected value of the time needed to reach Γ from i , $i = 1, \dots, n$, using policy π . If π is a proper policy, that is, a policy under which there is a positive probability of reaching Γ from $1, \dots, n$, then it follows at once from the definition of the expected time vector that $X(\pi)$ satisfies the relation :

$$X(\pi) = P(\pi) X(\pi) + 1 \quad (21)$$

where 1 is an n-vector with unity components.

A policy π^0 is defined to be optimal if for all π :

$$X(\pi^0) < X(\pi) \quad (22)$$

where (22) means that every component of $X(\pi^0)$ is less than or equal to the corresponding component of $X(\pi)$. For convenience, $X(\pi^0)$ will be denoted by X^0 . Thus :

$$X^0 = \underset{\pi}{\text{Min}} X(\pi) \quad (23)$$

if this minimum exists. (Note that the existence of this minimum cannot be taken for granted since $X(\pi)$ is a vector and the relation $<$ between vectors defines a partial rather than complete ordering).

As shown in [12], X^0 does exist if there exists at least one proper policy. Then, on applying the operator $\underset{\pi}{\text{Min}}$ to both sides of (21) and observing that (i) in the right-hand member of (21) only π , is present in the first row, π_2 in the second row, etc. ; and (ii) :

$$\underset{\pi}{\text{Min}} [p(\pi) X(\pi) + 1] = \underset{\pi}{\text{Min}} [P(\pi) X^0 + 1] \quad (24)$$

we obtain the functional equation :

$$X^0 = \underset{\pi}{\text{Min}} (P(\pi) X^0 + 1) \quad (25)$$

which is satisfied by X^0 .

It is convenient to write this equation in the symbolic form :

$$X^0 = T(X^0) \quad (26)$$

where T is a nonlinear mapping from R^n (space of n-tuples of real numbers) into R^n defined by the right-hand member of (25). This places in evidence the fact that X^0 is a fixed point of T, which, as is shown in [12], exists and is unique.

The fixed point of T can be determined by starting with an arbitrary n-vector X and forming the sequence $T(X), T^2(X), T^3(X), \dots$ which converges to X^0 . If X is an arbitrary expected time vector (i. e., if X satisfies (21) for some π) then it is easy to show that the above sequence is monotone. An alternative way of determining X^0 is furnished by the "policy improvement" technique of R. A. Howard [13]. In this case, the elements of the approximating sequence to X^0 are expected time vectors, whereas the elements of the sequence $T(X), T^2(X), \dots$ do not, in general, have this pro-

party. Actually, in many real-life control applications it may be practicable to determine the solution of (25) by an analog circuit employing diodes and resistors. A discussion of analog techniques of this type may be found in the book by Dennis [14].

5 - COMPETITIVE CASE -

In the more general case where the motion of the target is controlled by an input v through the conditional probability distribution function $p_{i,v}(j)$, we have what is known in game theory as a stochastic game [15]. The corresponding reduced problem differs from that considered in the preceding section in the following respects. (1) The transition probability from i to j depends on both u and v , rather than on u alone; thus, it has the form $p_{i,u,v}(j)$ rather than $p_{i,u}(j)$. (2) The optimal policies for the pursuer and the target are in general mixed rather than pure. This means that a policy for the pursuer will in general have the form of a $n \times 2$ matrix π^p whose i^{th} row is a 2-vector p^i whose first and second components are respectively, the probabilities of choosing $u = 0$ and $u = 1$ when A is in state i . Similarly, a policy for the target will have the form of an $n \times 2$ matrix π^t whose i^{th} row is a 2-vector q^i whose first and second components are, respectively, the probabilities of choosing $v = 0$ and $v = 1$ when A is in state i . (3) Let P^{ij} denote the 2×2 matrix defined below:

$$P^{ij} = \begin{bmatrix} p_{i,0,0}(j) & p_{i,0,1}(j) \\ p_{i,1,0}(j) & p_{i,1,1}(j) \end{bmatrix} \quad (27)$$

where $p_{i,u,v}(j)$ is the conditional probability of a transition from i to j when the pursuer applies u and the target applies v , and let C^i denote a 2×2 matrix whose 1,1 element is equal to 1-probability of reaching Γ from i when the pursuer applies 0 and the target applies 0 at i , and similarly for other elements of C^i . In terms of these matrices, the analog of (21) for the competitive case reads:

$$X_i = 1 + \sum_{j=1}^n p^i P^{ij} q^j X_j, \quad i = 1, \dots, n \quad (28)$$

where X_i denotes the expected time of reaching Γ from i when the pursuer chooses its input at i in accordance with the probability distribution p^i ($p^i =$ row vector with two components, $p_1^i = \text{Prob}\{u = 0\}$ and $p_2^i = \text{Prob}\{u = 1\}$) and the target chooses its input at i in accordance with the probability distribution q^i ($q^i =$ column 2-vector with $q_1^i = \text{Prob}\{v = 0\}$ and $q_2^i = \text{Prob}\{v = 1\}$). (4) Whereas in the non-competitive case an optimal policy for the pursuer is one that minimizes X , in the competitive case optimal policies for the pursuer and the target yield a saddle (minimax) point for X , that is, a point (π_0^p, π_0^t) such that:

$$X(\pi_0^p, \pi^t) \leq X(\pi_0^p, \pi_0^t) \leq X(\pi^p, \pi_0^t) \quad (29)$$

for all π^p and π^t , where we write $X(\pi^p, \pi^t)$ to exhibit the dependence of X on the policies of the pursuer and the target.

Finally, in the competitive case the analog of the functional equation (25) is the set of functional equations:

$$X_i^o = \text{Min}_{p^i} \text{Max}_{q^i} (1 + \sum_{j=1}^n p^i P^{ij} q^j X_j), \quad i = 1, \dots, n. \quad (30)$$

or more compactly:

$$X^o = T(X^o) \quad (31)$$

where $X^o = (X_1^o, \dots, X_n^o)$ is the expected time vector and T is a mapping of R into itself defined by the relation:

$$Y = T(X) \iff Y_i = \text{Min}_{p^i} \text{Max}_{q^i} (1 + \sum_{j=1}^n p^i P^{ij} q^j X_j) \quad (32)$$

$$i = 1, \dots, n.$$

To illustrate how (31) can be solved for X^o , we shall consider a simple special case where $C^i > 0$ for $i = 1, \dots, n$ (i.e., all the elements of C^i are positive for each i). This case as such is

not of practical interest since it implies that Γ is reachable in one step from every state of the system. However, the method used to deal with this case can readily be extended to the more general case where Γ is reachable with positive probability, but not necessarily in one step, from every state i , $i = 1, \dots, n$.

Essentially, we shall show that T is a contraction mapping, which implies that (31) has a unique fixed point X^0 . For this purpose, we define the norm of a vector X in R^n by the relation :

$$\|X\| = \text{Max}_i |X_i|, \quad i = 1, \dots, n.$$

Then, on letting X' and X'' be two arbitrary vectors in R^n and denoting by Y' and Y'' their respective images, we have :

$$\begin{aligned} \|Y' - Y''\| &= \text{Max}_i |Y'_i - Y''_i| \\ &= \text{Max}_i \left| \text{Min}_{p^i} \text{Max}_{q^i} Z'_i - \text{Min}_{p^i} \text{Max}_{q^i} Z''_i \right| \end{aligned} \quad (33)$$

where :

$$Z'_i = 1 + \sum_{j=1}^n p^i P^{ij} q^i X'_j \quad (34)$$

and :

$$Z''_i = 1 + \sum_{j=1}^n p^i P^{ij} q^i X''_j \quad (35)$$

Now for any functions of p^i we can write :

$$\left| \text{Min}_{p^i} f(p^i) - \text{Min}_{p^i} g(p^i) \right| \leq \text{Max}_{p^i} |f(p^i) - g(p^i)| \quad (36)$$

and similarly :

$$\left| \text{Max}_{q^i} f(q^i) - \text{Max}_{q^i} g(q^i) \right| \leq \text{Max}_{q^i} |f(q^i) - g(q^i)| \quad (37)$$

On applying these inequalities to (33), we find :

$$\begin{aligned} \|Y' - Y''\| &\leq \text{Max}_i \text{Max}_{p^i} \text{Max}_{q^i} |Z'_i - Z''_i| \\ &\leq \text{Max}_i \text{Max}_{p^i} \text{Max}_{q^i} p^i \left(\sum_{j=1}^n P^{ij} (X'_j - X''_j) \right) q^i \\ &\leq \text{Max}_i \text{Max}_{p^i} \text{Max}_{q^i} p^i \left(\sum_{j=1}^n P^{ij} \|X' - X''\| \right) q^i \\ &\leq \|X' - X''\| \text{Max}_i \text{Max}_{p^i} \text{Max}_{q^i} p^i \left(\sum_{j=1}^n P^{ij} \right) q^i \end{aligned} \quad (38)$$

Now by virtue of the assumption about the C^i , we have :

$$\sum_{j=1}^n P^{ij} \leq (1 - \epsilon) E \quad (39)$$

where E is an $n \times n$ matrix whose elements are unity and ϵ is a positive number less than unity. Therefore :

$$\begin{aligned} \|Y' - Y''\| &\leq (1 - \epsilon) \|X' - X''\| \text{Max}_i \text{Max}_{p^i} \text{Max}_{q^i} p^i E q^i \\ &\leq (1 - \epsilon) \|X' - X''\| \end{aligned} \quad (40)$$

Since :

$$\text{Max}_i \text{Max}_{p^i} \text{Max}_{q^i} p^i E q^i = 1$$

The inequality (40) shows that T is a contraction, and hence by the Banach fixed point theorem [16] (31) has a unique fixed point X° . As in the non-competitive case, this fixed point can be determined by starting with an arbitrary point X in R^n and forming the sequence $T(X)$, $T^2(X)$,... which converges to X° .

This concludes our brief discussion of the competitive case as well as our sketchy exposition of some of the elementary properties of stochastic finite-state systems. It is very likely that such systems will be playing an increasingly important role in control theory and it will not be long before we know much more about them than we do at the present time.

REFERENCES

- [1] R. E. BELLMAN - "On the theory of dynamic programming", Proc. Nat. Acad. Sci., Vol. 38, pp. 716-719, 1952.
- [2] V. G. BOLTYANSKII, R. V. GAMKRELIDZE and L. S. PONTRYAGIN - "On the theory of optimal processes", Dokl. Akad. Nauk SSSR, (N. S.), Vol. 110, pp. 7-10, 1956.
- [3] L. W. NEUSTADT - "Synthesizing time optimal control systems", Jour. Math. Analysis and Appl., Vol. 1, No. 4, pp. 484-493, 1960.
- [4] A. E. BRYSON and W. F. DENHAM - "A Steepest-ascent method for solving optimum programming problems", Jour. of Appl. Mech., Vol. 29, Series E, pp. 247-257, June 1962.
- [5] H. J. KELLEY - "Gradient theory of optimal flight paths", Jour. of the Amer. Rocket Soc., Vol. 30, pp. 947-953, 1960.
- [6] Y. C. HO - "A successive approximation technique for optimal control systems subject to input saturation", Jour. of Basic Eng., Vol. 84, Series D, No. 1, pp. 33-40, 1962.
- [7] J. H. EATON - "An iterative solution to time-optimal control", Jour. of Math. Analysis and Appl., Vol. 5, pp. 329-344, Octobre 1962.
- [8] L. A. ZADEH - "Optimal control problems in discrete-time systems", Computer Control Systems Technology (edited by C. T. Leondes), McGraw-Hill Book Co., pp. 389-414, 1961.
- [9] C. E. SHANNON - "The mathematical theory of communication", Bell Syst. Tech. Jour., Vol. 27, pp. 379-423 and 623-656, 1948.
- [10] J. CARLYLE - "Equivalent stochastic sequential machines", Tech. Report No. 415, Electronics Research Lab., Univ. of Calif., Berkeley, Calif., 1961. To be published in the Jour. of Math. Analysis and Appl.
- [11] E. F. MOORE - "Gedanken experiments on sequential machines", Automata Studies, Annals of Mathematical Studies No. 34, Princeton Univ. Press, pp. 129-153, 1956.
- [12] J. H. EATON and L. A. ZADEH - "Optimal pursuit strategies in discrete-state probabilistic systems", Proc. Joint Automatic Control Conference, Paper No. 61-JAC-11, June 1961. Published in Jour. Basic Eng., Vol. 84, Series D, pp. 23-29, March 1962.
- [13] R. A. HOWARD - "Dynamic programming and Markov processes", Technology Press and Wiley, New-York, N. Y., 1960.
- [14] J. DENNIS - "Mathematical programming and electrical networks", Technology Press and Wiley, New-York, N. Y., 1959.
- [15] L. S. SHAPLEY - "Stochastic games", Proc. Nat. Acad. Sci., Vol. 39, pp. 1095-1100, 1953.
- [16] A. N. KOLMOGOROV and S. V. FOMIN - "Elements of the theory of functions and functional analysis", Graylock Press, Rochester, N. Y., 1957.