MANAGEMENT SCIENCE
Vol. 28, No. 1, January 1982
Printed in U.S.A.

State of the Art

A SURVEY OF PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES: THEORY, MODELS, AND ALGORITHMS*

GEORGE E. MONAHAN†

This paper surveys models and algorithms dealing with partially observable Markov decision processes. A partially observable Markov decision process (POMDP) is a generalization of a Markov decision process which permits uncertainty regarding the state of a Markov process and allows for state information acquisition. A general framework for finite state and action POMDP's is presented. Next, there is a brief discussion of the development of POMDP's and their relationship with other decision processes. A wide range of models in such areas as quality control, machine maintenance, internal auditing, learning, and optimal stopping are discussed within the POMDP-framework. Lastly, algorithms for computing optimal solutions to POMDP's are presented.

(MARKOV DECISION PROCESSES; PARTIALLY OBSERVABLE; SURVEY)

1. Introduction

This paper surveys models and algorithms dealing with partially observable Markov decision processes (POMDP's). A POMDP is a generalization of a Markov decision process (MDP) which permits uncertainty regarding the state of a Markov process and allows state information acquisition. Howard [25] described movement in an MDP as a frog in a pond jumping from lily pad to lily pad. Adapting Vazsonyi's [72] analogy in a discussion of stochastic automata, we can view the setting of a POMDP as a fog shrouded lily pond. The frog is no longer certain about which pad it is currently on. Before jumping, the frog can obtain information about its current location. In the following paragraphs we will show that the generalization is nontrivial and admits a wide range of important decision problems arising in many contexts.

The generalization of MDP's to POMDP's is significant in problem settings where state uncertainty is a central issue that can not be assumed away. Examples in such diverse areas as machine maintenance, quality control, learning theory, internal auditing, optimal stopping, and others given in Section 4 illustrate the wide range of problems that can be modeled as POMDP's. The key feature of all these models is the presence of state uncertainty and its impact on the optimal choice of actions. It will be shown that such uncertainty can often have surprising consequences on the structure of optimal decision rules.

Partially observable models are typically more difficult to analyze than their MDP counterparts. The added generality is not a free good. In many applications incorporating the theory of (perfectly observable) MDP's, a primary goal of the analysis is to determine structural properties of the optimal policy (and the optimal value function). (See, e.g., Heyman and Sobel [23, Chapters VI–VII].) Economically appealing assump-

*Accepted by Marcel F. Neuts, former Departmental Editor; received November 6, 1980. This paper has been with the author 1 month for 1 revision.

†Georgia Institute of Technology.

0025-1909/82/2801/0001$01.25

tions regarding elements of the model are translated into intuitively appealing structural results. As a classic example, convexity of the one period cost function in a stochastic multi-period inventory problem with setup costs yields (s, S)-type optimal ordering policies [28]. In perfectly observable machine replacement problems conditions on the one-period transition matrix, equivalent to first-order stochastic dominance, translates into repair policies that are characterized by a single parameter, i.e., are of the control-limit type [15].

In POMDP models structural results such as those suggested above are much more difficult to obtain. All of the intricacies found in perfectly observable sequential decision problems remain. Additional complications are added due to two sources of potential error in determining the current underlying (core) state. The first is the uncertainty about the initial state of the unobservable process. The second and most significant is the possible error in the information regarding the underlying state of the process. The presence of the two types of uncertainty in the model typically destroys structural properties. In most applications of POMDP's, structured policies are optimal only when it is possible to obtain perfect core state information. When there are imperfect observations the structure of the optimal policy is invariably lost. Examples of this phenomenon in very simple two-state models are given in §4.

The lack of structure of many optimal policies in POMDP models may not be a serious operational issue. Of course the use of structural results in algorithms for computing optimal policies is ruled out. There may, however, be sub-optimal structured policies that are "good enough" when balanced against computational effort. This is an area for further research.

The generalization of MDP's to POMDP's also results in added computational difficulties. In a finite state MDP, an optimal policy can be expressed in simple tabular form, listing optimal actions for each state. When state uncertainty is introduced into the same model, we have a POMDP with an enlarged set of states. The optimal policy is now defined over a continuum of states. The path-breaking work of Sondik has mitigated many of the problems inherent in the computation of optimal policies for POMDP's. His algorithms are discussed in some detail in §5. Efficient computational procedures exist for short, finite horizon POMDP's. Less efficient procedures exist for infinite horizon problems.

The partially observable Markov decision process is formally presented in §2. The main result is summarized as follows: although a partially observable process is not Markovian (in general), the POMDP can be formulated as a Markov decision process with an enlarged state space, namely the space of probability distributions over the underlying (partially observable) states. The usual dynamic programming recursions for both the finite and discounted infinite horizon models are presented.

A brief history of the development of POMDP's is given in §3. We also examine the relationship between POMDP's and stochastic automata, certain Bayesian decision processes, and various continuous time, partially observable stochastic processes.

§4 contains a survey of models which either have been or could be cast in the partially observable framework developed in §2. A detailed description of some machine replacement/quality control models is given in the first subsection. Applications of POMDP models in several other contexts, such as accounting, optimal stopping, and learning are then presented.

In §5, various computational procedures for solving POMDP's are discussed. The preponderance of this section is devoted to the finite and infinite horizon algorithms (and their variations) developed by Sondik.

Notational conventions are as follows: $I \equiv \{0, 1, 2, \ldots\}$, $I_+ \equiv \{1, 2, \ldots\}$, $\Pr\{\cdot\}$ denotes the probability of the event $\{\cdot\}$, $\mathbb{R}^N$ denotes the N-fold Cartesian product of the real line, $\mathbb{R}$, and $|\mathcal{Q}|$ denotes the number of elements in the set $\mathcal{Q}$.

2. The Finite State Partially Observable Markov Decision Process

In this section the finite state and action space version of the partially observable Markov decision process is presented.

Let X_t be a random variable defined on a sample space Ω, where $t \in I$; assume X_t takes on values in the finite set $\mathfrak{N} \equiv \{1, \ldots, N\}$. The stochastic process $\{X_t, t \in I\}$, called the *core process*, is assumed to be a finite state Markov chain with stationary $N \times N$ transition probability matrix $P = [p_{ij}]$, $i, j \in \mathfrak{N}$. The core process is completely described by P and the initial distribution over $\mathfrak{N}$, denoted by $\pi(0) = (\pi_1(0), \ldots, \pi_N(0))$, where $\pi_i(0) \equiv \Pr\{X_0 = i\}$, $i = 1, \ldots, N$. The core process is not directly observable; that is, the realization of X_t is not determinable with certainty at time t.

Associated with X_t is a random variable Y_t which takes on values in a finite "message" space $\mathfrak{M} \equiv \{1, \ldots, M\}$. By observing Y_t at time t, information regarding the true value of X_t is obtained. The probabilistic relationship between X_t and Y_t is known to the decision maker. Suppose that if $X_t = i$, an observation will have message k with probability q_{ik}, i.e.,

$$q_{ik} \equiv \Pr\{Y_t = k \mid X_t = i\} \quad \text{for } i \in \mathfrak{N}, k \in \mathfrak{M}. \tag{2.1}$$

Define the $N \times M$ *information matrix* as $Q = [q_{ik}]$, $i \in \mathfrak{N}, k \in \mathfrak{M}$. The stochastic process $\{Y_t, t \in I\}$ is called the *observation process*.

A decision structure is now defined which incorporates the core and observation processes. Assume that the decision maker can control both the observation and core processes by choosing actions. Let $\mathcal{Q}$ be a finite set denoting all of the actions available to the decision maker. Let $P(a) = [p_{ij}(a)]$ denote the "law of motion" of the core process when action $a \in \mathcal{Q}$ is chosen. That is, if i is the current state and action a is chosen, the core process moves to a new state j with probability $p_{ij}(a)$, $i, j \in \mathfrak{N}$. Similarly, let $Q(a) = [q_{ij}(a)]$ denote the relationship between the observation and core processes when $a \in \mathcal{Q}$ is chosen.

Let $m_t \in \mathfrak{M}$ and $a_t \in \mathcal{Q}$ denote the value of Y_t observed and the action taken at time t, respectively. The data available for decision making at time t is denoted by $d_t \equiv (\pi(0), m_1, a_1, \ldots, a_{t-1}, m_t)$.

Define $\pi_i(t) \equiv \Pr\{X_t = i \mid d_t\}$ and let

$$\pi(t) = [\pi_1(t), \ldots, \pi_N(t)];$$

$$\pi(t) \in \mathcal{S}_N \equiv \left\{ x \in \mathbb{R}^N : \sum_{i=1}^{N} x_i = 1, x_i \geqslant 0, i = 1, \ldots, N \right\}$$

is called the *information vector*. Using Bayes' formula, the transformation of the information vector from time t to $t+1$ is specified as:

$$T_i(\pi(t) \mid j, a_t) \equiv \pi_i(t+1) = \Pr\{X_{t+1} = i \mid d_{t+1} = (d_t, a_t, m_{t+1} = j)\}$$

$$= \frac{q_{ij}(a_t) \sum_{k \in \mathfrak{N}} p_{ki}(a_t)\pi_k(t)}{\sum_{l \in \mathfrak{N}} q_{lj}(a_t) \cdot \sum_{k \in \mathfrak{N}} p_{kl}(a_t)\pi_k(t)}, \qquad i = 1, \ldots, N, \tag{2.2}$$

where $q_{ij}(a_t)$ and $p_{ki}(a_t)$ are the (i, j)th and (k, i)th elements of $Q(a_t)$ and $P(a_t)$, respectively.

The following result is readily established (e.g., see [9], [65], [68]):

$\pi(t)$ summarizes all of the information necessary
for making decisions at time t.

It is customary to denote by s_t, say, all the information required for decision making at time t. Then, from the result cited above, $s_t = \pi(t)$. The following theorem is also well-known (e.g., see [4], [6], [48], [59], [65]):

THEOREM 2.1. *For any fixed sequence of actions $a_1, \ldots, a_t \in \mathcal{A}$, the sequence of probabilities $\{\pi(t), t \in I\}$ is a Markov process; that is, if $\Gamma \subset \mathbb{S}_N$, then*

$$\Pr\{\pi(t+1) \in \Gamma \mid \pi(0), \ldots, \pi(t), a_t\} = \Pr\{\pi(t+1) \in \Gamma \mid \pi(t), a_t\}.$$

With these results, the POMDP can be converted into an equivalent (completely observable) Markov decision process.

Note. The core process was defined on a finite state space. Since that process is unobservable, an equivalent observable Markov process is now defined on an uncountable state space, namely the $(N-1)$-simplex in $\mathbb{R}^N$.

For notational convenience, let

$$\gamma(\pi(t), j, a_t) = \sum_{i \in \mathfrak{N}} q_{ij}(a_t) \cdot \sum_{k \in \mathfrak{N}} p_{ki}(a_t)\pi_k(t).$$

Then $\gamma(\pi(t), j, a_t) = \Pr\{Y_{t+1} = j \mid s_t = \pi(t), a_t\}$, which is the denominator of (2.2).

Assume that there is a reward function, say $r: \mathfrak{N} \times \mathcal{A} \to \mathbb{R}$, where $r_i(a_t)$ is the immediate expected reward that is earned at time t if the core process is in state i and action a_t is taken; the expectation is with respect to the conditional probability measures associated with the core and observation processes. The immediate reward could depend upon the current state of the core process, the next state of the core process (that is, there may be a reward associated with transitions from state i to state j in the core process), the outcome of the observation, and the action taken. Notationally,

$$r_i(a_t) = \sum_{j=1}^{N} \sum_{k=1}^{M} R(i, j, k, a_t) p_{ij} q_{jk}$$

where $R: \mathfrak{N} \times \mathfrak{N} \times \mathfrak{M} \times \mathcal{A} \to \mathbb{R}$ is a bounded function, with $R(i, j, k, a_t)$ representing the immediate reward when action a_t is taken, the core process is in state i, moves to state j, and output k is observed.

For ease of notation let $r(a_t) \equiv [r_1(a_t), \ldots, r_N(a_t)]$ and "$\cdot$" denote the usual inner product operator. Then, if the state of the POMDP is s_t and action a_t is taken, an immediate expected reward of

$$E\big[r(X_t, a_t) \mid s_t = \pi\big] = \pi \cdot r(a_t) \tag{2.3}$$

is obtained.

A schematic representation of the decision process is given in Figure 1 (cf. Sondik [65, Figure 2.5]).

Note. Some authors view the sequence of events comprising a POMDP slightly differently. In some models an observation is taken and then the transition to the new state is made. In this presentation, movement in the chain is followed by an observation. Although the formula for updating the information vector, eqn. (2.2), may be slightly different, the two views are equivalent.

Let the function $\delta_t: \mathbb{S}_N \to \mathcal{A}$ denote a *decision rule* which indicates the action, $\delta_t(\pi) \in \mathcal{A}$ to take at time t when the current state is $s_t = \pi$. A policy (or strategy) δ, is defined as a sequence of decision rules, $\delta = \{\delta_1, \ldots, \delta_t, \ldots\}$. A strategy is said to be (non-randomized) *stationary* if it is a single function $\alpha: \mathbb{S}_N \to \mathcal{A}$, where, for any $\pi \in \mathbb{S}_N$, $\alpha(\pi)$ denotes the action to take when the state of the process is π.

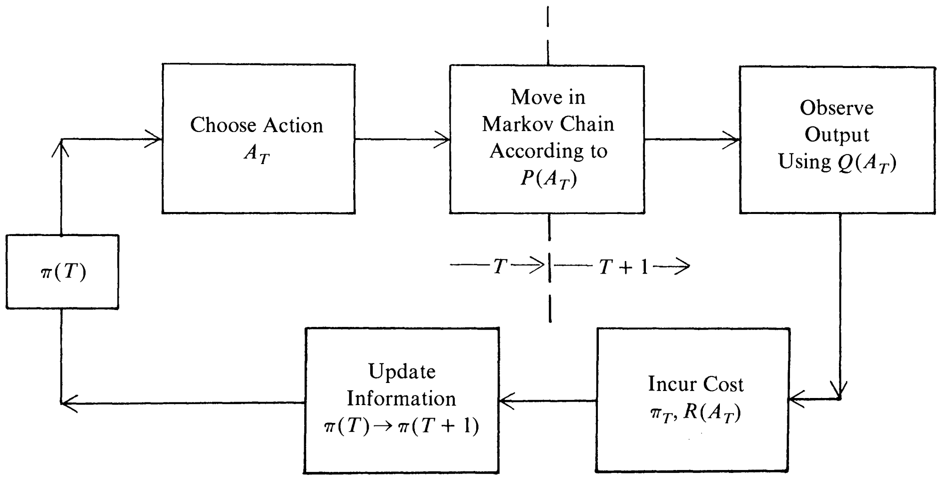
Choose Action A_T
Move in Markov Chain According to $P(A_T)$
Observe Output Using $Q(A_T)$
$\pi(T)$
—T→|—$T+1$→
Update Information $\pi(T) \to \pi(T+1)$
Incur Cost $\pi_T, R(A_T)$

FIGURE 1. A Partially Observable Markov Decision Process.

Given strategy δ is employed and the process starts at $\pi \in \mathbb{S}_N$, define the *expected discounted infinite horizon reward of the* POMDP as:

$$V_\delta(\pi) = E_\delta\left[\sum_{t=0}^{\infty} \beta^t r(X_t, a_t) \mid s_0 = \pi\right], \qquad \pi \in \mathbb{S}_N, \tag{2.4}$$

where $0 \leqslant \beta < 1$ is the discount factor, and E_δ denotes the conditional expectation operation given δ.

The *optimal value function* $V_\beta(\cdot)$ can then be defined as:

$$V_\beta(\pi) = \operatorname*{Sup}_{\delta} V_\delta(\pi), \qquad \pi \in \mathbb{S}_N.$$

A strategy δ^* is said to be *β-optimal* if

$$V_\beta(\pi) = V_{\delta^*}(\pi) \quad \text{for all } \pi \in \mathbb{S}_N. \tag{2.5}$$

The objective of the decision maker is to determine a β-optimal strategy.

A well-known result concerning the optimal value function is now presented. Let $T(\pi \mid j, a) = [T_1(\pi \mid j, a), \ldots, T_N(\pi \mid j, a)]$ where $T_i(\pi \mid j, a)$ is defined in (2.2).

THEOREM 2.2. (*See, e.g., Ross* [52, *Theorem* 6.1].) *The infinite horizon β-optimal value function $V_\beta(\pi)$ defined in* (2.5) *satisfies the following recursion*:

$$V_\beta(\pi) = \operatorname*{Max}_{a \in \mathcal{Q}}\left\{\pi \cdot r(a) + \beta \sum_{j \in \mathfrak{M}} V_\beta[T(\pi \mid j, a)]\gamma(j \mid \pi, a)\right\} \tag{2.6}$$

for $\pi \in \mathbb{S}_N$, where $r(a)$ is defined as above.

The finite horizon analog of (2.6) can be defined recursively as:

$$V_\beta^0(\pi) = \pi \cdot r(0)$$

$$V_\beta^n(\pi) = \operatorname*{Max}_{a \in \mathcal{Q}}\left\{\pi \cdot r(a) + \beta \sum_{j \in \mathfrak{M}} V_\beta^{n-1}[T(\pi \mid j, a)]\gamma(j \mid \pi, a)\right\}$$

$$\text{for } \pi \in \mathbb{S}_N, n \in I_+, \tag{2.7}$$

where $r_i(0)$ is the terminal reward received when the core process is in state i, $i = 1, \ldots, N$. For $n \in I_+$, $V_\beta^n(\pi)$ denotes the maximum discounted expected reward that can be obtained given that the process is currently in state π and there are n periods remaining before the decision process must end.

3. Development of POMDP's and Related Literature

The problem of controlling random process (including Markov processes) with incomplete state information was initially studied by Sirjaev [61] and Dynkin [17]. Wald's [73] pioneering work on sequential sampling may be thought of as a special type of POMDP. Blackwell [10] developed an entropy measure for a partially observable Markov chain. Drake [16] developed the first explicit POMDP model. Striebel [68] proved the sufficiency of the information vector for a wide class of stochastic control problems. About the same time, Astrom [6], [7] and Aoki [4], [5] also formulated finite horizon POMDP's in the context of stochastic control problems. Generalizations of their work followed. Sawaragi and Yoshikawa [59] developed the theory of POMDP's with an uncountable action space and a countable core process state space. Rhenius [48] considered POMDP's where both the action and core process state spaces were Borel spaces. Furukawa [21] also considered a POMDP with an arbitrary core process state space and a finite action space. Striebel [69] generalized the Astrom-Aoki control model to more general state and action spaces. Hinderer [24] studied non-stationary POMDP's which have a more general reward structure than the model considered in this paper. In [85], White and Harrington studied the value function associated with any given (not necessarily optimal) policy, in a POMDP framework. They gave conditions which insured that the value function does not diminish as observation quality improves. Iosifescu and Mandl [29] and Platzman [43] developed conditions under which undiscounted infinite horizon POMDP's are well-defined. Issues dealing with the effect of information acquisition on the conditional distributions over the core states were studied by Rudemo [54] and Platzman [44]. Kaijser [30] developed conditions which insured that the limiting conditional state distribution converged to a measure which is independent of the initial state distribution.

Sondik, in his thesis [65] and subsequent papers [64], [66], was the first to address and resolve the computational difficulties associated with POMDP's. His algorithms for computing solutions to the finite and infinite horizon discounted problems are discussed in §5. White [77, 78] generalized the POMDP to allow for a semi-Markov core process. He extended Sondik's computational procedure to compute policies for finite horizon POMDP's with a semi-Markov core process.

A number of papers in the literature have explored conditions which insure that optimal policies have certain structural characteristics, such as monotonicity and/or control-limit form. Albright [1] presented conditions under which the optimal policy for a POMDP with a two-state core process would be monotone in the information vector. White [84] gave conditions which yield monotone optimal policies for finite horizon POMDP's where there is either perfect observability or no observability. He then demonstrated how these structural results simplify the computation of the optimal policy. Other papers dealing with structural results in certain machine replacement problems are discussed in the next section.

The theory of POMDP's is now being used to aid in the solution of non-POMDP problems. White and Kim [86] developed algorithms for finding the set of all pure, stationary nonrandomized strategies for vector criterion MDP's. Hsu and Marcus [26] studied the problem of the decentralized control of a Markov chain. The movement of the chain depends upon its current state and the actions of two or more decentralized decision makers. Each decision maker chooses an action given local information about the state of the unobservable chain, but agrees to share this information in the next period. This information pattern is referred to as One Step Delay Sharing. The problem is formulated as a POMDP, and results from the theory of POMDP's are used to establish the existence of an optimal stationary policy, and to develop algorithms for computing such policies. White and Schussel [87] used the theory of POMDP's to compute bounds and sub-optimal policies for multi-module MDP's. (A multi-module MDP is a system of MDP's that are linked together only through the cost structure.)

One of the main characteristics of the POMDP is the transformation of the information vector from period to period via Bayes' rule [see (2.2)]. There is a body of literature dealing with Bayesian control of sequential decision processes which is only indirectly related to the POMDP's considered in this paper; see e.g., [21], [39], [49], [55], [71], and [76]. In this literature, elements of the decision process are unknown. The decision maker may not know, for example, the transition probability matrix governing the movement of the process. Information regarding the parameters of the objects in the model is obtained. In a POMDP, however, all the elements of the decision process are assumed to be known. Only information regarding the current state of the unobservable core process is obtained.

There is a literature dealing with the acquisition of information for various *continuous time* partially observable stochastic processes. The interested reader is directed to see, for example, [2], [3], [8], and [20].

Brooks and Leondes [12] considered a special type of MDP with one stage information delay and computed the marginal cost associated with the delay. Although this is an MDP with incomplete state information, the form of the information available permits the problem to be transformed into a (perfectly observable) MDP.

Finally, it should be pointed out that a POMDP is a special case of a stochastic sequential machine (SSM) (see, e.g., Paz [40]). Using the notation of Section 2, an SSM is defined as a quadruple, $(\mathcal{N}, \mathcal{A}, \mathcal{M}, \{A(m \mid a)\})$, where $\{A(m \mid a)\}$ is a finite set containing $|\mathcal{A}| \cdot M$ square matrices each of order N, such that $a_{ij}(m \mid a) \geqslant 0$ for all i and j and

$$\sum_{m=1}^{M} \sum_{j=1}^{N} a_{ij}(m \mid a) = 1, \qquad i = 1, \ldots, N,$$

and

$$A(m \mid a) = \left[a_{ij}(m \mid a) \right].$$

The SSM is a generalization of the POMDP model presented in §2, in that $a_{ij}(m \mid a)$ represents the probability that the core process moves to state j and the message variable has value m given the core process is currently in state i and action a is taken.

The theory of probabilistic automata (which includes the study of SSM's) has not yet been specialized to the study of POMDP's. However, Platzman [42], [45] considered ideas such as state reduction and state equivalence found in that theory, in his development of an algorithm to compute approximately optimal policies for infinite horizon POMDP's. His algorithm is discussed in §5.

4. Models Incorporating the Theory of POMDP's

In this section models incorporating the theory presented in §2 will be discussed. Although many of the models presented were not formulated explicitly as POMDP's, they all deal with the optimal control of a random process based on incomplete information.

A. POMDP *Models of Machine Replacement / Quality Control Problems*

The quality control models in the literature can be classified on the basis of the source and degree of partial information. For simplicity, a general version of a two-state model is presented using the notation of Section 2. By placing restrictions on the general model, many of the models in the literature can then be discussed. The restriction to two core states is done for ease of exposition. Many of the models discussed below were formulated with three or more core states.

The core process represents the condition of a machine which is deteriorating over time. The true condition of the machine is not known with certainty. There are two sources of information regarding the condition of the machine. Firstly, information can be obtained by observing the machine's output. Secondly, information can be obtained by actually inspecting the machine. The observation process of the POMDP consists of data generated from these sources. The actions available in any period are: do nothing (perhaps observe the machine's output), inspect the machine or inspect the machine's output, and repair (replace) the machine.

Let $\mathcal{Q} = \{a_1, a_2, a_3\}$, where a_1 denotes doing nothing, a_2 denotes inspecting, and a_3 repairing. The transition and observation matrices are defined for the two-state process (state 1 ≡ "good" condition, state 2 ≡ "bad" condition) as:

$$P(a_1) = \begin{bmatrix} 1-\gamma & \gamma \\ 0 & 1 \end{bmatrix}, \qquad Q(a_1) = \begin{bmatrix} \alpha_1 & 1-\alpha_1 \\ \alpha_2 & 1-\alpha_2 \end{bmatrix},$$

$$0 \leqslant \gamma \leqslant 1, \text{ and } 0 \leqslant \alpha_i \leqslant 1, i = 1, 2,$$

$$P(a_2) = \begin{bmatrix} 1-\gamma & \gamma \\ 0 & 1 \end{bmatrix}, \qquad Q(a_2) = \begin{bmatrix} \nu_1 & 1-\nu_1 \\ 1-\nu_2 & \nu_2 \end{bmatrix},$$

$$0 \leqslant \nu_i \leqslant 1, i = 1, 2,$$

$$P(a_3) = \begin{bmatrix} 1-\gamma & \gamma \\ 1-\gamma & \gamma \end{bmatrix}, \qquad Q(a_3) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}.$$

Note. An information matrix with a column of 1's denotes an observation process that provides no information since the message observed is independent of the core state; the identity matrix denotes perfect information since there is a one-to-one relationship between messages and core states.

The identical rows of $P(a_3)$ denote the possible deterioration of a new machine.

Girshick and Rubin [22] were the first to consider a variation of the problem given above. If observation of the machine's output is costly (or destructive) a rule should indicate not only when to repair the machine but should also indicate which items to inspect. In their model, $\alpha_1 = \alpha_2 = 1$ (implying no information is available if the "do nothing" action is selected), and $\nu_1 = \nu_2 = 1$ (implying perfect information is available if the "inspection" action is selected). Conjectures they made concerning the form of the optimal maintenance policy were shown to be false (via a counter-example) by Taylor [70], who considered a replacement model in a general setting.

Klein [33] also considered a variation of the non-100% inspection case. His approach was somewhat different in that a new completely observable problem was formulated which modeled periodic inspection. Decisions were of the form "repair now, but do not inspect in the next m periods."

Based on the completely observable model of Derman [15] ($\alpha_1 = \alpha_2 = 1$, a_2 not permitted), Eckles [18] and Ross [53] formulated POMDP's for a problem similar to

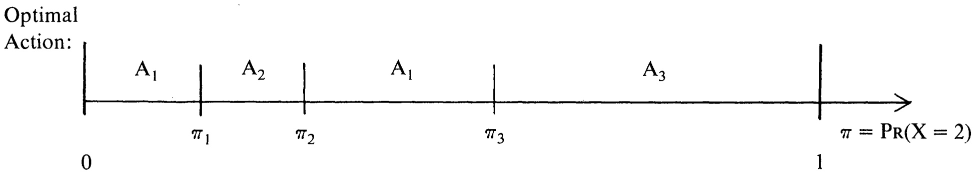

FIGURE 2. Ross' Two-State Optimal Inspection/Replacement Policy.

the non-100% inspection problem. In terms of the general model, all three actions are permitted, and $\alpha_1 = \alpha_2 = \nu_1 = \nu_2 = 1$ (no information is available if action a_1 is taken, perfect information regarding the true condition of the machine is available if action a_2 is taken).

Ross [53] characterized the optimal inspection/maintenance policy. In particular, he showed that for the two-state core process, the optimal policy has at most four regions. (See Figure 2). Ross also gave conditions on the parameters which would insure that $\pi_1 = \pi_2$, i.e., that inspecting would never be optimal.

Ehrenfeld [19] also examined various aspects of the Derman-Ross model. He considered the possibility that inspection is not perfect ($\nu_i \neq 1, i = 1, 2$) but was unable to establish conditions that would insure a well-structured policy.

White [80, 81] studied a problem which is very similar to the general maintenance problem. As a special case he considered a model where $\nu_1 = \nu_2 = 1$ but $\alpha_i \neq 1$, $i = 1, 2$, thus generalizing Ross' model. He proved that the optimal policy has the form depicted in Figure 2. The general partially observable model ($\nu_i \neq 1, \alpha_i \neq i, i = 1, 2$) was also discussed. However, as in the Ehrenfeld paper, the characterization of the optimal policy remained an open question.

Conditions that guaranteed optimal control-limit policies for a model with imperfect observability were ultimately introduced by White [83] and represent a significant contribution to the literature. The new condition that was required is difficult to interpret. The demonstration of the control-limit structure used Porteus' [47] results on the optimality of structured policies in sequential decision problems. White [82] also demonstrated the optimality of structured policies for the special cases of perfect and no observability in a machine replacement setting.

Rosenfield [50], [51] considered yet another variation of the general model. He defined a process which has a state space consisting of pairs of nonnegative integers denoted by (i, k) where i is the condition of the machine and k is the number of periods which have elapsed since it was known for certain that the machine was in state i (and hence is somewhat similar to the Klein model discussed earlier). Rosenfield proved that an optimal maintenance policy is monotonic in the following sense: the optimal policy is defined by control-limit numbers $k^*(i)$, $i = 1, \ldots, N$, which are nondecreasing in i, where, for each state (i, k), repair is done only if $k > k^*(i)$.

Wang [74], [75] discussed various forms of the general model (all of them precluding any form of inspection action a_2) under weaker conditions on the parameters. He proved that a control-limit-type policy is still optimal. A procedure he used to compute such a policy is discussed in the next section.

Pierskalla and Voelker [41] give an excellent review of maintenance models which includes a section on models with incomplete information.

B. *Other* POMDP *Models*

Kaplan [31] applied the basic results of the machine inspection/replacement problem to a cost control problem in accounting. He assumed that an operating segment of a firm can be in one of two states: state 1 indicates that the costs incurred by the segment are "in control," meaning that management can not affect (reduce) the costs;

state 2 indicates that costs being incurred are "out of control"—management action can be taken to reduce costs. Management has only two alternatives: it can do nothing or it can take corrective action at some cost. As in the analogous machine replacement problem, the optimal policy was shown to be of the control-limit type.

Analogously, Hughes [27] modeled the internal control of a corporate control system as a POMDP. Using Ross' [53] quality control model, Hughes viewed the core process as the level of effectiveness of internal control. Information pertaining to the effectiveness of control can be obtained through an internal audit. The actions available are: do nothing, audit (corresponding to inspecting in the Ross model), and restore (analogous to the replace action in the quality control context). Hughes did not develop any new alternative results for the POMDP used in this audit-timing context.

Karush and Dear [32] formulated a dynamic programming model of a learning process which can be classified as a POMDP. A subject is to be taught m items in the course of n trials. The subject is assumed to be either conditioned (C) or unconditioned (U) with respect to each of the m items; the state space of the core process consists of the 2^m m-vectors representing all of the combinations of m-tuples of U's and C's and is unobservable. An action consists of presenting item i to the subject on trial k, $k = 1, \ldots, n$. The observation process consists of responses of the subject when presented with an item or a trial. The state of the POMDP is $\Lambda = (\lambda_1, \ldots, \lambda_m)$, where λ_i is the (posterior) probability that the subject is conditioned to item i, $i = 1, \ldots, m$. The objective is to maximize a terminal reward which is a function of Λ. Karush and Dear proved that the following *myopic* policy is optimal: at trial k, given the current POMDP-state vector $\Lambda^k = (\lambda_1^k, \ldots, \lambda_m^k)$, present item i^* where $\lambda_{i^*}^k = \mathrm{Min}_i \lambda_i^k$; that is present the item which currently has the highest probability of being conditioned.

Smallwood [62] also developed a simple two core state POMDP model of optimal teaching strategies. Core states represent the student's state of knowledge. One teaching strategy presents the stimulus along with the correct response. This causes the student to move from the no-knowledge state to the knowledge state with a known probability. A second teaching strategy presents the stimulus and requires a response from the student. The response, of course, may or may not reflect the true state of knowledge of the student. Given costs per trial for each strategy and given a terminal cost for ending the teaching session with the student in the no-knowledge state, Smallwood proved the piecewise-linearity of the infinite horizon value function and described methods for computing the optimal teaching strategy.

Pollock [46] developed a two-state core process POMDP to model optimal search effort for an object that moves between two regions. Before each move, the decision maker can choose which region to examine in order to locate the object. Under various conditions that correspond to forms of perfect observability, he proved that the optimal search rules had special structure. Special structure could not be obtained for general parameter values.

Smallwood, et al. [63] used POMDP concepts in the development of methodology for the analysis of health-care systems. They defined (unobservable) patient states and related observable states (symptoms, diagnostic data, etc.). Physician states correspond to states in the POMDP, that is, they are physician-specific distributions over patient states. Different patient state-physician state pairs are defined for various problems arising in the health-care area, including the design of individual medical facilities, regional health systems, and the funding of health-service programs.

Given a simple network consisting of two (geographically separated) computers, Segall [60] studied the problem of where to locate a common data file. He assumed that the demand rate for the data at one of the computer sites is an unobservable finite-state Markov chain, while the demand rate at the other site is a deterministic function of time. Given costs of data storage at each site and transmission costs, a

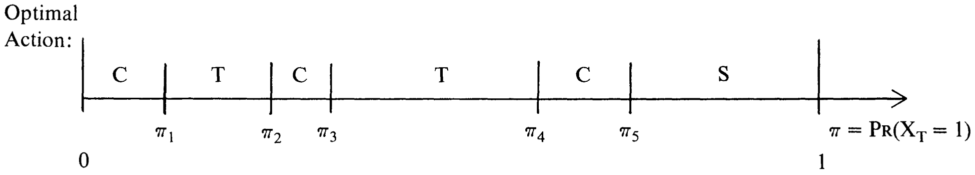

FIGURE 3. An Optimal Policy, $\delta^*(\pi)$.

POMDP model was formulated to determine the optimal file location over time. Structural properties of the optimal policy were not examined.

In another computer network context, Platzman [42] formulated the slotted ALOHA problem (see [34], [35]) as a POMDP. In this problem, remote terminals communicate with a central computer via a channel that can carry at most one message per time interval. If two or more terminals attempt to transmit messages simultaneously, none of the messages are transmitted. Individual terminals must decide when to transmit based on imperfect knowledge of the status of other terminals on the network. Observations available for decision making in each time interval in this context are the outcomes of a transmission attempt.

White [79] applied the theory of POMDP's to design questionnaires in situations where responses may not be truthful.

Monahan [37] formulated a discrete-time problem of stopping in a partially observable binary-valued Markov chain as a POMDP. In this context, $X_t \in \{0, 1\}$ represents the reward received by the decision maker if the process is stopped at time t. Before deciding to stop or continue, the decision maker can sequentially purchase additional information regarding the value of X_t. The actions that are available are S, T, and C which denote "stop", "test" (purchase more information), and "continue" (forego X_t and consider X_{t+1}). In [37] it was shown that the optimal policy may be highly unstructured. Let $\pi = \Pr\{X_t = 1\}$ denote the probability that the reward at t is "good" and let $\delta^*(\pi)$ denote the optimal policy. It is possible for $\delta^*(\pi)$ to have the form depicted in Figure 3 when the information regarding the current reward is imperfect. However, when the information indicates the core process state without error, $\delta^*(\pi)$ has the form depicted in Figure 2; see [38]. This is another example of a model where structural properties of the optimal policy can be determined only when perfect state information is available.

Finally, we point out that there are other extensive classes of models which can be classified as POMDP's but will not be discussed here. They include models of search for a hidden object (discrete search models—see, e.g., Stone [67]), sequential sampling problems (see DeGroot [14] and references therein), and two- and multi-armed bandit problems (also DeGroot [14]).

5. Algorithms for Solving Partially Observable Processes

In this section computational procedures for solving POMDP's are discussed. The most significant work in this area has been done by Sondik [65], who developed Howard-like value-determination, policy-improvement algorithms for solving both finite and infinite horizon POMDP's. Two of Sondik's algorithms are described in general terms below. A modification of the finite horizon algorithm due to White [84] is also presented. Other computational procedures used to solve particular POMDP models are then mentioned.

A. Sondik's "One-Pass" Algorithm

The Sondik "one-pass" algorithm ([64], [65]) is used to compute the optimal policy and value function for finite horizon POMDP's. The one-pass algorithm exploits the structure of the finite horizon optimal value function $V^n(\cdot)$ given in (2.7). It is straightforward to show that $V^n(\cdot)$ is piecewise-linear and convex in its argument. Using the notation of §2, let

$$A_0 = \{r(0)\}, \qquad A_n = \left\{\alpha : \alpha = r(a) + \sum_{k \in \mathfrak{M}} P(a)Q^k(a)\alpha_k, \alpha_k \in A_{n-1}, a \in \mathcal{A}\right\},$$

where $Q^k(a)$ is the $N \times N$ diagonal matrix formed from the kth column of $Q(a)$; that is, $Q_{ii}^k(a) = Q_{ik}(a)$, $i \in \mathfrak{N}$ and $k \in \mathfrak{M}$. Then

$$V^n(\pi) = \operatorname{Max}\{\pi \cdot \alpha : \alpha \in A_n\}. \tag{5.1}$$

For any $n \in I$, A_n is a finite set. However some of the α-coefficients in the set may be dominated by others and can be removed. To find the minimal set of α-coefficients that define $V^n(\cdot)$, solve the following linear program for each $\alpha \in A_n$:

$$\operatorname*{Min}_{\pi}\{x - \pi \cdot \alpha : x \geqslant \pi \cdot \alpha', \alpha' \in A_n, \pi \in \mathbb{S}_N\}.$$

If $x \neq 0$, remove α from A_n.

The optimal strategy when the current state is π and there are n periods remaining is now easily determined: find the index j^* that maximizes $\pi \cdot \alpha_j$, $\alpha_j \in A_n$. Then $\delta_n^*(\pi) = a_{j^*}$.

White [84] modified the one-pass algorithm to exploit known structural properties of the optimal policy for certain classes of POMDP's, thus making the algorithm more efficient. The modified algorithm restricts the space of feasible policies to those which are (in some sense) *isotone* (monotonically nondecreasing). Additional structure is also placed on the POMDP. The core process state space is assumed to be a partially ordered space $(n, \leqslant_n)$, where n is countable. The order relation $\leqslant_n$ induces a partial order on $\mathbb{S}_N$, denoted $<_\pi$; i.e., $(\mathbb{S}_N, <_\pi)$ is a partially ordered space. The action space $\mathcal{A}$, is assumed to be a finite linearly (totally) ordered space $(\mathcal{A}, \leqslant_A)$. If $\pi_1, \pi_2 \in \mathbb{S}_N$ are such that $\pi_1 <_\pi \pi_2$ implies $\delta(\pi_1) \leqslant_A \delta(\pi_2)$, then $\delta(\cdot)$ is called an *isotone policy*. Conditions which guarantee optimal isotone policies also insure that the optimal infinite horizon value functions are monotone; i.e., for $\pi_1, \pi_2 \in \mathbb{S}_N$, if $\pi_1 <_\pi \pi_2$, then $V^n(\pi_1) \leqslant V^n(\pi_2)$, $n = 0, 1, \ldots$. The White modification of the one-pass algorithm uses the added structure in two ways. Firstly, regions of the state space over which the optimal value function is linear can be enlarged. Secondly, using the relation $\leqslant_A$ eliminates the need to calculate boundaries which will ultimately never be used to identify subsets over which a particular action is optimal.

B. *Sondik's Discounted, Infinite Horizon Algorithm*

Sondik [66] developed a Howard-like policy iteration algorithm to compute ϵ-optimal policies for the infinite horizon POMDP. *Finitely-transient* (f.t.) policies play a fundamental role in the algorithm. In general terms a policy is f.t. if after a finite number of transitions the resulting state is not one at which the policy is discontinuous, independent of both the current state and the sequences of messages observed. The significance of f.t. policies is that the infinite horizon discounted value function associated with such a policy is piecewise-linear and the optimal infinite horizon policy can be computed with the one-pass algorithm in a finite number of iterations. Therefore, as in the one-pass algorithm, piecewise-linearity of the value function permits the computation of both the optimal value function and optimal policy.

Unfortunately, not all stationary policies are f.t. Sondik [66] defined an approximation so that all stationary policies are (*almost*) *f.t.* The result is an approximate value function that is piecewise-linear.

The Sondik algorithm is a policy iteration algorithm which uses the one-pass algorithm in the value-determination step.

C. Platzman's Algorithm for Computing Suboptimal Infinite Horizon Policies

Platzman [45] developed an algorithm for computing approximately optimal policies for infinite horizon POMDP's. His work was motivated by the following considerations. In Sondik's one-pass algorithm, the number of elements in A_n explodes as n tends to infinity. Of course, when a policy is finitely-transient, the number of elements in A_n is finite for all n. However, since finite transience may be difficult to verify in practice, Platzman exploited an idea attributed to Drake [16] to insure finiteness of A_n for all n. The decision maker is restricted to consideration of only a finite number of the most recent observations and actions when choosing an action. The notion that so-called finite memory policies may be adequate in many decision making contexts is a prime motivation in Platzman's algorithm. A finite-memory, randomized strategy is selected which indicates the action to take and specifies the next memory state as a function of the current memory state. Thus the decision maker is modeled as a probabilistic automaton, or equivalently, another POMDP. Selecting the optimal strategy amounts to solving a finite-dimensional nonlinear program. Performance bounds are given which indicate how close the current solution is to the global optimum.

D. Other Computational Procedures

Satia and Lave [56] developed an implicit enumeration algorithm for computing ϵ-optimal solutions to the finite horizon POMDP in a finite number of iterations. They also briefly discussed using the control-limit structure of the optimal policy in the Girshick-Rubin machine replacement problem. They report a computation time of 110 seconds required to determine the control-limit for a two state, two action problem, which would seem to indicate that the algorithm is not very efficient.

Wang [74], [75] developed a special purpose computational procedure for determining optimal policies for certain finite-state machine replacement problems. The models he considered allow for only two actions (do nothing, replace); inspection is not allowed. The procedure described appears to be quite efficient for solving these special problems with more than two unobservable states.

Buckman and Miller [13] presented an algorithm for solving Kaplan's [31] optimal investigation problem discussed in §4. They formulated the problem as a regenerative stopping problem [11] and exploited structural properties to improve the computation of the optimal policy. Miller [36] used similar techniques to compute solutions to the Rosenfield [50], [51]-type maintenance model discussed in §4.

Sawaki [57] and Sawaki and Ichikawa [58] pointed out the rather obvious fact that the optimal infinite horizon discounted POMDP value function can be approximated arbitrarily closely by a (sufficiently large) finite horizon POMDP which is piecewise-linear and admits piecewise-constant optimal policies. The practical significance of their results for efficiently computing optimal policies is dubious since the number of linear segments can very quickly exceed the storage capacity of any existing computer.[1]

[1] The author would like to thank Loren K. Platzman for his many helpful comments on an earlier version of this paper and for suggesting some of the applications and references.

References

1. Albright, S., "Structural Results for Partially Observable Markov Decision Processes," *Operations Res.*, Vol. 27 (1979), pp. 1041–1053.
2. Anderson, R. F. and Friedman, A., "Optimal Inspections in a Stochastic Control Problem with Costly Observations," *Math. Operations Res.*, Vol. 2 (1977), pp. 155–190.
3. ——— and ———, "Optimal Inspections in a Stochastic Control Problem with Costly Observations, II," *Math. Operations Res.*, Vol. 3 (1978), pp. 67–81.

4. AOKI, M., "Optimal Control of Partially Observable Markovian Control Systems," *J. Franklin Inst.*, Vol. 280 (1965), pp. 367–386.
5. ———, *Optimization of Stochastic Systems*, Academic Press, New York, 1967.
6. ASTROM, K., "Optimal Control of Markov Processes with Incomplete State Information," *J. of Math. Analysis and Appl.*, Vol. 10 (1965), pp. 174–205.
7. ———, "Optimal Control of Markov Processes with Incomplete State Information, II. The Convexity of the Loss Function," *J. of Math. Analysis and Appl.*, Vol. 26 (1969), pp. 403–406.
8. BATHER, J., "An Optimal Stopping Problem with Costly Information," *Bull. Inst. Internat. Statist.*, Vol. 45, Book 3 (1973), pp. 9–24.
9. BERTSEKAS, D., *Dynamic Programming and Stochastic Control*, Academic Press, New York, 1976.
10. BLACKWELL, D., "The Entropy of Functions of Finite-State Markov Chains," in *Information Theory, Statistical Decision Functions, Random Processes: Transactions of the First Prague Conference, 1956*, Publishing House of Czechosolvak Academy of Sciences, Prague, 1957.
11. BREIMAN, L., "Stopping-Rule Problems," Chapt. 10 in *Applied Combinatorial Mathematics*, E. Beckenback, ed., Wiley, New York, 1964.
12. BROOKS, D. AND LEONDES, C., "Markov Decision Processes with State-Information Lag," *Operations Res.*, Vol. 20 (1972), pp. 904–907.
13. BUCKMAN, A. G. AND MILLER, B. L., "Optimal Investigation as a Regenerative Stopping Problem," Western Management Sci. Instit., UCLA, Working Paper No. 289, 1979.
14. DEGROOT, M., *Optimal Statistical Decisions*, McGraw-Hill, New York, 1970.
15. DERMAN, C., "Optimal Replacement Rules when Changes of State are Markovian," in *Mathematical Optimization Techniques*, R. Bellman, ed., Univ. of California Press, Berkeley, Calif., 1963.
16. DRAKE, A., *Observation of a Markov Process Through a Noisy Channel*, unpublished Sc.D. Thesis, Dept. of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Mass., 1962.
17. DYNKIN, E., "Controlled Random Sequences," *Theory Probability Appl.*, Vol. 10 (1965), pp. 1–14.
18. ECKLES, J., "Optimum Maintenance with Incomplete Information," *Operations Res.*, Vol. 16 (1968), pp. 1058–1067.
19. EHRENFELD, S., "On a Sequential Markovian Decision Procedure with Incomplete Information," *Comput. and Operations Res.*, Vol. 3 (1976), pp. 39–48.
20. FRIEDMAN, A., "Optimal Stopping for Random Evolution of Multidimensional Poisson Processes with Partial Information," in *Stochastic Analysis*, Friedman, A., and Pinskey, M., eds., Academic Press, New York, 1978.
21. FURUKAWA, N., "A Bayes Controlled Process," *Mem. Fac. Sci., Kyushu Univ. Ser. A*, Vol. 21 (1968), pp. 249–258.
22. GIRSHICK, M. AND RUBIN, H., "A Bayes' Approach to a Quality Control Model," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 114–125.
23. HEYMAN, D. AND SOBEL, M., *Stochastic Models in Operations Research, Vol. II: Stochastic Optimization*, McGraw-Hill, New York (forthcoming).
24. HINDERER, K., *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*, Springer-Verlag, Berlin, 1970.
25. HOWARD, R., *Dynamic Programming and Markov Processes*, The M.I.T. Press, Cambridge, Mass., 1960.
26. HSU, K. AND MARCUS, S., "Decentralized Control of Finite State Markov Processes," *Proceedings 19th IEEE Conf. Dec. and Control*, Dec. 1980, pp. 143–148.
27. HUGHES, J., "Optimal Internal Audit Timing," *Accounting Rev.*, Vol. LII (1977), pp. 56–58.
28. IGLEHART, D., "Optimality of (s, S) Policies in the Infinite Horizon Dynamic Inventory Problem," *Management Sci.*, Vol. 9 (1963), pp. 259–267.
29. IOSIFESCU, M. AND MANDL, P., "Application Des Systèmes à Liaisions Complètes a un Problème de Réglage," *Rev. Roum. Math. Pures et Appl.*, Vol. XI (1966), pp. 533–539.
30. KAIJSER, J., "A Limit Theorem for Partially Observed Markov Chains," *Ann. Probability*, Vol. 3 (1975), pp. 677–696.
31. KAPLAN, R., "Optimal Investigation Strategies with Imperfect Information," *J. Accounting Res.*, Vol. 7 (1969), pp. 32–43.
32. KARUSH, W. AND DEAR, R., "Optimal Strategy for Item Presentation in Learning Models," *Management Sci.*, Vol. 13 (1967), pp. 773–785.
33. KLEIN, M., "Inspection-Maintenance-Replacement Schedules Under Markovian Deterioration," *Management Sci.*, Vol. 9 (1962), pp. 25–32.
34. KLEINROCK, L. AND LAM, S., "Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation," *IEEE Trans. Comm.*, Vol. COM-23 (1975), pp. 410–423.
35. LAM, S. AND KLEINROCK, L., "Packet Switching in a Multiaccess Broadcast Channel: Dynamic Control Procedures," *IEEE Trans. Comm.*, Vol. COM-23 (1975), pp. 891–904.
36. MILLER, B. L., "Countable State Average Cost Regenerative Stopping Problems," Western Management Sci. Instit., UCLA, Working Paper No. 288 (1979).

37. MONAHAN, G. E., "Optimal Stopping in a Partially Observable Markov Process with Costly Information," *Operations Res.*, Vol. 28 (1980), pp. 1319–1334.
38. ———, "Optimal Stopping in a Partially Observable, Binary-valued Markov Chain with Perfect, Costly Information," (forthcoming) *J. Appl. Probability*, Vol. 19 (1982).
39. NAHMIAS, S., "A Sequential Decision Problem with Partial Information," *Cahiers du Cente d'Etudes de Recherche Operationnelle*, Vol. 17 (1975), pp. 53–64.
40. PAZ, A., *Introduction to Probabilistic Automata*, Academic Press, New York, 1971.
41. PIERSKALLA, W. AND VOELKER, J., "A Survey of Maintenance Models: The Control and Surveillance of Deteriorating Systems," *Naval Res. Logist. Quart.*, Vol. 23 (1976), pp. 353–388.
42. PLATZMAN, L., *Finite-Memory Estimation and Control of Finite Probabilistic Systems*, unpublished Ph.D. thesis, Department of Electrical Engineering and Computer Science, M.I.T.; also M.I.T. Electronic Systems Laboratory Technical Report ESL-R-723, Cambridge, Mass., 1977.
43. ———, "Optimal Infinite-Horizon Undiscounted Control of Finite Probabilistic Systems," *SIAM J. Control and Optimization*, Vol. 18 (1980), pp. 362–380.
44. ———, "State-Estimation of Partially-Observed Markov Chains: Decomposition, Convergence, and Component Identification," mimeograph, March, 1978.
45. ———, "A Feasible Computational Approach to Infinite-Horizon Partially-Observed Markov Decision Problems," mimeograph, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Ga., January, 1981.
46. POLLOCK, S., "A Simple Model of Search for a Moving Target," *Operations Res.*, Vol. 18 (1970), pp. 883–903.
47. PORTEUS, E., "On the Optimality of Structured Policies in Countable Stage Decision Processes," *Management Sci.*, Vol. 22 (1975), pp. 148–157.
48. RHENIUS, D., "Incomplete Information in Markovian Decision Models," *Ann. Statist.*, Vol. 2 (1974), pp. 1327–1334.
49. RIEDER, U., "Bayesian Dynamic Programming," *Advances Appl. Probability*, Vol. 7 (1975), pp. 720–736.
50. ROSENFIELD, D., "Markovian Deterioration with Uncertain Information," *Operations Res.*, Vol. 24 (1976), pp. 141–155.
51. ———, "Markovian Deterioration with Uncertain Information—A More General Model," *Naval Res. Logist. Quart.*, Vol. 23 (1976), pp. 389–406.
52. ROSS, S., *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, Calif., 1970.
53. ———, "Quality Control Under Markovian Deterioration," *Management Sci.*, Vol. 17 (1971), pp. 587–596.
54. RUDEMO, M., "State Estimation for Partially Observed Markov Chains," *J. Math. Anal. Appl.*, Vol. 44 (1973), 581–611.
55. SATIA, J. AND LAVE, R., "Markovian Decision Processes with Uncertain Transition Probabilities," *Operations Res.*, Vol. 21 (1973), pp. 728–740.
56. ——— AND ———, "Markovian Decision Processes with Probabilistic Observation of States," *Management Sci.*, Vol. 20 (1973), pp. 1–13.
57. SAWAKI, K., "Piecewise-Linear Markov Decision Processes with an Application to Partially Observable Models," in Hartley, R. et al., eds., *Recent Advances in Markov Decision Processes*, Academic Press, New York, 1980, pp. 245–260.
58. ——— AND ICHIKAWA, A., "Optimal Control for Partially Observable Markov Decision Processes Over an Infinite Horizon," *J. Operations Res. Soc. Japan*, Vol. 21 (1978), pp. 1–15.
59. SAWARAGI, Y. AND YOSHIKAWA, T., "Discrete-Time Markovian Decision Processes with Incomplete State Observation," *Ann. Math. Statist.*, Vol. 41 (1970), pp. 78–86.
60. SEGALL, A., "Dynamic File Assignment in a Computer Network," *IEEE Trans. Auto. Control*, Vol. AC-21 (1976), pp. 161–173.
61. SIRJAEV, A., "On the Theory of Decision Functions and Control of a Process of Observation Based on Incomplete Information," *Selected Translations in Math. Stat. and Prob.*, Vol. 6 (1966), pp. 162–188.
62. SMALLWOOD, R., "The Analysis of Economic Teaching Strategies for a Simple Learning Model," *J. Math. Psych.*, Vol. 8 (1971), pp. 285–301.
63. ———, SONDIK, E. AND OFFENSEND, F., "Toward an Integrated Methodology for the Analysis of Health-Care Systems," *Operations Res.*, Vol. 19 (1971), pp. 1300–1322.
64. ——— AND ———, "The Optimal Control of Partially Observable Markov Processes over a Finite Horizon," *Operations Res.*, Vol. 21 (1973), pp. 1071–1088.
65. SONDIK, E., *The Optimal Control of Partially Observable Markov Processes*, unpublished Ph.D. dissertation, Stanford University, 1971.
66. ———, "The Optimal Control of Partially Observable Markov Processes Over the Infinite Horizon: Discounted Costs," *Operations Res.*, Vol. 26 (1978), pp. 282–304.
67. STONE, L., *Theory of Optimal Search*, Academic Press, New York, 1975.

68. Striebel, C., "Sufficient Statistics in the Control of Stochastic Systems," *J. Math. Anal. Appl.*, Vol. 12 (1965), pp. 576–592.

69. ———, *Optimal Control of Discrete Time Stochastic Systems*, Springer-Verlag, Berlin, 1975.

70. Taylor, H., "Markovian Sequential Replacement Processes," *Ann. Math. Statist.*, Vol. 38 (1966), pp. 871–890.

71. Van Hee, K., "Bayesian Control of Markov Chains," *Mathematical Centre Tract 95*, Amsterdam, The Netherlands, 1978.

72. Vazsonyi, A., "Information Systems in Management Science—The Use of Mathematics for Management Information Systems II," *Interfaces*, Vol. 6, (1976), pp. 42–46.

73. Wald, A., *Sequential Analysis*, Wiley, New York, 1947; republished by Dover, New York, 1973.

74. Wang, R., "Computing Optimal Quality Control Policies—Two Actions," *J. Appl. Probability*, Vol. 13 (1976), pp. 826–832.

75. ———, "Optimal Replacement Policy Under Unobservable States," *J. Appl. Probability*, Vol. 14 (1977), pp. 340–348.

76. Wessels, J., *Decision Rules in Markovian Decision Processes with Incompletely Known Transition Probabilities*, dissertation, University of Technology, Eindhoven, 1968.

77. White, C., "Cost Equality and Inequality Results for a Partially Observed Stochastic Optimization Problem," *IEEE Trans. Systems, Man, and Cybernetics*, Vol. SMC-5 (1975), pp. 576–582.

78. ———, "Procedures for the Solution of a Finite-Horizon, Partially Observed, Semi-Markov Optimization Problem," *Operations Res.*, Vol. 24 (1976), pp. 348–358.

79. ———, "Optimal Diagnostic Questionnaries Which Allow Less Than Truthful Responses," *Information and Control*, Vol. 32 (1976), pp. 61–74.

80. ———, "A Markov Quality Control Process Subject to Partial Observation," *Management Sci.*, Vol. 23 (1977), pp. 843–852.

81. ———, "Optimal Inspection and Repair of a Production Process Subject to Deterioration," *J. Operational Res. Soc.*, Vol. 29 (1978), pp. 235–243.

82. ———, "Bounds on Optimal Cost for a Replacement Problem with Partial Observation," *Naval Res. Logist. Quart.*, Vol. 26 (1979), pp. 415–422.

83. ———, "Optimal Control-Limit Strategies for a Partially Observed Replacement Problem," *Internat. J. Systems Science*, Vol. 10 (1979), pp. 321–331.

84. ———, "Monotone Control Laws for Noisy, Countable-State Markov Chains," *European J. Operational Res.*, Vol. 5 (1980), pp. 124–132.

85. ——— and Harrington, D., "Application of Jensen's Inequality for Adaptive Suboptimal Design," *J. Opt. Theory and Appl.*, Vol. 32 (1980), pp. 89–100.

86. ——— and Kim, K., "Solution Procedures for Solving Vector Criterion Markov Decision Processes," *J. Large-Scale Systems*, Vol. 1 (1980), pp. 129–140.

87. ——— and Schussel, K., "Suboptimal Design for Large-Scale, Multi-Module Systems," *Operations Res.* (to appear).