ON THE AVERAGE COST OPTIMALITY EQUATION AND THE STRUCTURE OF OPTIMAL POLICIES FOR PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES*

Emmanuel Fernández-Gaucherand, Aristotle Arapostathis,
and Steven I. Marcus
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, Texas 78712-1084

Abstract

We consider partially observable Markov decision processes with finite or countably infinite (core) state and observation spaces and finite action set. Following a standard approach, an equivalent completely observed problem is formulated, with the same finite action set but with an *uncountable* state space, namely the space of probability distributions on the original core state space. By developing a suitable theoretical framework, it is shown that some characteristics induced in the original problem due to the countability of the spaces involved are reflected onto the equivalent problem. Sufficient conditions are then derived for solutions to the average cost optimality equation to exist. We illustrate these results in the context of machine replacement problems. Structural properties for average cost optimal policies are obtained for a two state replacement problem; these are similar to results available for discount optimal policies. The set of assumptions used compares favorably to others currently available.

Key Words: Optimal control, Markov chains, partial observability, average cost, optimality equation, structured optimal policies.

* This research was supported in part by the Advanced Technology Program of the State of Texas, in part by the Air Force Office of Scientific Research under Grant AFOSR-86-0029, in part by the National Science Foundation under Grant ECS-8617860, and in part by the Air Force Office of Scientific Research (AFSC) under Contract F49620-89-C-0044.

1. Introduction

Since the pioneering work of Bellman [BEL1] in the mid fifties, the field of Markovian decision processes (MDP) has received sustained attention over the years, giving rise to the development of a rich theory as presented in, e.g., [BE], [BS], [DY], [HLM1], [HS], [KV], [RO3]. Furthermore, numerous applications have been proposed within the realm of, e.g., operations research, economics, computer and communication networks [DJW1], [DJW2], [DJW3], [STH], [WW]. Using a stochastic approach, these models address the problem of decision-making under uncertainty. If the state of the process is concealed from the decision-maker, i.e., only *partial observations* are available, then he/she is further faced with the problem of *estimating the state* from the available information. Moreover, some parameters of the model may be unknown and/or time-varying, and thus the need may arise to treat the problem from an *adaptive control* standpoint [BEL2], [HLM1], [KV], [VHE].

We study here partially observable MDP with finite or countably infinite state and observation spaces, and finite action set [MO1]. Following a standard approach, an associated completely observed problem is formulated, with the same finite action set but with an *uncountable* state space, namely the space of probability distributions on the original state space. For several optimality criteria, these two problems are equivalent, in the sense of equal optimal costs [BS], [HLM1]. In particular, we consider the average cost criterion. For these problem, the particular structure of the problems considered is advantageously used, instead of using general results from the theory of MDP with general (Borel) state space. We develop a theoretical framework based on the notion of *invariant* subsets of an MDP. Sufficient conditions are then given, for the existence of solutions to the associated optimality equation. The conditions developed compare favorably with respect to other available results, since they are in many cases easier to verify and appear to be significantly less restrictive in general. We illustrate our results in the context of two-state machine replacement problems, and with the aid of the optimality equation, structural results are obtained for the optimal policy. These results for the optimization problem are important in their own right, and furthermore, they are indispensable for the study of the adaptive control problem based on an *enforced certainty equivalence principle,* c.f. [AM], [FAM1], [FAM2], [GE2], [HM], [MFA].

The paper is organized as follows. In Section 2 we give the statement of the problem and collect the main results in the literature concerning partially observable MDP, with both finite horizon and discounted cost optimization criteria, viewing these within the framework of Borel state space MDP. In Section 3, the average cost optimality equation is discussed, and some related results are reviewed. Section 4 presents an extension of the standard vanishing discount approach to the partially observable situation, leading to verifiable conditions that guarantee the existence of solutions to the average cost optimality equation. In Section 5, monotone MDP are discussed, and it is shown how the existence of *reset* actions leads easily to the verification of conditions under which our results hold. In Section 6, a two-state replacement problem is studied in detail, and the corresponding average cost optimality equations are obtained for different special cases. Using these, structural results are then obtained for average cost optimal policies, similar to known results for discounted cost optimal policies. Finally, Section 7 presents a discussion of some related results currently available in the literature, e.g. [OMK], [PL], and our conclusions.

2. Partially Observable Markov Decision Processes (POMDP)

We are interested in those problems of discrete time optimal stochastic control in which the state dynamics are governed by a Markovian law, and only partial observations of the state are available, i.e., POMDP. Specifically, we consider problems within the following framework. Let X and Y be countable, linearly ordered sets; thus without loss of generality, $X = \{0, \ldots, N_X\}$ if X is finite, for some nonnegative integer $N_X \in I\!N_0 := \{0\} \cup I\!N$, or $X = I\!N_0$ otherwise, and similarly for Y. Also, let U be a finite, linearly ordered set, i.e., $U = \{0. \ldots, N_U\}, N_U \in I\!N_0$. We endow these sets with their respective order topologies, which in this situation are equal to their discrete topologies, denoted as, e.g., 2^X. In general, for a topological space W, $\mathcal{B}(W)$ will denote its Borel σ-algebra. For the above spaces we thus have, e.g., $\mathcal{B}(X) = 2^X$. The system's state process will be modelled as a finite state controlled Markov chain with ("core") state space X and action space U, and with transition matrices $\{P(u)\}_{u\in U}$. Assume for the moment that there is an underlying probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Hence, the core state process is given by a random process $\{x_t\}$ on $(\Omega, \mathcal{F}, \mathcal{P})$, $t \in I\!N_0$, where for a sequence of U-valued random variables $\{u_k\}_{k=0}^{t-1}$ on $(\Omega, \mathcal{F}, \mathcal{P})$, called the controls (or decisions), we have

$$\mathcal{P}\{x_{t+1} = j | x_t = i, x_{t-1}, \ldots, x_0; u_t = u, u_{t-1}, \ldots, u_0\}$$

$$= [P(u)]_{i,j} =: p_{i,j}(u), \quad t \in I\!N_0.$$

Only partial observations of $\{x_t\}_{t\in I\!N_0}$ are available in the form of a random process $\{y_t\}_{t\in I\!N_0}$ taking values in the observation space Y. The core and observation processes are related by

$$\mathcal{P}\{y_{t+1} = y | y_t, \ldots, y_1; x_{t+1} = i, \ldots, x_0; u_t = u, \ldots, u_0\}$$

$$= \mathcal{P}\{y_{t+1} = y | x_{t+1} = i, u_t = u\} =: q_{i,y}(u), \quad t \in I\!N_0,$$

which leads to the definition of a collection of observation matrices $\{Q(u)\}_{u \in U}$ such that

$$Q(u) := [q_{i,y}(u)]_{i \in X, y \in Y}.$$

The sequence of events is assumed as follows: initially the system is in state x_0, a decision u_0 is made, a transition to a state x_1 has occurred by the beginning of time epoch 1, and a first observation y_1 becomes available; at the beginning of time epoch $t \in I\!N$ the system is in state x_t, observation y_t becomes available, and action u_t is taken; transition to a state x_{t+1} has occurred by the beginning of time epoch $t+1$, another observation y_{t+1} becomes available, and then a new decision u_{t+1} is made; and so on.

It is assumed that the probability distribution of the initial state, $p_0 := [\mathcal{P}\{x_0 = i\}]_{i \in X} \in \Delta$, is available for decision making, where $\Delta := \{p \in I\!R^X : p^{(i)} \geq 0, \mathbf{1}'p = 1\}$; here, $\mathbf{1}$ denotes the (column) vector in $I\!R^X$ with all components equal to one, "prime" denotes transposition, and $p^{(i)}$ denotes the i^{th} component of p. We endow Δ with the topology induced by the metric $d(\cdot\,,\cdot)$ given by

$$d(p_1, p_2) := \sum_{i \in X} \left| p_1^{(i)} - p_2^{(i)} \right| = \|p_1 - p_2\|_1,$$

where $\|\cdot\|_1$ is the standard ℓ_1-norm on $I\!R^X$. Note that when X is a finite set, Δ is compact. The following is easily shown.

Lemma 2.1: (Δ, d) is a Polish space, i.e., it is a complete and separable metric space.

Recursively define the *history spaces*

$$\begin{aligned} H_0 &:= \Delta, \\ H_t &:= H_{t-1} \times U \times Y, \qquad t \in I\!N, \\ H_\infty &:= H_0 \times (U \times Y)^\infty, \end{aligned}$$

each equipped with its respective product topology. An element $h_t \in H_t, t \in I\!N_0$, is called an *observable history* and represents the information available for decision making at time epoch t. Recall that a Borel space is a Borel subset of a Polish space. It is straightforward to show that X, Y, U, and, by Lemma 2.1, Δ are Borel spaces, and hence so is $H_t, t \in I\!N_0$ (see [BS, p.119]).

A *Borel measurable stochastic kernel* $\mu_t(\cdot\,|\,\cdot)$ on U given H_t is a collection of (discrete) probability distributions $\{\mu_t(\cdot\,|h_t) : h_t \in H_t\}$ on $(U, 2^U)$ such that for each $B \in 2^U$, $\mu_t(B|\,\cdot)$ is a measurable function on H_t. An admissible control *law, policy* or *strategy* μ is a sequence of stochastic kernels $\{\mu_t(\cdot\,|\,\cdot)\}_{t\in I\!N_0}$, or $\{\mu_0(\cdot\,|\,\cdot), \ldots, \mu_n(\cdot\,|\,\cdot)\}$ for a finite horizon; $\mu_t(\cdot\,|\,\cdot)$ is called the control law at time t. If for each $h_t \in H_t, t \in I\!N_0$, $\mu_t(\cdot\,|h_t)$ is concentrated at a point in U, then $\mu_t(\cdot\,|\,\cdot)$ is said to be *nonrandomized;* similarly for a strategy μ. Thus, we can view a nonrandomized strategy μ as a sequence of measurable maps $\mu_t : H_t \to U$.

Let $c : X \times U \to I\!R$ be a given (measurable) map; $c(x,u)$ is interpreted as the cost incurred given that the system was in state x and control action u was selected. The following assumption is made for the rest of the paper.

Assumption 2.1: $0 \le c(x,u)$, for all $x \in X$ and $u \in U$.

Let $\Omega := (X \times U) \times (X \times Y \times U)^\infty$, which is endowed with the respective product topology, and let an initial distribution p_0 and an admissible strategy μ be given. Then there is a (unique) probability measure $\mathcal{P}^\mu_{p_0}$ on $(\Omega, \mathcal{B}(\Omega))$ induced by p_0 and the strategy μ, which satisfies the following consistency conditions (see [BS, p.140-144 and 249]):

$$\mathcal{P}^\mu_{p_0}\{x_0\} = p_0; \quad \mathcal{P}^\mu_{p_0}\{u_t = u\,|\,h_t\} = \mu_t(u\,|\,h_t), \quad h_t \in H_t;$$

$$\mathcal{P}^\mu_{p_0}\{x_{t+1} = j|x_t = i, x_{t-1}, \ldots, x_0; u_t = u, u_{t-1}, \ldots, u_0\} = p_{i,j}(u), \quad t \in I\!N_0.$$

Denote by $E^\mu_{p_0}$ expectation with respect to $\mathcal{P}^\mu_{p_0}$, or an appropiate marginal. Then, to each admissible strategy μ and initial distribution p_0, the following expected costs are associated.

Finite Horizon:

$$J_\beta(\mu, p_0, n) := E^\mu_{p_0}\left[\sum_{t=0}^{n} \beta^t c(x_t, u_t)\right], \; n \in I\!N_0, \quad 0 < \beta. \qquad (FH)$$

Discounted Cost:

$$J_\beta(\mu, p_0) := \lim_{n\to\infty} E^\mu_{p_0}\left[\sum_{t=0}^{n} \beta^t c(x_t, u_t)\right], \; 0 < \beta < 1. \qquad (DC)$$

Average Cost:

$$J(\mu, p_0) := \limsup_{n\to\infty} E^{\mu}_{p_0}\left[\frac{1}{n+1}\sum_{t=0}^{n} c(x_t, u_t)\right]. \qquad (AC)$$

The *optimal control (or decision) problem* is that of selecting an (optimal) admissible strategy such that one of the above criteria is minimized over all admissible strategies, for all $p_0 \in \Delta$. The optimal (DC) *value function* is obtained as $\Gamma_\beta(p_0) := \inf_\mu \{J_\beta(\mu, p_0) : \mu$ is an admissible strategy$\}$, for each $p_0 \in \Delta$. Similarly denote by $\Gamma_\beta(\cdot\,, n)$ and $\Gamma(\cdot)$ the optimal cost functions for the (FH) problem with horizon $n \in I\!N_0$, and the (AC) problem, respectively.

During the last two decades, it has been rigorously shown that a partial or imperfect observations stochastic control problem can be converted into an equivalent problem with perfect observations [AS1], [BS], [HLM1], [SY], where the new state must be an *information state* [KV]. Let $p_t \in I\!N_0$ denote the conditional probability distribution of the (core) state process, whose i^{th} component is given by

$$p_t^{(i)} := \mathcal{P}^{\mu}_{p_0}\{x_t = i | y_t, \ldots, y_1; u_{t-1}, \ldots u_0\}, \quad t \in I\!N,$$

and p_0 is the given initial distribution. Then, assuming that $\mathcal{P}^{\mu}_{p_0}\{h_t, u_t, y_{t+1} = y\} \neq 0 \quad a.s.$, for $t \in I\!N_0$ and for each $y \in Y$, and using Bayes' rule, it is easily shown that (e.g. see [AS1], [KV, Sect. 6.6])

$$p_{t+1} = \sum_{y\in Y} \frac{\overline{Q}_y(u_t)P'(u_t)p_t}{\mathbf{1}'\overline{Q}_y(u_t)P'(u_t)p_t} \cdot I\,[y_{t+1} = y], \quad t \in I\!N_0, \qquad (2.1)$$

where $I[A]$ denotes the indicator function of the event A and the matrices $\overline{Q}_y(u)$ are given by $\overline{Q}_y(u) :=$ diag $\{q_{i,y}(u)\}$. If $\mathcal{P}^{\mu}_{p_0}\{h_t, u_t, y_{t+1} = y\} = 0 \quad a.s.$, then the corresponding term for $y_{t+1} = y$ in (2.1) is defined conveniently. Note that p_t is a function of (y_t, p_{t-1}, u_{t-1}), i.e., it is recursively computable given the most recent information.

A *separated* admissible law for time epoch t is a Borel measurable stochastic kernel $\mu_t(\cdot\,|\,\cdot)$ on U given Δ. Separated admissible strategies are defined in the obvious way. A *nonrandomized* separated admissible strategy is thus viewed as a sequence of measurable maps $\mu_t : \Delta \to U$. When $\mu_t(\cdot\,|\,\cdot) = \mu(\cdot\,|\,\cdot)$ for all values of t, then the policy is said to be *stationary.* A nonrandomized separated admissible law μ can be regarded as a nonrandomized admissible law via $h_t \mapsto \mu_t(p_t)$, where p_t is obtained from h_t by applying (2.1) recursively. Let

$$\overline{T}(y, p, u) := \overline{Q}_y(u)P'(u)p, \qquad y \in Y, p \in \Delta, u \in U;$$

$$V(y,p,u) := \mathbf{1}'\overline{T}(y,p,u); \qquad T(y,p,u) := \overline{T}(y,p,u)/V(y,p,u).$$

Then $V(y,p,u)$ is interpreted as the (one-step ahead) conditional probability of the observation being y given an *a priori* distribution p for the core state, under decision u. Likewise, $T(y,p,u)$ is interpreted as the *a posteriori* conditional probability distribution of the core state given decision u was made, observation y obtained, and an *a priori* distribution p. That is, for any admissible policy μ and any initial distribution p_0,

$$\begin{aligned}\mathcal{P}^{\mu}_{p_0}\{y_{t+1} = y | u_t = u, \ldots, u_0; y_t, \ldots, y_1\}& \\ = \mathcal{P}^{\mu}_{p_0}\{y_{t+1} = y | p_t; u_t = u\} &= V(y, p_t, u),\\ [\mathcal{P}^{\mu}_{p_0}\{x_{t+1} = i | u_t = u, \ldots, u_0; y_{t+1} = y, y_t, \ldots, y_1\}]_{i \in X} &= T(y, p_t, u).\end{aligned}$$

In this notation, (2.1) can be written compactly in the form

$$p_{t+1} = \sum_{y \in Y} T(y, p_t, u_t) \cdot I\left[y_{t+1} = y\right]. \tag{2.2}$$

2.1. An Equivalent Borel State Space MDP

We present some well known properties of the process $\{p_t\}$; e.g., see [AS1], [FG], [SY].

Lemma 2.2: (i) For any fixed sequence of actions $\{u_0, u_1, \ldots\} \subseteq U$, the controlled process $\{p_t\}$ is Markovian.

(ii) The transition kernel for the controlled Markov process $\{p_t\}$ is given by

$$\begin{aligned}\mathcal{P}^{\mu}_{p_0}\{p_{t+1} \in B | p_t = p; u_t = u\} &= \sum_{y \in Y} V(y,p,u) I[T(y,p,u) \in B] \\ &=: \mathcal{K}(B|p,u), \quad B \in \mathcal{B}(\Delta).\end{aligned} \tag{2.3}$$

Since p_t can be recursively updated via (2.1), given the observable history, then $\{p_t\}$ is a *completely observable* controlled Markov process, the state space of which is Δ, an *uncountable* Borel space. Let $\overline{c}(\cdot\,,\,\cdot)$ be defined on $\Delta \times U$ as $\overline{c}(p,u) := p'[c(i,u)]_{i \in X}$. In addition to Assumption 2.1, the following is assumed to hold for the rest of the paper.

Assumption 2.2: For all $p \in \Delta$ and $u \in U$, $\overline{c}(p,u) < \infty$.

Note that the above holds whenever X is finite, or when $c(\cdot\,,\,\cdot)$ is bounded. Then a *completely observable,* finite horizon, optimal control problem, with state space Δ, can be formulated as finding a separated admissible strategy which minimizes, over all admissible strategies,

$$J'_\beta(\mu, p_0, n) := E^\mu_{p_0}\left[\sum_{t=0}^{n} \beta^t \overline{c}(p_t, u_t)\right], \qquad (FH')$$

for all $p_0 \in \Delta$. The problems (DC') and (AC') are defined similarly. It is well known that a separation principle holds for the problems listed above, i.e. these problems have been shown to be *equivalent,* in the sense of equal minimum costs, to their corresponding counterparts in the original POMDP, c.f. [AS1], [BE], [BS], [HLM1], [KV], [SY]. Thus optimal cost are denoted as before, e.g., $\Gamma_\beta(\cdot)$.

The above equivalent formulation falls within the framework of MDP with *general Borel state space* (BMDP), as studied in, e.g., [BS], [HLM]: (Δ, d) is a Borel space, by Lemma 2.1, and $\mathcal{K}$ in (2.3) is a stochastic kernel on Δ given $\Delta \times U$. Henceforth, given a POMDP, specified by $\{X, U, Y, Q, P, c\}$, we will refer to it by its equivalent formulation, specified by $\{\Delta, U, \mathcal{K}, \overline{c}\}$, the latter being viewed as a BMDP, and thus results from this general theory can be used in our context.

The specification of $\{Q(u)\}$ affects the attainable optimal costs for $\{\Delta, U, \mathcal{K}, \overline{c}\}$, via its influence in $\mathcal{K}$ of (2.3). For the case when X and Y are finite sets, we say that the decision process is *completely observable* (CO) if $Q(u) = I$, for all $u \in U$, where I is the identity matrix, and *completely unobservable* (CU) if

$$Q(u) = \frac{1}{N_X + 1}\big[\,\mathbf{1} \;\;|\;\; \cdots \;\;|\;\; \mathbf{1}\,\big], \quad \forall u \in U.$$

Thus in a CO problem, observations give complete "information" on the core state, and in a CU problem, observations do not convey any "information" about the core state. Denote by $\Gamma^{(co)}_\beta(\cdot)$ and $\Gamma^{(cu)}_\beta(\cdot)$ the value functions associated with these decision processes, respectively, and similarly for problems with a finite horizon. Also, recursively define sets of $(N_X + 1)$-dimensional vectors as follows

$$A_{-1} = \{(0, 0, \ldots, 0)'\},$$

$$A_n = \{\big[c(i,u)\big]_{i\in X} + \beta P(u) \sum_{y\in Y} \overline{Q}_y(u)\alpha_y : \alpha_y \in A_{n-1}, u \in U\}, \quad n \in I\!N_0,$$

and note that the cardinality of these sets obeys the recursion

$$|A_n| \le |U| \cdot |A_{n-1}|^{|Y|}, \quad n \in I\!N.$$

Some important results are summarized below.

Theorem 2.1: (i) For an (FH') decision problem, $\Gamma_\beta(\cdot\,,n)$ is concave in $p \in \Delta$, for all $n \in I\!N_0$ and $0 < \beta$. In addition, Bellman's optimality equation holds

$$\Gamma_\beta(p,n) = \min_{u\in U}\Big\{\overline{c}(p,u) + \beta \sum_{y\in Y} V(y,p,u)\Gamma_\beta(T(y,p,u),n-1)\Big\},$$

where any (nonrandomized) separated stationary policy which attains the minimum above is optimal. Also, if X and Y are finite sets, then $\Gamma_\beta(\cdot\,,n)$ is piecewise linear, and it can be computed as

$$\Gamma_\beta(p,n) = \min_{\alpha\in A_n}\{p'\alpha\}.$$

Furthermore

$$\Gamma_\beta^{(co)}(p,n) \le \Gamma_\beta(p,n) \le \Gamma_\beta^{(cu)}(p,n).$$

(ii) For a (DC') decision problem, for all $0 < \beta < 1$, $\Gamma_\beta(\cdot)$ is concave. In addition, Bellman's infinite horizon optimality equation holds

$$\Gamma_\beta(p) = \min_{u\in U}\Big\{\overline{c}(p,u) + \beta \sum_{y\in Y} V(y,p,u)\Gamma_\beta(T(y,p,u))\Big\}, \tag{2.4}$$

where any (nonrandomized) separated stationary policy which attains the minimum above is optimal. Also, if X and Y are finite sets, then

$$\Gamma_\beta^{(co)}(p) \le \Gamma_\beta(p) \le \Gamma_\beta^{(cu)}(p).$$

Remark 2.1: The optimality equations of (a) and (b) are obtained from the general theory of BMDP [BS], [HLM1]. For the other results, see [AS1], [AS2], [BE], [SO1], [SO2], [SS].

3. The Average Cost Optimality Equation

When a discounted cost criterion is used in a Markov decision problem with bounded costs and general Borel state space, the corresponding dynamic programming operator exhibits nice contractive properties which allow the development of a rather complete theory. The same cannot be said of the average cost criterion, for which it is much more difficult to obtain functional characterizations, and solutions to these, for the optimal costs and policies. Stringent ergodicity conditions [HLM1], [HMC], or equicontinuity assumptions [RO1] are usually required to obtain such characterizations, which in our situation are as described below.

Definition 3.1: If there are real-valued functions J and h on Δ, such that, for all $p \in \Delta$,

$$J(p) + h(p) = \min_{u \in U} \left\{ \overline{c}(p,u) + \sum_{y \in Y} V(y,p,u) h(T(y,p,u)) \right\}, \tag{3.1}$$

then the pair (J, h) is said to be a solution to the *average cost optimality equation* (ACOE).

The ACOE of (3.1) is a specialization, to our situation, of the corresponding equation arising in general BMDP, c.f. [HLM1]. The following result can be obtained easily from Theorem 2.2 in [HLM1, pp. 53-55].

Theorem 3.1: Suppose that (J, h) is a solution to the ACOE, and that for each admissible policy μ and each $p_0 \in \Delta$, $J(\cdot)$ and $h(\cdot)$ are integrable with respect to $\mathcal{P}^{\mu}_{p_0}$ and

$$\lim_{n \to \infty} E^{\mu}_{p_0} \left[\frac{h(p_n)}{n} \right] = 0. \tag{3.2}$$

Then

(i) for each admissible policy μ and each $p_0 \in \Delta$

$$\limsup_{n \to \infty} \frac{1}{n+1} E^{\mu}_{p_0} \left[\sum_{t=0}^{n} J(p_t) \right] \leq J(\mu, p_0), \tag{3.3}$$

and if μ is a (nonrandomized) stationary separated policy such that $\mu(p)$ attains the minimum in (3.1), equality is attained in (3.3);

(ii) if there a constant $\Gamma^* \in I\!R$ such that $J(p) = \Gamma^*$, for all $p \in \Delta$, then $\Gamma^* = \Gamma(p)$, for all $p \in \Delta$, and if μ^* is a (nonrandomized) stationary separated policy such that $\mu^*(p)$ attains the minimum in (3.1), then μ^* is average cost optimal.

Remark 3.1: An interpretation of (i) above is that if J is taken as the cost function to define the BMDP $\{\Delta, U, \mathcal{K}, J\}$, then for *any* admissible μ, the average cost assessed under the cost function J does not exceed that under cost function $\overline{c}$.

A pair (Γ^*, h) is said to be a *bounded* solution to the ACOE if $\Gamma^* \in I\!\!R$, and $h(\cdot)$ is a bounded function on Δ. Given the results above, naturally there has been considerable interest in finding conditions which guarantee the existence of a bounded solution (Γ^*, h) to the ACOE, for then (3.2) is satisfied trivially, and Theorem 3.1(ii) applies. However, the type of sufficient conditions available in the literature, for such a solution to the ACOE to exist, impose a very restrictive recurrence structure on the model under *every* stationary policy, c.f. [CC1], [CC2], [HLM1], [HMC], [RO1], [RO3], [TH]. For the case of countable state space MDP and bounded costs, Cavazos-Cadena has shown in [CC2] that commonly used sufficient conditions are extremely restrictive, in that these not only guarantee the existence of a bounded solution to the ACOE for *any* bounded cost function $c(\cdot\,,\,\cdot)$, but they also do so for a whole *family* of MDP. For general state space MDP, Hernández-Lerma et al. [HMC] have given a comprehensive account of recurrence conditions used for the purpose above, and relations among them. On the other hand, it can be shown that a *necessary* condition for the existence of a bounded solution to the ACOE is that the following *uniform boundedness* condition holds.

(UB) There is a constant $M > 0$ such that

$$\left|\Gamma_\beta(p) - \Gamma_\beta(\overline{p})\right| \leq M, \qquad \forall 0 < \beta < 1,\ \ \forall p, \overline{p} \in \Delta.$$

Theorem 3.2: Suppose there is a bounded solution (Γ^*, h) to the ACOE. Then condition (UB) is satisfied with $M = 2 \cdot sp(h)$, where

$$sp(h) := \sup_{p\in\Delta}\{h(p)\} - \inf_{p\in\Delta}\{h(p)\}.$$

For the case when X and Y are both finite, the result above can be inferred from work by Platzman [PL], and for general BMDP, with cost *not necessarily bounded,* it has been shown by Fernández-Gaucherand et al. [FAM4], [FG].

4. The Vanishing Discount Approach Revisited

It is well known that for countable state space MDP with a bounded cost function, a uniform boundedness condition, similar to (UB), is sufficient for bounded solutions to the corresponding ACOE to exist, c.f. [KV, pp.163-165], [BE, pp.311-312], [RO3, pp.95-96]. This is shown using the *vanishing discount approach,* introduced by Taylor [TA] in the context of some replacement problems. In this section, we give conditions that allow the partition of Δ into *countable* subsets, such that the process $\{p_t\}$ will remain in the particular subset containing the given initial distribution p_0. We then use a vanishing discount approach to show that, if in addition, condition (UB) holds, then there exists a bounded solution to the ACOE; furthemore if instead of condition (UB) a weaker condition $(UBGT)$ holds, then there exists a *possibly unbounded* solution to the ACOE.

Note that for fixed $p \in \Delta$ and $u \in U$, the support of $\mathcal{K}(\cdot\,|\,p,u)$ is countable: if $B(p,u) := \{T(y,p,u)\,|\,y \in Y\}$, which is a countable set since Y is countable, then $\mathcal{K}(B(p,u)|p,u) = 1$. Thus, at any time epoch $t \in I\!\!N_0$, the set of possible *next states* for p_t is the set $\bigcup_{u\in U} B(p_t,u)$, which is countable since U is finite. Therefore, even though Δ is an uncountable Borel space, the BMDP $\{\Delta, U, \mathcal{K}, \bar{c}\}$ has a very special structure. This has also been noticed previously by others, c.f. [AS1, p. 187], [SO1, p. 19-20], [PL, p. 369]. However, the first formulation that extensively exploited this fact is that in [FAM3], where it was shown how the analysis could be reduced to a *countable* state space. We substantially improve upon this formulation in the sequel.

Definition 4.1: (i) For each $p \in \Delta$ the sets of *ancestors* and *descendants* of p are defined, respectively, as

$$A_p := \Big\{ s \in \Delta : \exists n \in I\!\!N_0, y^{n+1} = \{y_1, \ldots, y_{n+1}\} \subseteq Y, \\ u^n = \{u_0, u_1, \ldots, u_n\} \subseteq U, \text{ for which } p = T(y^{n+1}, s, u^n) \Big\},$$

and

$$D_p := \Big\{ s \in \Delta : \exists n \in I\!\!N_0, y^{n+1} = \{y_1, \ldots, y_{n+1}\} \subseteq Y, \\ u^n = \{u_0, u_1, \ldots, u_n\} \subseteq U, \text{ for which } s = T(y^{n+1}, p, u^n) \Big\},$$

where the maps above are given recursively as

$$T(y^1, \,\cdot\,, u^0) = T(y_1, \,\cdot\,, u_0),$$
$$T(y^{n+1}, \,\cdot\,, u^n) = T(y_{n+1}, T(y^n, \,\cdot\,, u^{n-1}), u_n); \quad n \in I\!\!N.$$

(ii) The set of *relatives* of $p \in \Delta$ is defined as

$$R_p^{(1)} := A_p \cup \{p\} \cup D_p.$$

Then, $R_p^{(1)}$ contains all the points that either lead to or are derived from p, in any finite number of steps, via repeated applications of $T(y, \cdot, u)$, for any combination of observations in Y and actions in U. Our definition of the sets D_p is an extension, to our context, of Doob's concept of *consequent sets* [DB, p.206], and one may intuitively think of D_p as the "reachable" set, starting at p. Furthermore, note that D_p is a countable set, but A_p is an uncountable set, in general. The following condition will be assumed to derive our main results; it will then be shown that a weaker condition can be used, and the same results are obtained.

(C) For all $y \in Y$, $u \in U$, and $p \in \Delta$, $T^{-1}(y, p, u)$ is a countable set.

Lemma 4.1: Assume that condition (C) holds. Then $R_p^{(1)}$ is a countable set, for any $p \in \Delta$.

Proof: By its definition D_p is countable, for any $p \in \Delta$, since $T(y, \cdot, u)$ is a well defined map for any $y \in Y$ and $u \in U$. Furthermore, since the countable union of countable sets is itself countable, then A_p is countable, under the given assumptions. This gives the result. Q.E.D.

Remark 4.1: Clearly, condition (C) holds if $T(y, \cdot, u)$ is an injective map, for any $y \in Y$ and $u \in U$. For the case when X and Y are both finite, conditions for $T(y, \cdot, u)$ to be injective are given in the following lemma.

Lemma 4.2: Assume that both X and Y are finite sets, and let $y \in Y$ and $u \in U$ be given. The map $T(y, \cdot, u)$ is injective if $\overline{Q}_y(u)$ and $P(u)$ are both nonsingular.

Proof: Let $p, \overline{p} \in \Delta$, then

$$\begin{aligned} & T(y,p,u) = T(y,\overline{p},u) \\ \Leftrightarrow\ & \overline{T}(y,p,u) = \frac{V(y,p,u)}{V(y,\overline{p},u)}\overline{T}(y,\overline{p},u) \\ \Leftrightarrow\ & \overline{Q}_y(u)P'(u)\left(p - \frac{V(y,p,u)}{V(y,\overline{p},u)}\overline{p}\right) = \underline{0} \\ \Leftrightarrow\ & V(y,\overline{p},u)p = V(y,p,u)\overline{p} \end{aligned}$$

Since $\mathbf{1}'p = \mathbf{1}'\overline{p} = 1$, then $V(y,\overline{p},u)p = V(y,p,u)\overline{p}$ if and only if $p = \overline{p}$. Hence the result follows. Q.E.D.

Definition 4.2: For $p \in \Delta$, define its *genealogical tree* GT_p as

$$GT_p := \bigcup_{n \in I\!\!N} R_p^{(n)},$$

where the sets $R_p^{(n)}$ are defined recursively as

$$R_p^{(n+1)} := \bigcup_{s \in R_p^{(n)}} R_s^{(1)}, \quad n \in I\!\!N.$$

Then if condition (C) holds, $R_p^{(n)}$ is a countable set, for each $n \in I\!\!N$, by Lemma 4.1, and thus GT_p is also a countable set.

The following result will be used in subsequent sections.

Lemma 4.3: Let $q, s, p \in \Delta$.

(i) If $q \in R_p^{(n)}$ and $s \in R_q^{(m)}$, for some $n, m \in I\!\!N$, then $s \in R_p^{(n+m)}$.

(ii) If $q \in R_p^{(n)}$, then $p \in R_q^{(n)}$.

(iii) If $q \in GT_p$, then $GT_q = GT_p$.

Proof: (i) Fix $n \in I\!\!N$ arbitrarily. If $q \in R_p^{(n)}$ and $s \in R_q^{(1)}$, then it follows from the definition of $R_p^{(n+1)}$ that $s \in R_p^{(n+1)}$. Proceeding by induction, suppose that for some $m \in I\!\!N$, if $q \in R_p^{(n)}$ and $s \in R_q^{(m)}$, then $s \in R_p^{(n+m)}$. Now let $q \in R_p^{(n)}$ and $s \in R_q^{(m+1)}$; then there exists $r \in R_q^{(m)}$ such that $s \in R_r^{(1)}$. By the induction hypothesis, we have that $r \in R_p^{(n+m)}$, and thus $s \in R_p^{(n+m+1)}$, completing the induction procedure.

(ii) It is clear that if $q \in R_p^{(1)}$, then $p \in R_q^{(1)}$. Proceeding by induction, suppose that for some $n \in I\!\!N$, if $q \in R_p^{(n)}$, then $p \in R_q^{(n)}$, . Let $q \in R_p^{(n+1)}$; then there exists $s \in R_p^{(n)}$ such that $q \in R_s^{(1)}$. Since $s \in R_q^{(1)}$, and by the induction hypothesis $p \in R_s^{(n)}$, then by (i) we conclude that $p \in R_q^{(n+1)}$, completing the induction procedure.

(iii) Let $q \in GT_p$; then $q \in R_p^{(n)}$, for some $n \in I\!\!N$. Let $s \in GT_q$; then $s \in R_q^{(m)}$, for some $m \in I\!\!N$. Hence, by (i) we have that $s \in R_p^{(n+m)}$, and thus $s \in GT_p$. Therefore, we have that $GT_q \subseteq GT_p$. Now, let $s \in GT_p$; then $s \in R_p^{(m)}$, for some $m \in I\!\!N$. Since by (ii) $p \in R_q^{(n)}$, then we conclude from (i) that $s \in R_q^{(n+m)}$, and thus $s \in GT_q$. Therefore, we have that $GT_p \subseteq GT_q$, and hence $GT_q = GT_p$. Q.E.D.

4.1. Controlled Sub-MDP

Let us consider the BMDP specified by $\{\Delta, U, \mathcal{K}, \overline{c}\}$.

Definition 4.3: Let $B \in \mathcal{B}(\Delta)$.

(i) B is said to be *positively invariant* if $D_p \subseteq B$, for all $p \in B$.

(ii) B is said to be *invariant* if it is positively invariant, and if $A_p \subseteq B$, for all $p \in B$.

The above definition of positively invariant sets is an extension, to our context, of the concept as introduced by Doob [DB, p.206]. Very importantly, note that if $B \in \mathcal{B}(\Delta)$ is a positively invariant set, then $\{B, U, \mathcal{K}, \overline{c}\}$ is also a BMDP. Since $B \subseteq \Delta$, then $\{B, U, \mathcal{K}, \overline{c}\}$ will be called a sub-MDP of $\{\Delta, U, \mathcal{K}, \overline{c}\}$. The concept of sub-MDP has been used also by Kurano [KU2],[KU3], in a different context but for similar purposes.

For any $p \in \Delta$, it is clear from our definitions that D_p is the *smallest* positively invariant set containing p, and that GT_p is the *smallest* invariant set containing p. Therefore $\{D_p, U, \mathcal{K}, \overline{c}\}$ and, if condition (C) holds, $\{GT_p, U, \mathcal{K}, \overline{c}\}$ are *countable* state space sub-MDP of $\{\Delta, U, \mathcal{K}, \overline{c}\}$, where for $p_1, p_2 \in D_p$, their state transition matrices have elements given by $\mathcal{K}(\{p_2\} \,|\, p_1, u)$. The idea is then to appropriately use known results for countable state space MDP, in the analysis of the uncountable state space MDP specified by $\{\Delta, U, \mathcal{K}, \overline{c}\}$.

Under condition (UB), it follows that for any $\overline{p} \in \Delta$, $\big|h_\beta(\cdot)\big| \leq M$, uniformly in $\beta \in (0,1)$, where for each $0 < \beta < 1$ and $p \in \Delta$,

$$h_\beta(p) := \Gamma_\beta(p) - \Gamma_\beta(\overline{p}). \tag{4.1}$$

For our purposes, the following conditions can be used instead of condition (UB).

$(UBGT)$ There is a $\overline{p} \in \Delta$, such that for each $p \in \Delta$

$$|h_\beta(s)| \leq M_p, \;\; \forall 0 < \beta < 1, \quad \forall s \in GT_p,$$

for some constant $M_p > 0$.

(UBD) There is a $\overline{p} \in \Delta$, such that for each $p \in \Delta$

$$|h_\beta(s)| \leq M_p, \;\; \forall 0 < \beta < 1, \quad \forall s \in D_p,$$

for some constant $M_p > 0$.

Obviously, (UB) implies $(UBGT)$, and $(UBGT)$ implies (UBD). The next result is similar to those given in [SEN], for countable state space MDP, and in [HL], for BMDP.

Lemma 4.4: Assume that condition (UBD) holds. Then for each $p \in \Delta$, there is a constant $\overline{M}_p > 0$ such that

$$0 \leq (1-\beta)\Gamma_\beta(p) \leq \overline{M}_p, \qquad \forall 0 < \beta < 1.$$

Proof: By Assumption 2.1 and the optimality equation (2.4), we have that for any $u \in U$

$$\begin{aligned} 0 \leq (1-\beta)\Gamma_\beta(p) &\leq \overline{c}(p,u) + \beta \sum_{y \in Y} V(y,p,u)\Gamma_\beta(T(y,p,u)) - \beta\Gamma_\beta(p) \\ &= \overline{c}(p,u) + \beta \sum_{y \in Y} V(y,p,u)h_\beta(T(y,p,u)) - \beta h_\beta(p), \end{aligned}$$

for each $p \in \Delta$ and $0 < \beta < 1$. Thus

$$0 \leq (1-\beta)\Gamma_\beta(p) \leq \overline{c}(p,u) + 2M_p. \qquad \text{Q.E.D.}$$

Since for each $p \in \Delta$, the MDP specified by $\{D_p, U, \mathcal{K}, \overline{c}\}$ has a countable state space, then under assumption (UBD) a *vanishing discount approach* may be followed to obtain the next result.

Theorem 4.1: Assume that condition (UBD) holds, and for any $p \in \Delta$ consider the MDP specified by $\{D_p, U, \mathcal{K}, \overline{c}\}$. Then

(i) there is a constant $\Gamma^* \in I\!R$ and a function $h_{D_p} : D_p \to [-M_p, M_p]$ such that

$$\Gamma^* + h_{D_p}(s) = \min_{u \in U}\left\{\overline{c}(s,u) + \sum_{y \in Y} V(y,s,u)h_{D_p}(T(y,s,u))\right\}, \ \forall s \in D_p; \qquad (4.2)$$

(ii) any (nonrandomized) separated stationary policy that attains the minimum in (4.2) is average cost optimal, one such policy exists, and the minimal average cost is $\Gamma(s) = \Gamma^*$, for all $s \in D_p$;

(iii) we have that

$$(1-\beta)\Gamma_\beta(\overline{p}) \underset{\beta \uparrow 1}{\longrightarrow} \Gamma^*.$$

Proof: Let $\{\beta_n\} \subseteq (0,1)$ be a given sequence, such that $\beta_n \uparrow 1$. By Lemma 4.4 and via the Bolzano-Weierstrass Theorem [BA, p.108], with no loss in generality we may assume that $\{\beta_n\}$ is such that $(1-\beta_n)\Gamma_{\beta_n}(\overline{p})$ converges to a point $\Gamma^* \in [0, \overline{M}_{\overline{p}}]$, that is

$$(1-\beta_n)\Gamma_{\beta_n}(\overline{p}) \underset{\beta_n \uparrow 1}{\longrightarrow} \Gamma^*.$$

Now, let $p \in \Delta$ be chosen and consider the set D_p. For any $s \in D_p$, rewrite the discounted cost optimality equation (2.4) as

$$(1-\beta_n)\Gamma_{\beta_n}(\overline{p}) + h_{\beta_n}(s) = \min_{u \in U}\left\{ \overline{c}(s,u) + \beta_n \sum_{y \in Y} V(y,s,u) h_{\beta_n}(T(y,s,u)) \right\}. \quad (4.3)$$

Since $\left|h_{\beta_n}(s)\right| \leq M_p$, for all $s \in D_p$, then via the Bolzano-Weierstrass Theorem and a Cantor Diagonalization argument [BA, p.24], there is a subsequence $\beta_{n_k} \uparrow 1$ such that

$$h_{\beta_{n_k}}(s) \underset{\beta_{n_k} \uparrow 1}{\longrightarrow} h_{D_p}(s), \;\; \forall s \in D_p,$$

for some function $h_{D_p} : D_p \to [-M_p, M_p]$. Then taking limits in (4.3) as $\beta_{n_k} \uparrow 1$, and using the Bounded Convergence Theorem [BA, p.242], [RY, p.81-82], equation (4.2) is obtained, proving (i). Since $h_{D_p}(\cdot)$ is bounded on D_p, and since U is finite, then by standard arguments, e.g., see [BE, p.311-312], [KV, p.163-165], [RO3, p.95-96], [HLM1, p.52-55], (ii) follows. Finally, since $\beta_n \uparrow 1$ was arbitrary, and since by (ii) Γ^* gives the optimal average cost, (iii) follows. Q.E.D.

Remark 4.2: From the proof of Theorem 4.1, note that Γ^* *does not* depend on the particular set D_p considered, and thus it gives the optimal average cost for $\{\Delta, U, \mathcal{K}, \overline{c}\}$. Therefore condition (UBD) induces a uniformity among all the positively invariant sets D_p, in the sense of equal optimal average costs.

4.2. A Solution to the ACOE

We have shown in Theorem 4.1 that if condition (UBD) holds, then there is a bounded solution to the ACOE *on each* invariant set D_p, for all $p \in \Delta$. Furthermore, under this assumption, the same optimal average cost is attained on each invariant set D_p. However, the method of proof used in Theorem 4.1 does not lead to a well defined function $h(\cdot)$ on all of Δ, since even if for $p_1, p_2 \in \Delta$, such that $D_{p_1} \bigcap D_{p_2} = \phi$, it may be the case that $GT_{p_1} = GT_{p_2}$. Next, an equivalence relation is defined on Δ, which is used to circumvent this problem.

Definition 4.4: Let $p, q \in \Delta$. We say that $p \sim q$ if $GT_p = GT_q$.

It follows trivially that "$\sim$" defines an *equivalence relation* on Δ, see [RY, p. 22]. Furthermore, from Lemma 4.3(iii), we see that the equivalence classes defined by "$\sim$" are the sets GT_p. Hence, if condition (C) holds, the collection of equivalence classes $\Delta/\sim$ is a partition of Δ into *countable* subsets. Therefore, if conditions (C) and $(UBGT)$ hold, the method of proof of Theorem 4.1 can be used to obtain the following.

Theorem 4.2: Assume that conditions (C) and $(UBGT)$ hold. Then

(i) there is a solution (Γ^*, h), with $\Gamma^* \in I\!\!R$, to the ACOE

$$\Gamma^* + h(p) = \min_{u \in U}\left\{\overline{c}(p,u) + \sum_{y \in Y} V(y,p,u)h(T(y,p,u))\right\}, \forall p \in \Delta; \tag{4.4}$$

(ii) $h : \Delta \to I\!\!R$ is such that for each $p \in \Delta$, $\left|h(s)\right| \leq M_p$, for all $s \in GT_p$;

(iii) any (nonrandomized) stationary separated policy that attains the minimum in (4.4) is average cost optimal, one such policy exists, and the minimal average cost is $\Gamma(p) = \Gamma^*$, for all $p \in \Delta$;

(iv) we have that

$$(1-\beta)\Gamma_\beta(\overline{p}) \xrightarrow[\beta \uparrow 1]{} \Gamma^*.$$

Proof: Let $\{\beta_n\} \subseteq (0,1)$ be a given sequence, such that $\beta_n \uparrow 1$, and as in Theorem 4.1, assume with no loss in generality that

$$(1-\beta_n)\Gamma_{\beta_n}(\overline{p}) \xrightarrow[\beta_n \uparrow 1]{} \Gamma^*.$$

As in the proof of Theorem 4.1(i), a function $h_{GT_p} : GT_p \longrightarrow [-M_p, M_p]$ can be obtained, for each $p \in \Delta$, such that (Γ^*, h_{GT_p}) is a solution to the ACOE on GT_p. Since Δ is partitioned into the sets GT_p, a function $h : \Delta \longrightarrow I\!\!R$ can be well defined as

$$h(p) := h_{GT_p}(p), \quad \forall p \in \Delta.$$

Then, (Γ^*, h) is a solution to the ACOE on all of Δ, showing (i). Furthermore, under condition $(UBGT)$, parts (ii)-(iv) follow as in Theorem 4.1. Q.E.D.

Remark 4.3: Note that if condition (UB) holds, then from Theorem 4.2(ii), we conclude that a bounded solution to the ACOE exists. On the other hand, condition $(UBGT)$ guarantees the existence of a *possibly unbounded* solution to the ACOE;

the latter resembles other recent results in the area of MDP with an average cost criterion, e.g., [BOR], [GM], [HL], [SEN].

The results in Theorem 4.1 and Theorem 4.2 were derived under the assumption that condition (C) held. These results remain valid under the assumption that the following weaker condition holds.

(C') The action set U can be partitioned as $U = U_1 \cup U_2$, such that:
(i) $T^{-1}(y, p, u)$ is a countable set, for all $y \in Y$, $u \in U_1$, and $p \in \Delta$,
(ii) the set $\{T(y, p, u) | y \in Y, p \in \Delta, u \in U_2\}$ is finite.

Corollary 4.1: Assume that conditions (C') and $(UBGT)$ hold. Then the conclusions of Theorem 4.2 hold.

Proof: Let $\{\beta_n\} \subseteq (0, 1)$ be a given sequence, such that $\beta_n \uparrow 1$, and as in Theorem 4.1, assume with no loss in generality that

$$(1 - \beta_n)\Gamma_{\beta_n}(\overline{p}) \underset{\beta_n \uparrow 1}{\longrightarrow} \Gamma^*.$$

By our hypotheses, $\big\{h_{\beta_n}(T(y, p, u)|y \in Y, p \in \Delta, u \in U_2\big\}$ is a finite set, which is bounded by $M := \min\big\{M_{T(y,p,u)}|y \in Y, p \in \Delta, u \in U_2\big\}$, uniformly in β_n, and thus there is a subsequence $\beta_{n_k} \uparrow 1$ such that

$$h\beta_{n_k}(T(y, p, u)) \underset{\beta_{n_k} \uparrow 1}{\longrightarrow} h(T(y, p, u)).$$

Then using $\{\beta_{n_k}\}$, the proofs of Theorem 4.1 and Theorem 4.2 can be paralleled for the terms corresponding to $u \in U_1$ in (4.3), leading to well defined limits $h(p)$, for all $p \in \Delta$. Q.E.D.

Remark 4.4: The set in (ii) of condition (C') may also be taken to be countable, but then condition (UB) must hold in order to obtain the results of Corollary 4.1.

Next, a simple generalization of the concept of Blackwell optimality, as given in [BE, p.339], is formulated for particular actions. This result will be needed in subsequent sections.

Lemma 4.5: Assume that condition (UBD) holds. Then

(i) for a given $p \in \Delta$, it is average cost optimal to take action "a" at p, denoted as $\mu^*(p) = a$, if there is a sequence $\{\beta_n\} \subseteq (0, 1)$, with $\beta_n \uparrow 1$, such that it is β_n-discount optimal to take action "a" at p, denoted as $\mu^*_{\beta_n}(p) = a$;

(ii) if $\mu^*(p) = a$ is the *only* average cost optimal action to take at $p \in \Delta$, then there is a sequence $\{\beta_n\} \subseteq (0,1)$, with $\beta_n \uparrow 1$, such that it is β_n-discount optimal to take action $\mu^*_{\beta_n}(p) = a$ at $p \in \Delta$.

Proof: (i) Let $\{\beta_n\} \subseteq (0,1)$, with $\beta_n \uparrow 1$, and assume that $\mu^*_{\beta_n}(p) = a$ attains equality in (4.3). Then, by taking limits along an appropriate subsequence $\{\beta_{n_k}\}$, as in the proof of Theorem 4.1, we conclude that action $\mu^*(p) = a$ attains equality in (4.1), and hence is average cost optimal, by Theorem 4.1 and Remark 4.2.
(ii) Suppose that $\mu^*(p) = a$ is the only action which is average cost optimal at $p \in \Delta$, but that there does not exists a sequence $\{\beta_n\} \subseteq (0,1)$, with $\beta_n \uparrow 1$, such that $\mu^*_{\beta_n}(p) = a$. Then, since the action set is finite, for each sequence $\{\beta_n\} \subseteq (0,1)$, with $\beta_n \uparrow 1$, there is a subsequence $\beta_{n_k} \uparrow 1$, such that $\mu^*_{\beta_{n_k}}(p) = u$, for some $u \neq a$, which by part (ii) leads to a contradiction. Q.E.D.

5. Monotone POMDP

Let $\{\Delta, U, \mathcal{K}, \overline{c}\}$ be the (equivalent) specification of a POMDP. Suppose that a partial order "$\prec$" has been defined on Δ, and let "$\prec_u$" denote the linear order given for U. We use the notation $\{(\Delta, \prec), (U, \prec_u), \mathcal{K}, \overline{c}\}$ to make explicit what specific order relations are being used.

Definition 5.1: Consider $\{(\Delta, \prec), (U, \prec_u), \mathcal{K}, \overline{c}\}$, and let $p_1, p_2 \in \Delta$. We say that:

(i) the value functions are *monotone* if

$$p_1 \prec p_2 \Rightarrow \Gamma_\beta(p_1) \leq \Gamma_\beta(p_2), \quad \forall\ 0 < \beta < 1,$$

(ii) a (nonrandomized) stationary separated policy μ is *monotone* if

$$p_1 \prec p_2 \Rightarrow \mu(p_1) \prec_u \mu(p_2).$$

Two frequently used partial orders on Δ are the *stochastic dominance* $\prec_{st}$ and the *monotone likelihood ratio* $\prec_{lr}$, defined below.

Definition 5.2: Let $p_1, p_2 \in \Delta$; we say that:

(i) $p_1 \prec_{st} p_2$ if $\displaystyle\sum_{i \geq q} p_1^{(i)} \leq \sum_{i \geq q} p_2^{(i)}$, for all $q \in X$, and

(ii) $p_1 \prec_{lr} p_2$ if $p_1^{(j)} p_2^{(i)} \leq p_1^{(i)} p_2^{(j)}$, for all $i, j \in X$ such that $i \leq j$.

Remark 5.1: When $N_X = 1$, it follows easily that $\prec_{st}$ and $\prec_{lr}$ are equivalent *total* orders on Δ; furthermore, by (uniquely) identifying each $p \in \Delta$ with its second component, then $\prec_{st}$ and $\prec_{lr}$ are equivalent to the standard order in $I\!R$.

Let e^j denote the element of Δ with the j^{th} component equal to 1, $j \in X$; thus, e.g., $e^0 = (1, 0, 0, \ldots)$. The following is easily shown.

Lemma 5.1: (i) If $p_1, p_2 \in \Delta$ and $p_1 \prec_{lr} p_2$, then $p_1 \prec_{st} p_2$.

(ii) For all $p \in \Delta$, $e^0 \prec_{lr} p$.

Structural properties of sequential decision processes, e.g., the monotonicity notions defined above, give very useful information that may be advantageously used in several ways. For example, monotonicity of the value functions and optimal strategies may be exploited to design efficient computational algorithms, or it may suggest suboptimal strategies, which may be simple to implement but yield a near optimal performance. In a very general framework, Porteus has given conditions that imply that *both* value functions and optimal control strategies exhibit given structural properties [KPO], [PO1], [PO2]. As will be shown, monotonicity properties of the value function and the availability of *reset actions* can be combined to yield that condition (UB) holds.

5.1. Reset Actions

An action $u_j \in U$ is called a reset action if, for some $j \in X$, $T(y, p, u_j) = e^j$, for all $y \in Y$ and $p \in \Delta$. This corresponds to the core state of the system being j, with probability one, at the next time epoch after action u_j has been taken. Hence $P'(u_j)$ has all columns equal to e^j. This type of action arises naturally in manufacturing systems subject to inspection, maintenance, and replacement. The following results derive from the work of Sondik [SO1].

Lemma 5.2: (i) If there exists a reset action $u_j \in U$, then

$$\Gamma_\beta(p) - \Gamma_\beta(e^j) \le \overline{c}(p, u_j), \quad \forall p \in \Delta, \tag{5.1}$$

(ii) if X is finite, and for each $j \in X$ there is a corresponding reset action, then for each $\beta \in (0, 1)$ there exists $J \in X$ such that

$$0 \le \Gamma_\beta(p) - \Gamma_\beta(e^J) \le M, \quad \forall p \in \Delta,$$

where $M := \max\big\{c(i, u) \,|\, i \in X, u \in U\big\}$.

Proof: (i) We have that

$$\begin{aligned}\Gamma_\beta(p) &= \min_{u\in U}\left\{\overline{c}(p,u)+\beta\sum_{y\in Y}V(y,p,u)\Gamma_\beta(T(y,p,u))\right\}\\ &\le \overline{c}(p,u_j)+\Gamma_\beta(e^j),\end{aligned}$$

and the result follows. (ii) Recall that $\Gamma_\beta(\cdot)$ is concave, for each $\beta\in(0,1)$, and thus it attains its minimum at a (vertex) e^j, since X is finite. Let $J\in X$ be such that the minimum is attained at e^J, and note that it depends on $\beta\in(0,1)$, in general. Since costs are assumed to be positive, then

$$\begin{aligned}0\le\Gamma_\beta(p)-\Gamma_\beta(e^J)&\le\overline{c}(p,u_J)+\beta\Gamma_\beta(e^J)-\Gamma_\beta(e^J)\\ &\le M-(1-\beta)\Gamma_\beta(e^J)\le M,\end{aligned}$$

and thus the result follows as above, see [SO1, p.195-196]. Q.E.D.

Remark 5.2: Note that if $\Gamma_\beta(\cdot)$ is monotone with respect to $\prec_{lr}$, and if there is an action $u_0\in U$ that resets the state to e^0, then $0\le\Gamma_\beta(p)-\Gamma_\beta(e^0)\le\overline{c}(p,u_0)$, uniformly in $\beta\in(0,1)$. Thus, if $\overline{c}(s,u_0)\le M_p$, for all $s\in GT_p$, for some constant $M_p>0$, then condition $(UBGT)$ holds. Furthermore, note that when X is finite, a constant $M>0$ exists such that $\overline{c}(p,u_0)\le M$, for all $p\in\Delta$, and thus condition (UB) holds.

Models with a *replacement* action that resets the system to an "as new" state e^0 have been considered in, e.g., [AKU], [LO2], [OKM], [OMK], [RO2], [WA1], [WA2], [W1], [W2], [W3], [W4], [W5]. Related problems are those considered in [FI], where a reset action to a most desirable state is available, and in [HW], where (maintenance) reset actions u_j are available for all $j\ne0$, with X a finite set. Also, two-state replacement problems have been considered in the context of *adaptive control* in [FAM1], [FAM2], [GE2], [HM], [MFA]. Condition (C') is satisfied for most of the cases considered in, e.g., [AKU], [FAM1], [FAM2], [FI], [HW], [MFA], [RO2], [W1], [WA1], [WA2], and thus the results of Theorem 4.2 and Corollary 4.1 are applicable to a large number of the models above. Some of the models above do not satisfy condition (C'); however, a special case of each of these, namely when $\overline{Q}_y(u)$ and $P(u)$ are injective, for all $y\in Y$ and all non-reset actions u, leads to condition (C') being satisfied, by Lemma 4.2.

6. The Two-State Replacement Problem

We consider now in detail a replacement problem for a production unit in a manufacturing system, that may be in either of two states: a good (0) or a failed (1) state. The three available actions are to produce with the current unit (0), to produce with the current unit and at the same time inspect it (1), or to replace it with a new machine (2), where the associated costs are $c(0,0) = 0$, $c(1,0) = C$, $c(i,1) = I$ and $c(i,2) = R$, $i = 0,1$, with the natural assumption that $0 < C < I < R$. If initially the unit is not failed, and no replacement actions are taken, it will fail at a random time; the model of the POMDP takes the form

$$P(u) = \begin{bmatrix} 1-\theta & \theta \\ 0 & 1 \end{bmatrix}, u = 0,1; \quad P(2) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix};$$

$$Q(u) = \begin{bmatrix} q_u & 1-q_u \\ 1-q_u & q_u \end{bmatrix}, u = 0,1,2,$$

where $\theta \in [0,1]$ and $q_u \in [0.5,1]$. The cases $q_u = 1/2$ and $q_u = 1$ correspond to the completely unobservable (CU) and completely observable (CO) processes, respectively; furthermore, by the symmetry of $Q(u)$ we need only consider q_u as specified; since for $u = 2$ the state of the machine is good w.p.1 by the next decision epoch, then the value of q_2 is unimportant. Also it is natural to expect θ to be some small positive number.

The conditional probability vector can be written as $p_t = [1-\rho_t, \rho_t]'$, where ρ_t is the conditional probability of the unit being in failed. Hence, the problem reduces to a scalar one, in terms of ρ_t. Given an *a priori* probability ρ of the unit being failed, we have the following one-step ahead conditional probabilities for the observations (with obvious modifications in the definitions for this scalar case), for $1/2 \leq q_u \leq 1$, $u = 0,1$:

$$V(0,\rho,u) = q_u(1-\rho)(1-\theta) + (1-q_u)[\rho(1-\theta)+\theta],$$

$$V(1,\rho,u) = (1-q_u)(1-\rho)(1-\theta) + q_u[\rho(1-\theta)+\theta],$$

and when these quantities are nonzero, the *a posteriori* conditional probabilities of the system being in the bad state are given by

$$T(0,\rho,u) = \frac{(1-q_u)[\rho(1-\theta)+\theta]}{V(0,\rho,u)},$$

$$T(1,\rho,u) = \frac{q_u[\rho(1-\theta)+\theta]}{V(1,\rho,u)},$$

$$T(y,\rho,2)=0;\quad y=0,1,\ \rho\in[0,1].$$

For $q_u=1$, $u=0,1$, we get that $V(0,\rho,u)=0$, for $\rho=1$ or $\theta=1$, hence we *define* $T(0,\rho,u)=1$; also $V(1,\rho,u)=0$, for $\rho=\theta=0$ and thus we *define* $T(1,\rho,u)=0$. Otherwise, we have for $q_u=1$, $u=0,1$

$$T(y,\rho,u)=y.$$

If $q_u={}^1\!/_2$, we get

$$T(y,\rho,u)=\rho(1-\theta)+\theta=1-(1-\rho)(1-\theta)=:T(\rho),$$

for $y=0,1$. Note that if $\theta=0$, then any $\rho\in[0,1]$ is a fixed point of $T(\cdot)$, which says that our initial knowledge of the state of the system never changes in this situation; if $\theta\neq 0$, the unique fixed point of $T(\cdot)$ is $\overline{\rho}=1$, this can be seen by iterating the map

$$T^0(\rho):=\rho,$$
$$T^n(\rho):=T(T^{n-1}(\rho))=1-(1-\rho)(1-\theta)^n;\quad n\in I\!N,$$

and by noticing that under motion by $T^n(\cdot)$, ρ_t monotonically tends towards $\overline{\rho}=1$, i.e. $\rho<T(\rho)$, $\rho\neq 1$.

With $\rho_0:=[0\ 1]p_0$, we have

$$\rho_{t+1}=T(1,\rho_t,u_t)y_{t+1}+T(0,\rho_t,u_t)(1-y_{t+1}),$$

where u_t is the decision made at time epoch t. We now study some important properties of $T(y,\,\cdot\,,u),y,u\in\{0,1\}$ for the case when $q_u\in(0.5,1)$, see also [AKU], [W1]. Define

$$f_0^{(u)}(\rho):=T(0,\rho,u)-\rho,\qquad f_1^{(u)}(\rho):=T(1,\rho,u)-\rho.$$

Thus, the roots of $f_0^{(u)}(\cdot)$ and $f_1^{(u)}(\cdot)$ are the fixed points of $T(0,\,\cdot\,,u)$ and $T(1,\,\cdot\,,u)$, respectively. The pertinent quadratic equation for $f_0^{(u)}(\cdot)$ is

$$(\xi_0^{(u)})^2(2q_u-1)(1-\theta)-\xi_0^{(u)}\left[(2q_u-1)(1-\theta)+(1-q_u)\theta\right]+(1-q_u)\theta=0,$$

and for $\theta\neq 1$ its roots are

$$\overline{\xi}_0^{(u)}=1;\qquad \xi_0^{(u)}=\frac{(1-q_u)\theta}{(2q_u-1)(1-\theta)}.$$

For $\theta = 1$, both roots are equal to 1. Replacing q_u by $(1 - q_u)$ above, the expressions corresponding to $f_1^{(u)}(\cdot)$ are obtained.

Lemma 6.1: Let $q_u \in (0.5, 1)$; if $\theta \in (0, 1)$, the following holds:

(i) $\big|V(y, \rho, u) - V(y, \rho', u)\big| \leq (2q_u - 1)(1 - \theta)|\rho - \rho'|$, for $u = 0, 1$ and $y = 0, 1$;

(ii) $\big|T(y, \rho, u) - T(y, \rho', u)\big| \leq (\frac{q_u}{1-q_u})(1 - \theta)|\rho - \rho'|$, for $u = 0, 1$, and $y = 0, 1$;

(iii) $T(y, \cdot, u)$ is monotone increasing in $[0, 1)$, for each $u = 0, 1$ and $y = 0, 1$;

(vi) $T(0, \rho, u) < T(1, \rho, u)$, $\rho \in [0, 1)$, $u = 0, 1$;

(v) $\rho < T(1, \rho, u)$, for $\rho \in [0, 1)$, $u = 0, 1$;

(vi) if $\frac{1}{2-\theta} < q_u$, then $\xi_0^{(u)} \in [0, 1)$, $\rho_1 < T(0, \rho_1, u)$ and $T(0, \rho_2, u) < \rho_2$, for $\rho_1 \in [0, \xi_0^{(u)})$ and $\rho_2 \in (\xi_0^{(u)}, 1)$, $u = 0, 1$;

furthermore, if $\theta \neq 0$, we have that

(vii) if $q_u \leq \frac{1}{2-\theta}$, then $\xi_0^{(u)} \geq 1$ and $\rho < T(0, \rho, u)$ for $\rho \in [0, 1)$, $u = 0, 1$.

(viii) $T(0, 0, 0) < \xi_0^{(0)}$; also $\xi_0^{(0)} < \xi_0^{(1)}$ if and only if $q_1 < q_0$.

Proof: Since $V(y, \rho, u) = [q_u(1 - \theta) + (1 - q_u)\theta] + \rho(1 - \theta)(1 - 2q_u)$, then part (i) follows directly. Also, from the latter expression we see that $V(y, \rho, u)$ is decreasing in θ, since $1 - 2q_u < 0$; thus $V(y, \rho, u) < 1 - q_u$, the last quantity being the value of $V(y, \rho, u)$ when $\theta = 1$. Then (ii) follows by simple algebraic manipulations on the numerator of the pertinent expression. Part (iii) is shown by using the following fact: If a function $h : D \subseteq I\!R \to I\!R$ satisfies the property that, for some scalars a, b, c, d, $h(x) = (ax + b)/(cx + d)$ with $cx + d \neq 0$ for all $x \in D$, then h is monotone increasing (nondecreasing) on D if $bc < ad$ ($bc \leq ad$). By a similar argument, considering q_u as the variable, and since $(1 - q_u) < q_u$ for $q_u \in (0.5, 1)$, (iv) also follows [W1]. We have that

$$\xi_1^{(u)} = \frac{q_u\theta}{(1 - 2q_u)(1 - \theta)},$$

is one of the fixed points of $f_1^{(u)}(\cdot)$, and since $q_u \in (0.5, 1)$, then $\xi_1^{(u)} \leq 0$. It is easy to see then that $f_1^{(u)}(\rho) > 0$ for $\rho \in (\xi_1^{(u)}, 1)$, giving part (v). Writing

$$\xi_0^{(u)} = \frac{(1 - q_u)\theta}{(2q_u - 1)(1 - \theta)} = \frac{(1 - q_u)\theta}{(1 - q_u)\theta + [q_u(2 - \theta) - 1]},$$

then we have that $0 \leq \xi_0^{(u)} < 1$ for $1/2 \leq \frac{1}{2-\theta} < q_u < 1$ and, if $\theta \neq 0$, $1 \leq \xi_0^{(u)}$ for $1/2 < q_u \leq \frac{1}{2-\theta}$. Then parts (vi) and (vii) follow by analyzing the sign of $f_0^{(u)}(\cdot)$ on

$[0,1)$. Finally, part (viii) follows by simple algebraic operations on the expressions for the given quantities. Q.E.D.

Remark 6.1: When $\theta = 1$ (not a very interesting situation), it follows that $T(0,\,\cdot\,,u) = T(1,\,\cdot\,,u) = 1$ on $[0,1]$, for $u = 0,1$. Also, (v) and (vi) above can be interpreted as meaning that an observation of the process being in the bad state is trusted more than an observation of the process being in the good state.

The CU case, i.e. $q_0 = q_1 = 1/2$, has a deterministic nature in the sense that $T(\cdot)$ is fixed and ρ_t moves towards $\rho = 1$ monotonically (if $\theta \neq 0$). Also, when $\theta \neq 0$ and $1/2 < q_u \leq \frac{1}{2-\theta}$, we have from (vii) in Lemma 6.1 that the ρ_t moves monotonically towards $\rho = 1$; since θ is to be expected to be positive but small, then this behavior is due to the process being nearly completely unobservable.

The following is the cost structure for the equivalent problem:

$$\begin{aligned} \overline{c}(\rho,0) &= \rho C &&: \text{ Produce,} \\ \overline{c}(\rho,1) &= I &&: \text{ Inspect,} \\ \overline{c}(\rho,2) &= R &&: \text{ Replace.} \end{aligned}$$

Thus $0 \leq \overline{c}(\rho,u) \leq R$, for all $\rho \in [0,1]$, $u = 0,1,2$, satisfying Assumption 2.2 trivially. Furthermore, as shown in [W1, Corollary 5.6], we have that $\Gamma_\beta(\cdot)$ is monotone nondecreasing in $\rho \in [0,1]$, and thus monotone with respect to $\prec_{lr}$. Hence, for any $\rho \in [0,1]$

$$0 \leq \Gamma_\beta(0) \leq \Gamma_\beta(\rho) \leq \Gamma_\beta(1) \leq R + \beta\Gamma_\beta(0). \tag{6.1}$$

The rightmost inequality becomes an equality if it is optimal to replace at $\rho = 1$. The latter is guaranteed if [W1]

$$R \leq \frac{C(1+\beta\theta)}{1-\beta(1-\theta)}.$$

Hence, from our discussion of monotone MDP with reset actions, we conclude that condition (UB) holds.

For the case when $1/2 \leq q_0, q_1 < 1$ and $\theta \in (0,1)$, we obtain from Lemma 4.2 that $T(y,\,\cdot\,,u)$ is injective, for $y,u \in \{0,1\}$; also, we have that $T(y,\rho,2) = 0$, for all y,ρ, and thus the term corresponding to $u = 2$ in the discounted cost optimality equation is given by $R + \beta\Gamma_\beta(0)$. Hence, condition (C') holds with $U_2 = \{2\}$, and applying Corollary 4.1 the corresponding ACOE is obtained as

$$\begin{aligned}\Gamma^* + h(\rho) = \min\Bigg\{ \rho\, C + \sum_{y=0}^{1} V(y,\rho,0)\, h(T(y,\rho,0))\,; \\ I + \sum_{y=0}^{1} V(y,\rho,1)\, h(T(y,\rho,1))\,; \quad R \Bigg\},\end{aligned} \tag{6.2}$$

where $\overline{\rho} = 0$ was taken as a reference state. Also, when $1/2 \leq q_0 < q_1 = 1$ we similarly obtain, with $U_2 = \{1,2\}$, that

$$\begin{aligned}\Gamma^* + h(\rho) = \min\Bigg\{ \rho\, C + \sum_{y=0}^{1} V(y,\rho,0)\, h(T(y,\rho,0))\,; \\ I + V(1,\rho,1)\, h(1)\,; \quad R \Bigg\},\end{aligned} \tag{6.3}$$

is the ACOE, where $\overline{\rho} = 0$.

6.1. The Structure of Optimal Policies

For the two-state replacement problem, we wish to determine structural properties of average cost optimal policies, by using (6.2) and (6.3). Recall from Theorem 2.1 that $h_\beta(\cdot)$ is concave; furthermore, it can be shown that $h_\beta(\cdot)$ is nondecreasing in $\rho \in [0,1]$, c.f. [W1]. In [RO2], [W1], structural properties where obtained for *discounted cost* optimal policies, by taking advantage of the concavity of $h_\beta(\cdot)$. Even though it can be established by other methods [OMK], by our method of proof we cannot conclude that the function $h(\cdot)$ in Theorem 4.2 is concave, since its values at different points are obtained as a limits of values of $h_\beta(\cdot)$ along *possibly different* sequences $\beta_n \uparrow 1$. This precludes us from following a similar method as in [RO2], [W1]. However, the Bolzano-Weierstrass Theorem allows us to follow a different approach; we exclude the uninteresting cases $\theta = 0$, and $\theta = 1$. The following condition will be needed for some of our results.

(R) There exists $0 < \beta < 1$ such that

$$R < \frac{C(1+\beta\theta)}{1-\beta(1-\theta)}.$$

Theorem 6.1: Let $\theta \in (0,1)$; then:

(i) If every average cost optimal policy replaces at $\rho \in [0,1]$, then every average cost optimal policy replaces in the interval $[\rho, 1]$.

(ii) Let $\xi := \min\{\xi_0^{(1)}, T(0,0,0)\}$; then it is average cost optimal to produce in the interval $[0,\xi)$.

(iii) If $\xi_0^{(1)} < T(0,0,0)$, then there is an average cost optimal policy that does not replace in the interval $[\xi_0^{(1)}, T(0,0,0))$.

(iv) If $q_1 \leq q_0$, then there is an average cost optimal policy that does not inspect at any $\rho \in [0,1]$.

(v) If condition (R) holds, then there exists a number $\xi \leq \alpha_R < 1$ such that it is average cost optimal to replace in the interval $[\alpha_R, 1]$.

(vi) If condition (R) holds, and $q_0 < q_1 = 1$, then there are numbers $0 \leq \alpha_P \leq \alpha_I \leq \alpha_R < 1$, such that it is average cost optimal to produce for $\rho \in [0,\alpha_P) \cup [\alpha_I, \alpha_R)$, to inspect for $\rho \in [\alpha_P, \alpha_I)$, and to replace for $\rho \in [\alpha_R, 1]$.

(vii) If condition (R) does not hold, then it is optimal to produce for all $\rho \in [0,1]$.

Proof: (i) Since the term in the β-discounted cost optimality equation corresponding to the action to replace is constant, and since $\Gamma_\beta(\cdot)$ is concave, then it is simple to show that the β-discounted replace *region,* i.e., the set of points in $[0,1]$ at which it is β-discounted optimal to replace, is either empty, or an interval $[\alpha_R(\beta), 1]$, with $0 < \alpha_R(\beta) < 1$, see [AKU], [LO1], [RO2], [W1]. Now, if the action to replace is the only average cost optimal action to take at $\rho \in [0,1]$, then there is a sequence $\beta_n \uparrow 1$ such that it is β_n-discount optimal to replace at ρ, by Lemma 4.5(ii), but then every β_n-discount optimal policy replaces in the interval $[\rho, 1]$, and the result follows by Lemma 4.5(i).

(ii) It is known that it is always discount optimal to produce at $\rho = 0$, and thus it is average cost optimal to produce at this point, by Lemma 4.5(i). We consider next the nontrivial case when $\xi \neq 0$. From [AKU, Lemma 2], it is known that it is not β-discount optimal to inspect at $\rho \in [0, \xi_0^{(1)})$, for any $0 < \beta < 1$. Thus it is average cost optimal to either produce or replace at each ρ in this interval, by Lemma 4.5(i). Let $\rho \in [0,\xi)$, and suppose that it is not average cost optimal to produce at ρ, then it is average cost optimal to replace in $[\rho, 1]$, by (i). By Lemma 6.1(ii), it is seen that the average cost accrued by this policy is $\Gamma^* = {}^{R}\!/_{2}$, since $T(y,\rho,0) \geq \xi$, $y = 0,1$. But then a policy that produces in $[0,\xi)$ and replaces in $[\xi, 1]$, also accrues an average cost of $\Gamma^* = {}^{R}\!/_{2}$, and hence it is optimal to produce at ρ, giving the result.

(iii) Suppose that there is an average cost optimal policy which takes action to replace at some $\rho \in [\xi, T(0,0,0))$. Then as in (ii), it is seen that a policy that produces in $[0, T(0,0,0))$ accrues the same average cost, giving the result.

(iv) It is known that for this situation it is not β-discount optimal to inspect at any $\rho \in [0,1]$, for any $0 < \beta < 1$, [AKU], [W1]. Then the result follows by Lemma 4.5(i).

(v) For $0 < \beta < 1$, it is known that every β-discount optimal policy replaces in a neighborhood of $\rho = 1$ if and only if the cost to replace is such that

$$0 < R < \frac{C(1+\beta\theta)}{1-\beta(1-\theta)} =: H(\beta), \tag{6.4}$$

(see [AKU] and [W1]). Furthermore, as in the proof of Lemma 6.1(i), it is easily shown that $H(\cdot)$ is monotone increasing on $(0,1)$. Then, $H(\beta) \uparrow \frac{C(1+\theta)}{\theta}$, and by our assumption there is an $0 < \varepsilon < 1$ such that (6.4) holds for all $1 - \varepsilon \leq \beta < 1$. Let $\{\beta_n\} \subseteq (1-\varepsilon, 1)$ be such that $\beta_n \uparrow 1$; then by Lemma 4.5(i) it follows that it is average cost optimal to replace at $\rho = 1$. Furthermore, we claim that there is a number $\xi \leq \alpha_R < 1$ such that it is average cost optimal to replace in the interval $[\alpha_R, 1]$. We argue by contradiction: suppose that $\alpha_R = 1$; we have that $W_\beta(\rho) - W_\beta(0) \to C\rho/\theta$, for $\rho \neq 0$, and $(1-\beta)W_\beta(0) \to C$, as $\beta \uparrow 1$, where $W_\beta(\cdot)$ is the β-discounted cost accrued by the policy that produces for all $\rho \in [0,1)$ (see [W1] for an expression of $W_\beta(\cdot)$). Then, by direct substitution into (6.2), it is shown that $\Gamma^* := C$ and $h(\rho) := C\rho/\theta$ solve the ACOE, under our hypothesis that, for $\rho \in [0,1)$,

$$C\rho + \sum_{y=0}^{1} V(y,\rho,0)\left[\frac{CT(y,\rho,0)}{\theta}\right] = C\rho + \frac{C\left[\rho(1-\theta)+\theta\right]}{\theta} \leq R.$$

Then, letting $\rho \uparrow 1$ we obtain $\frac{C(1+\theta)}{\theta} \leq R$, which contradicts the hypothesis. This establishes (v).

(vi) If $q_1 = 1$, then to each $0 < \beta < 1$ there correspond numbers $0 < \alpha_P(\beta) \leq \alpha_I(\beta) \leq \alpha_R(\beta) < 1$ such that it is β-optimal to produce for $\rho \in [0, \alpha_P(\beta)) \cup [\alpha_I(\beta), \alpha_R(\beta))$, to inspect for $\rho \in [\alpha_P(\beta), \alpha_I(\beta))$, and to replace for $\rho \in [\alpha_R(\beta), 1]$. Then, given a sequence $\beta_n \uparrow 1$, there is a subsequence $\{\beta_{n_k}\}$ such that, as $\beta_{n_k} \uparrow 1$

$$\alpha_P(\beta_{n_k}) \longrightarrow \alpha_P, \quad \alpha_I(\beta_{n_k}) \longrightarrow \alpha_I, \quad \alpha_R(\beta_{n_k}) \longrightarrow \alpha_R,$$

and $0 \leq \alpha_P \leq \alpha_I \leq \alpha_R \leq 1$. Also, by similar arguments as in (v), we have that $\alpha_R < 1$.

(vii) Finally, if condition (R) does not hold, then it is β-discount optimal to produce for all $\rho \in [0,1]$, see [W1]. The result follows by Lemma 4.5(i). Q.E.D.

Remark 6.2: Theorem 6.1 extends the following results found in the literature.

(i) Those obtained by Ross in [RO2], for a particular model with $q_0 = 1/2$, $q_1 = 1$, where results similar to those in (vi) of Theorem 6.1 are obtained by showing that $\{h_\beta(\cdot)\}$ is uniformly bounded and equicontinuous.

(ii) Results by Wang [WA1], similar to those in (v) of Theorem 6.1, obtained for the model of [RO2], but with no inspection actions allowed.

(iii) Results by Andriyanov et al. [AKU] and Lovejoy [LO2], where structural results are given for discounted cost optimal policies.

(iv) Several results by White, e.g. those in [W1], where discounted costs are considered for the infinite horizon, and also equicontinuity of $\{h_\beta(\cdot)\}$ is shown for some values of q_0, q_1, and θ, but structural results are not given for the average cost optimal policy; and those in [W2], [W3] and [W5], where a restriction to a countable set in Δ is imposed, for the average cost case.

(v) Results by Georgin [GE2], who established structural properties of the average cost optimal policy and of $h(\cdot)$ for the case when $q_0 = 1/2$ and $q_1 = 1$, as studied in [RO2].

(vi) Results by Ohnishi et al. [OMK], where analogs to (i) and (vi) in Theorem 6.1 were obtained, for a model with an arbitrary, but finite, number of states, and perfect observations under inspection.

(vii) Results by Albright, who obtained analogs to (i) and (iv), for a model with $q_u = q$, for all $u \in U$, $q > 1/2$, but allowing for possibly uncountably many actions.

7. Comments on Other Approaches

In the study of POMDP, a prevalent approach in the past has been to view these as Borel state space MDP, and then use results from this general theory, c.f. [GE2], [HM], [HW], [RO2], [SY]. One shortcoming of this approach, specially when an average cost criterion is used, is that the types of conditions that are usually employed in order to derive results for general BMDP models, are very restrictive or difficult to verify. For example, for MDP with a countable state space, much research has been devoted to the problem of finding bounded solutions to the ACOE, and necessary and sufficient conditions for solutions to exist have been recently given by Cavazos-Cadena [CC1]. However, for BMDP this problem is much more difficult; considerable research has been carried out in this area, c.f. [FAM4], [GE1], [GS], [HL], [HLM1], [HLM2], [HMC], [KU1], [KU2], [KU3], [RO1], [WIJ]. In general, the available results require ergodicity conditions that must hold under *every* stationary policy and/or *every* possible initial state. These types of conditions are very stringent and difficult to verify, in general; see [HMC] for a comprehensive survey of such conditions.

We illustrate the restrictiveness of some of these ergodicity conditions, within the context of the problems with resetting actions considered in this paper. For $p \in \Delta$, $u \in U$, and $B \in \mathcal{B}(\Delta)$, recall that $\mathcal{K}(B|p,u)$, as given in (2.3), is the transition kernel for p_{t+1}, given that $p_t = p$ and $u_t = u$. Very importantly, recall that $\mathcal{K}(\cdot\,|\,p,u)$ is supported on the set $\{T(y,p,u)|y \in Y\}$, a key fact used in the developments in this paper. Now define a finite signed measure on $\mathcal{B}(\Delta)$ as

$$Q(\cdot\,|\,p,\overline{p},u) := \mathcal{K}(\cdot\,|\,p,u) - \mathcal{K}(\cdot\,|\,\overline{p},u)$$

Then, it is easily seen that, in general,

$$\|Q(\cdot\,|\,p,\overline{p},u)\| = 2$$

for all $p \neq \overline{p}$, where $\|\cdot\|$ denotes the total variation of $Q(\cdot\,|p,\overline{p},u)$ [HLM1], [RY]. For example, for the two-state replacement problems considered in previous sections, $u = 0, 1$ are actions for which the above holds.

In [RO1], the existence of a bounded solution to the ACOE is shown, under an equicontinuity assumption imposed upon the family $\{h_\beta(\cdot)\}_{\beta\in(0,1)}$, in addition to a uniform boundedness condition with respect to β. However, for the intended purposes, it suffices that $\{h_{\beta_n}(\cdot)\}$ be uniformly bounded and equicontinuous, for some sequence $\beta_n \uparrow 1$. It is also important to note that it suffices to show equicontinuity of $\{h_{\beta_n}(\cdot)\}$ with respect to *any* metric $D(\cdot\,,\,\cdot)$ on Δ, such that (Δ, D) is separable, allowing therefore the use of the Arzela-Ascoli Theorem [RY, p. 177-179]. However, equicontinuity with respect to the usual metric $d(\cdot\,,\,\cdot)$ of Lemma 2.1 is difficult to verify, even in simple situations with considerable structure as in [RO2], [W1]. Now, if $|h_{\beta_n}(\cdot)| < M < \infty$, for all $p \in \Delta$ and some $\beta_n \uparrow 1$, it is easy to see that

$$|h_{\beta_n}(p) - h_{\beta_n}(\overline{p})| \leq \max_{u\in U}\Big\{|\overline{c}(p,u) - \overline{c}(\overline{p},u)| + M\|Q(\cdot\,|p,\overline{p},u)\|\Big\}$$

and thus if

$$(GS) \qquad \|Q(\cdot\,|p,\overline{p},u)\| \underset{\overline{p}\to p}{\longrightarrow} 0, \quad \text{for each } u \in U$$

then $\{h_{\beta_n}(\cdot)\}$ is equicontinuous. This, combined with arguments similar to those in [RO1], was used by Gubenko and Statland [GS, Theorem 10] to show the existence of a bounded solution to the ACOE. However, from our comments above, condition (GS) does not hold for many problems for which the conditions in Theorem 4.2 and Corollary 4.1 are satisfied. For the same reasons, related approaches based on span-contractive operators [HLM1, p.56-61] cannot be used for such problems. In [RO1] conditions proposed by Taylor [TA] to show equicontinuity of $\{h_\beta(\cdot)\}$ for a replacement process are used. However, as pointed out in [GS, p.59], these imply that (GS) above is satisfied.

Given a uniform boundedness condition on $\{h_\beta(\cdot)\}$, Georgin proves the existence of a bounded solution to the ACOE [GE1, Proposition 3], by first giving conditions under which $\{h_\beta(\cdot)\}$ is an equicontinuous family. However, for these results to be applicable, it has to be possible to represent the transition kernel as

$$(GE) \qquad \mathcal{K}(B|p,u) = \int_B \xi(\overline{p}|u,p)\mu(d\overline{p})$$

where $\xi(\cdot\,|u,p)$ is a density and $\mu(\cdot)$ is a probability measure on $\mathcal{B}(\Delta)$. However, since $\mathcal{K}(\cdot\,|\,p,u)$ is supported on the set $\{T(y,p,u)|y \in Y\}$, then a representation as (GE) is not possible if $\{T(y,p,u)|y \in Y\} \neq \{T(y,p,\overline{u})|y \in Y\}$ for some $u \neq \overline{u}$. Hence, (GE) is not satisfied for a large number of problems for which the conditions in Theorem 4.2 and Corollary 4.1 are satisfied.

For the case when the core state space X is finite and condition (UB) holds, the finite dimensionality of Δ and the concavity of $h_\beta(\cdot)$ are used in [PL], [OMK] to find a sequence $\beta_n \uparrow 1$, such that a bounded solution (Γ^*, h) to the ACOE is obtained via the vanishing discount approach, by taking limits as $\beta_n \uparrow 1$. These results rely critically on X being finite and the concavity of $h_\beta(\cdot)$, neither of which is needed for the results of Theorem 4.2 to hold. Furthermore, as shown in [FG], $\{h_\beta(\cdot)\}$ is an equicontinuous family, with respect to a metric on Δ used in [PL] for other purposes. The topology on Δ induced by this metric can be shown to be separable and thus, contrary to what is claimed in [PL], this situation falls within the formulation in [RO1].

In this paper, we have followed an approach which more advantageously uses the particular structure of the countably supported transition kernels associated with POMDP. From the above and the results obtained in the paper, we see that it is indeed beneficial to use the structure of the particular uncountable state problem at hand. However, these are problems with very special characteristics, and there are many important open questions in the general situation [HMC].

Acknowledgments

We wish to thank the referees for many helpful comments, and specially to one referee who suggested the use of an equivalence relation in the proof to Theorem 4.2. Also, we are grateful to Prof. O. Hernández-Lerma for providing us with several preprints of his work, and for many personal communications on the subject.

References

[AKU] V.A. Andriyanov, I.A. Kogan and G.A. Umnov, "Optimal Control of a Partially Observable Discrete Markov Process," *Aut. Remot. C.,* **4**, 1980, 555-561.

[AL] S.C. Albright, "Structural Results for Partially Observable Markov Decision Processes," *Opns. Res.,* **27**, 1979, 1041-1053.

[AM] A. Arapostathis and S.I. Marcus, "Analysis of an Identification Algorithm Arising in the Adaptive Estimation of Markov Chains," *Math. Control, Signals, and Systems,* **3**, 1990, 1-29.

[AS1] K.J. Åström, "Optimal Control of Markov Processes with Incomplete State Information," *J. Math. Anal. Appl.,* **10**, 1965, 174-205.

[AS2] K.J. Åström, "Optimal Control of Markov Processes with Incomplete State Information, II. The Convexity of the Loss Function," *J. Math. Anal. Appl.,* **26**, 1969, 403-406.

[BA] R.G. Bartle, *The Elements of Real Analysis*, 2nd ed., John Wiley, New York, 1976.

[BE] D.P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Models,* Prentice-Hall, Englewood Cliffs, 1987.

[BEL1] R. Bellman, "A Markovian Decision Problem," *J. Math. Mech.,* **6**, 1957, 679-684.

[BEL2] R. Bellman, *Adaptive Control Processes: A Guided Tour*, Princeton University Press, Princeton, 1961.

[BOR] V.S. Borkar, "Control of Markov Chains with Long-Run Average Cost Criterion: the Dynamic Programming Equations," *SIAM J. Control Optim.* **27**, 1989, 642-657.

[BS] D.P. Bertsekas and S.E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.

[CC1] R. Cavazos-Cadena, "Necessary and Sufficient Conditions for a Bounded Solution to the Optimality Equation in Average Reward Markov Decision Chains," *Syst. Control Lett.,* **10**, 1988, 71-78.

[CC2] R. Cavazos-Cadena, "Necessary Conditions for the Optimality Equation in Average-Reward Markov Decision processes," *Appl. Math. Opt.,* **19**, 1989, 97-112.

[DB] J.L. Doob, *Stochastic Processes,* John Wiley, New York, 1953.

[DJW1] D.J. White, "Real Applications of Markov Decision Processes," *Interfaces,* **15**, 1985, 73-83.

[DJW2] D.J. White, "Further Real Applications of Markov Decision Processes," *Interfaces,* **18**, 1988, 55-61.

[DJW3] D.J. White, "A Selective Survey of Hypothetical Applications of Markov Decision Processes," Technical Report, Dept. of Systems Engineering, University of Virginia, Charlottesville, Virginia.

[DY] E.B. Dynkin and A.A. Yushkevich, *Controlled Markov Processes*, Springer-Verlag, New York, 1979.

[FAM1] E. Fernández-Gaucherand, A. Arapostathis, and S.I. Marcus, "On the Adaptive Control of a Partially Observable Markov Decision Process," *Proc. 27th IEEE Conf. on Decision and Control,* Austin, Texas, 1988, 1204-1210.

[FAM2] E. Fernández-Gaucherand, A. Arapostathis and S.I. Marcus, "On the Adaptive Control of a Partially Observable Binary Markov Decision Process," in *Advances in Computing and Control,* W.A. Porter, et al., eds., Lecture Notes in Control and Information Sciences, Vol. 130, Springer-Verlag, New York, 1989, 217-228.

[FAM3] E. Fernández-Gaucherand, A. Arapostathis, and S.I. Marcus, "On Partially Observable Markov Decision Processes with an Average Cost Criterion," *Proc. 28th IEEE Conf. on Decision and Control,* Tampa, Florida, 1989, 1267-1272.

[FAM4] E. Fernández-Gaucherand, A. Arapostathis, and S.I. Marcus, "Remarks on the Existence of Solutions to the Average Cost Optimality Equation in Markov Decision Processes," preprint, 1990 (submitted for publication).

[FG] E. Fernández-Gaucherand, "Estimation and Control of Partially Observable Markov Decision Processes," Ph.D. dissertation, Electrical and Computer Engineering Dept., The University of Texas at Austin, 1990.

[FI] C.H. Fine, "A Quality Control Model with Learning Effects," *Opns. Res.,* **36**, 1988, 437-444.

[GE1] J.-P. Georgin, "Contrôle des Chaines de Markov Sur des Espaces Arbitraires," *Ann. Inst. H. Poincaré,* **14**, Sect. B, 1978, 255-277.

[GE2] J.-P. Georgin, "Estimation et Contrôle des Chaines de Markov Sur des Espaces Arbitraires," *Lecture Notes Math.,* **636**, Springer-Verlag, Berlin, 1978, 71-113.

[GM] M.K. Ghosh and S.I. Marcus, "Ergodic Control of Markov Chains," to appear in *Proc. 29th IEEE Conf. on Decision and Control,* Honolulu, Hawaii, 1990.

[GS] L.G. Gubenko and E.S. Statland, "On Controlled, Discrete-Time Markov Decision Processes," *Theory Probab. Math. Statist.,* **7**, 1975, 47-61.

[HL] O. Hernández-Lerma, and J.B. Lasserre, "Average Cost Optimal Policies for

Markov Control Processes with Borel State Space and Unbounded Costs," LAAS-Report 90067, LAAS-CNRS, Toulouse, France, 1990.

[HLM1] O. Hernández-Lerma, *Adaptive Markov Control Processes*, Springer Verlag, New York, 1989.

[HLM2] O. Hernández-Lerma, "Harris-Recurrent Markov Control Processes," preprint, 1989.

[HM] O. Hernández-Lerma and S.I. Marcus, "Adaptive Control of Markov Processes with Incomplete State Information and Unknown Parameters," *J. Optim. Theory Appl.,* **52**, 1987, 227-241.

[HMC] O. Hernández-Lerma, R. Montes-de-Oca, and R. Cavazos-Cadena, "Recurrence Conditions for Markov Decision Processes with Borel State Space: A Survey," preprint, 1990.

[HS] D.P. Heyman and M.J. Sobel, *Stochastic Models in Operations Research, Vol. II: Stochastic Optimization,* McGraw-Hill, New York, 1984.

[HW] W.J. Hopp and S.C. Wu, "Multiaction Maintenance Under Markovian Deterioration and Incomplete Information," *Naval Res. Logist. Quart.,* **35**, 1988, 447-462.

[KPO] D. Kreps and E. Porteus, "On the Optimality of Structured Policies in Countable Stage Decision Processes, II: Positive and Negative Problems," *SIAM J. Appl. Math.,* **32**, 1977, 457-466.

[KU1] M. Kurano, "Markov Decision Processes with a Borel Measurable Cost Function: The Average Case," *Math. Oper. Res.,* **11**, 1986, 309-320.

[KU2] M. Kurano, "The Existence of a Minimum Pair of State and Policy for Markov Decision Processes Under the Hypothesis of Doeblin," *SIAM J. Control Optim.,* **27**, 1989, 296-307.

[KU3] M. Kurano, "Average Cost Markov Decision Processes under the Hypothesis of Doeblin," Report #9, Dept. of Mathematics, Faculty of Education, Chiba University, Japan, 1990.

[KV] P.R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*, Prentice-Hall, Englewood Cliffs, 1986.

[LO1] W.S. Lovejoy, "On the Convexity of Policy Regions in Partially Observed Systems," *Opns. Res.,* **35**, 1987, 619-621.

[LO2] W.S. Lovejoy, "Some Monotonicity Results for Partially Observed Markov Decision Processes," *Opns. Res.,* **35**, 1987, 736-743.

[MFA] S.I. Marcus, E. Fernández-Gaucherand, A. Arapostathis, "Analysis of an Adaptive Control Scheme for a Partially Observed Markov Decision Process," to appear in *Proc. 24th Annual Conf. on Information Sciences and Systems,* Princeton University, 1990.

[MO] G.E. Monahan, "A Survey of Partially Observable Markov Decision Processes: Theory, Models, and Algorithms," *Management Sci.,* **28**, 1982, 1-16.

[OKM] M. Ohnishi, H. Kawai and H. Mine, "An Optimal Inspection and Replacement Policy under Incomplete State Information," *European J. Operational Res.,* **27**, 1986, 117-128.

[OMK] M. Ohnishi, H. Mine and H. Kawai, "An Optimal Inspection and Replacement Policy under Incomplete State Information: Average Cost Criterion," in *Stochastic Models in Reliability Theory,* S. Osaki and Y. Hatoyama, eds., Lect. Notes Econ. Math. Syst. #235, Springer-Verlag, Berlin, 1984, 187-197.

[PL] L.K. Platzman, "Optimal Infinite-Horizon Undiscounted Control of Finite Probabilistic Systems," *SIAM J. Control Optim.,* **18**, 1980, 362-380.

[PO1] E.L. Porteus, "On the Optimality of Structured Policies in Countable Stage Decision Processes," *Management Sci.* **22**, 1975, 148-157.

[PO2] E.L. Porteus, "Conditions for Characterizing the Structure of Optimal Strategies in Infinite-Horizon Dynamic Programs," *J. Optim. Theory Appl.* **36**, 1982, 419-432.

[RO1] S.M. Ross, "Arbitrary State Markovian Decision Processes," *Ann. Math. Stat.,* **39**, 1968, 2118-2122.

[RO2] S.M. Ross, "Quality Control Under Markovian Deterioriation," *Management Sci.,* **17**, 1971, 587-596.

[RO3] S.M. Ross, *Introduction to Stochastic Dynamic Programming*, Academic Press, New York, 1983.

[RY] H.L. Royden, *Real Analysis*, 2nd. ed., Macmillan, New York, 1968.

[SEN] L.I. Sennott, "Average Cost Optimal Stationary Policies in Infinite State Markov Decision Processes with Unbounded Costs," *Opns. Res.,* **37**, 1989, 626-633.

[SO1] E.J. Sondik, "The Optimal Control of Partially Observable Markov Processes," Ph.D. dissertation, Electrical Engineering Dept., Stanford University, 1971.

[SO2] E.J. Sondik, "The Optimal Control of Partially Observable Markov Processes over the Infinite Horizon: Discounted Costs," *Opns. Res.,* **26**, 1978, 282-304.

[SS] R.D. Smallwood and E.J. Sondik, "The Optimal Control of Partially Observable Markov Process Over a Finite Horizon," *Opns. Res.,* **21**, 1973, 1071-1088.

[STH] S. Stidham, “Scheduling, Routing, and Flow Control in Stochastic Networks,” in *Stochastic Differential Systems, Stochastic Control Theory and Applications*, W. Fleming and P.L. Lions, eds., The IMA Volumes in Mathematics and Its Applications, **10**, Springer-Verlag, Berlin, 1988, 529-561.

[SY] Y. Sawaragi and T. Yoshikawa, “Discrete-Time Markovian Decision Processes with Incomplete State Observations,” *Ann. Math. Stat.,* **41**, 1970, 78-86.

[TA] H.M. Taylor, “Markovian Sequential Replacement Processes,” *Ann. Math. Statist.,* **38**, 1965, 1677-1694.

[TH] L.C. Thomas, “Connectedness Conditions for Denumerable State Markov Decision Processes,” in *Recent Developments in Markov Decision Processes*, R. Hartley, L.C. Thomas, D.J. White, eds., Academic Press, London, 1980.

[VHE] K.M. VanHee, *Bayesian Control of Markov Chains,* Math. Centre Tracts, **95**, Mathematisch Centrum, Amsterdam, 1978.

[W1] C.C. White, “A Markov Quality Control Process Subject to Partial Observation,” *Management Sci.,* **23**, 1977, 843-852.

[W2] C.C. White, “Optimal Inspection and Repair of a Production Process Subject to Deterioration,” *J. Operational Res. Soc.,* **29**, 1978, 235-243.

[W3] C.C. White, “Bounds on Optimal Cost for a Replacement Problem with Partial Observation,” *Naval Res. Logist. Quart.,* **26**, 1979, 415-422.

[W4] C.C. White, “Optimal Control-Limit Strategies for a Partially Observed Replacement Problem,” *Int. J. Systems Sci.,* **10**, 1979, 321-331.

[W5] C.C. White, “Monotone Control Laws for Noisy, Countable-State Markov Chains,” *European J. Operational Res.,* **5**, 1980, 124-132.

[WA1] R. Wang, “Computing Optimal Quality Control Policies – Two Actions,” *J. Appl. Prob.,* **13**, 1976, 826-832.

[WA2] R. Wang, “Optimal Replacement Policy with Unobservables States,” *J. Appl. Prob.,* **14**, 1977, 340-348.

[WIJ] J. Wijngaard, “Stationary Markovian Decision Problems and Perturbation Theory of Quasi-Compact Linear Operators,” *Math. Oper. Res.,* **2**, 1977, 91-102.

[WW] C.C. White and D.J. White, “Markov Decision Processes,” *European J. Operational Res.,* **39**, 1989, 1-16.