EXTENSIONS AND CHARACTERIZATION OF OPTIMAL MAINTENANCE POLICIES FOR MULTISTATE PARTIALLY OBSERVED MARKOVIAN SYSTEMS

BY

MOHAMMAD MANSOUR FADEL ALDURGAM

A Dissertation Presented to the

DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

In

SYSTEMS ENGINEERING

June 2009

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This dissertation, written by MOHAMMAD MANSOUR FADEL ALDURGAM under the direction of his dissertation advisor and approved by his dissertation committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY IN Industrial & Systems Engineering.

Dissertation Committee

22/7/2009
Prof. Salih O. Duffuaa
Dissertation Advisor

24.06.09
Prof. Shokri Z. Selim
Member

Jun 24, 2009
Prof. Moustafa El-Shafei
Member

24/6/2009
Prof. Mohammed Ben-Daya
Member

24/6/09
Dr. Umar M. Al-Turki
Member

24, 09
Department Chairman

Dean of Graduate Studies

27/7/09
Date

Dedication

To my father and my mother,

I dedicate this work

Acknowledgment

All gratitude, praise and glory to Allah Almighty for giving me patience, capability and opportunity to finish this work. Without his help and will, nothing is accomplished. Blessing and peace be upon our leader Muhammad, his family, his companions and those that follow his guidance until the last day.

My deepest appreciation goes to my thesis advisor, Prof. Salih O. Duffuaa, for his invaluable guidance, encouragement and endless support. The completion of this work is credited to his tireless support and priceless ideas. He gave me the knowledge and research experience that will never be forgotten.

My great thanks to my thesis committee members for their suggestions and valuable comments. Special thanks for Prof. Moustafa El-Shafei for sparing much time for me.

I would like to thank King Fahd university of Petroleum & Minerals for the support and the opportunity provided to me to pursue my Ph.D. Special thanks are due to the Systems Engineering Department and all the faculty members for their encouragement and their direct and indirect help.

Special thanks goes for the following people for the valuable comments and inspiring-ideas which help me a lot in achieving my research objectives: Dr. Julie Simmons Ivy (North Carolina State University) for giving me an office hour to discuss my research work and for her valuable suggestions, my friend, Dr. Mohammad Jarrar (KFUPM) for his suggestions and insightful ideas and to my brother Mohannad AlDurgam (University of Toronto) for his on call availability and assistance to me.

No words can express my gratitude to my wonderful family, My Father, Mother, brothers and sister. Their encouragements and personal sacrifices are truly appreciated and will be remembered forever.

TABLE OF CONTENTS

List of Tables

List of Figures

ملخص بحث

درجة الدكتوراه في الفلسفة

الأسم	:	محمد بن منصور فاضل الضرغام
عنوان الأطروحة	:	تمثيل سياسات الصيانة المثلى للأنظمة الماركوفية متعددة الحالات جزئية المشاهدة
التخصص الدقيق	:	هندسة نظم
التاريخ	:	يونيو 2009

يهدف هذا البحث إلى نمذجة و تمثيل سياسات الصيانة المثلى للأنظمة معقدة التركيب و التي تتهالك مع مرور الزمن. توجد مثل هذه الأنظمة في المؤسسات الصناعية و غيرها و قد استحوذت على اهتمام الباحثين في هذا المجال لعدة عقود. إن الأسلوب المتبع للنمذجة الرياضية لمثل هذه الأنظمة هو تمثيلها كأنظمة متعددة الحالات. باستخدام النماذج الماركوفية كاملة وجزئية المشاهدة.

ان الهدف الرئيس لهذا البحث هو نمذجة و تمثيل قرارات الصيانة المثلى للأنظمة متعددة الحالات و المشاهدة جزئيا على عناصر المجال (مصفوفات أحادية البعد تمثل الحالة المتوقعة للنظام) و المرتبة وفق الهيمنة العشوائية أحادية الرتبة. تم استحداث بعض الشروط الجديدة و التي تضمن وجود سياسات صيانة مثلى و منتظمة. إن هذه الشروط تتميز بضمان استمرارية و بقاء عناصر المجال مرتبة وفق الهيمنة العشوائية أحادية الرتبة على مدى الإطار الزمني للحل. كأسلوب جديد لدراسة هذه المسألة، تمت دراسة العلاقة بين الترتيب الجزئي أحادي الرتبة و بعض الترتيبات الجزئية الأخرى و التي لم تؤخذ بعين الاعتبار من قبل لحل هذه المسألة، مثل ترتيب الخطورة العكسية و هيمنة العناصر الكلية. و بهدف الإيضاح تم عرض نموذج رياضي لنظام ذو حالتين ومن ثم طور للحالة العامة.

تم ربط سياسات الصيانة والإنتاج و الجودة عن طريق نموذج رياضي ماركوفي يهدف إلى تعظيم الكفاءة الكلية للنظام. إن مصطلح " الكفاءة الكلية" للأنظمة يمثل مقياس متكامل لمدى جاهزية النظم للإنتاج على مدار الساعة و معدل الجودة. استخدم النموذج لإيجاد سياسات الصيانة والإنتاج المثلى للأنظمة وتم إثبات إن الدالة الهدفية لهذا النموذج دالة مقعرة مكونة من قطع خطية مستقيمة.

تمت أيضا نمذجة وتمثيل اثر الخطأ في القياس على سياسات الصيانة المثلى للنظم الماركوفية متعددة الحالات و المشاهدة جزئيا. وقد اشتُقت معادلة - باستخدام قانون بيز- لتوقع حلة النظام المقترح و الممثل بسلسلة ماركوفية ثلاثية الطبقات و جزئية المشاهدة. وتم إثبات أن الدالة الهدفية لهذا النموذج بوجود أخطاء القياس دالة مقعرة مكونة من قطع خطية مستقيمة و متصلة. و في النهاية دُرست العلاقة بين نوعية المعلومات أو القراءات المقاسة للاستدلال على حالة النظام و أثر الخطأ في القياس على سياسات الصيانة المثلى.

درست الأطروحة حالة وجود رأي الخبراء كأحد معطيات المسألة بحيث يكون رأي خبراء الصيانة هو الأفضل للربط بين نوعية المشاهدات أو القراءات المأخوذة من النظام و الحالة الحقيقية للنظام. لذا تم افتراض وجود عملية اتخاذ قرار لنظام ماركوفي مشاهد جزئيا و ذلك بوجود مشاهدات غير واضحة ثم تم اقتراح اقتران عضوية للمشاهدات غير الواضحة من اجل إعادة تكوين المصفوفة التي تمثل العلاقة بين المشاهدات و الحالة الحقيقية للنظام و قد عرضت بعض الأمثلة لتوضيح هذه الحالة. ختمت الأطروحة بملخص لنتائج البحث و مقترحات للبحوث المستقبلية.

DISSERTATION ABSTRACT

NAME : Mohammad Mansour Fadel AlDurgam

TITLE OF STUDY : Extensions and Characterization of Optimal Maintenance Policies for Multistate Partially Observed Markovian Systems

MAJOR FIELD : Systems Engineering

DATE OF DEGREE : June 2009

This research aims at modeling and characterizing optimal maintenance policies for complex deteriorating systems. Such systems widely exist in manufacturing enterprises where it has been gaining the interest of researchers for many decades. Representing complex systems as multi-state systems has been the trend in the literature (Derman, 1962), (Hopp and Wu, 1990) and (Maillart, 2006). Markov Decision Process (MDP) and Partially Observed Markov Decision Process (POMDP) have been used as the general frame to model and represent such systems.

The main objective of this work is to characterize optimal maintenance policies for multistate systems over the systems state occupancy vectors ordered by the first order stochastic dominance. New set of conditions to guarantee the existence of optimal threshold-type maintenance policies are provided. The main advantage of the developed conditions is ensuring the first order stochastic dominance to survive conditioning. As a new approach, this is achieved by developing new relations with other useful partial orders which were not considered for this problem before, namely, the reverse hazard rate and the component wise dominance partial orders. For the sake of illustration, a two-state model is provided first, and then it is extended to the case of n-states.

In order to link maintenance, operation and quality, a new model within the POMDP framework is formulated. The model uses Overall Systems Effectiveness (OSE) as a criterion. OSE combines systems availability, process rate and quality rate in a composite criterion. This provides a mechanism that ties maintenance and operation through the process rate. In such situation the optimal action will be a maintenance action coupled with a specific system speed that reflects the process rate.

Condition-based maintenance is usually based on measuring or observing systems conditions; however, measurements are not error free. The impact of measurement errors on the POMDP optimal maintenance polices is formulated and studied. A new Bayesian update for a three layers hidden Markov model is provided and proved to be a sufficient statistic. Also, the objective function for the POMDP problem is shown to be piecewise linear convex one. The relation between observations quality and the impact of measurement errors is discussed.

For the case where an expert opinion better relates observed signals to the true underlying state of the system is considered. A POMDP with fuzzy observations is assumed. A fuzzy membership function is provided and utilized to fuzzify the state observations matrix. The application of the fuzzy membership function and the significance of this scenario are illustrated by examples. Finally, the dissertation is concluded by a summary of the contributions and suggestions for future research.

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

This research considers complex multistate systems that deteriorate over time. The process of deterioration reduces systems effectiveness, productivity, and quality. To restore systems effectiveness, maintenance actions are needed. Maintenance actions include inspections, different repairs, and replacement. Such actions are taken to avoid the system operating in undesirable states. Doing too much maintenance means extra cost; on the other hand, not doing maintenance systems will end up in a failure or operating in an undesirable state. The challenge is to model such systems to determine optimal maintenance actions that minimize overall systems cost or maximizes profit or overall systems effectiveness. Such systems exist in manufacturing and service industry.

As an example consider an electric arc furnace that is operated continuously. If the furnace is left operating for long time, its performance will deteriorate. On the other hand, stopping an electric arc furnace to characterize its exact state is very costly, also while operating; an electric arc furnace provides output signals like quality of the output product which is related to the true state of the furnace. Another example is provided by Ivy (2005) where quality of spot welds in cars assembly process reflects the alignment of the electrodes involved in the spot welding process, stopping the process too frequently is not feasible also producing too many imperfect spot welds is a quality-related loss . This is a basic scenario in statistical quality control applications where an out of control process is sometimes related to a failure in the production system.

The rest of this chapter is organized as follows: a general statement of the problem is provided in Section 1.2 and the approaches to the multistate machine maintenance problem are provided in Section 1.3. An MDP based real life decision making system for road pavement management is provided in Section 1.4 and motivation for the work in this dissertation is provided in Section 1.5. The dissertation objectives are provided in Section 1.6 and the contribution made in this dissertation is outlined in Section 1.7. Finally, the dissertation organization is provided in Section 1.8.

1.2 GENERAL STATEMENT OF THE PROBLEM

In this Dissertation, a multistate system or machine is considered over a finite horizon. The states of the system range from 1 to n representing as good as new up to failed state respectively. The system states are assumed to be controllable states, that is, a decision maker can enhance the state of the system by means of a set of available control actions A. The control actions are taken at discrete time epochs over the horizon considered. Control actions can be as simple as do nothing (a_0), minimal repair (a_i), or replacement of the whole system (a_n), which is assumed to renew the system. Actions effects differ based on the action type. If the system is left with no maintenance (a_0) it is assumed that the system will keep deteriorating.

Different repair actions are assumed to improve the system state. For example, if maintenance action a is applied, system movement to an improved state is governed by P^a.

Also, it is assumed that the system true state is not directly observable by the decision maker. Instead, only noise-corrupted information is received. This information is assumed to be probabilistically related to the true or actual system state. This is represented by a, possibly action-dependent, state observation transition matrix R^a with r_{jk}^a elements.

Since system states are partially observable, the decision maker is assumed to make his decisions based on the belief state. This is nothing but a state occupancy vector:

$$\pi = [\pi_1 \ \pi_2 \ \dots \ \pi_n \,]$$

Where, the number of elements in π equals the number of the system states. As will be illustrated later on, this state occupancy vector is updated at each time epoch whenever a decision has to be made. As in the literature, the updated state occupancy vector is usually referred to as the Bayesian update.

In general, the reward criterion of the underlying system is assumed to be state/action dependent structure. That is, g(i, a) reward will be received if action a is taken when the system true underlying state was i. The objective is the optimal policy selection for every possible state of the system at any given time epoch.

The problem of determining and characterizing optimal maintenance policies for such systems has attracted the attention of many researchers over the span of the last five decades. Fixing some maintenance action for a given system state at given time stage over some planning horizon is called a “maintenance policy”. If there is a value (system state number, belief state, age or others) in the decision domain only prior or beyond which certain policy is adopted, this value is called a “threshold”. The policy structure when threshold values exist is called a “threshold policy”. This type of policy is of interest in the maintenance literature due to its ease of application and, sometimes, computational advantage.

In general, this is the problem considered in this dissertation. Different settings of this problem are considered and precisely stated in Chapters 4, 5 and 6.

1.3 APPROACHES TO THE PROBLEM

Representing complex dynamic systems as multi-state systems has been widely reported in literature (Derman, 1962), (Hopp and Wu, 1990) and (Maillart, 2006), where the system is represented by means of "state" and "state transition mechanism". Most of the models in literature let the state to represent a different degree of deterioration of the system, while "state transition mechanism" differs based on the assumed stochastic process. Also, cost or profit/revenue components vary in the available multi-state systems maintenance models.

Assuming that a stochastic process governing the state transition mechanism is Markovian has been widely reported in the literature, this is for its mathematical tractability; mainly, due to its forgetfulness property. Forgetfulness property; simply, means that the future value of a random variable depends only on its current value.

A Discrete time Markov chain is a stochastic process whose state changes at discrete points in time. State transition in a Markov chain depends only on the current state of the system on hand. When the state transitions of a Markov chain can be influenced by some control action(s) this dynamic decision making framework is called an MDP. This assumes that the true state of the system is observed or known certainly at each stage. On the other hand, there are cases where the true state of a system is not known or observed directly. Instead, the decision maker receives noise-corrupted information, which is probabilistically related to the true state of the system on hand.

Under these new settings, the MDP is known as POMDP. This is discussed in more details in Chapter 2.

MDP and POMDP have been used as frameworks to model complex multistate multistage systems. Many examples from the literature can be found in Chapter 3. In this dissertation the general POMDP framework is adopted as the general modeling approach.

1.4 MDP-BASED PAVEMENT MANAGEMENT SYSTEM

In this example, adopted from Puterman (2005), a real life example of MDP model is provided to reflect the applicability of this decision making framework.

the Arizona Department of Transportation (ADOT) (the decision maker) had to manage a 7,400 miles road network (system). To define the "states" of this system, the 7,400 roads network was divided into 7,400 one mile segment. Each segment can be characterized by roughness (three levels), percentage of cracking (three levels), change in cracking since last year (three levels) and an index describing the last time for maintenance operation and the level of maintenance (5 levels). Hence, each segment can have ($3\times3\times3\times5=135$) levels with 120 out of 135 feasible combinations. The whole road system was divided into nine sub networks, with a dynamic model developed for each category.

"Transition probabilities" and "costs" were found using statistical analysis and the opinion of expertise. Each state had limited subsequent possible states and thus 97% of the transition probabilities where equivalent to zero. Action dependent costs and state/action dependent costs for the routine maintenance actions were assumed in the model.

Table 1-1 Different action costs for road pavement problem (Puterman, 1994)

Action Index	Action Description	Constructing Cost $/yd^2
1	Routine Maintenance (RM)	
2	Seal Coat	0.55
3	ACFC	0.75
4	ACFC + AR	2.05
5	ACFC + HS	1.75
6	1.5 inch Ac	1.575
7	1.5 inch Ac + AR	2.875
8	1.5 inch Ac + HS	2.575
9	2.5 inch AC	2.625
10	2.5 inch Ac + AR	3.925
11	2.5 inch Ac + HS	3.625
12	3.5 inch AC	3.675
13	3.5 inch AC + AR	4.975
14	3.5 inch AC + HS	4.675
15	4.5 inch AC	4.725
16	5.5 inch AC	5.775
17	Recycling (6 inch AC)	6.3

ACFC: Asphalt Concrete Fine Coat, AR: Asphalt Rubber, HS: Heater Scarifier, AC: Asphalt Concrete

From the table above we notice the existence of action-dependent costs, except for the Routine Maintenance (RM) which was reported to have 13 categories and each has a system-state (in terms of roughness and percentage of cracking)/action dependent costs.

The objective was to minimize the average cost per unit time with some road quality constraints. For example, 80% of high traffic roads must have roughness level not exceeding 165 inches/mile.

As a result of applying this systematic maintenance actions allocation procedure between (1980-1984), 115 million dollars savings over 5 years were achieved. The same model was modified and applied in Kansas, Finland and Saudi Arabia.

1.5 MOTIVATION

The following points provide the motivation behind pursuing research in this problem:

1. To the best of our knowledge, there are very few researchers who considered characterizing the general multistate POMDP maintenance policies, namely, White (1979, a), Lovejoy (1987) and Ivy and Pollack (2005). The researchers have provided conditions and results that characterize the existence of optimal threshold-type policies. The review of the literature indicated that there is a room to provide new conditions that ensure the existence of threshold-type policies under different partial orders.
2. Most of the multi-state models in the literature either assume cost minimization or profit maximization as an optimality criteria. Thus, there is a need to consider different or more comprehensive criteria. The suggested criterion is to maximize Overall Systems Effectiveness (OSE). This criterion ties together, availability, process rate and quality rate. It provides a room to tie maintenance and operation through speed of production.
3. The information gathered to infer about the state of the system is not error free. Modeling the effect and impact of the measurement errors within the POMDP has not been addressed in the literature. A model is suggested to assess the impact of

the error on the optimal maintenance polices. Also the case of fuzzy observations is needed to be addressed.

In order to address the issues raised in 1-3 above, the objectives of this dissertation have been formulated and stated in the next section.

1.6 DISSERTATION OBJECTIVES

The objectives of this dissertation are summarized by the following points:

- To develop conditions to ensure the existence of threshold-type policies under the First Order Stochastic Dominance (FSD) partial order.
- To formulate the POMDP existing models for determining optimal maintenance policies in order to maximize Overall System Effectiveness (OSE).
- To study the effect of measurement errors on the maintenance models formulated as POMDP.
- To propose or implement a computational procedure to obtain the optimal solution for the developed model.
- To explore possible fuzzification of the maintenance models developed in POMDP framework (fuzzy measurements).

The contributions of this dissertation are provided in the next section.

1.7 DISSERTATION CONTRIBUTION

The main contribution of this dissertation are a new set of new conditions on the parameters of the POMDP model to ensure the existence of threshold-type maintenance polices over the belief space ordered by the first order stochastic dominance. This is achieved by utilizing newly established and existing relations between the first order stochastic dominance partial order and other partial orders (the reverse hazard rate and the component-wise dominance). The conditions and this new approach make the first order stochastic dominance survives conditioning. This has the advantage of enlarging the set of the belief space elements over which the optimal solutions of the POMDP problem can be characterized. This dissertation also provides a POMDP model to maximize Overall Systems Effectiveness (OSE) as a criterion (Availability $\times$ Process Rate $\times$ Quality Rate) (Nakajima, 1988). This criterion provides a link between maintenance and operation through the process rate of the system or the equipment. The output of such a model is a maintenance action with a speed level or, in other words, a maintenance action with a specific load on the system. The concept of measurement errors is introduced within the POMDP framework. A model is provided to assess the impact of measurement errors. The POMDP framework is modified to adjust the effect of possible measurement errors, by using the concept of three layers hidden Markov model. The effect of the information quality (system signals) on measurement errors is also illustrated as well. Finally, the case of fuzzy observations within the POMDP framework is considered. Here, it is assumed that a system state is better judged by an expert opinion than POMDP systems observations or signals. A fuzzy membership function is provided to utilize such an opinion in updating system belief state. Some examples are provided to illustrate the concept.

1.8 DISSERTATION ORGANIZATION

This rest of this dissertation is organized as follows: some theoretical background for this dissertation is provided in Chapter 2 followed by the literature review in Chapter 3. A two-state POMDP model is presented in Chapter 4 and extended to an n-state POMDP model in Chapter 5. Chapter 6 describes a model that maximizes Overall Systems Effectiveness (OSE) and the process of modeling measurement errors and observations fuzzification are addressed in Chapter 7. The dissertation conclusion and directions for future research are the subject of Chapter 8.

CHAPTER 2

THEORITICAL BACKGROUND

2.1 INTRODUCTION

The research in this dissertation requires knowledge in several areas for the representation and modeling of multistate deteriorating systems. This knowledge include mainly: POMDP, Partial Orders and some knowledge of Fuzzy Logic. The background needed for the dissertation is presented in this chapter.

The rest of this chapter is organized as follows: the definition of partial orders, in addition to some examples, are provided in Section 2.2 followed by a formal definition of both the MDP and POMDP and their elements in Section 2.3. A typical mathematical model for the POMDP decision making framework is provided in Section 2.4 followed by a description and illustration of the value iteration algorithm in Section 2.5. a brief background on fuzzy logic is introduced in Section 2.6.

2.2 PARTIAL ORDERS DEFINITIONS AND EXAMPLES

An order ($\leq$) is a partial order for a set, if and only if for any three members of the set $\{\pi, \pi' \; and \; \pi''\}$ the following three properties hold:

1. Reflexivity: this means $\pi \leq \pi$ implies $\pi \leq \pi$
2. Anti-symmetry: if $\pi \leq \pi'$ and $\pi' \leq \pi$ then $\pi = \pi'$
3. Transitivity: if $\pi \leq \pi'$ and $\pi' \leq \pi''$ then $\pi \leq \pi''$

To illustrate consider the following examples:

Example 2.1: (First Order Stochastic Dominance (FSD))

This is a basic partial order, and it is widely applied in decision making theory and its applications. A probability vector $\pi \geq_{SD} \pi'$ (π stochastically dominates π' in FSD sense) if and only if:

$$\sum_i^n \pi_i \leq \sum_i^n \pi_i' \qquad \forall\, i\ \in S$$

It can be rewritten as follows:

$$\sum_0^i \pi_i \geq \sum_0^i \pi_i' \qquad \forall\, i\ \in S$$

This is to be discussed in more detail, in addition to its different special forms, in Chapter 5 (Section 5.3).

Example 2.2

Vectors can not be ordered based on their lengths because the anti-symmetry relation is violated. For illustration, consider $\pi = [1 \quad 0 \quad 0]$ and $\pi' = [0 \quad 0 \quad 1]$. $\pi \leq \pi'$ and $\pi' \leq \pi$, but $\pi \neq \pi'$

Example 2.3

The strict less than or greater than relation can not be a base for partial orders, because the reflexivity relation will be violated. Other examples of partial orders are provided in Chapter 5 (Section 5.3).

2.3 MDP AND POMDP

Formally, MDP can be defined as a dynamic decision making framework that aims at optimally controlling a Markov stochastic process over a given number of future stages, such that, a set of available control actions influence the state transition of the Markov chain at each stage. Where, an MDP can be formulated as a stochastic dynamic program.

The elements of an MDP are:

1. State Set (S): this is a set containing all the possible states of a multi-state system $\{1, 2, \dots, n\}$. 1 represents an as good as new system and n represents a failed system. The other states represent the different degrees of system deterioration.
2. Action set (A): a controller has the option to select an action from the action set in order to influence the state of the system. An action can range from do nothing to replace the whole system.
3. Policy $\delta(s)$: when action is assigned to each possible state of the system, this is called policy.
4. Optimal Policy $\delta^*(s)$: when action is assigned to each possible state of the system such that some criteria is minimized or maximized, this is called an optimal policy.
5. State Transition Matrices: for each action there is a probability transition matrix governing the state transition of a system (P^a).
6. Reward criteria: this can depend on the current or next state or both, control action, or all. This depends on the objective of the model.

An MDP is assumed to work as follows:

- Control action is taken
- Gain or loss takes place (usually, function of the action taken and the current state of the system)

- System is moved to a new state (following a Markov Chain)
- Next stage is started

A PODMP is a generalization of the MDP where the true/actual state of the system is not known exactly to the controller, instead, a noise-corrupted output is received from the system. This error-prone output signal is assumed to be probabilistically related to the true/actual state of the system. Hence, it is “partially observed”.

The elements of a POMDP are similar to that in an MDP, plus the following two elements:

1. State occupancy vector or belief state (π): this is a probability vector, where, the elements of this vector provide different states occupancy probabilities. For instance, π_i: is the probability that the system is currently in state $i \in S$. Also, the collection of all the state occupancy vectors is called belief space $\Pi(S)$.
2. Observations set (O): for each state of the system, there are a number of output signals can be observed. The set that contains all possible outcome signals is referred to as the observations set $\{1,2, \dots m\}$.
3. States-observations transition matrix (R): given that a system is currently occupying state i, there is probability corresponding to each possible output signal form the observation set.

Mainly, a POMDP decision making framework consists of the following steps:

- Control action is taken
- Gain or loss takes place (due to any or all of the state transition, control action, and observation of the system)

- Some signal is observed from the system
- State occupancy vector is updated using Byes rule.
- Next stage is started

A POMDP is actually nothing but a core process, triggered by action-based state transition matrices P^a and an observation process triggered by the R matrix. Each observation is probabilistically related to the states of the core process. This is what Figure 2.1 shows.

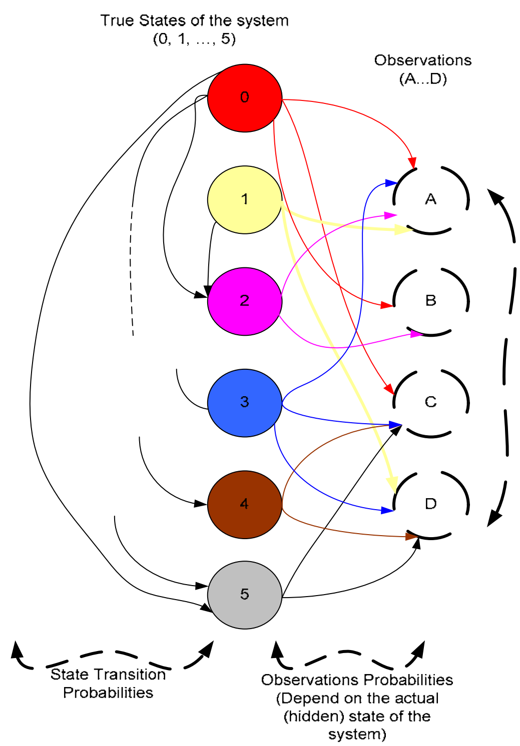


Figure 2.1 Core process and observation process in POMDP. Sample arrows shown, represent probabilities of transitions

There are many applications of POMDP reported in the literature. Some of them follow:

- Finance
- Marketing
- Health care
- Operations management
- Robotics
- Others

In the POMDP context, similar to any other decision making process, there is cost and profit tradeoff. In its simplest form, for a maintenance application, performing to excessive maintenance will be costly. On the other hand, not doing maintenance in a frequent-enough rate will lead to many unexpected failures and this leads to losses due to lost production opportunity and the higher maintenance costs corresponding to sudden stops. This is illustrated by Figure 2.2 next.

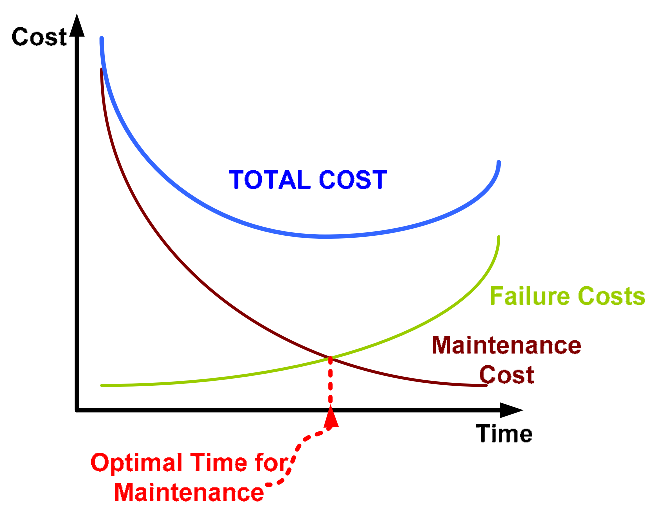


Figure 2.2 Maintenance and failure costs tradeoff

2.4 MATHEMATICAL MODEL OF THE POMDP

In this section a typical mathematical model for the POMDP decision making framework, discussed in Section 2.3, is provided. A POMDP is a dynamic decision making framework for hidden Markov chains (Markov chain with error-corrupted state information). Follows is the nomenclature used to describe the POMDP formally:

S	System's state set {1,2,…n}
i, j	Elements of S
O	Observations set with components {1,2, … m}
k	Element of O
A	Totally ordered actions set available to the decision maker
a_i	Element of A, that can range from a_0 (do nothing) to a_n (replace)
$\mathrm{g}(.\,, a)$	Reward the system generates if it is in state s and action a was taken
π	State occupancy vector, or belief state
$V_t^*(\pi, a)$	The optimal value function at time t given belief state vector π and maintenance action a
P^a	System's state transition matrix corresponding to action a $(n \times n)$
P_{ij}^a	Probability that the system will move from state i to state j if action a was taken
R	An $(n \times m)$ state observation transition matrix subject to action a
r_{jk}	And entry in the R^a matrix which gives the probability that observation k will be observed if the system has moved to state j and action a has been taken.
$T(\pi, a, k)$	Posterior state occupancy probability vector
$\sigma(\pi, a)$	The probability vector of the system observations $\{m\}$ when action a is taken and the system is believed to be in π
$\sigma(k; \pi, a)$	The k^{th}component of $\sigma(\pi, a)$
β	Discount factor

The expected total discounted reward over t time horizons can be expressed using the following Bellman recursive equation:

$$V_t(\pi, a) = \sum_{i \in S} \pi_i \mathrm{g}(i, a) + \beta \times \sum_{k \in o} \sigma(k; \pi, a) V_{t-1}^*\big(T(\pi, a, k)\big) \tag{2.1}$$

This recursive equation (2.1) is nothing but the expected instantaneous reward as a function of the current belief state and the action taken plus the expected reward over the remaining $t-1$ time horizons, with respect to $\pi, a,$ and k. With T is the updated belief state vector and $\sigma(k;\pi,a)$ is the probability of observing observation k given π, and a. Starting with a belief state π, a control action is to be taken, suppose it a, then applying Byes rule, it can be shown that (Smallwood and Sondik, 1973):

$$T_j(\pi;\ a,k) = \frac{\sum_i^n \pi_i p_{ij}^a r_{jk}}{\sigma(k;\pi,a)}$$

Where, π is the current belief state = $[\pi_1, \pi_2, \dots, \pi_n]^T$ and $\sigma(k;\pi,a) = \sum_i^n \sum_j^n \pi_i p_{ij}^a r_{jk}$ is the probability of observing k (total probability theorem).

As a result of some action (a) and observed output signal (k) from the partially observable Markov process the updated state occupancy vector becomes:

$$T(\pi;\ a,k) = [T_1(\pi,a,k), T_2(\pi,a,k), \dots, T_n(\pi,a,k)]^T$$

For some optimal action (a^*) the optimal value function becomes:

$$V_t^*(\pi) = max_a\{V_t(\pi,a)\} = \sum_{i\in S} \pi_i \mathrm{g}(i,a^*) + \beta \sum_{k\in o} \sigma(k;\pi,a^*) V_{t-1}^*\big(T(\pi,a^*,k)\big) \quad (2.2)$$

Finally, the main objective of a finite stage POMDP is the optimal action selection at any time stage, for every belief state in the belief space at that stage.

2.5 THE VALUE ITERATION ALGORITHM

For an MDP, the value iteration algorithm is an algorithm used to solve MDP and POMDP models for a finite horizon. It is just a straight forward application of Bellman optimality principle, namely, the optimal solution for t horizons is the optimal for 1 horizon plus the expected optimal for the remaining $t-1$ horizons. This is done recursively backwards as follows.

Given a reward criterion (state/action dependent for example), a set of possible control actions and their corresponding transition matrices, conduct the following backward recursive calculations:

1. Calculate the expected reward for $t = 1$ (only one horizon left) using all the possible actions for each state i.
2. Then for each state select the action a_i^* that gives the maximum reward that is consisting of $V_1^*(i)$ and $V_{t-1}^*(j)$ conditioned on the system movement from state i to state j.

This is nothing but Bellman's optimality principle. Consider the following illustrative example for a detailed illustration:

Example 2.4: Value iteration algorithm for 3-state, 3-stages, and 2-actions system.

Consider a three-state system ($S = \{0,1,2\}$), with two available control actions ($A = \{a_0, a_1\}$). Each of the two actions is assumed to affect the system state stochastically by means of a homogeneous state transition matrix as follows:

$$P^{a_0} = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0 & 0.7 & 0.3 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P^{a_1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

The reward criteria are assumed to be state/action dependent ones as follows:

For a_0:

$g(1, a_0) = 8, g(2, a_0) = 5$ and $g(3, a_0) = 1$

For a_1:

$g(1, a_1) = 3, g(2, a_1) = 4$ and $g(3, a_1) = 5$

Notice here that, for this totally observed system (MDP), equations 2.1 and 2.2 reduces to the two following equations respectively:

$$\mathrm{V_t}(i, a) = \mathrm{g}(i, a) + \sum_{j=1}^{n} p_{ij} V_{t-1}^*(j) \text{ where } j \in S$$

With $V_t^*(i) = \max_a\{V_t(i,a)\}$

Tables 2.1, 2.2 and 2.3 provide three iterations of value iteration algorithm for t =1, 2 and 3 respectively.

Table 2-1 The value iteration algorithm t=1

i	a_0	a_1	$V_1^*(i)$	a^*
1	8	3	8	a_0
2	5	4	5	a_0
3	**1**	**5**	**5**	a_1

Table 2-2 The value iteration algorithm t=2

i	a_0	a_1	$V_1^*(i)$	a^*
1	8+ 8*0.9 +5*0.05+5*0.05 = 15.7	3+ 8*1 +5*0+5*0 = 11	15.7	a_0
2	5+ 8*0 +5*0.7+5*0.3 = 10	4+ 8*1 +5*0+5*0 = 12	12	a_1
3	**1+ 8*0 +5*0+5*1 = 6**	**5+ 8*1 +5*0+5*0 = 13**	**13**	a_1

Table 2-3 The value iteration algorithm t=3

i	a_0	a_1	$V_1^*(i)$	a^*
1	8+ 15.7*0.9 +12*0.05+13*0.05 = 23.38	3+ 15.7*1 +12*0+13*0 = 18.7	23.38	a_0
2	5+ 15.7*0 +12*0.7+13*0.3 = 17.3	4+ 15.7*1 +12*0+13*0 = 19.7	19.70	a_1
3	**1+ 15.7*0 +12*0+13*1 = 14**	**5+ 15.7*1 +12*0+13*0 = 20.7**	**20.70**	a_1

For three horizons ($t = 3$), the results indicate that it is optimal to take the first action (a_0) if the system is in state 1 and the second action (a_1) if the system in states 2 or 3.

In chapters 5 and 6, the same Value Iteration algorithm will be used to provide the solutions numerically for the POMDP models there, over a discretized belief space of the state occupancy vectors. To correct for the fact that the actions don't map back to the discretized belief space, bilinear interpolation is used. A MATLAB code for solving POMDPs for 3-state, n-action and m-observation system is provided in the Appendix.

2.6 SOME BACKGROUND ON FUZZY LOGIC

Fuzzy logic has been introduced to deal with judgments and the world vagueness. Basically, it extends the Boolean logic, where, instead of binary membership values for sets, there are partial membership values, the concept of partial truth and partial falseness are introduced. This is done by means of fuzzy membership function. A membership function is a function that assigns values from 0 to 1 as a membership value for a given set. 0 represents the absolute falseness and 1 the truth. Values in between represents different degrees of membership to a set. Consider the famous height example. Suppose that people who are taller than 6 inches are regarded tall. Then based on Boolean logic a person who is 5.9 inches is not considered tall. In fuzzy logic any height is considered tall to some extent. This is nothing but the concept of membership. Hence, for the set of tall people every element of the set has a membership value $\in [0,1]$.

Following are the basic steps providing the general framework of fuzzy logic:

1. Fuzzification: conversion of real inputs to fuzzy set values by means of suitable membership function(s).
2. Inference mechanism: this has to do with the (if then and else) rules, the operators (and, not, or …) used to execute the logic, and the implication mechanism (this provides the output as a fuzzy membership function). Zadeh, Godel and Mamadani are famous implication techniques.
3. Defuzzification: this is where real membership value from the inference mechanism output is obtained. Some famous ways are center of gravity and maximum value as defuzzification mechanisms.

CHAPTER 3

LITERATURE SURVEY

3.1 INTRODUCTION

This literature review considers systems that deteriorate over time. Such systems exist in manufacturing and service industry. The process of deterioration reduces systems effectiveness, productivity, and quality. To restore systems effectiveness, maintenance actions are required. Maintenance actions include inspections, repairs or replacements. Such actions are taken to avoid the system operating in undesirable states. The challenge is to model such systems to determine optimal maintenance actions that minimize overall systems cost or maximizes overall systems effectiveness. The literature has many models that address these types of systems. A brief review of the relevant literature will be provided below. For detailed surveys on classical maintenance models, refer to McCall (1965), Pierskalla and Vollker (1976), Valdez-Flores and Feldman (1989), and Wang (2002). Surveys of integrated maintenance and production control models are provided by Ben-Daya and Rahim (2001) and Budai, *et. al.* (2006). And for a survey on POMDP models and algorithms, refer to Monahan (1982).

3.2 MUULTISTATE MARKOVIAN AND PARTIALLY OBSERVED MARKOVIAN DECISION MODELS

Derman (1962) derived one of the earliest maintenance models of a multi-state system. In his important model, Derman has shown that for a multi-state system with state transition mechanism governed by a Markov chain, the optimal replacement policy is control limit policy in one of the system states (state-dependent policy), the model is an MDP model because the true state of the system is assumed to be observable. Only maintenance action costs are considered in Derman's model. Kolesar (1966) extended Derman's model, but with a state occupancy cost added to the model. The same result of control limit policy was obtained.

In Hopp and Wu (1990) it has been shown that for a multi-state totally observable system, the optimal maintenance policy has a monotonic control-limit structure. The result was obtained after assuming a monotone reward functions and monotone differences between the different reward function.

Reward criteria in the MDP settings can vary as reported in the literature, Puterman (2005) summarizes the different reward criteria for MDP models as follows:

- Fixed amount of reward received at a fixed or a random point in time received in the current stage prior to the next stage.
- Received continuously during a given time stage
- A random amount depending on a system's state at the next stage of time.
- Or, combination of the above.

These cost structures can also be adopted for POMDP models. Mainly, state-action dependent reward is usually reported in the literature. Smallwood and Sondik (1973) used a reward that depends on system's current state, next state, and the observation obtained from the system.

Kim and Gen (1993) developed optimal replacement policies under the concept of fuzzy cost data for a two-state two-action POMDP model.

In some systems the state is not known with certainty; instead, noise-corrupted information on the system's state is all what is available. This opens the door for POMDP as a model for such cases.

POMDP has very wide areas of application such as: Financial, medical, communications… maintenance is a major area of POMDP applications. Many authors considered models where the system subject to maintenance is Multi-state and the true state of the system is partially observed. Monahan (1982) Surveys POMDPs theory and he presents a few POMDP articles devoted to maintenance and non-maintenance applications.

Pierskalla and Voelker (1976) suggested seven dimensional classifications of maintenance models:

1. states of the system: age, degree of deterioration, number failed components, number of operating components, cumulative wear number of shocks
2. Actions available: repair, replacement, opportunistic replacement, periodic inspections, etc.
3. Time horizon: finite and infinite, discrete or continuous.
4. knowledge of the system: complete or incomplete information
5. Nature of the model: deterministic, stochastic.
6. Objectives of the system: minimize cost or unavailability.

7. Method of solution: LP, DP, etc.

White (1979, b) and Maillart (2005, 2006) classify POMDP problems into the following categories, based on the availability of true information regarding the true state of the system:

1. No Observations (NO): this means the controller does not receive any information regarding the state of the system; thus, the true state of the system is completely unobserved.
2. Perfect Observations (PO): in this case, the controller has the option to know the true state of the system with some cost. But if the state is assumed to be correctly observed every time epoch, then the model reduces to that of MDP.
3. Imperfect Observations (IO): this is the general case, where the controller receives only noise-corrupted information; also, in this case the observation times are adaptively scheduled. Usually, in IO it is possible to know the true state of the system after the replace action which renews the system.

It is shown in the literature that the Bayesian update of a POMDP is a sufficient statistic; thus, a POMDP can be reformulated as a MDP whose possible states are nothing but the different possible state occupancy vectors. Where, the Bayesian update is shown to be a sufficient statistic in which all of the history of a POMDP in terms of actions and observations is embedded in. see for example, Striebel (1965) and Smallwood and Sondik (1973).

In one of the early key papers that characterize the structure of the belief space, Ross (1971) has shown that, for a discrete time, finite stage, partially observed Markov chain, and with three possible maintenance actions available for the controller, namely:

- Do nothing: System will keep deterioration
- Inspect: Reveals the true state of the system on hand
- Replace: Takes the system back to the as good as new state (State 0)

The optimal discounted and cost-minimizing value function is piecewise linear concave function. Also, Ross has proved that the regions of the belief space corresponding to both, the inspect and replace actions, are convex regions. Also, Ross has shown that when n=2, the optimal maintenance policy has a counterintuitive At Most Four Region policy (AM4R) as follows:

$0 \le p \le p_1$	Do nothing
$p_1 \le p \le p_2$	Inspect
$p_2 \le p \le p_3$	Do nothing
$p_3 \le p \le 1$	Replace

Where p is the probability the system is in state 0.

In Smallwood and Sondik (1973), it has been shown that for a discrete time and finite horizon, the optimal value profit function is piecewise-linear convex function over the space of state occupancy vectors. This encouraged the development of an algorithm well known as “Sondik’s one pass algorithm” presented in the same paper.

Although, a POMDP is equivalent to an MDP with an enlarged state space, namely, the belief space which is continuous. Still, research has provided some geometrical characterizations of the newly formulated problem that enabled the evaluation and characterization of the solutions.

Later on, Hughes (1978) considered the case of n=2 in Ross’s model, above, with the assumption of maintenance action should be proceeded by an inspection action. As a result, it was shown that optimal maintenance policy has the structure of AM2R policy.

Maillart (2005) considered cost rate minimization for multi-state POMDP. Both cases of Perfect information and imperfect information were considered under obvious failures (failures that are apparent to the controller opposite to silent failure which can not be discovered unless inspection action is done). Maillart has shown that for three control actions available, namely: do nothing, inspect, and replace, the optimal value function is piecewise linear concave in the state occupancy vectors ordered by First order Stochastic Dominance and the optimal policy is of MA4R policy form. The contribution of Maillart's work is considering obvious failures and consideration of n-state system.

With the state of the system defined by two indices(i, k); such that, k time units ago, the system was truly known to be in state i; Rosenfield (1976) considered Ross's n state system, and showed that an AM4R policy exists if the matrix of the system's states transition is restricted to be upper triangular and totally positive of order 2.

White extended the results of Ross and Rosenfield, such that, it is shown in White (1978) that the optimal maintenance policy is an AM4R policy in the space of state occupancy vectors ordered by first order stochastic dominance. In his work, White extended the result by Ross by considering n-state system. And gave the same result due to Rosenfield who restricted the P matrix to be upper triangular matrix and TP2 with a different less restrictive assumption.

Kuo (2006) provided a POMDP model to determine the optimal sampling size and maintenance policy for a finite horizon POMDP model. Unlike, other POMDP maintenance models that assume fixed or a continuous sampling scheme, Kuo considered the sampling size as a decision variable. Hence, the optimal sampling size is considered as well as the optimal maintenance schedules.

For a two state systems ($S = \{0, 1\}$); it is usually easier to derive and observe structural results for the POMDP model. This is because the state occupancy vector can be summarized by only one number, for instance p which is the probability that the unobserved system is in state 0 this means that the probability of being in state two = *1-p*. Grosfeld-Nir has worked on the two-state POMDP problem. In Grosfeld-Nir (1996) it is shown that for a two state (good and bad) POMDP with silent failure and under the availability of two actions (do nothing and replace). A control limit policy exists, that is, do nothing if the probability of being in the good state is less than a threshold value and replace if the probability exceeds the threshold value. Also, for the uniformly-distributed observations case, it is shown that the control limit value as a function in time remaining in the finite horizon is not Monotone and this is a counterintuitive result. Next, Grosfeld-Nir (2007) proved that for two state POMDP model, the dominance expectation, which is weaker than first order stochastic dominance and implied by it is sufficient for the optimal maintenance policy to be of threshold type policy. That is, for the available do nothing and replace actions, it is doing nothing if the probability that the system is in the good state exceeds a threshold value.

One research direction that is related to POMDP's is obtaining structural results as follows:

1. Optimal value function: Monotonicity and/or convexity or concavity.
2. Optimal policy: Monotonicity and/or number of decision regions, and control limit.
3. Decision Regions (resembled by state occupancy vectors): convexity and/or

number of regions per policy.

To obtain structural results for the n-state partially observed Markov processes, far to our knowledge, three partial orders are reported to be applied in the literature, namely:

1. Stochastic Dominance
2. Monotone Likelihood Ratio
3. Marginal Monotonicity

Structural results for optimal policies are useful in two ways, namely, developing efficient computational algorithms, and ease of application of the optimal policies.

In the literature, very few people addressed structural results of the POMDP in its general settings. Usually, people derived structural results for POMDPs assuming certain partial order. In the following three paragraphs, the literature that addresses partial orders-based structural results of the POMDPs is presented. All of them assume silent failures, where a silent failure is an unobserved failure.

Extending the two conditions by Derman (1962) which guarantee the control-limit type optimal maintenance policy, White (1979, a) presents three conditions which are sufficient for the optimal maintenance policy to be control-limit policy under the assumption of a POMDP governing the system.

White (1980, a) provides conditions that guarantee the existence of monotone optimal control policies over the space of state occupancy vectors ordered by the stochastic dominance partial order ($\geq_S$)for the m-state, n-action totally observed Markov process (MDP) and for the completely unobserved case. With two counter examples, White showed that the monotonicity conditions for the completely observed case are not strong enough for the more general case of completely unobserved case; and the conditions of the completely unobserved case are not strong enough for the most general case of partially observed system.

Later on, and based on White (1979, a), Lovejoy (1987) derived an important structural for discrete-time, finite-horizon POMDP, Lovejoy provides two monotonicity results for the objective function and for the optimal policy, that is, optimal value function has a monotonic structure over the space of state occupancy vectors ordered by Monotone Likelihood Ratio (MLR). Also, he showed that the optimal policy is monotone. Sufficient conditions that guarantee these monotonicity results were derived. The main point in Lovejoy's model is the choice of MLR partial order which enabled deriving conditions that guarantee its survival against conditioning. Lovejoy's result can be regarded as a generalization of Albright (1979) where two state systems are considered. Also, it is noteworthy to mention that in the two state contexts stochastic dominance is equivalent to the monotone likelihood ratio order.

Ivy and Pollack (2005) and Ivy (1998) provides sufficient conditions for the POMDP model, under the marginal monotonicity partial order, to have a monotone piecewise linear concave function. Also, it is shown that optimal maintenance policies are monotone for the n-state n-action POMDP model. Actions effects where assumed to be perfect.

In our proposed research we will be considering the Reverse Hazard partial order to characterize the optimal maintenance policy for n-state m-action POMDP machine maintenance models.

As described before, a POMDP reduces to MDP with the belief space (all possible state occupancy vectors) as the state set, which is countable infinite. This makes the solution of such problems difficult. The main contribution in this regard was due to Smallwood and Sondik (1973) where it was shown that the optimal value function of a POMDP problem is pricewise linear function. They used this result to provide Sondik's one pass algorithm, where it is enough to explore some belief state vectors to characterize the optimal value function. Later on, many algorithms have been developed such as: White's algorithm, Cheng's linear support, Zhang incremental pruning and others. Refer to Lovejoy (1991) for an excellent survey on POMDP solution techniques. Other useful surveys can be found in White (1991) and Monhan (1982).

Availability in multi-state systems models

In addition to possible differences in the governing stochastic process, control actions, number of states, and process observability, maintenance models can be classified based on the underlying objective function. The objective can be cost or unavailability minimization (Pierskalla and Voelker, 1976). In Virtanen (1977), the availability for a multi-state system has been generalized, such that, two generalized availability detentions are suggested, namely, mean availability of the capacity (the expected value of the proportional levels of performance at a given point in time), and availability of levels of performance (probability that the level of performance of the system at time t is at least equal to some fixed value). A three-state system has been analyzed in (Chen and Trivedi, 2001). The system behavior in the steady state is characterized for the embedded Markov chain, where, system's failure time, repair time, maintenance time, and the maintenance interval are all generally distributed. The model determines the optimal replacement interval such that a certain availability target is achieved. Later on, Cao *et. al.* (2002) extended the model of Chen and Trivedi (2001) for a system with n outages. Kuo (2006) provided a POMDP model to determine the optimal sampling size and maintenance policy for a finite horizon POMDP model. AlDurgam and Duffuaa (2009) provided a mathematical model for three-state POMDP model, the model maximizes OSE, such that, each state of the system has a fixed process rate and quality rate level. In this paper, this model will be extended to reflect the components of OSE on the POMDP framework in an explicit way for an n-states system. Where, The concept of OEE was developed in Total Productive Maintenance (TPM). TPM is a Japanese philosophy for maintenance. According to Nakajima (1988), OEE is defined as:

$$OEE = Availability\ \times Process\ Rate\ \times Quality\ Rat$$

3.3 MEASUREMENT ERRORS MODELING

The significance of quality measurement error has been widely reported in quality related literature. Case *et. al*. (1977) considered the effect of quality measurement errors on variables acceptance sampling plans and quantify the cost effect of such measurement errors.

Duffuaa (1996) studied the effect of measurement errors on complete repeat inspection plans, for multi-characteristic critical components. Where, the statistical and economic impact of measurement errors has been investigated. it has been shown by Duffuaa that both, type I and type II errors have a significant effects on the repeat inspection sampling plans.

Duffuaa and Khan (2005) considered quality measurement errors in designing acceptance sampling plans for multi-characteristic components. It was assumed that there exist several types of measurement errors by means of misclassification. For an incoming good, or rework or a scrapped part it was possible to have six types of classifications errors. Namely, good as scrapped or rework, rework as scrapped or good, and scrapped as rework or good. the authors considered the statistical as well as the economical effect of these type of errors.

Also, due to its significance, measurement error is an essential factor in any statistical control chart design. Basically, a control chart is nothing but a repetitive hypothesis testing of a process mean and/or variation. Where the limits of the control chart are determined based on some conflicting costs, which include: costs of sampling, costs of fixing the causes of assignable causes, and type I and type II errors costs. In any control chart design, type I and type II errors costs are basic ingredients. This indicates the significance of quality measurement errors. For examples: Duncan (1956) designed the first quality control chart assuming an exponential distributed times between out of control states. Banerjee and Rahim (1988) considered a Weibull shock model. Whereas, Rahim and Banerjee (1993) considered a generally distributed, time between shocks with an increasing hazard rate for their control chart design.

Hong and Elsayed (1999) considered an extended version of the classical can filling model by Hunter and Kartha's (1977). Quality measurement error is also assumed to exist in the model, where, the effect of the measurement errors reflects on the optimal solution of the problem.

Similarly, Duffuaa and Siddiqui (2003) considered the same process targeting problem, but with multi-class screening. Also, the measurement errors reflects on the quality of the solution obtained.

The literature review shows there is a room to extend the work in this area by establishing the existence of optimal maintenance polices for finite horizon, multi-state, multi-action stochastic partially observable system under different partial orders, investigating the effect of measurement errors, and using other comprehensive objectives such as maximizing Overall System Effectiveness (OSE).

3.4 CONCLUSION

This chapter has surveyed the literature of multi-state deteriorating systems from different points of view, with the focus given to structural results of MDP and POMDP models. The literature shows that there is a room to extend the work in this area. For examples, by establishing the existence of optimal maintenance polices for finite horizon, multi-state, multi-action stochastic partially observable system under different partial orders, investigating the effect of measurement errors, and using other comprehensive objectives such as maximizing Overall System Effectiveness (OSE).

CHAPTER 4

STRUCTURED OPTIMAL MAINTENANCE POLICIES FOR TWO-STATE MACHINE MAINTENANCE PROBLEM

4.1 INTRODUCTION

In this chapter a two-state POMDP model is provided. This is to facilitate the understanding of the n-state model provided next in Chapter 5. The goal is to derive conditions for the POMDP problem parameters to ensure the existence of optimal threshold-type policy. A policy is nothing but the action that will be selected at any time for a given belief state vector.

The rest of this chapter is organized as follows: the notation used in this chapter and a precise statement of the problem are provided in Section 4.2, In order to make the chapter self-contained, some important partial orders are defined in Section 4.3, also, some implications relations are highlighted form the literature in addition to new ones being developed to help in deriving the results of this chapter and Chapter 5, in Section 4.4, the developed mathematical model is presented, in Section 4.5. the optimal maintenance policy is characterized and, finally, some concluding remarks are provided in Section 4.6.

Next, a formal definition of the problem is provided for a two-state POMDP model.

4.2 NOMENCLATURE AND STATEMENT OF THE PROBLEM

In this section the nomenclature used in this chapter is presented in Subsection 4.2.1 and Subsection 4.2.2 provides a precise statement of the problem.

4.2.1 NOMENCLATURE

Follows is the nomenclature used throughout this chapter.

S	System's state set {1,2}
i, j	Elements of S
O	Observations set with components {1=good, 2=bad}
k	Element of O
A	Totally ordered actions set available to the decision maker
a_i	Element of A, that is either a_0 (do nothing) or a_1 (replace)
$\mathrm{g}(.\,, a)$	Reward the system generates if it is in state S and action a was taken
π	State occupancy vector, or belief state
$V_t^*(\pi, a)$	The optimal value function at time t given belief state vector π and maintenance action a
P^a	System's state transition matrix corresponding to action a (2×2)
$P^a(i)$	The i^{th}raw of the P matrix
P_{ij}^a	Probability that the system will move from state i to state j if action a was taken
R^a	A (2×2) state observation transition matrix subject to action a
$r(k)$	The k^{th}column of the matrix R.
$r(j)$	The j^{th}raw of the R^a matrix
r_{jk}	And entry in the R^a matrix which gives the probability that observation k will be observed if the system has moved to state j and action a has been taken.
$T(\pi, a, k)$	Posterior state occupancy probability vector
(π, a)	The probability vector of the system observations $\{1,2\}$ when action a is taken and the system is believed to be in π
$\sigma(k; \pi, a)$	The k^{th}component of $\sigma(\pi, a)$
β	Discount factor

Next, a precise statement of the problem is provided.

4.2.2 STATEMENT OF THE PROBLEM

In this chapter a two-state, two-action partially observable Markovian system is considered. With the state set defined as $S = \{0, 1\}$. State 1 means that the system is up; whereas, state 2 represents the system when it is failed. There are two control actions available to the decision maker, namely: do nothing (a_0) and replace (a_1). When a control action is taken the system state might change depending on the control action taken and the current underlying state of the system. This can be represented by the following 2×2 state transition matrices corresponding to both actions:

$$\mathrm{P}^{a_0} = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \end{array}, \quad \mathrm{P}^{a_1} = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \end{array}$$

For instance, p_{ij}^{a}: represents the probability that the system will move to state j given that it was actually in state i before maintenance action a is taken, where, $i\ and\ j\ \in S$

This, also, can be represented by the following state transition diagrams (Figure 4.1):

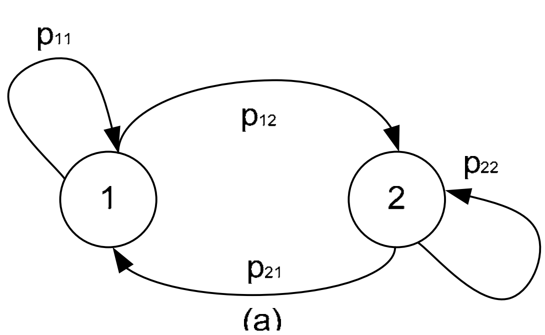


Figure 4.1 State-transition diagrams for the (a) do nothing and (b) replace actions.

For instance, Figure 4.1 (b) illustrates that the replace action is assumed to be perfect; that is, up on replacement, the system is moved to state 1 irrespective to the underlying previous state of the system.

The decision maker takes maintenance actions at discrete points in time depending on the true underlying state of the system, where, it is assumed that the state of the system is not directly observable to the decision maker; instead, only noise-corrupted information is available to him. This information is assumed to depend only on the true underlying state of the system. Hence, this can be represented by a state observations transition matrix as follows:

$$\mathrm{R} = \begin{array}{c c} & \begin{array}{cc} Good & Bad \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \end{array}$$

Notice that the observations set is defined as $O = \{Good = 1, Bad = 2\}$, where, r_{jk} represents the probability that the decision maker receives outcome k given that the system is actually in state j, $j \in S$ and $k \in O$. For instance, r_{11} represents the probability of receiving the signal (process output) "good" if the system is actually in state 1.

Regarding the reward structure for this problem, at any time stage, it will be assumed that the reward gained due to both control actions is state-action dependent reward as follows

$$g(a_0) = [\mathrm{g}(1, a_0) \quad \mathrm{g}(2, a_0)],\ g(a_1) = [\mathrm{g}(1, a_1) \quad \mathrm{g}(2, a_1)]$$

This reward structure has two vectors with two elements each. For illustration, $\mathrm{g}(1, a_0)$ represents the reward obtained if action a_0 is taken when the system was actually in state 1.

Now, having the states of the system indirectly observed by the decision maker, through observations, the decision maker takes the control actions based on the belief state or the state occupancy vector. This is nothing but a vector with dimension equal to the number of system states, with the i^{th} element of the vector, for instance, giving the probability of the system being actually in state i. This state occupancy vector has to be updated at every time stage before actions are taken. This is usually referred to as the Bayesian update. The Bayesian update takes the system previous state, control action and observations as inputs. The output is a new state occupancy vector. Mathematically, for a two state system It can be shown that the Bayesian update of the POMDP model can be represented by the following equation. The proof is a special case of that presented in Chapter 7 (Proposition 1) by letting $R^e = I$ (the identity matrix).

$$T = [T_1(\pi, a, k) \;\; T_2(\pi, a, k)]$$

$$\text{Such that, } T_j(\pi, a, k) = \frac{\sum_{i=1}^{2} \pi_i p_{ij} r_{jk}}{\sum_{j=1}^{2} \sum_{i=1}^{2} \pi_i p_{ij} r_{jk}} \;\; j \in S, k \in O$$

Where, π represents the initial belief state vector.

4.3 SOME PARTIAL ORDERS DEFINITIONS AND RELATIONS

Some important partial orders are presented in this section. Also, two new propositions are provided to characterize the relations between the different partial orders. This will be very useful in developing the results in this chapter and Chapter 5.

Definition 1: First Order Stochastic Dominance (FSD) $(\geq_{SD})$

This is a basic partial order, and it is widely applied in decision making theory and its applications.

A probability vector $\pi \geq_{SD} \pi'$ (π stochastically dominates π' in FSD sense) if and only if:

$\sum_i^n \pi_i \leq \sum_i^n \pi_i' \qquad\qquad \forall\, i \in S$

This means that the probability of observing outcome i or more in a given random process is greater or equal for the probability vector π than that of π'. Basically, this is how first order stochastic dominance is presented in the financial applications, where, in these applications the more is better.

In the application of this paper, the states of the system are ordered from best to worst starting form state 1 to state n. According to this ordering (less is better), FSD can be expressed as follows:

$$\sum_0^i \pi_i \geq \sum_0^i \pi_i' \qquad \forall\, i\ \in S$$

Next, definitions of five stronger forms of the FSD are provided.

Definition 2: Monotone Likelihood Ratio ($\geq_{MLR}$)

A probability vector π is $\geq_{MLR} \pi'$ (π dominates π' in MLR sense) if and only if:

$$\frac{\pi_{i'}}{\pi_i} \geq \frac{\pi'_{i'}}{\pi'_i} \qquad \forall\, i\ \in S\ with\, i' \leq i$$

This is presented in the literature as a stronger form of the first order stochastic dominance

Definition 3: Monotone Probability Ratio ($\geq_{MPR}$)

Hopkins and Kornienko (2007) defined this partial order as: a probability vector π is $\geq_{MPR} \pi'$ (π dominates π' in MPR sense) if and only if:

$$\frac{\sum_0^{i'} \pi_{i'}}{\sum_0^i \pi_i} \geq \frac{\sum_0^{i'} \pi'_{i'}}{\sum_0^i \pi'_i} \qquad \forall\, i' \leq i,\ i\ and\ i' \in S$$

Definition 4: Reverse Hazard ($\geq_{RHazard}$)

Hopkins and Kornienko (2007) defined this partial order as: a probability vector π is $\leq_{RHazard} \pi'$ (π dominates π' in reverse hazard sense) if and only if:

$$\frac{\pi_i}{\sum_0^i \pi_i} \leq \frac{\pi'_i}{\sum_0^i \pi'_i} \qquad \forall\, i\ \in S$$

Definition 5: Component-wise Dominance Partial Order ($\geq_{CD}$)

A probability vector π is $\geq_{CD1} \pi'$ (π dominates π' component-wise) if and only if:

$$\pi_i \geq \pi'_i \qquad\qquad i = 1, 2, \dots n-1$$

with $\pi_n \leq \pi'_n$ since a probability vector's components should add up to one. Another possible variation is the case where $\pi_1 \geq \pi'_1$ and $\pi_i \leq \pi'_i$ for $i = 2, \dots n$ this variation is expressed as follows: $\pi \geq_{CD2} \pi'$ It is easily verifiable that both of these forms imply first order stochastic dominance.

Figure 4.2 shows the implications relations between the different partial orders presented in this section; where, relation 1 is shown by Proposition 1, relation 2 is shown in Hopkins and Kornienko (2007), 3 and 4 are shown in Eeckhoudt and Gollier (1995), relation 5 is shown by propositions 2 in this paper, and finally, relation 6 is very easy to verify from the definitions. It is easily verifiable, from Figure 4.2, that all the partial orders defined in this section are different forms of the first order stochastic dominance, but in a stronger sense.

Definition 6: Marginal Monotonicity Partial Order ($\geq_{MM}$)

A probability vector π is $\geq_{MM} \pi'$ (π dominates π' in marginal monotonicity sense) if and only if for one $i \in \{1, 2, \dots n-1\}$: $\pi_i \geq \pi'_i$ and

$$\pi_j = \pi'_j \qquad\qquad \forall\ j = 1, 2, \dots n-1, j \neq i$$

Notice here that: $\pi_n \leq \pi'_n$ since a probability vector's components should add up to one. i.e. a movement in one of the Cartesian directions.

$$\begin{tikzcd}
\pi \geq_{CD2} \pi' \arrow[d, "1"] & \pi \geq_{MLR} \pi' \arrow[d, "3"] & \pi \geq_{CD1} \pi' \arrow[d, "5"] & \pi \geq_{MM} \pi' \arrow[l, "6"'] \\
\pi \leq_{RHazard} \pi' \arrow[r, "2"] & \pi \geq_{MPR} \pi' \arrow[r, "4"] & \pi \geq_{SD} \pi' &
\end{tikzcd}$$

Figure 4.2 Partial orders implications relations

Proposition 1

$\leq_{\text{CD2}} \rightarrow \leq_{RHazard}$ and does not imply $\leq_{MLR}$ as Figure 4.2 shows.

Proof:

Let π and π' any arbitrary probability vectors with $\pi \geq_{\text{CD2}} \pi'$, for $\pi \leq_{\text{RHazard}} \pi'$ to hold true the following relation should hold:

$$\frac{\pi_i}{\pi_1+\pi_2+\cdots+\pi_i} \leq \frac{\pi_i'}{\pi_1'+\pi_2'+\cdots+\pi_i'} \quad \forall\, i \in S$$

This inequality is true for $i = 1$.

For $i \geq 2$, we have the following:

Whereas, for the numerator

$\pi_i \leq \pi_i'$ for $i = 2, \dots n$, by the definition of $\leq_{\text{CD2}}$

Whereas, for the denominator of the inequality:

Since $\pi_i \leq \pi_i'$ for $i = 2, \dots n$, by the definition of $\leq_{\text{CD2}}$, then:

$\pi_{i+1} + \pi_{i+2} + \cdots + \pi_n \leq \pi_{i+1}' + \pi_{i+2}' + \cdots + \pi_n'$ is true

$1 - (\pi_{i+1} + \pi_{i+2} + \cdots + \pi_n) \geq 1 - (\pi_{i+1}' + \pi_{i+2}' + \cdots + \pi_n')$

$\pi_1 + \pi_2 + \cdots + \pi_i \geq \pi_1' + \pi_2' + \cdots + \pi_i'$

This gives:

$$\frac{\pi_i}{\pi_1+\pi_2+\cdots+\pi_i} \leq \frac{\pi_i'}{\pi_1'+\pi_2'+\cdots+\pi_i'} \quad \forall\, i \in S$$

Based on the definitions of partial orders provided and noticing that probability vectors components should add up to one, the result follows. Finally, notice that: $\pi = [0.5 \quad 0.2 \quad 0.3]$ and $\pi' = [0.2 \quad 0.4 \quad 0.4]$ is a counter example where $\pi \geq_{CD2} \pi'$ but not $\geq_{MLR}$.

Proposition 2

$\leq_{\text{CD1}} \rightarrow \leq_{SD}$ and does not imply $\leq_{MLR}$ as Figure 4.2 shows.

Proof:

From the definitions, it can be easily shown that $\leq_{\text{CD1}}$ implies first order stochastic dominance. And by the following counterexample: $\pi = [0.3 \quad 0.3 \quad 0.4]$ and $\pi' = [0.2 \quad 0.1 \quad 0.7]$ where $\pi \geq_{CD1} \pi'$ but not $\geq_{MLR}$.

Next, in Section 4.4, this two state decision making problem will be modeled as a POMDP.

4.4 MODEL FORMULATION

As illustrated in Chapter 2, the problem on hand can be modeled, using Bellman optimality equation, as follows:

For a_0, the expected reward value with t time horizons left can be represented as follows:

$$V_t(\pi, a_0) = \sum_{i \in S} \pi_i \mathrm{g}(i, a_0) + \beta \times \{\sum_{k \in O} \sigma(k; \pi, a_0) \times V_{t-1}^*(T(\pi, a_0, k))\}$$

$$= \pi_1 \mathrm{g}(1, a_0) + \pi_2 \mathrm{g}(2, a_0) + \beta \times \{\sum_{k \in O} \sigma(k; \pi, a_0) \times V_{t-1}^*(T(\pi, a_0, k))\}$$

Noticing that a state occupancy vector (π) components should add up to one, the equation above can be expressed as follows. Hence, the reward function reduces to a function in one variable that is π_1 as follows:

$$= \pi_1 \mathrm{g}(1, a_0) + (1 - \pi_1) \mathrm{g}(2, a_0) + \beta \times \{\sum_{k \in O} \sigma(k; \pi, a_0) \times V_{t-1}^*(T(\pi, a_0, k))\} \quad (4.1)$$

Where:

$$T_j(\pi, a, k) = \frac{\sum_{i=1}^{2} \pi_i p_{ij} r_{jk}}{\sum_{j=1}^{2} \sum_{i=1}^{2} \pi_i p_{ij} r_{jk}} \; j \in S, k \in O$$

$$T_1(\pi, a_0, k) = \frac{\sum_{i=1}^{2} \pi_i p_{i1} r_{1k}}{\sum_{j=1}^{2} \sum_{i=1}^{2} \pi_i p_{ij} r_{jk}} = \frac{\pi_1 p_{11} r_{11} + \pi_2 p_{21} r_{11}}{\pi_1 p_{11} r_{11} + \pi_2 p_{21} r_{11} + \pi_1 p_{12} r_{21} + \pi_2 p_{22} r_{21}}$$

$$T_2(\pi, a_0, k) = \frac{\sum_{i=1}^{2} \pi_i p_{i2} r_{2k}}{\sum_{j=1}^{2} \sum_{i=1}^{2} \pi_i p_{ij} r_{jk}} = 1 - T_1(\pi, a_0, k)$$

$$\forall\, k \in O$$

$\sigma(k; \pi, a) = \sum_{j=1}^{2} \sum_{i=1}^{2} \pi_i p_{ij} r_{jk}$, the denominator of the Bayesian update.

Equation 4.1 means that if action a_0 is taken, that is, the system is left to deteriorate without any maintenance action, with t time horizons left, an instantaneous reward will be gained. This reward is function of both the current state of the system and the action taken. In addition, an expected discounted quantity representing the future is gained as well. The expectation for the future discounted reward is done with respect to the different possible next time epoch observations, such that, the probability of observing an outcome k takes place with probability $\sigma(k; \pi, a_0)$. This probability also shows that observing an outcome k depends on the belief state vector π and the action taken.

Assuming that replaced systems return as good as new, and noticing that $\sigma(\pi, a)$ is a probability vector whose components $(\sigma(k; \pi, a), k \in O)$ add up to one. Then, for the replace action (a_1), the expected reward over t time horizons left can be expressed as follows:

$$V_t(\pi, a_1) = \sum_{i \in S} \pi_i \mathrm{g}(i, a_1) + \beta \times \{\sum_{k \in O} \sigma(k; \pi, a_0) \times V_{t-1}^*(e_1)) =$$

$$\pi_1 \mathrm{g}(1, a_1) + (1 - \pi_1) \mathrm{g}(2, a_1) + \beta \times \{V_{t-1}^*(e_1)\} \tag{4.2}$$

For the replace action, the Bayesian update is easily verifiable to be equivalent to:

$T = e_1$, i.e:

$T_1(\pi, a_1, k) = 1$

$T_2(\pi, a_1, k) = 0$

$\forall\, k \in O$

At any time stage, the optimal reward function is the maximum of both reward functions corresponding to both of the available control actions. Mathematically, this can be represented as follows

$V_t^*(\pi) = \max\ \{V_t(\pi, a_0), V_t(\pi, a_1)\}$

$V_t(\pi, a_0), V_t(\pi, a_1)$ are expressed by Equations 4.1 and 4.2. respectively. Hence, the objective is to find the optimal course of action (a) at any point of time t, for any belief state π, represented by π_1 $(since\ \pi_2 = 1 - \pi_1,$ as illustrated before). This is refrred to usually as the optimal policy.

Next, the model of this section is analyzed in Section 4.5.

4.5 ANALYSIS AND RESULTS

In this section, some conditions are derived to ensure the existence of optimal threshold-type maintenance policy for the two state system model provided in Section 4.4.

Figure 4.3, below, represents a threshold-type (or cut-off) optimal maintenance policy, where, it is optimal to replace the system in the region to the left of π_1^* (the $[0, \pi_1^*]$ interval) and to do nothing otherwise. Structured type of policies is preferred for its ease of applications, and due to the computational advantage they provide.

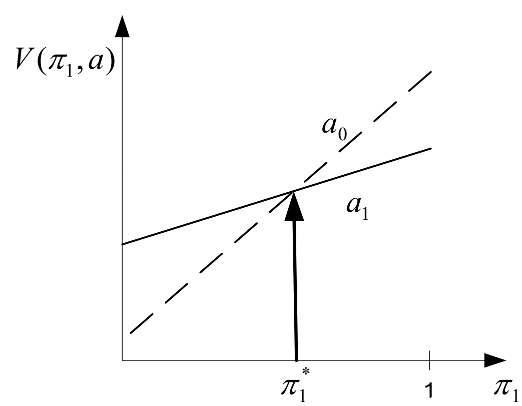


Figure 4.3 Optimal threshold policy for two-state, two-action system

As illustrated previously, belief state vectors of two-state systems can be represented by only one of the two state occupancy probabilities of the vector, and since the summation of probabilities of a state occupancy vector should add up to one, the other probability value of the vector can be obtained easily. Hence, we need only to deal with one variable; as such, we have a totally ordered set. Finally, for two state systems, all the partial orders reduce to be the same.

For two state systems the threshold structure is very well established for our problem. In this chapter, the same result will be presented to facilitate the understanding of the n-state system presented in Chapter 5. Follows, are some Lemmas that will be needed for establishing the main result of this chapter.

Lemma 1:

If X is a countable completely ordered set and π and $\pi'$are elements of $\Pi(S)$ then $\pi \geq_{SD} \pi'$ if and only if $\sum_{i=1}^{n} \pi_i f(i) \geq \sum_{i=1}^{n} \pi_i' f(i)$ for every $f: X \to R$ non-increasing on X

Proof:

Refer to Stoyan (1983) for the proof.

Lemma 2:

If $a \geq a'$ and $b \geq b'$, where $\{a, a', b,$ and $b'\} \in [0,1]$, then $ab + a'b' \geq ab' + a'b$

Proof:

This Lemma is a straight forward extension of Lemma 1.

Let $a = a' + \Delta_1$, and $b = b' + \Delta_2$, such that, both Δ_1 and $\Delta_2 \geq 0$, then substituting in the inequality above yields:

$(a' + \Delta_1)(b' + \Delta_2) + a'b' \geq (a' + \Delta_1)b' + a'(b' + \Delta_2)$

$\equiv$

$a'b' + a'\Delta_2 + b'\Delta_1 + \Delta_1\Delta_2 + a'b' \geq a'b' + b'\Delta_1 + a'b' + a'\Delta_2$

Hence $\Delta_1\Delta_2 \geq 0$ and the result follows.

Lemma 3:

If $p_{11}p_{22} \geq p_{12}p_{21}$, then $T(\pi, a_0, k) \geq_{SD} T(\pi', a_0, k)$ $\forall\, \pi \geq_{SD} \pi'$, and $k \in O$

Proof:

By the definition of stochastic dominance, $T(\pi, a_0, k) \leq_{RHazard} T(\pi', a_0, k)$ if and only if $T_1(\pi, a_0, k) \geq T_1(\pi', a_0, k)$, writing this explicitly:

$$\frac{\pi_1 p_{11} r_{1k} + \pi_2 p_{21} r_{1k}}{\pi_1 p_{11} r_{1k} + \pi_2 p_{21} r_{1k} + \pi_1 p_{12} r_{2k} + \pi_2 p_{22} r_{2k}} \geq \frac{\pi_1' p_{11} r_{1k} + \pi_2' p_{21} r_{1k}}{\pi_1' p_{11} r_{1k} + \pi_2' p_{21} r_{1k} + \pi_1' p_{12} r_{2k} + \pi_2' p_{22} r_{2k}}$$

$\equiv$

$$\pi_1 p_{11} r_{1k} \pi_1' p_{11} r_{1k} + \pi_1 p_{11} r_{1k} \pi_2' p_{21} r_{1k} + \pi_1 p_{11} r_{1k} \pi_1' p_{12} r_{2k} + \pi_1 p_{11} r_{1k} \pi_2' p_{22} r_{2k}$$
$$+ \pi_2 p_{21} r_{1k} \pi_1' p_{11} r_{1k} + \pi_2 p_{21} r_{1k} \pi_2' p_{21} r_{1k} + \pi_2 p_{21} r_{1k} \pi_1' p_{12} r_{2k} + \pi_2 p_{21} r_{1k} \pi_2' p_{22} r_{2k} \geq$$

$$\pi_1 p_{11} r_{1k} \pi_1' p_{11} r_{1k} + \pi_2 p_{21} r_{1k} \pi_1' p_{11} r_{1k} + \pi_1 p_{12} r_{2k} \pi_1' p_{11} r_{1k} + \pi_2 p_{22} r_{2k} \pi_1' p_{11} r_{1k}$$
$$+ \pi_1 p_{11} r_{1k} \pi_2' p_{21} r_{1k} + \pi_2 p_{21} r_{1k} \pi_2' p_{21} r_{1k} + \pi_1 p_{12} r_{2k} \pi_2' p_{21} r_{1k} + \pi_2 p_{22} r_{2k} \pi_2' p_{21} r_{1k}$$

$\equiv$

$$\pi_1 \pi_2' p_{11} p_{22} + \pi_1' \pi_2 p_{21} p_{12} \geq \pi_1' \pi_2 p_{22} p_{11} + \pi_1 \pi_2' p_{12} p_{21}$$

Noticing that $\pi \geq_{SD} \pi'$ implies $\pi_1 \pi_2' \geq \pi_1' \pi_2$ the result follows by Lemma 2. Next, a sufficient condition for Lemma 3 condition is provided

Lemma 4:

If $P(1) \geq_{CD2} P(2)$ then $p_{11}p_{22} \geq p_{12}p_{21}$

Proof:

The proof is straightforward by the definition of first order stochastic dominance.

Lemma 5:

If $r_{11} \times r_{22} \geq r_{12} \times r_{21}$, then $T(\pi, a_0, k=1) \geq_{SD} T(\pi, a_0, k=2) \;\; \forall \, \pi \in \Pi(S)$

Proof:

By the definition of first order stochastic dominance, $T(\pi, a_0, k=1) \geq_{SD} T(\pi, a_0, k=2)$ if and only if $T_1(\pi, a_0, k=1) \geq T_1(\pi', a_0, k=2)$, writing this explicitly:

$$\frac{\pi_1 p_{11} r_{11} + \pi_2 p_{21} r_{11}}{\pi_1 p_{11} r_{11} + \pi_2 p_{21} r_{11} + \pi_1 p_{12} r_{21} + \pi_2 p_{22} r_{21}} \geq \frac{\pi_1^{'} p_{11} r_{12} + \pi_2^{'} p_{21} r_{12}}{\pi_1^{'} p_{11} r_{12} + \pi_2^{'} p_{21} r_{12} + \pi_1^{'} p_{12} r_{22} + \pi_2^{'} p_{22} r_{22}}$$

$\equiv$

$$\pi_1 p_{11} r_{11} \pi_1^{'} p_{11} r_{12} + \pi_1 p_{11} r_{11} \pi_2^{'} p_{21} r_{12} + \pi_1 p_{11} r_{11} \pi_1^{'} p_{12} r_{22} + \pi_1 p_{11} r_{11} \pi_2^{'} p_{22} r_{22}$$

$$+ \pi_2 p_{21} r_{11} \pi_1^{'} p_{11} r_{12} + \pi_2 p_{21} r_{11} \pi_2^{'} p_{21} r_{12} + \pi_2 p_{21} r_{11} \pi_1^{'} p_{12} r_{22} + \pi_2 p_{21} r_{11} \pi_2^{'} p_{22} r_{22} \geq$$

$$\pi_1 p_{11} r_{11} \pi_1^{'} p_{11} r_{12} + \pi_2 p_{21} r_{11} \pi_1^{'} p_{11} r_{12} + \pi_1 p_{12} r_{21} \pi_1^{'} p_{11} r_{12} + \pi_2 p_{22} r_{21} \pi_1^{'} p_{11} r_{12}$$

$$+ \pi_1 p_{11} r_{11} \pi_2^{'} p_{21} r_{12} + \pi_2 p_{21} r_{11} \pi_2^{'} p_{21} r_{12} + \pi_1 p_{12} r_{21} \pi_2^{'} p_{21} r_{12} + \pi_2 p_{22} r_{21} \pi_2^{'} p_{21} r_{12}$$

$\equiv$

$$r_{11} r_{22} [\pi_1 p_{11} \pi_1^{'} p_{12} + \pi_1 p_{11} \pi_2^{'} p_{22} + \pi_2 p_{21} \pi_1^{'} p_{12} + \pi_2 p_{21} \pi_2^{'} p_{22}] \geq$$

$$r_{21} r_{12} [\pi_1 p_{12} \pi_1^{'} p_{11} + \pi_2 p_{22} \pi_1^{'} p_{11} + \pi_1 p_{12} \pi_2^{'} p_{21} + \pi_2 p_{22} \pi_2^{'} p_{21}]$$

Noticing that $\pi_1 \pi_2^{'} \geq \pi_2 \pi_1^{'}$, and by assumption that $P(1) \geq_{CD2} P(2)$ yielding $p_{11} p_{22} \geq p_{21} p_{12}$

Then by Lemma 2 $\pi_1 p_{11} \pi_2^{'} p_{22} + \pi_2 p_{21} \pi_1^{'} p_{12} \geq \pi_2 p_{22} \pi_1^{'} p_{11} + \pi_1 p_{12} \pi_2^{'} p_{21}$, then

$$[\pi_1 p_{11} \pi_1^{'} p_{12} + \pi_1 p_{11} \pi_2^{'} p_{22} + \pi_2 p_{21} \pi_1^{'} p_{12} + \pi_2 p_{21} \pi_2^{'} p_{22}] \geq$$

$$[\pi_1 p_{12} \pi_1^{'} p_{11} + \pi_2 p_{22} \pi_1^{'} p_{11} + \pi_1 p_{12} \pi_2^{'} p_{21} + \pi_2 p_{22} \pi_2^{'} p_{21}]$$

Hence, if $r_{11} r_{22} \geq r_{21} r_{12}$, the result will follow by applying Lemma 2 again.

Next, a sufficient condition for Lemma 5 condition is provided.

Lemma 6:

If $R(1) \geq_{CD1} R(2)$ then $r_{11}r_{22} \geq r_{12}r_{21}$

Proof:

The proof is straightforward by the definition of first order stochastic dominance.

Lemma 7:

If $p_{11}p_{22} \geq p_{12}p_{21}$ and $r_{11}r_{22} \geq r_{21}r_{12}$, then $\sigma(k; \pi, a) \geq_{SD} \sigma(k; \pi', a)$ $\forall\, \pi \geq_{SD} \pi'$ in $\Pi(S)$.

Proof:

The proof is special case of that in Proposition 1 in Lovejoy (1987) we restate it here for two-state systems:

By definition of $\geq_{SD}$ it will be sufficient to show that $\sigma(1; \pi, a) \geq \sigma(1; \pi', a)$, for $\pi \geq_{SD} \pi'$.

$$\sigma(k; \pi, a) = \sum_{j=1}^{2} \sum_{i=1}^{2} \pi_i p_{ij} r_{jk}$$

$$\sigma(1; \pi, a) = (\pi_1 p_{11} + \pi_2 p_{21}) r_{11} + (\pi_1 p_{12} + \pi_2 p_{22}) r_{21}$$

$$\sigma(1; \pi', a) = \left(\pi_1^{'} p_{11} + \pi_2^{'} p_{21}\right) r_{11} + \left(\pi_1^{'} p_{12} + \pi_2^{'} p_{22}\right) r_{21}$$

Note that $(\pi_1 p_{11} + \pi_2 p_{21}) \geq \left(\pi_1^{'} p_{11} + \pi_2^{'} p_{21}\right)$ and $(\pi_1 p_{12} + \pi_2 p_{22}) \leq \left(\pi_1^{'} p_{12} + \pi_2^{'} p_{22}\right)$ by Lemma 1. Also note that πP is a probability vector. Then:

$(\pi_1 p_{11} + \pi_2 p_{21}) r_{11} + (\pi_1 p_{12} + \pi_2 p_{22}) r_{21} \geq \left(\pi_1^{'} p_{11} + \pi_2^{'} p_{21}\right) r_{11} + \left(\pi_1^{'} p_{12} + \pi_2^{'} p_{22}\right) r_{21}$

by Lemma 1 again, and the result follows.

Lemma 8:

If $g(i, a)$ is non-increasing on i in S, then $V_1(\pi, a)$ is monotonically-non decreasing over $\Pi(S)$ ordered by $\geq_{SD}$. Hence, $V_1(\pi, a) \geq V_1(\pi', a)\ \forall\ \pi \geq_{SD} \pi'$ in $\Pi(S)$ and $\forall\ a \in A$

Proof:

The proof is a straight forward application of Lemma 1.

Proposition 3:

If

1. $g(i, a)$ is non-increasing on i in S
2. $P(1) \geq_{CD2} P(2)$
3. $R(1) \geq_{CD1} R(2)$

Then $V_t^*(\pi)$ is monotonically non-decreasing over $\Pi(S)$ ordered by $\geq_{SD}$. Hence, $V_t(\pi, a_0) \geq V_1(\pi', a_0)\ \forall\ \pi \geq_{SD} \pi'$

Proof:

By induction:

By assumption 1 and Lemma 8, the result is true for $t = 1$.

Now, assuming that the result is true for $t - 1$

Then for any $a \in A$

$V_t(\pi, a_0) = \sum_{i \in S} \pi_i g(i, a_0) + \beta \times \left\{\sigma(1; \pi, a_0) \times V_{t-1}^*\big(T(\pi, a_0, 1)\big) + \sigma(2; \pi, a_0) \times V_{t-1}^*(T(\pi, a_0, 2))\ \right\} \geq \sum_{i \in S} \pi_i^{'} g(i, a_0) + \beta \times \left\{\sigma\big(1; \pi_i^{'}, a_0\big) \times V_{t-1}^*\big(T(\pi_i^{'}, a_0, 1)\big) + \sigma\big(2; \pi_i^{'}, a_0\big) \times V_{t-1}^*\big(T(\pi_i^{'}, a_0, 2)\big)\ \right\} = V_t(\pi', a_0)$

$\sum_{i \in S} \pi_i g(i, a_0) \geq \sum_{i \in S} \pi_i^{'} g(i, a_0)$ by Lemma 1

$T(\pi, a_0, 1) \geq T(\pi_i^{'}, a_0, 1)$ and $T(\pi, a_0, 2) \geq T(\pi_i^{'}, a_0, 2)$ by assumption 2 and Lemma 3

$\sigma(1; \pi, a_0) \geq \sigma\big(1; \pi_i^{'}, a_0\big)$ by assumptions 1 and 2 and Lemma 7

By the assumption that the result is true for $t - 1$ and applying Lemma 1 on the quantity between the brackets the result of this Lemma follows.

The same monotonicity result for $V_t(\pi, a_1)$ follows trivially.

Hence, $V_t^*(\pi) = \max\ \{V_t(\pi, a_0), V_t(\pi, a_1)\}$ is monotonically non-decreasing as well.

Now the main result of this chapter, namely, the threshold policy can be established as follows:

Proposition 4:

If $\mathrm{g}(i, a_0) - \mathrm{g}(i, a_1)$ is either increasing or decreasing then $V_t(\pi, a_0) - V_t(\pi, a_1)$ is either increasing or decreasing, hence, an optimal threshold policy exists.

Proof:

Writing explicitly:

$V_t(\pi, a_0) - V_t(\pi, a_1) = \sum_{i\in S} \pi_i \mathrm{g}(i, a_0) - \sum_{i\in S} \pi_i \mathrm{g}(i, a_1) + \beta \times \{\sigma(1; \pi, a_0) \times$

$V_{t-1}^*(T(\pi, a_0, 1)\big) + \sigma(2; \pi, a_0) \times V_{t-1}^*(T(\pi, a_0, 2)) - V_{t-1}^*(e_1)\}$

The proof follows in a similar way as that in Proposition 3.

4.6 CONCLUSION

This chapter established the result of optimal threshold policy for two-state, two-actions partially observed system over the belief space ordered by the first order stochastic dominance for a finite horizon. The result is similar to others in the literature as in Albright (1979) and Lovejoy (1987), with a different approach followed here, that is, forcing the Bayesian update to survive conditioning by following the reverse hazard rate partial order. This is supposed to be appreciated more in the n-state model (next in Chapter 5) since for two-states all the partial orders discussed here are equivalent. The results provided here are supposed to facilitate the understanding of next chapter results.

CHAPTER 5

STRUCTURED OPTIMAL MAINTENANCE POLICIES FOR N-STATE MACHINE MAINTENANCE PROBLEM

5.1 INTRODUCTION AND MOTIVATION

In this chapter, the multi-state machine maintenance problem modeled as a Partially Observed Markov Decision Process (POMDP) over a finite horizon is considered. The optimal maintenance policies over the space of state occupancy vectors ordered by the first order stochastic dominance are characterized. A new set of sufficient conditions on the cost parameters, the core Markov process, and the observations process of the POMDP model are provided to ensure the existence of threshold-type optimal maintenance policy. This is achieved by utilizing newly established and existing relations between the first order stochastic dominance partial order and other partial orders (the reverse hazard rate and the component-wise dominance). The conditions make the first order stochastic dominance survives conditioning. This has the advantage of enlarging the set of the belief space elements over which the optimal solutions of the POMDP problem can be characterized. Finally, examples are provided to demonstrate the contribution of our model and results compared to existing results in the literature.

The main motivation behind this work is that, in spite of the wide range of POMDP models in the literature, very few structural results are reported in the literature. The main results reported center around monotonicity results which lead to obtaining optimal threshold-type control policies and structured policies (Krishnamurthy and Djoni, 2007). In fact, this is due to the complexity of the problem. Structured policies are usually obtained over partially ordered belief space. As far as to our knowledge, only three papers contributed directly to this direction, namely: White (1979, a), Lovejoy (1987), and Ivy and Pollack (2005) which has been already discussed in Chapter 3. The importance of developing optimal maintenance policies can not be underestimated. Optimal maintenance policies improve systems output quality and increase systems availability. In this chapter, the theory in the literature is extended, such that, new conditions are developed to ensure and characterize the existence of optimal maintenance policies; for multi-state systems with the state occupancy vectors ordered by first order stochastic dominance. In addition, examples are shown to demonstrate the utility of the results compared to other results in the literature. This chapter extends the results of Chapter 4 to an n-state systems.

Functions that are defined over sets with sets members belonging to multi-dimensional spaces, are usually easier to be characterized over partially ordered sets. The difference between partially and totally ordered sets, is that, for a totally ordered set it is necessary to have every two elements in the set related to each other (e.g. the numbers along the real numbers line). On the other hand, this is not necessary for partial orders (several examples are to be provided next).

The rest of this chapter is organized as follows: the notation used in this chapter and a precise statement of the problem are provided in Section 5.2, in Section 5.4, the developed mathematical model is presented, in Section 5.5 the optimal policies for Section 5.4 model are characterized, Illustrative examples are presented in Section 5.5

Finally, some concluding remarks are provided in Section 5.6.

5.2 NOMENCLATURE AND STATEMENT OF THE PROBLEM

In this section the nomenclature used in this chapter is presented in Subsection 5.2.1 and Subsection 5.2.2. provides a precise statement of the problem.

5.2.1 NOMENCLATURE

Follows is the nomenclature used throughout this chapter.

S	System's state set {1,2,…n}
i, j	Elements of S
O	Observations set with components {1,2, … m}
k	Element of O
A	Totally ordered actions set available to the decision maker
a_i	Element of A, that can range from a_0 (do nothing) to a_n (replace)
g$(.,a)$	Reward the system generates if it is in state S and action a was taken
π	State occupancy vector, or belief state
$V_t^*(\pi, a)$	The optimal value function at time t given belief state vector π and maintenance action a
P^a	System's state transition matrix corresponding to action a $(n \times n)$
$P^a(i)$	The i^{th}raw of the P matrix
P_{ij}^a	Probability that the system will move from state i to state j if action a was taken
R^a	An $(n \times m)$ state observation transition matrix subject to action a
$r(k)$	The k^{th}column of the matrix R.
$r^a(j)$	The j^{th}raw of the R^a matrix
r_{jk}^a	And entry in the R^a matrix which gives the probability that observation k will be observed if the system has moved to state j and action a has been taken.
$\geq_{SD}$	First-order stochastic dominance partial order
$\geq_{MPR}$	Monotone probability ratio partial order
$\geq_{RHazard}$	Reverse hazard rate partial order
$\geq_{MLR}$	Monotone likelihood ratio partial order
$\geq_{MM}$	Marginal Monotonicity
$\geq_{CD}$	Component-wise dominance partial order with different forms like $\geq_{CD1}$ and $\geq_{CD2}$
$T(\pi, a, k)$	Posterior state occupancy probability vector
$\sigma(\pi, a)$	The probability vector of the system observations $\{m\}$ when action a is taken and the system is believed to be in π

$\sigma(k; \pi, a)$	The k^{th}component of $\sigma(\pi, a)$
β	Discount factor
$argmax_a \{V_t^*(\pi)\}$	The value of a which maximizes the quantity inside the brackets
$*.$	Element by element array multiplication

Next, a precise statement of the problem is presented.

5.2.2 STATEMENT OF THE PROBLEM

In this Dissertation, a multi-state system or machine is considered over a finite horizon. The states of the system range from 1 to n representing as good as new up to failed state respectively. The system states are assumed to be controllable states, that is, a decision maker can enhance the state of the system by means of a set of available control actions. The control actions are taken at discrete time epochs over the horizon considered. Control actions can be as simple as do nothing (a_0), minimal repair (a_i), or replacement of the whole system (a_n), which is assumed to renew the system. Actions effects differ based on the action type. If the system is left with no maintenance (a_0) it is assumed that the system will keep deteriorating stochastically. That is, starting from a certain state i it will be more likely that the system will deteriorate to a worse state, which is still better than the case of the system starting form another state $j \geq i$ (more deterioration level). This is expressed by the following partial order.

$$P^{a_0}(i) \geq_{CD2} P^{a_0}(j) \quad \forall\ i \geq j \quad in\ S$$

Different repair actions are assumed to improve the system state. For example, if maintenance action a is applied, when the system was actually in state i, then it will be more likely that the system will move to state $j \leq i$ (better state) than if action a' is applied where $a \geq a'$(a involves more maintenance activities than a'). This is expressed by the following partial order.

$$P^a(i) \geq_{CD2} P^{a'}(i) \quad \forall\, a \geq a' \;\; in\ A$$

Also, it is assumed that the system true state is not directly observable by the decision maker. Instead, only noise-corrupted information is received. This information is assumed to be probabilistically related to the true or actual system state. This is represented by a, possibly action-dependent, state observation transition matrix R^a with r_{jk}^a elements.

Since system states are partially observable, the decision maker is assumed to make his decisions based on the belief state. This is nothing but a state occupancy vector:

$$\pi = [\pi_1 \;\; \pi_2 \;\; \dots \;\; \pi_n\,]$$

Where, the number of elements in π equals the number of the system states. As will be illustrated later on, this state occupancy vector is updated at each time epoch whenever a decision has to be made. As in the literature, the updated state occupancy vector is usually referred to as the Bayesian update.

The reward criterion of the underlying system is assumed to be state/action dependent structure. That is, $g(i, a)$ reward will be received if action a is taken when the system true underlying state was i.

The objective function is to maximize the expected reward form the described system, by means of the proper selection among the different available maintenance actions at every time epoch for a given finite planning horizon.

The problem defined in this section is modeled mathematically, next, in Section 5.3.

5.3 MATHEMATICAL MODEL

In this section the machine maintenance problem for an n-state system is considered and modeled as a POMDP for the case of imperfect observation with multiple control actions available. The following two main assumptions are made to address the problem. First, the decision maker does not have any feedback regarding the true/actual state of the system other than noise-corrupted information. The information may include: output product quality, vibration or others. This is a common assumption in the literature. Second, there is a fixed control action for, at least one of the possible observations, which is selected, without loss of generality, to be the m^{th} observation. Also, and without loss of generality the corresponding action is assumed to be the replacement action. In many practical systems we may find such a situation where a certain signals call for some specified action to be taken; for example, a blue smoke from a car indicates that the car needs over-hall.

The objective function used in this model for action a taken at time $n-t$ with t time epochs left is expressed as follows:

$$V_t(\pi,a) = \sum_{i\in S}\pi_i \mathrm{g}(i,a) + \beta \times \{\sum_{k\in O-1}\sigma(k;\pi,a) \times V_{t-1}^{*}(T(\pi,a,k)) + \sigma(m;\pi,a) \times [\sum_{i\in S}\pi_i * \mathrm{g}\left(i,a_{replace}\right) + V_{t-2}^{*}(e_1)]\ \} \tag{5.1}$$

Where; $T_j(\pi,a,k) = \frac{\sum_i \pi_i p_{ij}^a r_{jk}^a}{\sum_i \sum_j \pi_i p_{ij}^a r_{jk}^a}$ is the j^{th} element of the Bayesian update.

Equation 5.1 means that after an action a is taken at t time epochs remaining, instantaneous reward is gained, which is function of both the current state and the action taken plus a discounted recursive part representing the future, replacement which takes one time epoch will be done next with probability $\sigma(m;\pi,a)$. Otherwise the best course of action will be taken for all other possible observations.

Recalling the assumption that replaced systems return as good as new, and noticing that $\sigma(k;\pi,a)$ is a probability vector whose components add up to one. Then, for the replace action (a_n), equation 1 becomes:

$$V_t(\pi,a_n) = \sum_{i\in S}\pi_i \mathrm{g}(i,a_n) + \beta \times V_{t-1}^*(e_1) \tag{5.2}$$

At any time stage, the optimal reward function is the maximum of both reward functions corresponding to both of the available control actions. Mathematically, this can be represented as follows

$$V_t^*(\pi) = \max_{\mathrm{i}}\ \{V_t(\pi,a_i)\} \tag{5.3}$$

With

$$a^* = argmax_a\ \{V_t^*(\pi)\} \tag{5.4}$$

Next, some theoretical results are provided for this POMDP model, where, it has been characterized over the space of state occupancy vectors ordered by the first order stochastic dominance.

5.4 THEORITICAL RESULTS

Before starting developing this chapter theoretical results, it is noteworthy to mention that unlike the case of the two-state model presented in Chapter 4, we will be dealing with multi-state systems ($n \geq 3$). As illustrated by Chapter 4 model, for n=2 it is easy to characterize the optimal value function and the corresponding optimal policies because the problem can be represented by one variable that is the probability of being in state 1 or even 2, hence the belief space reduces to a totally ordered one. In this chapter, we will be dealing with vectors instead of variables, where, in such characterizing the problem over partially ordered spaces becomes the choice. Next we restate Propositions 1 and 2 from the previous chapter since they will be referred to in this chapter as well.

Proposition 1

$\leq_{\mathrm{CD2}} \rightarrow \leq_{RHazard}$ and does not imply $\leq_{MLR}$ as Figure 1 shows.

Proof:

See proposition 1 proof in Chapter 4.

Proposition 2

$\leq_{\mathrm{CD1}} \rightarrow \leq_{SD}$ and does not imply $\leq_{MLR}$ as Figure 1 shows.

Proof:

See proposition 2 proof in Chapter 4.

The main results of this chapter are given in propositions 3 and 4 in this section, and the following of Lemmas are necessary to prove them.

Lemma 1 below is restated again from the previous chapter since it will be used extensively in the proofs of this chapter.

Lemma 1:

If X is a countable completely ordered set and π and $\pi'$are elements of $\Pi(S)$ then $\pi \geq_{\mathrm{SD}} \pi'$ if and only if $\sum_{i=1}^{n} \pi_i f(i) \geq \sum_{i=1}^{n} \pi_i' f(i)$ for every $f: X \rightarrow R$ non-increasing on X

Proof:

Refer to Stoyan (1983) for the proof.

Lemma 2:

For any a in A, if $P^a(i') \geq_{CD2} P^a(i) \ \forall\, i \geq i'$, then:

$$(\pi^T . P^a)^T \leq_{RHazard} ({\pi'}^T . P^a)^T, \quad \forall\, \pi \geq_{SD} \pi'$$

Proof:

Since $\pi \geq_{SD} \pi'$, and also, $P^a(i') \geq_{CD2} P^a(i)\ \ \forall\, i \geq i'$, then by Lemma 1: $(\pi^T.P^a)_j^T \leq (\pi'^T.P^a)_j^T$ for any $j = 2 \dots n$

And because $(\pi^T.P^a)^T$ elements must add up to one, then $(\pi^T.P^a)_1^T \geq (\pi'^T.P^a)_1^T$. This can be expressed as:

$(\pi^T.P^a)^T \geq_{CD2} ({\pi'}^T.P^a)^T$

Then, since $\leq_{\mathrm{CD2}} \;\rightarrow\; \leq_{RHazard}$, by Proposition 1, the result follows.

Lemma 3:

If $P^a(i) \geq_{CD2} P^{a'}(i), \forall\, i\ \in S$ and $P^a(i') \geq_{CD2} P^a(i),\ \forall\, i \geq i'$, then:

$(\pi^T.P^a)^T \leq_{RHazard} (\pi^T.P^{a\prime})^T$ for any $a \geq a'$ in A

Proof:

This result follows in a similar way as in Lemma 2 proof. That is:

$$(\pi^T.P^a)^T \geq_{\mathrm{CD2}} (\pi^T.P^{a\prime})^T,\ \ \forall\, a \geq a'$$

And since $\leq_{\mathrm{CD2}}$ implies $\leq_{\mathrm{Rhazard}}$ (by Proposition 1), then:

$(\pi^T.P^a)^T \leq_{\mathrm{RHazard}} (\pi^T.P^{a\prime})^T, \forall\, a \geq a'$

Hence, the result follows.

Lemma 4:

if $\mathrm{g}(i, a)$ is non-increasing in $i \in S$, then:

$V_1^*(\pi, a)$ is non-decreasing over the state occupancy vectors $\Pi(S)$ ordered by $\leq_{SD}$

Proof

The proof is a straight forward application of Lemma 1.

Lemma 5

$T(\pi, a, k)$ is $\leq_{RHazard}$ non-decreasing in $k \in O-1$, for any $\pi \in \Pi(S)$, for any a in A if and only if:

$$(\pi^T.P^a)^T.*\, r(k') \leq_{RHazard} (\pi^T.P^a)^T.*\, r(k) \qquad \forall\, k \geq k'$$

Where:

.*: Array multiplication (element by element), for an illustration:

Let $a = \begin{bmatrix} a1 \\ a2 \\ . \\ . \\ an \end{bmatrix}$ and let $b = \begin{bmatrix} b1 \\ b2 \\ . \\ . \\ bn \end{bmatrix}$

Then, $a.*\, b = \begin{bmatrix} a1b1 \\ a2b2 \\ . \\ . \\ anbn \end{bmatrix}$

. : Dot product

$a.b = a1b1 + a2b2 + \cdots + anbn$

Hence,

$r(k)$: is the K^{th} column of the matrix R.

$$(\pi^T.P)^T = \begin{bmatrix} \sum_i^n \pi_i p_{i1} \\ \sum_i^n \pi_i p_{i2} \\ . \\ . \\ \sum_i^n \pi_i p_{in} \end{bmatrix}$$

$(\pi^T.P)_j^T$: Is the j^{th} element of $(\pi^T.P)^T$

Proof:

Writing the $\leq_{RHazard}$ expression explicitly for the Bayesian update:

For any $k \geq k', k\ and\ k' \in O-1$.

$$\frac{\frac{\sum_i^n \pi_i p_{il} r_{lk\prime}}{\sigma(k\prime;\pi,a)}}{\frac{\sum_i^n \pi_i p_{i1} r_{1k\prime}}{\sigma(k\prime;\pi,a)}+\frac{\sum_i^n \pi_i p_{i2} r_{2k\prime}}{\sigma(k\prime;\pi,a)}+\cdots\frac{\sum_i^n \pi_i p_{il} r_{lk\prime}}{\sigma(k\prime;\pi,a)}}$$

$$\leq \frac{\frac{\sum_i^n \pi_i p_{il} r_{lk}}{\sigma(k;\pi,a)}}{\frac{\sum_i^n \pi_i p_{i1} r_{1k}}{\sigma(k;\pi,a)}+\frac{\sum_i^n \pi_i p_{i2} r_{2k}}{\sigma(k;\pi,a)}+\cdots\frac{\sum_i^n \pi_i p_{il} r_{lk}}{\sigma(k;\pi,a)}}$$

This inequality should be true $\forall\ l \in\ S$

Now, multiplying the LHS of the inequality by $\frac{\sigma(k\prime;\pi,a)}{\sigma(k\prime;\pi,a)}$ and the right hand side by $\frac{\sigma(k;\pi,a)}{\sigma(k;\pi,a)}$ gives:

$$\frac{\sum_i^n \pi_i p_{il} r_{lk\prime}}{\sum_i^n \pi_i p_{i1} r_{1k\prime}+\sum_i^n \pi_i p_{i2} r_{2k\prime}+\cdots\sum_i^n \pi_i p_{il} r_{lk\prime}}$$

$$\leq \frac{\sum_i^n \pi_i p_{il} r_{lk}}{\sum_i^n \pi_i p_{i1} r_{1k}+\sum_i^n \pi_i p_{i2} r_{2k}+\cdots\sum_i^n \pi_i p_{il} r_{lk}}$$

This can be written in vector format as:

$$(\pi^T.P^a)^T.*\,r(k') \leq_{RHazard} (\pi^T.P^a)^T.*\,r(k) \qquad \forall\ k \geq k'$$

Similarly, if

$$(\pi^T.P^a)^T.*\,r(k') \leq_{RHazard} (\pi^T.P^a)^T.*\,r(k) \qquad \forall\ k \geq k'$$

is true, then again writing this inequality explicitly and multiplying the LHS of the inequality by $\frac{\sigma(k\prime;\pi,a)}{\sigma(k\prime;\pi,a)}$ and the right hand side by $\frac{\sigma(k;\pi,a)}{\sigma(k;\pi,a)}$ will give $T(\pi,a,k)$ is $\leq_{RHazard}$ non-decreasing in $k \in O-1$, for any $\pi \in \Pi(S)$, for any a in A.

Next, a sufficient condition for Lemma 5 condition is provided.

Lemma 6:

If $r(k') \geq_{MLR} r(k)\ \forall\ k' < k$ in $O-1$, then

$(\pi^T . P^a)^T .* r(k') \leq_{RHazard} (\pi^T . P^a)^T .* r(k) \qquad \forall\, k \geq k'$

Proof:

Let $(\pi^T . P)^T = a, r(k') = b, and\ r(k) = c$, then:

Writing the expression for Lemma 5 explicitly:

$$\frac{a_l b_l}{a_1 b_1 + a_2 b_2 + \cdots + a_l b_l} \leq \frac{a_l c_l}{a_1 c_1 + a_2 c_2 + \cdots + a_l c_l}$$

Inverting the inequality:

$$\frac{a_1 b_1}{a_l b_l} + \frac{a_2 b_2}{a_l b_l} + \cdots + \frac{a_{l-1} b_{l-1}}{a_l b_l} \geq \frac{a_1 c_1}{a_l c_l} + \frac{a_2 c_2}{a_l c_l} + \cdots + \frac{a_{l-1} c_{l-1}}{a_l c_l}$$

Having $\frac{b_1}{b_l} \geq \frac{c_1}{c_l}$, $\frac{b_2}{b_l} \geq \frac{c_2}{c_l}$… for all $l \in S$ implies the result of this Lemma.

Lemma 7:

$T(\pi, a, k)$ is $\geq_{RHazard}$ non-increasing in $\pi \in \Pi(S)$ $(\geq_{RHazard})$ for any k in $O - 1$ for any a in A, if and only if:

$(\pi^T . P)^T .* r(k) \leq_{RHazard} (\pi'^T . P)^T .* r(k) \qquad \forall\, \pi \geq_{SD} \pi'$

Proof:

Writing the $\leq_{RHazard}$ expression explicitly for the Bayesian update:

For any $\pi \geq_{SD} \pi'$ in $\Pi(S)$

$$\frac{\frac{\sum_i^n \pi_i p_{il} r_{lk}}{\sigma(k;\pi,a)}}{\frac{\sum_i^n \pi_i p_{i1} r_{1k}}{\sigma(k;\pi,a)} + \frac{\sum_i^n \pi_i p_{i2} r_{2k}}{\sigma(k;\pi,a)} + \cdots \frac{\sum_i^n \pi_i p_{il} r_{lk}}{\sigma(k;\pi,a)}}$$

$$\leq \frac{\frac{\sum_i^n \pi'_i p_{il} r_{lk}}{\sigma(k;\pi\prime,a)}}{\frac{\sum_i^n \pi'_i p_{i1} r_{1k}}{\sigma(k;\pi\prime,a)} + \frac{\sum_i^n \pi'_i p_{i2} r_{2k}}{\sigma(k;\pi\prime,a)} + \cdots \frac{\sum_i^n \pi'_i p_{il} r_{lk}}{\sigma(k;\pi\prime,a)}}$$

Now, multiplying LHS sides of the inequality above by $\frac{\sigma(k;\pi,a)}{\sigma(k;\pi,a)}$ and the RHS by $\frac{\sigma(k;\pi\prime,a)}{\sigma(k;\pi\prime,a)}$ gives:

$$\frac{\sum_i^n \pi_i p_{il} r_{lk}}{\sum_i^n \pi_i p_{i1} r_{1k} + \sum_i^n \pi_i p_{i2} r_{2k} + \cdots \sum_i^n \pi_i p_{il} r_{lk}}$$

$$\leq \frac{\sum_i^n \pi_i' p_{il} r_{lk}}{\sum_i^n \pi_i' p_{i1} r_{1k} + \sum_i^n \pi_i' p_{i2} r_{2k} + \cdots \sum_i^n \pi_i' p_{il} r_{lk}}$$

This can be written, in vector format, as follows:

$$(\pi^T . P)^T .* \, r(k) \leq_{RHazard} (\pi'^T . P)^T .* \, r(k) \qquad \forall \, \pi \geq_{SD} \pi'$$

Similarly, if

$$(\pi^T . P)^T .* \, r(k) \leq_{RHazard} (\pi'^T . P)^T .* \, r(k) \qquad \forall \, \pi \geq_{SD} \pi'$$

is true, then again writing this inequality explicitly and multiplying the LHS sides of the inequality above by $\frac{\sigma(k;\pi,a)}{\sigma(k;\pi,a)}$ and the RHS by $\frac{\sigma(k;\pi\prime,a)}{\sigma(k;\pi\prime,a)}$will give $T(\pi, a, k)$ is $\geq_{RHazard}$ non-increasing in $\pi \in \Pi(S)$ $(\geq_{RHazard})$ for any k in $O - 1$ for any a in A

Next, a sufficient condition for Lemma 7 condition is provided.

Lemma 8:

if $P^a(i') \geq_{CD2} P^a(i) \qquad \forall \; i' \leq i$

and if $r(j') \geq_{CD1} r(j) \qquad \forall \; j' < j \quad in \; S$

Then:

$$(\pi^T . P)^T .* \, r(k) \leq_{RHazard} (\pi'^T . P)^T .* \, r(k) \qquad \forall \, \pi \geq_{SD} \pi'$$

Proof:

Since $P^a(i') \geq_{CD2} P^a(i) \qquad \forall \, i \geq i'$

This implies: $(\pi^T . P)^T \geq_{CD2} ({\pi'}^T . P)^T \qquad \forall \, \pi \leq_{SD} \pi'$

As shown by Lemma 2.

Let $(\pi^T . P)^T = a, ({\pi'}^T . P)^T = b, and \; r(k) = c$, then:

Writing the expression for Lemma 7 explicitly we will have:

$$\frac{a_l c_l}{a_1 c_1 + a_2 c_2 + \cdots + a_l c_l} \leq \frac{b_l c_l}{b_1 c_1 + b_2 c_2 + \cdots + b_l c_l}$$

For all $l = 2 \ldots n, a_l \leq b_l$ since $a \geq_{CD2} b$ then for the numerator $a_l \leq b_l$. For the denominator, since $a \geq_{SD} b$ is implied by $a \geq_{CD2} b$ having c_l non-increasing as l increases and applying Lemma 1. will lead to the denominator of the left hand side always greater than that of the right hand side. Hence, the result follows.

Lemma 9:

$T(\pi, a, k)$ is $\geq_{RHazard}$ non-increasing on $a\ in\ A$ for any k $\in O - 1$ and for any $\pi \in \Pi(S)$, if and only if:

$$(\pi^T.P^a)^T.*\, r^a(k) \leq_{RHazard} (\pi^T.P^{a'})^T.*\, r^{a'}(k) \qquad \forall\, a \geq\ a'$$

Proof:

Writing the $\leq_{RHazard}$ expression explicitly for the Bayesian update:

For any $a \geq\ a'$ in A

$$\frac{\frac{\sum_i^n \pi_i p_{il}{}^a r_{lk}{}^a}{\sigma(k;\pi,a)}}{\frac{\sum_i^n \pi_i p_{i1}{}^a r_{1k}{}^a}{\sigma(k;\pi,a)}+\frac{\sum_i^n \pi_i p_{i2}{}^a r_{2k}{}^a}{\sigma(k;\pi,a)}+\cdots\frac{\sum_i^n \pi_i p_{il}{}^a r_{lk}{}^a}{\sigma(k;\pi,a)}}$$

$$\leq \frac{\frac{\sum_i^n \pi_i p_{il}{}^{a\prime} r_{lk}{}^{a\prime}}{\sigma(k;\pi,a\prime)}}{\frac{\sum_i^n \pi_i p_{i1}{}^{a\prime} r_{1k}{}^{a\prime}}{\sigma(k;\pi,a\prime)}+\frac{\sum_i^n \pi_i p_{i2}{}^{a\prime} r_{2k}{}^{a\prime}}{\sigma(k;\pi,a\prime)}+\cdots\frac{\sum_i^n \pi_i p_{il}{}^{a\prime} r_{lk}{}^{a\prime}}{\sigma(k;\pi,a\prime)}}$$

Now, multiplying LHS sides of the inequality above by $\frac{\sigma(k;\pi,a)}{\sigma(k;\pi,a)}$ and the RHS by $\frac{\sigma(k;\pi,a')}{\sigma(k;\pi,a')}$ gives:

$$\frac{\sum_i^n \pi_i p_{il}{}^a r_{lk}{}^a}{\sum_i^n \pi_i p_{i1}{}^a r_{1k}{}^a+\sum_i^n \pi_i p_{i2}{}^a r_{2k}{}^a+\cdots\sum_i^n \pi_i p_{il}{}^a r_{lk}{}^a}$$

$$\leq \frac{\sum_i^n \pi_i p_{il}{}^{a\prime} r_{lk}{}^{a\prime}}{\sum_i^n \pi_i p_{i1}{}^a r_{1k}{}^a+\sum_i^n \pi_i p_{i2}{}^{a\prime} r_{2k}{}^{a\prime}+\cdots\sum_i^n \pi_i p_{il}{}^{a\prime} r_{lk}{}^{a\prime}}$$

This can be written, in vector format, as follows:

$$(\pi^T.P^a)^T.*\, r^a(k) \leq_{RHazard} (\pi^T.P^{a\prime})^T.*\, r^{a\prime}(k) \qquad \forall\, a \geq a'$$

Similarly, if

$$(\pi^T.P^a)^T.*\, r^a(k) \leq_{RHazard} (\pi^T.P^{a\prime})^T.*\, r^{a\prime}(k) \qquad \forall\, a \geq a'$$

is true, then again writing this inequality explicitly and multiplying the LHS sides of the inequality above by $\frac{\sigma(k;\pi,a)}{\sigma(k;\pi,a)}$ and the RHS by $\frac{\sigma(k;\pi,a')}{\sigma(k;\pi,a')}$ will give $T(\pi, a, k)$ is $\geq_{RHazard}$ non-increasing on $a \; in \; A$ for any k $\in O - 1$ and for any $\pi \in \Pi(S)$

Next, a sufficient condition for Lemma 9 condition is provided

Lemma 10:

If

1. $P^a(i') \geq_{CD2} P^a(i) \qquad \forall \; i' \leq i \; in \; S, \forall a \; in \; A$
2. $r^a(j') \geq_{CD1} r^a(j) \qquad \forall \quad j' \leq j \; in \; S, \forall a \; in \; A$
3. $r^a(j) \geq_{CD1} r^{a'}(j) \qquad \forall \quad j \; in \; S, \forall a \geq a' \; in \; A$

Then:

$$(\pi^T . P^a)^T .* \, r^a(k) \leq_{RHazard} (\pi^T . P^{a'})^T .* \, r^{a'}(k) \qquad \forall \, a \geq \, a'$$

Proof:

Utilizing the result of Lemma 3, the proof is a straight forward extension of Lemma 8 proof.

Lemma 11:

If $P^a(i') \geq_{CD2} P^a(i), \; \forall \, i \geq i' \; in \; S$, and $r(j') \geq_{CD1} r(j), \forall \, j' < j \; in \; S$, then:

$\sigma(\pi, a) \geq_{SD} \sigma(\pi', a)$ for any $\pi \geq \pi'$

Proof:

For any l 1, 2 …n and $a \in A$

$\sum_{k \leq l} r^a_{jk}$ is non-increasing as j increases in S by since $r(j') \geq_{CD1} r(j)$ implies $r(j') \geq_{SD} r(j)$

Also, since $P^a(i') \geq_{CD2} P^a(i)$ implies $P^a(i') \geq_{SD} P^a(i) \quad \forall \, i' < i$ then:

$\sum_{j\in S} P_{ij}^a \sum_{k\le l} r_{jk}^a$ is non-increasing as $i \in S$ increases by Lemma 1.

Consequently, again by applying Lemma 1, $\sigma(\pi, a) \ge_{SD} \sigma(\pi', a)$ for $\pi \ge_{SD} \pi'$.

Proposition 3:

if

1. $g(i, a)$ is non-decreasing as i decreases $\forall a \in A$.
2. $r(k') \ge_{MLR} r(k) \qquad \forall\ k' < k$ in O
3. $r(j') \ge_{CD1} r(j) \qquad \forall\ j' < j \quad in\ S$
4. $P^a(i') \ge_{CD2} P^a(i) \qquad \forall\ i' \le i\ \ \forall a$

Then $V_t^*(\pi, a)$, as defined in equation 1, is a non-decreasing function over the space of state occupancy vectors ordered by $\ge_{SD}$

Proof:

Inductively, by Lemma 4. The result is true for t=1. Assuming that V_{t-1}^* is non-decreasing in $\Pi(s)$ ordered by $\ge_{SD}$

$$V_t^*(\pi) \ge \sum_{i\in S} \pi_i g(i, a') + \beta \times \{\sum_{k\in O-1} \sigma(k; \pi, a') \times V_{t-1}^*(T(\pi, a', k)) +$$

$$\sigma(m; \pi, a') \times \left[\sum_{i\in S} \pi_i g\left(i, a_{replace}\right) + V_{t-2}^*(e_1)\right] \} \quad \ge$$

$$\sum_{\in S} \pi_i g(i, a') + \beta \times \{\sum_{k\in O-1} \sigma(k; \pi', a') \times V_{t-1}^*(T(\pi, a', k)) + \sigma(m; \pi', a') \times$$

$$\left[\sum_{i\in S} \pi_i g\left(i, a_{replace}\right) + V_{t-2}^*(e_1)\right] \}$$

By Lemma 1, Lemma 6 and Lemma 11:

$$\ge \sum_{i\in S} \pi_i g(i, a') + \beta \times \{\sum_{k\in O-1} \sigma(k; \pi', a') \times V_{t-1}^*(T(\pi', a', k)) + \sigma(m; \pi', a') \times$$

$$\left[\sum_{i\in S} \pi_i g\left(i, a_{replace}\right) + V_{t-2}^*(e_1)\right] \} = V_t^*(\pi')$$

By Lemma 7, and because $\le_{\mathrm{RHazard}} \rightarrow \le_{SD}$ for the Bayesian update. Finally,

Noticing that: $V_{t-1}^*(T(\pi, a', k)) \ge \sum_{i\in S} \pi_i * g\left(i, a_{replace}\right) + V_{t-2}^*(e_1)$, applying Lemma 1 again the result follows.

Note:

$\mathrm{g}(i, a_{replace})$ is non-increasing as i increases. This may represent the salvage value of replaced systems. Where, the less deteriorated the system the higher salvage value it has

Corollary 1

Similar to Corollary 1 in Lovejoy, the columns of the observation matrix may be permutated with no further restrictions. This gives a very reasonable applicability of our model where at least one observation can be found with fixed action

Corollary 2

For the case of k=2, there is no need to assume condition 3 in Proposition 3; since it will follow automatically by assumption 2.

Proposition 4:

If $V_t^*(\pi, a)$ has isotone differences on $\pi \times a \to R$ for $a \in \{a_0 = do\ nothing, a_1 = replace\}$, then there exist an optimal threshold policy with two regions one for a_0 and a_1 over $\Pi(S)$ ordered by $\geq_{SD}$. This is a similar result to those in White (1979, a) and Lovejoy (1987) subject to our new developed conditions.

Proof:

Assuming $\mathrm{g}(i, a_0) - \mathrm{g}(i, a_1)$ is non-increasing as i increases in S.

$V_t^*(\pi, a_0) - V_t^*(\pi, a_1) = \sum_{i \in S} \pi_i(\mathrm{g}(i, a_0) - \mathrm{g}(i, a_1)) + \beta \times \{\sum_{k \in O-1} \sigma(k; \pi, a_0) \times$

$V_{t-1}^*(T(\pi, a_0, k)) + \sigma(m; \pi, a_0) \times [\sum_{i \in S} \pi_i * \mathrm{g}(i, a_1) + V_{t-2}^*(e_1)]\} - \beta \times V_{t-1}^*(e_1)$

is non-increasing as i increases.

From the proof of proposition 3 the following quantity is non-increasing in i on S.

$$\beta \times \{\sum_{k \in O-1} \sigma(k; \pi, a_0) \times V_{t-1}^*(T(\pi, a_0, k)) + \sigma(m; \pi, a_0) \times [\sum_{i \in S} \pi_i * \mathrm{g}(i, a_1) +$$

$V^*_{t-2}(e_1)] \}$

Also since $V^*_{t-1}(e_1)$ not a function in π, the result follows.

Lemma 12:

Assuming:

1. $P^a \geq_{CD2} P^{a'}$ $\quad\quad \forall\, a \geq a'\ in\ A$
2. $R^a \geq_{CD1} R^{a'}$ $\quad\quad \forall j \in S$
3. $r^a(j') \geq_{CD1} r^a(j)$ $\quad\quad \forall j \geq j'\ in\ S$

It can be shown that $\sigma(\pi, a) \geq_{SD} \sigma(\pi, a')$ for $a \geq a'$

Proof:

The proof of this Lemma follows in a similar way as in Lemma 11 proof.

The following proposition provides lower bound on the optimal policy of the POMDP problem considered. Let:

$\alpha(\pi) = argmax_{a \in A}\{\sum_{i \epsilon S} \pi_i \mathrm{g}(i, a) : a \in A\} \quad\quad \pi \in \Pi(S).$

Lemma 13 (Lovejoy (1987)):

let V and V' be two real-valued functions on A, and let $A^*$and A'^* be the sets of actions that maximize V and V', respectively, on A. if for every $a \geq a'$, in A.

$$V(\pi, a) - V(\pi, a') \geq V'(\pi, a) - V'(\pi, a')$$

Then for every $a^* \in A$ there exists an $a'^* \in A'^*$ there exists an $a^* \in A^*$ with $a^* \in a'^*$.

Proposition 5

The following conditions, developed earlier, are sufficient for $\alpha(\pi)$ to provide lower bound on the optimal policy for the POMDP. With the problem approximated with one horizon problem.

1. $g(i, a)$ is non-increasing on i in S $\forall$ $a \in A$
2. $r(k') \geq_{MLR} r(k)$ $\quad \forall$ $k' \leq k$ in $O-1$
3. $r(j') \geq_{CD1} r(j)$ $\quad \forall$ $j' \leq j$ $in\ S$
4. $P^a(i') \geq_{CD2} P^a(i)$ $\quad \forall$ $i \geq i'$ $in\ S$
5. $P^a \geq_{CD2} P^{a'}$ $\quad \forall$ $a \geq a'$ $in\ A$
6. $R^a \geq_{CD1} R^{a'}$

Proof:

Hence for any $\pi \in \Pi(S)$, $a \geq a'$ in A, and $t = 1 \ldots n$

$$\left\{\sum_{k \in O-1} \sigma(k; \pi, a) \times V^*_{t-1}(T(\pi, a, k)) + \sigma(m; \pi, a) \times \left[\sum_{i \in S} \pi_i * \mathrm{g}\left(i, a_{replace}\right) + V^*_{t-2}(e_1)\right]\right\} \geq$$

$$\left\{\sum_{k \in O-1} \sigma(k; \pi, a') \times V^*_{t-1}(T(\pi, a, k)) + \sigma(m; \pi, a') \times \left[\sum_{i \in S} \pi_i * \mathrm{g}\left(i, a_{replace}\right) + V^*_{t-2}(e_1)\right]\right\}$$

By applying Lemma 1, noticing:

Condition 2 implies $T(\pi, a, k) \geq_{RHazard}$ increasing on k in $O-1$,which implies $\geq_{SD}$

Conditions 1- 4 imply V^*_t is non-decreasing on the belief space ordered by $\geq_{SD}$ as in Proposition 3.

Conditions 3, 5, and 6 imply the result of Lemma 12.

$$\geq \left\{\sum_{k \in O-1} \sigma(k; \pi, a') \times V^*_{t-1}(T(\pi, a', k)) + \sigma(m; \pi, a') \times \left[\sum_{i \in S} \pi_i * \mathrm{g}\left(i, a_{replace}\right) + V^*_{t-2}(e_1)\right]\right\}$$

By conditions 3, 5, and 6 which implies $T(\pi, a, k) \geq_{SD}$ increasing on a in A as by Lemma 10. And applying the result of Lemma 1 again.

Thus,

$h(\pi, a, V_{t-1}^*) - h(\pi, a', V_{t-1}^*) \geq \sum_{i \in S} \pi_i g(i, a) - \sum_{i \in S} \pi_i g(i, a')$

hence, the result follows from Lemma 13.

5.5 NUMERICAL EXAMPLES

This section provides two examples to demonstrate the results developed in this chapter.

Example 5.1

In Lovejoy (1987) an example has been presented to demonstrate the results in that paper. We consider the same example parameters, where, Lovejoy's conditions are satisfied and our conditions violated. We demonstrate that under Lovejoy's conditions with change in the partial order to $\geq_{RHazard}$, the optimal reward function is not monotone.

Consider two actions with Identity matrix as the transition matrix for both. With the following reward criteria and observation matrix:

$g(a_0) = [2 \;\; 0 \;\; 0]$

$g(a_1) = [1 \;\; 1 \;\; 1]$

$\pi \; = [0.0143 \quad 0.9857 \qquad 0]$

$\pi' = [0.0025 \quad 0.2406 \quad 0.7569]$

$$R = \begin{bmatrix} 0.001 & 0.7 & 0.299 \\ 0.1 & 0.8 & 0.1 \\ 0.299 & 0.7 & 0.001 \end{bmatrix}$$

For two horizons the optimal reward, using equation 5.1, will be:

$V_2^*(\pi) = 1 + 0.0986 \times V_1^*[\, 0.0001 \quad 0.9999 \quad 0] + 0.7986 \times$

$V_1^*[\, 0.0125 \quad 0.9875 \quad 0] + 0.1028 \times V_1^*[\, 0.0416 \quad 0.9584 \quad 0] = 1.9057$

$V_2^*(\pi') =$

$1 + 0.2504 \times V_1^*[\ 0.0000 \quad 0.0961 \quad 0.9039] + 0.7241 \times$

$V_1^*[\ \ 0.0024 \quad 0.2658 \quad 0.7317] + 0.0256 \times V_1^*[\ \ 0.0294 \quad 0.9410 \quad 0.0296] = 1.9759$

Notice that $\pi \leq_{RHazard} \pi'$ whereas $V_2^*(\pi) \leq V_2^*(\pi')$, hence the optimal value function is not monotone over the $\leq_{RHazard}$ partial order with Lovejoy's conditions satisfied.

Example 5.2

In this example Value Iteration algorithm is used to provide the solutions numerically for Equation 5.4 with the assumptions of propositions 3 and 4 holding true. The solutions are provided over a discretized belief space of the state occupancy vectors. To correct for the fact that the actions don't map back to the discretized belief space, bilinear interpolation is used. With three control actions available, namely: do nothing, repair, and replace $(a_0, a_1,\ and\ a_2)$ respectively, consider the following reward, transition, and observation data for a three-state machine maintenance POMDP model.

$g(a_0) = [4\ \ 2\ \ 0]$

$g(a_1) = [3\ \ 2\ \ 0]$

$g(a_2) = [2\ \ 2\ \ 2]$

$$P^{a_0} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0 & 0.6 & 0.4 \end{bmatrix}, P^{a_1} = \begin{bmatrix} 0.8 & 0.2 & 0.0 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.6 & 0.3 \end{bmatrix}, P^{a_2} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

The observations matrix is assumed to be action independent:

$$R = \begin{bmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \\ 0.3 & 0.7 \end{bmatrix}$$

For this three state system example, notice that the collection of all possible (π_1, π_2) points (the belief space) are represented by the simplex $\{0 \leq \pi_1 \leq 1, 0 \leq \pi_2 \leq 1, \pi_1 +$

$\pi_2 \leq 1\}$. The objective is to find action which maximizes the expected total reward over some finite planning horizon. Figures 5.1-5.5 provide optimal policy regions for different numbers of time epochs left from a given planning horizon. for example, assuming that time epochs are weeks and the planning horizon is 52 weeks, then Figure 5.1 provides the optimal maintenance policy regions for the 52 week. Notice that the replace action has an at most one region.

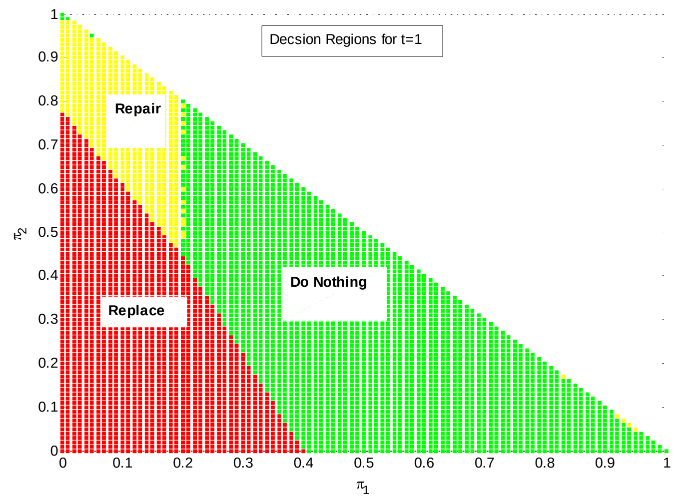


Figure 5.1 Optimal policy regions for t=1

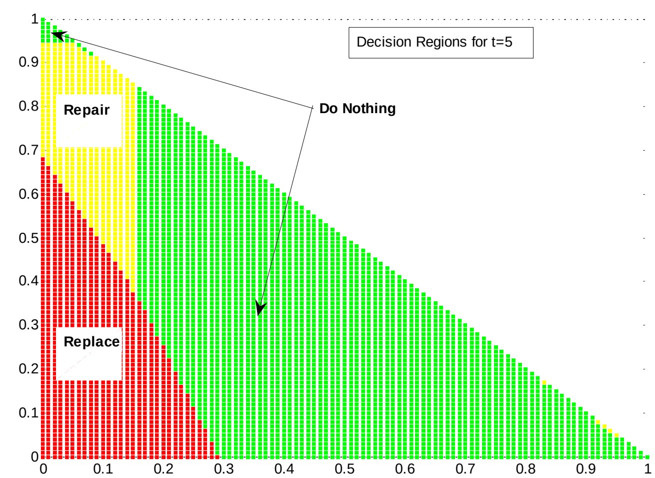


Figure 5.2 Optimal policy regions for t=5

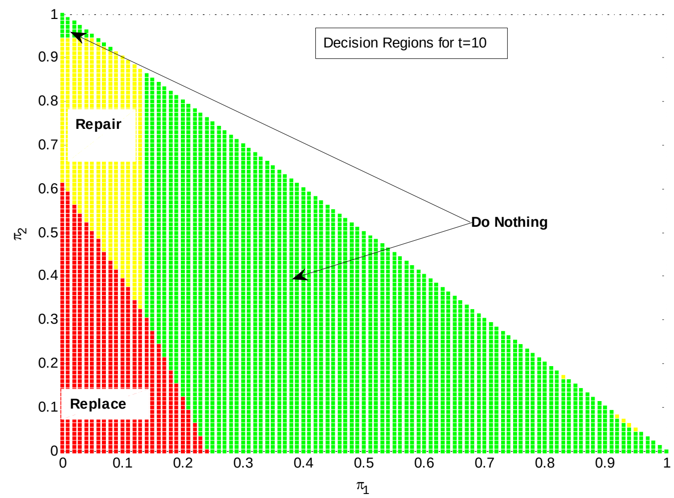


Figure 5.3 Optimal policy regions for t=10

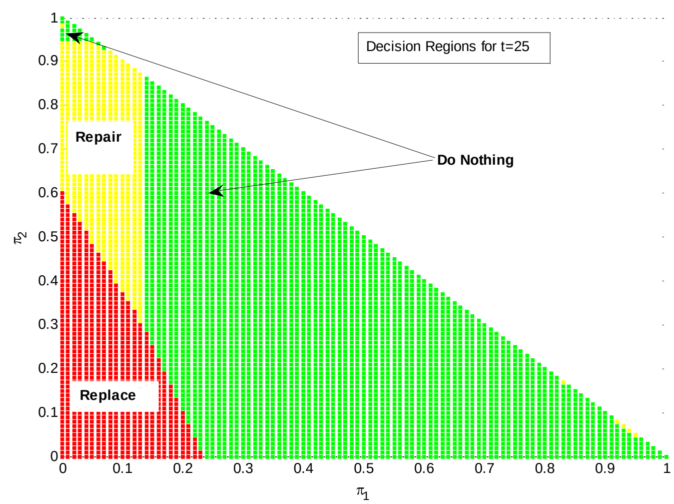


Figure 5.4 Optimal policy regions for t=26

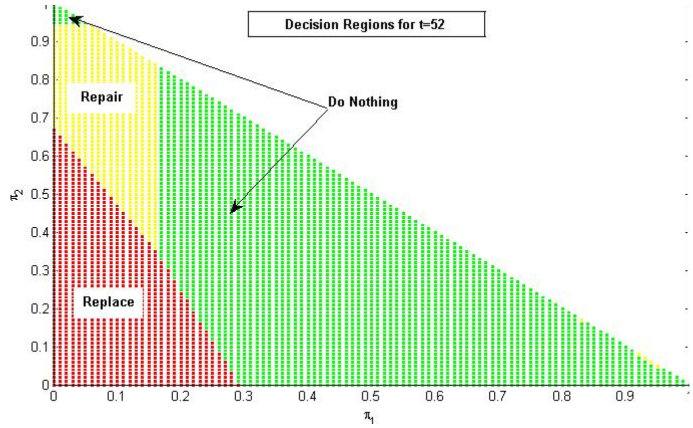


Figure 5.5 Optimal policy regions for t=52

5.6 CONCLUSION

In this chapter, the problem of determining the optimal maintenance for a multi-state system with an increasing failure rate is considered. The problem is formulated as a POMDP. New sufficient conditions which guarantee the existence of optimal threshold-policy over the space of state occupancy vectors ordered by first order stochastic dominance are developed. The main advantage of the developed conditions is ensuring the first order stochastic dominance to survive conditioning. This is achieved by developing new relations with other useful partial orders which were not considered for this problem before. This is supposed to give one more alternative for applications that requires these types of structured policies. Hence, an alternative set of conditions, which can be applied subject to a tradeoff with other conditions existing in the literature, is provided.

CHAPTER 6

MAXIMIZING OVERALL SYSTEMS EFFECTIVENESS (OSE) FOR A PARTIALLY OBSERVED MARKOV DECISION PROCESS (POMDP)

6.1 INTRODUCTION

In this chapter the problem of determining the optimal maintenance policy for a multi-state, multi-stage machine maintenance problem is considered. This problem has been formulated in the literature as a Partially Observed Markov Decision Process (POMDP) over a finite horizon. A new formulation that ties maintenance and operation explicitly within the POMDP framework is provided. The new formulation maximizes Overall Systems Effectiveness (OSE) for an n-state system with multiple speeds and maintenance actions. The model selects an optimal maintenance policy with an operation level indicated by the production speed level. The application of the model and the effect of measurement error on the optimal maintenance and operation policies are demonstrated by numerical examples.

The deterioration mechanism of the different systems depends on many factors that include maintenance actions and system operation reflected by the speed level. Systems deterioration may lead to unavailability, speed losses and reduction in the quality of the output. Maintenance actions take place to control systems deterioration and bring back the systems to a better operating state. Excessive maintenance may result in high costs and not performing enough maintenance will cause a system to drift and operate in an undesirable state, where system effectiveness is low. For example, consider an ammonia production plant, where, there are usually many pumps and compressors involved in the production process. Scheduled maintenance and inspections of these equipments are usually performed periodically, where; the production process must be stopped. On the other hand, if such inspections and maintenance operations are not performed a pump or a compressor failure might cause a sudden failure. Such failures might need, some times, few weeks to have the process back to its full production capacity (due to spare parts unavailability, maintenance labor skill issues or other factors). In some instances, operation requires higher level of production to meet demand. In such cases, the system will be operated in a higher speed. This causes the system to deteriorate faster. System deterioration increases the percentage of products off specification, or in other words, reduces quality rates. It can be easily seen from this ammonia plant example that a tradeoff exists between maintenance actions, operation level as reflected by speed, deterioration level and output quality. The problem of determining optimal maintenance action and operation levels is a complex decision making problem.

The motivation for the work in this chapter stems from the need to develop an optimal realistic maintenance policies for complex deteriorating systems that, explicitly, ties maintenance, operation and quality. This is achieved by extending the concept of Overall Equipment Effectiveness (OEE) to systems and using it as an objective function in a POMDP framework. The concept of OEE was developed in Total Productive Maintenance (TPM). TPM is a Japanese philosophy for maintenance. According to Nakajima (1988), OEE is defined as:

$$OEE = Availability \times Process\ Rate \times Quality\ Rate \qquad (6.1)$$

The literature review in Chapter 3 indicates that there is a need for a mathematical model that integrates maintenance and operation in POMDP framework. Also, a need exists to address the impact of measurement errors. In this chapter, this model will be extended to reflect the components of OSE on the POMDP framework in an explicit way for an n-states system. The developed model maximizes the expected OSE.

The rest of this chapter is organized as follows: the nomenclature used to develop the n-state model to maximize OSE and a precise statement for the problem are provided in Section 6.2. Then, a mathematical model to maximize OSE is presented in Section 6.3. Section 6.4 provides an illustrative example for the application of the model. Section 6.5 concludes the chapter with some remarks.

6.2 NOMENCLATURE AND PRECISE STATEMENT OF THE PROBLEM

In this section the Nomenclature to be used in this chapter models is presented in Subsection 6.2.1. Then, a precise statement of the problem is provided in Subsection 6.2.2.

6.2.1 NOMENCLATURE

Follows is the nomenclature that will be used throughout this chapter

S	System's state set {1,2,…n}
i, j, k	Elements of S
q_r	Quality rate of the system or percent of products within specifications
O	System signals set with components {1,2, … m}
x	Quality level output, element of O (a binary random variable)
A	The set of maintenance actions available to the decision maker
a_i	Element of$A$, that can range from a_0 (do nothing) to a_n (replace). a (without an index) represents an arbitrary maintenance action
$A(a\alpha)$	Availability indices vector
$A(i, a\alpha)$	An element of $A(a\alpha)$ corresponding to action a and system state i, $i \in S$

$g(a\alpha)$	OSE revenue vector
α	The set of production speed actions available to the decision maker
α_i	Element of α
P^a	An $(n \times n)$ system state transition matrix corresponding to maintenance action a
P^a_{ij}	Probability that the system will move to state j if action a was taken, given the system was in state i
P^α	An $(n \times n)$ system state transition matrix if the system is operated at speed level α
P^α_{jk}	Probability that the system will move to state k if speed action α was taken, given the system was in state j
$P^{a\alpha}$	An $(n \times n)$ system state transition matrix = $P^a \times P^\alpha$
$P^{a\alpha}_{ik}$	Probability that the system will move to state k given it was in i if actions a and α were taken
R	An $(n \times m)$ state observation transition matrix
r_{kx}	And entry in the R matrix which gives the probability that observation x will be observed if the system is in state k
$r(1)$	The column in the R matrix corresponding to the within-specifications process output
$T(\pi, a, \alpha, x)$	Posterior state occupancy probability vector given π, a, α and x
$\sigma(x; \pi, a, \alpha)$	The probability of observing outcome x given π, a and α
$argmax_a \{\}$	The value of a which maximizes the quantity inside the brackets
β	Discount factor

6.2.2 STATEMENT OF THE PROBLEM

Consider a deteriorating system that can be, at any point in time $(t = 1, 2, \ldots N)$, in any of the states $\{1, 2, \ldots, n\}$. State 1 represents a new system, whereas; state n represents a failed system. States $\{2, 3, \ldots, n-1\}$ represent increasing stages of system deterioration.

A decision maker has a set of maintenance and production speed control actions. Maintenance actions are taken to control the deterioration level of a system. Control actions can be as simple as do nothing (a_0), some minimal repair i (a_i), or replacement of the whole system (a_n), which is assumed to renew the system. Maintenance control actions are assumed to change the system state depending on the current underlying state and the action taken only. This can be represented by the Markovian transition matrix P^a for a maintenance action a, with elements p^a_{ij} (i and j $\in S$). For example, If a new system is left with no maintenance actions it will keep deteriorating following a Markov process, governed by a transition matrix P^{a_0}, where, $p_{ij}^{a_0}$ is the conditional probability of the system moving from state i to state j in the next time epoch. After selecting a maintenance control action, it is assumed that the decision maker will determine the speed level of the system. Speed control actions simply reflects production rate; where, it is assumed that the production speed (loading level of the system) affects the deterioration level of the system as well. Also, this can be represented by the Markovian transition matrix P^{α} for a speed level α, with elements p_{jk}^{α} with (j and k $\in S$). It can be easily verified that taking a maintenance action a followed by speed action α the state transition can be represented by the matrix $P^{a\alpha} = P^{a} \times P^{\alpha}$ with elements $p_{ik}^{a\alpha}$.

System states are assumed to be partially observed and this is only at discrete points of time. Hence, at time points $t = 1, 2, \dots N$, the decision maker observes a noise corrupted signal that is probabilistically related to the true state of the system on hand. The signal can be, for example, the system output quality, vibration level, or temperature. In this model, output quality is considered as the system signal, this output quality is also assumed to be dependent the system current state only. Thus, r_{kx} represent the conditional probability of observing output quality level x if the system is currently in state k. Since the states of the system are partially observed.

The decision maker's knowledge of a system state is assumed to be probabilistic, such that, π_i represents the probability that a system is currently in state i. Hence, the decision maker has only a belief state or a state occupancy vector, whose elements add up to one, defined as follows:

$$\pi = [\pi_1, \pi_2, \dots, \pi_n]$$

At each time epoch and before making a decision, the decision maker updates the belief state of the system on hand depending on the previous belief state, the maintenance action taken, the production speed level taken and the observation of the output quality. The tradeoffs of this problem are assumed to be as follows: each speed action will give a production rate level for the system, such that, the more the production speed the more the production rate, also, the more the deterioration rate. For maintenance actions, the more the maintenance activities the less deterioration of the system, but on the other hand the more the down time (production loss). Now, irrespective of the maintenance and/or production speed the outcome product quality by means of percent defectives is assumed to depend on the deterioration level of the system (system state).

6.3 MODEL FORMULATION

There are many applications of the POMDP reported in the literature, for examples: industrial, financial, marketing, artificial intelligence, medical, agricultural and many others. In this chapter, a POMDP model is provided to maximize the expected OSE over a finite horizon. Subsection 6.3.1 starts with introducing the POMDP framework, and then an OSE maximization model is provided in Subsection 6.3.2.

6.3.1 POMDP FRAMEWORK

A POMDP is a generalization of MDP (Markov Decision Process). An MDP is a dynamic decision making framework that aims at optimally controlling a Markov stochastic process over a given number of future stages, such that, a set of available control actions influence the state transition of the Markov chain at discrete points of time, with the states of the system being totally observed.

As illustrated in Chapter 2 POMDP is assumed to take place as follows:

1. Starting with a given belief state a control action $a \in A$ is taken
2. A gain or loss takes place ($\mathrm{g}(i, a)$)
3. System is moved to a new state $j \in S$
4. A signal x is observed form the system, which depends on the system state
5. Belief state is updated and the next stage is started

Here we slightly modify to include two actions to be taken in the first step, namely, a maintenance action, followed immediately by a speed action.

6.3.2 ELEMENTS OF THE OSE AND THE POMDP FRAMEWORK

The elements of OSE (Availability, process rate, and quality rate) can be reflected on the POMDP framework as follows:

Availability

In our multi-state settings, it is assumed that there exists an availability index$\in [0, 1]$ for each possible state of the system on hand and for each maintenance action. That is, state/action dependent index $\mathrm{A}(i, a\alpha)$. This can reflect several things, namely, the downtime caused by maintenance actions reflected by unavailability. It can reflect system unavailability for a given state(s) as well and, also that, there can be an interaction effects between production speed used and the maintenance level performed. Basically, this index has to do with maintenance operations so it can be used as $\mathrm{A}(i, a)$ independent of α. An example is a three-state system with the following indices $\mathrm{A}(a\alpha) = [0.8\ \ 0.8\ \ 0]$. This indicates that action a followed by α causes 20% unavailability due to downtime. And the third state indicates an unavailable system.

Hence, the expected availability of a multi-state system may be represented as follows:

$$Availability = \sum_i \pi_i \, \mathrm{A}(i, a\alpha)$$

As another example, for the do nothing action (a_0), where, the n^{th} state is the only state where the system is unavailable, availability can be represented mathematically as the probability of being at any state other than the n^{th} state as follows:

$$Availability = 1 - \pi_n$$

Process Rate

Practically, process rate is decision variable or control action. In this sense, a set of speed actions are assumed to be available to the decision maker where each of these actions will make the system operate at some speed level (α_i) which will yield certain process rate (Pr^{α_i}).

Also, there will be an $n \times n$ state transition matrix $P^{\propto}$ reflecting the assumption, that, deterioration rate depends on the loading level (speed level) at which a system operates and the state of the underlying system, for example, new systems are less likely to deteriorate compared to an already deteriorated ones when both are operated at the same loading level.

An entry of $P^{\propto}$ is P_{jk}^{α}, namely, the conditional probability of the system moves from state j to state k if the system is operated at speed level α, with j and $k \in S$.

As illustrated previously, it can be easily verified that taking a maintenance action a followed by speed action α the state transition can be represented by the matrix $P^{a\alpha} = P^{a} \times P^{\alpha}$ with elements $p_{ik}^{a\alpha}$.

Quality rate

As described in Section 2, for a POMDP, r_{kx} is the conditional probability of observing x

given that the true state of the system is j. As a random variable, X can be used as an unbiased estimator of the quality rate of the products produced.

For discrete random variable X the probability of observing outcome x, at any time, can be expressed as follows:

$$\sigma(x;\pi,a,\alpha)=\sum_i\sum_j\sum_k\pi_i\,p_{ij}^{a}p_{jk}^{\alpha}r_{kx}=\sum_i\sum_k\pi_i p_{ik}^{a\alpha}r_{kx}$$

For within and off specifications, letting $x=1$ for within specifications and 0 for off specifications will give the following as an expected value for the quality rate of the system:

$$E(q_r)=\sigma(1;\pi,a,\alpha)=\sum_i\sum_j\sum_k\pi_i\,p_{ij}^{a}p_{jk}^{\alpha}r_{k1}$$

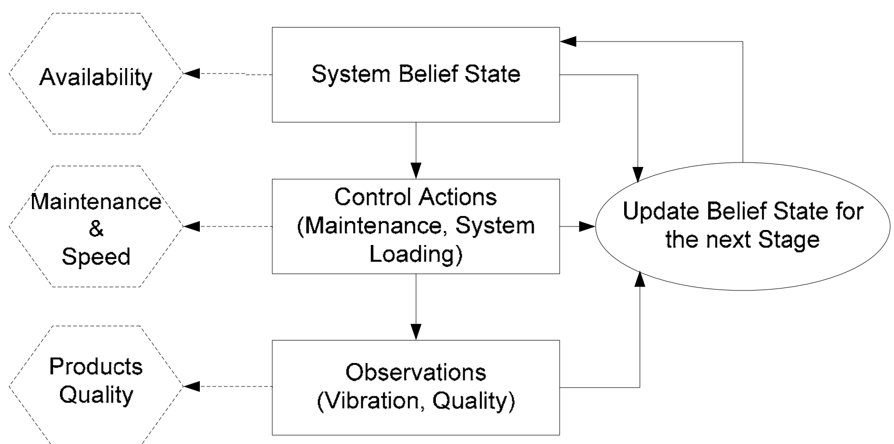


Figure 6.1 OSE elements reflected on the POMDP framework

Figure 6.1 demonstrates the relationship between the POMDP framework and OSE elements; where, Availability is function of the deterioration level of the system resembled by its belief state. Production speeds, as well as, maintenance actions are control actions. The observations received by the process controller represent an unbiased estimator of the underlying quality level of the process outcomes in terms of percent of defectives. At any time stage, maintenance actions and speed levels affect the deterioration level of the system. This provides a new belief state of the system depending on the resulting quality outcome as well. This will result in a new availability of the system. The objective is to maximize OSE such that a tradeoff is made among available control actions (policy) for a given planning horizon.

Combining the three ingredients of the OSE discussed before together can be done as shown by the following equation:

$$OSE_t(\pi, a, \alpha) =$$

$$\sum_i \sum_j \sum_k \sum_x x.(\pi_i\, p_{ij}^a p_{jk}^\alpha r_{kx}) p_r^\alpha \mathrm{A}(i, a\alpha) + \beta \sum_x \sigma(x; \pi, a, \alpha)\, OSE_{t-1}^*(T[\pi, a, \alpha, x]) \quad (6.2)$$

Equation 6.2. gives the expected OSE over t time horizons, with maintenance action a and speed control action α taken at the first time epoch, with the recursion part reflecting the expected future OSE with respect to the future observed signals.

with:

$$T = [T_1(\pi, a, \alpha, x), T_2(\pi, a, \alpha, x), \dots, T_n(\pi, a, \alpha, x)]$$

Such that $T_k(\pi, a, \alpha, x)$ is the conditional probability that the system move to state k starting from state i, given the output x, actions a and α are taken and the current belief state is π.

$$T_k(\pi, a, \alpha, x) = \frac{\sum_i \sum_j \pi_i p_{ij}^a p_{jk}^\propto r_{kx}}{\sum_i \sum_j \sum_k \pi_i p_{ij}^a p_{jk}^\propto r_{kx}}\ with\ i, j, \text{and } k \in S, \text{and } x \in O \quad (6.3)$$

Equivalently:

$$T_k(\pi, a, \alpha, x) = \frac{\sum_i \pi_i p_{ik}^{a\alpha} r_{kx}}{\sum_i \sum_k \pi_i p_{ik}^{a\alpha} r_{kx}}\ with\ i, k \in S, and\ x \in O \quad (6.4)$$

Where, $p_{ik}^{a\propto}$ is the ik^{th} element of the $P^{a\alpha}$ matrix resulting from multiplying P^a and P^α

matrices.

Equation 6.2 can be rewritten as follows for output classified as either within or off specification, where, $x = 1$ for within specifications and 0 for off specifications:

$$OSE_t(\pi, a, \alpha) =$$

$$\sum_i \sum_k (\pi_i \, p_{ik}^{a\alpha} r_{k1}) p_r^{\alpha} \mathrm{A}(i, a) + \beta \sum_x \sigma(x; \pi, a, \alpha) \, OSE_{t-1}^{*}(T[\pi, a, \alpha, x]) \qquad (6.5)$$

Hence, the challenge is to find the optimal maintenance and speed actions at every time horizon t as follows:

$$argmax_{a,\alpha} \, \{OSE_t^{*}(\pi)\} \qquad (6.6)$$

with

$$OSE_t^{*}(\pi) = max_{a,\alpha}[OSE_t(\pi, a, \alpha)]$$

Next, an illustrative example is provided.

6.4 NUMERICAL EXAMPLE

Consider a three state machine with three maintenance actions: do nothing (a_0), repair (a_1) and replace(a_2). And, two speed actions: low speed (α_0) and high speed (α_1). The following are the state transition matrices corresponding to these actions:

$$P^{a_0} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0 & 0.4 & 0.6 \end{bmatrix}, P^{a_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, P^{a_2} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$P^{\alpha_0} = \begin{bmatrix} 0.99 & 0.01 & 0 \\ 0 & 0.98 & 0.02 \\ 0 & 0 & 1 \end{bmatrix}, P^{\alpha_1} = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.9 & 0.1 \\ 0 & 0 & 1 \end{bmatrix},$$

The observations matrix is assumed to be action independent, with two outcomes good (1st column), and bad (2nd column) as follows.

$$R = \begin{bmatrix} 0.95 & 0.05 \\ 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$$

Notice, from Equations 6.2 and 6.5, that, for an action a followed by an α the resulting OSE revenue vector $g(a\alpha)$ can obtained easily from the different transition matrices of the problem as follows.

$$g(a\alpha) = P^{a}P^{\alpha}r(1) *. \mathrm{A}(a\alpha) \times p_r^{\alpha}$$

Where, $r(1)$ is the column of the R matrix corresponding to the within specifications outcome.

Thus, equation 6.2 becomes

$$OSE_t(\pi, a, \alpha) =$$

$$\sum_i \pi_i g(a\alpha, i) + \beta \sum_x \sigma(x; \pi, a, \alpha)\, OSE_{t-1}^{*}(T[\pi, a, \alpha, x]) \qquad (6.2)'$$

With $g(a\alpha, i)$ is the i^{th} component of the $g(a\alpha)$ vector.

$$\mathrm{A}(a_0\alpha_0) = [1.0 \quad 1.0 \quad 0]$$

$$\mathrm{A}(a_0\alpha_1) = [1.0 \quad 1.0 \quad 0]$$

$$\mathrm{A}(a_1\alpha_0) = [0.7 \quad 0.7 \quad 0]$$

$$\mathrm{A}(a_1\alpha_1) = [0.55 \quad 0.55 \quad 0]$$

$$\mathrm{A}(a_2\alpha_0) = [0.6 \quad 0.6 \quad 0]$$

$$\mathrm{A}(a_2\alpha_1) = [0.45 \quad 0.45 \quad 0]$$

$$p_r^{\alpha_0} = 5/6$$

$$p_r^{\alpha_1} = 1$$

Using equation $(6.2)'$ together with the corresponding reward vectors, for example:

$$g^T(a_0\alpha_0) = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0 & 0.4 & 0.6 \end{bmatrix} \times \begin{bmatrix} 0.99 & 0.01 & 0 \\ 0 & 0.98 & 0.02 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 0.95 \\ 0.6 \\ 0.2 \end{bmatrix} *. \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \times \frac{5}{6}$$

$$g(a_0\alpha_1) = [0.6675 \quad 0.3627 \quad 0]$$

The solutions for various horizons lengths of the provided example are provided below. The figures provide the decision maker with a policy graphs representing the optimal maintenance and speed control actions for all possible belief states. If we assume that the complete horizon is 52 weeks then, for instance, the optimal policy regions for 10 time horizons left or the 43rd week, are provided by Figure 6.6. It can be noticed from the figure that it is optimal to replace the system and operate it at high speed if the system is believed to be in state 1 with a probability of 0.48 or less (approximately). Also, repair and low speed is the optimal course of action if the probability of being in state 1is greater or equal to 0.7 (approximately). Otherwise, no maintenance has to be performed and the system should be operated at high speed. This is the optimal policy graph to be adopted by the decision maker for the system parameters given in this example for t=10. Similarly, the other figures provide the optimal maintenance and speed policy regions for different time points in the 52 weeks planning horizon. For examples, Figures 6.8, 6.9 and 6.10 provide the optimal policy regions for the 3rd, 2nd and 1st weeks in the 52 weeks planning horizon.

The Value Iteration algorithm is used to provide the solutions numerically, over a discretized belief space of the state occupancy vectors. To correct for the fact that the actions don't map back to the discretized belief space, bilinear interpolation is used. Consider the following reward, transition, and observation data for a three-state machine maintenance POMDP model.

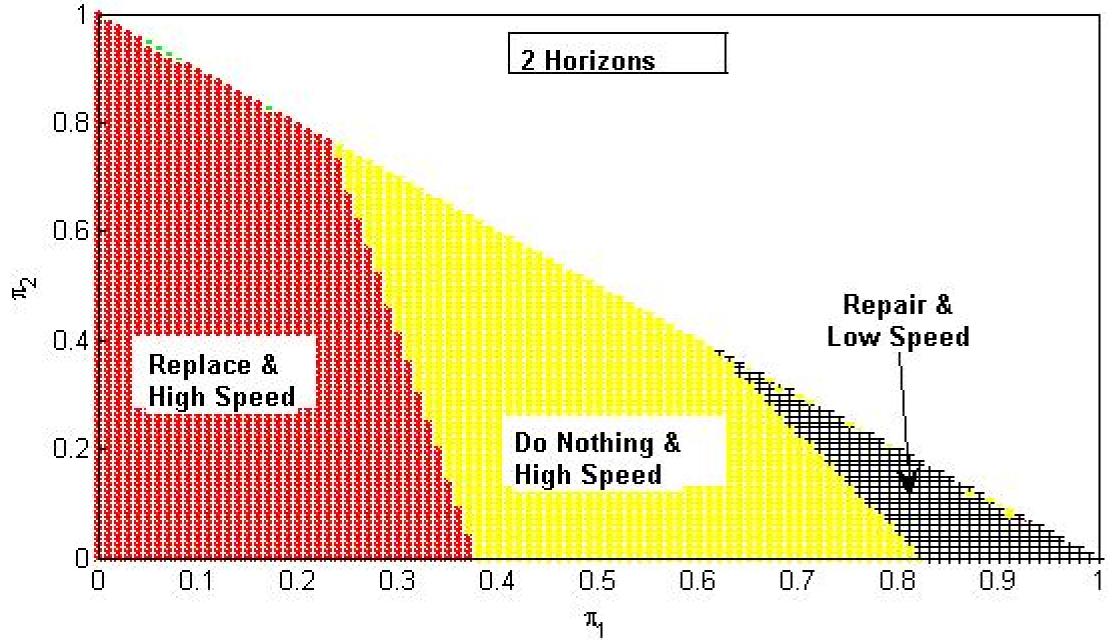


Figure 6.2 Optimal Policy Regions for t=2

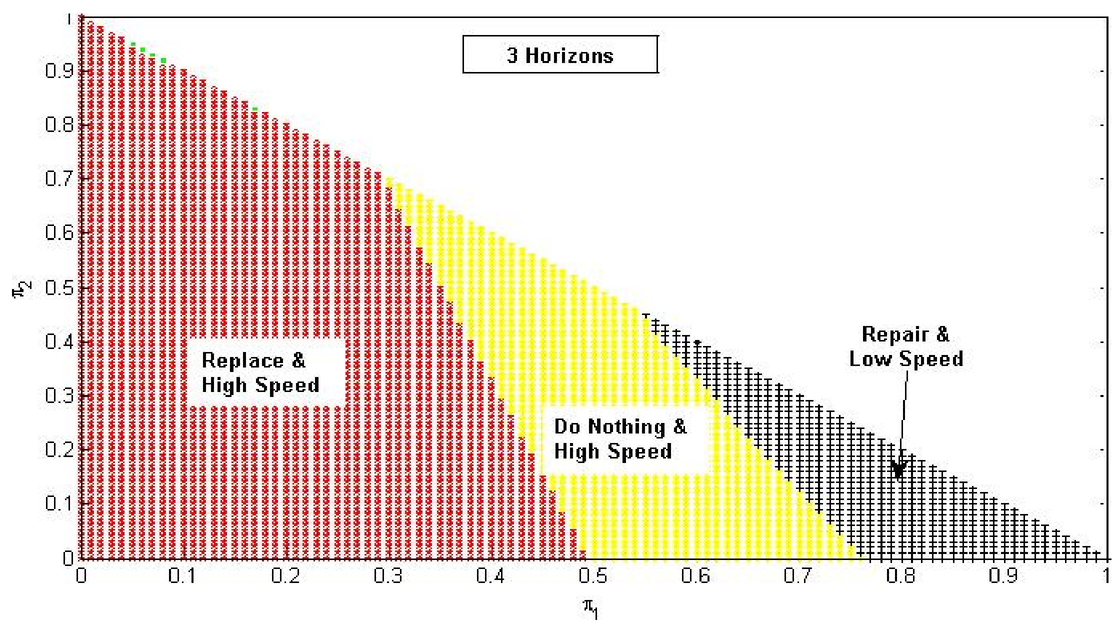


Figure 6.3 Optimal Policy Regions for t=3

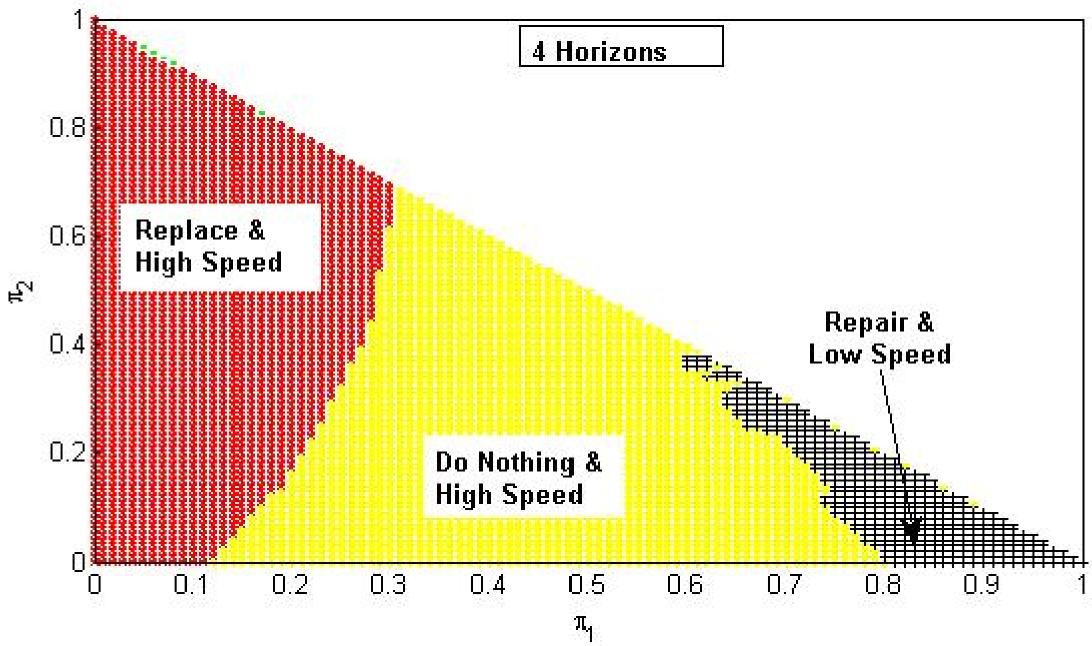


Figure 6.4 Optimal Policy Regions for t=4

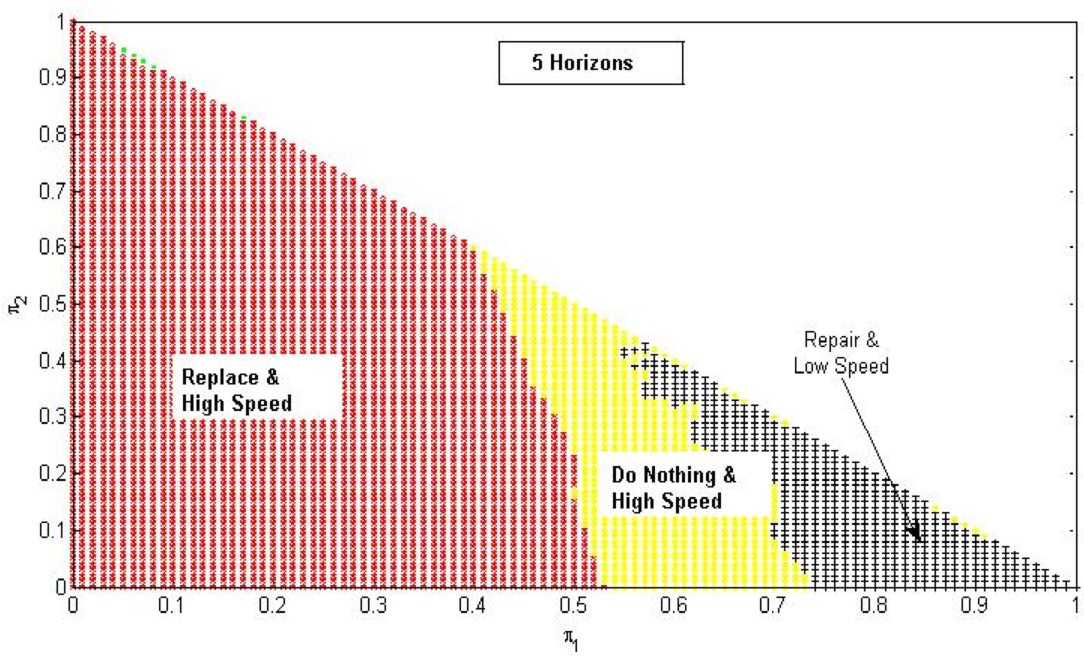


Figure 6.5 Optimal Policy Regions for t=5

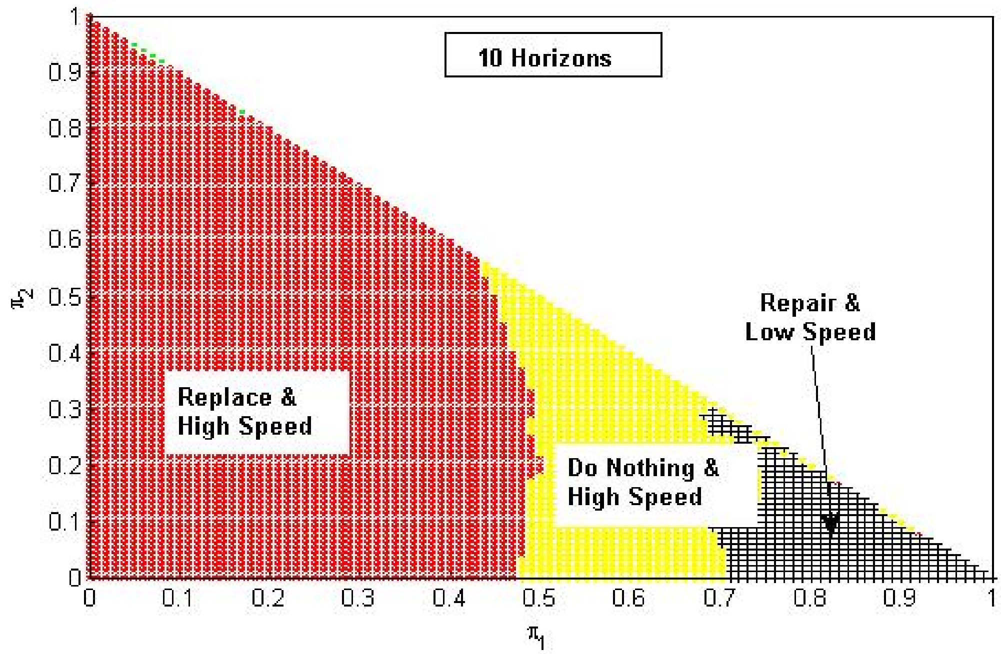


Figure 6.6 Optimal Policy Regions for t=10

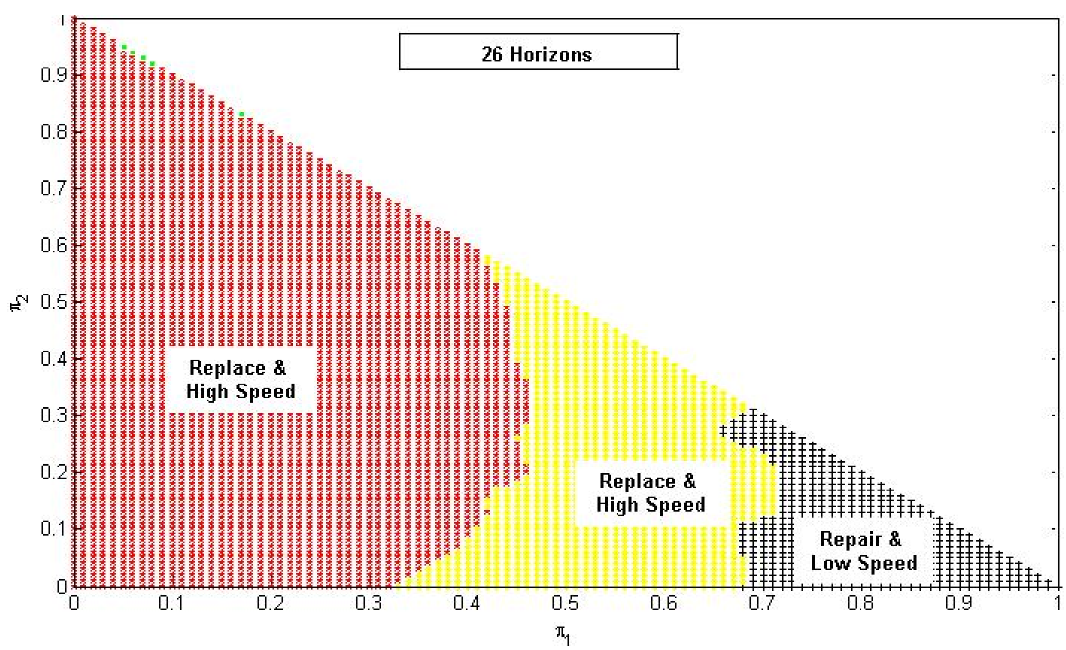


Figure 6.7 Optimal Policy Regions for t=26

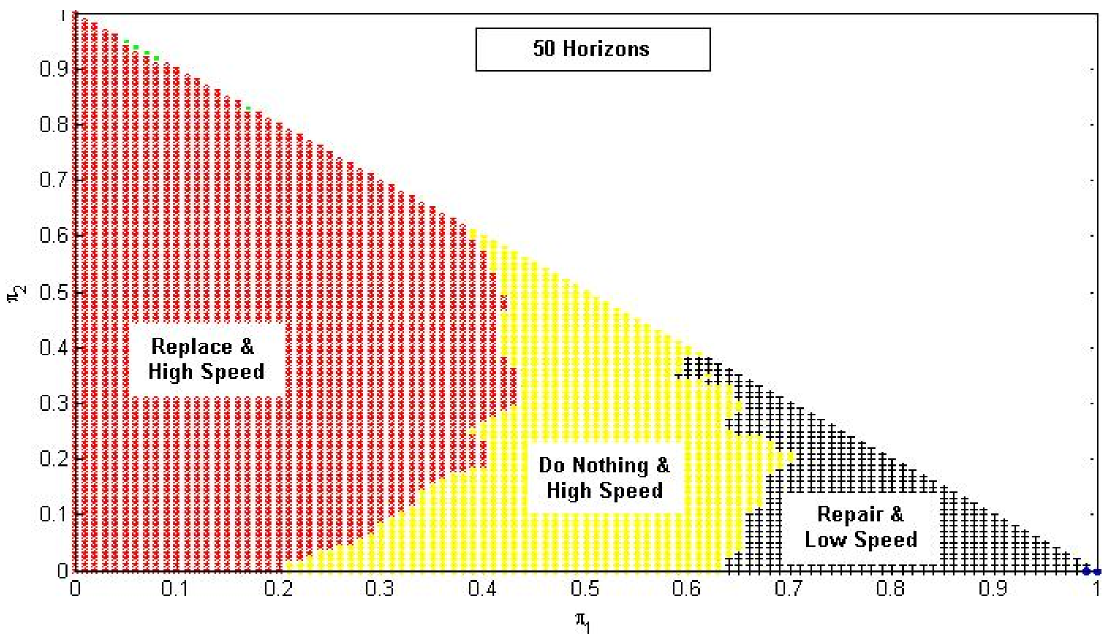


Figure 6.8 Optimal Policy Regions for t=50

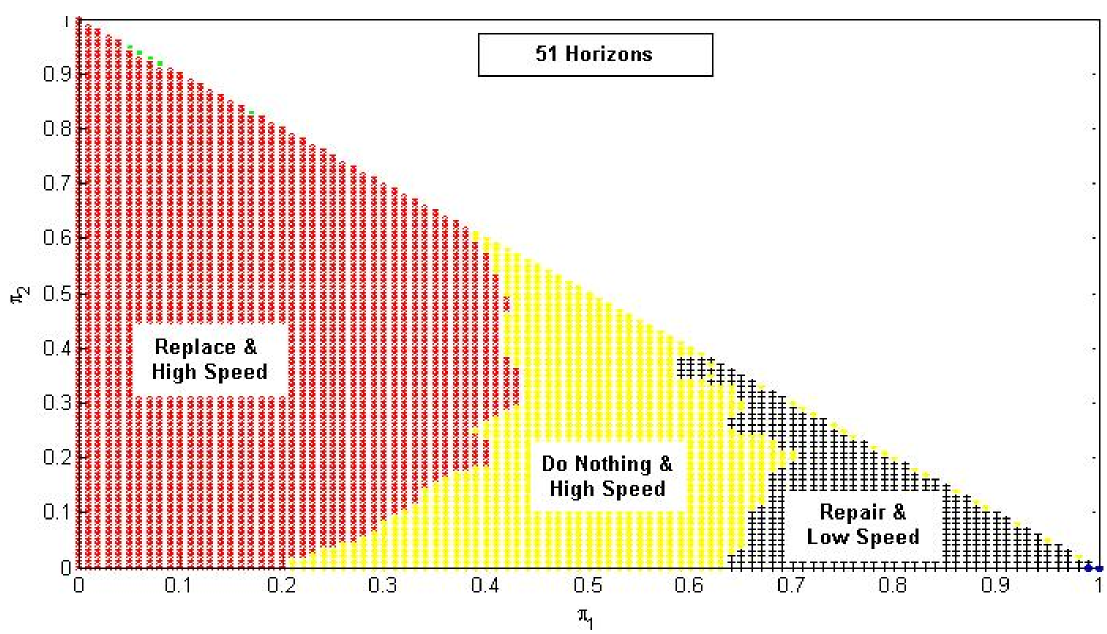


Figure 6.9 Optimal Policy Regions for t=51

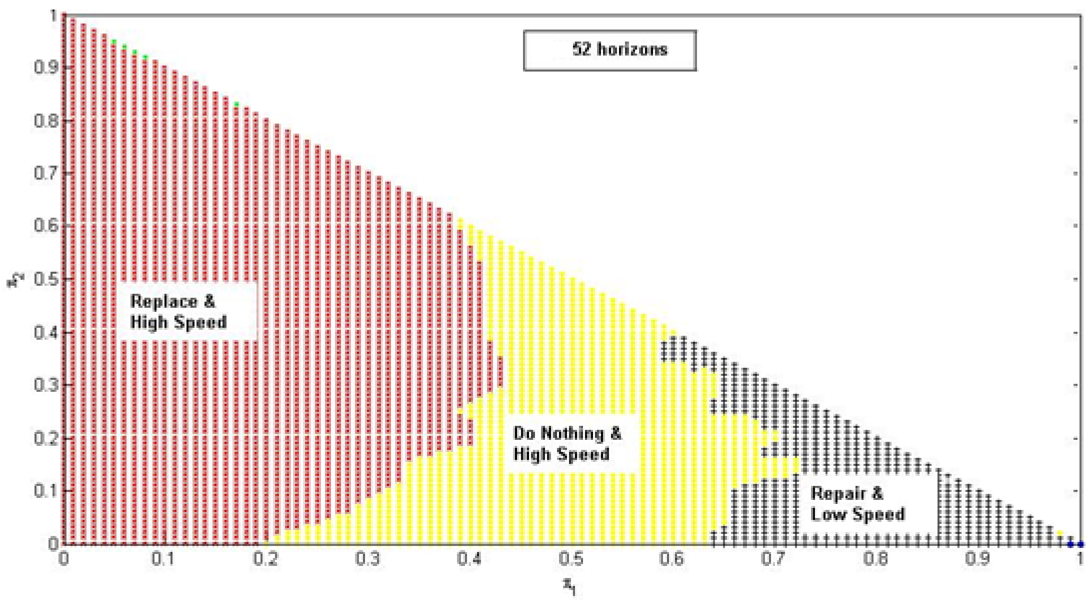


Figure 6.10 Optimal Policy Regions for t=52

6.5 CONCLUSION

In this chapter a model that maximizes OSE and ties maintenance and production decisions has been formulated within the POMDP. The model provides maintenance and operation decision makers with a tool to develop and assess the tradeoff among alternative maintenance actions. In addition, the model was extended to assess the effect of measurement errors. The model assumes the existence of estimates of transition probabilities and clear identification of different system states. The model can provide a framework for better decision making in the areas of maintenance and operation. The work in this chapter can be extended by considering other objective functions and understanding more the effect operation on system deterioration.

CHAPTER 7

MEASUREMENT ERRORS AND FUZZY OBSERVATIONS

7.1 INTRODUCTION

In this chapter, two problems are considered. First, is the effect of measurement errors on the optimal maintenance polices for the POMDP models, which to the best of our knowledge was not addressed in the literature before. Second, is the case of fuzzy POMDP observations.

A mathematical model that incorporates the effect of measurement errors is provided. This is achieved by using a three layers hidden Markov Model. Hence, this model is suitable to correct for the effect of measurement errors. It has been shown that the resulting Bayesian update of the developed model is a sufficient statistic and the optimal value function is piecewise-linear convex.

Although POMDP itself can represent an error prone MDP; but far to our knowledge, error prone signals for POMDP was not considered before. In spite of the fact that, output quality measures are widely accepted to be the signal received in the POMDP context and the significance of quality measurement errors as reflected by the amount of research dealing with it.

The significance of quality measurement errors, for example, is reported in the different fields of quality control as illustrated by in Chapter 3. Outcome products quality of the production processes is reported in the literature as a possible output signal of the

POMDP models; see for examples Smallwood and Sondik (1973) and Ivy and Pollack (2005). As far to our knowledge, outcome quality measurement errors was never considered for POMDP models. In this chapter the quality measurement errors is formulated and practical examples are suggested from the practice, and numerical examples are suggested as well.

As an other application of measurement errors in POMDP models is the following extension for the example by Smallwood and Sondik (1973), where, to evaluate different teaching approaches it is stated that students' grades are the output signal form their true underlying level of understanding. Here we add: in some exams, more than a grader grade the same exam sheet and if there was a variation in their evaluation a third person grades the same sheet. This is another practical application where signals measurement error can fit in the POMDP framework.

Unlike the POMDP models in the literature which assume a two-layer hidden Markov model, where as described in Chapter 2, one Markov chain represents the deterioration mechanism and the second is for the observation which is probabilistically related to the true state of the system. Here, it is assumed that a measured value of output signal is also probabilistically related to the actual signal value due to measurement errors. Hence, a three-layer hidden Markov chain exists.

Then, fuzzy logic is integrated to the POMDP framework by fuzzifying the state observations matrix. This is a typical scenario where relation between the system state and the output signal is better judged by an expert opinion.

The rest of this chapter is organized as follows: Section 7.2. provides the nomenclature and the measurement errors model formulation, Section 7.3. provides some illustrative examples for measurement errors, a different view point on the problem is provided in Section 7.4., where, fuzzy observations are considered, Section 7.5. provides some examples on the application of fuzzy observations and, in Section 7.6., a new technique to calculate the Bayesian update with continuous fuzzy observations is provided. Finally, in Section 7.7., some concluding remarks are provided.

7.2 MEASUREMENT ERRORS: NOMENCLATURE AND MODEL FORMULATION

In this section the nomenclature that will be used throughout this chapter is provided. Also, the mathematical model incorporating the effect of measurement errors is provided as well.

7.2.1 NOMENCLATURE

Follows is the nomenclature that will be used throughout this chapter

S	System's state set {1,2,…n}
i, j	Elements of S
O	System signals set with elements {1,2, … m}
x	System signal or observation, element of O
$O^{'}$	Observations set with elements {1,2, … m}
y	Error-prone signal measure of x, element of $O^{'}$
A	The set of maintenance actions available to the decision maker
a_i	Element of A, that can range from a_0 (do nothing) to a_n (replace)
P^a	System's state transition matrix $(n \times n)$ corresponding to maintenance action a
P_{ij}^a	Probability that the system will move from state i to state j if action a was taken
	An $(n \times m)$ state signal transition matrix
r_{jx}	And entry in the R matrix which gives the probability that observation x will be observed if the system has moved to state j
R^e	An $(m \times l)$ output signal and output signal measures transition matrix
r_{xy}^e	And entry in the R^e matrix which gives the probability that measurement y will be taken for the actual output signal x
R'	An $(n \times l)$ state signal measurement transition matrix = $R \times R^e$
r_{jy}'	And entry in the $R^{'}$ matrix which gives the probability that measurement y will be taken for the actual output signal x

R''	Fuzzified R matrix
f_l	Fuzzy membership function for the set low
f_m	Fuzzy membership function for the set medium
f_h	Fuzzy membership function for the set high
$T(\pi, a, x)$	Posterior state occupancy probability vector given π, a and x
$\sigma(x; \pi, a)$	The probability of observing output x given π and a
$argmax_a \{\}$	The value of a which maximizes the quantity inside the brackets
β	Discount factor
α_x and β_x	Fuzzy membership parameters

7.2.2 MODEL FORMULATION AND ANALYSIS

In this section a typical POMDP maximization model, as the one provided in Chapter 2 is extended to include the effect of measurement errors. It will be assumed that the signal received by the decision maker in the POMDP framework is error-prone as well, hence this measurement is probabilistically related to the true state of the POMDP signal.

To model this scenario, instead of observing an output signal x, it is assumed that the controller receives y; where, the probability of observing y is function in the underlying true outcome x. Mathematically, $f(Y) = f(Y/X)$. For discrete X and Y, and assuming that measurement errors depend only on the observed output quality, the observation process can be modeled as a Markov chain with R^e as a transition matrix with r^e_{xy} as the conditional probability of observing outcome y while the true underlying one is x, with $x \in O$ and $y \in O'$.

The following Equation (7.1) represents the POMDP maximization model discussed in Chapter 2.

$$V_t(\pi, a) = \sum_i \pi_i g(i, a) + \beta \sum_x \sigma(x; \pi, a) V^*_{t-1}(T[\pi, a, x]) \quad (7.1)$$

Hence, the problem on hand can be modeled as a three-layer hidden Markov model as follows:

$$V_t(\pi,a) = \sum_i \pi_i g(i,a) + \beta \sum_y \sigma(y;\pi,a)\, V_{t-1}^*(T[\pi,a,y]) \qquad (7.2)$$

The difference between equations 7.1 and 7.2, is the recursion part, where in the later, it is conditioned on the error-prone quality measure y of the actual outcome quality level x.

Such that, $\sigma(y;\pi,a) = \sum_i \sum_j \sum_x \pi_i p_{ij}^a r_{jx} r_{xy}^e$

For example, the system above adds one more hidden Markov chain as an extension to the student learning example by Smallwood and Sondik. Specifically, teaching approach (action) brings the student from state i to state j with P_{ij} and based on the student's response to some questions, his output in the exam is x with probability of r_{jx}, here we want to add an error component, that is, the student's grade will differ if different teachers grad his paper. So that, we will assume the teacher's grading (measurement) is error prone. The student outcome will be y instead of x with a probability of r_{xy}^e. Practically speaking, in some cases, more than one grader are assigned to grade the same paper for each student and it happens that results might differ from a grader to another.

Also it can be shown that the following Bayesian update is a sufficient statistic, which should be used instead with the model in Section 4.2. to adjust for the measurement errors.

$$T_j(\pi,a,y) = \frac{\sum_i \sum_x \pi_i p_{ij}^a r_{jx} r_{xy}^e}{\sigma(y;\pi,a)} \; with\; i,j \in S, x \in O, and\; y \in O' \qquad (7.3)$$

With $r_{jx} \in R$ and $r_{xy}^e \in R^e$, O' is the set of possible erroneous values that is measured.

Again, the objective is to find

$$a^* = argmax\{V_t^*(\pi)\} \qquad (7.4)$$

Where,

$$V_t^*(\pi) = max_a[V_t(\pi,a)] \qquad (7.5)$$

With $V_t(\pi,a)$ given by equation 7.2.

Proposition 1:

The Bayesian update given by Equation (7.3) is sufficient statistic to describe the system state for the three layers hidden Markov Model.

Proof:

The Bayesian update of the three layers POMDP as given by Equation 7.3 is sufficient statistic for the three layers hidden Markov Model

$$T_j(\pi, a, y) = \frac{\sum_i \sum_x \pi_i p_{ij}^a r_{jx} r_{xy}^e}{\sum_i \sum_k \sum_x \pi_i p_{ij}^a r_{jx} r_{xy}^e} \; with \; i, j \in S, x \in O, and \; y \; \in O'$$

Proof:

The proof is a straight forward extension for that in Smallwood and Sondik (1973).

P_{ij}^a: Probability that the system will move to state j given it is in state i and action a is taken.

r_{jx}: Probability that, true output x is revealed from the system if system state is j.

r_{xy}^e: Probability of observing outcome y as a measurement if the actual outcome is x.

At any point of time the history of the hidden Markov process on hand can be captured by the following vector

$$h(t) = [r(t), r^e(t), a(t), h(t-1)] \qquad \text{(I)}$$

The probability of being in state j at any point of time is expressed as follows:

$$T_k(t) = P_r\{S_t = j | h(t)\} \qquad \text{(II)}$$

Substituting (I) in (II) yields:

$$T_k(t) = \frac{\sum_x P_r\{S(t)=j, r(t)=x, r^e(t)=y, a(t), h(t-1)\}}{\sum_x P_r\{r(t)=x, r^e(t)=y, a(t), h(t-1)\}}$$

$$= \frac{\sum_x P_r\{S(t)=j, r(t)=x, r^e(t)=y | \; a(t), h(t-1)\}}{\sum_x P_r\{r(t)=x, r^e(t)=y | a(t), h(t-1)\}}$$

Conditioning on $S(t-1)$ and expanding joint probabilities in the numerator into multiplications of conditional probabilities yields:

$$T_k(t) = \frac{\sum_i \sum_x C1 \times C2 \times C3 \times C4}{\sum_x P_r\{r(t)=x, r^e(t)=y, a(t), h(t-1)\}}$$

Where:

$C1 = P_r\{S(t-1) = i | a(t), h(t-1)\}$

$C2 = P_r\{S(t) = j | S(t-1) = i, a(t), h(t-1)\}$

$C3 = P_r\{r(t) = x | S(t-1) = i, S(t) = j, a(t), h(t-1)\}$

$C4 = P_r\{r^e(t) = y | S(t-1) = i, S(t) = j, r(t) = x, a(t), h(t-1)\}$

Hence at any time t for time independent process, we have:

$$T_k(\pi, a, y) = \frac{\sum_i \sum_x \pi_i p_{ij}^a r_{jx} r_{xy}^e}{\sum_i \sum_j \sum_x \pi_i p_{ij}^a r_{jx} r_{xy}^e}$$

And the result follows.

Proposition 2:

$V_t^*(\pi)$ given by equation 7.5 is piecewise linear convex function

Proof:

In this Appendix, the result of Smallwood and Sondik (1973) of piecewise linear convex optimal reward function is extended to a three-layer hidden Markov model.

$V_t^*(\pi)$ given by Equation 7.5 is piecewise linear convex function and can be expressed as follows:

$$V_t^*(\pi) = max_a[\sum_{i=1}^n \gamma_i^a(t)\pi_i]$$

For a set of vectors $\gamma^a(t) = \{\gamma_1^a(t), \gamma_2^a(t), \dots, \gamma_n^a(t)\}$

Such that $a \in A$, namely, the set of all control actions A.

Proof: (by induction)

The result is true for $t = 1$ as follows:

$V_1^*(\pi) = max_a[\sum_i \pi_i g(i,a)]$ (I)

Where, $\gamma_i^a(1) = g(1,a)$

Then, assuming that $V_t^*(\pi)$ is piecewise linear convex function, then:

$V_t^*(T[\pi,a,y]) = max_a\left[\sum_{j=1}^{n} \gamma_j^a\,(t) T_j(\pi,a,y)\right]$ (II)

then

$V_{t+1}^*(\pi) = max_a\{\sum_i \pi_i g(i,a) + \beta \sum_y \sigma(y;\pi,a)\, V_{t-1}^*(T[\pi,a,y])\}$ (III)

Substituting equation (II) in (III) gives

$= max_a\{\sum_i \pi_i g(i,a) + \beta \sum_y \sigma(y;\pi,a)\, max_a\left[\sum_{j=1}^{n} \gamma_j^a\,(t) T_j(\pi,a,y)\right]\}$ (IV)

Now, from Equation (7.3)

$T_j(\pi,a,y) = \frac{\sum_i \sum_x \pi_i p_{ij}^a r_{jx} r_{xy}^e}{\sigma(y;\pi,a)}$

Again, substituting Equation (7.3) in (IV) gives:

$max_a\{\sum_i \pi_i g(i,a) + \beta \sum_y max_a\left[\sum_{j=1}^{n} \gamma_j^a\,(t)(\sum_i \sum_x \pi_i p_{ij}^a\, r_{jx} r_{xy}^e)\right]\}$ (V)

Then the result follows for any t.

7.3 NUMERICAL ILLUSTRATIONS FOR MEASUREMENT ERRORS

In this section, two examples are provided to illustrate the effect of measurement errors.

Example 7.1:

This is an example of a POMDP model with measurement errors (Equations 7.2-7.4). Consider a three state machine with three maintenance actions: do nothing a_0 , repair a_1 and replace a_2. It is assumed that outcome production quality is the signal observed from the process. The following are the reward vectors and the transition matrices corresponding to the different maintenance actions:

$g(a_0\,) = [4\ \ 2\ \ 0]$

$g(a_1) = [3\ \ 2\ \ 0]$

$(a_2) = [2\ \ 2\ \ 2]$

$$P^{a_0} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0 & 0.4 & 0.6 \end{bmatrix}, P^{a_1} = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.6 & 0.3 \end{bmatrix}, P^{a_2} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

The observations matrix is assumed to be action independent, with three product quality outcomes good (1st column), having defects (2nd column) and defective (3rd column).

$$R = \begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

It is assumed that the type of error in measurement that can take place is misclassifying the output signal as an other one, for example, classifying good outcome as having defects or defective. The error is assumed to be uniformly distributed. Follows are three error transition matrices which correspond to the 5%, 10% and 20% errors.

$$\mathrm{R}^{e_1} = \begin{bmatrix} 0.95 & 0.025 & 0.025 \\ 0.025 & 0.95 & 0.025 \\ 0.025 & 0.025 & 0.95 \end{bmatrix}$$

$$\mathrm{R}^{e_2} = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}$$

$$\mathrm{R}^{e_3} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

Follows are the solutions for various horizons lengths of the provided example. Value Iteration algorithm is used to provide the solutions numerically, over a discretized belief space of the state occupancy vectors. To correct for the fact that the actions don't map back to the discretized belief space, bilinear interpolation is used.

Policy regions for the cases of no error and the existence of measurement error are given in the figures below. It can be seen from the optimal policy regions that measurement errors has an effect on the optimal policy regions. For example, consider any of the four horizons say the 26th horizon (Figures 7.9-7.12). It can be seen from the figures that the more the error the more the difference from the case with no error (Figure 7.9) for the 26 horizons. Also it can be seen from all the given horizons (2, 10, 26 and 52) that the area of the policy region corresponding to the replace action increases slightly as the error increases. In fact, measurement errors do not show a drastic effect on the optimal policy regions. This is due to the fact that $R \times R^e$ is very close to the original *R* matrix as shown below:

R$\times$ R^e:

No Error

$$\begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

5% Error

$$\begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \times \begin{bmatrix} 0.95 & 0.025 & 0.025 \\ 0.025 & 0.95 & 0.025 \\ 0.025 & 0.025 & 0.95 \end{bmatrix} = \begin{bmatrix} 0.6725 & 0.3025 & 0.0259 \\ 0.2100 & 0.4875 & 0.3025 \\ 0.1175 & 0.1175 & 0.7650 \end{bmatrix}$$

10% Error

$$\begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \times \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix} = \begin{bmatrix} 0.6451 & 0.3050 & 0.0509 \\ 0.2200 & 0.4750 & 0.3050 \\ 0.1350 & 0.1350 & 0.7300 \end{bmatrix}$$

20% Error

$$\begin{bmatrix}0.7 & 0.3 & 0\\0.2 & 0.5 & 0.3\\0.1 & 0.1 & 0.8\end{bmatrix}\times\begin{bmatrix}0.8 & 0.1 & 0.1\\0.1 & 0.8 & 0.1\\0.1 & 0.1 & 0.8\end{bmatrix}=\begin{bmatrix}0.5901 & 0.3101 & 0.1008\\0.2400 & 0.4500 & 0.3100\\0.1700 & 0.1700 & 0.6600\end{bmatrix}$$

Hence the Bayesian update as a function of R and R^e, or equivalently $R = R \times R^e$, is not highly affected.

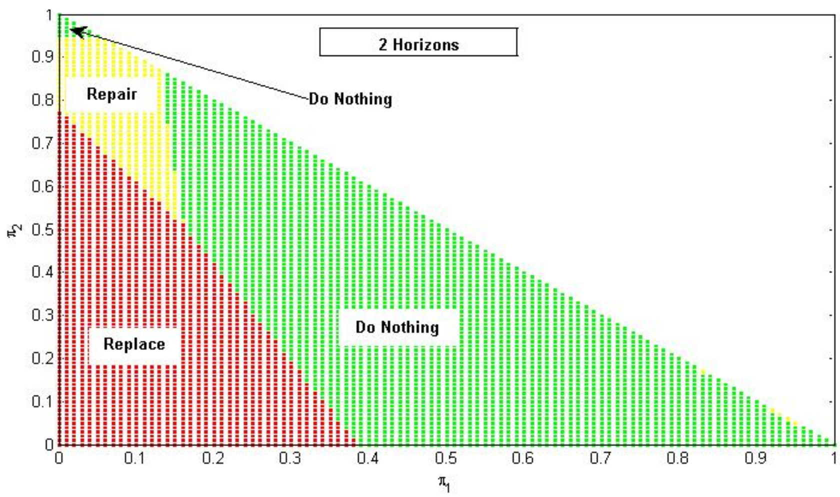


Figure 7.1 Optimal Policy Regions for t=2

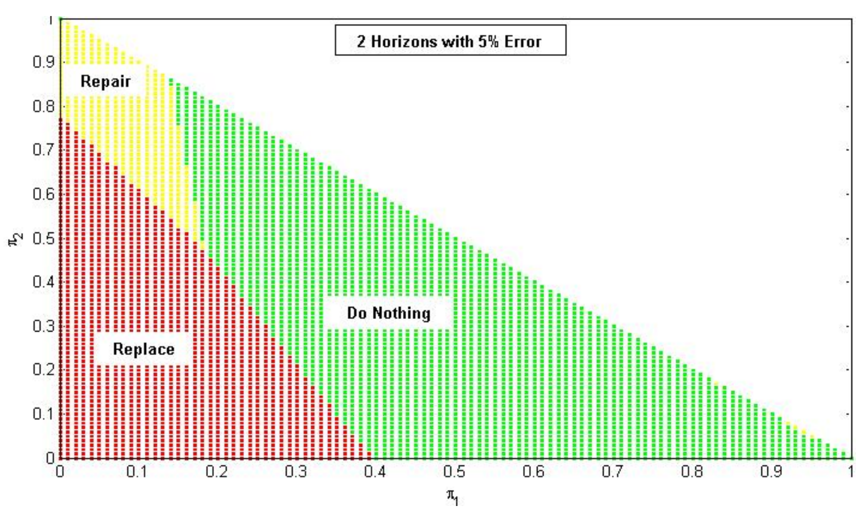


Figure 7.2 Optimal Policy Regions for t=2 with 5% Error

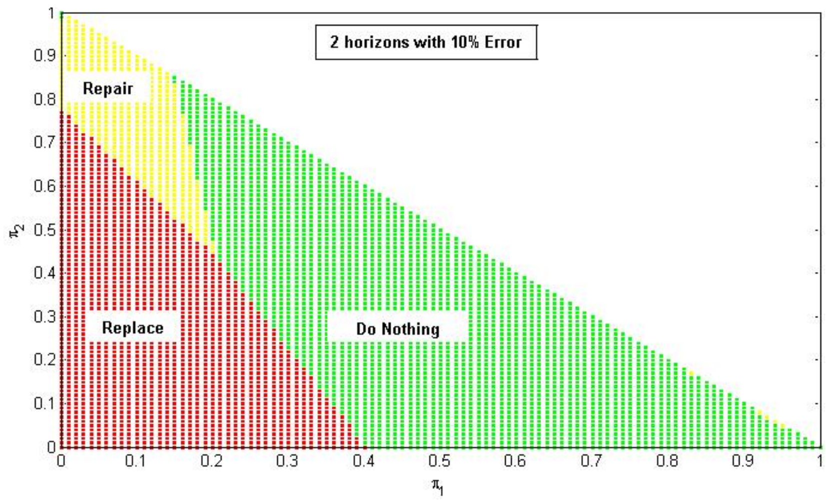


Figure 7.3 Optimal Policy Regions for t=2 with 10% Error

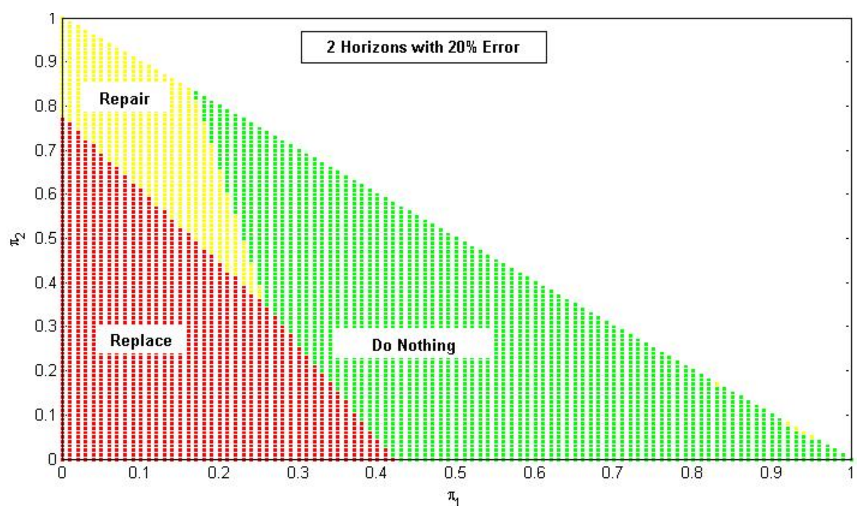


Figure 7.4 Optimal Policy Regions for t=2 with 20% Error

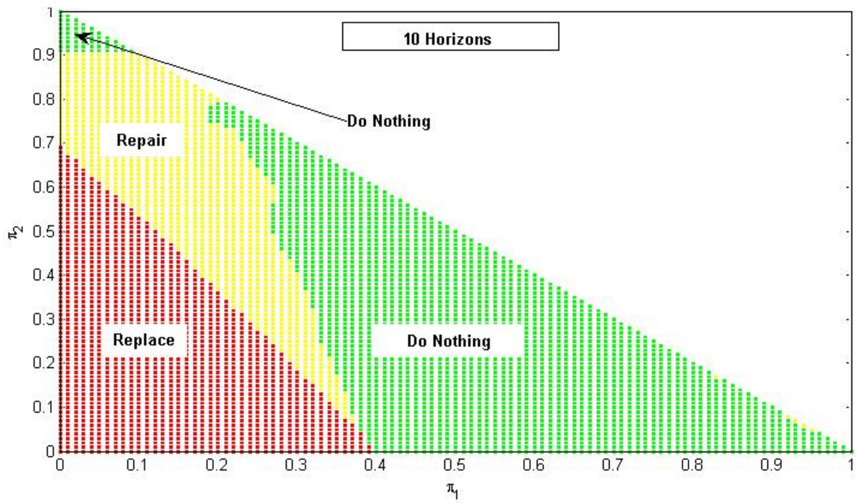


Figure 7.5 Optimal Policy Regions for t=10

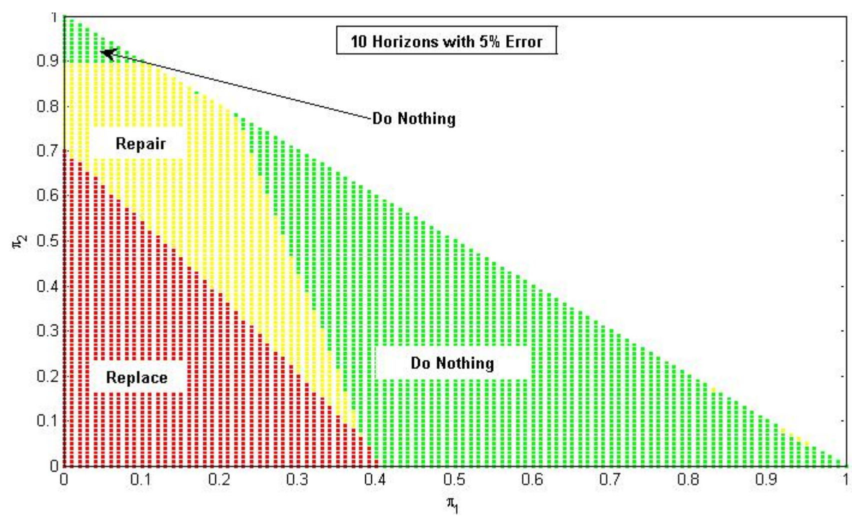


Figure 7.6 Optimal Policy Regions for t=10 with 5% Error

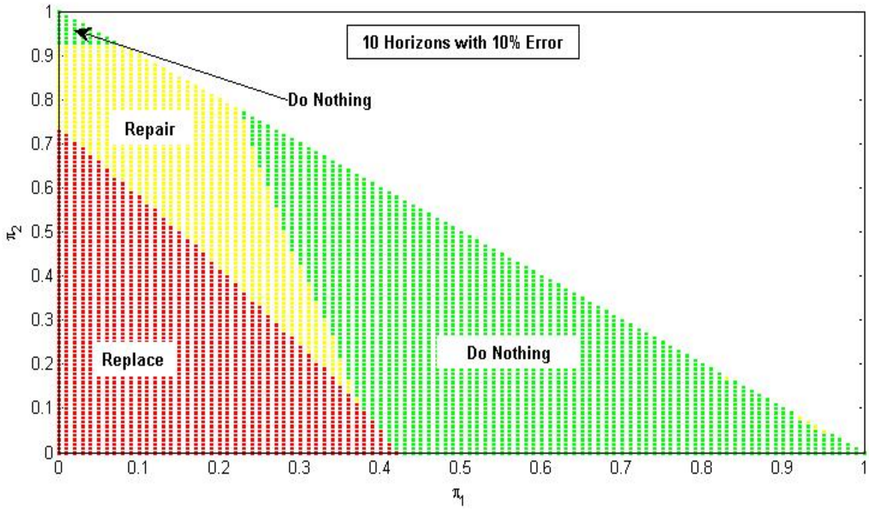


Figure 7.7 Optimal Policy Regions for t=10 with 10% Error

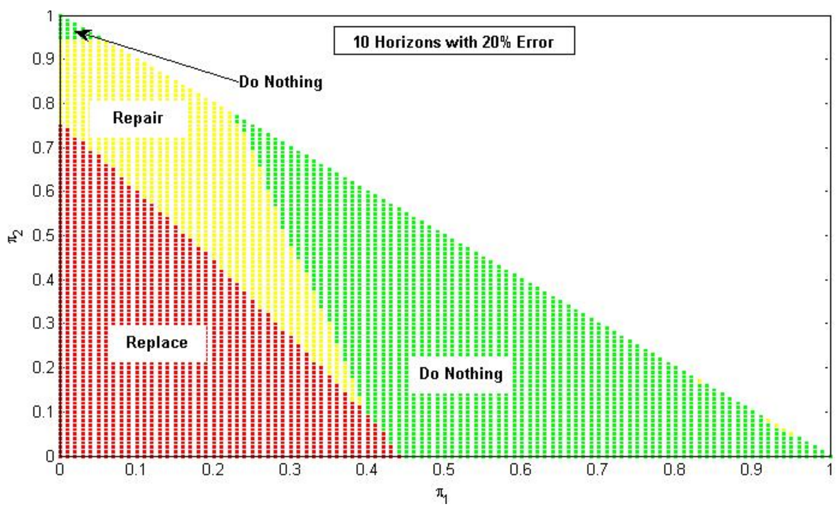


Figure 7.8 Optimal Policy Regions for t=10 with 20% Error

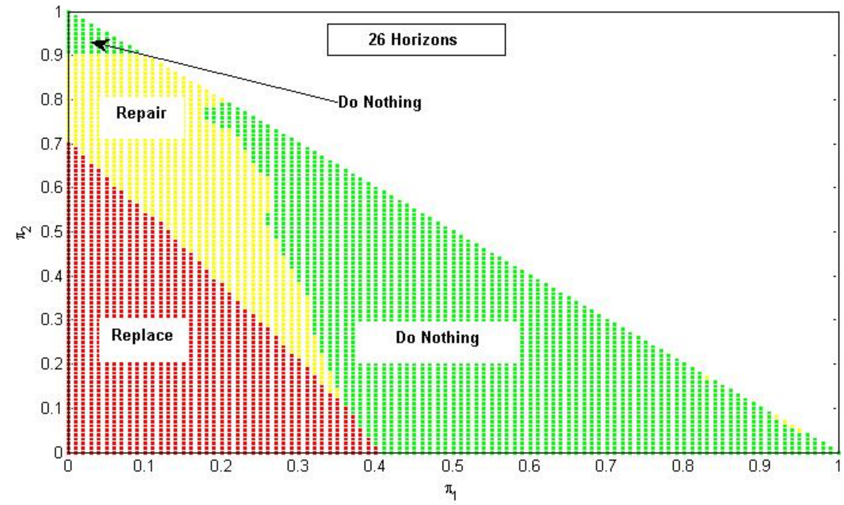


Figure 7.9 Optimal Policy Regions for t=26

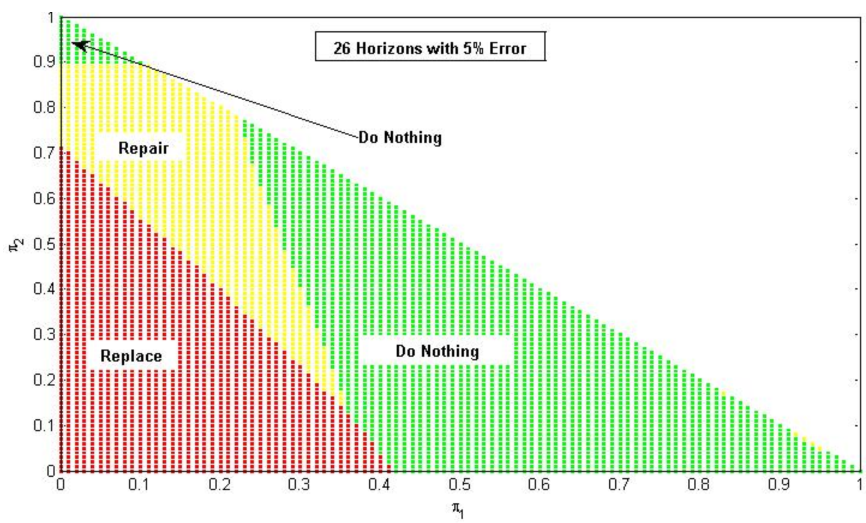


Figure 7.10 Optimal Policy Regions for t=26 with 5% Error

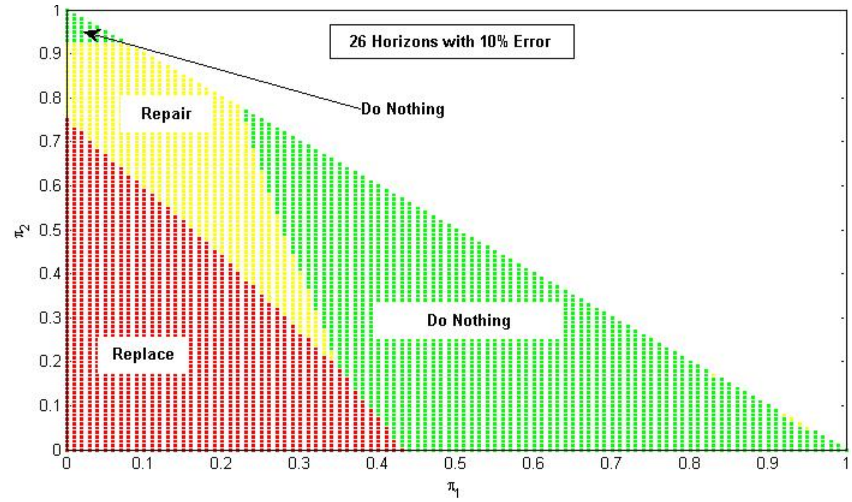


Figure 7.11 Optimal Policy Regions for t=26 with 10% Error

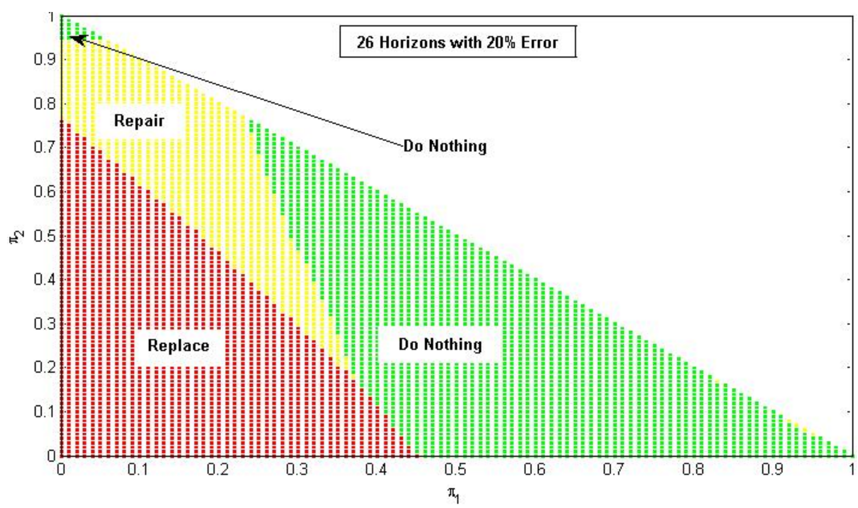


Figure 7.12 Optimal Policy Regions for t=26 with 20% Error

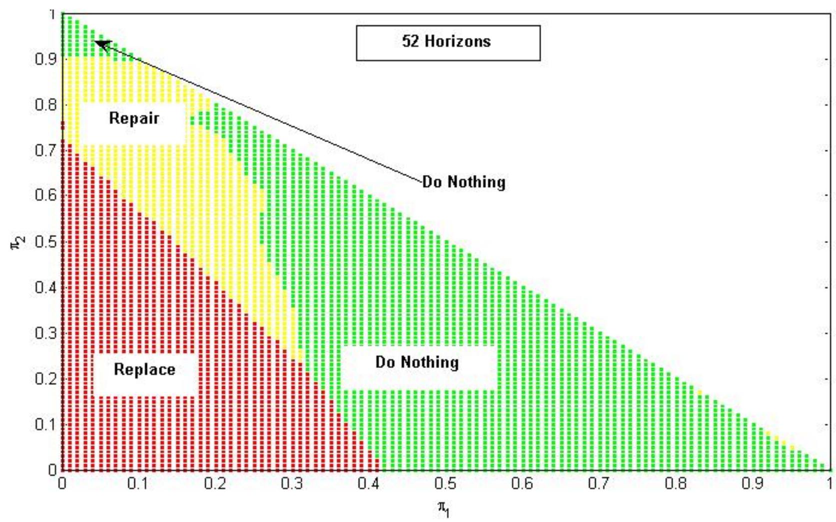


Figure 7.13 Optimal Policy Regions for t=52

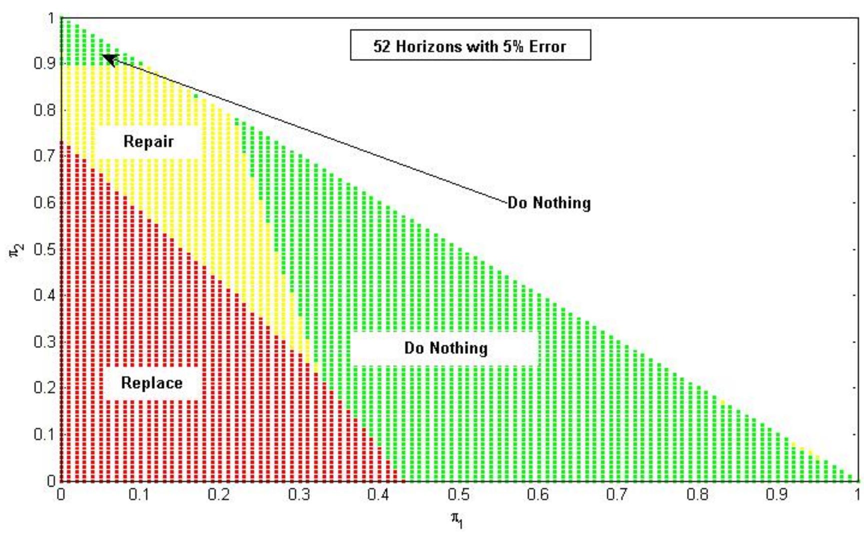


Figure 7.14 Optimal Policy Regions for t=52 with 5% Error

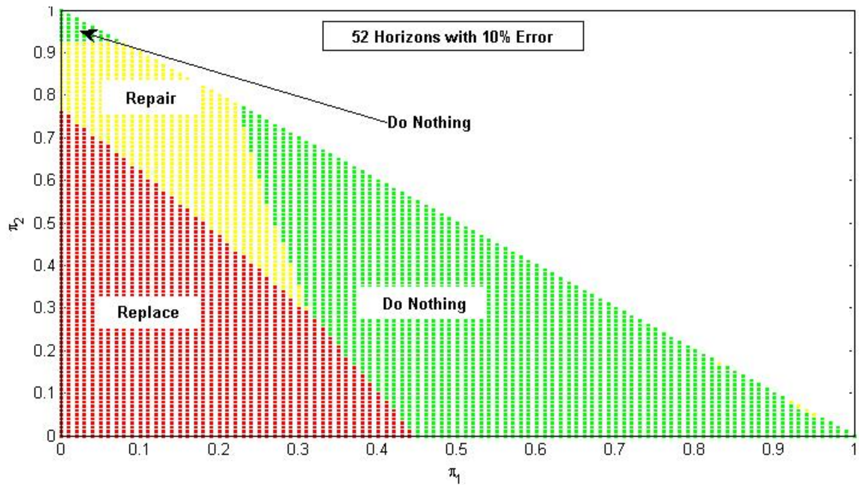


Figure 7.15 Optimal Policy Regions for t=52 with 10% Error

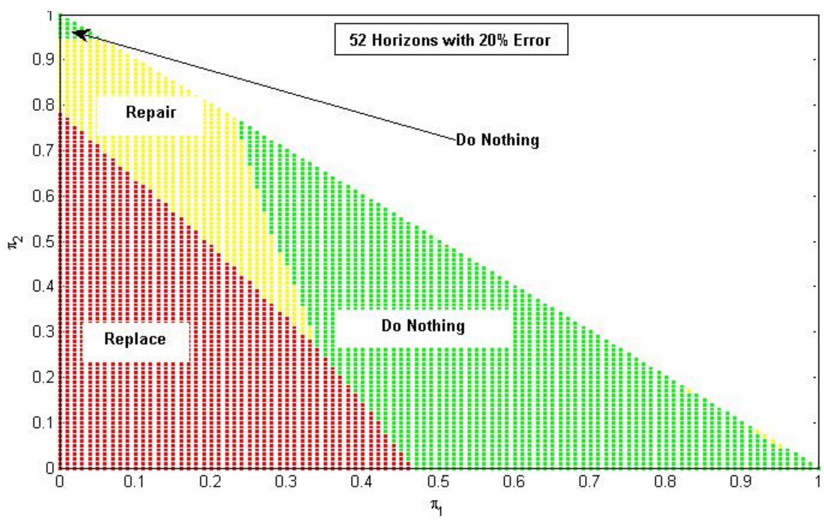


Figure 7.16 Optimal Policy Regions for t=52 with 20% Error

Example 7.2:

This example demonstrates the relationship between the quality of the signal received and the measurement errors. To do this we consider two systems. One that is close to being non-observable (POMDP with no information) and the other provides better signal quality. Hence, the first system has a close to being identical *R* matrix rows.

Scenario 1:

Consider a three-state, three-action and three-observations POMDP model with the following data:

With three control actions available, namely: do nothing, repair, and replace ($a_0, a_1,$ $and\ a_2$) respectively, consider the following reward, transition and observation data for this POMDP system.

$g(a_0) = [\,4 \quad 2 \quad 1\,]$

$g(a_1) = [\,3 \quad 1.5 \quad 1]$

$g(a_2) = [\,1 \quad 1 \quad 1]$

$$P^{a_0} = \begin{bmatrix} 0.9 & 0.07 & 0.03 \\ 0 & 0.6 & 0.4 \\ 0 & 0.1 & 0.9 \end{bmatrix}, P^{a_1} = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.4 & 0.6 & 0 \\ 0.1 & 0.8 & 0.1 \end{bmatrix}, P^{a_2} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$R = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.5 & 0.3 & 0.2 \\ 0.4 & 0.3 & 0.3 \end{bmatrix}$$

Scenario 2: this scenario has the same data as in scenario 1in addition to assuming the existence of measurement errors as follows:

$$\mathrm{R}^e = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.1 & 0.7 & 0.2 \\ 0.1 & 0.8 & 0.1 \end{bmatrix}$$

Scenario 3:

This is same to scenario 1 except the existence of a different *R* matrix.

$g(a_0) = [\,4 \quad 2 \quad 1\,]$

$g(a_1) = [\,3 \quad 1.5 \quad 1]$

$g(a_2) = [\,1 \quad 1 \quad 1]$

$$P^{a_0} = \begin{bmatrix} 0.9 & 0.07 & 0.03 \\ 0 & 0.6 & 0.4 \\ 0 & 0.1 & 0.9 \end{bmatrix}, P^{a_1} = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.4 & 0.6 & 0 \\ 0.1 & 0.8 & 0.1 \end{bmatrix}, P^{a_2} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$R = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.7 & 0.3 & 0 \\ 0 & 0.4 & 0.6 \end{bmatrix}$$

Notice the R matrix of scenario 1 is close to the no information case (identical rows) unlike the case of this scenario.

Scenario 4: here we have similar to scenario 3 but assuming the existence of measurement errors as follows:

$$\mathrm{R}^e = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.1 & 0.7 & 0.2 \\ 0.1 & 0.8 & 0.1 \end{bmatrix}$$

Next, solutions for these four scenarios are provided for three horizons.

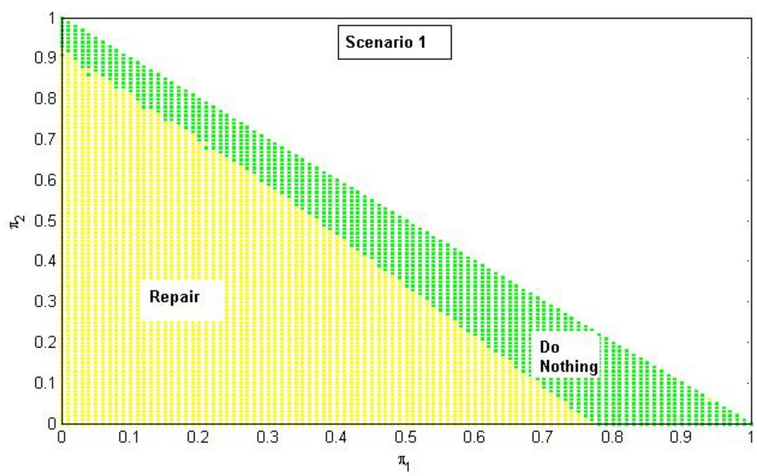


Figure 7.17 Optimal policy regions for t=3 under scenario 1

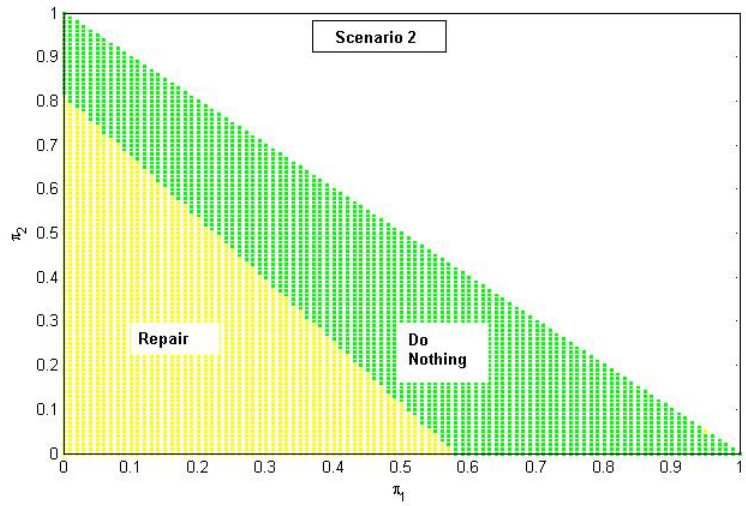


Figure 7.18 Optimal policy regions for t=3 under scenario 2

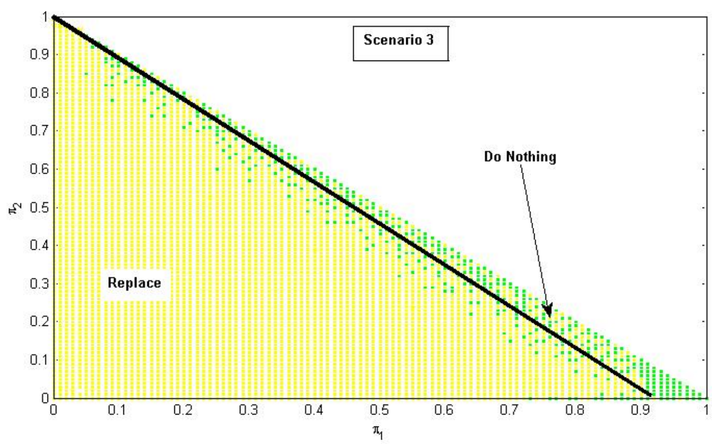


Figure 7.19 Optimal policy regions for t=3 under scenario 3

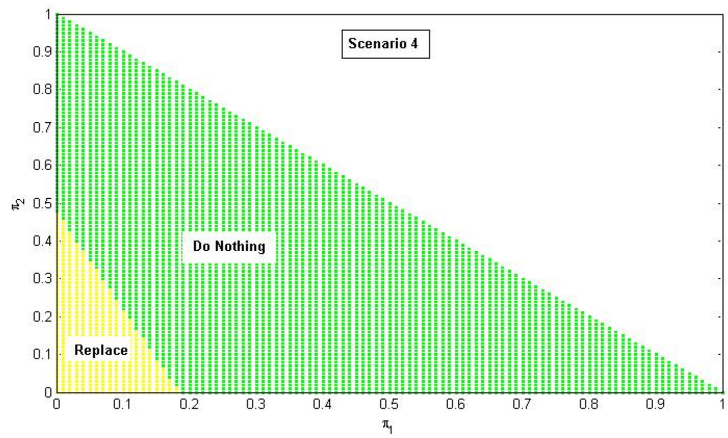


Figure 7.20 Optimal policy regions for t=3 under scenario 4

We notice the bigger difference between scenarios 3 and 4 in comparison to scenarios 1 and 2. Noticing also, that, scenarios 2 and 4 add the same amount of error to scenarios 1 and 3 respectively, with all other parameters of the problem fixed. This will guide us to the following result:

Result 1

For the no information case POMDP models, there is no effect of the measurement errors.

Proof

In practical sense, this result is trivial since there are no measurements and no measurements errors as well. But it is proved here mathematically for the POMDP model.

It can be easily verified that the Bayesian update given by equation 7.3, namely

$$T_j(\pi, a, y) = \frac{\sum_i \sum_x \pi_i p_{ij}^a \, r_{jx} r_{xy}^e}{\sum_i \sum_j \sum_x \pi_i p_{ij}^a r_{jx} r_{xy}^e}$$

Can be rewritten as follows:

$$T_j(\pi, a, y) = \frac{\sum_i \pi_i p_{ij}^a r'_{jy}}{\sum_i \sum_j \pi_i p_{ij}^a r'_{jy}}$$

Where, r'_{jy} is the jy element of the R' matrix, where:

$$R' = R \times R^e$$

Recalling that the POMDP with no information has an R matrix with identical rows. Then, for such an R matrix, with whatever the value of R^e the resulting R' will remain a matrix with identical rows i.e. a totally unobservable system as well. And the same Bayesian update will result regardless errors exist or not.

For an illustration, consider the following 3×3 state-observations matrix:

Let $R = \begin{bmatrix} a & b & 1-a-b \\ a & b & 1-a-b \\ a & b & 1-a-b \end{bmatrix}$

Then for any arbitrary $R^e = \begin{bmatrix} c & d & 1-c-d \\ e & f & 1-e-f \\ g & h & 1-g-h \end{bmatrix}$

$$R' = R \times R^e = \begin{bmatrix} a & b & 1-a-b \\ a & b & 1-a-b \\ a & b & 1-a-b \end{bmatrix} \times \begin{bmatrix} c & d & 1-c-d \\ e & f & 1-e-f \\ g & h & 1-g-h \end{bmatrix}$$

$$r'_{11} = r'_{21} = r'_{31} = ac + be + (1-a-b)g$$

$$r'_{12} = r'_{22} = r'_{32} = ad + bf + (1-a-b)h$$

$$r'_{13} = r'_{23} = r'_{33} = a(1-c-d) + b(1-e-f) + (1-a-b)(1-g-h)$$

Hence, the new system with error remains equivalent to the no information case and the optimal policy regions will not change since the Bayesian update will remain the same irrespective of the other problem parameters.

7.4 FUZZY ANALYSIS

This application deals with the scenario where, all output signals (Outcome products quality levels) are classified into conforming units or non-conforming based on an expert judgment, also, it will be assumed that the true state of the machine is actually related to the output as judged by the expert. We will consider the application where the state observation matrix is reduced into $n \times 2$. This is achieved by means of membership function inspired by the concept of Taguchi loss function concept. The reduced state observations matrix can then be used within the POMDP framework as usual. This will have the value of reducing the computation level drastically.

For the following state observation transition matrix, with columns are ordered from best quality output at the left to the worst at the right.

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{1m} \\ r_{21} & r_{22} & r_{23} & r_{2m} \\ r_{31} & r_{32} & r_{33} \ldots & r_{3m} \\ & \vdots & & \\ r_{n1} & r_{n2} & r_{n3} & r_{nm} \end{bmatrix}$$

The objective is to lump these columns into two columns corresponding to "good" and "bad" as outcomes. This can be done using the following membership functions.

$$\mu(conforming) = \left(\frac{r_{1x}}{\alpha_x}\right)^{\beta_x} \tag{7.6}$$

$$\mu(nonconforming) = \left(\frac{r_{1m}}{\alpha_{m-x+1}}\right)^{\beta_{m-x+1}} \tag{7.7}$$

Equation 7.6 gives the membership value of the elements to the set good for an outcome signal x, whereas, equation 2 gives the membership value for the set bad.

To better illustrate the concept of fuzzy observations consider the following example

7.5 NUMERICAL ILLUSTRATIONS FOR FUZZY OBSERVATIONS

This section provides some examples to illustrate the application of fuzzifying the observations matrix.

Example 7.3:

Consider the following state observations transition matrix that gives the conditional probabilities of receiving any of a 5 possible outcomes for any of a system three states:

$$R = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 & 0 \\ 0.7 & 0.1 & 0.1 & 0.1 & 0 \\ 0.1 & 0.1 & 0.2 & 0.3 & 0.3 \end{bmatrix}$$

Let $\beta_x = x$ and $\alpha_x = x$

Hence,

for r_{11}, $\mu(conforming) = \left(\frac{0.9}{1}\right)^1 = 0.9$

for r_{11}, $\mu(nonconforming) = \left(\frac{0.9}{5}\right)^5 = 0.000189$

for r_{12}, $\mu(conforming) = \left(\frac{0.1}{2}\right)^2 = 0.0025$

for r_{12}, $\mu(nonconforming) = \left(\frac{0.1}{4}\right)^5 = 0$

we proceed similarly for all the elements of the R matrix.

Now, in order to defuzzify the resulting discrete membership functions, we simply take the summation of all the membership values as follows:

$$d(conforming) = \left(\frac{r_{11}}{\alpha_1}\right)^{\beta_1} + \left(\frac{r_{12}}{\alpha_2}\right)^{\beta_2} + \cdots + \left(\frac{r_{1m}}{\alpha_m}\right)^{\beta_m} \quad (7.8)$$

$$d(nonconforming) = \left(\frac{r_{1m}}{\alpha_1}\right)^{\beta_1} + \left(\frac{r_{1m-1}}{\alpha_2}\right)^{\beta_2} + \cdots + \left(\frac{r_{11}}{\alpha_m}\right)^{\beta_m} \quad (7.9)$$

Finally to obtain the $n \times 2$ state observation transition matrix, the rows of the resulting conforming/nonconforming membership matrix are normalized such that the sum of every raw adds up to 1.

Using this membership function the total number of observations is reduced into within or off specifications. This is suitable for attribute type of produced products.

$$\mu(Conforming) = \begin{bmatrix} 0.9 & 0.0025 & 0 & 0 & 0 \\ 0.7 & 0.0025 & 0 & 0 & 0 \\ 0.1 & 0.0025 & 0.0003 & 0 & 0 \end{bmatrix}$$

$$\mu(nonconforming) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0025 & 0 \\ 0 & 0 & 0.0003 & 0.0025 & 0.3 \end{bmatrix}$$

$$d_1(Conforming/nonconforming) = \begin{bmatrix} 0.9025 & 0.00019 \\ 0.70254 & 0.00259 \\ 0.10283 & 0.3228 \end{bmatrix}$$

$$R'' = \begin{bmatrix} 0.99979 & 0.00021 \\ 0.99633 & 0.00367 \\ 0.24159 & 0.75841 \end{bmatrix}$$

Following is an example of a POMDP model, with two scenarios that differ only in the state observations transition matrices. Scenario 1 uses the R matrix whereas scenario 2 uses R''.

Example 7.4:

In this example, we have the same scenario of example 7.3, but with Let $\beta_x = 1$ and $\alpha_x = x$

Similar to the approach in the previous example we will have the following:

$$\mu(Conforming) = \begin{bmatrix} 0.9 & 0.05 & 0 & 0 & 0 \\ 0.7 & 0.05 & 0.03333 & 0.025 & 0 \\ 0.1 & 0.05 & 0.06667 & 0.075 & 0.06 \end{bmatrix}$$

$$\mu(nonconforming) = \begin{bmatrix} 0.18 & 0.025 & 0 & 0 & 0 \\ 0.14 & 0.025 & 0.03333 & 0.05 & 0 \\ 0.02 & 0.025 & 0.006667 & 0.15 & 0.3 \end{bmatrix}$$

$$d_1(Conforming/nonconforming) = \begin{bmatrix} 0.95 & 0.205 \\ 0.80833 & 0.24833 \\ 0.35167 & 0.56167 \end{bmatrix}$$

$$R'' = \begin{bmatrix} 0.82251 & 0.17749 \\ 0.76498 & 0.23502 \\ 0.38504 & 0.61496 \end{bmatrix}$$

Example 7.5:

Consider three states, three control actions POMDP model. The possible control actions are: do nothing, repair, and replace ($a_0, a_1,$ and a_2 respectively). Following are the reward criteria and the transition matrices for POMDP model.

$g(a_0) = [8\ \ 4\ \ 0]$

$g(a_1) = [5\ \ 3\ \ 1]$

$g(a_2) = [1\ \ 1\ \ 1]$

$$P^{a_0} = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.6 & 0.4 \\ 0 & 0.3 & 0.7 \end{bmatrix}, P^{a_1} = \begin{bmatrix} 0.8 & 0.2 & 0.0 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}, P^{a_2} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Regarding the state observations matrix, three scenarios are to be considered:

Scenario 1:

This is the base scenario, where, the matrix the original matrix of Example 7.3 is used as the state observations matrix

$$R = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 & 0 \\ 0.7 & 0.1 & 0.1 & 0.1 & 0 \\ 0.1 & 0.1 & 0.2 & 0.3 & 0.3 \end{bmatrix}$$

Scenario 2

The R matrix in scenario 1 is fuzzified as in example 7.3 as well.

$$R = \begin{bmatrix} 0.99979 & 0.00021 \\ 0.99633 & 0.00367 \\ 0.24159 & 0.75841 \end{bmatrix}$$

Scenario 3

The R matrix in scenario 7.3 is fuzzified as in example 7.4.

$$R = \begin{bmatrix} 0.82251 & 0.17749 \\ 0.76498 & 0.23502 \\ 0.38504 & 0.61496 \end{bmatrix}$$

Based on Scenarios 1, 2, and 3, the difference between the optimal policies regions is obvious. This difference is due to the fact that a POMDP model states are inferred statistically from the problem parameters.

Following are the optimal policy regions developed for these three scenarios over different time horizons:

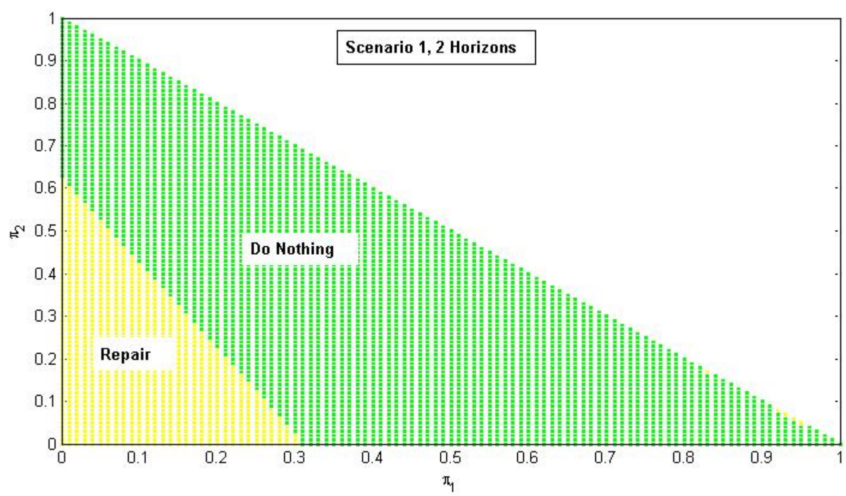


Figure 7.21 Optimal policy regions for t=2 under scenario 1

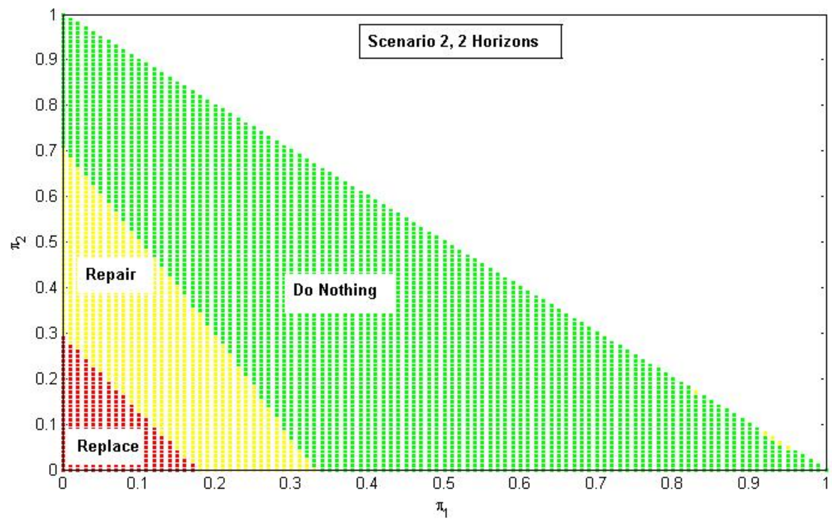


Figure 7.22 Optimal policy regions for t=2 under scenario 2

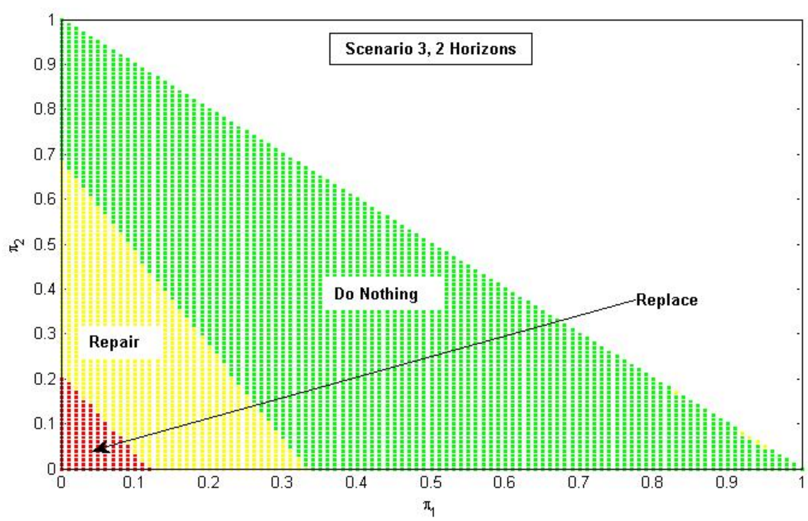


Figure 7.23 Optimal policy regions for t=2 under scenario 3

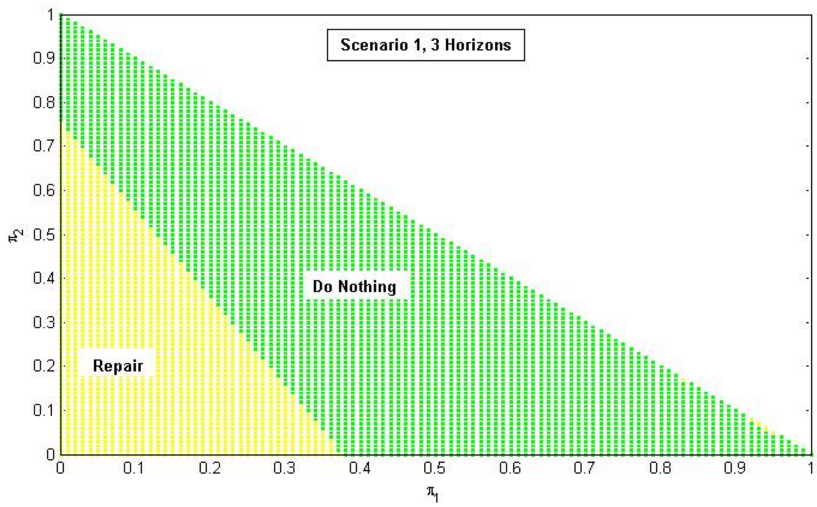


Figure 7.24 Optimal policy regions for t=3 under scenario 1

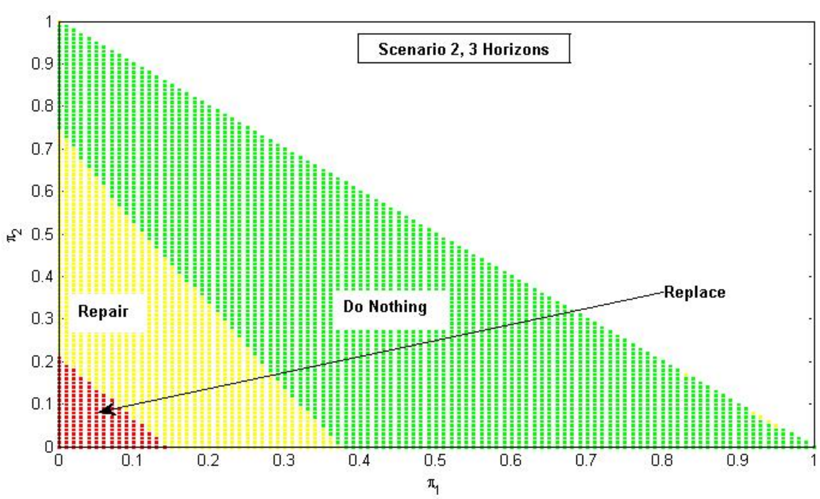


Figure 7.25 Optimal policy regions for t=3 under scenario 2

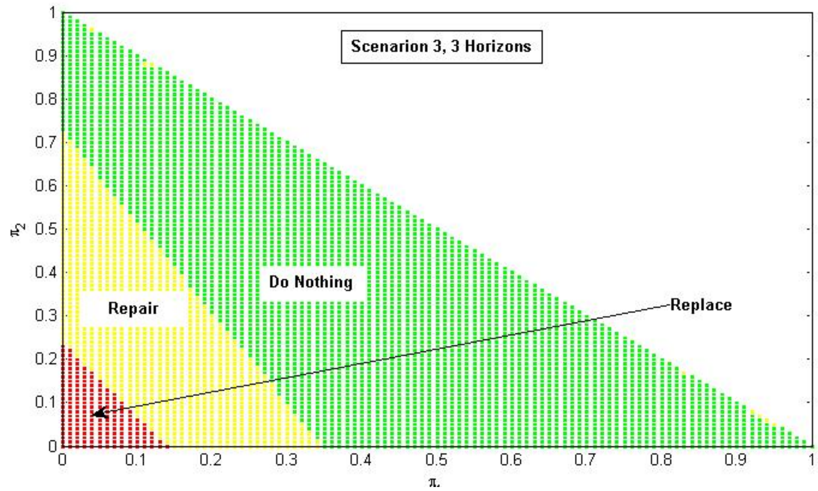


Figure 7.26 Optimal policy regions for t=3 under scenario 3

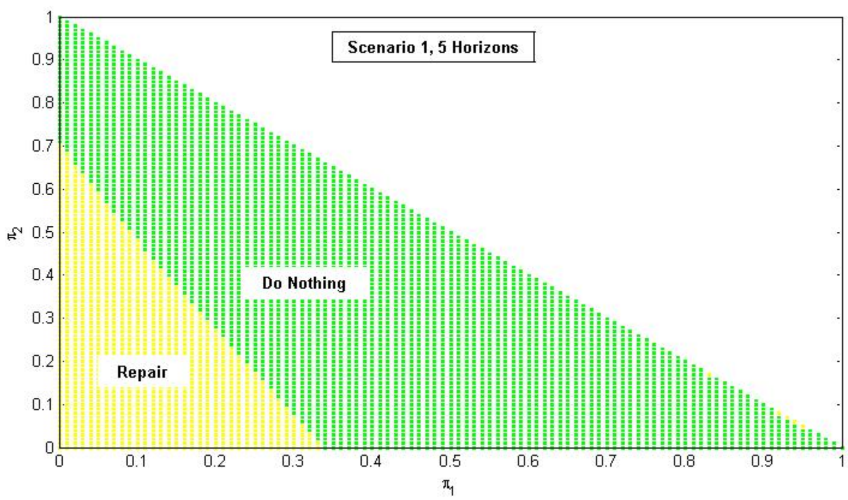


Figure 7.27 Optimal policy regions for t=5 under scenario 1

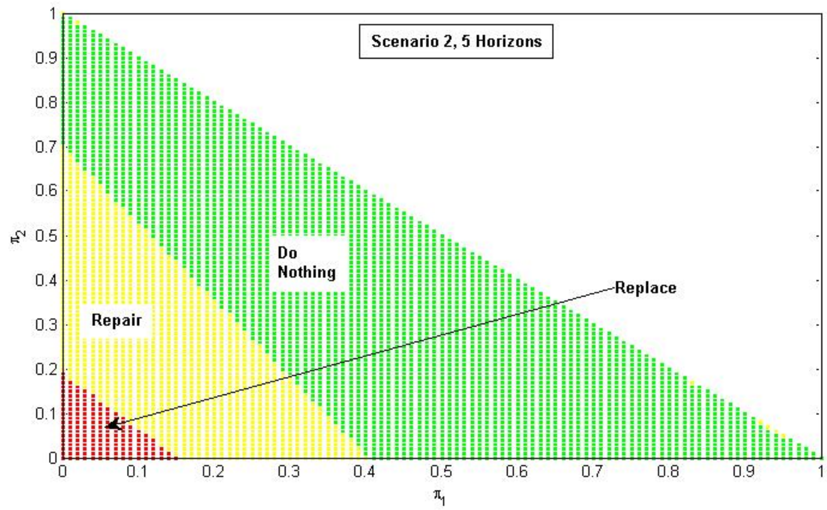


Figure 7.28 Optimal policy regions for t=5 under scenario 2

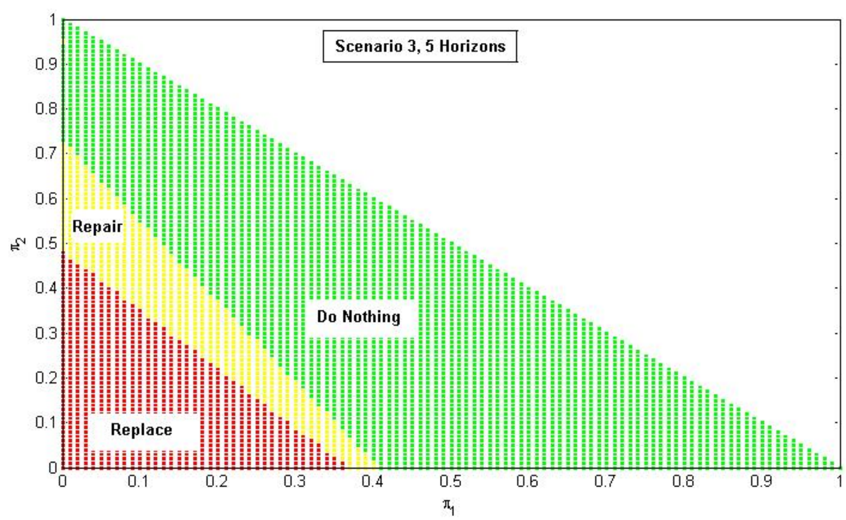


Figure 7.29 Optimal policy regions for t=5 under scenario 3

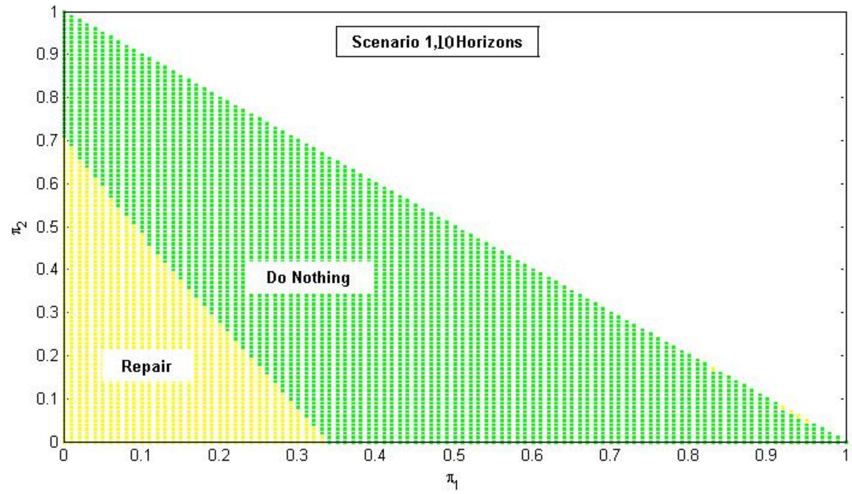


Figure 7.30 Optimal policy regions for t=10 under scenario 1

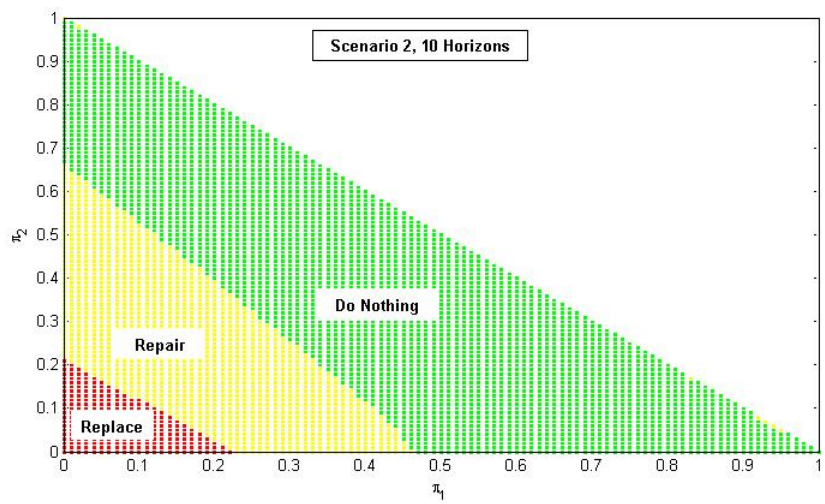


Figure 7.31 Optimal policy regions for t=10 under scenario 2

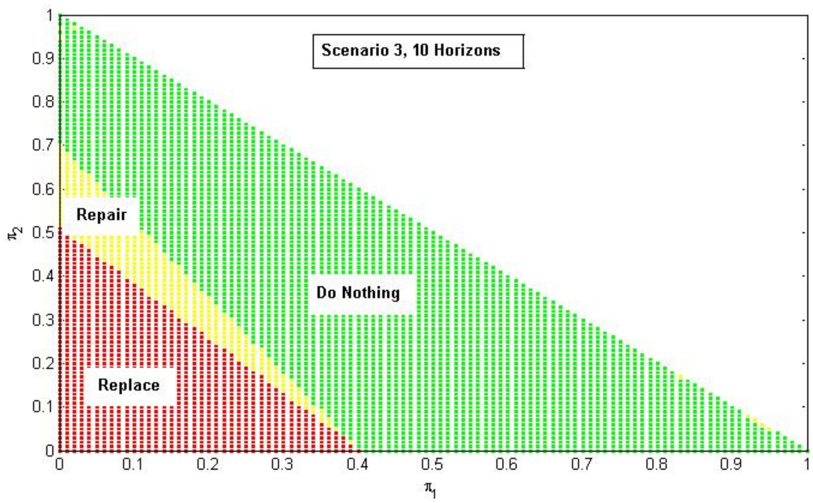


Figure 7.32 Optimal policy regions for t=10 under scenario 3

It is obvious from the different scenarios that the optimal cutoff points changed clearly due to the different policies. This indicates the importance of fuzzy logic approach when expert opinion is needed.

7.6 POMDP with Continuous Fuzzy Observations

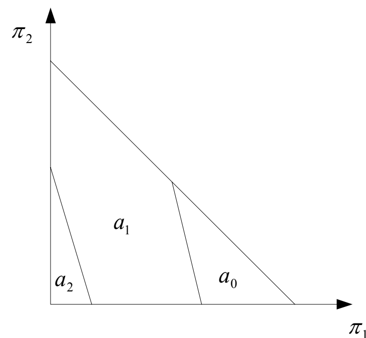
In this section a new technique for determining the optimal policy based on a fuzzy continuous observations received by the decision maker. To illustrate the concept the case of three possible fuzzy sets is discussed. Assume that the signal x received by the decision maker, say the temperature of the system on hand can be low, medium or high as follows:

$$x \in \{low\ (l) = 1, medium\ (m) = 2\ and\ high\ (h) = 3\}$$

Given an optimal policy simplex, say for a three state system for some horizon length t, the decision maker usually took the action based on the updated belief state. As discussed before, The Bayesian update is calculated using the following formula:

$$T_j(\pi, a, x) = \frac{\sum_i \pi_i p_{ij}^a r_{jl}^a}{\sum_i \sum_j \pi_i p_{ij}^a r_{jl}^a}$$

For three state systems, once the Bayesian update value is found the optimal action is determined, from a graph similar to the following as in the examples provided by Chapters 5 and 6.

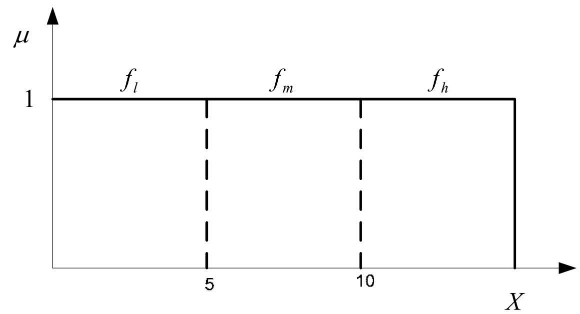


7.33 Optimal policy regions for a three state system

Usually in POMDP the observations are discrete. Hence, continuous scale observations are discretized. For instance, assume $x \in [0 - 15)$ and discretized uniformly over three equivalent intervals. We will have:

$$\begin{Bmatrix} x = l & if \;\; x \in [0,5) \\ x = m & if \;\; x \in [5,10) \\ x = h & if \;\; x \in [10,15) \end{Bmatrix}$$

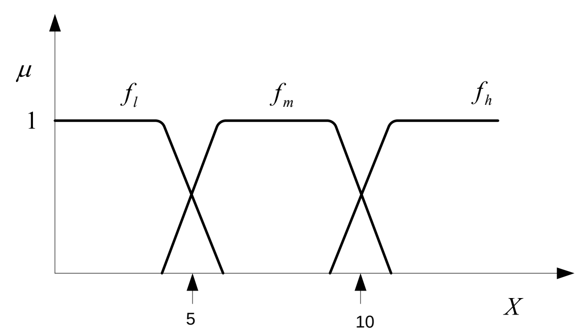
Let f_l be the fuzzy membership function for the fuzzy set low, f_m for the set medium and f_h for the set high. Thus in the fuzzy terminology, x is represented by the following crisp limits membership function.



7.34 Non-overlapping Fuzzy membership functions

But assume the readings 4.99 and 5.01 were obtained, in fact, there is no much difference to claim that 4.99 is low and 5.01 is medium as in, say, 0.5 and 8.5. Hence the concept of fuzzy observations is suggested to calculate the Bayesian update as follows:

Consider, for example the following membership function. Notice that for x= 4.99 and x=5.01, each of the observations has a membership value of low and medium = 0.5 for the low and medium observations which makes more sense.



7.35 Overlapping Fuzzy membership functions

Practically speaking, once the decision maker receives a signal from the system the Bayesian update can be calculated using the following equation:

$$\hat{T}_j(\pi, a, x) = \mu_{x=l} \times \frac{\sum_i \pi_i p_{ij}^a r_{jl}^a}{\sum_i \sum_j \pi_i p_{ij}^a r_{jl}^a} + \mu_{x=m} \times \frac{\sum_i \pi_i p_{ij}^a r_{jm}^a}{\sum_i \sum_j \pi_i p_{ij}^a r_{jm}^a} + \mu_{x=h} \times \frac{\sum_i \pi_i p_{ij}^a r_{jh}^a}{\sum_i \sum_j \pi_i p_{ij}^a r_{jh}^a} \quad (7.10)$$

Then we normalize to obtain a probability vector as follows:

$$T_j(\pi, a, x) = \hat{T}_j(\pi, a, x) / (\textstyle\sum_j \hat{T}_j(\pi, a, x)) \quad (7.11)$$

This approach is supposed to better evaluate the value of the belief state when the signal received by the decision maker is fuzzy signal.

7.7 CONCLUSION

In this chapter, measurement errors for the signals received from POMDP models hidden states is formulated. Sufficient Bayesian update is derived to correct for the effect of measurement errors. The significance of measurement errors has been discussed through illustrative examples. The concept of fuzzy observations is introduced for the POMDP framework. The case of classifying systems signals into conforming and nonconforming is provided assuming that the sets of conforming and nonconforming are fuzzy sets. This is done by fuzzifying the given state observations transition matrices. This can fit in applications where system true states relate better to the fuzzy output sets rather than the crisp ones usually used in the POMDP framework.

CHAPTER 8

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

8.1 SUMMARY

In this dissertation the problem of determining optimal maintenance policies for a multi-state, multi-stage deteriorating systems is addressed. The maintenance decision making problem of these systems is modeled as a POMDP. Optimal maintenance policies are characterized over the space of state occupancy vectors ordered by First order Stochastic Dominance (FSD). This has been achieved by developing a new set of conditions using different approach than that in the literature. The idea behind our approach is to let the Bayesian update of the problem follow the reverse hazard rate partial order. When used for this purpose, the reverse hazard rate partial order makes the FSD survives conditioning (notice that the reverse hazard implies the FSD). This was possible to do by developing new relations between the FSD and other partial orders (reverse hazard rate and component-wise dominance partial order). Sufficient conditions to guarantee our result of threshold optimal maintenance policies have been developed. If the conditions developed in this dissertation (Chapter 6) are satisfied then the result of optimal threshold maintenance policy follows for all the partial orders defined in Chapter 4. This is because all of these partial orders imply the FSD. On the other hand, our conditions provide an alternative set of new conditions compared to other results in the literature (see White (1978) and Lovejoy (1987)). Not only the developed conditions are useful for maintenance applications, but also it is useful for any application that can be modeled as a POMDP.

On the other hand, a model has been developed to maximize Overall Systems Effectiveness (OSE). This represents a combined productivity, quality and availability measure. The model provides maintenance and operation decision makers with a tool to develop and assess the tradeoff among alternative maintenance and production rate actions. This model provides an alternative criterion than the usual cost minimization/profit maximization models in the literature. Also this model has the advantage of being explicit in terms of the POMDP problem parameters and there is no need to estimate cost parameters.

Measurement errors in the context of POMDP models are modeled as a three-layer hidden Markov model, a Bayesian update is derived for the problem, where, it is shown to be a sufficient statistic. Also, the optimal value function for the three-layer hidden Markov model is shown to be piecewise linear convex one. Two real life examples are provided to emphasize the importance of the developed model, namely and briefly, the student learning process and quality measurement errors for POMDP systems. The relation between the quality of the signal received form the POMDP system and the effect of the measurement errors on the optimal policy regions is studied. It has been shown the less observable the system the less the effect of measurement errors. Hence, for totally unobserved systems it has been shown that there is no effect for measurement errors which sounds trivial.

The case of fuzzy observations is introduced to the POMDP framework. This is a typical scenario where relation between the system state and the output signal is better judged by an expert opinion. This is achieved by means of introducing a fuzzy membership function to fuzzify the available state observations matrix. The result of fuzzifying then defuzzifying and normalizing is a state observations matrix where the number of observations is only two, namely, good and bad. This reduces the amount of computation needed. Also, the significance of the expert decisions are reflected on the resulting optimal maintenance policies by some examples, see Chapter 7. Future research directions are provided next.

8.2 FUTURE RESEARCH DIRECTIONS

The following are suggestions for future research directions:

- For POMDP models, one possible future research directions can be characterizing the optimal maintenance policies over newly developed partial orders. This can be achieved by considering partial orders possessing certain features that facilitate the derivation of reasonable conditions on the problem parameters to ensure optimal threshold-type policies. For examples, the Monotone Probability Ratio and the Reverse Hazard partial orders discussed in Chapter 4.
- Another research direction is to consider developing optimal maintenance policies using non-Markovian models. The Markovian property is usually assumed in modeling complex multi-state systems due to its mathematical tractability. However, not all real systems possess this feature. Let's for example consider medical applications. History is a very essential component for medications decisions taken by doctors, for example, the disease history of the patients is usually kept track of. Even some times disease history is kept track of on the family level. Hence, future decisions depend on previous states of a patient as well as previous treatments he/she received.

- The more the number of states in the POMDP problem, the more the state transition probabilities that need estimation. This is really difficult to achieve in some real life scenarios. An efficient way of lumping the underlying system states might be a possible future research direction.
- In our developed models it has been assumed that all the transition matrices of the POMDP models are time homogeneous. It would be interesting to consider time-inhomogeneous POMDP reflecting more the actual dynamics of real world systems.
- Other quality measures can be a future direction of research within the POMDP framework. For example quality-maintenance integrated models can be studied where sample size and sampling costs or inspection plans are explicitly integrated in the model.
- Providing more applications to POMDP models is also a possible future research direction.
- Instead of presenting the optimal policy regions over the space of state occupancy vectors, it would be more convenient to provide the optimal policy regions over the observations (signals) directly, as well as, developing sufficient conditions to guarantee the existence of optimal threshold polices in terms of the observed signals of the POMDP.
- The effect of measurement errors can be further studied and represented by other distributions. This representation can be casted on the Markovian framework to study the impact of the error.
- The OSE model provided in Chapter 6 can be further enhanced by considering a

constrained version of the problem, for example adding constraints on the production speed to meet a certain demand forecast. And considering another maximization criterion like maximizing the expected OSE per unit time.

- A possible research direction is considering other possible fuzzy components within the POMDP framework. This may include fuzzy system states and fuzzy control actions. Also, continuous fuzzy observations can be considered and this implies the existence of an efficient solution algorithm because this will add more complexity to the problem.

APPENDIX A

Matlab Code for the Value Iteration Algorithm

```
clear;
clc;
ga1=[4 ,2.5 ,1];
ga2=[3 ,2.2 ,1];
ga3=[2.5 ,2.5 ,2.5];

%P=[.5 .5 0; .3 .2 .1 ; .3 0 .7];
P1=[.7 .2 0.1; 0 .6 .4 ; 0 0.6 .4];
P2=[.8 .2 0; .1 .6 0.3 ; .1 .6 .3];
P3=[1 0 0; 1 0 0 ; 1 0 0];

R=[0.7 0.3; 0.5 0.5; 0.3 0.7];

k=[1 2];

[v1]=vt1ai(ga1);
[v2]=vt1ai(ga2);
[v3]=vt1ai(ga3);


[vstar]= findvstar(v1,v2,v3);
vstarfinal=vstar;

vstarfinal_1=vstar;


for t=2:1:2
[v11]=updatev(vstarfinal_1, vstarfinal,v1, P1, R, k);
[v22]=updatev(vstarfinal_1,vstarfinal,v2, P2, R, k);
[v33]=updatev(vstarfinal_1,vstarfinal,v3, P3, R, k);

vstarfinal_1=vstarfinal;
vstarfinal=findvstar(v11,v22,v33);
[vpolicy]= findpolicymatrix(v11,v22,v33,vstarfinal);
end

vpolicy;
```

```
plotpolicysimplex(vpolicy)
function [v]= vt1ai(g)

i=1;
j=1;

for x2=0:0.01:1
    for y2=0:0.01:1-x2

v(i,j)=g(1)*x2+g(2)*y2+g(3)*(1-x2-y2);

pi(i)=x2;
fi(j)=y2;

j=j+1;
    end
    j=1;
    i=i+1;
end



function [w]=findvstar (v1,v2,v3)

v12 = max(v1,v2);
w=max(v12,v3);



function [vv1]=updatev(vstar_1, vstar,v1,P, R, k)

vstar_1=vstar_1;
vstar=vstar;
P=P;
R=R;
k=k;
v1=v1;

i=1;
j=1;
for x2=0:0.01:1
    for y2=0:0.01:1-x2

pi(i)=x2;
fi(j)=y2;

[rec]= recours(vstar_1, vstar,pi(i),fi(j),P, R, k);
```

```
vv1(i,j)=v1(j,i)+rec;


rec;
v1(i,j);

j=j+1;
    end
    j=1;
    i=i+1;
end

vv1=vv1';



function [tttt]= bayesianupdte(xx1,xx2,P, R, kk)



PI=[xx1 xx2 1-xx1-xx2];


Rcolumn=[R(1,kk) R(2,kk) R(3,kk)];

denomm=PI*P*Rcolumn';
numerr=PI*P.*Rcolumn;
tttt=numerr/denomm;



function
[betsactionmatrix]=findpolicymatrix(v1,v2,v3,vstar);

v1=v1;
v2=v2;
v3=v3;
vstar=vstar;

for i=1:1:101
    for j=1:1:102-i

        if (v1(i,j)>=vstar(i,j))
  betsactionmatrix(i,j)=1;
        elseif(v2(i,j)>=vstar(i,j))
```

```
    betsactionmatrix(i,j)=2;
          else
    betsactionmatrix(i,j)=3;
          end
      end
  end

  function []=plotpolicysimplex(vpolicy)
  l=1;
          m=1;

  for i=0:.01:1
      for j=0:.01:1


          if vpolicy (l,m)==0

plot(i,j,'square','MarkerFaceColor','white','MarkerEdgeColo
r', 'w','MarkerSize',3)
                   hold on
                elseif vpolicy (l,m)==1

plot(i,j,'square','MarkerFaceColor','green','MarkerEdgeColo
r', 'w','MarkerSize',3)
                    hold on
                elseif vpolicy (l,m)==2

plot(i,j,'square','MarkerFaceColor','yellow','MarkerEdgeCol
or', 'w','MarkerSize',3)
                    hold on
                else
plot(i,j,'square','MarkerFaceColor','red','MarkerEdgeColor'
, 'w','MarkerSize',3)
                   hold on
                end
  m=m+1;
      end
      m=1;
      l=l+1;
  end
  hold off
```

REFERENCES

AlDurgam, M., Duffuaa, S. (2009), "Maximizing Overall System Effectiveness (OSE) for Three-State, Partially Observable System", *Proceedings of the Third International Conference on Modeling, Simulation and Applied Optimization,* Sharjah,U.A.E January 20-22.

Albright, S. C., 1979, "Structural Results for Partially Observable Markov Decision Process", *Operations Research*, Vol. 27, No. 5, 1041-1053.

Banerjee, P.K., Rahim, M.A. (1988), "Economic design of $\bar{x}$ control charts under Weibull shock models", *Technometrics*, Vol. 30 No.4, pp.407-14.

Cao, Y., Sun, H., Trivedi, K., Han, J. (2002), "System Availability with Non-exponentially Distributed Outages", *IEEE Transactions on Reliability,* Vol. 51, No. 2, pp. 193-198.

Ben-Daya, M., Rahim, M., 2001, "Integrated Production, Quality & Maintenance Models: an Overview", in M. Rahim and M. Ben-Daya (eds), Integrated models in production planning, inventory, quality, and maintenance. Kluwer Academic Publishers, 3-28.

Budai, G., Dekker, R., Nicolai, R., 2006, "A Review of Planning Models for Maintenance & Production", *Econometric Institute Report* 2006-44.

Case, Kenneth E., Bennett, G.Kemble, "Economic Effect of Measurement Error on Variables Acceptance Sampling", *International Journal of Production Research*, Vol. 15, Issue 2, March 1977, Pages 117-128

Cassandra, A.,1998, "A Survey of POMDP Applications", Microelectronics and Computer Technology Corporation, Austin, TX. Available online (Last time accessed Feb, 2009): http://www.pomdp.org/pomdp/papers/applications.pdf

Chen, D., Trivedi, K. (2001), "Analysis of Periodic Preventive Maintenance with General System Failure Distribution", *In Proc. 2001 Pacific Rim Int. Symp. Dependable Computing*, Seoul, Korea, 103–110.

Derman, C., 1962, "On Sequential Decisions and Markov Chains", *Management Science*, Vol. 9, No. 1, 16–24.

Duffuaa, S.O. (1996), Impact of inspection errors on performance measures of a complete repeat inspection plan, *International Journal of Production Research*. Vol. 34, Issue 7, Pages 2035-2049.

Duffuaa, S.O. and Khan, M. (2005), "Impact of inspection errors on the performance measures of a general repeat inspection plan", *International Journal of Production*

Research, Vol. 43, Issue 23, Pages 4945-4967

Duffuaa, S.O., Siddiqui, A.W. (2003), "Process targeting with multi-class screening and measurement error", *International Journal of Production Research*. Vol. 41, Issue 7, pp. 1373-1391.

Duncan, A.J. (1956), "The economic design of $\bar{x}$ charts used to maintain current control of a process", *Journal of American Statistical Association*, Vol. 51 No.274, pp.228-42.

Eeckhoudt, L., Gollier, C., 1995, "Demand for Risky Assets and the Monotone Probability Ratio Order", *Journal of Risk and Uncertainty*, Vol. 11, 113-122.

Grosfeld-Nir, A., 1996, "A Two-State Partially Observable Markov Decision Process with Uniformly Distributed Observations", *Operations Research*, Vol. 44, No. 3, pp. 458-463.

Hong, S.H. and Elsayed, E.A. (1999), "The optimum mean for processes with normally distributed measurement error", *Journal of Quality Technology,* Vol. 31, Issue. 3, pp. 338-344.

Hopp, W., Wu, S., 1990, "Machine Miantenance with Multiple Maintenance Actions", *IIE Transactions,* Vol. 22, No. 3, pp. 226-233.

Hopkins, E., Kornienko, T., 2007, "Cross and Double Cross: Comparative Statics in First Price Auctions", *The B.E. Journal of Theoretical Economics,* Vol. 7, Iss. 1, Topics, Article 19.

Hughes, J., 1978, "A Note on Quality control under Markovian deterioration", *Operations Research*, Vol. 28, No. 2, 421-424.

Hunter, W. G. and Kartha, C.P. (1977), "Determining The Most Profitable Target Value For A Production Process**",** *Journal of Quality Technology*, Vol. 9, No. 4, pp. 176-181.

Ivy, J. S., 1998 "Maintenance and Replacement Policies for a Multi-State Deteriorating Process with Probabilistic Monitoring", *PhD Dissertation*, University of Michigan Ann Arbor.

Ivy, J., Pollack, S., 2005, "Marginally Monotonic Maintenance Policies for a Multi-state Deteriorating Machine with Probabilistic Monitoring, and Silent Failures", *IEEE Transactions on Reliability*, Vol. 54, Iss. 3, pp. 489-497.

Jaraiedi, Majid, Herrin, Gary D. (1985), "Effect of Human Inspector Error on Sample Plan Design", Proceedings - Fall Industrial Engineering Conference (Institute of Industrial Engineers), pp. 436-439.

Kim, C. E. and Gen, M. (1993), “Replacement policy for a partially observable markov decision process model using fuzzy data”, *Computers and industrial engineering*. Vol. 25, Issue, 1-4. pp. 435-438.

Kim, M.J. and Makis, V. (2009), “Optimal maintenance policy for a multi-state deteriorating system with two types of failures under general repair”, *computers and Industrial Engineering*, Article in press.

Kuo, Y. (2006), “Optimal Adaptive Control Policy for Joint Machine Maintenance and Product Quality Control”, *European Journal of Operational Research*, Vol. 171, pp. 586-597.

Kolesar, P., 1966, “Minimum Cost Replacement under Markovian Deterioration”, *Management Science*, Vol. 12, No. 9, 694–706.

Krishnamurthy, V., Djonin, D. V., 2007, “Structured Threshold Policies for Dynamic Sensor Scheduling—A Partially Observed Markov Decision Process Approach”, *IEEE Transactions on Signal Processing*, Vol. 55, Iss. 10, pp. 4938-4957.

Kuo, Y. (2006), “Optimal Adaptive Control Policy for Joint Machine Maintenance and Product Quality Control”, *European Journal of Operational Research*, Vol. 171, pp. 586-597.

Lovejoy, W., 1987, “Some Monotonicity Results for Partially Observed Markov Decision Process”, *Operations Research*, Vol. 35, No. 5, pp. 736-743.

Lovejoy, W., 1991, “A survey of algorithmic methods for partially observed Markov decision processes”, *Annals of Operations Research*, Vol. 28, No. 1, pp. 47-65.

Maillart, L., 2005, “Optimal Observation and Preventive Maintenance Schedules for Partially Observed Multi-State Deterioration Systems with obvious Failures”, *PhD Dissertation*, University of Michigan.

Maillart, L., 2006, “Maintenance policies for systems with condition monitoring and obvious failures”, *IIE Transactions*, Vol.38, No.6, pp. 463-475.

McCall, J.J., 1965, “Maintenance Policies for Stochastically Failing Equipment: a Survey”, *Management Science*, Vol. 11, 493-525.

Monahan, G., 1982, “A Survey of Partially Observable Markov Decision Processes: Theory, Models, and Algorithms”, *Management Science*, Vol. 28, No. 1, pp. 1-16.

Nakajima, S. (1988), “Introduction to TPM: Total Productive Maintenance”, *Productivity Press, Cambridge, MA*.

Panagiotidou, S. and Tagaras, G. (2007), “Optimal preventive maintenance for equipment with two quality states and general failure time distributions”, *European Journal of Operational Research* Vol. 180, Issue 1, pp. 329-353

Pierskalla, W.P., Vollker, J.E., 1976, “A Survey of Maintenance Models: the Control and Surveillance of Deteriorating Systems”, *Naval Research Logistics Quarterly*, Vol. 23, 353-388.

Puterman, M., 2005, “Markov Decision Process, Discrete Stochastic Dynamic Programming”, John Wiley & Sons Inc.

Rahim, M.A. (1989), " Determination of optimal design parameters of joint $\bar{x}$ and R charts ", *Journal of Quality Technology*, Vol. 21, No.1, pp.65-70.

Rahim, M.A. and Banerjee, P.K. (1993), "A generalized model for the economic design of $\bar{x}$-control charts for production system with increasing failure rate and early replacement", *Naval Research Logistics*, Vol. 40 No.6, pp. 787-809.

Rosenfield, D., 1976, “Markovian Deterioration with Uncertain Information”, *Operations Research*, Vol. 24, No. 1, 141-155.

Ross, S., 1971, “Quality control under Markovian deterioration”, *Management Science*, Vol. 17, No. 9, 587-596.

Smallwood, R. and Sondik, E., 1973, “The Optimal Control of Partially Observable Markov Processes Over a Finite Horizon”, *Operations Research*, Vol. 21, No. 5, pp. 1071-1088.

Stoyan, D., 1983, “Comparison Methods for Queues and Other Stochastic Models”, John Wiley & Sons, New York.

Striebel, C., 1965, “Sufficient Statistics in the Control of Stochastic Systems”, *J. Math. Anal*. Appl., Vol. 12, pp. 576-592.

Virtanen, I. (1977), “Generalized Availability Characteristics for Systems with States of Reduced Efficiency”, *Lawrence Symposium on Systems and Decision Sciences*, Berkeley (Calif.), USA, October 3-4.

Valdez-Flores, C., Feldman, R.M., 1989, “A Survey of Preventive Maintenance Models for Stochastically Deteriorating Single-unit Systems”, *Naval Research Logistics* (1989), Vol. 36, pp. 419–446.

Wang, H., 2002, “A Survey of Maintenance Policies of Deteriorating systems”, *European Journal of Operational Research*, Vol. 139, Iss. 3, 469–489.

White, C., 1978, “Optimal Inspection and Repair of a Production Process Subject to

Deterioration”, *journal of the operational research society,* Vol. 29, 235-243.

White, C., 1979 a, “Optimal Control-limit Strategies for a Partially Observed Replacement Problem”, *Int. J. Syst. Sci.*, Vol. 10, No. 3, pp. 321-331.

White, C., 1979 b, “Bounds on Optimal Cost for a replacement Problem with Partial Observation”, *Naval Research Logistics*, Vol. 26, pp. 415-422.

White, C., 1980 a, “Monotone Control Laws for Noisy, Countable-State Markov Chains”, *European Journal of Operational Research*, Vol. 5, pp. 124-132.

White, C., 1980 b, “Structured Policy Results for Single Stage Decisionmaking Under Uncertainty”, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-10, No. 12, pp. 891-894.

White, C., 1991, “A survey of solution techniques for the partially observed Markov decision process”, *Annals of Operations Research*, Vol. 32, No. 1, pp. 215-230.

Mohammad AlDurgam is a PhD candidate and Lecturer-B in the Systems Engineering (SE) Department, King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia. He obtained his B.Sc. and M.Sc. degrees from the University of Jordan, Amman-Jordan, in Industrial Engineering in 2002 and 2005, respectively. His research interests include stochastic systems modeling, decision making under uncertainty, optimization using metaheuristics and simulation applications in manufacturing. He has approached many Jordanian Industrial Companies and participated in several projects and case studies. Mohammad covers lab lectures of the simulation course for the undergraduates of the SE department in the KFUPM using ARENA^{TM} software.