PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES WITH APPLICATIONS

BY

DALE J. HOCKSTRA

TECHNICAL REPORT NO. 156
SEPTEMBER 28, 1973

SUPPORTED BY THE ARMY AND NAVY
UNDER CONTRACT NO0014-67-A-0112-0052 (NR-042-002)
WITH THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted for any purpose of the United States Government
Distribution of this Document is Unlimited

DEPARTMENT OF OPERATIONS RESEARCH
AND
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES WITH APPLICATIONS

by

DALE J. HOCKSTRA

TECHNICAL REPORT NO. 156

September 28, 1973

SUPPORTED BY THE ARMY AND NAVY
UNDER CONTRACT N00014-67-A-0112-0052 (NR-042-002)
WITH THE OFFICE OF NAVAL RESEARCH

Gerald J. Lieberman, Project Director

Reproduction in Whole or in Part is Permitted
for any Purpose of the United States Government

Distribution of this document is unlimited

DEPARTMENT OF OPERATIONS RESEARCH
and
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
NON TECHNICAL SUMMARY

This paper examines a class of partially observable Markov decision processes with applications to the general problem of diagnosis and treatment. Included are a broad formulation, basic results on the existence of optimal policies, general characteristics of the optimal policy, and more precise descriptions of the optimal policy for two special cases. The reader interested in computational algorithms should also refer to the paper by Sondik [24].

Consider a system which may be in any one of several states. At least one of the states is desirable, e.g., a smoothly operating machine or a healthy person. However, other of the states are undesirable. For example, the machine may have various types of breakdowns and the human can contract various diseases. Unfortunately, it is often the case that the state of the system is not directly observable, i.e., the decision maker can not simply look at the system and ascertain which state it is in.

The decision maker would like to have the system in the desirable state. He has various actions available to further this goal. These consist of "tests" which give him information about the state of the system and "treatments" which will restore the system from some undesirable state to the desirable one if properly applied. Each of these actions has potential costs and benefits associated with it. For example, disassembly of some component of a machine involves wages for a repairman,
lost production time, the potential of restoring the machine to perfect condition, and the possibility that nothing was wrong with the component resulting in wasted effort. The problem is to determine the optimal sequence of actions to be performed. In other words, what actions should be taken and in what order so as to maximize the expected return?

The problem would be fairly simple if the state of the system were directly observable. One would apply the appropriate "treatment" for the situation and thus restore the system to the desirable state. Unfortunately, as indicated above, this is often not the case. Hence this paper assumes that the decision maker only knows the chance that the system is in a given state, i.e., a probability distribution over the possible states of the system. The decision maker must decide on the best action with this partial information. As a result of his action, he gains further information about the state of the system which he then uses to revise his estimates of the probabilities of the possible states. This is formally accomplished by using Bayes Theorem.

The other major feature of the model is that it allows the state of nature to be dynamic. In other words, the underlying state of the system may change during the course of the decision process. As an example, a patient with the flu might get pneumonia. This is formally modeled via a Markov chain.

The end result is a model which is a discrete time Markov decision process with a continuous state space, a finite action space, and a special transition structure. To keep the expected reward finite it is assumed that there is a positive probability of termination at
any stage. Under reasonable assumptions, it can be shown that the region where a given "treatment" is optimal is star-shaped at an appropriate corner of the state space and that the region where a given "test" is optimal is halo-shaped at the appropriate corner.

This paper was originally motivated by the problem of medical diagnosis and treatment. As a result, it is formulated in terms of that problem. However, the person interested in other applications needs only to replace a few key words to read the paper. For example, "disease" would be replaced by "state of nature" and "doctor" would be replaced by "decision maker". Section 7.2 discusses the transition to the machine maintenance problem in some detail.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>INTRODUCTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.1 The Problem</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.2 Literature Survey</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1.3 The Relevance of Medical Models</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>1.4 Summary of the Dissertation</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>FORMULATION AND EXISTENCE.</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2.1 Basic Information Structure</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2.2 Formulation as a Markov Decision Process.</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>2.3 Critique of the Formulation -- Observation of Termination.</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>2.4 Critique of the Formulation -- The State of Health.</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>2.5 Existence of an Optimal Policy.</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>2.6 Classes of Policies</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>2.7 Randomized Actions.</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>BINARY STRUCTURE</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>3.1 The Problem with Analysis of the General Model</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>3.2 Binary Models</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>3.3 The Scaling Theorem</td>
<td>49</td>
</tr>
<tr>
<td>4</td>
<td>MODEL I.</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>4.1 Introduction.</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>4.2 The Transition Structure.</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>4.3 The Reward Structure.</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>4.4 The Optimal Policy for Model I.</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>4.5 The Case ( w = 0 )</td>
<td>77</td>
</tr>
<tr>
<td>5</td>
<td>MODEL II</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>5.1 Formulation</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>5.2 The Optimal Policy for Model II</td>
<td>83</td>
</tr>
<tr>
<td>6</td>
<td>LEXICOGRAPHIC IMPROVEMENT.</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>6.1 The Conjecture.</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>6.2 Results</td>
<td>93</td>
</tr>
<tr>
<td>7</td>
<td>CONCLUSION</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>7.1 Summary</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>7.2 Application to Machine Maintenance.</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>7.3 Further Research.</td>
<td>106</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>PAGE</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>A SYMBOLS</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>B EXAMPLE OF NON-CONVEXITY</td>
<td>114</td>
<td></td>
</tr>
<tr>
<td>C ADJACENCY RESULTS</td>
<td>119</td>
<td></td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>124</td>
<td></td>
</tr>
</tbody>
</table>
### LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>DESCRIPTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chain of Events</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>Observation of Termination</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>Structure of $T^n_i$</td>
<td>54</td>
</tr>
<tr>
<td>4</td>
<td>Conjectured Optimal Policy</td>
<td>66</td>
</tr>
<tr>
<td>5</td>
<td>Optimal Policy for Model I</td>
<td>66</td>
</tr>
<tr>
<td>6</td>
<td>Structure of $H^n_i$</td>
<td>69</td>
</tr>
<tr>
<td>7</td>
<td>Optimal Policy for Model II</td>
<td>91</td>
</tr>
<tr>
<td>8</td>
<td>Lexicographic Improvement</td>
<td>94</td>
</tr>
<tr>
<td>9</td>
<td>Parallel Correspondence</td>
<td>94</td>
</tr>
<tr>
<td>10a</td>
<td>Hyperplanes on $e_i e_k$</td>
<td>96</td>
</tr>
<tr>
<td>10b</td>
<td>Hyperplanes on $z_k$</td>
<td>96</td>
</tr>
<tr>
<td>11</td>
<td>Lexicographic Situation</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>Non-Convexity</td>
<td>118</td>
</tr>
<tr>
<td>13</td>
<td>Illustration of Theorem I</td>
<td>122</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1. The Problem.

One hundred years ago the practice of medicine was a relatively simple affair. The stereotyped image of the country doctor with his black bag riding in a horse drawn buggy to sit by the bedside of the patient still remains. In fact it is a very appealing alternative to the chromium gadgetry, antiseptic smell and hurried impersonality which seems to characterize medicine today. And yet modern medicine could hardly be otherwise. The explosion of knowledge and technology which has characterized all the sciences in recent years certainly applies to medicine. Today's physician is blessed with a much greater understanding of disease mechanisms than his predecessor of a century ago. He has at his disposal a myriad of diagnostic tests and therapeutic procedures unknown in 1873.

It is doubtful that anyone would claim medicine of 1873 vintage to be superior to modern day practice. At the same time the complexity of modern medicine has brought with it the common attendant problems. There is a tremendous volume of potential information about the patient. Which information should be collected, and once collected, how should it be used? What is the relationship between the individual patient and the larger medical environment, e.g., the hospitals or clinics? Is there a tradeoff between spiraling costs and the associated benefits? On a larger spectrum there is presently controversy as to whether good medical care is a privilege of those who can afford it or whether
it is a basic human right. With the increasing popularity of the later view there has been greater governmental involvement (e.g., Medicare) and consequently, greater demand for medical services. This leads to questions such as the design of regional medical programs and the efficacy of multi-phasic testing procedures. These in turn have implications for the individual patient-doctor encounter.

As a result, at least partially, of these considerations there has been an increase in recent years in the application of analytical techniques to questions of medical care. One area which has been studied is the process of diagnosis and treatment at the individual level, i.e., between a single patient and his physician. Several different models of diagnosis and treatment have been proposed; these are summarized and briefly discussed in the next section.

Even at this level of specificity, the problem of diagnosis and treatment is complex and multi-faceted. For example, some diseases are very well defined entities in the sense that there are precise physiological conditions which describe them, e.g., mitral insufficiency is a heart condition where a certain valve fails to close completely. Other diseases, however, are very nebulous entities. For example, systemic lupus erythematosus is defined as the presence of any four out of fourteen manifestations and may vary greatly from patient to patient [1, p. 645]. No model covers all facets of diagnosis and treatment and it is doubtful there will be one which does, at least in the near future.

A survey of the available models, however, reveals that most of them are static in the sense that they assume the patient's state
remains unchanged, at least during the diagnostic process. It seems clear, however, that diseases change over time, either with respect to the disease entity present, or the symptoms it exhibits. In particular, a given disease may exhibit different degrees of severity, which may require different treatments. Hence, from a practical point of view, a single disease can be subdivided into several distinct medical entities. This paper attempts to take this facet of diagnosis and treatment into account. Specifically, a class of models is developed which allow the state of the patient to change according to a stochastic process during the course of diagnosis and treatment. The model is a terminating Markov decision process.

It should be noted from the outset that the model developed here concentrates on the dynamic facet of diagnosis and treatment. It extends the available analytic results in this area. However, it is not fully general; such a model would not permit the derivation of interesting results and demand more information than is medically available.

1.2. Literature Survey.

If one were to isolate a single germinal paper on the analytical analysis of diagnosis and treatment it would probably be the 1959 article of Ledley and Lusted [17]. Since that point there have been numerous models proposed; they can be roughly categorized as (1) logical flow models, (2) static Bayesian models, (3) pattern recognition models, and (4) sequential decision models. For a slightly different classification and summary of the literature through 1968 refer also to the survey article by Ledley [16].
Logical flow models attempt to mimic the decision process of
the physician in a rigorous, logically spelled out manner. Generally
they consist of some sort of flow chart which gives the next action
to be taken depending on the outcomes to previous actions. Fries [9]
has applied the method to the diagnosis of arthritic diseases while
Tufo and Burger [21] have proposed its use as an efficient method of
utilizing paramedical personnel in a military clinic. Fries noted a
96% rate of agreement with the clinical diagnosis [9, pp. 650] indicating
that the method can work quite successfully. One problem with the flow
chart method is that of determining the logical order in which to
perform tests and treatments. Also, once determined, there is no
opportunity for individualization to a specific patient.

Perhaps the most widely applied method is the static Bayesian
model. These models generally apply only to diagnosis. It is assumed
that the patient has one of a given collection of diseases, usually
including the possibility of being healthy. The physician knows the
conditional probabilities of observing various symptoms given the diseases
and also the prior probability distribution of the possible diseases,
usually their incidence rate in the population. A fixed block of
symptom information is collected on the patient. Bayes theorem is then
used to determine the posterior distribution of the diseases given the
patient symptoms. This is offered as a differential diagnosis for the
the method to congenital heart disease. Since then it has been applied
to the diagnosis of bone tumors [18], thyroid function [19], and liver
disease [4]. The rate of correct diagnosis has generally matched that
of an expert clinician [26, pg. 367]. Problems with the method include
the difficulty of achieving probabilistic independence in the medical
world [28] and the fact that the same information is collected on all
patients. There is also a tendency to "overdiagnose" in the sense that
the posterior probabilities are very close to zero or one [14].

Pattern recognition techniques essentially amount to defining
symptom patterns associated with various diseases, often called symptom
complexes, via separating surfaces, usually hyperplanes, in some multi-
dimensional space. Specht [25] has applied the method to the analysis
of electrocardiograms while Kulikowski [15] has studied the diagnosis
of hyperthyroidism. Success rates on the order of 90% correct diagnoses
were observed by Kulikowski [15, pg. 177]. The method seems to be most
successful when used as a screening device among a small number of
subsets of a population. For more diseases the number of patterns
increases rapidly, especially since there are often several symptom
complexes associated with a given disease.

The final category of model proposed for diagnosis and treatment
is the sequential decision model. This area can be further subdivided
into those models which are decision analytic in flavor versus those
is perhaps the best example to date of the former group. He analyzes
the problem of pleural effusion with respect to both diagnosis and
treatment. Ginsberg and Offensend [12] apply decision analysis to the
problem of a child with a difficult spinal disorder.

The models on the dynamic programming side select the actions to
be performed, tests and treatments, in a sequential fashion. They are
generally based on Bayes theorem as the static Bayesian models are, only now the information collected on a patient is selected according to the needs of the individual patient. Gorry and Barnett [13] use a heuristic dynamic programming scheme which looks only one period in the future to diagnose congenital heart disease. Using the same data as Warner, et al. [30] they were able to achieve the same level of diagnosis with substantially fewer tests per patient [13, pg. 504].

Hockstra and Miller [14] utilized a similar scheme to diagnose valvular heart disease with the modification that the next test was selected using a game theoretic procedure allowing for multiple value systems. Sondik [24] and Smallwood and Sondik [23] formulate a model similar to the one proposed in this paper but only mention its application to medicine. Differences between their model and the one presented here will be indicated where appropriate.

The present model belongs to the sequential decision category. It differs from the previous models in that a dynamic rather than a static patient is assumed. Mathematically the proposed model is a Markov decision process; Derman's book [6] gives a good summary. Sondik [24, p. 7] presents a summary of the theoretical literature on partially observable Markov decision processes; his paper should be included therein. Finally, a paper whose results on the form of the optimal policy are similar in spirit to those given here is Ross [22].
1.3. The Relevance of Medical Models.

One feature shared by all the models of the preceding section is the fact that they are normative rather than descriptive. They do not attempt to discover the actual thought processes of the physician during the course of patient management. At best these would differ substantially from one doctor to another. Rather the models attempt to define what a "rational" man ought to do given the information and alternatives available. They provide a standardized, consistent approach to a class of problems faced by a physician.

Hence these models should be viewed as a tool of the physician, not as a replacement for him. The notion of the computer replacing the physician sounds very much like the inhuman utopia of Orwell's 1984, and is at best a questionable goal. The use of a computer model as a diagnostic/therapeutic agent in the absence of a doctor has been tried on a very limited scale. For example, the shipboard computer of the Japanese supertanker Seiko Maru has a diagnostic/therapeutic routine for use by crew members [2]. Obviously, however, this is a special case.

What then is the function of models of diagnosis and treatment? There are several possibilities. First, the model can be used as another tool just as the X-ray machine or electrocardiogram is. It is not too far fetched to envision the day when the physician has a computer terminal in his office. In difficult cases, or simply as a check on his own decisions, the doctor could use the model as a consultant, a source of additional information. Unfortunately the presently existing models are not yet sufficient for this task. There will also be a moderately large amount of resistance on the part of medical people to this suggestion; doctors are a notoriously conservative group.
A second possible function of a model of diagnosis and treatment is as a research tool. Because of their mathematical nature almost all of the models require data in a precise, standardized form. This has forced a reevaluation of medical knowledge which has lead to some new insights in the field of medicine. For example, in attempting to define symptoms associated with systemic lupus erythematosus it was noticed that there were two distinct groups of patients suggesting two different physiological entities [8]. The need for precise data has also stimulated development of improved, systematic medical record keeping [10].

Finally, these models can be used as a valuable pedagogical device. Patients can be hypothetically generated by a simple simulation scheme. The student is then asked to diagnose and recommend treatment for the simulated patient. He can instantaneously compare his actions with those recommended by the model. This can provide a first or second year medical student with practice in clinical decision making on a scale heretofore impossible. It cannot replace actual experience in the ward but it can contribute to the prospective physician's education substantially, just as flight simulators are valuable for pilots.

Realistically, models of diagnosis and treatment will probably have their major use in research and teaching, rather than in private practice, at least for the next few years. This should not be interpreted as indicating that they have minimal value, however. It merely indicates the relationship between the medical and mathematical worlds. While mathematics in general and operations research in particular have made great strides in recent years, it would be remiss to argue that any of the models of diagnosis and treatment, or all of them combined, capture all of the important complexities of medicine. The model proposed
and analyzed here, as well as those which preceded it, narrows the gap but does not close it. In the process, however, more is learned about mathematics and medicine, and the end product is not without use. It seems that if operations research is going to address itself to some of today's significant problems this is the inevitable state of affairs.


This dissertation is divided into seven chapters. Chapter 2 presents the basic theoretical framework along with existence results and those results on the form of the optimal policy which follow without any additional structural assumptions. Also included are critiques of two of the major assumptions of the formulation. Chapter 3 then develops the major assumption which allows stronger results on the form of the optimal policy to be verified.

Based on this, Chapters 4 and 5 analyze two specific models and derive similar results for each. The major difference between the two cases is the status of the "healthy" state. Chapter 4 assumes that it is an observable, terminating state while Chapter 5 views it as a non-observable, absorbing state. Chapter 6 considers a problem common to both of the specific models. The question studied, however, is tangential to the main flow of the thesis and hence Chapter 6 may be omitted by the reader without loss of continuity. Chapter 7 is a summary along with potential applications in non-medical areas and suggestions for further research. There are also three Appendices.
Appendix A is a symbol list, B gives a computed example of non-convexity and C gives some partial results extending the basic policy characterization.
CHAPTER 2
FORMULATION AND EXISTENCE

2.1. Basic Information Structure.

Consider the following model of medical diagnosis and treatment. Suppose there is a finite set $D$ of possible states indexed by $i$. Included in $D$ are various diseases and other medically related patient states, for example health and death. Assume $D$ has $M$ elements, i.e., $D = \{1, 2, \ldots, M\}$. There is a second finite set $A$ of actions comprised possibly of (a) tests giving information about the patient's unknown disease, (b) treatments for various diseases, (c) sending the patient home, or (d) observing him (doing nothing for some period of time). $A$ may also include combinations of individual actions, e.g., a compound testing procedure. Let $a \in A$ be an arbitrary action. For each $a \in A$ there is a set $\Theta(a)$ of possible outcomes to action $a$. Assume that each $\Theta(a)$ is a finite set and denote an arbitrary outcome in $\Theta(a)$ by the index $j$. As an example, action $a$ could represent taking a throat culture and the outcomes could be the presence or absence of a certain bacteria. The assumption of a finite number of outcomes for each action is, in some cases, a simplification of medical reality. For example, blood pressure is a continuous variable which must be discretized to be incorporated in this framework; e.g., low, normal, or high. Little is lost in practice by this as very fine discretizations may be made.

Assume there is a prior probability distribution giving some guess at the chance the patient has the various possible diseases.
Let $P_t(i)$ be the probability of state $i \in D$ at time $t$. This may simply be the proportion of the population which exhibits disease $i$ or it may be an opinion by the physician based on specific, previously gathered information related to the patient.

Further assume that the conditional probabilities $P(j|i, a)$ are known, i.e., the likelihood of observing outcome $j \in \Theta(a)$ if the patient is in state $i$ and action $a$ is taken. This is more or less the way medical knowledge is presented in the medical literature, though quite often not in exact numerical form [8].

Consider the following scenario. Action $a$ is taken at time $t$. At some later time, $t+1$, the outcome $j \in \Theta(a)$ becomes known. This gives further information about the patient's state; information valid at the time, $t$, action $a$ was taken. This can be expressed using Bayes Theorem

$$P_t(i|j, a) = \frac{P(j|i, a) P_t(i)}{\sum_{k \in D} P(j|k, a) P_t(k)}.$$

Verbally, $P_t(i|j, a)$ is the probability that the patient was in state $i$ at time $t$, given that action $a$ was taken and outcome $j \in \Theta(a)$ was observed.

However, $j$ is observed at time $t+1$ and not at time $t$.

During the intervening interval the patient's state may spontaneously change. Assume that this change occurs according to a Markov chain which may depend on the action taken. Let $P_{i\ell}(a)$ represent the probability that the patient's state changes from $\ell$ to $i$ during the time required for action $a$. Then the probability that the patient
is in state \( i \) at time \( t+1 \) after observing outcome \( j \) to action \( a \) as

\[
(2) \quad P_{t+1}(i|j, a) = \sum_{\ell \in D} p_{\ell i}(a) P_t(\ell|j, a) .
\]

Combining equations (1) and (2) yields

\[
(3) \quad P_{t+1}(i|j, a) = \frac{\sum_{\ell \in D} p_{\ell i}(a) P(j|\ell, a) P_t(\ell)}{\sum_{k \in D} P(j|k, a) P_t(k)}.
\]

This is the appropriate probability of state \( i \) at time \( t+1 \), i.e., the time \( j \in \Theta(a) \) is observed.

Given the outcome \( j \), the posterior probabilities of all the states in \( D \) are desired. To facilitate this, matrix notion is introduced. Let \( x \in \mathbb{R}^M \) be the vector of state probabilities, \( x = (P_t(1), \ldots, P_t(M)) \) so that \( x_i \geq 0 \) and \( \sum_{i=1}^M x_i = 1 \). The time dependence of \( x \) has been suppressed so \( x_i \) refers to the \( i \)th component of \( x \) in the usual sense. Treat \( x \) as a column vector. Define

\[
(4) \quad z_{k\ell}(j) = p_{\ell k}(a) P(j|\ell, a)
\]

and define the matrix \( Z(j) \) by

\[
(5) \quad (Z(j))_{k\ell} = z_{k\ell}(j) .
\]
Note that the dependence of $Z$ on action $a$ has been suppressed since this can be thought of as being included in $j$. This is done to keep the notation manageable later. The vector of posterior probabilities depends on the outcome $j \in \Theta(a)$. It is denoted by $x(j)$ and can be written

$$ x(j) = \frac{Z(j)\mathbf{x}}{P(j)} \tag{6} $$

where

$$ P(j) = \sum_{k \in D} P(j \mid k, a)x_k. \tag{7} $$

$x(j)$ is the vector of state probabilities which is the best guess at the patient's state at time $t+1$, i.e., after taking action $a \in A$ and observing $j \in \Theta(a)$. The time subscript, $t+1$, has also been suppressed from $x(j)$ to simplify notation.

$Z(j)$ is not a stochastic matrix, nor even a substochastic one as can easily be shown by example.

2.2. **Formulation as a Markov Decision Process.**

Let $S = \{x \mid x \in \mathbb{R}^M, x_i \geq 0, \sum x_i = 1\}$. Then equation (6) describes a transition scheme between elements of $S$, i.e., between probability distributions over the states in $D$. Further, the transition from $x$ at time $t$ to $x(j)$ at time $t+1$ occurs with probability $P(j)$ as given in (7). This suggests a Markov decision process model of medical diagnosis and treatment.
It was assumed above that the patient's state could change spontaneously during the course of diagnosis and treatment. It is also possible that the decision process stops at some point. The patient might become healthy, either by accident or design. He could die, or go to another doctor, or the physician could decide that the patient is healthy and send him home. In any case, the decision process ceases for our purposes. To model this, include within the set $D$ a number of terminating states. Formally, partition $D$ into two subsets $D_1$ and $D_2$ where $D_1 = \{1, 2, \ldots, N\}$ and $D_2 = \{N+1, N+2, \ldots, M\}$. $D_1$ represents the set of actual diseases while $D_2$ consists of the terminating states. For example, $D_2$ could consist of the states health and death as $N+1$ and $N+2$ respectively. It should be noted at this point that health could either be in $D_1$ or in $D_2$; Section 2.4 addresses this question.

Assume that the states in $D_2$ are absorbing in the probabilistic sense so that in terms of the spontaneous Markov transition mechanism,

$$p_{ii}(a) = 1; \quad i \in D_2, a \in A.$$ 

It has been assumed that the patient's disease cannot be observed directly; instead an outcome is observed which yields information about the patient's state. It seems, however, that cessation of the decision process ought to be observable. There are several possibilities which result in stopping the process. The physician may simply choose to stop, the patient may spontaneously enter one of the terminating states in $D_2$, or an outcome may be observed which would indicate that further
choices are either irrelevant or predetermined. For example, diagnosis of the disease as a result of some test would likely put a halt to further testing. Similarly, successful treatment would end the process.

To model this in a manner consistent with the previous analysis, partition the set of outcomes $\theta(a)$ into two disjoint subsets, $\theta_C(a)$ and $\theta_T(a)$ as follows. $\theta_C(a)$ consists of those outcomes to action a which yield information regarding the patient's disease and do not imply cessation of the decision process. $\theta_T(a)$ contains those outcomes which imply that the decision process stops. $\theta_T(a)$ is further subdivided into two subsets $\theta_{T1}(a)$ and $\theta_{T2}(a)$. $\theta_{T1}(a)$ consists of outcomes which yield information regarding the patient's disease and stop the process. $\theta_{T2}(a)$ is that set of outcomes which imply that the patient has spontaneously entered one of the terminating states in $D_2$. For example, failure of a treatment would be in $\theta_C(a)$, success of the treatment would be in $\theta_{T1}(a)$, and the information that the patient died would be in $\theta_{T2}(a)$. In essence this says that the decision process continues if $j \in \theta_C(a)$ is observed and stops if $j \in \theta_T(a)$ is observed. Termination, even to an absorbing state in $D_2$, is determined by observation of an outcome in $\theta(a)$ rather than by direct observation. This is consistent with the basic information structure of Section 2.1. However, as a consequence, termination to a state in $D_2$ is not instantaneously known to the decision maker. This problem is further laid out in the scenario of events described below and critiqued in Section 2.3.

Assume that decision points are scattered discretely through time, i.e., represent the diagnosis and treatment process by a series of decision points. Since the transition scheme depends upon the action
and outcome it need not be the case that these decision points are equally spaced in real time. From a medical point of view they probably are not. It is assumed that the time required to perform action \( a \) is known and that \( p_{k_1}(a) \) takes this into account. Thus time \( t \) represents some decision point and time \( t+1 \) represents the following decision point. The process proceeds as follows. At time \( t \) the decision process is in state \( x \), the physician's state of knowledge, and some action \( a \in A \) is chosen. This incurs some immediate cost \( r(x,a) \). \( r(x,a) \) is the expected cost of performing action \( a \) if the state is \( x \). If \( r_1(a) \) is the cost associated with performing action \( a \) when the patient is in state \( i \in D \), then

\[
(9) \quad r(x,a) = \sum_{i=1}^{M} r_i(a) x_i = r(a)x
\]

is the expected cost of performing action \( a \) given state \( x \). \( r(a) \) is a row vector. Note that the expected cost is a linear function of \( x \).

The physician can not see the actual cost \( r_1(a) \) incurred for if he did he would have additional information regarding the patient's state. Hence assume that the costs and rewards are recorded somewhere and observed only at the end of the decision process. There are some practical difficulties involving utility theory in determining \( r_1(a) \); see Section 7.3 for a brief discussion. These are circumvented in Chapters 4 and 5 by assuming that the cost depends only on the action.

At this point there are two possibilities, either the patient was in one of the disease states in \( D_1 \) at time \( t \) or he was in one of the terminating states in \( D_2 \). If the former is the case, some outcome \( j \in \theta_C(a) \cup \theta_T(a) \) is observed at a later time \( t+1 \). Since \( \theta_T(a) \)
is the set of outcomes associated with terminating states it is impossible to observe \( j \in \Theta_{T_2}(a) \). To formalize this let

\[
(10) \quad P(j|i, a) = 0 ; \quad j \in \Theta_{T_2}(a), \ i \in D_1, \ a \in A.
\]

Hence suppose \( j \in \Theta_C(a) \). This gives further information about the patient's disease so the state is updated via equation (6). Since further decisions are required a new action \( b \in A \) may now be chosen and the process repeats. On the other hand, suppose \( j \in \Theta_{T_1}(a) \). Again this gives further information about the patient's disease so the state is updated via equation (6). In addition the information is sufficient to make further decisions fruitless, hence stopping the decision process. A reward of \( U(j) x(j) \) accrues where \( U(j) \in \mathbb{R}^M \) and is considered a row vector. Again if \( U_i(j) \) is the reward or penalty associated with observing outcome \( j \in \Theta_{T_1}(a) \) if the patient has disease \( i \), then the expected reward is

\[
\sum_{i=1}^{M} U_i(j) x_i(j) = U(j)' x(j).
\]

Notice that the expected reward is a linear function of the updated state. The linearity is a result of the physician's state of knowledge, namely \( x \), and is not an additional assumption. Hence the cost and reward scheme is quite general even though it has a lot of structure. The specific models of Chapters 4 and 5 illustrate this well.

The other major possibility is that the patient was in one of the terminating states, \( D_{2'} \), at time \( t \). In this case an outcome \( j \in \Theta_{T_2}(a) \) occurs immediately after action \( a \) is selected. Hence there is no intervening time period as in the case of disease states. No outcome
\( j \in \theta_C(a) \cup \theta_{T1}(a) \) can occur. Hence assume

\[
(11) \quad P(j \mid i, a) = 0 ; \quad j \in \theta_C(a) \cup \theta_{T1}(a) , \ i \in D_2 , \ a \in A .
\]

Further assume that \( \theta_{T2}(a) \) is independent of \( a \in A \), i.e., \( \theta_{T2}(a) = \theta_{T2} \), and consists of exactly one outcome for each of the terminating states in \( D_2 \). Observation of \( j \in \theta_{T2} \) indicates the patient is in state \( N+j \in D_2 \). Hence for \( j \in \theta_{T2} \) define

\[
(12) \quad P(j \mid N+j , a) = 1 ; \quad a \in A
\]

\[
P(j \mid i, a) = 0 ; \quad i \neq j , \ a \in A .
\]

In other words, outcome \( j \in \theta_{T2} \) will be observed if and only if the patient is in state \( N+j \in D_2 \). Suppose that \( e_k \) is the \( M \) vector whose \( k \)th component is one with zeroes elsewhere. Then observation of \( j \in \theta_{T2} \) implies that the posterior distribution is \( e_{N+j} \).

**Lemma 1:** If \( x_{N+j} > 0 \) then \( x(j) = e_{N+j} \) for \( j \in \theta_{T2} \).

**Proof:** Note that \( j \in \theta_{T2} \) is possible only if \( x_{N+j} > 0 \) for if \( x_{N+j} = 0 \) then \( P(j) = 0 \). Now

\[
x(j) = \frac{Z(j)x}{P(j)} = \frac{Z(j)x}{\sum_{k=1}^{M} P(j \mid k,a)x_k} = \frac{Z(j)x}{x_{N+j}} .
\]

But \( z_{k\ell}(j) = p_{\ell k}(a) P(j \mid \ell, a) = 0 \) for \( \ell \neq N+j \) since \( P(j \mid \ell, a) = 0 \). Also, \( z_{k,N+j}(j) = p_{N+j,k}(a) = 0 \) for \( k \neq N+j \) since state \( N+j \) is
absorbing. Hence the only element of $Z(j)$ which is non-zero is $e_{N+j,i}^{N+j} - 1$. Thus

$$x(j) = \frac{Z(j)x}{x_{N+j}} = \frac{(0, \ldots, 0, x_{N+j}, 0, \ldots, 0)^c}{x_{N+j}} = e_{N+j}. \quad \|$$

Observation of $j \in \theta_{T2}$ stops the decision process and, as in the case where $j \in \theta_{T1}(a)$, there is an expected reward/penalty of $U(j) x(j)$. Since $x(j) = e_{N+j}$ the expected reward is $U(j) x(j) = U(j) e_{N+j} = U_{N+j}(j)$. This is effectively the same as a scalar return of $U_{N+j}(j)$; vector notation is maintained to emphasize the linear form and facilitate certain proofs. If, for example, $N+j$ represents death then $U_{N+j}(j)$ is the penalty for allowing the patient to die.

The entire chain of events is summarized in Figure 1.

---

Figure 1: Chain of Events
The reward structure proposed may seem arbitrary at this point. As pointed out above, however, it is non-restrictive and will allow for the pursuit of a variety of goals. In Chapters 4 and 5 these are basically avoidance of death and restoration of the patient to health. By eliminating the reward for health one is left with a model which essentially maximizes the probability of the patient living. Other considerations could include certainty of diagnosis, dollar cost, or pain involved.

2.3. Critique of the Formulation -- Observation of Termination.

A consequence of the hypothesized chain of events is that spontaneous transition to one of the terminating states is not observed instantaneously upon its occurrence. Suppose action \( a \) is chosen at time \( t \) and that termination to \( N+k \in D_2 \) occurs during the period between \( t \) and \( t+1 \). At time \( t+1 \), \( j \in \theta_C(a) \cup \Theta_{T1}(a) \) is observed because this output, though observed at \( t+1 \), is determined at \( t \). Hence at \( t+1 \) the state of the process is \( x(j) \). If \( j \in \theta_C(a) \) a new action \( b \in A \) is selected at time \( t+1 \). Immediately afterwards \( k \in \Theta_{T2} \) is observed indicating that the patient is in the terminating state \( N+k \). In reality, \( b \) would never even be performed. If, on the other hand, \( j \in \Theta_{T1}(a) \) the decision process stops anyhow. Figure 2 illustrates the situation graphically.

The major difficulty is the fact that a new action \( b \) has to be selected before the output \( k \in \Theta_{T2} \) indicates that the patient has entered state \( N+k \). Medically this implies a scenario similar to the following. The doctor orders action \( a \) during morning rounds (time \( t \)).
Figure 2: Observation of Termination

The action, say some test, is performed and then during the day the patient dies. Before evening sounds (time \( t+1 \)) the doctor receives output \( j \in \theta_C(a) \cup \theta_{T1}(a) \) which either stops the process or prompts him to select \( b \in A \), say a treatment. He writes up the order for this and gives it to the nurse at the beginning of evening rounds. She promptly informs him that the patient has died.

Unquestionably this is stretched from reality. One would expect the attending physician to be informed immediately if the patient were to die. Why not assume that if the patient makes a transition to a terminating state during the period the doctor is immediately informed via \( j \in \theta_{T2} \)? If no such transition occurs then \( j \in \theta_C(a) \cup \theta_{T1}(a) \) is
observed at the end of the period. The problem with this is that it results in a dynamic programming recursion which is non-linear in \( x \). This arises because the optimal \( n-1 \) period reward at \( x(j) \) is multiplied by both the probability of observing \( j \) and the probability of non-termination during the period. Each of these is linear in \( x \); the end result is not. This would not harm the existence results but, as will be seen, almost all the results on the form of the optimal policy depend on a linear recursion.

A possible remedy would be to divide the decision period into two parts. During the first part the patient either makes a transition to a terminating state which is immediately observed or he remains in the same disease state. If he has not terminated then an action is selected, output observed and state updated during the second half of the period. Only transitions to other disease states would be allowed during the second part of the period. This would probably allow immediate observation of termination while maintaining linearity in the recursion. It is questionable, though, whether this is any more realistic than the proposed model.

On the positive side, the present model may be interpreted in a reasonable fashion. Recall that \( k \in \theta_{12} \) is observed immediately after selecting action \( b \) so the delay can be made arbitrarily short. Every action in \( A \) can be thought of as including a test for termination. Immediately upon selection of an action the patient is tested to determine whether or not he is in a terminating state. If so, then some \( k \in \theta_{12} \) is the immediate output. If not, the rest of the action is performed and the appropriate output is observed later. This is
consistent with the assumption that all of the real time events have been compressed into decision points.

Further, the reward structure can be defined in a way which yields an appropriate reward stream. Refer to Figure 2 and suppose \( j \in \varnothing_C(a) \). Then action \( b \) is selected resulting immediately in \( k \in \varnothing_{T2} \). As indicated above, \( b \) would realistically never be performed, and yet \( r(x(j), b) \) accrues. Recall that \( r_i(b) \) was defined as the immediate cost of performing action \( b \) if the patient was in state \( i \). Define \( r_i(b) = 0 \) for all \( i \in D_2 \) and all \( b \in A \). In this way there is no cost for observing \( k \in \varnothing_{T2} \) and determining that the patient is in state \( N+k \). Returnwise this is the same as never performing action \( b \).

Similarly suppose \( j \in \varnothing_{T1}(a) \). Then the process terminates, but without the physician knowing that the patient is in state \( N+k \). However \( U(j) x(j) \) can be defined in such a way as to maintain a very reasonable reward stream. See Chapter 4 for one way this can be worked out in detail.

2.4. Critique of the Formulation -- The State of Health.

A second problem with the formulation of Section 2.2 is the status of the healthy state. If it is assumed that health is one of the terminating states in \( D_2 \) then it must be an absorbing state and there is a single output which is its indicator. The first conclusion is reasonable since the model is designed to consider the course of a single illness rather than a lifetime of medical problems. The second implication,
however, is problematical. There is no medical test which indicates that a person is healthy. Health may be essentially defined as the absence of disease. When a doctor pronounces someone healthy it means that the tests performed and observations made did not indicate any illness, though it is still possible the patient is sick. Clearly, health is not precisely defined from an operative point of view. Two people may both be healthy and yet appear different.

The problem then is whether the healthy state should be considered a terminating state and hence grouped among such obviously observable states as death and going to another doctor or whether it should fall into some other category. The specific model of Chapter 4 assumes that health is a terminating state. Chapter 5, on the other hand, treats health as an absorbing disease state and hence as not directly observable. This is appealing since there is no requirement that there be an output in $\Theta T_2$ indicating that the patient is well. Further, one of the possible actions is to terminate the diagnosis and treatment process, i.e., send the patient home. Intuitively, one would expect this to happen in the real world if the patient were healthy.

A problem with this formulation, however, involves the interpretation of a sufficient condition for the existence of a return function and optimal policy. Basically one assumes that there is a strictly positive probability that the patient enters one of the terminating states from any disease state. If health is one of the terminating states then this is very reasonable, since there is usually some chance that the patient will get well. Suppose on the other hand, health is an absorbing disease state. From a medical point of view it may be
more difficult to argue that a terminating state is reachable from every disease state. Further, it is impossible to reach any terminating state from the healthy state, thus violating the assumption of a positive termination probability. From a practical point of view, however, one could assume that there is a very small chance of termination from the healthy state without altering the results.

2.5. **Existence of an Optimal Policy**

The purpose of this section is to show that an optimal policy exists for the general model of Section 2.2, and further, that it is stationary. Since the state space $S$ is a continuum there are several possible ways one might proceed.

The fixed point, contraction map arguments of Denardo [5] apply but they suffer from the fact that only stationary policies are considered. In general, a policy $\pi$ is a sequence of functions $\delta_i : S \rightarrow A$, $i = 1, 2, \ldots$. $\delta_1(x)$ is the action to be taken if the process is in state $x$ in period 1, $\delta_2(x)$ is the action to be taken if in state $x$ in period 2, etc. A stationary policy is one such that $\delta^i = \delta$ for $i = 1, 2, \ldots$. Clearly, a stationary policy is desirable from a practical point of view as a non-stationary policy would be quite difficult to use. One could simply assume that since only stationary policies are practical, the optimum among the class of stationary policies is sufficient. Then Denardo's arguments imply the existence of a stationary optimal policy.

A more general approach would be to use the "transient policy condition" of Veinott [29]. A policy is transient if the series of
n-fold transition matrices converges [29, p. 1636], implying that the expected reward associated with a transient policy is finite. Veinott proves that if every stationary policy is transient then there is a stationary policy which maximizes the expected return over all policies. Unfortunately, his paper only analyzes the finite state case, i.e., $S$ a finite set. The model of this paper has a finite set of underlying states but due to the partial observability, the state space $S$ must be considered as a simplex, and hence non-finite. It is not clear whether the transient policy condition extends to the case of $S$ infinite.

A third possible approach is that of Blackwell [3]. His paper allows for more general state and action spaces, but has a more restrictive formulation in that a discount factor is used. His Theorem 7b [3, p. 234] shows that as long as $A$ is finite there is an optimal stationary policy, even though $S$ is infinite.

In order to be able to apply Blackwell's result the remainder of this section shows that under certain termination conditions, the general model of Section 2.2 behaves like a discounted process. More precisely, the expected return for any policy is shown to be bounded above by a convergent series which has a form essentially identical to that of a standard discounted process. As a trivial consequence it will also be obvious that the expected reward of any stationary policy is finite, i.e., the transient policy condition essentially holds.

Assume that there is a positive probability of entering some terminating state in $D_2$ from any other state. Precisely,
One could use a more general $N$-stage assumption, i.e., that there is a positive probability of falling into $D_2$ after $N$ stages. This permits a slight increase in realism at the expense of complexity. Equation (13) implies that the decision process terminates in finite time with probability one, which from a medical point of view, is obvious.

Let $\pi = (\delta^1, \delta^2, \ldots)$ be any policy. If the process is in state $x$ in period $n$ and $\delta^n$ is followed the immediate expected return is $\bar{r}(x, \delta^n)$ where

$$\bar{r}(x, \delta^n) = r(\delta^n(x)) \cdot x + \sum_{j \in \Theta_T(a)} U(j) x(j) P(j).$$

Hence the total expected reward associated with the policy $\pi$ is

$$V_\pi(x) = \bar{r}(x, \delta^1) + \sum_{n=1}^{\infty} \left\{ \sum_{j^1 \in \Theta_C(\delta^1)} \sum_{j^2 \in \Theta_C(\delta^2)} P(j^1) P(j^2) \right\} \cdots \sum_{j^n \in \Theta_C(\delta^n)} P(j^n) \bar{r}(x(j^n), \delta^{n+1}).$$

Note that, for brevity, $j^n \in \Theta_C(\delta^n)$ is used rather than $j^n \in \Theta(\delta^n(x))$. The problem is to show that this is essentially the same as the expected reward for a discounted process.
Let

\[ \alpha_n(x, \pi) = \sum_{j^n \in \Theta_C(\delta^n)} p(j^n) \leq 1. \]

Note that \( \alpha_n \) depends on \( x \) and on the sequence of actions and outcomes which occur. Assume \( \alpha_n(x, \pi) > 0 \). If \( \alpha_n(x, \pi) = 0 \) then \( V_\pi(x) \) is a finite series and hence trivially convergent. Then define

\[ (15) \quad \bar{V}(j^n) = \frac{p(j^n)}{\alpha_n(x, \pi)} . \]

Thus \( \sum_{j^n \in \Theta_C(\delta^n)} \bar{V}(j^n) = 1 \). Substituting in (14),

\[ (16) \quad V_\pi(x) = \bar{r}(x, \delta^1) + \sum_{n=1}^{\infty} \left\{ \sum_{j^1 \in \Theta_C(\delta^1)} \alpha_1(x, \pi) \bar{V}(j^1) \right\} \]

\[ \ldots \sum_{j^n \in \Theta_C(\delta^n)} \alpha_n(x, \pi) \bar{V}(j^n) \bar{r}(x(j^n), \delta^{n+1}) \right\} . \]

The next lemma shows that \( \alpha_n(x, \pi) \) may be bounded away from 1 for \( n = 2, 3, 4, \ldots \). As a result it will be possible to bound the expression for \( V_\pi(x) \) by an expression for a discounted process. Recall that \( \sum_{j \in D_2} p_{ij}(a) > 0 \) for \( i \in D_1, a \in A \). Since both \( D_1 \) and \( A \) are finite sets let

\[ p = \inf_{i \in D_1, a \in A} \left\{ \sum_{j \in D_2} p_{ij}(a) \right\} \quad \text{and note that} \quad p > 0 . \]
Lemma 2: For \( n = 2, 3, \ldots \)

\[ \sup \alpha_n(x, \pi) \leq 1 - p \]

where the \( \sup \) is over all \( x \in S \) and all preceding actions and outcomes.

Proof: Consider the case \( n = 2 \), arbitrary \( x \) and arbitrary \( j^1 \in \theta_C(\delta^1) \). Then

\[
\sum_{j^2 \in \theta_{T^2}} p(j^2) = \sum_{i \in D^2} x_i(j^1)
\]

\[
= \sum_{i \in D^2} \left\{ \frac{\sum_{\ell \in D} p_{\ell j^1}(\delta^1(x)) P(j^1|\ell, \delta^1(x)) x_{\ell}}{\sum_{k \in D} P(j^1|k, \delta^1(x)) x_k} \right\}
\]

\[
= \frac{\sum_{\ell \in D} \sum_{i \in D^2} p_{\ell j^1}(\delta^1(x)) P(j^1|\ell, \delta^1(x)) x_{\ell}}{\sum_{k \in D} P(j^1|k, \delta^1(x)) x_k}
\]

\[
\geq \frac{\sum_{\ell \in D} p P(j^1|\ell, \delta^1(x)) x_{\ell}}{\sum_{k \in D} P(j^1|k, \delta^1(x)) x_k} = p
\]

Hence for any \( x \) and \( j^1 \in \theta_C(\delta^1) \),

\[
\alpha_2(x, \pi) = \sum_{j^2 \in \theta_C(\delta^2)} P(j^2) \leq 1 - \sum_{j^2 \in \theta_{T^2}} P(j^2) \leq 1 - p < 1.
\]
Thus \( \sup \alpha_2(x, \pi) \leq 1 - p \). The proof for \( i \geq 3 \) is identical with \( 2 \) replaced by \( i \) and \( 1 \) replaced by \( i - 1 \) everywhere.

To facilitate the comparison with a discounted process let

\[
\beta_n = \sup \alpha_n(x, \pi) \leq 1 - p. \quad \text{Then}
\]

\[
(17) \quad V_\pi(x) \leq \tilde{r}(x, \delta) + \sum_{n=1}^{\infty} \left\{ \beta_1 \beta_2 \cdots \beta_n \sum_{j_1 \in \theta_G(\delta)} \tilde{F}(j_1) \right. \\
\quad \left. \cdots \sum_{j_n \in \theta_G(\delta)} \tilde{F}(j_n) \tilde{r}(x(j_n), \delta^{n+1}) \right\}.
\]

Now \( \sup_{n \geq 2} \beta_n \leq \sup_{n \geq 2} (1 - p) = 1 - p \) so finally

\[
(18) \quad V_\pi(x) \leq \tilde{r}(x, \delta) + \sum_{n=1}^{\infty} (1 - p)^{n-1} \left\{ \sum_{j_1 \in \theta_G(\delta)} \tilde{F}(j_1) \right. \\
\quad \left. \cdots \sum_{j_n \in \theta_G(\delta)} \tilde{F}(j_n) \tilde{r}(x(j_n), \delta^{n+1}) \right\}.
\]

This is the expression for a standard discounted Markov decision process where the discounting is delayed one period. Hence the original expression (14) behaves like a discounted process even though the discount factor is hidden and variable. As a result Blackwell’s theorem implies that there is an optimal stationary policy \( \pi^* = (\delta^*, \delta^*, ...) \) [3, p. 234]. For future convenience this is simply denoted by \( \delta^* \).

Note that equation (18) implies that a transient policy type condition holds, i.e., the expected reward associated with every stationary policy is finite. Further, \( \delta^* \) satisfies the optimality
equation [3, p. 232], namely if \( V^* \) is the optimal return function then

\[
V^*(x) = \sup_{a \in A} \left\{ r(a)x + \sum_{j \in \Theta_T(a)} U(j) x(j) P(j) + \sum_{j \in \Theta_C(a)} V^*(x(j)) P(j) \right\}.
\]

This will be useful in later sections.

2.6. **Classes of Policies.**

In the previous section the existence of a stationary optimal policy \( \hat{\pi}^* \) and an optimal return function \( V^* \) was shown. This section develops a few of their basic properties.

First recall that a partition \( T \) of \( S \) is a set of subsets of \( S \) such that

\[
T = \{ T_1, T_2, \ldots \}
\]

\[
T_i \cap T_j = \emptyset \quad \text{for } i \neq j
\]

\[
\bigcup_{i} T_i = S
\]

The partitions considered here also satisfy

\[
T_i \text{ is a convex set, } \quad i = 1, 2, 3, \ldots
\]
Let $V^n(x)$ be the optimal return for an $n$ period decision process if the starting state is $x$. $V^n(\cdot)$ obeys the dynamic programming recursion

\begin{equation}
V^n(x) = \max_{a \in A} r(a)x + \sum_{j \in \mathcal{G}_a} V^{n-1}(x(j)) P(j) + \sum_{j \in \mathcal{G}_T(a)} U(j) x(j) P(j)
\end{equation}

It is a well known fact that $V^n(\cdot) \rightarrow V^*(\cdot)$ as $n \rightarrow \infty$.

**Theorem 1:**
1. $V^n(\cdot)$ is piecewise linear on a finite partition $T^n$ of $S$ satisfying (20) - (23), i.e., $V^n(x) = b_1 x$ for $x \in T^n_1$ where $b_1 \in \mathbb{R}^M$, and
2. $V^n(\cdot)$ is convex.

**Comment:** $b_1$ is assumed to be a row vector.

**Proof:** Proceed by induction.

\begin{align*}
V^1(x) &= \max_{a \in A} \left\{ r(a)x + \sum_{j \in \mathcal{G}_T(a)} U(j) x(j) P(j) \right\} \\
&= \max_{a \in A} \left\{ r(a)x + \sum_{j \in \mathcal{G}_T(a)} U(j) x(j) \right\}.
\end{align*}

The right hand expression is a linear function of $x$. Thus $V^1(x)$ can be expressed in the form $V^1(x) = \max_{b \in B^1} bx$ where $b \in \mathbb{R}^M$ and $B^1$
is a finite set. The maximum of a finite set of linear functions is clearly piecewise linear and convex. Also, the set of points \( Q \) such that \( b_i x \) is the maximal hyperplane,

\[
Q = \{ x \in S \mid b_i x \geq bx \text{ for all } b \in B^1 \},
\]

is defined by finitely many linear inequalities and hence is convex. Thus a finite partition \( T^1 \) of \( S \) may be chosen such that \( T^1 \) satisfies (20) - (23) and \( V^1(x) = b_i x \) for \( x \in T^1 \). The only possible difficulty in verifying the first part of the theorem occurs when \( b_i x = b_j x \) for \( i \neq j \), i.e., on the boundaries of \( Q \). At this point, in order to assure that each of the elements of the partition is convex, care must be taken in choosing which hyperplane is chosen as the maximum along the boundary. A lexicographic scheme will resolve this.

To advance the induction assume the result is true for \( n-1 \). By hypothesis, \( V^{n-1}(x) = \max_{b \in B^{n-1}} bx \) where \( b \in \mathbb{R}^n \) and \( B^{n-1} \) is finite. Thus

\[
\begin{align*}
  r(a) \cdot x + \sum_{j \in \Theta_C(a)} V^{n-1}(x(j)) P(j) + \sum_{j \in \Theta_T(a)} U(j) x(j) P(j) \\
  = r(a) \cdot x + \sum_{j \in \Theta_C(a)} \max_{b \in B^{n-1}} bx(j) P(j) + \sum_{j \in \Theta_T(a)} U(j) x(j) P(j) \\
  = r(a) \cdot x + \sum_{j \in \Theta_C(a)} \max_{b \in B^{n-1}} bZ(j)x + \sum_{j \in \Theta_T(a)} U(j) Z(j)x
\end{align*}
\]

which is a piecewise linear, convex function. Hence it may be expressed as \( \max_{c \in C(a)} c \cdot x \) where \( c \in \mathbb{R}^n \) and \( C(a) \) is a finite set. Note that \( c \in C(a) \).
for each \( a \in A \) there is a set of hyperplanes. Each of these hyperplanes is said to be associated with action \( a \) since \( a \) would be the first action performed. Now

\[
V^n(x) = \max_{a \in A} \left\{ r(a) \cdot x + \sum_{j \in \theta_c(a)} V^{n-1}(x(j)) P(j) + \sum_{j \in \theta_T(a)} U(j) x(j) P(j) \right\}
\]

\[
= \max_{a \in A} \left\{ \max_{c \in C(a)} c \cdot x \right\}.
\]

\( V^n(\cdot) \) may thus be expressed as the maximum of a finite collection of hyperplanes. Hence it is piecewise linear and convex and, exactly as in the case \( n = 1 \), a finite partition \( T^n \) satisfying (20) - (23) is associated with \( V^n(\cdot) \).

The fact that \( V^n(\cdot) \) is piecewise linear implies that for an \( n \)-period model the optimal policy can be characterized by a finite partition \( T^n \) of \( S \) with a single action being optimal in each \( T^n_1 \). This is because each hyperplane is associated with some \( a \in A \) as in the proof above. In general there will be several \( T^n_1 \) where the same action \( a \in A \) is optimal. Unfortunately, it is not clear that there is an upper bound on the number of elements in \( T^n \) as \( n \rightarrow \infty \). Hence consider three classes of policies:

I. all stationary policies,

II. all stationary policies which induce a countable partition on \( S \),

III. all stationary policies which induce a finite partition on \( S \).

A stationary policy \( \delta \) "induces" a partition \( T \) if there is a partition \( T \) of \( S \) satisfying (20) - (23) such that
(25) \[ x \in T_i \Rightarrow \delta(x) = a(i) \in A, \quad i = 1, 2, \ldots \]

The next two theorems show that there is an optimal class II policy.

**Theorem 2:** \( V^*(\cdot) \) is piecewise linear and convex on a countable partition \( T^* \) of \( S \) which satisfies (20) - (23), i.e., \( V^*(x) = b_i x \) for \( x \in T_i^* \).

**Proof:** Let \( T^n \) be the partition of \( S \) associated with \( V^n(\cdot) \) satisfying (20) - (23). Note that \( T^n \) is finite for each \( n \). Define the collection of sets \( T \) (which will be the desired partition) by

\[
(26) \quad T^* = \{ T^n_i | T^n_i = \bigcap_{n=1}^{\infty} T^n_{k(n)} \text{ where } T^n_{k(n)} \in T^n \}.
\]

Verbally, \( T \) is the set of all possible intersections of elements of the partitions \( T^n \) taken one at a time. Since each \( T^n \) is finite, \( T \) is countable.

It is first necessary to show that \( T^* \) is a partition.

\( T^*_i \cap T^*_i' = \emptyset \) for \( i \neq i' \) for if not then there is some \( x \in T^*_i \cap T^*_i' \).

Let \( \ell \) be the first index such that \( k(\ell) \) associated with \( T^*_i \) is not equal to \( k(\ell) \) associated with \( T^*_i' \). Then \( x \in T^\ell_{k(\ell)} \) and \( x \in T^\ell_{k(\ell)} \) which is a contradiction. Clearly \( S \supseteq \bigcup T^*_i \). Suppose \( x \in S \). Then \( x \in T^n_{k(n)} \) for each \( n = 1, 2, \ldots \) and some \( k(n) \).

Hence \( x \in \bigcap_{n=1}^{\infty} T^n_{k(n)} \) which is an element of \( T^* \). Thus \( \bigcup T^*_i \supseteq S \).

Now for every \( x \in T^*_i \), \( x \in T^n_{k(n)} \), \( n = 1, 2, \ldots \). Hence \( V^n(x) = b_i k(n)^* x \). For each such \( x \), \( V^n(x) \to V^*(x) \) so
(27) \[ V^*(x) = \lim_{n \to \infty} b_k(n) \cdot x = \left( \lim_{n \to \infty} b_k(n) \right) \cdot x = b_i \cdot x \]

Finally condition (23) holds since the intersection of convex sets is convex. \( V^*(\cdot) \) is convex since each \( V^n(\cdot) \) is.

Theorem 3: Suppose \( V^*(\cdot) \) is piecewise linear and convex on a countable partition \( T^* \) of \( S \) satisfying (20) - (23). Then the policy \( \delta^*(\cdot) \) defined by

\[
\delta^*(x) \in \arg \max_{a \in A} \left\{ r(a) \cdot x + \sum_{j \in T_C(a)} V^*(x(j)) P(j) + \sum_{j \in T_T(a)} U(j) x(j) P(j) \right\}
\]

is an optimal policy and can be chosen to satisfy (25), i.e.,

\( \delta^*(x) = a(i) \) for \( x \in T^*_1 \).

Proof: Since \( V^*(\cdot) \) is piecewise linear and convex, \( V^*(x) = \max_{b \in B^*} bx \)

\( B^* \) countable, where the maximum is attained. Thus for every \( x \in T^*_1 \),

\[
\max_{a \in A} \left\{ r(a) \cdot x + \sum_{j \in T_C(a)} V^*(x(j)) P(j) + \sum_{j \in T_T(a)} U(j) x(j) P(j) \right\} = \max_{a \in A} \left\{ r(a) \cdot x + \max_{b \in B} bx(j) P(j) + \sum_{j \in T_T(a)} U(j)x(j)P(j) \right\} = b_i \cdot x
\]
Since this is true for every $x \in T_1^x$ there must be some $a(i)$ which achieves the maximum for every $x \in T_1^x$, i.e.,

$$b_i x = r(a(i)) x + \sum_{j \in \Theta_c(a(i))} b(j) Z(j) x + \sum_{j \in \Theta_I(a(i))} U(j) Z(j) x.$$ 

Hence $\delta^*$ can be chosen so that $\delta^*(x) = a(i)$ for $x \in T_1$ and $\delta^*$ is clearly optimal.

By Theorem 2 $V^*$ is piecewise linear on a countable partition $T^*$ of $S$. By Theorem 3 there is a stationary $\delta^*$ which satisfies conditions (20) - (23) and (25). Hence,

**Corollary 1:** There is an optimal class II policy.

2.7. **Randomized Actions**

In the previous sections the existence of an optimal policy with certain properties was shown. In case there was more than one optimal action for $x$ a specific $a \in A$ was chosen so that certain convexity requirements were met. Now consider all actions which are optimal at $x$. Let

$$I(x) = \{a \in A | V(x, a) = V^*(x)\}$$

(28)

where

(29) \[ V(x, a) = r(a) x + \sum_{j \in \Theta_c(a)} V^*(x(j)) P(j) + \sum_{j \in \Theta_I(a)} U(j) x(j) P(j). \]
A randomized action, $\alpha$, is a probability distribution over $A$, i.e.,
$\alpha_a$ represents the probability of choosing action $a$. Hence $\alpha_a \geq 0$
and $\sum_{a \in A} \alpha_a = 1$. Any randomization of optimal actions is optimal.

**Lemma 2:** If $\alpha$ is a randomized action then $\alpha$ is optimal for $x$
if and only if $\alpha_a > 0$ implies $a \in I(x)$.

**Proof:** Trivial.

**Theorem 4:** The set of optimal actions at $x$ is a convex polyhedron
whose vertices are $a \in I(x)$.

**Proof:** $a \in I(x)$ implies $a$ is optimal at $x$. Now $\alpha$ is optimal
at $x$ if and only if it is a randomization of elements of $I(x)$,
i.e., it lies in the convex hull of $I(x)$. But the convex hull
of a finite set is a convex polyhedron. It is obvious by construction
that $a \in I(x)$ is a vertex.

Finally, the set of optimal actions as a function of the
generalized state has a closed graph.

**Theorem 5:** If $a \in I(x^\ell)$, $\ell = 1, 2, \ldots$ and $\lim_{\ell \to \infty} x^\ell = x \in S$
then $a \in I(x)$.

**Proof:** $a \in I(x^\ell)$ implies $V^\star(x^\ell) = V(x^\ell, a)$. But both $V^\star(\cdot)$ and
$V(\cdot, a)$ are continuous functions of $x$. Taking the limit yields
$V^\star(x) = V(x, a)$ implying $a \in I(x)$.
CHAPTER 3

BINARY STRUCTURE

3.1. The Problem with Analysis of the General Model

The formulation of the previous chapter encompasses a variety of possible specific models. Sondik [24] and Smallwood and Sondik [23] studied a model similar to this one which, however, makes use of a discount factor rather than termination probabilities. They also reverse the order of the application of Bayes Theorem and the Markov chain, i.e., an action is chosen, the Markov state transition occurs, and then the output is generated based on the state after the transition. This differs from the present model where the output depends on the original state. Their formulation is excellent for some cases; see the discussion of machine maintenance in Section 7.2. However, it seems more appropriate to use the present formulation in the medical context. Sondik [24] derives a computational algorithm of the policy iteration variety but does not analyze the form of the optimal policy. It appears that additional structure is required to derive further theoretical results. Very reasonable examples can be exhibited which refute several of the obvious conjectures regarding the form of the optimal policy. The example of Appendix B shows that the region where action $a$ is optimal need not be convex, even under restrictive assumptions. In addition, Sondik showed that the region need not even be connected in the general case.
It appears that most of the difficulties are due to the exceedingly large number of ways the hyperplanes forming \( v^N(\cdot) \) and \( v^*(\cdot) \) can interact. The sequences of possible actions and outcomes blossom out into a tree which, in the words of Raiffa [21], rapidly becomes a "bushy mess". The number of possible elements in the partition \( T^N \) increases equally quickly.

To alleviate this, a version of the model proposed in Chapter 2 is now considered which has, in a certain sense, a binary rather than a tree-like structure. A basic result called the scaling theorem is given here. Chapters 4 and 5 then analyze two specific cases.

### 3.2. Binary Models

As in Chapter 2, assume that there are \( N \) possible diseases which the patient can have, indexed by \( i = 1, 2, \ldots, N \). These may include multiple disease states, e.g., a broken leg with lacerations. In addition there are \( M-N \) possible terminating states which are labeled \( N+1, N+2, \ldots, M \). The actions available are basically tests and treatments for various diseases. From a macroscopic point of view each of these can be thought of as having two outcomes, success and failure. If action \( a \) is a treatment then it either cures the patient or does not. If it is a test for some disease then it either identifies the disease in the patient or indicates that he does not have the disease. This becomes very reasonable if each test is considered as a collection of medical procedures to identify whether or not the patient has a given disease.
If each action essentially results in success or failure then it is also reasonable to assume that the decision process continues until success occurs, at which point it stops. If a treatment is applied to the patient and he is cured there is no incentive to continue the diagnosis and treatment process. If a test identifies the disease that the patient has then one would expect the physician to treat that disease. This, however, can be thought of as a fairly standardized procedure. Hence for all practical purposes the decision process stops when the disease is identified.

The basic ideas of a binary structure then are that each action has two basic outcomes, success and failure, and that the decision process stops whenever success occurs. While this is not completely general it seems quite reasonable and not too restrictive in the medical context. It certainly provides a reasonable starting point for the development of results on the form of the optimal policy.

Mathematically the necessary assumption for a binary structure is that \( \theta_C(a) \) be a singleton for all \( a \in A \), i.e., \( \theta_C(a) = \{ f \} \). In other words, there is only one outcome, failure, which implies the continuation of the decision process. Note that this assumption is slightly weaker than the previous paragraph since it allows for several possible successful outcomes as well as outcomes indicating that the process has spontaneously entered one of the terminating states.

The mathematical implications of the assumption that \( \theta_C(a) \) is a singleton for all \( a \in A \) may be informally described as follows. Suppose \( x \) is the state at the start of the decision process and some policy \( \pi \) is being followed. In the first period some action \( a_1 \) is performed. If \( f \in \theta_C(a_1) \) occurs then the process moves to the new state \( x(f) \) at which time another action \( a_2 \) is performed in period 2.
Note that only one action is possible each period. Hence for a given x, the policy could be described by the sequence of actions to be followed if failure occurs in the preceding period, i.e., \((a_1, a_2, \ldots)\). Now \(V^n(x)\) is the maximum of the expected rewards for all possible n-period policies. Equivalently, it is the maximum of the expected returns associated with all possible sequences of n actions. In other words, if \(H^n(a_1, a_2, \ldots, a_n)(x)\) is the expected return starting in x and using \(a_k\) in period k if \(a_{k-1}\) fails in period k-1 then

\[
V^n(x) = \max_{(a_1, a_2, \ldots, a_n)} H^n(a_1, a_2, \ldots, a_n)(x).
\]

The \(H^n(a_1, \ldots, a_n)(\cdot)\) turn out to be the hyperplanes which form \(V^n(\cdot)\). This form for \(V^n(\cdot)\) is useful in characterizing the optimal policy.

More formally, refer to the recursive arguments of Section 2.6 where it was proved that \(V^n(\cdot)\) is piecewise linear on a convex partition of S. Recall that

\[
(1) \quad V^n(x) = \max_{a \in A} \left\{ r(a)x + \sum_{j \in \Theta(a)} \max_{b \in B^{n-1}} bZ(j)x + \sum_{j \in \Theta(a)} U(j)z(j)x \right\}.
\]

Now if \(\Theta(a)\) has only the single outcome of failure in it, i.e., \(\Theta(a) = \{f\}\) then (1) reduces to

\[
(1') \quad V^n(x) = \max_{a \in A} \left\{ r(a)x + \max_{b \in B^{n-1}} bZ(f)x \right\}.
\]
\[ V^n(x) = \max_{a \in A} \left\{ r(a)x + \max_{b \in B^{n-1}} bZ(f)x + \sum_{j \in \Theta_T(a)} U(j) Z(j)x \right\} \]

\[ = \max_{a \in A, b \in B^{n-1}} \left\{ r(a)x + bZ(f)x + \sum_{j \in \Theta_T(a)} U(j) Z(j)x \right\} . \]

The expression in \{ \} is a hyperplane. Since \[ V^n(x) = \max_{b \in B^n} bx \]
it is clear that the elements of \( B^n \) may be indexed by a single action to be performed in the first period and by an element of \( B^{n-1} \). In other words, the hyperplanes which form \( V^n(\cdot) \) depend on an action \( a \) and an element \( b \) from \( B^{n-1} \). But now the elements of \( B^{n-1} \) depend on some other action and the elements of \( B^{n-2} \) so that \( B^n \) depends on a sequence of two actions and the elements of \( B^{n-2} \). Continuing, one finally obtains the fact that the elements of \( B^n \), and hence the hyperplanes forming \( V^n(\cdot) \), may be indexed by a sequence of \( n \) actions, \( (a_1, a_2, \ldots, a_n) \). Intuitively this corresponds to the fact that given \( x \) the entire sequence of actions to be taken over \( n \) periods yielding \( V^n(x) \) may be described simply by giving the action to be taken in period \( k \) if the actions in the previous \( k-1 \) periods fail.

In line with the above discussion define a binary policy as one for which \( \delta^n(x) = a_n \) for all \( x \in S \). In other words, action \( a_n \) is taken in period \( n \) regardless of what the state is. Since there is only one action per period, an \( n \)-period binary policy can be denoted by the sequence \( (a_1, a_2, \ldots, a_n) \). This means that action \( a_1 \) is taken in period 1. If it fails then action \( a_2 \) is taken in period 2. If \( a_2 \) also fails then \( a_3 \) is performed in period 3 and so on.
$H^n(a_1, a_2, \ldots, a_n)(x)$ be defined as the expected n-period return associated with the binary policy $(a_1, \ldots, a_n)$ if the starting state is $x$. If it can be shown that $H^n(a_1, \ldots, a_n)(\cdot)$ is a hyperplane then from the above discussion it is clear that $H^n(a_1, a_2, \ldots, a_n)(x)$ corresponds to $bx$ for some $b \in B^n$.

In other words, as $(a_1, \ldots, a_n)$ range over all possible n-period binary policies the $H^n(a_1, a_2, \ldots, a_n)(\cdot)$ are the hyperplanes which form $V^n(\cdot)$. Hence

$$V^n(x) = \max_{(a_1, \ldots, a_n)} H^n(a_1, a_2, \ldots, a_n)(x).$$

Showing that $H^n(a_1, a_2, \ldots, a_n)(\cdot)$ is a hyperplane is simple. $H^n$ obeys the recursion

$$H^n(a_1, \ldots, a_n)(x) = r(a_1)x + \sum_{j \in T(a_1)} U(j)Z(j)x + P(f)H^{n-1}(a_2, \ldots, a_n)(x(f)).$$

Now if $H^{n-1}$ is a linear function then

$$H^n(a_1, \ldots, a_n)(x) = r(a_1)x + \sum_{j \in T(a_1)} U(j)Z(j)x + H^{n-1}(a_2, \ldots, a_n)(Z(f)x)$$

implying that $H^n$ is a linear function. Hence it only remains to show that $H^1(a_1)(x)$ is a hyperplane. To do this it is necessary to define a terminal reward structure. This was sidestepped in Chapter 2.
since the infinite horizon properties do not depend on it. Now assume
that if the process terminates by running out of time when the patient
has disease i there is a reward/penalty of $U^0_i$. In Chapters 4 and 5
these values are based on a consideration of what might happen to the
patient if he is left with disease i. Then letting $U^0 = (U^0_1, \ldots, U^0_N)$
the expected terminal reward is $U^0 x$ if $x$ is the final state.

Hence

$$H^1(a)(x) = r(a)x + \sum_{j \in \Theta_T(a')} U(j) Z(j)x + U^0 x(f) P(f)$$

$$= r(a)x + \sum_{j \in \Theta_T(a')} U(j) Z(j)x + U^0 Z(f)x$$

which is clearly linear in $x$.

To summarize, under the assumption of binary structure, i.e.,
that $\Theta_0(a) = \{f\}$,

$$V^n(x) = \max_{(a_1, \ldots, a_n)} H^n(a_1, \ldots, a_n)(x)$$

where $H^n(a_1, \ldots, a_n)(x)$ is the expected return associated with
the binary policy $(a_1, \ldots, a_n)$. The $H^n(a_1, \ldots, a_n)(\cdot)$ are
the hyperplanes which determine $V^n(\cdot)$. As in the previous chapter,$H^n(a_1, \ldots, a_n)(\cdot)$ is associated with its first action, $a_1$. Thus
if

$$V^n(x) = H^n(a_1^*, a_2^*, \ldots, a_n^*)(x)$$
then $a^*_1$ is the optimal action to be taken in the first period at $x$.

Also, if $a$ is the optimal first action at $x$ for an $n$-period problem then

$$v^n(x) = H^n(a_1^*, a_2^*, \ldots, a_n^*)(x)$$

for some $(a_1^*, a_2^*, \ldots, a_n^*)$. Let $\delta^n$ be the optimal $n$-period decision rule, i.e., $\delta^n : S \rightarrow A$ and for each $x \in S$, $\delta^n(x)$ is the optimal action to be taken in the first period of an $n$-period problem.

In the following chapters $\delta^n$ will be studied to gain insight into the form of $\delta^*$, the optimal stationary policy for the infinite horizon problem. For now, note that $\delta^n(x) = a$ if and only if

$$v^n(x) = H^n(a, a_2^*, \ldots, a_n^*)(x) \quad \text{for some } (a_2^*, \ldots, a_n^*).$$

Hence the optimal decision rule and return function may be analyzed by considering only binary policies and the expected returns associated with them.

It is possible to get an explicit expression for $H^n(a_1, \ldots, a_n)(\cdot)$. For a single period,

$$H^1(a_1)(x) = r(a_1)x + \sum_{j \in T(a_1)} U(j) Z(j)x + U^0 Z(f)x \quad (4)$$

Similarly,

$$H^2(a_1, a_2)(x) = r(a_1)x + r(a_2) Z(f, a_2)x + \sum_{j \in T(a_1)} U(j) Z(j, a_1)x$$

$$+ \sum_{j \in T(a_2)} U(j) Z(j, a_2) Z(f, a_1)x + U^0 Z(f, a_2) Z(f, a_1)x \quad (5)$$

47
where here the dependence of $Z$ on the action as well as the outcome has been made explicit since the multiplication of several $Z$'s together could otherwise be confusing. In other words, $Z(j,a) = Z(j)$ where $j \in \Theta(a)$. For notational convenience let

\begin{equation}
\Omega(a_1, \ldots, a_k) = Z(f, a_k) Z(f, a_{k-1}) \cdots Z(f, a_1).
\end{equation}

The following generalizes (4) and (5).

**Lemma 1:**

\[ H_n(a_1, \ldots, a_n)(x) = \sum_{k=1}^{n} r(a_k) \Omega(a_1, \ldots, a_{k-1})x \]

\[ + \sum_{k=1}^{n} \sum_{j \in \Theta(a_k)} U(j) Z(j, a_k) \Omega(a_1, \ldots, a_{k-1})x \]

\[ + U \Omega(a_1, \ldots, a_n)x. \]

**Proof:** By induction. Equations (4) and (5) show that the result is true for $n = 1$ and $n = 2$. Hence assume it holds for $n-1$. Then
\[ H^n(a_1, \ldots, a_n)(x) \]
\[ = r(a_1)x + \sum_{j \in T(a_1)} U(j) Z(j, a_1)x + \sum_{k=2}^{n} r(a_k) \Omega(a_2, \ldots, a_{k-1}) Z(f, a_1)x \]
\[ + \sum_{k=2}^{n} \sum_{j \in T(a_k)} U(j) Z(j, a_k) \Omega(a_2, \ldots, a_{k-1}) Z(f, a_1)x \]
\[ + U^0 \Omega(a_2, \ldots, a_n) Z(f, a_1)x \]
\[ = \sum_{k=1}^{n} r(a_k) \Omega(a_1, \ldots, a_{k-1})x + \sum_{k=1}^{n} \sum_{j \in T(a_k)} U(j) Z(k, a_k) \Omega(a_1, \ldots, a_{k-1})x \]
\[ + U^0 \Omega(a_1, \ldots, a_n)x. \]

3.3. The Scaling Theorem

The major result of this chapter can be called a "scaling" theorem. The idea is that in most cases the values of \( x_i \) for \( i \in D_2 \) (the terminating states) are irrelevant in determining the optimal action. The result shows that the optimal policy can be characterized by setting \( x_{N+1} = x_{N+2} = \cdots = x_M = 0 \) which will be convenient later. It also implies that each of the partition elements \( T^n_i \) is pyramid-shaped in a certain sense. The following lemma is mathematically useful.
Lemma 2: Suppose $x \in S$ and $j \in \theta_G(a) \cup \theta_{T1}(a)$. Then

$$Z(j)x = Z(j) (x_1, \ldots, x_N, 0, \ldots, 0)^t.$$

Proof: $[Z(j)]_{k\ell} = p_{k\ell}(a) P(j|\ell, a)$. For $j \in \theta_G(a) \cup \theta_{T1}(a)$ and

$\ell \in D_2$, i.e., $\ell = N+1, N+2, \ldots, M$, $P(j|\ell, a) = 0$. Hence the last

M-N columns of $Z(j)$ are zero, implying that the last M-N components

of $x$ never contribute to the multiplication.

For $x \in S$ let $d(x) = 1 - \sum_{i \in D_2} x_i = \sum_{i \in D_1} x_i$.

Theorem 1: (Scaling Theorem). Suppose $x \in S$ and $d(x) \neq 0$. Let

$y = d^{-1}(x) (x_1, x_2, \ldots, x_N, 0, \ldots, 0)^t$. Suppose $r_i^0(a) = r_i^0(b)$

for $i \in D_2$ and all $a, b \in A$. Then

$$H^n(a_1^*, a_2^*, \ldots, a_n^*) (x) \geq H^n(a_1, a_2, \ldots, a_n) (x)$$

if and only if

$$H^n(a_1^*, a_2^*, \ldots, a_n^*) (y) \geq H^n(a_1, a_2, \ldots, a_n) (y).$$

Comment: The hypothesis that $r_i(a) = r_i(b)$ is trivially satisfied

since it was previously assumed that $r_i(a) \equiv 0$ for $i \in D_2$ and $a \in A$.

Proof: Suppose that $H^n(a_1^*, \ldots, a_n^*) (x) \geq H^n(a_1, \ldots, a_n)(x)$. Applying Lemma 1 and separating the terms with $k = 1$ out yields
\[ r(a_1^*)x + \sum_{j \in \Theta_T(a_1^*)} U(j) Z(j, a_1^*) x + \sum_{k=2}^{n} [r(a_k^*) \Omega(a_1^*, \ldots, a_{k-1}^*) + \sum_{j \in \Theta_T(a_k^*)} U(j) Z(j, a_k^*) \Omega(a_1^*, \ldots, a_{k-1}^*)] x + U^0 \Omega(a_1^*, \ldots, a_n^*) x \]
\[ \geq r(a_1^*)x + \sum_{j \in \Theta_T(a_1)} U(j) Z(j, a_1^*) x + \sum_{k=2}^{n} [r(a_k^*) \Omega(a_1^*, \ldots, a_{k-1}^*) + \sum_{j \in \Theta_T(a_k^*)} U(j) Z(j, a_k^*) \Omega(a_1^*, \ldots, a_{k-1}^*)] x + U^0 \Omega(a_1^*, \ldots, a_n^*) x . \]

Now by Lemma 2 and linearity,
\[ \Omega(a_1^*, \ldots, a_{k-1}^*) x = \Omega(a_1^*, \ldots, a_{k-1}^*) d(x)y, \quad \text{for } k \geq 2. \]

Hence \( x \) can be replaced by \( d(x)y \) in the third term on both sides of the inequality without changing the inequality. Similarly,
\[ U^0 \Omega(a_1^*, \ldots, a_n^*) x = U^0 \Omega(a_1^*, \ldots, a_n^*) d(x)y \]
so \( x \) can be replaced by \( d(x)y \) in the final term of both sides of the expression.

Consider now the first term of both sides of the inequality,
\[ r(a_1^*)x = r(a_1^*) d(x)y + \sum_{i \in D_2} r_i(a_1^*) x_i \]
and
\[ r(a_1) x = r(a_1) d(x)y + \sum_{i \in D_2} r_i(a_1) x_i . \]
But \( r_1(a_1^x) = r_1(a_1) \) for \( i \in D_2 \) so substituting the above expressions in place of the first terms and then subtracting \( \sum_{i \in D_2} r_1(a_1) x_i \) from both sides effectively replaces \( x \) by \( d(x)y \) in the first term.

Finally examine the second term of both sides. For \( j \in \Theta_{T_1}(a_1) \),

\[
\sum_{j \in \Theta_{T_1}} U(j) Z(j, a_1) x = \sum_{j \in \Theta_{T_1}(a_1)} U(j) Z(j, a_1) d(x)y
\]

by Lemma 2. A similar replacement holds for \( a_1^x \). For \( j \in \Theta_{T_2} \),

\( Z(j, a)x = e_{N+j} \) no matter what \( a \) is. Hence

\[
\sum_{j \in \Theta_{T_2}} U(j) Z(j, a_1^x) x = \sum_{j \in \Theta_{T_2}} U(j) Z(j, a_1) x.
\]

Subtract this quantity from both sides of the inequality. Note also that

\[
\sum_{j \in \Theta_{T_2}} U(j) Z(j, a_1^x) d(x)y = 0 = \sum_{j \in \Theta_{T_2}} U(j) Z(j, a_1) d(x)y
\]

so again \( x \) can be replaced by \( d(x)y \) in the second term of the inequality.

The end result of these replacements is
\[
\begin{align*}
& r(a_1^*) \, d(x) y + \sum_{j \in \Theta_T(a_1^*)} U(j) \, Z(j, a_1^*) \, d(x) y + \sum_{k=2}^n [r(a_k^*) \, \Omega(a_1^*, \ldots, a_{k-1}^*) \\
& + \sum_{j \in \Theta_T(a_k^*)} U(j) \, Z(j, a_k^*) \, \Omega(a_1^*, \ldots, a_{k-1}^*)] \, d(x) y \\
& + U^0 \, \Omega(a_1^*, \ldots, a_n^*) \, d(x) y \\
\geq& \ r(a_1) \, d(x) y + \sum_{j \in \Theta_T(a_1)} U(j) \, Z(j, a_1) \, d(x) y \\
& + \sum_{k=2}^n [r(a_k) \, \Omega(a_1, \ldots, a_{k-1}) + \sum_{j \in \Theta_T(a_k)} U(j) \, Z(j, a_k) \, \Omega(a_1, \ldots, a_{k-1})] \\
& \cdot d(x) y + U^0 \, \Omega(a_1, \ldots, a_n) \, d(x) y.
\end{align*}
\]

Since both sides of the inequality are linear multiplication by the scalar \( d^{-1}(x) \) yields

\[
H^n(a_1^*, \ldots, a_n^*)(y) \geq H^n(a_1, \ldots, a_n)(y).
\]

To prove the "if" part merely reverse all steps.

Since \( V^n(x) = \max_{(a_1, \ldots, a_n)} H^n(a_1, \ldots, a_n)(x) \), Theorem 1 implies that if \( V^n(y) = H^n(a_1^*, \ldots, a_n^*)(y) \) then \( V^n(x) = H^n(a_1^*, \ldots, a_n^*)(x) \).

Hence the optimal \( n \)-period policy can be completely characterized by describing the optimal policy for those \( x \in S \) for which \( \sum_{i \in D_2} x_i = 0 \).

The only case not covered by Theorem 1 is when \( d(x) = 0 \), i.e., when \( \sum_{i \in D_2} x_i = 1 \). But this is equivalent to saying that the process has terminated with probability 1. Hence no decision is necessary.
Theorem 1 implies that each element of the partition $T^n$ associated with $V^n$ is pyramid-shaped. This is difficult to visualize geometrically since $D_2$ in general has several elements. The idea, however, can be made clear by considering the case where $D_2$ is a singleton. Then each $T^n_1$ is a "pyramid" whose point is $e_{N+1}$. Further, the cross section of $T^n_1$ associated with $x_{N+1}$ equal to a constant is a scaled down version of the cross section associated with $x_{N+1} = 0$. Figure 3 illustrates the situation.

Figure 3: Structure of $T^n_1$
This pyramid shape can be mathematically formalized.

**Corollary 1:** Suppose \( x \in T^n_i \) associated with \( V^n(\cdot) \). Then

\[
\alpha x + \sum_{i \in D_2} \beta_i \cdot e_i \in T^n_i \\
\text{for } \alpha > 0, \beta_i \geq 0, \text{ and } \alpha + \sum_{i \in D_2} \beta_i = 1.
\]

Stated slightly differently, Corollary 1 says that if action \( a \) is optimal at \( x \) then it is also optimal at \( \alpha x + \sum_{i \in D_2} \beta_i \cdot e_i \).
CHAPTER 4
MODEL I

4.1. Introduction

This chapter formulates and analyzes a specific case of the general class of models presented in Chapters 2 and 3. Basically it assumes that there is one test and one treatment for each possible disease. In addition there are two terminating states, health and death. Three costs are considered, the actual cost of performing an action, a reward for restoring the patient to health and a penalty for allowing him to die.

The next two sections formulate the transition and reward structures respectively. Sections 4 and 5 then analyze the form of the optimal policy under sufficient conditions which have reasonable medical interpretations.

As mentioned previously, Chapter 5 analyzes another specific case which treats health as an unobservable disease state. Hence these chapters may be viewed as two specializations of the general framework, each of which has positive and negative attributes and is more or less appropriate in a given medical context.

4.2. The Transition Structure

Model I assumes that there are exactly $N + 2 = M$ underlying states. $D_1$ consists of $N$ diseases while $D_2$ has two elements, health and death, denoted $N+1$ and $N+2$ respectively. The set $A$ contains

56
actions, i.e., \( A = \{0, 1, \ldots, 2N\} \). Action 0 consists of doing nothing for some specified period of time. This can be thought of as corresponding to the physician's statement "come back and see me in a week". Obviously several waiting or observation actions could be considered, each for some different amount of time. For simplicity only one is considered here. Action \( a = i \) for \( i = 1, 2, \ldots, N \) is a binary test for disease \( i \). In other words, action \( a = i \) informs the physician that the patient either does or does not have disease \( i \). Action \( a = N+i \) for \( i = 1, 2, \ldots, N \) is a treatment for disease \( i \). The treatment will cure disease \( i \) but have no effect on the patient if he does not have disease \( i \). The physician will know, however, that the patient does not have disease \( i \) due to the failure of the treatment.

These assumptions imply a special structure for the set of possible observations \( \Theta(a) \) and the conditional probabilities \( P(j|i, a) \). Recall that \( \Theta(a) = \Theta_C(a) \cup \Theta_{TL}(a) \cup \Theta_{T2} \). Since \( D_2 \) has two elements, \( \Theta_{T2} = \{1, 2\} \) where outcome 1 indicates the patient is healthy while outcome 2 indicates the patient is dead. As before \( P(j|i, a) = 1 \) for \( i = N+j, j \in \Theta_{T2} \) and zero otherwise.

Consider the tests and treatments, actions 1, 2, \ldots, 2N. Each of these has two outcomes, failure and success, denoted \( f \) and \( s \) respectively. A test succeeds if it indicates that the patient has the associated disease and fails if it indicates that the patient does not have the disease. Similarly a treatment succeeds if the patient is cured and fails if he is not. In case of success there is little motivation to continue the decision process; either the patient is cured or his disease is identified so the appropriate cure may be applied.
Formally assume that \( \theta_C(a) = \{f\} \) and \( \theta_T(a) = \{s\} \) for \( a = 1, 2, \ldots, 2N \). Since failure occurs if the action does not correspond with the disease,

\[
P(f| i, a) = 1 \quad \text{a} \neq i \text{ or } N+i,
\]

\[
P(f| i, a) = 0 \quad a = i \text{ or } N+i.
\]

Similarly success is observed only if the test or treatment corresponds to the disease so,

\[
P(s| i, a) = 0 \quad \text{a} \neq i \text{ or } N+i,
\]

\[
P(s| i, a) = 1 \quad a = i \text{ or } N+i.
\]

The one exceptional case is the action of observing the patient, \( a = 0 \). Here there is no concept of success or failure; the decision process merely continues at some later time. Hence let \( \theta_C(0) = \{f\} \) and \( \theta_T(0) = \emptyset \), the empty set. Thus \( f \) always occurs unless the patient is in a terminating state, so

\[
P(f| i, 0) = i, \quad i \in D_1.
\]

For convenience all of the conditional probabilities are summarized below:
a = 0:
\[ P(f|i, 0) = i \quad i \in D_1 \]
\[ P(f|i, 0) = 0 \quad i \in D_2 \]
\[ P(j|i, 0) = 0 \quad j \in \Theta_{T2}, \quad i \in D_1 \text{ or } j \neq i \in D_2 \]
\[ P(j|i, 0) = 1 \quad j \in \Theta_{T2}, \quad i = N+j \in D_2 \]

a \neq 0:
\[ P(f|i, a) = 0 \quad a = i \text{ or } N+i \text{ or } i \in D_2 \]
\[ P(f|i, a) = 1 \quad a \neq i \text{ or } N+i, \quad i \in D_1 \]
\[ P(s|i, a) = 0 \quad a \neq i \text{ or } N+i \text{ or } i \in D_2 \]
\[ P(s|i, a) = 1 \quad a = i \text{ or } N+i, \quad i \in D_1 \]
\[ P(j|i, a) = 0 \quad j \in \Theta_{T2}, \quad i \in D_1 \text{ or } j \neq i \in D_2 \]
\[ P(j|i, a) = 1 \quad j \in \Theta_{T2}, \quad i = N+j \in D_2 . \]

In addition to the conditional probabilities of outcomes given actions and states of nature, the Markov transition probabilities \( p_{ij}(a) \) must be specified. Assume that no test interferes with the course of disease. Then the spontaneous changes in state if some test is performed are the same as if nothing was done. Hence let

\[ p_{ij}(a) = p_{ij}(0) = p_{ij} , \quad a = 1, 2, \ldots, N . \]

Further, assume that none of the treatments are destructive. If the patient has the associated disease the treatment will cure him, otherwise the treatment has no effect.

Formally let
Note that only the \( i \)-th row of the transition matrix associated with treatment \( N+i \) is different from that associated with test \( i \).

This completely describes the transition scheme for Model I since for \( j \in \Theta(a) \),

\[
x(j) = \frac{Z(j) x}{P(j)}
\]

where \( Z(j) \) and \( P(j) \) are defined in terms of the conditional probabilities and the Markov transition probabilities. Two simple lemmas verify that the transition scheme behaves as postulated. First, the posterior distribution associated with successful treatment implies health with probability one.

**Lemma 1:** If \( x_k > 0 \), \( k \in D_1 \) then \( x(s) = e_{N+1} \) for \( s \in \Theta_T(N+k) \).

**Proof:**

\[
x(s) = \frac{Z(s) x}{P(s)} \text{ where } P(s) = \sum_{\ell=1}^{N-2} P(s|\ell, N+k) x_{\ell} = x_k \text{ by (1)},
\]

Also \( Z_{ij}(s) = p_{ij}(N+k) P(s|j, N+k) = 0 \) for \( j \neq k \). In the \( k \)-th column of \( Z \), \( p_{ki}(N+k) = 0 \) unless \( i = N+1 \) in which case \( p_{k,N+1}(N+k)=1 \).

Hence \( Z_{N+1,k}(s) = 1 \) and all other elements of \( Z(s) \) are zero. Thus \( Z(s)x = (0, \ldots, 0, x_k, 0) \) which implies \( x(s) = e_{N+1} \).

Second, the posterior distribution if treatment \( N+k \) fails is exactly the same as the posterior distribution if test \( k \) had failed instead.
Lemma 2: If $x_k + x_{N+1} + x_{N+2} \neq 1$ then $x(f)$ for $f \in \theta_{C}(N+k)$ is exactly equal to $x(f)$ for $f \in \theta_{C}(k)$.

Proof: For $f \in \theta_{C}(N+k)$,

$$P(f) = \sum_{\ell=1}^{N+2} P(f|\ell, N+k)x_{\ell} = \sum_{\ell=1}^{N+2} P(f|\ell, k)x_{\ell} = P(f)$$

for $f \in \theta_{C}(k)$. Now for $f \in \theta_{C}(N+k)$,

$$Z_{1j}(f, N+k) = p_{ji}(N+k) P(f|j, N+k) = p_{ji} P(f|j, k)$$

for all $i, j$ except when $j = k$. In case $j = k$, $P(f|k, N+k) = 0$ so $Z_{1k}(f, N+k) = 0$ for all $i$. However, $Z_{1k}(f, k) = 0$ also so $Z(f, N+k) = Z(f, k)$. Hence $x(f)$ is the same whether $f \in \theta_{C}(N+k)$ or $f \in \theta_{C}(k)$.

Since treatment $N+k$ has the same diagnostic power as test $k$ plus the ability to heal the patient it might appear that test $k$ would never be used. The advantages of treatment must be considered in light of the cost structure, however, since treatment may be much more expensive than a test.

4.3. The Reward Structure

The general idea of the reward structure is the following. Associated with the performance of action $a$ there is an immediate cost $c_a$. If the patient then dies a penalty $-m$ is assessed. If, on the
other hand, the patient is restored to health via successful treatment or via accident a reward $w$ accrues. $w$ and $-m$ are obviously simplifications of very complex realities. $w$ may represent the collection of fees, prestige in the medical community and the publication of a journal article while $-m$ may include considerations such as loss of prestige or a malpractice suit. Evaluation of $-m$ and $w$ may be quite difficult. That is, however, a problem in utility analysis which, though important, is left outside the scope of this paper. See the comments in Section 7.3. Also refer to Ginsberg [11] and Ginsberg and Offensend [12] for a discussion of utility analysis in a medical context.

Recall that $r(x,a) = r(a)x$. Since this is viewed as the expected cost of performing action $a$ when the state of knowledge is $x$, $r_k(a)$ is the cost of performing action $a$ if the patient has disease $k$.

Hence let

$$r(a) = (c_{a_1}, c_{a_2}, \ldots, c_{a_7}, 0, 0), \quad c_{a_k} \leq 0.$$  \hspace{1cm} (4)

This means that the cost, $c_{a_k}$, of action $a$ is incurred if the patient is in one of the disease states but not if he is in one of the terminating states. This is logical since the action would never be performed if the decision process has terminated.

It seems reasonable that a treatment for a disease would cost more than a test for the disease. Hence assume that

$$c_{N+1} \leq c_{i_i}, \quad i = 1, 2, \ldots, N.$$  \hspace{1cm} (5)
This makes tests attractive since failure of a test will be less costly than failure of the corresponding treatment. It also seems reasonable that doing nothing ought to be the least expensive action. Let

\( c_1 \leq c_0 \leq 0, \quad i = 1, 2, \ldots, N. \)

For the purposes of this chapter it will be assumed that \( c_0 = 0. \)

The reward vectors associated with termination must also be defined. If \( j = 1 \in \theta_{T2} \) is observed the patient is healthy. Hence let

\( U(1) = (0, \ldots, 0, w, 0), \quad w \geq 0. \)

Using Lemma 2.1,

\[ U(1) x(1) = (0, \ldots, 0, w, 0) e_{N+1} = w, \]

which is the appropriate reward for attaining the healthy state.

Similarly for \( j = 2 \in \theta_{T2} \) let

\( U(2) = (0, \ldots, 0, -m), \quad m \geq 0, \)

so that \( U(2) x(2) = -m, \) the penalty, for death.

Now suppose that treatment \( N+k \) is performed and success is observed. This means that \( x(s) = e_{N+1} \) by Lemma 1. Hence define
(9) \[ U(s) = (0, \ldots, 0, w, 0), \quad s \in \Theta_{T_k}(N+k), \quad k = 1, 2, \ldots, N, \]
so that \( U(s) x(s) = w. \) Note that \( U(s) = U(l). \) If, on the other hand, test \( k \) is performed and success is observed then the patient had disease \( k \) when the test was performed and the decision process stops. Medically one would then expect to treat disease \( k \) via \( N+k \in A. \)
Hence append a fictitious period to the process in which treatment \( N+k \) is applied. This costs an amount \( c_{N+k} \) to begin with. If the patient still has disease \( k \) the treatment will heal him, resulting in a return of \( w. \) If he has made a transition to disease \( i \neq k \) the treatment will have no effect. During the fictitious period the patient may die with probability \( p_{1,N+2} \) or become healthy with probability \( p_{1,N+1}, \)
receiving \( -m \) or \( w \) respectively. Hence for \( s \in \Theta_{T_l}(k), \) \( k = 1, 2, \ldots, N, \)
\[ U(s) = (c_{N+k} + wp_{1,N+1} - mp_{1,N+2}, \ldots, c_{N+k} + wp_{k-1,N+1} - mp_{k-1,N+2}, \ldots, c_{N+k} + wp_{N,N+1} - mp_{N,N+2}, w, -m). \]
Thus \( U(s) x(s) \) represents the appropriate expected reward described above since \( x_i(s) \) is the probability the patient is in state \( i \) after the successful outcome to test \( k \) is observed.

Finally consider \( U^0, \) the terminal reward vector. This represents the case where the end of the decision process is reached without achieving successful diagnosis or treatment. No treatment would be applied in this case, since the patient's disease is unknown. To be consistent with the above argument nothing would be done during the
fictitious period implying that any reward received is due to chance. If the patient has disease \( k \) he will get healthy with probability \( F_{k,N+1} \) and die with probability \( P_{k,N+2} \). Thus define

\[
U^0 = (w_{p_1,N+1} - m_{p_1,N+2}, \ldots, w_{p_N,N+1} - m_{p_N,N+2}, w, -m)
\]

This completes the formulation of Model I. In the next section some of the implications of this formulation are analyzed.

4.4. The Optimal Policy for Model I

Since there is one test and one treatment for each disease it seems reasonable that treatment \( N+i \) ought to be optimal for those states of knowledge which have a high probability of disease \( i \), i.e., those \( x \) near \( e_i \). Test \( i \) should be optimal for cases where disease \( i \) is the likely candidate but with insufficient certainty to apply treatment \( N+i \). The physician might observe only in cases where there is a high degree of uncertainty. Hence one might conjecture that the area where treatment \( N+i \) is optimal is a convex region with \( e_i \) as a corner point. The optimal region for test \( i \) might form a boundary around the region where treatment \( N+i \) is optimal so that the union of the two regions is convex. Figure 4 illustrates this conjecture policy for three diseases.

Unfortunately, this conjecture is too strong as shown in Appendix B. The convexity conditions fail. What can be shown is that under reasonable hypotheses the region where treatment \( N+i \) is optimal is star-shaped [31, p. 57] at \( e_i \). The region where test \( i \) is optimal then forms a halo around the treatment region. A set \( R \) is said to be star-shaped at a
point $e$ if $x \in R$ implies that $\alpha x + (1-\alpha)e \in R$ for all $\alpha$ such that $0 \leq \alpha \leq 1$. A set $R$ is halo-shaped at $e$ if $x_1, x_2 \in R$ and $x_2 = \beta x_1 + (1-\beta)e$ for some $\beta$, $0 \leq \beta \leq 1$ imply that $\alpha x_1 + (1-\alpha)x_2 \in R$ for all $\alpha$ such that $0 \leq \alpha \leq 1$. Figure 5 illustrates the situation.
The proofs of the results are based on recursively analyzing the specific form that \( H^n(a_1, \ldots, a_n)(\cdot) \) exhibits under the hypotheses of Model I. Recall first, however, that the scaling theorem (Theorem 3.1) implies that \( x_{N+1} \) and \( x_{N+2} \) may be set equal to zero without any loss of generality. This assumption will be made for the remainder of this chapter unless otherwise noted.

Consider \( H^n(N+i, a_2, \ldots, a_n)(e_i) \), i.e., the situation where the correct treatment is applied to a patient who has disease \( i \). A simple examination of \( Z(f) \) will show that \( Z(f)e_i \) is the zero vector. Hence \( \Omega(N+i, a_2, \ldots, a_{k-1})e_i \) is the zero vector for \( k \geq 2 \).

Also \( Z(1)e_i = Z(2)e_i = (0, \ldots, 0)^t \) for \( 1, 2 \in \Theta(N+i) \).

Application of these facts along with the definitions of \( U(s) \) and \( r(N+i) \) in Lemma 3.1 yields

\[
(H^n(N+i, a_2, \ldots, a_n)(e_i) = r(N+i)e_i + U(s)Z(s)e_i
\]

\[
= c_{N+i} + w
\]

This expression is independent of \( (a_2, \ldots, a_n) \), hence all the hyperplanes associated with treatment \( N+i \) have the same value at \( e_i \).

Intuitively, it makes no difference what you plan to do later; if the patient has disease \( i \) and you treat for it now, the return is the same. This significant fact holds for \( i = 1, 2, \ldots, N \). Similarly for test \( i \),

67
(13) \[ H^n(i, a_2, \ldots, a_n)(e_i) \]
\[ = c_i + (1 - p_{i,N+1} - p_{i,N+2})c_{N+1} + w(p_{ii} + p_{i,N+1} + \sum_{j \neq i}^{N} p_{ij}p_{j,N+1}) \]
\[ - m(p_{i,N+2} + \sum_{j=1}^{N} p_{ij}p_{j,N+2}), \quad i = 1, 2, \ldots, N. \]

Since the coefficient of \( w \) in equation (13) is less than one and \( c_{N+1} \leq c_i \),

(14) \[ H^n(N+i, a_2, \ldots, a_n)(e_i) \geq H^n(i, a_2, \ldots, a_n)(e_i) \].

Thus, if the patient has disease \( i \) it is better to immediately treat him than to test for the disease and then apply treatment.

Equations (12) and (13) also provide the key to proving the original propositions. Suppose treatment \( N+i \) is optimal at \( x \). Equation (12) implies that the situation resembles that of Figure 6 below.

If action \( N+i \) is optimal at \( e_i \) then the uppermost of the hyperplanes associated with \( N+i \) will dominate all other hyperplanes in some region around \( e_i \). Suppose treatment \( N+i \) is optimal at \( x \) so that

\[ V^n(x) = H^n(N+i, a_2^*, \ldots, a_n^*)(x) \]

for some \( (a_2^*, \ldots, a_n^*) \). Then, as the picture suggests, the entire line segment between \( x \) and \( e_i \) is associated with the uppermost
treatment hyperplane, implying that treatment $N+i$ is optimal for any point between $x$ and $e_i$. The immediate task, then, is to determine conditions which are sufficient to guarantee that treatment $N+i$ is optimal at $e_i$. As before, let $\delta^{n}(\cdot)$ be the optimal $n$-period decision rule.

**Theorem 1:** For Model I suppose

\[(15) \quad c^{N+i} + w \geq w(p_{1,N+1} + \sum_{j=1}^{N} p_{ij} p_{j,N+1}) - m(p_{1,N+2} + \sum_{j=1}^{N} p_{ij} p_{j,N+2})\]

and
\[(16) \quad c_{N+i} + w \geq \sum_{j=1}^{N} (c_{N+j} + w)p_{ij} + wp_{i,N+1} - mp_{i,N+2} \]

for \( i = 1, 2, \ldots, N \). Then

1. \( V^n(e_i) = c_{N+i} + w, \ i = 1, 2, \ldots, N \),
2. \( \delta^n(e_i) = N+i, \ i = 1, 2, \ldots, N \), and
3. \( V^n(e_{N+i}) = w, \ V^n(e_{N+2}) = -m \).

**Proof:** Proceed by induction. For \( n = 1 \) and \( i = 1, 2, \ldots, N \)

\[ V^1(e_i) = \max_{a \in A} H^1(a)(e_i) \]

\[ = \max \begin{cases} H^1(N+i)(e_i) = c_{N+i} + w & a = N+i \[ \vdots \] \end{cases} \]

Now \( H^1(i)(e_i) \leq H^1(N+i)(e_i) \) by (14). Also

\[ \Omega(a)e_i = Z(f)e_i = (p_{i1}, \ldots, p_{i,N+2})^T \quad \text{for} \ a \neq i. \]

Thus

\[ U^0 \Omega(a)e_i = w(p_{i,N+1} + \sum_{j=1}^{N} p_{ij} p_{j,N+1}) - m(p_{i,N+2} + \sum_{j=1}^{N} p_{ij} p_{j,N+2}) \].

70
Since \( c_0 \geq c_a \) for all \( a \in A \),

\[
V^1(e_i) = \max \left\{ \begin{array}{ll}
    c_{N+i} + w = H_{N+i}(e_i) & \text{if } a = N+i \\
    w(p_{i,N+1} + \sum_{j=1}^{N} p_{ij}p_{j,N+1}) - m(p_{i,N+2} + \sum_{j=1}^{N} p_{ij}p_{j,N+2}) & \text{if } a = 0
\end{array} \right.
\]

\[= c_{N+i} + w\]

by (15). Hence \( \delta^1(e_i) = N+i \). It is trivial to show that \( V^1(e_{N+1}) = w \)
and \( V^1(e_{N+2}) = -m \).

Now suppose the result is true for \( n-1 \). Then

\[
V^n(e_i) = \max_{a \in A} \{ r(a)e_i + U(1)Z(1)e_i + U(2)Z(2)e_i + U(s)Z(s)e_i + V^{n-1}(Z(f)e_i) \}
\]

\[= \max \left\{ \begin{array}{ll}
    c_{N+i} + w & \text{if } a = N+i \\
    c_a + V^{n-1}(Z(f)e_i) & \text{if } a \neq i, N+i
\end{array} \right.
\]

But \( Z(f)e_i = (p_{i1}, \ldots, p_{i,N+2})^t \) for all \( a \neq i, N+i \) so

\[V^n(e_i) = \max \left\{ \begin{array}{ll}
    c_{N+i} + w & \text{if } a = N+i \\
    V^{n-1}((p_{i1}, \ldots, p_{i,N+2})^t) & \text{if } a = 0
\end{array} \right.
\]

Now \( V^{n-1}(\cdot) \) is convex so
\[ v^{n-1}(p_{i1}, \ldots, p_{i,N+2}) \leq \sum_{j=1}^{N+2} p_{ij} v^{n-1}(e_j) \]
\[ = \sum_{j=1}^{N} p_{ij} (c_{N+j} + w) + wp_{i,N+1} - mp_{i,N+2} \]
\[ \leq c_{N+i} + w \quad \text{by (16).} \]

Hence \( v^n(e_i) = c_{N+i} + w = h^n(N+i, a_2, \ldots, a_n)(e_i) \) and \( s^n(e_i) = N+i \).

Again \( v^n(e_{N+1}) = w \) and \( v^n(e_{N+2}) = -m \) trivially.

A couple of comments about the hypotheses of Theorem 1 are in order. Inequality (16) has the following medical interpretation. Suppose the patient has disease \( i \). The left side of (16) is simply the reward for immediately treating for disease \( i \) and consequently curing the patient. The right side of (16), on the other hand, is the expected reward associated with doing nothing the first period and then treating the disease the patient has next period. The inequality thus says that there is no point in waiting around if the patient has disease \( i \); the benefits of cure and/or avoidance of death outweigh the cost of performing the treatment. Medically this seems rather reasonable. Inequality (15) represents the same thing when there is only a single decision period remaining.

Conditions (15) and (16) may be respectively replaced by

\[ c_{N+i} \geq -m(p_{i,N+2} + \sum_{j=1}^{N} p_{ij} p_{j,N+2}) \quad \text{(17)} \]

and

\[ (m+w) p_{i,N+2} + (1 - p_{ii})c_{N+i} \geq 0 \quad \text{(18)} \]
These are sufficient (but not necessary) conditions and they are somewhat simpler. However, they lack the above medical interpretation.

In a "static" case the conditions simplify even further. Suppose that the patient's disease either stays the same, goes away, or kills him. Then

\[ p_{i1} + p_{i,N+1} + p_{i,N+2} = 1. \]

In this case (16) becomes

\[ c_{N+1} + w \geq p_{i1}(c_{N+1} + w) + wp_{i,N+1} - mp_{i,N+2} \]

which is true over a large range of reasonable values for \( c_{N+1}, m, w, \) and \( p_{i1}. \)

Theorem 1 may be extended to the infinite horizon case very simply. In doing so inequality (15) can be dropped from the hypothesis. Intuitively, (15) is only used to start the induction; as \( n \to \infty \) this washes out.

**Corollary 1:** Suppose for Model I that (16) holds for \( i = 1, 2, \ldots, N. \) Then

1. \( v^*(e_i) = c_{N+1} + w, \ i = 1, 2, \ldots, N, \)
2. \( v^*(e_i) = N+1, \ i = 1, 2, \ldots, N, \) and
3. \( v^*(e_{N+1}) = w \) and \( v^*(e_{N+2}) = -m. \)

**Proof:** It must be shown that the proposed values satisfy the appropriate contraction map, namely
\[ V^*(e_i) = \max_{a \in A} (r(a)e_i + U(1) Z(1)e_i + U(2) Z(2)e_i + U(s) Z(s)e_i + V^*(Z(e_i)) \],

But this is exactly what was shown in extending the induction from \( n-1 \) to \( n \) in Theorem 1.

Theorem 1 can now be used to prove the star-shapedness of the treatment regions. Recall from Chapter 2 that there was a finite, convex partition \( T^n \) associated with \( V^n(x) \) and \( \delta^n(x) \). Along the same lines let

\[ T^n(a_1, \ldots, a_n) = \{ x \in S | V^n(x) = H^n(a_1, \ldots, a_n)(x) \} \]

\( T^n(a_1, \ldots, a_n) \) is the set of points where \( H^n(a_1, \ldots, a_n)(x) \) is the maximal hyperplane, thus it corresponds to the set of points where \( V^n(x) = b_i x \) for some specific \( b_i \in B^n \). But this was the partition element \( T^n_i \). Actually the \( T^n(a_1, \ldots, a_n) \) are not quite the partition sets of Chapter 2, rather they are the \( T^n_i \) with their boundaries included.

**Theorem 2:** For Model I suppose \((i5)\) and \((16)\) hold for \( i = 1, 2, \ldots, N \). Then if \( \delta^n(x) = N+i \), \( \delta^n(\alpha x + (1-\alpha)e_i) = N+i \) for \( 0 \leq \alpha \leq 1 \), \( i = 1, 2, \ldots, N \) and \( n = 1, 2, \ldots \).

**Proof:** \( V^n(x) = \max H^n(a_1, \ldots, a_n)(x) \) where the maximum is over all possible sequences \( (a_1, \ldots, a_n) \). Now \( \delta^n(x) = N+i \) implies that \( V^n(x) = H^n(N+i, a^*_2, \ldots, a^*_n)(x) \) for some \( (a^*_2, \ldots, a^*_n) \), hence...
\( x \in T^n(N+i, a_2^*, \ldots, a_n^*) \). By Theorem 1, \( e^n_i = N+i \) so \( e_i \in T^n(N+i, a_2^*, \ldots, a_n^*) \) for some \((a_2^*, \ldots, a_n^*)\). But equation (12) implies that \( H^n(N+i, a_2^*, \ldots, a_n^*)(e_i) \) has the same value for all \((a_2^*, \ldots, a_n^*)\), so \( e_i \in T^n(N+i, a_2^*, \ldots, a_n^*) \). Since \( T^n(N+i, a_2^*, \ldots, a_n^*) \) is defined by a set of linear inequalities it is convex, hence \( \alpha x + (1-\alpha)e_i \in T^n(N+i, a_2^*, \ldots, a_n^*) \) for \( 0 \leq \alpha \leq 1 \). Thus \( e^n(\alpha x + (1-\alpha)e_i) = N+i \).

Theorem 2 says that the optimal n-period policy has treatment regions which are star-shaped at \( a_i \). As \( n \to \infty \) the number of hyperplanes becomes countably infinite but their structure remains the same. Hence the following corollary holds.

**Corollary 2:** For Model I suppose (16) holds for \( i = 1, 2, \ldots, N \). Then if \( e^n(x) = N+i \), \( e^n(\alpha x + (1-\alpha)e_i) = N+i \) for \( 0 \leq \alpha \leq 1 \) and \( i = 1, 2, \ldots, N \).

While Theorem 2 and Corollary 2 describe the form of the optimal policy for the treatment regions, they say nothing about the test regions. As indicated, one would expect the region where test 1 is optimal to form a halo around the region where treatment \( N+i \) is optimal (refer to Figure 5). Specifically the following can be shown.

**Theorem 3:** For Model I suppose \( x_1, x_2 \in S \) with \( x_2 = \beta x_1 + (1-\beta)e_i \) for some \( \beta, 0 \leq \beta \leq 1 \). Then \( e^n(\alpha x_1 + (1-\alpha)x_2) = i \) for all \( \alpha \), \( 0 \leq \alpha \leq 1 \).
Proof: By equation (13)

\[ H^n(i, a_2, \ldots, a_n)(e_i) = \]

\[ = c_i + (1 - p_{i,N+1} - p_{i,N+2})c_{N+i} + w(p_{ii} + \sum_{j \neq i} N_{i,j}p_{i,j} + \sum_{j=1}^{N} p_{i,j}p_{j,N+1}) \]

\[ - m(p_{i,N+2} + \sum_{j=1 \neq i}^{N} p_{i,j}p_{j,N+2}) \]

for all \( (a_2, \ldots, a_n) \).

Since \( \delta^n(x_1) = i \), \( V^n(x_1) = H^n(i, a_2^*, \ldots, a_n^*)(x_1) \) for some \((a_2^*, \ldots, a_n^*)\).

Combining these facts, \( H^n(i, a_2^*, \ldots, a_n^*)(\cdot) \) must be maximal among the hyperplanes associated with test \( i \) along \( e_1x_1 \), i.e., \( H^n(i, a_2^*, \ldots, a_n^*)(x) \geq H^n(i, a_2, \ldots, a_n)(x) \) for all \((a_2, \ldots, a_n)\) and \( x \) on the segment \( e_1x_1 \). Hence \( V^n(x_2) = H^n(i, a_2^*, \ldots, a_n^*)(x_2) \). Now \( V^n(\cdot) \) is convex so

\[ V^n(\alpha x_1 + (1-\alpha)x_2) \leq \alpha V^n(x_1) + (1-\alpha) V^n(x_2) \]

\[ = H^n(i, a_2^*, \ldots, a_n^*) (\alpha x_1 + (1-\alpha)x_2). \]

Since \( H^n(i, a_2^*, \ldots, a_n^*)(\cdot) \leq V^n(\cdot) \) equality holds in the above expression implying \( \delta^n(\alpha x_1 + (1-\alpha)x_2) = i. \)

Theorem 3 says that if test \( i \) is optimal at \( x_1 \) and at \( x_2 \) where \( x_2 \) is on the line between \( x_1 \) and \( e_i \) then test \( i \) is optimal for any point on the line between \( x_1 \) and \( x_2 \). Hence the test region is halo-shaped.
Corollary 3: For Model I, suppose $x_1, x_2 \in S$ with $x_2 = \beta x_1 + (1-\beta)e_1$ for some $\beta, 0 \leq \beta \leq 1$ and suppose $\delta^*(x_1) = \delta^*(x_2) = i$, $1 \leq i \leq N$. Then $\delta^*(\alpha x_1 + (1-\alpha)x_2) = i$ for $0 \leq \alpha \leq 1$.

Theorems 2 and 3 along with their corollaries provide a basic characterization of the form of the optimal policy under fairly general hypotheses. One possible extension which they do not consider is the question of whether the regions where treatment $N+1$ and test $i$ are optimal respectively are adjacent to each other. Appendix C presents some partial results suggestive of further research on this extension.

4.5. The Case $w = 0$

A potentially interesting subcase of Model I is when $w = 0$. Medically this corresponds to the assumption that avoidance of death, or some other major trauma, is the criterion by which decisions are made. This is not too unrealistic. Some medical cases exhibit "death-dominance" when the probability of death and/or the relative severity of the death consequence outweigh all other considerations.

The results of the previous section carry over.

Corollary 4: Suppose $w = 0$ in Model I and that

\[ c_{N+1} \geq \sum_{j=1}^{N} (c_{N+1}^{(j)}) p_{ij} - m p_{ij, N+2}, \quad i = 1, 2, \ldots, N. \]

Then if $\delta^*(x) = N+1$, $\delta^*(\alpha x + (1-\alpha)e_1) = N+1$ for $0 \leq \alpha \leq 1$, and if $\delta^*(x_1) = \delta^*(x_2) = i$ with $x_2 = \beta x_1 + (1-\beta)e_1$ for some $\beta \in [0,1]$ then $\delta^*(\alpha x_1 + (1-\alpha)x_2) = i$, $0 \leq \alpha \leq \frac{1}{i}$, $i = 1, 2, \ldots, N$. 

77
Remark: Inequality (21) is a rather weak one. Since avoidance of disaster is the sole concern it seems that \( m \) should be very large in comparison to \( c_{N+1} \). Essentially (21) requires that the cost of performing a treatment be less than the expected value of disaster if the treatment were not done. While this is simply a special case of Model I it helps to illustrate the nature of the sufficient conditions for the star-shaped treatment and halo-shaped test regions and adds credence to those results.
5.1. **Formulation**

Model II is very similar to Model I in many respects. Hence this formulation only stresses the points of divergence rather than giving a repetitive reformulation.

As indicated in Section 2.4 the major difference between the two models is the status of the healthy state. Model I assumed that health was a terminating state in $D_2$ and hence was observable in the sense that there was an outcome in $T_2$ which indicated that the patient was in the healthy state. Medically it can be argued that this is not the case. There is no single outcome which indicates that a person is healthy; instead health is defined as the absence of disease. Certainly a major medical problem is the determination of health, i.e., when to cease tests and treatments based on the physician's belief that sufficient information has been gathered to reasonably believe that the patient is healthy.

In Model II this is explicitly taken into account by assuming that health is an absorbing disease state. Formally, assume

$$D = \{1, 2, \ldots, N, N+1, N+2\} \text{ where } D_1 = \{1, 2, \ldots, N, N+1\} \text{ and } D_2 = \{N+2\}. \text{ As before states } 1 \text{ through } N \text{ are various disease states, state } N+1 \text{ is health, and state } N+2 \text{ is death. Since health is assumed to be an absorbing state,}

\begin{equation}
\begin{array}{c}
\nu_{N+2, N+1} = 1, \\
a \in A
\end{array}
\end{equation}
Also since the set of terminating states is now the singleton \( \{N+2\} \),
the set \( \theta_{T2} \) of outcomes indicating that the patient is in a terminating
state has only one element. For the sake of consistency with the previous
chapter let \( \theta_{T2} = \{2\} \).

All of the actions, tests, treatments and the option of observing
the patient, which were available in Model I are available in Model II.
The conditional probabilities and transition probabilities, \( P(\|i, a) \)
and \( p_{ik}(a) \) respectively, remain unchanged with three exceptions. First,
\( j = 1 \in \theta_{T2} \) no longer exists. Second, the probability of the negative
outcome, \( f \), to either a test or observation of the patient is one if the
patient is healthy, i.e.,

\[
(2) \quad P(f|N+1, a) = 1, \quad a = 0, 1, \ldots, N.
\]

Third, the probability of successful treatment is one for a healthy
patient since there is nothing to cure, i.e.,

\[
(3) \quad P(s|N+1, a) = 1, \quad a = N+1, N+2, \ldots, 2N.
\]

One new action, denoted \( 2N+1 \), is available in Model II. This
consists of the option of sending the patient home and terminating the
decision process. It seems reasonable that the physician halt the
diagnosis and treatment process if the patient is healthy. There is no
need for any further tests or treatments, or even observation. Hence
sending the patient home is analogous to treating for the healthy state.

Since sending the patient home automatically terminates the
decision process the set of outcomes \( \theta(2N+1) \) may be defined by
\[ \theta(2N+1) = \theta_C(2N+1) \cup \theta_{T1}(2N+1) \cup \theta_{T2} \] where \( \theta_C(2N+1) = \emptyset, \theta_{T1}(2N+1) = \{ s \} \) and \( \theta_{T2} = \{ 2 \} \). In other words, there are only two outcomes, \( s \) indicating that the patient was sent home and \( 2 \) indicating that the patient had died. The appropriate conditional probabilities are

\[\begin{align*}
P(s \mid i, 2N+1) &= 1, \quad i \neq N+2, \\
P(2 \mid N+2, 2N+1) &= 1.
\end{align*}\]

No medical procedure is applied to the patient. It thus seems reasonable that the "natural" transition probabilities between diseases apply to action \( 2N+1 \), i.e.,

\[ p_{ik}(2N+1) = p_{ik}, \quad \text{for all } i, k. \]

This may seem irrelevant due to the nature of the action but it is important in defining the reward structure in the case where a non-healthy patient is incorrectly sent home.

As in the previous chapter there are three returns, the cost of performing an action, a penalty for allowing the patient to die and a reward for restoring him to health. Let \( c_a \leq 0 \) be the cost of action \( a \). Recall that \( r(a) \cdot x \) is the expected cost of performing action \( a \). Since \( c_a \) now accrues for any underlying state except death, i.e., for any state in \( D_1 \), define

\[ r(a) = (c_a, c_a, \ldots, c_a, 0) \in \mathbb{R}^{N+2}. \]
Suppose a treatment is performed and it is successful. Then \( U(s) \) as defined in Model I remains appropriate so

\[
U(s) = (0, \ldots, 0, w_j, 0), \quad s \in \theta_T(k), \quad k = N+1, N+2, \ldots, 2N.
\]

On the other hand, suppose test \( k \) is successfully performed. Since health is not observable, treatment \( N+k \) will be performed in the fictitious period unless the patient has died. If the patient starts the fictitious period in state \( i \), the expected return is \( c_{N+k} + w_{p_{1,N+1}} - m_{1,N+2} \). Thus define

\[
(7) \quad U(s) = (c_{N+k} + w_{p_{1,N+1}} - m_{1,N+2}, \ldots, c_{N+k} + w_{p_{k-1,N+1}} - m_{k-1,N+2}, c_{N+k} + w_j, w_j - m) \quad s \in \theta_T(k), \quad k = 1, 2, \ldots, N.
\]

The terminal reward remains the same as in Model I,

\[
(8) \quad U^0 = (w_{p_{1,N+1}} - m_{1,N+2}, \ldots, w_{p_{N,N+1}} - m_{N,N+2}, w_j, -m).
\]

Similarly, for \( j = 2 \in \theta_{T2} \)

\[
(9) \quad U(2j) = (0, \ldots, 0, 0, 0, -m).
\]
Hence it only remains to define $U(s)$ for $s$ associated with sending the patient home in order to completely define the reward structure for Model II. Suppose the physician decides to send the patient home and the patient is not dead. Then $s \in \Theta_T(2N+1)$ is "observed". The period associated with the choice of action $2N+1$ may then be considered a fictitious period. If the patient is restored to health or healthy to begin with a reward $w$ accrues. If he dies there is a penalty $-m$. If he is sent home with some disease assume there is a penalty $-m'$ where $m' < m$. The idea here is that there is a penalty for stopping the decision process if the patient is still sick, but the penalty is smaller than that associated with letting the patient die. The probabilities associated with these various possibilities are simply those given in $x(s)$ since $p_{ij}(2N+1) = p_{ij}$. Hence define

$$U(s) = (-m', -m', \ldots, -m', w_r, -m) , \quad s \in \Theta_T(2N+1).$$

Then $U(s) x(s) P(s)$ represents the appropriate expected return associated with sending the patient home.

5.2. The Optimal Policy for Model II

The previous section, along with Sections 4.1 and 4.2, defines the transition and reward structures for Model II. This section presents a basic characterization of the optimal policy. Results analogous to Theorems 4.2 and 4.3 on star-shaped treatment regions and halo-shaped test regions are presented. Since the methods are parallel to those of Chapter 4, most proofs are abbreviated or omitted.
Once again the relationships between the $H^n(a_1, \ldots, a_n)(x)$ are analyzed. Recall that $H^n(a_1, \ldots, a_n)(x)$ is the n-period expected return if the starting state is $x$ and if action $a_k$ is used in period $k$ given that $a_{k-1}$ failed in period $k-1$. From Lemma 3.1

(11) $H^n(a_1, \ldots, a_n)(x)$

$$= \sum_{k=1}^{n} r(a_k) \Omega(a_1, \ldots, a_{k-1})x$$

$$+ \sum_{k=1}^{n} \sum_{j \in \Theta(a_k)} U(j) Z(j) \Omega(a_1, \ldots, a_{k-1})x$$

$$+ U^0 \Omega(a_1, \ldots, a_n)x.$$ 

Suppose, however, $a_{\ell} = 2N+1$, i.e., send the patient home. Then $a_{\ell+1}, a_{\ell+2}, \ldots, a_n$ are irrelevant since they would never be performed. The above formula can be made consistent with this case by adopting the convention that $Z(f, 2N+1) = 0$ so that

$$\Omega(a_1, \ldots, a_{k-1}) = Z(f, a_{k-1}) Z(f, a_{k-2}) \cdots Z(f, a_1) = 0$$

for $k = \ell+1, \ell+2, \ldots, n$.

Now suppose the patient has disease $i$, $1 \leq i \leq N$ and treatment $N+1$ is performed, i.e., $a_1 = N+1$. Then from equation (11) with appropriate substitutions,

(12) $H^n(N+1, a_2, \ldots, a_n)(e_i) = c_{N+1} + w_i$, $i = 1, 2, \ldots, N$.
Similarly,

\[(13) \quad \mathbb{H}^n(i, a_2, \ldots, a_n)(e_1) = c_1 + (1 - P_{i, N+2})^N + w(P_{i, i} + P_{i, N+1} + \sum_{j=1}^{N} P_{i, j}P_{j, N+1})
\]

\[= m(P_{i, N+2} + \sum_{j=1}^{N} P_{i, j}P_{j, N+2}) \quad ; \quad i = 1, 2, \ldots, N.\]

In Model II it is also possible that the patient is healthy since \(N+1 \in D_1\). Suppose \(x = e_{N+1}\) and \(a_1 = 2N+1\). Then

\[(14) \quad \mathbb{H}^n(2N+1, a_2, \ldots, a_n)(e_{N+1}) = r(2N+1)e_{N+1} + U(s)Z(s) e_{N+1}
\]

\[= c_{2N+1} + U(s)Z_{x, N+1}(s)
\]

\[= c_{2N+1} + w\]

since \(P_{N+1, N+1}(2N+1) = 1\). Thus the return for sending a healthy patient home is \(c_{2N+1} + w\).

In Model I it was assumed that the cost of observation was zero, i.e., \(c_0 = 0\), though none of the results depended on the assumption. It seems clear, however, that sending the patient home is cheaper than observing the patient for some period. This is due to the fact that observation of the patient involves continuing the decision process. One could thus think of \(c_0\) as the cost of the next office visit. Hence, for Model II assume that
\[
(15) \quad c_a \leq c_0 \leq c_{2N+1} \leq 0; \quad a = 1, 2, \ldots, 2N.
\]

The first theorem gives sufficient conditions which guarantee that treatment \(N+i\) is optimal at \(e_i\) and that sending the patient home is optimal at \(e_{N+1}\).

**Theorem 1:** For Model II suppose

\[
(16) \quad c_{N+i} + w \geq c_0 + w(p_{i,N+1} + \sum_{j=1}^{N} p_{ij} p_{j,N+1})
\]
\[
- m(p_{i,N+2} + \sum_{j=1}^{N} p_{ij} p_{j,N+2}),
\]

\[
(17) \quad c_{N+i} + w \geq c_{2N+1} + wp_{i,N+1} - m\sum_{j=1}^{N} p_{ij} - mp_{i,N+2},
\]

and

\[
(18) \quad c_{N+i} + w \geq c_0 + \sum_{j=1}^{N} p_{ij}(w + c_{N+j}) + p_{i,N+1}(w + c_{2N+1}) = mp_{i,N+2}
\]

for \(i = 1, 2, \ldots, N\). Then

1. \(V^n(e_i) = w + c_{N+i}, \quad i = 1, 2, \ldots, N, N+1,\)
2. \(u^n(e_i) = N+i, \quad i = 1, 2, \ldots, N+1,\) and
3. \(V^n(e_{N+2}) = -m.\)

**Proof:** The argument is basically the same as that for Theorem 4.1. Proceed by induction. For \(n = 1,\) and \(i = 1, 2, \ldots, N,\)
$$V^1(e_i) = \max_{a \in A} \mathcal{H}^1(a)(e_i)$$

$$= \begin{cases} 
  c_{N+1} + w ; & a = N+1 \\
  \mathcal{H}^1(i)(e_i) ; & a = i \\
  c_a + U^2 \Omega(a)e_i ; & a \neq i, N+1, 2N+1 \\
  c_{2N+1} + U(s) Z(s, 2N+1)e_i ; & a = 2N+1 
\end{cases}$$

Now $c_{N+1} + w \geq \mathcal{H}^1(i)(e_i)$ and $U(s) Z(s, 2N+1)e_i = (-m', ..., -m', w, -m) Z(s, 2N+1)e_i$. But $z_{ji}(s, 2N+1) = p_{ij} p(s|i, 2N+1) = p_{ij}$ so

$$U(s) Z(s, 2N+1)e_i = -m' \sum_{j=1}^{N} p_{ij} + p_{i,N+1} w - p_{i,N+2} m .$$

Hence

$$V^1(e_i) = \max \left\{ \begin{array}{ll}
  c_{N+1} + w ; & a = N+1 \\
  c_0 + w(p_{i,N+1} + \sum_{j=1}^{N} p_{ij} p_{j,N+1}) - m(p_{i,N+2} + \sum_{j=1}^{N} p_{ij} p_{j,N+2}) ; & a = 0 \\
  c_{2N+1} + wp_{i,N+1} - m' \sum_{j=1}^{N} p_{ij} - mp_{i,N+2} ; & a = 2N+1 
\end{array} \right\}$$

$$= c_{N+1} + w$$

by (16) and (17).

Thus $\delta^1(e_i) = N+1, i = 1, 2, ..., N$. Now
\[ V^1(e_{N+1}) = \max \begin{cases} c_{2N+1} + w; & a = 2N+1 \\ c_a + U(0) \Omega(a)e_{N+1}; & a = 0, 1, \ldots, N \\ c_a + U(s) Z(s,a)e_{N+1}; & a = N+1, \ldots, 2N. \end{cases} \]

But \( \Omega(a) = Z(s,a) \) and \( z_{j,N+1}(s,a) = p_{N+1,j}(a) P(s|N+1,a) = 0 \) for \( a = 0, 1, \ldots, N \) unless \( j = N+1 \). Hence \( U^0 \Omega(a)e_{N+1} = U^0 e_{N+1} = w \).

Similarly \( U(s) Z(s,a)e_{N+1} = w \) so

\[ V^1(e_{N+1}) = \max \begin{cases} c_{2N+1} + w j; & a = 2N+1 \\ c_a + w; & a \neq 2N+1 \end{cases} \]

\[ = c_{2N+1} + w \quad \text{by (15).} \]

Hence \( \delta^1(e_{N+1}) = 2N+1 \). It is trivial to show that \( V^1(e_{N+2}) = -m \).

Now suppose the result is true for \( n-1 \). Then for \( 1 \leq i \leq N \),

\[ V^n(e_i) = \max_{a \in A} \{ r(a)e_i + U(0) Z(0)e_i + U(s) Z(s)e_i + V^{n-1}(Z(f)e_i) \} \]

\[ = \max \begin{cases} c_{N+i} + w; & a = N+i \\ c_a + V^{n-1}(Z(f)e_i); & a \neq i, N+i, 2N+1 \\ c_{2N+1} + U(s) Z(s,2N+1)e_i; & a = 2N+1 \end{cases} \]

\[ = \max \begin{cases} c_{N+i} + w; & a = N+i \\ c_0 + V^{n-1}(p_{i1}, \ldots, p_{i,N+2}); & a = 0. \end{cases} \]
But \( v^{n-1}(\cdot) \) is convex so

\[
c_0 + v^{n-1}(p_{i1}, \ldots, p_{i,N+2}) \leq c_0 + \sum_{j=1}^{N+2} p_{ij} v^{n-1}(e_j)
\]

\[
\leq c_0 + \sum_{j=1}^{N} p_{ij}(c_{N,j} + w) + p_{i,N+1}(c_{2N+1} + w) - p_{i,N+2}^m
\]

\[
\leq c_{N+1} + w \quad \text{by (18).}
\]

Hence \( v^n(e_{i1}) = c_{N+1} + w \) and \( s^n(e_{i1}) = N+1, i = 1, 2, \ldots, N. \)

For the healthy state

\[
v^n(e_{N+1}) = \max \begin{cases} 
  c_{2N+1} + w; & a = 2N+1 \\
  c_a + v^{n-1}(Z(f,a)e_{N+1}); & a = 0, 1, \ldots, N \\
  c_a + U(s)Z(s,a)e_{N+1}; & a = N+1, \ldots, 2N
\end{cases}
\]

\[
= \max \begin{cases} 
  c_{2N+1} + w; & a = 2N+1 \\
  c_a + v^{n-1}(e_{N+1}); & a = 0, \ldots, N \\
  c_a + w; & a = N+1, \ldots, 2N
\end{cases}
\]

But \( v^{n-1}(e_{N+1}) = c_{2N+1} + w \leq w \) so \( v^n(e_{N+1}) = c_{2N+1} + w \) and \( s^n(e_{N+1}) = 2N+1. \) Finally \( v^n(e_{N+2}) = -m \) trivially.

Just as in Chapter 4, Theorem 1 may be directly extended to the infinite horizon case by noting that the appropriate functional equation is satisfied. In the process the hypotheses corresponding to the 1-period model may be dropped since they no longer matter.
**Corollary 1:** For Model II suppose (18) holds for \( i = 1, 2, \ldots, N \).

Then

1. \( V^*(e_i) = w + c_{N+i}, \ i = 1, 2, \ldots, N, N+1, \)
2. \( \delta^*(e_i) = N+i, \ i = 1, 2, \ldots, N+1, \) and
3. \( V^*(e_{N+2}) = -m. \)

Corollary 1 gives sufficient conditions under which the intuitively appropriate action is optimal at each corner of the state space \( S \).

Moreover the value of all the hyperplanes associated with treatment \( N+i \) is the same at \( e_i \). The same holds for the hyperplanes associated with sending the patient home, action \( 2N+1 \), at \( e_{N+1} \). Hence the same method of proof used in Theorem 4.2 may be applied here to show that the treatment regions and the region where sending the patient home is optimal are star-shaped.

**Theorem 2:** For Model II suppose (18) holds for \( i = 1, 2, \ldots, N \).

Then if \( \delta^*(x) = N+i, \delta^*(\alpha x + (1-\alpha)e_i) = N+i \) for \( 0 \leq \alpha \leq 1 \) and \( i = 1, 2, \ldots, N, N+1. \)

Note that Theorem 2 applies to the region where action \( 2N+1 \), sending the patient home, is optimal as well as to the treatment regions. In that sense it is a slightly stronger result than its counterpart in the previous chapter.

In the same fashion as before, the test regions form halos around the corresponding treatment regions.

**Theorem 3:** For Model II suppose \( x_1, x_2 \in S \) with \( x_2 = \beta x_1 + (1-\beta)e_i \) for some \( \beta, 0 \leq \beta \leq 1 \), and some \( i, 1 \leq i \leq N \). Suppose \( \delta^*(x_1) = \delta^*(x_2) = i. \)

Then \( \delta^*(\alpha x_1 + (1-\alpha)x_2) = i \) for all \( \alpha, 0 \leq \alpha \leq 1. \)
Proof: The proof is essentially identical to Theorem 4.3.

Unfortunately Theorem 3 does not include the action 0, namely observing the patient. This is due to the fact that \( H^n(0, a_2, \ldots, a_n)(e_{n+1}) \) depends upon the sequence \( (a_2, \ldots, a_n) \). Hence there is no common value to work from.

Theorems 2 and 3 yield an optimal policy which can be described verbally as follows. Each corner of \( S \) representing one of the diseases is the vertex of a star-shaped region in which treating the patient for that disease is optimal. Surrounding each of these treatment regions is a halo-shaped region where testing for the respective disease is optimal. At the corner of \( S \) representing health, \( e_{n+1} \), there is a star-shaped region where sending the patient home is optimal. Action 0, the option of doing nothing while continuing the decision process, i.e., observing the patient for some period, fills in the gaps around the other regions. Figure 7 is suggestive of the situation.

As in the previous chapter, Theorems 2 and 3 leave one question unanswered. Are the test regions directly adjacent to their respective treatment regions as Figure 7 suggests? Refer to Appendix C for some partial results.

![Figure 7: Optimal Policy for Model II](image_url)
CHAPTER 6
LEXICOGRAPHIC IMPROVEMENT

6.1. The Conjecture

This chapter presents some results which extend the basic
characterization of the optimal policy given in the previous two chapters.
The theorems are incomplete but are suggestive of further research. The
reader may omit this chapter and skip directly to the conclusion without
loss of continuity should he desire. The results presented here apply
to either Model I or Model II since they depend only on the fact that

\[ V^*(x) = \sup_{a_1, a_2, \ldots} H(a_1, a_2, \ldots)(x) \]

where \( H(a_1, a_2, \ldots)(\cdot) \) is the infinite horizon hyperplane associated
with the binary policy \((a_1, a_2, \ldots)\).

Suppose \( \delta^*(x) = N+i \) and that \( y \) is "closer" to \( e_i \), though
not necessarily on the line segment between \( x \) and \( e_i \). Does \( \delta^*(y) = N+i \)?
The numerical example of Appendix B suggests that Euclidean distance is
inappropriate. Hence lexicographic improvement is considered, i.e., the
component by component difference between \( y \) and \( x \). First a precise
definition is needed.

**Definition 1**: \( y \in S \) is lexicographically greater than or equal to
\( x \in S \) with respect to state \( i \), written \( y \succeq_i x \), if and only if,
1. \( y_i \geq x_i \), and
2. \( y_k \leq x_k \) for \( k = 1, 2, \ldots , M; k \neq i \).

The conjecture can now be made explicit. Suppose \( y \geq x \) and \( \delta^*(x) = N+i \). Under what conditions will \( \delta^*(y) = N+i \), i.e., when will treatment \( N+i \) be optimal at \( y \)?

6.2. Results

The analysis proceeds along the following lines. Suppose \( y \geq x \) and treatment \( N+i \) is optimal at \( x \). By trading off one component of \( x \) with \( x_i \) at a time, a piecewise linear path connecting \( x \) and \( y \) can be created. Further, each of the pieces will be parallel to one of the edges of \( S \) which emanates from \( e_i \). The idea is to show that \( N+i \) is optimal at each corner on the path. This is done by comparing the positions of the hyperplanes along each segment of the piecewise linear path with their positions along the corresponding parallel edge of \( S \).

Since the sections of the hyperplanes along the segments are essentially straight lines their "slopes" may be examined. Because the segments and the edges of \( S \) are parallel the hyperplanes have the same slopes in both cases, though in absolute magnitude the hyperplanes may have moved up or down. By imposing certain conditions on the "slopes" the desired result will be achieved. Figure 8 illustrates the situation.

It is first necessary to show that the section of \( \mathbb{R}^n \) defined by the line \( \overline{zx} \) (see Figure 8) has the same slope as the section defined by the line \( e_i \cdot e_k \) parallel to \( \overline{zx} \), where \( z = x + \beta(e_i - e_k) \) for some \( \beta \geq 0 \).
Figure 8: Lexicographic Improvement

**Lemma 1:** Suppose \( \ell_1 \) and \( \ell_2 \) are two parallel lines in \( \mathbb{R}^M \) and \( f(\cdot) \) is a linear function, \( f: \mathbb{R}^M \to \mathbb{R} \). Any transversal \( \ell_3 \) of \( \ell_1 \) and \( \ell_2 \) defines a 1-1 correspondence between points \( x_1 \) on \( \ell_1 \) and points \( x_2 \) on \( \ell_2 \); see Figure 9. Let \( x_1 \) and \( x_2 \) be any such points, then \( f(x_2) = f(x_1) + c \) where \( c \in \mathbb{R} \) is a constant.

Figure 9: Parallel Correspondence
Proof: Let \( x_1^0 \) and \( x_2^0 \) be the points at the intersection of \( l_1 \) with \( l_2 \) respectively. Define \( c = f(x_2^0) - f(x_1^0) \). Now
\[
f(x_1) = f(x_1^0 + (x_1 - x_1^0)) = f(x_1^0) + f(x - x_1^0)
\]
since \( f \) is linear. Similarly \( f(x_2) = f(x_2^0) + f(x_2 - x_2^0) \). But \( x_1 - x_1^0 = x_2 - x_2^0 \) by the correspondence. Hence
\[
f(x_2) - f(x_1) = f(x_1^0) + f(x_1 - x_1^0) - f(x_2^0) - f(x_2 - x_2^0) = c. \]

Lemma 1 shows that the sections of \( H^n \) corresponding to the lines \( zx \) and \( e_1 e_k \) differ only in absolute magnitude; they have the same slopes.

Some things are known about the relative positions of the various hyperplanes on the line \( e_1 e_k \); one would like to use that to say something about their positions on the line \( zx \).

Figure 10a schematically suggests the situation on the line \( e_1 e_k \). The lines \( g_1 \) and \( g_2 \) are sections of the hyperplanes \( H(N+i, e_2, \ldots)(\cdot) \), i.e., those associated with treatment \( N+i \), while the \( f \)'s are sections of other hyperplanes. Moving to the line segment \( zx \) the slopes of the sections remain constant but the lines move up or down. Sufficient conditions are desired so that Figure 10b accurately represents the situation, namely that if \( N+i \) is optimal at \( x \) (since \( g_2 \) is maximal) then one of the \( H^n \) associated with \( N+i \) will be optimal at \( z \) (\( g_1 \) in the picture).

Theorem 1: Let \([k, i]\) be an interval on the real line with \( i > k \) and suppose there is a collection of linear functions on \([k, i]\) with the following properties.
Figure 10a: Hyperplanes on \( e_i e_k \)

Figure 10b: Hyperplanes on \( zx \)
1. \( f_\alpha(x) = b_\alpha x + c_\alpha, \alpha = 1, 2, \ldots, \)

2. \( g_\beta(x) = d_\beta x + e_\beta, \beta = 1, 2, \ldots, \)

3. \( \sup_\alpha \{b_\alpha\} = b \leq d = \inf_\beta \{d_\beta\}. \)

Let \( \hat{f}_\alpha(x) = f_\alpha(x) + h_\alpha, \alpha = 1, 2, \ldots \) and \( \hat{g}_\beta(x) = g_\beta(x) + j_\beta, \) \( \beta = 1, 2, \ldots \), where \( h_\alpha, j_\beta \in \mathbb{R}. \) If \( k \leq x \leq l \) and

\[ \sup_\alpha \{\hat{f}_\alpha(x), \hat{g}_\beta(x)\} = \hat{g}_\beta(x) \quad \text{for some } \beta^* \]

then for \( x \leq z \leq l \)

\[ \sup_\alpha \{\hat{f}_\alpha(z), \hat{g}_\beta(z)\} = \hat{g}_\beta(z) \quad \text{for some } \beta. \]

**Proof:** Suppose \( \sup_\alpha \{\hat{f}_\alpha(x), \hat{g}_\beta(x)\} = \hat{g}_\beta(x). \) Then

\[ d_{\beta^*} x + e_{\beta^*} + j_{\beta^*} \geq b_\alpha x + c_\alpha + h_\alpha \quad \text{for } \alpha = 1, 2, \ldots. \]

Now \( z = x+t \) where \( t \geq 0. \) Thus

\[ g_{\beta^*}(z) = d_{\beta^*}(x+t) + e_{\beta^*} + j_{\beta^*}, \quad \text{and} \]

\[ f_\alpha(z) = b_\alpha(x+t) + c_\alpha + h_\alpha, \quad \alpha = 1, 2, \ldots. \]

But \( b_\alpha \leq d \leq d_{\beta^*} \) so \( b_\alpha t \leq d_{\beta^*} t \) since \( t \geq 0. \) Combining this with (1) yields \( \hat{g}_{\beta^*}(z) \geq \hat{f}_\alpha(z) \) for \( \alpha = 1, 2, \ldots. \) Hence

\[ \sup_\alpha \{\hat{f}_\alpha(z), \hat{g}_\beta(z)\} = \sup_\beta \{\hat{g}_\beta(z)\} = \hat{g}_\beta(z) \quad \text{for some } \beta. \]
Now let the $k$ in Theorem 1 correspond to $e_k$ and $i$ correspond to $e_i$. The hyperplanes of the form $H(N+i, a_2, \ldots)(e_i)'$ correspond to the $g_{\alpha}$. Their slopes, corresponding to $d_{\alpha}$, are (rise/run)

$$\frac{[H(N+i, a_2, \ldots)(e_i) - H(N+i, a_2, \ldots)(e_k)]}{\sqrt{2}}.$$ 

Similarly the hyperplanes of the form $H(a_1, a_2, \ldots)(e_i)'$ where $a_1 \neq N+i$ correspond to the $f_{\alpha}$. Their slopes, corresponding to the $b_{\alpha}$, are

$$\frac{[H(a_1, a_2, \ldots)(e_i) - H(a_1, a_2, \ldots)(e_k)]}{\sqrt{2}}.$$ 

The following theorem is the result.

**Theorem 2:** Suppose

$$\sup_{a_1 \neq N+i} [H(a_1, a_2, \ldots)(e_i) - H(a_1, a_2, \ldots)(e_k)] \leq \inf_{a_2, a_3, \ldots} [H(N+i, a_2, \ldots)(e_i) - H(N+i, a_2, \ldots)(e_k)]$$

for all $k \neq i$ and $k \notin D_2$. Also suppose $\delta^*(x) = N+i$ and $y \geq x$. Then $\delta^*(y) = N+i$ for $i \in D_1$.

**Proof:** By the scaling theorem assume $x_i = 0$ for $i \in D_2$. Define a sequence of points $z^1, z^2, \ldots, z^p$ inductively by
\[ z^1 = x + (x_1 - y_1)e_1 - (x_1 - y_1)e_1 \]
\[ z^2 = z^1 + (x_2 - y_2)e_1 - (x_2 - y_2)e_2 \]
\[ \vdots \]
\[ z^p = y \]

i.e., trade off one component at a time against the ith component in order to create a piecewise linear path between \( x \) and \( y \). The number of points will be one less than the number of diseases in \( D_i \). Note that \( z^k_{\geq I} \) is parallel to \( e_1 \) and \( e_2 \). Consider \( z^1 \).

Inequality (2) divided by \( \sqrt{2} \) plus Lemma 1 imply that the hypotheses of Theorem 1 hold. Since \( h*(x) = N+1 \), Theorem 1 implies that \( h*(z^1) = N+1 \).

Now apply the same argument to \( z^2 \) using \( h*(z^1) = N+1 \) to get \( h*(z^2) = N+1 \). Continuing, \( h*(z^1) = h*(z^2) = \cdots = h*(y) = N+1 \).

The hypotheses of Theorem 2 seem very complicated. Attempts to express (2) in terms of \( m, w, c_i \), and \( p_{ij} \) have not met with any success. However, (2) does have an intuitive explanation. Since \( h*(e_i) = N+k \) under weak conditions, (2) is trivially true for \( e_i = N+k \).

Now suppose \( a_i \neq N+k \). Then \( a_i \) is the "wrong" action at both \( e_i \) and \( e_k \). The left side of (2) is thus the difference in expected return for doing a wrong action at both \( e_i \) and \( e_k \). The right side, however, can be thought of as the difference in returns from doing the "correct" action at \( e_i \) versus an incorrect action at \( e_k \). It seems very reasonable that the later difference ought to be larger than the former.
Graphically inequality (2) implies that the slopes of all the hyperplanes not associated with action \textit{N+1} be less than the slope of the uppermost hyperplane associated with treatment \textit{N+1} along \(\overline{e_i, e_k}\). Figure 11 illustrates the situation.

![Diagram illustrating lexicographic situation]

\textbf{Figure 11: Lexicographic Situation}

In a manner essentially identical to Theorem 2 the result may be extended to include tests as well as treatments.

\textbf{Theorem 3:} Suppose

\begin{equation}
(3) \sup_{a_i \neq i, N+1} [H(a_1, a_2, \ldots)(e_i) - H(a_1, a_2, \ldots)(e_k)] 
\leq \inf_{a_1 = i, N+1} [H(a_1, a_2, \ldots)(e_i) - H(a_1, a_2, \ldots)(e_k)]
\end{equation}

for all \(k \neq i, k \notin D_2\). Also suppose \(\delta^*(x) = i\) and \(y \geq_1 x\). Then \(\delta^*(y) = i\) or \(N+1, i = 1, 2, \ldots, N\).
Theorem 3 says that given (3), if test $i$ is optimal at $x$ then either test $i$ or treatment $N_i$ is optimal at any point lexicographically greater than $x$. 
CHAPTER 7
CONCLUSION

7.1. Summary

This dissertation has formulated and analyzed a partially observable Markov decision process model of medical diagnosis and treatment. It was assumed that the process of a physician diagnosing and treating a single patient could be represented as a sequence of interrelated decisions. The patient may have any one of a finite collection of diseases. Unfortunately, the physician is not able to directly observe the patient's disease. Instead he must make inferences and decisions based on partial information, e.g., the results of tests or the symptoms of the patient. This was modeled by assuming that the physician observes one of a finite collection of outcomes to each action he takes.

In addition to partial information it was assumed that the patient's disease is dynamic, during the diagnosis and treatment process. This was incorporated by allowing the patient's state to change between every action and the observation of the outcome according to a Markov chain. It may be reasonably questioned whether changes in states of health are Markovian; in fact they probably do depend on the patient's history to some extent. The Markov chain, however, seems to be a reasonable first order approximation.

A general model along the above lines was constructed in Chapter 2. The model is essentially a continuous state, finite action Markov decision process. The existence of an optimal policy was shown using standard arguments. Basic properties of the model such as the piecewise linearity
and convexity of the return function were then developed. The results of Chapter 2, however, do not give a detailed characterization of the optimal policy. In order to get this it was assumed that only one of the outcomes associated with each action resulted in the continuation of the decision process. This implies that the $n$-period return function is expressable as the maximum of a finite collection of hyperplanes which can be indexed in a simple fashion. Two specializations of this stronger structure were presented and analyzed in Chapters 4 and 5. It was shown that the state space could be divided into regions with a single action optimal in each, and that under reasonable sufficient conditions the location and shape of the regions agreed with what would intuitively be expected. For example, it is optimal to treat a patient for a disease if the probability is high that he has that disease.

The mathematical model, while presented in terms of diseases, tests and treatments, is applicable to other situations. One of these is explored in the next section. Finally, some suggestions for further research are given in Section 7.3.


The class of models analyzed in this thesis may be viewed in a fairly general light. Essentially they involve a system which can be in one of finitely many states which are not directly observable. Several actions are available to the controller of the system, including actions which give him information about the underlying state of the system and actions which allow him to alter the state of the system to some preferred state if certain conditions prevail.
One of the classic examples of such a system is the problem of machine maintenance. Suppose that a complex machine may be in any one of several states of partial or complete failure due to various causes. The manager of the machine would like it to be in perfect condition; in that way it will perform its function properly. Machines being what they are, however, it is clear that parts wear out and breakdowns occur. The manager's problem then is analogous to that of the physician; he must determine what is wrong and correct it. A large computer is a good illustration of the type of machine system involved. On a more personal level there is the example of one's automobile.

More precisely, assume that the machine may either be in perfect condition or in one of finitely many failure states. These become the underlying states of nature. Unfortunately the machine can not tell the manager the cause of failure; rather it exhibits a set of symptoms. Hence, just as in medicine, the process is not directly observable. The state of the machine may be represented by a probability distribution over the underlying states.

If left alone the machine would gradually deteriorate over time until it reached some state of total failure. It is well known that this phenomenon can be modeled by a Markov chain giving transition probabilities among the underlying states. See, for example, Derman [7].

A finite collection of actions is available to the manager. These consist of various testing procedures which determine the cause of breakdown and various repair procedures for different situations, along with the possibility of doing nothing for some specified period. Clearly these actions cost some amount to perform. If the machine is in perfect
condition there is a profit to be had and if it fails there are opportunity costs and increased repair costs. The problem is to determine what the best sequence of maintenance actions is.

At this point there are two possibilities. The first looks microscopically at a single failure episode, analogous to the medical situation. Assume that the machine is in some state of failure. The manager must determine what is wrong and remedy it, at which point the decision process is over. For many machines diagnostic and repair procedures require that the machine be shut down, or operate only for testing purposes. Thus it seems that the state of the machine would be static rather than dynamic. This can be accommodated in the model by setting \( p_{11} = 1 \). In this case, there are essentially two relevant returns, the costs of performing the actions and opportunity cost associated with down-time. If one assumes that there is one diagnostic procedure and one repair procedure for each failure state along with the option of restoring the machine to production then the model of Chapter 5 may be applied. With appropriate simplification of the sufficient conditions caused by the static state assumption it can be shown that the repair and return to production options have star-shaped regions while the diagnostic options have halo-shaped ones.

The second possibility is to consider the maintenance plan for the machine over a long period of time. At periodic intervals the manager has the option of diagnostic or repair actions. He may also decide to do nothing for the interval, i.e., continue operating the machine. When the machine is in operation it produces a profit which depends on its condition, the better the condition the higher the profit.
Diagnostic and repair procedures cost money and if the machine totally fails there is also a penalty in addition to the repair bill.

In this case the decision process does not really terminate. This, combined with the economic nature of the consequences, makes a discount factor more appealing than termination probabilities as a method to insure the finiteness of the reward streams. A machine diagnostic procedure may consist of disassembly and inspection of a given part, hence the result of the procedure is immediately known. This may be contrasted to a blood test where the results become known at some time after the sample is taken. As a result it may be appropriate in the machine maintenance problem to reverse the order of the application of Bayes Theorem and the Markov chain in determining the transition structure, i.e., the machine operates for some period, then some action is taken, an outcome is observed and the state distribution is revised.

While the two modifications suggested above are non-trivial they do not really alter the basic structure of the decision process. Results on the form of the optimal policy similar to those given for the medical case should hold.

For a further discussion of the machine maintenance example see Sondik [24] and Smallwood and Sondik [23]. These models are also similar to one presented by Ross [22].

7.3. **Further Research**

Obviously there are many directions in which this research could be continued. The sufficient conditions for lexicographic improvement need to be worked into a form which is verifiable from the parameters
of the problem. The same holds for the sufficient conditions which guarantee that the test and treatment regions are adjacent to each other. One would like to have conditions which would yield enough structure to allow an analysis of the form of the optimal policy without requiring binary structure, i.e., without requiring that each action have only a single failure outcome. A related problem is the determination of conditions which imply that a finite sequence of actions is optimal even though infinite sequences are available. The question is suggested by the fact that from a medical perspective there may be infinite sequences of actions possible but these are rarely, if ever, optimal, e.g., continual observation of a patient. Medical events occur in real time, not at discrete decision points. Hence another possible extension would be to the continuous time case.

In terms of formulation, one interesting possibility would be to consider a decision process which consists of two simultaneous Markov processes. One process, corresponding to the patient's symptoms would be completely observable while the second, corresponding to the patient's disease would be unobservable. The state of the observable (symptom) process would probabilistically depend on the state of the unobservable (disease) process. Actions taken would influence the state of the unobservable process, hence indirectly affecting the state of the observable process.

There are also questions of application. Determination of the necessary probabilities may be difficult due to the log-book style common in medical record keeping. Vast quantities of data exist but they are not in usable form. See Fries [10] for some new developments and comments in this area. An even more difficult problem may be the
determination of the appropriate costs and rewards. In Chapters 4 and 5 it was assumed that the cost \( c_a \) of the action, a reward \( w \) for restoring the patient to health, and a penalty \( -m \) for letting him die were the only relevant costs and rewards. This is somewhat simplistic, however, as it seems reasonable that the costs associated with various actions depend on the benefits which may accrue as a result of performing the action. For example, a shot of penicillin will do little for the common cold and yet it cures pneumonia. Thus it would seem that there is less value in giving penicillin to a patient with a cold than to one with pneumonia. Equivalently the cost of giving penicillin to a patient with a cold is higher than the cost of giving it to a patient with pneumonia. The magnitude of the difference may depend on the feelings of the patient involved. Thus one would like to develop some sort of utility function to determine the appropriate values for \( c_a \), \( w \), and \( m \), or more generally for \( r_i(a) \) and \( U_i(j) \), the immediate and terminal returns respectively associated with performing action \( a \) and observing outcome \( j \in \Theta(a) \) if the patient has disease \( i \). From the above comments, however, it is clear that \( r_i(a) \) and \( U_i(j) \) are related. Hence the questions posed to determine the utility must be developed in such a way as to separate the two and avoid double counting of relevant factors. For example, a shot of penicillin given to a patient with pneumonia involves a dollar cost, some discomfort, a perceived potential of cure, a risk of allergic reaction, and the reality of health if the treatment is successful. It is quite difficult to separate the value of potential success, presumably associated with \( r_i(a) \) from the value of realized success, presumably associated with \( U_i(j) \). The example also points out the fact that the utilities are multidimensional. The interested reader should
see Ginsberg [11] and Ginsberg and Offensend [12]. The whole problem of utility determination needs much more work, especially in the medical context.

Finally, there is the potential of applications to other, non-medical problems. The model is appropriate to any case involving the diagnosis and repair of a system which has several possible ways of failing. This is a fairly common situation in the industrial world. The machine maintenance example of the previous section is illustrative. Industrial specializations of the general model will probably be easier to formulate and work with since the costs and rewards involved may be precisely formulated in monetary terms, thus bypassing the utility difficulties mentioned above.
## APPENDIX A

### SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>set of diseases</td>
</tr>
<tr>
<td>$A$</td>
<td>set of available actions</td>
</tr>
<tr>
<td>$a$</td>
<td>element of $A$, an action</td>
</tr>
<tr>
<td>$\Theta(a)$</td>
<td>set of outcomes to action $a$</td>
</tr>
<tr>
<td>$P_t(i)$</td>
<td>probability of disease $i$ at time $t$</td>
</tr>
<tr>
<td>$P(j</td>
<td>i,a)$</td>
</tr>
<tr>
<td>$P_t(i</td>
<td>j,a)$</td>
</tr>
<tr>
<td>$P_{|i}(a)$</td>
<td>probability of transition from disease $|$ to disease $i$ during action $a$</td>
</tr>
<tr>
<td>$P_{t+1}(i</td>
<td>j,a)$</td>
</tr>
<tr>
<td>$x$</td>
<td>vector of disease probabilities</td>
</tr>
<tr>
<td>$z_{k\ell}(j)$</td>
<td>element of matrix defining updated state if $j$ is observed</td>
</tr>
<tr>
<td>$Z(j)$</td>
<td>matrix defining updated state if $j$ is observed</td>
</tr>
<tr>
<td>$x(j)$</td>
<td>vector of posterior disease probabilities after $j$ is observed</td>
</tr>
</tbody>
</table>
\( P(j) \)  
probability of observing \( j \)

\( s \)  
state space, set of probability vectors over diseases

\( D_1 \)  
subset of \( D \) containing disease states

\( D_2 \)  
subset of \( D \) containing terminating states

\( \theta_C(a) \)  
outcomes to action \( a \) which allow decision process to continue

\( \theta_T(a) \)  
outcomes to action \( a \) which imply that the decision process stops

\( \theta_{T1}(a) \)  
outcomes to action \( a \) which stop process but do not imply patient has entered one of the terminating states

\( \theta_{T2}(a) \)  
outcomes to action \( a \) which stop process by implying patient is in a terminating state

\( r(x,a) \)  
expected cost of performing action \( a \) if state is \( x \)

\( r_1(a) \)  
cost of performing action \( a \) if patient has disease \( i \)

\( r(a) \)  
vector of costs

\( U_i(j) \)  
reward/cost for observing outcome \( j \) if patient has disease \( i \)

\( U(j) \)  
vector of rewards

\( t \)  
transpose of a vector or matrix

\( \delta^i \)  
decision rule used in period \( i \)

\( \pi \)  
a policy, sequence of decision rules

\( \tilde{r}(x,\delta^n) \)  
total immediate expected return if state is \( x \) and decision rule \( \delta^n \) is used
\( V_\pi(x) \) expected reward if policy \( \pi \) is followed starting in state \( x \)

\( \alpha_n(x,\pi) \) probability of continuation in period \( n \)

\( \pi^* \) the optimal policy

\( S^* \) optimal stationary policy

\( V^*(x) \) optimal return function

\( V^n(x) \) optimal \( n \)-period return function

\( T \) partition of \( S \)

\( T_i \) element of the partition \( T \)

\( T^h \) partition of \( S \) associated with the \( n \)-period optimal return function

\( B^n \) set of vectors defining hyperplanes which yield \( V^n(\cdot) \)

\( b \) element of \( B^n \), a row vector

\( T^* \) partition of \( S \) associated with \( V^*(\cdot) \)

\( B^* \) set of vectors which define the hyperplanes forming \( V^*(\cdot) \)

\( I(x) \) set of optimal actions at \( x \)

\( V(x,a) \) return function for action \( a \) determined by \( V^*(\cdot) \)

\( a \) a randomized action

\( f \) element of \( \theta_o(a) \) denoting failure of the action

\( H^n(a_1,\ldots,a_n)(\cdot) \) hyperplane giving \( n \)-period expected return if binary policy \( (a_1, \ldots, a_n) \) is followed

\( (a_1,\ldots,a_n) \) binary policy with action \( a_k \) in period \( k \) of \( a_{k-1} \) fails in period \( k-1 \)
\( U^o \) \quad terminal reward vector
\( U^o_i \) \quad return if \( n \)-period process unsuccessfully ends in disease \( i \)
\( \Omega(a_1, \ldots, a_k) \) \quad \( k \) stage transition matrix if failure occurs each period
\( d(x) \) \quad sum of disease probabilities
\( s \) \quad successful outcome to an action
\( e_i \) \quad vector with 1 as the \( i \)th component and zeroes elsewhere
\( c_a \) \quad cost of action \( a \)
\( -m \) \quad penalty for death of the patient
\( w \) \quad reward for health
\( T^n(a_1, \ldots, a_n) \) \quad set of points where the binary policy \( (a_1, \ldots, a_n) \) is optimal
\( -m' \) \quad penalty for sending a sick patient home
\( H(a_1, a_2, \ldots)(x) \) \quad infinite horizon expected return from the binary policy \( (a_1, a_2, \ldots) \) if \( x \) is the starting state
\( \geq_i \) \quad lexicographically greater than or equal to with respect to component \( i \)
APPENDIX B

EXAMPLE OF NON-CONVEXITY

The purpose of this appendix is to present a counter-example showing that the optimal policy need not consist of convex regions. Specifically the example shows that the region where treatment for disease 1 is optimal for a two stage decision process is not convex.

Consider the framework of Model I with the following additional assumptions. There are three diseases, i.e., \( D_1 = \{1,2,3\} \). The patient either remains in the same disease state or dies, hence

\[
p_{ii} + p_{i,N+2} = 1, \quad i \in D_1.
\]

Further assume that \( w = 0 \) and \( c_a = 0 \) for all \( a \in A \), i.e., there is no reward for restoring the patient to health and all actions cost nothing.

Suppose \( m > 0 \) and \( p_{i,N+2} > 0, \ i \in D_1 \). Under these assumptions it is trivial to show that inequalities 4.15 and 4.16 hold. Hence by Theorem 4.1 treatment \( i \) is optimal at \( e_i, \ i = 1, 2, 3 \). Since treatment for disease \( i \) costs the same as testing for disease \( i \) or doing nothing, the later options are never optimal since treatment does everything they do plus more. Formally one can show that for any hyperplane \( H^2(a_1, a_2)(\cdot) \), there is a hyperplane \( H^2(a_1', a_2')(\cdot) \) which dominates \( H^2(a_1, a_2)(\cdot) \) at \( e_i, \ i = 1, 2, 3 \) with \( a_1', a_2' \) both treatments.

As a result the two period return function, and hence the two-period decision rule, may be described by
\[ v^2(x) = \max_{a_1, a_2 \in A'} H^2(a_1, a_2)(x) \]

where \( A' = \{1, 2, 3\} \) is the set of the three treatments. (This is slightly different notation than that used in Chapter 4.) Note that there are nine possible hyperplanes which form \( v^2(\cdot) \).

Setting \( x_{N+1} = x_{N+2} = 0 \) by the scaling theorem and simplifying,

\[ H^2(a_1, a_2)(x) = -m_{N+2} x - m_{N+2} \Omega(a_1)x + U^0 \Omega(a_1, a_2)x \]

Upon making the appropriate substitutions for \( \Omega(a_1), \Omega(a_1, a_2) \) and \( U^0 \) the nine hyperplanes which define \( v^2(\cdot) \) may be represented by three of the form \( H(a, a)(\cdot) \) with

1. \( H(a, a)(x) = -m_{k, N+2}(1 + p_{kk} + p_{kk}^2)x_k - m_{L, N+2}(1 + p_{Ll} + p_{Ll}^2)x_l, \)
   \( k, l \neq a \)

and six of the form \( H(a_1, a_2)(\cdot), a_1 \neq a_2, \) with

2. \( H(a_1, a_2)(x) = -m_a a_2_{N+2} x_{a_2} - m_{k, N+2}(1 + p_{kk} + p_{kk}^2)x_k, k \neq a_1, a_2 \)

Every hyperplane of the form (1) is dominated by one of the form (2) so that

\[ v^2(x) = \max_{a_1, a_2 \in A'} H^2(a_1, a_2)(x) \]
Assume that \( m = 1 \), \( p_{11} = .5 \), \( p_{22} = .7 \) and \( p_{33} = .9 \). Then the coefficients for the six relevant hyperplanes are given in the table below.

<table>
<thead>
<tr>
<th>((a_1, a_2))</th>
<th>(x_1)</th>
<th>(x_2)</th>
<th>(x_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>0</td>
<td>-.3</td>
<td>-.271</td>
</tr>
<tr>
<td>1,3</td>
<td>0</td>
<td>-.657</td>
<td>-.1</td>
</tr>
<tr>
<td>2,1</td>
<td>-.5</td>
<td>0</td>
<td>-.271</td>
</tr>
<tr>
<td>2,3</td>
<td>-.875</td>
<td>0</td>
<td>-.1</td>
</tr>
<tr>
<td>3,1</td>
<td>-.5</td>
<td>-.657</td>
<td>0</td>
</tr>
<tr>
<td>3,2</td>
<td>-.875</td>
<td>-.3</td>
<td>0</td>
</tr>
</tbody>
</table>

For example \( H^2(1,2)(x) = 0x_1 - .3x_2 - .271x_3 \).

To determine the optimal decision rule one needs to determine the intersection of all possible pairs of hyperplanes in the surface of \( S \). This amounts to solving 15 systems of equations of the form

\[
H^2(a_1, a_2)(x) = H^2(a_1', a_2')(x)
\]

\[
x_1 + x_2 + x_3 = 1.
\]

Since there are two linear equations in three unknowns the solution is a line in the surface of \( S \). The line may be graphed by determining two points on it. This was done for \( H^2(1,2)(\cdot) \) and \( H^2(1,3)(\cdot) \) intersecting all other hyperplanes. The results are given below.
\[ H^2(a_1, a_2)(\cdot) = H^2(a_1', a_2')(\cdot) \]

<table>
<thead>
<tr>
<th></th>
<th>1, 2</th>
<th>1, 3</th>
<th>(1, 0, 0)</th>
<th>(0, 0.3238, 0.6762)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2</td>
<td>1, 3</td>
<td>2, 1</td>
<td>(0, 0, 1)</td>
<td>(0.375, 0.625, 0)</td>
</tr>
<tr>
<td>1, 2</td>
<td>2, 3</td>
<td>(0.2553, 0.7447, 0)</td>
<td>(0.1635, 0, 0.8365)</td>
<td></td>
</tr>
<tr>
<td>1, 2</td>
<td>3, 1</td>
<td>(0.3514, 0, 0.6486)</td>
<td>(0, 0.4315, 0.5685)</td>
<td></td>
</tr>
<tr>
<td>1, 2</td>
<td>3, 2</td>
<td>(0.2364, 0, 0.7636)</td>
<td>(0, 0.2065, 0.7935)</td>
<td></td>
</tr>
<tr>
<td>1, 3</td>
<td>2, 1</td>
<td>(0.5678, 0.4322, 0)</td>
<td>(0, 0.2065, 0.7935)</td>
<td></td>
</tr>
<tr>
<td>1, 3</td>
<td>2, 3</td>
<td>(0.4288, 0.5712, 0)</td>
<td>(0, 0.4315, 0.5685)</td>
<td></td>
</tr>
<tr>
<td>1, 3</td>
<td>3, 1</td>
<td>(0.1666, 0, 0.8334)</td>
<td>(0, 0.3238, 0.6762)</td>
<td></td>
</tr>
<tr>
<td>1, 3</td>
<td>3, 2</td>
<td>(0.1025, 0, 0.8973)</td>
<td>(0, 0.2897, 0.7103)</td>
<td></td>
</tr>
</tbody>
</table>

These points were carefully plotted on graph paper and the lines drawn in. Now the area where either \( H^2(1, 2)(\cdot) \) or \( H^2(1, 3)(\cdot) \) dominates all other \( H^2(a_1, a_2)(\cdot) \) is the region where \( \delta^2(x) = 1 \).

This area is shown darkly outlined in Figure 12. The dotted line divides the region into the two areas where \( H^2(1, 2)(\cdot) \) and \( H^2(1, 3)(\cdot) \) are maximal respectively. Clearly the entire region is non-convex, i.e., \( \{x \in S | \delta^2(x) = 1 \} \) is not a convex set.
Figure 12: Non-Convexity
APPENDIX C

ADJACENCY RESULTS

As mentioned in Section 4.4, one possible extension of the basic results on the form of the optimal policy is to show that the test and treatment regions are directly adjacent to each other. More precisely, under what conditions will the regions where test \( i \) and treatment \( N+i \) respectively are optimal be adjacent to each other as shown in Figure 5. The following theorem partially answers the question.

**Theorem 1:** The Model I suppose inequalities (4.15) and (4.16) hold for \( i = 1, 2, \ldots, N \). Also suppose that

\[
(1) \quad c_i + (1 - p_{i,N+1} - p_{i,N+2}) c_{N+1} + \sum_{j=1}^{N} p_{ij} \sum_{j \neq 1}^{N} p_{i,j,N+1} \\
- m(p_{i,N+2} + \sum_{j=1}^{N} p_{ij} p_{j,N+2}) \geq v^{n-1}(p_{i1}, \ldots, p_{i,N+2})^t
\]

for \( i = 1, 2, \ldots, N \). Then if \( \delta^n(x) = i \) either \( \delta^n(\alpha x + (1-\alpha) e_i) = i \) or \( \delta^n(\alpha x + (1-\alpha) e_i) = N+i, i = 1, 2, \ldots, N \).

**Comment:** Theorem 1 says that if test \( i \) is optimal at \( x \) then either test \( i \) or treatment \( N+i \) is optimal at any point on the line between \( x \) and \( e_i \). From the proof (following) it can be
seen that the role of (1) is to insure that testing for disease \( i \) is the second best option available at \( e_i \), hence it dominates doing nothing. From a medical standpoint this seems quite reasonable. If for some reason immediate treatment is impossible then test for disease \( i \), observe the positive outcome which stops the decision process and then treat the patient. Doing nothing for one period and then treating the next period runs the risk of failure of the treatment due to a state change which would force the decision process to continue with further costs incurred. On the other hand, if the disease state is fairly stable it may be better to save the cost of testing and do nothing since the risk of failure is very small. Inequality (1) says that the former situation holds. So far, (1) has eluded reduction to an easily verifiable form, hence Theorem 1 should be viewed as a suggestive starting point rather than a definitive result.

**Proof:** Using Theorem 4.2 as a guide,

\[
V^n(e_i) = \max \begin{cases} 
H^n(i, a_2, \ldots, a_n)(e_i) ; & a = N+i \\
H^n(i, a_2, \ldots, a_n)(e_i) ; & a = i \\
c_a + V^{n-1}(p_{i1}, \ldots, p_{i, N+2}) ; & a \neq i, N+i 
\end{cases}
\]

where \( H^n(N+i, a_2, \ldots, a_n) = c_{N+i} + w \) and
\[ H^n(i, a_2, \ldots, a_n)(e_1) \]
\[ = c_i + (1 - p_{i,N+1} - p_{i,N+2}) c_{N+1} \]
\[ + w(p_{i1} + p_{i,N+1} + \sum_{j=1}^{N} p_{ij} p_{j,N+1}) \]
\[ - m(p_{i,N+2} + \sum_{j=1, j \neq i}^{N} p_{ij} p_{j,N+2}) \cdot \]

Now by (1) and \( c_a \leq 0 \),

\[ H^n(i, a_2, \ldots, a_n)(e_1) \geq c_a + V^{n-1}(p_{11}, \ldots, p_{i,N+2}) \cdot \]

Hence test \( i \) dominates all strategies except treatment \( N+1 \) at \( e_1 \). Since \( \delta^n(x) = i, H^n(i, a_2^*, \ldots, a_n^*)(x) = V^n(x) \) for some \( a_2^*, \ldots, a_n^* \). By an argument exactly analogous to Theorem 4.3,

\[ H^n(i, a_2^*, \ldots, a_n^*)(y) \geq H^n(a_1, a_2, \ldots, a_n)(y) \]

for all \( y = \alpha x + (1-\alpha)e_1 \) and all \( (a_1, a_2, \ldots, a_n) \) with \( a_1 \neq i \). In other words, test \( i \) dominates all actions other than possibly \( N+1 \) on the segment \( x e_1 \). Thus \( \delta^n(\alpha x + (1-\alpha)e_1) = i \) or \( N+1 \).

Figure 13 illustrates the situation.

Theorem 1 may be extended to the infinite horizon case.
Corollary 1: If inequalities (4.16) and (1) hold with $V^{n-1}$ replaced by $V^*$ for $i = 1, 2, \ldots, N$ then $\delta^*(x) = i$ implies $\delta^*(\alpha x + (1-\alpha)e_i) = i$ or $N+i, i = 1, 2, \ldots, N$.

A similar result also holds for Model II.

Corollary 2: If (5.13) holds for $i = 1, 2, \ldots, N$ and if

$$c_i + (1 - p_{i,N+2})c_{N+1} + w(p_{ii} + p_{i,N+1} + \sum_{j=1}^{N} p_{ij} p_{j,N+1})$$

$$- m(p_{i,N+2} + \sum_{j=1}^{N} p_{ij} p_{j,N+2}) \geq V^*(p_{i1}, \ldots, p_{i,N+2})^t$$
for $i = 1, 2, \ldots, N$ then $\delta(x) = i$ implies $\delta(\alpha x + (1-\alpha)e_i) = i$
or $N \setminus i$ for $i = 1, 2, \ldots, N$. 


PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES WITH APPLICATIONS

Technical Report

HOCKSTRA, Dale J.

September 28, 1973

N00014-67-A-0112-0052

a.

NF-042-002

c.

Distribution of this document is unlimited.

This study examines a class of partially observable sequential decision models motivated by the process of machine maintenance and corrective action or medical diagnosis and treatment. Emphasis is placed on the dynamics of the state, i.e., the possibility that the machine (disease) state changes during the decision process. This is incorporated in the form of a Markov chain. It is also assumed that the state is only indirectly observable via outputs probabilistically related to the state. The end result is a model which is a discrete time Markov decision process with a continuous state space, a finite action space, and a special transition structure.

The model is formulated in general terms and the existence of an optimal policy and some of its basic properties are proved without further assumptions. More precise results on the form of the optimal policy are then given for two special cases based on reasonable sufficient conditions.
MARKOV DECISION PROCESSES
PARTIALLY OBSERVABLE
OPTIMAL POLICY
DIAGNOSIS AND TREATMENT
MACHINE MAINTENANCE

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.16 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIBING AGENCY: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b. 5a. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system number, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number(s) by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (other than those imposed by security classification, using standard statements such as:

1. Quarterly requested may obtain copies of this report from DDC.

2. "Foreign announcement and dissemination of this report by DDC is not authorized."

3. "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through"

4. "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through"

5. "All distribution of this report is controlled. Qualified DDC users shall request through"

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

10. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

11. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

12. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code, name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of indices, rates, and weights is optional.