INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Stochastic optimal control of $G/M/1$ queueing system with breakdowns

Hou, Yunkui, Ph.D.
The Ohio State University, 1991
STOCHASTIC OPTIMAL CONTROL OF G/M/1 QUEUEING SYSTEM WITH BREAKDOWNS

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

by
Yunkui Hou, B. Sc., M. Sc.

The Ohio State University
1991

Dissertation Committee:
Robert Bartoszynski
Michael Fligner
Dennis Pearl

Approved by

Adviser
Department of Statistics
ACKNOWLEDGMENTS

I express sincere appreciation to Professor Robert Bartoszynski for his guidance, continues encouragement, and patience throughout the research. His support and confidence in me was invaluable. Thanks goes to the other members of my advisory committee, Professors Dennis Pearl and Michael Fligner, for their suggestions and comments. Finally, I want to thank my wife Xiang, and my daughter Lui for their understanding and support.
VITA

October 5, 1956 .................................. Born - Fenzhen, China

1982 ............................................................. B.Sc. (Geodetic Science)

Wuhan Technical University of Surveying
and Mapping, Wuhan, China

1986 .............................................. M. Sc. (Geodetic Science)

The Ohio State University.

1989 ............................................................ M. Sc. (Statistics)

The Ohio State University.

1986-PRESENT ...................................... Teaching Associate,

The Ohio State University.

FIELDS OF STUDY

Major Field: Statistics

Studies in Applied Probability Models and Stochastic Processes: Professor

Robert Bartoszynski

Studies in Inference Theory: Professors Douglas Wolfe and Sue Leurgans

Studies in Probability Theory: Professor Boirs Pittel
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Acknowledgements</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vita</td>
<td>iii</td>
</tr>
</tbody>
</table>

## CHAPTER PAGE

<table>
<thead>
<tr>
<th>I. Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 A Review of Dynamic Programming</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Semi-Markov Decision Processes And Stochastic Control of Queueing processes</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Cost Structure And Cost Criterion</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 General Set Up</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Cost Structure And Cost Criterion</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. The Existence of Optimal Stationary Policy</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Dynamic Programming</td>
<td>18</td>
</tr>
<tr>
<td>3.1.1 Some Known Results in Dynamic Programming</td>
<td>19</td>
</tr>
<tr>
<td>3.2 Semi-Markov Decision Processes</td>
<td>22</td>
</tr>
<tr>
<td>3.3 Existence of The Discounted Cost Optimal Stationary Policy</td>
<td>26</td>
</tr>
<tr>
<td>3.3.1 Background</td>
<td>26</td>
</tr>
<tr>
<td>3.3.2 The Contraction Property</td>
<td>30</td>
</tr>
<tr>
<td>3.3.3 The Optimal Equation</td>
<td>37</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Stochastic control theory deals with dynamic systems described by differential or difference equations. The theory aims at answering questions of analysis and synthesis (Astrom, 1970, p6).

Analysis – what are the statistical properties of the system variables?

Parametric optimization – suppose that we are given a system and a regulator with a given structure but with unknown parameters. How are the parameters to be adjusted in order to optimize the system with respect to a given criterion?

Stochastic optimal control – given a system and a criterion, find a control law which minimizes the criterion.

Stochastic control relies heavily on the concepts and techniques of dynamic programming.
1.1 A Review of Dynamic Programming

Dynamic programming is a theory of multistage decision processes. The name dynamic programming was coined because in many cases, it leads to a solution in the form of a program for a computer. The technique of backward induction introduced by Arrow, Blackwell and Girshick (1949) for developing sequential methods may be regarded as a precursor of dynamic programming. The first explicit formulation of dynamic programming outside the game context is given by Bellman (1957). After the computational breakthrough by Howard (1960), some of the methods which predate the formal concept of dynamic programming, such as backward induction and method of successive approximation, became an important part of dynamic programming.

Whittle (1982,1983) has provided an extensive survey of dynamic programming and stochastic control. Ross (1986) is an introductory book in stochastic dynamic programming, which requires only the knowledge of conditional expectations. In a two part book, Bertsekas and Shreve (1978) provide a unified and mathematically rigorous theory for a broad class of dynamic programming and discrete time optimal control problems. The first part resolves structural questions where measurability of various objects is of no essential concern. Part two resolves the measurability questions associated with stochastic optimal control problems.
1.2 Semi-Markov Decision Processes And Stochastic Control of Queueing Processes

The dynamic programming technique was generalized to semi-Markov decision processes by allowing the transition interval between states to be random. (see, among others, Pyke (1961), Jewell (1963), and Ross (1970))

The difficulty of applying these techniques to the control of queueing processes is that the cost in each transition in queueing processes is not uniformly bounded, as is required by the dynamic programming technique. Because of this difficulty, Sobel (1969) investigates the optimal stationary policy in a restricted set of policies. Lippman (1973) studies the existence of an optimal stationary policy in the control of queueing processes, with the assumption that the state space is one dimensional and the state \( x \) can only be transformed to \( x+1 \) or \( x-1 \). With these assumptions, Lippman (1973) shows that there exists an optimal stationary policy when the customers arrive according to a Poisson process and the service time is exponential.

In this study, the control of a queueing process with breakdowns is investigated; the existence of an optimal stationary policy is proved. In chapter four, the theories developed in chapter three are applied to a situation which is similar to the situation studied by Lippman (1973). The results obtained in this study are more general in the sense that: (i) The distribution of the service time is general with finite second moment, we do not even need to know the form of the distribution. (ii) The system is subject to
breakdowns, and therefore the change of states cannot be restricted as in Lippman (1973).
CHAPTER II

COST STRUCTURE AND CONTROL CRITERION

2.1 General Set Up

A problem of interest in the practical application of the theory of queues is how the interruptions to the serving of customers affects the behavior of the queue. One such case is when the service station is subject to random breakdowns.

Suppose we have a service station with a single server, and customers arrive according to a homogeneous Poisson process with parameter \( \lambda \). Each customer requires a random service time \( t_s \) with finite second moment. While operating, the station is subject to random breakdowns which come according to a homogeneous Poisson process with parameter \( \xi \). On breaking down, the station becomes inoperative and requires a random period \( t_r \) with finite second moment to repair. During this repairing period, customers still arrive according to the same Poisson process with parameter \( \lambda \) but no service is provided.

The economics of the system operation is influenced by the various costs involved. When the system is on, serving customers or not, costs, such as power, heat, maintenance, etc., may be incurred. We call this kind of cost the maintenance cost.
Activating a dormant system may involve power surges, equipment charges or the process of getting familiar with the working situation if the server is human. The associated cost is called start-up cost. Another kind of cost, called delay cost, is incurred for delaying the customers in the system.

The central problem in this study is to find an operating rule which minimizes the operation cost. By operating rule, we mean a decision making criterion, or a policy to be defined later. For the moment, we shall let a policy be denoted by \( \pi \).

2.2 Cost Structure and Control Criteria

Let \( N = \{0, 1, 2, \ldots \} \) and let \( S = \{(x, \delta) : x \in N, \delta \in \{0, 1, 2\}\} \) denote the state space of the system. Then elements of \( S \) are pairs \( (x, \delta) \), and \( s = (x(t), \delta(t)) \) denotes the state of the process at time \( t \). Here \( x(t) \) is the number of customers in the system at time \( t \), including the one receiving service, and \( \delta(t) = 0, 1, 2 \) represents the situations that the system is off, the system is on and is capable of rendering service to customers, and the system is on but is incapacitated by breakdowns.

Both the customers and the breakdowns arrive according to a homogeneous Poisson process and we have no control over them. But we can control the service station. For example, when there is no customer in the system, we may want to turn it off to save maintenance cost. When there are customers in the system and the system is capable of rendering service to customers, we may want to continue the service. When a breakdown occurs, we may choose different repairing rates according to how many customers are waiting for service. The controls imposed to the system are called actions. At the end of
each transition period, the state of the process is observed and an action is taken according to some operating rule and the state of the process; this action may be a different action or may be simply a repetition of the previous one. As an example, when the transition period is ended by a completion of service and there are customers in the system, we may chose the action "serve the next customer", which is just a repetition of the previous action. When the system is off and the transition period is ended by an arrival of a customer (when the system is off, the only event causing transition of the system is the arrival of a customer), we may choose the action "leave the system off". This is again a repetition of the previous action, which could be either "turn the system off" or "leave the system off" (these two actions are considered the same since the results of these two actions are the same, that is, the system is off when either of these two actions is taken). But if we choose the action "serve the first customer in queue," then we have chosen a different action from the previous one. Again notice that we do not distinguish the actions "serve the first customer in the queue" and "serve the next customer in the queue" since when either of these two actions is taken, the customer in the front of the queue will be served. As another example, Heyman (1968) studied the queueing control model with start-up and shut-down cost (without breakdowns). The operating rule is to start up the system and serve when there are K customers present and shut it down when the system is empty. Under this operating rule, when the system is on, we observe the state at the moment of each service completion. If the queue is not empty, then we take the action "serve the next customer in queue"; otherwise we take the action "shut down the system". If the system is off, then we observe the number of customers at the moment of an arrival. If the number of customers is K, then we take the
If the process (system) is in state \( s \in S \) and an action, say \( a \), is taken, then independent of the past, two things occur

(i) The process transfers to a new state according to the transition probability \( P(s' \mid s, a) \).

(ii) Some cost is incurred according to some cost criteria.

Conditional on the event that the next state is \( s' \), the transition time \( \tau \) is a random variable with probability distribution function \( F(\tau \mid s, a, s') \). For example, if the transition from \( s \) to \( s' \) is ended by an arrival of a customer with \( \delta = 0 \) and the action \( a \) is to leave the system off, then \( F(\tau \mid s, a, s') = 1 - e^{-\lambda t} \). The cost is accumulated according to the function \( C(t \mid s, a, s') \), depending on the action taken, the states \( s, s' \) and the clock time since the beginning of the last transition. We will assume that \( C(0 \mid s, a, s') = 0 \) and denote the total cost of one transition at the end of a transition interval by \( C(s, a, s') \).

A special linear cost occurs when there is a fixed cost \( R(s, a) \) and a cost rate \( r(s, a, s') \) per unit time so that

\[
C(t \mid s, a, s') = \begin{cases} 
0 & \text{if } t = 0 \\
R(s, a) + r(s, a, s')t & \text{if } 0 < t < \tau(s, a, s') 
\end{cases} \tag{2-2-1}
\]

where \( \tau(s, a, s') \) is the length of the transition interval from \( s \) to \( s' \) when action \( a \) is taken at state \( s \).
Therefore, if action $a$ is taken at state $s$, the expected cost in the transition is

$$
\tilde{C} = R(s, a) + \sum_{s' \in S} p(s'|s, a) \int_{0}^{\infty} t \cdot d\tau|s, a, s')
$$

(2.2.2)

and the expected length of the transition interval is

$$
\bar{T}(s, a) = \sum_{s' \in S} p(s'|s, a) \int_{0}^{\infty} d\tau|s, a, s').
$$

(2.2.3)

To ensure that with probability one, only a finite number of transitions take place in a finite time, we assume that there exist $\delta > 0$, $\varepsilon > 0$ such that for every action $a$ and every state $s \in S$

$$
F(\delta \mid s, a, s') < 1 - \varepsilon.
$$

(2.2.4)

Let $T' = \{ t' \geq 0 : s(t') \neq s(t') \}$ and let $(t'_0, t'_1, \ldots)$ be the ordered sequence of elements in $T'$, where $t'_0 = 0$ and $t'_k$ is the epoch of the $k^{th}$ change of state. Therefore, the length of the $k^{th}$ transition interval is $\tau_k = t'_k - t'_{k-1}$. Notice the elements in $T'$ can be ordered into a sequence since we assume that only a finite number of transitions can take place in finite time, which ensures that every element in $T'$ is an isolated point. We will also assume that actions can only be taken at epochs $t'_i \in T'$.

If $C(t)$ denotes the total cost incurred up to time $t$, then our goal is to find an optimal policy $\pi$, to minimize
\[ \overline{V}(\pi, s) = \limsup_{T \to \infty} E_{\pi} \left\{ \frac{1}{T} \int_0^T C(dt) | S(0) = s \right\} \quad \forall s \in S \quad (2-2-5) \]

where the subscript \( \pi \) means the process is controlled by policy \( \pi \). The concept of policy will be made precise in Chapter 3. Here it suffices to say that a policy is a rule which tells us what action to take at each decision time, depending on the observed and past states, and the past actions. Here \( S_0 = S(0) \) is the initial state of the process and \( S_n \) will be used to denote the state of the process between \( n^{th} \) and \( (n+1)^{th} \) action.

The criteria of minimizing \( \overline{V}(\pi, s) \) are called average cost criteria and any policy \( \pi^\ast \) such that

\[ \overline{V}(\pi^\ast, s) = \inf_{\pi \in \Pi} \overline{V}(\pi, s) \quad \forall s \in S \]

is called the optimal average cost policy if such \( \pi^\ast \) exists. Here \( \Pi \) is the set of all possible policies.

To establish another cost criterion, let \( C_n, \tau_n, n = 1, 2, \ldots \) denote the cost incurred in the \( n^{th} \) transition and the length of the \( n^{th} \) transition interval respectively and define
\[ V(\pi, s) = \limsup_{N \to \infty} \frac{E_\pi(\sum_{n=1}^{N} C_n | S_0 = s)}{N} \]

Thus, \( \bar{V}(\pi, s) \) and \( V(\pi, s) \) both represent, in some sense, the average expected cost. While \( \bar{V} \) is more appealing, \( V(\pi, s) \) is easier to work with. Ross (1970) shows that under certain conditions the two criteria are equivalent.

For any initial state \( s \in S \), let

\[ T = \inf \{ t > 0, S(0) = s, S(t) = s \} \]

and

\[ N = \min \{ n > 0, S_0 = s, S_n = s \} \]

Hence, \( T \) is the time of the first return to state \( s \) and \( N \) is the number of transitions it takes. In the queueing application, \( T \) is just the length of a busy cycle and \( N \) is the number of transitions in the busy cycle. Clearly, the continuation of the process beyond \( T \) is a probabilistic replica of the whole process starting at 0. Processes with such property are called regenerative processes and \( T \) is called the regeneration point or cycle point.
The regenerative property implies the existence of further times $T_1, T_2, \ldots$, having the same property as $T$ and $\{T, T_1, T_2, \ldots\}$ forms a renewal process. Based on this property, Ross (1970) gives the following theorem.

**THEOREM (2 - 1) (Ross, 1970)**

Assume condition (2 -2 -4) holds. If $f$ is a stationary policy and if

$$E_f(T \mid S_0 = s) < \infty,$$

then

$$
\bar{V}(f, s) = V(f, s) = \frac{E_f(C(T) \mid S_0 = s)}{E_f(T \mid S_0 = s)}
$$

where $S_0 = S(0)$ and $C(T)$ is the cost incurred in the period $T$, that is, in one busy cycle\(^*\).

**Proof.** Ross (1970).

The following definition gives another criterion which will be used to prove the existence of an optimal stationary policy.

**Definition.** The total expected $\alpha$-discounted cost associated with policy $\pi$ is defined by

$$V_\alpha(\pi, s) = E_\pi\left\{ \int_0^{\infty} e^{-\alpha t} C(dt) \mid S(t_0) = s \right\} \quad (2 - 2 - 7)$$

\(^*\) A stationary policy $\pi$ is of the form $\pi = \{f, f, \ldots\}$. Therefore, we say a stationary policy $\pi$ is a map from state space to action space such that when the state is $s$, the action $f(s)$ will be taken. For this reason, we may use $f$ to denote a stationary policy.
where \(0<\alpha<1\) is the discount factor.

The expected discounted cost in the transition from \(s\) to \(s'\) given that the process enters state \(s\) at \(t(s)\) and that action \(a\) is taken is

\[
\rho(\alpha, s, a, s', t(s)) = \int_0^{t(s) + \tau} dF(\tau|s, a, s') \int e^{-\alpha t_c} (dt|s, a, s'). \quad (2-2-8)
\]

Therefore, the expected one transition cost given that the process entered state \(s\) at \(t(s)\) and the action \(a\) is taken is

\[
\rho(\alpha, s, a, t(s)) = \sum_{s' \in S} p(s'|s, a) \rho(\alpha, s, a, s', t(s))
\]

\[
= \sum_{s' \in S} p(s'|s, a) \int_0^{t(s) + \tau} dF(\tau|s, a, s') \int e^{-\alpha t_c} (dt|s, a, s') . \quad (2-2-9)
\]

If the linear cost structure is assumed, then (2-2-9) can be simplified as

\[
\rho(\alpha, s, a, t(s)) = e^{-\alpha t(s)} \rho(\alpha, s, a) \quad (2-2-10)
\]

with

\[
\rho(\alpha, s, a) = R(s, a) + \sum_{s' \in S} p(s'|s, a)r(s, a, s') \int_0^\tau dF(\tau|s, a, s') \int e^{-\alpha t_c} dt
\]

\[
= R(s, a) + \sum_{s' \in S} p(s'|s, a)\rho(\alpha, s, a, s') \quad (2-2-11)
\]
where
\[ p(\alpha, s, a, s') = r(s, a, s') \int_0^\infty dF(\tau|s, a, s') e^{-\alpha \tau} d\tau \]
\[ = \frac{1}{\tau(s, a, s')}[1 - \beta(\alpha, s, a, s')] \] (2-2-12)

and
\[ \rho(\alpha, s, a, s') = \int_0^\infty e^{-\alpha \tau} dF(\tau|s, a, s') \] (2-2-13)

We see that \( \beta(\alpha, s, a, s') \) is just the Laplace-Stieltjes transformation of the conditional distribution of the transition interval from \( s \) to \( s' \) given that action \( a \) is taken at state \( s \), and \( \rho(\alpha, s, a) \) is the expected discounted cost incurred in one transition given action \( a \) is taken at state \( s \). This cost will be further discounted by \( e^{-\alpha \tau(s)} \), where \( \tau(s) \) is the time at which the transition starts. In order to use the dynamic programming technique, we need:

**LEMMA (2-1)**

Under the assumption of the linear cost structure given in (2-1-1), the total expected \( \alpha \)-discounted cost associated with policy \( \pi \) can be written as

\[ V_\alpha(\pi, s) = E_\pi \left\{ \sum_{i=0}^{\infty} e^{-\alpha(\tau_i + \tau_{i+1} + \cdots + \tau_{i+j})} \rho(\alpha, s_i, \pi_i) | s_0 = s \right\} \]
\[ = E_\pi \left\{ [\rho(\alpha, s_0, \pi_0) + \sum_{i=1}^{\infty} \prod_{j=1}^{i} \beta(\alpha, s_{j-1}, \pi_j, s_j) \rho(\alpha, s_i, \pi_i)] | s_0 = s \right\} \] (2-2-14)
where $s_i$ is the state of the process at the beginning of the $(i+1)^{th}$ transition, $\pi_i$ is the action chosen according to policy $\pi$, and $\tau_i$ is the length of the $(i+1)^{th}$ transition interval with $\tau_1=0$.

Proof. Under the assumption of (2 - 2 - 1), the total expected $\alpha$-discounted cost can be written as

$$V_\alpha(\pi, s) = \mathbb{E}_\pi \left\{ \sum_{i=0}^{\infty} e^{-\alpha(\tau_{i-1}+\tau_0+\tau_1+\ldots+\tau_{i-1})} \{ [R(s_i, \pi_i) + r(s_i, \pi_i, s_{i+1}) \int_{0}^{\tau_i} e^{-\alpha t} dt] | s_0 = s \} \right\}$$

Let $H_i = \{ s_0, \pi_0, \tau_0, s_1, \pi_1, \ldots, s_i, \pi_i \}$ denote the history up to the beginning of the $(i+1)^{th}$ transition. Then

$$V_\alpha(\pi, s) = \mathbb{E}_\pi \left\{ \sum_{i=0}^{\infty} E_\pi \left\{ e^{-\alpha(\tau_{i-1}+\tau_0+\tau_1+\ldots+\tau_{i-1})} [R(s_i, \pi_i) + r(s_i, \pi_i, s_{i+1}) \int_{0}^{\tau_i} e^{-\alpha t} dt] | H_i \} | s_0 = s \right\}$$

$$= \mathbb{E}_\pi \left\{ \sum_{i=0}^{\infty} E_\pi \left\{ e^{-\alpha(\tau_{i-1}+\tau_0+\tau_1+\ldots+\tau_{i-1})} E \left\{ [R(s_i, \pi_i) + r(s_i, \pi_i, s_{i+1}) \int_{0}^{\tau_i} e^{-\alpha t} dt] | H_i \} | s_0 = s \right\} \right\}$$

$$= \mathbb{E}_\pi \left\{ \sum_{i=0}^{\infty} E_\pi \left\{ e^{-\alpha(\tau_{i-1}+\tau_0+\tau_1+\ldots+\tau_{i-1})} E \left\{ [R(s_i, \pi_i) + r(s_i, \pi_i, s_{i+1}) \int_{0}^{\tau_i} e^{-\alpha t} dt] | s_i, \pi_i, \tau_i \} | s_0 = s \right\} \right\}$$

$$= \mathbb{E}_\pi \left\{ \sum_{i=0}^{\infty} e^{-\alpha(\tau_{i-1}+\tau_0+\tau_1+\ldots+\tau_{i-1})} E \left\{ [R(s_i, \pi_i) + r(s_i, \pi_i, s_{i+1}) \int_{0}^{\tau} e^{-\alpha t} dt] | s_i, \pi_i \} | s_0 = s \right\} \right\}$$
\[ \begin{align*}
&= r(s_i, \pi_i, s_{i+1}) + \int_0^\tau dF(s_i, \pi_i, s_{i+1}) |s_i, \pi_i) |s_0 = s \\
&= E_\pi \left\{ \sum_{i=0}^\infty e^{-\alpha(t_0 + \tau_1 + \ldots + \tau_{i-1})} \rho(\alpha, s_i, \pi_i) |s_0 = s \right\} \\
&= E_\pi \left\{ [\rho(\alpha, s_0, \pi_0) + \sum_{i=0}^\infty e^{-\alpha(t_0 + \tau_1 + \ldots + \tau_{i-1})} \rho(\alpha, s_i, \pi_i)] |s_0 = s \right\} \\
&= E_\pi \left\{ [\rho(\alpha, s_0, \pi_0) \\
&\quad + \sum_{i=0}^\infty e^{-\alpha(t_0 + \tau_1 + \ldots + \tau_{i-2})} \rho(\alpha, s_i, \pi_i) E \{ e^{-\alpha t_i} |s_{i-1}, \pi_{i-1}, s_i \} ] |s_0 = s \right\} \\
&= E_\pi \left\{ [\rho(\alpha, s_0, \pi_0) \\
&\quad + \sum_{i=0}^\infty e^{-\alpha(t_0 + \tau_1 + \ldots + \tau_{i-2})} \rho(\alpha, s_i, \pi_i) \beta(\alpha, s_{i-1}, \pi_{i-1}, s_i)] |s_0 = s \right\}
\end{align*} \]

where the fifth equality is due to (2-2-11) and the last one is due to (2-2-13).

Conditioning on \((H_{i-2}, s_{i-1}, \pi_{i-1})\) and repeating this procedure, we get

\[ V_\alpha(\pi, s) = E_\pi \left\{ \sum_{i=0}^\infty E \{ E \{ E \{ \ldots E \{ \rho(\alpha, s_i, \pi_i) \beta(\alpha, s_{i-1}, \pi_{i-1}, s_i) |H_{i-2}, s_{i-1}, \pi_{i-1} \} |H_{i-3}, s_{i-2}, \pi_{i-2} \} \ldots \} |s_0 = s \} \right\} \]
\[ \Omega \]

\[ \left\{ \sum_{S=0}^{s} \left( \sum_{i=1}^{\infty} \prod_{x_0} \mathcal{P}(\omega, \alpha) \delta \left( \sum_{i=1}^{\infty} (0 \times s_0 \times 0 \times \alpha, \alpha, \alpha) \right) \right) \right\}_{\Omega} = \]

1
CHAPTER III

THE EXISTENCE OF THE OPTIMAL STATIONARY POLICY

3.1 Dynamic Programming

Dynamic programming is a mathematical abstraction of situations requiring a sequence of interrelated decisions. The outcome of each decision is not fully predictable but can be observed before the next decision is made. Each decision transforms the current situation into a new situation under which further decisions will be made. A certain cost is involved in each decision and the object is to minimize the cost.

Dynamic Programming has served as a very useful mathematical technique for many years in the areas of engineering, mathematics and the social sciences. It was Bellman (1957), however, who realized soon after the appearance of Wald's work in sequential analysis, that Dynamic Programming techniques could be developed into a systematic tool for optimization and noticed the fact that sequential decision problems with a multiplicative cost functional can be treated by Dynamic Programming techniques.
This technique was extensively developed by Howard (1960, 1965), further analyzed by Blackwell (1962, 1964, 1965), Derman (1965, 1966, 1967) and others, and generalized to Semi-Markov decision process by allowing the transition interval between states to be random by, among others, Jewell (1963) and Ross (1970).

3.1.1 Principle of Optimality

Clearly, the choice of a decision that minimizes the present cost might be punished by a higher further cost; thus, instead of viewing decisions in isolation, Dynamic Programming technique selects at each stage a decision that minimizes the sum of the current stage cost and the best cost that can be expected from further stages. This idea of balancing the desire to minimize the cost of present against a desire to avoid the possibility of higher cost in the future stages is the principle of optimality originally stated by Bellman (1957) as follows:

Principle of optimality: An optimal policy has the property that, whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

3.1.2 Some Known Results in Discrete Dynamic Programming

The discounted dynamic programming with bounded expected per stage cost is by far the simplest infinite horizon problem. This is due to the contraction property induced by the presence of the discount factor. Many authors have contributed to the analysis of discounted dynamic programming with bounded expected per stage cost, most notably Bellman (1957), Howard (1960) and Blackwell (1964). Taylor (1965)
has considered the average cost criterion for the special case for inventory and replacement systems allowing for an infinite state space. In the well known paper "Discounted Dynamic Programming," Blackwell (1964) gives the following results:

THEOREM (3 - 1)

1. There need not exist an $\varepsilon$-optimal policy $\pi$. That is, there are situations in which there is an $\varepsilon > 0$ such that for every policy $\pi$, there is a policy $\pi'$ such that

$$V_\alpha(\pi', s) \leq V_\alpha(\pi, s) - \varepsilon$$

for some $s \in S$.

2. There always exists a $(p, \varepsilon)$-optimal stationary policy $\pi^*$. That is, for any probability distribution $P$ on $S$ and any $\varepsilon > 0$, there is a stationary policy $\pi^*$, such that for every $\pi$

$$P( V_\alpha(\pi, s) < V_\alpha(\pi^*, s) - \varepsilon ) = 0$$

3. Not every policy $\pi$ need be dominated within $\varepsilon$ by a stationary policy $\pi^*$, that is, there are situations where there is a policy $\pi$ and an $\varepsilon > 0$ such that for every stationary policy $\pi^*$

$$V_\alpha(\pi^* s) > V_\alpha(\pi, s) + \varepsilon$$

for some $s \in S$.

4. If the action space $A$ is countable, then there is an $\varepsilon$-optimal stationary $\pi^*$. That is, for all $\varepsilon > 0$, there exists a stationary policy $\pi^*$ such that for every policy $\pi$

---

# Blackwell considered reward instead of cost.
If the action space $A$ is finite, then there is an optimal stationary policy $\pi^*$; that is, there exists a stationary policy $\pi^*$, such that $\forall \pi$

$$V_{\alpha}(\pi^*, s) \leq V_{\alpha}(\pi, s) \quad \forall s \in S$$

(6) If there is an optimal policy $\pi$, there is one which is stationary.


Derman (1966) studied the average cost criteria and has proved that under the assumption of finite action space and bounded expected per stage cost, an optimal stationary policy, while it always exists for the discounted criteria, does not necessarily always exist for the average cost criteria. Motivated by the policy iterative procedure, Derman (1966) gives the following sufficient condition for the existence of an optimal stationary policy for the average cost criteria.

**THEOREM (3 - 2)**

If the action space is finite and the expected one-transition cost is uniformly bounded, and if there exists a bounded set of numbers $\{g, h_i\} \forall i \in I$ satisfying

$$g + h_i = \min \{ w(i, a) + \sum_{j \in I} p(j | i, a) \} \quad i \in I$$

then there exists a stationary policy $\pi^*$ such that for any $i \in I$ and every policy $\pi \in \Pi$
\[ g = V(\pi^*, i) \leq V(\pi, i) \quad (3-1-2) \]

\( \pi^* \) is the policy which for each \( i \), prescribes the action that minimizes the right hand side of (3-1-1).

Here \( w(i, a) \) is the expected cost incurred in one transition when action \( a \) is taken at state \( i \in \mathcal{I} \) with \( i \) being the number of customers and \( \mathcal{I} \) is the state space, assumed to be countable.

**Proof.** Derman (1966).

The difficulty in applying the above theorem is that the existence of a bounded solution of the functional equation (3 - 2 - 1) cannot be checked directly. Derman's paper (1966), however, in conjunction with a later paper of Derman and Veinott (1967) shows that a sufficient condition for the existence and uniqueness of solution to the functional equation (3 - 2 - 1) is: (i) For each stationary policy, the resulting Markov chain is positive recurrent and, (ii) There exists some state (say 0) and a constant \( T < \infty \) such that \( M_{i,0}(\pi) < T \) for any state \( i \in \mathcal{I} \) and any stationary policy \( \pi \in \Pi \), where \( M_{i,0}(\pi) \) denotes the mean recurrence time from state \( i \) to state 0 when using policy \( \pi \). Ross (1968) simplifies this condition and shows that the average cost optimal policy is a limit point of discounted cost policies.

### 3.2 Semi-Markov Decision Processes

A semi-Markov decision process, which is an extension of discrete dynamic programming, is defined by the following five objects:
(i) State space $S$, the set of all possible states of the process. An element $s \in S$ is called a state variable. When the state variable is $s$, we say the state is $s$.

(ii) Action space, $A = \bigtimes_{s \in S} A_s$, where $A_s$ is the set of possible actions at state $s$. An element $a \in A$ is called an action. An action is also called a control, or a decision. Therefore, the action space is also called the control space or the decision space. In this study, the action space is assumed to be finite.

(iii) Transition probability $p(s' | s, a)$, defined as the conditional probability that next state is $s'$ given action $a$ is taken at state $s$.

(iv) Distribution function of transition interval $F(s' | s, a)$, being the conditional probability distribution of the sojourn time between transitions, given that action $a$ is taken at state $s$ and the next state is $s'$.

(v) Cost $C(t | s, a, s')$ incurred up to time $t$ since the last transition. This is the cost accumulated up to $t$ since the beginning of the last transition, given that action $a$ is taken at state $s$ and the next state is $s'$.

An action chosen by a policy may depend on the history of the process up to that point, or it may even be randomized in the sense that action $a$ will be chosen with a probability distribution. In this case, $\pi_i$ is a conditional probability distribution. A non-randomized policy $\pi$ is a sequence $\{\pi_0, \pi_1, \ldots\}$ such that

$$\pi_i : H \to A$$
where \( H = (S \times A \times S \ldots) \), and \( \pi_t \) is measurable with respect to the \( \sigma \) algebra generated by 
\( (s_0, a_0, s_1, a_1, \ldots, s_{i-1}, a_{i-1}, s_i) \). (Bartoszynski, 1971).

That is, a policy \( \pi \) is a sequence \( \{\pi_0, \pi_1, \ldots\} \) of decision rules where the \( i \)th decision \( \pi_i \) tells how to select an action. Therefore, using policy \( \pi \) means that, if we find the process in state \( s \) at \( i \)th transition, then we choose action \( \pi_i(s) \).

A policy \( \pi \) is said to be stationary if it is non-randomized and the action it chooses at time \( t \) depends only on the state of the process at time \( t \), regardless of what happened previous to \( t \). In other words, a policy \( \pi \) is said to be stationary if there exists a map

\[
f: S \rightarrow A
\]

with the interpretation that when the state is \( s \), then policy always chooses action \( f(s) \). Therefore, the space, say \( Q \), of all stationary policies is just the class of policies \( \pi = (f, f, f \ldots) \), where \( f: S \rightarrow A \), and \( Q \) may be identified with the class of all such functions \( f \).

Clearly, the class \( Q \) of all stationary policies \( f: S \rightarrow A \) is a subspace of \( \Pi \), the space of all possible policies. However, we will show later that, similar to the dynamic programming problems with uniformly bounded expected per transition cost, under certain conditions, the optimal policy is a member of \( Q \) even without the assumption of uniformly bounded expected per transition cost.
An important reason that the stationary policy is of particular interest is that under the control of a stationary policy, the process \{S(t): t \geq 0\} is a semi-Markov process with transition probability \( p(s'|s, f(s)) \).

**DEFINITION:** A policy \( \pi^* \) is \( \alpha \)-optimal if

\[
V_{\alpha}(\pi^*, s) = V_{\alpha}(s) \quad \forall \, s \in S.
\]

where

\[
V_{\alpha}(s) = \inf_{\pi \in \Pi} V_{\alpha}(\pi, s) \quad \forall \, s \in S. \tag{3 - 2 - 1}
\]

**THEOREM (3 - 3)**

If there exists a \( M < \infty \) such that \( C(s, a) \leq M \), then the optimal value \( V_{\alpha}(s) \) satisfies

\[
V_{\alpha}(s) = \max \left\{ \rho(\alpha, s, a) + \sum_{s' \in S} p(s'|s, a) \beta(\alpha, s, a, s') V_{\alpha}(s') \right\} \tag{3 - 2 - 2}
\]

where

\[
\rho(\alpha, s, a) = R(s, a) + \sum_{s' \in S} p(s'|s, a) \int_0^\infty \int_0^\tau e^{-\omega d} dt \int dF(\tau|s, a, s')
\]

\[
= R(s, a) + \sum_{s' \in S} p(s'|s, a) \rho(\alpha, s, a, s')
\]

is the expected one transition discounted cost incurred when action \( a \) is taken at state \( s \) and
\[ p(a, s, a, s') = \int_{0}^{\infty} e^{-\alpha \tau} dF(\tau|s, a, s') \]

is the Laplace-Steiljes transformation of the transition interval \( \tau \).

Equation (3 - 2 - 2) is the Bellman functional equation or optimality equation of the semi-Markov decision process. Comparing (3 - 2 - 2) with the optimality equation of the discrete Dynamic Programming, we see \( \beta(\alpha, s, a, s') \) plays the role of a discount factor. The proof of the theorem is similar to that of the optimality equation of discrete Dynamic Programming, which can be found in, for example, Ross (1970). A complete proof of the theorem will be given later without the assumption of uniformly bounded expected per transition cost.

### 3.3 Existence of The Discounted Cost Optimal Stationary Policy

#### 3.3.1 Background

The object of this section is to find an optimal policy for the queueing control problem stated in the previous chapter via the discounted semi-Markov decision process, some basic results of which are briefly stated in the previous sections. The results there, however, cannot be used here since in this study, the expected per transition cost is not uniformly bounded. In the queueing control model studied here, the expected per transition cost is proportional to the number of customers in the system (the number of customers in the service station). However, the cost in each transition is certainly bounded by a polynomial in \( x \), the number of customers in the system at the beginning of the transition period.
Lippman (1973) first investigated the situation that the expected per transition cost is bounded by a polynomial in \( x \), the number of customers in the system. He assumed that only states \( x+1 \) and \( x-1 \) are accessible from \( x \) in one transition; that is, for each state \( x \) and each action \( a \) we have*

\[
P(\text{x+1} | \text{x}, a) + P(\text{x-1} | \text{x}, a) = 1 \quad \forall \, \text{x} \in S \text{ and } \forall \, \text{a} \in A.
\]

Lippman (1973) shows that if the service time distribution is exponential and decisions are also made at all times of arrival, then there exists an average cost optimal stationary policy (the system is not subject to breakdowns).

In the rest of this study, we will assume that there exists a constant \( M > 0 \) such that the cost \( C(s, a, s') \) satisfies

\[
\max_{s' \in S} \max_{a \in A_s} C(s, a, s') \leq M(x+1) \quad s = (x, \delta).
\]

which implies

\[
\rho(a, s, a) \leq M(x+1).
\]

**LEMMA (3-1)**

There exists a constant \( 0 < \beta < 1 \) such that \( \forall \, s, s' \in S \text{ and } \forall \, a \in A_s \)

\[
\beta(\alpha, s, a, s') \leq \beta.
\]

* The state space in Lippman (1973) is \( S = \{0, 1, 2, \ldots, \} \).
Proof. By (2 - 2 - 4) we know that there exist $\delta > 0$ and $\varepsilon > 0$ such that

$$\beta(\alpha, s, a, s') = \int_{0}^{\infty} e^{-\alpha t} dF(t|s, a, s')$$

$$\delta \quad \infty$$

$$= \int_{0}^{\delta} e^{-\alpha t} dF(t|s, a, s') + \int_{\delta}^{\infty} e^{-\alpha t} dF(t|s, a, s')$$

$$\leq F(\delta|s, a, s') + e^{-\alpha \delta} [1 - F(\delta|s, a, s')]$$

$$= F(\delta|s, a, s') [1 - e^{-\alpha \delta}] + e^{-\alpha \delta}$$

$$\leq 1 - e + ee^{-\alpha \delta}$$

$$= \beta < 1.$$ 

On the other hand

$$\beta(\alpha, s, a, s') = \int_{0}^{\infty} e^{-\alpha t} dF(t|s, a, s')$$

$$\delta \quad \infty$$

$$= \int_{0}^{\delta} e^{-\alpha t} dF(t|s, a, s') + \int_{\delta}^{\infty} e^{-\alpha t} dF(t|s, a, s')$$

$$\geq e^{-\alpha \delta} F(\delta|s, a, s') + e^{-\alpha \delta} [1 - F(\delta|s, a, s')]$$

$$= e^{-\alpha \delta} > 0,$$ 

where the second inequality is due to condition (2 - 2 - 4). QED.
LEMMA (3 - 2)

If \( u = \{ u(s) \}_{s \in S}, v = \{ v(s) \}_{s \in S} \) are such that

\[
\sup_{s \in S} \frac{|u(s)|}{x+1} < \infty \quad \text{and} \quad \sup_{s \in S} \frac{|v(s)|}{x+1} < \infty,
\]

then

\[
d(u, v) = \sup_{s \in S} u(s) - v(s)
\]

is a metric,

where \( s = (x, \delta), \quad x = 0, 1, 2, \ldots \), \( \delta = 0, 1, 2. \)

Proof. Clearly, \( d(u, v) = d(v, u) \) and \( d(u, v) = 0 \) if and only if \( u(s) = v(s) \forall s \in S. \)

Now let \( \omega = \{ \omega(s) \}_{s \in S} \) be such that

\[
\sup_{s \in S} \frac{|\omega(s)|}{x+1} < \infty.
\]

Then

\[
d(u, v) = \sup_{s \in S} \frac{|u(s) - v(s)|}{x+1}
\]

\[
= \sup_{s \in S} \frac{|u(s) - \omega(s) + \omega(s) - v(s)|}{x+1}
\]

\[
\leq \sup_{s \in S} \frac{|u(s) - \omega(s)|}{x+1} + \sup_{s \in S} \frac{|\omega(s) - v(s)|}{x+1}
\]

\[
= d(u, \omega) + d(\omega, v)
\]
where the inequality is due to the fact that $\sup_{s \in S} |A(s) + B(s)| \leq \sup_{s \in S} |A(s)| + \sup_{s \in S} |B(s)|$.

QED.

3.3.2 The Contraction Property

We start by defining a complete metric space

$$B = \{ \nu = (\nu(s))_{s \in S} : S \to \mathbb{R} \text{ such that } \sup_{s \in S} \frac{|\nu(s)|}{x+1} < \infty, s = (x, \delta) \}$$

with metric

$$d(\nu, \nu) = \sup_{s \in S} \frac{|\nu(s) - \nu'(s)|}{x+1} \quad \forall \nu, \nu' \in B. \quad (3.3.4)$$

DEFINITION: Let $B$ be a metric space with metric $d(\cdot, \cdot)$. Then a map

$$\Gamma : B \to B$$

is said to be a contraction mapping if there exists a number $\beta$ such that $0 < \beta \leq 1$ and

$$d(\Gamma(x), \Gamma(y)) \leq \beta d(x, y) \quad \forall x, y \in B.$$

For any $a \in A_s$ and for any $s \in S$, define a map $T_a$ by

$$T_a \nu(s) = \rho(\alpha, s, a) + \sum_{s' \in S} p(s'|s, a) \int_0^\tau e^{-\alpha \tau} \nu(s')dF(\tau \mid s, a, s')$$

$$= \rho(\alpha, s, a) + \sum_{s' \in S} p(s'|s, a)\beta(\alpha, s, a, s')\nu(s'), \quad (3.3.5)$$
LEMMA (3-3)

The map $T_a$ defined by (3-3-5) is a map from $B$ to itself.

Proof. By Lemma (3-1), we have

$$\sum_{s' \in S} p(s' | s, a) \beta(\alpha, s, a, s') u(s') \leq \beta \sum_{s' \in S} \frac{\mu(s')}{x'+1} (x'+1)p(s' | s, a)$$

$$\leq \beta \sup_{s' \in S} \frac{\mu(s')}{x'+1} \sum_{s' \in S} (x'+1)p(s' | s, a). \quad (3-3-6)$$

Notice

$$\sum_{s' \in S} (x'+1)p(s' | s, a) = 1 + \sum_{s' \in S} x'p(s' | s, a)$$

$$= 1 + \sum_{\delta=0}^{x} \sum_{x'=0}^{x} x' P(x' | (x, \delta), a) P(\delta') (x, \delta), a)$$

$$\leq \max \{ x, x+1, x+2, x+1+N_r \}$$

$$\leq x+2+N_r. \quad (3-3-7)$$

where $x$, $x+1$, and $x+2$ correspond to the situation where the transition is ended by a completion of service, an interruption of breakdown and an arrival of a new customer.

The last term is for the situation where the action can only be taken at the epochs of the completion of a service if the system is on and the arrival of a new customer if the system is off, and
\[ N_r = \xi E(t_e) + 1 \] \[ E(t_s) \lambda \]

is the expected number of arrivals in the period of one service completion.

Substituting (3-3-6) and (3-3-7) into (3-3-5), we have

\[ T_a u(s) \leq \rho(\alpha, s, a) + \beta \sup_{s \in S} \frac{h(s)}{x + 1} \sum_{s' \in S} (x + 1)p(s'|s,a) \]

\[ \leq M(x + 1) + \beta(x + 2 + N_r) \sup_{s \in S} \frac{h(s)}{x + 1} < \infty. \]

where \( M \) is defined in (3-3-1). QED.

For every fixed \( s \in S \), we now define another map \( T \) by

\[ T u = \min_{a \in A_s} T_a u \quad \forall u \in B. \]

Clearly, \( T \) is a map from \( B \) to itself since \( T u \leq T_a u \) for all \( u \in B \).

THEOREM (3-4)

There exists an integer \( J \) such that

\[ T^J : B \to B \text{ is a contraction mapping,} \]

where \( T^J \) is defined inductively by \( T^1 = T, T^{n+1} = T(T^n) \).

Proof. \( \forall u, v \in B \), and \( \forall s = (x, \delta) \in S \)
\[ T_0(s) - T_V(s) = \min_{a \in A_s} \left\{ \rho(\alpha, s, a) + \sum_{s' \in S} p(s' | s, a) \beta(\alpha, s, a, s') U(s') \right\} \]

\[ - \min_{a \in A_s} \left\{ \rho(\alpha, s, a) + \sum_{s' \in S} p(s' | s, a) \beta(\alpha, s, a, s') V(s') \right\} \]

Let \( a_0 \in A_s \) be the action such that

\[ T_{a_0} U(s) = \rho(\alpha, s, a_0) + \sum_{s' \in S} p(s' | s, a_0) \beta(\alpha, s, a_0, s') U(s') \]

\[ = \min_{a \in A_s} \left\{ \rho(\alpha, s, a) + \sum_{s' \in S} p(s' | s, a) \beta(\alpha, s, a, s') U(s') \right\} . \]

(3 - 3 - 9)

Substituting (3 - 3 - 9) into (3 - 3 - 8) gives

\[ T_0 - T_V \leq \sum_{s' \in S} p(s' | s, a_0) \beta(\alpha, s, a_0, s')[U(s') - V(s')] \]

\[ = \sum_{s' \in S} p(s' | s, a_0) \beta(\alpha, s, a_0, s') \frac{[U(s') - V(s')] \cdot x' + 1}{x' + 1} \]

\[ \leq \beta \sup_{s' \in S} \frac{[U(s') - V(s')] \cdot x' + 1}{x' + 1} \sum_{s' \in S} (x' + 1) p(s'|s, a_0) \]

\[ = \beta d(u, v) \sum_{s' \in S} (x' + 1) p(s'|s, a_0) \]

\[ \leq \beta d(u, v) (x+2+N_\rho). \]

(3 - 3 - 10)

where the last inequality is due to (3 - 3 - 7) and \( s=(x, \delta) \).
Interchanging \( u \) and \( v \), we obtain

\[
TV(s) - Tu(s) \leq \beta d(u,v) \cdot (x+2+N_r).
\] (3 - 3 - 11)

Combining (3 - 3 - 10) and (3 - 3 - 11) we have

\[
|TV(s) - Tu(s)| \leq \beta d(u,v) \cdot (x+2+N_r).
\] (3 - 3 - 12)

and

\[
T^2u(s) - T^2v(s) = \min_{a \in A_s} \{ \rho(\alpha, s, a) + \sum_{s' \in S} p(s'|s, a)\beta(\alpha, s, a, s')Tu(s') \}
\]

\[
- \min_{a \in A_s} \{ \rho(\alpha, s, a) + \sum_{s' \in S} p(s'|s, a)\beta(\alpha, s, a, s')Tv(s') \}.
\]

Again choose the action, say \( a_1 \in A_s \), such that

\[
\rho(\alpha, s, a_1) + \sum_{s' \in S} p(s'|s, a_1)\beta(\alpha, s, a_1, s')Tv(s')
\]

\[
= \min_{a \in A_s} \{ \rho(\alpha, s, a) + \sum_{s' \in S} p(s'|s, a)\beta(\alpha, s, a, s')Tv(s') \}
\]

and similarly

\[
T^2u(s) - T^2v(s) \leq \sum_{s' \in S} p(s'|s, a_1)\beta(\alpha, s, a_1, s')|TV(s') - Tu(s')| + \beta \sum_{s' \in S} p(s'|s, a_1)|Tu(s') - Tv(s')|
\]
\[ \beta^2 d(u,v) \sum_{s \in S} (x'+2+N_r)p(s'|s, a_1) \]

\[ \leq \beta^2 d(u,v) \cdot \{x+2(2+N_r)\}, \]

where the second, third and the last inequalities are due to (3-3-3), (3-3-12) and (3-3-7) respectively.

Again, the symmetry of \( u \) and \( v \) implies

\[ |T^2 u(s) - T^2 v(s)| \leq \beta^2 d(u,v) \cdot \{x+2(2+N_r)\}. \quad (3-3-13) \]

Now assume that

\[ |T^k u(s) - T^k v(s)| \leq \beta^k d(u,v) \cdot \{x+k(2+N_r)\}. \]

Then

\[ T^{k+1} u(s) - T^{k+1} v(s) = T \cdot T^k u(s) - T \cdot T^k v(s) \]

\[ = \min_{a \in A_s} \left\{ p(\alpha, s, a) + \sum_{s' \in S} p(s' | s, a) \beta(\alpha, s, a, s') T^k u(s') \right\} \]

\[ - \min_{a \in A_s} \left\{ p(\alpha, s, a) + \sum_{s' \in S} p(s' | s, a) \beta(\alpha, s, a, s') T^k v(s') \right\} \]

\[ \leq \beta \sum_{s \in S} p(s' | s, a_k+1) |T^k u(s') - T^k v(s')| \]

\[ \leq \beta^{k+1} d(u,v) \sum_{s \in S} [x'+k(2+N_r)]p(s'|s, a_1) \]
where $a_{k+1} \in A_s$ is such that

$$
\rho(\alpha, s, a_{k+1}) + \sum_{s' \in S} p(s' \mid s, a_{k+1}) \beta(\alpha, s, a_{k+1}, s') T^k v(s')
$$

$$
= \min_{a \in A_s} \left\{ \rho(\alpha, s, a) + \sum_{s' \in S} p(s' \mid s, a) \beta(\alpha, s, a, s') T^k v(s') \right\}
$$

and the second inequality is due to the induction hypothesis.

Using the symmetry of $u$ and $v$ once more we obtain

$$
|T^k u(s) - T^k v(s)| \leq \beta^{k+1} d(u, v) \cdot [x + (k+1)(2+N_r)].
$$

Thus we have proved by induction that $\forall k \geq 1$

$$
|T^k u(s) - T^k v(s)| \leq \beta^k d(u, v) \cdot [x + k(2+N_r)]. \tag{3 - 3 - 14}
$$

Dividing by $x+1$ and taking the supremum in both sides of (3 - 3 -14) gives

$$
d(T^k u, T^k v) = \sup_{s \in S} \frac{|T^k u(s) - T^k v(s)|}{x+1}
$$

$$
\leq \beta^k (u, v) \sup_{s \in S} \frac{x+k(2+N_r)}{x+1}
$$

$$
\leq \beta^k d(u, v) \cdot [1+k(2+N_r)]. \tag{3 - 3 - 15}
$$

Choosing $J = \min \{ K : \beta^k [1+k(2+N_r)] \leq \beta \}$ gives
\[ d(T^j u, T^j v) \leq \beta d(u, v). \]  \hspace{1cm} (3-3-16)

QED.

3.3.3 The Optimality Equation

Since the development in this section depends on the Banach fixed point theory, we report here this well known theory. The proof can be found in any Functional analysis book (see, for example, Edwards (1965)).

**THEOREM (3-5)**

Let \((X, r)\) be a complete semimetric space, and \(T\) a continuous map of \(X\) into itself. If for some natural number \(J\), \(T^J\) is a contraction, then there exists a unique fixed point \(x^* \in X\); that is, \(Tx^* = x^*\).

Since \(T\) is a continuous map (which can be established immediately from (3-3-12)), by the fixed point theorem (3-5), there exists a unique fixed point of \(T\); that is, there exists an unique \(u^* \in B\) such that

\[ Tu^* = u^* \quad \forall \ s \in S, \]  \hspace{1cm} (3-3-17)

or more explicitly,

\[
u^* = \min_{\alpha \in A} \left\{ \rho(\alpha, s, a) + \sum_{s' \in S} p(s' \mid s, a) \beta(\alpha, s, a, s') u^*(s') \right\}
\]

\[
= \min_{\alpha \in A} \left\{ \rho(\alpha, s, a) + \sum_{s' \in S} p(s' \mid s, a) u^*(s') \int_{0}^{\infty} e^{-\alpha \tau} dF(\tau \mid s, a, s') \right\}.
\]
This is exactly the same form of the functional equation given in (3 - 2 - 2). However, it is important to realize that we cannot conclude at this point that the fixed point in (3 - 3 - 17) is the optimal value function \( V_\alpha \) of (3 - 2 - 1). Two relevant questions are whether \( u^*(s) \) can be attained, or even approximated, by \( V_\alpha(\pi, s) \) for some policy \( \pi \), and whether \( u^*(s) \) is the optimal value function \( V_\alpha(s) \). In order to prove that \( u^* \) in (3 - 3 - 17) is actually the optimal value function \( V_\alpha \) of (3 - 2 - 1), we need the following theorem.

**THEOREM (3 - 6)**

The fixed point of \( T \) is the optimal value function of (3 - 2 - 1); that is, 

\[
u^*(s) = V_\alpha(s) = \inf_{\pi \in \Pi} V_\alpha(\pi, s) \quad \forall s \in S. \tag{3-3-18}\]

**Proof.** We first show that both \( T_\alpha \) and \( T \) are monotone functions.

For \( \nu, \nu' \in \mathcal{B} \), we write \( \nu \geq \nu' \) if \( \nu(s) \geq \nu'(s) \) for all \( s \in S \).

Now suppose \( \nu \geq \nu' \). Then for any \( s \in A_\alpha \) and for any \( s' \in S \)

\[
T_\alpha \nu(s) - T_\alpha \nu'(s) = \sum_{s' \in S} p(s', s, a, s') [\nu(s') - \nu'(s')]
\]

\[
\geq \beta \min_{s \in S} \{ \nu(s') - \nu'(s') \}
\]

\[
\geq 0,
\]

and
\[ T_u(s) - T_v(s) = \min_{a \in A_s} T_a u(s) - \min_{a \in A_s} T_a v(s) \]

\[ \geq T_{a_0} u(s) - T_{a_0} v(s) \geq 0, \]

where \( a_0 \) is the action such that \( T_{a_0} u(s) = \min_{a \in A_s} T_a u(s) \) for any given \( s \in S \).

where \( a_0 \) is the action such that \( T_{a_0} v(s) = \min_{a \in A_s} T_a v(s) \) for any given \( s \in S \).

Let \( \pi = \{ \pi_0, \pi_1, \pi_2, \ldots \} \) be an arbitrary policy. Then the value \( V_\alpha(\pi, s) \) of the cost functional corresponding to the policy \( \pi \) with initial state \( s \) is given by

\[ V_\alpha(\pi, s) = E_\pi \left\{ \sum_{i=0}^{\infty} e^{-\alpha(\tau_0 + \tau_1 + \ldots + \tau_i)} \rho(\alpha, s_i, \pi_i) \mid s_0 = s \right\}, \tag{3-3-19} \]

where \( s_i \) and \( \pi_i(s_i) \) are the state at the beginning of the \( i \)th transition and the action chosen according to the policy \( \pi \), and

\[ \rho(\alpha, s_i, \pi_i) = R(s_i, \pi_i) \]

\[ + \sum_{s' \in S} p(s' \mid s_i, \pi_i) \int_0^\tau dF(r \mid s_i, \pi_i, s') \int_0^\tau e^{-\alpha t} dt \]

\[ = R(s_i, \pi_i) + \sum_{s' \in S} p(s' \mid s_i, \pi_i) \rho(\alpha, s_i, \pi_i, s') \]

is the one transition expected cost under policy \( \pi \) (note that this cost is subject to further discount).

Equation (3-3-19) can be rewritten as

\[ V_\alpha(\pi, s) = \rho(\alpha, s, \pi_0) + \sum_{s' \in S} p(s' \mid s, \pi_0) V_\alpha(\pi, s'), \tag{3-3-20} \]
where

\[ \rho(\alpha, s, \pi_0) \] is the expected cost incurred in the first transition using policy \( \pi \)

and

\[
\tilde{V}_{\alpha, \pi}(s, s') = E_\pi \left\{ \sum_{i=1}^{\infty} e^{-\alpha (\tau_0 + \tau_1 + \cdots + \tau_{i-1})} \rho(\alpha, s_i, \pi_i) | s_1 = s' \right\} \tag{3-3-21}
\]

is the expected discounted cost incurred after the first transition.

Therefore, we have

\[
\tilde{V}_{\alpha, \pi}(s, s') = \int_0^\infty V_{\alpha}(s', \pi) e^{-\alpha t} dF(t | s, \pi, s') \leq \int_0^\infty V_{\alpha}(s') e^{-\alpha t} dF(t | s, \pi_i, s') = V_{\alpha}(s') \int_0^\infty e^{-\alpha t} dF(t | s, \pi, s') = \beta(\alpha, s, \pi_0, s') V_{\alpha}(s') \quad \forall \ s \in S , \tag{3-3-22}
\]

where

\[ V_{\alpha}(s') = \inf_{\pi \in \Pi} V_{\alpha}(\pi, s') \quad \forall \ s' \in S . \]

Hence
\[ V_\alpha(\pi, s) = \rho(\alpha, s, \pi_0) + \sum_{s' \in S} p(s'|s, \pi_0) V_\alpha(\pi, s') \]

\[ \geq \rho(\alpha, s, \pi_0) + \sum_{s' \in S} p(s'|s, \pi_0) \beta(\alpha, s, \pi_0, s') V_\alpha(s') \]

\[ \geq \min_{\pi \in \Lambda_\alpha} \{ \rho(\alpha, s, \pi_0) + \sum_{s' \in S} p(s'|s, \pi_0) \beta(\alpha, s, \pi_0, s') V_\alpha(s') \} \]

\[ = TV_\alpha(s), \quad (3 - 3 - 23) \]

where the first inequality is due to (3 - 3 - 22) and the last equality is by the definition of \( T \).

Taking infimum in both sides of (3 - 3 - 23) over all policies, we have

\[ \inf_{\pi \in \Pi} V_\alpha(\pi, s) = V_\alpha(s) \geq TV_\alpha(s) \quad \forall s \in S. \quad (3 - 3 - 24) \]

By the monotonicity of map \( T \), we have

\[ V_\alpha(s) \geq TV_\alpha(s) \geq T^2 V_\alpha(s) \geq \cdots \geq T^m V_\alpha(s) \quad \forall s \in S, \quad \forall m \geq 1. \quad (3 - 3 - 25) \]

Since \( \nu^* \) is the fixed point of \( T \) and \( V_\alpha \in \mathcal{B}, \) therefore

\[ \lim_{k \to \infty} T^k V_\alpha(s) = \nu^*(s) \quad \forall s \in S; \]

that is
\[ V_\alpha(s) \geq V^*(s) \quad \forall s \in S \quad (3.3.26) \]

The proof of the reverse inequality is much more tedious and technical.

For any sequence \( \{\varepsilon_k\} \) with \( \varepsilon_k > 0 \), let \( \mu = (\mu_0, \mu_1, \ldots) \) be a policy such that \( \forall s \in S \) and \( \forall k \geq 0 \)

\[ T\mu^*(s) + \varepsilon_k \geq T\mu^*_\kappa(s). \quad (3.3.27) \]

For the existence of such policies, see Bertsekas (1976) and Hinder (1970).

We can again write \( T\mu^*_\kappa(s) \) as

\[ T\mu^*_\kappa(s) = \rho(\alpha, s, \mu_\kappa) + \sum_{s' \in S} p(s'|s, \mu_\kappa)\beta(\alpha, s, \mu_\kappa, s')V^*(s') \]

and (2.2.14) as

\[ V_\alpha(\pi, s) = \lim_{N \to \infty} E_\pi \{ \sum_{i=0}^{N-1} e^{-\alpha(\tau_1 + \tau_2 + \cdots + \tau_i)} \rho(\alpha, s_i, \pi_i) | s_0 = s) \} \]

\[ = \lim_{N \to \infty} E_\pi V_{\alpha, N-1}(s). \]

We rewrite \( E_\pi V_{\alpha, N-1}(s) \) as

\[ E_\pi V_{\alpha, N-1}(s) = E_\pi \{ \rho(\alpha, s, \pi_0) \]
\[
\sum_{i=1}^{N-1} \prod_{j=1}^{l} \beta(\alpha, s_{j-1}, \pi_j, s_j) \rho(\alpha, s_i, \pi_i) | s_0 = s \}
\]

\[\leq \liminf_{N \to \infty} \prod_{j=1}^{N} \beta(\alpha, s_{j-1}, \pi_j, s_j) \ u^*(s_N) + V_{\alpha, N-1}(\pi, s)\]

\[\leq \liminf_{N \to \infty} \prod_{j=1}^{N} \beta(\alpha, s_{j-1}, \mu_j, s_j) \ u^*(s_N) + V_{\alpha, N-1}(\mu, s)\]

(3 - 3 - 28)

Where (3 - 3 - 18) is obtained by replacing \(\infty\) by \(N-1\) in (2 - 2 - 14) with \(\tau_j\) being the length of the \(j\)th transition interval, and \(\rho(\alpha, s, \pi_0)\) is the cost incurred in the first transition if action \(\pi_0\) is taken according the policy \(\pi\).

Under this notation, we have

\[V_{\alpha}(s) = \inf_{\pi \in \Pi} \lim_{N \to \infty} E_{\pi} V_{\alpha, N-1}(\pi, s)\]

\[\leq \inf_{\pi \in \Pi} \liminf_{N \to \infty} E_{\pi} \{ \prod_{j=1}^{N} \beta(\alpha, s_{j-1}, \pi_j, s_j) \ u^*(s_N) + V_{\alpha, N-1}(\pi, s) \}\]

\[\leq \liminf_{N \to \infty} E_{\mu} \{ \prod_{j=1}^{N} \beta(\alpha, s_{j-1}, \mu_j, s_j) \ u^*(s_N) + V_{\alpha, N-1}(\mu, s) \},\]

(3 - 3 - 29)

since

\[E_{\mu} \{ \prod_{j=1}^{N} \beta(\alpha, s_{j-1}, \mu_j, s_j) \ u^*(s_N) + V_{\alpha, N-1}(\mu, s) \}\]

\[= E_{\mu} \{ \prod_{j=1}^{N-1} \beta(\alpha, s_{j-1}, \mu_j, s_j) \ [\beta(\alpha, s_{N-1}, \mu_N, s_j) u^*(s_N) \]

...
Let us observe that

\[ E_{\mu}\{ [\beta(\alpha, s_{N-1}, \mu_{N-1}, s_N)u^*(s_N)+\rho(\alpha, s_{N-1}, \mu_{N-1})]\mid s_{N-1}\} \]

\[ = p(\alpha, s, \mu_{N-1}) + \sum_{s_n \in S} p(s_n\mid s_{N-1}, \mu_{N-1})\beta(\alpha, s_{N-1}, \mu_{N-1}, s_n)u^*(s_n) \]

\[ = \mu_{N-1}^* u^*(s_{N-1}). \]  \tag{3-3-31}

Therefore, by substituting (3-3-30) and (3-3-16) into (3-3-29) and noting that \( u^* \) is a fixed point of \( T \) we have

\[ E_{\mu}\{ \prod_{j=1}^{N} \beta(\alpha, s_{j-1}, \mu_j, s_j) u^*(s_N) + V_{\alpha,N-1}(\mu, s) \} \]

\[ = E_{\mu}\{ \prod_{j=1}^{N-1} \beta(\alpha, s_{j-1}, \mu_j, s_j) T_{\mu}^{*}\mid s_{N-1}\} + V_{\alpha,N-2}(s) \}

\[ \leq E_{\mu}\{ \prod_{j=1}^{N-1} \beta(\alpha, s_{j-1}, \mu_j, s_j) (Tu^*(s_{N-1})+\epsilon_{N-1}) + V_{\alpha,N-2}(\mu, s) \} \]
\[
\begin{align*}
&\leq E_{\mu}\left( \prod_{j=1}^{N-1} \beta(\alpha, s_{j-1}, \mu_j, s_j) \, T \psi^*(s_{N-1}) + V_{\alpha, N-2}(\mu, s_1) \right) + \beta^{N-1} e_{N-1} \\
&= E_{\mu}\left( \prod_{j=1}^{N-1} \beta(\alpha, s_{j-1}, \mu_j, s_j) \, \psi^*(s_{N-1}) + V_{\alpha, N-2}(\mu, s_1) \right) + \beta^{N-1} e_{N-1},
\end{align*}
\]

where the second inequality is due to (3 - 3 - 3).

Similarly, we have
\[
\begin{align*}
&\leq E_{\mu}\left( \prod_{j=1}^{N-1} \beta(\alpha, s_{j-1}, \mu_j, s_j) \, \psi^*(s_{N-1}) + V_{\alpha, N-2}(\mu, s_1) \right) \leq \psi^*(s) + \sum_{k=0}^{N-1} \beta^k e_k.
\end{align*}
\]

Using this argument repeatedly, we obtain the following:
\[
\begin{align*}
&\leq E_{\mu}\left( \prod_{j=1}^{N} \beta(\alpha, s_{j-1}, \mu_j, s_j) \, \psi^*(s_N) + V_{\alpha, N-1}(\mu, s_N) \right) \leq \psi^*(s) + \sum_{k=0}^{N-1} \beta^k e_k.
\end{align*}
\]

Substituting (3 - 3 - 32) and (3 - 3 - 33) into (3 - 3 - 29), we have
\[
\begin{align*}
V_{\alpha}(s) &\leq \psi^*(s) + \lim_{N \to \infty} \sum_{k=0}^{N-1} \beta^k e_k.
\end{align*}
\]
Since the sequence \( \{\varepsilon_k\} \) is arbitrary, we can select \( \{\varepsilon_k\} \) so that \( \lim_{N \to \infty} \sum_{k=0}^{N-1} \beta^k \varepsilon_k \) is arbitrarily close to zero. Therefore

\[ V_\alpha(s) \leq u^*(s) \quad \forall s \in S \quad (3 - 3 - 35) \]

and the result follows by combining (3 - 3 - 26) and (3 - 3 - 35). QED.

3.3.4 The Existence of Stationary Policy

We have already showed that the optimal value function is the unique fixed point of the map \( T \). Therefore, (3 - 3 - 17) is the optimality equation, and any function \( \psi \in \mathcal{B} \) satisfying (3 - 3 - 17) is the unique optimal value function and the corresponding policy is optimal. The remaining question now is whether this optimal policy can be a stationary one.

THEOREM (3 - 7)

There exists a stationary policy \( f \) such that the fixed point \( \psi \) of \( T_f \) defined by

\[
T_f \psi(s) = \rho(\alpha, s, f(s)) + \sum_{s' \in S} p(s' | s, f(s)) \psi(s') \int e^{-\alpha \tau} dF(\tau | s, f(s), s')
\]

\[
= \rho(\alpha, s, f(s)) + \sum_{s' \in S} p(s' | s, f(s)) \psi(s') \beta(\alpha, s, f(s), s'), \quad (3 - 3 - 35)
\]

satisfies
\( u(s) = u^*(s) = V_a(s) \quad \forall \ s \in S. \quad (3-3-36) \)

Proof. For any policy \( \pi = (\pi_0, \pi_1, \cdots) \), the finiteness of the action space \( A \) implies that the sequence \( \{ T_{\pi_n}u(s) \} \) contains only finitely many different elements and therefore \( \forall \ a \in A \), there exists an \( n \) such that
\[
T_a u(s) = T_{\pi_n} u(s) \quad \forall \ s \in S
\]
and
\[
Tv(s) = \min_{a \in A} T_a u(s) = \min_n T_{\pi_n} u(s) \quad \forall \ s \in S,
\]
also
\[
u^*(s) = \min_n T_{\pi_n} u^*(s) \quad \forall \ s \in S. \quad (3-3-37)
\]

Take a partition of \( S \) by \( S = \bigcup_n G_n \), where
\[G_n = \{ s \in S : T_{\pi_n} u^*(s) = u^*(s), T_{\pi_j} u^*(s) > u^*(s), j=0, 1, 2, \ldots, n-1 \}\]

Thus, \( n \) is the smallest \( i \) such that \( T_{\pi_i}u^*(s) = u^*(s) \). We now define \( f = \pi_n \) on \( G_n \) and therefore
\[
T_f u^*(s) = u^*(s).
\]

On the other hand, \( u \) is the fixed point of \( T_f \). Thus,
\( u(s) = v^*(s) \quad \forall s \in S \quad \text{QED.} \)

Remark: The stationary policy so obtained is called \( \pi \)-generated by Blackwell (1966).

The original definition is stated as follows:

Definition: For any non-randomized policy \( \pi = \{\pi_0, \pi_1, \ldots\} \) we say that \( f: S \to A \) is \( \pi \)-generated if there exists a partition of \( S \) into Borel sets \( S_1, S_2, \ldots \) such that \( f = \pi_n \) on \( S_n \).

3.4 Average Cost Optimal Policy

We have derived the optimal policy for the discounted cost criteria in the previous sections. In this section, we will establish the existence of the optimal policy for the average cost criteria via the results of the previous sections.

When the cost incurred in one-transition is uniformly bounded, conditions for the existence of the average cost optimal stationary policy are given by Ross (1970). Lippman (1973) investigated the situation when the one transition cost is bounded by a polynomial of the state variable \( x \). Sennott (1989) which deals with average cost, discrete time Markov decision process, where the assumption of a polynomial bound is relaxed but the author assumes the existence of a non-negative \( M_i \) such that for every \( i \) and every \( 0 < \alpha < 1 \),

\[ h_\alpha(i) = v_\alpha(i) - v_\alpha(0) \leq M_i. \]

Also, for every \( i \), there exists an action \( a(i) \) such that \( \Sigma P(j|i, a(i))M_j < \infty \). and for all \( a \in A \) \( \Sigma P(j|i, a)M_j < \infty \).
The following theorem is an extension of Ross (1970) where the expected per transition cost is uniformly bounded, also the cost structure and state space are different from this study. But the proof of Ross is still valid except for the change of notations.

**THEOREM (3-8)**

If there exists a bounded function $h(s)$ and a constant $g$ such that

$$h(s) = \min_{a \in A_s} \left\{ \tilde{C}(s, a) + \sum_{s' \in S} p(s' | s, a) \tau(s, a, s') h(s) - g \tau(s, a) \right\}, \quad (3-4-1)$$

then there exists a policy (average cost) $\pi^*$ such that

$$g = V(s, \pi^*) = \inf_{\pi \in \Pi} V(\pi, s) \quad \forall s \in S, \quad (3-4-2)$$

and $\pi^*$ is any policy which, for each $s \in S$, prescribes an action which minimizes the right side of (3-4-1), where

$$\tilde{C}(s, a) = R(s, a) + \sum_{s' \in S} p(s' | s, a) \tau(s, a, s') \int_0^\infty t \, dF(t | s, a, s'),$$

and

$$\tau(s, a) = \sum_{s' \in S} p(s' | s, a) \int_0^\infty t \, dF(t | s, a, s').$$
are the expected cost in one transition and the length of the transition interval respectively, when action \(a\) is taken at state \(s\).

Proof. Let \(G_n = \{s_0, a_0, s_1, a_1, \ldots, s_n, a_n\}\) denote the history of the process up to the \(n\)th transition. Then for any policy \(\pi\),

\[
E_\pi\left\{\sum_{i=0}^{N-1} h(s_{i+1}) - E_\pi(h(s_{i+1}) | G_i)\right\} = 0
\]

But

\[
E_\pi(h(s_{i+1}) | H_i) = \sum_{s_{i+1} \in S} h(s') p(s_{i+1} | s_i, a_i)
\]

\[
= C(s_i, a_i) + \sum_{s_{i+1} \in S} h(s') p(s_{i+1} | s_i, a_i) - C(s_i, a_i) + g\tau(s_i, a_i) - g\tau(s_i, a_i)
\]

\[
\geq \min_{a \in A_{s_i}} \{ C(s_i, a) + \sum_{s_{i+1} \in S} h(s') p(s_{i+1} | s_i, a_i) - g\tau(s_i, a) - C(s_i, a_i) + g\tau(s_i, a_i) \}
\]

\[
= h(s_i) - C(s_i, a_i) + g\tau(s_i, a_i)
\]

(3-4-3)

with equality for \(\pi^*\).

Hence, by taking expectation in both sides of (3-4-3) we have

\[
E_\pi\left\{h(s_{i+1}) - h(s_i) + C(s_i, a_i) - g\tau(s_i, a_i)\right\} \geq 0
\]

and therefore
or equivalently

$$
E_{\pi} \left\{ \sum_{i=0}^{N-1} \left[ h(s_{i+1}) - h(s_i) + \bar{C}(s_i, a_i) - g(t(s_i, a_i)) \right] \right\} \geq 0,
$$

(3 - 4 - 4)

with equality for $\pi^\ast$.

Since

$$
\tau(s_i, a_i) = \sum_{s' \in \mathcal{S}} p(s' \mid s_i, a_i) \int_0^\infty \, dF(t \mid s_i, a_i, s')
$$

$$
\geq \sum_{s' \in \mathcal{S}} p(s' \mid s_i, a_i) \int_0^\infty \frac{dF(t \mid s_i, a_i, s')}{\delta}
$$

$$
\geq \delta \sum_{s' \in \mathcal{S}} p(s' \mid s_i, a_i) [1 - F(\delta \mid s_i, a_i, s')]
$$

$$
\geq \delta \varepsilon,
$$

where the last inequality is due to condition (2 - 2 - 4).

we have
Now letting $N \to \infty$, using the boundedness of $h(s)$ and (3 - 4 - 5) we have

$$
E_{\pi} \sum_{i=0}^{N-1} \tilde{\tau}(s_i, a_i) \geq (N-1)\delta \varepsilon .
$$

(3 - 4 - 5)

for all values of $s_0$ and with equality for $\pi^*$.

QED.

The following theorem gives a sufficient condition for the existence of an average cost optimal stationary policy, which is again an extension of the uniformly bounded expected one transition cost situation, but the proof is much more technical and tedious.

**THEOREM (3 -9)**

Assume that there exists a state, say $s^0$, and a function $Z(s)$ such that

$$
|h_\alpha(s)| = |V_\alpha(s) - V_\alpha(s^0)| \leq Z(s), \forall s \in S \quad \text{and} \quad \forall 0 < \alpha < 1.
$$

If for some stationary $\alpha$-discounted policy $f_\alpha$,

$$
\sum_{s' \in S} p(s'|s, f_\alpha(s))Z(s') < \infty ,
$$

then there exists a stationary policy $\pi^*$ such that
\[ g\tau(s, a) + h(s) = \overline{C}(s, f(s)) + \sum_{s' \in S} p(s'|s, f(s))h(s), \]  
\hspace{1cm} (3 - 4 - 6) 

where

\[ g = \lim_{n \to \infty} \alpha_n V\alpha_n(s^0) \]  
\hspace{1cm} (3 - 4 - 7) 

and

\[ h(s) = \lim_{n \to \infty} (V\alpha_n^+(s) - V\alpha_n^-(s^0)) \]  
\hspace{1cm} (3 - 4 - 8) 

for some subsequences \( \{n'\} \) and \( \{n''\} \) of \( \{n\} \) such that \( \alpha_{n'} \downarrow 0 \) and \( \alpha_{n''} \downarrow 0 \).

Proof. Since \( A_S \) is finite for each \( s \in S \), the set \( A = \bigcap_{s \in S} A_s \) is compact by the Tichonoff theorem. For all sequences \( \{\alpha_n\} \) such that \( 0 < \alpha_n < 1 \) and \( \alpha_n \downarrow 0 \), let \( \{f_{\alpha_n}\} \) be the \( \alpha \)-discounted stationary policy*. Then \( f_{\alpha_n} \in A \) \( \forall n \). The compactness of \( A \) guarantees the existence of a subsequence \( \{\alpha_{n(k)}\} \) with \( \alpha_{n(k)} \downarrow 0 \) and a stationary policy \( f \in A \) such that

\[ f_{\alpha_{n(k)}} \to f \in A \]  
\hspace{1cm} as \( k \to \infty \).

Now let \( f_{\alpha} \) denote the \( \alpha \)-discounted optimal stationary policy. Then

* The existence of such policies has already been proved in the previous sections, for example, we can always take the optimal stationary policy established in section 3.3.3.
\[ h_{\alpha}(s) = V_{\alpha}(s) - V_{\alpha}(s^0) \]

\[ = C(s, f_{\alpha}(s)) + \sum_{s' \in S} p(s' \mid s, f_{\alpha}(s))\beta(\alpha, s, f_{\alpha}(s), s')V_{\alpha}(s) - V_{\alpha}(s^0) \]

\[ = C(s, f_{\alpha}(s)) + \sum_{s' \in S} p(s' \mid s, f_{\alpha}(s))\beta(\alpha, s, f_{\alpha}(s), s')[V_{\alpha}(s) - V_{\alpha}(s^0)] \]

\[ - V_{\alpha}(s^0)[1 - \sum_{s' \in S} p(s' \mid s, f_{\alpha}(s))\beta(\alpha, s, f_{\alpha}(s), s')] \]

\[ = C(s, f_{\alpha}(s)) + \sum_{s' \in S} p(s' \mid s, f_{\alpha}(s))\beta(\alpha, s, f_{\alpha}(s), s')h_{\alpha}(s') \]

\[ - V_{\alpha}(s^0)[1 - \beta(\alpha, s, f_{\alpha}(s))] \]  

(3 - 4 - 9)

where

\[ \beta(\alpha, s, f_{\alpha}(s), s') = \sum_{s' \in S} p(s' \mid s, f_{\alpha}(s))\beta(\alpha, s, f_{\alpha}(s), s') \]

\[ = \sum_{s' \in S} p(s' \mid s, f_{\alpha}(s)) \int_0^{\infty} e^{-\alpha t} dF(t \mid s, f_{\alpha}(s), s') \]

\[ = \sum_{s' \in S} p(s' \mid s, f_{\alpha}(s)) \int_0^{\infty} [1 + \alpha t + o(\alpha)] dF(t \mid s, f_{\alpha}(s), s') \]

\[ = 1 + \alpha \sum_{s' \in S} p(s' \mid s, f_{\alpha}(s)) \int_0^{\infty} dF(t \mid s, f_{\alpha}(s), s') + o(\alpha) \]

\[ = 1 + \alpha \tau(s, f_{\alpha}(s)) + o(\alpha). \]  

(3 - 4 - 10)
Substituting (3.4.10) into (3.4.9), we obtain

\[ h_\alpha(s) = \bar{C}(s, f_\alpha(s)) + \sum_{s' \in S} p(s' \mid s, f_\alpha(s)) \beta(\alpha, s, f_\alpha(s), s') h_\alpha(s') \]

\[ - \alpha \bar{\tau}(s, f_\alpha(s)) V_\alpha(s^0) + o(\alpha), \]

or

\[ \alpha V_\alpha(s^0) = \frac{1}{\bar{\tau}(s, f_\alpha(s))} \left[ \bar{C}(s, f_\alpha(s)) + \sum_{s' \in S} p(s' \mid s, f_\alpha(s)) \beta(\alpha, s, f_\alpha(s), s') h_\alpha(s') \right] \]

\[ + o(\alpha). \quad (3.4.11) \]

Both \( \bar{C}(s, f_\alpha(s)) \) and \( h_\alpha(s) \) are bounded by functions of \( x \) for all \( 0 < \alpha < 1 \) and

\[ \sum_{s' \in S} p(s' \mid s, f_\alpha(s)) \beta(\alpha, s, f_\alpha(s), s') h_\alpha(s') \]

\[ \leq \sum_{s' \in S} p(s' \mid s, f_\alpha(s)) Z(s') < \infty \quad \forall \ 0 < \alpha < 1, \quad (3.4.12) \]

which implies

\[ \alpha V_\alpha(s^0) \leq M \quad \forall \ 0 < \alpha < 1, \]

where
\[
M = \inf_{s \in S} \frac{1}{\bar{\tau}(s, \alpha(s))} \{ C(s, f_{\alpha}(s)) + \sum_{s' \in S} p(s' \mid s, f_{\alpha}(s))Z(s') + h_{\alpha}(s) \}.
\]

Therefore we can choose a further subsequence \( \{ \alpha_{n'} \} \) of the subsequence \( \{ \alpha_{n(k)} \} \) along which \( f_{\alpha_{n(k)}} \rightarrow \tilde{f} \) such that

\[
\alpha_{n'}V_{\alpha_{n'}}(s^0) \rightarrow g \quad \text{as } n'
\]

Again, the fact that \( h_{\alpha}(s) \leq Z(s) \quad \forall \alpha < 1 \) and \( \forall s \in S \) implies we can find another subsequence \( \{ \alpha_{n''} \} \) of \( \{ \alpha_{n'} \} \) such that

\[
h_{\alpha_{n''}}(s) = V_{\alpha_{n''}}(s) - V_{\alpha_{n''}}(s^0) \rightarrow h(s) \quad \text{as } n'' \uparrow \infty \quad (\alpha_{n'} \downarrow 0).
\]

There remains to show that

\[
\lim_{n'' \uparrow \infty} \sum_{s' \in S} p(s' \mid s, f_{\alpha_{n''}}(s), \beta(\alpha_{n''}, s, f_{\alpha_{n''}}(s), s')h_{\alpha_{n''}}(s') = \sum_{s' \in S} p(s' \mid s, \tilde{f}(s))h(s').
\]

(3-4-13)

We shall first observe that if \( x \), the number of customers in the system is large enough, then \( h_{\alpha}(s) \) is increasing in \( x \). More importantly, when \( x \) is large enough, \( h_{\alpha}(s) \) is a monotone function of \( \alpha \), because \( h_{\alpha}(s) = V_{\alpha}(s) - V_{\alpha}(s^0) \) is just the relative cost due to the difference of the initial state \( s \) from \( s^0 \). Therefore, when \( x \) is large enough, \( h_{\alpha}(s) > 0 \) and is monotone decreasing in \( \alpha \).

Now let \( D > 0 \) be such that

\[
h_{\alpha}(s) > 0 \quad \forall x \geq D \text{ and } \forall 0 < \alpha < 1 \quad s = (x, s).
\]
Since $\beta(\alpha, s, f_\alpha(s), s')h(s') > 0$ and is decreasing in $\alpha$, therefore the function $\beta(\alpha_n, s, f_{\alpha_n}(s), s')h_{\alpha_n}(s')$ is monotone increasing in $n'$ and

$$\lim_{n' \to \infty} \sum_{s' \in S} p(s'| s, f_{\alpha_n}(s))\beta(\alpha_n, s, f_{\alpha_n}(s), s')h_{\alpha_n}(s')$$

$$= \sum_{\delta = 0}^{D} \sum_{x' = 0}^{D} \lim_{n' \to \infty} \sum_{x' = 0}^{D} p(s'| s, f_{\alpha_n}(s))\beta(\alpha_n, s, f_{\alpha_n}(s), s')h_{\alpha_n}(s')$$

$$+ \sum_{\delta = 0}^{D} \lim_{n' \to \infty} \sum_{x' = D+1}^{D} p(s'| s, f_{\alpha_n}(s))\beta(\alpha_n, s, f_{\alpha_n}(s), s')h_{\alpha_n}(s')$$

$$= \sum_{\delta = 0}^{D} \sum_{x' = 0}^{D} p(s'| s, f(s))h(s')$$

$$+ \sum_{\delta = 0}^{D} \lim_{n' \to \infty} \sum_{x' = D+1}^{D} p(s'| s, f(s))\beta(\alpha_n, s, f_{\alpha_n}(s), s')h_{\alpha_n}(s')$$

$$+ \sum_{\delta = 0}^{D} \lim_{n' \to \infty} \sum_{x' = D+1}^{D} [p(s'| s, f_{\alpha_n}(s)) - p(s'| s, f(s))]\beta(\alpha_n, s, f_{\alpha_n}(s), s')h_{\alpha_n}(s')$$

(3 - 4 - 14)

Notice that the last term is well defined because

$$\sum_{x' = D+1}^{\infty} [p(s'| s, f_{\alpha_n}(s)) - p(s'| s, f(s))]\beta(\alpha_n, s, f_{\alpha_n}(s), s')h_{\alpha_n}(s') < \infty,$$

where the second equality in (3 - 4 - 15) is due to the fact that $\beta(\alpha_n, s, f_{\alpha_n}(s), s') \uparrow 1$ as $n'' \uparrow \infty$ ($\alpha_n \downarrow 0$).

Now using the monotone convergence theorem we have
\[
\lim_{n' \to \infty} \sum_{x' = D+1}^{\infty} p(s'|s, f(s)) \beta(\alpha_{n''}, s, f\alpha_{n''}(s), s') h\alpha_{n''}(s')
\]
\[
= \sum_{x' = D+1}^{\infty} p(s'|s, f(s)) h(s')
\]  
(3-4-15)

Also
\[
\lim_{n' \to \infty} \sum_{x' = D+1}^{\infty} [p(s'|s, f\alpha_{n''}(s)) - p(s'|s, f(s))] \beta(\alpha_{n''}, s, f\alpha_{n''}(s), s') h\alpha_{n''}(s')
\]
\[
\leq \limsup_{n'' \to \infty} \sum_{x' = D+1}^{\infty} |p(s'|s, f\alpha_{n''}(s)) - p(s'|s, f(s))| Z(s')
\]
\[
\leq \sum_{x' = D+1}^{\infty} \limsup_{n'' \to \infty} |p(s'|s, f\alpha_{n''}(s)) - p(s'|s, f(s))| Z(s')
\]
\[
= 0,
\]  
(3-4-16)

where the last inequality is due to Fatou's Lemma.

Substituting (3-4-15) and (3-4-16) into (3-4-14) gives (3-4-13) and the desired result follows by letting \( \alpha \downarrow 0 \) along the sequence \( \{\alpha_{n''}\} \) in (3-4-9). The optimality of \( f(s) \) follows from theorem (3-8). QED.
CHAPTER IV

OPTIMIZATION OF A G/M/1 QUEUE WITH BREAKDOWNS

4.1 Model Formulation And Optimal Policy

In this chapter, we will find an optimal policy for the queueing system described in chapter two. This system can be controlled by turning it on or off, by choosing a different repairing rate when a breakdown arrives, or a combination of the two. We will study the first case only. In a very recent paper by Federgruen and So (1990), the second case is studied. The existence of an optimal stationary policy for the third case is proved in chapter three, but the exact form of the optimal policy needs further study. For convenience, we recollect the notations here.

Consider a single server queue with customers arriving according to a homogeneous Poisson process with parameter $\lambda$. Each customer requires a random service time $t_s$ with finite second moment. When the system is rendering service to customers, it is subject to random breakdowns which arrive according to a homogeneous Poisson process with parameter $\xi$. Each breakdown requires a random time $t_r$ with finite second moment to repair. If no control is imposed on the system, then the repair process starts immediately after the breakdown of the system and the service to the interrupted...
customer resumes as soon as the repair is completed and no loss of service is involved, that is, the service station undergoes a cycle of two parts: for a period of time it is operative, then it breaks down, and after the completion of repair, the cycle starts again.

The system is controlled by turning on the system at any point when the system is off and turning it off at service completion epochs during a busy period.

We now impose the cost structure in the following manner:

(i) The start-up cost $C_s$. This is the cost incurred for starting up a dormant system.

(ii) The maintenance cost rate $C_m$. This is the cost incurred in keeping the system on.

(iii) The delay cost rate $C_d$. This is the cost incurred for delaying the customers in the system.

**THEOREM (4-1)**

If the system can be turned off only at epochs of service completion when the system is on, and can be turned on only at epochs of new arrival when the system is off, then there is an optimal stationary policy of the form:

$$f(s) = \begin{cases} 
0 & \text{if } x=0 \\
\delta & \text{if } 0 < x < R \\
1 + \delta & \text{if } x \geq R
\end{cases}$$

(4.1.1)

for some positive integer $R$. 
where $1 \vee \delta = \max \{1, \delta \}$ and

$$f(s) = 0$$ means to turn (or leave) the system off.

$$f(s) = 1$$ means to turn (or leave) the system on.

$$f(s) = 2$$ means to clear the breakdown.

$\delta = 0, 1, 2$ represents the situations where the system is off, the system is on and is capable of rendering service to customers, and the system is on but is incapacitated by breakdowns.

Proof. The optimality of the stationary policy $f$ in the class of stationary policies is proved by Sobel (1969). The desired results follows from Theorems (3-7) and (3-8).

Remark: Sobel (1969) shows that for a single server GI/G/1 queue, among the class of stationary policies, the optimal policy is to shut down the system when no customer remains. Start it up when a queue of length $R$ develops and provide service until the system is empty. The result is obtained with random walk arguments. In the model of this study, the system is also subject to breakdowns, but we can reduce it to a M/G/1 queue with the service time being made up of two parts: the time spent on actual service to customers and the time spent on repairing breakdowns. Also notice that at the epochs of service completion, the system is in proper working order with probability one since the event that a breakdown arrives at the moment of a service completion has probability zero.
COROLLARY (4-1)

If the system can be turned on or off at any time \( t' \in T' = \{ t \geq 0 : s(t) \neq s(t-) \} \), the stationary policy \( \pi \) in (4-1) is still optimal.

Proof. Let \( T'' = \{ t \geq 0 : x(t-) - x(t) = 1 \} \cup \{ t \geq 0 : x(t) - x(t-) = 1 \text{ and } \delta(t) = 0 \} \) and \( \{ t''_0, t''_1, \ldots \} \) be the sequence of ordered elements in \( T'' \) with \( t''_0 = 0 \), and \( t''_k \) being the \( k \)th opportunity to turn the system on or off under the control of the stationary policy \( \pi \) in theorem (4-1) (notice the elements of \( T'' \) can be ordered since \( T' \supset T'' \)).

Now for \( t_0' \in T' \) such that \( t_0' \in T'' \), we have \( \delta(t_0') \neq 0 \) since \( T' \cap \{ \delta = 0 \} \in T'' \). The event which happened at \( t_0' \) cannot be a completion of service since \( t_0' \in T'' \). Therefore, at the moment of \( t_0' \), there is at least one customer in the system whose service is not completed. If we do not shut down the system, then it is equivalent to the situation in which no control will be imposed until the service of the leading customer is completed. If we take the action "turn the system off", then this customer will be held until the system again starts rendering service to customers and an extra delay cost is added to this customer. The policy choosing this action is thus dominated by the policy which will choose the actions to complete the service of this customer before shutting down. Therefore an optimal policy takes actions only at \( t'' \in T'' \). QED.

4.2 The Residual Life Distribution

While the server is rendering service to a customer, new customers arrive according to a homogeneous Poisson process. The residual life of service time is defined as the remaining actual service time required by the leading customer, that is, the time interval
between the arrival of the new customer and the termination of the service of the leading customer if there are no breakdowns in this period.

**LEMMA (4-1)**

Let \( W_s, t_s \) denote the residual life of service time and the service time respectively. Then the density and expected value of the residual life are given by:

\[
g_s(w) = \frac{1}{E(t_s)} \int_0^\infty \! dF_s(u) \quad (4 - 2 - 1)
\]

and

\[
E(W_s) = E(t_s) \frac{1 + \gamma_s^2}{2} \quad (4 - 2 - 2)
\]

where

\[
\gamma_s = \frac{\text{Var}(t_s)^{0.5}}{E(t_s)} \quad (4 - 2 - 3)
\]

is the coefficient of variation of \( t_s \) and \( F(\cdot) \) is the distribution function of the service time.

Proof. Since customers arrive at the service station according to a homogeneous Poisson process, given there is a customer who arrived during the service period, the arriving time is uniformly distributed and therefore

\[
P(W_s \geq w) = \int_w^\infty P(t_s \geq t) dt.
\]
Since

\[ 1 = P(W_s \geq 0) = C \int_0^\infty P(t_s \geq t) dt = CE(t_s), \]

we have

\[ P(W_s \geq w) = \frac{1}{E(t_s)} \int_0^\infty P(t_s \geq t) dt \quad (4 - 2 - 4) \]

and

\[ g_s(w) = -\frac{d}{dw} P(W_s \geq w) = \frac{1}{E(t_s)} \int_0^\infty dF_s(u). \]

Also

\[ E(W_s) = \int_0^\infty P(W_s \geq w) dw \]

\[ = \frac{1}{E(t_s)} \int_0^\infty dw \int_0^t P(t_s \geq t) dt. \]

Using Fubini Theorem, we obtain the following by changing the order of integration:

\[ E(W_s) = \frac{1}{E(t_s)} \int_0^t \int_0^\infty P(t_s \geq t) dt dw \]

\[ = \frac{1}{E(t_s)} \int_0^t P(t_s \geq t) dt, \quad (4 - 2 - 5) \]
Since

\[ E(W_s) \leq E(t_s) < \infty, \]

we can integrate (4 - 2 - 5) by parts and obtain

\[ E(W_s) = \frac{1}{2E(t_s)} \int_0^\infty t^2 dF_s(t) = \frac{\text{Var}(t_s) + E^2(t_s)}{2E(t_s)} \]

\[ = E(t_s) \frac{1 + \gamma^2}{2}. \]

QED.

Remark: Equation (4 - 2 - 2) can also be obtained from the result of Cox and Smith (1954) by noticing the fact that there is no mutual dependence between the initiation of service to a customer and the arrival of another customer and therefore, we can assume the process started long time ago.

The residual life for the repairing time \( t_r \) is defined similarly and is denoted by \( W_r \). Clearly, the density and the expected value of the residual life of the repairing time \( t_r \) are

\[ g_r(w) = \frac{1}{E(t_r)} \int_u^\infty dF_r(u) \]  

(4 - 2 - 6)

and

\[ E(W_r) = E(t_r) \frac{1 + \gamma^2}{2}, \]  

(4 - 2 - 7)

where
\[
\gamma_r = \frac{\text{Var}(t_r)^{0.5}}{E(t_r)}
\]

is the coefficient of variation of the repairing time \( t_r \).

### 4.3 Busy Period And Busy Cycle

#### 4.3.1 The Differential Equation of The Queueing System

The following notations will be used to derive the differential equation of the queue system.

- \( X(t) \): number of customers in the system at time \( t \).
- \( Z(t) \): time spent on serving the leading customer up to time \( t \).
- \( Y(t) \): time spent on repairing the current breakdowns up to \( t \). \( Y(t)=0 \) if \( \delta(t^-) \neq 2 \).

\[
\varphi_n(u, t) = P\{Z(t) \leq u, X(t) = n | \delta(t) = 1\}
\]

\[
\psi_n(u, v, t) = \frac{\partial}{\partial u} P\{Z(t) \leq u, Y(t) \leq v, X(t) = n | \delta(t) = 2\}
\]

\[
U(t) = P\{\delta(t) = 0, X(t) = 0\}
\]

Let \( \eta(t) \) and \( \zeta(t) \) be the intensity functions of service time and repairing time respectively. Then

\[
F_S(t) = 1 - \exp\left\{ - \int_0^t \eta(\tau) d\tau \right\}, \quad (4-3-1)
\]
and

\[ F_r(t) = 1 - \exp\left\{ - \int_0^t \zeta(\tau) d\tau \right\}. \] (4 - 3 - 2)

Using the continuity arguments concerning the motion of the system in \((t, t+\Delta)\), we have, to the first order in \(\Delta\)

\[ \phi_n(u, t+\Delta) = \phi_n(u, t) \left[ 1 - (\lambda + \eta(u) + \xi)\Delta \right] + \lambda \Delta \phi_{n-1}(u, t) \]

\[ + \Delta \int_0^\infty \psi_n(u, v, t) \zeta(t) dt \] (4 - 3 - 3)

\[ \psi_n(u, v+\Delta, t+\Delta) = [1 - (\lambda + \zeta(u))] \psi_n(u, v, t) + \lambda \Delta \psi_{n-1}(u, v, t) \]

\[ + \Delta \int_0^\infty \eta(u) \phi_1(u, t) du \] (4 - 3 - 4)

\[ U(t+\Delta) = [1 - \lambda \Delta] U(t) + \Delta \int_0^\infty \eta(u) \phi_1(u, t) du \] (4 - 3 - 5)

Rearranging (4 - 3 - 3) - (4 - 3 - 5) and dividing by \(\Delta\) in both sides, then letting \(\Delta \rightarrow 0\), we obtain

\[ \frac{\partial \phi_n(u, t)}{\partial t} + \frac{\partial \phi_n(u, t)}{\partial u} = (\lambda + \eta(u) + \xi) \phi_n(u, t) + \lambda \phi_{n-1}(u, t) \]

\[ + \int_0^\infty \zeta(u) \psi_n(u, u, t) du \] (4 - 3 - 6)

\[ \frac{\partial \psi_n(u, v, t)}{\partial t} + \frac{\partial \psi_n(u, v, t)}{\partial u} = -[\lambda + \zeta(u)] \psi_n(u, v, t) + \lambda \psi_{n-1}(u, v, t) \] (4 - 3 - 7)
\[
\frac{\partial U(t)}{\partial t} = -\lambda U(t) + \int_0^\infty \eta(u)\varphi_1(u, t)\,du
\]  
(4-3-8)

Letting \( t \to \infty \), we obtain the differential equations at steady state:

\[
\frac{d\varphi_n(u)}{du} = (\lambda + \eta(u) + \xi)\varphi_n(u) + \lambda \varphi_{n-1}(u) + \int_0^\infty \zeta(u)\psi_n(u, u)\,du
\]  
(4-3-9)

\[
\frac{\partial \psi_n(u,v)}{\partial v} = -[\lambda + \zeta(u)]\psi_n(u,v) + \lambda \psi_{n-1}(u,v)
\]  
(4-3-10)

\[
\lambda U(t) = \int_0^\infty \eta(u)\varphi_1(u)\,du.
\]  
(4-3-11)

We also have the following boundary conditions:

\[
\psi_n(0, v) = 0
\]

\[
\psi_n(u, 0) = \xi \varphi_n(u)
\]

\[
\varphi_n(0) = \int_0^\infty \eta(u)\varphi_{n-1}(u)\,du + \lambda U I_{\{n=1\}}.
\]

The first boundary condition is due to the fact that the system is subject to breakdowns only when it is rendering service to customers and \( I_{\{n=1\}} \) is an indicator function.

\*
\* the condition that ensures the existence of a stationary distribution is 
\( \lambda E(t_3)[1+\zeta E(t_4)] < 1 \), which is given in (4-3-33).
This system of differential equations is quite difficult to solve even if the analytic solution exists. The computation of the busy period based on this system of equations is much more complicated. Fortunately, we have an alternative way which does not require solving the differential equations.

4.3.2 Computation of The Busy Period by Iteration

In the queueing theory, a busy period is defined as beginning with the arrival of a customer to an idle channel and ending when the channel next becomes idle. A busy cycle is the sum of a busy period and an adjacent idle period. When the queueing system is subject to breakdowns and is controlled by policy \( f \) given in \( (4 - 1 - 1) \), we define a busy period as the period when the system is on and a busy cycle as the sum of the adjacent on and off periods.

DEFINITION: The traffic intensity \( \rho \) of a queueing system is defined by (see, e.g. Karlin and Taylor 1981)

\[
\rho = \frac{\text{expected length of service time per customer}}{\text{expected length of interarrival time}};
\]

\( \rho \) is also called the utilizing factor or busy fraction since it is by definition, a measure of the average use of the service facility.

We now define \( \rho_s \) and \( \rho_r \) by

\[
\rho_s = \lambda E(t_s) \quad (4 - 3 - 12)
\]

\[
\rho_r = \xi E(t_r) \quad (4 - 3 - 13)
\]
with the interpretation that $\rho_s$ and $\rho_r$ are busy fraction of serving customers and repairing breakdowns.

For a $G/M/1$ queue with arrival rate $\lambda$ and service time distribution function $B(\cdot)$, the distribution function of a busy period $G(\cdot)$ satisfies (see, e.g Cox, 1961)

$$G^*(s) = B^*(\lambda + s - \lambda G^*(s)) \quad (4 - 3 - 14)$$

where

$$G^*(s) = \int_0^\infty e^{-sh} dG(h) \quad (4 - 3 - 15)$$

and

$$B^*(s) = \int_0^\infty e^{-st} dB(t) \quad (4 - 3 - 16)$$

are the Laplace-Stieljes transformation of $G(\cdot)$ and $B(\cdot)$.

Therefore, the expected length of a busy period $L$ is given by

$$E(L) = -\left. \frac{dG^*(s)}{ds} \right|_{s=0} = -G^{**}(0). \quad (4 - 3 - 17)$$

Since

$$\frac{dG^*(s)}{ds} = B^{**}[s + \lambda - G^*(s)], \quad ((4 - 3 - 18))$$

we have
\[ E(L) = - \frac{dG^*(s)}{ds} \bigg|_{s=0} = -B^*(0)[1 + \lambda E(L)] \]

or

\[ E(L) = - \frac{B^*(0)}{1 + \lambda B^*(0)} = \frac{\mu}{1 - \lambda \mu} \quad (4 - 3 - 19) \]

where

\[ \mu = -B^*(0) = \int_0^\infty dB(t) \quad (2 - 3 - 20) \]

is the expected length of service required by one customer.

We now define another queueing process as follows:

Suppose we have a single server service station with two kinds of customers, primary and ordinary. The ordinary customers arrive at the station according to a homogeneous Poisson process with parameter \( \lambda \) and require a random service time \( t_0 \) with finite second moment. During the period of rendering service to ordinary customers, primary customers arrive at the station according to a homogeneous Poisson process with parameter \( \xi \) and require a random service time \( t_r \) with finite second moment. The service to the ordinary customer is interrupted by the arriving primary customer and is resumed immediately when the service to the primary customer is completed.

This queueing process clearly has the same busy period as the queueing process we are studying. Since the length of the busy period is independent of queue disciplines,
we will adopt the following discipline to compute the busy period of the process defined above:

Every busy period starts by serving the ordinary customers in the queue when the queue size reaches \( R \) and the server will keep serving the ordinary customers until no one left in the system. The primary customers who arrived in this period will form their own queue without interrupting the service to ordinary customers. When there is no ordinary customers left in the system, the server starts serving the primary customers who arrived in the period of serving the ordinary customers and the arriving ordinary customers will form their own queue (during this period, no primary customer will arrive). After the service to the primary customers is completed, the server starts serving the ordinary customers arrived during this period and another cycle starts. The busy period of the service station ends when there neither kind of customers is in the system.

Let \( T_{s(i)}, K^{(i)}, T_{r(i)} \) denote the length of the \( i \)th period of serving the ordinary customers, the number of primary customers arrived during this period and the time required to complete the service of these primary customers. Then

\[
E(K^{(i)}) = E\{E(K^{(i)} | T_{s(i)})\} = \xi E(T_{s(i)}) \quad (4-3-21)
\]

and

\[
E(T_{r(i)}) = E\{E(T_{r(i)} | K^{(i)})\} = E(t_r)E(T_{s(i)}) \quad (4-3-22)
\]

Since the server keeps serving the ordinary customers until their queue is empty, the expected length of \( T_{s(1)} \) is actually the expected value of the busy period of a M/G/1
queue and, since the distribution of a busy period induced by R customers is just the convolution of distributions of n busy periods each induced by one customer, therefore, from (4 - 3 - 19) we have

\[ E(T_{s(1)}) = R \frac{E(t)}{1 - \lambda E(t)} = \frac{R E(t)}{1 - \rho_s} = R \tau_b. \quad (4 - 3 - 23) \]

Thus

\[ E(K(1)) = \xi E(T_{s(1)}) = R \xi \tau_b \quad (4 - 3 - 24) \]

and

\[ E(T_{r(1)}) = E(t_r)E(T_{s(1)}) = R \xi \tau_b E(t_r) \]

\[ = R \rho_r \tau_b. \quad (4 - 3 - 25) \]

During the period of \( T_{r(1)} \), there are a random number \( R^{(2)} \) of ordinary customers who arrived and the expected length of the busy period of serving the ordinary customers induced by these \( R^{(2)} \) ordinary customers is

\[ E(T_{s(2)}) = E \{ E(T_{s(2)}|R^{(2)}) \} = \tau_b E(R^{(2)}) \]

\[ = \tau_b E \{ E(R^{(2)}|T_{r(1)}) \} = \tau_b \lambda E(T^{(1)}) \]

\[ = \lambda R \rho_r \tau_b^2 \quad (4 - 3 - 26) \]

where the last equality is due to (4 - 3 - 25).

Similarly, the expected number of primary customers who arrived during \( T_{s(2)} \) is
\( E(K^{(2)}) = E\{E(K^{(2)}|T_s^{(2)})\} = \xi E(T_s^{(2)}) \)

\[ = \lambda R \xi \rho_r t_b^2 \quad (4-3-27) \]

and the expected time required to complete the service of these \( K^{(2)} \) primary customers is

\( E(T_r^{(2)}) = E\{E(T_r^{(2)}|K^{(2)})\} = E(t_r)E(K^{(2)}) \)

\[ = \lambda R (\rho_r t_b)^2 . \quad (4-3-28) \]

Repeating this procedure, we obtain

\( E(T_s^{(m)}) = R t_b (\lambda \rho_r t_b)^{m-1} \quad (4-3-29) \)

and

\( E(T_r^{(m)}) = R \lambda^{m-1}(\rho_r t_b)^m . \quad (4-3-30) \)

Now we assume the parameters satisfy the following condition:

\[ \rho_s(1+\rho_r) = \lambda E(t_s)[1+\xi E(t_r)] < 1 . \quad (4-3-31) \]

This is a sufficient condition for the existence of stationary distribution of the process.

Let \( t_s^b \) and \( t_r^b \) denote the expected length of periods that the station is serving ordinary and primary customers in a busy period. Then
The expected length of busy period $B_R$ of the station is

$$B_R = t_{s}^{b} + t_{r}^{b} = R \frac{(1+\rho_{r})E(t_{s})}{1-(1+\rho_{r})\rho_{s}}, \quad (4 - 3 - 34)$$

where the subscript $R$ indicates that a busy period starts when a queue of length $R$ develops.

In the queueing model with breakdowns, we can interpret $\mu_b = (1+\rho_{r})E(t_{s})$ as the average time the system spent on each customer. Let $b = \lambda(1+\rho_{r})E(t_{s})$. Then $b$ is the busy fraction of the system and $(4 - 3 - 34)$ can be written as

$$B_R = R \frac{\mu_b}{1-b}. \quad (4 - 3 - 35)$$

The expected busy cycle of the system is

$$\Phi = B_R + \frac{R}{\lambda} = \frac{R}{\lambda(1-b)}. \quad (4 - 3 - 46)$$
4.4 Expected Queueing Time And Queue Length

The queueing time of a customer is defined as the time that the customer spent in the system, including both the time of waiting in the system and the time of receiving service. At steady state, we have

\[ P(\delta = 0) = \frac{R}{\lambda \Phi} = 1 - b \quad (4 - 4 - 1) \]

and

\[ P(\delta \neq 0) = 1 - P(\delta = 0) = b \quad (4 - 4 - 2) \]

\[ P(\delta = 1) = \frac{b\lambda}{\Phi} = \lambda E(t_b) = \rho_s \quad (4 - 4 - 3) \]

\[ P(\delta = 2) = \frac{b^2}{\Phi} = \rho_s \rho_{tr} \quad (4 - 4 - 4) \]

LEMMA (4-2)

Let \( N \) be the number of customers in the queue whom an arriving customer finds. Then under the control of policy (4 - 1 - 1), at steady state, we have

\[ P(N=i, \delta = 0) = \frac{1 - b}{R} \quad i = 0, 1, \ldots, R - 1. \quad (4 - 4 - 5) \]

Proof. Let \( t_i \) be the arrival epoch of the \( i \)th customer and let \( t \) be the epoch that an arbitrary customer arrives at the system in a busy cycle. Given \( t \leq t_k \), we know \( t = t_i \) for some \( i \leq k \). and

\[ P(N \geq i | \delta = 0) = P(t \geq t_i | t \leq t_k) \]
\[
= \mathbb{P}(t_i \leq t \leq t_k | t \leq t_k)
\]
\[
= \mathbb{E}\{\mathbb{P}(t_i \leq t \leq t_k, t_1, t_2, \ldots t_k)\}
\]
\[
= \mathbb{E}\left\{\frac{\sum_{j=1}^{i-1} \tau_j}{R}\right\}
\]
\[
\sum_{j=1}^{i} \tau_j
\]

where

\[
\tau_j = t_j - t_{j-1} \quad j = 1, 2, \ldots, k.
\]

and the last equality of (4 - 4 - 6) is due to the fact that given a customer arrived before \(t_k\), the arrival time is uniformly distributed in \((0, t_k)\).

Since \(\tau_k\)'s are i.i.d (independent identically distributed) \(\exp(\lambda)\), we have

\[
1 = \mathbb{E}\left\{\frac{\sum_{j=1}^{i-1} \tau_j}{R}\right\} = R \mathbb{E}\left\{\tau_i^{-1} \sum_{j=1}^{R} \tau_j\right\} \quad \forall \ i = 1, 2, \ldots, R
\]

or

\[
\mathbb{E}\left\{\tau_i^{-1} \sum_{j=1}^{R} \tau_j\right\} = \frac{1}{R} \quad \forall \ i = 1, 2, \ldots, R.
\]  

(4 - 4 - 7)

Therefore
\[ P(N \geq i \mid \delta = 0) = (K-i-1) E \left\{ \tau_i^{-1} \sum_{j=1}^{R} \tau_j \right\} = \frac{R - i - 1}{R} \]

\[ \forall \ i = 0, 1, 2, \ldots, R-1. \quad (4-4-8) \]

Similarly,

\[ P(N \geq i + 1 \mid \delta = 0) = (K-i-2) E \left\{ \tau_i^{-1} \sum_{j=1}^{R} \tau_j \right\} = \frac{R - i - 2}{R} \]

\[ \forall \ i = 0, 1, 2, \ldots, R-2. \quad (4-4-9) \]

\[ P(N=i \mid \delta = 0) = P(N \geq i \mid \delta = 0) - P(N \geq i + 1 \mid \delta = 0) = \frac{1}{k} \]

\[ \forall \ i = 0, 1, 2, \ldots, R. \quad (4-4-10) \]

Thus

\[ P(N=i, \delta = 0) = P(N=i \mid \delta = 0)P(\delta = 0) = \frac{1-b}{R} \]

\[ \forall \ i = 0, 1, 2, \ldots, R-1. \quad (4-4-11) \]

QED.

Let \( Q \) be the expected number of customers that an arriving customer finds at steady state. Then

\[ Q = \sum_{i=0}^{R-1} i P(N=i, \delta = 0) + \sum_{i=0}^{\infty} i P(N=i, \delta = 1) + \sum_{i=0}^{\infty} i P(N=i, \delta = 2). \quad (4-4-12) \]
Now let us consider the time that a customer has to wait until his service is completed, that is, the time elapsed between his arrival and the termination of his service. This waiting time can be decomposed into four parts:

(i) The time between the arrival of the customer and the start of the system to render service to customers in the queue, if the arriving customer finds that the system is turned off and the queue size is less than $R-1$.

(ii) The time to complete the service of the customer in service if the arriving customer finds the server is serving a customer. This is the residual time of the service time $t_s$.

(iii) The time to complete the repair of a breakdown if the arriving customer finds the server is dealing with a breakdown. This is the residual time of the repairing time $t_r$.

(iv) The time elapsed between the departure of the customer in service and the departure of the new customer.

During the residual time of service, the system breaks down randomly from time to time and the expected number of breakdowns in this period is given by

$$E(D_w) = E\{E(D_w | W_s)\} = \xi E(W_s)$$

$$= \xi E(t_s) \frac{1 + \gamma W_s^2}{2}, \quad (4-4-13)$$

where $D_w$ is the number of breakdowns in the period of $W_s$. 
Let $T_d$ denote the time required to repair the $D_w$ breakdowns. Then we have

$$E(T_d) = E\{E(T_d | D_w)\} = E(t_r)E(D_w)$$

$$= \rho_r E(t_s) \frac{1+\gamma_s^2}{2}. \quad (4-4-14)$$

Thus the expected time required to clear the customers already in the service is

$$E(W) = E(T_d) + E(W_s) = (1+\rho_r)E(t_s) \frac{1+\gamma_s^2}{2} \quad (4-4-15)$$

and the expected waiting time of the customer who finds $i$ customers in the system and the server serving the leading customer is

$$E(t_w | N = i, \delta = 1) = (1+\rho_r)E(t_s) \frac{1+\gamma_s^2}{2} + i(1+\rho_r)E(t_s) \quad (4-4-16)$$

where the second term of the right side of $(4-3-16)$ is obtained as follows:

If there were no breakdowns, the expected time to complete $i$ customers is $iE(t_s)$. But during this period, the expected number of breakdowns is $iE(t_s)\gamma_s^2$ and the result follows by adding the expected service time and the expected repairing time. Similarly

$$E(t_w | N = i, \delta = 2) = E(t_r) \frac{1+\gamma_r^2}{2} + E(t_w | N = i, \delta = 1)$$

$$= E(t_r) \frac{1+\gamma_r^2}{2} + (1+\rho_r)E(t_s) \frac{1+\gamma_s^2}{2} + i(1+\rho_r)E(t_s) \quad (4-4-17)$$

and
\[ E(t_w | N=i, \delta=0) = \frac{R-1}{\lambda} + (1+ i)(1+p_r)E(t_s) \quad \forall i = 0, 1, \ldots, R-1. \]

(4 - 4 - 18)

Therefore, the expected waiting time for an arbitrary customer is given by

\[
E(W) = \sum_{i=0}^{R-1} E(t_w | N=i, \delta=0) P(N=i, \delta=0) \\
+ \sum_{i=1}^{\infty} E(t_w | N=i, \delta=1) P(N=i, \delta=1) + \sum_{i=2}^{\infty} E(t_w | N=i, \delta=2) P(N=i, \delta=2)
\]

\[
= \sum_{i=0}^{R-1} \left[ \frac{R-i-1}{\lambda} + (1+i)(1+p_r)E(t_s) \right] P(N=i, \delta=0)
\]

\[
+ \sum_{i=1}^{\infty} \left[ (1+p_r)E(t_s) \frac{1+\gamma_s^2}{2} + i(1+p_r)E(t_s) \right] P(N=i, \delta=1)
\]

\[
+ \sum_{i=1}^{\infty} \left[ E(t_s) \frac{1+\gamma_s^2}{2} + (1+p_r)E(t_s) \frac{1+\gamma_s^2}{2} + i(1+p_r)E(t_s) \right] P(N=i, \delta=2)
\]

\[
= \sum_{i=0}^{R-1} \left[ \frac{R-i-1}{\lambda} + (1+p_r)E(t_s) \right] P(N=i, \delta=0)
\]

\[
+ \sum_{i=1}^{\infty} \left[ (1+p_r)E(t_s) \frac{1+\gamma_s^2}{2} \right] P(N=i, \delta=1)
\]

\[
+ \sum_{i=1}^{\infty} \left[ E(t_s) \frac{1+\gamma_s^2}{2} + (1+p_r)E(t_s) \frac{1+\gamma_s^2}{2} \right] P(N=i, \delta=2)
\]
Substituting (4 - 4 - 5) and (4 - 4 -12) into (4 - 4 - 19), we obtain

\[
E(W) = \frac{(1-b)(R-1)}{2\lambda} + (1+p_r)(1-b)E(t_o) + (1+p_r)E(t_s) \frac{1+\gamma^2}{2} P(\delta=1)
+ \left[ E(t_r) \frac{1+\gamma^2}{2} + (1+p_r)E(t_s) \frac{1+\gamma^2}{2} \right] P(\delta=2) + (1+p_r)E(t_s)Q
\]

\[
= \frac{(1-b)(R-1)}{2\lambda} + (1+p_r)(1-b)E(t_o) + (1+p_r)E(t_s) \frac{1+\gamma^2}{2} P(\delta=1)
+ E(t_r) \frac{1+\gamma^2}{2} P(\delta=2) + (1+p_r)E(t_s)Q. \tag{4 - 4 - 20}
\]

Substituting (4 - 4 - 3) and (4 - 4 - 4) into (4 - 4 - 20), we obtain

\[
E(W) = \frac{(1-b)(R-1)}{2\lambda} + (1+p_r)(1-b)E(t_o) + (1+p_r)E(t_s) \frac{1+\gamma^2}{2} b
+ E(t_r) \frac{1+\gamma^2}{2} \rho_r \rho_s + (1+p_r)E(t_s)Q. \tag{4 - 4 - 21}
\]

Now we can use Little's formula (Little, 1961)

\[
Q = \lambda E(W) \tag{4 - 4 - 22}
\]

to obtain
If \( R=1, \xi=0 \), which is the situation for \( M/G/1 \) queue without breakdown and control, then (4-3-23) reduces to

\[
Q = \frac{(1-b)(R-1)}{2\lambda} + b(1-b) + \rho_s^2 \cdot 1 + \gamma_s^2 \cdot \frac{1}{2} + \lambda E(t_r) \cdot \frac{1+\gamma_r^2}{2} \rho_r \rho_s + bQ
\]

or

\[
Q = \frac{R-1}{2} + b + \frac{1+\gamma_s^2}{2(1-b)} b^2 + \frac{\lambda E(t_r)}{2(1-b)} \rho_r \rho_s.
\] (4-4-23)

If \( R=1, \xi=0 \), which is the situation for \( M/G/1 \) queue without breakdown and control, then (4-3-23) reduces to

\[
Q = \rho_s + \frac{1+\gamma_s^2}{2(1-\rho_s)} \rho_s^2
\] (4-4-24)

and this is known as Pollaczek's formula.

4.5 Optimization

Using theorem (2-1), the long run expected average cost of the queueing system under the control of policy (4-1-1) can be written as

\[
C_f(R) = \frac{C(\Phi)}{\Phi}
\] (4-5-1)

where \( \Phi \) is the expected length of a busy cycle at steady state given in (4-3-46) and \( C(\Phi) \) is the expected cost incurred in this period.

Since the delay cost per unit time is proportional to the average number of customers in the system, we have

\[
C_f(R) = \frac{C_s}{\Phi} + QC_d + bC_m
\]
The optimal value $R$ is any positive integer value $R^*$ such that

$$C_f(R^*) = \min_{R \in I} \{ C_f(R) \}$$

where $I$ is the set of all positive integers.

**LEMMA (4-3)**

The cost function $C_f(R)$ is convex in $R$.

Proof. $C_f(R)$ can be written as

$$C_f(R) = \frac{\lambda (1-b)C_s}{R} + L(R),$$

where $L(R)$ is a linear function of $R$.

Since $\frac{1}{R}$ is convex in $R$, therefore

$$C_f(\alpha R_1+(1-\alpha)R_2) = \frac{\lambda (1-b)C_s}{\alpha R_1+(1-\alpha)R_2} + L(\alpha R_1+(1-\alpha)R_2)$$

$$\leq \alpha \frac{\lambda (1-b)C_s}{R_1} + (1-\alpha) \frac{\lambda (1-b)C_s}{R_2} + \alpha L(R_1)+(1-\alpha) L(R_2)$$

$$= \alpha C_f(R_1)+(1-\alpha)C_f(R_2).$$

QED.

Thus, $R^*$ must satisfy the following conditions:

$$C_f(R^*+1) \geq C_f(R^*)$$
After some simplification, we obtain

\[
R^* (R^*-1) \leq \frac{2\lambda(1-b)C_s}{C_d} 
\]  
(4-5-6)

\[
R^* (R^*+1) \geq \frac{2\lambda(1-b)C_s}{C_d} 
\]  
(4-5-7)

The smallest integer R* which satisfies (4-5-6) and (4-5-7) is the optimal value.

An alternative way of finding the optimal value is as follows: If we treat R as a continuous value, then we can find the optimal value R⁰ of R by solving \( \frac{dC_f(R)}{dR} = 0 \).

The optimal value so obtained is not necessarily an integer. But since \( C_f(R) \) is a convex function of R, the optimal integer value of R is one of the integers surrounding \( R^0 \) (e.g., French et al., 1986, p131). Therefore, we have

\[
R^* = [R^0] \quad \text{or} \quad R^* = [R^0] + 1
\]

where
Remark: As we have noticed, the optimal value $R^*$ is independent of the maintenance cost rate $C_m$. This is because when we decide to control the process by turning it on or off, the busy fraction $b$ is a constant no matter what value $R^*$ is as long as no loss of service is involved. Then the question that arises naturally is what is the upper bound, $C^u$, of $C_m$, such that for all $C_m \leq C^u$, the optimal policy is "keep the system on all the time". To answer this question we need to compute the long run expected average cost without control. We denote this cost by $C^*$, so that

$$C^* = Q_1 C_d + C_m$$  \hspace{1cm} (4-5-9)$$

where

$$Q_1 = b + \frac{1+\rho_s^2}{2(1-b)} b^2 + \frac{\lambda E(t_0)}{2(1-b)} \rho_s \rho_s$$  \hspace{1cm} (4-5-10)$$

is the expected queue length when the system is always on. Which is obtained by putting $R = 1$ in (4-4-23).

Let

$$\sigma = C_f(R^*) - C^*$$

$$= \frac{\lambda (1-b) C_s}{R^*} - (1-b) C_m + \frac{R^*-1}{2} C_d.$$  \hspace{1cm} (4-5-11)$$
Therefore, we control the system if $\sigma < 0$ and leave it on all the time otherwise. That is, we control the system when

\[ C_m \geq \frac{\lambda C_s}{R^*} + \frac{R^*-1}{2(1-b)} C_d. \]  

(4 - 5 - 12)

4.6 Summary

In this dissertation, I studied the control of a M/G/1 queueing process subject to breakdowns with a linear cost function. The existence of the optimal stationary policy for the average cost criterion is proved in chapter three. This optimal stationary policy is a limit point of the optimal stationary policy for the discounted cost criterion as the discount factor $\alpha$ goes to zero. In chapter four, a simplified case is studied and is compared with the existing results. A simple formula to calculate the optimal starting value $R$ is derived. To use this result, we only need to know the first two moments of the service and repairing time, instead of the exact distributions. One important feature of this result is the closed and simple form of the solution. Also, the results obtained here is more general than that of Lippman's (1973) in the sense that (1). The system is subject to breakdowns in this study. (2). The distribution of service time is general in this study, but Lippman (1973) assumes the distribution of the service time is exponential.
REFERENCES


