stichting
mathematisch
centrum

AFDELING MATHEMATISCH BESLISKUNDE BN 24/74 AUGUST

G. DE LEVE AND H.C. TIJMS

A GENERAL MARKOV DECISION METHOD, WITH APPLICATIONS TO CONTROLLED QUEUEING SYSTEMS

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

AMS (MOS) subject classification scheme (1970): 90C45, 90C40, 93C60, 60K25, 60J0.

CONTENTS

0. INTRODUCTION 1 - 2

1. MODEL AND METHOD 3 - 27

1.1. The materials of the model 3 - 6

1.2. A formula for the average cost and a system of functional equations 6 - 11

1.3. Basic tools 12 - 17

1.4. The solution techniques 17 - 19

1.5. Proofs of Theorem 1.5 and the Lemmas 1.1 and 1.2 20 - 22

1.6. Convergence results 22 - 27

2. APPLICATIONS TO CONTROLLED QUEUEING SYSTEMS 28 - 44

2.1. A switch-over policy for controlling the workload in the M|M|1 queue with switch-over costs 28 - 35

2.2. An algorithm for the switch-over policy controlling the queue size in the M|G|1 queue with two service types and switch-over costs 35 - 44

APPENDIX 45 - 46

REFERENCES 47 - 48

0. INTRODUCTION.

In this paper we consider a Markov decision model studied by De Leve (1964). This model is a generalization of the finite-state semi-Markov decision model (cf. Howard (1960) and Jewell (1963)) and allows for general state spaces and for controlling the system at each point of time. The optimality criterion is the *long-run average cost*. The original presentation in De Leve (1964) is difficult accessible due to the generality of the model considered. This paper treats the general Markov decision model under the assumption that the decision processes are regenerative processes as is the case in almost any application. Under this assumption a self-contained exposition of the model will be given with proofs that have been considerably simplified. Emphasis will be laid on the presentation of a policy iteration method.

It is characterisic for our model to consider any decision process as a superposition of a so-called natural process and interventions made in certain states of the natural process. The natural process could be considered as a process describing the evolution of the state of the system when no interventions are made. In solving any particular problem we have first to specify the natural process where to a certain extent we are free in the choice of this process provided that the result of the natural process and the control by a policy agrees with the "reality" of the problem. The natural process underlies the policy iteration method and for a particular problem the final form of this method is determined by the choice of the natural process. Since the structure of the problem considered is reflected in the natural process, this structure will be exploited by the policy iteration method. This may result in a simple algorithm. An iteration step of the policy iteration method consists not only of the familiar value determination operation and policy improvement operation but also of a cutting operation. It was notified by Weeda (1974a) that the cutting operation is in fact the optimal stopping of a Markov process.

Chapter 1 deals with the model. In section 1.1 we define the materials of the model. Section 1.2 discusses a formula for the average cost of a (stationary) policy. This formula in itself may be useful, in particular in controlled inventory and queueing systems. In section 1.3 we give the basic tools for the solution techniques. These techniques consisting of a direct and an iterative approach are summarized in section 1.4. In section 1.5 we prove several results stated in section 1.3. Convergence results for the policy iteration method are established in section 1.6. In chapter 2 we give two applications to controlled queueing systems. In section 2.1 we apply the formula for the average cost to a control policy that switches from rate 1 to rate 2 when the amount of work in the system exceeds the level y_1 and switches from rate 2 to rate 1 when the work in the system falls to the level $y_2 < y_1$. In section 2.2 we derive a very simple algorithm for an M|G|1 queue with two service types where the decision to use which service type is based on the queue size. Finally, in the appendix we state some results for discrete-time Markov processes with a general state space.

Throughout this paper the words set and function serve as abbreviations for Borel set and real-valued Baire function. The phrase "the system enters the set A at time t" means that the system is in the set A at time t but not at time t^-.

1. MODEL AND METHOD

1.1. *The materials of the model.*

This section formulates the materials of the model. For any particular problem these materials have to be first specified before the actual solution of the problem can be started.

PROPERTY 1. *There is a state space X such that at each point of time the state of the system can be described by a point in X, where X is a subset of a finite dimensional Euclidean space.*

PROPERTY 2. *There is stochastic process called the natural process. This process has X as state space and could be considered as a process describing the evolution of the state of the system when no interventions are made. The natural process is a strong Markov process having stationary transition probabilities, and sample paths which are almost surely right continuous and have a finite number of discontinuities in any finite time interval.*

We note that in most applications the choice of the state space and the natural process involves the use of the supplementary variable technique. The natural process will be controlled by interventions.

PROPERTY 3. *For each state* $x \in X$ *there is a finite set* $D(x)$ *of feasible decisions in state* x*, where between null-decisions and interventions is distinguished. A null-decision is a decision that does not disturb the natural process. An intervention is a decision that interrupts the natural process and causes an instantaneous (possibly random) change of the state of the system.*

For any $x \in X_0$ and intervention $d \in D(x)$, let [*)]

$\underline{T}_{x,d}$ = the state into which the system is transferred instantaneously by the intervention d in state x,

[*)] In this paper random variables are underlined.

where

$$X_0 = \{x \in X \mid D(x) \text{ contains an intervention}\}$$

Observe that we may assume that a transition caused by an intervention takes no time because at each point of time the state of the system is defined. In most applications the effect of an intervention is deterministic. The elements of the properties 1-3 have to be chosen in such a way that the following property holds.

PROPERTY 4. *The states in which the null-decision is not feasible constitute a non-empty closed set* A_0 *(say) such that for each initial state, with probability* 1, *the natural process will eventually reach the set* A_0. *Further, with probability* 1, *any intervention in a state of* A_0 *causes an instantaneous transition to a state outside* A_0.

PROPERTY 5. *In the natural process there is incurred a cost at rate* $c_1(x)$ *when the system is in state* x, *and there is an immediate cost* $c_2(x,y)$ *at time* t *when the natural process is in state* x *at time* t^- *and is in state* y *at time* t *where* $x \neq y$. *There is incurred an immediate decision cost* $c_3(x,d)$ *when in state* x *the intervention* $d \in D(x)$ *is made. The functions* c_1, c_2 *and* c_3 *are non-negative.**)

In the next property we introduce the quantities $k(x;d)$ and $t(x;d)$. It will appear hereafter that in our model these quantities play the same role as the one-step expected costs and transition times in the semi-Markov decision model. The sets A_{01} and A_{02} introduced in property 6 are only used to define the functions $k(x;d)$ and $t(x;d)$ and may be freely chosen.

PROPERTY 6. *Choose two non-empty closed sets* $A_{01} \subseteq A_0$ *and* $A_{02} \subseteq A_0$ *such that for each initial state, with probability* 1, *the natural process will eventually reach* A_{0i} *for* $i = 1,2$. *Let* $k_0(x) = 0$ *for* $x \in A_{01}$, *and, for* $x \notin A_{01}$, *let* $k_0(x)$ *be the expected cost incurred up to and including the first epoch at which the system enters the set* A_{01} *when the system is sub-*

---

*) The non-negativity assumption is made for convenience.

jected to the natural process and is in state x *at epoch* 0. *For any* $x \in X_0$ *and intervention* $d \in D(x)$, *let* $k_1(x;d) = c_3(x,d) + Ek_0(\underline{T}_{x,d})$. *That is,* $k_1(x;d)$ *the expected cost incurred up to and including the first epoch at which the system enters* A_{01} *when at epoch* 0 *intervention* d *is made in state* x *and after this intervention the system is subjected to the natural process with the state resulting from this intervention as initial state.* *Similarly, let* $t_0(x) = 0$ *for* $x \in A_{02}$, *and, for* $x \notin A_{02}$, *let* $t_0(x)$ *be the expectation of the first epoch at which the system enters the set* A_0 *when the system is subjected to the natural process and is in state* x *at epoch* 0. *For any* $x \in X_0$ *and intervention* $d \in D(x)$, *let* $t_1(x;d) = Et_0(\underline{T}_{x,d})$. *It is assumed that the functions* k_0, k_1, t_0 *and* t_1 *are finite. For any* $x \in X_0$ *and intervention* $d \in D(x)$, *let*

$$k(x;d) = k_1(x;d) - k_0(x) \text{ and } t(x;d) = t_1(x;d) - t_0(x).$$

The class of policies we will consider is denoted by Z. This class has the following property

PROPERTY 7. *Any policy* $z \in Z$ *is a function that adds to each state* $x \in X$ *a single decision* $z(x) \in D(x)$. *The states in which policy* $z \in Z$ *prescribes an intervention constitute a closed set* A_z *such that* $\Pr\{\underline{T}_{x,z(x)} \in A_z\} = 0$ *for all* $x \in A_z$ *and* $\Pr\{\underline{T}_{x,z(x)} \in A\}$ *is a Baire function of* $x \in A_z$ *for any set A.* *)

REMARK 1.1. The process resulting from the control of the natural process by a policy $z \in Z$ is called the *decision process* of policy z. Between two successive interventions the behaviour of the decision process is described by the natural process. It is characteristic for our model to regard any decision process as a superposition of the natural process and interventions made in certain states. As a consequence of this view we have some freedom in choosing the natural process for the particular problem to be solved provided that the resulting decision processes agree with the "reality" of that problem. Moreover, this view enables us to exploit fully

---

*) Observe that $A_0 \subseteq A_z \subseteq X_0$ for all $z \in Z$.

any structure of the problem considered. It will appear hereafter that the choice of the natural process is determinative for the final form of the policy iteration method.

Finally, we introduce the following notation. For any state $x \in X$ and closed set $A \supseteq A_0$, let

$\underline{S}[x,A]$ = the first state in the set A taken on by the natural process starting from state x.

Observe that, by $A \supseteq A_0$ and property 4, the random variable $\underline{S}[x,A]$ is well-defined. Also, observe that $\underline{S}[x,A] = x$ when $x \in A$.

1.2. *A formula for the average cost and a system of functional equations.*

This section derives a formula for the average cost of a policy from Z and discusses the system of functional equations to be solved in the value determination operation. Unless stated otherwise, we assume that a fixed policy $z \in Z$ is used. We make the following assumptions.

ASSUMPTION 1. *For any policy $z \in Z$ there are positive numbers δ_z and ε_z such that under policy z for each initial state $x \in A_z$ the probability that the time until the next return to A_z exceeds δ_z is at least ε_z.*

This assumption implies that, with probability 1, the number of interventions is finite in any finite time interval.

We now introduce a Markov chain embedded in the decision process. Given that at epoch 0 the system is in state $x \in A_z$, let $\underline{I}_0 = x$, and for $n \geq 1$, let $\underline{I}_n$ be the state of the n-th entry of the decision process into the set A_z. Using property 2, it can be shown that $\{\underline{I}_n, n \geq 0\}$ is a discrete-time Markov process with state space A_z, cf. part II of De Leve (1964). Denote by

$$p^k(x,A,z) = \Pr\{\underline{I}_k \in A \mid \underline{I}_0 = x\} \qquad k \geq 0,$$

the k-step transition probability function of the Markov chain $\{\underline{I}_n, n \geq 0\}$. We write $p^1(x,A,z) = p(x,A,z)$.

ASSUMPTION 2. *For any policy* $z \in Z$ *there is some state* s_z *(say) such that*

$$\Pr\{\underline{I}_n = s_z \text{ for some } n \geq 1 \mid \underline{I}_0 = x\} = 1 \qquad \textit{for all } x \in A_z$$

and

$$E(\underline{N} \mid \underline{I}_0 = s_z) < \infty \textit{ where } \underline{N} = \inf\{n \geq 1 \mid \underline{I}_n = s_z\}.$$

Now, by Theorem A.1 in the appendix, the Markov chain $\{\underline{I}_n\}$ has a unique stationary probability distribution $Q(\cdot,z)$ (say) where

$$Q(A,z) = \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n} p^k(x,A,z) \qquad \text{for all } x \text{ and } A. \tag{1.1}$$

and

$$Q(A,z) = \int_{A_z} p(y,A,z)Q(dy,z) \qquad \text{for all } A. \tag{1.2}$$

ASSUMPTION 3. *For any policy* $z \in Z$,

(a) $$\int_{A_z} k_0(x)Q(dx,z) < \infty \textit{ and } \int_{A_z} t_0(x)Q(dx,z) < \infty$$

(b) *For each initial state* $x \in X$ *holds that under policy* z *both the time until the first return of the decision process to state* s_z *and the cost incurred during this time have a finite expectation.*

Now, let $\underline{Z}(t)$ be the total cost incurred during $[0,t)$.

THEOREM 1.1. *For each initial state,* $\underline{Z}(t)/t$ *converges for* $t \to \infty$ *both in expectation and with probability* 1 *to*

$$g(z) = \int_{A_z} k(x;z(x))Q(dx,z) \Big/ \int_{A_z} t(x;z(x))Q(dx,z). \text{ *)} \tag{1.3}$$

---

*) By relation (1.9) below, $\int t(x;z(x))Q(dx,z) > 0$ for all $z \in Z$.

PROOF. Let $\underline{T}_0=0$, and, for $n \geq 1$, let $\underline{T}_n$ be the epoch of the nth entry of the decision process into the set A_z. For any $n \geq 0$, let $\underline{K}_n$ be the cost incurred between the epochs $\underline{T}_n$ and $\underline{T}_{n+1}$, where $\underline{K}_n$ includes the immediate decision cost incurred at epoch $\underline{T}_n$ but not the immediate decision cost incurred at epoch $\underline{T}_{n+1}$, and $\underline{K}_n$ includes any cost of entering A_z at epoch $\underline{T}_{n+1}$ but not any cost of entering A_z at epoch $\underline{T}_n$ (cf. property 5). Let $\tau(x,z) = E(\underline{T}_{n+1}-\underline{T}_n \mid \underline{I}_n=x)$, and let $\kappa(x,z) = E(\underline{K}_n \mid \underline{I}_n=x)$ for $x \in A_z$.

Consider first the case where the initial state is s_z. Following the proof of Theorem 7.5 in Ross (1970) and using the assumptions 1–2 and 3(b), we get

$$(1.4) \qquad \lim_{t\to\infty} E\underline{Z}(t)/t = \lim_{n\to\infty} n^{-1}\sum_{i=0}^{n} E\underline{K}_i / \lim_{n\to\infty} n^{-1}\sum_{i=0}^{n} E(\underline{T}_{i+1}-\underline{T}_i),$$

where both the numerator and the denumerator of the right side of (1.4) are finite. Now, using the non-negativity of $\kappa(\cdot,z)$ and $\tau(\cdot,z)$, relation (1.1) and Proposition 17 on p.231 in Royden (1968), we get

$$(1.5) \qquad \infty > \lim_{n\to\infty} n^{-1}\sum_{i=0}^{n} E\underline{K}_i \geq \int_{A_z} \kappa(x,z)Q(dx,z),$$

$$(1.6) \qquad \infty > \lim_{n\to\infty} n^{-1}\sum_{i=0}^{n} E(\underline{T}_{i+1} - \underline{T}_i) \geq \int_{A_z} \tau(x,z)Q(dx,z).$$

By (1.4)-(1.6) and relation (A.7) in the appendix, we next get

$$(1.7) \qquad \lim_{t\to\infty} E\underline{Z}(t)/t = \int_{A_z} \kappa(x,z)Q(dx,z)/\int_{A_z} \tau(x,z)Q(dx,z)$$

We shall now prove that

$$(1.8) \qquad \int_{A_z} \kappa(x,z)Q(dx,z) = \int_{A_z} k(x;z(x))Q(dx,z),$$

$$(1.9) \qquad \int_{A_z} \tau(x,z)Q(dx,z) = \int_{A_z} t(x;z(x))Q(dx,z).$$

In property 6 we have introduced the sets A_{01} and A_{02} and the functions k_0, k_1, t_0 and t_1. By $A_z \supseteq A_0$ we have $A_z \supseteq A_{0i}$ for i = 1,2. Using this and the definitions of the functions k_0, k_1, t_0, t_1, κ and τ, it is easy to see

$$(1.10) \qquad k_1(x;z(x)) = \kappa(x,z) + \int_{A_z} k_0(y)p(x,dy,z) \qquad \text{for all } x \in A_z,$$

$$(1.11) \qquad t_1(x;z(x)) = \tau(x,z) + \int_{A_z} t_0(y)p(x,dy,z) \qquad \text{for all } x \in A_z.$$

Now integrate both sides of each of the relations (1.10) and (1.11) with respect to $Q(\cdot,z)$. Using the non-negativity of the functions involved, assumption 3(a) and the relations (1.2), (1.5) and (1.6), we get after an interchange of the order of integration (1.8) and (1.9). By (1.7) - (1.9), $E\underline{Z}(t)/t$ converges to g(z) as $t \to \infty$. However, using assumption 3(b), it is easy to verify that $\lim_{t\to\infty} E\underline{Z}(t)/t$ is independent of the initial state. Moreover, by Theorem 3.16 in Ross (1970), we have for each initial state, with probability 1, $\lim_{t\to\infty}\underline{Z}(t)/t$ equals $\lim_{t\to\infty}E\underline{Z}(t)/t$. This ends the proof.

The quantity g(z) represents the *long-run average (expected) cost per unit time* when policy z is used. This quantity is independent of the initial state. A policy $z^* \in Z$ is called *optimal* when $g(z^*) \le g(z)$ for all $z \in Z$.

The average cost g(z) can also be found by solving a system of functional equations. In solving this system we obtain in addition a function that will be used to improve policy z.

ASSUMPTION 4. *For any policy* $z \in Z$ *there is a finite number* α_z *such that*

$$(1.12) \qquad E(\underline{N} \mid \underline{I}_0 = x) \le \alpha_z \textit{ for all } x \in A_z \textit{ where } \underline{N} = \inf\{n \ge 1 \mid \underline{I}_n = s_z\}.$$

ASSUMPTION 5. *For any policy* $z \in Z$ *there is a finite number* β_z *such that the functions* $k(x;z(x))$ *and* $t(x;z(x))$ *are bounded by* β_z *for* $x \in A_z$.

Consider now the following system of functional equations,

$$v(x) = k(x;z(x)) - gt(x;z(x)) + E\{v(\underline{I}_1)|\underline{I}_0 = x\}, \qquad x\in A_z. \tag{1.13}$$

For any bounded solution $\{g,v(x) \mid x\in A_z\}$ to (1.13), define

$$v(x) = Ev(\underline{S}[x,A_z]) \qquad \text{for } x\notin A_z. \tag{1.14}$$

where $\underline{S}[x,A]$ is defined on p.6.

THEOREM 1.2. (a) *Let* $g = g(z)$, *and, for* $x \in A_z$, *let*

$$v(x) = \sum_{n=0}^{\infty} \int_{A_z} \{k(y;z(y)) - gt(y;z(y))\}\, \hat{p}^n(x,dy,z), \tag{1.15}$$

where $\hat{p}^0(x,A,z) = 1$ *for* $x \in A$, $\hat{p}^0(x,A,z) = 0$ *for* $x \notin A$, *and*

$$\hat{p}^n(x,A,z) = \Pr\{\underline{I}_n \in A,\ \underline{I}_k \neq s_z \text{ for } 1 \le k \le n \mid \underline{I}_0 = x\} \qquad \text{for } n\geq 1. \tag{1.16}$$

Then $v(s_z) = 0$ *and* $\{g,v(x) \mid x\in A_z\}$ *is a bounded solution to* (1.13).
(b) *For any bounded solution* $\{g,v(x)\}$ *to* (1.13) *holds* $g = g(z)$.
(c) *For any two bounded solutions* $\{g,v_1(x)\}$ *and* $\{g,v_2(x)\}$ *to* (1.13) *there is a constant* c *such that* $v_1(x) - v_2(x) = c$ *for all* $x \in A_z$.
(d) *Let* y *be an arbitrary state in* X, *then together* (1.13) *and* (1.14) *have a unique bounded solution with* $v(y) = 0$.

PROOF. (a) We first observe that, by $E\underline{N} = \sum_0^\infty \Pr\{\underline{N} > n\}$,

$$E(\underline{N} \mid \underline{I}_0 = x) = \sum_{n=0}^{\infty} \hat{p}^n(x,A_z,z) \qquad \text{for } x\in A_z. \tag{1.17}$$

By the assumptions 4-5 and relation (1.17), the function $v(x)$ is bounded. From (A.3) we have $Q(A,z) = \sum_0^\infty \hat{p}^n(s_z,A,z)/E(\underline{N} \mid \underline{I}_0 = s_z)$. Together this and (1.3) imply that $v(s_z) = 0$. Using $v(s_z) = 0$, assumption 5 and the relation

$$\hat{p}^n(x,A,z) = \int_{A_z\setminus\{s_z\}} \hat{p}^{n-1}(y,A,z)\,p(x,dy,z) \qquad \text{for } n\geq 1. \tag{1.18}$$

we next find that $\{g,v(x)\}$ satisfies (1.13).

(b) Integrating both sides of (1.13) with respect to $Q(\cdot,z)$, and using the relations (1.2) and 1.3), we get (b).

(c) Using part (b), we have $v_1(x) - v_2(x) = \int\{v_1(y) - v_2(y)\}p(x,dy,z)$ for $x \in A_z$. Iterate this equality n times and average over n. Letting $n \to \infty$ and using (1.1), we get $v_1(x) - v_2(x) = \int\{v_1(y) - v_2(y)\}\ Q(dy,z)$ for $x \in A_z$ which proves (c).

(d) This assertion follows easily from (a)-(c).

REMARK 1.2. When we replace in Theorem 1.2 bounded by finite, the assumptions 4 and 5 can be weakened somewhat (cf. Derman and Veinott (1967)). To avoid overburdening the text, Theorem 1.2 is not stated with maximum generality.

REMARK 1.3. The following relation may be useful in solving (1.13) and (1.14),

$$v(x) = k(x;z(x)) - gt(x;z(x)) + Ev(\underline{T}_{x,z(x)}) \qquad \text{for } x \in A_z. \tag{1.19}$$

This relation follows from (1.13), (1.14) and the fact that the intervention z(x) in state x causes an instantaneous transition to the state $\underline{T}_{x,z(x)}$ outside A_z, see property 7.

REMARK 1.4. In fact we need only to solve (1.13) in order to obtain a solution to (1.13) and (1.14). *) The dimension of the system of equations (1.13) is equal to the dimension of the embedded set of A_z. However the dimension of A_z is determined by the choice of the natural process because an intervention is a decision which interrupts the natural process. Therefore, to keep the number of equations to be solved as small as possible it may be advantageous to make "obvious optimal decisions" part of the natural process.

---

*) Sometimes (1.13) can be solved by solving a system of equations on a set which in its turn is embedded in A_z. Embedding may be very useful to reduce the number of equations to be actually solved.

1.3. *Basic tools.*

This section gives the basic tools for the solution techniques. Now, fix a policy $z_1 \in Z$ and a bounded solution $\{g(z_1), v(z_1;x) \mid x \in X\}$ to (1.13) and (1.14) with $z = z_1$. To introduce the tools for improving policy z_1, define for any $x \in X$ and $d \in D(x)$,

$$(1.20) \qquad v(d.z_1;x) = \begin{cases} v(z_1;x) & \text{for } d = \text{null-decision}, \\ k(x;d) - g(z_1)t(x;d) + Ev(z_1;\underline{T}_{x,d}) & \text{otherwise.} \end{cases}$$

Furthermore, for any policy $z \in Z$, let

$$(1.21) \qquad v([z]z_1;x) = \begin{cases} v(z(x).z_1;x) & \text{for } x \in A_z, \\ Ev([z]z_1;\underline{S}[x,A_z]) & \text{for } x \notin A_z. \end{cases}$$

REMARK 1.5. In this remark we give a number of obvious relations that will be frequently used in the sequel. By (1.19) and (1.20),

$$(1.22) \qquad v(z_1(x).z_1;x) = v(z_1;x) \qquad \text{for all } x \in X$$

Distinguishing between $x \in A_z$ and $x \notin A_z$, we get from (1.22), (1.21) and (1.14),

$$(1.23) \qquad v([z_1]z_1;x) = v(z_1;x) \qquad \text{for all } x \in X.$$

Further, let V be any closed set with $V \supseteq A_{z_1}$. Then using (1.14), the strong Markov property of the natural process and the theorem of conditional expectation

$$(1.24) \qquad v(z_1;x) = Ev(z_1;\underline{S}[x,V]) \qquad \text{for all } x \in X.$$

Similarly, let W be any closed set with $W \supseteq A_z$. Then by (1.21),

$$(1.25) \qquad v([z]z_1;x) = Ev([z]z_1; \underline{S}[x,W]) \qquad \text{for all } x \in X.$$

We now prove the following main theorem.

THEOREM 1.3. *Let policy* $z \in Z$ *be such that* $v([z]z_1;x) \le v(z_1;x)$ *for all* $x \in X_0$. *Then* $g(z) \le g(z_1)$. *This assertion remains true when the inequality signs are reversed.*

PROOF. Since $v([z]z_1;x) \le v(z_1;x)$ for $x \in X_0$, it follows from (1.25) with $W = X_0$ and (1.24) with $V = X_0$ that

$$v([z]z_1;x) = Ev([z]z_1;\ \underline{S}[x,X_0]) \le Ev(z_1;\ \underline{S}[x,X_0]) = v(z_1;x), \qquad x \in X.$$

Hence, for all $x \in A_z$,

$$(1.26) \qquad k(x;z(x)) - g(z_1)t(x;z(x)) + Ev([z]z_1;\ \underline{T}_{x,z(x)}) \le$$

$$\le k(x;z(x)) - g(z_1)t(x;z(x)) + Ev(z_1;\underline{T}_{x,z(x)}).$$

By (1.20) and (1.21), the right side of (1.26) equals $v([z]z_1;x)$. Since, with probability 1, the state $\underline{T}_{x,z(x)}$ does not belong to A_z, we have by (1.21) that

$$(1.27) \qquad Ev([z]z_1;\ \underline{T}_{x,z(x)}) = \int_{A_z} v([z]z_1;y)p(x,dy,z) \qquad \text{for } x \in A_z.$$

Hence, by (1.26), for all $x \in A_z$,

$$(1.28) \qquad k(x;z(x)) - g(z_1)t(x;z(x)) + \int_{A_z} v([z]z_1;y)p(x,dy,z) \le$$

$$\le v([z]z_1;x).$$

Now, integrate both sides of (1.28) with respect to $Q(\cdot,z)$. Using the boundedness of the functions k, t and v, and using the relations (1.2) and (1.3) we get after an interchange of the order of integration $g(z) \le g(z_1)$. Clearly, this proof carries over when the inequality signs are reversed.

Theorem 1.3 has the following corollary.

THEOREM 1.4. *If* $v(z_1;x) = \min_{z \in Z} v([z]z_1;x)$ *for all* $x \in X_0$, *then policy* z_1 *is optimal.*

This theorem provides us a *direct approach* for determining an optimal policy. We refer to De Leve, Tijms and Weeda (1970) for an application. However, in most cases an optimal policy can only be found by an iterative approach. We shall now discuss a policy iteration method. When we want to improve policy z_1 Theorem 1.4 suggests to look for a policy $z_2 \in Z$ satisfying

$$(1.29) \qquad v([z_2]z_1;x) = \min_{z \in Z} v([z]z_1;x) \qquad \text{for all } x \in X_0.$$

Observe that, by (1.23) and Theorem 1.3, $g(z_2) \leq g(z_1)$ when z_2 satisfies (1.29). Also, by Theorem 1.4, z_2 is optimal when $z_2 = z_1$.

We shall now prove that a policy z_2 satisfying (1.29) can be found by two operations. The first one is the *policy improvement operation* which adds to each state $x \in X_0$ a decision $d \in D(x)$ for which $v(d.z_1;x)$ is minimal, where $d = z_1(x)$ is chosen when this decision minimizes $v(d.z_1;x)$. In this way we obtain a policy z_1'. It is assumed that $z_1' \in Z$. By the construction of z_1' and the fact that $v(d.z_1;x)$ assumes the same value for both $d = z_1(x)$ and d = null-decision (see (1.20) and (1.22)), we have

$$(1.30) \qquad A_{z_1'} \supseteq A_{z_1}.$$

Although $g(z_1') \leq g(z_1)$ (see Theorem 1.5 below) it will be clear from (1.30) that we need a second operation which determines a policy whose set of intervention states is contained in $A_{z_1'}$. To formulate this *cutting operation* define for any policy $z \in Z$ and closed set $A \supseteq A_0$,

$$(1.31) \qquad v(A.[z]z_1;x) = Ev([z]z_1;\ \underline{S}[x,A]) \qquad \text{for } x \in X.$$

REMARK 1.6. It may be helpful to interprete $v(A.[z]z_1;x)$ as the expected stopping cost for initial state x when the natural process is stopped at the states of the set A and there is a cost of $v([z]z_1;y)$ for stopping at state y.

REMARK 1.7. In this remark we give some obvious relations that will be needed in the sequel. By definition (1.31), (1.32) $\quad v(A.[z]z_1;x) = v([z]z_1;x) \qquad$ for all $x \in A$,

and, by (1.21) and (1.32),

(1.33) $\quad v(A_z.[z]z_1;x) = v([z]z_1;x) \qquad$ for all $x \in X$.

Further, let V be any closed set with $V \supseteq A$. Then, by (1.31) and the theorem of conditional expectation,

(1.34) $\quad v(A.[z]z_1;x) = Ev(A.[z]z_1;\ \underline{S}[x,V]) \qquad$ for all $x \in X$.

THEOREM 1.5. *Let* $z \in Z$ *be such that* $A_z \supseteq A_{z_1}$ *and* $v([z]z_1;x) \leq v(z_1;x)$ *for all* $x \in A_z$. *Let* A *be any closed set with* $A_0 \subseteq A \subseteq A_z$ *such that* $v(A.[z]z_1;x) \leq$ $\leq v([z]z_1;x)$ *for all* $x \in A_z$. *Suppose that policy* $z_A \in Z$ *where* $z_A(x) = z(x)$ *for* $x \in A$, *and* $z_A(x) =$ *null-decision, otherwise. Then,* $g(z_A) \leq g(z_1)$.

Observe that, by taking $A = A_z$ and $z = z_1'$ in Theorem 1.5 and using (1.33), we get $g(z_1') \leq g(z_1)$. Next we introduce

ASSUMPTION 6. *There is a closed set* A' *with* $A_0 \subseteq A' \subseteq X_0$ *such that, for all* $x \in X_0$, $v(A'.[z_1']z_1;x) \leq v(B.[z_1']z_1;x)$ *for any closed set* B *with* $A_0 \subseteq B \subseteq X_0$.

We note that, by (1.34) with $V = X_0$, it is easy to verify that under assumption 6 the inequality in this assumption holds for all $x \in X$. The set A' can be interpreted as an optimal stopping set for the optimal stopping of the natural process when this process must be stopped at the states of A_0, may be stopped at the states of X_0, and must be continued at the states of $X \backslash X_0$ and there is a cost of $v([z_1']z_1;x)$ for stopping at state x. By a well known result in the theory of optimal stopping (cf. chapter 7 in Derman (1970)), assumption 6 holds when $X_0 \backslash A_0$ is finite.

In section 1.5 we prove

LEMMA 1.1. *Let* A' *be as in assumption* 6. *Suppose that* $A' \subseteq A_{z_1'}$ *and that policy* $z_2 \in Z$, *where* $z_2(x) = z_1'(x)$ *for* $x \in A'$, *and* $z_2(x) =$ *null-decision, otherwise. Then, policy* z_2 *satisfies* (1.29) *and* $g(z_2) \leq g(z_1)$.

We shall now construct a set A' satisfying the conditions of Lemma 1.1.

ASSUMPTION 7. *For any closed set* A *with* $A_0 \subseteq A \subseteq X_0$ *holds that for each initial state, with probability* 1*, the number of times where the natural process enters* A *before it enters* A_0 *is finite.*

In section 1.5 we prove

LEMMA 1.2. *Let* u(x) *be a bounded function on* X. *Let* A_1 *and* A_2 *be closed sets with* $A_0 \subseteq A_i \subseteq X_0$ *for* i=1,2. *For* i=1,2 *and* $x \in X$, *let* $v_i(x) = Eu(\underline{S}[x,A_i])$, *and let* $v(x) = Eu(\underline{S}[x,A_1 \cap A_2])$. *Suppose that* $v_i(x) \leq u(x)$ *for* i = 1,2 *and all* $x \in X_0$. *Then* $v(x) \leq v_i(x)$ *for* i = 1,2 *and all* $x \in X$.

Next define R as the class of all closed sets A' satisfying the conditions of assumption 6. Also, let K be the class of all closed sets A such that $A_0 \subseteq A \subseteq X_0$ and $v(A.[z_1']z_1;x) \leq v([z_1']z_1;x)$ for all $x \in X_0$. By (1.33) with $z = z_1'$,

(1.35) $$A_{z_1'} \in K \text{ and } R \subseteq K.$$

Taking $u(x) = v([z_1']z_1;x)$ in Lemma 1.2 and using $R \subseteq K$, we easily get

(1.36) $$A_1 \cap A_2 \in R \text{ when } A_1 \in R \text{ and } A_2 \in K.$$

ASSUMPTION 8. *The intersection* A^* *of all sets in* R *belongs to* R.

Observe that, by (1.35) and (1.36), this assumption holds when R is finite. Now define $A'_{z_1'}$ as the intersection of all sets in K. This set is further characterized in the next theorem where it is proved that this set is the desired set.

THEOREM 1.6. (a) $A'_{z_1'} \subseteq A_{z_1'}$ *and* $A'_{z_1'} = A^*$.
(b) $A'_{z_1'}$ *is the smallest closed set* A *with* $A_0 \subseteq A \subseteq A_{z_1'}$ *such that, for any closed set* B *with* $A_0 \subseteq B \subseteq A_{z_1'}$,

(1.37) $\qquad v(A.[z_1']z_1;x) \le v(B.[z_1']z_1;x)$ $\qquad$ *for all* $x \in A_{z_1'}$.

(c) $A_{z_1'}'$ *is the smallest closed set* A *with* $A_0 \subseteq A \subseteq A_{z_1'}$ *such that*

(1.38) $\qquad v(A.[z_1']z_1;x) \le v([z_1']z_1;x)$ $\qquad$ *for all* $x \in A_{z_1'}$.

PROOF. (a) By (1.35), $A_{z_1'} \in K$ and $A^* \in K$, so, $A_{z_1'}' \subseteq A_{z_1'}$ and $A_{z_1'}' \subseteq A^*$. Let $B \in K$. Then, by (1.36), $A^* \cap B \in K$, so, $B \supseteq A^*$. Hence $A_{z_1'}' \supseteq A^*$ which proves (a).
(b) By the second assertion of (a) and the assumptions 6-8, we have $A = A_{z_1'}$ satisfies (1.37). Now, let A be any closed set satisfying the conditions of (b). Taking $B = A_{z_1'}$ in (1.37) and using (1.33), yields $v(A.[z_1']z_1;x) \le$ $\le v([z_1']z_1;x)$ for all $x \in A_{z_1'}$. Next, by (1.34) and (1.25) with $V = W = A_{z_1'}$, this equality holds for all $x \in X$. Hence $A \in K$ which proves (b).
(c) The proof of (c) is very similar to that of (b).

THEOREM 1.7. *Define* $z_2(x) = z_1'(x)$ *for* $x \in A_{z_1'}'$, *and* $z_2(x)$ = *null-decision, otherwise. Suppose that policy* $z_2 \in Z$. *Then* z_2 *satisfies* (1.29) *and* $g(z_2) \le g(z_1)$. *Policy* z_2 *is optimal when* $z_2 = z_1$.

PROOF. The theorem follows from Theorem 1.6(a), assumption 8 and lemma 1.1.

It is important to note that Theorem 1.6(b) states that $A_{z_1'}'$ is the smallest optimal stopping set for the optimal stopping of the natural process when this process must be stopped at the states of A_0, may be stopped at the states of $A_{z_1'}$, and must be continued at the states of $X \setminus A_{z_1'}$, and there is a cost of $v(z_1'(x).z_1;x)$ for stopping at state x. Since in general practical problems have a special structure and this structure is reflected in the natural process, in most applications the determination of the above optimal stopping set turns out to be rather simple.

1.4. *The solution techniques.*

This section summarizes the solution techniques. In solving any particular problem, we first have to specify the properties 1-6 for this problem.

Next an optimal policy may be obtained by a direct approach or an iterative one. In most cases the iterative approach must be used.

DIRECT APPROACH (cf. Theorem 1.3)

Determine a policy $z^* \in Z$ such that $v(z^*;x) = \min_{z \in Z} v([z]z^*;x)$ for all $x \in X_0$.

ITERATIVE APPROACH (*the policy iteration method*).

Let z_n be the policy obtained at the end of the (n-1)-th iteration step (the first step is started with an arbitrary policy z_1). The n-th step of the policy iteration method proceeds as follows.

(a) *Value determination operation.* Determine a bounded solution $\{g(z_n), v(z_n;x)\}$ to (1.13) and (1.14) with $z = z_n$.

(b) *Policy improvement operation.* Construct policy z_n' by adding to each state $x \in X_0$ a decision $d \in D(x)$ for which

$$v(d.z_n;x) = k(x;d) - g(z_n)t(x;d) + Ev(z_n;\underline{T}_{x,d})$$

is minimal, where $z_n'(x) = z_n(x)$ is chosen when $z_n(x)$ is a minimizing decision.

(c) *Cutting operation (optimal stopping).* Determine the smallest optimal stopping set $A'_{z_n'}$ for the optimal stopping of the natural process when this process must be stopped at the states of A_0, may be stopped at the states of $A_{z_n'}$, and must be continued at the states of $X \setminus A_{z_n'}$, and there is a cost of $v(z_n'(x).z_n;x)$ for stopping at state x. Define policy z_{n+1} by

$$z_{n+1}(x) = \begin{cases} z_n'(x) & \text{for } x \in A'_{z_n'}, \\ \text{null-decision,} & \text{otherwise.} \end{cases}$$

This policy iteration method generates a sequence $\{z_n, n \geq 1\}$ of policies where it is assumed that $z_n, z_n' \in Z$ for all $n \geq 1$. By Theorem 1.7 we have $g(z_{n+1}) \leq g(z_n)$ for all $n \geq 1$. Also, policy z_k is optimal when

$z_{k+1} = z_k$. In section 1.6 we shall give conditions under which

$$\lim_{n\to\infty} g(z_n) = \inf_{z\in Z} g(z).$$

REMARK 1.8. In this remark we consider a *modified policy iteration method* where the policy improvement operation is not applied to all states and a stopping set is determined yielding a lower stopping cost than stopping immediately. By Theorem 1.5 a policy $f_1 \in F$ can also be improved to a policy $f_2 \in F$ as follows:

(a) Determine a bounded solution $\{g(f_1), v(f_1;x)\}$ to (1.13) and (1.14) with $z = f_1$.

(b) Construct a policy $\hat{f}_1 \in F$ such that $A_{\hat{f}_1} \supseteq A_{f_1}$ and $v(\hat{f}_1(x).f_1;x) \le v(f_1;x)$ for all $x \in A_{\hat{f}_1}$.

(c) Determine a set A with $A_0 \subseteq A \subseteq A_{\hat{f}_1}$ such that (cf. also remark 1.6)

$$v(A.[\hat{f}_1]f_1;x) \le v(\hat{f}_1(x).f_1;x) \qquad \text{for all } x\in A_{\hat{f}_1},$$

that is, the stopping of the natural process at the states of A is at least as good as the immediate stopping of the natural process when there is a cost of $v(\hat{f}_1(x).f_1;x)$ for stopping at state x. Also, the set A must be determined such that policy $f_2 \in Z$ where

$$f_2(x) = \begin{cases} \hat{f}_1(x) & \text{for all } x\in A, \\ \text{null-decision}, & \text{otherwise.} \end{cases}$$

This modified policy iteration method may be used to generate a sequence $\{f_n\}$ of structured policies, see section 2.2 for an example. Also, the modified policy iteration method is very often computationally more attractive than the above policy iteration method, see Weeda (1974a). Finally, we note that in case the sequence $\{f_n\}$ converges to a policy $f \in F$ this policy is optimal when a single iteration step of the above policy iteration method applied to f yields again f.

1.5. *Proofs of Theorem 1.5 and the Lemmas 1.1 and 1.2.*

Proof of Theorem 1.5. We first prove $v(A.[z]z_1;x) = v([z_A]z_1;x)$ for all $x \in X$. For $x \in A$ this equality follows immediately from the relations (1.32) and (1.21) and that the fact that $z_A(x)$ is an intervention which equals $z(x)$. For $x \notin A$ we have by (1.34) with $V = A$ and (1.21),

$$v(A.[z]z_1;x) = Ev(A.[z]z_1;\underline{S}[x,A]) = Ev([z_A]z_1;\underline{S}[x,A]) = v([z_A]z_1;x).$$

Next we prove $v(A.[z]z_1;x) \le v(z_1;x)$ for all $x \in X$. By the conditions of the Theorem, this inequality holds for $x \in A_z$. Since $A \subseteq A_z$ and $A_{z_1} \subseteq A_z$ it follows from (1.34) and (1.24) with $V = A_z$ that, for $x \notin A_z$,

$$v(A.[z]z_1;x) = Ev(A.[z]z_1;\underline{S}[x,A_z]) \le Ev(z_1;\underline{S}[x,A_z]) = v(z_1;x).$$

Together the above relations yield $v([z_A]z_1;x) \le v(z_1;x)$ for all $x \in X$, so, by Theorem 1.3, $g(z_A) \le g(z_1)$.

Proof of Lemma 1.1. By the first part of the proof of Theorem 1.5,

$$v(A'.[z_1']z_1;x) = v([z_2]z_1;x) \qquad \text{for all } x \in X. \tag{1.39}$$

We shall next prove that, for all $x \in X$,

$$v(z_1;x) \ge v([z_1']z_1;x). \tag{1.40}$$

Clearly, by (1.21) and the construction of z_1', this inequality holds for $x \in A_{z_1'}$. Next it follows from (1.24) with $V = A_{z_1'}$ and (1.21) that (1.40) holds for all $x \in X$. By the construction of z_1' we have $v(z(x).z_1';x) \ge$ $\ge v(z_1'(x).z_1;x)$ for all $x \in X$ and $z \in Z$. Distinguishing between $x \in A_{z_1'}$ and $x \notin A_{z_1'}$ it now follows from the latter inequality, (1.40) and the definitions (1.20) and (1.21) that, for any policy $z \in Z$,

$$v([z]z_1;x) \ge v([z_1']z_1;x) \qquad \text{for all } x \in A_z. \tag{1.41}$$

By (1.31), (1.41) and (1.21), for all $z \in Z$ and $x \in X$,

$$(1.42)\qquad v(A_z.[z_1']z_1;x) \le Ev([z]z_1;\underline{S}[x,A_z]) = v([z]z_1;x).$$

Assume now to the contrary that $v([z_0]z_1;x_0) < v([z_2]z_1;x_0)$ for some $z_0 \in Z$ and $x_0 \in X_0$. Together this inequality, (1.39) and (1.42) contradict the inequality in assumption 6. Hence z_2 satisfies (1.29). This implies $g(z_2) \le g(z_1)$ as observed below relation (1.29).

Proof of Lemma 1.2. For reasons of symmetry if suffices to prove $v_1(x) \ge v(x)$ for all $x \in X$. Clearly, this inequality holds with the equality sign for $x \in A_1 \cap A_2$. Let $P(B|x,A) = \Pr\{\underline{S}(x,A) \in B\}$. Now fix $x \in A_1^c$ where $A^c = X \setminus A$. Using the fact that $u(y) \ge v_2(y)$ for all $y \in A_1$, we get

$$\begin{aligned} v_1(x) &= \int_{A_2} u(y_1)P(dy_1|x,A_1) + \int_{A_2^c} u(y_1)P(dy_1|x,A_1) \\ &\ge \int_{A_2} u(y_1)P(dy_1|x,A_1) + \int_{A_2^c} P(dy_1|x,A_1)\{\int_{A_1} u(y_2)P(dy_2|y_1,A_2) + \\ &\qquad + \int_{A_1^c} u(y_2)P(dy_2|y_1,A_2)\}. \end{aligned}$$

Using the fact that $u(y) \ge v_1(y)$ for all $y \in A_2$, we next get

$$\begin{aligned} &\int_{A_2^c} P(dy_1|x,A_1) \int_{A_1^c} u(y_2)P(dy_2|y_1,A_2) \ge \\ &\ge \int_{A_2^c} P(dy_1|x,A_1) \int_{A_1^c} P(dy_2|y_1,A_2)\{\int_{A_2} u(y_3)P(dy_3|y_2,A_1) + \\ &\qquad + \int_{A_2^c} u(y_3)P(dy_3|y_2,A_1)\}. \end{aligned}$$

Continuing in this way yields for $n = 2,3,\ldots$

$$\begin{aligned} v_1(x) &\ge \int_{A_2} u(y_1)P(dy_1|x,A_1) + \textstyle\sum_{k=1}^{n-1} \int_{B_1^c} P(dy_1|x,B_0) \ldots \\ &\qquad \ldots \int_{B_k^c} P(dy_k|y_{k-1},B_{k-1}) \int_{B_{k+1}} u(y_{k+1})P(dy_{k+1}|y_k,B_k) + c_n, \end{aligned}$$

where

$$c_n = \int_{B_1^c} P(dy_1|x,B_0) \dots \int_{B_n^c} u(y_n)P(dy_n|y_{n-1},B_{n-1}),$$

$$B_{2k} = A_1, \text{ and } B_{2k+1} = A_2 \qquad \text{for } k=0,1,\dots .$$

By assumption 7 and the boundedness of $u(\cdot)$, $\lim_{n\to\infty} c_n = 0$. Further, for any set B,

$$P(B|x,A_1\cap A_2) = P(B\cap A_2|x,A_1) + \int_{A_2^c} P(dy_1|x,A_1)P(B\cap A_1|y_1,A_2) +$$
$$+ \int_{A_2^c} P(dy_1|x,A_1) \int_{A_1^c} P(dy_2|y_1,A_2)P(B\cap A_2|y_2,A_1) + \dots .$$

Using these relations we have $\int u(y)P(dy|x,A_1\cap A_2)$ equals the limit of the right side of the latter inequality as $n\to\infty$. Hence $v_1(x) \geq v(x)$ for all $x \in A_1^c$. For reasons of symmetry, $v_2(x) \geq v(x)$ for all $x \in A_2^c$. From this we get $v_1(x) = u(x) \geq v_2(x) \geq v(x)$ for all $x \in A_1 \setminus (A_1\cap A_2)$. The proof is now complete.

1.6. *Convergence results.*

This section gives conditions under which $\lim_{n\to\infty} g(z_n) = \inf_{z\in Z} g(z)$, where $\{z_n\}$ is the sequence of policies generated by the policy iteration method, see section 1.4. To our knowledge the only other paper dealing with the convergence of a policy iteration method for the case of a non-finite state space is that of Derman (1966) where different conditions for convergence are given.

ASSUMPTION 9. *There are finite numbers* α *and* β *such that* $\alpha_z \leq \alpha$ *and* $\beta_z \leq \beta$ *for all* $z \in Z$ *where* α_z *and* β_z *are introduced in the assumptions* 4 *and* 5.

THEOREM 1.8. *Suppose that* $\lim_{n\to\infty}\{v([z_{n+1}]z_n;x) - v(z_n;x)\} = 0$ *for all* $x \in X_0$. *Then* $g^* = \inf_{z\in Z} g(z)$.

PROOF. We first observe that, by Theorem 1.2(c) and (1.21), for any $n \geq 1$ the difference $v([z_{n+1}]z_n;x) - v(z_n;x)$ is independent of the particular solution $v(z_n;x)$. It follows from assumption 9 and (1.17) that, for any $z \in Z$, the particular solution (1.15) of (1.13) is bounded by $\alpha\beta\{1+g(z)\}$. Using this and the fact that the sequence $\{g(z_n)\}$ is non-increasing, we have $v([z_{n+1}]z_n;x) - v(z_n;x)$ is uniformly bounded in x and n. Next we observe that, by (1.24) and (1.25) with $V = W = X_0$ and the bounded convergence theorem, we get

$$\text{(1.43)} \qquad \lim_{n\to\infty}\{v([z_{n+1}]z_n;x) - v(z_n;x)\} = 0 \qquad \text{for all } x\in X.$$

Now, fix $z \in Z$. By theorem 1.7 and (1.29), $v([z]z_n;x) \geq v([z_{n+1}]z_n;x)$ for all $x \in X_0$ and $n \geq 1$. Using (1.25) with $W = X_0$, we have that this inequality holds for all $x \in X$. Hence, for all $x \in A_z$ and $n \geq 1$,

$$k(x;z(x)) - g(z_n)t(x;z(x)) + Ev([z]z_n;\underline{T}_{x,z(x)}) \geq$$

$$\geq k(x;z(x)) - g(z_n)t(x;z(x)) + Ev(z_n;\underline{T}_{x,z(x)}) + Ev([z_{n+1}]z_n;\underline{T}_{x,z(x)}) +$$

$$- Ev(z_n;\underline{T}_{x,z(x)}).$$

Using the definitions (1.20) and (1.21), we next get, for all $x \in A_z$ and $n \geq 1$,

$$k(x,z(x)) - g(z_n)t(x;z(x)) + \int_{A_z} v([z]z_n;y)p(x,dy,z) \geq$$

$$\geq v([z]z_n;x) + Ev([z_{n+1}]z_n;\underline{T}_{x,z(x)}) - Ev(z_n;\underline{T}_{x,z(x)}).$$

Integrate both sides of this inequality with respect to $Q(\cdot,z)$. Letting $n\to\infty$, and using the relations (1.2), (1.3) and (1.43) and the bounded convergence theorem, we get $g^* \geq g(z)$. This ends the proof.

We need the following lemma.

LEMMA 1.3. *Let* $\{v_n, n \geq 1\}$ *be a bounded sequence such that for any* $\varepsilon > 0$ *there is an integer* $N(\varepsilon)$ *such that* $v_{n+m} \leq v_n + \varepsilon$ *for all* $n,m \geq N(\varepsilon)$. *Then* $\lim_{n\to\infty} v_n$ *exists.*

PROOF. Let $v = \liminf_{n\to\infty} v_n$, and let $V = \limsup_{n\to\infty} v_n$. Choose $\varepsilon > 0$. Then, $V \leq v_n + \varepsilon$ for all $n \geq N(\varepsilon)$, so, $V \leq v + \varepsilon$ which proves the lemma since ε was arbitrary.

ASSUMPTION 10. *The state* s_z *introduced in assumption 2 is independent of* $z \in Z$ *and equals* s^* *(say). The set* $D(s^*)$ *consists of a single intervention.*

THEOREM 1.9. $g^* = \inf_{z \in Z} g(z)$.

PROOF. Using Theorem 1.2(c), it is easy to see that for any $n \geq 1$ the policy z_{n+1} is independent of the particular choice of the solution $v(z_n;x)$ of (1.13) with $z = z_n$. Hence it is no restriction to assume that $v(z_n;s^*) = 0$ for all $n \geq 1$, cf. Theorem 1.2(d). Since the sequence $\{g(z_n)\}$ is non-increasing, it now follows from assumption 9 and (1.14) - (1.17) that

$$(1.44) \qquad |v(z_n;x)| \leq \alpha\beta\{1 + g(z_1)\} \qquad \text{for all } x \in X \text{ and } n \geq 1.$$

We shall now prove

$$(1.45) \qquad \lim_{n\to\infty} v(z_n;x) \text{ exists and is finite for all } x \in X.$$

Suppose that (1.45) holds. Writing $v([z_{n+1}]z_n;x) - v(z_n;x) = v([z_{n+1}]z_n;x) + - v(z_{n+1};x) + v(z_{n+1};x) - v(z_n;x)$, and distinguishing between $x \in A_{z_{n+1}}$ and $x \notin A_{z_{n+1}}$, it easily follows from the relations (1.45), (1.21) - (1.19) and (1.14) and the bounded convergence theorem that

$$(1.46) \qquad \lim_{n\to\infty} \{v([z_{n+1}]z_n;x) - v(z_n;x)\} = 0 \qquad \text{for all } x \in X.$$

Hence, by Theorem 1.8, it suffices to prove (1.45). Now, fix n. By Theorem 1.7, (1.29) and (1.23), $v(z_n;x) \geq v([z_{n+1}]z_n;x)$ for all $x \in X_0$, so, by (1.24) and (1.25), (1.47) $$v(z_n;x) \geq v([z_{n+1}]z_n;x) \qquad \text{for all } x\in X.$$

Put for abbreviation

(1.48) $$a(x) = k(x,z_{n+1}(x)) - g(z_n)t(x;z_{n+1}(x)) \qquad \text{for } x\in A_{z_{n+1}}.$$

Since $z_{n+1}(s^*) = z_n(s^*)$ and $v(z_n;s^*) = 0$, it follows from (1.19) with $z = z_n$ that

(1.49) $$a(s^*) + Ev(z_n;\underline{T}_{s^*,z_{n+1}(s^*)}) = 0.$$

Fix $x \in A_{z_{n+1}}$. From (1.47), (1.21), (1.20), (1.16) and (1.49),

$$v([z_{n+1}]z_n;x) = a(x) + Ev(z_n;\underline{T}_{x,z_{n+1}(x)}) \geq a(x) + Ev([z_{n+1}]z_n;\underline{T}_{x,z_{n+1}(x)}) =$$

$$= a(x) + \int_{A_{z_{n+1}}} \{a(y) + Ev(z_n;\underline{T}_{y,z_{n+1}(y)})\}p(x,dy,z_{n+1}) =$$

$$= a(x) + \int_{A_{z_{n+1}}} \{a(y) + Ev(z_n;\underline{T}_{y,z_{n+1}(y)})\}\hat{p}^1(x,dy,z_{n+1}).$$

Continuing in this way and using repeatedly (1.47) and (1.49), we get, for $k \geq 1$,

$$v([z_{n+1}]z_n;x) \geq \sum_{j=0}^{k} \int_{A_{z_{n+1}}} a(y)\hat{p}^j(x,dy,z_{n+1}) +$$

$$+ \int_{A_{z_{n+1}}} v(z_n;\underline{T}_{y,z_{n+1}(y)})\hat{p}^k(x,dy,z_{n+1}).$$

Letting $k \to \infty$, and using (1.44), (1.17) and assumption 2, we get

(1.50) $$v([z_{n+1}]z_n;x) \geq \sum_{j=0}^{\infty} \int_{A_{z_{n+1}}} a(y)\hat{p}^j(x,dy,z_{n+1}) \qquad \text{for all } x\in A_{z_{n+1}}.$$

Let $\phi_n(x)$ be equal to the right side of (1.50), and let $\psi_n(x)$ be equal to the right side of (1.15) with $z = z_{n+1}$. Observe that ϕ_n and ψ_n are bounded.

Also, let $P_n(B|x) = \Pr\{\underline{S}[x,A_{z_{n+1}}] \in B\}$. Since $v([z_{n+1}]z_n;x) =$ $= Ev([z_{n+1}]z_n;\underline{S}[x,A_{z_{n+1}}])$ for all $x \in X$, we have by (1.47) and (1.50)

$$(1.51) \qquad v(z_n;x) \geq \int_{A_{z_{n+1}}} \phi_n(y)P_n(dy|x) \qquad \text{for all } x\in X.$$

Since $v(z_{n+1};s^*) = 0$ it follows from Theorem 1.2(a),(d) and (1.14) that

$$(1.52) \qquad v(z_{n+1};x) = \int_{A_{z_{n+1}}} \psi_n(y)P_n(dy|x) \qquad \text{for all } x\in X.$$

Let $\Delta_n = g(z_n) - g(z_{n+1})$. Observe that $\Delta_n \geq 0$. Now, by (1.52), (1.51), assumption 9 and (1.17), for any $n \geq 1$,

$$v(z_{n+1};x) - v(z_n;x) \leq \int_{A_{z_{n+1}}} \{\psi_n(y) - \phi_n(y)\}P_n(dy|x) =$$

$$= \Delta_n \int_{A_{z_{n+1}}} P_n(dy|x) \sum_{j=0}^{\infty} \int_{A_{z_{n+1}}} t(u;z_{n+1}(u))\hat{p}^j(y,du,z_{n+1}) \leq \Delta_n \alpha\beta$$

for all $x \in X$. Hence, for all $x \in X$,

$$v(z_{n+m};x) - v(z_n;x) \leq \{g(z_n) - g(z_{n+m})\}\alpha\beta \qquad \text{for all } n,m\geq 1.$$

Using (1.44), Lemma 1.3 and the fact that the sequence $\{g(z_n)\}$ is bounded and convergent, relation (1.45) now follows. This completes the proof.

REMARK 1.9. A convergence proof can also be given when there is some fixed state which is a regeneration state for each decision process and in which the null-decision is the only feasible decision.

For the case where X_0 is finite it is shown in Weeda (1974b) in all generality that the policy iteration method converges to an optimal policy after a finite number of steps.

REMARK 1.10. Consider the special case where $A_0 = X_0$ and consequently $A_z = X_0$ for all $z \in Z$. Then, using (1.45)-(1.46), (1.14) and the finiteness of $D(x)$, it follows from $v([z_{n+1}]z_n;x) = \min_d\{k(x,d)-g(z_n)t(x;d)+Ev(z_n;\underline{T}_{x,d})\}$

for all $x \in X_0$ and $n \geq 1$ that there exists a bounded function $v^*(x)$, $x \in X_0$ such that

$$(1.53) \qquad v^*(x) = \min_{d \in D(x)} \{k(x;d) - g^* t(x;d) + \int_{X_0} v^*(y) p(dy|x,d)\}, \qquad x \in X_0,$$

where $p(A|x,d)$ is the probability that the next intervention state belongs to the set A when in state x the intervention d is made. Moreover, by a standard argument as used in the proof of Theorem 1.3, any policy $z_0 \in Z$ such that $z_0(x)$ minimizes the right side of (1.53) for all $x \in X_0$ is optimal.

2. APPLICATIONS TO CONTROLLED QUEUEING SYSTEMS

2.1. *A switch-over policy for controlling the workload in the M|M|1 queue with switch-over costs*

2.1.0. *Introduction*

We consider a service station with a single server where jobs arrive in accordance with a Poisson process with rate λ. Each job involves an amount of work. The amounts of work of the jobs are known upon arrival and are independently sampled from an exponential distribution with mean $1/\mu$. At any time the server may choose between the service rates 1 and 2. When the server is in service and uses service rate i an amount of work σ_i will be processed per unit time, $i = 1,2$. It is assumed that $\sigma_2 > \sigma_1 > \lambda/\mu$. Define the workload at time t as the total amount of work remaining to be processed in the system at time t, $t \geq 0$. The server provides service when the system is not empty and uses the following switch-over policy. The server switches from rate 1 to rate 2 when the workload exceeds the level y_1 and switches from rate 2 to rate 1 when the workload falls to the level y_2, where y_1 and y_2 are given numbers with $0 < y_2 < y_1$. It is assumed that it takes no time to switch from one service rate to another.

The following costs are incurred. There is a holding cost of $h > 0$ per unit work in the system per unit time. When the server is busy and uses service rate i there is a service cost at rate $r_i \geq 0$, $i = 1,2$. There is a service cost at rate $r_0 \geq 0$ when the system is empty. The cost of switching from rate 1 to rate 2 is $K \geq 0$ (any cost of switching from rate 2 to rate 1 is assumed to be included in K).

Denote the above switch-over policy as the (y_1,y_2) policy. We shall determine an expression for the average cost of the (y_1,y_2) policy. Related work was done by Thatcher (1968) who derived by busy-period analysis an expression for the average cost of the (y_1,y_2) policy with $y_1 = y_2$ for the M|G|1 queue with no switch-over costs.

An expression for the average cost of the (y_1,y_2) policy will be derived from Theorem 1.1 of section 1.2. To do this, we first have to specify the properties 1-6 in section 1.1. This will be done in section 2.1.2.

Some preliminaries are given in section 2.1.1. Finally, in section 2.1.3 the embedded Markov process $\{\underline{I}_n\}$ corresponding to the (y_1,y_2) policy is studied and the expression for the average cost is given.

2.1.1. *Preliminaries*

Consider the M|G|1 queue where jobs arrive in accordance with a Poisson process with rate λ and the amounts of work involved by the jobs are independent, positive random variables having a common probability distribution function $F(x)$ with finite first moment β and finite second moment $\beta^{(2)}$. When the system is not empty the server provides service where an amount of work σ is processed per unit time. Assume that $\lambda\beta/\sigma < 1$. For any $t \geq 0$, let $\underline{v}(t)$ be the total amount of work remaining to be processed in the system at time t. In queueing theory the process $\{\underline{v}(t),\ t \geq 0\}$ is often called the virtual waiting time process. Further, let $\underline{\tau} = \inf\{t \geq 0 \mid \underline{v}(t) = 0\}$. We have the following lemma (cf. Theorem 4 in Thatcher (1968) and Tijms (1974)),

LEMMA 2.1. *For any* $x > 0$,

$$E(\underline{\tau} \mid \underline{v}(0) = x) = x/\sigma(1-\lambda\beta/\sigma) \tag{2.1}$$

and

$$E\left[\int_0^{\tau} \underline{v}(t)dt \mid \underline{v}(0) = x\right] = x^2/2\sigma(1-\lambda\beta/\sigma) + \lambda x\beta^{(2)}/2\sigma^2(1-\lambda\beta/\sigma)^2. \tag{2.2}$$

PROOF. For the completeness we prove the second relation. The first one is well-known. For any $x > 0$, let $\underline{n}_x$ be the number of arrivals in $(0,x/\sigma)$. Let $w = \int_0^\infty w(x)F(dx)$ where $w(x)$ is defined as the left side of (2.2). Then, by well-known properties of the Poisson process and (2.1),

$$E\left[\int_0^{\underline{\tau}} \underline{v}(t)dt \mid \underline{v}(0) = x,\ \underline{n}_x = n\right] =$$

$$= x^2/2\sigma + n\beta x/2\sigma + nw + \textstyle\sum_{k=1}^{n}(n-k)\beta\{\beta/\sigma(1-\lambda\beta/\sigma)\},$$

so, by unconditioning on $\underline{n}_x$, $w(x) = x^2/2\sigma + \lambda\beta x^2/2\sigma^2 + \lambda x w/\sigma + \lambda^2x^2\beta^2/\sigma^3(1-\lambda\beta/\sigma)$ for all $x > 0$. Together this and $w = \int_0^\infty w(x)F(dx)$ yield (2.2).

2.1.2. *The natural process, the decisions, and the k- and t-functions*

The natural process and the feasible decisions will be of course specified to measure the (y_1,y_2) policy. Before doing this, we make the following observations. The natural process and the intervention must be chosen in such a way that the result of the natural process and the control by the interventions agrees with the process describing the workload when the (y_1,y_2) policy is used. However, these choices determine the set A_0. In its turn the set A_0 is determinative for the calculation of the k- and t-functions. It will be obvious that we shall try to choose the natural process and the interventions in such a way that the resulting set A_0 allows for a simple calculation of the k- and t-functions. Clearly, a convenient choice for the natural process is one where the server never switches from one service rate to another. Given this choice it will also be clear that the states corresponding to the cases where the system becomes empty are attractive states to set up the sets A_{01} and A_{02}. However, the state corresponding to the situation where the system becomes empty while the server is adjusted to service rate 1 is not an intervention state for the (y_1,y_2) policy. Nevertheless, by a generally usuable trick, we can achieve that this state is an intervention state. We can define the natural process such that in the natural process the service station is closed down forever when the system becomes empty while the server is adjusted to rate 1. This has as a consequence that we also have to introduce both a fictitious intervention which re-opens the station immediately and a fictitious state to which the system is instantaneously transferred by this intervention. All this can be done provided that the result of the natural process and the control by the interventions agrees with the process describing the workload under the (y_1,y_2) policy, cf. remark 1.1 of section 1.1. These observations will be used in the specification of the properties 1-3.

We choose as state space

$$X = \{u \mid u \text{ real}, u \geq 0\} \cup \{u' \mid u \text{ real}, u \geq 0\} \cup \{\bar{0}\}.$$

State u(u') corresponds to the situation where the workload equals u and the server is adjusted to rate 1 (2). In addition, in state 0 the station is closed down. State $\bar{0}$ corresponds to the situation where the workload is zero, the station is open and the server is adjusted to rate 1.

The natural process is chosen such that in the natural process the server never switches from one service rate to another. For any initial state u' we define the natural process as the process describing the workload when always service rate 2 is used. For initial state $u > 0$ the natural process is defined as the process describing the workload under the use of service rate 1 until the system becomes empty. Then the natural process closes down the station and assumes state 0 which is an absorbing state for the natural process. When the initial state is $\bar{0}$ the natural process stays in this state until the next job arrives. Then the natural process assumes state y when this job involves an amount of work y.*)

Next we define the possible decisions. In each state either the null-decision or the intervention $d = 1$ is the only possible decision. Both in state u with $0 < u < y_1$, state u' with $u > y_2$ and in state $\bar{0}$ the null-decision is the only feasible decision. The null-decision does not disturb the natural process. In the other states the intervention $d = 1$ is the only possible decision. The intervention $d = 1$ in state u' with $0 \le u \le y_2$ prescribes to switch from rate 2 to rate 1 and causes an instantaneous transition to state u when $u > 0$ and to state $\bar{0}$ when $u = 0$. The intervention $d = 1$ in state u with $u \ge y_1$ prescribes to switch from rate 1 to rate 2 and causes an instantaneous transition to state u'. Finally the intervention $d = 1$ in state 0 prescribes to re-open the station and causes an instantaneous transition to state $\bar{0}$.

Now, it will be clear that the result of this natural process and the control by the above decisions agrees with the process describing the workload under the (y_1,y_2) policy. Also, by the above choices,

$$A_0 = \{0\} \cup \{u \mid u \ge y_1\} \cup \{u' \mid 0 \le u \le y_2\}.$$

Since $\sigma_2 > \sigma_1 > \lambda/\mu$ we have for each initial state that, with probability 1, the natural process will eventually reach one of the states 0 and 0', so the set A_0 satisfies the conditions of property 4 in section 1.1. The fact

*) The above discussion shows that the choice of the natural process requires some practice just as the choice of the contour in complex integration.

that the states 0 and 0' belong to A_0 has as a consequence that the task of calculating the k- and t-functions is an easy one. To do this, we choose

$$A_{01} = A_{02} = \{0\} \cup \{0'\}.$$

For $u \geq 0$ and $i = 1,2$, let

$$\tau_i(u) = \frac{u}{\sigma_i(1-\lambda/\mu\sigma_i)} \quad \text{and} \quad w_i(u) = \frac{u^2}{2\sigma_i(1-\lambda/\mu\sigma_i)} + \frac{\lambda u}{\mu^2\sigma_i^2(1-\lambda/\mu\sigma_i)^2}\ .$$

Now, using Lemma 2.1, it easily follows that (see property 6 in section 1.1)

$$k_0(u) = hw_1(u) + r_1\tau_1(u),\ t_0(u) = \tau_1(u) \qquad \text{for } u\geq 0,$$

$$k_0(u') = hw_2(u) + r_2\tau_2(u),\ t_0(u') = \tau_2(u) \qquad \text{for } u\geq 0,$$

$$k_0(\bar{0}) = r_0/\lambda + \int_0^\infty k_0(y)\mu e^{-\mu y}dy,\ t_0(\bar{0}) = 1/\lambda + \int_0^\infty t_0(y)\mu e^{-\mu y}dy,$$

$$k(u;1) = K + k_0(u') - k_0(u),\ t(u;1) = t_0(u') - t_0(u) \qquad \text{for } u\geq y_1,$$

$$k(u';1) = k_0(u) - k_0(u'),\ t(u';1) = t_0(u) - t_0(u') \qquad \text{for } 0<u\leq y_2,$$

$$k(0;1) = k_0(\bar{0}),\ t(0;1) = t_0(\bar{0})$$

2.1.4. *The average cost of the (y_1,y_2) policy*

Clearly, the class Z of policies consists of the single policy $z = (y_1,y_2)$. Also, $A_z = A_0$. It will be obvious that assumption 1 in section 1.2 is satisfied. Now we consider the embedded Markov chain $\{\underline{I}_n\}$, see section 1.2. This Markov chain describes the workload at the epochs at which the system enters the set A_z using the (y_1,y_2) policy. Since $\sigma_2 > \sigma_1 > \lambda/\mu$ it follows that assumption 2 holds when we take $s_z = 0$. Let $Q(\cdot)$ be the unique stationary probability distribution of $\{\underline{I}_n\}$. For ease of notation, put $Q_0 = Q(\{0\})$, $Q(v) = Q(\{u \mid u\geq v\})$ for $v\geq y_1$, and $Q_2 = Q(\{y_2'\})$. To determine these probabilities, define, for all $0<u<y_1$ and $v\geq y_1$,

$p(u,v)$ = probability that the state of the first entry of the natural process into the set $\{0\} \cup \{x \mid x\geq y_1\}$ belongs to the set $\{x \mid x\geq v\}$ given that the initial state is u.

Also, let $p_0(u) = 1 - p(u,y_1)$ for $0 < u < y_1$. Now, by (1.2) in section 1.2,

$$Q_0 = Q_2 p_0(y_2) + Q_0 \int_0^{y_1} p_0(y)\mu e^{-\mu y} dy, \quad Q_2 = Q(y_1), \quad \text{and}$$

$$Q(v) = Q_2 p(y_2,v) + Q_0\{e^{-\mu v} + \int_0^{y_1} p(y,v)\mu e^{-\mu y} dy\} \qquad \text{for all } v \geq y_1.$$

Considering what can happen in a very small time interval, we get for any $v \geq y_1$

$$\frac{\partial p(u,v)}{\partial u} = (\lambda/\sigma_1)[-p(u,v) + \int_0^{y_1-u} p(u+x,v)\mu e^{-\mu x} dx + e^{-\mu(v-u)}], \qquad 0<u<y_1.$$

Routine analysis involving the use of Laplace transforms yields after some algebra

$$p(u,v) = \lambda e^{-(\mu v - \lambda y_1/\sigma_1)} [e^{(\mu\sigma_1-\lambda)(u/\sigma_1)} - 1][\mu\sigma_1 - \lambda e^{-(\mu\sigma_1-\lambda)(y_1/\sigma_1)}]^{-1}$$

for all $0 < u < y_1$ and $v \geq y_1$. From $p_0(u) = 1 - p(u,y_1)$,

$$p_0(u) = [\mu\sigma_1 - \lambda e^{-(\mu\sigma_1-\lambda)(y_1-u)/\sigma_1}][\mu\sigma_1 - \lambda e^{-(\mu\sigma_1-\lambda)(y_1/\sigma_1)}]^{-1}, \qquad 0<u<y_1.$$

The formula for $p_0(u)$ was also found in Keilson (1963). Using these results we next get after some algebra

$$Q_0 = c^{-1}\{\mu\sigma_1 - \lambda e^{-(\mu\sigma_1-\lambda)(y_1-y_2)/\sigma_1}\}$$

$$Q_2 = c^{-1}(\mu\sigma_1-\lambda)e^{-(\mu\sigma_1-\lambda)(y_1/\sigma_1)}$$

$$q(v) = c^{-1}\mu(\mu\sigma_1-\lambda)e^{-(\mu v-\lambda y_1/\sigma_1)} \quad \text{for all } v \geq y_1, \text{ where } q(v) = -\partial Q(v)/\partial v,$$

with $c = \mu\sigma_1 + (2\mu\sigma_1 - 2\lambda)e^{-(\mu\sigma_1-\lambda)(y_1/\sigma_1)} - \lambda e^{-(\mu\sigma_1-\lambda)(y_1-y_2)/\sigma_1}$. It is easy to check that assumption 3 holds. Now, by Theorem 1.1, the average cost of the (y_1,y_2) policy is equal to

$$g(y_1,y_2) = \frac{k(0;1)Q_0 + \int_{y_1}^{\infty} k(u;1)q(u)du + k(y_2';1)Q_2}{t(0;1)Q_0 + \int_{y_1}^{\infty} t(u;1)q(u)du + t(y_2';1)Q_2}.$$

Each of the quantities in this expression has been explicitly determined. We further refrain from working out this expression. To end, we consider the special case where the switch-over cost $K=0$ and the y-policy is used. The y-policy switches from rate 1 to rate 2 when the workload exceeds the level y and switches from rate 2 to rate 1 when the workload falls to the level y. Putting $K=0$ and $y_1=y_2=y$ in the above expression for $g(y_1,y_2)$, we find after some algebra that the average cost of the y-policy is equal to

$$g(y) = [(\gamma_1+\gamma_2 y)e^{-(\mu\sigma_1-\lambda)(y/\sigma_1)} + \gamma_3][\delta_1 e^{-(\mu\sigma_1-\lambda)(y/\sigma_1)} + \delta_2]^{-1}$$

where

$$\gamma_1 = h(\alpha_2-\alpha_1)/\mu + h\lambda(\alpha_2^2-\alpha_1^2)/\mu + r_2\alpha_2 - r_1\alpha_1,$$

$$\gamma_2 = h(\alpha_2-\alpha_1),\quad \gamma_3 = r_0/\lambda + h\alpha_1/\mu + h\lambda\alpha_1^2/\mu + r_1\alpha_1,$$

$$\delta_1 = \alpha_2 - \alpha_1,\quad \delta_2 = 1/\lambda + \alpha_1,\ \text{and}\ \alpha_i = 1/(\mu\sigma_i-\lambda)\quad \text{for } i=1,2.$$

Putting the derivative $g'(y) = 0$ yields after some algebra

$$y + \lambda(\sigma_2-\sigma_1)(\mu\sigma_1-\lambda)^{-1}(\mu\sigma_2-\lambda)^{-1}\{e^{-(\mu\sigma_1-\lambda)(y/\sigma_1)} - 1\} = b \tag{2.3}$$

where $b = (h\mu\sigma_1)^{-1}(\mu\sigma_1-\lambda)\{r_0 + (r_2\sigma_1-r_1\sigma_2)/(\sigma_2-\sigma_1)\}$. By $\sigma_2>\sigma_1>\lambda/\mu$ the function $g'(y)$ is strictly increasing for $y\geq 0$ with $g'(y)\to\infty$ as $y\to\infty$. Using this it follows that for $b>0$ the function $g(y)$ is minimal for the unique positive y^* (say) satisfying (2.3). Moreover, using the fact that $1-e^{-x}\leq x$ for $x\geq 0$, it is easily derived from (2.3) that

$$y^* \leq (h\mu\sigma_2)^{-1}(\mu\sigma_2-\lambda)\{r_0 + (r_2\sigma_1-r_1\sigma_2)/(\sigma_2-\sigma_1)\}.$$

In case $b \leq 0$ the function $g(y)$ is minimal for $y = 0$. We note that (2.3) agrees with relation (2) on p.80 in Thatcher (1968).

Below we give the optimal y^* and $g(y^*)$ for a number of numerical examples with $\mu = 1$, $\sigma_1 = 4$, $h = 5$, $r_0 = 0$, $r_1 = 10$ and $r_2 = 15$.

	$\sigma_2 = 5$		$\sigma_2 = 4.5$		$\sigma_2 = 4.25$	
λ	y^*	$g(y^*)$	y^*	$g(y^*)$	y^*	$g(y^*)$
3	.759	16.297	1.874	19.370	3.872	21.361
3.25	.665	18.863	1.566	23.330	3.103	26.764
3.50	.572	22.027	1.260	28.800	2.342	35.044
3.75	.479	26.144	.954	37.268	1.580	50.402
3.90	.423	29.340	.768	45.341	1.117	69.302

2.2. *An algorithm for the switch-over policy controlling the queue size in the M|G|1 queue with two service types and switch-over costs.*

2.2.0. *Introduction*

Consider a single-server station where customers arrive in accordance with a Poisson process with rate λ. The service times of the customers are independent random variables. The server provides service when there are customers present where the server may choose between two service types. When the server uses service type 1 the service time of a customer is exponentially distributed with mean $1/\mu$, while for service type 2 the service time of a customer has a general distribution with finite first moment β and finite second moment $\beta^{(2)}$. It is assumed that $\lambda\beta < 1$. The server may switch from service type 1 to service type 2 both at arrival epochs and service completion epochs. The service of a customer starts anew when it is interrupted for switching to service type 2. When the server is adjusted to service type 2 it is only allowed to switch to service type 1 at service completion epochs. It takes no time to switch from one service type to another.

The following costs are considered. There is a holding cost at rate

$h \cdot i$ when i customers are in the system, where $h > 0$. A service cost at rate $r_i \geq 0$ applies when the server is busy and uses service type i (=1,2), while a service cost at rate r_0 applies when the system is empty. The cost of switching from service type 1 to service type 2 is $K \geq 0$ (it is assumed that any cost of switching from service type 2 to service type 1 is included in K).

In solving this problem we will only consider policies that always use service type 1 when the system is empty and always use service type 2 when at least N customers are present where N is some fixed positive integer. This seems no restriction in practical problems provided that N is chosen sufficiently large. We are interested in policies of the following form. The server switches from service type 1 to service type 2 when the number of customers present reaches from below the level i_1 and the server switches from service type 2 to service type 1 when the number of customers present falls to the level i_2, where i_1 and i_2 are given integers with $0 \leq i_2 < i_1 \leq N$. Such a policy will be called an (i_1,i_2) policy. We shall give two very simple finite algorithms which each generate a sequence of (i_1,i_2) policies where any policy of the sequence has a lower average cost than its predecessor. The algorithms do not involve the solving of any system of linear equations. We note that for the case where $K=0$ and the service time under service type 2 is also exponential it was shown in Crabill (1972) and Lippman (1973) that under general conditions an (i_1,i_2) policy with $i_1=i_2+1$ is average cost optimal among the class of all possible policies.

In section 2.2.1 we choose the state space, the natural process and the feasible decisions. Further, in this section we determine the k- and t-functions. The algorithms are given in section 2.2.2.

2.2.1. *The state space, the natural process, the decisions, and the k- and t-functions.*

We choose as state space

$$\underset{\sim}{X} = \{i \mid i = 0,1,\ldots,N\} \cup \{i' \mid i=0,1,\ldots\} \cup$$
$$\cup \{(0',0),(i',t) \mid i=1,2,\ldots \text{ and } t \geq 0\}.$$

State i corresponds to the situation where i customers are in the system and the server is adjusted to service type 1, while state i' corresponds to the situation where i customers are in the system and a service has been just completed by the use of service type 2. State (i',t) with $i \geq 1$ corresponds to the situation where i customers are present, the server is busy and uses service type 2, and the elapsed service time of the customer being served is t. State (0',0) corresponds to the situation where the system is empty and the server is adjusted to service type 2.

We choose the natural process in such a way that in the natural process the server never switches from service type 2 to service type 1 and switches from service type 1 to service type 2 only when a customer arrives who finds N other customers present. Now, for any initial state (i',t) the natural process is defined as the process describing jointly the number of customers present and the elapsed service time of the customer being served (if any) when service type 2 is always used. When the initial state is i' the natural process makes an instantaneous transition to state (i',0). For initial state i the natural process describes the number of customers present when service type 1 is used until the arrival of a customer who finds N other customers present. At the arrival epoch of this customer the natural process makes an instantaneous transition to state ((N+1)',0).

Next we choose the feasible decisions. The only two possible decisions are the null-decision and the intervention $d = 1$. The null-decision is the only possible decision in both the states (i',t), the states i' with $i \geq N$ and the state 0. In the states 0' and N we take the intervention $d = 1$ as the only possible decision. In the other states both the null-decision and the intervention $d = 1$ are possible. The null-decision does not disturb the natural process. The intervention $d = 1$ in state i prescribes to switch to service type 2 and causes an instantaneous transition to state (i',0), while the intervention $d = 1$ in state i' prescribes to switch to service type 1 and causes an instantaneous transition to state i.

$$A_0 = \{N\} \cup \{0'\}.$$

Observe that this set satisfies the conditions of property 4 in section 1.1.

Also, observe that, by the above definitions, there are only a finite number of states in which an intervention is possible. Using this it is easy to see that the assumptions 1-8 in chapter 1 are satisfied.

To determine the k- and t-functions, we choose

$$A_{01} = A_{02} = A_0.$$

Before determining these functions, we make the following observations:
(i) Consider the $M|G|1$ queue where the arrival rate is λ and the service time of a customer has finite first moment β and finite second moment $\beta^{(2)}$ with $\lambda\beta < 1$. It is well-known that the expected length of one busy period and the expected total amount of time spent by customers in the system during one busy period are given by

$$\beta/(1-\lambda\beta) \quad \text{and} \quad \beta/(1-\lambda\beta) + \lambda\beta^{(2)}/2(1-\lambda\beta)^2,$$

respectively.
(ii) Consider the $M|M|1$ queue where the arrival rate is λ and the service time of a customer has mean $1/\mu$. Assume that there is a holding cost of h per customer per unit time spent in the system and a service cost at rate r_1 (r_0) when the server is busy (idle). Given that at epoch 0 there are i customers present, let $\tau(i)$ be the expectation of the first epoch at which N customers are present, and let $c(i)$ be the expected holding and service cost incurred up to this epoch, $i = 0,\dots,N$. Now, from

$$\tau(i) = (\lambda+\mu)^{-1}\{1 + \lambda\tau(i+1) + \mu\tau(i-1)\} \qquad \text{for } 1 \le i < N,$$

$$c(i) = (\lambda+\mu)^{-1}\{r_1 + hi + \lambda c(i+1) + \mu c(i-1)\} \qquad \text{for } 1 \le i < N,$$

$$\tau(0) = 1/\lambda + \tau(1),\ c(0) = r_0/\lambda + c(1),\ \text{and } \tau(N) = c(N) = 0,$$

it is routine to derive that, for $0 \le i \le N$, (cf. pp.313-317 in Feller (1957))

$$\tau(i) = \begin{cases} (\lambda-\mu)^{-1}[N - i + \mu(\lambda-\mu)^{-1}\{(\mu/\lambda)^N - (\mu/\lambda)^i\}] & \text{if } \lambda \ne \mu, \\ (2\lambda)^{-1}[N(N+1) - i(i+1)] & \text{if } \lambda = \mu, \end{cases}$$

$$c(i) = (\lambda-\mu)^{-1}[h(N^2-i^2)/2 + \{r_1 - h(\lambda+\mu)/2(\lambda-\mu)\}\{N-i\} +$$

$$+ \{-r_0 + r_1\lambda/(\lambda-\mu) - h\lambda\mu/(\lambda-\mu)^2\}\{(\mu/\lambda)^N - (\mu/\lambda)^i\}] \quad \text{if } \lambda\neq\mu,$$

$$c(i) = \lambda^{-1}[h(N^3-i^3)/6 + r_1(N^2-i^2)/2 + (r_0-r_1/2-h/6)(N-i)] \quad \text{if } \lambda=\mu.$$

It is now easy to give the k- and t-functions, see property 6 of section 1.1. Using the results in (i) and (ii), it follows that

$$k_0(i) = c(i), \; t_0(i) = \tau(i) \quad \text{for } 0\le i\le N,$$

$$k_0(i') = hi\{\beta/(1-\lambda\beta) + \lambda\beta^{(2)}/2(1-\lambda\beta)^2\} + \{hi(i+1)/2 + r_2 i\}\beta/(1-\lambda\beta) \quad \text{for } i\ge 0,$$

$$t_0(i') = i\beta/(1-\lambda\beta) \quad \text{for } i\ge 0.$$

Hence, by $k(i;1) = K + k_0(i') - k_0(i)$, $t(i;1) = t_0(i') - t_0(i)$, $k(i';1) = k_0(i) - k_0(i')$, and $t(i';1) = t_0(i) - t_0(i)$, we find

$$k(i;1) = K + k(i), \quad t(i;1) = t(i) \quad \text{for } 1\le i\le N, \tag{2.4}$$

$$k(i';1) = -k(i), \quad t(i';1) = -t(i) \quad \text{for } 0\le i<N, \tag{2.5}$$

where, for $0 \le i \le N$,

$$k(i) = hi\{\beta/(1-\lambda\beta) + \lambda\beta^{(2)}/2(1-\lambda\beta)^2\} + \{hi(i+1)/2 + r_2 i\}\beta/(1-\lambda\beta) - c(i),$$

$$t(i) = i\beta/(1-\lambda\beta) - \tau(i).$$

2.2.2. *Algorithms*

First we shall specify the equations (1.13) and (1.14) for the (i_1,i_2) policy. Now, fix an (i_1,i_2) policy with $0 \le i_2 < i_1 \le N$. The set of intervention states for this policy is given by $\{i \mid i_1 \le i \le N\} \cup \{i' \mid 0 \le i \le i_2\}$. Clearly, the embedded Markov chain $\{\underline{I}_n\}$ corresponding to the (i_1,i_2) policy assumes alternatively the states i_1 and i_2. The equations (1.13) and (1.14) yield for the (i_1,i_2) policy,

$$v(i) = k(i;1) - gt(i;1) + v(i_2) \text{ for } i_1 \le i \le N, \quad v(i) = v(i_1) \text{ for } 0 \le i \le i_1,$$

$$v(i') = k(i';1) - gt(i';1) + v(i_1) \text{ for } 0 \le i \le i_2, \quad v(i) = v(i_2) \text{ for } i \ge i_2.$$

To obtain a unique solution to this system of equations, we put $v(i_2) = 0$ (cf. Theorem 1.2(d)). Using (2.4) and (2.5), we then find

$$(2.6) \qquad g(i_1,i_2) = \{K + k(i_1) - k(i_2)\}/\{t(i_1) - t(i_2)\},$$

$$v(i) = K + k(i) - g(i_1,i_2)t(i) \quad \text{for } i_1 \le i \le N, \quad v(i) = v(i_1) \text{ for } 0 \le i \le i_1,$$

$$v(i') = -k(i) + g(i_1,i_2)t(i) \text{ for } 0 \le i \le i_2, \quad v(i') = 0 \text{ for } i \ge i_2.$$

Observe that, by (2.6), we have found an explicit expression for the average cost of an (i_1,i_2) policy. Also, observe that this expression is independent of N.

Denote the (i_1,i_2) policy by f_1. We shall now give a procedure that derives from policy f_1 an improved policy $f_2 = (k_1,k_2)$ of the same structure. This procedure will be based on the modified policy iteration method outlined in remark 1.8 of section 1.4. First we determine a policy $\hat{f}_1 = (j_1,j_2)$ with $i_2 \le j_2 < j_1 \le i_1$ by applying the policy improvement operation in the following way. Let j_2 be the largest integer such that $i_2 < j_2 < i_1$ and $v(1.f_1;i') < v(i') = 0$ for all $i_2 < i \le j_2$ if such an integer exists, and let $j_2 = i_2$, otherwise. Then let j_1 be the smallest integer such that $j_2 < j_1 < i_1$ and $v(1.f_1;i) < v(i) = v(i_1)$ for all $j_1 \le i < i_1$ if such an integer exists, and let $j_1 = i_1$, otherwise. Next we determine a policy $f_2 = (k_1,k_2)$ with $0 \le k_2 \le j_2 < j_1 \le k_1 \le N$ such that the stopping of the natural proces at the states of the set $A = \{i \mid k_1 \le i \le N\} \cup \{i' \mid 0 \le i \le k_2\}$ yields a lower expected stopping cost than the immediate stopping of the natural process at the states of the set $\hat{A} = \{i \mid j_1 \le i \le N\} \cup \{i' \mid 0 \le i \le j_2\}$ when there is a cost of $v(1.f_1;x)$ for stopping at state $x \in \hat{A}$. To do this, we observe that for initial state $i' \neq 0'$ the state of the first entry of the natural process into the set $\{j' \mid j \le k\}$ with $0 \le k < i$ is the state k', while for initial state $i < N$ the state of the first entry of the natural process into the set $\{j \mid j \ge k\}$ with $i < k \le N$ is the state k. Now the policy

$f_2 = (k_1,k_2)$ is constructed as follows. Let k_2 be the largest value of i for which $v(1.f_1;i')$ is minimal on $[0,j_2]$, and let k_1 be the smallest value of i for which $v(1.f_1;i)$ is minimal on $[j_1,N]$.

So far we have not specified the quantities $v(1.f_1;i)$ and $v(1.f_1;i')$. Using the fact that $v((i',0)) = v(i_2) = 0$ for $i > i_2$ (see (1.14)), it follows from (1.21) that

$$v(1.f_1;i) = K + k(i) - g(i_1,i_2)t(i) \qquad \text{for } i_2<i\leq N,$$

$$v(1.f_1;i') = -k(i) + g(i_1,i_2)t(i) + v(i_1) \qquad \text{for } 0\leq i<i_1.$$

Using this, it is straightforward to verify from the construction of $f_2 = (k_1,k_2)$ that

$$K+k(k_1)-g(i_1,i_2)t(k_1) \leq v(i_1) \text{ and } -k(k_2)+g(i_1,i_2)t(k_2)+v(i_1) \leq 0,$$

where the equality sign in the first (second) inequality holds only when $k_1 = i_1$ $(k_2=i_2)$. In case the (k_1,k_2) policy is unequal to the (i_1,i_2) policy it now follows from (2.6) that

$$\text{(2.7)} \qquad g(k_1,k_2) = \{K+k(k_1)-k(k_2)\}/\{t(k_1)-t(k_2)\} < g(i_1,i_2),$$

so the above procedure constructs a policy f_2 whose average cost is lower than that of policy f_1 in case $f_2 \neq f_1$.

We shall now formulate two algorithms.

Algorithm 1

Step 0. For the (i_1,i_2) policy compute $g(i_1,i_2) =$
$= \{K+k(i_1)-k(i_2)\}/\{t(i_1) - t(i_2)\}$ and $v(i_1) = K+k(i_1)-g(i_1,i_2)t(i_1)$.
Step 1. In case either $i_2+1 = i_1$ or $-k(i_2+1)+g(i_1,i_2)t(i_2+1)+v(i_1) \geq 0$ when $i_2 < i_1 - 1$, let $j_2 = i_2$. Otherwise, let j_2 be the largest integer with $i_2 < j_2 < i_1$ such that $-k(i)+g(i_1,i_2)t(i) < 0$ for all $i_2 < i \leq j_2$.
Step 2. In case either $i_1 = j_2+1$ or $K+k(i_1-1)-g(i_1,i_2)t(i_1-1) \geq v(i_1)$ when $i_1 > j_2+1$, let $j_1 = i_1$. Otherwise, let j_1 be the smallest integer with

$j_2 < j_1 < i_1$ such that $K + k(i) - g(i_1,i_2)t(i) < v(i_1)$ for all $j_1 \le i < i_1$.
Step 3. Let k_2 be the largest integer for which $-k(i) + g(i_1,i_2)t(i)$ is minimal on $[0,j_2]$, and let k_1 be the smallest integer for which $k(i) - g(i_1,i_2)t(i)$ is minimal on $[j_1,N]$.
Step 4. If $k_1 \ne i_1$ or $k_2 \ne i_2$, go to step 0 with the (k_1,k_2) policy. Otherwise, the algorithm has been converged.

Since there are only a finite number of policies, it follows from (2.7) that this algorithm converges after a finite number of iterations to an (i_1^*,i_2^*) policy, say. Using relation (1) on p.106 in Derman (1970) it is easy to formulate conditions under which the (i_1^*,i_2^*) policy is optimal among the class Z of policies. These conditions can be numerically checked when the algorithm has been converged. We do not discuss these conditions which involve the probability distribution of the number of arrivals during a service according to service type 2.

It appears from computational considerations that the following algorithm requires in general less iterations than algorithm 1 above.

Algorithm 2.

The steps 0-2 and 4 are as in algorithm 1.
Step 3. Compute $g(j_1,j_2) = \{K + k(j_1) - k(j_2)\}/\{t(j_1) - t(j_2)\}$. Let k_2 be the largest integer for which $-k(i) + g(j_1,j_2)t(i)$ is minimal on $[0,j_2]$, and let k_1 be the smallest integer for which $k(i) - g(j_1,j_2)t(i)$ is minimal on $[j_1,N]$.

It is straightforward to verify from this algorithm that

$$g(i_1,i_2) \ge g(j_1,j_2) \ge g(k_1,k_2).$$

where the equality sign in the first [second] inequality holds only when $(j_1,j_2) = (i_1,i_2)$ $[(k_1,k_2) = (j_1,j_2)]$. Consequently algorithm 2 is finite too. It should be noted that algorithm 2 can also be derived directly from the modified policy iteration method outlined in remark 1.8 of section 1.4. In fact an iteration of algorithm 2 consists of both an iteration in which only a policy improvement operation is performed and an iteration in which only a cutting operation is performed.

REMARK 2.1. An examination of the algorithms 1 and 2 shows that it is immaterial when we replace the functions $k(i)$ and $t(i)$ by $k(i)+c_1$ and $t(i)+c_2$ for any constants c_1 and c_2. This observation may be useful when the functions $k(i)$ and $t(i)$ are very large as will be the case when $\lambda < \mu$ and N is very large.

REMARK 2.2. The finite machine repair problem with a single repairman, two possible repair times from which repair time 1 is exponential, and switch-over costs can be solved in an almost identical way. In this finite source problem the k- and t-functions must be computed by solving a system of linear equations. We note that for the finite machine problem with two exponential repair times and no switch-over costs Crabill (1973) has given conditions under which a switch-over policy (i_1,i_2) with $i_1 = i_2+1$ is average cost optimal among the class of all possible policies.

We have applied algorithm 2 to a number of numerical examples with $\mu = 1.1$, $\beta = .6$, $\beta^{(2)} = .72$, $h = 1$, $r_0 = 0$, $r_1 = 5$, $r_2 = 40$ en $N = 40$. In each example we have started the algorithm with policy $(N/2,0)$. In table 1 we give the results of the iterations of algorithm 2 for the example with $\lambda = 1$ and $K = 25$. For the other examples we give in table 2 the finally obtained policy (i_1^*,i_2^*), the average cost $g(i_1^*,i_2^*)$, and the number n of iterations required.

Table 1. The iterations when $\lambda = 1$ and $K = 25$.

iteration	(i_1,i_2)	$g(i_1,i_2)$	(j_1,j_2)	$g(j_1,j_2)$	(k_1,k_2)
1	(20,0)	12.3450	(20,16)	12.2797	(20,7)
2	(20,7)	12.0501	(13,9)	12.0395	(17,8)
3	(17,8)	11.9479	(15,9)	11.9424	(16,9)
4	(16,9)	11.9363	(16,9)	11.9363	(16,9)

Table 2. The policy (i_1^*,i_2^*).

	K = 0			K = 25			K = 50		
λ	(i_1^*,i_2^*)	$g(i_1^*,i_2^*)$	n	(i_1^*,i_2^*)	$g(i_1^*,i_2^*)$	n	(i_1^*,i_2^*)	$g(i_1^*,i_2^*)$	n
.8	(20,19)	6.2994	2	(25,17)	6.3013	5	(27,17)	6.3019	4
.9	(15,14)	8.4254	5	(20,12)	8.4655	4	(21,12)	8.4843	5
1.0	(12,11)	11.7220	4	(16,9)	11.9363	4	(17,8)	12.0505	4
1.1	(10,9)	16.1431	4	(13,6)	16.6396	3	(14,6)	16.9288	3
1.2	(8,7)	21.3958	2	(11,5)	22.1864	3	(12,4)	22.6408	4

REMARK 2.3. For the case where the service time under service type 2 is also exponential we were able to check numerically for each of the examples above that the following conditions guaranteeing the optimality of the (i_1^*,i_2^*) policy among the class Z of policies are satisfied,

$$-k(i)+g(i_1^*,i_2^*)t(i)+v(i_1^*) \le \sum_{j=0}^{i_2^*-i+1} \{-k(i-1+j)+g(i_1^*,i_2^*)t(i-1+j)\}p_j \qquad \text{for all } 0<i<i_2^*$$

and

$$K+k(i)-g(i_1^*,i_2^*)t(i) \le \lambda(\lambda+\mu)^{-1}\{K+k(i+1)-g(i_1^*,i_2^*)t(i+1)\}+$$
$$+\mu(\lambda+\mu)^{-1}\{K+k(i-1)-g(i_1^*,i_2^*)t(i-1)\} \qquad \text{for all } i_1^*<i<N,$$

where $p_j = (\lambda\beta)^j/(1+\lambda\beta)^{j+1}$ is the probability of j arrivals during one service according to service type 2.

APPENDIX

The appendix gives some results for discrete-time Markov processes with a general state space. Consider a Markov chain $\underline{x}_0, \underline{x}_1, \underline{x}_2, \ldots$ with stationary transition probability function $p(\cdot,\cdot)$ on $(S,\mathcal{B})$ where the state space S is a Borel set of a finite dimensional Euclidean space and $\mathcal{B}$ is the class of all Borel sets in S. For any $n \geq 0$, let $p^n(\cdot,\cdot)$ be the n-step transition probability function of the Markov chain. That is, $p^n(x,A) = \Pr\{\underline{x}_n \in A \mid \underline{x}_0 = x\}$. We assume that there is some state s (say) such that

(A.1) $$\Pr\{\underline{x}_n = s \text{ for some } n \geq 1 \mid \underline{x}_0 = x\} = 1 \qquad \text{for all } x \in S,$$

(A.2) $$E(\underline{N} \mid \underline{x}_0 = s) < \infty \text{ where } \underline{N} = \inf\{n \geq 1 \mid \underline{x}_n = s\}.$$

Let $\hat{p}^0(x,A) = 1$ for $x \in A$, let $\hat{p}^0(x,A) = 0$ for $x \notin A$, and let

$$\hat{p}^n(x,A) = \Pr\{\underline{x}_n \in A,\ \underline{x}_k \neq s \text{ for } 1 \leq k \leq n \mid \underline{x}_0 = x\} \quad \text{for } n \geq 1.$$

For any set $A \in \mathcal{B}$, define

(A.3) $$Q(A) = \sum_{n=0}^{\infty} \hat{p}^n(s,A)/E(\underline{N} \mid \underline{x}_0 = s).$$

Observe that, by $E\underline{N} = \sum_0^{\infty} \Pr\{\underline{N} > n\}$,

(A.4) $$E(\underline{N} \mid \underline{x}_0 = s) = \sum_{n=0}^{\infty} \hat{p}^n(s,S),$$

so, $Q(\cdot)$ is a probability distribution. We note that $Q(A)$ can be interpreted as the ratio of the expected number of visits of the Markov chain to the set A before returning to state s and the expected number of transitions needed to return to state s starting from state s.

THEOREM A.1. *For any* $A \in \mathcal{B}$

(A.5) $$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n} p^k(x,A) = Q(A) \qquad \textit{for all } x \in S$$

and

$$\text{(A.6)} \qquad Q(A) = \int_S p(x,A)Q(dx).$$

Further, Q *is the unique stationary probability distribution of the Markov chain* $\{\underline{x}_n\}$. *Also, when* $\underline{x}_0 = s$,

$$\text{(A.7)} \qquad \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n} Ef(\underline{x}_k) = \int_S f(x)Q(dx)$$

for any Baire function f *such that* $\int |f(x)|Q(dx)$ *is finite.*

PROOF. For any $x \in S$, let $f_0(x) = 0$, and let $f_n(x) = \Pr\{\underline{N} = n \mid \underline{x}_0 = x\}$ for $n \geq 1$. By (A.1), $\sum_0^\infty f_n(x) = 1$ for all x. Clearly, for any x and A (cf. p.365 in Feller (1966)),

$$\text{(A.8)} \qquad p^n(x,A) = \hat{p}^n(x,A) + \sum_{k=0}^{n} p^{n-k}(s,A)f_k(x) \qquad \text{for } n \geq 0.$$

For $x = s$ this relation is a renewal equation. By (A.2) and (A.4), both $\sum nf_n(s)$ and $\sum \hat{p}^n(s,A)$ are finite. Now, by applying the Key Renewal Theorem (see p.292 in Feller (1957)), for any $A \in \mathcal{B}$,

$$\text{(A.9)} \qquad \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n} p^k(s,A) = \sum_{n=0}^{\infty} \hat{p}^n(s,A)/\sum_{n=0}^{\infty} nf_n(s) = Q(A).$$

Since $\sum_0^\infty f_n(x) = 1$ and $\hat{p}^n(x,A) \to 0$ as $n \to \infty$ for all x and A, relation (A.5) now follows from (A.8) and (A.9). Using (A.5) it is easy to verify that Q satisfies the steady state equation (A.6) (cf. pp.133-134 in Breiman (1968)). Since the Markov chain $\{\underline{x}_n\}$ has no two disjoint closed sets, Q is the unique probability distribution satisfying (A.6), see Theorem 7.16 in Breiman (1968). To prove (A.7), let m be a finite measure on $(S,\mathcal{B})$ such that $m(A) > 0$ if and only if $s \in A$. Then, by (A.1), $m(A) > 0$ implies $\Pr\{\underline{x}_n \in A \text{ for some } n \geq 1 \mid \underline{x}_0 = x\} = 1$ for all $x \in S$. Consequently, the Markov chain $\{\underline{x}_n\}$ satisfies the recurrence condition of Harris (cf. pp.206-207 in Jain (1966)). Relation (A.7) now follows from Theorem 3.3 in Jain (1966).

REFERENCES

BREIMAN, L. (1968), *Probability*, Addison-Wesley, Reading, Massachusetts.

CRABILL, T.B. (1972), Optimal control of a service facility with variable exponential service time and constant arrival rate, *Management Sci.* 18, 560-566.

CRABILL, T.B. (1973), Optimal control of a maintenance system with variable service rates (to appear in *Operations Res.*)

DERMAN, C. (1966), Denumerable state Markovian decision processes-average cost criterion, *Ann. Math. Statist.* 37, 1545-1554.

DERMAN, C. and A.F. VEINOTT, Jr. (1967), A solution to a countable system of equations arising in Markovian decision processes, *Ann. Math. Statist.* 38, 582-584.

DERMAN, C. (1970), *Finite state Markovian decision processes*, Academic Press, New York.

FELLER, W. (1957), *An introduction to probability theory and its applications*, Vol.1. (2nd ed.), Wiley, New York.

FELLER, W. (1966), *An introduction to probability theory and its applications*, Vol.2, Wiley, New York.

HOWARD, R.A. (1960), *Dynamic programming and Markov processes*, Technology Press, Cambridge, Massachusetts.

JAIN, N.C. (1966), Some limit theorems for general Markov processes, *Z. Wahrscheinlichkeitstheorie verw. Geb.* 6, 206-223.

JEWELL, W.S. (1963), Markov renewal programming, I: Formulation, finite return models, II : Infinite return models, *Operations Res.* 6, 938-972.

LEVE, G. de (1964), *Generalized Markovian decision processes, Part I: Model and method, Part II : Probabilistic background*, Mathematical Centre Tracts No.3 and 4, Mathematisch Centrum, Amsterdam.

LEVE, G. de, TIJMS, H.C. and P.J. WEEDA (1970), *Generalized Markovian decision processes, applications*, Mathematical Centre Tract No.5, Mathematisch Centrum, Amsterdam.

LIPPMAN, S.A. (1973), Semi-Markov decision processes with unbounded rewards, *Management Sci.* 19, 717-731.

KEILSON, J. (1963), A gambler's ruin type problem in queueing theory, *Operations Res.* 11, 570-576.

ROSS, S.M. (1970), *Applied probability models with optimization applications*, Holden-Day, Inc., San Francisco.

ROYDEN, H.L. (1968), *Real Analysis* (2nd ed.), MacMillan, New York.

THATCHER, R.M. (1968), Optimal single channel service policies for stochastic arrivals, ORC 68-16, Operations Research Center, Univ. of California, Berkeley.

TIJMS, H.C. (1974), An alternative proof of the formula for the average workload cost for the D-policy in the M|G|1 queue, Report BW 31/74, Mathematisch Centrum, Amsterdam.

WEEDA, P.J. (1974a), Generalized Markov programming with a finite state semi-Markov process as natural process (extended abstract), Report BW 32/74, Mathematisch Centrum, Amsterdam.

WEEDA, P.J. (1974b) Generalized Markov programming with a finite state semi-Markov process as natural process, Report BW 34/74, Mathematisch Centrum, Amsterdam.