Adv. Appl. Prob. 9, 316–335 (1977)
Printed in Israel

A GENERAL MARKOV DECISION METHOD II: APPLICATIONS

G. DE LEVE, A. FEDERGRUEN AND
H. C. TIJMS, *Mathematisch Centrum, Amsterdam*

Abstract

In a preceding paper [2] we have introduced a new approach for solving a wide class of Markov decision problems in which the state-space may be general and the system may be continuously controlled. The criterion is the average cost. This paper discusses two applications of this approach. The first application concerns a house-selling problem in which a constructor builds houses which may be sold at any stage of the construction and potential customers make offers depending on the stage of the construction. The second application considers an $M/M/c$ queueing problem in which the number of operating servers can be controlled by turning servers on or off.

MARKOV DECISION PROBLEMS; AVERAGE COST; GENERAL STATE SPACE; CONTINUOUS CONTROL; APPLICATIONS; HOUSE-SELLING PROBLEM; $M/M/c$ QUEUEING PROBLEM WITH VARIABLE NUMBER OF SERVERS

1. Introduction

In a preceding paper [2] we have introduced a new approach for solving a wide class of Markov decision problems with the average cost as criterion, including problems in which the state-space is general and the system can be continuously controlled. This paper discusses two applications of this approach. Each of these applications will be illustrated with numerical results.

The first application concerns a house-selling problem in which a constructor builds houses which may be sold at any stage of the construction and potential customers make offers depending on the stage of the construction. From the optimality equation given in [2], an integral–differential equation for the curve determining an optimal policy for accepting offers is derived.

The second application considers the well-known $M/M/c$ queueing problem in which the number of servers turned on is variable. Using a general policy-iteration method developed in [2], we derive a special policy-iteration algorithm which exploits the structure of this problem and calculates an optimal policy

Received in revised form 4 November 1976.

within a certain class of structured policies for controlling the number of servers turned on.

In this paper we will follow the notation introduced in [2].

2. A house-selling problem

2.1 *Introduction.* Consider a building contractor constructing identical houses which may be sold in any stage of the construction. The construction time that is needed to perform a fraction u of the total building of a house has a gamma probability distribution function with density

$$g(t \mid u) = \Gamma(cu)^{-1} a^{cu} t^{(cu-1)} e^{-at}, \quad t \geqq 0,$$

where $a, c > 0$. Observe that this distribution has mean cu/a and that the distribution of the sum of the construction time of a fraction u_1 and that of a fraction u_2 has the same distribution as the construction time of a fraction $u_1 + u_2$, cf. p. 46 in Feller [3].

Potential customers for the houses arrive in accordance with a Poisson process with rate λ. Each potential customer makes an offer where the amount of money offered has a probability distribution function $F(\cdot \mid y)$ with finite mean when a fraction y of the total construction has been completed. If the offer is accepted, the house is sold and the contractor immediately starts with the construction of a new house. In case the building of a house is completed without any offer having been accepted in the meantime, the house will always be sold for an amount K. Finally, there are building costs at rate $b(y)$ when a fraction y of the total construction of the house has been completed.

Using the optimality equation (16) of [2] we shall characterize the structure of an average-cost optimal policy and show that this policy is in fact determined by an integral–differential equation. This will be done in Section 2.3 after we have specified the elements 1–6 of [2] in Section 2.2. Finally, in Section 2.4 we give some numerical results.

2.2 *The elements.* We first note that the state-space, the natural process and the feasible decisions must be chosen such that Element 4 of [2] applies. To achieve this, a convenient choice of the natural process is one in which the constructor accepts every offer and no new construction is started once a house is sold. As will become clear hereafter this choice will involve the introduction of an artificial state E (say) which is taken to be absorbing for the natural process. We now choose as state-space

$$X = \{y \mid 0 \leqq y < 1\} \cup \{(y_1, y_2) \mid 0 \leqq y_1 < 1,\ y_2 \geqq 0\} \cup \{E\}.$$

State y corresponds to the situation where a house is under construction and a fraction y of the total construction has been completed, while no offer is currently made. State (y_1, y_2) corresponds to the situation where an offer of size y_2 is made for a house of which a fraction y_1 of the total construction has been completed. State E corresponds to the situation where no house is under construction. The natural process is chosen as follows. Starting from state y^0 the natural process moves along the states y with $y^0 \leqq y < 1$ according to the 'building process' described above until either an offer is made or the total construction of the house is completed without any offer having been made in the meantime. In case of an offer of size y_2 in state y the natural process jumps to state (y, y_2) while in case of completion of the construction the natural process jumps to state E. The natural process starting from state (y_1, y_2) jumps immediately to state E. We take state E as an absorbing state for the natural process (e.g. imagine that in the natural process the contractor gives up his work in state E). Observe that in this natural process any offer is accepted. We next choose the feasible decisions. For each state y the only feasible decision is the null-decision which leaves the natural process untouched. For any state (y_1, y_2) the feasible decisions consist of the null-decision which prescribes acceptance of the offer and causes an instantaneous transition to state E, and the intervention $d = 1$ which prescribes refusal of the offer and causes an instantaneous transition to state y_1. The only feasible decision in state E is the intervention $d = 1$ which prescribes starting with a new construction and causes an instantaneous change to state 0. The following costs are associated with the natural process and the interventions. In the natural process there is incurred a cost at rate $b(y)$ when the natural process is in state y. Further, when the natural process makes a transition to state (y_1, y_2) a cost of $-y_2$ is incurred and when the natural process makes a transition to state E after completion of a construction a cost of $-K$ is incurred. Finally, by the above choices, there is no cost associated with any intervention.

Now, for any policy, the superimposition of the natural process and the interventions prescribed by that policy agrees with the evolution of the system resulting from the specific control as executed by the decision-maker. Clearly, Element 4 of [2] applies with

$$A_0 = \{E\}.$$

We choose $A_{01} = A_{02} = A_0$ in order to determine the k- and t-functions, see [2]. Clearly, for all $(y_1, y_2) \in X$,

$$t((y_1, y_2); 1) = t_0(y_1) - t_0((y_1, y_2)), \quad k((y_1, y_2); 1) = k_0(y_1) - k_0((y_1, y_2)),$$

and, furthermore,

$$t(E; 1) = t_0(0) \quad \text{and} \quad k(E; 1) = k_0(0).$$

Since the natural process starting from state (y_1, y_2) immediately jumps to state

E, we have for all (y_1, y_2),

$$t_0((y_1, y_2)) = 0 \quad \text{and} \quad k_0((y_1, y_2)) = -y_2.$$

Further, $t_0(y) = E\min[A, T(y)]$ for all $0 \leqq y < 1$ where A and $T(y)$ are independent random variables such that A is exponentially distributed with mean $1/\lambda$ and the construction time $T(y)$ has a gamma distribution with density $g(\cdot \mid 1-y)$. We find for $0 \leqq y < 1$,

$$t_0(y) = \frac{1}{\lambda}\left\{1 - \left(\frac{a}{a+\lambda}\right)^{c(1-y)}\right\}.$$

To determine the function $k_0(y)$, we first make the following observation. The building costs incurred between stages y_0 and y_1 of the construction are given by, for all $0 \leqq y_0 < y_1 < 1$,

$$\lim_{n\to\infty} \sum_{i=0}^{n} b\left(y_0 + \frac{i}{n}(y_1 - y_0)\right)\frac{c}{a}\left(\frac{y_1 - y_0}{n}\right) = \frac{c}{a}\int_{y_0}^{y_1} b(v)\,dv.$$

Further, for any initial state y with $0 \leqq y < 1$, let the random variable X_y be equal to 1 when the total construction is completed before a first offer occurs, and let X_y be equal to the stage of the construction at the epoch of the first offer, otherwise. It is routine to verify that, for all $0 \leqq y < 1$,

$$\Pr\{X_y = 1\} = \left(\frac{a}{a+\lambda}\right)^{c(1-y)} \quad \text{and} \quad \Pr\{X_y \leqq u\} = 1 - \left(\frac{a}{a+\lambda}\right)^{c(u-y)}$$
$$\text{for} \quad y \leqq u < 1.$$

Let $h(u \mid y)$ be the derivative of $\Pr\{X_y \leqq u\}$ with respect to u. Then for all $0 \leqq y < 1$,

$$h(u \mid y) = c\ln\left(1 + \frac{\lambda}{a}\right)\left(\frac{a}{a+\lambda}\right)^{c(u-y)} \quad \text{for} \quad y < u < 1.$$

Now, we have by the choice of the natural process that, for all $0 \leqq y < 1$,

$$k_0(y) = E\left[\frac{c}{a}\int_y^{X_y} b(v)\,dv\right] - \int_y^1 \left\{\int_0^\infty v\,dF(v \mid u)\right\} h(u \mid y)\,du - K\Pr\{X_y = 1\}$$

from which we get after some algebra

$$k_0(y) = \frac{c}{a}\int_y^1 b(u)\left(\frac{a}{a+\lambda}\right)^{c(u-y)} du - \alpha(y) - \left(\frac{a}{a+\lambda}\right)^{c(1-y)} K, \qquad 0 \leqq y < 1,$$

where

$$\text{(2.1)} \qquad \alpha(y) = c\ln\left(1 + \frac{\lambda}{a}\right)\int_y^1 \left\{\int_0^\infty v\,dF(v \mid u)\right\}\left(\frac{a}{a+\lambda}\right)^{c(u-y)} du, \quad 0 \leqq y < 1.$$

Note that by the above choices of the natural process and the feasible decisions the functions $k(x; z(x))$ and $t(x; z(x))$ are bounded for $x \in A_z$ for each policy z that has the property of accepting at each stage of the construction any offer which exceeds some critical (possibly very large) value. The assumptions A1–A4 in [2] are satisfied for this class Z of policies.

2.3 *Characterization of an optimal policy.* In this section we shall derive from the optimality equation (16) of [2] the existence and the structure of an average-cost optimal policy. Moreover, we shall find that in fact such a policy is determined by an integral–differential equation.

Now, let z^* be any policy of Z. Denote by $\{g(z^*), v(z^*; x) \;\; x \in X\}$ the unique solution to the equations (8)–(9) with $z = z^*$ of [2] such that

$$v(z^*; E) = 0. \tag{2.2}$$

Since the intervention $d = 1$ in state (y_1, y_2) causes an instantaneous transition to state y_1, it follows from relation (11) of [2] and the above formulas for the functions k and t that

$$\begin{aligned} v(z^*; (y_1, y_2)) &= k((y_1, y_2); 1) - g(z^*)t((y_1, y_2); 1) + v(z^*; y_1) \\ &= y_2 + R(z^*; y_1) + v(z^*; y_1) \quad \text{for all} \quad (y_1, y_2) \in A_{z^*}, \end{aligned} \tag{2.3}$$

where

$$\begin{aligned} R(z^*; y) = \frac{c}{a} \int_y^1 b(u) \left(\frac{a}{a+\lambda}\right)^{c(u-y)} du - \alpha(y) - \left(\frac{a}{a+\lambda}\right)^{c(1-y)} K \\ - \frac{g(z^*)}{\lambda} \left\{1 - \left(\frac{a}{a+\lambda}\right)^{c(1-y)}\right\} \quad \text{for} \quad 0 \leqq y < 1. \end{aligned} \tag{2.4}$$

By relation (9) of [2] and the fact that the natural process starting from state (y_1, y_2) jumps to the intervention state E, we have

$$v(z^*; (y_1, y_2)) = v(z^*; E) = 0 \quad \text{for} \quad (y_1, y_2) \notin A_{z^*}. \tag{2.5}$$

Finally, by relation (11) of [2],

$$v(z^*; E) = k(E; 1) - g(z^*)t(E; 1) + v(z^*; 0) = R(z^*; 0) + v(z^*; 0). \tag{2.6}$$

Now, let $z \in Z$. Then, by virtue of the fact that the only possible intervention is $d = 1$, it follows from the relations (11) and (13)–(14) of [2] that

$$v([z]z^*; E) = k(E; 1) - g(z^*)t(E; 1) + v(z^*; 0) = v(z^*; E) \tag{2.7}$$

and

$$\begin{aligned} v([z]z^*; (y_1, y_2)) &= k((y_1, y_2); 1) - g(z^*)t((y_1, y_2); 1) + v(z^*; y_1) \\ &= y_2 + R(z^*; y_1) + v(z^*; y_1) \quad \text{for all} \quad (y_1, y_2) \in A_z. \end{aligned} \tag{2.8}$$

Further, by definition (14) of [2] and the relations (2.2) and (2.7),

$$v([z]z^*;(y_1,y_2)) = v([z]z^*;E) = 0 \quad \text{for all} \quad (y_1,y_2) \notin A_z. \tag{2.9}$$

We shall now prove that a policy $z^* \in Z$ satisfies the optimality equation (see (16) of [2])

$$v(z^*;x) = \min_{z \in Z} v([z]z^*;x) \quad \text{for all} \quad x \in X_0 \tag{2.10}$$

if and only if for policy z^* the following inequalities hold:

$$y_2 + R(z^*;y_1) + v(z^*;y_1) \leqq 0 \quad \text{for all} \quad (y_1,y_2) \in A_{z^*} \tag{2.11}$$

$$y_2 + R(z^*;y_1) + v(z^*;y_1) \geqq 0 \quad \text{for all} \quad (y_1,y_2) \notin A_{z^*}. \tag{2.12}$$

To prove this, we first observe that, by relation (15) of [2] and (2.7), the optimality equation (2.10) is equivalent to

$$v([z]z^*;(y_1,y_2)) \geqq v(z^*;(y_1,y_2)) \quad \text{for all} \quad (y_1,y_2) \in X \quad \text{and all } z \in Z. \tag{2.13}$$

Suppose first that (2.13) holds. To establish (2.11), we observe that for any state $(y_1,y_2) \in A_{z^*}$ we can find a policy $z \in Z$ such that $(y_1,y_2) \notin A_z$, so, by (2.3), (2.9) and (2.13), we get (2.11). Also, for any state $(y_1,y_2) \notin A_{z^*}$ we can find a policy $z \in Z$ such that $(y_1,y_2) \in A_z$ so, by (2.5), (2.8) and (2.13), we get (2.12). Next assume that (2.11)–(2.12) hold. To verify (2.13), fix $z \in Z$. For $(y_1,y_2) \notin A_z$, we get (2.13) from (2.9), (2.5), (2.3) and (2.11). For $(y_1,y_2) \in A_z$ we get (2.13) from (2.8), (2.3), (2.5) and (2.12).

We now have proved that a policy $z^* \in Z$ for which (2.11)–(2.12) hold is optimal. Moreover, we can conclude that such a policy z^* is determined by a function $s(y_1)$, $0 \leqq y_1 < 1$ such that

$$A_{z^*} = \{(y_1,y_2) \mid y_2 \leqq s(y_1)\}. \tag{2.14}$$

Furthermore,

$$s(y_1) = -R(z^*;y_1) - v(z^*;y_1). \tag{2.15}$$

Since we know the structure of A_{z^*} we can express $v(z^*;y_1)$ in the function $s(\cdot)$. To do this, we first observe that, by (2.3), (2.5) and (2.14)–(2.15), for all (y_1,y_2)

$$v(z^*;(y_1,y_2)) = \begin{cases} y_2 - s(y_1) & \text{for } y_2 \leqq s(y_1), \\ 0 & \text{for } y_2 \geqq s(y_1). \end{cases} \tag{2.16}$$

Using relation (12) of [2] with $V = \{(y_1,y_2)\} \cup \{E\}$, (2.2) and (2.16), we get

$$
\begin{aligned}
v(z^*;y) &= \int_y^1 \left\{ \int_0^\infty v(z^*;(u,v))\,dF(v\mid u) \right\} h(u\mid y)\,du + v(z^*;E)\Pr\{X_y=1\} \\
&= c\ln\left(1+\frac{\lambda}{a}\right) \int_y^1 \left\{ \int_0^{s(u)} (v-s(u))\,dF(v\mid u) \right\} \left(\frac{a}{a+\lambda}\right)^{c(u-y)} du, \\
&\qquad\qquad 0 \leqq y < 1.
\end{aligned} \tag{2.17}
$$

From this relation and (2.15), we get for $0 \leqq y_1 < 1$,

$$
s(y_1) = - R(z^*;y_1) - c\ln\left(1+\frac{\lambda}{a}\right) \int_{y_1}^1 \left\{ \int_0^{s(u)} (v-s(u))\,dF(v\mid u) \right\} \left(\frac{a}{a+\lambda}\right)^{c(u-y_1)} du.
$$

Differentiating this formula and using (2.1) and (2.4), we get, after some algebra,

$$
\begin{aligned}
s'(y_1) &= \frac{c}{a} b(y_1) \\
&\quad + c\ln\left(1+\frac{\lambda}{a}\right) \left\{ \int_{s(y_1)}^\infty (s(y_1)-v)\,dF(v\mid y_1) - \frac{g(z^*)}{\lambda} \right\}, \quad 0 \leqq y_1 < 1.
\end{aligned} \tag{2.18}
$$

Using the fact that $\lim_{y\to 1} v(z^*;y) = 0$ (see (2.17)) and the relations (2.2) and (2.6), we have the boundary conditions

$$
s(0) = 0 \quad \text{and} \quad s(1) = K. \tag{2.19}
$$

The integral–differential equation (2.18) and the boundary conditions (2.19) determine both the curve $s(\cdot)$ giving the optimal policy z^* and the minimal average cost $g(z^*)$.

2.4 *Numerical results.* In this section we give some numerical results for the case where $F(\cdot\mid y_1)$ is a gamma distribution with density

$$
\frac{\{n\lambda(y_1)\}^n}{(n-1)!} v^{n-1} e^{-n\lambda(y_1)v}, \qquad v \geqq 0,
$$

where n is a positive integer and $\lambda(y_1)$ is a given function. Observe that the mean and the variance of this distribution are equal to $1/\lambda(y_1)$ and $1/n\{\lambda(y_1)\}^2$. By a well-known relation between the Poisson distribution and the gamma distribution, we have

$$
\begin{aligned}
\int_{s(y_1)}^\infty (s(y_1)-v)\,dF(v\mid y_1) = e^{-n\lambda(y_1)s(y_1)} \Bigg\{ & s(y_1) \sum_{j=0}^{n-2} [n\lambda(y_1)s(y_1)]^j/j\,! \\
& - \frac{1}{\lambda(y_1)} \sum_{j=0}^{n-1} [n\lambda(y_1)s(y_1)]^j/j\,! \Bigg\}.
\end{aligned}
$$

Hence the relation (2.18) reduces to a differential equation with unknown parameter $g(z^*)$. To solve this differential equation with the boundary conditions (2.19), we have used a computer program developed [11] for parameter estimation in differential equations. In Table 1 we give some numerical results.

TABLE 1.

$\lambda = 2$, $a = 1$, $c = 1$, $K = 2$, $b(y) = 1$ and $\lambda(y) = 1/(3y + 0.01)$

n	$g(z^*)$	y	0.15	0.30	0.45	0.60	0.75	0.90
1	−3.243	$s(y)$	0.401	0.770	1.106	1.405	1.665	1.881
5	−2.816	$s(y)$	0.371	0.721	1.048	1.350	1.622	1.861
10	−2.729	$s(y)$	0.365	0.712	1.037	1.339	1.614	1.858

3. An $M/M/c$ queueing problem with a variable number of servers

3.1 *Introduction.* We consider the $M/M/c$ queueing problem studied by McGill [6], where the number of servers operating can be adjusted at arrival and service completion epochs. The customers arrive in accordance with a Poisson process with rate λ, and there are c independent servers available each having an exponentially distributed service time with mean $1/\mu$. It is assumed that the lowest possible traffic intensity $\lambda/c\mu$ is less than 1. The cost structure includes a holding cost of $h > 0$ per customer in the system per unit time, an operating cost of $w > 0$ per server turned on per unit time and a switch-over cost of $K(a, b)$ when the number of servers turned on is adjusted from a to b. We assume that

$$K(a,b) = k^+ .(b-a) \text{ when } a < b \text{ and } K(a,b) = k^- .(a-b) \text{ when } a \geqq b,$$

where $k^+, k^- \geqq 0$. This problem has been treated amongst others by Bell [1], Lippman [5], McGill [6], Robin [8] and Sobel [9]: cf. also Sobel [10]. It was shown by Lippman [5] that there is an integer M such that an average-cost optimal policy has all c servers turned on or left on when M or more customers are present. We henceforth only consider the following finite class C of stationary policies with this property. A policy in C is characterized by integers $s(i)$, $S(i)$, $t(i)$ and $T(i)$ for $i = 0, 1, \cdots$ such that

(a) $-1 \leqq s(i) < S(i) \leqq T(i) < t(i) \leqq c + 1$ for all $i \geqq 0$, where $s(i) = c - 1$, $S(i) = T(i) = c$ and $t(i) = c + 1$ for all $i \geqq M$,

(b) $s(i) \leqq s(i+1)$ and $t(i) \leqq t(i+1)$ for all $i \geqq 0$.

Under this policy the number of servers operating is adjusted both at arrival and service completion epochs. If there are i customers present and k servers turned on, the number of servers on is adjusted upward to $S(i)$ when $k \leqq s(i)$, is kept unaltered when $s(i) < k < t(i)$ and is adjusted downward to $T(i)$ when $k \geqq t(i)$.

It is a famous conjecture that there is an average-cost optimal policy which belongs to the class C and has the additional property that $S(i) = s(i) + 1$ and $T(i) = t(i) - 1$ for all i.

In this paper a special policy iteration algorithm will be developed whic locates an average-cost optimal policy. This algorithm generates within the clas C a sequence of improved policies, and in all examples tested the algorithr converged to an optimal policy with $S(i) = s(i) + 1$ and $T(i) = t(i) - 1$ for all (We note that the algorithm may also be used for locating an average-co: optimal policy within the class C for the case of general switch-over costs.) Th algorithm exploits the structure of the particular queueing problem. Thi appears especially in the value-determination part of the algorithm in which th size of the system of linear equations to be solved is of the order $2M$ independent of c. In addition the algorithm does not require any truncation c the state-space, i.e. no approximation of the infinite-capacity problem to a finit one is needed. These facts compare favourably with the policy iteratio algorithm in Howard [4] in which Nc linear equations must be solved in th value-determination part, where the integer N arises from the truncation of th state-space and denotes the maximum number of customers allowed in th system. We may expect that $N \gg M$, especially when $\lambda / c\mu$ is close to 1, in whicl case a large choice of N is required in order to obtain a fair approximation of th infinite-capacity problem, whereas the estimate of M tends to be small since i this case an optimal policy tends to have all c servers on with relatively fev customers in the system.

In Section 3.2 we specify the basic Elements 1–6 of [2] which are crucial for th algorithm and we determine some absorption probabilities which underly th transition probabilities of the embedded decision processes. In Section 3.3 w derive the system of linear equations to be solved in the value-determinatio operation. Finally, in Section 3.4 we present the algorithm and give som numerical results.

3.2 *The elements.* In choosing the state-space, the natural process and th feasible decisions, similar considerations to those in the first application will pla a role. In order to obtain a set A_0 which has the desired properties and furthe allows for computationally tractable k- and t-functions, we will choose th Elements 1–3 in such a way that in the natural process c servers will always b turned on when the number of customers is larger than M and, moreover, th states in which no customers are present are intervention states for any policy The latter can always be achieved by choosing these states absorbing for th natural process, e.g. imagine that in the natural process the system is close down forever when the system becomes empty. This choice involves th introduction of both (artificial) interventions for these states and (artificial) state to which the system is transferred by these interventions.

After these introductory remarks, we now choose as state-space

$$X = \{(i, s) \mid i = 0, 1, \cdots; s = 0, 1, \cdots, c\} \cup \{(\bar{0}, \bar{s}) \mid s = 0, 1, \cdots, c\},$$

where state (i, s) with $i \geqq 1$ corresponds to the situation where i customers are present and there are s servers turned on of which $\min(i, s)$ servers provide service. The state $(0, s)$ corresponds to the situation where no customers are present, there are s servers turned on and the servers are not available for any future service, while state $(\bar{0}, \bar{s})$ corresponds to the same situation except that the servers are now available for future service. We choose the natural process as follows. For both initial state (i, s) with $1 \leqq i \leqq M$ and initial state (i, s) with $i \neq 0$ and $s = c$ the natural process stays in state (i, s) until the next epoch at which an arrival or service completion occurs, after which the natural process assumes either state $(i + 1, s)$ or $(i - 1, s)$ depending upon whether an arrival or service completion occurs first, so for these initial states the number of servers on is left unaltered in the natural process. For initial state (i, s) with $i > M$ and $s \neq c$ the natural process jumps immediately to state (i, c), i.e. for this initial state the number of servers is adjusted upward to c in the natural process. The states $(0, m)$, $m = 0, \cdots, c$ are chosen as absorbing states for the natural process, whereas the natural process starting from state $(\bar{0}, \bar{s})$ stays in this state until the next arrival epoch at which the natural process assumes state $(1, s)$.

We next choose the sets of feasible decisions. For state (i, s) with $1 \leqq i \leqq M - 1$ and $s \neq c$ the set of feasible decisions consists of the decisions $d = 0, 1, \cdots, c$ where decision d prescribes to adjust the number of servers turned on from s to d and causes an instantaneous transition to state (i, d). Observe that for this state (i, s) the decision $d = s$ is the null-decision and any decision $d \neq s$ is an intervention. In state (M, s) with $s \neq c$ we choose as the only possible decision the intervention $d = c$ which prescribes an upward adjustment of the number of servers to c and causes an instantaneous transition to state (M, c). In each of the states $(0, s)$, $0 \leqq s \leqq c$, the set of feasible decisions consists of the interventions $d = 0, \cdots, c$ where the intervention d prescribes to 'reactivate' the servers and to adjust the numbers of servers on from s to d and causes an instantaneous transition to state $(\bar{0}, \bar{d})$. Finally, in the states (M, c) and $(\bar{0}, \bar{s})$ for $0 \leqq s \leqq c$ we take the null-decision as the only possible decision. The cost structure is as follows. In the natural process a holding cost at rate $h \,.\, i$ and an operating cost at rate $w \,.\, s$ are incurred when there are i customers present and s servers turned on. An intervention cost of $K(s, d)$ is incurred when the intervention d is made in any state in which s servers are on.

Now, for any policy, the superimposition of the natural process and the interventions prescribed by that policy agree with the evolution of the system resulting from the specific control as executed by the decision-maker. Using the fact that $\lambda / c\mu < 1$, it follows that Element 4 of [2] applies with

$$A_0 = \{(0, s) \mid s = 0, \cdots, c\} \cup \{(M, s) \mid s = 0, \cdots, c - 1\}.$$

To determine the k- and t-functions introduced in Element 5 of [2], we choose

$A_{01} = A_{02} = A_0$. From the definitions of these functions, it follows that, for any state (i, s) with $i \neq 0$ and intervention d,

$$t((i,s);d) = t_0((i,d)) - t_0((i,s)),$$

$$k((i,s);d) = K(s,d) + k_0((i,d)) - k_0((i,s)).$$

Further, for any state $(0, s)$ and $d = 0, \cdots, c$,

$$t((0,s);d) = t_0((\bar{0},\bar{d})) - t_0((0,s)),$$

$$k((0,s);d) = K(s,d) + k_0((\bar{0},\bar{d})) - k_0((0,s)).$$

We shall now calculate the functions t_0 and k_0 as far as needed. Fix s with $s \neq c$. Then

$$(3.1)\quad t_0((i,s)) = \begin{cases} (\lambda + i\mu)^{-1}[1 + i\mu t_0((i-1,s)) + \lambda t_0((i+1,s))], & 1 \leqq i \leqq s, \\ (\lambda + s\mu)^{-1}[1 + s\mu t_0((i-1,s)) + \lambda t_0((i+1,s))], & s \leqq i \leqq M-1, \end{cases}$$

with $t_0((0,s)) = t_0((M,s)) = 0$. For ease of notation, denote by $h_0((i,s))$ the component of $k_0((i,s))$ in which the expected holding costs are represented, i.e.

$$k_0((i,s)) = swt_0((i,s)) + h_0((i,s)).$$

We have

$$(3.2)\quad h_0((i,s)) = \begin{cases} (\lambda + i\mu)^{-1}[hi + i\mu h_0((i-1,s)) + \lambda h_0((i+1,s))], & 1 \leqq i \leqq s, \\ (\lambda + s\mu)^{-1}[hi + s\mu h_0((i-1,s)) + \lambda h_0((i+1,s))], & s \leqq i \leqq M-1, \end{cases}$$

with $h_0((0,s)) = h_0((M,s)) = 0$. We now discuss briefly the solution of (3.1). The solution of (3.2) proceeds in the same way. We refer to Miller [7] for details. The equation for $t_0((i,s))$ is a second-order linear difference equation with non-constant coefficients for $i \leqq s$ and constant coefficients for $i \geqq s$. The solution of the equation with constant coefficients is standard. To solve the equation with non-constant coefficients, multiply both sides of this equation by $\lambda + i\mu$ and consider the equation for $\Delta t_0(i) = t_0((i+1,s)) - t_0((i,s))$. This equation is a first-order linear difference equation and a particular solution may be found by using the method of parameter variation. We find for the case of $\lambda / s\mu \neq 1$,

$$t_0((i,s)) = \begin{cases} a_1(i) + \beta_1 b(i) + \alpha_1 & \text{for} \quad 0 \leqq i \leqq s, \\ c_1(i) + \delta_1 d(i) + \gamma_1 & \text{for} \quad s \leqq i \leqq M, \end{cases}$$

where

$$a_1(i) = -\sum_{t=0}^{i-1}\sum_{j=0}^{t}\frac{t!(\mu/\lambda)^{t-j}}{\lambda j!}, \quad b(i) = \sum_{j=0}^{i-1}(\mu/\lambda)^j j! \tag{3.3}$$

$$c_1(i) = (i-M)/(s\mu-\lambda), \quad d(i) = (s\mu/\lambda)^i - (s\mu/\lambda)^M. \tag{3.4}$$

By the boundary conditions $t_0((0,s)) = t_0((M,s)) = 0$, we have $\alpha_1 = \gamma_1 = 0$. The constants β_1 and δ_1 follow by considering (3.1) for $i = s$ and substituting the above explicit expressions for $t_0((i,s))$ with $i = s-1$, s and $s+1$ where there are two possibilities for $t_0((s,s))$. To save space, we omit the formulas for these constants. For the same reason, we omit the expression for $t_0((i,s))$ when $\lambda/s\mu = 1$.

Similarly, we find for the case of $\lambda/s\mu \neq 1$,

$$h_0((i,s)) = \begin{cases} a_2(i) + \beta_2 b(i) & \text{for} \quad 0 \leqq i \leqq s, \\ c_2(i) + \delta_2 d(i) & \text{for} \quad s \leqq i \leqq M, \end{cases}$$

where

$$a_2(i) = -h\sum_{t=0}^{i-1}\sum_{j=1}^{t}\frac{t!(\mu/\lambda)^{t-j}}{\lambda(j-1)!}, \quad c_2(i) = \frac{h(i^2-M^2)}{2(s\mu-\lambda)} + \frac{h(s\mu+\lambda)(i-M)}{2(s\mu-\lambda)^2}. \tag{3.5}$$

The constants β_2 and δ_2 follow by the same considerations as above.

Next we determine the functions $t_0((i,c))$ and $h_0((i,c))$ where h_0 is defined as above. Clearly,

$$t_0((i,c)) = (\lambda + i\mu)^{-1}[1 + i\mu t_0((i-1,c)) + \lambda t_0((i+1,c))], \quad 1 \leqq i \leqq c, \tag{3.6}$$

with $t_0((0,c)) = 0$. To give a recursive relation for $t_0((i,c))$ for $i \geqq 1$, we make the following observation. Using the 'memoryless' property of the exponential distribution, it is easily seen that the time needed to reduce the number of customers from $i \geqq c$ to $i-1$ by using c exponential servers having each mean service time $1/\mu$ is distributed as the length of one busy period in the $M/M/1$ queue with arrival rate λ and mean service time $1/c\mu$. This implies

$$t_0((i,c)) = 1/(c\mu-\lambda) + t_0((i-1,c)) \quad \text{for} \quad i \geqq c. \tag{3.7}$$

Using $t_0((0,c)) = 0$, we get that the solution to (3.6) is given by

$$t_0((i,c)) = a_1(i) + \xi_1 b(i) \quad \text{for} \quad 0 \leqq i \leqq c,$$

where $a_1(i)$ and $b(i)$ are defined in (3.3) and the constant ξ_1 follows by using (3.7) with $i = c$. Next we find

$$h_0((i,c)) = (\lambda + i\mu)^{-1}[hi + i\mu h_0((i-1,c)) + \lambda h_0((i+1,c))], \quad 1 \leqq i \leqq c, \tag{3.8}$$

with $h_0((0,c)) = 0$. Using the fact that for the above $M/M/1$ queue the total expected amount of time spent by the customers in the system during one busy period equals $c\mu/(c\mu-\lambda)^2$ (observe that the ratio of this quantity and the expected length of one busy cycle gives the average number of customers present), we find

$$h_0((i,c)) = \frac{h(i-1)}{c\mu-\lambda} + \frac{hc\mu}{(c\mu-\lambda)^2} + h_0((i-1,c)) \quad \text{for} \quad i \geqq c. \tag{3.9}$$

Using $h_0((0,c))=0$, we find that the solution to (3.8) is given by

$$h_0((i,c)) = a_2(i) + \xi_2 b(i)$$

where $a_2(i)$ and $b(i)$ are given in (3.5) and (3.3) and the constant ξ_2 follows by using (3.9) with $i=c$.

We end this section by determining some absorption probabilities which underly the one-step transition probabilities of the embedded decision processes. For any integers i, s, L and R with $0 \leqq L \leqq i \leqq R \leqq M$, $R \neq L$ and $0 \leqq s \leqq c$, define $p(i,s,L,R)$ as the probability that the natural process starting from state (i,s) will assume state (R,s) before state (L,s). Suppress for the moment the dependence of p on L, R and s and write $p(i,s,L,R)=p(i)$. Since in the natural process the number of servers on is not changed as long as not more than M customers are present, we find

$$p(i) = \begin{cases} (\lambda + i\mu)^{-1}[i\mu p(i-1) + \lambda p(i+1)] & \text{for} \quad i \leqq s, \\ (\lambda + s\mu)^{-1}[s\mu p(i-1) + \lambda p(i+1)] & \text{for} \quad i \geqq s, \end{cases} \tag{3.10}$$

with $p(L)=0$ and $p(R)=1$. We give only the solution when $\lambda/s\mu \neq 1$ and we distinguish between three cases.

Case 1. $L \geqq s$. Then we find the solution of the classical ruin problem,

$$p(i,s,L,R) = \{(s\mu/\lambda)^i - (s\mu/\lambda)^L\}/\{(s\mu/\lambda)^R - (s\mu/\lambda)^L\} \quad \text{for all } i.$$

Case 2. $R \leqq s$. Then

$$p(i,s,L,R) = \left\{\sum_{j=L}^{i-1} (\mu/\lambda)^j j\,!\right\} \Big/ \left\{\sum_{j=L}^{R-1} (\mu/\lambda)^j j\,!\right\} \quad \text{for all } i.$$

Case 3. $L < s < R$. Then

$$p(i,s,L,R) = \begin{cases} 1 + \eta_1\{(s\mu/\lambda)^i - (s\mu/\lambda)^R\} & \text{for } s \leqq i \leqq R \\ \eta_2 \sum_{j=L}^{i-1} (\mu/\lambda)^j j\,! & \text{for } L \leqq i \leqq s, \end{cases}$$

where the constants η_1 and η_2 follow by the same considerations as before.

3.3 *The system of equations for a policy of the class C.* Fix policy $z \in C$. In this section we shall specify for policy z the system of equations (8)–(9) introduced in [2]. We recall that policy z is characterized by integers $s(i)$, $S(i)$, $t(i)$ and $T(i)$ for $i = 0, \cdots, M-1$ (see Section 3.1) and we observe that its set of intervention states is given by $A_z = \{(i,s) \mid i \geqq 1, \quad s \leqq s(i) \quad \text{or} \quad s \geqq t(i)\} \cup \{(0,s) \mid 0 \leqq s \leqq c\}$. By the structure of policy z we have that after any intervention the system assumes one of the states $(i, S(i))$, $(i, T(i))$ or $(\bar{0}, \bar{s})$ where $1 \leqq i \leqq M$ and $s(0) < s < t(0)$. This fact will have as a consequence that in the value-determination procedure we need only to solve $2M + t(0) - s(0) - 3$ linear equations. Before showing this, we note that, by the monotonicity properties of policy z, the set A_z will be entered in one of the states $(L(s), s)$ and $(R(s), s)$ with $0 \leqq s \leqq c$ where

$$L(s) = \max\{i \mid 1 \leqq i \leqq M,\ t(i) \leqq s\} \text{ if } s \geqq t(0), \text{ and } L(s) = 0,$$

(3.11) otherwise,

$$R(s) = \min\{i \mid 1 \leqq i \leqq M,\ s(i) \geqq s\} \text{ if } s < c, \text{ and } R(c) = \infty.$$

That is, for s servers turned on, $L(s)$ denotes the largest queue size for which policy z prescribes either a reduction of the number of servers on or at least their 'reactivation', whereas $R(s)$ denotes the smallest queue size for which policy z prescribes an upward adjustment of the number of servers turned on.

We now specify the equations for the average cost g and the relative values $v((i,s))$ with $(i,s) \in A_z$. By relation (11) in [2], we have for $1 \leqq i \leqq M$

$$\begin{aligned} v((i,s)) &= k((i,s); S(i)) - gt((i,s); S(i)) + v((i, S(i))), \quad s \leqq s(i), \\ v((i,s)) &= k((i,s); T(i)) - gt((i,s); T(i)) + v((i, T(i))), \quad s \geqq t(i), \end{aligned} \tag{3.12}$$

whereas for the intervention states $(0,s)$, $0 \leqq s \leqq c$, we find

$$\begin{aligned} v((0,s)) &= k((0,s); S(0)) - gt((0,s); S(0)) + v((\bar{0}, \overline{S(0)})), \quad s \leqq s(0), \\ v((0,s)) &= k((0,s); T(0)) - gt((0,s); T(0)) + v((\bar{0}, \overline{T(0)})), \quad s \geqq t(0), \\ v((0,s)) &= k((0,s); s) - gt((0,s); s) + v((\bar{0}, \bar{s})), \quad \text{otherwise}. \end{aligned} \tag{3.13}$$

Letting $p(i, s, L(s), R(s))$ for $s < c$ be defined as in Section 3.2 and letting $p(i, c, L(c), R(c)) = 0$, it follows from relation (9) in [2] that, for state $(i,s) \notin A_z$,

$$\begin{aligned} v((i,s)) = {} & p(i, s, L(s), R(s)) v((R(s), s)) \\ & + \{1 - p(i, s, L(s), R(s))\} v((L(s), s)). \end{aligned} \tag{3.14}$$

Further, using the fact that $L(s) = 0$ for $s(0) < s < t(0)$, we find

$$\begin{aligned} v((\bar{0}, \bar{s})) = {} & p(1, s, 0, R(s)) v((R(s), s)) \\ & + \{1 - p(1, s, 0, R(s))\} v((0,s)) \quad \text{for} \quad s(0) < s < t(0). \end{aligned} \tag{3.15}$$

The equations for the remaining relative values will not be needed and are omitted.

It now follows that we get $2M+t(0)-s(0)-3$ linear equations in the $2M+t(0)-s(0)-2$ unknowns g, $v((i,S(i)))$, $v((i,T(i)))$ and $v((\bar{0},\bar{s}))$ with $1 \leqq i \leqq M-1$ and $s(0)<s<t(0)$ by taking the equations (3.15) and the equations (3.14) with both $s=S(i)$ and $s=T(i)$ and by substituting in the right-hand sides of these equations the corresponding equations for $v((R(s),s))$ and $v((L(s),s))$, cf. (3.12)–(3.13). To determine these unknowns uniquely, we put one of the relative values equal to zero (see Theorem 2 in [2]), e.g. put $v((M-1,T(M-1)))=0$. Once the above $2M+t(0)-s(0)-3$ linear equations have been solved, we can next compute any of the required $v(x)$ from (3.12)–(3.14).

3.4 *The algorithm.* We shall now present a policy-iteration algorithm which generates a sequence of policies belonging to the class C of structured policies. Before specifying the details of this algorithm, we first give a general outline of the algorithm which is based on the modified policy iteration method given in Section 5 of [2].

Algorithm (a) *Value-determination procedure.* Solve for the current policy $z \in C$ with parameters $s(i)$, $S(i)$, $t(i)$ and $T(i)$ the above-described system of $2M+t(0)-s(0)-3$ linear equations.

(b) *Policy-improvement procedure.* Determine a policy $z' \in C$ with parameters $s'(i)$, $S'(i)$, $t'(i)$ and $T'(i)$ where $s'(i) \geqq s(i)$ and $t'(i) \leqq t(i)$.

(c) *Cutting-procedure.* Determine a policy $z'' \in C$ with parameters $s''(i)$, $S''(i)$, $t''(i)$ and $T''(i)$ where $S''(i)=S'(i)$, $T''(i)=T'(i)$, $s''(i) \leqq s'(i)$ and $t''(i) \geqq t'(i)$.

(d) If $z''=z$, stop, otherwise, go to (a).

We now give in detail the policy-improvement and the cutting procedure.

Policy-improvement procedure

Suppose that we have solved for policy z the system of $2M+t(0)-s(0)-3$ linear equations as described in Section 3.3. For the obtained solution, denote by $g(z)$ the average cost of policy z and denote by $v(z;x)$ the relative value for state x (as already noted, once we have solved the embedded system of equations described in Section 3.3 any required $v(z;x)$ follows immediately from one of the relations (3.12)–(3.14)). Since we want to obtain a policy $z' \in C$, we have to apply the policy-improvement procedure of the modified policy-iteration algorithm given in Section 5 of [2]. Before doing this, we note that for any state (i,s) with $0 \leqq i \leqq M-1$ and any decision $d \in D((i,s))$ (cf. definition

(13) in [2] and Section 3.2),

$$v(d\,.\,z\,;(i,s)) = K(s,d) + \psi_i(d) - k_0((i,s)) + g(z)t_0((i,s))$$

where

$$\psi_i(d) = \begin{cases} k_0((i,d)) - g(z)t_0((i,d)) + v(z\,;(i,d)) & \text{for} \quad i \geqq 1, \\ k_0((\bar{0},\bar{d})) - g(z)t_0((\bar{0},\bar{d})) + v(z\,;(\bar{0},\bar{d})) & \text{for} \quad i = 0. \end{cases}$$

Further, we recall that in the policy-improvement procedure any intervention prescribed by policy z cannot be replaced by the null-decision but only by another intervention. Since in the states $(0,s)$, $0 \leqq s \leqq c$ the null-decision is not feasible as opposed to the states (i,s) with $i \geqq 1$, the two cases have to be considered in a slightly different way.

Fix first $1 \leqq i \leqq M-1$. Define d_i^* and d_i^{**} as the smallest and the largest integer for which $K(0,d) + \psi_i(d)$ and $K(c,d) + \psi_i(d)$ are minimal on the interval $[s(i)+1, t(i)-1]$. Observe that d_i^* and d_i^{**} minimize $v(d\,.\,z\,;(i,0))$ and $v(d\,.\,z\,;(i,c))$ for $s(i) < d < t(i)$. It is straightforward to verify that $d_i^* \leqq d_i^{**}$. By the same reasoning as on p. 258 in Sobel [9], we find that, for all $0 \leqq s \leqq d_i^*$, the number d_i^* minimizes $K(s,d) + \psi_i(d)$ and hence $v(d\,.\,z\,;(i,s))$ for $s(i) < d < t(i)$. Hence, for all $0 \leqq s \leqq d_i^*$,

$$v(d_i^*\,.\,z\,;(i,s)) = \min_{s(i)<d<t(i)} v(d\,.\,z\,;(i,s)) \leqq v(z\,;(i,s)), \tag{3.16}$$

where the latter inequality follows from the fact that $v(d\,.\,z\,;x) = v(z\,;x)$ for $d = z(x)$. Similarly, we have for all $d_i^{**} \leqq s \leqq c$,

$$v(d_i^{**}\,.\,z\,;(i,s)) = \min_{s(i)<d<t(i)} v(d\,.\,z\,;(i,s)) \leqq v(z\,;(i,s)). \tag{3.17}$$

For $i = 0$ we determine the numbers d_0^* and d_0^{**} in the same way as above except that we now take $[0,c]$ as the minimization interval instead of $[s(i)+1, t(i)-1]$. Similar properties hold for d_0^* and d_0^{**} as for d_i^* and d_i^{**}.

It now follows that we obtain policy $z' \in C$ by taking $s'(i) = d_i^* - 1$, $S''(i) = d_i^*$, $t'(i) = d_i^{**} + 1$ and $T'(i) = d_i^{**}$ for $0 \leqq i \leqq M-1$.

The cutting procedure

Suppose we have performed part (b) of the algorithm and obtained policy z'. In addition we have obtained the function $v(z'(x)\,.\,z\,;x)$ for $x \in A_{z'}$. For ease of notation, we write $\tilde{v}(x) = v(z'(x)\,.\,z\,;x)$ for $x \in A_{z'}$.

For the natural process with a cost of $\tilde{v}(y)$ for stopping at state $y \in A_{z'}$ we shall now determine a set A with $A_0 \subseteq A \subseteq A_{z'}$ such that (a) the set A is a stopping set at least as good as the set $A_{z'}$ for each initial state $x \in A_{z'}$ (in fact this is trivially met for $x \in A$, so that verification is only needed for $x \in A_{z'} \backslash A$), (b) $A = A_{z''}$ for some $z'' \in C$. This will be done according to the principle outlined in Remark 4 of [2]. For a properly chosen sequence of states $x \in A_{z'}$ with $x \notin A_0$, we shall verify whether $A_{z'} \backslash \{x\}$ is a better stopping set than $A_{z'}$ or not for the natural process starting from state x. Next the intersection of all those sets which are better stopping sets will give the desired set A. Before we demonstrate how this principle can be developed into a simple procedure in our queueing problem, we first evaluate for $x = (i, s) \in A_{z'}$ the quantity $Q'_{is} = E\bar{v}(S_x)$, where S_x is the first entrance state of the natural process into the set $A_{z'} \backslash \{x\}$ when the initial state is x, cf. definition (18) in [2]. Consider first the case where $x = (i, s)$ with $s \leqq s'(i)$. Then the possible realizations of S_x are the states $(i+1, s)$ and $(i-1, s)$ if $s \leqq s'(i-1)$ and the states $(i+1, s)$ and $(L'(s), s)$ if $s > s'(i-1)$ where $L'(s)$ is defined by (3.11) with z replaced by z'. Using the definition of the absorption probability p given in Section 3.2, we find for state (i, s) with $s \leqq s'(i)$,

$$Q'_{is} = \begin{cases} [\lambda + \mu \min(i, s)]^{-1}[\lambda \bar{v}((i+1, s)) + \mu \min(i, s) \bar{v}((i-1, s))], \\ \hfill s \leqq s'(i-1), \\ p(i, s, L'(s), i+1)\bar{v}((i+1, s)) + \{1 - p(i, s, L'(s), i+1)\}\bar{v}((L'(s), s)), \\ \hfill s > s'(i-1). \end{cases}$$

Similarly, for state (i, s) with $s \geqq t'(i)$ we find

$$Q'_{is} = \begin{cases} [\lambda + \mu \min(i, s)]^{-1}[\lambda \bar{v}((i+1, s)) + \mu \min(i, s) \bar{v}((i-1, s))], \\ \hfill s \geqq t'(i+1), \\ p(i, s, i-1, R'(s))\bar{v}((R'(s), s)) + \{1 - p(i, s, i-1, R'(s))\}\bar{v}((i-1, s)), \\ \hfill s < t'(i+1), \end{cases}$$

where $R'(s)$ is defined by (3.11) with z replaced by z' and $p(\cdot, c, \cdot, \cdot) = 0$.

We can now describe the determination of the parameters $s''(i)$, $S''(i)$, $t''(i)$ and $T''(i)$ of policy $z'' \in C$. Recall that in the cutting procedure any intervention prescribed by policy z' cannot be replaced by a different intervention but only by the null-decision. Consequently the states $(0, s)$ for $0 \leqq s \leqq c$ need not to be considered in this procedure. Further, we have $S''(i) = S'(i)$, $T''(i) = T'(i)$, $s''(i) \leqq s'(i)$ and $t''(i) \geqq t'(i)$ for all i with $s''(0) = s'(0)$ and $t''(0) = t'(0)$. We determine the numbers $s''(i)$ for $i \geqq 1$ by calculating successively $s''(1), \cdots, s''(M-1)$ in the following way. For $i = 1, \cdots, M-1$, let $s''(i)$ be the largest value of s with $\max(0, s''(i-1)) \leqq s \leqq s'(i)$ such that $Q'_{is} \geqq \bar{v}((i, s))$ if such a value of s exists, otherwise let $s''(i) = s''(i-1)$. The numbers $t''(i)$ for $i \geqq 1$ are determined by calculating successively $t''(M-1), \cdots, t''(1)$. Let $t''(M) = c+1$. For $i = M-1, \cdots, 1$, let $t''(i)$ be the smallest value of s with $t'(i) \leqq s \leqq \min(c, t''(i+1))$ such that $Q'_{is} \geqq \tilde{v}((i,s))$ if such a value of s exists, otherwise let $t''(i) = t''(i+1)$. In this way we obtain a policy $z'' \in C$.

Remark 1. In any iteration step the above policy-improvement procedure yields a policy $z' \in C$ having the additional property that $S'(i) = s'(i)+1$ and $T'(i) = t'(i) - 1$ for all i. However, except for the final iteration step, the cutting procedure by its very design may generate policies in C without this property.

Remark 2. The above algorithm needs only a minor modification in order to locate an optimal policy among the class C of policies in case of general switch-over costs with the separability property $K(a,b) = k^+(b) + b^+(a)$ for $b > a$, $K(a,b) = k^-(b) + b^-(a)$ for $b < a$ and $K(a,b) = 0$ for $b = a$, where $k^+(\cdot)$, $k^-(\cdot)$, $b^+(\cdot)$ and $b^-(\cdot)$ are non-negative, $k^+(\cdot)$ is non-decreasing and k^- is non-increasing. Observe that this function $K(a,b)$ includes the case where the switch-over costs consist of a fixed adjustment cost plus linear costs as above. In order to apply the algorithm, only the policy-improvement part needs a slight modification. For all $i \geqq 0$ we determine the numbers d_i^* and d_i^{**} as before. We again find $d_i^* \leqq d_i^{**}$ for all i. However, we now find for $i \geqq 1$ that the relations (3.16) and (3.17) only hold for $0 \leqq s \leqq s(i)$ and $t(i) \leqq s \leqq c$, respectively. The parameters of the new policy z' are now obtained as follows. We choose $S'(i) = d_i^*$ and $T'(i) = d_i^{**}$ for all $i \geqq 0$ as before. The numbers $s'(i)$ are determined by calculating successively $s'(M-1), \cdots, s'(0)$. For $i = M-1, \cdots, 0$, let $s'(i)+1$ be the smallest value of s with

$$\begin{cases} s(i)+1 \leqq s \leqq \min(S'(i)-1, s(i+1)) & \text{if} \quad i \geqq 1 \\ 0 \leqq s \leqq \min(S'(0)-1, s(1)) & \text{if} \quad i = 0 \end{cases}$$

such that $v(S'(i).z;(i,s)) \geqq v(z;(i,s))$ if such a value of s exists, otherwise let $s'(i) = \min(S'(i)-1, s(i+1))$. The numbers $t'(i)$ are determined by calculating successively $t'(0), \cdots, t'(M-1)$. Let $t'(-1) = 0$. For $i = 0, \cdots, M-1$, let $t'(i)-1$ be the largest value of s with

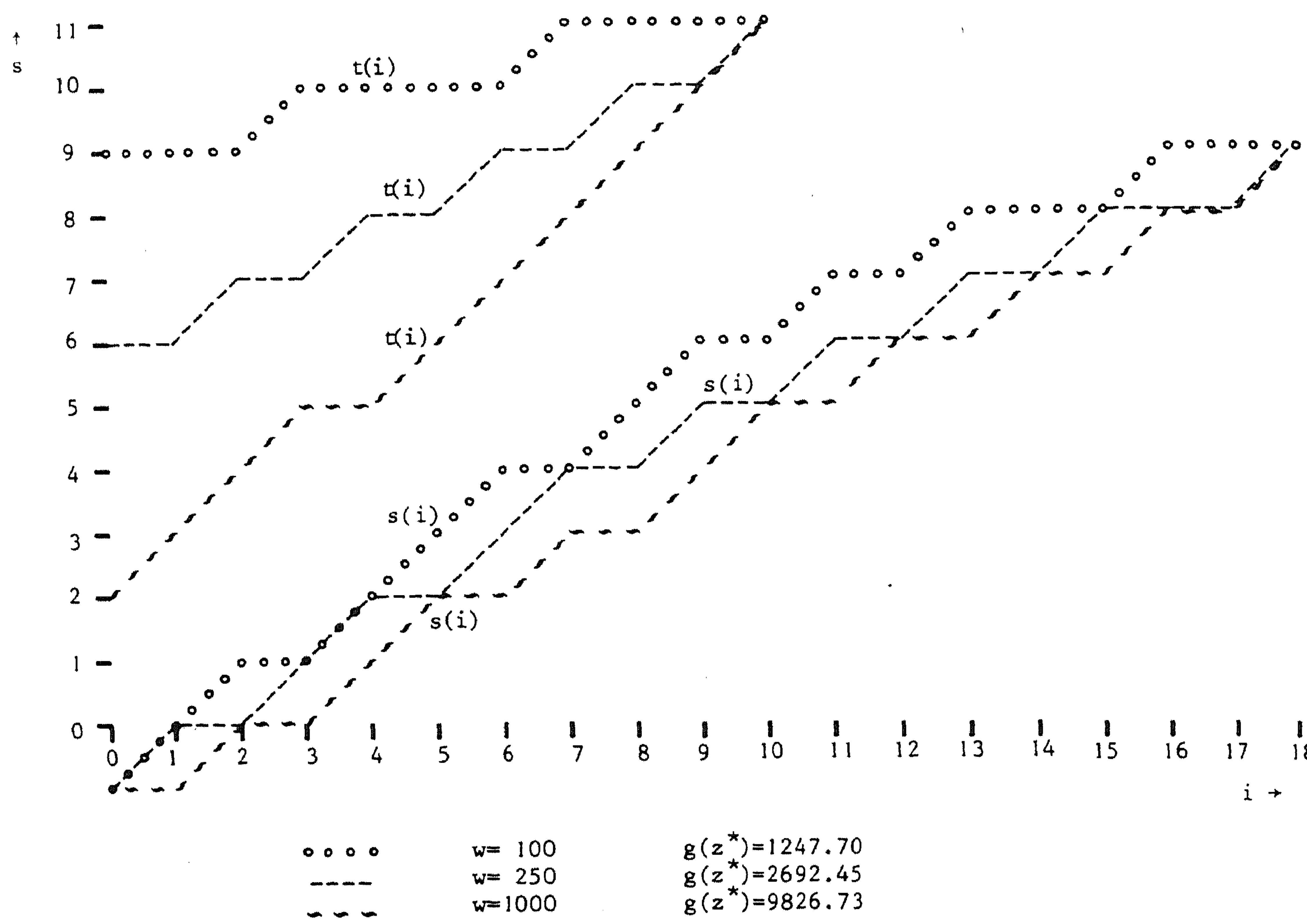
$$\begin{cases} \max(t'(-1), T'(0)+1) \leqq s \leqq c & \text{if } i = 0 \\ \max(t'(i-1), T'(i)+1) \leqq s \leqq t(i)-1 & \text{if } i \geqq 1 \end{cases}$$

such that $v(T'(i).z;(i,s)) \geqq v(z;(i,s))$ if such a value of s exists, otherwise let $t'(i) = \max(t'(i-1), T'(i)+1)$.

We were not able to show that the algorithm converges in a finite number of iteration steps to an optimal policy, although any step yields an improved policy. However, convergence appeared in all examples tested. After convergence of

TABLE 2.

$c = 10,\ \lambda = 9.5,\ \mu = 1,\ h = 10,\ k^{+} = k^{-} = 100.$



the algorithm to a policy z^* (say) we checked a criterion guaranteeing that policy z^* is optimal among the class of all stationary policies when this criterion is satisfied. This criterion is based on Theorem 8 in [2] and requires the verification that (a) $v(d.z^*;(i,s)) \geqq v(z^*;(i,s))$ for all (i,s) and all $d \in D((i,s))$, and (b) $Q^*_{is} \geqq v(z^*;(i,s))$ for all $(i,s) \in A_{z^*}$ with $1 \leqq i \leqq M-1$ where Q^*_{is} is defined as Q'_{is} above with z' replaced by z^*.

In all examples tested this criterion was satisfied and, consequently, an optimal policy was found.

In Table 2 we give for a number of numerical examples the minimal average cost $g(z^*)$ and optimal values for $s(i)$, $S(i)$, $t(i)$ and $T(i)$ where $S(i)$ and $T(i)$ are given by $S(i) = s(i) + 1$ and $T(i) = t(i) - 1$.

Acknowledgement

We are indebted to Mr. Rob van der Horst for most valuable help in computer programming.

References

[1] Bell, C. E. (1975) Turning off a server with customers present: is this any way to run an $M/M/c$ queue with removable servers? *Opns. Res.* **23**, 571–574.

[2] De Leve, G., Federgruen, A. and Tijms, H. C. (1977) A general Markov decision method, I: model and techniques. *Adv. Appl. Prob.* **9**,

[3] Feller, W. (1966) *An Introduction to Probability Theory and its Applications*, Vol. 2. Wiley, New York.

[4] Howard, R. A. (1971) *Dynamic Probabilistic Systems, Vol. II: Semi-Markov and Decision Processes.* Wiley, New York.

[5] Lippman, S. A. (1973) A new technique in the optimization of exponential queueing systems. Working Paper No. 211, Western Management Science Institute, University of California, Los Angeles.

[6] McGill, J. T. (1969) Optimal control of a queueing system with variable number of servers. Technical Report No. 2, Department of Operations Research, Stanford University.

[7] Miller, K. S. (1968) *Linear Difference Equations.* W. A. Benjamin Inc., New York.

[8] Robin, M. (1975) Contrôle optimal de files d'attente. Rapport de Recherche No. 117, IRIA, Rocqencourt, France.

[9] Sobel, M. J. (1974) Optimal operation of queues. In *Mathematical Methods in Queueing Theory*, ed. A. B. Clarke. Lecture Notes in Economics and Mathematical Systems **98**, Springer-Verlag, Berlin.

[10] Sobel, M. J. (1971) Production smoothing with stochastic demand II: infinite-horizon case. *Management Sci.* **17**, 724–735.

[11] Van Domselaar, V. and Hemker, P. W. (1975) Nonlinear parameter estimation in initial value problems. Mathematical Centre Report NW 18/75, Mathematisch Centrum, Amsterdam.