stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 36/74 AUGUST

P.J. WEEDA

ON THE RELATIONSHIP BETWEEN THE CUTTING OPERATION OF GENERALIZED MARKOV PROGRAMMING AND OPTIMAL STOPPING

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.

ON THE RELATIONSHIP BETWEEN THE CUTTING OPERATION OF GENERALIZED MARKOV PROGRAMMING AND OPTIMAL STOPPING.

by

P.J.Weeda

ABSTRACT

The principles of generalized Markov programming were developed by DE LEVE [4] to solve continuous time Markov decision problems under the long run average return criterion. Here we investigate the generalized Markov decision model that arises if the natural process is given by a finite state semi Markov process and interventions are restricted to the points of time just after a state transition.
The iteration method induced by the general iteration scheme of DE LEVE for this special model distinguishes three operations at each iteration step which are called respectively: the value determination-, the policy improvement - and the cutting operation. The first two are related to similar operations in the iteration methods of HOWARD [2] and JEWELL [3] and are directly amenable for computation. This however is not true for the third one. In this report the relationship between the cutting operation and optimal stopping for this special model is stated and proved. This relationship yields a useful algorithm for this operation.

KEY WORDS AND PHRASES: *Generalized Markov programming. Finite state Markov decision problems. Cutting operation. Optimal stopping.*

INTRODUCTION

In generalized Markovian decision processes, DE LEVE [4], the state of the system is described by a point in a finite dimensional Cartesian space at each point of time. For each initial state the evolution of the state of the system is assumed to be described by a homogeneous strong Markov process, called the *natural process*. The decisionmaker may interrupt the natural process in each state by an *intervention* which implies an instantaneous (possibly random) change of the state of the system. In each state the decisionmaker has a set of feasible interventions at his disposal, which may be uncountable. The only alternative to interventions is to leave the natural process untouched. This alternative is called the *nulldecision* in that state. With the exception of a nonempty subset of states, the nulldecision is feasible in each state. After an intervention the evolution of the system is again described by the natural process until the next intervention is effectuated. It is assumed that at most a finite number of interventions is taken in each finite timeperiod. Also a general iteration scheme, to be called here generalized Markov programming, is presented in DE LEVE [4]. It is proved there, that this scheme converges to a strategy which is optimal with respect to the class of stationary deterministic strategies in an infinite number of iteration steps. The optimality-criterion is to maximize the expected average return per unit of time in the long run. Some applications of the method are presented in DE LEVE, TIJMS & WEEDA [5].

In this paper we consider the special model that arises if the natural process is given by a *finite state semi Markov process* and the decisionmaker is only allowed to intervene at the points of time a state transition in the natural process has just occurred. The iteration method induced by the general iteration scheme for this special model is formulated. In agreement with the general scheme this iteration method distinguishes three operations per iteration step: the value determination -, the policy improvement - and the cutting operation. The iteration method for this model has the pleasant property of convergence within a finite number of steps. The attention in this paper is focused on the cutting operation. New is the relation between the cutting operation and *optimal stopping* which is stated and proved for the special model. This relation yields a method

which is directly amenable for computation and can be generally applied to problems satisfying this special model. It is hoped that the results will be useful in developing efficient methods for this cutting operation in the more general type of Markov decision problems covered by the iteration scheme of DE LEVE.

THE MODEL

Natural process

The natural process of this generalized Markov decision model is supposed to be given by a finite state semi Markov process. In a finite state semi Markov process the system makes random transitions among a finite number of states. Let J denote the set of states. If a transition to some state $i \in J$ has just occurred at time t, the system remains in state i until the next transition to a random state $\underline{j}$ *) occurs at a random time $t + \underline{\tau}_i$ where $\underline{\tau}_i$ is the sojourn time in state i. Sufficient information for our purposes about the behaviour of the process is provided by the triple (Q, u, h) where Q denotes the $|J|\times|J|$- matrix of transition probabilities q_{ij}, $i, j \in J$, satisfying $0 \leq q_{ij} \leq 1$ and $\sum_{j\in J} q_{ij} = 1$; $u>0$ denotes the $|J|$- dimensional vector of expected sojourn times and h denotes the $|J|$-dimensional vector with elements h_i $(-\infty<h_i<\infty)$ representing the expected return of the process during the sojourn time in state i including the transition to the next state.

Interventions and nulldecisions

In each state $i \in J$ the decisionmaker has a finite set of actions X(i) at his disposal consisting of interventions and at most one nulldecision, which is denoted by $x_0(i)$. The nulldecision leaves the state of the system unchanged, which implies here that the natural process remains untouched during the sojourn time in that state including the next state transition. The nulldecision satisfies

$$x_0(i) \notin X(i) \qquad \text{for } i \in A_0$$

*) Random variables are underlined.

where A_0 is a nonempty subset of states. Further A_0 and the matrix Q have to satisfy the requirement that the inverse exists of the matrix $(I-Q)_{\bar{A}_0}$ with entries $\delta_{ij}-q_{ij}$ for $i, j \in \bar{A}_0$, with δ_{ij} satisfying $\delta_{ii}=1$ and $\delta_{ij}=0$ for $j \neq i$. To each intervention $x \in X(i)$ is associated a probability distribution $p_{im}(x)$ of the state $\underline{m}$ into which the intervention leads and an expected cost $g_i(x)$. If the system assumes state $\underline{m} = m$ after an intervention then it remains in state m until the next transition in the natural process has occurred. The sojourn time in state m has expectation $u_m = E\,\underline{\tau}_m$. By the foregoing the nulldecision can be viewed as an intervention satisfying

$$p_{im}(x_0(i)) = \begin{cases} 1 & \text{if } i=m \\ 0 & \text{otherwise} \end{cases}$$

and

$$g_i(x_0(i)) = 0$$

Strategies

A stationary deterministic strategy Z makes use of the same action $Z(i) \in X(i)$ each time a transition to state i has just occurred. By a strategy of this type the state space is dichotomized into a set A_Z defined by

$$A_Z := \{i \in J : Z(i) \neq x_0(i)\}$$

and its complement. The definitions of A_0 and A_Z imply

(1) $$A_Z \supseteq A_0$$

THE ITERATION METHOD

Preliminary computations

---

*) Random variables are underlined

Compute:

a. The $|J|$- dimensional vector k_0 defined by

$$(k_0)_{\bar{A}_0} := (I - Q)^{-1}_{\bar{A}_0} (h)_{\bar{A}_0}$$

$$(k_0)_{A_0} := 0.$$

b. The $|J|$- dimensional vector t_0 defined by

$$(t_0)_{\bar{A}_0} := (I - Q)^{-1}_{\bar{A}_0} (u)_{\bar{A}_0}$$

$$(t_0)_{A_0} := 0.$$

c. The numbers $k(i,x)$ defined for each $x \in X(i)$ and $i \in J$ by

$$k(i,x) := -g_i(x) + \sum_{m \in J} p_{im}(x)\, k_0(m) - k_0(i).$$

d. The numbers $t(i,x)$ defined for each $x \in X(i)$ and $i \in J$ by

$$t(i,x) := \sum_{m \in J} p_{im}(x)\, t_0(m) - t_0(i).$$

The interpretation of the vectors k_0 and t_0 is as follows: Each element $k_0(i)$ ($t_0(i)$) represents the expected return (expected time elapsed) in the natural process with initial state $i \in \bar{A}_0$ until the first state in A_0 is assumed. The elements $k_0(i)$ ($t_0(i)$) for $i \in A_0$ vanish. The numbers $k(i,x)$ ($t(i,x)$) represent the difference in expected return (expected duration) between two stochastic walks. The first walk applies action $x \in X(i)$ in initial state i and is subsequently described by the natural process until the first state in the set A_0 is taken on. The second walk is completely described by the natural process from initial state i until the first state in A_0 is taken on. The definitions of $k(i,x)$ and $t(i,x)$ imply $k(i,x_0(i)) = t(i,x_0(i)) = 0$.

After these preliminary computations the iteration cycle is entered

with an arbitrarily chosen initial strategy. During each iteration step the following three operations are executed.

Value determination operation

Compute

a. The $|A_Z|$- dimensional vector k(Z) with elements k(i, Z(i)), $i \in A_Z$.

b. The $|A_Z|$- dimensional vector t(Z) with elements t(i, Z(i)), $i \in A_Z$.

c. The $|\bar{A}_Z| \times |A_Z|$- matrix $S(A_Z)$ defined by

$$S(A_Z) := (I - Q)^{-1}_{\bar{A}_Z} (Q)_{\bar{A}_Z A_Z}$$

where $(Q)_{\bar{A}_Z A_Z}$ denotes the $|\bar{A}_Z| \times |A_Z|$- matrix with entries q_{ij}, $i \in \bar{A}_Z$, $j \in A_Z$. The existence of the matrix $(I - Q)^{-1}_{\bar{A}_Z}$ is implied by the existence of $(I - Q)^{-1}_{\bar{A}_0}$ and relation (1).

d. The $|A_Z| \times |A_Z|$- matrix R(Z) defined by

$$R(Z) := P(Z)\, S(A_Z)$$

where P(Z) denotes the $|A_Z| \times |\bar{A}_Z|$- matrix with entries $p_{im}(Z(i))$, $i \in A_Z$, $m \in \bar{A}_Z$ *). R(Z) is the matrix of transition probabilities of the imbedded process defined by the states $i \in A_Z$.

*) It is assumed in generalized Markov programming that $p_{im}(Z(i))=0$ for $i, m \in A_Z$ for each stationary deterministic strategy Z.

e. The subvectors $(y(Z))_{A_Z}$ and $(v(Z))_{A_Z}$ by solving the following set of equations

$$(y(Z))_{A_Z} = R(Z)\ (y(Z))_{A_Z}$$

$$(v(Z))_{A_Z} = k(Z) - (y(Z))_{A_Z} \square\ t(Z) + R(Z)\ (v(Z))_{A_Z}$$

where the notations $a \square b$ stands for the vector with elements $a_i\ b_i$

A unique solution to this set is obtained by choosing in each ergodic set $K(\ell)$, $\ell=1,\ldots,L(Z)$ of the imbedded process an arbitrary state $i(\ell) \in K(\ell)$ for which we put $v_{i(\ell)}(Z) = 0$, $\ell=1,\ldots,L(Z)$.

f. The subvectors $(y(Z))_{\bar{A}_Z}$ and $(v(Z))_{\bar{A}_Z}$ from

$$(y(Z))_{\bar{A}_Z} = S(A_Z)\ (y(Z))_{A_Z}$$

$$(v(Z))_{\bar{A}_Z} = S(A_Z)\ (v(Z))_{A_Z}$$

Policy improvement operation

Compute

a. The $|J|$- dimensional vector y' with elements y'_i, $i \in J$ defined by

$$y'_i := \max_{x \in X(i)} \left[\sum_{j \in J} p_{ij}(x)\ y_j(Z)\right]$$

b. The subset $X_1(i)$ of $X(i)$ defined by

$$X_1(i) := \{x \in X(i) : \sum_{j \in J} p_{ij}(x)\ y_j(Z) = y'_i\}$$

c. The $|J|$- dimensional vector v' with elements v_i', $i \in J$ defined by

$$v_i' := \max_{x \in X_1(i)} [k(i,x) - y_i' \, t(i,x) + \sum_{j \in J} p_{ij}(x) \, v_j(Z)]$$

d. The subset $X_2(i)$ of $X_1(i)$ defined by

$$X_2(i) := \{x \in X_1(i) : k(i,x) - y_i' \, t(i,x) + \sum_{j \in J} p_{ij}(x) \, v_j(Z) = v_i'\}.$$

e. Strategy Z' defined by the following rule: Take $Z'(i) = Z(i)$ if $Z(i) \in X_2(i)$; otherwise take $Z'(i)$ equal to an arbitrary action from $X_2(i)$.

We note that at the computation of y' the nulldecision for a state $i \in A_Z \cap \bar{A}_0$ yields

$$\sum_{j \in J} p_{ij} \, (x_0(i)) \, y_j(Z) = y_i(Z)$$

while the intervention Z(i) yields

$$\sum_{j \in J} p_{ij} \, (Z(i)) \, y_j(Z) = y_i(Z).$$

The same holds at the computation of v'. Because the policy improvement operation implies $Z'(i) = Z(i)$ if $y_i' = y_i(Z)$ and $v_i' = v_i(Z)$ we conclude that in any case $Z'(i) \neq x_0(i)$ for $i \in A_Z$ or equivalently

$$A_{Z'} \supseteq A_Z \tag{2}$$

Cutting operation

Let A be an arbitrary set of states satisfying $A_0 \subseteq A \subseteq A_{Z'}$. Define the $|J|$- dimensional vectors y"(A) and v"(A) respectively by

$$\begin{cases} (y''(A))_{\bar{A}} := S(A) \, (y')_A \\ (y''(A))_A := (y')_A \end{cases} \tag{3}$$

and

$$(4)\qquad \begin{cases} (v''(A))_{\bar{A}} := S(A)(v')_A \\ (v''(A))_A := (v')_A \end{cases}$$

Let M be the collection of sets A satisfying either $y_i''(A) > y_i'$ or $y_i'(A) = y_i'$ and $v_i''(A) \geq v_i'$ for each $i \in A_{Z'}$.

Compute:

a. The set A* defined by

$$A^* := \bigcap_{A \in M} A$$

b. The strategy Z" defined by

$$Z''(i) := \begin{cases} Z'(i) & \text{for } i \in A^* \\ x_0(i) & \text{for } i \in \bar{A}^*. \end{cases}$$

If Z" = Z then the iteration cycle is terminated. Otherwise the value determination operation is reentered with Z := Z".

The following lemma is implied by a result of DE LEVE (see [4], page 57, lemma 3.2)

LEMMA 1. *If* $A_1, A_2 \in M$ *are two subsets of states then*

$$A_1 \cap A_2 \in M.$$

The following corollary to lemma 1 is not true in the general model considered in DE LEVE [4].

COROLLARY 1.

$$A^* \in M.$$

PROOF. The assertion follows directly from lemma 1 and the fact that M contains a finite number of sets. □

In the next section it will be shown that the set A* of the cutting operation is identical to the solution of the second of a sequence of two optimal stopping problems. The numerical solutions of these two optimal stopping problems are easily obtained by a specialized version of the policy iteration method of HOWARD [2].

THE CUTTING OPERATION AND OPTIMAL STOPPING

In this section we state and prove the relationship between the cutting operation of the preceding section and optimal stopping in a finite Markov chain. Primarily optimal stopping is reviewed.

Suppose that a finite Markov chain with set of states J is given. In each state $i \in J$ at most two actions x_0 and x_1 are feasible. If action x_0 is applied in state i then the original chain is continued at least until the next transition has occurred. If action x_1 is applied in state i then the chain is stopped and a return w_i is obtained. An optimal stopping problem in a finite Markov chain is completely defined by the quadruple (A_s, A_c, Q, w) where A_s is the (nonempty) subset of states in which only action x_1 is feasible; A_c is the (possible empty) set of states satisfying $\bar{A}_c \supset A_s$ and containing all the states in which only action x_0 is feasible; Q is the matrix of transition probabilities q_{ij} of the original chain and w is an $|\bar{A}_c|$- dimensional vector with elements $-\infty < w_i < \infty$. The matrix Q and the set A_s are required to imply the existence of the matrix $(I - Q)^{-1}_{\bar{A}_s}$.

The optimal stopping problem defined above can be considered as a finite state Markov decision problem if action x_1 in each state $i \in \bar{A}_c$ is interpreted as to make i an absorbing state with a return w_i received at each transition $i \to i$. Because a stationary deterministic strategy is optimal for a finite state Markov decision problem (see DERMAN [1] by example) the computation of an optimal strategy can be restricted to the class Z of this special type of strategies. Each strategy $Z \in Z$ in an optimal stopping problem dichotomizes the set of states J into a *feasible stopping set* B

defined by

$$B := \{i \in J : Z(i) = x_1\}$$

and its complement. Clearly there exists a 1-1 correpondence between the collection of feasible stopping sets $A_s \subseteq B \subseteq \bar{A}_c$ and the class of strategies Z. To each feasible stopping set B an *expected return vector* f(B) is associated, whose elements $f_i(B)$ represent the expected return for each initial state $i \in J$. The vector f(B) is calculated by solving the following set of equations

$$(5) \qquad \begin{aligned} (f(B))_B &= (w)_B \\ (f(B))_{\bar{B}} &= (Q)_{\bar{B}} (f(B)_{\bar{B}} + (Q)_{\bar{B}B} (f(B))_B \end{aligned}$$

The set (5) possesses a unique solution because the existence of $(I - Q)_{\bar{B}}^{-1}$ is implied by the existence of $(I - Q)_{\bar{A}_s}^{-1}$ for each set B satisfying $A_s \subseteq B \subset J$. If we write S(B) for the $|\bar{B}| \times |B|$-matrix $(I - Q)_{\bar{B}}^{-1} (Q)_{\bar{B}B}$ then the solution of (5) is given by

$$(6) \qquad \begin{aligned} (f(B))_B &= (w)_B \\ (f(B))_{\bar{B}} &= S(B)\ (w)_B \end{aligned}$$

An *optimal stopping set* (notation : B_m) satisfies for each feasible stopping set B

$$(7) \qquad f(B_m) \geq f(B)$$

An optimal stopping set can be calculated by a specialized version of the policy iteration method of HOWARD [2]. The iteration starts with an arbitrary feasible stopping set (strategy). At each step the following two operations are executed:

1. *Value determination operation*

Let B be the feasible stopping set (strategy Z) obtained at the preceding step. Solve the set of equations (5) in f(B).

2. *Policy improvement operation*

Compute:

a. The $|J|$- dimensional vector f' with elements f_i' defined by

$$f_i' := \begin{cases} \max [w_i, \sum_{j \in J} q_{ij} f_j(B)] & \text{for } i \in \bar{A}_s \cap \bar{A}_c \\ f_i(B) & \text{for } i \in A_s \cup A_c. \end{cases}$$

b. The feasible stopping set B' (strategy Z') by taking

$$Z'(i) \neq Z(i) \qquad \text{if } f_i' > f_i(B)$$

$$Z'(i) = Z(i) \qquad \text{if } f_i' = f_i(B)$$

These two operations are repeated until B' = B. This identity is obtained within a finite number of steps and implies the optimality of the feasible stopping set satisfying B' = B. The proofs of HOWARD and others imply the following lemma for an optimal stopping problem in a finite Markov chain.

LEMMA 2. *If* B *and* B' *are two feasible stopping sets, obtained at two successive steps of the policy iteration algorithm above, then we have either*

$$f' > f(B) \Rightarrow f(B') > f(B)$$

or

$$B' = B \Leftrightarrow B \text{ is optimal}$$

By the policy improvement operation we have that for an optimal stopping set (notation : B_m) $f_i(B_m)$ satisfies

$$(8) \qquad \begin{aligned} f_i(B_m) &= \sum_{j \in J} q_{ij} f_j(B_m) \geq w_i && \text{for } i \in \bar{B}_m \cap \bar{A}_c \\ f_i(B_m) &= w_i \geq \sum_{j \in J} q_{ij} f_j(B_m) && \text{for } i \in B_m \cap \bar{A}_s. \end{aligned}$$

Define:

$$(9) \qquad B_m^s := B_m \setminus \{ i \in B_m \cap \bar{A}_s : \sum_{j \in J} q_{ij} f_j(B_m) = w_i \}$$

and

$$(10) \qquad B_m^{\ell} := B_m \cup \{ i \in \bar{B}_m \cap \bar{A}_c : \sum_{j \in J} q_{ij} f_j(B_m) = w_i \}$$

The following lemma specifies the collection of optimal stopping sets.

LEMMA 3.

(a) *The feasible stopping sets* B_m^s *and* B_m^{ℓ} *satisfy* $f(B_m^s) = f(B_m^{\ell}) = f(B_m)$.

(b) *Each optimal stopping set* A *satisfies* $B_m^s \subseteq A \subseteq B_m^{\ell}$.

PROOF.

(a) By definition $f(B_m^s)$ satisfies (5). By (8) and (9) $f(B_m)$ satisfies

$$f_i(B_m) = \sum_{j \in J} q_{ij} f_j(B_m) \qquad \text{for } i \in \overline{B_m^s}$$

$$f_i(B_m) = w_i \qquad \text{for } i \in B_m^s$$

Because the solution to (5) is unique we have $f(B_m^s) = f(B_m)$.
By a similar argument: $f(B_m^{\ell}) = f(B_m)$.

(b) Relation (8) and the definitions (9) and (10) imply that the sets $\bar{A}_s \cap B_m^s$ and $\bar{A}_c \cap B_m^{\ell}$ are disjunct. Hence

$$(11) \qquad B_m^s = \{i \in \bar{A}_c \cap \bar{A}_s : \sum_{j \in J} q_{ij} f_j(B_m) < w_i \} \cup A_s$$

and

$$(12) \qquad B_m^{\ell} = \{i \in \bar{A}_c \cap \bar{A}_s : \sum_{j \in J} q_{ij} f_j(B_m) \leq w_i \} \cup A_s$$

Because A is optimal, B_m may be replaced by A in (8), (11) and (12). With this modification these relations imply $B_m^s \subseteq A \subseteq B_m^{\ell}$. □

In the sequel the expected return vector of a feasible stopping set B to $(A_0, \bar{A}_{Z'}, Q, y')$ will be denoted by $y''(B)$ in agreement with its definition (3) and relation (6). The vector $v''(B)$ represents the same for a feasible stopping set B to $(B_m^s(y'), B_m^{\ell}(y'), Q, v')$. At this point we are able to state the algorithm to compute A* based upon optimal stopping.

An algorithm for the cutting operation

Compute:

a. An optimal stopping set to $(A_0, \bar{A}_{Z'}, Q, y')$ (notation: $B_m(y')$) by the method of HOWARD.

b. The sets $B_m^s(y')$ and $B_m^{\ell}(y')$ defined respectively by

$$B_m^s(y') := B_m(y') \setminus \{i \in B_m(y') \cap \bar{A}_0 : \sum_{j \in J} q_{ij}\, y_j''(B_m(y')) = y_i'\}$$

and

$$B_m^{\ell}(y') := B_m(y') \cup \{i \in \overline{B_m(y')} \cap A_{Z'} : \sum_{j \in J} q_{ij}\, y_j''(B_m(y')) = y_i'\}$$

c. An optimal stopping set to $(B_m^s(y'), \overline{B_m^{\ell}(y')}, Q, v')$ (notation: $B_m(v')$) by the method of HOWARD.

d. The set $B_m^s(v')$ defined by

$$B_m^s(v') := B_m(v') \setminus \{i \in B_m(v') \cap \overline{B_m^s(y')} : \sum_{j \in J} q_{ij}\, v_j''(B_m(y')) = v_i'\}$$

Next we prove two lemmas which are required to prove the main result (theorem 1), on which this cutting algorithm is based.

LEMMA 4. Let A *and* B, $A \supset B$, *be two feasible stopping sets to* $(A_0, \bar{A}_{Z'}, Q, y')$ *as well as to* $(A_0, \bar{A}_{Z'}, Q, v')$. *Let* $y_i''(B) > y_i'$ *for* $i \in A \cap \bar{B}$. *Then a state* $k \in \bar{A}$ *satisfying* $y_k''(B) = y_k''(A)$ *also satisfies* $v_k''(B) = v_k''(A)$.

PROOF. The assumptions $A \supset B$ and $y_k''(B) = y_k''(A)$ imply

$$y_k''(B) = \sum_{j \in A} s_{kj}(A)\, y_j''(B) = y_k''(A) = \sum_{j \in A} s_{kj}(A)\, y_j'$$

Because $y''_j(B) > y'_i$ for $i \in A \cap \bar{B}$ we have

$$(13) \qquad \sum_{j \in B} s_{kj}(A) = 1$$

Because $A \supset B$, (13) and $v''_i(B) = v''_i(A) = v'_i$ for $i \in B$ we have

$$v''_k(B) = \sum_{j \in A} s_{kj}(A)\, v''_j(B) = \sum_{j \in B} s_{kj}(A)\, v''_j(B) =$$

$$= \sum_{j \in B} s_{kj}(A)\, v'_j = \sum_{j \in A} s_{kj}(A)\, v'_j = v''_k(A). \qquad \square$$

LEMMA 5.

$$B^s_m(v') \in M.$$

PROOF. By the optimality of $B^s_m(v')$ to $(A_0, \overline{A_{Z'}}, Q, y')$ and by lemma 3 we have

$$y''_i(B^s_m(v')) > y'_i \qquad \text{for } i \in \overline{B^\ell_m(y')} \cap A_{Z'}$$

and

$$y''_i(B^s_m(v')) = y'_i \qquad \text{for } i \in B^\ell_m(y')$$

By the optimality of $B^s_m(v')$ to $(B^s_m(y'), \overline{B^\ell_m(y')}, Q, v')$ we have

$$v''_i(B^s_m(v')) \geq v'_i \qquad \text{for } i \in B^\ell_m(y')$$

These relations imply the assertion. $\square$

The main result is now proved.

THEOREM 1.

$$A^* = B^s_m(v').$$

PROOF. Suppose A^* is not optimal with respect to $(A_0, \overline{A_{Z'}}, Q, y')$ then the method of HOWARD entered with A^* would yield a stopping set B after one iteration step which would satisfy $y''(B) > y''(A^*)$ by lemma 2. Because $A^* \in M$

implies $y_i''(A^*) \geq y_i'$ for $i \in A_{Z'}$, we have $y_i''(B) \geq y_i''(A^*) \geq y_i'$ for $i \in A_{Z'}$. Also we have $B \subset A^*$, because otherwise there would be at least one state $i \in A_{Z'}$, satisfying $y_i''(A^*) = \sum_{j \in J} q_{ij}\, y_j''(A^*) < y_i'$ contradicting $A^* \in M$. For each state $k \in \bar{B} \cap A_{Z'}$ satisfying $y_k''(B) = y_k''(A^*)$ we have by lemma 4 $v_k''(B) = v_k''(A^*)$ and because $A^* \in M$: $v_k''(B) = v_k''(A^*) \geq v_k'$. For $i \in B$ we have $y_i''(B) = y_i'$ and $v_i''(B) = v_i'$. By these arguments $B \in M$. However, $B \subset A^*$ and $B \in M$ contradict the definition of A^*. Hence A^* is optimal to $(A_0, A_{Z'}, Q, y')$ and by lemma 3 we have $B_m^s(y') \subseteq A^* \subseteq B_m^{\ell}(y')$.

Now suppose that A^* is optimal to $(A_0, A_{Z'}, Q, y')$ as is proved but that A^* is not optimal to $(B_m^s(y'), B_m^{\ell}(y'), Q, v')$. Then the method of HOWARD applied to $(B_m^s(y'), B_m^{\ell}(y'), Q, v')$ and entered with A^* would yield a stopping set C after one iteration step satisfying $C \subset A^*$ by the same argument as used above for B. Lemma 2 implies now $v''(C) > v''(A^*)$ and the optimality of C to $(A_0, A_{Z'}, Q, y')$ implies $y''(C) = y''(A^*)$. Hence because $A^* \in M$ also $C \in M$. But $C \subset A^*$ and $C \in M$ contradict the definition of A^*. Hence A^* is optimal to $(B_m^s(y'), B_m^{\ell}(y'), Q, v')$ implying $A^* \supseteq B_m^s(v')$ by the definition of $B_m^s(v')$ and lemma 3b. On the other hand the definition of A^* and lemma 5 imply $A^* \subseteq B_m^s(v')$. So we have the identity $A^* = B_m^s(v')$. □

REFERENCES

[1] DERMAN, C., *Finite State Markovian Decision Processes*, Mathematics in Science and Engineering, Volume 67, Academic Press, New York and London, 1970.

[2] HOWARD, R.A., *Dynamic programming and Markov Processes*, M.I.T. Press, Cambridge, Massachusetts, 1960.

[3] JEWELL, W.S., *Markov-Renewal Programming, I: Formulation, Finite Return Models, II: Infinite Return Models*, Operations Research 10 (1963) 938-972.

[4] LEVE, G. DE, *Generalized Markovian Decision Processes, Part I: Model and Method, Part II: Probabilistic Background*, Mathematical Centre Tracts 3 and 4, Amsterdam, 1964.

[5] LEVE, G. De, TIJMS, H.C. and WEEDA, P.J., *Generalized Markovian Decision Processes, Applications*, Mathematical Centre Tracts 5, Amsterdam, 1970.