MATHEMATICAL CENTRE TRACTS 92

FINITE GENERALIZED MARKOV PROGRAMMING

P.J. WEEDA

MATHEMATISCH CENTRUM AMSTERDAM 1979

AMS(MOS) subject classification scheme (1970): 90C45

ISBN 90 6196 158 0

CONTENTS

ACKNOWLEDGEMENTS

At the completion of this monograph I wish to thank Prof.Dr. G. de Leve for his stimulating accompaniment and supervision of this research, Prof.Dr. J. Wessels for the many improvements of the text and my former colleagues Prof.Dr. A. Hordijk and Dr. H.C. Tijms for the stimulating conversations on Markov decision theory.

Moreover, I thank the Mathematical Centre for the opportunity to publish this monograph in their series Mathematical Centre Tracts and all those at the Mathematical Centre who have contributed to its realization.

INTRODUCTION

In the generalized Markovian decision model, developed in [DE LEVE 1964], a system is considered which has a *general state space* and can be controlled *continuously*. There is an underlying stochastic process, called the *natural process*. It describes the evolution of the state of the system during the course of time if the system is left uncontrolled. A *reward structure* is associated with the natural process. A decisionmaker controls the system by making *interventions*, chosen from an arbitrary set of feasible interventions in each state. An intervention induces a reward and an instantaneous (possibly random) change of the state of the system. At each point of time either an intervention is made or the natural process is left untouched. In the latter case we speak of a *null decision*. It is further assumed that only a finite number of interventions may be made in any finite time interval. In this model only *stationary* policies are considered. A stationary policy associates with each state either an intervention from the set of interventions in that state or the null decision. A policy iteration principle, hereafter called the *general method*, has been developed in [DE LEVE 1964] which is proven to approximate the *maximum expected average reward per unit of time* sufficiently good if the number of iterations is large enough.

Applications of the general method (cf. [DE LEVE, TIJMS & WEEDA 1970]) are versions of the automobile insurance problem and the anticipation problem in production control, continuous time inventory models and a continuous version of the automobile replacement problem of [HOWARD 1960]. Although the general method has a large number of potential applications, its use is obstructed by the fact that it is not a ready-made technique. The only way to solve a problem numerically is to "translate" the principles of the method in terms of the problem and develop numerical procedures by exploiting the special structure of the problem. A relatively simple numerical solution has been obtained for the automobile insurance problem. The mentioned version of the production control problem is much harder to solve. Although numerical solutions have been obtained, not much can be said about their quality so far.

In this monograph the general model and method are considered under much stronger assumptions. In the first place the natural process is a *finite state Markov renewal process* induced by a semi Markov matrix.

Secondly the sets of interventions are *finite* sets. Thirdly, interventions can be made only at the epochs of state transitions of the natural process.

On the one hand, such a specialization yields stronger results like finite step convergence. On the other, since this specialization of the general method is still not a ready-made technique, computational procedures have to be developed which are preferably general within this special model.

The thus obtained model is called the *finite generalized Markov programming model*. It is defined in section 2.2. Each problem which satisfies the assumptions of this model can also be formulated as a finite Markov renewal decision problem and vice versa. A policy iteration method, which is predominantly a direct specialization of the general method, is presented in section 2.3. This policy iteration method has four operations, a *preparatory part*, a *policy evaluation operation*, a *policy improvement operation* and a *cutting operation*. The preparatory part is executed once and the remaining three operations at each iteration step. The second and third operation are different compared with corresponding operations in the well-known methods described in [HOWARD 1960] and [JEWELL 1963]. The preparatory part and the fourth operation do not form part of any other existing policy iteration method.

In chapter IV numerical procedures, which are general within this special model, are developed for the cutting operation. In the general method the cutting operation is formulated in terms of its target set: the intersection of all sets that do not worsen in a lexicographical sense a pair of vectors defined for each set, with respect to a certain given pair. It is proven that the target set is an *optimal cutting set*, i.e. a set that maximizes the lexicographical improvement with respect to the given pair. The computational procedures obtaining an optimal cutting set are based on lexicographically optimal stopping in a Markov chain with two arbitrary income vectors. These methods may serve as a guidance for developing techniques for more general special models as in [DE LEVE, TIJMS & FEDERGRUEN 1977]. Moreover in chapter IV less far-reaching procedures are developed in the sense that they aim at a *suboptimal cutting set*. A finite step convergence proof in a sufficiently general setting to cover all variants of chapter IV, is presented in chapter III.

In chapter V a numerical comparison among policy iteration methods is

performed. It involves the three main variants developed in chapter IV and the policy iteration method of Jewell/Howard. The class of problems which satisfy the assumptions of the finite generalized Markov programming model is partitioned into two subclasses. For the first subclass it is shown that all four methods are identical. The numerical investigation is performed on members of the second subclass for which all four methods are different. It includes two sets of randomly generated problems as well as a discrete version of the production control problem mentioned above.

The treatment of the finite generalized Markov programming model of section 2.2 is continued in Chapter VI. Discounting is invoked in section 6.3 and a policy iteration method maximizing the expected discounted reward vector for a fixed interest rate ρ is presented. In section 6.4 a *parametric* generalized Markov programming model is considered. In this model an intervention is assumed to take a small time $\varepsilon \geq 0$. During the time ε the natural process is "frozen". For the parametric model a *partial Laurent expansion* for the expected discounted reward vector for a fixed policy in the parameter ρ is obtained for each sufficiently small fixed $\varepsilon > 0$.

This Laurent expansion is utilized in chapter VII to develop a new type of sensitive optimality called *sensitive intervention time optimality*. It is shown that policies of this type exist and that they can be computed by applying the methods of this monograph. Moreover a procedure is specified for the computation of a *bias-optimal* policy in the finite generalized Markov programming model. The procedures are illustrated on a numerical example.

CHAPTER I

PRELIMINARIES

1.1 NOTATION AND SOME PROPERTIES OF MATRICES

Let $\mathbb{N}$ be the set of natural numbers and let $\mathbb{C}$ be the set of complex numbers. For $J \in \mathbb{N}\setminus\{0\}$ we denote by $\mathbb{C}^J$ the set of J-dimensional column vectors with components in $\mathbb{C}$. Similarly $\mathbb{C}^{J\times J}$ denotes the set of J×K-matrices with entries in $\mathbb{C}$. Matrix and vector operations are defined in the usual way.

The equation in $u \in \mathbb{C}^J$ for fixed $B \in \mathbb{C}^{J\times K}$ and $\lambda \in \mathbb{C}$ given by

$$(1.1;1) \qquad Bu = \lambda u$$

has a nonzero solution for $n \in \{1,\ldots,J\}$ distinct values of λ called the *eigenvalues* of B, cf. [KATO 1966, p.37]. The set of all eigenvalues of B is called the *spectrum* of B and is denoted by $\sigma(B)$.

The *norm* of B, notation: $\|B\|$, is defined by

$$(1.1;2) \qquad \|B\| = \max_{i\in\{1,\ldots,J\}} \sum_{j=1}^{J} |B_{ij}|.$$

For each matrix $B \in \mathbb{C}^{J\times J}$ $\lim_{n\to\infty} \|B^n\|^{1/n}$ exists (cf. [KATO 1966, p.27]) and is called the *spectral radius* of B, notation: spr(B). Between spr(B) and $\sigma(B)$ the following identity holds (cf. [KATO 1966, p.38]):

$$(1.1;3) \qquad spr(B) = \max_{\lambda\in\sigma(B)} |\lambda|.$$

Let $v \in \mathbb{C}^J$, $B \in \mathbb{C}^{J\times J}$ and $\lambda \in \mathbb{C}$. Consider the inhomogeneous equation in $u \in \mathbb{C}^J$

$$(1.1;4) \qquad [\lambda I - B]u = v,$$

where I is the J×J identity matrix. This equation has a unique solution u if and only if $\lambda \notin \sigma(B)$. Then the inverse $[\lambda I - B]^{-1}$ exists and the solution

of (1.1;4) is given by

$$(1.1;5) \qquad u = [\lambda I - B]^{-1} v.$$

The matrix function $R_\lambda(B)\colon \mathbb{C}\backslash\sigma(B) \longrightarrow \mathbb{C}^{J\times J}$ defined by $R_\lambda(B) = [\lambda I - B]^{-1}$ is called the *resolvent* of matrix B. For particular $\lambda,\mu \notin \sigma(B)$ the matrices $R_\lambda(B)$ and $R_\mu(B)$ satisfy the *resolvent equation*, cf. [KATO 1966, p.36]

$$(1.1;6) \qquad R_\lambda(B) - R_\mu(B) = (\lambda-\mu)R_\lambda(B)R_\mu(B).$$

If, moreover, $\mathrm{spr}((\mu-\lambda)R_\mu(B)) < 1$ then $R_\lambda(B)$ has the Neumann series expansion

$$(1.1;7) \qquad R_\lambda(B) = \sum_{n=0}^{\infty} (\lambda-\mu)^n (R_\mu(B))^{n+1}.$$

Let S be a subset of $\{1,\ldots,J\}$. Then $\bar{S} \overset{d}{=} \{1,\ldots,J\}\backslash S$ and $|S|$ denotes the number of elements in S. Let $B \in \mathbb{C}^{J\times J}$ and let T,S be nonempty subsets of $\{1,\ldots,J\}$. The matrices with entries B_{ij}, $i,j \in S$, respectively B_{ij}, $i \in S$, $j \in T$, are denoted by B_S and B_{ST}. Similarly for $u \in \mathbb{C}$, the vector with components u_i, $i \in S$ is denoted by u_S.

Let $\mathbb{R}^J$ denote the real J-dimensional Euclidean space and let $\mathbb{R}^J_+$ denote the set of vectors $u \in \mathbb{R}^J$ with nonnegative components. For $u,v \in \mathbb{R}^J$, we write $u \geq v$ if $u_i \geq v_i$ for $i \in \{1,\ldots,J\}$ and $u > v$ if $u \geq v$ and $u \neq v$. The notation $\langle u,v\rangle$ denotes the scalar product of u and v. Sometimes the operation $\square$ between two vectors $u,v \in \mathbb{R}^J$ will be used; $u\square v$ denotes the vector with components $u_i v_i$. Observe that this operation is commutative, associative and distributive. The vector in $\mathbb{R}^J_+$ with all components equal to one is denoted by 1. The notation 1_S is reserved for the subvector of 1 corresponding to a nonempty set $S \subset \{1,\ldots,J\}$.

Let $\mathbb{R}^{J\times K}$ denote the set of real J×K-matrices. For $A \in \mathbb{R}^{J\times K}$ we write $A \geq 0$ if all entries are nonnegative, $A > 0$ if $A \geq 0$ and $A \neq 0$ and $A >> 0$ if all entries are positive. For $A,B \in \mathbb{R}^{J\times K}$ we write $A \geq (>)(>>)B$ if $A-B \geq (>)(>>)0$. The same notation applies to vectors.

A matrix $A \in \mathbb{R}^{J\times K}$ is called *lexicographically nonnegative*, notation: $A \succeq 0$, if for each row of A either all elements vanish or the first non-vanishing element is positive. A is *lexicographically semi-positive*, notation $A \succ 0$ if $A \succeq 0$ and $A \neq 0$. For two matrices $A,B \in \mathbb{R}^{J\times K}$ we write $A \succeq (\succ)B$ if $A-B \succeq (\succ)0$.

1.2 SUBSTOCHASTIC MATRICES

A matrix $P \in \mathbb{R}_+^{J\times J}$ is called *substochastic* if

(1.2;1) $\quad \|P\| \leq 1.$

A well-known model described by a substochastic matrix and a vector $\alpha \in \mathbb{R}_+^J$ such that $\sum_{i=1}^J \alpha_i = 1$ is a *finite homogeneous Markov chain*. Let $(\Omega,\mathcal{H},\mathbb{P})$ be a probability space and let $M \overset{d}{=} \{1,\ldots,J\}$ be a finite index set. A sequence of possibly defective random variables $\underline{j}_n: \Omega \to M$ defined for $n \in \mathbb{N}$ with $\mathbb{P}\{\underline{j}_0=i\} = \alpha_i$ is a finite Markov chain if

(1.2;2) $\quad \mathbb{P}\{\underline{j}_{n+1}=i_{n+1} \mid \underline{j}_0= i_0,\ldots, \underline{j}_n= i_n\} = \mathbb{P}\{\underline{j}_{n+1}= i_{n+1} \mid \underline{j}_n= i_n\},$

whenever $\mathbb{P}\{\underline{j}_0 = i_0,\ldots, \underline{j}_n = i_n\} > 0$. A finite Markov chain is homogeneous if for some substochastic matrix P

(1.2;3) $\quad \mathbb{P}\{\underline{j}_{n+1} = j \mid \underline{j}_n = i\} = P_{ij} \quad$ for $n \in \mathbb{N}$,

whenever $\mathbb{P}\{\underline{j}_n=i\} > 0$. For a finite homogeneous Markov chain it is easily derived that for $u \geq 0$ and $P^0 \overset{d}{=} I$

(1.2;4) $\quad \mathbb{P}\{\underline{j}_n = j \mid \underline{j}_0 = i\} = (P^n)_{ij}.$

In the sequel, the Markov chains we consider are finite and homogeneous. Therefore, the adjectives finite and homogeneous can be omitted without confusion. To the indices $1,\ldots,J$ will be referred to as *states* and P will be called the *transition matrix*.

A state $i \in M$ *has access to* state j if for some $n \in \mathbb{N}$ $(P^n)_{ij} > 0$. A state having only access to itself is called *absorbing*. Two states having access to each other are said to *communicate*. The communication relation between two states is an *equivalence relation*.

For a substochastic matrix P the Cesaro sum of nonnegative powers of P

(1.2;5) $\quad P^* \overset{d}{=} \lim_{n\to\infty} (n+1)^{-1} \sum_{i=0}^{n} P^i$

exists and satisfies (cf. [DOOB 1953, p.175])

(1.2;6) $\quad P^* = P^*P^* = PP^* = P^*P.$

In the sequel P^* will be called the Cesaro sum of P. An important special case of a substochastic matrix is described by the following lemma (cf. [KEMENY & SNELL 1960]).

LEMMA 1.2.1. *For a substochastic matrix* P *the following statements are equivalent:*

(i) $\quad \mathrm{spr}(P) < 1;$

(ii) $\quad \|P^n\| < 1$ *for some* $n \in \mathbb{N}\setminus\{0\};$

(iii) $\quad \lim_{n\to\infty} P^n = 0;$

(iv) $\quad \sum_{n=0}^{\infty} P^n$ *converges;*

(v) $\quad [I-P]$ *is nonsingular;*

(vi) $\quad \|P^J\| < 1;$

(vii) $\quad P^* = 0.$

Each of the statements (i) - (vii) *implies* $[I-P]^{-1} = \sum_{n=0}^{\infty} P^n.$

A substochastic matrix is called *transient* if one of the statements of lemma 1.2.1 holds. Frequently the following trivial property of a transient matrix will be used.

LEMMA 1.2.2. *Every square submatrix of a transient matrix is again transient.*

Also the following lemma will be in frequent use (cf. [DENARDO & FOX 1968]).

LEMMA 1.2.3. *Let* P *be a transient matrix. Let* $s \in \mathbb{R}^J$ *be an arbitrary vector. For each* $u \in \mathbb{R}^J$ *define the operator* $L: \mathbb{R}^J \to \mathbb{R}^J$ *by* $Lu \overset{d}{=} s + Pu$. *Define* $L^n \overset{d}{=} L\,L^{n-1}$ *for* $n = 2,3,\ldots$ *with* $L^1 \overset{d}{=} L$. *Then* L *is a* J*-stage contraction and has the following properties:*

(i) $\quad \lim_{n\to\infty} L^n u$ *exists and is given by* $u^* = [I-P]^{-1} s;$

(ii) $\quad Lu = u \iff u = u^*;$

(iii) $\quad Lu \geq (>)(\leq)(<)u \Rightarrow u^* \geq (>)(\leq)(<)u.$

A substochastic matrix P is called *stochastic if*

(1.2;7) $\quad \sum_{j\in J} P_{ij} = 1 \quad$ for $i \in J.$

In this case $\mathrm{spr}(P) = 1$ and $P^* > 0$.

The set of states M of a Markov chain with stochastic matrix P can be

partitioned uniquely into two sets of states E and T. Each state $i \in E$ satisfies $P^*_{ii} > 0$ and each state $i \in T$ (T may be empty) satisfies $P^*_{ii} = 0$. The set E can be partitioned in a unique way into the equivalence classes (with respect to the communication relation) E_λ, $\lambda = 1,\ldots,2$ with $1 \leq L \leq |E|$. Such an equivalence class is called a *subchain*. The states in E are called *persistent* and those in T *transient*.

The following lemma (cf. [KEMENY & SNELL 1960]) relates and specifies the solution sets of the equations $Px = x$ and $P^*y = y$.

LEMMA 1.2.4. *If P is a stochastic matrix then the solution sets of the equations* $Px = x$ *and* $P^*y = y$ *are identical and their general solution is given by*

$$(1) \qquad x = \sum_{\lambda=1}^{L} \alpha_\lambda \phi_\lambda$$

where

α_λ *is an arbitrary constant for each fixed* $\lambda \in \{1,\ldots,L\}$,
ϕ_λ *is a vector in* $\mathbb{R}^J_+$ *for each fixed* $\lambda \in \{1,\ldots,L\}$ *such that* $(\phi_\lambda)_i = 1$ *for* $i \in E_\lambda$, $(\phi_\lambda)_i = 0$ *for* $i \in E \backslash E_\lambda$ *and the subvector* $(\phi_\lambda)_T$ *is given by*

$$(2) \qquad (\phi_\lambda)_T = [I_T - P_T]^{-1} P_{TE_\lambda} 1_{E_\lambda}.$$

If P is transient then $x = 0$ *is the only solution of* $Px = x$.

Some more properties of the matrix P^* are summarized in the following lemma (cf. [DOOB 1953, p.175-183]).

LEMMA 1.2.5. *Let P be a stochastic matrix with Cesaro sum* P^* *and subchains* E_λ, $\lambda = 1,\ldots,L$. *Let* P^*_i *be the* i^{th} *row of* P^* *for* $i \in M$. *Then*

(i) $P^*_i = P^*_j$ *for* $i,j \in E_\lambda$;
(ii) $P^*_{ij} = 0$ *for* $i \in E_\lambda$, $j \notin E_\lambda$;
(iii) $P^*_{ij} > 0$ *for* $i,j \in E_\lambda$;
(iv) *defining* $\pi_\lambda \overset{d}{=} P^*_i$ *for* $i \in E_\lambda$, $\lambda = 1,\ldots,L$ *then*

$$P^*_i = \sum_{\lambda=1}^{L} (\phi_\lambda)_i \pi_\lambda \qquad \text{for } i \in M,$$

where ϕ_λ *is defined in lemma* 1.2.4.

Next some relations between the spectra of a stochastic matrix P and some matrices associated with P are stated (cf. [VEINOTT 1969]).

LEMMA 1.2.6. *Let* P *be a stochastic matrix with* P^* *as its Cesaro sum. Let* $\bar{P} \overset{d}{=} P-I$ *and* $\sigma_1(P) \overset{d}{=} \sigma(P)\backslash\{1\}$. *Then*

(i) $\sigma(\bar{P}) = \{\lambda - 1,\ \lambda \in \sigma(P)\}$;

(ii) $\sigma(P-P^*) = \{0\} \cup \sigma_1(P)$;

(iii) $\sigma(\bar{P}-P^*) = \{-1\} \cup \{\lambda - 1,\ \lambda \in \sigma_1(P)\}$.

For a substochastic matrix P, the matrix function $H_\rho: \mathbb{C}\backslash\sigma(\bar{P}-P^*) \to \mathbb{C}^{J\times J}$ is defined by

$$H_\rho = R_\rho(\bar{P}-P^*) - (1+\rho)^{-1} P^*. \tag{1.2;8}$$

Observe that H_0 exists by lemma 1.2.6. H_ρ and H_0 satisfy the resolvent equation

$$H_\rho - H_0 = -\rho H_\rho H_0 \tag{1.2;9}$$

and for $|\rho| < [\mathrm{spr}(H_0)]^{-1}$, H_ρ can be expanded into the Neumann series

$$H_\rho = \sum_{n=0}^{\infty} (-\rho)^n H_0^{n+1}. \tag{1.2;10}$$

The resolvent $R_\rho(\bar{P}): \mathbb{C}\backslash\sigma(\bar{P}) \to \mathbb{C}^{J\times J}$ has the following Laurent expansion (cf. [VEINOTT & MILLER 1969])

$$R_\rho(\bar{P}) = \rho^{-1} P^* + H_\rho = \rho^{-1} P^* + \sum_{n=0}^{\infty} (-\rho)^n H_0^{n+1} \tag{1.2;11}$$

for $0 < |\rho| < [\mathrm{spr}(H_0)]^{-1}$.

The following lemmas are frequently used in the sequel (cf. [DENARDO & FOX 1968] and [DENARDO 1971] respectively).

LEMMA 1.2.7. *For a subchain* E_λ, $\lambda \in \{1,\ldots,L\}$, *of a stochastic matrix* P *and and a vector* $r \in \mathbb{R}^{|E_\lambda|}$

$$P_{E_\lambda} r \geq (\leq) r \Rightarrow P_{E_\lambda} r = r.$$

LEMMA 1.2.8. *Let* P *be a stochastic matrix and let* $b \in \mathbb{R}^J$ *satisfy* $P^*b = 0$. *Then* $x \in \mathbb{R}^J$ *satisfies* $[I-P]x = b$ *if and only if* $x = H_0 b + r$ *with* $r \in \mathbb{R}^J$ *satisfying* $P^*r = r$.

In the subsequent chapters the set of states M of a Markov chain will be

partitioned frequently. If $(A,\bar{A})$ is such a partition a Markov chain with the set A as set of states may be considered. Such a Markov chain will be called an A-*embedded Markov chain*. Whenever $\mathrm{spr}(P_{\bar{A}}) < 1$ its transition matrix P(A) is related to P by

$$(1.2;12) \qquad P(A) = P_A + P_{A\bar{A}}[I_{\bar{A}} - P_{\bar{A}}]^{-1} P_{\bar{A}A}.$$

1.3. MARKOV RENEWAL PROCESSES

In this section a summary of Markov renewal processes is given which establishes the notation and states some properties needed in subsequent chapters. For a more extensive treatment the reader is referred to [ÇINLAR 1969, 1975 resp.], unless stated otherwise.

Let $(\Omega,\mathcal{H},\mathbb{P})$ be a probability space and let $M = \{1,\ldots,J\}$ be a finite index set. On this probability space is for each $n \in \mathbb{N}$ defined a sequence of pairs of possibly defective random variables $\underline{j}_n: \Omega \to M$ and $\underline{t}_n: \Omega \to \mathbb{R}_+$ such that $0 = \underline{t}_0 \leq \underline{t}_1 \leq \underline{t}_2 \leq \ldots$. The sequence of pairs $(\underline{j}_n,\underline{t}_n)$ thus defined, forms a *Markov renewal process* if

(i) it has the Markov renewal property:

$$(1.3;1) \qquad \mathbb{P}\{\underline{j}_{n+1} = j,\ \underline{t}_{n+1} - \underline{t}_n \leq t \mid \underline{j}_0 = i_0,\ldots,\underline{j}_n = i_n;\ 0 = \underline{t}_0 \leq \underline{t}_1 \leq \tau_1 \leq \underline{t}_2 \leq \ldots \leq \underline{t}_n \leq \tau_n\}$$

$$= \mathbb{P}\{\underline{j}_{n+1} = j,\ \underline{t}_{n+1} - \underline{t}_n \leq t \mid \underline{j}_n = i_n\},$$

whenever $\mathbb{P}\{\underline{j}_0 = i_0,\ldots,\ \underline{j}_n = i_n,\ \underline{t}_0 = 0 \leq \underline{t}_1 \leq \tau_1 \leq \underline{t}_2 \leq \ldots \leq \underline{t}_n \leq \tau_n\} > 0$, and

(ii) it is homogeneous in time, which means that for $n \in \mathbb{N}$

$$(1.3;2) \qquad \mathbb{P}\{\underline{j}_{n+1} = j,\ \underline{t}_{n+1} - \underline{t}_n \leq t \mid \underline{j}_n = i\} = Q(t)_{ij}, \qquad i,j \in M$$

for some possibly defective distribution function $Q(t)_{ij}$ with support in $\mathbb{R}_+$ and $\sum_{j\in M} Q(t)_{ij} \leq 1$ for each fixed $t \in \mathbb{R}_+$. Assume further

$$(1.3;3) \qquad \mathbb{P}\{\underline{j}_0 = i,\ \underline{t}_0 = 0\} = 1, \qquad \text{for some } i \in M.$$

The family of probabilities $\{Q(t)_{ij},\ i,j \in M,\ t \in \mathbb{R}_+\}$ is called a *semi Markov kernel* and can be arranged in an $J\times J$-matrix Q with the functions

$Q(t)_{ij}$ as entries. Such a matrix is called a *semi Markov matrix*. For each fixed $t \in \mathbb{R}_+$ the numbers $Q(t)_{ij}$, $i,j \in M$, form a substochastic matrix $Q(t)$. A semi Markov matrix Q can, therefore, also be understood as a matrix function $Q(t): \mathbb{R}_+ \to \mathbb{R}_+^{J\times J}$. The name semi Markov matrix will be used in either sense and will be abbreviated by SMM.

For an SMM $Q(t)$ the limit $\lim_{t\to\infty} Q(t)_{ij}$ exists for $i,j \in M$ and the numbers P_{ij} are defined by

$$(1.3;4) \qquad P_{ij} \overset{d}{=} \lim_{t\to\infty} Q(t)_{ij} \qquad \text{for } i,j \in M.$$

The numbers P_{ij} also constitute a substochastic matrix. The Markov chain induced by it is called the *embedded Markov chain* of the Markov renewal process.

In relation to the possibly defective distribution functions $Q(t)_{ij}$, $i,j \in M$, are defined the functions $Q^{(n)}(t)_{ij}: \mathbb{R}_+ \to \mathbb{R}_+$ for $n = 1,2,\ldots$ and $i,j \in M$ such that

$$Q^{(1)}(t)_{ij} = Q(t)_{ij}, \qquad \text{for } t \in \mathbb{R}_+$$

and

$$(1.3;5) \qquad Q^{(n)}(t)_{ij} = \sum_{k\in M} \int_{u\in[0,t]} Q(du)_{ik}\, Q^{(n-1)}(t-u)_{kj}, \qquad \text{for } t \in \mathbb{R}_+.$$

Observe that the matrices $Q^{(n)}(t)$, $n = 1,2,\ldots$, with as entries the functions $Q^{(n)}(t)_{ij}$ are again semi Markov matrices as a consequence of the convolution theorem for (possibly defective) distribution functions (cf. [FELLER 1966, p.141]).

An SMM $Q(t)$ is called *normal* if

$$(1.3;6) \qquad \mathrm{spr}(Q(0)) < 1.$$

This property guarantees that the corresponding Markov renewal process developes beyond time zero. Normality will be postulated in the sequel of this section.

For fixed $i,j \in M$ the function $R(t)_{ij}: \mathbb{R}_+ \to \mathbb{R}_+$ is defined by

$$(1.3;7) \qquad R(t)_{ij} = \sum_{n=0}^{\infty} Q^{(n)}(t)_{ij}.$$

$R(t)_{ij}$ is easily recognized as a renewal function and represents the expected number of renewals of state j in the time interval $[0,t]$ if the initial

state is i. Observe that by (1.3;7) $R(t)_{ij}$ is right continuous and non-decreasing. The matrix function $t \to R(t)$ with entries $R(t)_{ij}$ is called the *Markov renewal matrix* (abbreviation: MRM) corresponding to $Q(t)$. The following lemma states some equivalences.

LEMMA 1.3.1. *The following statements are equivalent:*

(i) $R(0) < \infty$;

(ii) $R(t) < \infty$ *for* $t \in \mathbb{R}_+$;

(iii) $Q(t)$ *is normal.*

The following trivial properties of $Q(t)$ and $R(t)$ are frequently used.

LEMMA 1.3.2. *Let* $Q(t)$ *be a normal SMM with MRM* $R(t)$ *and* $P = \lim_{t\to\infty} Q(t)$ *a stochastic matrix. Let* E_λ *be some subchain of* P. *Then*

(i) $Q(t)_{ij} = 0$ *for all* $t \in \mathbb{R}_+$ *if and only if* $P_{ij} = 0$;

(ii) $R(t)_{ij} = 0$ *for all* $t \in \mathbb{R}_+$ *whenever* $i \in E_\lambda$ *and* $j \notin E_\lambda$.

The Laplace-Stieltjes transform of a normal SMM $Q(t)$ is denoted by $q(s)$. Each entry $q(s)_{ij}$, $i,j \in M$, of $q(s)$ is for $s \geq 0$ defined by

$$(1.3;8) \qquad q(s)_{ij} \overset{d}{=} \int_{t\in\mathbb{R}_+} e^{-st}\, Q(dt)_{ij}.$$

Observe that $\|q(s)\| < 1$ for $s > 0$. By an extension to finite matrices of the multiplication rule for Laplace-Stieltjes transforms of convolutions (cf. [FELLER 1966, p.411]), the transform of

$$(1.3;9) \qquad q^{(n)}(s) = q(s)\, q^{(n-1)}(s) = [q(s)]^n, \qquad \text{for } s \geq 0.$$

Observe that $\|q^{(n)}(s)\| < 1$ for $s > 0$. Since the sequence of partial sums $\sum_{n=0}^{m} q^{(n)}(s)_{ij}$, $m = 0,1,2,\ldots$ is bounded for $s > 0$ we have by the extended continuity theorem for Laplace-Stieltjes transforms (cf. [FELLER 1966, p.410])

$$(1.3;10) \qquad r(s)_{ij} \overset{d}{=} \sum_{n=0}^{\infty} q^{(n)}(s)_{ij} = \int_{t\in\mathbb{R}_+} e^{-st}\, R(dt)_{ij}, \qquad \text{for } s > 0.$$

Note that as a consequence of (1.3;9) and $\|q(s)\| < 1$ for $s > 0$

$$(1.3;11)\qquad r(s) = \sum_{n=0}^{\infty} [q(s)]^n = [I-q(s)]^{-1}$$

and

$$(1.3;12)\qquad r(s)q(s) = q(s)r(s).$$

The m^{th} *normalized moment* of $Q(t)$ is denoted by Q_m, $m \in \mathbb{N}$ with entries

$$(Q_m)_{ij} \stackrel{d}{=} \int_{t\in\mathbb{R}_+} (m!)^{-1}\, t^m\, Q(dt)_{ij},$$

provided that the integral is finite for all pairs $(i,j) \in M\times M$. If Q_m is finite then Q_ℓ is finite for $\ell < m$. Notice that $P = Q_0$ (cf. (1.3;4)). The following lemma presents the asymptotic expansion for $s \downarrow 0$ of $q(s)$ in its normalized moments (cf. [CHUNG 1974, p.168]).

LEMMA 1.3.3. *Suppose* Q_m *is finite for some* $m \in \mathbb{N}$. *Then* $q(s)$ *has the asymptotic expansion*

$$q(s) = \sum_{\ell=0}^{m} (-1)^\ell\, Q_\ell\, s^\ell + o(s^m), \qquad s \downarrow 0.$$

The following asymptotic properties of $r(s)$ for $s \downarrow 0$ are used in chapter VI (cf. [DENARDO 1971]).

LEMMA 1.3.4. *Let* $Q(t)$ *be a normal* SMM *with corresponding* MRM $R(t)$. *Let* $P = Q_0$ *be stochastic with Cesaro sum* P^*. *Let* $a \in \mathbb{R}^J$. *Then*

(i) $r(s) = O(1/s)$, $s \downarrow 0$;

(ii) $P^*a = 0$ *implies* $r(s)a = o(1/s)$.

A partial Laurent expansion for $r(s)$ in terms of the normalized moments of $Q(t)$ is provided by the following lemma (cf. [DENARDO 1971, p.484]).

LEMMA 1.3.5. *Let* $Q(t)$ *be a normal* SMM *and let* Q_{m+2} *be finite. Let* $P = Q_0$ *be substochastic with Cesaro sum* P^*. *Then* $r(s)$ *has a partial Laurent expansion in powers of* s *at the origin given by*

$$r(s) = \sum_{i=-1}^{m} s^i\, R_i + o(s^m), \qquad s \downarrow 0,$$

where the matrices $R_i \in \mathbb{R}^{J\times J}$, $i = -1,0,1,\ldots,m$ *are uniquely determined by*

the system of equations

$$(1) \qquad \begin{cases} [I-P]R_i = F_i \\ P^* Q_1 R_i = P^* G_i, \end{cases}$$

with

$$F_i \stackrel{d}{=} \begin{cases} I & \text{for } i = 0 \\ 0 & \text{for } i \neq 0 \end{cases} + \sum_{\ell=1}^{i+1} (-1)^{\ell} Q_{\ell} R_{i-\ell},$$

$$G_i \stackrel{d}{=} F_{i+1} + Q_1 R_i.$$

In case P *is transient then the finiteness of* Q_{m+1} *is required and* (1) *simplifies to*

$$(2) \qquad R_i \stackrel{d}{=} \begin{cases} [I-P]^{-1} F_i & \text{for } i > -1 \\ 0 & \text{for } i = -1. \end{cases}$$

To conclude this section the Markov renewal equation in $U(t)\colon \mathbb{R}_+ \to \mathbb{R}^J$, given by

$$(1.3;13) \qquad U(t) = G(t) + \int_{y\in[0,t]} Q(dy)U(t-y)$$

is considered where $G(t)\colon \mathbb{R}_+ \to \mathbb{R}^J$ is a given vector function and $Q(t)$ is a normal SMM. The solution space of (1.3;13) is constructed as follows. Let the norm in $\mathbb{R}^J$ be defined as $\|V\| = \max_{j\in J}|V_j|$. Let F_J be the collection of functions $U(t)\colon \mathbb{R}_+ \to \mathbb{R}$ such that

(i) U_j is Borel measurable for $j = 1,\ldots,J$;

(ii) $\|U(t)\|$ is bounded on finite intervals.

A property of the solution of (1.3;13) is specified by

LEMMA 1.3.6. *The Markov renewal equation* (1.3;13) *has a unique solution in* F_J *given by*

$$U(t) = \int_{y\in[0,t]} R(dy)G(t-y)$$

if $Q(t)$ *is a normal SMM and* $G(t) \in F_J$.

The Markov renewal equation (1.3;13) can be obviously extended to matrix functions. Each column vector function is the unique solution in F_J of an equation of the type (1.3;13). The name Markov renewal equation will also be used for this extension. Notice for example that a MRM R(t) is uniquely determined by

$$(1.3;14) \qquad R(t) = I + \int_{y\in[0,t]} Q(dy)R(t-y).$$

In this monograph, the vector function G(t) in (1.3;13) is assumed to be of bounded variation. In agreement with its economic interpretation, it will be called a *reward vector*. In order to define the normalized moments of G(t), let $G^+(t) \overset{d}{=} \max[G(t),0]$ and $G^-(t) \overset{d}{=} \max[-G(t),0]$ for $t \in \mathbb{R}_+$. If $\int_{t\in\mathbb{R}_+} (m!)^{-1} t^m G^+(dt) < \infty$ and $\int_{t\in\mathbb{R}_+} (m!)^{-1} t^m G^-(dt) < \infty$ then the m-th normalized moment $G_m = \int_{t\in\mathbb{R}_+} (m!)^{-1} t^m G(dt)$ exists and is finite.

The Laplace-Stieltjes transform g(s) of a reward vector G(t) exists for $s \geq 0$. An asymptotic expansion for g(s), $s \downarrow 0$, is provided by

LEMMA 1.3.7. *If a reward vector* G(t) *has a finite* k^{th} *normalized moment* G_k *for some* $k \geq 0$, *then* g(s) *has the asymptotic expansion*

$$g(s) = \sum_{\ell=0}^{k} (-1)^\ell G_\ell s^\ell + o(s^k), \qquad s \downarrow 0.$$

1.4. FINITE MARKOV RENEWAL DECISION MODELS

Markov renewal decision models with a finite number of states and a finite number of actions per state were considered among others by [HOWARD 1960, 1963], [JEWELL 1963], [DE CANI 1964], [SCHWEITZER 1965, 1969], [DENARDO & FOX 1968] and [DENARDO 1970, 1971].

Primarily a description of the general finite Markov renewal decision model (cf. [DENARDO 1971]) will be given. Its name will be abbreviated by MRD-model.

In the MRD-model a system is observed at a sequence of stochastic epochs $\underline{t}_n$, n = 0,1,2,... such that $\underline{t}_0 = 0 \leq \underline{t}_1 \leq \underline{t}_2 \ldots$. At each epoch the

system is in a certain state. Let $M = \{1,\ldots,J\}$ be the set of states. If at epoch $\underline{t}_n$ state i has been observed, the decision maker chooses an action k from a finite set of actions K(i). Each pair (i,k), $i \in M$, $k \in K(i)$ induces

(i) a set of J possibly defective distribution functions $Q(t)_{ij}^{k}$, $j \in M$ with support on $\mathbb{R}_+$ each representing the joint probability that the next observation epoch occurs not later than $\underline{t}_n + t$ and state $\underline{j}$ is the next observed state (implying $\sum_{j\in M} Q_{ij}(t) \leq 1$) and

(ii) a reward function $C(t)_i^k$ denoting the cumulative expected reward earned by the system during the interval $[\underline{t}_n, \min(t,\underline{t}_{n+1}))$.

Let $F \stackrel{d}{=} \times_{i\in M} K(i)$ and let $f \in F$. Each component f(i) of f satisfies $f(i) \in K(i)$. Hence f specifies a decision rule. If the same decision rule f is applied at each observation epoch, we speak of a stationary policy. The adjective stationary is omitted in the sequel. Each policy f induces a normal SMM Q(t;f) with entries $Q(t)_{ij}^{f(i)}$, $i,j \in M$ and reward vector C(t;f) with components $C(t)_i^{f(i)}$. Moreover, it is assumed that for $i \in M$, $k \in K(i)$ and $j \in M$

$$(1.4;1) \qquad \int_{t\in\mathbb{R}_+} t\, Q(dt)_{ij}^{k} < \infty.$$

Each policy $f \in F$ induces a vector function $\hat{V}(t;f)$ which is the unique solution in F_J of the Markov renewal equation

$$(1.4;2) \qquad \hat{V}(t;f) = C(t;f) + \int_{u\in[0,t]} Q(dt;f)\hat{V}(t-u;f).$$

Each component $\hat{V}(t;f)_i$ of $\hat{V}(t;f)$ is interpreted as the expected reward earned in the time interval [0,t] during the Markov renewal decision process induced by policy f with initial state $i \in M$.

If discounting is invoked in the MRD-model then a reward at time t is multiplied by a discount factor $e^{-\rho t}$ where ρ is the interest rate. Invoking discounting in this model corresponds to Laplace-Stieltjes transformation of the functions $\hat{V}(t;f)$, C(t;f) and Q(t;f). Their Laplace-Stieltjes transforms are denoted by $\hat{v}(\rho;f)$, $c(\rho;f)$ and $q(\rho;f)$ and exist at least for real $\rho > 0$. A component $\hat{v}(\rho;f)_i$ of $\hat{v}(\rho;f)$ then represents the expected discounted reward in the time interval $[0,\infty)$ during the Markov renewal decision process induced by policy $f \in F$ with initial state i. $\hat{v}(\rho;f)$ uniquely satisfies (1.4;2) in transformed form given by

(1.4;3) $\quad \hat{v}(\rho;f) = c(\rho;f) + q(\rho;f)\hat{v}(\rho;f) \quad$ for $\rho > 0$.

For fixed $\rho > 0$ (1.4;3) represents a set of J linear equations with the components $\hat{v}(\rho;f)_i$, $i \in M$ as unknowns. A policy iteration method can be developed (cf. [JEWELL 1963]) which computes a policy maximizing $\hat{v}(\rho;f)$ over $f \in F$ for any fixed $\rho > 0$. Such a policy is called ρ-*optimal*.

In [DENARDO 1971] ρ is considered as a variable. By substitution of the asymptotic expansions for $\rho \downarrow 0$ of $q(\rho;f)$ (cf. lemma 1.3.3) and $c(\rho;f)$ (which is similar to the one for $g(s)$ in lemma 1.3.7) in equation (1.4;3) a partial Laurent expansion for $\hat{v}(\rho;f)$ is obtained which is given by

$$(1.4;4) \qquad \hat{v}(\rho;f) = \sum_{m=-1}^{n} \hat{V}_m(f)\rho^m + o(\rho^n), \quad \rho \downarrow 0$$

whenever the normalized moments $Q_{n+2}(f)$ and $C_{n+1}(f)$ of $Q(t;f)$ and $C(t;f)$ respectively are finite for some $n \in \mathbb{N}$. The system of linear equations in $(\hat{V}_m(f), \hat{V}_{m+1}(f))$, $m = -1,0,1,\ldots,n$ given by

$$(1.4;5) \qquad \begin{cases} [I-Q_0(f)]\hat{V}_m(f) = (-1)^m C_m(f) + \sum_{k=1}^{m+1} (-1)^k Q_k(f)\hat{V}_{m-k}(f) \\ [I-Q_0(f)]\hat{V}_{m+1}(f) = (-1)^{m+1} C_{m+1}(f) - Q_1(f)\hat{V}_m(f) + \\ \qquad\qquad + \sum_{k=2}^{m+2} (-1)^k Q_k(f)\hat{V}_{m+1-k}(f), \end{cases}$$

with $C_{-1}(f) \overset{d}{=} 0$, has a solution which is unique in $\hat{V}_m(f)$.

The coefficient $\hat{V}_{-1}(f)$ in 1.4;4) represents the expected average reward vector per time unit in the long run for a fixed policy $f \in F$. It will be called the gain of policy f. A policy $f^* \in F$ such that $V_{-1}(f^*) \geq$ $\geq V_{-1}(f)$ for $f \in F$ exists and is called *gain-optimal*. A gain-optimal policy which maximizes also $\hat{V}_0(f)$ over the set of gain-optimal policies is called *bias-optimal*.

The computation of the gain $V_{-1}(f)$ for a policy $f \in F$ requires the matrix $Q_0(f)$ and the vectors $T(f) \overset{d}{=} Q_1(f)1$ and $C_0(f)$. Hence if only the data

$$(1.4;6) \qquad \begin{cases} (Q_0)^k_{ij} = Q(\infty)^k_{ij} & \text{for } i,j \in M,\ k \in K(i) \\ T^k_i \overset{d}{=} \sum_{j\in M}(Q_1)^k_{ij} = \sum_{j\in M} \int_{t\in\mathbb{R}_+} tQ(dt)^k_{ij} & \text{for } i \in M,\ k \in K(i) \\ (C_0)^k_i = C(\infty)^k_i & \text{for } i \in M,\ k \in K(i) \end{cases}$$

of the MRD-model are specified then the gain can be computed for each policy. If the data of the MRD-model are restricted to $(Q_0)_{ij}^k$, T_i^k, $(C_0)_i^k$ for $k \in K(i)$, $i,j \in M$ and moreover

$$(1.4;7) \qquad T_i^k > 0 \quad \text{for } k \in K(i), \quad i \in M,$$

then we speak of the *undiscounted* MRD-*model*. If the data of the MRD-model are restricted to $(Q_0)_{ij}^k$, T_i^k, $(C_0)_i^k$ for $k \in K(i)$, $i,j \in M$ and

$$(1.4;8) \qquad T_i^k \geq 0$$

holds with equality sign for some pair (i,k), $i \in M$, $k \in K(i)$ and if moreover for each policy $f \in F$ and subchain $E_\lambda(f)$, $\lambda = 1,\ldots,L(f)$, of $Q_0(f)$

$$(1.4.9) \qquad T_i^{f(i)} > 0 \quad \text{for some } i \in E_\lambda(f),$$

then we shall use the name *undiscounted* MRD-*model with interventions*. Observe that assumption (1.4;7) is a sufficient and assumption (1.4;9) is a necessary and sufficient condition for the normality of the SMM $Q(t;f)$ for each $f \in F$ in the MRD-model.

Usually the gain $g(f) \overset{d}{=} V_{-1}(f)$ is computed by solving the system of equations in $g(f)$ and $u(f)$

$$(1.4;10) \qquad \begin{cases} g(f) = Q_0(f)g(f) \\ \\ u(f) = Q_0(f)u(f) + C_0(f) - T(f)\square g(f). \end{cases}$$

(1.4;10) is a simplification of (1.4;5) for $m = -1$. A solution $(g(f),u(f))$ of (1.4;10) is unique in $g(f)$ but not in $u(f)$. A particular solution of (1.4;10) is obtained by defining

$$(1.4;11) \qquad u(f)_{i(\lambda)} \overset{d}{=} 0 \quad \text{for } \lambda = 1,\ldots,L(f)$$

where $i(\lambda)$ is an arbitrarily chosen state in subchain $E_\lambda(f)$.

In all three models a gain-optimal policy can be computed by

METHOD 1.4.1. The following computational steps are to be executed:

(i) Fix an initial policy $f_1 \in F$.

(ii) Compute a particular solution $(g(f_1),u(f_1))$ to (1.4;10) for policy f_1 by taking for each state $i(\lambda)$, $\lambda = 1,\ldots,L(f_1)$ the state with the largest index in subchain $E_\lambda(f_1)$.

(iii) Compute for each $i = 1,\ldots,J$ the action sets

$$K_1(i) \stackrel{d}{=} \{k \in K(i): \sum_{j\in M} (Q_0)^k_{ij}\, g(f_1)_j = \max_{k\in K(i)} \sum_{j\in M} (Q_0)^k_{ij}\, g(f_1)_j\}$$

and

$$K_2(i) \stackrel{d}{=} \{k \in K_1(i): (C_0)^k_i - g_i T^k_i + \sum_{j\in M} (Q_0)^k_{ij}\, u(f_1)_j = \max_{k\in K_1(i)} [(C_0)^k_i - g_i T^k_i + \sum_{j\in M} (Q_0)^k_{ij}\, u(f_1)_j]\}.$$

(iv) Choose a policy $f \in F$ such that $f(i) = f_1(i)$ whenever $f_1(i) \in K_2(i)$ and $f(i) \in K_2(i)$ arbitrarily otherwise. If $f(i) = f_1(i)$ for $i \in M$ then stop. If $f_1 \neq f$ then redefine f_1 by $f_1 \stackrel{d}{=} f$ and repeat operations (ii), (iii) and (iv).

Next the well-known conditions for the optimality of a policy (cf. [DENARDO & FOX 1968]) in the undiscounted MRD-model are stated in

THEOREM 1.4.2. *A policy* f^* *satisfying for* $i \in M$ *and* $k \in K(i)$

$$\sum_{j\in M} (Q_0)^k_{ij}\, g(f^*)_j \leq g(f^*)_i,$$

with

$$\sum_{j\in M} (Q_0)^k_{ij}\, g(f^*)_j = g(f^*)_i$$

implying

$$(C_0)^k_i - g(f^*)T^k_i + \sum_{j\in M} (Q_0)^k_{ij}\, u(f^*)_j \leq u(f^*)_i,$$

is gain-optimal, i.e. $g(f) \leq g(f^*)$ *for* $f \in F$.

REMARK 1.4.3. If each policy $f \in F$ has only one subchain then the conditions of theorem 1.4.2 are reduced to

$$(C_0)^k_i - g(f^*)T^k_i + \sum_{j\in M} (Q_0)^k_{ij}\, u(f^*)_j \leq u(f^*)_i \quad \textit{for } i \in M,\ k \in K(i).$$

CHAPTER II

FINITE GENERALIZED MARKOV PROGRAMMING MODELS

2.1. INTRODUCTION

The generalized Markov programming model, introduced in [DE LEVE 1964], generalizes the Markov decision models of [HOWARD 1960] and [JEWELL 1963]. The system considered has a general state space and can be controlled continuously. There is an underlying stochastic process, called the *natural process*, which describes the evolution of the state of the system during the course of time if the system is left uncontrolled. The decisionmaker controls it by making *interventions*. An intervention causes an instantaneous (possibly random) change of the state of the system. At each point of time either an intervention can be taken or the natural process is left untouched. In the last case we speak of a *null decision*. A *policy* assigns to each state either an intervention from a set of feasible interventions in that state or the null decision in that state. The process resulting from the natural process and a policy is called a *decision process*. For each fixed policy the decision process is assumed to have the property that only a finite number of interventions may be taken in any finite time interval. Also in [DE LEVE 1964] a policy iteration method has been developed and is proven to approximate the minimum expected average cost per unit of time sufficiently good if the number of iterations is large enough.

Applications of the model and the method are presented in [DE LEVE & WEEDA 1968] and in [DE LEVE, TIJMS & WEEDA 1970]. Recently in [DE LEVE, TIJMS & FEDERGRUEN 1977] the general model is treated under the simplifying assumption that any decision process has a fixed regeneration state. The authors also provide some applications of this approach.

In the sequel the general model is considered under the much stronger assumption that the natural process is a finite state Markov renewal process and the sets of interventions are finite sets. This finite model is described in section 2.2. In section 2.3 a policy iteration method is presented

which is predominantly a direct specialization of the general method to this finite model. Under these stronger assumptions a stronger result as finite step convergence is possible and proved in chapter III. Further such a specialization makes sense because the general method is no ready made technique and presents no computational procedures for its operations. Particularly the so-called cutting operation indicates no specific computational procedure. Such a procedure is developed in chapter IV. Furthermore, the finite step convergence proof of chapter III is given in a more general setting then is needed for the policy iteration method of section 2.3. This is done to allow for variants on this cutting operation. These variants are worked out in chapter IV and are computationally investigated in chapter V. Also the existing methods of HOWARD and JEWELL are included in this investigation.

2.2. A FINITE GENERALIZED MARKOV PROGRAMMING MODEL

A system is controlled by a decision maker. Let $\Psi \overset{d}{=} \{1,\ldots,J'\}$ be the set of observable states of the system. If the decision maker leaves the system untouched the evolution in time of the system is described by a stochastic process called the *natural process*.

ASSUMPTION 2.2.1. The natural process is a Markov renewal process induced by an SMM $N(t)$ with $\|N(0)\| < 1$, $N_1 \overset{d}{=} \int_{t\in\mathbb{R}_+} t\, N(dt) < \infty$ and associated with it a reward vector $G(t)$.

The natural process is assumed to possess some additional properties specified by

ASSUMPTION 2.2.2. There exists a unique nonempty set $A_0 \subset \Psi$ such that

(i) $N(t)_{ij} = 0$ for $i \in A_0$, $j \in \Psi$;

(ii) $G(t)_i = 0$ for $i \in A_0$;

(iii) $\sum_{j\in\Psi} (N_0)_{ij} = 1$ for $i \notin A_0$ with $N_0 \overset{d}{=} \int_{t\in\mathbb{R}_+} N(dt)$;

(iv) $(N_0)_{\overline{A_0}}$ is transient.

The decision maker may only interrupt the natural process at one of its observation epochs. An interruption immediately followed by an instantaneous (possibly random) change of the state of system, is called an *intervention*.

ASSUMPTION 2.2.3. An intervention x in state $i \in \Psi$ induces

(i) a family of probabilities $\{P^x_{ij},\ j = 1,\ldots,J'\}$ with $\sum_{j\in\Psi} P^x_{ij} = 1$ each representing the probability that j is the state observed after applying intervention x in state i,

(ii) a reward $G^x_i \in \mathbb{R}$.

The only possible alternative to applying an intervention in some state is to leave the natural process untouched until its next observation epoch. This alternative is called the *nulldecision* in that state.

ASSUMPTION 2.2.4. The nulldecision $x_0(i)$ is feasible for $i \in \bar{A}_0$ and not for $i \in A_0$.

From assumption 2.2.1 it is clear that the nulldecision $x_0(i)$ in induces the family of probabilities $\{N(t)_{ij},\ j \in \Psi,\ t \in \mathbb{R}_+\}$ and the reward function $G(t)_i$.

Immediately after applying an intervention the state of the system is again observed and followed by either a new intervention or the nulldecision in the new state. This implies that in this finite model a sequence of interventions at one epoch is permitted rather than only one in the general model (cf. [DE LEVE 1964]).

The sets of actions at the disposal of the decision maker are specified by

ASSUMPTION 2.2.5. In each state $i \in \Psi$ a finite set of actions X(i) is given such that

(i) X(i) consists of the interventions in state i if $i \in A_0$;

(ii) X(i) consists of the interventions and the nulldecision $x_0(i)$ in state i if $i \notin A_0$.

In the sequel the symbol x will be used for an action (here either an intervention or a nulldecision). The notation $x_0(i)$ will not be used unless the nulldecision is specifically meant.

In this model the set of policies Z is given by the Cartesian product

$$Z \overset{d}{=} \mathop{\mathsf{X}}_{i\in\Psi} X(i). \qquad (2.2;1)$$

The process resulting from the control of the natural process by a policy $z \in Z$ is called the *decision process* corresponding to policy z. Further, to each policy $z \in Z$ corresponds a partition $(A(z), \overline{A(z)})$ of Ψ with

$$(2.2;2) \qquad A(z) \overset{d}{=} \{i \in \Psi\colon z(i) \neq x_0(i)\}.$$

An element of $A(z)$ is called an *intervention state*. The SMM associated with policy $z \in Z$ will be denoted by $S(t;z)$. Notice that entry $S(t;z)_{ij}$ of $S(t;z)$ is given by

$$(2.2;3) \qquad S(t;z)_{ij} = \begin{cases} N(t)_{ij} & \text{for} \quad i \notin A(z),\ j \in \Psi \\ P_{ij}^{z(i)} & \text{for} \quad i \in A(z),\ j \in \Psi \end{cases}.$$

The numbers $P_{ij}^{z(i)}$, $i \in A(z)$, $j \in \Psi$, can be arranged in an $|A(z)| \times J'$-matrix denoted by $P(z)_{A(z)\Psi}$. In partitioned form $S(t;z)$ is then given by

$$(2.2;4) \qquad S(t;z) = \begin{bmatrix} N(t)_{\overline{A(z)}\Psi} \\ \hline P(z)_{A(z)\Psi} \end{bmatrix}.$$

ASSUMPTION 2.2.6. The SMM $S(t;z)$ is normal for each $z \in Z$.

REMARK 2.2.7. Assumption 2.2.6 is made to prevent that the decision process for each fixed policy $z \in Z$ does not develop beyond some observation epoch. To prevent this several conditions are possible which are either necessary or sufficient or both for assumption 2.2.6 to hold. By the assumptions 2.2.2 and 2.2.3, assumption 2.2.6 implies the existence of a nonempty set A_1 such that

$$(1) \qquad X(i) = \{x_0(i)\} \quad \text{for } i \in A_1.$$

The existence of A_1 is a necessary condition. A sufficient condition for the validity of assumption 2.2.6 is

$$(2) \qquad P(z)_{A(z)} = 0 \quad \text{for } z \in Z.$$

A more general form of condition (2) is assumed in the general model (cf. [DE LEVE 1964]).

A necessary and sufficient condition is

$$(3) \qquad \mathrm{spr}(P(z)_{A(z)}) < 1 \quad \text{for } z \in Z.$$

In chapter VI the model of this section is parametrized by the assumption

that each intervention takes a small time $\varepsilon > 0$. ε is then considered as a variable. Then assumption 2.2.6 can be dropped at the expense of the appearance of an extra term in the principal part of a partial Laurent series expansion of the expected discounted reward vector in terms of ε and the interest rate ρ.

2.3. POLICY ITERATION IN THE FINITE GENERALIZED MARKOV PROGRAMMING MODEL

In the sequel the model of section 2.2 will be referred to as the *finite* GMP-*model*. If the data of the natural process in the finite GMP-model are restricted to the 0-th moment N_0 of $N(t)$, the 0-th moment G_0 of $G(t)$ and the row sums of the 1-st moment N_1 of $N(t)$ given by $\tau \overset{d}{=} N_1\, 1$, then the model will be called the *undiscounted* GMP-*model*.

Summarizing we have in the undiscounted GMP-model for the natural process

(2.3;1) $\quad (N_0)_{ij} = 0$ for $j \in \Psi$; $\quad (G_0)_i = 0$; $\quad \tau_i = 0 \qquad$ for $i \in A_0$

(2.3;2) $\quad (N_0)_{ij} \geq 0$ for $j \in \Psi$; $\quad \sum_{j \in \Psi} (N_0)_{ij} = 1 \qquad$ for $i \notin A_0$

(2.3;3) $\quad (G_0)_i \in \mathbb{R}$; $\quad 0 < \tau_i < \infty \qquad$ for $i \notin A_0$

(2.3;4) $\quad \mathrm{spr}((N_0)_{\overline{A_0}}) < 1.$

For convenience assumption 2.2.6 is sometimes replaced by the equivalent

(2.3;5) $\quad \mathrm{spr}(P(z)_{A(z)}) < 1 \qquad$ for $z \in Z$.

For the finite GMP-model a policy iteration method is developed which computes a policy maximizing the gain, i.e. the expected average reward per unit of time in the long run. The data of the undiscounted GMP-model are sufficient for this computation. In preparation it is necessary to define several quantities related to the natural process and the decision process for a fixed policy $z \in Z$. Some properties of these quantities are developed. These results are partly used to prove some properties of the solution of a set of equations arising in one of the operations of the

policy iteration method and partly stated in preparation of the development in the chapters VI and VII.

The policy iteration method is to a considerable extent a specialization of a general scheme given in [DE LEVE 1964]. However, it has the facility built in to handle policies which imply a sequence of interventions at the observation epochs.

The method is iterative and consists of a preparatory part and four operations per iteration step. The first two operations are respectively related to the policy evaluation and policy improvement operation of method 1.4.1. For the third operation, called the *cutting operation*, no specific computational procedure is indicated by the general scheme. Computational methods for this cutting operation are developed in chapter IV and investigated on their computational merits in chapter V.

Primarily two vectors related to the natural process are specified in definitions 2.3.1 and 2.3.2.

DEFINITION 2.3.1. The vector $k_0 \in \mathbb{R}^{J'}$ is defined by

(1) $$(k_0)_{\overline{A_0}} \overset{d}{=} [I_{\overline{A_0}} - (N_0)_{\overline{A_0}}]^{-1} (G_0)_{\overline{A_0}}$$

(2) $$(k_0)_{A_0} \overset{d}{=} 0.$$

DEFINITION 2.3.2. The vector $t_0 \in \mathbb{R}^{J'}$ is defined by

(1) $$(t_0)_{\overline{A_0}} \overset{d}{=} [I_{\overline{A_0}} - (N_0)_{\overline{A_0}}]^{-1} \tau_{\overline{A_0}}$$

(2) $$(t_0)_{A_0} \overset{d}{=} 0.$$

Notice that $(k_0)_i$ ($(t_0)_i$) represents the expected total reward incurred (expected total time elapsed) during the natural process with initial state $i \notin A_0$ until the set A_0 is entered.

The next two quantities involve one action and the natural process.

DEFINITION 2.3.3. For each action $x \in X(i)$ and state $i \in \Psi$ a number $k(i,x) \in \mathbb{R}$ is defined. For x intervention in state $i \notin A_1$

$$k(i,x) \overset{d}{=} G_i^x + \sum_{j \in \Psi} P_{ij}^x (k_0)_j - (k_0)_i$$

and for the nulldecision in state $i \notin A_0$

$$k(i,x_0(i)) \overset{d}{=} 0.$$

DEFINITION 2.3.4. For each action $x \in X(i)$ and state $i \in \Psi$ a number $t(i,x) \in \mathbb{R}$ is defined. For x intervention in state $i \notin A_1$

$$t(i,x) \overset{d}{=} \sum_{j \in \Psi} p_{ij}^x (t_0)_j - (t_0)_i$$

and for the nulldecision in state $i \notin A_0$

$$t(i,x_0(i)) \overset{d}{=} 0.$$

The term $-(k_0)_i$ $(-(t_0)_i)$ appearing in the right-hand member of the definition of $k(i,x)$ $(t(i,x))$ is interpreted above. The term $G_i^x + \sum_{j \in \Psi} p_{ij}^x (k_0)_j$ $(\sum_{j \in \Psi} p_{ij}^x (t_0)_j)$ represents the expected total reward incurred (expected total time elapsed) if in initial state i intervention x is taken and after that the evolution of the system is described by the natural process until the set A_0 is entered. This interpretation covers also the case that in state i the nulldecision $x_0(i)$ is taken.

Next two matrices related to the natural process and a set of states A are introduced and some of their properties are summarized in the lemmas 2.3.5 and 2.3.6.

LEMMA 2.3.5. *Let* A *be a given set of states such that* $A_0 \subseteq A \subset \Psi$. *Then the matrix* $[I_{\bar{A}} - (N_0)_{\bar{A}}]$ *is nonsingular and its inverse*

$$R_0(\bar{A}) \overset{d}{=} [I_{\bar{A}} - (N_0)_{\bar{A}}]^{-1}$$

exists.

PROOF. Because $\overline{A_0} \supseteq \bar{A}$, $(N_0)_{\overline{A_0}}$ transient (cf. assumption 2.2.2 (iv)) implies $(N_0)_{\bar{A}}$ transient by lemma 1.2.2. The assertion follows then by lemma 1.2.1. □

LEMMA 2.3.6. *For a given set* A *satisfying* $A_0 \subseteq A \subset \Psi$ *the matrix* $U_0(\bar{A}) \in \mathbb{R}^{|\bar{A}| \times |A|}$ defined by

$$U_0(\bar{A}) \overset{d}{=} R_0(\bar{A}) \; (N_0)_{\bar{A}A}$$

has the following properties

(i) $U_0(\bar{A}) \geq 0$;

(ii) $U_0(\bar{A})\ 1_A = 1_{\bar{A}}$.

PROOF. By the lemmas 2.3.5 and 1.2.1 the entries of the matrix $R_0(\bar{A})$ are infinite sums of nonnegative real numbers implying $R_0(\bar{A}) \geq 0$. Because also $(N_0)_{\bar{A}A} \geq 0$ assertion (i) easily follows. By assumption 2.2.2(iii) and because $\bar{A}_0 \supseteq \bar{A}$

$$(N_0)_{\bar{A}}\ 1_{\bar{A}} + (N_0)_{\bar{A}A}\ 1_A = 1_{\bar{A}}.$$

Hence

$$\begin{aligned} U_0(\bar{A})\ 1_A &= R_0(\bar{A})(N_0)_{\bar{A}A}\ 1_A \\ &= R_0(\bar{A})[I_{\bar{A}} - (N_0)_{\bar{A}}]\ 1_{\bar{A}} \\ &= I_{\bar{A}}\ 1_{\bar{A}} = 1_{\bar{A}}\ , \end{aligned}$$

completing the proof. □

Observe that $R_0(\bar{A})$ is the 0^{th} moment of the MRM corresponding to SMM $N(t)_{\bar{A}}$ (cf. lemma 1.3.5 (2)). Each entry $U_0(\bar{A})_{ij}$, $i \notin A$, $j \in A$, is interpreted as the probability that state j is the first state taken on in the set A by the natural process with initial state $i \notin A$.

The following lemma states two relations expressing respectively $(k_0)_{\bar{A}}$ in $(k_0)_A$ and $(t_0)_{\bar{A}}$ in $(t_0)_A$ for a set of states A, $A_0 \subseteq A \subset \Psi$.

LEMMA 2.3.7. *For each set of states* A *such that* $A_0 \subseteq A \subset \Psi$, $(k_0)_{\bar{A}}$ *and* $(t_0)_{\bar{A}}$ *satisfy respectively*

(i) $(k_0)_{\bar{A}} = R_0(\bar{A})(G_0)_{\bar{A}} + U_0(\bar{A})(k_0)_A$;

(ii) $(t_0)_{\bar{A}} = R_0(\bar{A})\tau_{\bar{A}} + U_0(\bar{A})(t_0)_A$.

PROOF. Premultiplying (1) of definition 2.3.1 by $[I_{\bar{A}_0} - (N_0)_{\bar{A}_0}]$ shows that

$$(k_0)_{\bar{A}_0} = (G_0)_{\bar{A}_0} + (N_0)_{\bar{A}_0}(k_0)_{\bar{A}_0}.$$

Since $\overline{A_0} \supseteq \bar{A}$ and $(k_0)_{A_0} = 0$

$$(1) \qquad (k_0)_{\bar{A}} = (G_0)_{\bar{A}} + (N_0)_{\bar{A}}(k_0)_{\bar{A}} + (N_0)_{\bar{A}A}(k_0)_A.$$

By the lemmas 2.3.5 and 2.3.6, (1) is equivalent to

$$(2) \qquad (k_0)_{\bar{A}} = R_0(\bar{A})\,(G_0)_{\bar{A}} + U_0(\bar{A})\,(k_0)_A .$$

The relation (ii) follows by replacing k_0 and G_0 by t_0 and τ respectively. □

Some quantities related to the decision process for a fixed policy $z \in Z$ are specified in

DEFINITION 2.3.8. For each fixed policy $z \in Z$ the following quantities are defined.

(i) The vectors $k(z)$ and $t(z) \in \mathbb{R}^{J'}$ with components $k(i,z(i))$ and $t(i,z(i))$, $i \in \Psi$.

(ii) The matrix $S_0(z) \in \mathbb{R}^{J' \times J'}$ by

$$S_0(z) \overset{d}{=} \begin{bmatrix} (N_0)_{\overline{A(z)}\,\Psi} \\ \hline P(z)_{A(z)\,\Psi} \end{bmatrix} .$$

The Cesaro sum of nonnegative powers of $S_0(z)$ is defined in the usual way and denoted by $S_0^*(z)$. Observe that the matrix $S_0(z)$ is stochastic by assumption 2.2.2 (iii) and assumption 2.2.3. The set of persistent states is denoted by $E(z)$ and the set of all transient states by $F(z)$. The subchains of $S_0(z)$ are denoted by $E_\lambda(z)$, $\lambda = 1,\ldots,L(z)$.

(iii) The vectors $G_0(z) \in \mathbb{R}^{J'}$ and $\tau(z) \in \mathbb{R}_+^{J'}$ by

$$G_0(z) \overset{d}{=} \begin{bmatrix} (G_0)_{\overline{A(z)}} \\ \hline G_{A(z)} \end{bmatrix} \quad \text{and} \quad \tau(z) \overset{d}{=} \begin{bmatrix} \tau_{\overline{A(z)}} \\ \hline 0 \end{bmatrix}$$

respectively, where $G_{A(z)}$ denotes the vector with components $G_i^{z(i)}$, $i \in A(z)$, induced by the decision process of policy z.

(iv) The matrix $P_0(A(z)) \in \mathbb{R}_+^{|A(z)| \times |A(z)|}$ of the $A(z)$-embedded chain, by

$$P_0(A(z)) \overset{d}{=} P(z)_{A(z)} + P(z)_{A(z)\overline{A(z)}}\, U_0(\overline{A(z)}) .$$

Notice that $P_0(A(z))$ is stochastic by lemma 2.3.6 (ii) and assumption 2.2.3. The Cesaro sum of nonnegative powers of $P_0(A(z))$ is defined in

the usual way and denoted by $P_0^*(A(z))$. The matrix $H_0(A(z)) \in \mathbb{R}^{|A(z)|\times|A(z)|}$ is defined by (cf. section 1.2)

$$H_0(A(z)) \overset{d}{=} [I_{A(z)} - P_0(A(z)) - P_0^*(A(z))]^{-1} - P_0^*(A(z)).$$

The set of persistent (transient) states of the A(z)-embedded Markov chain is denoted by $E(A(z))(F(A(z)))$. The subchains of $P_0(A(z))$ are denoted by $E_\lambda(A(z))$, $\lambda = 1,\ldots,L'(z)$, where $L'(z)$ is the number of subchains of $P_0(A(z))$.

(v) The matrix $\Gamma_0(z) \in \mathbb{R}^{J'\times|A(z)|}$ by

$$\Gamma_0(z) \overset{d}{=} \begin{bmatrix} U_0(\overline{A(z)}) \\ \hline P_0(A(z)) \end{bmatrix}.$$

The i^{th} row of $\Gamma_0(z)$ is easily interpreted as the probability distribution of the first future intervention state given initial state $i \in \Psi$.

The next lemma shows the equivalence of the solution sets of the equations $\mu = S_0(z)\mu$ and $\mu = \Gamma_0(z)\mu_{A(z)}$ for a fixed policy $z \in Z$.

LEMMA 2.3.9. *For a vector* $\mu \in \mathbb{R}^{J'}$ *we have for fixed* $z \in Z$

$$\mu = S_0(z)\mu \iff \mu = \Gamma_0(z)\mu_{A(z)}.$$

PROOF. By the lemmas 2.3.5 and 2.3.6 and definition 2.3.8 (ii) we have

$$\mu_{\overline{A(z)}} = S_0(z)_{\overline{A(z)}}\ \mu_{\overline{A(z)}} + S_0(z)_{\overline{A(z)}\ A(z)}\ \mu_{A(z)}$$

$$= (N_0)_{\overline{A(z)}}\ \mu_{\overline{A(z)}} + (N_0)_{\overline{A(z)}\ A(z)}\ \mu_{A(z)},$$

which is equivalent to

$$\mu_{\overline{A(z)}} = U_0(\overline{A(z)})\mu_{A(z)}.$$

Using definition 2.3.8 (ii) and (iv)

$$\mu_{A(z)} = S_0(z)_{A(z)}\ \mu_{A(z)} + S_0(z)_{A(z)\ \overline{A(z)}}\ \mu_{\overline{A(z)}} =$$

$$= P(z)_{A(z)} \mu_{A(z)} + P(z)_{A(z)\overline{A(z)}} U_0^{(\overline{A(z)})} \mu_{A(z)}$$

$$= P_0(A(z))\mu_{A(z)}.$$

The use of definition 2.3.8 (v) completes the proof. □

The next lemma proves that the number of subchains in the decision process of a fixed policy $z \in Z$ equals the number of subchains of the A(z)-embedded Markov chain.

LEMMA 2.3.10. *For each fixed policy* $z \in Z$

$$L'(z) = L(z).$$

PROOF. Since $(N_0)_{\overline{A(z)}}$ is transient $E_\lambda(z) \cap A(z) \neq \emptyset$ for $\lambda = 1,\ldots,L(z)$ and $L'(z) \geq L(z)$. On the other hand by (2.3;5) $E_\lambda(z) \cap \overline{A(z)} \neq \emptyset$ for $\lambda = 1,\ldots,L(z)$ and $L'(z) \leq L(z)$. Hence $L'(z) = L(z)$. □

The lemmas 2.3.11, 2.3.12 and 2.3.13 establish some relations between $S_0^*(z)$, $P_0(A(z))$, $k(z)$ and $t(z)$.

LEMMA 2.3.11. *For each fixed policy* $z \in Z$ *we have*

(i) $S_0^*(z)_{\Psi A(z)} P(z)_{A(z)\overline{A(z)}} R_0^{(\overline{A(z)})} = S_0^*(z)_{\Psi\overline{A(z)}}$;

(ii) $S_0^*(z)_{\Psi A(z)} P_0(A(z)) = S_0^*(z)_{\Psi A(z)}$.

PROOF. By partitioning the matrices $S_0(z)$ and $S_0^*(z)$ the relation $S_0^*(z)\, S_0(z) = S_0^*(z)$ implies the four relations

(1) $$S_0^*(z)_{\overline{A(z)}} = S_0^*(z)_{\overline{A(z)}} (N_0)_{\overline{A(z)}} + S_0^*(z)_{\overline{A(z)}A(z)} P(z)_{A(z)\overline{A(z)}}$$

(2) $$S_0^*(z)_{\overline{A(z)}A(z)} = S_0^*(z)_{\overline{A(z)}} (N_0)_{\overline{A(z)}A(z)} + S_0^*(z)_{\overline{A(z)}A(z)} P(z)_{A(z)}$$

(3) $$S_0^*(z)_{A(z)\overline{A(z)}} = S_0^*(z)_{A(z)\overline{A(z)}} (N_0)_{\overline{A(z)}} + S_0^*(z)_{A(z)} P(z)_{A(z)\overline{A(z)}}$$

(4) $$S_0^*(z)_{A(z)} = S_0^*(z)_{A(z)\overline{A(z)}} (N_0)_{\overline{A(z)}A(z)} + S_0^*(z)_{A(z)} P(z)_{A(z)}.$$

Solving $S_0^*(z)_{A(z)\overline{A(z)}}$ from (3) and $S_0^*(z)_{\overline{A(z)}}$ from (1), which is permitted because $(N_0)_{\overline{A(z)}}$ is transient, yields

(5) $$S_0^*(z)_{\overline{A(z)}} = S_0^*(z)_{\overline{A(z)}A(z)} P(z)_{A(z)\overline{A(z)}} R_0(\overline{A(z)})$$

and

(6) $$S_0^*(z)_{A(z)\overline{A(z)}} = S_0^*(z)_{A(z)} P(z)_{A(z)\overline{A(z)}} R_0(\overline{A(z)}).$$

(5) and (6) constitute assertion (i). Substitution of (6) in (4), and (5) in (2) yields respectively

(7) $$S_0^*(z)_{A(z)} = S_0^*(z)_{A(z)} P_0(A(z))$$

and

(8) $$S_0^*(z)_{\overline{A(z)}A(z)} = S_0^*(z)_{\overline{A(z)}A(z)} P_0(A(z)).$$

(7) and (8) constitute assertion (ii). □

LEMMA 2.3.12. *For each policy* $z \in Z$ *the vector* $t(z)$ *satisfies*
(i) $t(z)_{A(z)} = [P_0(A(z)) - I_{A(z)}] (t_0)_{A(z)} + P(z)_{A(z)\overline{A(z)}} R_0(\overline{A(z)}) \tau_{\overline{A(z)}}$;
(ii) $S_0^*(z)t(z) = S_0^*(z)_{\Psi\overline{A(z)}} \tau_{\overline{A(z)}}$

PROOF. By its definition $t(z)_{A(z)}$ satisfies

(1) $$t(z)_{A(z)} = P(z)_{A(z)\Psi} t_0 - (t_0)_{A(z)}$$

$$= P(z)_{A(z)\overline{A(z)}} (t_0)_{\overline{A(z)}} + [P(z)_{A(z)} - I_{A(z)}](t_0)_{A(z)}.$$

By substitution of the relation (ii) of lemma 2.3.7 into (1) implies assertion (i). To prove assertion (ii) notice that by lemma 2.3.11 (ii), assertion (i) and lemma 2.3.11 (i)

$$S_0^*(z)t(z) = S_0^*(z)_{\Psi A(z)} t(z)_{A(z)}$$

$$= S_0^*(z)_{\Psi A(z)} P(z)_{A(z)\overline{A(z)}} R_0(\overline{A(z)}) \tau_{\overline{A(z)}}$$

$$= S_0^*(z)_{\Psi\overline{A(z)}} \tau_{\overline{A(z)}}$$ □

LEMMA 2.3.13. *For each policy* $z \in Z$ *the vector* $k(z)$ *satisfies*
(i) $k(z)_{A(z)} = [P_0(A(z)) - I_{A(z)}](k_0)_{A(z)} + G_0(z)_{A(z)} +$
$+ P(z)_{A(z)\overline{A(z)}} R_0(\overline{A(z)}) G_0(z)_{\overline{A(z)}}$;
(ii) $S_0^*(z)k(z) = S_0^*(z)G_0(z)$.

PROOF. By its definition $k(z)_{A(z)}$ satisfies

$$(1) \qquad k(z)_{A(z)} = G_0(z)_{A(z)} + P(z)_{A(z)\Psi}\, k_0 - (k_0)_{A(z)} =$$

$$= G_0(z)_{A(z)} + P(z)_{A(z)\overline{A(z)}}\,(k_0)_{\overline{A(z)}} + [P(z)_{A(z)} - I_{A(z)}](k_0)_{A(z)}.$$

Substitution of lemma 2.3.7 (i) into (1) implies assertion (i). To prove assertion (ii) observe that by lemma 2.3.11 (ii), assertion (i) and lemma 2.3.11 (i)

$$S_0^*(z)k(z) = S_0^*(z)_{\Psi A(z)}\, P(z)_{A(z)\overline{A(z)}}\, R_0(\overline{A(z)})G_0(z)_{\overline{A(z)}} +$$

$$+ S_0^*(z)_{\Psi A(z)}\, G_0(z)_{A(z)} =$$

$$= S_0^*(z)_{\Psi\overline{A(z)}}\, G_0(z)_{\overline{A(z)}} + S_0^*(z)_{\Psi A(z)}\, G_0(z)_{A(z)} =$$

$$= S_0^*(z)G_0(z). \qquad \square$$

The following theorems are concerned with the general solution to a set of equations arising in the policy evaluation operation of the policy iteration method.

THEOREM 2.3.14. *Let $z \in Z$ be a fixed policy. Let for each fixed $\lambda \in \{1,\ldots,L(z)\}$ $\mu = \phi_\lambda(A(z))$ be the solution to the equation $\mu = P_0(A(z))\mu$ with $\phi_\lambda(A(z))_i = 1$ for $i \in E_\lambda(A(z))$ and $\phi_\lambda(A(z))_i = 0$ for $i \in E(A(z))\backslash E_\lambda(A(z))$. Then the set of equations*

$$(1) \qquad \begin{cases} \text{(a)} & y(z)_{A(z)} = P_0(A(z))y(z)_{A(z)} \\ \text{(b)} & w_{A(z)} = P_0(A(z))w_{A(z)} + (k(z) - y(z)\square t(z))_{A(z)} \end{cases}$$

has a solution $(y(z)_{A(z)}, w_{A(z)})$. The vector $y(z)_{A(z)}$ is uniquely given by

$$(2) \qquad y(z)_{A(z)} = \sum_{\lambda=1}^{L(z)} y_\lambda(z)\, \phi_\lambda(A(z)).$$

The scalars $y_\lambda(z)$ are given by

$$(3) \qquad y_\lambda(z) = \frac{\langle S_0^*(z)_i, k(z)\rangle}{\langle S_0^*(z)_i, t(z)\rangle} \qquad \text{for } \lambda = 1,\ldots,L(z),$$

where $S_0^*(z)_i$ *is the row of matrix* $S_0^*(z)$ *corresponding to an arbitrary state* $i \in E_\lambda(A(z))$. *The general solution for* $w_{A(z)}$ *is given by*

$$(4) \qquad w_{A(z)} = H_0(A(z))(k(z) - y(z)\square t(z))_{A(z)} + \sum_{\lambda=1}^{L(z)} \beta_\lambda \phi_\lambda(A(z)),$$

where the β_λ, $\lambda = 1,\ldots,L(z)$ *are arbitrary constants.*

PROOF. By lemma 1.2.4 the general solution of (1)(a) is given by

$$(5) \qquad y(z)_{A(z)} = \sum_{\lambda=1}^{L(z)} \alpha_\lambda \phi_\lambda(A(z)),$$

where the α_λ, $\lambda = 1,\ldots,L(z)$ are constants still to be specified. To specify the α_λ, (1)(b) is premultiplied by the matrix $S_0^*(z)_{\Psi A(z)}$ to yield by lemma 2.3.11 (iii)

$$(6) \qquad S_0^*(z)_{\Psi A(z)}(k(z) - y(z)\square t(z))_{A(z)} = 0.$$

Substitution of (5) in (6) yields for each $i \in E_\lambda(A(z))$ with $\lambda \in \{1,\ldots,L(z)\}$ fixed

$$(7) \qquad \langle S_0^*(z)_i, k(z)\rangle - \alpha_\lambda \langle S_0^*(z)_i, t(z)\rangle = 0.$$

Since by lemma 2.3.11 (ii) and $\tau(z)_i > 0$ for $i \in \overline{A(z)}$

$$(8) \qquad \langle S_0^*(z)_i, t(z)\rangle = \langle S_0^*(z)_i, \tau(z)\rangle > 0,$$

α_λ is uniquely determined by (7).

Defining $y_\lambda(z) = \alpha_\lambda$ completes the proof of (3). By lemma 1.2.8 with $b \overset{d}{=} (k(z) - y(z)\square t(z))_{A(z)}$ and by lemma 1.2.4 the general solution of (1)(b) is given by (4). □

COROLLARY 2.3.15. *A particular solution of* (1)(b) *of theorem* 2.3.14 *is obtained if in each subchain* $E_\lambda(A(z))$, $\lambda = 1,\ldots,L(z)$, *one arbitrary state* $i(\lambda)$ *is chosen for which* $w_{i(\lambda)} = 0$. *The scalars* β_λ *in* (4) *are in this case specified by*

$$(1) \qquad \beta_\lambda = -[H_0(A(z))(k(z) - y(z)\square t(z))]_{i(\lambda)}.$$

THEOREM 2.3.16. *If the constants* β_λ, $\lambda = 1,\ldots,L(z)$ *in theorem 2.3.14 are specified then the system of equations in* $y(z)$ *and* w

$$(1)\qquad \begin{cases} y(z) = \Gamma_0(z)\ y(z)_{A(z)} \\ w \quad = \Gamma_0(z)\ w_{A(z)} + k(z) - y(z)\square t(z) \end{cases}$$

has a unique solution.

PROOF. The subvectors $y(z)_{A(z)}$ and $w_{A(z)}$ are uniquely determined by theorem 2.3.14 and corollary 2.3.15. By the lemmas 2.3.5 and 2.3.6 also $y(z)_{\overline{A(z)}}$ and $w_{\overline{A(z)}}$ are uniquely determined by

$$y(z)_{\overline{A(z)}} = U_0(\overline{A(z)})\ y(z)_{A(z)}$$

and

$$w_{\overline{A(z)}} = U_0(\overline{A(z)})\ w_{A(z)},$$

completing the proof. □

Finally a policy iteration method for the finite GMP-model is stated. In the sequel the method will be referred to as GMP1.

METHOD 2.3.17 (GMP1). The method consists of the following five main operations:

(i) Preparatory part. Compute the vectors k_0 and t_0 and the real numbers $k(i,x)$ and $t(i,x)$ for $i \in \Psi$, $x \in X(i)$. Fix an initial policy $z \in Z$.

(ii) Policy evaluation operation. Compute the particular solution $(y(z),w(z))$ to the system of equations of theorem 2.3.16 obtained by taking for each state $i(\lambda)$, $\lambda = 1,\ldots,L(z)$, the state with the largest index in subchan $E_\lambda(A(z))$.

(iii) Policy improvement operation. Introduce the notation

$$(S_0)^x_{ij} \overset{d}{=} \begin{cases} (N_0)_{ij} & \text{for } x = x_0(i),\ i \notin A_0 \\ P^x_{ij} & \text{for } x \in X(i),\ x \neq x_0(i),\ i \in \Psi \end{cases},$$

and

$$X_i(i) \overset{d}{=} \begin{cases} X(i)\backslash\{x_0(i)\} & \text{for } i \in A(z)\backslash A_0 \\ X(i) & \text{otherwise.} \end{cases}$$

Compute:

(1) the vector $\hat{y} \in \mathbb{R}^J$ with components $\hat{y}_i$ with

$$\hat{y}_i \overset{d}{=} \max_{x \in X_1(i)} \left[\sum_{j \in \Psi} (S_0)^x_{ij} \, y(z)_j \right] ,$$

(2) the set of actions in each state $i \in \Psi$ given by

$$X_2(i) \overset{d}{=} \left\{ x \in X_1(i) : \sum_{j \in \Psi} (S_0)^x_{ij} \, y(z)_j = \hat{y}_i \right\} ,$$

(3) the vector $\hat{w} \in \mathbb{R}^J$ with components $\hat{w}_i$ given by

$$\hat{w}_i \overset{d}{=} \max_{x \in X_2(i)} \left[k(i,x) - \hat{y}_i t(i,x) + \sum_{j \in \Psi} (S_0)^x_{ij} \, w(z)_j \right] ,$$

(4) the set of actions in each state $i \in \Psi$ given by

$$X_3(i) \overset{d}{=} \left\{ x \in X_2(i) : k(i,x) - \hat{y}_i t(i,x) + \sum_{j \in \Psi} (S_0)^x_{ij} \, w(z)_j = \hat{w}_i \right\},$$

(5) policy $\hat{z} \in Z$ such that $\hat{z}(i) = z(i)$ whenever $z(i) \in X_3(i)$. If $z(i) \notin X_3(i)$ then choose $\hat{z}(i) \in X_3(i)$ arbitrarily.

(iv) Cutting operation. Let A satisfy $A_0 \subseteq A \subseteq A(\hat{z})$. For each set A define the vectors $y'(A)$ and $w'(A)$ respectively by

$$y'(A) = \begin{bmatrix} U_0(\bar{A}) \\ \hline I_A \end{bmatrix} \hat{y}$$

and

$$w'(A) = \begin{bmatrix} U_0(\bar{A}) \\ \hline I_A \end{bmatrix} \hat{w} \quad .$$

Define the collection of sets

$$M(\hat{z}) \overset{d}{=} \{ A_0 \subseteq A \subseteq A(\hat{z}) : [y'(A)_{A(\hat{z})}, w'(A)_{A(\hat{z})}] \geq [\hat{y}_{A(\hat{z})}, \hat{w}_{A(\hat{z})}] \}.$$

Define the set

$$A^*(\hat{z}) \overset{d}{=} \bigcap_{A \in M(\hat{z})} A$$

and policy z' such that

$$z'(i) = \begin{cases} \hat{z}(i) & \text{for } i \in A(z') \\ x_0(i) & \text{for } i \in \overline{A(z')} \end{cases} ,$$

with A(z') defined by

$$A(z') = A^*(\hat{z}) \cup \{i \in \overline{A^*(\hat{z})} \cap A(\hat{z}): y'(A^*(\hat{z}))_i = \hat{y}_i \text{ and } w'(A^*(\hat{z}))_i = \hat{w}_i\}.$$

(v) If $z'(i) = z(i)$ for $i \in \Psi$ then stop. Otherwise redefine z equal to z' and repeat operations (ii)...(v).

REMARK 2.3.18.

(i) A large part of the computation of k(z) and t(z) for a particular policy z is done in the policy independent preparatory part.

(ii) The collection $M(\hat{z})$ is nonempty because $A(\hat{z}) \in M(\hat{z})$.

(iii) $A(z'), A^*(\hat{z}) \in M(\hat{z})$ is proved in chapter IV.

(iv) No specific computational procedure is indicated in operation (iv) to compute the sets $A^*(\hat{z})$ and A(z'). We shall return to this problem in chapter IV.

Finally the relationship between the undiscounted GMP-model and the undiscounted MRD-model with interventions is exhibited.

Let Π_0 denote the class of problems satisfying the assumptions of the undiscounted MRD-model with interventions (cf. section 1.4). Let Π_2 denote the class of problems satisfying the assumptions of the undiscounted GMP-model. Then it is immediately verified that $\pi \in \Pi_2$ implies $\pi \in \Pi_0$. Conversely let $\pi \in \Pi_0$. A transformation Θ to be applied on $\pi \in \Pi_0$ is specified in

DEFINITION 2.3.19. The transformation Θ transforms a problem $\pi \in \Pi_0$ into a problem $\Theta(\pi)$ by the following steps:

(i) Define a nonempty set A_0 such that

$$M \supseteq A_0 \supseteq \{i \in M: T_i^k = 0 \text{ for } k \in K(i)\}.$$

If the right-hand set is empty then A_0 may be any nonempty subset of M.

(ii) For each $i \in M \backslash A_0$ if $A_0 \neq M$ specify a particular action $k \in K(i)$ satisfying $T_i^k > 0$. Denote this action by $x_0(i)$.

(iii) Extend the set of states to

$$\Psi \stackrel{d}{=} M \cup \{m = (i,k): T_i^k > 0,\ k \in K(i) \backslash \{x_0(i)\},\ i \in M\}.$$

(iv) Define a substochastic matrix N_0 with entries

$$(N_0)_{mj} \stackrel{d}{=} \begin{cases} (Q_0)_{ij}^k & \text{for } m = (i,k) \in \Psi\backslash M,\ j \in M \\ & \text{or } m = i \in M\backslash A_0,\ k = x_0(i),\ j \in M, \\ 0 & \text{otherwise,} \end{cases}$$

a vector τ with components

$$\tau_m \stackrel{d}{=} \begin{cases} T_i^k & \text{for } m = (i,k) \in \Psi\backslash M \text{ or } m = i \in M\backslash A_0,\ k = x_0(i) \\ 0 & \text{otherwise} \end{cases}$$

and a reward vector G_0 with components

$$(G_0)_m \stackrel{d}{=} \begin{cases} (C_0)_i^k & \text{for } m = (i,k) \in \Psi\backslash M \text{ or } m = i \in M\backslash A_0,\ k = x(i) \\ 0 & \text{otherwise.} \end{cases}$$

(v) For each pair (i,k), $i \in M$, $k \in K(i)\backslash\{x_0(i)\}$ such that $T_i^k > 0$, an intervention x in state i is defined inducing a reward $G_i^x = 0$ and a deterministic transformation to the state $(i,k) \in \Psi\backslash M$.

(vi) The set $X(i)$, $i \in \Psi$, is defined by

$$X(i) = \begin{cases} K(i) & \text{for } i \in M \\ \{x_0(i)\} & \text{for } i \in \Psi\backslash M. \end{cases}$$

It is easily shown that $\Theta(\pi) \in \Pi_2$ at least if $A_0 = M$. Depending on the structure of the problem under consideration also other choices of A_0 are possible to guarantee that $\Theta(\pi) \in \Pi_2$. The freedom of choice of the set A_0 can then be exploited in the computation of a gain-optimal policy. Observe that in this respect the number of states in Ψ is maximal and the cutting operation is superfluous if $A_0 = M$ is chosen.

CHAPTER III

ON THE CONVERGENCE OF GMP-SCHEMES

3.1. INTRODUCTION

In this chapter a more general version of the policy iteration method GMP1 for the finite GMP-model is considered. This version is called a GMP-*scheme*. It contains GMP1 and other variants, to be specified in Chapter IV as special cases.

A GMP-scheme has the preparatory part and the policy evaluation operation in common with GMP1 but replaces operations (iii), (iv) and (v) by an operation called a *compound policy improvement operation*. The computations to be executed in this operation are not specified but only the properties are stated which are required to obtain the final result of section 3.2: convergence within a finite number of steps to a gain-optimal policy. The properties imposed upon this compound policy improvement operation are specified by definitions 3.1.1 - 3.1.3.

DEFINITION 3.1.1. Let $z,z' \in Z$ be policies. Let $(y(z),w(z))$ be the solution to the system of equations in the policy evaluation operation of GMP1, obtained in the way indicated by corollary 2.3.15, for a particular policy z. The components of the vectors $y((z')z),w((z')z) \in \mathbb{R}^{J'}$ are for $i \in A(z')$ defined by

$$(1) \qquad y((z')z)_i \stackrel{d}{=} \sum_{j \in \Psi} S_0(z')_{ij}\, y(z)_j$$

and

$$(2) \qquad w((z')z)_i \stackrel{d}{=} \sum_{j \in \Psi} S_0(z')_{ij}\, w(z)_j + k(z')_i - y((z')z)_i\, t(z')_i.$$

The components of $y((z')z)$ and $w((z')z)$ for $i \notin A(z')$ are defined by

$$(3) \qquad y((z')z)_i \stackrel{d}{=} \sum_{j \in \Psi} S_0(z')_{ij}\, y((z')z)_j$$

and

$$(4) \qquad w((z')z)_i \overset{d}{=} \sum_{j\in\Psi} S_0(z')_{ij}\, w((z')z)_j .$$

Observe that in (1) and (2) the vectors $y(z)$ and $w(z)$ are involved while (3) and (4) are relations among the components of the vectors $y((z')z)$ and $w((z')z)$ respectively. This difference is typical for a GMP-scheme.

DEFINITION 3.1.2. For each policy $z \in Z$ the sets of policies $D(z)$ and $D^*(z)$ are defined by

$$D(z) = \{z' \in Z: [y((z')z), w((z')z)] \succeq [y(z), w(z)]\}$$

and

$$D^*(z) = \{z' \in Z: [y((z')z), w((z')z)] \succ [y(z), w(z)]\}.$$

Observe that $D(z)$ is nonempty because $z \in D(z)$ for $z \in Z$. However $D^*(z)$ may be empty.

DEFINITION 3.1.3. An operation computing a policy $z' \in D(z)$ and redefining z by $z \overset{d}{=} z'$ is called a *compound policy improvement operation*. An iterative method consisting of the preparatory part of GMP1, the policy evaluation operation of GMP1 and a compound policy improvement operation is called a *GMP-scheme*.

Definition 3.1.3 specifies a minimum requirement for a method to be called a GMP-scheme. Two additional conditions, required for finite step convergence to a gain-optimal policy, are given in

DEFINITION 3.1.4. A GMP-scheme is called

(i) *distinctive* if $z' \in D^*(z)$ whenever $D^*(z) \neq \emptyset$,

(ii) *preserving* if $z'(i) = z(i)$ whenever $y((z')z)_i = y(z)_i$ and $w((z')z)_i = w(z)_i$.

3.2. A FINITE STEP CONVERGENCE PROOF FOR DISTINCTIVE AND PRESERVING GMP-SCHEMES

Because the policy evaluation operation of GMP1 is maintained in a GMP-scheme the definitions and results of section 2.3 apply to a GMP-scheme. Only some additional notation is needed. For each policy $z \in Z$ we define

$$(3.2;1) \qquad E(\overline{A(z)}) \overset{d}{=} E(z) \cap \overline{A(z)}$$

and

(3.2;2) $\quad F(\overline{A(z)}) \overset{d}{=} F(z) \cap \overline{A(z)}$.

In relation to the compound policy improvement operation two additional vectors are specified by

DEFINITION 3.2.1. The vectors $\bar{y}((z')z)$ and $\bar{w}((z')z) \in \mathbb{R}^{J'}$ are defined by

(1) $\quad \bar{y}((z')z) = S_0(z')y((z')z)$

and

(2) $\quad \bar{w}((z')z) = S_0(z')w((z')z) + k(z') - y((z')z)\Box t(z')$.

The following theorem specifies some properties of a GMP-scheme.

THEOREM 3.2.2. *Let, for a given policy* $z \in Z$, $z' \in D(z)$ *be the policy obtained by the compound policy improvement operation of a GMP-scheme. Then*

(i) $\bar{y}((z')z)_i = y((z')z)_i = \bar{y}((z')z)_j = y((z')z)_j$ *for* $i,j \in E_\lambda(z')$ *and each fixed* $\lambda \in \{1,\ldots,L(z')\}$;

(ii) $\bar{w}((z')z)_i \geq w((z')z)_i$ for $i \in E(A(z'))$;

(iii) $y(z')_{E(z')} \geq y((z')z)_{E(z')} \geq y(z)_{E(z')}$;

(iv) $y(z')_{E(z')} = y(z)_{E(z')}$ *implies* $\bar{w}((z')z)_{E(z')} = w((z')z)_{E(z')}$;

(v) $y(z')_{F(z')} \geq y((z')z)_{F(z')} \geq y(z)_{F(z')}$;

(vi) $y(z') \geq y(z)$.

If in definition 3.1.2 *the symbol* $\geq$ *is replaced by* $\leq (=)$ *then* (i) - (vi) *hold with reversed inequality signs (equality signs).*

PROOF. By the definitions 3.1.1 and 3.2.1 we have

(1) $\quad \bar{y}((z')z) = S_0(z')y((z')z) \geq y((z')z) \geq y(z)$.

Applying lemma 1.2.7 to (1) per subchain implies assertion (i). Assertion (i) implies

(2) $\quad \sum_{j \in \Psi} S_0(z')_{ij}[y((z')z)_j - y(z)_j] = 0 \qquad$ for $i \in E(A(z'))$

and consequently

(3) $\quad \sum_{j \in \Psi} S_0(z')_{ij}[w((z')z)_j - w(z)_j] \geq 0 \qquad$ for $i \in E(A(z'))$.

Adding $k(z')_i - y((z')z)_i t(z')_i$ to both sides of (3) yields

$$(4)\qquad \bar{w}((z')z)_i = \sum_{j\in\Psi} S_0(z')_{ij} w((z')z)_j + k(z')_i - y((z')z)_i t(z')_i$$

$$\geq \sum_{j\in\Psi} S_0(z')_{ij} w(z)_j + k(z')_i - y((z')z)_i t(z')_i$$

$$= w((z')z)_i \qquad \text{for } i \in E(A(z')),$$

which completes the proof of assertion (ii). By the definitions 3.1.1 and 3.2.1 we have for $i \notin A(z')$

$$(5)\qquad \bar{w}((z')z)_i = \sum_{j\in\Psi} S_0(z)_{ij} w((z')z)_j = w((z')z)_i;$$

(4) and (5) imply

$$(6)\qquad \bar{w}((z')z_{E(z')} = S_0(z')_{E(z')} w((z')z)_{E(z')} +$$

$$+ [k(z') - y((z')z) \square t(z')]_{E(z')}$$

$$\geq w((z')z)_{E(z')}.$$

Premultiplying (6) by $S_0^*(z')_{E(z')}$ yields

$$(7)\qquad S_0^*(z')_{E(z')} [k(z') - y((z')z) \square t(z')]_{E(z')} \geq 0,$$

implying the left-hand part of assertion (iii). The right-hand part of assertion (iii) is implied by $z' \in D(z)$. To prove assertion (iv) notice that $y(z')_{E(z')} = y(z)_{E(z')}$ implies the equality sign in (7) and also

$$(8)\qquad S_0^*(z')_{E(z')} [\bar{w}((z')z) - w((z')z)]_{E(z')} = 0.$$

By lemma 1.2.5 (ii) and (iii) applied to the matrix $S_0^*(z)$ (8) implies assertion (iv). To prove (v) notice that for $y((z')z)_{F(z')}$ we have by (1) and assertion (iii)

$$(9)\qquad y((z')z)_{F(z')} \leq$$

$$\leq S_0(z')_{F(z')} y((z')z)_{F(z')} + S_0(z')_{F(z')E(z')} y((z')z)_{E(z')}$$

$$\leq S_0(z')_{F(z')} y((z')z)_{F(z')} + S_0(z')_{F(z')E(z')} y(z')_{E(z')}.$$

$y(z')_{F(z')}$ satisfies by theorem 2.3.16 and lemma 2.3.9

$$(10) \qquad y(z')_{F(z')} = S_0(z')_{F(z')} y(z')_{F(z')} + S_0(z')_{F(z')E(z')} y(z')_{E(z')}.$$

The application of lemma 1.2.3 (iii) with $u^* \overset{d}{=} y(z')_{F(z')}$ and $u \overset{d}{=} y((z')z)_{F(z')}$ yields $y(z')_{F(z')} \geq y((z')z)_{F(z')}$ and therefore assertion (v). Assertion (vi) is the immediate consequence of assertions (iii) and (v). Observe that if in the definition of D(z) the $\succeq$ sign is replaced by $\preceq(=)$ the proof remains valid with reversed inequality signs (equality signs throughout). □

In the remaining part of this section, let $\{z_n, n = 1,2,\ldots\}$ be a sequence of policies generated by a GMP-scheme, where z_1 is arbitrarily chosen.

LEMMA 3.2.3. *For* $n \geq M_0$, M_0 *some finite natural number, we have in a GMP-scheme*

(i) $y(z_{n+1}) = \bar{y}((z_{n+1})z_n) = y((z_{n+1})z_n) = y(z_n)$;

(ii) $\bar{w}((z_{n+1})z_n)_i = w((z_{n+1})z_n)_i$ *for* $i \in E(z_{n+1})$.

PROOF. Since Z is a finite set only a finite number of strict improvements in y(z) is possible. This implies assertion (i). Assertion (ii) is an immediate consequence of theorem 3.2.2 (iv). □

LEMMA 3.2.4. *For* $n \geq M_0$, *where* M_0 *is specified by lemma* 3.2.3, *we have in a preserving* GMP-*scheme*

$$(1) \qquad \begin{cases} \bar{w}((z_{n+1})z_n)_i = w((z_{n+1})z_n)_i = w(z_n)_i \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } i \in E(z_{n+1}). \\ z_{n+1}(i) = z_n(i) \end{cases}$$

PROOF. Fix $n \geq M_0$ and let B be the possibly empty set

$$(1) \qquad B \overset{d}{=} \{i \in \Psi: \bar{w}((z_{n+1})z_n)_i = w((z_{n+1})z_n)_i = w(z_n)_i\}.$$

Let $E_\lambda(z_{n+1})$ be an arbitrary subchain of $S_0(z_{n+1})$. By lemma 3.2.3 (ii) we have for $i \in E_\lambda(A(z_{n+1}))$

$$(2)\qquad \sum_{j\in E_\lambda(z_{n+1})} S_0(z_{n+1})_{ij}[w((z_{n+1})z_n)_j - w(z_n)_j] = 0.$$

(2) and lemma 3.2.3 (ii) imply that $j \in B$ for some $j \in E_\lambda(z_{n+1})$ satisfying $S_0(z_{n+1})_{ij} > 0$. Hence $B \neq \emptyset$ and $B \cap E_\lambda(z_{n+1}) \neq \emptyset$. By lemma 3.2.3 (i) and the preserving condition we have $z_{n+1}(i) = z_n(i)$ for $i \in B$.
By (1) we have for $j \in B \cap A(z_{n+1})$

$$(3)\qquad \sum_{\ell\in\Psi} S_0(z_{n+1})_{j\ell}[w((z_{n+1})z_n)_\ell - w(z_n)_\ell] = 0$$

and for $j \in B \cap \overline{A(z_{n+1})}$ since $z_{n+1}(j) = z_n(j)$

$$(4)\qquad \sum_{\ell\in\Psi} S_0(z_{n+1})_{j\ell} w((z_{n+1})z_n)_\ell = w((z_{n+1})z_n)_j = w(z_n)_j =$$

$$= \sum_{\ell\in\Psi} S_0(z_{n+1})_{j\ell}\, w(z_n)_\ell \ .$$

(4) implies that (3) holds for $j \in B$. Let $B_\lambda \overset{d}{=} B \cap E_\lambda(z_{n+1})$. For $j \in B_\lambda$ we have then

$$(5)\qquad \sum_{\ell\in E_\lambda(z_{n+1})} S_0(z_{n+1})_{j\ell}[w((z_{n+1})z_n)_\ell - w(z_n)_\ell] = 0,$$

implying $k \in B_\lambda$ for each $k \in E_\lambda(z_{n+1})$ such that $S_0(z_{n+1})_{jk} > 0$. This implies $B_\lambda = E_\lambda(z_{n+1})$. Since $E_\lambda(z_{n+1})$ was arbitrary this completes the proof. □

LEMMA 3.2.5. *For* $n \geq M_1$, $M_1 \geq M_0$ *some finite natural number and* M_0 *specified by lemma* 3.2.3, *we have for a preserving* GMP-*scheme*

(i) $E(z_{n+1}) = E(z_n)$;

(ii) $w(z_{n+1})_i = w(z_n)_i$ *for* $i \in E(z_n)$.

PROOF. By lemma 3.2.4 we have $z_{n+1}(i) = z_n(i)$ for $i \in E(z_{n+1})$, $n \geq M_0$. This implies $E(z_{n+1}) \subseteq E(z_n)$. Since the set of policies Z is finite, assertion (i) follows. Since for each subchain $E_\lambda(z_n)$, $\lambda \in \{1,\ldots,L(z_n)\}$, $n \geq M_1$ we have $y(z_{n+1})_{E_\lambda(z_n)} = y(z_n)_{E_\lambda(z_n)}$ by lemma 3.2.3. By assertion (i) and the policy evaluation operation of method 2.3.17 the states $i(\lambda)$, $\lambda \in \{1,\ldots,L(z_n)\}$ are identical for $n \geq M_1$. This implies assertion (ii). □

LEMMA 3.2.6. *For* $n \geq M_2$, $M_2 \geq M_1$ *some finite natural number and* M_1 *specified by lemma* 3.2.5, *a preserving* GMP-*scheme converges, i.e. it has*

$$z_{n+1} = z_n \qquad \text{for } n \geq M_2.$$

PROOF. For $n \geq M_1$, the lemmas 3.2.3, 3.2.4 and 3.2.5 imply $F(z_{n+1}) = F(z_n)$. Since $y(z_{n+1}) = y(z_n)$ we have $w((z_{n+1})z_n)_i > w(z_n)_i$ whenever $z_{n+1}(i) \neq z_n(i)$ by definition 3.1.4 (ii). By the definitions 3.1.2 and 3.2.1 and lemma 3.2.3 we have for $n \geq M_1$

$$\bar{w}((z_{n+1})z_n) = S_0(z_{n+1})w((z_{n+1})z_n) + k(z_{n+1}) - y(z_{n+1}) \,\square\, t(z_{n+1}) \geq \tag{1}$$

$$\geq w((z_{n+1})z_n) \geq w(z_n).$$

By the lemmas 3.2.4 and 3.2.5 (ii)

$$w(z_{n+1})_{E(z_n)} = w((z_{n+1})z_n)_{E(z_n)} = w(z_n)_{E(z_n)} \tag{2}$$

implying that $w((z_{n+1})z_n)_{F(z_{n+1})}$ satisfies

$$w((z_{n+1})z_n)_{F(z_{n+1})} \leq S_0(z_{n+1})_{F(z_{n+1})}\, w((z_{n+1})z_n)_{F(z_{n+1})} + \tag{3}$$

$$+ S_0(z_{n+1})_{F(z_{n+1})E(z_{n+1})}\, w(z_{n+1})_{E(z_{n+1})} +$$

$$+ [k(z_{n+1}) - y(z_{n+1}) \,\square\, t(z_{n+1})]_{F(z_{n+1})}.$$

Applying lemma 1.2.3 (iii) with $u \overset{d}{=} w((z_{n+1})z_n)_{F(z_{n+1})}$ and $u^* \overset{d}{=} w(z_{n+1})_{F(z_{n+1})}$ yields for $n \geq M_1$

$$w(z_n)_{F(z_{n+1})} \leq w((z_{n+1})z_n)_{F(z_{n+1})} \leq w(z_{n+1})_{F(z_{n+1})}.$$

Hence $z_{n+1}(i) \neq z_n(i)$ for $i \in F(z_{n+1})$ also implies $w(z_{n+1})_i > w(z_n)_i$. Since the number of policies is finite we have the assertion. □

Finally the convergence of a distinctive and preserving GMP-scheme to a gain-optimal policy is the subject of

THEOREM 3.2.7. *A distinctive and preserving GMP-scheme converges within a finite number of steps to a policy* z^* *satisfying*

$$y(z^*) \geq y(z) \qquad \text{for } z \in Z. \tag{1}$$

PROOF. Since in a distinctive and preserving GMP-scheme $z_{n+1} \in D^*(z_n)$ at

each step and the number of policies is finite we have $z_{n+1} = z_n$ and $D^*(z_n) = \emptyset$ for $n \geq M_2$. Defining $z^* \stackrel{d}{=} z_{M_2}$ we have for each policy $z \in Z$

(2) $$[y((z)z^*), w((z)z^*)] \preceq [y(z^*), w(z^*)].$$

The assertion follows by theorem 3.2.2 and (2). □

REMARK 3.2.8. The finite step convergence proof, given here for a GMP-scheme, differs in some respects from a standard proof of method 1.4.1 (cf. [DENARDO & FOX 1968]). The modifications implied by the structure of the compound policy improvement operation are rather straightforwardly. Only the argument to prove the implication $y(z_{n+1}) = y(z_n) \Rightarrow w((z_{n+1})z_n)_i = w(z_n)_i$ for $i \in E(z_{n+1})$ is necessarily different and relies on the preserving property postulated in lemma 3.2.4. Further the proof as a whole is constructed in such a way that a property of the convergence is exhibited. It shows that primarily convergence in the subchains is obtained and afterwards in the transient states.

CHAPTER IV

CUTTING METHODS AND OPTIMAL STOPPING

4.1. INTRODUCTION

In this chapter three policy iteration methods are developed for the finite GMP-model. All three methods are proven to be distinctive and preserving GMP-schemes. Theorem 3.2.7 then proves their convergence within a finite number of steps to a gain-optimal policy. An investigation of their computational performance is presented in chapter V.

The first method considered is GMP1 (method 2.3.17). As remarked in section 2.3 (cf. remark 2.3.18) the formulation of the cutting operation of GMP1 does not provide a specific computational procedure. To develop such a procedure the relationship between the cutting operation of GMP1 and the more basic problem of optimal stopping in a finite Markov chain is exposed in section 4.3. An optimal stopping problem in a finite Markov chain will be abbreviated by OSP with plural form OSPs. It is shown that the sets $A(z')$ and $A(z^*)$ of GMP1 are stopping sets which are optimal in a lexicographical sense to a coupled pair of OSPs. In preparation some properties of coupled pairs of OSPs are proven in section 4.2.

In section 4.4 the second method, denoted by GMP2, is developed. It is identical to the operations of GMP1 with the exception of the cutting operation, which is replaced by a suboptimal cutting method. A suboptimal cutting method computes, roughly speaking, a stopping set, which is "better" in a lexicographical sense than the set $A(\hat{z})$.

Finally in section 4.5 a third method, GMP3, is presented, which uses a suboptimal cutting method involving the vectors $y(z)$ and $w(z)$ of the current policy z. In GMP3 the policy improvement operation and the cutting operation can be combined to a single operation, which is identical to operation (iii) of method 2.3.17 with $X_1(i) \stackrel{d}{=} X(i)$ for all $i \in \Psi$.

4.2. OPTIMAL STOPPING AND OPTIMAL CUTTING

In this section some properties of a coupled pair of OSPs are proven. Primarily a brief review of optimal stopping in a Markov chain is given. For a discussion of this model the reader is referred to [BREIMANN 1964] and for a treatment as an optimal gain problem to [DERMAN 1970]. In [EMRICH 1970] algorithms computing optimal stopping policies are developed. In [HORDIJK, POTHARST & RUNNENBURG 1972] optimal stopping problems with a countable state space are considered. This latter study is based on and includes several extensions of material treated in [DYNKIN & JUSCHKEWITSCH 1969].

Suppose a Markov chain with set of states $M = \{1,\ldots,J\}$ and a substochastic matrix P is given. In each state $i \in M$ one may choose among at most two alternatives. The first alternative is to stop in that state, implying an income λ_i. The second alternative is to continue and earn nothing. The objective is to find a policy specifying the alternative per state such that the expected income is maximized. Such a policy is called an *optimal stopping policy*.

An optimal stopping problem is summarized by the 4-tuple

$$(4.2;1) \qquad (A_s, A_c, P, \lambda),$$

where A_s (A_c) is the largest nonempty (possibly empty) set of states with stopping (continuing) as the only feasible alternative. It is assumed that A_s and A_c are disjoint and that the set A_s and the matrix P satisfy

$$(4.2;2) \qquad \operatorname{spr}(P_{\overline{A_s}}) < 1.$$

The above formulation of an OSP implies that λ_i is not defined for $i \in A_c$. Occasionally λ_i will be defined for $i \in M$ in which case also the notation (4.2;1) will be used.

An OSP can be formulated as a problem satisfying the MRD-model of section 1.4 with $g(f) = 0$ for each policy $f \in F$. A set of states B satisfying $A_s \subseteq B \subseteq \overline{A_c}$ is called a *feasible stopping set*. The collection of feasible stopping sets is denoted by $\mathcal{B}$. Obviously there exists a one-to-one correspondence between feasible stopping sets and policies. The following method is a particular version of the policy iteration method of [HOWARD 1960] adapted to OSPs. This method computes an optimal stopping set.

METHOD 4.2.1. The following computational steps are executed:

(i) Fix an initial feasible stopping set B.

(ii) Solve the following set of equations in the components of the expected income vector $\eta(B)_i$, $i \in M$

$$(1) \quad \begin{cases} \eta(B)_i = \sum_{j \in M} P_{ij}\, \eta(B)_j & \text{for } i \in \bar{B}, \\ \eta(B)_i = \lambda_i & \text{for } i \in B. \end{cases}$$

(iii) Construct a feasible stopping set B' by the following rule. Define $i \in B'$ if and only if one of the following three conditions is satisfied:

(a) $i \in A_s$;

(b) $i \in \bar{B} \cap \overline{A_c}$ and $\lambda_i > \sum_{j \in M} P_{ij}\, \eta(B)_j$;

(c) $i \in B \cap \overline{A_s}$ and $\lambda_i \geq \sum_{j \in M} P_{ij}\, \eta(B)_j$.

(iv) If $B' = B$ then an optimal stopping set is obtained. Otherwise redefine $B \overset{d}{=} B'$ and return to step (ii).

REMARK 4.2.2. The above method is somewhat simplified if $\overline{A_c}$ is chosen as initial feasible stopping set. Then operation (iii) can be simplified by defining $i \in B'$ if and only if one of the following two conditions is satisfied:

(a) $i \in A_s$;

(b) $i \in B \cap \overline{A_s}$ and $\lambda_i \geq \sum_{j \in M} P_{ij}\, \eta(B)_j$.

This simplification is based on the fact that $B' \subset B$ at each nonterminal step in this case.

Since $P_{\overline{A_s}}$ is transient, $P_{\bar{B}}$ is transient for each feasible stopping set B. Hence $[I_{\bar{B}} - P_{\bar{B}}]^{-1}$ exists for $B \in \mathcal{B}$. In the sequel of this chapter a matrix $W(\bar{B}) \in \mathbb{R}^{|\bar{B}| \times |B|}$ will be used, defined by

$$(4.2;3) \qquad W(\bar{B}) \overset{d}{=} [I_{\bar{B}} - P_{\bar{B}}]^{-1}\, P_{\bar{B}B}.$$

Since $\sum_{j \in M} P_{ij} \leq 1$ for $i \in \overline{A_s}$ we have $W(\bar{B}) \geq 0$ and $W(\bar{B})\, 1_B \leq 1_{\bar{B}}$ (cf. lemma 2.3.6). The system of equations in $\eta(B)$ (cf. method 4.2.1 (ii)) is then equivalent to

$$(4.2;4) \qquad \begin{cases} \eta(B)_{\bar{B}} = W(\bar{B})\, \lambda_B, \\ \eta(B)_B = \lambda_B. \end{cases}$$

DEFINITION 4.2.3. An ordered pair of OSPs (A_s,A_c,P,λ) and (A_s',A_c',P',λ') is called *coupled* if $A_s = A_s'$, $A_c = A_c'$ and $P = P'$. The expected income vectors corresponding to a stopping set A which is feasible to both OSPs are denoted by $\eta(A)$ and $\nu(A)$ respectively.

LEMMA 4.2.4. *Let* B,C *satisfy* $C,B \in \mathcal{B}$ *and* $C \supseteq B$ *for the coupled pair of* OSPs (A_s,A_c,P,λ) *and* (A_s,A_c,P,μ). *Let* $\gamma,\delta \in \mathbb{R}^M$ *be given vectors with components satisfying respectively*

$$(1)\qquad \begin{cases} \sum_{j\in M} P_{ij}\,\gamma_j \geq \gamma_i & \text{for } i \in \bar{C} \\ \gamma_i \leq \lambda_i & \text{for } i \in B, \end{cases}$$

$$(2)\qquad \begin{cases} \sum_{j\in M} P_{ij}\,\delta_j \geq \delta_i & \text{for } i \in \bar{C} \\ \delta_i \leq \mu_i & \text{for } i \in B. \end{cases}$$

Then

(i) *if for* $i \in \bar{B} \cap C$ *either*

$$(3)\qquad \sum_{j\in M} P_{ij}\,\gamma_j > \gamma_i,$$

or

$$(4)\qquad \sum_{j\in M} P_{ij}\,\gamma_j = \gamma_i \text{ and } \sum_{j\in M} P_{ij}\,\delta_j \geq \delta_i$$

then we have

$$(5)\qquad [\eta(B),\nu(B)] \succeq [\gamma,\delta].$$

The assertion remains true in the following cases:

(ii) *The symbols* $\geq$ *and* $\succeq$ *are replaced in* (4) *and* (5) *by* $>$ *and* $\succ$ *respectively.*

(iii) *The symbols* $\succ$ *and* $>$ *are reversed in assertion* (i).

(iv) *The symbols* $\succ$ *and* $>$ *are reversed in assertion* (ii).

(v) *Equality signs hold throughout.*

PROOF. By (1), combined with either (3) or (4), we have

$$(6)\qquad P_{\bar{B}}\,\gamma_{\bar{B}} + P_{\bar{B}B}\,\lambda_B \geq \gamma_{\bar{B}}.$$

In the setting of lemma 1.2.3 (iii) we define $u^* = \eta(B)_{\bar{B}}$ and $u = \gamma_{\bar{B}}$. Then (6) implies

$$(7)\qquad \eta(B)_{\bar{B}} \geq \gamma_{\bar{B}}.$$

Since $\eta(B)_B = \gamma_B \geq \lambda_B$ by (1) we have

(8) $\qquad \eta(B) \geq \gamma.$

Consider the OSP $(A_s, A_c, P, \theta\lambda+\mu)$ where $\theta > 0$ is sufficiently large to guarantee for $i \in \bar{B} \cap C$

(9) $\qquad \sum_{j\in M} P_{ij}[\theta\gamma_j + \delta_j] \geq \theta\gamma_i + \delta_i.$

(9) implies by an argument similar to the one used above

(10) $\qquad \theta\eta(B) + \nu(B) \geq \theta\gamma + \delta.$

(8) and (10) imply (5). To prove assertion (ii) notice that (9) holds with strict inequality implying (10) and (5) with strict inequality. The other assertions follow by introducing the appropriate changes. □

LEMMA 4.2.5. *Let the coupled pair of* OSPs (A_s, A_c, P, λ) *and* (A_s, A_c, P, μ) *be given. Let the vectors* $\alpha, \beta, \gamma, \delta \in \mathbb{R}^M$ *satisfy*

(1) $\qquad [\alpha, \beta] \underline{\succ}(\succ) [\gamma, \delta].$

Moreover let for some state $i \in \overline{A_s} \cap \overline{A_c}$ *either*

(2) $\qquad \sum_{j\in M} P_{ij}\, \alpha_j < \lambda_i,$

or

(3) $\qquad \sum_{j\in M} P_{ij}\, \alpha_j = \lambda_i$ *and* $\sum_{j\in M} P_{ij}\, \beta_j \leq \mu_i.$

Then for such a state

(i) *either*

(4) $\qquad \sum_{j\in M} P_{ij}\, \gamma_j < \lambda_i,$

or

(5) $\qquad \sum_{j\in M} P_{ij}\, \gamma_j = \lambda_i$ *and* $\sum_{j\in M} P_{ij}\, \delta_j \leq \mu_i.$

The assertion remains true in the following cases:

(ii) *The symbol* $\leq$ *is replaced by* $<$ *in* (3) *and* (5).

(iii) *The symbols* $\succ$ *and* $>$ *are reversed in assertion* (i).

(iv) *The symbols* $\succ$ *and* $>$ *are reversed in assertion* (ii).

(v) *Equality signs hold throughout.*

PROOF. (1) implies $\alpha \geq \gamma$. Hence (1), (2) and (3) imply

$$\sum_{j\in M} P_{ij}\,\gamma_j \leq \sum_{j\in M} P_{ij}\,\alpha_j \leq \lambda_i.$$

If $\sum_{j\in M} P_{ij}\,\gamma_j = \lambda_i$ then $\sum_{j\in M} P_{ij}(\gamma_j-\alpha_j) = 0$ implying $\sum_{j\in M} P_{ij}(\delta_j-\beta_j) \leq 0$. Hence

$$\sum_{j\in M} P_{ij}\,\delta_j \leq \sum_{j\in M} P_{ij}\,\beta_j \leq \mu_i,$$

completing assertion (i). The assertions (ii)...(v) follow straight forwardly by introducing the appropriate changes. □

LEMMA 4.2.6. *Let* B,C *satisfy* $C,B \in \bar{\mathcal{B}}$ *and* $C \supseteq B$ *for a coupled pair of* OSPs (A_s,A_c,P,λ) *and* (A_s,A_c,P,μ). *Let* $\alpha,\beta \in \mathbb{R}^M$ *be given vectors with components satisfying respectively*

(1)
$$\begin{cases} \sum_{j\in M} P_{ij}\,\alpha_j \leq \alpha_i & \text{for } i \in \bar{B} \\ \alpha_i \overset{\geq}{=} \lambda_i & \text{for } i \in B \end{cases}$$

and

(2)
$$\begin{cases} \sum_{j\in M} P_{ij}\,\beta_j \leq \beta_i & \text{for } i \in \bar{B} \\ \beta_i \overset{\geq}{=} \mu_i & \text{for } i \in B. \end{cases}$$

Then

(i) *if for* $i \in \bar{B} \cap C$ *either*

(3) $\alpha_i > \lambda_i$,

or

(4) $\alpha_i = \lambda_i$ *and* $\beta_i \geq \mu_i$,

then we have

(5) $[\alpha,\beta] \succeq [\eta(C),\nu(C)]$.

The assertion remains true in the following cases:

(ii) *The symbols* $\geq$ *in* (4) *and* $\succeq$ *in* (5) *are replaced by* $>$ *and* $\succ$ *respectively.*

(iii) *The inequality signs including* $\succ$ *are reversed in assertion* (i).

(iv) *The inequality signs including* $\succ$ *are reversed in assertion* (ii).

(v) *Equality signs hold throughout.*

PROOF. By (1), combined with either (3) or (4), we have

(6) $\alpha_{\bar{C}} \geq P_{\bar{C}C}\ \alpha_C + P_{\bar{C}}\ \alpha_{\bar{C}} \geq P_{\bar{C}C}\ \lambda_C + P_{\bar{C}}\ \alpha_{\bar{C}}$.

In the setting of lemma 1.2.3 (iii) define $u^* = \eta(C)_{\bar{C}}$ and $u = \alpha_{\bar{C}}$. Then (6) implies

(7) $\alpha_{\bar{C}} \geq \eta(C)_{\bar{C}}$.

Since $\alpha_C \geq \lambda_C = \eta(C)_C$ by (1), combined with either (3) or (4), we have

(8) $\alpha \geq \eta(C)$.

Consider the OSP $(A_s,A_c,P,\theta\lambda+\mu)$ where $\theta > 0$ is sufficiently large to guarantee for $j \in \bar{B} \cap C$

(9) $\theta\alpha_j + \beta_j \geq \theta\lambda_j + \mu_j$.

Then (9) implies by the same argument

(10) $\theta\alpha + \beta \geq \theta\eta(C) + \nu(C)$.

(8) and (10) imply (5). To prove assertion (ii) notice that (9) holds with strict inequality implying (10) and (5) with strict inequality. The other assertions follow by introducing the appropriate changes. □

The following theorem presents a sufficient condition for a feasible stopping set to be optimal in a lexicographical sense to a coupled pair of OSPs. In the terminology of the cutting operation of the GMP-model such a set will be called an *optimal cutting set*.

THEOREM 4.2.7. *Let* $A \in \mathcal{B}$ *for the coupled pair of* OSPs (A_s,A_c,P,λ) *and* (A_s,A_c,P,μ) *satisfy for* $i \in \bar{A} \cap \overline{A_c}$ *either*

(1) $\sum_{j\in M} P_{ij}\ \eta(A)_j > \lambda_i$,

or

(2) $\sum_{j\in M} P_{ij}\ \eta(A)_j = \lambda_i$ and $\sum_{j\in M} P_{ij}\ \nu(A)_j \geq \mu_i$

and for $i \in A \cap \overline{A_s}$ *either*

(3) $\sum_{j\in M} P_{ij}\ \eta(A)_j < \lambda_i$,

or

(4) $\sum_{j\in M} P_{ij}\ \eta(A)_j = \lambda_i$ and $\sum_{j\in M} \nu(A)_j \le \mu_i$.

Then

(5) $[\eta(B),\nu(B)] \preceq [\eta(A),\nu(A)]$ for $B \in \mathcal{B}$.

PROOF. Let $B \in \mathcal{B}$ and $B' \stackrel{d}{=} A \cup B$. Then $B' \supseteq A$. Let in the setting of lemma 4.2.6 (i) $C \stackrel{d}{=} B'$, $B \stackrel{d}{=} A$, $\alpha \stackrel{d}{=} \eta(A)$ and $\beta \stackrel{d}{=} \nu(A)$. Then by this lemma

(6) $[\eta(A),\nu(A)] \succeq [\eta(B'),\nu(B')]$.

Let in the setting of lemma 4.2.5 (i) $\alpha \stackrel{d}{=} \eta(A)$, $\beta \stackrel{d}{=} \nu(A)$, $\gamma \stackrel{d}{=} \eta(B')$ and $\delta \stackrel{d}{=} \nu(B')$. Then (3), (4) and (6) imply by this lemma that $\eta(A)$ and $\nu(A)$ may be replaced by $\eta(B')$ and $\nu(B')$ in (3) and (4). Notice that $A \cap \bar{B} = B' \cap \bar{B}$. Hence for $i \in B' \cap \bar{B}$ we have either

$$\sum_{j\in M} P_{ij}\ \eta(B')_j < \lambda_i$$

or

$$\sum_{j\in M} P_{ij}\ \eta(B')_j = \lambda_i \text{ and } \sum_{j\in M} P_{ij}\ \nu(B')_j \le \mu_i.$$

In the setting of lemma 4.2.4 (iii) let $C \stackrel{d}{=} B'$, $B \stackrel{d}{=} B$, $\gamma \stackrel{d}{=} \eta(B')$ and $\delta \stackrel{d}{=} \nu(B')$. Then this lemma implies

(7) $[\eta(B),\nu(B)] \preceq [\eta(B'),\nu(B')]$

The proof is completed by combining (6) and (7). □

LEMMA 4.2.8. *Let* B_k, $k = 1,2$, *be feasible stopping sets to the coupled pair of* OSPs (A_s,A_c,P,λ) *and* (A_s,A_c,P,μ) *such that* $B_1 \neq B_2$. *Let* $B^* \stackrel{d}{=} B_1 \cup B_2$ *and* $B_* \stackrel{d}{=} B_1 \cap B_2$. *Then*

(1) $[\eta(B_k),\nu(B_k)]_{B^*} \succeq(\succ)(=) [\lambda,\mu]_{B^*}$ *for* $k = 1,2$

implies

(2) $[\eta(B_*),\nu(B_*)] \succeq(\succ)(=) [\eta(B_k),\nu(B_k)] \succeq(\succ)(=) [\eta(B^*),\nu(B^*)]$ *for* $k = 1,2$.

PROOF. Let in the setting of lemma 4.2.6 (i), $C \stackrel{d}{=} B^*$, $B \stackrel{d}{=} B_k$, $\alpha \stackrel{d}{=} \eta(B_k)$ and $\beta \stackrel{d}{=} \nu(B_k)$. Then (1) implies the right-hand part of (2) by this lemma. To

prove the lefthand part, let in the setting Lemma 4.2.4(i) $C \stackrel{d}{=} B_k$, $B \stackrel{d}{=} B_*$, $\gamma_i \stackrel{d}{=} \max_{k=1,2}[\eta(B_k)_i]$ for $i \in M$ and $\delta_i \stackrel{d}{=} \max_{k=1,2}[\nu(B_k)_i]$ for $i \in M$. By (1) we have for $i \in \bar{B}_k$, $k = 1,2$,

(3) $$\eta(B_k)_i = \sum_{j\in M} P_{ij}\, \eta(B_k)_j \le \sum_{j\in M} P_{ij}\, \gamma_j$$

and

(4) $$\nu(B_k)_i = \sum_{j\in M} P_{ij}\, \nu(B_k)_j \le \sum_{j\in M} P_{ij}\, \delta_j$$

(3) and (4) imply immediately

(5) $$\sum_{j\in M} P_{ij}\, \gamma_j \ge \gamma_i \quad \text{and} \quad \sum_{j\in M} P_{ij}\, \delta_j \ge \delta_i \quad \text{for } i \in \overline{B^*}.$$

For $i \in \bar{B}_k \cap B_{3-k}$, $k = 1,2$, we have by (1)

(6) $$\lambda_i \le \gamma_i = \eta(B_k)_i \le \sum_{j\in M} P_{ij}\, \gamma_j$$

and provided that $\nu(B_k)_i \ge \mu_i$

(7) $$\mu_i \le \delta_i = \nu(B_k)_i \le \sum_{j\in M} P_{ij}\, \delta_j$$

If $\nu(B_k)_i < \mu_i = \delta_i$ for some $i \in \bar{B}_k \cap B_{3-k}$, which may only be true if also $\eta(B_k)_i > \lambda_i$, then we have for sufficiently large θ

(8) $$\theta\eta(B_k)_i + \nu(B_k)_i \ge \theta\lambda_i + \mu_i = \theta\lambda_i + \delta_i.$$

From (4) and (8) we obtain

(9) $$\sum_{j\in M} P_{ij}\, \delta_j - \delta_i \ge \nu(B_k)_i - \delta_i \ge \theta(\lambda_i - \eta(B_k)_i)$$

Relation (6) implies

(10) $$\theta(\lambda_i - \eta(B_k)_i) \ge \theta(\lambda_i - \sum_{j\in M} P_{ij}\, \gamma_j) \ge \theta(\gamma_i - \sum_{j\in M} P_{ij}\gamma_j)$$

From (6), (7), (9) and (10) it follows that for $i \in \bar{B}_k \cap B_{3-k}$, $k = 1,2$

(11) $$\theta\gamma_i + \delta_i \le \sum_{j\in M} P_{ij}\, (\theta\gamma_j + \delta_j)$$

(5) and (11) show that the assumptions of Lemma 4.2.4(i) are satisfied. As a consequence we have for $k = 1,2$,

(12) $$[\eta(B_*),\nu(B_*)] \succeq [\gamma,\delta] \ge [\eta(B_k),\nu(B_k)]$$

The assertions with $\succeq$ replaced by $\succ$ and $=$ respectively follow by invoking the appropriate modifications. □

The next lemma characterizes the collection of feasible stopping sets to a coupled pair of OSPs having identical expected income vectors.

LEMMA 4.2.9. *Let $B \in \mathcal{B}$ be a feasible stopping set to the coupled pair of* OSPs (A_s,A_c,P,λ) *and* (A_s,A_c,P,μ). *Let $\mathcal{D}(B)$ be the collection of feasible stopping sets given by*

$$\mathcal{D}(B) \stackrel{d}{=} \{A \in \mathcal{B}: \eta(A) = \eta(B),\ \nu(A) = \nu(B)\}.$$

Moreover let

$$B^+ \stackrel{d}{=} \bigcup_{A \in \mathcal{D}(B)} A \quad \textit{and} \quad B_- \stackrel{d}{=} \bigcap_{A \in \mathcal{D}(B)} A .$$

Then

(i) $B^+ \in \mathcal{D}(B)$;

(ii) $B_- \in \mathcal{D}(B)$;

(iii) $C \in \mathcal{D}(B) \iff B_- \subseteq C \subseteq B^+$.

PROOF. If $\mathcal{D}(B)$ contains only B then necessarily $B = B^+ = B_-$ and (i), (ii) and (iii) trivially hold. Let $\mathcal{D}(B)$ contain at least two sets B_1 and B_2. Let $B^* \stackrel{d}{=} B_1 \cup B_2$ and $B_* \stackrel{d}{=} B_1 \cap B_2$. Then by lemma 4.2.8 in the equality version we have for k = 1,2

$$[\eta(B_*),\nu(B_*)] = [\eta(B_k),\nu(B_k)] = [\eta(B^*),\nu(B^*)].$$

This implies that if $B_1, B_2 \in \mathcal{D}(B)$ also $B^*, B_* \in \mathcal{D}(B)$. Since $\mathcal{D}(B)$ is a finite collection, repetition of this argument yields (i) and (ii). To prove (iii) notice that $A \in \mathcal{D}(B)$ implies $B_- \subseteq A \subseteq B^+$ by (i) and (ii). Conversely, notice that by assertions (i) and (ii)

$$[\eta(B^+),\nu(B^+)] = [\eta(B_-),\nu(B_-)] = [\eta(B),\nu(B)]. \tag{1}$$

Hence we have for $i \in B^+ \cap \bar{A}$

$$\lambda_i = \textstyle\sum_{j \in M} P_{ij}\ \eta(B^+)_i$$

and

$$\mu_i = \textstyle\sum_{j \in M} P_{ij}\ \nu(B^+)_j = \nu(B^+)_i.$$

Let $C \stackrel{d}{=} B^+$, $B \stackrel{d}{=} A$, $\gamma \stackrel{d}{=} \eta(B^+)$ and $\delta = \nu(B^+)$ in the setting of lemma 4.2.4 (v). Then this lemma implies here

$$[\eta(A),\nu(A)] = [\eta(B^+),\nu(B^+)]. \tag{2}$$

(1) and (2) complete the proof. □

4.3. SOME PROPERTIES OF GMP1

In this section we prove by means of theorem 4.3.3, preceded by the lemmas 4.3.1 and 4.3.2, that method 2.3.17 (GMP1) converges to a gain-optimal policy within a finite number of steps. Also a method is presented which computes the set A(z'), the ultimate goal of the cutting operation of GMP1. This method 4.3.5 is related to the coupled pair of OSPs

(4.3;1) $\quad (A_0, \overline{A(\hat{z})}, N_0, \hat{y})$

and

(4.3;2) $\quad (A_0, \overline{A(\hat{z})}, N_0, \hat{w})$.

It computes per iteration step the expected income vectors y'(B) and w'(B) for the current feasible stopping set B and constructs a new set B' with improved (in a lexicographical sense) expected income vectors. At each step the vectors y'(B) and w'(B) can both be obtained by inverting only one matrix. The method converges to the set A(z') within a finite number of steps as is proven by theorem 4.3.6 and 4.3.7.

An alternative to method 4.3.5 is to solve OSP (4.3;1) followed by a second OSP which maximizes with respect to $\hat{w}$ among the collection of optimal stopping sets of (4.3;1). Also this method is presented in this section (method 4.3.8).

LEMMA 4.3.1. *The sets* $A^*(\hat{z})$ *and* $A(z')$ *defined in* GMP1 *satisfy*

$$A^*(\hat{z}), A(z') \in M(\hat{z}).$$

PROOF. Let $B_1, B_2 \in M(\hat{z})$. B_1 and B_2 are feasible stopping sets to OSP (4.3;1) as well as to OSP (4.3;2). By lemma 4.2.8 in the $\underset{\sim}{\Upsilon}$-version

$$B_* \overset{d}{=} B_1 \cap B_2 \in M(\hat{z}).$$

Because $M(\hat{z})$ contains a finite number of sets this implies $A^*(\hat{z}) \in M(\hat{z})$. Applying lemma 4.2.4 (v) with $C \overset{d}{=} A(z')$, $B \overset{d}{=} A^*(\hat{z})$, $\gamma \overset{d}{=} y'(A(z'))$ and $\delta \overset{d}{=} w'(A(z'))$ we have

(1) $\quad [y'(A(z')), w'(A(z'))] = [y'(A^*(\hat{z})), w'(A^*(\hat{z}))]$

implying $A(z') \in M(\hat{z})$. □

LEMMA 4.3.2. *The method* GMP1 *satisfies at each iteration step*

$$[y'(A(z')),w'(A(z'))] \succeq [y(z),w(z)]$$

with the equality sign holding if and only if $z' = z$.

PROOF. By the policy improvement operation of GMP1 we have

$$(1) \qquad [y(z),w(z)]_{A(\hat{z})} \preceq [\hat{y},\hat{w}]_{A(\hat{z})}$$

where the equality sign holds if and only if $\hat{z} = z$. In the setting of lemma 4.2.6 (iii) let $C \overset{d}{=} A(\hat{z})$, $B \overset{d}{=} A(z)$, $\alpha \overset{d}{=} y(z)$ and $\beta \overset{d}{=} w(z)$. By this lemma, (1) implies

$$(2) \qquad [y(z),w(z)] \preceq [y'(A(\hat{z})),w'(A(\hat{z}))].$$

By the cutting operation of GMP1, we have since $A^*(\hat{z}) \in M(Z)$ by lemma 4.3.1

$$(3) \qquad [y'(A^*(\hat{z})),w'(A^*(\hat{z}))]_{A(\hat{z})} \succeq [y'(A(\hat{z})),w'(A(\hat{z}))]_{A(\hat{z})}.$$

Let, in the setting of lemma 4.2.4 (iii), $C \overset{d}{=} A(\hat{z})$, $B \overset{d}{=} A^*(\hat{z})$, $\gamma \overset{d}{=} y'(A(\hat{z}))$ and $\delta \overset{d}{=} w'(A(\hat{z}))$. Then by this lemma and (1) of lemma 4.3.1 we have

$$(4) \qquad [y'(A(z')),w'(A(z'))] \succeq [y'(A(\hat{z})),w'(A(\hat{z}))].$$

By Lemma 4.2.9 the equality sign holds in (4) and (2) if and only if $\hat{z} = z'$. (2) and (4) imply the assertion. □

THEOREM 4.3.3. *The method* GMP1 *converges within a finite number of steps to a gain-optimal policy.*

PROOF. In the setting of a GMP-scheme we have $y((z')z) = y'(Z(z'))$ and $w((z')z) = w'(A(z'))$. By lemma 4.3.2 $z' \in D^*(z)$ unless $z' = z$. Hence GMP1 is distinctive. By the policy improvement operation and the cutting operation GMP1 is also preserving. Theorem 3.2.7 completes the proof. □

DEFINITION 4.3.4. Let $(A_0,\overline{A(\hat{z})},N_0,\hat{y})$ and $(A_0,\overline{A(\hat{z})},N_0,\hat{w})$ be a coupled pair of OSPs with $\mathcal{B}$ denoting the collection of feasible stopping sets. Then a set $B_{ocs} \in \mathcal{B}$ satisfying for $B \in \mathcal{B}$

$$[y'(B_{ocs}),w'(B_{ocs})] \succeq [y'(B),w'(B)]$$

is called an *optimal cutting set*. The collection of optimal cutting sets is denoted by

$$\mathcal{D}_{ocs} \overset{d}{=} \{B \in \mathcal{B}: y'(B) = y'(B_{ocs}) \text{ and } w'(B) = w'(B_{ocs})\}.$$

Next a method is presented which computes a feasible stopping set B^{**}. In theorem 4.3.6 B^{**} is proved to be the optimal cutting set with the largest number of states.

METHOD 4.3.5. The following computational steps are executed:

(i) Fix $B = A(\hat{z})$ as initial feasible stopping set to the OSPs $(A_0, \overline{A(\hat{z})}, N_0, \hat{y})$ and $(A_0, \overline{A(\hat{z})}, N_0, \hat{w})$.

(ii) Compute the matrix $U_0(\bar{B})$ and the vectors $y'(B)$ and $w'(B)$ (cf. method 2.3.17, operation (iv)).

(iii) Compute the set B' such that $i \in \overline{B'}$ whenever

(1) $i \in \overline{A(\hat{z})}$,

(2) $i \in B \cap \overline{A_0}$ satisfies either

$$\sum_{j \in \Psi} (N_0)_{ij}\, y'(B)_j > \hat{y}_i,$$

or

$$\begin{cases} \sum_{j \in \Psi} (N_0)_{ij}\, y'(B)_j = \hat{y}_i \\ \sum_{j \in \Psi} (N_0)_{ij}\, w'(B)_j > \hat{w}_i \end{cases}$$

(iv) If $B' = B$ then define $B^{**} = B$ and stop. Otherwise redefine $B \overset{d}{=} B'$ and return to step (ii).

THEOREM 4.3.6. *Method* 4.3.5 *converges within a finite number of steps to a feasible stopping set* B^{**} *satisfying*

$$B^{**} = \bigcup_{B \in \mathcal{D}_{ocs}} B.$$

PROOF. Method 4.3.5 generates a sequence of feasible stopping sets satisfying

$$B_0 = A(\hat{z}) \supseteq B_1 \supseteq B_2 \supseteq \cdots .$$

The method stops at the smallest value of n such that $B_n = B_{n+1}$. Then $B^{**} = B_n$. At each nonterminal step we have for $i \in B_n \cap \overline{B_{n+1}}$ either

(1) $\quad \sum_{j\in\Psi} (N_0)_{ij}\, y'(B_n)_j > \hat{y}_i,$

or

(2) $\quad \sum_{j\in\Psi} (N_0)_{ij}\, y'(B_n)_j = \hat{y}_i$ and $\sum_{j\in\Psi} (N_0)_{ij}\, w'(B_n)_j > \hat{w}_i.$

In the setting of lemma 4.2.4 (ii) let $C \overset{d}{=} B_n$, $B \overset{d}{=} B_{n+1}$, $\gamma \overset{d}{=} y'(B_n)$ and $\delta \overset{d}{=} w'(B_n)$. This lemma implies then for each nonterminal n

(3) $\quad [y'(B_n),w'(B_n)] \prec [y'(B_{n+1}),w'(B_{n+1})].$

Since the collection $\mathcal{B}$ is finite, the method converges within a finite number of steps to some set $B^{**} \in \mathcal{B}$ satisfying for $i \in B^{**} \cap \overline{A_0}$ either

(4) $\quad \sum_{j\in\Psi} (N_0)_{ij}\, y'(B^{**})_j < \hat{y}_i,$

or

(5) $\quad \sum_{j\in\Psi} (N_0)_{ij}\, y'(B^{**})_j = \hat{y}_i$ and $\sum_{j\in\Psi} (N_0)_{ij}\, w'(B^{**})_j \leq \hat{w}_i.$

Notice that for $i \in \overline{B^{**}} \cap A(\hat{z})$ (1) and (2) are satisfied for some $n \in \mathbb{N}$. In the setting of lemma 4.2.5 (iii) with $\alpha \overset{d}{=} y'(B_n)$, $\beta \overset{d}{=} w'(B_n)$, $\gamma \overset{d}{=} y'(B^{**})$, $\delta \overset{d}{=} w'(B^{**})$, (1), (2) and (3) of this lemma are satisfied for $i \in \overline{B^{**}} \cap A(\hat{z})$ and some $n \in \mathbb{N}$. Hence (1) and (2) hold for $i \in \overline{B^{**}} \cap A(\hat{z})$ with B_n replaced by B^{**}. By this modification of (1) and (2) and the relations (4) and (5) we have by theorem 4.2.7

(6) $\quad [y'(B),w'(B)] \preceq [y'(B^{**}),w'(B^{**})] \quad$ for $B \in \mathcal{B}$,

implying $B^{**} \in \mathcal{D}_{ocs}$. For any set $B \in \mathcal{B}$ such that $B \cap \overline{B^{**}} \neq \emptyset$ the $\prec$-sign holds in (6). Hence $B \in \mathcal{D}_{ocs}$ implies $B \subseteq B^{**}$. By lemma 4.2.9 the assertion follows. □

The following theorem shows the relationship between the set B^{**} and the sets $A(z')$ and $A^*(\hat{z})$ used in method 2.3.17 (GMP1).

THEOREM 4.3.7. *Let* B^{**} *be the optimal cutting set specified by theorem* 4.3.6. *Let* B_{**} *be defined by*

$$B_{**} \overset{d}{=} B^{**} \setminus \{i \in \overline{A_0} \cap B^{**}: \hat{y}_i = \sum_{j\in\Psi} (N_0)_{ij}\, y'(B^{**})_j \text{ and } \hat{w}_i = \sum_{j\in\Psi} (N_0)_{ij}\, w'(B^{**})_j\}.$$

Then

(i) $B_{**} = A^*(\hat{z})$,

(ii) $B^{**} = A(z')$.

PROOF. By theorem 4.3.6 and the definitions of B^{**} and B_{**}

$$(1) \qquad [y'(B^{**}),w'(B^{**})]_{A(\hat{z})} = [y'(B_{**}),w'(B_{**})]_{A(\hat{z})} \succeq [\hat{y},\hat{w}]_{A(\hat{z})},$$

implying $B^{**},B_{**} \in M(\hat{z})$. Hence also $B^{**},B_{**} \supseteq A^*(\hat{z})$ (cf. method 2.3.17 (iv)). Suppose $A^*(\hat{z}) \subset B_{**}$. Since B^{**} satisfies (4) and (5) of theorem 4.3.6 for $i \in \overline{A^*(\hat{z})} \cap B_{**}$ and so does by lemma 4.2.8 B_{**} with strict inequality in (5) of theorem 4.3.6, we have for $i \in \overline{A^*(\hat{z})} \cap B_{**}$ either

$$(2) \qquad \sum_{j\in\Psi} (N_0)_{ij}\, y'(B_{**})_j < \hat{y}_i$$

or

$$(3) \qquad \sum_{j\in\Psi} (N_0)_{ij}\, y'(B_{**})_j = \hat{y}_i \text{ and } \sum_{j\in\Psi} (N_0)_{ij}\, w'(B_{**})_j < \hat{w}_i.$$

Now lemma 4.2.4 (iv) is applicable with $C \overset{d}{=} B_{**}$, $B \overset{d}{=} A^*(\hat{z})$, $\gamma \overset{d}{=} y'(B_{**})$ and $\delta \overset{d}{=} w'(B_{**})$ yielding

$$(4) \qquad [y'(A^*(\hat{z})),w'(A^*(\hat{z}))] \prec [y'(B_{**}),w'(B_{**})].$$

Let in the setting of lemma 4.2.5 (ii) $\alpha \overset{d}{=} y'(B_{**})$, $\beta \overset{d}{=} w'(B_{**})$, $\gamma \overset{d}{=} y'(A^*(\hat{z}))$, $\delta \overset{d}{=} w'(A^*(\hat{z}))$. Then by this lemma we may replace B_{**} by $A^*(\hat{z})$ in (2) and (3). Then (2) and (3) imply $A^*(\hat{z}) \notin M(\hat{z})$, a contradiction. Hence $A^*(\hat{z}) = B_{**}$ and moreover by lemma 4.2.9 $A(z') = B^{**}$. □

Alternatively the set $A(z')$ can be computed by solving a sequence of two separate OSPs. The first OSP maximes the expected income with respect to $\hat{y}$ and the second computes among its optimal stopping sets the set which maximizes the expected income with respect to $\hat{w}$ and contains the largest number of states.

METHOD 4.3.8. The set $A(z')$ can be computed by the following four steps.

Compute:

(i) An optimal stopping set B_{opt} to OSP $(A_0,\overline{A(\hat{z})},N_0,\hat{y})$ by method 4.2.1.

(ii) The sets $(B_{opt})^+$ and $B_{opt})^-$ defined respectively by

$$(B_{opt})^+ = B_{opt} \cup \{i \in A(\hat{z}) \cap \overline{B_{opt}} : \sum_{j\in\Psi} (N_0)_{ij}\, y'(B_{opt})_j = \hat{y}_i\}$$

and

$$(B_{opt})_- = B_{opt} \setminus \{i \in \overline{A_0} \cap B_{opt} : \sum_{j\in\Psi} (N_0)_{ij}\, y'(B_{opt})_j = \hat{y}_i\}.$$

(iii) An optimal stopping set A_{opt} to OSP $((B_{opt})_-, \overline{(B_{opt})^+}, N_0, \hat{w})$ by method 4.2.1.

(iv) The set A(z') by

$$A(z') = A_{opt} \cup \{i \in (B_{opt})^+ \cap \overline{A_{opt}} : \sum_{j \in \Psi} (N_0)_{ij} \, w'(B_{opt})_j = \hat{w}_i\}.$$

The question, which of the two methods is the most efficient, is not resolved here. However if all policies have exactly one subchain, both methods reduce to solving the OSP (4.3;2).

4.4. SUBOPTIMAL CUTTING METHODS

In this section cutting methods are introduced which do not necessarily lead to an optimal cutting set. Such methods will be called *suboptimal cutting methods*. A policy iteration method having the operations (i), (ii), (iii) and (v) in common with method 2.3.17 and operation (iv) of method 2.3.17 replaced by a suboptimal cutting method, is temporarily denoted by GMP2. Primarily it is shown that such a policy iteration method satisfies the requirements of theorem 3.2.7. Furthermore a particular suboptimal cutting method is presented which requires no matrix inversion. The method GMP2, referred to in chapter V, contains this particular suboptimal cutting method as cutting operation.

The next definition specifies the goal of a suboptimal cutting method.

DEFINITION 4.4.1. Consider the coupled pair of OSPs $(A_0, \overline{A(\hat{z})}, N_0, \hat{y})$ and $(A_0, \overline{A(\hat{z})}, N_0, \hat{w})$. A feasible stopping set B_{scs} satisfying either

(i) $B_{scs} \overset{d}{=} A(\hat{z})$ if and only if for $A_0 \subseteq B \subseteq A(\hat{z})$

$$[\eta(A(\hat{z})), \nu(A(\hat{z}))] \succeq [\eta(B), \nu(B)],$$

or

(ii) each state $i \in \overline{B_{scs}} \cap A(\hat{z})$ satisfies either

$$\sum_{j \in \Psi} (N_0)_{ij} \, y'(B_{scs})_j > \hat{y}_i$$

or

$$\sum_{j \in \Psi} (N_0)_{ij} \, y'(B_{scs})_j = \hat{y}_i \text{ and } \sum_{j \in \Psi} (N_0)_{ij} \, w'(A(\hat{z}))_j > \hat{w}_i$$

is called a *suboptimal cutting set*.

The policy obtained at the end of each iteration step of GMP2 will be denoted by z". Notice that for a given policy $z \in Z$ its successor z" is uniquely specified only for a particular suboptimal cutting method.

The following lemma and theorem describe the convergence properties of GMP2.

LEMMA 4.4.2. *The method* GMP2 *satisfies at each iteration step*

$$[y'(A(z")),w'(A(z"))] \succeq [y(z),w(z)]$$

with the equality sign holding if and only if $z" = z$.

PROOF. According to definition 4.4.1 we have either $A(z") = A(\hat{z})$ implying $z" = \hat{z}$ or $A(z") \subset A(\hat{z})$. In the latter case we have for $i \in \overline{A(z")} \cap A(\hat{z})$ by definition 4.4.1 either

$$\sum_{j\in\Psi} (N_0)_{ij}\, y'(A(z"))_j > \hat{y}_i,$$

or

$$\sum_{j\in\Psi} (N_0)_{ij}\, y'(A(z"))_j = \hat{y}_i \text{ and } \sum_{j\in\Psi} (N_0)_{ij}\, w'(A(z"))_j > \hat{w}_i.$$

Let in the setting of lemma 4.2.6 (ii) $\alpha \overset{d}{=} y'(A(z"))$, $\beta \overset{d}{=} w'(A(z"))$, $C \overset{d}{=} A(\hat{z})$, $B \overset{d}{=} A(z")$. By this lemma

(1) $$[y'(A(z")),w'(A(z"))] \succ [y'(A(\hat{z})),w'(A(\hat{z}))].$$

As in the proof of lemma 4.3.2 the policy improvement operation of GMP2 implies

(2) $$[y(z),w(z)] \preceq [y'(A(\hat{z})),w'(A(\hat{z}))]$$

with the equality sign holding if and only if $\hat{z} = z$. The assertion follows from (1) and (2). □

THEOREM 4.4.3. *The method* GMP2 *converges within a finite number of steps to a gain-optimal policy.*

PROOF. In the setting of a GMP-scheme we have $y((z')z) = y'(A(z"))$ and $w((z')z) = w'(A(z"))$. By lemma 4.4.2 $z" \in D^*(z)$ holds in GMP2 unless $z" = z$. Hence GMP2 is distinctive. By the policy improvement operation and the construction of a suboptimal cutting set, GMP2 is also preserving. Hence the requirements of theorem 3.2.7 are satisfied, completing the proof. □

A suboptimal cutting set can be obtained by truncating method 4.3.5 at some iteration step. However at least one matrix inversion is required. For the initial feasible stopping set $A(\hat{z})$ the computation of the vectors $y'(A(\hat{z}))$ and $w'(A(\hat{z}))$ requires the matrix $\left[I_{\overline{A(\hat{z})}} - (N_0)_{\overline{A(\hat{z})}}\right]^{-1}$. A suboptimal cutting set, the computation of which does not require a matrix inversion, is the subject of

LEMMA 4.4.4. *Consider the coupled pair of* OSPs $(A_0,\overline{A(\hat{z})},N_0,\hat{y})$ *and* $(A_0,\overline{A(\hat{z})},N_0,\hat{w})$. *Let the set* $B \in \mathcal{B}$ *satisfy for* $i \in A(\hat{z}) \cap \bar{B}$ *either*

$$(1) \qquad \sum_{j\in\Psi} (N_0)_{ij}\, \hat{y}_j > \hat{y}_i,$$

or

$$(2) \qquad \sum_{j\in\Psi} (N_0)_{ij}\, \hat{y}_j = \hat{y}_i \quad \text{and} \quad \sum_{j\in\Psi} (N_0)_{ij}\, \hat{w}_j > \hat{w}_i,$$

then B *is a suboptimal cutting set.*

PROOF. Let in the setting of lemma 4.2.5 (iii) $\alpha \overset{d}{=} \hat{y}$, $\beta \overset{d}{=} \hat{w}$, $\gamma \overset{d}{=} y'(A(\hat{z}))$, $\delta \overset{d}{=} w'(A(\hat{z}))$ then (1) and (2) imply for $i \in A(\hat{z}) \cap \bar{B}$ either

$$(3) \qquad \sum_{j\in\Psi} (N_0)_{ij}\, y'(A(\hat{z}))_j > \hat{y}_i,$$

or

$$(4) \qquad \sum_{j\in\Psi} (N_0)_{ij}\, y'(A(\hat{z}))_j = \hat{y}_i \text{ and } \sum_{j\in\Psi} (N_0)_{ij}\, w'(A(\hat{z}))_j > \hat{w}_i.$$

In the setting of lemma 4.2.4 (ii) let $B \overset{d}{=} B$, $C \overset{d}{=} A(\hat{z})$, $\gamma \overset{d}{=} y'(A(\hat{z}))$, $\delta \overset{d}{=} w'(A(\hat{z}))$. Then this lemma implies

$$[y'(B),w'(B)] \succ [y'(A(\hat{z})),w'(A(\hat{z}))],$$

excluding possibility (i) of definition 4.4.1. Let in the setting of lemma 4.2.5 (ii) $\alpha \overset{d}{=} y'(B)$, $\beta \overset{d}{=} w'(B)$, $\gamma \overset{d}{=} y'(A(\hat{z}))$, $\delta \overset{d}{=} w'(A(\hat{z}))$. Then this lemma implies that $y'(A(\hat{z}))$ and $w'(A(\hat{z}))$ may be replaced by $y'(B)$ and $w'(B)$ in (3) and (4). Hence B satisfies possibility (ii) of definition 4.4.1, completing the proof. □

The method, referred to as GMP2 in the sequel, has as cutting operation the computation of a particular suboptimal cutting set B. This set satisfies (1) and (2) of lemma 4.4.4 and moreover for $i \in \overline{A_0} \cap B$ either

$$(4.4;1) \qquad \sum_{j\in\Psi} (N_0)_{ij}\, \hat{y}_j < \hat{y}_i,$$

or

(4.4;2) $\quad \sum_{j\in\Psi} (N_0)_{ij}\, \hat{y}_j = \hat{y}_i$ and $\sum_{j\in\Psi} (N_0)_{ij}\, \hat{w}_j \leq \hat{w}_i$.

4.5. A THIRD SPECIAL VERSION OF A GMP-SCHEME

Finally in this section a third special case of a GMP-scheme is considered. It has the operations (i), (ii), (iii) and (v) in common with GMP1 and GMP2. Its cutting operation computes a set $A_0 \subseteq C \subseteq A(\hat{z})$ such that if $A(\hat{z}) \neq C$ for $i \in A(\hat{z}) \cap \bar{C}$ either

(4.5;1) $\quad \sum_{j\in\Psi} (N_0)_{ij}\, y(z)_j > \hat{y}_i$,

or

(4.5;2) $\quad \sum_{j\in\Psi} (N_0)_{ij}\, y(z)_j = \hat{y}_i$ and $\sum_{j\in\Psi} (N_0)_{ij}\, w(z_j) > \hat{w}_i$

and for $i \in \overline{A_0} \cap C$ either

(4.5;3) $\quad \sum_{j\in\Psi} (N_0)_{ij}\, y(z)_j < \hat{y}_i$,

or

(4.5;4) $\quad \sum_{j\in\Psi} (N_0)_{ij}\, y(z)_j = \hat{y}_i$ and $\sum_{j\in\Psi} (N_0)_{ij}\, w(z)_j \leq \hat{w}_i$.

The GMP-scheme specified in this way will be denoted by GMP3. A finite step convergence proof of GMP3 is covered by theorem 4.4.3. To see this let, in the setting of lemma 4.2.5 (iv), $\alpha \overset{d}{=} y(z)$, $\beta \overset{d}{=} w(z)$, $\gamma \overset{d}{=} y'(A(\hat{z}))$ and $\delta \overset{d}{=} w'(A(\hat{z}))$. By this lemma, (4.5;1) and (4.5;2) imply (3) and (4) of lemma 4.4.4. Replacing B by C in the remaining part of the proof of lemma 4.4.4 shows that also C is a suboptimal cutting set. If $A(\hat{z}) = C$ then only (4.5;3) and (4.5;4) hold. $A(\hat{z})$ is then a suboptimal cutting set by (2) of lemma 4.3.2 and definition 4.4.1(i). Consequently theorem 4.4.3 can be applied to establish finite step convergence in both cases.

Notice that GMP3 is obtained from GMP1 if in operation (iii) $X_1(i) \overset{d}{=} X(i)$ for all $i \in \Psi$ and the operations (iv) and (v) are omitted.

CHAPTER V

A NUMERICAL COMPARISON AMONG POLICY ITERATION METHODS

5.1. INTRODUCTION

In section 2.3 it has been shown that a problem satisfying the assumptions of the undiscounted MRD-model with interventions can be transformed to a problem satisfying the assumptions of the undiscounted GMP-model and vice versa. For this purpose a transformation θ has been introduced in definition 2.3.19.

For the numerical investigation in this chapter a related transformation T is needed to establish a similar relationship between the undiscounted GMP-model and the undiscounted MRD-model where interventions are excluded by condition (1.4;7). As a consequence, the class of problems which satisfy the assumptions of the undiscounted GMP-model can be partitioned into two subclasses. For one subclass the methods GMP1, 2 and 3 applied to one of its members and the method Jewell/Howard (method 1.4.1) applied to the transformed problem are identical. For the other subclass this is not true. The numerical comparison is performed on members of this latter subclass. For this purpose two sets of randomly generated problems are solved by the methods GMP1,2,3 and Jewell/Howard in section 5.3.

In section 5.4 a production control problem is formulated and solved numerically by each of the four methods.

Finally in section 5.5 some general conclusions are drawn on the numerical results.

5.2. SOME CONNECTIONS BETWEEN THE UNDISCOUNTED MRD- AND GMP-MODELS

In this section we display some connections between the undiscounted MRD-model of section 1.4 and the undiscounted GMP-model of section 2.3 and also between the policy iteration method of Jewell/Howard (method 1.4.1)

and the special cases GMP1, GMP2 and GMP3 of a GMP-scheme, discussed in chapter IV. Primarily a relationship between the two models is established. This is done by defining two transformations T and T'.

Let Π_1 denote the class of problems satisfying the assumptions of the undiscounted MRD-model. Then for each problem $\pi \in \Pi_1$ the data $(Q_0)^k_{ij}$, T^k_i, $(C_0)^k_i$ for $k \in K(i)$, $i,j \in M$ are given and condition (1.4;7) is satisfied. Let Π_2 denote the class of problems satisfying the assumptions of the undiscounted GMP-model. Then for each problem $\pi \in \Pi_2$ the data N_0, G_0 and τ are given and the conditions (2.3;1)-(2.3;5) and the assumptions 2.2.3-2.2.5 are satisfied. A subclass $\Pi_{21} \subset \Pi_2$ is specified in

DEFINITION 5.2.1. A problem $\pi \in \Pi_2$ satisfies $\pi \in \Pi_{21}$ if the following additional conditions are imposed on π:

(i) There exists a unique nonempty set $A \subset \Psi$ such that $x_0(i) \notin X(i)$ for $i \in A$ and $X(i) = \{x_0(i)\}$ for $i \in \bar{A}$.

(ii) For each pair (i,x) with $i \in A$, $x \in X(i)$ we have $G^x_i = 0$ and $P^x_{ij} = 1$ for some state $j \in \bar{A}$.

(iii) $(N_0)_{ij} = 0$ for $i,j \in \bar{A}$.

The two transformations T and T' are specified by the definitions 5.2.2 and 5.2.3.

DEFINITION 5.2.2. The transformation T transforms a problem $\pi \in \Pi_1$ into a problem $T(\pi)$ by defining

(i) The set of states of problem $T(\pi)$ by

$$\Psi \stackrel{d}{=} M \cup \{m=(i,k): i \in M,\ k \in K(i)\}.$$

(ii) A substochastic matrix N_0 with entries

$$(N_0)_{mj} \stackrel{d}{=} \begin{cases} (Q_0)^k_{ij} & \text{for } m = (i,k) \in \Psi\backslash M,\ j \in M, \\ 0 & \text{otherwise,} \end{cases}$$

a vector τ with components

$$\tau_m \stackrel{d}{=} \begin{cases} T^k_i & \text{for } m = (i,k) \in \Psi \setminus M, \\ 0 & \text{otherwise,} \end{cases}$$

and a reward vector G_0 with components

$$(G_0)_m \stackrel{d}{=} \begin{cases} (C_0)_i^k & \text{for } m = (i,k) \in \Psi \setminus M \\ 0 & \text{otherwise.} \end{cases}$$

(iii) Exactly one action x for each $k \in K(i)$, $i \in M$ such that each pair (i,x) induces a reward $G_i^x = 0$ and a family of probabilities $\{P_{ij}^x, j \in \Psi\}$ such that $P_{i,(i,k)}^x = 1$ and $P_{ij}^x = 0$ for $j \neq (i,k)$.

(iv) The set $X(i)$ for $i \in \Psi$ by

$$X(i) \stackrel{d}{=} \begin{cases} K(i) & \text{for } i \in M \\ \{x_0(i)\} & \text{for } i \in \Psi \setminus M. \end{cases}$$

DEFINITION 5.2.3. The transformation T' transforms a problem $\pi \in \Pi_{21}$ into a problem $T'(\pi)$ by defining

(i) Exactly one action k for each intervention $x \in X(i)$, $i \in A$ such that

$$(Q_0)_{ij}^k \stackrel{d}{=} \sum_{\ell \in \bar{A}} P_{i\ell}^x (N_0)_{\ell j},$$

$$T_i^k \stackrel{d}{=} \sum_{\ell \in \bar{A}} P_{i\ell}^x \tau_\ell$$

and

$$(C_0)_i^k \stackrel{d}{=} \sum_{\ell \in \bar{A}} P_{i\ell}^x (G_0)_\ell.$$

(ii) The set of actions $K(i) \stackrel{d}{=} X(i)$ for $i \in A$.

(iii) The set of states $M \stackrel{d}{=} A$.

It is easily verified that the transformed problem $T(\pi)$ for $\pi \in \Pi_1$ satisfies the conditions (2.3;1)-(2.3;5), the assumptions 2.2.3-2.2.5 and the additional conditions specified in definition 5.2.1 with A = M. Hence

(5.1;1) $\quad T(\pi) \in \Pi_{21} \quad$ for $\pi \in \Pi_1$.

Similarly problem $T'(\pi)$ satisfies

(5.1;2) $\quad T'(\pi) \in \Pi_1 \quad$ for $\pi \in \Pi_{21}$

by verifying the assumptions of the undiscounted MRD-Model (cf. section 1.4). Moreover, it is easily proved that

(5.1;3) $\quad T'(T(\pi)) = \pi \quad$ for $\pi \in \Pi_1$

and

$$(5.1;4) \qquad T(T'(\pi)) = \pi \quad \text{for } \pi \in \Pi_{21},$$

or, equivalently, T and T' are each others inverse and establish a bijective mapping between Π_1 and Π_{21}.

The main reason for considering subclass Π_{21} is lying in the fact that the computations to be executed by the method Jewell/Howard (in the sequel abbreviated by J/H) if applied to a problem $\pi \in \Pi_1$ are, except for a minor modification, identical to those to be executed by GMP1 if applied to problem $T(\pi) \in \Pi_{21}$. This fact is proven in

THEOREM 5.2.4. *Let* $\pi \in \Pi_1$. *If*

(i) *in the definition of the set* $K_2(i)$ *for* $i \in M$ *in operation* (iii) *of method* 1.4.1 (J/H), g_i *is replaced by* $\max_{k \in K(i)} \sum_{j \in M} (Q_0)^k_{ij}\, g_j$,

(ii) *the states* $i(\lambda)$ *in the operations* (ii) *of the method* J/H *and* GMP1 *are identically chosen for corresponding policies,*

(iii) *corresponding initial policies are chosen,*

then the computations to be executed by J/H *on* π *and* GMP1 *on* $T(\pi)$ *are identical.*

PROOF. For problem $T(\pi) \in \Pi_{21}$ we have $A = A_0 = A(z)$ for $z \in Z$ and $(N_0)_{\overline{\overline{A_0}}} = 0$ by definition 5.2.1. This implies $(k_0)_{\overline{\overline{A_0}}} = (G_0)_{\overline{\overline{A_0}}}$ and $(t_0)_{\overline{\overline{A_0}}} = \tau_{\overline{\overline{A_0}}}$ by definitions 2.3.1 and 2.3.2. Hence the computation of k_0 and t_0 in the preparatory part of GMP1 is superfluous. The fact that $A = A(z)$ for $z \in Z$ implies that also the cutting operation is superfluous. Furthermore for corresponding intervention x and action k

$$(1) \qquad k(i,x) = \sum_{j \in \bar{A}} p^x_{ij} (k_0)_j - (k_0)_i = \sum_{j \in \bar{A}} p^x_{ij} (G_0)_j = (C_0)^k_i,$$

$$(2) \qquad t(i,x) = \sum_{j \in \bar{A}} p^x_{ij} (t_0)_j - (t_0)_i = \sum_{j \in \bar{A}} p^x_{ij} \tau_j = T^k_i$$

and

$$(3) \qquad \sum_{\ell \in \bar{A}} p^x_{i\ell} (N_0)_{\ell j} = (Q_0)^k_{ij}.$$

Since $(N_0)_{\bar{\bar{A}}} = 0$,

(4) $$U_0(\bar{A}) = [I_{\bar{A}} - (N_0)_{\bar{A}}]^{-1} (N_0)_{\bar{A}A} = (N_0)_{\bar{A}A} .$$

By (1)-(4) it is easily verified that for corresponding policies z and f under the conditions (i), (ii) and (iii)

$$P_0(A(z)) = Q_0(f)$$

$$y(z)_{A(z)} = g(f) \quad ; \qquad w(z)_{A(z)} = u(f).$$

$$y(z)_{\overline{A(z)}} = \hat{y}_{\overline{A(z)}} \quad ; \qquad w(z)_{\overline{A(z)}} = \hat{w}_{\overline{A(z)}}.$$

$$X_2(i) = K_1(i) \quad ; \qquad X_3(i) = K_2(i),$$

and

$$z' = f_1,$$

completing the proof. □

5.3. NUMERICAL RESULTS FOR A CLASS OF RANDOMLY GENERATED PROBLEMS

From theorem 5.2.4 it follows that a numerical comparison between algorithms based on J/H and GMP1 makes no sense for problems in the subclass Π_{21}. This conclusion can be extended to GMP2 and GMP3 because these methods differ only in the cutting operation with GMP1 and this operation is superfluous for problems in Π_{21}. If one of the conditions of definition 5.2.1 is violated then theorem 5.2.4 is not valid. Hence it seems appropriate to investigate the subclass $\Pi_2 \backslash \Pi_{21}$. In this section a numerical comparison is performed on randomly generated problems, which are members of a subclass $\Pi_{22} \subset \Pi_2 \backslash \Pi_{21}$. This subclass Π_{22} is specified by

DEFINITION 5.3.1. A problem $\pi \in \Pi_2 \setminus \Pi_{21}$ is a member of subclass Π_{22} if it has the following properties

(i) All interventions imply deterministic transformations of the state of the system. For convenience a positive natural number is assigned to the symbol x, which is equal to the index of the state occupied after the intervention.

(ii) For each pair of interventions $x_1, x_2 \in X(i)$ there exists an intervention $x_2 \in X(x_1)$. The associated rewards $G_i^{x_1}$, $G_{x_1}^{x_2}$ and $G_i^{x_2}$ satisfy

$$G_i^{x_1} + G_{x_1}^{x_2} < G_i^{x_2}.$$

Next a transformation T" is introduced, which transforms a problem $\pi \in \Pi_{22}$ into a problem in Π_1.

DEFINITION 5.3.2. The transformation T" transforms a problem $\pi \in \Pi_{22}$ into a new problem T"(π) by defining

(i) Its set of states M identical to the set of states Ψ of problem π.

(ii) Exactly one action k for each intervention $x \in X(i)$, $i \in \Psi$ such that

$$(Q_0)_{ij}^k = P_{ix}^x (N_0)_{xj} = (N_0)_{xj},$$

$$(C_0)_i^k = P_{ix}^x (G_0)_x + G_i^x = (G_0)_x + G_i^x,$$

$$T_i^k = P_{ix}^x \tau_x = \tau_x.$$

(iii) Exactly one action k for each nulldecision such that

$$(Q_0)_{ij}^k = (N_0)_{ij},$$

$$(C_0)_i^k = (G_0)_i,$$

$$T_i^k = \tau_i.$$

REMARK 5.3.3. It is easily verified that problem T"(π) satisfies the assumptions of the undiscounted MRD-model. Hence T"(π) $\in \Pi_1$. The sets of policies of the two problems $\pi \in \Pi_{22}$ and T"(π) $\in \Pi_1$ are identical. However, the implications of two corresponding policies may not be identical. For example, a policy $z \in Z$ for problem π may have z(i) = intervention and z(z(i)) = intervention for some state $i \in \Psi$. Then for problem T"(π) the matrix $Q_0(f)$ for the policy f corresponding to z, may satisfy $Q_0(f) \neq S_0(z)$. Hence it is not immediately clear that a policy for problem T"(π) corresponding to an optimal policy for problem π is also optimal for T"(π). However, the conditions of definition 5.3.1 guarantee the existence of an optimal pair of corresponding policies. This fact is proven by theorem 5.3.8 preceded by the lemmas 5.3.4-5.3.7.

LEMMA 5.3.4. *Let* $\pi \in \Pi_{22}$ *and let* $z_1, z_2 \in Z$ *be two policies satisfying*

(i) $S_0(z_1)_{A(z_1)} \neq 0$,

(ii) $z_1(i) = z_2(i)$ *for* $i \in \Psi$ *except for one state* $\ell \in A(z_2)$,

(iii) $A(z_1) = A(z_2)$,

(iv) $z_2(\ell) = z_1(z_1(\ell))$ *with* $z_2(\ell) \in \overline{A(z_2)}$ *and* $z_1(\ell) \in A(z_1)$.

Then

$$z_2 \in D^*(z_1).$$

PROOF. By the assumptions on z_1 and z_2 it follows that

$$y(z_1)_\ell = y(z_1)_{z_1(\ell)} = y(z_1)_{z_1(z_1(\ell))} = y(z_1)_{z_2(\ell)} = y((z_2)z_1)_\ell.$$

If $z_2(i) = z_1(i)$ for $i \in A(z_1)$, $i \neq \ell$ then $y((z_2)z_1)_i = y(z_1)_i$ for $i \in A(z_1)$ and also for $i \in \overline{A(z_1)}$, since

$$y((z_2)z_1)_i = \sum_{j\in\Psi} U_0(\overline{A(z)})_{ij}\, y((z_2)z_1)_j =$$

$$= \sum_{j\in\Psi} U_0(\overline{A(z)})_{ij}\, y(z_1)_j = y(z_1)_i.$$

By property (ii) of definition 5.3.1 we have

$$G_\ell^{z_1(\ell)} + G_{z_1(\ell)}^{z_2(\ell)} < G_\ell^{z_2(\ell)},$$

implying

(1) $$k(\ell,z_1(\ell)) + k(z_1(\ell),z_2(\ell)) =$$

$$= G_\ell^{z_1(\ell)} + (k_0)_{z_1(\ell)} - (k_0)_\ell + G_{z_1(\ell)}^{z_2(\ell)} + (k_0)_{z_2(\ell)} - (k_0)_{z_1(\ell)} <$$

$$< G_\ell^{z_2(\ell)} + (k_0)_{z_2(\ell)} - (k_0)_\ell =$$

$$= k(\ell,z_2(\ell)).$$

Also we have

(2) $$t(\ell,z_1(\ell)) + t(z_1(\ell),z_2(\ell)) =$$

$$= (t_0)_{z_1(\ell)} - (t_0)_\ell + (t_0)_{z_2(\ell)} - (t_0)_{z_1(\ell)} =$$

$$= t(\ell,z_2(\ell)).$$

Hence by (1) and (2)

$$w(z_1)_\ell = k(\ell,z_1(\ell)) - y(z_1)_\ell\, t(\ell,z_1(\ell)) + w(z_1)_{z_1(\ell)} =$$

$$= k(\ell,z_1(\ell)) - y(z_1)_\ell\, t(\ell,z_1(\ell)) + k(z_1(\ell),z_2(\ell)) +$$

$$- y(z_1)_\ell\, t(z_1(\ell),z_2(\ell)) + w(z_1)_{z_2(\ell)}$$

$$< k(\ell,z_2(\ell)) - y(z_1)_\ell\, t(\ell,z_2(\ell)) + w(z_1)_{z_2(\ell)}$$

$$= w((z_2)z_1)_\ell.$$

For the remaining states $i \neq \ell$ we have $w(z_1)_i = w((z_2)z_1)_i$ for $i \in A(z_1)$ and for $i \in \overline{A(z_1)}$

$$w((z_2)z_1)_i = \sum_{j\in A(z_1)} U_0(\overline{A(z_1)})_{ij}\, w((z_2)z_1)_j \geq$$

$$\geq \sum_{j\in A(z_1)} U_0(\overline{A(z_1)})_{ij}\, w(z_1)_j = w(z_1)_i.$$

Hence $y((z_2)z_1) = y(z_1)$ and $w((z_2)z_1) > w(z_1)$ implying $z_2 \in D^*(z_1)$. □

LEMMA 5.3.5. *Let one of the methods* GMP1,2,3 *be applied to a problem* $\pi \in \Pi_{22}$, *yielding a gain-optimal policy* z^*. *Then* z^* *satisfies*

$$S_0(z^*)_{A(z^*)} = 0.$$

PROOF. Suppose the contrary. Then a policy z' can be constructed such that the assumptions (i)...(iv) of lemma 5.3.4 are satisfied with z_1 and z_2 replaced by z^* and z' respectively. Hence $z' \in D^*(z^*)$ contradicting the optimality of z^* (cf. theorem 3.2.7). □

LEMMA 5.3.6. *For a problem* $\pi \in \Pi_{22}$ *with transform* $T''(\pi) \in \Pi_1$, *we have for* $i \in \overline{A_1}$ *and intervention* $x \in X(i)$ *in* π *and corresponding action* $k \in K(i)$ *in* $T''(\pi)$

(1) $$k(i,x) = (C_0)^k_i + \sum_{j\in M} (Q_0)^k_{ij}(k_0)_j - (k_0)_i$$

and

(2) $$t(i,x) = T^k_i + \sum_{j\in M} (Q_0)^k_{ij}(t_0)_j - (t_0)_i.$$

PROOF. Applying transformation T" we obtain easily

$$k(i,x) = G_i^x + (k_0)_x - (k_0)_i =$$

$$= G_i^x + (G_0)_x + \sum_{j\in\Psi} (N_0)_{xj}(k_0)_j - (k_0)_i =$$

$$= (C_0)_i^k + \sum_{j\in M} (Q_0)_{ij}^k (k_0)_j - (k_0)_i$$

and

$$t(i,x) = (t_0)_x - (t_0)_i =$$

$$= \tau_x + \sum_{j\in\Psi} (N_0)_{xj}(t_0)_j - (t_0)_i =$$

$$= T_i^k + \sum_{j\in M} (Q_0)_{ij}^k (t_0)_j - (t_0)_i. \qquad \square$$

LEMMA 5.3.7. *Let* $z \in Z$ *be a policy for problem* $\pi \in \Pi_{22}$ *satisfying* $S_0(z)_{A(z)} = 0$. *Let* $(y(z),w(z))$ *be a particular solution of the system of equations of theorem* 2.3.16. *Then*

$$(g(f),u(f)) \stackrel{d}{=} (y(z),w(z) + k_0 - y(z) \square t_0)$$

is a solution of the system of equations (1.4;10) *for the policy* f *in problem* $T"(\pi)$ *corresponding to* z.

PROOF. For $y(z)_i$ and $w(z)_i$ we have for $i \in \Psi$ respectively

(1) $$y(z)_i = \sum_{j\in\Psi} S_0(z)_{ij}\, y(z)_j$$

and

(2) $$w(z)_i = k(i,z(i)) - y(z)_i\, t(i,z(i)) + \sum_{j\in\Psi} S_0(z)_{ij}\, w(z)_j.$$

Applying transformation T" to (1) and (2) and lemma 5.3.6 to (2) yields for $i \in M$

(3) $$y(z)_i = \sum_{j\in M} Q_0(f)_{ij}\, y(z)_j$$

and

(4) $$w(z)_i + (k_0)_i - y(z)_i (t_0)_i =$$

$$= C_0(f)_i - y(z)_i T(f)_i + \sum_{j\in M} Q_0(f)_{ij}[w(z)_j + (k_0)_j - y(z)_j (t_0)_j].$$

Identifying (3) and (4) with (1.4;10) completes the proof. □

For a problem $\pi \in \Pi_{22}$ the following theorem proves that the policy f^* for problem $T''(\pi)$, corresponding to a gain-optimal policy z^* for π, is gain-optimal for $T''(\pi)$.

THEOREM 5.3.8. *Let z^* be a gain-optimal policy for problem $\pi \in \Pi_{22}$ obtained by one of the methods* GMP1,2,3. *Let f^* be the policy in the setting of problem $T''(\pi) \in \Pi_1$, which corresponds to z^*. Then f^* is gain-optimal for problem $T''(\pi)$.*

PROOF. From the construction of the methods GMP1,2,3 it follows that for $i \in \Psi$, $x \in X(i)$ and an optimal policy z^* for problem π

$$(1) \qquad \begin{cases} \sum_{j\in\Psi} (S_0)^x_{ij}\, y(z^*)_j \le y(z^*)_i \\ k(i,x) - y(z^*)_i t(i,x) + \sum_{j\in\Psi} (S_0)^x_{ij}\, w(z^*)_j \le w(z^*)_i. \end{cases}$$

For $i \in \bar{A}_1$, $x \in X(i)\setminus\{x_0(i)\}$ it follows by the lemmas 5.3.6, 5.3.7 and (1) that

$$(2) \qquad \sum_{j\in\Psi} (S_0)^x_{ij}\, y(z^*)_j =$$

$$= y(z^*)_x \ge \sum_{j\in\Psi} (N_0)_{xj}\, y(z^*)_j =$$

$$= \sum_{j\in\Psi} (Q_0)^k_{ij}\, g(f^*)_j$$

and

$$(3) \qquad k(i,x) - y(z^*)_i\, t(i,x) + \sum_{j\in\Psi} (S_0)^x_{ij}\, w(z^*)_j =$$

$$= k(i,x) - y(z^*)_i\, t(i,x) + w(z^*)_x \ge k(i,x) - y(z^*)_i\, t(i,x) +$$

$$+ \sum_{j\in\Psi} (N_0)_{xj}\, w(z^*)_j =$$

$$= (C_0)^k_i - g(f^*)_i\, T^k_i + \sum_{j\in\Psi} (Q_0)^k_{ij}\, u(f^*)_j - (k_0)_i + y(z^*)(t_0)_i$$

for corresponding action k and policy f^* in problem $T''(\pi)$. Observe that (2) and (3) hold with equality for $i \in \bar{A}_0$ and $x = x_0(i)$. Substitution of (2) and (3) in (1) yields the optimality conditions for the undiscounted MRD-model (cf. theorem 1.4.2).

REMARK 5.3.9. The idea behind the transformation T" has been applied by [HOWARD 1960] to the automobile replacement problem. This problem also satisfies the conditions of definition 5.3.1. In practice these conditions are often satisfied. Another example is presented in section 5.4.

The methods GMP1,2,3 and J/H are compared for two sets of problems $S_1, S_2 \subset \Pi_{22}$. S_1 contains 10 problems with 10 states and S_2 contains 5 problems with 50 states. Each problem satisfies $A_0 = \{J\}$ and the triple (N_0, G_0, τ) is obtained as follows. Each row of the matrix $(N_0)_{\bar{A}_0\Psi}$ is obtained by generating J random numbers and dividing by their sum. The vectors G_0 and τ are obtained by generating $|\bar{A}_0|$ random numbers which are multiplied by a common factor. The numbers G_i^x are obtained by generating J random points in the unit square and taking G_i^x equal to the Eucledian distance between point i and point x after which these quantities are multiplied by a common factor.

To evaluate the effect of the initial policy on the results, two initial policies were used for problems from the set S_1. These are

$$z_1(i) = \begin{cases} x_0(i) & \text{for } i = 1 \\ 1 & \text{otherwise,} \end{cases} \tag{5.3;3}$$

and

$$z_1(i) = \begin{cases} 1 & \text{for } i = J \\ x_0(i) & \text{otherwise.} \end{cases} \tag{5.3;4}$$

In the tables 5.3.10...5.3.12 the following results are presented for each method:

(1) the number of iteration steps per problem (NIS),
(2) the total number of iteration steps for the whole set (TNIS),
(3) the total execution time in seconds (TET),
(4) the average execution time per step (AET).

The results are taken from [WEEDA 1974].

TABLE 5.3.10

Results for the set S_1 with initial policy (5.3;3)

Method	NIS										TNIS	TET	AET
	1	2	3	4	5	6	7	8	9	10			
GMP1	5	5	6	4	5	4	4	5	6	1	45	36	.80
GMP2	5	4	5	4	4	4	4	4	5	1	40	29	.71
GMP3	4	3	5	4	3	3	3	3	3	1	32	24	.75
J/H	3	2	3	3	3	2	2	2	2	1	23	20	.85

TABLE 5.3.11

Results for the set S_1 with initial policy (5.3;4)

Method	NIS										TNIS	TET	AET
	1	2	3	4	5	6	7	8	9	10			
GMP1	4	4	5	5	4	4	3	4	5	4	42	33	.80
GMP2	4	3	4	5	4	3	3	3	4	3	36	27	.75
GMP3	3	2	4	5	3	2	2	2	2	2	27	22	.83
J/H	3	2	3	4	3	2	2	2	2	2	25	25	.98

TABLE 5.3.12

Results for the set S_2 with initial policy (5.3;3)

Method	NIS					TNIS	TET	AET
	1	2	3	4	5			
GMP1	6	6	5	6	5	28	480	17.1
GMP2	5	5	4	5	4	23	315	13.5
GMP3	4	4	4	5	3	20	295	14.8
J/H	3	3	3	3	3	15	610	41.2

5.4. A PRODUCTION CONTROL PROBLEM

In this section the methods GMP1,2,3 and J/H are applied to a production control problem, which is in subclass Π_{22}. In order to apply method J/H to this problem, transformation T" (cf. definition 5.3.2) is used. A continuous time version of this problem has been solved numerically in [DE LEVE, TIJMS & WEEDA 1970].

A product can be produced at m+1 production rates, which are denoted by r = 0,1,...,m. r = 0 corresponds to the situation that the production is switched off. r > 0 corresponds to production at a rate of r units of product per unit of time. The demand per unit of time is Poisson distributed with parameter λ and supplied by the stock s, available at the beginning of

each unit time period. If the demand exceeds the available stock, the shortage is replenished by an emergency purchase. M denotes the maximum stock level. Stockholding costs are c_1 per unit of time and per unit of product in stock at the end of each unit time period. An emergency purchase costs c_2 per unit product. Production cost amounts to $c_3 r$ per unit of time for production rate r. Changing the production rate from r' to r" costs an amount b(r',r"). The criterion is to minimize the expected average cost per unit of time. Successively the basic notions of the undiscounted generalized Markov programming model are specified for this problem.

The set of states

$$\Psi \stackrel{d}{=} \{i=(r,s):\ r=0,1,\ldots,m,\quad s=0,1,\ldots,M\}.$$

The natural process and the set A_0

The set A_0 is defined by

$$A_0 \stackrel{d}{=} \{i=(r,M):\ r=1,\ldots,m\} \cup \{(0,0)\} \cup \{(1,0)\}.$$

The natural process describes the mutations in the stock level at a fixed production rate. Let $i=(r,s) \notin A_0$ be an arbitrary initial state of the natural process. Then production is continued at fixed rate r and the random stock $\underline{s}'$ after one unit time period is given by

$$\underline{s}' = \begin{cases} s+r-\underline{k} & \text{if } 0 < s+r-\underline{k} < M \\ M & \text{if } s+r-\underline{k} \geq M \\ 0 & \text{if } s+r-\underline{k} \leq 0 \end{cases},$$

where $\underline{k}$ denotes the random demand in that period with probability

$$a_k \stackrel{d}{=} \mathbb{P}\{\underline{k}=k\} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

We have then

$$\mathbb{P}\{\underline{s}'=M \mid (r,s)\} = \mathbb{P}\{\underline{k} \leq s+r-M\} = \sum_{k=0}^{s+r-M} a_k,$$

$$\mathbb{P}\{\underline{s}'=0 \mid (r,s)\} = \mathbb{P}\{s+r-\underline{k} \leq 0\} = \mathbb{P}\{\underline{k} \geq r+s\} = \sum_{k=s+r}^{\infty} a_k,$$

$$\mathbb{P}\{\underline{s}'=s' \mid (r,s)\} = \mathbb{P}\{\underline{k}=r+s-s'\} = a_{r+s-s'} \qquad 0 < s' < M.$$

For $i \in (r,s) \in \bar{A}_0$ the probabilities $(N_0)_{ij}$ are given by

$$(N_0)_{ij} = \begin{cases} \sum_{k=0}^{s+r-M} a_k & \text{if } j = (r,M) \\ \sum_{k=s+r}^{\infty} a_k & \text{if } j = (r,0) \\ a_{r+s-s'} & \text{if } j = (r,s'),\ 0 < s' < M \\ 0 & \text{otherwise.} \end{cases}$$

For $i = (r,s) \in \bar{A}_0$ we have $\tau_i = 1$ and

$$(G_0)_i = -c_1 M \sum_{k=0}^{s+r-M} a_k - c_1 \sum_{s'=1}^{M-1} a_{r+s-s'}\, s' + \\ + c_2 \sum_{k=s+r+1}^{\infty} a_k (s+r-k) - c_3 r.$$

For $i = (r,s) \in A_0$ we have by condition (2.3;1) $(N_0)_{ij} = 0$ for $j \in \Psi$, $\tau_i = 0$ and $(G_0)_i = 0$.

Interventions

The interventions in this problem change the production rate but maintain the stock level. Hence they imply deterministic transformations of the state of the system. In each state $i = (r,s) \in \Psi$ m interventions are feasible, which are denoted by $x = (r',s)$ with $r \neq r'$. The intervention cost G_i^x is given by

$$G_i^x = b(r,r') \quad \text{for } i = (r,s),\ x = (r',s)\ r \neq r'.$$

For $r = r'$ we take $b(r,r') \stackrel{d}{=} 0$ and for $r,r',r'' \in \{0,1,\ldots,m\}$ the number $b(r,r')$ are assumed to satisfy

$$b(r,r') < b(r,r'') + b(r'',r').$$

Observe that the conditions of definition 5.3.1 are satisfied.

Three different numerical examples of this problem are solved by each of the four considered methods.

Numerical example 1

$M = 20,\ m = 3,\ c_1 = .2,\ c_2 = 15,\ c_3 = 1,\ \lambda = 1.2$ and

$$b = \begin{bmatrix} 0 & 2 & 2 & 2 \\ 1 & 0 & 2 & 2 \\ 1 & 1 & 0 & 2 \\ 1 & 1 & 1 & 0 \end{bmatrix}.$$

Numerical example 2

$M = 20,\ m = 3,\ c_1 = .2,\ c_2 = 15,\ c_3 = 1,\ \lambda = 1.7$ and

$$b = \begin{bmatrix} 0 & 3 & 3 & 3 \\ 3 & 0 & 3 & 3 \\ 3 & 3 & 0 & 3 \\ 3 & 3 & 3 & 0 \end{bmatrix}.$$

Numerical example 3

$M = 25,\ m = 3,\ c_1 = .2,\ c_2 = 15,\ c_3 = 1,\ \lambda = 1.9$ and

$$b = \begin{bmatrix} 0 & 5 & 5 & 5 \\ 5 & 0 & 5 & 5 \\ 5 & 5 & 0 & 5 \\ 5 & 5 & 5 & 0 \end{bmatrix}.$$

The first two numerical examples use the same initial policy given by

$$z_1(r,s) = \begin{cases} (3,0) & \text{for } r = 0,1 \text{ and } s = 0 \\ (0,20) & \text{for } r = 1,2,3 \text{ and } s = 20 \\ x_0(r,s) & \text{otherwise.} \end{cases}$$

The third numerical example uses the initial policy given by

$$z_1(r,s) = \begin{cases} (3,0) & \text{for } r = 0,1 \\ (0,25) & \text{for } r = 1,2,3 \\ x_0(r,s) & \text{otherwise.} \end{cases}$$

The computational performance of the four methods on these three numerical examples is summarized in table 5.4.1. The number of iteration steps and the execution time are abbreviated by NIS(i) and ET(i) for problem i = 1,2,3 respectively.

TABLE 5.4.1

Results for the three numerical examples

Method	NIS(1)	NIS(2)	NIS(3)	ET(1)	ET(2)	ET(3)
GMP1	6	6	7	105(+4)*)	119(+4)	220(+6.4)
GMP2	6	4	4	98(+4)	73(+4)	117(+6.4)
GMP3	5	6	6	98(+4)	119(+4)	189(+6.4)
J/H	6	6	8	170	171	349

In addition, information about the convergence of y(z) and the optimal policies are given in the tables 5.4.2 ... 5.4.7.

TABLE 5.4.2

Convergence of $y(z_n)$ for example 1

z_n	$y(z_n)$			
n	GMP1	GMP2	GMP3	J/H
1	-3.674	-3.674	-3.674	-3.674
2	-2.836	-2.836	-2.710	-2.710
3	-2.484	-2.470	-2.489	-2.438
4	-2.346	-2.340	-2.351	-2.360
5	-2.339	-2.339	-2.339	-2.341
6	-2.339	-2.339		-2.339

TABLE 5.4.3

Optimal policy for example 1

r=0		r=1		r=2		r=3	
r=0	r'	s	r'	s	r'	s	r'
0,1	3	0,1	3	0,1,2,3	2	0,1,2	3
2	2	2,...,7	1	4,5,6	1	3,...,7	1
3,...,20	0	8,...,20	0	7,...,20	0	8,...,20	0

*) The numbers between parentheses are the times required for the computation of the vectors k_0 and t_0.

TABLE 5.4.4

Convergence of $y(z_n)$ for example 2

z_n	$y(z_n)$			
n	GMP1	GMP2	GMP3	J/H
1	-4.453	-4.453	-4.453	-4.453
2	-3.653	-3.560	-3.600	-3.600
3	-3.392	-3.267	-3.400	-3.380
4	-3.293	-3.249	-3.249	-3.291
5	-3.260		-3.249	-3.249
6	-3.249		-3.249	-3.249

TABLE 5.4.5

Optimal policy for example 2

r=0		r=1		r=2		r=3	
s	r'	s	r'	s	r'	s	r'
0,1	3	0,1	3	0	3	0,...,6	3
2,3	2	2	2	1,...,7	2	7,...,10	1
4,...,20	0	3,...,13	1	8,9,10	1	11,...,20	0
		14,...,20	0	11,...,20	0		

TABLE 5.4.6

Convergence of $y(z_n)$ for example 3

z_n	$y(z_n)$			
n	GMP1	GMP2	GMP3	J/H
1	-5.147	-5.147	-5.147	-5.147
2	-4.313	-4.196	-4.550	-4.550
3	-4.007	-3.742	-4.099	-4.094
4	-3.850	-3.733	-3.797	-3.770
5	-3.741		-3.733	-3.744
6	-3.733		-3.733	-3.733
7	-3.733			-3.733
8				-3.733

TABLE 5.4.7

Optimal policy for example 3

r=0		r=1		r=2		r=3	
s	r'	s	r'	s	r'	s	r'
0,1	3	0,1	3	0	3	0,...,4	3
2,3,4	2	2,3,4	2	1,...,10	2	5,6	2
5,...,25	0	5,...,18	1	11,12	1	7	3
		19,...,25	0	13,...,25	0	8,...,12	1
						13,...,25	0

5.5. SOME CONCLUSIONS

In this section some general conclusions are drawn from the numerical results of the sections 5.3 and 5.4. In table 5.5.1 the total execution time (TET) over all elements of the three sets of problems S_1, S_2 and PCP (the latter denoting the three numerical examples of the production control problem) is presented. The 4 methods are ordered in decreasing performance. Each problem of the set S_1 is counted twice, once for initial policy (5.2;2) and once for initial policy (5.2;3).

TABLE 5.5.1

The total execution time (TET) in sec.

S_1	TET	S_2	TET	PCP	TET
J/H	45	GMP3	295	GMP2	288
GMP3	46	GMP2	315	GMP3	406
GMP2	56	GMP1	480	GMP1	444
GMP1	69	J/H	610	J/H	690

From the results of table 5.5.1 we conclude that for problems of relatively small size the method J/H is preferred over the GMP-methods. This preference is reversed for problems of a larger size such as those in S_2 and PCP. In table 5.5.2 the average execution time per iteration step is exhibited.

TABLE 5.5.2

The average execution time (AET) per iteration step

S_1	AET	S_2	AET	PCP	AET
GMP3	.74	GMP2	14	GMP2	21
GMP3	.78	GMP3	15	GMP1	23
GMP1	.80	GMP1	17	GMP3	24
J/H	.94	J/H	41	J/H	35

It is observed in table 5.5.2 that the method J/H consumes the largest amount of time per iteration step. This is mainly due to the difference between the policy evaluation operations of a GMP-scheme and method J/H. The computation time for inverting a square matrix is a third degree polynomial

in its size. Since its coefficients are positive the inversion of 2 matrices of size $|\overline{A(z)}|$ and $\overline{A(z)}$ respectively will be less time consuming than the inversion of one matrix of size $|A(z)| + |\overline{A(z)}|$ as in method J/H.

A second contributor to the overall execution time is the number of iteration steps. In table 5.5.3 the total number of iteration steps is presented for the sets S_1, S_2 and PCP.

TABLE 5.5.3

The total number of iteration steps (TIS)

S_1	TIS	S_2	TIS	PCP	TIS
J/H	48	J/H	15	GMP2	14
GMP3	59	GMP3	20	GMP3	17
GMP2	76	GMP2	23	GMP1	19
GMP1	87	GMP1	28	J/H	20

Since the GMP-methods take about the same computation time per step, a secondary criterion to distinguish between them is the number of iteration steps. Although it is true that larger sample sizes are needed to draw more definite conclusions, it remains remarkable that the number of steps is non-decreasing in the order GMP3-2-1 uniformly for all randomly generated problems. For the set PCP this order is GMP2-3-1 except for the first example.

If the methods GMP1,2 and 3 are used with a common initial policy z and z', z" and z"' are respectively its successors then, by the results of chapter IV,

$$A(z') \subseteq A(z") \subseteq A(z"').$$

However, a "deeper" cutting operation implies neither a systematically larger nor a systematically smaller increase in the gain. This observation is confirmed by the results presented in the tables 5.5.4 and 5.5.5 for the three problems of the set PCP. Recall that in the experiments of section 5.4 the same initial policy z_1 is used per problem for each of the methods GMP1, 2 and 3. In table 5.5.4 the deepness of the cutting operation is defined by the ratio

$$\frac{|A(\hat{z}_1)| - |A(z_2)|}{|A(\hat{z}_1)|} .$$

The increase $y(z_2) - y(z_1)$ in the gain is presented in table 5.5.5.

TABLE 5.5.4

The deepness of the cutting method

Example	Method		
	GMP1	GMP2	GMP3
1	.43	.43	0
2	.61	.48	0
3	.44	.39	0

TABLE 5.5.5

The increase in the gain $y(z_2) - y(z_1)$

Example	Method		
	GMP1	GMP2	GMP3
1	.838	.838	.964
2	.800	.893	.853
3	.834	.951	.597

CHAPTER VI

GMP-MODELS WITH DISCOUNTING

6.1. INTRODUCTION

This chapter builds on the finite GMP-model of section 2.2. In section 6.2 some additional quantities are defined. In section 6.3 discounting is invoked. The value of a reward obtained at time t is multiplied by a factor $e^{-\rho t}$ where ρ is a fixed real positive number which is interpreted as an interest rate. A policy iteration method, computing a policy maximizing the expected discounted reward vector over the set of policies Z for a fixed interest rate ρ is presented.

In section 6.4 a parametric GMP-model (abbreviated as PGMP-model) is considered. The interest rate ρ is no longer fixed. Moreover, an intervention is assumed to take a small variable time $\varepsilon \geq 0$ and to induce a reward that is an infinitely differentiable function of the parameter ε. During the time ε the natural process is temporarily "frozen". For this model a partial Laurent expansion for the expected discounted reward vector for a fixed policy in the parameter ρ is obtained for sufficiently small fixed $\varepsilon > 0$. A numerical example is worked out in section 6.5.

In the PGMP-model assumption 2.2.6 of the finite GMP-model is dropped. The PGMP-model permits a unified treatment of problems which satisfy the assumptions of the finite GMP-model, including assumption 2.2.6 or not.

Moreover, the model can be applied to decision problems in which certain actions take a small but not exactly specified amount of time. It is of interest in that case to obtain a policy which is optimal for sufficiently small non-negative values of ε. For this purpose a new type of sensitive optimality is introduced with respect to the parameter ε in chapter VII.

6.2. SOME ADDITIONAL QUANTITIES IN THE FINITE GMP-MODEL

In this section some quantities are derived from the basic assumptions

of the finite GMP-model introduced in section 2.2. Also the corresponding Laplace-Stieltjes transforms are defined, because of their use in GMP-models with discounting. In the sequel a Laplace-Stieltjes transform is abbreviated by LST with plural form LSTs.

Frequently we meet in this chapter the convolution between a MRM $R(t)$ and a vector function $B(t) \in F_{J'}$ given by

$$(6.2;1) \qquad \int_{y\in[0,t]} R(dy)B(t-y).$$

For convolution the notation $*$ is used so that (6.2;1) becomes

$$(6.2;2) \qquad R(t) * B(t).$$

Observe that for $B(t)$ also a matrix function with column vectors in $F_{J'}$ can be substituted. Notice also that (6.2;1) can be viewed as the unique solution in $F_{J'}$ of the Markov renewal equation (cf. section 1.3) in $V(t)$ given by

$$(6.2;3) \qquad V(t) = B(t) + Q(t) * V(t).$$

Related to the SMM $N(t)$ and reward vector $G(t)$ of the natural process are the quantities specified by

DEFINITION 6.2.1. Let A be a given set such that $A_0 \subseteq A \subset \Psi$. Define

(i) the MRM $R(t;\bar{A})$ by

$$R(t;\bar{A}) \stackrel{d}{=} \sum_{m=0}^{\infty} N^{(m)}(t)_{\bar{A}},$$

where $N^{(m)}(t)_{\bar{A}}$ denotes the m^{th} convolution of $N(t)_{\bar{A}}$;

(ii) the matrix function $U(t;\bar{A})$ by

$$U(t;\bar{A}) \stackrel{d}{=} R(t;\bar{A}) * N(t)_{\bar{A}A};$$

(iii) the vector function $K(t;\bar{A})$ by

$$K(t;\bar{A})_{\bar{A}} \stackrel{d}{=} R(t;\bar{A}) * G(t)_{\bar{A}}$$

and

$$K(t;\bar{A})_{A} \stackrel{d}{=} 0.$$

An entry $U(t;\bar{A})_{ij}$ of $U(t;\bar{A})$ is interpreted as the probability that j is the first state taken on in the set A on or before time t by the natural process with initial state $i \in \bar{A}$.

A component $K(t;\bar{A})_i$ of $K(t;\bar{A})$ represents the expected reward earned during the natural process with initial state $i \in \bar{A}$. The natural process stops at time t or at the epoch the set A is entered for the first time depending on which of the two events occurs first.

The LSTs of $R(t;\bar{A})$, $U(t;\bar{A})$ and $K(t;\bar{A})$ are denoted by $r(\rho;\bar{A})$, $u(\rho;\bar{A})$ and $k(\rho;\bar{A})$ respectively and exist at least for $\rho > 0$. However, $r(\rho;\bar{A})$ also exists for $\rho = 0$ since $(N_0)_{\bar{A}}$ is transient using lemma 1.2.1. The same is true for $u(\rho;\bar{A})$ and $k(\rho;\bar{A})$. By their definitions and the application of the multiplication rule for LSTs (cf. [FELLER 1960], p.411) we have

$$(6.2;4) \qquad u(\rho;\bar{A}) = r(\rho;\bar{A})n(\rho)_{\bar{A}A} \qquad \text{for } \rho \geq 0$$

and

$$(6.2;5) \qquad k(\rho;\bar{A}) = r(\rho;\bar{A})g(\rho)_{\bar{A}} \qquad \text{for } \rho \geq 0.$$

Next a reward associated with action $x \in X(i)$ in state $i \in \Psi$ is introduced, which is related to k(i,x) of definition 2.3.3.

DEFINITION 6.2.2. For each action $x \in X(i)$ and state $i \in \Psi$ a function $K(t)_i^x$ is defined by

$$K(t)_i^x = \begin{cases} 0 & \text{for } x = x_0(i) \\ G_i^x + \sum_{j\in\Psi} P_{ij}^x K(t;\bar{A}_0)_j - K(t;\bar{A}_0)_i & \text{for } x \neq x_0(i). \end{cases}$$

Notice that according to this definition the relation with k(i,x) is $k(i,x) = \lim_{t\to\infty} K(t)_i^x$. The term $K(t;\bar{A}_0)_i$ is interpreted under definition 6.2.1, taking $A = A_0$. The expression $G_i^x + \sum_{j\in\Psi} P_{ij}^x K(t;\bar{A}_0)_j$ represents the expected reward during a stochastic walk with initial state $i \in \bar{A}_1$ at $t = 0$. In the initial state i intervention $x \in X(i)$ is taken, which transfers the system to a state j with probability P_{ij}^x and induces a reward G_i^x. From state j on the evolution of the system is described by the natural process. It stops either at time t or at the epoch the set A_0 is entered for the first time, depending on which of the two events occurs first. Thus $K(t)_i^x$ represents the difference in expected reward of two stochastic walks each with state i as initial state. Notice that this interpretation is also valid for $x = x_0(i)$.

The LST of $K(t)_i^x$ is denoted by $k(\rho)_i^x$ and satisfies for $\rho \geq 0$

$$(6.2;6) \qquad k(\rho)_i^x = \begin{cases} 0 & \text{for } x = x_0(i) \\ G_i^x + \sum_{j \in \Psi} P_{ij}^x \, k(\rho;\bar{A}_0)_j - k(\rho;\bar{A}_0)_i & \text{for } x \neq x_0(i). \end{cases}$$

The next quantities are related to the decision process for a fixed policy $z \in Z$.

DEFINITION 6.2.3. For each fixed policy $z \in Z$ are defined

(i) the matrix function

$$P(t;A(z)) \stackrel{d}{=} P(z)_{A(z)} + P(z)_{A(z)\overline{A(z)}} U(t;\overline{A(z)}),$$

(ii) the matrix function

$$\Gamma(t;z) \stackrel{d}{=} \begin{bmatrix} U(t;\overline{A(z)}) \\ \hline P(t;A(z)) \end{bmatrix},$$

(iii) the vector function

$$G(t;z)_i \stackrel{d}{=} \begin{cases} G(t)_i & \text{for } i \in \overline{A(z)} \\ G_i^{z(i)} & \text{for } i \in A(z), \end{cases}$$

and

(iv) the vector function $K(t;z)$ with components $K(t)_i^{z(i)}$, $i \in \Psi$.

The LSTs of $P(t;A(z))$, $\Gamma(t;z)$ and $K(t;z)$ are denoted by $p(\rho;A(z))$, $\gamma(\rho;z)$ and $k(\rho;z)$ respectively. We have for $\rho \geq 0$

$$(6.2;7) \qquad p(\rho;A(z)) = P(z)_{A(z)} + P(z)_{A(z)\overline{A(z)}} \, u(\rho;\overline{A(z)}),$$

$$(6.2;8) \qquad \gamma(\rho;z) = \begin{bmatrix} u(\rho;\overline{A(z)}) \\ \hline p(\rho;A(z)) \end{bmatrix}$$

and

$$k(\rho;z)_i = \begin{cases} 0 & \text{if } z(i) = x_0(i) \\ G_i^{z(i)} + \sum_{j \in \Psi} P_{ij}^{z(i)} \, k(\rho;\overline{A_0})_j - k(\rho;\overline{A_0})_i & \text{otherwise.} \end{cases}$$

In the finite GMP-model the following integral equation arises for

a fixed policy $z \in Z$

$$(6.2;9) \qquad V(t;z) = K(t;z) + \Gamma(t;z) * V(t;z)_{A(z)}.$$

Since $P(t;A(z))$ is a normal SMM by assumption 2.2.6 and $K(t;z) \in F_J$, since F_J is closed under addition and scalar multiplication, the part of equation (6.2;9) in $V(t;z)_{A(z)}$

$$(6.2;10) \qquad V(t;z)_{A(z)} = K(t;z)_{A(z)} + P(t;A(z)) * V(t;z)_{A(z)}$$

is a Markov renewal equation and has a unique solution in $F_{|A(z)|}$. $V(t;z)_{\overline{A(z)}}$ is uniquely determined by

$$(6.2;11) \qquad V(t;z)_{\overline{A(z)}} = U(t;\overline{A(z)}) * V(t;z)_{A(z)}.$$

Hence (6.2;9) has a unique solution in F_J, for each $z \in Z$. In the sequel $V(t;z)$ denotes this solution.

By Laplace-Stieltjes transformation of (6.2;9) it follows that the LST of $V(t;z)$, denoted by $v(\rho;z)$, uniquely satisfies for $z \in Z$ and $\rho > 0$

$$(6.2;12) \qquad v(\rho;z) = k(\rho;z) + \gamma(\rho;z)v(\rho;z)_{A(z)}.$$

With the exception of a policy independent term specified below, $V(t;z)_i$ ($v(\rho;z)_i$) is interpreted as the expected reward in $[0,t]$ (the expected discounted reward in $[0,\infty)$) earned during the decision process of policy $z \in Z$ with initial state i.

To specify the policy independent term the connection with the Markov renewal equation (1.4;2) in the MRD-model is exhibited. In the notation of the finite GMP-model (1.4;2) becomes

$$(6.2;13) \qquad \hat{V}(t;z) = G(t;z) + S(t;z) * \hat{V}(t;z).$$

Since $\hat{V}(t;z)_i$ represents the true expected reward in $[0,t]$ for initial state i and fixed policy $z \in Z$, the following lemma shows that this policy independent term is given by $K(t;\overline{A_0})$.

LEMMA 6.2.4. *The unique solutions* $V(t;z)$ *and* $\hat{V}(t;z)$ *respectively of the Markov renewal equations* (6.2;9) *and* (6.2;13) *satisfy for a fixed policy*

$z \in Z$

$$V(t;z) = \hat{V}(t;z) - K(t;\overline{A}_0).$$

PROOF. By elimination of $G(t;z)_{A(z)}$ from

$$\hat{V}(t;z)_{A(z)} = G(t;z)_{A(z)} + P(z)_{A(z)\Psi}\, \hat{V}(t;z)$$

and

$$K(t;z)_{A(z)} = G(t;z)_{A(z)} + P(z)_{A(z)\Psi}\, K(t;\overline{A}_0) - K(t;\overline{A}_0)_{A(z)}$$

and by elimination of $G(t;z)_{\overline{A(z)}}$ from

$$\hat{V}(t;z)_{\overline{A(z)}} = G(t;z)_{\overline{A(z)}} + N(t)_{\overline{A(z)}\Psi} * \hat{V}(t;z)$$

and

$$K(t;\overline{A}_0)_{\overline{A(z)}} = G(t;z)_{\overline{A(z)}} + N(t)_{\overline{A(z)}\Psi} * K(t;\overline{A}_0)$$

it follows that (6.2;13) is equivalent to

(1) $$\hat{V}(t;z) - K(t;\overline{A}_0) = K(t;z) + S(t;z) * [\hat{V}(t;z) - K(t;\overline{A}_0)].$$

Since

$$V(t;z)_{\overline{A(z)}} = U(t;\overline{A(z)}) * V(t;z)_{A(z)}$$

is equivalent to

$$V(t;z)_{\overline{A(z)}} = N(t)_{\overline{A(z)}\Psi} * V(t;z),$$

equation (6.2;9) is equivalent to

(2) $$V(t;z) = K(t;z) + S(t;z) * V(t;z).$$

The assertion is a consequence of (1) and (2). □

6.3. POLICY ITERATION IN THE DISCOUNTED GMP-MODEL

In this section a reward earned at time t is discounted by a factor $e^{-\rho t}$, where $\rho > 0$ is the interest rate. Invoking discounting in the finite GMP-model is equivalent to applying Laplace-Stieltjes transformation to the

normal SMM N(t) and the reward vector G(t) of the natural process. A policy iteration method is presented which computes a policy maximizing the expected discounted reward vector $v(\rho;z)$ for a fixed interest rate $\rho > 0$ over the set of policies Z. This method is a special case of the method GMP1. Notice that the elements $n(\rho)_{ij}$ of $n(\rho)$ can be interpreted as probabilities and moreover that

(6.3;1) $\quad \|n(\rho)\| < 1 \quad$ for $\rho > 0$.

The system of equations (1) of theorem 2.3.16 remains valid in a simplified version. The matrix $\Gamma_0(z)$ is replaced by $\gamma(\rho;z)$ and the vector $k(z)$ by $k(\rho;z)$. Since $P(t;A(z))$ is normal, (6.3;1) implies that $p(\rho;A(z))$ is a transient matrix. By lemma 1.2.4 $y(z)_{A(z)} = 0$ is the only solution of the equation $y(z)_{A(z)} = p(\rho;A(z))\ y(z)_{A(z)}$. This implies that the system of equations (1) of theorem 2.3.16 can be reduced to the single equation

$$w = k(\rho;z) + \gamma(\rho;z)w_{A(z)},$$

which is exactly (6.2;12) if we identify w with $v(\rho;z)$. Moreover, all operations concerning $y(z)$, $\hat{y}$ and y' in GMP1 can be omitted. In summary method 2.3.17 then becomes

METHOD 6.3.1. The method consists of the following main operations:

(i) Preparatory part. Compute the vector $k(\rho;\overline{A}_0)$ from (6.2;5) and the number $k(\rho)_i^x$ for each $i \in \Psi$, $x \in X(i)$ from (6.2;6). Fix an initial policy $z \in Z$.

(ii) Policy evaluation operation. Compute the unique solution of (6.2.;12) in $v(\rho;z)$.

(iii) Policy improvement operation. Introduce the notation

$$s(\rho)_{ij} \stackrel{d}{=} \begin{cases} n(\rho)_{ij} & \text{for } x = x_0(i) \\ p_{ij}^x & \text{for } x \neq x_0(i) \end{cases}$$

and

$$X_2(i) \stackrel{d}{=} \begin{cases} X(i) & \text{for } i \in \overline{A(z)} \cup A_0 \\ X(i)\setminus\{x_0(i)\} & \text{for } i \in A(z) \setminus A_0. \end{cases}$$

Compute:

(1) the vector $\hat{v} \in \mathbb{R}^{J'}$ with components $\hat{v}_i$ given by

$$\hat{v}_i \overset{d}{=} \max_{x \in X_2(i)} [k(\rho)_i^x + \sum_{j \in \Psi} s(\rho)_{ij}^x \, v(\rho;z)_j],$$

(2) the set of actions

$$X_3(i) \overset{d}{=} \{x \in X_2(i) : k(\rho)_i^x + \sum_{j \in \Psi} s(\rho)_{ij}^x \, v(\rho;z)_j = \hat{v}_i\},$$

(3) policy $\hat{z} \in Z$ such that $\hat{z}(i) = z(i)$ whenever $z(i) \in X_3(i)$. If $z(i) \notin X_3(i)$ then put $\hat{z}(i)$ equal to an arbitrary element of $X_3(i)$.

(iv) Cutting operation. Compute (cf. section 4.3)

(1) an optimal stopping set B_{opt} to OSP $(A_0, \overline{A(\hat{z})}, n(\rho), \hat{v})$,

(2) the set $(B_{opt})^+$,

(3) policy z' such that

$$z'(i) = \begin{cases} \hat{z}(i) & \text{for } i \in (B_{opt})^+ \\ x_0(i) & \text{otherwise.} \end{cases}$$

(v) If $z'(i) = z(i)$ for $i \in \Psi$ then stop. Otherwise redefine z equal to z' and repeat operations (ii) ... (v).

6.4. A PARTIAL LAURENT EXPANSION FOR THE EXPECTED DISCOUNTED REWARD VECTOR IN A PARAMETRIC GMP-MODEL

In this section the GMP-model of section 2.2 is parametrized by the assumption that each intervention takes a time $\varepsilon \geq 0$, where ε is considered as a variable. The partial Laurent expansion in ρ for the expected discounted reward vector is derived for sufficiently small fixed $\varepsilon > 0$. Primarily the asymptotic expansions for $\rho \downarrow 0$ of $n(\rho)$, $g(\rho)$, $r(\rho;\bar{A})$, $u(\rho;\bar{A})$ and $k(\rho;\bar{A})$ are presented. After that the parametric GMP-model is defined. Finally, after some preparatory results, the partial Laurent expansion is developed in theorem 6.4.16.

The lemmas 6.4.1-6.4.7 state the asymptotic expansions for $\rho \downarrow 0$ of various quantities of the finite GMP-model in their normalized moments.

LEMMA 6.4.1. *If* N_k *is finite for some* $k \in \mathbb{N}$ *then* $n(\rho)$ *has the asymptotic expansion*

$$n(\rho) = \sum_{m=0}^{k} (-\rho)^m N_m + o(\rho^k), \quad \rho \downarrow 0.$$

PROOF. The assertion is an immediate consequence of lemma 1.3.3. □

LEMMA 6.4.2. *If* G_k *is finite for some* $k \in \mathbb{N}$ *then* $g(\rho)$ *has the asymptotic expansion*

$$g(\rho) = \sum_{m=0}^{k} (-\rho)^m G_m + o(\rho^k), \quad \rho \downarrow 0.$$

PROOF. The assertion is an immediate consequence of lemma 1.3.7. □

LEMMA 6.4.3. *If* $(N_{k+1})_{\bar{A}}$ *is finite for some* $k \in \mathbb{N}$ *and* $A_0 \subseteq A \subset \Psi$ *then* $r(\rho;\bar{A})$ *has the asymptotic expansion*

$$r(\rho;\bar{A}) = \sum_{m=0}^{k} \rho^m R_m(\bar{A}) + o(\rho^k), \quad \rho \downarrow 0,$$

where $R_{-1}(\bar{A}) \stackrel{d}{=} 0$, $R_0(\bar{A}) = [I_{\bar{A}} - (N_0)_{\bar{A}}]^{-1}$ *and*

$$R_m(\bar{A}) = R_0(\bar{A})[\textstyle\sum_{\ell=1}^{m+1}(-1)^{\ell}(N_{\ell})_{\bar{A}} R_{m-\ell}(\bar{A})] \quad \text{for } m = 1,\ldots,k.$$

PROOF. The assertion is a consequence of lemma 1.3.5 and the fact that $(N_0)_{\bar{A}}$ is transient.

LEMMA 6.4.4. *If* $(N_{k+1})_{\bar{A}}$ *and* $(N_k)_{\bar{A}A}$ *are finite for some* $k \in \mathbb{N}$ *and* $A_0 \subseteq A \subset \Psi$ *then* $u(\rho;\bar{A})$ *has the asymptotic expansion*

$$(1) \qquad u(\rho;\bar{A}) = \sum_{m=0}^{k} \rho^m U_m(\bar{A}) + o(\rho^k), \quad \rho \downarrow 0,$$

where the coefficients $U_m(\bar{A}) \in \mathbb{R}^{|\bar{A}|\times|A|}$ *are uniquely determined by*

$$(2) \qquad U_m(\bar{A}) = \sum_{\ell=0}^{m} (-1)^{\ell} R_{m-\ell}(\bar{A})(N_{\ell})_{\bar{A}A} \qquad \text{for } m = 0,1,\ldots,k.$$

PROOF. The assertion follows by invoking the results of the lemmas 6.4.1 and 6.4.3 in (6.2;4) and working out the product.

LEMMA 6.4.5. *If* $(N_{\ell+1})_{\bar{A}}$ *and* $(G_{\ell})_{\bar{A}}$ *are finite for some* $\ell \in \mathbb{N}$ *and* $A_0 \subseteq A \subset \Psi$ *then* $k(\rho;\bar{A})_{\bar{A}}$ *has the asymptotic expansion*

$$k(\rho;\bar{A})_{\bar{A}} = \sum_{m=0}^{\ell} \rho^m K_m(\bar{A})_{\bar{A}} + o(\rho^{\ell}), \quad \rho \downarrow 0,$$

where the coefficients $K_m(\bar{A})_{\bar{A}} \in \mathbb{R}^{|\bar{A}|}$ *are uniquely determined by*

$$K_m(\bar{A})_{\bar{A}} = \sum_{i=0}^{m} (-1)^i R_{m-i}(\bar{A})(G_i)_{\bar{A}} \quad \text{for } m = 0,1,\ldots,\ell.$$

PROOF. The assertion follows by invoking the results of the lemmas 6.4.2 and 6.4.3 in (6.2;5) and working out the product. □

Observing that $k(\rho;\bar{A})_A = 0$, the expansion of lemma 6.4.5 is extended to the whole state space by taking $K_m(\bar{A})_A = 0$ for $m = 0,1,\ldots,\ell$.

LEMMA 6.4.6. *If* N_{k+1} *is finite for some* $k \in \mathbb{N}$ *then* $p(\rho;A(z))$ *has for each* $z \in Z$ *the asymptotic expansion*

$$p(\rho;A(z)) = \sum_{m=0}^{k} P_m(A(z))\rho^m + o(\rho^k), \quad \rho \downarrow 0,$$

where the coefficients $P_m(A(z)) \in \mathbb{R}^{|A(z)|\times|A(z)|}$ *are given by*

$$P_m(A(z)) = P(z)_{A(z)\overline{A(z)}}\, U_m(\overline{A(z)}) + \begin{cases} P(z)_{A(z)} & \text{for } m = 0 \\ 0 & \text{for } m = 1,\ldots,k. \end{cases}$$

PROOF. The assertion follows by substitution of the result of lemma 6.4.4 in (6.2;7). □

LEMMA 6.4.7. *If* N_{k+1} *is finite for some* $k \in \mathbb{N}$ *then* $\gamma(\rho;z)$ *has for each fixed policy* $z \in Z$ *the asymptotic expansion*

$$\gamma(\rho;z) = \sum_{m=0}^{k} \rho^m \Gamma_m(z) + o(\rho^k), \quad \rho \downarrow 0,$$

where the coefficients $\Gamma_m(z) \in \mathbb{R}^{J'\times|A(z)|}$ *are given by*

$$\Gamma_m(z) = \begin{bmatrix} U_m(\overline{A(z)}) \\ \hline P_m(A(z)) \end{bmatrix}.$$

PROOF. The assertion follows from the lemmas 6.4.4 and 6.4.6. □

Next a parametric GMP-model is introduced in

DEFINITION 6.4.8. A *parametric GMP-model* (abbreviation: PGMP-model) is defined by extending the finite GMP-model as follows.

(i) Assumption 2.2.6 is dropped.

(ii) Each intervention $x \in X(i)$, $i \in \Psi$ induces

(1) after $\varepsilon \geq 0$ time units an instantaneous (possibly random) trans-

formation of the state of the system to a state $\underline{j}$ with probability p^x_{ij},

(2) an infinitely differentiable function $G(\varepsilon)^x_i$ for each $\varepsilon \in \mathbb{R}$ represented by its Taylor expansion at $\varepsilon = 0$ given by

$$G(\varepsilon)^x_i = \sum_{j=0}^{\infty} (j!)^{-1} \varepsilon^j G_j(0)^x_i,$$

where $G_j(0)^x_i$ is the j-th derivative of $G(\varepsilon)^x_i$ at $\varepsilon = 0$,

(3) a function $K(t,\varepsilon)^x_i$ defined by

$$K(t,\varepsilon)^x_i = \begin{cases} G(\varepsilon)^x_i + \sum_{j\in\Psi} p^x_{ij} K(t-\varepsilon;\bar{A}_0)_j & \text{for } t \geq \varepsilon \\ -K(t;\bar{A}_0)_i & \\ 0 & \text{otherwise.} \end{cases}$$

The notation of the finite GMP-model is extended to the PGMP-model by

DEFINITION 6.4.9. For each fixed policy $z \in Z$ define

(i) the matrix function $P(t,\varepsilon;A(z))$ by

$$P(t,\varepsilon;A(z)) \stackrel{d}{=} \begin{cases} P(t-\varepsilon;A(z)) & \text{for } t \geq \varepsilon \\ 0 & \text{otherwise,} \end{cases}$$

(ii) the matrix function $\Gamma(t,\varepsilon;z)$ by

$$\Gamma(t,\varepsilon;z) \stackrel{d}{=} \begin{bmatrix} U(t;\overline{A(z)}) \\ \hline P(t,\varepsilon;A(z)) \end{bmatrix},$$

(iii) the vector function $K(t,\varepsilon;z)$ with components

$$K(t,\varepsilon;z)_i \stackrel{d}{=} \begin{cases} K(t,\varepsilon)^{z(i)}_i & \text{for } i \in A(z) \\ 0 & \text{for } i \notin A(z) \end{cases}$$

and

(iv) the vector function $G(\varepsilon;z)$ by its components

$$G(\varepsilon;z)_i \stackrel{d}{=} \begin{cases} G(\varepsilon)^{z(i)}_i & \text{for } i \in A(z) \\ 0 & \text{for } i \notin A(z). \end{cases}$$

The LSTs with respect to t of the first three of these quantities are respectively denoted by $p(\rho,\varepsilon;z)$, $\gamma(\rho,\varepsilon;z)$ and $k(\rho,\varepsilon;z)$. It follows easily that

(6.4.1) $\quad p(\rho,\varepsilon;A(z)) = e^{-\rho\varepsilon}p(\rho;A(z)) \qquad$ for $\rho,\varepsilon \in \mathbb{R}_+$

and

(6.4;2) $\quad k(\rho,\varepsilon;z)_{A(z)} = e^{-\rho\varepsilon}[G(\varepsilon;z)_{A(z)} + P(z)_{A(z)}\Psi^{k(\rho;\overline{A_0})}]-k(\rho;\overline{A_0})_{A(z)}$

for $\rho,\varepsilon \in \mathbb{R}_+$.

Observe that $k(\rho,\varepsilon;z)_{\overline{A(z)}} = 0$.

In the sequel some results are derived for the PGMP-model.

LEMMA 6.4.10. *For each fixed policy* $z \in Z$ *and fixed* $\varepsilon \geq 0$ *the function* $e^{\rho\varepsilon}\, k(\rho,\varepsilon;z)_{A(z)}$ *has the following expansion*

$$e^{\rho\varepsilon}\, k(\rho,\varepsilon;z)_{A(z)} = \sum_{k=0}^{m} \rho^k \sum_{\ell=0}^{\infty} \varepsilon^{\ell}\, K_{k\ell}(z)_{A(z)} + o(\rho^m), \qquad \rho \downarrow 0,$$

if N_{m+1} *and* G_m *are finite for some* $m \in \mathbb{N}$. *The coefficients* $K_{k\ell}(z)_{A(z)}$ *are given by*

$$K_{k\ell}(z)_{A(z)} = \begin{cases} (\ell!)^{-1} G_{\ell}(0;z)_{A(z)} & \text{for } k = 0 \text{ and } \ell \geq 0 \\ 0 & \text{otherwise} \end{cases} +$$

$$+ \begin{cases} P(z)_{A(z)}\Psi\, K_k(\bar{A}_0) & \text{for } \ell = 0 \text{ and } k \geq 0 \\ 0 & \text{otherwise} \end{cases} +$$

$$- \begin{cases} K_{k-\ell}(\bar{A}_0)_{A(z)}\,(\ell!)^{-1} & \text{for } k \geq \ell \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

PROOF. The assertion follows straightforwardly from lemma 6.4.5 and relation (6.4;2). □

The following lemma extends lemma 6.2.4 to the PGMP-model.

LEMMA 6.4.11. *In the PGMP-model the integral equation*

$$V(t,\varepsilon;z) = K(t,\varepsilon;z) + \Gamma(t,\varepsilon;z) * V(t,\varepsilon;z)_{A(z)}$$

has for each fixed $\varepsilon > 0$ *and policy* $z \in Z$ *a unique solution in* $F_{J'}$.

PROOF. The assertion follows since $K(t,\varepsilon;z) \in F_{J'}$ and $P(t,\varepsilon;z)$ is a normal SMM for each fixed $\varepsilon > 0$. □

In the sequel $V(t,\varepsilon;z)$ denotes the solution specified by lemma 6.4.11 for fixed $\varepsilon > 0$. Its LST with respect to t, $v(\rho,\varepsilon;z)$, satisfies for $\rho > 0$, $\varepsilon > 0$

$$(6.4;3) \qquad v(\rho,\varepsilon;z) = k(\rho,\varepsilon;z) + \gamma(\rho,\varepsilon;z)v(\rho,\varepsilon;z)_{A(z)}.$$

The interpretations of $V(t,\varepsilon;z)_i$ and $v(\rho,\varepsilon;z)_i$, $i \in \Psi$ are related to those given for $V(t;z)_i$ and $v(\rho;z)_i$ in an obvious way. In preparation of theorem 6.4.16 some additional results are needed. The first is related to the natural process in any GMP-model.

LEMMA 6.4.12. *Let the set* A *satisfy* $A_0 \subseteq A \subset \Psi$. *Then*

$$-U_1(\bar{A})1_A = R_0(\bar{A})\tau_{\bar{A}}.$$

PROOF. By the lemmas 6.4.3 and 6.4.4

$$\begin{aligned} -U_1(\bar{A}) &= -\sum_{\ell=0}^{1} (-1)^{\ell} R_{1-\ell}(\bar{A})(N_\ell)_{\bar{A}A} \\ &= -R_1(\bar{A})(N_0)_{\bar{A}A} + R_0(\bar{A})(N_1)_{\bar{A}A} \\ &= R_0(\bar{A})(N_1)_{\bar{A}} R_0(\bar{A})(N_0)_{\bar{A}A} + R_0(\bar{A})(N_1)_{\bar{A}A}. \end{aligned}$$

Postmultiplying by 1_A and applying lemma 2.3.6 (ii) yields

$$-U_1(\bar{A})\,1_A = R_0(\bar{A})[(N_1)_{\bar{A}}\,1_{\bar{A}} + (N_1)_{\bar{A}A}\,1_A] = R_0(\bar{A})\tau_{\bar{A}},$$

completing the proof. □

LEMMA 6.4.13. *Let* $b(z),q(z) \in \mathbb{R}^{|A(z)|}$ *be given vectors and let* $S_0^*(z)_{A(z)}\, b(z) = 0$. *Let* $P(t,\varepsilon;A(z))$ *be a normal* SMM *for* $\varepsilon = 0$ *and let* $P_1(A(z))$ *be finite. Then the system of equations in* $y_{A(z)}$ *and* $w_{A(z)}$ *given by*

$$(1) \qquad \begin{cases} (a) & y_{A(z)} - P_0(A(z))y_{A(z)} = b(z) \\ (b) & w_{A(z)} - P_0(A(z))w_{A(z)} = q(z) + P_1(A(z))y_{A(z)} \end{cases}$$

has a unique solution in $y_{A(z)}$. *This solution is given by*

$$(2) \qquad y_{A(z)} = H_0(A(z))b(z) + \sum_{\lambda=1}^{L(z)} \alpha_\lambda \phi_\lambda(A(z))$$

where $H_0(A(z))$ *is specified by definition* 2.3.8 (iv), $\phi_\lambda(A(z))$ *is defined similarly as in theorem* 2.3.14 *and the scalars* $\alpha_\lambda(z)$, $\lambda = 1,\ldots,L(z)$ *are given by*

$$(3) \qquad \alpha_\lambda(z) = \frac{\langle S_0^*(z)_i, d(z)\rangle}{\langle S_0^*(z)_i, t(z)\rangle} \qquad \text{for } i \in E_\lambda(z).$$

The vector $d(z) \in \mathbb{R}^{J'}$ *in* (3) *is defined by*

$$d(z)_{A(z)} \overset{d}{=} q(z) + P_1(A(z))H_0(A(z))b(z)$$

$$d(z)_{\overline{A(z)}} \overset{d}{=} 0.$$

PROOF. By lemma 1.2.8 the general solution to (1)(a) is given by

$$y_{A(z)} = H_0(A(z))b(z) + \beta(z)$$

where $\beta(z) \overset{d}{=} \sum_{\lambda=1}^{L(z)} \alpha_\lambda(z)\phi_\lambda(A(z))$.
For subchain $E_\lambda(A(z))$ we have then

$$\beta(z)_{E_\lambda(A(z))} = \alpha_\lambda(z)1_{E_\lambda(A(z))} \quad \text{for } \lambda = 1,\ldots,L(z).$$

By applying successively the lemmas 6.4.6, 6.4.12, 2.3.11 (ii) and 2.3.12 (ii) we have for subchain $E_\lambda(A(z))$

$$(4) \qquad -S_0^*(z)_{E_\lambda(z)E_\lambda(A(z))}\ P_1(E_\lambda(A(z))\beta(z)_{E_\lambda(A(z))} =$$

$$= \alpha_\lambda(z)[-S_0^*(z)_{E_\lambda(z)E_\lambda(A(z))}\ P(z)_{E_\lambda(A(z))E_\lambda(\overline{A(z)})} \cdot$$

$$\cdot\ U_1(\overline{A(z)})_{E_\lambda(\overline{A(z)})E_\lambda(A(z))}1_{E_\lambda(A(z))}]$$

$$= \alpha_\lambda(z)[S_0^*(z)_{E_\lambda(z)E_\lambda(A(z))}\ P(z)_{E_\lambda(A(z))E_\lambda(\overline{A(z)})} \cdot$$

$$\cdot\ R_0(\overline{A(z)})_{E_\lambda(\overline{A(z)})}\tau_{E_\lambda(\overline{A(z)})}]$$

$$= \alpha_\lambda(z)[S_0^*(z)_{E_\lambda(z)E_\lambda(\overline{A(z)})}\ \tau_{E_\lambda(\overline{A(z)})}]$$

$$= \alpha_\lambda(z)[S_0^*(z)_{E_\lambda(z)}\ t(z)_{E_\lambda(z)}].$$

By premultiplying the $E_\lambda(A(z))$-part of (1)(b) with $S_0^*(z)_{E_\lambda(z)E_\lambda(A(z))}$ and substituting (4), it follows that for subchain $E_\lambda(z)$, $\lambda = 1,\ldots,L(z)$

$$\alpha_\lambda(z)[S_0^*(z)_{E_\lambda(z)}t(z)_{E_\lambda(z)}] = S_0^*(z)_{E_\lambda(z)}d(z)_{E_\lambda(z)},$$

which implies (3) and completes the proof. □

The following lemma extends the result of lemma 6.4.13 to the set of states Ψ.

LEMMA 6.4.14. *Let* $b(z),q(z) \in \mathbb{R}^{J'}$ *and let* $S_0^*(z)_{A(z)}b(z)_{A(z)} = 0$. *Let* $P(t,\varepsilon;A(z))$ *be a normal* SMM *for* $\varepsilon = 0$ *and let* $P_1(A(z))$ *be finite. Then the system of equations in* (y,w) *given by*

$$(1)\qquad \begin{cases} (a) & y - \Gamma_0(z)y_{A(z)} = b(z) \\ (b) & w - \Gamma_0(z)w_{A(z)} = q(z) + \Gamma_1(z)y_{A(z)} \end{cases}$$

has a unique solution in y. $y_{A(z)}$ *is specified by lemma* 6.4.13 (2) *and* (3) *and* $y_{\overline{A(z)}}$ *by*

$$(2)\qquad y_{\overline{A(z)}} = U_0(\overline{A(z)})y_{A(z)} + b(z)_{\overline{A(z)}}.$$

PROOF. The $A(z)$-part of the assertion follows immediately from lemma 6.4.13. The $\overline{A(z)}$-part of (1) is identical to (2) and uniquely determines $y_{\overline{A(z)}}$. □

Observe that the system (1) of lemma 6.4.14 with $b(z) \overset{d}{=} 0$ is closely related to the system of equations considered in theorem 2.3.16. The following corollary states the connection between their solutions.

COROLLARY 6.4.15. *The system of equations in* (y,w) *given by*

$$(1)\qquad \begin{cases} y - \Gamma_0(z)y_{A(z)} = 0 \\ w - \Gamma_0(z)w_{A(z)} = k(z) - y\square t(z) \end{cases}$$

and

$$(2)\qquad \begin{cases} y - \Gamma_0(z)y_{A(z)} = 0 \\ w - \Gamma_0(z)w_{A(z)} = k(z) + \Gamma_1(z)y_{A(z)} \end{cases}$$

have identical and unique solutions in y.

PROOF. The assertion follows from theorem 2.3.16 and lemma 6.4.14 with $q(z) = k(z)$ and $b(z) = 0$. □

In preparation of theorem 6.4.16, a distinction among policies in Z is introduced. A policy is called *normal* if $P(t,\varepsilon;A(z))$ is a normal SMM for $\varepsilon = 0$ and *non-normal* otherwise. For a non-normal policy we have

$$(6.4;4) \qquad E_\lambda(A(z)) = E_\lambda(z) \qquad \text{for some } \lambda \in \{1,\dots,L(z)\}.$$

A subchain satisfying (6.4;5) is called a *non-normal subchain.*

THEOREM 6.4.16. *Let* N_{m+3} *and* G_{m+1} *be finite for some* $m \in \mathbb{N}$. *Then for a fixed policy* $z \in Z$, $v(\rho,\varepsilon;z)$ *has a partial Laurent expansion in* ρ, *which is for each sufficiently small fixed* $\varepsilon > 0$ *given by*

$$(1) \qquad v(\rho,\varepsilon;z) = \sum_{h=-1}^{m} \rho^h \sum_{\ell=-1}^{\infty} \varepsilon^\ell V_{h\ell}(z) + o(\rho^m), \quad \rho \downarrow 0.$$

For fixed $h \in \{-1,0,1,\dots,m\}$ *and* $\ell \in \{-1,0,1,\dots\}$ *and normal z the system of equations*

$$(2) \qquad \begin{cases} (a) \quad V_{h\ell}(z) - \Gamma_0(z)V_{h\ell}(z)_{A(z)} = K'_{h,\ell}(z) + \Gamma_1(z)V_{h-1,\ell}(z)_{A(z)} + \\ \qquad\qquad - \overset{0}{V}_{h-1,\ell-1}(z) \\ (b) \quad V_{h+1,\ell}(z) - \Gamma_0(z)V_{h+1,\ell}(z)_{A(z)} = K'_{h+1,\ell}(z) + \Gamma_1(z)V_{h,\ell}(z)_{A(z)} + \\ \qquad\qquad - \overset{0}{V}_{h,\ell-1}(z) \end{cases}$$

has a solution $(V_{h,\ell}(z), V_{h+1,\ell}(z))$ *which is unique in* $V_{h,\ell}(z)$. *Define* $V_{h,\ell}(z) = 0$ *otherwise. In* (2) $\overset{0}{V}_{h\ell}(z)$ *and* $K'_{h\ell}(z)$ *are respectively defined by*

$$(3) \qquad \overset{0}{V}_{h\ell}(z) = \begin{bmatrix} 0 \\ \text{---------} \\ V_{h\ell}(z)_{A(z)} \end{bmatrix}$$

and

$$(4) \qquad K'_{h\ell}(z) = K_{h\ell}(z) + \sum_{k=2}^{h+1} \Gamma_k(z)V_{h-k,\ell}(z)_{A(z)} - \sum_{k=2}^{\min(h,\ell)+1} \cdot$$

$$\cdot (k!)^{-1} \overset{0}{V}_{h-k,\ell-k}(z).$$

For non-normal z (2)(b) *has to be replaced by*

$$(2)(c)\qquad V_{h+1,\ell+1}(z)E_\lambda(z) - \Gamma_0(z)E_\lambda(z)V_{h+1,\ell+1}(z)E_\lambda(z) =$$

$$= K'_{h+1,\ell+1}(z)E_\lambda(z) - V_{h,\ell}(z)E_\lambda(z)$$

for each non-normal subchain $E_\lambda(z)$.

PROOF. Consider equation (6.4;3) in $v(\rho,\varepsilon;z)$. It will be proven that the expansion

$$(5)\qquad v(\rho,\varepsilon;z) = \sum_{h=-1}^{m}\rho^k\sum_{\ell=-1}^{\infty}\varepsilon^\ell V_{h\ell}(z) + f_m(\rho,\varepsilon)$$

satisfies (6.4;3) with $f_m(\rho,\varepsilon) = o(\rho^m)$, $\rho \downarrow 0$ for each fixed and sufficiently small $\varepsilon > 0$ if the vectors $V_{h\ell}(z)$, $h \in \{-1,0,1,\ldots,m\}$, $\ell \in \{-1,0,1,\ldots\}$ are uniquely determined by (2).

If the A(z)-part of (6.4;3) is multiplied by $e^{\rho\varepsilon}$ and the expansions of $e^{\rho\varepsilon}p(\rho,\varepsilon;A(z)) = p(\rho;A(z))$ (cf. lemma 6.4.6 and relation 6.4;1), $e^{\rho\varepsilon}k(\rho,\varepsilon;z)_{A(z)}$ (cf. lemma 6.4.10), $v(\rho,\varepsilon;z)$ (cf. (5)) and $e^{\rho\varepsilon}$ are inserted then we obtain

$$(6)\qquad \left\{\sum_{k=0}^{\infty}\frac{(\rho\varepsilon)^k}{k!}\right\}\left\{\sum_{h=-1}^{m}\rho^h\sum_{\ell=-1}^{\infty}\varepsilon^\ell V_{h\ell}(z)_{A(z)}\right\} + e^{\rho\varepsilon}f_m(\rho,\varepsilon)_{A(z)} =$$

$$= \sum_{h=0}^{m+1}\rho^h\sum_{\ell=0}^{\infty}\varepsilon^\ell K_{h\ell}(z)_{A(z)} + \sum_{k=0}^{m+2}P_k(A(z))\rho^k\sum_{h=-1}^{m}\rho^h \cdot$$

$$\cdot\sum_{\ell=-1}^{\infty}\varepsilon^\ell V_{h\ell}(z)_{A(z)} + p(\rho;A(z))f_m(\rho,\varepsilon)_{A(z)} + g_1(\rho,\varepsilon),$$

where $g_1(\rho,\varepsilon) = o(\rho^{m+1})$, $\rho \downarrow 0$ for each sufficiently small fixed $\varepsilon > 0$ (cf. the lemmas 6.4.6 and 6.4.10). Observe that

$$(7)\qquad \sum_{k=0}^{\infty}(\rho\varepsilon)^k(k!)^{-1}\sum_{h=-1}^{m}\rho^h\sum_{\ell=-1}^{\infty}\varepsilon^\ell V_{h\ell}(z)_{A(z)} =$$

$$= \sum_{h=-1}^{m}\rho^h\sum_{\ell=-1}^{\infty}\varepsilon^\ell\sum_{k=0}^{\min(h,\ell)+1}V_{h-k,\ell-k}(z)_{A(z)} +$$

$$+ \rho^{m+1}\sum_{\ell=-1}^{\infty}\varepsilon^\ell\sum_{k=1}^{\min(m+1,\ell)+1}(k!)^{-1}V_{m+1-k,\ell-k}(z)_{A(z)} + g_2(\rho,\varepsilon)$$

and

$$(8)\qquad \sum_{k=0}^{m+2} P_k(A(z))\rho^k \sum_{h=-1}^{m} \rho^h \sum_{\ell=-1}^{\infty} \varepsilon^\ell V_{h\ell}(z)_{A(z)} =$$

$$= \sum_{h=-1}^{m} \rho^h \sum_{\ell=-1}^{\infty} \varepsilon^\ell \sum_{k=0}^{h+1} P_k(A(z))V_{h-k,\ell}(z)_{A(z)} +$$

$$+ \rho^{m+1} \sum_{\ell=-1}^{\infty} \varepsilon^\ell \sum_{k=1}^{m+2} P_k(A(z))V_{m+1-k,\ell}(z)_{A(z)} + g_3(\rho,\varepsilon),$$

where $g_2(\rho,\varepsilon) = o(\rho^{m+1})$ and $g_3(\rho,\varepsilon) = o(\rho^{m+1})$ for $\rho \downarrow 0$ and sufficiently small fixed $\varepsilon > 0$. Substitution of (7) and (8) in (6) yields

$$(9)\qquad \sum_{h=-1}^{m} \rho^h \sum_{\ell=-1}^{\infty} \varepsilon^\ell \Big\{ \sum_{k=0}^{\min(h,\ell)+1} (k!)^{-1} V_{h-k,\ell-k}(z)_{A(z)} +$$

$$- \sum_{k=0}^{h+1} P_k(A(z))V_{h-k,\ell}(z)_{A(z)} \Big\} +$$

$$+ \rho^{m+1} \Big[\sum_{\ell=-1}^{\infty} \varepsilon^\ell \Big\{ \sum_{k=1}^{\min(m+1,\ell)+1} (k!)^{-1} V_{m+1-k,\ell-k}(z)_{A(z)} +$$

$$- \sum_{k=1}^{m+2} P_k(A(z))V_{m+1-k,\ell}(z)_{A(z)} \Big] +$$

$$+ \Big[e^{\rho\varepsilon} I_{A(z)} - p(\rho;A(z)) \Big] f_m(\rho,\varepsilon)_{A(z)} =$$

$$= \sum_{h=0}^{m} \rho^h \sum_{\ell=0}^{\infty} \varepsilon^\ell K_{h\ell}(z)_{A(z)} + \rho^{m+1} \sum_{\ell=0}^{\infty} \varepsilon^\ell K_{m+1,\ell}(z)_{A(z)} + g(\rho,\varepsilon),$$

where $g(\rho,\varepsilon) \overset{d}{=} g_1(\rho,\varepsilon) - g_2(\rho,\varepsilon) + g_3(\rho,\varepsilon)$.
If the vectors $V_{h\ell}(z)_{A(z)}$ satisfy

$$(10)\qquad \sum_{k=0}^{\min(h,\ell)+1} (k!)^{-1} V_{h-k,\ell-k}(z)_{A(z)} - \sum_{k=0}^{h+1} P_k(A(z))V_{h-k,\ell}(z) =$$

$$= K_{h\ell}(z)_{A(z)}$$

for $h \in \{-1,0,1,\ldots,m\}$, $\ell \in \{-1,0,1,\ldots\}$ then (9) implies

$$(11)\qquad \Big[e^{\rho\varepsilon} I_{A(z)} - p(\rho;A(z)) \Big] f_m(\rho,\varepsilon)_{A(z)} =$$

$$= \rho^{m+1} \Big[\sum_{\ell=0}^{\infty} \varepsilon^\ell K_{m+1,\ell}(z)_{A(z)} + \sum_{\ell=-1}^{\infty} \varepsilon^\ell \sum_{k=1}^{m+2} P_k(A(z)) \cdot$$

$$\cdot V_{m+1-k,\ell}(z)_{A(z)} \Big] +$$

$$- \rho^{m+1} \Big[\sum_{\ell=-1}^{\infty} \varepsilon^\ell \sum_{k=1}^{\min(m+1,\ell)+1} (k!)^{-1} V_{m+1-k,\ell-k}(z)_{A(z)} \Big] + g(\rho,\varepsilon).$$

Substitution of (3) and (4) in (10) yields the A(z)-part of (2)(a). The A(z)-part of (2)(b) is obtained if h is replaced by h+1 in (10). Lemma

6.4.13 guarantees the unique solution for $V_{h,\ell}(z)_{A(z)}$ of the A(z)-part of (2) if the vectors $V_{i,j}(z)$, $i \leq h-1$ and $j \leq \ell$ are known. To show that $f_m(\rho,\varepsilon)_{A(z)} = o(\rho^m)$, $\rho \downarrow 0$ (4) is substituted in (11), yielding

$$(12) \qquad [e^{\rho\varepsilon} I_{A(z)} - p(\rho;A(z))]\, f_m(\rho,\varepsilon)_{A(z)} =$$

$$= e^{\rho\varepsilon}[I_{A(z)} - p(\rho,\varepsilon;A(z))]\, f_m(\rho,\varepsilon)_{A(z)} =$$

$$= \rho^{m+1}\Big[\sum_{\ell=-1}^{\infty} \varepsilon^{\ell}\Big\{K'_{m+1,\ell}(z)_{A(z)} + P_1(A(z))V_{m\ell}(z)_{A(z)} +$$

$$- V_{m,\ell-1}(z)_{A(z)}\Big\}\Big] + g(\rho,\varepsilon).$$

Since $p(\rho,\varepsilon;A(z))$ is transient for each fixed $\rho > 0$ and $\varepsilon > 0$, lemma 1.2.1 (ii) implies that (12) is equivalent to

$$(13) \qquad f_m(\rho,\varepsilon)_{A(z)} = e^{-\rho\varepsilon}\{[I_{A(z)} - p(\rho,\varepsilon;A(z))]^{-1}\, g(\rho,\varepsilon)\} +$$

$$+ e^{-\rho\varepsilon}\Big\{\rho^{m+1} \sum_{\ell=-1}^{\infty} \varepsilon^{\ell}[I_{A(z)} - p(\rho,\varepsilon;A(z))]^{-1} \cdot$$

$$\cdot\, [K'_{m+1,\ell}(z)_{A(z)} + P_1(A(z))V_{m\ell}(z)_{A(z)} - V_{m,\ell-1}(z)_{A(z)}]\Big\}.$$

Applying (1.3;11) and lemma 1.3.4 yields for each fixed $\varepsilon > 0$

$$(14) \qquad [I_{A(z)} - p(\rho,\varepsilon;A(z))]^{-1} = O(1/\rho),\ \rho \downarrow 0,$$

and

$$(15) \qquad [I_{A(z)} - p(\rho,\varepsilon;A(z))]^{-1}[K'_{m+1,\ell}(z)_{A(z)} + P_1(A(z))V_{m\ell}(z)_{A(z)} -$$

$$- V_{m,\ell-1}(z)_{A(z)}] = o(1/\rho),\ \rho \downarrow 0.$$

Substitution of (14) and (15) in (13) and the observation that $g(\rho,\varepsilon) = o(\rho^{m+1})$, $\rho \downarrow 0$ for each sufficiently small fixed $\varepsilon > 0$ yields

$$(16) \qquad f_m(\rho,\varepsilon)_{A(z)} = e^{-\rho\varepsilon}\{O(1/\rho)g(\rho,\varepsilon) + \rho^{m+1}\, o(1/\rho),\ \rho \downarrow 0\} =$$

$$= o(\rho^m),\ \rho \downarrow 0,$$

for each sufficiently small fixed $\varepsilon > 0$, completing the A(z)-part of the assertion for a normal policy z.

For $v(\rho,\varepsilon;z)_{\overline{A(z)}}$ we have by (6.4;4)

$$(17) \qquad v(\rho,\varepsilon;z)_{\overline{A(z)}} = u(\rho;\overline{A(z)})v(\rho,\varepsilon;z)_{A(z)}$$

Inserting (1) of lemma 6.4.4 and the A(z)-part of (5) yields

$$(18) \qquad v(\rho,\varepsilon;z)_{\overline{A(z)}} = \left[\sum_{h=0}^{m+1} \rho^h U_h(\overline{A(z)}) + o(\rho^{m+1})\right] \cdot \left[\sum_{\ell=-1}^{\infty} \varepsilon^\ell \sum_{h=-1}^{m} \rho^h V_{h\ell}(z)_{A(z)} + o(\rho^m)\right], \quad \rho \downarrow 0.$$

for sufficiently small fixed $\varepsilon > 0$. By developing the product in the right-hand member of (18)

$$(19) \qquad v(\rho,\varepsilon;z)_{\overline{A(z)}} = \sum_{h=-1}^{m} \rho^h \sum_{\ell=-1}^{\infty} \varepsilon^\ell \sum_{k=0}^{h+1} U_k(\overline{A(z)})V_{h-k,\ell}(z)_{A(z)} + o(\rho^m), \ \rho \downarrow 0$$

is obtained implying the assertion for a normal policy z.

For a non-normal policy z we have for each non-normal subchain $E_\lambda(z)$

$$(20) \qquad P_k(E_\lambda(z)) = 0 \quad \text{for } k > 0.$$

Hence in (2)(b) the term $\Gamma_1(z)_{E_\lambda(z)} V_{h,\ell}(z)_{E_\lambda(z)}$ vanishes and $V_{k\ell}(z)_{E_\lambda(z)}$ cannot be determined by (2). Instead of the equation for $V_{h+1,\ell}(z)$ we take the one for $V_{h+1,\ell+1}(z)$, yielding (2)(c) for non-normal subchain $E_\lambda(z)$. By taking $P_1(E_\lambda(A(z)) = -I_{E_\lambda(A(z))}$ in lemma 6.4.13 it follows that $V_{h\ell}(z)_{E_\lambda(z)}$ is uniquely determined by (2)(a) and (2)(c) in this case.

REMARK 6.4.17. The partial Laurent expansion for the expected discounted reward vector $\hat{v}(\rho;f)$ in the MRD-model (cf. [DENARDO 1971]) is obtained if $\varepsilon = 0$ is substituted in theorem 6.4.16 for a normal policy z. Then we obtain

$$(1) \qquad v(\rho,0;z) = \sum_{h=-1}^{m} \rho^h V_{h,0}(z) + o(\rho^m), \quad \rho \downarrow 0$$

where $V_{h,0}(z)$ satisfies

$$(2) \qquad V_{h,0}(z) - \Gamma_0(z)V_{h,0}(z)_{A(z)} = K_{h,0}(z) + \sum_{k=1}^{h+1} \Gamma_k(z) \cdot V_{h-k,0}(z)_{A(z)}$$

and is uniquely determined by (2) and the equation in $V_{h+1,0}(z)$. The coefficient $\hat{V}_h(f)$ of the expansion in the MRD-model is obtained from

$$(3) \qquad [I-Q_0(f)]\hat{V}_h(f) = (-1)^h C_h(f) + \sum_{k=1}^{h+1} (-1)^k Q_k(f)\hat{V}_{h-k}(f)$$

and a similar equation in $\hat{V}_{h+1}(f)$. To display the connection between (3) and (2) observe that the following notational identities hold for corresponding z and f

$$Q_0(f) = S_0(z) = \left[\begin{array}{c|c} (N_0)_{\overline{A(z)}} & (N_0)_{\overline{A(z)}A(z)} \\ \hline P(z)_{A(z)\overline{A(z)}} & P(z)_{A(z)} \end{array}\right]$$

$$Q_h(f) = S_h(z) = \left[\begin{array}{c|c} (N_h)_{\overline{A(z)}} & (N_h)_{\overline{A(z)}A(z)} \\ \hline 0 & 0 \end{array}\right] \qquad h = 1,2,\ldots$$

$$C_h(f) = \left[\begin{array}{c} G_h \\ \hline G_0(0;z)\delta_h \end{array}\right] \qquad h = 0,1,\ldots$$

where $\delta_h = 1$ if $h = 0$ and $\delta_h = 0$ otherwise. The derivation will be restricted to an indication of the main steps. The relation

$$(4) \qquad k(\rho;\bar{A}_0)_{\overline{A(z)}} = k(\rho;\overline{A(z)})_{\overline{A(z)}} + u(\rho;\overline{A(z)})k(\rho;\bar{A}_0)_{A(z)}$$

is used, which implies the following relation between the moments (whenever they exist)

$$(5) \qquad K_h(\bar{A}_0)_{\overline{A(z)}} = K_h(\overline{A(z)})_{\overline{A(z)}} + \sum_{k=0}^{h+1} U_k(\overline{A(z)})K_{h-k}(\bar{A}_0)_{A(z)}.$$

Using (5) and lemma 6.4.10 we have for $K_{h,0}(z)_{A(z)}$

$$(6) \qquad K_{h,0}(z)_{A(z)} = G_0(0;z)_{A(z)}\delta_h + P(z)_{A(z)\overline{A(z)}}K_h(\overline{A(z)})_{\overline{A(z)}} + \\ - K_h(\bar{A}_0)_{A(z)} + \\ + \sum_{k=0}^{h+1} P_k(A(z))K_{h-k}(\bar{A}_0)_{A(z)}.$$

Inserting (5) and (6) into (2) yields

$$(7)\quad \begin{cases} (a)\ \bar{V}_{h,0}(z)_{\overline{A(z)}} - U_0(\overline{A(z)})\bar{V}_{h,0}(z) = K_h(\overline{A(z)})_{\overline{A(z)}} + \\ \qquad\qquad + \sum_{k=1}^{h+1} U_k(\overline{A(z)})\bar{V}_{h-k,0}(z)_{A(z)} \\ (b)\ \bar{V}_{h,0}(z)_{A(z)} - P_0(\overline{A(z)})\bar{V}_{h,0}(z)_{A(z)} = G_0(0;z)_{A(z)}\delta_h + P(z)_{A(z)\overline{A(z)}} \cdot \\ \qquad\qquad \cdot \left[K_h(\overline{A(z)})_{\overline{A(z)}} + \sum_{k=1}^{h+1} \cdot \right. \\ \qquad\qquad \left. \cdot\ U_k(\overline{A(z)})\bar{V}_{h-k,0}(z)_{A(z)} \right] \end{cases}$$

where $\bar{V}_{h,0}(z) \overset{d}{=} V_{h,0}(z) + K_{h,0}(\bar{A}_0)$. Substitution of (7)(a) in (7)(b) immediately yields for $\bar{V}_{h,0}(z)_{A(z)}$ the equation

$$(8)\qquad \bar{V}_{h,0}(z)_{A(z)} - P(z)_{A(z)\Psi}\,\bar{V}_{h,0}(z) = G_0(0;z)_{A(z)}\,\delta_h.$$

Inserting the moment expansions of the lemmas 6.4.1-6.4.4 and the relationship

$$(9)\qquad \sum_{k=0}^{h+1} U_k(\overline{A(z)})\ \bar{V}_{h-k,0}(z)_{A(z)} =$$

$$= R_0(\overline{A(z)})\ \sum_{k=0}^{h+1} (-1)^k (N_k)_{\overline{A(z)}\Psi}\ \bar{V}_{h-k}(z)$$

yields

$$(10)\qquad \bar{V}_{h,0}(z)_{\overline{A(z)}} - (N_0)_{\overline{A(z)}\Psi}\,\bar{V}_{h,0}(z) =$$

$$= (-1)^h (G_h)_{\overline{A(z)}} + \sum_{k=1}^{h+1} (-1)^k (N_k)_{\overline{A(z)}\Psi}\ \bar{V}_{h-k,0}(z).$$

It follows that (8) and (10) are identical to (3) which implies

$$\bar{V}_{h,0}(z) = \hat{V}_h(f).$$

A direct derivation of a partial Laurent expansion in the MRD-model corresponding to the one of theorem 6.4.16 has been given in [WEEDA 1976].

6.5. NUMERICAL EXAMPLE

Consider the following problem

$$\Psi = \{1,2,3\}$$

$$A_0 = \{1\}$$

Natural process

$$n(\rho) = \begin{bmatrix} 0 & 0 & 0 \\ e^{-\rho} & 0 & 0 \\ \tfrac{1}{2}e^{-\rho/2} & 0 & \tfrac{1}{2}e^{-\rho/2} \end{bmatrix}; \qquad g(\rho) = \begin{bmatrix} 0 \\ 2e^{-\rho} \\ e^{-\frac{1}{2}\rho} \end{bmatrix} .$$

Interventions

$$X(i) = \begin{cases} \{x_0(i)\} & \text{for } i = 2,3 \\ \{1,2,3,4\} & \text{for } i = 1 \end{cases}$$

x	P^x_{11}	P^x_{12}	P^x_{13}	$G_0(0)^x_1$	$G_1(0)^x_1$
1	½	½	0	1	0
2	0	1	0	2	1
3	½	0	½	1	0
4	0	0	1	2	0

$$G_j(0)^x_1 = 0 \qquad \text{for } j > 1, \quad x \in X(1).$$

Results are given for the vectors $V_{-1,0}(z)$, $V_{-1,1}(z)$ and $V_{0,0}(z)$, $z \in Z$.

For $r(\rho;\bar{A}_0)$, $u(\rho;\bar{A}_0)$ and $k(\rho;\bar{A}_0)$ we have

$$(6.5;1) \qquad r(\rho;\bar{A}_0) = [I_{\bar{A}_0} - n(\rho)_{\bar{A}_0}]^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & (1 - \tfrac{1}{2}e^{-\frac{1}{2}\rho})^{-1} \end{bmatrix}$$

$$(6.5;2) \qquad u(\rho;\bar{A}_0) = r(\rho;\bar{A}_0)n(\rho)_{\bar{A}_0 A_0} = \begin{bmatrix} e^{-\rho} \\ e^{-\frac{1}{2}\rho}(2-e^{-\frac{1}{2}\rho})^{-1} \end{bmatrix}$$

$$(6.5;3) \qquad k(\rho;\bar{A}_0) = r(\rho;\bar{A}_0)g(\rho)_{\bar{A}_0} = 2\begin{bmatrix} e^{-\rho} \\ e^{-\frac{1}{2}\rho}(2-e^{-\frac{1}{2}\rho})-1 \end{bmatrix}.$$

By the expansions for $u(\rho;\bar{A}_0)$ and $k(\rho;\bar{A}_0)$ we have (cf. the lemmas 6.4.4 and 6.4.5)

$$(6.5;4) \qquad U_0(\bar{A}_0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \quad U_1(\bar{A}_0) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}; \quad U_2(\bar{A}_0) = \begin{bmatrix} \frac{1}{2} \\ \frac{3}{4} \end{bmatrix}$$

$$(6.5;5) \qquad K_0(\bar{A}_0) = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}; \quad K_1(\bar{A}_0) = \begin{bmatrix} 0 \\ -2 \\ -2 \end{bmatrix}.$$

Denoting the policy with $z(1) = \ell$ by z_ℓ for $\ell = 1,2,3,4$ we have

$$(6.5;6) \qquad Z = \{z_1, z_2, z_3, z_4\}.$$

Each policy $z \in Z$ is normal and has one subchain. In this example we have

$$(6.5;7) \qquad E(A(z)) = A(z) = A_0 = \{1\} \qquad \text{for } z \in Z$$

and hence

$$(6.5;8) \qquad P_0(A(z)) = [1].$$

Using theorem 6.4.16, $V_{-1,0}(z)$, $V_{-1,1}(z)$ and $V_{0,0}(z)$ in this example are respectively obtained from

$$(6.5;9) \qquad \begin{cases} 0 = K'_{0,0}(z)_1 + P_1(A(z))_{11}\, V_{-1,0}(z)_1 \\ V_{-1,0}(z)_{\overline{A(z)}} = U_0(\overline{A(z)})\, V_{-1,0}(z)_1 \end{cases},$$

$$(6.5;10) \qquad \begin{cases} 0 = K'_{0,1}(z)_1 + P_1(A(z))_{11}\, V_{-1,1}(z)_1 - V_{-1,0}(z)_1 \\ V_{-1,1}(z)_{\overline{A(z)}} = U_0(\overline{A(z)})\, V_{-1,1}(z)_1 \end{cases}$$

and

$$(6.5;11) \qquad \begin{cases} 0 = K'_{1,0}(z)_1 + P_1(A(z))\, V_{0,0}(z)_1 - V_{0,-1}(z)_1 \\ V_{0,0}(z)_{\overline{A(z)}} = U_0(\overline{A(z)})\, V_{0,0}(z)_1 + U_1(\overline{A(z)})\, V_{-1,0}(z)_1 \end{cases}.$$

It is easily verified that since $K'_{0,-1}(z) = 0$ we have $V_{0,-1}(z) = 0$. The computations are summarized in the tables 6.5.1, 6.5.2 and 6.5.3.

TABLE 6.5.1

Ingredients for and result of the computation of $V_{-1,0}(z)$

Policy	$P_1(A(z))_{11}$	$K'_{0,0}(z)_1$	$V_{-1,0}(z)$
z_1	$-\frac{1}{2}$	2	[4,4,4]
z_2	-1	4	[4,4,4]
z_3	$-\frac{1}{2}$	2	[4,4,4]
z_4	-1	4	[4,4,4]

TABLE 6.5.2

Ingredients for and result of the computation of $V_{-1,1}(z)$

Policy	$P_1(A(z))_{11}$	$K'_{0,1}(z)_1$	$V_{-1,1}(z)$
z_1	$-\frac{1}{2}$	0	[-8,-8,-8]
z_2	-1	1	[-3,-3,-3]
z_3	$-\frac{1}{2}$	0	[-8,-8,-8]
z_4	-1	0	[-4,-4,-4]

TABLE 6.5.3.

Ingredients for and the result of the computation $V_{0,0}(z)$

Policy	$P_1(A(z))_{11}$	$P_2(A(z))_{11}$	$K'_{1,0}(z)_1$	$V_{0,0}(z)$
z_1	$-\frac{1}{2}$	$\frac{1}{4}$	0	[0,-4,-4]
z_2	-1	$\frac{1}{2}$	0	[0,-4,-4]
z_3	$-\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{2}$	[1,-3,-3]
z_4	-1	$\frac{3}{4}$	1	[1,-3,-3]

Observe that moreover (6.5;4) and (6.5;5) are needed. Notice also that $P_2(A(z))_{11}$ is required to obtain $K'_{1,0}(z)_1$.

CHAPTER VII

SENSITIVE OPTIMALITY IN THE PGMP-MODEL

7.1 INTRODUCTION

Sensitive discount optimality criteria were introduced in [VEINOTT & MILLER 1969] and [VEINOTT 1969] in finite state and action discrete time Markov decision problems with a substochastic transition matrix for each policy. In these papers a lexicographical method is developed which computes a sensitive discount optimal policy in a finite number of computations. An alternative computational approach has been given in [DENARDO 1970]. It performs the computation of bias-optimal policies by solving a sequence of three discrete Markov decision problems by policy iteration or linear programming. The basic tool is the Laurent expansion for the expected discounted reward vector for a fixed policy based upon (1.2;11).

The results in [VEINOTT & MILLER 1969] and [VEINOTT 1969] were recently generalized by [ROTHBLUM 1975] to discrete Markov decision problems having for each policy a non-negative transition matrix with spectral radius not exceeding one. In this case the principle part of the Laurent expansion contains ν terms with $1 \leq \nu \leq J$. Other generalizations are treated by [SLADKÝ 1974] (equivalence between sensitive discount and sensitive averaging criteria) and by [HORDIJK & SLADKÝ 1977] (extension to a countable state space). The partial Laurent expansion (1.4;4) obtained by [DENARDO 1971] is closely related to the special case $\varepsilon = 0$ and a normal policy in the partial Laurent expansion of theorem 6.4.16 as remark 6.4.17 exhibits.

Theorem 6.4.16 is used in this chapter in two ways. Its first use is the development of a new kind of sensitive optimality in the PGMP-model, which has also significance in the MRP-model with interventions.

It is called *sensitive intervention time optimality*. In section 7.2 sensitive intervention time optimal policies are defined and some of their properties established. Their computation is considered in section 7.3.

It is shown that this task can be accomplished by solving a sequence of problems, each satisfying the conditions of the undiscounted GMP-model, by means of one of the methods GMP 1,2,3.

Its second use is in the computation of bias-optimal policies in the PGMP-model by means of one of the methods GMP 1,2,3. The method suggested here uses an idea developed in [DENARDO 1970] for the computation of bias-optimal policies for discrete Markov decision problems. The procedures in the sections 7.3 and 7.4 are illustrated on the numerical example of section 6.5.

7.2 SENSITIVE INTERVENTION TIME OPTIMALITY

In this section sensitive intervention time optimal policies are defined in the PGMP-model and some of their properties are considered. In the sections 7.2 and 7.3 it is tacitly assumed that the matrix N_2 is finite. Primarily a Laurent expansion in ε is derived from theorem 6.4.16 in

LEMMA 7.2.1. *Let $z \in Z$ be a fixed policy. Then the limit*

$$v_0(\varepsilon;z) \stackrel{d}{=} \lim_{\rho \downarrow 0} \rho\, v(\rho,\varepsilon;z)$$

exists for each fixed $\varepsilon \in (0,\varepsilon_0)$ for some sufficiently small $\varepsilon_0 > 0$ and $v_0(\varepsilon;z)$ has the Laurent expansion

$$(1) \qquad v_0(\varepsilon;z) = \sum_{j=-1}^{\infty} \varepsilon^j V_{-1,j}(z) \qquad \text{for } 0 < \varepsilon < \varepsilon_0.$$

PROOF. By multiplying (1) of theorem 6.4.16 with ρ and taking the limit we obtain (1) of this lemma for fixed ε in $(0,\varepsilon_0)$. By the uniqueness theorem for Laurent expansions (1) holds on the whole interval $0 < \varepsilon < \varepsilon_0$. □

Observe that the domain of the function $v_0(\varepsilon;z)$ can be extended by analytic continuation to the annular domain $0 < |\varepsilon| < \varepsilon_0$, $\varepsilon \in \mathbb{C}$ by defining $v_0(\varepsilon;z) \stackrel{d}{=} \sum_{j=-1}^{\infty} \varepsilon^j V_{-1,j}(z)$ on the points for which it is not yet defined by lemma 7.2.1.

DEFINITION 7.2.2. A policy $z^* \in Z$ is (ρ,ε)-*optimal* if for fixed $\rho > 0$ and fixed $\varepsilon > 0$

$$v(\rho,\varepsilon;z^*) \geq v(\rho,\varepsilon;z) \qquad \text{for } z \in Z.$$

The set of (ρ,ε)-optimal policies will be denoted by $D_{\rho,\varepsilon}$.

DEFINITION 7.2.3. A policy $z^* \in Z$ is $(0,\varepsilon)$-*optimal* if for fixed $\varepsilon > 0$

$$v_0(\varepsilon;z^*) \geq v_0(\varepsilon;z)$$

The set of $(0,\varepsilon)$-optimal policies is denoted by $D_{0,\varepsilon}$.

DEFINITION 7.2.4. A policy $z^* \in Z$ is *sensitive intervention time optimal* if it is $(0,\varepsilon)$-optimal on some real interval $0 < \varepsilon < \mu$ where μ is positive. The set of these policies is denoted by D_0.

REMARK 7.2.5. Observe that the PGMP-model with ε fixed corresponds to a MRP-model where each intervention is replaced by an action with time ε. Hence (ρ,ε)-optimality and $(0,\varepsilon)$-optimality in the PGMP-model are equivalent with ρ-optimality respectively gain-optimality on the MRP-model. Consequently the sets $D_{\rho,\varepsilon}$ and $D_{0,\varepsilon}$ are non-empty.

Some properties of sensitive intervention time optimal policies are considered in

THEOREM 7.2.6. *Let Δ_0 denote the set of policies, which are $(0,\varepsilon_n)$-optimal for some sequence $\{\varepsilon_n \downarrow 0, n = 1,2,\ldots\}$ depending on the policy. Then*

(i) *Δ_0 is non-empty.*

(ii) *If $|\Delta_0| > 1$ then for any pair of policies $\gamma,\delta \in \Delta_0$*

$$v_0(\varepsilon;\gamma) \equiv v_0(\varepsilon;\delta) \qquad \text{for } 0 < |\varepsilon| < \mu(\gamma,\delta) \tag{1}$$

where $\mu(\gamma,\delta)$ is some real positive number.

(iii) $D_0 = \Delta_0$.

PROOF. To prove (i) consider a sequence $\{\nu_n \downarrow 0, n = 1,2,\ldots\}$. For each n D_{0,ε_n} is nonempty. Since Z is finite, some policy is $(0,\nu_{n_k})$-optimal on a subsequence $\{\nu_{n_k} \downarrow 0, k = 1,2,\ldots\}$ which implies $\Delta_0 \neq \emptyset$. Suppose Δ_0 contains at least two policies γ and δ. Let $i \in \Psi$ be an arbitrary state. By lemma 7.2.1 $v_0(\varepsilon;z)_i$ is analytic and hence continuous on $0 < \varepsilon < \varepsilon_0$. Since $\gamma,\delta \in \Delta_0$ we have $v_0(\varepsilon;\gamma)_i \geq v_0(\varepsilon;\delta)_i$ for $\varepsilon \in \{\varepsilon_k^{(i)} \downarrow 0, k = 1,2,\ldots\}$ and $v_0(\varepsilon;\delta)_i \geq v_0(\varepsilon;\gamma)_i$ for $\varepsilon \in \{\nu_\ell^{(i)}, \ell = 1,2,\ldots\}$. Hence there exist subsequences $\{\varepsilon_{k_m}^{(i)} \downarrow 0, m = 1,2,\ldots\}$ and $\{\nu_{\ell_m}^{(i)} \downarrow 0, m = 1,2,\ldots\}$ such that $\nu_{\ell_m-1}^{(i)} < \varepsilon_{k_m}^{(i)} < \nu_{\ell_m}^{(i)} < \varepsilon_{k_m+1}^{(i)}$ with $v_0(\varepsilon;\gamma)_i \geq v_0(\varepsilon;\delta)_i$ for $\varepsilon = \varepsilon_{k_m}^{(i)}$ and $v_0(\varepsilon;\gamma)_i \leq v_0(\varepsilon;\delta)_i$ for $\varepsilon = \nu_{\ell_m}^{(i)}$ for $m = 1,2,\ldots$. By the intermediate value theorem $v_0(\varepsilon;\gamma)_i = v_0(\varepsilon;\delta)_i$ for $\varepsilon \in \{\alpha_m^{(i)} \downarrow 0, m = 1,2,\ldots\}$ where $\alpha_m^{(i)}$

satisfies $\varepsilon_{k_m}^{(i)} \le \alpha_m^{(i)} \le \nu_{\ell_m}^{(i)}$ for each m. Defining

$$g(\varepsilon)_i \overset{d}{=} [v_0(\varepsilon;\gamma)_i - v_0(\varepsilon;\delta)_i]$$

then $g(\varepsilon)_i = 0$ for $\varepsilon \in \{\alpha_m^{(i)} \downarrow 0,\ m = 1,2,\ldots\}$. Since $g(\varepsilon)_i$ is analytic for $0 \le |\varepsilon| < \mu^{(i)}(\gamma,\delta)$ this implies $g(\varepsilon)_i \equiv 0$ for $0 \le |\varepsilon| < \mu^{(i)}(\gamma,\delta)$. Consequently $v_0(\varepsilon;\gamma)_i \equiv v_0(\varepsilon;\delta)_i$ for $0 < |\varepsilon| < \mu^{(i)}(\gamma,\delta)$ and (1) holds for $\mu(\gamma,\delta) \overset{d}{=} \min_{i\in\Psi}[\mu^{(i)}(\gamma,\delta)]$.

To prove (iii) observe that definition 7.2.4 implies $\Delta_0 \supseteq D_0$. If $|\Delta_0| = 1$ then obviously either $D_0 = \Delta_0$ or $D_0 = \emptyset$. By assertion (ii) the same holds if $|\Delta_0| > 1$. However $D_0 = \emptyset$ implies that some policy $\eta \notin \Delta_0$ is $(0,\beta_n)$-optimal on some sequence $\{\beta_n \downarrow 0,\ n = 1,2,\ldots\}$ contradicting the definition of Δ_0. This completes the proof. □

To facilitate the computation of sensitive intervention time optimal policies, the following sets of policies are defined,

DEFINITION 7.2.7. Let $Z(-1,-2) \overset{d}{=} Z$. The sets of policies $Z(-1,\ell)$, $\ell = -1,0,1,\ldots$ are defined by

$$Z(-1,-1) \overset{d}{=} \{z^* \in Z:\ V_{-1,-1}(z^*) \ge V_{-1,-1}(z) \quad \text{for } z \in Z\}$$

and for $\ell = 0,1,\ldots$

$$Z(-1,\ell) \overset{d}{=} \{z^* \in Z(-1,\ell-1):\ V_{-1,\ell}(z^*) \ge V_{-1,\ell}(z) \text{ for } z \in Z(-1,\ell-1)\}.$$

Definition 7.2.7 implies immediately

(7.2;1) $\quad Z(-1,\ell-1) \supseteq Z(-1,\ell) \qquad$ for $\ell = 0,1,\ldots$

By (7.2;1) the sets $Z(-1,\ell)$, $\ell = -1,0,1,\ldots$ form a monotone sequence and we may introduce

DEFINITION 7.2.8. The set of policies $Z(-1,\infty)$ is given by

$$Z(-1,\infty) \overset{d}{=} \lim_{\ell\to\infty} Z(-1,\ell).$$

The relation between the sets D_0 and $Z(-1,\infty)$ is established in

THEOREM 7.2.9. *In the* PGMP-*model we have*

$$Z(-1,\infty) = D_0.$$

PROOF. The assertion follows from the equivalence between

(1) $\quad v_0(\varepsilon;\gamma) - v_0(\varepsilon;\delta) \geq 0 \qquad$ for $0 < \varepsilon < \mu(\gamma,\delta)$

$\mu(\gamma,\delta)$ some real positive number and

(2) $\quad [V_{-1,-1}(\gamma), V_{-1,0}(\gamma), \ldots] \succeq [V_{-1,-1}(\delta), V_{-1,0}(\delta), \ldots]$

for two polices $\gamma,\delta \in Z$. □

As a consequence of theorem 7.2.9 and (7.2;1) we have

(7.2;2) $\quad Z(-1,\ell) \neq \emptyset \qquad$ for $\ell = -1,0,1,\ldots$

7.3 ON THE COMPUTATION OF SENSITIVE INTERVENTION TIME OPTIMAL POLICIES

According to the results of the preceeding section a sensitive intervention time optimal policy in the PGMP-model can be obtained by solving a sequence of optimization problems ξ_ℓ, $\ell = -1,0,1,\ldots$ maximizing $V_{-1,\ell}(z)$ over $Z(-1,\ell-1)$. If $Z(-1,\ell)$ contains only one policy for some ℓ then this policy is sensitive intervention time optimal and only a finite number of problems ξ_ℓ have to be solved. Otherwise $Z(-1,\infty)$ contains at least two policies and the procedure is not finite. To solve problem ξ_ℓ theorem 6.4.16 is used as a starting-point.

If $z \in Z(-1,\ell-1)$ is a normal policy then theorem 6.4.16 implies that a solution $(V_{-1,\ell}(z), V_{0,\ell}(z))$ of the system of equations

$$(7.3;1) \qquad \begin{cases} V_{-1,\ell}(z) - \Gamma_0(z) V_{-1,\ell}(z)_{A(z)} = 0 \\ V_{0,\ell}(z) - \Gamma_0(z) V_{0,\ell}(z)_{A(z)} = K'_{0,\ell}(z) + \Gamma_1(z) V_{-1,\ell}(z)_{A(z)} - \overset{0}{V}_{-1,\ell-1}(z) \end{cases}$$

is unique in $V_{-1,\ell}(z)$. Observe that $\overset{0}{V}_{-1,\ell-1}(z)$ is a known and common vector for $z \in Z(-1,\ell-1)$. By corollary 6.4.15 the same solution for $V_{-1,\ell}(z)$ is obtained if (7.3;1) is replaced by

$$(7.3;2) \qquad \begin{cases} V_{-1,\ell}(z) - \Gamma_0(z) V_{-1,\ell}(z)_{A(z)} = 0 \\ V_{0,\ell}(z) - \Gamma_0(z) V_{0,\ell}(z)_{A(z)} = K''_{0,\ell}(z) - V_{-1,\ell}(z) \square t(z) \end{cases}$$

where

$$K''_{0,\ell}(z) \stackrel{d}{=} K'_{0,\ell}(z) - \overset{0}{V}_{-1,\ell-1}(z).$$

Hence for a normal policy $z \in Z(-1,\ell-1)$ the vector $V_{-1,\ell}(z)$ can be obtained by solving the system of equations in the policy evaluation operation of a GMP-scheme with $k(z)$ replaced by $K''_{0,\ell}(z)$.

In the special case $\ell = -1$ observe that if z is normal $K''_{0,\ell}(z) = 0$, implying by lemma 6.4.14 with $b(z) = q(z) = 0$ that

$$V_{-1,-1}(z) = 0 \qquad \text{for } z \text{ normal.} \tag{7.3;3}$$

Hence if Z contains only normal policies then $Z(-1,-1) = Z$ and problem ξ_{-1} can be omitted.

If Z contains a non-normal policy z then $E_\lambda(z) = E_\lambda(A(z))$ for some subchain $E_\lambda(z)$. According to theorem 6.4.16, $V_{-1,\ell}(z)_{E_\lambda(z)}$ can be uniquely solved from the system of equations

$$\begin{cases} V_{-1,\ell}(z)_{E_\lambda(z)} - \Gamma_0(z)_{E_\lambda(z)}\, V_{-1,\ell}(z)_{E_\lambda(z)} = 0 \\ V_{0,\ell+1}(z)_{E_\lambda(z)} - \Gamma_0(z)_{E_\lambda(z)} V_{0,\ell+1}(z)_{E_\lambda(z)} = K'_{0,\ell+1}(z)_{E_\lambda(z)} - V_{-1,\ell}(z)_{E_\lambda(z)} \end{cases} \tag{7.3;4}$$

The system (7.3;4) shows that $V_{-1,-1}(z)_{E_\lambda(z)} \neq 0$ in general for a non-normal subchain $E_\lambda(z)$.

Returning to the case that Z contains only normal policies, let policy $z^* \in Z(-1,\ell)$ be obtained by solving ξ_ℓ. Then $V_{-1,\ell}(z_1) = V_{-1,\ell}(z_2)$ for arbitrary $z_1, z_2 \in Z(-1,\ell)$. Denote this common vector by $V^*_{-1,\ell}$. Further define for intervention $x \in X(i)$ and state $i \in \Psi$

$$(K''_{0,\ell})^x_i \stackrel{d}{=} \begin{cases} k(i,x) & \text{for } \ell = 0 \\ G_\ell(0)^x_i/\ell! & \text{for } \ell > 0 \end{cases} \tag{7.3;5}$$

Finally define simultaneously the action sets $X^{(\ell)}(i)$ for $i \in \Psi$ and $\ell = 0,1,2,\ldots$

$$X^{(\ell)}(i) \stackrel{d}{=} \{x \in Y^{(\ell-1)}(i): \textstyle\sum_{j\in\Psi} (S_0)^x_{ij} (V^*_{-1,\ell})_j = (V^*_{-1,\ell})_i \text{ and} \tag{7.3;6}$$

$$(K''_{0,\ell})^x_i - (V^*_{-1,\ell})_i t(i,x) + \textstyle\sum_{j\in\Psi} (S_0)^x_{ij} w^1_\ell(z^*)_j = w^1_\ell(z^*)_i\},$$

where $(V^*_{-1,\ell}, w^1_\ell(z^*))$ is a particular solution of (7.3;2) for policy z^*, the set of policies for $\ell = 0,1,2,\ldots$

$$\tilde{Z}(-1,\ell) \overset{d}{=} \underset{i\in\Psi}{\times} X^{(\ell)}(i), \tag{7.3;7}$$

the set of states for $\ell = 0,1,2,\ldots$

$$\Lambda(-1,\ell) \overset{d}{=} \underset{z\in\tilde{Z}(-1,\ell)}{\cup} E(z), \tag{7.3;8}$$

and the action sets $Y^{(\ell)}(i)$ for $i \in \Psi$, $\ell = -1,0,1,\ldots$

$$\begin{cases} Y^{(-1)}(i) \overset{d}{=} X(i) \\ Y^{(\ell)}(i) \overset{d}{=} \begin{cases} X^{(\ell)}(i) & \text{for } i \in \Lambda(-1,\ell) \\ \{x \in Y^{(\ell-1)}(i) : \sum_{j\in\Psi} (S_0)_{ij} (V^*_{-1,\ell})_j = (V^*_{-1,\ell})_i\} & \text{otherwise.} \end{cases} \end{cases} \tag{7.3;9}$$

Observe that for $\ell = 0,1,2,\ldots$

$$Z(-1,\ell) = \underset{i\in\Psi}{\times} Y^{(\ell)}(i), \tag{7.3;10}$$

$$\tilde{Z}(-1,\ell) \subseteq Z(-1,\ell) \tag{7.3;11}$$

and

$$\Lambda(-1,\ell) = \underset{z\in Z(-1,\ell)}{\cup} E(z). \tag{7.3;12}$$

The necessary information for the computation of a policy $z^{**} \in Z(-1,\ell+1)$ is contained in the solution of problem ξ_ℓ which specifies a policy $z^* \in Z(-1,\ell)$ and the vectors $V^*_{-1,\ell}$ and $w^1_\ell(z^*)$. The following method specifies the computation.

METHOD 7.3.1. Given the solution of problem ξ_ℓ, a policy $z^{**} \in Z(-1,\ell+1)$ is obtained in the following two steps

(i) Compute the sets $E(z)$ for $z \in \tilde{Z}(-1,\ell)$ and their union $\Lambda(-1,\ell)$. For this purpose a method developed in [FOX & LANDI 1968] can be used, which identifies the subchains of a stochastic matrix and hence $E(z)$ per policy.

(ii) Solve problem $\xi_{\ell+1}$ by applying one of the methods GMP 1,2,3 with the terminating policy z^* for problem ξ_ℓ as initial policy. Problem $\xi_{\ell+1}$

has the sets $Y^{(\ell)}(i)$, $i \in \Psi$ as action sets. Moreover the numbers $(K''_{0,\ell+1})^x_i$ specified by (7.3;5) replace the numbers $k(i,x)$ for $x \in Y^{(\ell)}(i)$ and $i \in \Psi$. This yields policy z^{**}.

The computation of a sensitive intervention time optimal policy is now illustrated on the numerical example of section 6.5. Primarily $V_{-1,0}(z)$ is obtained for $z \in Z$ by the policy evaluation operation of GMP 1. It is easily verified that

$$t_0 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad t(z_\ell)_1 = \begin{cases} \frac{1}{2} & \text{for } \ell = 1,3 \\ 1 & \text{for } \ell = 2,4 \end{cases}$$

$V_{-1,0}(z)$ is then obtained from the system of equations

$$\begin{cases} 0 = k(z)_1 - V_{-1,0}(z)_1 \; t(z)_1 \\ V_{-1,0}(z)_{\overline{A(z)}} = U_0(\overline{A(z)}) \; V_{-1,0}(z)_1 \end{cases}$$

which results in $V_{-1,0}(z) = V^*_{-1,0} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}$ for each $z \in Z$. Next problem ξ_1 is solved with

$$(K''_{0,1})^x_1 = G_1(0)^x_1 = \begin{cases} 1 & \text{for } x = 2 \\ 0 & \text{otherwise} \end{cases}$$

and the action sets

$$Y^{(0)}(i) = \begin{cases} \{1,2,3,4\} & \text{for } i = 1 \\ \{x_0(i)\} & \text{otherwise.} \end{cases}$$

Since in this problem $\Lambda(-1,0) = \Psi$, step (ii) of method 7.3.1 is sufficient for this computation. Taking z_1 as initial policy, $V_{-1,1}(z_1)$ is obtained from

$$\begin{cases} 0 = K'_{0,1}(z_1)_1 - V_{-1,0}(z_1)_1 - V_{-1,1}(z_1)_1 \; t(z)_1 \\ V_{-1,1}(z_1)_{\overline{A(z_1)}} = U_0(\overline{A(z_1)}) \; V_{-1,1}(z_1)_1 \end{cases}$$

Since $K'_{0,1}(z_1)_1 = G_1(0;z_1)_1 = 0$ we obtain

$$V_{-1,1}(z_1) = \begin{bmatrix} -8 \\ -8 \\ -8 \end{bmatrix}.$$

In view of the fact that each policy has only one subchain the policy improvement operation can be restricted to the computation of $\hat{w}_i$. This yields

$$\hat{w}_i = \sum_{j\in\phi} (N_0)_{ij} w_1^1(z)_j = 0 \qquad \text{for } i = 2,3$$

and

$$\hat{w}_1 = \max_{x\in X_1(1)} [G_1(0)_1^x - V_{-1,1}(z)_1 t(1,x) + \sum_{j\in\Psi} (S_0)_{1j}^x w_1^1(z)_j]$$

$$= \max_{x\in X_1(1)} [G_1(0)_1^x + 8\, t(1,x)] = \max[4,9,4,8] = 9.$$

The maximum is obtained for $x = 2$. Since $A(z) = A_0$ for $z \in Z$, no cutting operation is needed. Hence the next policy in the iteration is z_2. In the same way we obtain

$$V_{-1,1}(z_2) = \begin{bmatrix} -3 \\ -3 \\ -3 \end{bmatrix}$$

and

$$\hat{w}_i = \begin{cases} \sum_{j\in\Psi} (N_0)_{ij} w_1^1(z)_j = 0 & \text{for } i = 2,3 \\ \max_{x\in X_1(i)} [G_1(0)_i^x + 3\, t(i,x)] = \max[\frac{3}{2},4,\frac{3}{2},3] = 4 & \text{for } i = 1 \end{cases}$$

implying that policy $z_2 \in Z(-1,1)$ is sensitive intervention time optimal in agreement with the result of table 6.5.2.

7.4 ON THE COMPUTATION OF BIAS-OPTIMAL POLICIES

The Laurent expansion of theorem 6.4.16 can also be used to compute sensitive discount optimal policies. In this section only the computation of bias-optimal policies is considered. The question is whether an unaltered use of GMP-schemes for this computation remains possible.

In this section we assume that Z contains only normal policies. This implies $V_{-1,-1}(z) = 0$ for $z \in Z$ (cf. (7.3;3)). Further we substitute $\varepsilon = 0$ in the partial Laurent expansion for the PGMP-model of theorem 6.4.16 which becomes then (1) of remark 6.4.17. As a consequence we have in fact returned

to the finite GMP-model but retain the notation and use the results of the PGMP-model. The coefficient vectors $V_{h,0}(z)$, $h = 1,0,1,\ldots$ of this expansion are uniquely determined by the system of equations consisting of

$$(7.4;1) \qquad V_{h,0}(z) - \Gamma_0(z)V_{h,0}(z)_{A(z)} = K_{h,0}(z) + \sum_{k=1}^{h+1} \Gamma_k(z)V_{h-k,0}(z)_{A(z)}$$

and the corresponding equation in $V_{h+1,0}(z)$.

The computation of bias-optimal policies involves only the coefficient vectors $V_{-1,0}(z)$ and $V_{0,0}(z)$. Each particular solution $(y_{-1}(z),w_0(z))$ of the system of equations

$$(7.4;2) \qquad \begin{cases} (a)\ y_{-1}(z) - \Gamma_0(z)\ y_{-1}(z)_{A(z)} = 0 \\ (b)\ w_0(z) - \Gamma_0(z)\ w_0(z)_{A(z)} = k(z) + \Gamma_1(z)y_{-1}(z)_{A(z)} \end{cases}$$

has the property that $y_{-1}(z) = V_{-1,0}(z)$.

Consider the system of equations in $(V_{0,0}(z),V_{1,0}(z))$ given by

$$(7.4;3) \qquad \begin{cases} (a)\ V_{0,0}(z) - \Gamma_0(z)V_{0,0}(z)_{A(z)} = k(z)+\Gamma_1(z)V_{-1,0}(z)_{A(z)} \\ (b)\ V_{1,0}(z)-\Gamma_0(z)V_{1,0}(z)_{A(z)} = K_{1,0}(z)+\Gamma_1(z)V_{0,0}(z)_{A(z)}+\Gamma_2(z)V_{-1,0}(z)_{A(z)} \end{cases}$$

For a given particular solution $(V_{-1,0}(z),\ w_0(z))$ of (7.4;2) let $y_0(z) \stackrel{d}{=} V_{0,0}(z)-w_0(z)$. By subtraction of (7.4;2) (b) from (7.4;3) (a) it follows that $y_0(z)$ can be uniquely obtained from the system of equations in $(y_0(z),w_1(z))$ given by

$$(7.4;4) \qquad \begin{cases} (a)\ y_0(z) - \Gamma_0(z)y_0(z)_{A(z)} = 0 \\ (b)\ w_1(z) - \Gamma_0(z)w_1(z)_{A(z)} = K^*_{1,0}(z) + \Gamma_1(z)y_0(z) \end{cases}$$

where

$$(7.4;5) \qquad K^*_{1,0}(z) \stackrel{d}{=} K_{1,0}(z) + \Gamma_1(z)w_0(z)_{A(z)} + \Gamma_2(z)V_{-1,0}(z)_{A(z)}.$$

Observe the similarity between the systems (7.4;2) and (7.4;4). The following lemma exhibits the relationship between the solution of (7.4;2) and the one of the corresponding policy evaluation operation of a GMP-scheme.

LEMMA 7.4.1. *Let the system of equations in* $(y_{-1}(z),w_0(z))$ *be given by* (7.4;2) *and one in* $(y^1_{-1}(z),w^1_0(z))$ *by*

$$(1)\qquad \begin{cases} (a)\ y^1_{-1}(z) - \Gamma_0(z)y^1_{-1}(z)_{A(z)} = 0 \\ (b)\ w^1_0(z) - \Gamma_0(z)w^1_0(z)_{A(z)} = k(z) - y^1_{-1}(z) \square t(z). \end{cases}$$

Then corresponding particular solutions satisfy

$$(2)\qquad y^1_{-1}(z)_i = y_{-1}(z)_i \qquad \text{for } i \in \Psi$$

$$(3)\qquad w^1_0(z)_i = w_0(z)_i + y^1_{-1}(z)_i(t_0)_i \qquad \text{for } i \in E(z).$$

PROOF. Corollary 6.4.15 implies $y^1_{-1}(z)_i = y_{-1}(z)_i$ for $i \in \Psi$. It remains to prove (3). This is done for an arbitrary subchain $E_\lambda(z)$. For $(t_0)_{E_\lambda(\overline{A(z)})}$ the following relation holds (cf. lemma 2.3.7 (ii))

$$(4)\qquad (t_0)_{E_\lambda(\overline{A(z)})} = U_0(A(z))_{E_\lambda(\overline{A(z)})E_\lambda(A(z))}(t_0)_{E_\lambda(A(z))} + \\ + R_0(\overline{A(z)})_{E_\lambda(\overline{A(z)})}\tau_{E_\lambda(\overline{A(z)})}$$

Hence

$$(5)\qquad (t_0)_{E_\lambda(\overline{A(z)})} \square y_{-1}(z)_{E_\lambda(\overline{A(z)})} + \\ - U_0(\overline{A(z)})_{E_\lambda(\overline{A(z)})E_\lambda(A(z))}(t_0)_{E_\lambda(A(z))} \square y_{-1}(z)_{E_\lambda(A(z))} = \\ = R_0(\overline{A(z)})_{E_\lambda(\overline{A(z)})}\tau_{E_\lambda(\overline{A(z)})} \square y_{-1}(z)_{E_\lambda(\overline{A(z)})}.$$

Addition of (5) and the $E_\lambda(\overline{A(z)})$-part of (7.4;2) (b) yields

$$(6)\qquad [w_0(z) + t_0 \square y_{-1}(z)]_{E_\lambda(\overline{A(z)})} + \\ - U_0(\overline{A(z)})_{E_\lambda(\overline{A(z)})E_\lambda(A(z))}[w_0(z) + t_0 \square y_{-1}(z)]_{E_\lambda(A(z))} = \\ = U_1(\overline{A(z)})_{E_\lambda(\overline{A(z)})E_\lambda(A(z))}y_{-1}(z)_{E_\lambda(A(z))} + \\ + R_0(\overline{A(z)})_{E_\lambda(\overline{A(z)})}\tau_{E_\lambda(\overline{A(z)})} \square y_{-1}(z)_{E_\lambda(\overline{A(z)})} = 0.$$

Further by lemma 2.3.12 (i), which may also be applied here,

(7) $$P_1(E_\lambda(A(z))y_{-1}(z)_{E_\lambda(A(z))} =$$

$$= - P(z)_{E_\lambda(A(z))E_\lambda(\overline{A(z)})}R_0(\overline{A(z)})_{E_\lambda(\overline{A(z)})}{}^{\tau}E_\lambda(\overline{A(z)}) \square\, y_{-1}(z)_{E_\lambda(\overline{A(z)})} =$$

$$= [t(z)_{E_\lambda(A(z))} - [I-P_0(E_\lambda(A(z)))](t_0)_{E_\lambda(A(z))}] \square\, y_{-1}(z)_{E_\lambda(A(z))} =$$

$$= [t \quad \square\, y_{-1}(z)]_{E_\lambda(A(z))} +$$

$$- [I-P_0(E_\lambda(A(z)))]_{E_\lambda(A(z))}(t_0)_{E_\lambda(A(z))} \square\, y_{-1}(z)_{E_\lambda(A(z))}.$$

Substitution of (7) in (7.4;2) (b) yields

(8) $$[w_0(z) + t_0 \square\, y_{-1}(z)]_{E_\lambda(A(z))} +$$

$$- P_0(E_\lambda(A(z)))[w_0(z) + t_0 \square\, y_{-1}(z)]_{E_\lambda(A(z))} =$$

$$= [k(z) - y_{-1}(z) \square\, t(z)]_{E_\lambda(A(z))}$$

(6) and (8) imply the assertion. □

Suppose a gain optimal policy $z^* \in Z(-1,0)$ is obtained by using one of the methods GMP 1,2 or 3. In section 7.3 this optimization problem has been denoted by ξ_0. In addition, the solution of ξ_0 specifies the vectors $V^*_{-1,0}$ and $w^1_0(z^*)$. Based on these data we define for each $x \in X^{(0)}(i)$, $i \in \Psi$

$$(K^*_{1,0})^x_i \triangleq \begin{cases} - \sum_{j\in\Psi}(N_1)_{ij}w^1_0(z^*)_j + \sum_{j\in\Psi}(N_2)_{ij}(V^*_{-1,0})_j & \text{for } x = x_0(i) \\ \sum_{j\in\Psi}P^x_{ij}\, K_1(\overline{A_0})_j - K_1(\overline{A_0})_i & \text{otherwise.} \end{cases}$$

Notice that $(K^*_{1,0})^{z(i)}_i \neq K^*_{1,0}(z)_i$ in this notation. The next step to be performed is concerned with a second optimization problem denoted by ξ^*. Problem ξ^* is solved by means of one of the methods GMP 1,2 or 3. It is obtained by replacing $k(i,x)$ by $(K^*_{1,0})^x_i$ and $X(i)$ by $X^{(0)}(i)$ in the setting of problem ξ_0. The policy evaluation operation of the solution methods computes a particular solution $(y_0(z), w^1_1(z))$ to the system of equations

$$\begin{cases} y_0(z) - \Gamma_0(z)y_0(z)_{A(z)} = 0 \\ w_1^1(z) - \Gamma_0(z)w_1^1(z)_{A(z)} = K_{1,0}^*(z) - y_0(z) \,\square\, t(z) \end{cases}$$

where $K_{1,0}^*(z)$ is defined by (7.4;5). In the policy improvement operation $k(i,x)$ is replaced by $(K_{1,0}^*)_i^x$.

The solution of ξ^* specifies a policy $\hat{z}$ which is gain-optimal for $i \in \Psi$ and bias-optimal for $i \in \Lambda(-1,0)$. An improvement upon policy $\hat{z}$ can only be obtained in the states that are transient for all policies $z \in Z(-1,0)$, which may be done by returning to problem ξ_0 with initial policy $\hat{z}$ and policy evaluation (7.4;2) where for $w_0(\hat{z})$ is substituted $y_0(\hat{z}) + w_0(z^*)$. The method proposed to compute a bias-optimal policy in the PGMP-model is summarized in

METHOD 7.4.2. A bias-optimal policy z^{**} is obtained by the following computational steps.

(i) Problem ξ_0 is solved by applying GMP 1,2, or 3, specifying the vectors $V_{-1,0}^*$, $w_0^1(z^*)$ and policy $z^* \in Z(-1,0)$.

(ii) Compute the set of actions $X^{(0)}(i)$.

(iii) Compute $w_0(z^*)_i$ for $i \in E(z^*)$ by using the relation $w_0(z^*)_i = w_0^1(z^*)_i + - (V_{-1,0}^*)_i(t_0)_i$ and for $i \in F(z^*)$ by using (7.4;2).

(iv) Solve problem ξ^* by applying GMP 1,2 or 3 with initial policy z^*. This results in policy $\hat{z}$ and vector $y_0(\hat{z})$. Compute the vector $w_0(\hat{z}) \stackrel{d}{=} y_0(\hat{z}) + w_0(z^*)$.

(v) Return to problem ξ_0 and use GMP 1,2 or 3 with policy evaluation operation (7.4;2) to improve policy $\hat{z}$. This yields policy z^{**}.

REMARK 7.4.3.

(i) Observe that there is no need to compute the set $\Lambda(-1,0)$ in this procedure.

(ii) The system of equations (7.4;2) is used in the steps (iii) and (v). In either case $U_1(\overline{A(z)})y_{-1}(z)_{A(z)}$ may be replaced by $-R_0(\overline{A(z)})(N_1)_{\overline{A(z)}\Psi}y_{-1}(z)$ (cf. (10) of remark 6.4.17) if this is computationally convenient.

(iii) The computation of $K_{1,0}^*(z)$ may be simplified by using the relation

$$U_1(\overline{A(z)})w_0(z)_{A(z)} + U_2(\overline{A(z)})V_{-1,0}(z)_{A(z)} =$$

$$= R_0(\overline{A(z)}) \left[-(N_1)_{\overline{A(z)}\Psi}w_0(z) + (N_2)_{\overline{A(z)}\Psi}V_{-1,0}(z)\right].$$

The computation of a bias-optimal policy by means of method 7.4.2 is now illustrated on the numerical example of section 6.5. Assuming that step 1 of method 7.4.2 terminates with policy z_1. Then

$$V^*_{-1,0} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}; \quad w^1_0(z_1) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}; \quad Z(-1,0) = \{z_1, z_2, z_3, z_4\}$$

In step 2 the action set $X^{(0)}(i)$ is determined which is given here by

$$X(i) = \begin{cases} \{1,2,3,4\} & \text{for } i = 1 \\ \{x_0(i)\} & \text{otherwise.} \end{cases}$$

Since

$$N_1 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} \end{bmatrix} \qquad N_2 = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{8} & 0 & \frac{1}{8} \end{bmatrix} .$$

It follows from (7.4;6) that

$$(K^*_{1,0})^x_i = \begin{cases} -1 & \text{for } i = 1 \text{ and } x = 1,3 \\ -2 & \text{for } i = 1 \text{ and } x = 2,4 \\ 2 & \text{for } i = 2 \text{ and } x = x_0(i) \\ 1 & \text{for } i = 3 \text{ and } x = x_0(i) \end{cases}$$

In step 3 $w_0(z_1)$ is computed from $w^1_0(z_1)$. By lemma 7.4.1 it follows that

$$w_0(z_1)_i = w^1_0(z_1)_i - y_{-1}(z_1)_i \cdot (t_0)_i \quad \text{for } i = 1,2.$$

Since state 3 is transient for this policy, $w_0(z_1)_3$ is obtained from

$$w_0(z_1)_3 = U_0(\overline{A(z)})_{31}\, w_0(z_1)_1 + U_1(\overline{A(z)})_{31}\, y_{-1}(z)_1 .$$

This yields

$$w_0(z_1) = \begin{bmatrix} 0 \\ -4 \\ -4 \end{bmatrix} .$$

Next step 4 is applied with initial policy z_1. The vector $y_0(z_1)$ is obtained from the equations

$$\begin{cases} K^*_{1,0}(z_1)_1 - y_0(z_1)_1 t(z_1)_1 = 0 \\ y_0(z_1)_{\overline{A(z_1)}} = U_0(\overline{A(z_1)})y_0(z_1)_1 \end{cases}$$

and the vector $w^1_1(z_1)$ from

$$\begin{cases} w^1_1(z_1)_1 \overset{d}{=} 0 \\ w^1_1(z_1)_{\overline{A(z)}} = U_0(\overline{A(z)})w^1_1(z_1)_1 + K^*_{1,0}(z_1)_{\overline{A(z)}} \end{cases}$$

This yields, noting that $K^*_{1,0}(z) = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$,

$$y_0(z_1) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \; ; \qquad w^1_1(z_1) = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$$

Observe that indeed $V_{0,0}(z_1) = y_0(z_1) + w_0(z_1)$, (cf. table 6.5).

Applying the policy improvement operation we obtain for $\hat{w}_1$

$$\hat{w}_1 = \max_{x=1,2,3,4} [(K^*_{1,0})^x_1 - y_0(z_1)t(1,x) + \textstyle\sum_{j\in\Psi} (S_0)^x_{1j} w^1_1(z_1)_j]$$

$$= \max[-1+\tfrac{1}{2}.2,\ -2+1.2,\ -1+\tfrac{1}{2}.3,\ -2+1.3]$$

$$= \max[0, 0, \tfrac{1}{2}, 1] = 1.$$

Since only the null decision is feasible in state 2 and 3 this shows that policy z_4 is the next policy. By the policy evaluation operation we obtain

$$y_0(z_4) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \; ; \qquad w^1_1(z_4) = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix} .$$

Applying the policy improvement operation yields for $\hat{w}_1$

$$\hat{w}_1 = \max_{x \in X(1)} [(K^*_{1,0})^x_1 - y_0(z_4)_1 t(1,x) + \sum_{j \in \Psi} (S_0)^x_{1j} w^1_1(z_4)_j] =$$

$$= \max[-1-\tfrac{1}{2}+\tfrac{1}{2}.2,\ -2-1+1.2,\ -1-\tfrac{1}{2}+\tfrac{1}{2}.3,\ -2-1+1.3] =$$

$$= \max[-\tfrac{1}{2}, -1, 0, 0] = 0.$$

This implies that z_3 and z_4 are both bias-optimal in agreement with table 6.5.3.

REFERENCES

BLACKWELL, D., *Discrete dynamic programming,* Ann. Math. Stat. 33, 719-726 (1962).

BREIMAN, L., *Stopping rule problems,* in E.F. BECKENBACH (ed.), Applied combinatorial mathematics, J. Wiley & Sons, Inc., New York (1964).

CHUNG, K.L., *A course in probability theory,* Harcourt, Brace and World, New York, 2nd edition (1974).

CINLAR, E., *Markov renewal theory,* Adv. Appl. Prob. 1, 123-187 (1969).

CINLAR, E., *Markov renewal theory: a survey,* Man. Sc. 21, 727-752 (1975).

DECANI, J.S., *A dynamic programming algorithm for embedded Markov chains when the planning horizon is at infinity,* Man. Sc. 10, 716-733 (1964).

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

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

DE LEVE, G., A. FEDERGRUEN & H.C. TIJMS, *A general Markovian decision method I: Model and techniques, A general Markov decision method II: Applications,* Adv. Appl. Prob. 9, 296-335 (1977).

DE LEVE, G. & P.J. WEEDA, *Driving with Markov programming,* ASTIN Bull. 5, 62-86 (1968).

DENARDO, E.V., *Contraction mappings in the theory underlying dynamic programming,* SIAM Rev. 9, 165-177 (1967).

DENARDO, E.V., *Computing bias-optimal policies in a discrete-time Markov decision problem,* Oper. Res. 18, 279-289 (1970).

DENARDO, E.V., *Markov renewal programs with small interest rates,* Ann. Math. Stat. 42, 477-496 (1971).

DENARDO, E.V. & B. FOX, *Multichain Markov renewal programs,* SIAM J. Appl. Math. 16, 468-487 (1968).

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

DOOB, J., *Stochastic Processes,* Wiley, New York (1953).

DYNKIN, E.B. & A.A. JUSCHKEWITSCH, *Sätze und Aufgaben über Markoffsche Prozesse,* Springer-Verlag, Berlin (1969).

EMRICH, O., *Algorithmen zur Bestimmung optimaler Stoppregeln bei endlichen Markoff-Ketten,* Diss. Univ. Karlsruhe (1970).

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

FOX, B.L., *(g,w)-optimal in Markov renewal programs,* Man. Sci. 15, 210-212 (1968).

FOX, B.L. & D.M. LANDI, *An algorithm for identifying the ergodic subchains and transient states of a stochastic matrix,* Comm. ACM 11, 619-621 (1968).

HORDIJK, A., *Dynamic programming and Markov potential theory,* Mathematical Centre Tract No. 51, Mathematisch Centrum, Amsterdam (1974).

HORDIJK, A., R. POTHARST & J.TH. RUNNENBURG, *Optimaal stoppen van Markov-ketens,* MC Syllabus 19, Mathematisch Centrum, Amsterdam (1973).

HORDIJK, A. & K. SLADKÝ, *Sensitive optimality criteria in countable state dynamic programming,* Math. of O.R. 2, 1-14 (1977).

HOWARD, R.A., *Dynamic programming and Markov processes,* Technology Press of M.I.T., Cambridge, Mass. (1960).

HOWARD, R.A., *Semi-Markovian decision processes,* Proc. Int. Stat. Inst., Ottawa (1963).

JEWELL, W.S., *Markov-renewal programming I: formulation, finite return models;Markov-renewal programming II: infinite return models, example,* Oper. Res. 11, 938-971 (1963).

KATO, T., *Perturbation theory for linear operators,* Springer-Verlag, New York (1966).

KEMENY, J.G. & J.L. SNELL, *Finite Markov chains,* Van Nostrand, Princeton (1961).

MILLER, B.L. & A.F. VEINOTT, *Discrete dynamic programming with a small interest rate,* Ann. Math. Stat. 40, 366-370 (1969).

ROTHBLUM, U.G., *Multiplicative Markov decision chains,* Ph.D. Dissertation Dept. Of Oper. Res., Stanford University, Stanford, Ca. 94305 (1974).

SCHWEITZER, P.J., *Perturbation theory and Markovian decision processes,* Ph.D. Dissertation, MIT (1965).

SCHWEITZER, P.J., *Perturbation theory and undiscounted Markov renewal programming,* Oper. Res. 17, 716-727 (1969).

SLADKÝ, K., *On the set of optimal controls for Markov chains with rewards,* Kybernetika 10, 350-367 (1974).

TIJMS, H.C., *Analysis of (s,S) inventory models,* Mathematical Centre Tract No. 40, Mathematisch Centrum, Amsterdam (1972).

VEINOTT, A.F., *Discrete dynamic programming with sensitive discount optimality criteria,* Ann. Math. Stat. 40, 1635-1660 (1969).

WEEDA, P.J., *On the relationship between the cutting operation of generalized Markov programming and optimal stopping,* Report BW 36/74, Mathematisch Centrum, Amsterdam (1974).

WEEDA, P.J., *Some computational experiments with a special generalized Markov programming model,* Report BW 37/74, Mathematisch Centrum, Amsterdam (1974).

WEEDA, P.J., *Technical aspects of the iterative solution of the automobile insurance problem,* Report BN 27/75, Mathematisch Centrum, Amsterdam (1975).

WEEDA, P.J., *Sensitive time and discount optimality in Markov renewal decision problems with instantaneous actions,* Report BW 59/76, Mathematisch Centrum, Amsterdam (1976).

OTHER TITLES IN THE SERIES MATHEMATICAL CENTRE TRACTS

A leaflet containing an order-form and abstracts of all publications mentioned below is available at the Mathematisch Centrum, Tweede Boerhaavestraat 49, Amsterdam-1005, The Netherlands. Orders should be sent to the same address.

MCT 1 T. VAN DER WALT, *Fixed and almost fixed points*, 1963. ISBN 90 6196 002 9.

MCT 2 A.R. BLOEMENA, *Sampling from a graph*, 1964. ISBN 90 6196 003 7.

MCT 3 G. DE LEVE, *Generalized Markovian decision processes, part I: Model and method*, 1964. ISBN 90 6196 004 5.

MCT 4 G. DE LEVE, *Generalized Markovian decision processes, part II: Probabilistic background*, 1964. ISBN 90 6196 005 3.

MCT 5 G. DE LEVE, H.C. TIJMS & P.J. WEEDA, *Generalized Markovian decision processes, Applications*, 1970. ISBN 90 6196 051 7.

MCT 6 M.A. MAURICE, *Compact ordered spaces*, 1964. ISBN 90 6196 006 1.

MCT 7 W.R. VAN ZWET, *Convex transformations of random variables*, 1964. ISBN 90 6196 007 X.

MCT 8 J.A. ZONNEVELD, *Automatic numerical integration*, 1964. ISBN 90 6196 008 8.

MCT 9 P.C. BAAYEN, *Universal morphisms*, 1964. ISBN 90 6196 009 6.

MCT 10 E.M. DE JAGER, *Applications of distributions in mathematical physics*, 1964. ISBN 90 6196 010 X.

MCT 11 A.B. PAALMAN-DE MIRANDA, *Topological semigroups*, 1964. ISBN 90 6196 011 8.

MCT 12 J.A.TH.M. VAN BERCKEL, H. BRANDT CORSTIUS, R.J. MOKKEN & A. VAN WIJNGAARDEN, *Formal properties of newspaper Dutch*, 1965. ISBN 90 6196 013 4.

MCT 13 H.A. LAUWERIER, *Asymptotic expansions*, 1966, out of print; replaced by MCT 54 and 67.

MCT 14 H.A. LAUWERIER, *Calculus of variations in mathematical physics*, 1966. ISBN 90 6196 020 7.

MCT 15 R. DOORNBOS, *Slippage tests*, 1966. ISBN 90 6196 021 5.

MCT 16 J.W. DE BAKKER, *Formal definition of programming languages with an application to the definition of ALGOL 60*, 1967. ISBN 90 6196 022 3.

MCT 17 R.P. VAN DE RIET, *Formula manipulation in ALGOL 60, part 1*, 1968. ISBN 90 6196 025 8.

MCT 18 R.P. VAN DE RIET, *Formula manipulation in ALGOL 60, part 2*, 1968. ISBN 90 6196 038 X.

MCT 19 J. VAN DER SLOT, *Some properties related to compactness*, 1968. ISBN 90 6196 026 6.

MCT 20 P.J. VAN DER HOUWEN, *Finite difference methods for solving partial differential equations*, 1968. ISBN 90 6196 027 4.

MCT 21 E. WATTEL, *The compactness operator in set theory and topology*, 1968. ISBN 90 6196 028 2.

MCT 22 T.J. DEKKER, *ALGOL 60 procedures in numerical algebra, part 1*, 1968. ISBN 90 6196 029 0.

MCT 23 T.J. DEKKER & W. HOFFMANN, *ALGOL 60 procedures in numerical algebra, part 2*, 1968. ISBN 90 6196 030 4.

MCT 24 J.W. DE BAKKER, *Recursive procedures*, 1971. ISBN 90 6196 060 6.

MCT 25 E.R. PAERL, *Representations of the Lorentz group and projective geometry*, 1969. ISBN 90 6196 039 8.

MCT 26 EUROPEAN MEETING 1968, *Selected statistical papers, part I*, 1968. ISBN 90 6196 031 2.

MCT 27 EUROPEAN MEETING 1968, *Selected statistical papers, part II*, 1969. ISBN 90 6196 040 1.

MCT 28 J. OOSTERHOFF, *Combination of one-sided statistical tests*, 1969. ISBN 90 6196 041 X.

MCT 29 J. VERHOEFF, *Error detecting decimal codes*, 1969. ISBN 90 6196 042 8.

MCT 30 H. BRANDT CORSTIUS, *Excercises in computational linguistics*, 1970. ISBN 90 6196 052 5.

MCT 31 W. MOLENAAR, *Approximations to the Poisson, binomial and hypergeometric distribution functions*, 1970. ISBN 90 6196 053 3.

MCT 32 L. DE HAAN, *On regular variation and its application to the weak convergence of sample extremes*, 1970. ISBN 90 6196 054 1.

MCT 33 F.W. STEUTEL, *Preservation of infinite divisibility under mixing and related topics*, 1970. ISBN 90 6196 061 4.

MCT 34 I. JUHÁSZ, A. VERBEEK & N.S. KROONENBERG, *Cardinal functions in topology*, 1971. ISBN 90 6196 062 2.

MCT 35 M.H. VAN EMDEN, *An analysis of complexity*, 1971. ISBN 90 6196 063 0.

MCT 36 J. GRASMAN, *On the birth of boundary layers*, 1971. ISBN 90 6196 064 9.

MCT 37 J.W. DE BAKKER, G.A. BLAAUW, A.J.W. DUIJVESTIJN, E.W. DIJKSTRA, P.J. VAN DER HOUWEN, G.A.M. KAMSTEEG-KEMPER, F.E.J. KRUSEMAN ARETZ, W.L. VAN DER POEL, J.P. SCHAAP-KRUSEMAN, M.V. WILKES & G. ZOUTENDIJK, *MC-25 Informatica Symposium*, 1971. ISBN 90 6196 065 7.

MCT 38 W.A. VERLOREN VAN THEMAAT, *Automatic analysis of Dutch compound words*, 1971. ISBN 90 6196 073 8.

MCT 39 H. BAVINCK, *Jacobi series and approximation*, 1972. ISBN 90 6196 074 6.

MCT 40 H.C. TIJMS, *Analysis of (s,S) inventory models*, 1972. ISBN 90 6196 075 4.

MCT 41 A. VERBEEK, *Superextensions of topological spaces*, 1972. ISBN 90 6196 076 2.

MCT 42 W. VERVAAT, *Success epochs in Bernoulli trials (with applications in number theory)*, 1972. ISBN 90 6196 077 0.

MCT 43 F.H. RUYMGAART, *Asymptotic theory of rank tests for independence*, 1973. ISBN 90 6196 081 9.

MCT 44 H. BART, *Meromorphic operator valued functions*, 1973. ISBN 90 6196 082 7.

MCT 45 A.A. BALKEMA, *Monotone transformations and limit laws*, 1973. ISBN 90 6196 083 5.

MCT 46 R.P. VAN DE RIET, *ABC ALGOL, A portable language for formula manipulation systems, part 1: The language*, 1973. ISBN 90 6196 084 3.

MCT 47 R.P. VAN DE RIET, *ABC ALGOL, A portable language for formula manipulation systems, part 2: The compiler*, 1973. ISBN 90 6196 085 1.

MCT 48 F.E.J. KRUSEMAN ARETZ, P.J.W. TEN HAGEN & H.L. OUDSHOORN, *An ALGOL 60 compiler in ALGOL 60, Text of the MC-compiler for the EL-X8*, 1973. ISBN 90 6196 086 X.

MCT 49 H. KOK, *Connected orderable spaces*, 1974. ISBN 90 6196 088 6.

MCT 50 A. VAN WIJNGAARDEN, B.J. MAILLOUX, J.E.L. PECK, C.H.A. KOSTER, M. SINTZOFF, C.H. LINDSEY, L.G.L.T. MEERTENS & R.G. FISKER (Eds), *Revised report on the algorithmic language ALGOL 68*, 1976. ISBN 90 6196 089 4.

MCT 51 A. HORDIJK, *Dynamic programming and Markov potential theory*, 1974. ISBN 90 6196 095 9.

MCT 52 P.C. BAAYEN (ed.), *Topological structures*, 1974. ISBN 90 6196 096 7.

MCT 53 M.J. FABER, *Metrizability in generalized ordered spaces*, 1974. ISBN 90 6196 097 5.

MCT 54 H.A. LAUWERIER, *Asymptotic analysis, part 1*, 1974. ISBN 90 6196 098 3.

MCT 55 M. HALL JR. & J.H. VAN LINT (Eds), *Combinatorics, part 1: Theory of designs, finite geometry and coding theory*, 1974. ISBN 90 6196 099 1.

MCT 56 M. HALL JR. & J.H. VAN LINT (Eds), *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry*, 1974. ISBN 90 6196 100 9.

MCT 57 M. HALL JR. & J.H. VAN LINT (Eds), *Combinatorics, part 3: Combinatorial group theory*, 1974. ISBN 90 6196 101 7.

MCT 58 W. ALBERS, *Asymptotic expansions and the deficiency concept in statistics*, 1975. ISBN 90 6196 102 5.

MCT 59 J.L. MIJNHEER, *Sample path properties of stable processes*, 1975. ISBN 90 6196 107 6.

MCT 60 F. GÖBEL, *Queueing models involving buffers*, 1975. ISBN 90 6196 108 4.

* MCT 61 P. VAN EMDE BOAS, *Abstract resource-bound classes, part 1*. ISBN 90 6196 109 2.

* MCT 62 P. VAN EMDE BOAS, *Abstract resource-bound classes, part 2*. ISBN 90 6196 110 6.

MCT 63 J.W. DE BAKKER (ed.), *Foundations of computer science*, 1975. ISBN 90 6196 111 4.

MCT 64 W.J. DE SCHIPPER, *Symmetric closed categories*, 1975. ISBN 90 6196 112 2.

MCT 65 J. DE VRIES, *Topological transformation groups 1 A categorical approach*, 1975. ISBN 90 6196 113 0.

MCT 66 H.G.J. PIJLS, *Locally convex algebras in spectral theory and eigenfunction expansions*, 1976. ISBN 90 6196 114 9.

* MCT 67 H.A. LAUWERIER, *Asymptotic analysis, part 2.*
ISBN 90 6196 119 X.

MCT 68 P.P.N. DE GROEN, *Singularly perturbed differential operators of second order*, 1976. ISBN 90 6196 120 3.

MCT 69 J.K. LENSTRA, *Sequencing by enumerative methods*, 1977.
ISBN 90 6196 125 4.

MCT 70 W.P. DE ROEVER JR., *Recursive program schemes: semantics and proof theory*, 1976. ISBN 90 6196 127 0.

MCT 71 J.A.E.E. VAN NUNEN, *Contracting Markov decision processes*, 1976.
ISBN 90 6196 129 7.

MCT 72 J.K.M. JANSEN, *Simple periodic and nonperiodic Lamé functions and their applications in the theory of conical waveguides*, 1977.
ISBN 90 6196 130 0.

MCT 73 D.M.R. LEIVANT, *Absoluteness of intuitionistic logic*, 1979.
ISBN 90 6196 122 x.

MCT 74 H.J.J. TE RIELE, *A theoretical and computational study of generalized aliquot sequences*, 1976. ISBN 90 6196 131 9.

MCT 75 A.E. BROUWER, *Treelike spaces and related connected topological spaces*, 1977. ISBN 90 6196 132 7.

MCT 76 M. REM, *Associons and the closure statement*, 1976. ISBN 90 6196 135 1.

MCT 77 W.C.M. KALLENBERG, *Asymptotic optimality of likelihood ratio tests in exponential families*, 1977 ISBN 90 6196 134 3.

MCT 78 E. DE JONGE, A.C.M. VAN ROOIJ, *Introduction to Riesz spaces*, 1977.
ISBN 90 6196 133 5.

MCT 79 M.C.A. VAN ZUIJLEN, *Empirical distributions and rankstatistics*, 1977.
ISBN 90 6196 145 9.

MCT 80 P.W. HEMKER, *A numerical study of stiff two-point boundary problems*, 1977. ISBN 90 6196 146 7.

MCT 81 K.R. APT & J.W. DE BAKKER (eds), *Foundations of computer science II*, part I, 1976. ISBN 90 6196 140 8.

MCT 82 K.R. APT & J.W. DE BAKKER (eds), *Foundations of computer science II*, part II, 1976. ISBN 90 6196 141 6.

MCT 83 L.S. VAN BENTEM JUTTING, *Checking Landau's "Grundlagen" in the AUTOMATH system*, 1979 ISBN 90 6196 147 5.

MCT 84 H.L.L. BUSARD, *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?) books vii-xii*, 1977.
ISBN 90 6196 148 3.

MCT 85 J. VAN MILL, *Supercompactness and Wallman spaces*, 1977.
ISBN 90 6196 151 3.

MCT 86 S.G. VAN DER MEULEN & M. VELDHORST, *Torrix I*, 1978.
ISBN 90 6196 152 1.

* MCT 87 S.G. VAN DER MEULEN & M. VELDHORST, *Torrix II*,
ISBN 90 6196 153 x.

MCT 88 A. SCHRIJVER, *Matroids and linking systems*, 1977.
ISBN 90 6196 154 8.

MCT 89 J.W. DE ROEVER, *Complex Fourier transformation and analytic functionals with unbounded carriers*, 1978. ISBN 90 6196 155 6.

* MCT 90 L.P.J. GROENEWEGEN, *Characterization of optimal strategies in dynamic games*, . ISBN 90 6196 156 4.

* MCT 91 J.M. GEYSEL, *Transcendence in fields of positive characteristic*, . ISBN 90 6196 157 2.

MCT 92 P.J. WEEDA, *Finite generalized Markov programming*, 1979. ISBN 90 6196 158 0.

MCT 93 H.C. TIJMS (ed.) & J. WESSELS (ed.), *Markov decision theory*, 1977. ISBN 90 6196 160 2.

MCT 94 A. BIJLSMA, *Simultaneous approximations in transcendental number theory*, 1978 . ISBN 90 6196 162 9.

MCT 95 K.M. VAN HEE, *Bayesian control of Markov chains*, 1978. ISBN 90 6196 163 7.

* MCT 96 P.M.B. VITÁNYI, *Lindenmayer systems: structure, languages, and growth functions*, . ISBN 90 6196 164 5.

* MCT 97 A. FEDERGRUEN, *Markovian control problems; functional equations and algorithms*, . ISBN 90 6196 165 3.

MCT 98 R. GEEL, *Singular perturbations of hyperbolic type*, 1978. ISBN 90 6196 166 1

MCT 99 J.K. LENSTRA, A.H.G. RINNOOY KAN & P. VAN EMDE BOAS, *Interfaces between computer science and operations research*, 1978. ISBN 90 6196 170 X.

MCT 100 P.C. BAAYEN, D. VAN DULST & J. OOSTERHOFF (Eds), *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1*, 1979. ISBN 90 6196 168 8.

MCT 101 P.C. BAAYEN, D. VAN DULST & J. OOSTERHOFF (Eds), *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2*, 1979. ISBN 90 9196 169 6.

MCT 102 D. VAN DULST, *Reflexive and superreflexive Banach spaces*, 1978. ISBN 90 6196 171 8.

MCT 103 K. VAN HARN, *Classifying infinitely divisible distributions by functional equations*, 1978 . ISBN 90 6196 172 6.

MCT 104 J.M. VAN WOUWE, *Go-spaces and generalizations of metrizability*, 1979. ISBN 90 6196 173 4.

* MCT 105 R. HELMERS, *Edgeworth expansions for linear combinations of order statistics*, . ISBN 90 6196 174 2.

MCT 106 A. SCHRIJVER (Ed.), *Packing and covering in combinatories*, 1979. ISBN 90 6196 180 7.

MCT 107 C. DEN HEIJER, *The numerical solution of nonlinear operator equations by imbedding methods*, 1979. ISBN 90 6196 175 0.

* MCT 108 J.W. DE BAKKER & J. VAN LEEUWEN (Eds), *Foundations of computer science III*, part I, . ISBN 90 6196 176 9.

* MCT 109 J.W. DE BAKKER & J. VAN LEEUWEN (Eds), *Foundations of computer science III*, part II . ISBN 90 6196 177 7.

MCT 110 J.C. VAN VLIET, *ALGOL 68 transput*, part I,1979 . ISBN 90 6196 178 5.

MCT 111 J.C. VAN VLIET, *ALGOL 68 transput*, part II: *An implementation model*, 1979. ISBN 90 6196 179 3.

AN ASTERISK BEFORE THE NUMBER MEANS "TO APPEAR"