KYBERNETIKA — VOLUME 16 (1980), NUMBER 6

Bounds on Discrete Dynamic Programming Recursions I

Models with Non-Negative Matrices

KAREL SLADKÝ

We consider (at discrete time points $n = 0, 1, \ldots$) a system whose utility vector at time n, denoted $x(n)$, obeys the dynamic programming recursion $x(n+1) = \max Q(f)\, x(n)$, where $Q(f)$ is a non-negative (in general reducible) matrix and symbol max is considered with respect to the decision vector f taken from a finite set F. We establish some bounds on the growth of the utility vector (that refine some previous results of Mandl (1970) and Rothblum (1977)) and suggest a policy iteration algorithm for maximizing the growth of the utility vector in the class of stationary policies.

1. INTRODUCTION

We shall consider at discrete time points $n = 0,1, \ldots$ a system with finite state space $I = \{1, 2, \ldots, N\}$ whose utility vector at time n, denoted $x(n)$ (column N-vector), obeys the following dynamic programming recursion

$$x(n + 1) = \max_{f \in F} Q(f)\, x(n) = Q(\hat{f}^{(n)})\, x(n)\,. \tag{1.1}$$

Here $x(0) > 0$ is given, $Q(f)$ is an $N \times N$ non-negative (in general reducible matrix) matrix depending on a decision vector f (i.e. N-vector whose i-th component $f(i)$ specifies the decision in state i), and F is a finite set of all decision vectors at each time point.

Following the usual terminology of dynamic programming, i-th component of the utility vector $x(n)$ will be called utility in state i at time n. So $[Q(f)]_i$ (i-th row of $Q(f)$) depends only on $f(i)$ and we suppose that the set F possesses a "product property", i.e. if $f_1, f_2 \in F$ then there exists also $f \in F$ such that $[Q(f_1)]_{i_1} = [Q(f)]_{i_1}$, $[Q(f_2)]_{i_2} = [Q(f)]_{i_2}$ for each pair $i_1, i_2 \in I$. Consequently, the vectorial maximum in (1.1) does always exist, and on denoting by $F(i)$ the set of all possible decisions in state i (so $f(i) \in F(i)$ for any $f \in F$) then $F \equiv \bigtimes_{i=1}^{N} F(i)$.

Remember that matrices, resp. (column) vectors, are denoted by capital, resp. small, letters. For matrix C $[C]_i$ denotes i-th row of C, $[C]_{ij}$ is reserved for the ij-th element of C. We write $C > 0$, resp. $C \gg 0$, if each element of C is non-negative, resp. positive, and $C > B$, resp. $C \gg B$, iff $C - B > 0$, resp. $C - B \gg 0$. Symbol C^T denotes the transpose of C. The same notations are also used for vectors. Symbol I, resp. e, is reserved for a unit matrix, resp. unit vector, of an appropriate dimension. Sometimes we shall need lexicographical ordering of matrices. We say that C is lexicographically greater than B (and write $C \succ B$) iff the first non-zero element of each row of $C - B$ is positive and $C \neq B$. Similarly, $C \succeq B$ if either $C \succ B$ or $C = B$.

A sequence of decision vectors, say $\pi \equiv (\dots, f^{(n)}, \dots f^{(0)})$, is called a (Markovian) policy and if $f^{(n)} \equiv f$ for each n then such a policy is called stationary. Observe that $f^{(n)}$ denotes the decision vector used at time n; for what follows it will be advantageous to used the above "opposite" time denotation.

Dynamic programming recursion (1.1) has several interesting interpretations:

(I) *Functional equations for total expected return of classical Markov decision chains* (cf. e.g. [2], [4], [6], [8], [15]).

Let us consider a classical Markov decision chain with $N - 1$ states and let $P(f)$, resp. $r(f) \geqq 0$, be the transition probability matrix, resp. vector of one-stage rewards, if decision vector $f \in F$ is selected. On setting

$$Q(f) = \begin{bmatrix} P(f) & r(f) \\ 0 & 1 \end{bmatrix}, \quad [x(0)]_N = 1$$

then by (1.1) $[x(n)]_N \equiv 1$ (dummy variable) and for $\tilde{x}(n) = v(n)$ (subvector of $x(n)$ containing the first $N - 1$ components of $x(n)$) we get (here $v(0) \geqq 0$ denotes the vector of "terminal rewards")

$$v(n+1) = r(\hat{f}^{(n)}) + P(\hat{f}^{(n)})\, r(\hat{f}^{(n-1)}) + \dots + P(\hat{f}^{(n)}) \dots P(\hat{f}^{(1)})\, r(\hat{f}^{(0)}) + \\ + P(\hat{f}^{(n)}) \dots P(\hat{f}^{(0)})\, v(0) = r(\hat{f}^{(n)}) + P(\hat{f}^{(n)})\, v(n)$$

(so for $i < N$ $[v(n)]_i = [x(n)]_i$ is the maximum total expected return in state i to be obtained in the n next transitions).

Similarly, if

$$Q(f) = \begin{bmatrix} P(f) & r(f) & 0 \\ 0 & 1 & e^T \\ 0 & 0 & J \end{bmatrix}$$

(here 0's always denote zero submatrices of appropriate dimensions) with $P(f)$, $r(f)$ being of dimension $N - l$ and J being an upper triangular matrix of dimension $l - 1$ whose each entry on or above the diagonal equals 1 (consequently also e^T has dimension $l - 1$) on setting for any $n - l < j \leqq N$ $[x(0)]_j = 1$ then by a simple calculation we get for each $j > N - l$ $[x(n)]_j = \binom{n + N - j}{N - j}$. Moreover, for $\tilde{x}(n)$ (subvector of $x(n)$ containing the first $N - l$ components of $x(n)$) we have

$$\tilde{x}(n + 1) = \binom{n + l - 1}{l - 1} r(\hat{f}^{(n)}) + P(\hat{f}^{(n)})\, \tilde{x}(n)\,.$$

So in this case $\tilde{x}(n)$ represents maximum of the appropriate cumulative total expected rewards if n transitions of the considered Markov decision chains are to be left.

(II) *Functional equation for "risk-sensitive" (multiplicative) Markov decision chains* (cf. [7]).

Let us again consider a Markov decision chain with N states and let $P(f)$, resp. $R(f)$, be the transition probability matrix, resp. the matrix of one-stage rewards, under decision f (so $[R(f)]_{ij}$ denotes the reward accrued to a transition from state i into state j). Now let us suppose that the sequence of successively earned rewards is evaluated according to some multiplicative utility function; i.e., let to a sequence of a successively earned outcomes $c_1, \ldots, c_n$ be assigned a utility function $\exp(-\gamma \sum_{t=1}^{n} c_t)$ (here given real number γ is called "risk aversion" coefficient). Clearly, in the stochastic case this problem cannot be transformed into the problem with additive utility by taking logarithms.

On denoting $x(n)$ the vector of maximum expected utilities in the n next transitions (so $[x(n)]_i$ denotes maximum expected utility in state i to be obtained in the n next transitions), an easy calculation reveals that $x(n)$ obeys dynamic programming recursion (1.1) where the elements of $Q(f)$ are given by

$$[Q(f)]_{ij} = [P(f)]_{ij} \exp(-\gamma [R(f)]_{ij})$$

and $x(0)$ denotes the vector of "terminal rewards" (here we suppose $x(0) > 0$).

(III) *Supervised linear economic models* (cf. [3], Chap. 16).

Let us consider at discrete time points an economic system with N industries and let $[Q(f)]_{ij}$ corresponds to the influence of the j-th industry on the i-th industry if decision f is selected. Similarly, let $[x(n)]_i$ describe the "state" of the i-th industry at time n and let $x(n+1) = Q(f)\, x(n)$ (here we tacitly assume that $x(n+1)$ depends also on f). Obviously, a sequence of decisions maximizing the "growth" of the whole economy (i.e. the growth of $\{x(n)\}$) must obey dynamic programming recursion (1.1).

(IV) *Controlled branching processes* (cf. [3], Chap. 16).

We consider at discrete time points a population consisting of N types of individuals and let $[x(n)]_i$ be the expected number of individuals of type i at time n. If $[Q(f)]_{ij}$ denotes the expected number of individuals of type i that arise under decision f from an individual of type j at the next considered time instant, the policy (i.e. sequence of decisions) maximizing the growth of the whole population as well as the maximum expected number of each type of the individuals is again given by the dynamic programming recursion (1.1).

(V) *Controlled growth of personnel in an organization.*

Let us consider an organization the personnel of which is divided into $N - 1$ ranks. We set $[x(n)]_N \equiv 1$ (dummy variable) and for $i = 1, 2, \ldots, N - 1$ we denote by $[x(n)]_i$ maximum expected number of personnel belonging at time n to rank i. Let $P^T(f)$ be a transpose of an $(N-1) \times (N-1)$ substochastic matrix $P(f)$ such that $[P(f)]_{ij}$ is the probability that under decision f a person belonging to rank i will promote or demote to rank j at the next considered time instant (so $1 - \sum_{j=1}^{N-1} [P(f)]_{ij}$ is the probability that a person belonging to rank i will be dismissed). Clearly, on introducing $(N-1)$-column vector $s(f)$ whose i-th component is the number of newly hired persons of rank i if decision f is selected,

$$x(n + 1) = \max_{f \in F} Q(f)\, x(n) \quad \text{with} \quad Q(f) = \begin{bmatrix} P^T(f) & s(f) \\ 0 & 1 \end{bmatrix}.$$

2. PRELIMINARIES

Let $\sigma(f)$ be the spectral radius of (a non-negative matrix) $Q(f)$. According to the well-known Perron - Frobenius theorem $\sigma(f)$ equals to the largest positive eigenvalue of $Q(f)$ and we can choose the corresponding eigenvector $u(f) > 0$. Recall that if $Q(f)$ is irreducible then even $u(f) \gg 0$ and $\sigma(f)$ is simple. Moreover, if $Q(f)$ is reducible, i.e., if by suitable permuting of rows and corresponding columns of $Q(f)$ it is possible to write (remember that in our matrix notation blanks will always denote zero submatrices of appropriate dimensions)

$$Q(f) = \begin{bmatrix} Q_{(11)}(f) & Q_{(12)}(f) \cdots & Q_{(1r)}(f) \\ & Q_{(22)}(f) \cdots & Q_{(2r)}(f) \\ & \ddots & \vdots \\ & & Q_{(rr)}(f) \end{bmatrix} \tag{2.1}$$

where each $Q_{(ii)}(f)$ itself is an irreducible matrix with spectral radius $\sigma_{(i)}(f)$, necessary and sufficient conditions for $u(f) \gg 0$ can be easily formulated by means of accessibility properties between irreducible classes of $Q(f)$. Similarly as in Markov chain theory we say that $Q_{(ii)}(f)$ is accessible to $Q_{(ll)}(f)$ iff there exists a sequence of integers $k_0 = i < k_1 < \ldots < k_p = l$ such that $Q_{(k_{j-1},k_j)}(f) > 0$ for all $j = 1, \ldots, p$ and $Q_{(ii)}(f)$ is called a basic, resp. non-basic, class of $Q(f)$ iff $\sigma_{(i)}(f) = \sigma(f)$, resp. $\sigma_i(f) < \sigma(f)$.

It can be shown (cf. [5], Theorem 7 of Chap. 13): $u(f) \gg 0$ if and only if

(i) $\quad \sigma_{(i)}(f) < \sigma(f) \Rightarrow Q_{(ij)}(f) \neq 0$ at least for one $j \neq i$; and

(ii) $\quad \sigma_{(i)}(f) = \sigma(f) \Rightarrow Q_{(ij)}(f) = 0$ for any $j \neq i$.

So (cf. (2.1)) $u(f) \gg 0 \Rightarrow \sigma_{(r)}(f) = \sigma(f)$ and on relabelling the irreducible classes of $Q(f)$ we may assume that for some $r' = 1, 2, \ldots, r$

$$\sigma_{(i)}(f) < \sigma(f) \Leftrightarrow i < r', \quad \sigma_{(i)}(f) = \sigma(f) \Leftrightarrow i \geqq r'.$$

Obviously, using the notion of accessibility instead of (i), (ii) it can be equivalently stated:

$u(f) \gg 0$ if and only if each non-basic, resp. basic, class of $Q(f)$ is accessible to some basic class, resp. is not accessible to any other irreducible class, of $Q(f)$.

In virtue of this fact diagonal submatrices of $Q(f)$ in (2.1) need not be the "largest" submatrices of $Q(f)$ having strictly positive eigenvectors corresponding to their spectral radii. However in the proof of Lemma 2.1 an algorithmic procedure will be given for suitable ordering the irreducible classes of $Q(f)$ to obtain an upper triangular matrix whose diagonal (possibly reducible) classes are the "largest" submatrices of $Q(f)$ having strictly positive eigenvectors. Remember that an irreducible class, say $Q_{(jj)}(f)$, belonging to some $Q_{ii}(f)$ (submatrix of $Q(f)$), will be called basic, resp. non-basic, class of $Q_{ii}(f)$ iff $\sigma_{(j)}(f) = \sigma_i(f)$, resp. $\sigma_{(j)}(f) < \sigma_i(f)$ ($\sigma_i(f)$ denotes spectral radius of $Q_{ii}(f)$). It holds:

Lemma 2.1. By possibly permuting rows and corresponding columns of (2.1) we can write

$$(2.1.1) \qquad Q(f) = \begin{bmatrix} Q_{11}(f) & Q_{12}(f) \ldots & Q_{1s}(f) \\ & Q_{22}(f) \ldots & Q_{2s}(f) \\ & \ddots & \vdots \\ & & Q_{ss}(f) \end{bmatrix}$$

where for $i = 1, 2, \ldots,$ (s depends on f)

$$(2.1.2) \qquad Q_{ii}(f)\, u_i(f) = \sigma_i(f)\, u_i(f)$$

with $\sigma_i(f)$, resp. $u_i(f) \gg 0$, being the spectral radius, resp. corresponding right eigenvector, of $Q_{ii}(f)$ (in general reducible) and

$$(2.1.3) \qquad \sigma_1(f) \geqq \sigma_2(f) \geqq \ldots \geqq \sigma_s(f)$$

with $\sigma_{i+1}(f) = \sigma_i(f)$ implying that each irreducible class of $Q_{ii}(f)$ is accessible to some basic of $Q_{i+1,i+1}(f)$.

Proof. Let (cf. (2.1)) $K = \{1, 2, \ldots, r\}$ and let $i \in K^{(1)} \subset K$ iff $\sigma_{(i)}(f) = \sigma(f)$. Now let $i \in K_1^{(1)} \subset K^{(1)}$ iff $Q_{(ii)}(f)$ is not accessible to any other basic class of $Q(f)$ (obviously $K_1^{(1)} \neq \emptyset$) and let us define recursively $\{K_m^{(1)} \subset K^{(1)}, m = 1, \ldots, p\}$ such that for $m > 1$ $i \in K_m^{(1)}$ iff $Q_{(ii)}(f)$ is accessible to some $Q_{(ll)}(f)$ with $l \in K_{m-1}^{(1)}$ and is not accessible to any $Q_{(ll)}(f)$ with $l \in K^{(1)} \setminus \bigcup_{n=1}^{m-1} K_n^{(1)}$. Let us further introduce

$\{\bar{K}_m^{(1)}, m = p, \ldots, 1\}$ such that $i \in \bar{K}_p^{(1)}$ iff $Q_{(ii)}(f)$ (possibly non-basic) is accessible to some $Q_{(ll)}(f)$ with $l \in K_p^{(1)}$ and for $m < p$ $i \in \bar{K}_m^{(1)}$ iff $i \notin \bigcup_{n=m+1}^{p} \bar{K}_n^{(1)}$ and $Q_{(ii)}(f)$ is accessible to some $Q_{(ll)}(f)$ with $l \in K_m^{(1)}$ (obviously $\bar{K}_m^{(1)} \supset K_m^{(1)}$).

From the above construction it is clear that the irreducible classes labelled by integers belonging to each $\bar{K}_m^{(1)}$ $(m = 1, \ldots, p)$ form the "largest" submatrices of $Q(f)$ having strictly positive eigenvectors corresponding to $\sigma_1(f) \equiv \sigma(f)$; if $i \in \bar{K}_m^{(1)}$ then $Q_{(ii)}(f)$ cannot be accessible to any $Q_{(ll)}(f)$ with $l \in \bar{K}_n^{(1)}$ and $m < n$ and if $i \in \bar{K}_m^{(1)}$ with $m > 1$ then $Q_{(ii)}(f)$ must be accessible to some $Q_{(ll)}(f)$ with $l \in K_{m-1}^{(1)}$. On possibly permuting the rows and corresponding columns of (2.1) belonging to $\bigcup_{m=1}^{p} \bar{K}_m^{(1)}$ it is possible to construct (cf. (2.1.1)) $Q_{11}(f), \ldots, Q_{pp}(f)$ (with $\sigma_1(f) = \ldots = \sigma_p(f) = \sigma(f)$) such that each $Q_{ii}(f)$ $(i = 1, \ldots, p)$ contains the "rows" labelled by integers from $\bar{K}_i^{(1)}$. Recalling the accessibility properties of irreducible classes belonging to $Q_{ii}(f)$ obviously $Q_{ij}(f) = 0$ for any $j < i \leqq p$.

Now cancel in (2.1) the rows (and columns) of $Q(f)$ belonging to diagonal classes labelled by integers from $\bigcup_{m=1}^{p} \bar{K}_m^{(1)}$ to obtain submatrix ${}^{(2)}Q(f)$ of $Q(f)$ (whose rows and columns contain only diagonal classes labelled by integers from ${}^{(2)}K = K \setminus \bigcup_{m=1}^{p} \bar{K}_m^{(1)}$), find its spectral radius (that must be less than $\sigma(f)$) and repeat the whole above procedure to obtain analogously defined $K^{(2)}$, $K_m^{(2)}$ and $\bar{K}_m^{(2)}$'s. Repeating further (if necessary) the above construction we find $\bar{K}_m^{(i)}$'s (where $K = \bigcup_{i,m} \bar{K}_m^{(i)}$ and $\bar{K}_m^{(i)} \cap \bar{K}_m^{(l)} = \emptyset$ if $i \neq l$ or $m \neq n$). By possibly permuting rows and corresponding columns $Q(f)$ can be written in an upper block-triangular form satisfying (2.1.1), (2.1.2) and (2.1.3). □

So in virtue of Lemma 2.1 for any $f \in F$ we can construct (by possibly permuting rows and corresponding columns of $Q(f)$) diagonal classes $Q_{ii}(f)$ as the "largest" submatrices of $Q(f)$ having strictly positive eigenvectors corresponding to $\sigma_i(f)$ and fulfilling also the remaining properties of Lemma 2.1. Similarly, let the state space $I = \bigcup_{i=1}^{s} I_i(f)$ (with $I_i(f) \cap I_k(f) = \emptyset$ for $i \neq k$) such that $j \in I_i(f)$ iff $[Q(f)]_{jj}$ belongs to $Q_{ii}(f)$ and let $\bar{I}_i(f) \subset I_i(f)$ denotes all j's belonging to some basic class of $Q_{ii}(f)$.

Now let us introduce index of class $Q_{ii}(f)$, denoted $v_i(f)$, in such a way that $v_s(f) = 1$ and for $i = 1, \ldots, s - 1$

$$v_i(f) = r \quad \text{iff} \quad \sigma_i(f) = \sigma_{i+1}(f) = \ldots = \sigma_{i+r-1}(f)$$

where either

$$\sigma_{i+r-1}(f) > \sigma_{i+r}(f) \quad \text{or} \quad i + r - 1 = s .$$

Obviously, if $\sigma_1(f) > \sigma_2(f) > \ldots > \sigma_s(f)$ then all $v_i(f) = 1$. Observe that in Lemma

2.1 (2.1.3) together with the subsequent condition on $\sigma_{i+1}(f) = \sigma_i(f)$ can be replaced by

$$(2.1.3') \qquad (\sigma_1(f)\,;\ \nu_1(f)) \succ (\sigma_2(f)\,;\ \nu_2(f)) \succ \ldots \succ (\sigma_s(f)\,;\ \nu_s(f))$$

(here symbol $\succ$ denotes lexicographically greater). It can be shown (cf. [12]) that $\nu_1(f)$ equals to the index of $Q(f)$ (i.e. $\nu_1(f)$ is the smallest integer n such that the null spaces of $(Q(f) - \sigma(f)\,I)^n$ and $(Q(f) - \sigma(f)\,I)^{n+1}$ coincide). However, this fact only motivates our terminology; it will not be used anywhere in the sequel.

Moreover, to each state of $Q(f)$, say j, we introduce spectral radius $\sigma(f, j)$, resp. index $\nu(f, j)$, by setting $\sigma(f, j) = \sigma_i(f)$, resp. $\nu(f, j) = \nu_i(f)$, where $j \in I_i(f)$. Obviously, $\sigma(f, j)$ equals to the maximum spectral radius of any irreducible class that is accessible from j and $\nu(f, j)$ is the maximum number of such irreducible classes that can be subsequently reached.

Throughout the paper we make the following general assumption.

Assumption GA. For any $f \in F$ $\sigma_i(f) > 0$ for each $i = 1, \ldots, s = s(f)$. (Obviously, by (2.1.2) it suffices only to assume $\sigma_s(f) > 0$).

In the following lemma some useful facts about non-negative matrices are summarized.

Lemma 2.2. It holds:

(1) $\sigma(f)$ is a continuous function of the elements of $Q(f)$ (of course, this need not be true for the corresponding eigenvector $u(f)$).

(2) If $Q(f_1) > Q(f_2)$ then $\sigma(f_1) \geqq \sigma(f_2)$ with a strict inequality if $Q(f_1)$ is irreducible.

(3) On setting $\tilde{Q}_{ii}(f) = (\sigma_i(f))^{-1}\, Q_{ii}(f)$ then

$$\lim_{n\to\infty} n^{-1} \sum_{m=0}^{n-1} (\tilde{Q}_{ii}(f))^m = Q^*_{ii}(f) > 0 \quad \text{always exists}$$

(with $[Q^*_{ii}(f)]_{jk} > 0$ for each pair $j, k \in I_i(f)$ belonging to the same basic class of $Q_{ii}(f)$) and $\sigma_i(f)\, Q^*_{ii}(f) = Q_{ii}(f)\, Q^*_{ii}(f)$. Moreover, if $Q_{ii}(f)$ is aperiodic (i.e. modulus of each eigenvector of $Q_{ii}(f)$ different from $\sigma_i(f)$ is less than $\sigma_i(f)$) then there exists $\lim_{n\to\infty} (\tilde{Q}_{ii}(f))^n = Q^*_{ii}(f)$. In case that $Q_{ii}(f)$ is periodic, on setting $\varkappa$ equal to the l.c.m. of the periods of all basic irreducible classes of $Q_{ii}(f)$, then for $l = 0, 1, \ldots, \varkappa - 1$ there exists

$$\lim_{n\to\infty} (\tilde{Q}_{ii}(f))^{n\varkappa+l} = {}^{(l)}Q^*_{ii}(f) \quad \text{where} \quad \sum_{l=0}^{\varkappa-1} {}^{(l)}Q^*_{ii}(f) = \varkappa\, Q^*_{ii}(f)\,.$$

(4) For $n \to \infty$ $(Q_{ii}(f))^n$ is bounded iff $\sigma_i(f) \leqq 1$ (in case that $\sigma_i(f) < 1$ then even $\lim_{n\to\infty} (Q_{ii}(f))^n = 0$).

(5) There exists unique v_i satisfying $(\sigma_i(f)\, I - Q_{ii}(f))\, v_i = h_i$ for given h_i such

that $[v_i]_j = [h_i]_j = 0$ for all $j \in \bar{I}_i(f)$. Moreover, if $h_i > 0$, resp. $h_i < 0$, then $v_i > 0$, resp. $v_i < 0$.

Proof. Continuity of $\sigma(f)$ is immediate as the solutions of the respective characteristic equation depend continuously on the elements of $Q(f)$ (cf. [11], Appendix K). The proof of part (2) can be found in [5] (cf. Theorem 6 of Chap. 13). To establish parts (3), (4) observe that by (2.1.2) $Q_{ii}(f)$ is positively similar to some stochastic matrix (here the similarity matrix T_i is diagonal with $T_i\, u_i(f) = e$) and so the above properties of $Q_{ii}(f)$ follow immediately from the well-known limiting properties of stochastic matrices. Using the same way of reasoning we can easily verify part (4). The proof of part (5) follows immediately on using well-known fact that for each non-basic class of $Q_{ii}(f)$, say $Q_{(kk)}(f)$, $(\sigma(f) I - Q_{(kk)}(f))^{-1} \gg 0$ always exists together with the accessibility properties between irreducible classes of matrix $Q_{ii}(f)$. □

3. POLICY ITERATION METHOD FOR MAXIMIZING GROWTH OF THE UTILITY VECTOR

Throughout this section we shall consider only the class of stationary policies. For given $f \in F$ we set for $n = 0, 1, \ldots$ with $x(0; f) = x(0) > 0$

$$x(n + 1; f) = Q(f)\, x(n; f) \tag{3.1}$$

and we shall examine the growth of $\{x(n; f),\ n = 0, 1, \ldots\}$. Furthermore, we suggest a policy iteration algorithm for finding decision vector $\hat{f} \in F$ maximizing the growth of $x(n; f)$ and show that for some suitable fixed ordering of diagonal classes of $Q(\hat{f})$ (such that (2.1.1), (2.1.2), (2.1.3') hold for $f \equiv \hat{f}$) also $Q(f)$ will be a block-triangular matrix for any $f \in F$.

Firstly, using the block-triangular structure of $Q(f)$ (cf. (2.1.1)) by means of Perron - Frobenius theorem we derive some bounds on $\{x(n; f)\}$. Of course, classical results of matrix theory (cf. e.g. Chap. 5 of [5]) enable to express $x(n; f)$ as a function of eigenvalues of $Q(f)$ but, on the other hand, the approach used in the proof of the following lemma can be easily adapted even for $x(n)$ obeying dynamic programming recursion (1.1) as it is indicated in Theorem 4.1.

Lemma 3.1. Let $x_i(n; f)$ be a subvector of $x(n; f)$ whose components are labelled from $I_i(f)$. Then for suitably chosen vector $u_i(f)$ (where $u_i(f) \gg 0$ satisfy (2.1.2)) and all $n = 1, 2, \ldots$

$$x_i(n; f) \leqq (\sigma_i(f))^n \binom{n + v_i(f) - 1}{v_i(f) - 1} u_i(f)\,. \tag{3.1.1}$$

Proof. First observe by (2.1.1), (2.1.2) that for suitably chosen $u'(f) = [u'_i(f)]_{i=1}^s \gg 0$ such that $u'(f) \geqq x(0)$ we can find an upper triangular matrix $K(f) \geqq Q(f)$

with $K_{ii}(f) = Q_{ii}(f)$ (for any $i = 1, \ldots, s$), $K_{ij}(f) = 0$ (for any pair $j < i$) and $K_{ij}(f) \geqq Q_{ij}(f)$ (for any $j > i$) such that for any $j > i$

$$Q_{ij}(f)\, u'_j(f) \leqq K_{ij}(f)\, u'_j(f) = \sigma_j(f)\, u'_i(f)\,. \tag{3.1.2}$$

The construction of $K_{ij}(f)$'s together with the choice of suitable $u'_i(f)$'s can be performed successively for $i = s - 1$, $j = s$; then for $i = s - 2$, $j = s, s - 1$; then for $i = s - 3$, $j = s, s - 1, s - 2$ etc. by selecting $u'_i(f) \geqq x_i(0)$ such that for $j = i + 1, \ldots, s$ $\sigma_j(f)\, u'_i(f) \geqq Q_{ij}(f)\, u'_j(f)$ and then "enlarging" the elements of $Q_{ij}(f)$ such that (3.1.2) will hold. So we get for all $i = 1, \ldots, s$ (cf. (2.1.2), (3.1.2))

$$x_i(1; f) \leqq \sigma_i(f)\, u'_i(f) + \sum_{j=i+1}^{s} K_{ij}(f)\, u'_j(f) = u'_i(f) \sum_{j=i}^{s} \sigma_j(f)$$

and by induction with respect to n we can verify that

$$x_i(n; f) \leqq u'_i(f) \sum_{m_i + \ldots + m_s = n} (\sigma_i(f))^{m_i} (\sigma_{i+1}(f))^{m_{i+1}} \ldots (\sigma_s(f))^{m_s} \tag{3.1.3}$$

(here the sum on the RHS of (3.1.3) is to be understood over all integers $m_k \geqq 0$ satisfying $\sum_{k=i}^{s} m_k = n$).

Observe that the induction step for (3.1.3) follows easily as by (2.1.1), (2.1.2), (3.1.2)

$$\begin{aligned} x_i(n + 1; f) &= \sum_{j=i}^{s} Q_{ij}(f)\, x_j(n; f) \leqq \\ &\leqq \sum_{j=i}^{s} Q_{ij}(f)\, u'_j(f) \Big[\sum_{m_j + \ldots + m_s = n} (\sigma_j(f))^{m_j} \ldots (\sigma_s(f))^{m_s} \Big] \leqq \\ &\leqq u'_i(f) \sum_{j=i}^{s} \sigma_j(f) \Big[\sum_{m_j + \ldots + m_s = n} (\sigma_j(f))^{m_j} \ldots (\sigma_s(f))^{m_s} \Big] = \\ &= u'_i(f) \sum_{m_i + \ldots + m_s = n+1} (\sigma_i(f))^{m_i} \ldots (\sigma_s(f))^{m_s}\,. \end{aligned}$$

Now if $v_i(f) = r$ (i.e. if $\sigma_i(f) = \ldots = \sigma_{i+r-1}(f) > \sigma_{i+r}(f)$) then on setting $\vartheta_i(f) = \sigma_{i+r}(f)/\sigma_i(f) < 1$ in virtue of (2.1.3) we get

$$\sum_{m_i + \ldots + m_s = n} (\sigma_i(f)^{m_i}) \ldots (\sigma_s(f))^{m_s} \leqq (\sigma_i(f))^n \sum_{m_i + \ldots + m_s = n} (\vartheta_i(f))^{(m_{i+r} + \ldots + m_s)}\,. \tag{3.1.3'}$$

As $\sum_{m_i + \ldots + m_s = n} 1 = \binom{n + s - i}{s - i}$ (this identity can be easily verified e.g. by induction with respect to s; for $s = i$ it holds trivially and the induction step is immediate by

$$\sum_{m_i + \ldots + m_s = s} 1 = \sum_{l=0}^{n} \Big(\sum_{m_i + \ldots + m_{s-1}} 1 \Big) = \sum_{l=0}^{n} \binom{l + s + 1 - i}{s - 1 - i} = \binom{n + s - i}{s - i} \Big)$$

we get

$$\sum_{m_i+\ldots+m_s=n} (\vartheta_i(f))^{(m_{i+r}+\ldots+m_s)} = \sum_{k=0}^{n} [\sum_{m_i+\ldots+m_{i+r-1}=n-k} 1] \cdot \tag{3.1.4}$$

$$\cdot [\sum_{m_{i+r}+\ldots+m_s=k} (\vartheta_i(f))^k] = \sum_{k=0}^{n} (\vartheta_i(f))^k \binom{n-k+r-1}{r-1} \binom{k+s-i-r}{s-i-r} \leqq$$

$$\leqq \binom{n+r-1}{r-1} \sum_{k=0}^{n} (\vartheta_i(f))^k \binom{k+s-i-r}{s-i-r}.$$

But for $\vartheta_i(f) \in (0, 1)$ the sum on the RHS of (3.1.4) is obviously nondecreasing with respect to n and there exists

$$\lim_{n\to\infty} \sum_{k=0}^{n} (\vartheta_i(f))^k \binom{k+s-i-r}{s-i-r} = \alpha_i(f) < \infty . \tag{3.1.4'}$$

Denoting $u_i(f) = \alpha_i(f)\, u_i'(f)$ by (3.1.3), (3.1.3'), (3.1.4), (3.1.4') we immediately get (3.1.1). □

Now we shall try to find a decision vector, say $\hat{f} \in F$, maximizing the above bounds on $\{x_i(n; f)\}$ for large n uniformly with respect to $f \in F$. To this order we suggest a policy iteration algorithm for finding decision vector $\hat{f} \in F$ such that for all $f \in F$, $j \in I$ $(\sigma(\hat{f}, j); \nu(\hat{f}, j)) \succeq (\sigma(f, j); \nu(f, j))$.

Algorithm 1. (*Policy Iteration Method for Maximizing Growth of the Utility Vector*).

Construct a (finite) sequence of decision vectors $f_0, f_1, \ldots, f_n, \ldots f_r \equiv \hat{f}$ with f_0 arbitrary and f_{n+1} derived from f_n in the following way:

(i) Find a decision vector $f \in F$ such that for each $j \in I$

$$f(j) = f_n(j) \tag{3.2}$$

iff $[Q(h)]_{jk} = 0$ for any $h \in F$ and each $k \in I$ with $(\sigma(f_n, k); \nu(f_n, k)) \succ \succ (\sigma(f_n, j); \nu(f_n, j))$ and set

$$f(j) = h(j) \tag{3.2'}$$

for $h \in F$ such that $[Q(h)]_{jk} \neq 0$ for some $k \in I$ satisfying $(\sigma(f_n, k); \nu(f_n, k)) \succeq \succeq (\sigma(f_n, l); \nu(f_n, l)) \succ (\sigma(f_n, j); \nu(f_n, j))$ for any $g \in F$ and each $l \in I$ with $[Q(g)]_{jl} \neq 0$ (observe that $[Q(h)]_j$ depends only on $h(j)$).

(ii) Let the rows and corresponding columns of $Q(f)$ be permuted so that (2.1.1), (2.1.2), (2.1.3') hold. Construct $f_{n+1} \in F$ in such a way that for each $i = 1, \ldots, s$

$$\gamma_i(f_{n+1}; f) = \max_{h\in F} \gamma_i(h; f) \tag{3.3}$$

with

$$[\gamma_i(f_{n+1}; f)]_j = 0 \Rightarrow f_{n+1}(j) = f(j) \tag{3.3'}$$

where

$$\gamma_i(h;f) = (Q_{ii}(h) - Q_{ii}(f))\, u_i(f) \geqq 0 \tag{3.4}$$

(here $Q_{ii}(h)$ as well as $Q_{ii}(f)$ contain exactly the elements labelled from $I_i(f)$, so $\gamma_i(f;f) = 0$ and, consequently, $\gamma_i(h;f) \geqq 0$ can always be found). Observe that $[\gamma_i(h;f_n)]_j$ depends only on $h(j)$ and set for $l = 1, \ldots, s = s(f_{n+1})$ $[u(f_{n+1})]_j = [u(f_n)]_j$ for one j of each basic class of any $Q_{ll}(f_{n+1})$ (where the elements of $Q_{ll}(f_{n+1})$ belong to $I_l(f_{n+1})$).

Then it holds

Theorem 3.2. In a finite number of policy improvement steps according to Algorithm 1 we obtain decision vector $f_r \equiv \hat{f} \in F$ that cannot be further improved. Then it holds for any $f \in F$ and all $j \in I$

$$(\sigma(\hat{f}, j)\,;\; v(\hat{f}, j)) \succeq (\sigma(f, j)\,;\; v(f, j))\,. \tag{3.2.1}$$

Moreover, by possibly permuting rows and corresponding columns of $Q(\hat{f})$,

$$Q(\hat{f}) = \begin{bmatrix} Q_{11}(\hat{f}) & Q_{12}(\hat{f}) \ldots & Q_{1s}(\hat{f}) \\ & Q_{22}(\hat{f}) \ldots & Q_{2s}(\hat{f}) \\ & \ddots & \vdots \\ & & Q_{ss}(\hat{f}) \end{bmatrix} \tag{3.2.2}$$

where (recall that $j \in I_i(\hat{f}) \Rightarrow \sigma(\hat{f}, j) = \sigma_i(\hat{f}),\ v(\hat{f}, j) = v_i(\hat{f})$)

$$(\sigma_1(\hat{f})\,;\; v_1(\hat{f})) \succ (\sigma_2(\hat{f})\,;\; v_2(\hat{f})) \succ \ldots \succ (\sigma_s(\hat{f})\,;\; v_s(\hat{f})) \tag{3.2.3}$$

and for $i = 1, 2, \ldots, s = s(\hat{f})$ and any $f \in F$

$$Q_{ii}(f)\, u_i(\hat{f}) \leqq Q_{ii}(\hat{f})\, u_i(\hat{f}) = \sigma_i(\hat{f})\, u_i(\hat{f}) \quad \text{with} \quad u_i(\hat{f}) \gg 0 \tag{3.2.4}$$

(the decomposition of $Q(f)$ for any $f \in F$ is considered according to (3.2.2); so the entries of $Q_{ii}(f)$ are labelled by integers from $I_i(\hat{f})$) and for any $k < i$

$$Q_{ik}(f) \equiv 0 \tag{3.2.5}$$

(observe that for an arbitrary element of $Q_{ik}(f)$, say $[Q_{ik}(f)]_{jl}$, we have $j \in I_i(\hat{f})$, $l \in I_k(\hat{f})$).

Before the proof of Theorem 3.2 we present two lemmas being the main ingredients to the policy iteration procedure of Algorithm 1.

Lemma 3.3. Let $f,\ g \in F$ satisfy (3.3), (3.3′) and (3.4) (with $g \equiv f_{n+1}$). Then for each $j \in I$

$$(\sigma(g, j)\,;\; v(g, j)) \succeq (\sigma(f, j)\,;\; v(f, j))\,. \tag{3.3.1}$$

Proof. Let us fix certain class $Q_{ii}(f)$ (cf. (2.1.1), (2.1.2)), set $Q'(f) \equiv Q_{ii}(f)$ and let $I' \equiv I_i(f)$. Similarly, let $Q'(g)$ be a submatrix of $Q(g)$ containing exactly all elements labelled by integers from I'. By possibly permuting rows and corresponding columns of $Q'(g)$ we may suppose (cf. (2.1.1), (2.1.2), (2.1.3')) that

$$Q'(g) = \begin{bmatrix} Q'_{11}(g) & Q'_{12}(g) \dots & Q'_{1p}(g) \\ & Q'_{22}(g) \dots & Q'_{2p}(g) \\ & \ddots & \vdots \\ & & Q'_{pp}(g) \end{bmatrix} \tag{3.3.2}$$

where for each $Q'_{kk}(g)$ $(k = 1, \dots, p)$

$$\sigma'_k(g)\, u'_k(g) = Q'_{kk}(g)\, u'_k(g)\,, \quad u'_k(g) \gg 0 \tag{3.3.3}$$

and for any $1 \leqq k < l \leqq p$

$$(\sigma'_k(g); \nu'_k(g)) \succ (\sigma'_l(g)\,;\; \nu'_l(g))\,. \tag{3.3.4}$$

Remember that here the values of spectral radius $\sigma'_k(g)$, resp. index $\nu'_k(g)$, are considered with respect to matrix $Q'(g)$ and by $\sigma'(g, j)$, resp. $\nu'(g, j)$, we denote the values of $\sigma'_k(g)$, resp. $\nu'_k(g)$, corresponding to $j \in I'$.

Now observe (cf. Lemma 2.1 and part (2) of Lemma 2.2) that for any $j \in I'$ $\sigma'_p(g) \leqq \leqq \sigma'(g, j) \leqq \sigma(g, j)$ and, by the definition of $Q'(f)$, $\sigma(f, j) = \sigma_i(f)$ for each $j \in I'$.

To verify (3.3.1) first we establish that for each $j \in I'$ (and, consequently, also for each $j \in I$ as $Q_{ii}(f) \equiv Q'(f)$ is chosen arbitrarily in $Q(f)$) $\sigma(g, j) \geqq \sigma(f, j)$ by showing that

$$\sigma'_p(g) \geqq \sigma_i(f)\,. \tag{3.3.5}$$

To this order observe that by (3.3.2), (3.3.3) and (2.1.2)

$$\begin{aligned} \sigma'_p(g)\,(u'_p(g) - u'_p(f)) + (\sigma'_p(g) - \sigma_i(f))\, u'_p(f) = \\ = Q'_{pp}(g)\,(u'_p(g) - u'_p(f)) + Q'_{pp}(g)\, u'_p(f) - \sum_{k=1}^{p} Q'_{pk}(f)\, u'_k(f) = \\ = Q'_{pp}(g)\,(u'_p(g) - u'_p(f)) + \gamma'_p(g; f) \end{aligned} \tag{3.3.6}$$

(here $u'_k(f)$, resp. $\gamma'_k(g; h)$, is a subvector of $u_i(f)$, resp. $\gamma_i(g; f)$, containing the components corresponding to $Q'_{kk}(g)$ and, similarly, $Q'_{pk}(f)$ contains the elements corresponding to $Q'_{pk}(g)$). By (3.3.6) (as $\gamma'_p(g; f) \geqq 0$ and we can choose $u'_p(g) > u'_p(f) \gg 0$) we conclude that $\sigma_i(f) > 0 \Rightarrow \sigma'_p(g) > 0$; so (cf. Assumption GA) $\sigma'_p(g) > 0$ and

$$Q'^*_{pp}(g) = \lim_{n \to \infty} n^{-1} \sum_{m=0}^{n-1} (\sigma'_p(g))^{-1}\, Q'_{pp}(g))^m$$

is well defined.

On premultiplying (3.3.6) by $Q'^*_{pp}(g)$ (cf. Lemma 2.2, part (3)) we immediately obtain

$$(\sigma'_p(g) - \sigma_i(f))\, Q'^*_{pp}(g)\, u'_p(f) = Q'^*_{pp}(g)\, \gamma'_p(g; f)\,. \tag{3.3.7}$$

As $u'_p(f) \gg 0$, $Q'^{*}_{pp}(g) > 0$ and $\gamma'_p(g;f) > 0$ (3.3.5) must be true and, consequently, for any $j \in I$

$$\sigma(g,j) \geqq \sigma(f,j)\,. \tag{3.3.8}$$

To finish the proof it only suffices to establish that for all j

$$\sigma(g,j) = \sigma(f,j) \Rightarrow \nu(g,j) \geqq \nu(f,j)\,. \tag{3.3.9}$$

To this order without loosing generality we may suppose that $j \in I_i(f) \equiv I'$ and that (3.3.9) will hold for any $j \in I_k(f)$ with $k > i$ (if possible). Furthermore, observe that if $\sigma(g,j) = \sigma(f,j)$ for some $j \in I'$ then (cf. (3.3.2), (3.3.5))

$$\sigma'_p(g) = \sigma_i(f) \tag{3.3.5'}$$

and (cf. definition of $\sigma(g,j)$ and part (2) of Lemma 2.2) at least one basic class of $Q'_{pp}(g)$ is not contained in some other irreducible class of $Q(g)$ as well as it is not accessible to any other irreducible class of $Q(g)$ whose spectral radius is greater then $\sigma_i(f)$.

Now let $\bar{I}'_p = \{j \in I' : j$ belongs to some basic class of $Q'_{pp}(g)\}$. If (3.3.5') holds then by part (3) of Lemma 2.2 from (3.3.7) we conclude that $[\gamma'_p(g;f)]_j = 0$ for any $j \in \bar{I}'_p$ and, consequently, by (3.3')

$$g(j) = f(j) \quad \text{for any} \quad j \in \bar{I}'_p\,. \tag{3.3.10}$$

So if (3.3.9) holds for each $j \in I_k(f)$ with $k > i$ by (3.3.10) we conclude that (3.3.9) must hold for any $j \in \bar{I}'_p$ and taking into account the structure of $Q'(g)$ (cf. (3.3.4)) (3.3.9) must hold for all $j \in I_i(f)$. □

Lemma 3.4. Let for decision vectors $f, g \in F$ $(f \neq g)$ and any $j \in I$ either

$$g(j) = f(j) \tag{3.4.1}$$

or

$$[Q(g)]_{jk} > 0 \tag{3.4.2}$$

for certain k with

$$(\sigma(f,k);\ \nu(f,k)) \succ (\sigma(f,j);\ \nu(f,j))\,. \tag{3.4.2'}$$

Then for any $j \in I$

$$(\sigma(g,j);\ \nu(g,j)) \succeq (\sigma(f,j);\ \nu(f,j)) \tag{3.4.3}$$

with a strict inequality at least for one $j \in I$.

Proof. First let the rows and corresponding columns of $Q(f)$ be permuted so that (2.1.1), (2.1.2) and (2.1.3′) hold and let us suppose that for some fixed $i = 1, \ldots, s$

(3.4.4) $$g(j) \neq f(j) \quad \text{for (possibly several)} \quad j \in I_i(f)\,, \quad \text{and}$$
$$g(j) = f(j) \quad \text{for any} \quad j \in I \setminus I_i(f)\,.$$

Obviously for all $j \in I_l(f)$ with $l > i$ $(\sigma(g, j);\ \nu(g, j)) = (\sigma(f, j);\ \nu(f, j))$ and by the definition of $\sigma(f, j)$, structure of class $Q_{ii}(f)$ and (3.4.2), (3.4.2′) also for any $j \in I_l(f)$ with $l \leqq i$

(3.4.5) $$\sigma(g, j) \geqq \sigma(f, j)\,.$$

To establish (3.4.3) it suffices to show that for any $j \in I_l(f)$ with $l \leqq i$

(3.4.6) $$\sigma(g, j) = \sigma(f, j) \Rightarrow \nu(g, j) \geqq \nu(f, j)$$

and in virtue of (3.4.4), (3.4.5) and the definition of $\nu(f, j)$ it suffices to verify (3.4.6) only for $j \in I_i(f)$. To this order let for

(3.4.7) $$\nu(g, j_0) = \min_{j \in I_i(f)} \{\nu(g, j) : \sigma(g, j) = \sigma(f, j)\}$$

hold

(3.4.7′) $$\nu(g, j_0) < \nu_i(f)\,.$$

Recalling the construction of $g \in F$ by (3.4.7), (3.4.7′) and definition of $\nu(g, j)$ there exists $l < i$ and $k_0 \in I_l(f)$, $j_1 \in I_i(f)$ (where under $g \in F$ j_0, resp. k_0, is accessible to k_0, resp. j_1) such that

(3.4.8) $$\sigma(g, j_0) = \sigma(g, k_0) = \sigma(g, j_1)$$

and

(3.4.8′) $$\nu(g, j_0) > \nu(g, k_0) > \nu(g, j_1)$$

that contradicts (3.4.7′); so (3.4.6) must hold. (3.4.5) together with (3.4.6) imply under condition (3.4.4) relation (3.4.3). Moreover, if (3.4.2), (3.4.2′) are satisfied for some $j \in I_i(f)$, $k \in I_l(f)$, say $j = j'$, $k = k'$, then by (3.4.3) written for $j = k'$ and the definition of $\sigma(f, j)$, $\nu(f, j)$ we immediately get

(3.4.9) $$(\sigma(g, j');\ \nu(g, j')) \succ (\sigma(f, j');\ \nu(f, j'))\,.$$

To finish the proof let us construct a (finite) sequence of decision vectors $\{g_n, n = 1, \ldots, s = s(f)\}$ with $g_1 \equiv f$ whose elements for $n > 1$ are defined by

(3.4.10) $$g_n(j) = g(j) \qquad \text{for all} \quad j \in I_n(f)$$

(3.4.10′) $$g_n(j) = g_{n-1}(j) \quad \text{for any} \quad j \in I \setminus I_n(f)$$

(so $g_s \equiv g$) and apply successively the above reasoning. □

 Now we can present

Proof of Theorem 3.2. Applying the results of Lemmas 3.3 and 3.4 to the policy iteration procedure of Algorithm 1 we construct a sequence of decision vectors $\{f_n, n = 0, 1, \ldots\}$ such that for any $j \in I$

$$(\sigma(f_{n+1}, j);\ v(f_{n+1}, j)) \geqq (\sigma(f_n, j);\ v(f_n, j))\,. \tag{3.2.6}$$

So to establish that the elements of $\{f_n\}$ cannot recur it suffices to show that if $f_{n+1} \neq f_n$ and for all $j \in I$

$$(\sigma(f_{n+1}, j);\ v(f_{n+1}, j)) = (\sigma(f_n, j);\ v(f_n, j)) \tag{3.2.7}$$

then for all $i = 1, \ldots, s = s(f_n)$

$$I_i(f_n) = I_i(f_{n+1}) \tag{3.2.8}$$

$$\bar{I}_i(f_{n+1}) \subset \bar{I}_i(f_n) \tag{3.2.8'}$$

together with (cf. (2.1.2))

$$u_i(f_{n+1}) \geqq u_i(f_n) \quad \text{and} \quad u(f_{n+1}) > u(f_n) \tag{3.2.9}$$

where for any $j \in \bar{I}_i(f_{n+1})$

$$[u_i(f_{n+1})]_j = [u_i(f_n)]_j \tag{3.2.9'}$$

(recall that if (3.2.8'), (3.2.9') hold then $u_i(f_{n+1})$ is unique).

Now (3.2.8) follows immediately by (3.2.7) and definition of $I_i(f)$; to verify (3.2.8') observe that by (3.2.8)

$$\sigma_i(f)\, u_i(f) = Q_{ii}(f)\, u_i(f) \tag{3.2.10}$$

will hold for $f = f_n, f_{n+1}$ with $\sigma_i(f_{n+1}) = \sigma_i(f_n)$. Using similar way of reasoning as in the proof of Lemma 3.3 by (3.2.10) written for $f = f_n, f_{n+1}$ we get

$$\sigma_i(f_{n+1})\,(u_i(f_{n+1}) - u_i(f_n)) = \\ = Q_{ii}(f_{n+1})\,(u_i(f_{n+1}) - u_i(f_n)) + \gamma_i(f_{n+1}; f_n) \tag{3.2.11}$$

where (observe that under condition (3.2.7) no improvement according to part (i) of Algorithm 1 can be performed)

$$\gamma_i(f_{n+1}; f_n) = (Q_{ii}(f_{n+1}) - Q_{ii}(f_n))\, u_i(f_n) \geqq 0 \tag{3.2.11'}$$

with a strict inequality at least for one i. On premultiplying (3.2.11') by $Q^*_{ii}(f_{n+1})$ (cf. part (3) of Lemma 2.2) we get

$$Q^*_{ii}(f_{n+1})\, \gamma_i(f_{n+1}; f_n) = 0 \Rightarrow [\gamma_i(f_{n+1}; f_n]_j = 0 \quad \text{for any} \quad j \in \bar{I}_i(f_{n+1})\,. \tag{3.2.12}$$

Consequently by (3.3′) and (3.2.12) for each $j \in \bar{I}_i(f_{n+1})$ $f_{n+1}(j) = f_n(j)$ that establishes (3.2.8′) and (3.2.9′). Now on applying part (5) of Lemma 2.2 to (3.2.11) by (3.2.9′) also (3.2.9) must hold.

So the elements of a sequence $\{f_n\}$ cannot recur and as there exists a finite number of decision vectors the sequence of successively improved decision vectors $\{f_n, n = 0, 1, \ldots\}$ must be finite (with $f_r \equiv \hat{f}$ being its last element).

As for $\hat{f} \equiv f_r$ no further improvement according to part (i) of Algorithm 1 is possible if (by suitable permuting rows and corresponding columns of $Q(\hat{f})$) $Q(\hat{f})$ is an upper block-triangular matrix and the same must also be true for any $Q(f)$. So (3.2.2), (3.2.3) and (3.2.5) must hold and if $Q'(f)$ denotes some irreducible class of $Q(f)$ all the elements of $Q'(f)$ must be labelled by integers from some (fixed) $I_i(\hat{f})$. As for $\hat{f}$ no improvement according to part (ii) of Algorithm 1 is possible, also (3.2.4) holds.

To finish the proof we only need to establish (3.2.1). To this order observe that for any $Q(f)$ and each $i = 1, \ldots, s = s(\hat{f})$ there exists $Q'_{ii}(f) = Q_{ii}(f) + H_{ii}(f)$ (where $H_{ii}(f) \geqq 0$) such that by (3.2.4) $Q'_{ii}(f)\, u_i(\hat{f}) = \sigma_i(\hat{f})\, u_i(\hat{f})$ with $u_i(\hat{f}) \gg 0$. Then by Perron - Frobenius theorem and part (2) of Lemma 2.2 $\sigma_i(f) \leqq \sigma_i(\hat{f})$ and if $\sigma_i(f) = \sigma_i(\hat{f})$ then each basic class of $Q_{ii}(f)$ cannot be accessible to any other irreducible class of $Q_{ii}(f)$. These facts together with the definition of $\sigma(f, j)$, $\nu(f, j)$ immediately imply (3.2.1). □

We shall finish this section by a simple but useful remark extending slightly similar results for stochastic matrices in [15] (cf. Lemma 2.3 in [15]). These facts are presented only for completeness; they will not be used anywhere in the sequel.

Recall that $\bar{I}_i(\hat{f})$ denotes all $j \in I_i(\hat{f})$ belonging to some basic class of $Q_{ii}(\hat{f})$ and let for any $f \in F$

$$\bar{I}_i(\hat{f}; f) \equiv \{j \in I_i(\hat{f}) : j \quad \text{belongs to some basic class of} \quad Q_{ii}(f)\}$$

(here $Q_{ii}(f)$ contains exactly the elements labelled by integers from $I_i(\hat{f})$). Similarly, let

$$\bar{I}_i^0(\hat{f}; f) \equiv \{j \in \bar{I}_i(\hat{f}; f) : \sigma_i(f) = \sigma_i(\hat{f})\}$$

and

$$\bar{I}_i^0(\hat{f}) = \{j \in I_i(\hat{f}) : \exists f \in F \quad \text{with} \quad \sigma_i(f) = \sigma_i(\hat{f}) \quad \text{and} \quad j \in \bar{I}_i(\hat{f}; f)\}$$

(obviously $\bar{I}_i^0(\hat{f}; f) \subset \bar{I}_i^0(\hat{f}) \neq \emptyset$ and $\sigma_i(f) = \sigma_i(\hat{f}) \Rightarrow \bar{I}_i(\hat{f}; f) \equiv \bar{I}_i^0(\hat{f}; f)$).

Remark 3.5. There need not exist any $f \in F$ such that $\bar{I}_i^0(\hat{f}) = \bar{I}_i^0(\hat{f}; f)$.
For example, let $I = \{1, 2, 3\}$, $F = \{f_1, f_2\}$ with

$$Q(f_1) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad Q(f_2) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix};$$

so (cf. Theorem 3.2) $s = 1, \hat{f} = f_1, f_2$ with $\sigma(\hat{f}) = \sigma(f_1) = 1, \nu_1(\hat{f}) = 1$ and $\bar{I}_1^0(\hat{f}) = I$, but $\bar{I}_1^0(\hat{f}, f_1) \neq \bar{I}_1^0(\hat{f}), \bar{I}_1^0(\hat{f}; f_2) \neq \bar{I}_1^0(\hat{f})$.

However by the definition of $\bar{I}_i^0(\hat{f})$ there exists $f_1, f_2, \ldots, f_p$ (p finite) such that $j \in \bar{I}_i^0(\hat{f}) \Rightarrow j \in \bar{I}_i^0(\hat{f}; f_{k_j})$ for (possibly several) $k_j = 1, \ldots, p$. Now let us construct matrix Q_{ii} such that (cf. (3.2.4))

$$[Q_{ii}]_j \begin{cases} = [Q_{ii}(\hat{f})]_j & \text{for any } j \in I_i(\hat{f}) \setminus \bar{I}_i^0(\hat{f}) \\ = \sum_{m=1}^{p} \alpha_{jm} [Q_{ii}(f_m)]_j & \text{for any } j \in \bar{I}_i^0(\hat{f}) \end{cases} \tag{3.5.1}$$

where for real numbers $\alpha_{jm} \geqq 0$ $(m = 1, \ldots, p)$

$$\alpha_{jm} \neq 0 \Leftrightarrow j \in \bar{I}_i^0(\hat{f}; f_m), \quad \text{and} \tag{3.5.1'}$$

$$\sum_{m=1}^{p} \alpha_{jm} = 1 . \tag{3.5.1''}$$

Moreover, let for any $j \in \bar{I}_i^0(\hat{f}; f)$

$$\bar{I}_{ij}^0(\hat{f}; f) \equiv \{k \in \bar{I}_i^0(\hat{f}; f) : j, k \text{ belong to the same basic class of } Q_{ii}(f)\}$$

and observe that by (3.2.4) and Perron - Frobenius theorem (we abbreviate $\sigma_i(\hat{f})$, resp. $u_i(\hat{f})$, by σ_i, resp. u_i)

$$j \in \bar{I}_i^0(\hat{f}; f) \Leftrightarrow [Q_{ii}(f)\, u_i]_k = \sigma_i [u_i]_k\,, \quad [Q_{ii}(f)]_{kl} = 0 \quad \text{for } \forall\, k \in \bar{I}_{ij}^0(\hat{f}; f)\,; \ \forall\, l \in I_i(\hat{f}) \setminus \bar{I}_{ij}^0(\hat{f}; f)\,. \tag{3.5.2}$$

Now by (3.5.1), (3.5.1′), (3.5.1″), (3.5.2) we get for any $f \in F$

$$Q_{ii} u_i = \sigma_i u_i \tag{3.5.2'}$$

where each $j \in \bar{I}_i^0(f)$ belongs to some irreducible class of Q_{ii} that is not accessible to any other irreducible class of Q_{ii} (so by (3.5.2) this class must be basic of Q_{ii}).

As Q_{ii} is obtained by a suitable convex combination of $Q_{ii}(f_m)$ with $f_m \in F$ (cf. (3.5.1), (3.5.1″)) we can equivalently say that there exists "randomized decision vector", say $\hat{f}_{rd}$, (selecting each $f_m \in F$ with given probability), such that for $Q_{ii}(\hat{f}_{rd})$

$$Q_{ii}(f)\, u_i \leqq Q_{ii}(\hat{f}_{rd})\, u_i = \sigma_i u_i \tag{3.5.2'}$$

and moreover $\bar{I}_i^0(\hat{f}) \equiv \bar{I}_i^0(\hat{f}; \hat{f}_{rd})$.

4. BOUNDS ON DYNAMIC PROGRAMMING RECURSIONS

In this section we establish some bounds on the utility vector $x(n)$ calculated from dynamic programming recursion (1.1). These results will also provide some bounds on non-homogeneous matrix products. Our approach heavily depends on the existence of a decision vector $\hat{f}$ maximizing (lexicographically) the pair $(\sigma(f, j); v(f, j))$ for any $j \in I$. Recall that by Theorem 3.2 such a decision vector, say $\hat{f}$, can be found by policy iterations (Algorithm 1). Moreover, in virtue of Theorem 3.2 (by possibly permuting rows and corresponding columns) $Q(\hat{f})$ is "upper block-triangular" (cf. (3.2.2)) and on using the same partition also for each $f \in F$

$$(4.1) \qquad Q(f) = \begin{bmatrix} Q_{11}(f) & Q_{12}(f) \dots & Q_{1s}(f) \\ & Q_{22}(f) \dots & Q_{2s}(f) \\ & \ddots & \vdots \\ & & Q_{ss}(f) \end{bmatrix}$$

where $s = s(\hat{f})$, each $Q_{ii}(f)$ contains the elements labelled by integers from $I_i(\hat{f})$ and $Q_{ik}(f) = 0$ for any $k < i$.

Remember that the partition of $Q(f)$ given by (4.1) will be currently used throughout this section. Similarly for any matrix, say C, symbol C_{mn} denotes a submatrix of C such that for its arbitrary entry, say $[C_{mn}]_{jk}$, we have $j \in I_m(\hat{f})$, $k \in I_n(\hat{f})$. The same convention will be also used for vectors. $\sigma_i(f)$ denotes the spectral radius of $Q_{ii}(f)$.

To simplify the notations we set $u \equiv u(\hat{f})$, $\sigma_i \equiv \sigma_i(\hat{f})$, $v_i \equiv v_i(\hat{f})$ and so (3.2.3) reads

$$(4.2) \qquad Q_{ii}(f)\, u_i \leqq Q_{ii}(\hat{f})\, u_i = \sigma_i u_i$$

(here u_i is a subvector of $u \gg 0$ whose elements belong to $I_i(\hat{f})$). Similarly, throughout this section $x_i(n)$ will denote a subvector of $x(n)$ (calculated from dynamic programming recursion (1.1)) whose components belong to $I_i(\hat{f})$.

Using an analogy with Lemma 3.1 the above facts will enable to produce some bounds on $\{x(n)\}$. It holds:

Theorem 4.1. For given $x(0) > 0$ there exists $u \gg 0$ (where for $i = 1, \dots, s = s(\hat{f})$ u_i satisfy (4.2)) such that

$$(4.1.1) \qquad x_i(n) \leqq \sigma_i^n \binom{n + v_i - 1}{v_i - 1} u_i .$$

Proof. Let us set for $i = 1, \dots, s$ $K_{ii} = Q_{ii}(\hat{f})$ and $K_{ij} = 0$ for any $j < i$. Furthermore, on using the decomposition according to (4.1), let us construct for any pair $j > i$ an auxiliary matrix M_{ij} such that $M_{ij} \geqq \max_{f \in F} Q_{ij}(f)$ (here symbol max is to be considered componentwise).

Using the same reasoning as in the proof of Lemma 3.1, on the base of M_{ij}'s we can choose for any pair $j > i$ vectors $u'_i \gg 0$ (such that $u'_i \geqq x_i(0)$) satisfying (4.2) and matrices K_{ij}'s in such a way that for any $f \in F$ $K_{ij} \geqq M_{ij} \geqq Q_{ij}(f)$ and

$$(4.1.2) \qquad Q_{ij}(f)\, u'_j \leqq K_{ij} u'_j = \sigma_j u'_i .$$

The proof can be accomplished by mimicking the steps of the proof of Lemma 3.1 if we replace (3.1.2) by (4.1.2) (observe that (3.1.3) is replaced by

$$(4.1.3)\qquad x_i(n) \leqq u_i' \sum_{m_i+\ldots+m_s=n} \sigma_i^{m_i} \sigma_{i+1}^{m_{i+1}} \ldots \sigma_s^{m_s} .) \qquad \square$$

In case that $s(\hat{f}) = 1$ the bounds on $x(n)$ established in Theorem 4.1 have a more simple form. As for $s(\hat{f}) = 1$ by (4.2) there exists $u \gg 0$ such that for any $f \in F$

$$(4.2')\qquad Q(f)\, u \leqq Q(\hat{f})\, u = \sigma u$$

Theorem 4.1 yields the following corollary (recall that (4.2') trivially holds if $Q(f)$ can be found irreducible).

Corollary 4.2. Let (4.2') hold for any $f \in F$ and let $x(0) \leqq u$. Then for all $n = 0, 1, \ldots$.

$$(4.2.1)\qquad x(n) \leqq \sigma^n u .$$

Now let us try to establish some bounds on $\prod_{m=0}^{n} Q(f_m)$ where $\{f_m, m = 0, 1, \ldots\}$ is an arbitrary sequence of decision vectors. To this order let us assume $x(0) \gg 0$, iterate (1.1) to get

$$(4.3)\qquad x(n) \geqq \prod_{m=0}^{n-1} Q(f_m)\, x(0)$$

and observe that then

$$x(0) \gg 0 \Rightarrow [\prod_{m=0}^{n-1} Q(f_m)]_{kl} \leqq \frac{[x(n)]_k}{[x(0)]_l} .$$

Then Theorem 4.1 implies some bounds on $\prod_{m=0}^{n} Q(f_m)$ that slightly extend similar results in [14]. These facts can be summarized as

Corollary 4.3. Let

$$^iQ(f) = \begin{bmatrix} Q_{ii}(f) & Q_{i,i+1}(f) & \ldots & Q_{is}(f) \\ & Q_{i+1,i+1}(f) & \ldots & Q_{i+1,s}(f) \\ & & \ddots & \vdots \\ & & & Q_{ss}(f) \end{bmatrix} .$$

Then for an arbitrary sequence of decision vectors, say $\{f_m, m = 0, 1, \ldots\}$,

$$(4.3.1)\qquad \sigma_i^{-n} \binom{n + \nu_i - 1}{\nu_i - 1} \prod_{m=0}^{n-1} {}^iQ(f_m)$$

must be uniformly bounded in n and, in particular, (cf. also [14])

$$(4.3.1') \qquad \sigma_1^{-n} n^{-\nu_1+1} \prod_{m=0}^{n-1} Q(f_m)$$

is uniformly bounded in n.

We shall finish this section by a useful remark to Corollary 4.3 concerning non-homogeneous products of non-negative matrices.

Remark 4.4. Let $\{Q^{(k)}, k = 1, \ldots, p\}$ be a (finite) set of $N \times N$ non-negative matrices and let for $n = 1, 2, \ldots$.

$$M(n) = Q^{(k_1)} Q^{(k_2)} \ldots Q^{(k_n)} \quad \text{where integers} \quad k_j \in \langle 1; p \rangle .$$

Define decision vectors f_k $(k = 1, \ldots, p)$ such that $Q(f_k) = Q^{(k)}$ and let F be the minimal set of decision vectors containing all f_k $(k = 1, \ldots, p)$ and possessing the "product property" mentioned in the Introduction. Then the sequence $\{\sigma^{-n} n^{-\nu+1} . M(n), n = 1, 2, \ldots\}$ is uniformly bounded in n (here $\sigma = \sigma_1(\hat{f})$, $\nu = \nu_1(\hat{f})$ satisfy conditions of Theorem 3.2 with respect to matrices $Q(f), f \in F$).

5. CONCLUSION AND DISCUSSION

The present paper deals with some properties of the utility vector $x(n)$ calculated from dynamic programming recursion (1.1). We have suggested a policy iteration method (cf. Algorithm 1) for finding a decision vector $\hat{f} \in F$ maximizing the growth of the utility vector in the class of stationary policies. Theorem 3.2 shows that the block-triangular matrix $Q(\hat{f})$ plays a dominant role for the set of matrices $\{Q(f), f \in F\}$ and extends some of the respective results known only for irreducible matrices (cf. [1], [3], [7], [9]). On the base of Theorem 3.2 it is not too difficult to find some bounds on the growth of the utility vector $x(n)$ (cf. Theorem 4.1 and Corollary 4.2). This extends corresponding results of [9], [10] and [14]. The obtained bounds on $x(n)$ enable also to produce some bounds on non-homogeneous products of non-negative matrices as it is shown in Corollary 4.3 and Remark 4.4.

The pioneer work in this direction and most of the interpretations of this dynamic programming problem (mentioned in the Introduction) is due to Bellman (cf. [1], [3]) but Bellman himself restricted his analysis to the "easy" case with $Q(f) \gg 0$ for each $f \in F$.

As it was shown in this paper (cf. Example I of Section 1) our model also includes classical Markov decision chains and the problem concerning the bounds on the vector of maximum total expected rewards $v(n)$ (calculated from the respective dynamic programming recursion) was intensively studied and solved in the dynamic programming literature (cf. [2], [4], [6], [8], [15]). Observe that in this case conditions of Theorem 3.2 are trivially fulfilled and Theorem 4.1 "only states" that for transient dynamic programming $v(n)$ is uniformly bounded in n (of course, the same must be also true for discounted dynamic programming) and that in the undiscounted dynamic programming model $v(n)$ can grow at most linearly with n. However, in the second part of this paper (cf. [18]) we shall refine the obtained bounds on $x(n)$. In particular, these refined results applied on classical Markov decision chains immediately imply the well-known bounds on $v(n)$ that were originally obtained in [4] and [8].

In the overwhelming dynamic programming literature considerably less attention was devoted to the general form of dynamic programming recursion (1.1). Besides the work of Bellman let us mention [9] (convergence radius of (1.1) if all $Q(f)$'s are irreducible; the obtained results correspond to those of Corollary 4.2), [10] (convergence radius of (1.1) for $Q(f)$ reducible; using the denotation of Theorem 4.1 in fact it was shown there that for any $\sigma < \sigma_i$ $\lim_{n\to\infty} \sigma^{-n} x_i(n) = 0$), [7] (formulation of the functional equation for "risk-sensitive" Markov decision chains, cf. Example II in the Introduction, and a policy iteration method for maximizing $\sigma(f)$ if all $Q(f)$'s are irreducible – the same policy iteration method was already mentioned in [9]) and [16], [17] (asymptotic properties of (1.1) under some specific assumptions).

Some properties of dynamic programming recursion (1.1) are closely connected with so called multiplicative Markov decision chains; i.e., Markov decision chains where the transition probability matrices are replaced by general non-negative matrices. We have already mentioned transient dynamic programming introduced in [20]. Moreover, on extending the methods used in [20] primarily for classical Markov decision chains, Rothblum (cf. [13], [14]) obtained interesting results (especially with respect to various sensitive optimality criteria) for the general case of multiplicative Markov decision chains. In particular, boundedness of (4.3.1′) was originally established by the methods different of ours in Section 3 of [14].

The material of this paper is a slightly revised form of the first three chapters of [19] and the paper itself presents some natural extensions of the corresponding results in [17]. In [17] dynamic programming recursion (1.1) is investigated under the condition that for any $f \in F$ $\sigma_1(f) > \sigma_2(f) > \ldots > \sigma_s(f)$, each $Q_{ii}(f)$ contains only one basic class and comparing with Section 3 of the presented paper no algorithm for finding $Q(f)$ is given.

Recently, using the methods different of ours, Zijms (cf. [21]) also showed the existence of a block-triangular matrix $Q(\hat{f})$ satisfying (3.2.3), (3.2.4) and in [22] presented a (different) proof of our Theorem 4.1.

(Received April 11, 1980.)

REFERENCES

[1] R. Bellman: On a Quasi-Linear Equation. Canad. J. Math. *8* (1956), 198–202.

[2] R. Bellman: A Markovian Decision Process. J. Math. Mech. *6* (1957), 670–684.

[3] R. Bellman: Introduction to Matrix Analysis. McGraw Hill, New York 1960. Second edition 1970.

[4] B. W. Brown: On the Iterative Method of Dynamic Programming in a Finite Space Discrete Time Markov Processes. Ann. Math. Stat. *36* (1965), 4, 1279—1285.
[5] F. R. Gantmakher: Teoriya matric. Second edition. Nauka, Moscow 1966.
[6] R. A. Howard: Dynamic Programming and Markov Processes. Technology Press and Wiley Press, New York 1960.
[7] R. A. Howard, J. E. Matheson: Risk-sensitive Markov Decision Processes. Manag. Sci. *18* (1972), 7, 357—369.
[8] E. Lanéry: Étude asymptotique des systèmes Markoviens à commande. Rev. Franc. Inf. Rech. Opér. *3* (1967), 1, 3—56.
[9] P. Mandl, E. Seneta: The Theory of Non-Negative Matrices in a Dynamic Programming Problem. Austral. J. Stat. *11* (1969), 2, 85—96.
[10] P. Mandl: Decomposable Non-Negative Matrices in a Dynamic Programming Problem. Czech. Math. J. *20* (1970), 3, 504—510.
[11] A. M. Ostrowski: Solution of Equations and System of Equations. Academic Press, New York 1960.
[12] U. G. Rothblum: Algebraic Eigenspaces of Nonnegative Matrices. Linear Algebra Applic. *12* (1975), 281—292.
[13] U. G. Rothblum: Multiplicative Markov Decision Chains. PhD Dissertation, Dept. Oper. Res., Stanford Calif. 1974.
[14] U. G. Rothblum: Normalized Markov Decision Chains II — Optimality of Nonstationary Policies. SIAM J. Control. Optim. *15* (1977), 2, 221—232.
[15] P. J. Schweitzer, A. Federgruen: The Asymptotic Behavior of Undiscounted Value Iteration in Markov Decision Problems. Mathem. Oper. Res. *2* (1977), 4, 360—381.
[16] K. Sladký: On Dynamic Programming Recursion for Multiplicative Markov Decision Chains. Math. Progr. Study *6* (1976), 216—226.
[17] K. Sladký: Successive Approximation Methods for Dynamic Programming Models. In: Proceedings of the Third Formator Symposium on the Analysis of Large-Scale Systems. (L. Bakule, J. Beneš, eds.) Academia, Prague 1979.
[18] K. Sladký: Bounds on Discrete Dynamic Programming Recursions II — Polynomial Bounds on Problems with Block-Triangular Structure. Submitted to Kybernetika.
[19] K. Sladký: On Functional Equations of Discrete Dynamic Programming with Non-Negative Matrices. Research Report No. 900, Institute of Information Theory and Automation, Prague 1978.
[20] A. F. Veinott, Jr.: Discrete Dynamic Programming with Sensitive Discount Optimality Criteria. Ann. Math. Stat. *40* (1969), 5, 1635—1660.
[21] W. H. M. Zijms: On Nonnegative Matrices in Dynamic Programming I. Memorandum Cosor 79—10, Eindhoven University of Technology, Eindhoven 1979.
[22] W. H. M. Zijms: Maximizing the Growth of the Utility Vector in a Dynamic Programming Model. Memorandum Cosor 79—11, Eindhoven University of Technology, Eindhoven 1979.

Ing. Karel Sladký, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.