Karel Sladký

Bounds on discrete dynamic programming recursions. II. Polynomial bounds on problems with block-triangular structure

Kybernetika, Vol. 17 (1981), No. 4, 310--328

Persistent URL: http://dml.cz/dmlcz/124353

KYBERNETIKA — VOLUME 17 (1981), NUMBER 4

BOUNDS ON DISCRETE DYNAMIC PROGRAMMING RECURSIONS II

Polynomial Bounds on Problems with Block-Triangular Structure

KAREL SLADKÝ

For dynamic programming models being of a specific block-triangular structure some polynomial bounds on utility vector $x(n)$ are established. Here $x(n)$ is calculated from dynamic programming recursion $x(n+1) = \max Q(f)\, x(n)$, where $Q(f)$ is a (not necessarily nonnegative) matrix having a specific structure and symbol max is considered with respect to the decision vector f taken from a finite set F. The presented results refine the bounds obtained in Part I and generalize some well-known results for classical Markov decision chains.

1. NOTATIONS AND PRELIMINARIES

Throughout the paper notations and terminology used in Part I of the present paper (cf. [7]) will be followed as close as possible.

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

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

Here $x(0) > 0$ is given, $Q(f)$ is an $N \times N$ matrix depending on a decision vector f (i.e. N-vector whose i-th component $f(i) \in F(i)$, $F(i)$ finite, specifies the decision in state i), and $F \equiv \bigtimes_{i=1}^{N} F(i)$ is a finite set of all decision vectors at each time point. Recall that the set F possesses an important "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 $i_1, i_2 \in I$ (here for any matrix C symbol $[C]_i$ denotes i-th row of C and similarly $[C]_{ij}$ is reserved for the ij-th element of C; the same notation is also used for vectors usually denoted by small letters).

In Part I of the present paper we have assumed that $Q(f) > 0$ for each $f \in F$ (i.e. $Q(f) \neq 0$ and each entry of $Q(f)$ nonnegative). Then on the base of Perron-Frobenius theorem we have shown that for any $f \in F$ the resulting matrix $Q(f)$ can be decomposed into an upper block-triangular form with specific properties. According to Lemma 2.1 of [7], by possibly permuting rows and corresponding columns, we can write for any $f \in F$

$$Q(f) = \begin{bmatrix} Q_{11}(f) & Q_{12}(f) & \dots & Q_{1s}(f) \\ 0 & Q_{22}(f) & \dots & Q_{2s}(f) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Q_{ss}(f) \end{bmatrix} \tag{1.2}$$

where for $i = 1, 2, \dots, s \equiv s(f)$ the elements of each diagonal submatrix $Q_{ii}(f)$ are labelled by integers from $I_i(f) \subset I$ (here $I = \bigcup_{i=1}^{s(f)} I_i(f)$ and $I_i(f) \cap I_k(f) = \emptyset$ for any $i \neq k$), and

$$Q_{ii}(f)\, u_i(f) = \sigma_i(f)\, u_i(f) \tag{1.2'}$$

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

$$(\sigma_1(f); \nu_1(f)) \succ (\sigma_2(f); \nu_2(f)) \succ \dots \succ (\sigma_s(f); \nu_s(f)) \tag{1.2''}$$

(symbol $\succ$ means lexicographically greater). $\nu_i(f)$ is the index of $Q_{ii}(f)$, i.e. the number of irreducible classes of $Q(f)$ having spectral radius $\sigma_i(f)$ and being subsequently accessible from any $j \in I_i(f)$. Observe that

$$\sigma_i(f) = \sigma_{i+1}(f) \;\Rightarrow\; \nu_i(f) = \nu_{i+1}(f) + 1 ,$$

and

$$\sigma_i(f) > \sigma_{i+1}(f) \quad \text{or} \quad i = s \;\Rightarrow\; \nu_i(f) = 1 .$$

Moreover, assuming only that for any $f \in F$ $\sigma_s(f) > 0$, in Part I of the present paper we have established for the set of nonnegative matrices $\{Q(f), f \in F\}$ the following result (cf. Theorem 3.2 in [7]) summarized as:

Proposition 1. There exists $\hat{f} \in F$ such that (by possibly permuting rows and corresponding columns) (1.2), (1.2′) and (1.2″) will hold for $f \equiv \hat{f}$, $s \equiv s(\hat{f})$. Moreover, using the same decomposition for any $Q(f), f \in F$; i.e. if

$$Q(f) = \begin{bmatrix} Q_{11}(f) & Q_{12}(f) & \dots & Q_{1s}(f) \\ 0 & Q_{22}(f) & \dots & Q_{2s}(f) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Q_{ss}(f) \end{bmatrix} \tag{1.3}$$

where $s \equiv s(\hat{f})$ and (for $i = 1, 2, \dots, s$) each $Q_{ii}(f)$ contains the elements labelled by integers from $I_i(\hat{f})$ (so for $[Q_{ik}(f)]_{jl}$ it holds $j \in I_i(\hat{f})$, $l \in I_k(\hat{f})$); then for any $i = 1, 2, \dots, s \equiv s(\hat{f})$ and any $f \in F$

$$Q_{ik}(f) = 0 \quad \text{for any} \quad k < i , \quad \text{and} \tag{1.3'}$$

$$Q_{ii}(f)\, u_i \leqq Q_{ii}(\hat{f})\, u_i = \sigma_i u_i \tag{1.4}$$

where $\sigma_i \equiv \sigma_i(\hat{f}) > 0$ and $u_i \equiv u_i(\hat{f}) \gg 0$ (similarly we abbreviate $\nu_i(\hat{f})$ by ν_i).

From now on, the same decomposition as that in Proposition 1 will be currently used for any $Q(f)$. Similarly, for any $N \times N$ square 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 often used for vectors. Writing a matrix relation symbol I, resp. e, will denote unit matrix, resp. unit vector, of an appropriate dimension.

In the further text we drop the assumption $Q(f) > 0$ and we shall only assume that each $Q(f)$ fulfils the assertions of Proposition 1. So throughout the paper we make:

Assumption A. There exists $\hat{f} \in F$ such that for any $Q(f)$ with $f \in F$ there exists a (fixed) decomposition by (1.3) satisfying conditions (1.3'), (1.4) such that $Q_{ii}(f) \geqq 0$ for any $f \in F$ and all $i = 1, \ldots, s \equiv s(\hat{f})$ with $Q_{ij}(f)$ not necessarily nonnegative for $j > i$. In particular, $Q(\hat{f})$ satisfies (1.2), (1.2'), (1.2") (for $f \equiv \hat{f}$ and $\sigma_i(\hat{f}) \equiv \sigma_i > 0$), and if $\sigma_{i+1}(\hat{f}) = \sigma_i(\hat{f})$ then each irreducible class of $Q_{ii}(\hat{f})$ is accessible to some basic class of $Q_{i+1,i+1}(\hat{f})$.

Remember that an irreducible class of $Q_{ii}(\hat{f})$, say $Q_{(jj)}(\hat{f})$ (here the decomposition of $Q(f)$ is according to (2.1) of [7] – symbol $Q_{(kk)}(f)$ is always reserved for an irreducible class of $Q(f)$), is accessible to an irreducible class of $Q_{i+1,i+1}(\hat{f})$, say $Q_{(ll)}(\hat{f})$, iff there exists a sequence of integers $k_0 = j < k_1 < \ldots < k_p = l$ such that $Q_{(k_{n-1},k_n)}(\hat{f}) \neq 0$ for any $n = 1, \ldots, p$ (observe that if $Q_{(jj)}(\hat{f})$ is a basic class of $Q_{ii}(\hat{f})$ then $Q_{(k_{n-1},k_n)}(\hat{f}) > 0$ for all $n > 1$).

Recall that $x_i(n)$ will denote a subvector of $x(n)$ (calculated from dynamic programming recursion (1.1)) whose components belong to $I_i(\hat{f})$ and sometimes it will be convenient to denote $F_i \equiv \bigtimes_{j \in I_i(\hat{f})} F(j)$, $\bar{x}_i(n) = \sigma_i^{-n} x_i(n)$ and introduce $\bar{Q}_{ij}(f) = \sigma_i^{-1} Q_{ij}(f)$ for any $f \in F$ and $i \leqq j$. Similarly, for $x(n; f)$ defined recursively by

$$(1.1') \qquad x(n+1; f) = Q(f)\, x(n; f) \quad \text{with} \quad x(0) = x(0; f)$$

(obviously, $x(n; f) = (Q(f))^n x(0)$), elements of each $x_i(n; f)$ $(i = 1, \ldots, s \equiv s(\hat{f}))$ are labelled by integers from $I_i(\hat{f})$ and $\bar{x}_i(n; f) = \sigma_i^{-n} x_i(n; f)$.

Observe that in virtue of (1.3), (1.3') of Proposition 1 it is possible to write (1.1) as

$$(1.5) \qquad \begin{bmatrix} x_1(n+1) \\ x_2(n+1) \\ \vdots \\ x_s(n+1) \end{bmatrix} = \begin{bmatrix} Q_{11}(f^{(n)}) & Q_{12}(f^{(n)}) & \ldots & Q_{1s}(f^{(n)}) \\ 0 & Q_{22}(f^{(n)}) & \ldots & Q_{2s}(f^{(n)}) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & Q_{ss}(f^{(n)}) \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \\ \vdots \\ x_s(n) \end{bmatrix}.$$

On the base of the results obtained in [7] and summarized in Proposition 1 of the present paper, assuming $Q(f) > 0$ some bounds on $x_i(n)$ can be established (cf. [7], Theorem 4.1 and Corollary 4.2). It can be easily verified that, for the considered (not necessarily nonnegative) matrices of the set $\{Q(f), f \in F\}$ fulfilling Assumption A, on replacing the elements of non-diagonal submatrices $Q_{ij}(f)$ $(i = 1, \ldots, s; j > i)$

by their absolute values the upper bounds on $x_i(n)$ obtained in Theorem 4.1 of [7] remain valid. These facts can be summarized as:

Proposition 2. For each $i = 1, \ldots, s \equiv s(\hat{f})$ there exist vectors $u_i \gg 0$ (depending on $x(0)$ and satisfying (1.4)) such that for all $n = 0, 1, \ldots$

$$x_i(n) \leqq \sigma_i^n \binom{n + \nu_i - 1}{\nu_i - 1} u_i . \tag{1.6}$$

In particular, if $s(\hat{f}) = 1$ and $u_1 = u \geqq x(0)$, $\sigma \equiv \sigma(\hat{f})$, then

$$x(n) \leqq \sigma^n u . \tag{1.6'}$$

Remark. Observe that by (1.6) for any $\varrho > \sigma_i$ there exist vectors $c_i' \ll 0$, $c_i'' \gg 0$ (depending on ϱ and $x(0)$) such that

$$c_i'\varrho^n \leqq x_i(n) \leqq c_i''\varrho^n \quad \text{for all} \quad n = 0,1, \ldots \tag{1.7}$$

and, consequently, for any $\varrho > \sigma_i$

$$\lim_{n \to \infty} \varrho^{-n} x_i(n) = 0 . \tag{1.7'}$$

In particular, for any sequence of decision vectors, say $\{f_j \in F, j = 0, 1, \ldots\}$, and any $\varrho > \sigma(\hat{f})$ the elements of the "product matrix" $\prod_{j=0}^{n} Q(f_j)$ must grow slower than ϱ^n (cf. also Corollary 4.3 of [7] for the details) and $\lim_{n \to \infty} \varrho^{-n} x(n) = 0$.

In fact, Proposition 2 shows that the growth of $\{x_i(n)\}$ (and, of course, also of $\{x(n)\}$) can be given by an exponential as well as a polynomial part. In Section 3 we refine the polynomial bounds on $\{x_i(n)\}$ by establishing their specific properties for the case with $\nu_i > 1$. To this order in Section 2 we extend well-known policy iteration methods for finding an optimal policy of a classical Markov decision chain.

2. ON POLICY ITERATION METHODS

In this section we present a constructive proof of the existence of a solution to the set of (non-linear) equations of a specific type. The solution will be obtained by a policy iteration method that can be thought of as a generalization of classical policy iteration methods for Markov decision chains. These results will be very useful in the further text for the construction of bounding polynomials on $\{x(n)\}$.

It can be easily recognized that a special case of these equations (cf. assumptions of Theorem 2.1 – if $Q_{ii}(f)$ is a stochastic matrix and $c_i^{(l)}(f) \equiv 0$ for any $f \in F$ and any $l < r - 1$) was treated in the dynamic programming literature in connection with sensitive optimality criteria (cf. [6], [12]) and the case with $Q_{ii}(f)$ stochastic and $r = 1$ turns out to be well-known for Markov decision chains with average reward optimality criteria (cf. [4], and especially [5], [11]).

Theorem 2.1. Let Assumption A hold, i.e., (1.3), (1.3′) and (1.4) of Proposition 1 are satisfied. Recalling that $\bar{Q}_{ii}(f) = \sigma_i^{-1} Q_{ii}(f)$ by (1.4) we get for any $f \in F$

$$\bar{Q}_{ii}(f)\, u_i \leqq \bar{Q}_{ii}(\hat{f})\, u_i = u_i \quad \text{where} \quad u_i \gg 0\,, \tag{2.1.1}$$

and let $c_i^{(l)}(f)$ (for $l = r - 1, \ldots, 1, 0$) be given vectors. Then there exist vectors $u_i^{(l)}$ (for $l = r, r - 1, \ldots, 0$) and a nonincreasing sequence of (non-empty) sets of decision vectors

$$F_i \equiv \tilde{F}_i^{(r+1)} \supset \tilde{F}_i^{(r)} \supset \ldots \supset \tilde{F}_i^{(1)} \supset \tilde{F}_i^{(0)} \neq \emptyset \tag{2.1.2}$$

such that on setting

$$\tilde{\varphi}_i^{(r)}(f) = (\bar{Q}_{ii}(f) - I)\, u_i^{(r)} \tag{2.1.3}$$

and for $l = r - 1, \ldots, 1, 0$

$$\tilde{\varphi}_i^{(l)}(f) = (\bar{Q}_{ii}(f) - I)\, u_i^{(l)} - u_i^{(l+1)} + c_i^{(l)}(f)\,, \tag{2.1.4}$$

then for any $l = r, r - 1, \ldots, 0$ it holds:

$$\tilde{\varphi}_i^{(l)}(f) \leqq 0 \quad \text{for each} \quad f \in \tilde{F}_i^{(l+1)}\,, \tag{2.1.5}$$

where

$$\tilde{F}_i^{(l)} = \{f \in \tilde{F}_i^{(l+1)} : \tilde{\varphi}_i^{(l)}(f) = 0\}\,, \tag{2.1.6}$$

and there exists $\tilde{f} \in F_i$ such that

$$\tilde{\varphi}_i^{(l)}(\tilde{f}) = 0 \quad \text{for any} \quad l = r, r - 1, \ldots, 0\,. \tag{2.1.7}$$

We shall prove Theorem 2.1 in a sequence of several lemmas. To simplify the notations, in Lemmas 2.2, 2.4 we shall often delete index i (so e.g. we write $Q(f)$, $u(f)$, $c^{(l)}(f)$ instead of $Q_{ii}(f)$, $u_i(f)$, $c_i^{(l)}(f)$) and assume that $\sigma_i = 1$.

Lemma 2.2. Let (cf. (2.1.1)) for some (fixed) $Q(f) > 0$ and $u \gg 0$

$$Q(f)\, u \leqq u\,. \tag{2.2.1}$$

Then there exists $Q^*(f) = \lim_{n\to\infty} n^{-1} \sum_{m=0}^{n-1} (Q(f))^m$ together with $Z(f) = (I - Q(f) + Q^*(f))^{-1}$ and for $l = r, r - 1, \ldots, 0$

$$u^{(l)}(f) = Q^*(f)\, c^{(l-1)}(f) - \sum_{k=l}^{r-1} (-Z(f))^{1+k-l} (I - Q^*(f))\, c^{(k)}(f) \tag{2.2.2}$$

(we set $c^{(-1)}(f) \equiv 0$) is a solution to the set of equations

$$(Q(f) - I)\, u^{(r)}(f) = 0 \tag{2.2.3}$$

$$(Q(f) - I)\, u^{(l)}(f) - u^{(l+1)}(f) + c^{(l)}(f) = 0 \tag{2.2.4}$$

where $l = r - 1, \ldots, 0$.

Moreover, under the normalizing condition

$$(2.2.4') \qquad Q^*(f)\, u^{(0)}(f) = 0$$

$u^{(l)}(f)$'s given by (2.2.2) are the unique solution to (2.2.3), (2.2.4).

Proof. To show the existence of $Q^*(f)$ and $Z(f)$, let T be a diagonal matrix (with positive on-diagonal entries) satisfying $Te = u$. Then by (2.2.1) $P(f) = T^{-1}\, Q(f)\, T$ is a substochastic matrix; so by well-known limiting properties of substochastic matrices $Q^*(f)$ must exist. Recalling (cf. e.g. p. 721 of [2]) the existence and basic properties of fundamental matrices in Markov chain theory, we can also easily verify that under condition (2.2.1) $Z(f) = (I - Q(f) + Q^*(f))^{-1}$ always exists and

$$(2.2.5) \qquad (Q(f) - I)\, Z(f) = Q^*(f) - I\,; \quad Q^*(f)\, Z(f) = Z(f)\, Q^*(f) = Q^*(f)\,.$$

As $Q(f)\, Q^*(f) = Q^*(f)$ on employing (2.2.5) by a direct calculation we can verify that $u^{(l)}(f)$'s given by (2.2.2) are the solutions to (2.2.3), (2.2.4). To verify that for $l = r, \ldots, 0$ these $u^{(l)}(f)$'s are unique, in case that $l < r$ let us assume that for $l = r, \ldots, l' + 1$ $u^{(l)}(f)$ are unique and that (2.2.4), for $l = l'$, $l = l' - 1$, can be fulfilled both for $u' \equiv u^{(l)}(f)$, $u'' \equiv u^{(l)}(f)$. Then by (2.2.3) if $l' = r$ or by (2.2.4) for $l = l' < r$, resp. $l = l' - 1$ multiplied by $Q^*(f)$, we get

$$(2.2.6) \qquad (Q(f) - I)\,(u' - u'') = 0\,,$$

resp.

$$(2.2.6') \qquad Q^*(f)\,(u' - u'') = 0\,.$$

As $(I - Q(f) + Q^*(f))^{-1} = Z(f)$ always exists, by (2.2.6), (2.2.6') we can immediately conclude that $u' = u''$ (notice that if condition (2.2.4') is not assumed then $u^{(0)}(f)$ need not be unique). □

Remark 2.3. It can be easily verified that $u^{(l)}(f)$'s given by (2.2.2) can be calculated recursively for $l = r - 1, \ldots, 1, 0$ by

$$(2.3.1) \qquad u^{(l)}(f) = Q^*(f)\, c^{(l-1)}(f) + Z(f)\,(c^{(l)}(f) - u^{(l+1)}(f))$$

where

$$(2.3.1') \qquad u^{(r)}(f) = Q^*(f)\, c^{(r-1)}(f)\,.$$

Moreover, if $\sigma(f) < 1$ (cf. part (4) of Lemma 2.2 in [7]) then $Q^*(f) = 0$ and by (2.3.1') $u^{(r)}(f) = 0$, and (cf. (2.2.4), (2.3.1)) for $l = r - 1, \ldots, 0$

$$(2.3.2) \qquad u^{(l)}(f) = Z(f)\,(c^{(l)}(f) - u^{(l+1)}(f)) = -\sum_{k=l}^{r-1} (-Z(f))^{1+k-l}\, c^{(k)}(f)$$

where $Z(f) = (I - Q(f))^{-1} > 0$.

The following lemma is the main ingredient to the policy iteration algorithm for finding $u_i^{(l)}$'s such that (2.1.5), (2.1.6) and (2.1.7) are satisfied. Recall that if $Q(g) > 0$ and $Q(g)\, u \leqq u$ for some $u \gg 0$, then $Q^*(g) = \lim_{n\to\infty} n^{-1} \sum_{m=0}^{n-1} (Q(g))^m$ always exists, and that $u^{(l)}(f)$, $u^{(l)}(g)$ will denote the solutions of (2.2.3), (2.2.4), (2.2.4') (written for argument $g \in F$ if $u^{(l)}(g)$ is considered).

Lemma 2.4. Let (cf. (2.1.1)) for $f, g \in F$ with $f \neq g$, $Q(f) > 0$, $Q(g) > 0$, and some $u \gg 0$

$$Q(f)\, u \leqq u\,, \quad Q(g)\, u \leqq u\,, \tag{2.4.1}$$

and denote

$$\tilde{\varphi}^{(r)}(g; f) = (Q(g) - I)\, u^{(r)}(f) \tag{2.4.2}$$

$$\tilde{\varphi}^{(l)}(g; f) = (Q(g) - I)\, u^{(l)}(f) - u^{(l+1)}(f) + c^{(l)}(g) \tag{2.4.3}$$

where $l = r - 1, \ldots, 0$ (observe that $\tilde{\varphi}^{(l)}(f; f) = 0$ for any l).

If for some $m = 1, \ldots, r$

$$(\tilde{\varphi}^{(m)}(g; f);\ \tilde{\varphi}^{(m-1)}(g; f)) \succ 0 \quad \text{with} \quad \tilde{\varphi}^{(m)}(g; f) > 0\,, \tag{2.4.4}$$

and

$$\tilde{\varphi}^{(l)}(g; f) = 0 \quad \text{for all} \quad l = m + 1, \ldots, r\,, \tag{2.4.4'}$$

then

$$u^{(l)}(g) = u^{(l)}(f) \quad \text{for any} \quad l = m + 2, \ldots, r\,, \tag{2.4.5}$$

$$Q^*(g)\, \tilde{\varphi}^{(m)}(g; f) > 0 \ \Rightarrow\ u^{(m+1)}(g) > u^{(m+1)}(f)\,, \tag{2.4.6}$$

$$Q^*(g)\, \tilde{\varphi}^{(r)}(g; f) = 0\,, \tag{2.4.7}$$

$$Q^*(g)\, \tilde{\varphi}^{(m)}(g; f) = 0 \ \Rightarrow\ u^{(m+1)}(g) = u^{(m+1)}(g) \text{ and } u^{(m)}(g) > u^{(m)}(f)\,. \tag{2.4.8}$$

Proof. Let $\Delta u^{(l)} = u^{(l)}(g) - u^{(l)}(f)$. (2.2.3), (2.2.4) together with (2.4.2), (2.4.3) immediately imply for $l = r, \ldots, 0$ (we set $\Delta u^{(r+1)} = 0$, $c^{(r+1)}(f) \equiv 0$)

$$\begin{aligned} \Delta u^{(l+1)} &= (Q(g) - I)\, u^{(l)}(g) + c^{(l)}(g) - u^{(l+1)}(f) = \\ &= (Q(g) - I)\, \Delta u^{(l)} + \tilde{\varphi}^{(l)}(g; f)\,. \end{aligned} \tag{2.4.9}$$

Premultiplying (2.4.9) written for $l \doteq l - 1$ by $Q^*(g)$ (recall that $Q^*(g)\, Q(g)) = Q^*(g))$, for $l = r, \ldots, 1$ we get

$$Q^*(g)\, \Delta u^{(l)} = Q^*(g)\, \tilde{\varphi}^{(l-1)}(g; f)\,. \tag{2.4.9'}$$

By (2.4.9), (2.4.9') we conclude (remember that under (2.4.1) $Z(g) = (I - Q(g) + Q^*(g))^{-1}$ always exists, $Z(g)\, Q^*(g) = Q^*(g)$ and $\Delta u^{(r+1)} \equiv 0$) that for $l = r, \ldots, 1$

$$\Delta u^{(l)} = Z(g)\, [-\Delta u^{(l+1)} + \tilde{\varphi}^{(l)}(g; f)] + Q^*(g)\, \tilde{\varphi}^{(l-1)}(g; f)\,. \tag{2.4.10}$$

Now on applying (2.4.10) successively for $l = r, \ldots, m + 1$ we can verify (2.4.5). Similarly, (2.4.6) follows immediately by (2.4.10) written for $l = m + 1$, and (2.4.7) is nothing else than (2.4.9′) written for $l = r + 1$.

To verify (2.4.8), first observe that if (2.4.4), (2.4.4′) hold, then by (2.4.5) and (2.4.10) (written for $l = m + 1$) we get

$$(2.4.11) \quad Q^*(g)\,\tilde{\varphi}^{(m)}(g; f) = 0 \;\Rightarrow\; u^{(m+1)}(g) = u^{(m+1)}(f) \text{ and } Q^*(g)\,\tilde{\varphi}^{(m-1)}(g; f) \geqq 0\,.$$

Moreover, if (2.4.4), (2.4.4′) hold by a more detailed analysis we can show that

$$(2.4.12) \qquad Q^*(g)\,\tilde{\varphi}^{(m)}(g; f) = 0 \;\Rightarrow\; Z(g)\,\tilde{\varphi}^{(m)}(g; f) > 0\,.$$

To justify (2.4.12) let us decompose $Q(g)$ such that

$$Q(g) = \begin{bmatrix} Q_{\mathrm{TT}}(g) & Q_{\mathrm{TR}}(g) \\ 0 & Q_{\mathrm{RR}}(g) \end{bmatrix}, \quad \text{so} \quad Q^*(g) = \begin{bmatrix} 0 & Q^*_{\mathrm{TR}}(g) \\ 0 & Q^*_{\mathrm{RR}}(g) \end{bmatrix}.$$

Here $Q_{\mathrm{RR}}(g)$, resp. $Q_{\mathrm{TT}}(g)$, contains all irreducible classes of $Q(g)$ whose spectral radii equal 1, resp. are less than 1; of course, $Q_{\mathrm{TT}}(g)$ or $Q_{\mathrm{RR}}(g)$ may be empty. Using the same decomposition for $Z(g)$, resp. $\tilde{\varphi}(g; f)$, i.e. on writing

$$Z(g) = \begin{bmatrix} Z_{\mathrm{TT}}(g) & Z_{\mathrm{TR}}(g) \\ 0 & Z_{\mathrm{RR}}(g) \end{bmatrix}, \quad \text{resp.} \quad \tilde{\varphi}(g; f) = \begin{bmatrix} \tilde{\varphi}_{\mathrm{T}}(g; f) \\ \tilde{\varphi}_{\mathrm{R}}(g; f) \end{bmatrix},$$

we immediately conclude that

$$(2.4.13) \qquad Z_{\mathrm{TT}}(g) = (I - Q_{\mathrm{TT}}(g))^{-1} > 0$$

with positive diagonal elements. Now condition (2.4.4) (i.e. $\tilde{\varphi}^{(m)}(g; f) > 0$) immediately implies (recall that $Q^*_{\mathrm{RR}}(g) > 0$ with at least positive diagonal elements)

$$(2.4.14) \qquad Q^*(g)\,\tilde{\varphi}^{(m)}(g; f) = 0 \;\Rightarrow\; \tilde{\varphi}^{(m)}_{\mathrm{R}}(g; f) = 0 \text{ and } \tilde{\varphi}^{(m)}_{\mathrm{T}}(g; f) > 0\,.$$

Consequently, (2.4.12) follows immediately by (2.4.13) and (2.4.14).

Now it only remains to verify the last inequality of (2.4.8). To this order it suffices to use (2.4.10) for $l = m$ and apply (2.4.11), (2.4.12) to establish

$$(2.4.15) \quad Q^*(g)\,\tilde{\varphi}^{(m)}(g; f) = 0 \;\Rightarrow\; \Delta u^{(m)} = Z(g)\,\tilde{\varphi}^{(m)}(g; f) + Q^*(g)\,\tilde{\varphi}^{(m-1)}(g; f) > 0;$$

so (2.4.8) must hold. □

Now we are in a position to present the policy iteration algorithm establishing:

Proof of Theorem 2.1. Construct a (finite) sequence of decision vectors $f_0, f_1, \ldots, f_p \equiv \hat{f}$ with f_0 arbitrary and f_{n+1} being obtained by the following improvement of f_n:

For given $f_n \in F$ let us calculate $u_i^{(l)}(f_n)$ $(l = r, r - 1, \ldots, 0)$ being the solution to (2.2.3), (2.2.4) for $f \equiv f_n$ (by Lemma 2.2 such a solution always exists). So (cf. (2.3.1), (2.3.1′)) $u_i^{(l)}(f_n)$ can be obtained recursively for $l = r, r - 1, \ldots, 0$ by

$$(2.1.8) \qquad u_i^{(l)}(f_n) = \bar{Q}^*_{ii}(f_n)\, c_i^{(l-1)}(f_n) + Z_{ii}(f_n)\,(c_i^{(l)}(f_n) - u_i^{(l+1)}(f_n))$$

where

(2.1.8′) $$u_i^{(r)}(f_n) = \bar{Q}_{ii}^*(f_n)\, c_i^{(r-1)}(f_n)$$

and $Z_{ii}(\cdot) = (I - \bar{Q}_{ii}(\cdot) + \bar{Q}_{ii}^*(\cdot))^{-1}$ with $\bar{Q}_{ii}^*(\cdot) = \lim_{n\to\infty} n^{-1} \sum_{m=0}^{n-1} (\bar{Q}_{ii}(\cdot))^m$.

On the base of $u_i^{(l)}(f_n)$ we can perform an improvement of f_n; i.e., we can select $f_{n+1} \neq f_n$ (if possible) such that

(2.1.9) $$(\tilde{\varphi}_i^{(r)}(f_{n+1}; f_n); \ldots; \tilde{\varphi}_i^{(0)}(f_{n+1}; f_n)) \succ 0\,.$$

Usually we choose f_{n+1} such that for any $g \in F$

(2.1.9′) $$(\tilde{\varphi}_i^{(r)}(f_{n+1}; f_n); \ldots; \tilde{\varphi}_i^{(0)}(f_{n+1}; f_n)) \geqq (\tilde{\varphi}_i^{(r)}(g; f_n); \ldots; \tilde{\varphi}_i^{(0)}(g; f_n))$$

where (cf. (2.4.2), (2.4.3))

(2.1.10) $$\tilde{\varphi}_i^{(r)}(g; f) = (\bar{Q}_{ii}(g) - I)\, u_i^{(r)}(f)$$

and for $l < r$

(2.1.10′) $$\tilde{\varphi}_i^{(l)}(g; f) = (\bar{Q}_{ii}(g) - I)\, u_i^{(l)}(f) - u_i^{(l+1)}(f) + c_i^{(l)}(g)\,.$$

Now, on applying Lemma 2.4 we get for any $l \leqq r$

(2.1.11) $$\tilde{\varphi}_i^{(l)}(f_{n+1}; f_n) = 0 \quad (l = r, \ldots, m+1): \text{ together with}$$

$$\tilde{\varphi}_i^{(m)}(f_{n+1}; f_n) > 0 \Rightarrow (u_i^{(r)}(f_{n+1}); \ldots; u_i^{(m)}(f_{n+1})) \succ (u_i^{(r)}(f_n); \ldots; u_i^{(m)}(f_n))$$

and, consequently, the elements of a sequence $\{f_n\}$ cannot recur. As F is finite in a finite number of policy improvement steps we obtain $f_p \equiv \tilde{f}$ that cannot be further improved and $u_i^{(l)} \equiv u_i^{(l)}(\tilde{f})$ satisfy (2.1.5), (2.1.6) and (2.1.7) of Theorem 2.1. □

Remark 2.5. On the base of the policy iteration method used in the proof of Theorem 2.1 by the arguments of Lemma 2.4 we can easily verify that the decision vector $\tilde{f}$ maximizes lexicographically the matrix $[u_i^{(r)}(f), \ldots, u_i^{(0)}(f)]$; i.e. for any $f \in F_i$

(2.5.1) $$(u_i^{(r)}(\tilde{f}); \ldots; u_i^{(0)}(\tilde{f})) \geqq (u_i^{(r)}(f); \ldots; u_i^{(0)}(f))\,.$$

3. POLYNOMIAL BOUNDS ON UTILITY VECTOR

In this section we establish some polynomial bounds on the utility vector $x(n)$ generated by dynamic programming recursion (1.1). In particular, if for some $i = 1, \ldots, s \equiv s(\hat{f})$ $\sigma_i = \ldots = \sigma_{i+q} = \ldots = \sigma_{i+r}$ with $\sigma_{i+r} > \sigma_{i+r+1}$ or $i + r = s$ (so $v_i = r + 1$) and for any $j = i + q, \ldots, i + r$ $\sigma_i^{-n} x_j(n)$ converges to some polynomial, in Theorem 3.1 we establish some asymptotic properties of $\{x_i(n)\}$ together with the polynomial bounds on $\{x_i(n)\}$ that are considerably better than the bounds mentioned in Proposition 2. Moreover, if there exists $\hat{f} \in F$ such that for any $j = i + q, \ldots, i + r$ $\lim_{n\to\infty} \sigma_i^{-n}[x_j(n) - x_j(n; \hat{f})] = 0$, then it is shown in Corollary 3.4 that, for suitably selected $\hat{f}^*$, also $x_i(n)$ can be well approximated by $x_i(n; \hat{f}^*)$.

As it is indicated in Section 4, the assumptions of Theorem 3.1 and Corollary 3.4 are always fulfilled if $x_i(n)$ correspond to the cumulative sums of one-stage rewards; so these assumptions are always satisfied for, in the literature widely discussed, dynamic programming recursions of classical Markov decision chains (cf. [7] Example 1 of Section 1). Moreover, it is shown in [9], [10] that these assumptions are always fulfilled if some aperiodicity conditions on $Q_{jj}(f)$ (with $j = i + q, \ldots, i + r$) hold; e.g. in case that for any $j = i + q, \ldots, i + r$ and any $f \in F$ there exists $\lim_{n\to\infty} \sigma_i^{-n} (Q_{jj}(f))^n$.

Before presenting Theorem 3.1, let us recall that $\bar{x}_i(n) = \sigma_i^{-n} x_i(n)$, $\bar{Q}_{ii}(f) = \sigma_i^{-1} Q_{ii}(f)$ and remember that for any vector, say c, c_j usually will denote its subvector whose components are labelled by integers from $I_j(\hat{f})$, $j = 1, \ldots, s \equiv s(\hat{f})$. On the base of our definition of σ_i (spectral radius), resp. ν_i (index), of $Q_{ii}(\hat{f})$ we immediately conclude (cf. Proposition 1)

$$\nu_i > 1 \Rightarrow \sigma_j = \sigma_i \quad \text{and} \quad \nu_j = \nu_i + i - j \quad \text{for} \quad j = i, \ldots, i + \nu_i - 1 . \tag{3.1}$$

For what follows, it will be useful to introduce for any $i = 1, \ldots, s \equiv s(\hat{f})$ and $j = i, \ldots, i + \nu_i - 1$

$$y_j^{(i)}(n) = \sigma_i^{-n} x_j(n) - \sum_{l=0}^{\nu_j - 1} \binom{n}{l} w_j^{(l)} , \tag{3.2}$$

where $w_j^{(l)}$ are (yet unspecified) vectors of an appropriate dimension, and set

$$y_j^{(i)}(n) = \sigma_i^{-n} x_j(n) \quad \text{for any} \quad j \geqq i + \nu_i . \tag{3.2'}$$

Observe that $y_j^{(i)}(n) = y_j^{(j)}(n)$, $\bar{x}_j(n) = \sigma_i^{-n} x_j(n)$, $\sigma_j = \sigma_i$ for any $i \leqq j < i + \nu_i$ and by (1.7') for any $j \geqq i + \nu_i$ $\lim_{n\to\infty} y_j^{(i)}(n) = 0$.

Theorem 3.1. Let $\nu_{i_0} > 1$, integer $p \in (i_0; i_0 + \nu_{i_0})$ and let vectors $w_j^{(l)}$ (for $l = 0, \ldots, \nu_j - 1$) in (3.2) be selected such that

$$\lim_{n\to\infty} y_j^{(i_0)}(n) = 0 \quad \text{for any} \quad j = p, \ldots, i_0 + \nu_{i_0} - 1 . \tag{3.1.1}$$

Then for any $i = i_0, \ldots, p - 1$ there exist vectors $w_i^{(l)}$ (where $l = 0, \ldots, \nu_i - 1$), that can be found by the policy iteration method used in the proof of Lemma 3.3, such that

$$\lim_{n\to\infty} \binom{n}{\nu_i - \nu_p}^{-1} y_i^{(i_0)}(n) = 0 . \tag{3.1.2}$$

Moreover, if the convergence in (3.1.1) is exponential, i.e. if there exist number $\lambda \in (0, 1)$ and vectors $c'_j \ll 0$, $c''_j \gg 0$ such that

$$(3.1.3) \qquad c'_j \lambda^n \leqq y_j^{(i_0)}(n) \leqq c''_j \lambda^n \quad \text{for all } n \text{ and any } \; j = p, \ldots, i_0 + \nu_{i_0} - 1\,,$$

then for any $i = i_0, \ldots, p - 1$ the elements of a (vector) sequence

$$\left\{ \binom{n}{\nu_i - \nu_p - 1}^{-1} y_j^{(i_0)}(n), \; n = 0, 1, \ldots \right\}$$

are uniformly bounded (observe that by (3.1) $\nu_i - \nu_p = p - i$).

Before presenting the proof of Theorem 3.1 we establish two lemmas.

Lemma 3.2. Let $\{a_i(n), n = 0, 1, \ldots\}$, resp. $\{\beta(n), n = 0, 1, \ldots\}$, be a sequence of vectors, resp. nonnegative numbers, satisfying for each $n = 0, 1, \ldots$

$$(3.2.1) \qquad a_i(n + 1) \leqq \bar{Q}_{ii}(f^{(n)})\, a_i(n) + \alpha''\, \beta(n)\, u_i$$

together with

$$(3.2.1') \qquad a_i(n + 1) \geqq \bar{Q}_{ii}(f)\, a_i(n) + \alpha'\, \beta(n)\, u_i$$

where $f^{(n)}, f \in F$, numbers $\alpha' < 0$, $\alpha'' > 0$, and vector $u_i \gg 0$ satisfies (1.4). Then

$$(3.2.2) \qquad \beta(n) = \lambda^n \quad \text{with} \quad \lambda \in (0, 1) \;\Rightarrow\; \{a_i(n)\} \quad \text{is bounded;}$$

$$(3.2.2') \qquad \lim_{n\to\infty} \beta(n) = 0 \;\Rightarrow\; \lim_{n\to\infty} n^{-1}\, a_i(n) = 0\,;$$

$$(3.2.3) \qquad \beta(n) = \binom{n}{k} \;\Rightarrow\; \left\{ \binom{n}{k + 1}^{-1} a_i(n) \right\} \quad \text{is bounded,}$$

and

$$(3.2.3') \qquad \lim_{n\to\infty} \binom{n}{k}^{-1} \beta(n) = 0 \;\Rightarrow\; \lim_{n\to\infty} \left[\binom{n}{k + 1}^{-1} a_i(n) \right] = 0\,.$$

Proof. Iterating (3.2.1), resp. (3.2.1′), we get by (1.4)

$$(3.2.4) \qquad a_i(n) \leqq \prod_{m=1}^{n} \bar{Q}_{ii}(f^{(n-m)})\, a_i(0) + \alpha'' \sum_{m=0}^{n-1} \beta(m)\, u_i\,,$$

resp.

$$(3.2.4') \qquad a_i(n) \geqq (\bar{Q}_{ii}(f))^n\, a_i(0) + \alpha' \sum_{m=0}^{n-1} \beta(m)\, u_i\,.$$

Moreover, by iterating (1.4) (as $u_i = (\bar{Q}_{ii}(\bar{f}))^n u_i \geqq \prod_{m=1}^{n} \bar{Q}_{ii}(f^{(n-m)})\, u_i$ and, similarly, $u_i \geqq (\bar{Q}_{ii}(f))^n u_i$ with $u_i \gg 0$), we conclude that $(\bar{Q}_{ii}(f))^n, \prod_{m=1}^{n} \bar{Q}_{ii}(f^{(n-m)})$ are uniformly bounded in n. So (3.2.2), resp. (3.2.2′), follows immediately by (3.2.4), (3.2.4′).

To establish (3.2.3), (3.2.3′) observe that

$$(3.2.5) \qquad \beta(m) = \binom{m}{k} \;\Rightarrow\; \sum_{m=k}^{n-1} \beta(m) = \binom{n}{k + 1}$$

(this fact can be easily verified by induction with respect to n). Consequently, (3.2.3), resp. (3.2.3′), follows immediately in virtue of boundedness of $(\bar{Q}_{ii}(f))^n, \prod_{m=1}^{n} \bar{Q}_{ii}(f^{(n-m)})$ by inserting (3.2.5) into (3.2.4), (3.2.4′). $\square$

The following lemma is the main ingredient to the proof of Theorem 3.1 (recall that decision vector $\hat{f}^{(n)}$ is defined by (1.1) or by (1.5)). For what follows, it will be useful (cf. (2.1.3), (2.1.4) of Theorem 2.1) for any $i = 1, \ldots, s \equiv s(\hat{f})$, each $f \in F$ and given $w_j^{(l)}$'s (where $j = i, \ldots, i + v_i - 1$; $l = 0, \ldots, v_j - 1$) to introduce

$$\varphi_i^{(v_i-1)}(f) = (\bar{Q}_{ii}(f) - I)\, w_i^{(v_i-1)} \tag{3.3}$$

and define for $l = v_i - 2, \ldots, 1, 0$

$$\varphi_i^{(l)}(f) = (\bar{Q}_{ii}(f) - I)\, w_i^{(l)} - w_i^{(l+1)} + \sum_{j=i+1}^{i+v_i-l-1} \bar{Q}_{ij}(f)\, w_j^{(l)} . \tag{3.3′}$$

Lemma 3.3. Let for some $i = 1, \ldots, s$ $v_i > 1$ and (cf. (3.2)) $w_j^{(l)}$'s (for $j = i + 1, \ldots$ $\ldots, i + v_i - 1$; $l = 0, \ldots, v_j - 1$) be given. So by (3.2) $y_j^{(i)}(n)$ are well-defined for all $j = i + 1, \ldots, i + v_i - 1$ and (3.2′) defines also $y_j^{(i)}(n)$ for any $j = i + v_i, \ldots$ $\ldots, s$.

Then there exist vectors $w_i^{(l)}$ (for $l = 0, \ldots, v_i - 1$), integer $n_i < \infty$ and decision vector $f^* \in F$ such that

$$y_i^{(i)}(n + 1) = \bar{Q}_{ii}(\hat{f}^{(n)})\, y_i^{(i)}(n) + \sum_{j=i+1}^{s} \bar{Q}_{ij}(\hat{f}^{(n)})\, y_j^{(i)}(n) + \sum_{l=0}^{v_i-1} \binom{n}{l} \varphi_i^{(l)}(\hat{f}^{(n)}) \tag{3.3.1}$$

$$y_i^{(i)}(n + 1) \leqq \bar{Q}_{ii}(\hat{f}^{(n)})\, y_i^{(i)}(n) + \sum_{j=i+1}^{s} \bar{Q}_{ij}(\hat{f}^{(n)})\, y_j^{(i)}(n) \quad \text{for any} \quad n \geqq n_i \tag{3.3.2}$$

$$y_i^{(i)}(n + 1) \geqq \bar{Q}_{ii}(f^*)\, y_i^{(i)}(n) + \sum_{j=i+1}^{s} \bar{Q}_{ij}(f^*)\, y_j^{(i)}(n) \quad \text{for any } n . \tag{3.3.3}$$

Proof. By (1.5) and (3.3), (3.3′) we get for arbitrary $w_i^{(l)}$'s and $w_j^{(l)}$'s with $w_j^{(l)} = 0$ for any $l \geqq v_i$ or $j \geqq i + v_i$

$$\begin{aligned} y_i^{(i)}(n + 1) &= \bar{x}_i(n + 1) - \sum_{l=0}^{v_i-1} \binom{n+1}{l} w_i^{(l)} = \\ &= \bar{Q}_{ii}(\hat{f}^{(n)})\, \bar{x}_i(n) + \sum_{j=i+1}^{s} \bar{Q}_{ij}(\hat{f}^{(n)})\, \sigma_i^{-n}\, x_j(n) - \sum_{l=0}^{v_i-1} \binom{n}{l} w_i^{(l)} - \sum_{l=1}^{v_i-1} \binom{n}{l-1} w_i^{(l)} = \\ &= \bar{Q}_{ii}(\hat{f}^{(n)})\, y_i^{(i)}(n) + \sum_{j=i+1}^{s} \bar{Q}_{ij}(\hat{f}^{(n)})\, y_j^{(i)}(n) + (\bar{Q}_{ii}(\hat{f}^{(n)}) - I) \\ &\quad \sum_{l=0}^{v_i-1} \binom{n}{l} w_i^{(l)} + \sum_{j=i+1}^{i+v_i-1} \bar{Q}_{ij}(\hat{f}^{(n)}) \sum_{l=0}^{v_j-1} \binom{n}{l} w_j^{(l)} - \sum_{l=1}^{v_i-1} \binom{n}{l-1} w_i^{(l)} = \\ &= \bar{Q}_{ii}(\hat{f}^{(n)})\, y_i^{(i)}(n) + \sum_{j=i+1}^{s} \bar{Q}_{ij}(\hat{f}^{(n)})\, y_j^{(i)}(n) + \\ &\quad + \binom{n}{v_i - 1} \varphi_i^{(v_i-1)}(\hat{f}^{(n)}) + \sum_{l=1}^{v_i-1} \binom{n}{l-1} \varphi_i^{(l-1)}(\hat{f}^{(n)}) \end{aligned} \tag{3.3.4}$$

(recall that $\bar{Q}_{ij}(f) = \sigma_i^{-1} Q_{ij}(f)$ and observe that, as by (3.1) $v_j = v_i - j + i$ and as $l - 1 \geqq v_j \Rightarrow w_j^{(l-1)} = 0$, by changing summation we get

$$\sum_{j=i+1}^{i+v_i-1} \bar{Q}_{ij}(\hat{f}^{(n)}) \sum_{l=0}^{v_j-1} \binom{n}{l} w_j^{(l)} = \sum_{l=1}^{v_i-1} \binom{n}{l-1} \sum_{j=i+1}^{i+v_i-l} \bar{Q}_{ij}(\hat{f}^{(n)})\, w_j^{(l-1)} \Big).$$

Now we show how to select $w_i^{(l)}$'s if $w_j^{(l)}$'s (for $j = i + 1, \ldots, i + v_i - 1$) are known. According to Theorem 2.1 on the base of given $w_j^{(l)}$'s (with $j = i + 1, \ldots, i + v_i - 1$; $l = 0, \ldots, v_j - 1$) we can construct by the policy iteration algorithm $w_i^{(l)}$'s (for $l = v_i - 1, \ldots, 0$) such that

$$\varphi_i^{(v_i-1)}(f) \leqq 0 \quad \text{for any} \quad f \in F_i \tag{3.3.5}$$

and for any $l = v_i - 2, \ldots, 0$

$$\varphi_i^{(l)}(f) \leqq 0 \quad \text{for any} \quad f \in F_i^{(l+1)} \tag{3.3.5'}$$

where $\{F_i^{(l)}, l = v_i - 1, \ldots, 0\}$ is defined recursively by

$$F_i^{(l)} = \{f \in F_i^{(l+1)} : \varphi_i^{(l)}(f) = 0\} \quad \text{with} \quad F_i^{(v_i)} \equiv F_i\,. \tag{3.3.6}$$

Moreover, by Theorem 2.1 an equality sign holds in (3.3.5), (3.3.5′) at least for one $f \in F_i^{(0)}$, say $f \equiv f^*$. Inserting (3.3.5), (3.3.5′) into (3.3.4) we get for any $f \in F_i$

$$\begin{aligned} y_i^{(i)}(n+1) &= \bar{Q}_{ii}(\hat{f}^{(n)})\, y_i^{(i)}(n) + \sum_{j=i+1}^{s} \bar{Q}_{ij}(\hat{f}^{(n)})\, y_j^{(i)}(n) + \\ &+ \sum_{l=0}^{v_i-1} \binom{n}{l} \varphi_i^{(l)}(\hat{f}^{(n)}) \geqq \bar{Q}_{ii}(f)\, y_i^{(i)}(n) + \sum_{j=i+1}^{s} \bar{Q}_{ij}(f)\, y_j^{(i)}(n) + \sum_{l=0}^{v_i-1} \binom{n}{l} \varphi_i^{(l)}(f) \end{aligned} \tag{3.3.7}$$

establishing (3.3.1). Recalling that the vectors $w_i^{(l)}$'s are selected such that (3.3.5), (3.3.5′) hold and denoting for any $f \in F_i$

$$s_i(f; n) = \sum_{l=0}^{v_i-1} \binom{n}{l} \varphi_i^{(l)}(f)\,, \tag{3.3.8}$$

each component of $s_i(f; n)$ is a polynomial in n and, in virtue of (3.3.5), (3.3.5′) and (3.3.6), its first non-vanishing coefficient must be negative. Consequently, for any $f \in F_i$ there exists finite integer $n_i(f)$ such that $s_i(f; n) \leqq 0$ for any $n \geqq n_i(f)$ and, as the set F_i is finite, there also exists $n_i < \infty$ such that

$$s_i(f; n) \leqq 0 \quad \text{for any} \quad n \geqq n_i\,, \quad f \in F_i\,. \tag{3.3.9}$$

(3.3.2) follows then by inserting (3.3.9) into the first part of (3.3.7). Recalling that $\varphi_i^{(l)}(f^*) = 0$ (for any $l = v_i - 1, \ldots, 1$) by (3.3.7) we immediately get (3.3.3). □

Now we are in a position to present:

Proof of Theorem 3.1. First recall that by (3.2′) and (1.7) there exist vectors $c_j' \ll 0$, $c_j'' \gg 0$ and a number $\lambda \in (0, 1)$ such that

$$c_j' \lambda^n \leqq y_j^{(i_0)}(n) \leqq c_j'' \lambda^n \quad \text{for any} \quad j \geqq i_0 + v_{i_0}, \ldots, s\,. \tag{3.1.4}$$

The proof of the theorem proceeds by induction for $i = p - 1, p - 2, \ldots, i_0$ and heavily employs (3.3.2), (3.3.3). By (3.1.1), (3.1.4) if $n \to \infty$ the last term on the RHS of (3.3.2), (3.3.3) tends to zero and so (3.1.2) for $i = p - 1$ follows immediately by (3.2.2′). To show that under (3.1.3) $\{n^{-1} y_{p-1}^{(i_0)}(n)\}$ is bounded, it suffices to select numbers $\alpha' < 0$, $\alpha'' > 0$ and $\lambda \in (0, 1)$ such that for $i = p - 1$ and any n

$$(3.1.5) \qquad \alpha' \lambda^n u_i \leqq \sum_{j=i+1}^{s} \bar{Q}_{ij}(f^*)\, y_j^{(i_0)}(n)\,, \quad \sum_{j=i+1}^{s} \bar{Q}_{ij}(\hat{f}^{(n)})\, y_j^{(i_0)}(n) \leqq \alpha'' \lambda^n u_i$$

and apply (3.2.2) to (3.3.2), (3.3.3) where the last term is replaced from (3.1.5).

Now let us suppose by induction argument that for some $i < p - 1$, $i \geqq i_0$ and each $j = i + 1, \ldots, p - 1$ $w_j^{(l)}$'s are selected such that

$$(3.1.6) \qquad \lim_{n \to \infty} \binom{n}{v_j - v_p}^{-1} y_j^{(i_0)}(n) = 0\,,$$

resp.

$$(3.1.7) \qquad \left\{ \binom{n}{v_j - v_p - 1}^{-1} y_j^{(i_0)}(n),\ n = 0, 1, \ldots \right\} \text{ is bounded.}$$

Observe that by (3.1) $v_j - v_p = p - j$ and that in (3.1.6), (3.1.7) it suffices to consider only $w_j^{(l)}$'s with $l = v_j - 1, \ldots, v_j - v_p$; by (3.3), (3.3′) these $w_j^{(l)}$'s are calculated only on the base of $w_k^{(l)}$'s with $k = p, \ldots, i_0 + v_{i_0} - 1$. So by (3.1.6) we immediately conclude

$$(3.1.8) \qquad \lim_{n \to \infty} \binom{n}{p - i - 1}^{-1} \sum_{j=i+1}^{s} \bar{Q}_{ij}(\hat{f}^{(n)})\, y_j^{(i_0)}(n) = \lim_{n \to \infty} \binom{n}{p - i - 1}^{-1} \sum_{j=i+1}^{s} \bar{Q}_{ij}(f^*)\, y_j^{(i_0)}(n) = 0\,.$$

Similarly, by (3.1.7) for suitably chosen numbers $\alpha' < 0$, $\alpha'' > 0$ we get

$$(3.1.9) \qquad \alpha' \binom{n}{p - i - 2} u_i \leqq \sum_{j=i+1}^{s} \bar{Q}_{ij}(f^*)\, y_j^{(i_0)}(n)\,, \quad \sum_{j=i+1}^{s} \bar{Q}_{ij}(\hat{f}^{(n)})\, y_j^{(i_0)}(n) \leqq \alpha'' \binom{n}{p - i - 2} u_i\,.$$

The induction proof of (3.1.2) can be easily completed by applying (3.2.3′) to (3.3.2), (3.3.3) and (3.1.8). Similarly, boundedness of $\left\{ \binom{n}{v_i - v_p - 1}^{-1} y_i^{(i_0)}(n) \right\}$ follows immediately by applying (3.2.3) to (3.3.2), (3.3.3) where the last term is replaced by (3.1.9). □

Now recall that $\bar{x}_i(n; f) = \sigma_i^{-n} x_i(n; f)$ (where $x_i(n; f)$ are defined recursively by (1.1′)) and remember that $\hat{f}^{(n)} \in F$ is given by (1.1). Theorem 3.1 implies the following:

Corollary 3.4. Let assumption (3.1.1) of Theorem 3.1 hold. Then for any $i = i_0, \dots, p-1$:

(i) There exists (finite) n_i such that for any $n \geqq n_i$

$$\varphi_i^{(l)}(\hat{f}^{(n)}) = 0 \quad \text{for} \quad l = v_i - 1, \dots, v_i - v_p = p - i \tag{3.4.1}$$

and, consequently, (cf. (3.3.1)) for any $n \geqq n_i$

$$y_i^{(i)}(n+1) = \bar{Q}_{ii}(\hat{f}^{(n)})\, y_i^{(i)}(n) + \sum_{j=i+1}^{s} \bar{Q}_{ij}(\hat{f}^{(n)})\, y_j^{(i_0)}(n) + \sum_{l=0}^{p-i-1} \binom{n}{l} \varphi_i^{(l)}(\hat{f}^{(n)}) . \tag{3.4.2}$$

(ii) If there exists $\tilde{f} \in F$ such that

$$\lim_{n\to\infty} [\bar{x}_j(n) - \bar{x}_j(n; \tilde{f})] = 0 \quad \text{for any} \quad j = p, \dots, i_0 + v_{i_0} - 1 , \tag{3.4.3}$$

then there also exists $\tilde{f}^* \in F$ (with $\tilde{f}^*(k) = \tilde{f}(k)$ for any $k \in I_m(\tilde{f})$ with $m \geqq p$) such that

$$\lim_{n\to\infty} \binom{n}{v_i - v_p}^{-1} \left[\bar{x}_i(n; \tilde{f}^*) - \sum_{l=0}^{v_i-1} \binom{n}{l} \tilde{w}_i^{(l)} \right] = 0 \tag{3.4.4}$$

where (cf. (3.2), (3.1.1)) $w_i^{(l)} \equiv \tilde{w}_i^{(l)}$ for $l = v_i - 1, \dots, v_i - v_p$, and

$$\lim_{n\to\infty} \binom{n}{v_i - v_p}^{-1} [\bar{x}_i(n) - \bar{x}_i(n; \tilde{f}^*)] = 0 . \tag{3.4.4'}$$

Moreover, if also (3.1.3) holds and the convergence in (3.4.3) is also exponential, then even the sequence

$$\left\{ \binom{n}{v_i - v_p - 1}^{-1} [\bar{x}_i(n) - \bar{x}_i(n; \tilde{f}^*)],\ n = 0, 1, \dots \right\}$$

is bounded.

Proof. To establish part (i) we shall employ (3.3.1). Premultiplying (3.3.1) by $\binom{n}{p-i}^{-1}$ (recall that $v_p - v_i = p - i$), then letting $n \to \infty$ and inserting from (3.1.1), (3.1.2), (3.1.4) we conclude that

$$\lim_{n\to\infty} \binom{n}{p-i}^{-1} \sum_{l=0}^{v_i-1} \binom{n}{l} \varphi_i^{(l)}(\hat{f}^{(n)}) = 0 . \tag{3.4.5}$$

As $\hat{f}^{(n)}$ are taken from a finite set F, by (3.4.5) we conclude (recall (3.3.5), (3.3.5')) that (3.4.1) must hold. (3.4.2) follows then immediately by (3.3.1) and (3.4.1).

To verify part (ii), mimicking the procedure used in the proof of Lemma 3.3 (cf.

(3.3.4)–(3.3.7)), we easily conclude that for any $i = i_0, \ldots, p - 1$ and some $\tilde{f}^* \in F$

$$(3.4.6) \qquad y_i^{(i_0)}(n + 1; \tilde{f}^*) = \bar{Q}_{ii}(\tilde{f}^*)\, y_i^{(i_0)}(n; \tilde{f}^*) + \sum_{j=i+1}^{s} \bar{Q}_{ij}(\tilde{f}^*)\, y_j^{(i_0)}(n; \tilde{f}^*)$$

where (cf. (3.2)) for $i = i_0, \ldots, i_0 + v_{i_0} - 1$

$$(3.4.6') \qquad y_i^{(i_0)}(n; \tilde{f}^*) = \bar{x}_i(n; \tilde{f}^*) - \sum_{l=0}^{v_i - 1} \binom{n}{l} \tilde{w}_i^{(l)}$$

$$(3.4.6'') \qquad y_j^{(i_0)}(n; \tilde{f}^*) = \sigma_{i_0}^{-n}\, x_j(n; \tilde{f}^*) \quad \text{for any} \quad j \geqq i_0 + v_{i_0} \,.$$

Here $\tilde{w}_i^{(l)}$'s together with $\tilde{f}^* \in F$ are calculated on the base of

$$(3.4.7) \qquad \tilde{w}_j^{(l)} \equiv w_j^{(l)} \quad (\text{for } j = p, \ldots, i_0 + v_{i_0} - 1,\ l = v_j - 1, \ldots, 0)$$

introduced in Theorem 3.1; so by (3.3), (3.3′) we immediately conclude that also

$$(3.4.7') \qquad \tilde{w}_i^{(l)} = w_i^{(l)} \quad \text{for any} \quad l = v_i - 1, \ldots, v_i - v_p \,.$$

In virtue of (3.1.1), (3.4.3), (3.4.6′), (3.4.7) and (3.4.6″), (1.7′) we get that

$$(3.4.8) \qquad \lim_{n \to \infty} y_j^{(i_0)}(n; \tilde{f}^*) = 0 \quad \text{for} \quad j \geqq p \,.$$

Now by (3.4.6), (3.4.8), using the same arguments as in the proof of Theorem 3.1, we conclude that

$$(3.4.9) \qquad \lim_{n \to \infty} \binom{n}{v_i - v_p}^{-1} y_i^{(i_0)}(n; \tilde{f}^*) = 0$$

and, consequently, (3.4.5) can be immediately verified by (3.1.2), (3.4.9).

In case that (3.1.3) together with exponential convergence in (3.4.3) is assumed, convergence in (3.4.9) is exponential and, similarly as in the proof of Theorem 3.1, we can verify boundedness of $\left\{\binom{n}{v_i - v_p - 1}^{-1} y_i^{(i_0)}(n; \tilde{f}^*)\right\}$. As by Theorem 3.1 also $\left\{\binom{n}{v_i - v_p - 1}^{-1} y_i^{(i_0)}(n)\right\}$ is bounded, in virtue of (3.4.7′) $\left\{\binom{n}{v_i - v_p - 1}^{-1} \cdot [\bar{x}_i(n) - \bar{x}_i(n; \tilde{f}^*)]\right\}$ must be bounded. □

4. CONCLUSION AND DISCUSSION

In the present paper we have established some polynomial bounds on the utility vector $x(n)$ calculated from dynamic programming recursion (1.1). In the dynamic programming literature analogous results were obtained only for Markov decision chains (cf. [1], [4] and especially [3] for the most complete results); i.e. (cf. Example 1 of Section 1 in [7]) for a very special case of our model with $s = 2$ and

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

where $P(f)$ is the transition probability matrix and $r(f)$ is the vector of one-stage rewards of the considered Markov decision chain. Then for $x(n) = \begin{bmatrix} v(n) \\ 1 \end{bmatrix}$, $v(0) = 0$ we get

$$v(n + 1) = r(\hat{f}^{(n)}) + P(\hat{f}^{(n)})\, v(n) = \sum_{m=0}^{n} \left(\prod_{k=1}^{m} P(\hat{f}^{(n+1-k)}) \right) r(\hat{f}^{(n-m)})$$

and $v(n)$ can be interpreted as the vector of maximum expected total rewards in the n next transition of the considered Markov decision chain. We can easily verify that, in this very special case, the assumptions of Theorem 3.1 are trivially fulfilled with $x_2(n) \equiv 1$, $\bar{x}_1(n) = x_1(n) \equiv v(n)$; so by Theorem 3.1 $y_1^{(1)}(n) = x_1(n) - nw_1^{(1)} - w_1^{(0)}$ is bounded. Observe that $w_1^{(1)} = \max_{f \in F} P^*(f)\, r(f)$ (the vector of maximum average rewards, $P^*(f)$ is the Cesaro limit of $P(f)$) can be found by the celebrated Howard's policy iteration algorithm. By the second part of Example I of Section 1 in [7] we can also easily verify that the assumptions of Theorem 3.1 are also fulfilled for functional equations corresponding to cumulative expected rewards of Markov decision chains (cf. [8] for a detailed discussion).

To obtain corresponding results for more general discrete dynamic programming models, we have heavily employed specific block-triangular structure of the considered dynamic programming problems. Remember that in [7] we have already shown that any discrete dynamic programming model with nonnegative matrices possesses this property (cf. Proposition 1). The results of [7] summarized in Proposition 2 present some "rough" bounds on the utility vector of this type of dynamic programming models. To obtain "finer" bounds we have generalized in Section 2 (cf. Theorem 2.1) policy iteration algorithm for finding maximum average reward of a classical Markov decision chain and extended in Section 3 the reasoning used in [3] to the general case with $s \neq 2$ and not necessarily stochastic $Q_{ii}(f)$'s. These "finer" polynomial bounds presented in Section 3 (cf. Theorem 3.1) are obtained under the assumptions that are trivially fulfilled in Markov programming.

In a companion paper [8] we show how this functional equations approach together with the obtained polynomial bounds on the utility vector $x(n)$ can be employed for classical and multiplicative Markov decision chains to establish a family of optimality criteria having a nice property that an optimal policy can be found in the class of stationary policies.

Moreover, it was shown in [9] that the assumptions of Theorem 3.1 are also always fulfilled if some aperiodicity conditions on $Q_{jj}(f)$ (with $j = p, \ldots, i_0 + v_{i_0} - 1$) hold; i.e., e.g. in case that for any $j = p, \ldots, i_0 + v_{i_0} - 1$ and any $f \in F$

$\lim\limits_{n\to\infty} \sigma_j^{-n}(Q_{jj}(f))^n$ exists. In particular, Theorem 4.3 of [9] asserts that, if $\sigma_i^{-n}\, x_j(n)$ (for all $j > i$) converges to some polynomial and this converegence is exponential, then $\bar{x}_i(n)$ is bounded by a polynomial of degree ν_i and, moreover, under some aperiodicity conditions on $\{Q_{ii}(f), f \in F\}$ also $\bar{x}_i(n)$ converges to some polynomial and this convergence is exponential. Observe that these facts immediately imply that, if aperiodicity conditions are imposed on all $Q_{jj}(f)$ with $j \geqq i$, then $\bar{x}_i(n)$ converges to some polynomial. These results, in a slightly generalized form, will be included into the forthcoming paper [10].

Recently, the same problem was independently studied by Zijms. In [13] the properties of generalized eigenvectors for the sets of nonnegative matrices are presented (these generalized eigenvectors well corresponds to our construction of vectors $w_i^{(l)}$'s based on the results of Section 2) and the main result of [14] establishes (under the assumptions that for all $f \in F$ and $i = 1, \ldots, s$ $Q_{ii}(f)$ are aperiodic) that $\bar{x}_i(n)$ converges to some polynomial and this convergence is exponential.

The following example shows that, if the assumptions of Theorem 3.1 are not satisfied, the respective averages of $\{\bar{x}_i(n)\}$ (with $\nu_i > 1$) need not converge to any polynomial.

Example. Let $Q \equiv Q(f)$ where

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ 0 & Q_{22} \end{bmatrix} \quad \text{with} \quad Q_{11} = Q_{22} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Q_{12} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

and $x(0) = [0\ 0\ 1\ 0]^{\mathrm{T}}$ (T denotes transpose). Then by a simple calculation from (1.1), (1.1′) we get for $x(n)$:

$$x(2m) = [0\ m\ 1\ 0]^{\mathrm{T}} \quad \text{and} \quad x(2m+1) = [m+1\ 0\ 0\ 1]^{\mathrm{T}}.$$

In virtue of the presented example we may conjecture that for the considered dynamic programming model (if no aperiodicity assumptions on $Q_{jj}(f)$'s are made) at least $\{\bar{x}_i(n\varkappa + m), n = 0, 1, \ldots\}$ (for some integers $m, \varkappa$ where $0 \leqq m < \varkappa$) will be bounded or even converge to some polynomial. Discussion of these problems is postponed into [10].

(Received December 19, 1980.)

REFERENCES

[1] R. Bellman: A Markovian decision process. J. Math. Mech. *6* (1957), 679–684.

[2] D. Blackwell: Discrete dynamic programming. Ann. Math. Statist. *33* (1962), 2, 719–726.

[3] B. G. Brown: On the iterative methods of dynamic programming in a finite space discrete time Markov processes. Ann. Math. Statist. *36* (1965), 4, 1279–1285.

[4] R. A. Howard: Dynamic Programming and Markov Processes. M.I.T. and Wiley Press, New York 1960.

[5] P. Mandl: Controlled Markov chains (in Czech). Kybernetika *6* (1969), Supplement, 1—74.
[6] K. Sladký: On the set of optimal controls for Markov chains with rewards. Kybernetika *10* (1974), 4, 350—367.
[7] K. Sladký: Bounds on discrete dynamic programming recursions I — Models with non-negative matrices. Kybernetika *16* (1980), 6, 526—547.
[8] K. Sladký: On the existence of stationary optimal policies in discrete dynamic programming. Submitted to Kybernetika.
[9] 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.
[10] K. Sladký: On functional equations of discrete dynamic programming. In preparation.
[11] A F. Veinott, Jr.: On finding optimal policies in discrete dynamic programming with no discounting. Ann. Math. Statist. *37* (1966), 5, 1284—1294.
[12] A. F. Veinott, Jr.: Discrete dynamic programming with sensitive discount optimality criteria. Ann. Math. Statist. *40* (1969), 5, 1635—1660.
[13] W. H. M. Zijms: Generalized Eigenvectors and Sets of Nonnegative Matrices. Memorandum Cosor 80 — 03, Eindhoven University of Technology, Eindhoven 1980.
[14] W. H. M. Zijms: Maximizing the Growth of the Utility Vector in a Dynamic Programming Model. Memorandum Cosor 80 — 04, Eindhoven University of Technology, Eindhoven 1980.

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.*