AN ANALITICAL METHOD TO CALCULATE THE ERGODIC AND DIFFERENCE MATRICES OF THE DISCOUTED MARKOV DECISION PROCESSES

Jan Kałuski

The Silesian Technical University of Gliwice

Departament of Organization and Management

Chair of Computer Science and Econometrics

ul.Roosevelt,$26 - 28$, $41 - 800$ Zabrze, Poland

Fax: (48 32) 237 21 77

Tel: (48 32) 2777 354

E–mail: j.kaluski@ia.polsl.gliwice.pl

Summary: In the work the analitical method to calculate the ergodic and difference matrices of finite state discouted Marcov decision processes is presented. On the basis well - known literature the result for overall discounted value, this one in interpretation of the calculated matrices is shown. The obtained results gives a possibility to distinguish the constant and variable parts of the overall discounted value.

The presented analitical method is illustrated by two simple examples. New performance index to discounted optimal Markov control problem is proposed.

1   Introduction

Marcov Decision Processes (MDP) are also called Controlled Marcov Processes or Marcov Processes wiht reward since 1960, when Howard [25] introduced them, ones became extremely attractive research tool and they found wide application in different technical and research disciplines. From the beginning it can be observed that the method is more and more improved [6, 9, 10, 11, 12, 16, 19]. The improvement concerns connections of Markov Processes with different methods of mathematical programming [3, 10, 11, 12, 17, 20, 22, 24, 25, 29, 30, 34, 35, 38, 42], their optimization and searching of computational methods of different matrices which are connected with stochastic Marcov matrix of transitions. Simultaneously with development of MDP (some scientists think that even earlier) developed discipline which is connected with game theory, namely - stochastic games [4, 5, 6, 15, 17, 21, 23, 28, 31, 33, 36, 37, 38, 40]. Quite not long ago on the basis of two topics: stochastic games and Markov Decision Processes arose new discipline connected with competition in Marcov Decision Processes (so called Competitive Marcov Decision Processes) [18].

Especially practical importance for development of mentioned methods has theory of irreducible Marcov Decision Chains with finite set of states and given

discount factor β [1]. This factor allows to calculate finite expected rewards which appear during different economic – technical processes in long period of time (theoretically without time limit). This will happen in case if mathematical model of these real processes is MDP.

Stochastic matrix of transitions P generates for irreducible Markov Chain with finite set of states very important matrices which can be used for analysis of these processes: ergodic, fundamental and potential. Very simple and at the same time strictly mathematical methods of calculation of these matrices in work [14, 18, 42] are widely discussed. Presented methods are in most cases numerical iterative algorithms which rely on Howard's algorithms. It exists well – known analytical method which calculates ergodic matrix for $t \to \infty$. But it is impossible to analyse the complete process with disturbances of the transient processs which happen during initial development period of Markov Decision Process with or without discount.

The goal of this paper is to present analytical method of calculation of ergodic matrix and so called difference matrices of Discounted Finite States Markov Decision Chain.

It allows to analyse the total process in $t \in [0, \infty)$ range through separation of two parts: constant which is represented by ergodic matrix and variable which is represented by difference matrices.

The paper is organized as follows: In section 2 reminded readers derivation of general formula for total expected rewards for $t \to \infty$, given matrix P and $\beta \in [0, 1)$. The derivation relies on known formula for total expected rewards if the input state was defined. In section 3 theorem about existing ergodic matrix and connected with it difference matrices is formulated and proved. These matrices always exist for $\beta < 1$. In section 4 two simple examples which illustrate computational method are solved. In section 5 new performance index used for optimization Discounted Markov Decision Processes is interpreted.

2 Total expected reward with discount

We consider ergodic Marcov Chain with finite set of states N and given stochastic matrix $P = [p_{ij}]$, $i, j = \overline{1, N}$. We have also one – step matrix of rewards $R = [r_{ij}]$, $i, j = \overline{1, N}$ which is controlled by Markov Chain. Let $\nu_i(n)$, $i, j = \overline{1, N}$, $n = 0, 1, 2, \ldots$ mean total process reward, if the input state was i-th state. The system will be closed after n steps (transition). Then we can show

---

[1]Discounted factor $\beta < 1$ means that value of reward unit which was achieved in moment $t = k$, in moment $t = k + n$ has value β^n.

(see [18, 25]) that the following recurrent formula for rewards is correct

$$\nu_i(n) = \sum_{j=1}^{N} P_{ij}\left[r_{ij} + \nu_j(n-1)\right]. \tag{1}$$

Other form of this formula is:

$$\nu_i(n) = \sum_{j=1}^{N} p_{ij}r_{ij} + \sum_{j}^{N} p_{ij}\nu_j(n-1), \tag{2}$$

where

$$q_i = \sum_{j=1}^{N} p_{ij}r_{ij}, \tag{3}$$

as one-step reward of process.
Now we can write

$$\nu_i(n) = q_i + \sum_{j=1}^{N} p_{ij}\nu_j(n-1). \tag{4}$$

After taking discount factor β into consideration we receive:

$$\nu_i(n,\beta) = q_i + \beta \sum_{j=1}^{N} p_{ij}\nu_j(n-1). \tag{5}$$

Let us write this formulae as vector

$$\nu(n,\beta) = q + \beta \cdot P \cdot \nu(n-1), \quad n = 0,1,2,\ldots \tag{6}$$

It is easy to notice that

$$\left.\begin{array}{l} \nu(1,\beta) = q + \beta P\nu(0) \\ \nu(2,\beta) = q + \beta P\nu(1) = q + \beta P\left(q + \beta P\nu(0)\right) = q + \beta Pq + \beta^2 P^2 \nu(0) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \nu(n,\beta) = q + \beta^n P^n \nu(0) + \sum_{n=1}^{n-1} \beta^n P^n q \end{array}\right\}. \tag{7}$$

Taking into consideration fact, that

$$q \equiv \beta^0 P^0 q \tag{8}$$

we receive

$$\nu\left(n, \beta\right) = \beta^n \cdot P^n \cdot \nu\left(0\right) + \sum_{n=0}^{n-1} \beta^n \cdot P^n \cdot q. \tag{9}$$

For $n \to \infty$ and $\beta < 1$ it follows to the next formula:

$$\nu_\infty\left(\beta\right) = \sum_{n=0}^{\infty} \beta^n \cdot P^n \cdot q = (I - \beta P)^{-1} \cdot q. \tag{10}$$

Formula (10) allows to calculate total expected rewards if factor β and starting state $i = 1, \ldots, N$ were given. We should pay attention to fact that this value is finite for $\beta < 1$. This formula is well– known in literature and often used for calculation of mentioned process rewards in long period of time in case if real process can be modelled by means of MDP. We can optimize decision process by means of formula (10) if we can choose different strategies of behaviour during analysis of real process [18, 25]. The choice of optimal decision means choice at i-th state of process so strategy of behaviour which gives maximum total expected reward. As we know the fact can be achieved using iterative Howard's algorithms for recurrent process (or it later version).

Discussed formula does not allow to analyse the process during the whole investigated period of time $t \in [0, \infty)$.

This problem was solved for discrete and continuous Markov processes without discount by Howard by means of z transformation (for discrete processes) and Laplace'a transformation (for continuos processes) for total rewards. After inverse transformation new formulae in explicit depend on n. The dependence is a sum of n–components. Ergodic matrix, which depends on n, is always the first component. Next $N - 1$ components are named difference matrices and they also depend on n. The sum of elements of each row for differential matrices is always equal zero. For $n \to \infty$, components approach zero. But the total expected reward approaches in this case infinity.

Next we show analytical method of calculation of ergodic matrix and difference matrices for Discounted Markov Chain with irreducible stochastic matrix P and finite set of states N which is based on approach proposed by Howard [25]. It allows to get finite total expected rewards $\nu_\infty\left(\beta\right)$ which can be characterised by two components: the first component represents finite part of constant reward which is connected with ergodic matrix and the second component represents finite part of variable reward which is connected with transient states of Markov Process. The part is a sum of rewards connected with difference matrices. So we can now analyse quality of investigated Markov Processes by comparison of constant and variable part of total reward in infinite period of time.

3 Method of calculation of ergodic and difference matrices

We consider dependence for total discounted rewards given by formula (10) again.

$$\nu_\infty(\beta) = (I - \beta P)^{-1} \cdot q. \tag{11}$$

It is not difficult to notice, that

$$(I - \beta P)^{-1} = \frac{1}{det\,(I - \beta P)} (I - \beta P)\,ad, \tag{12}$$

where $(I - \beta P)\,ad$ is an algebraically complement of matrix $(I - \beta P)$. Next we can write

$$(I - \beta P)\,ad = [D_{ji}(\beta)], \quad i, j = \overline{1, N} \tag{13}$$

where $D_{ji}(\beta) = (-1)^{j+i} \cdot M_{ji}(\beta)$, and $M_{ji}(\beta)$ is a minor of matrix $(I - \beta P)^T$, hence

$$(I - \beta P)^{-1} = \frac{[D_{ji}(\beta)]}{det\,(I - \beta P)}. \tag{14}$$

Theorem:
Let determinant of matrix $(I - \beta P)$ have real and singular roots, then for each stochastic matrix P and factor $\beta < 1$ exist such $\alpha_k \neq 0$, $k = 1, 2, \ldots, N$ that true is the following formula:

$$(I - \beta P)^{-1} = \frac{\left[D_{ji}^1\right]}{(1 - \alpha_1\beta)} + \frac{\left[D_{ji}^2\right]}{(1 - \alpha_2\beta)} + \ldots + \frac{\left[D_{ji}^N\right]}{(1 - \alpha_N\beta)}, \tag{15}$$

where

$$det\,(I - \beta P) = (1 - \alpha_1\beta)(1 - \alpha_2\beta)\ldots(1 - \alpha_N\beta)\ldots, \tag{16}$$

and

$D_{ji}^k, k = 1, 2, 3, \ldots, N$, are constant factors different from zero.

Proof:
From linear algebra results, that the determinant of matrix $(I - \beta P)$ always exist, it is bigger than zero and it is a polynomial of N - degree. Each polynomial of N - degree has exactly N different real roots. Hence we can show $det\,(I - \beta P)$ in form (16). Let prove (15) and write this formula in the following form:

$$\frac{[D_{ji}\,(\beta)]}{det\,(I - \beta P)} = \frac{\left[D_{ji}^1\right]}{(1 - \alpha_1\beta)} + \frac{\left[D_{ji}^2\right]}{(1 - \alpha_2\beta)} + \ldots + \frac{\left[D_{ji}^N\right]}{(1 - \alpha_N\beta)}. \qquad (17)$$

This formula shows decomposition of the left side of dependence (17) into sum of N partial fractions. Such a decomposition is always possible. If we want to calculate values of factors D_{ji}^k, $k = 1, 2, 3, \ldots, N$, we should solve $(N \times N)$ systems of equations:

$$\frac{D_{ji}\,(\beta)}{det\,(I - \beta P)} = \frac{D_{ji}^1}{1 - \alpha_1\beta} + \frac{D_{ji}^2}{1 - \alpha_2\beta} + \ldots + \frac{D_{ji}^N}{1 - \alpha_N\beta}, \quad i, j = \overline{1, N} \qquad (18)$$

it means

$$\frac{D_{ji}\,(\beta)}{det\,(I - \beta P)} = \frac{D_{ji}^1\,(1 - \alpha_2\beta)\ldots(1 - \alpha_N\beta) + D_{ji}^2\,(1 - \alpha_1\beta)\,(1 - \alpha_3\beta)\ldots}{det\,(I - \beta P)}$$
$$\frac{\ldots(1 - \alpha_N\beta) + \ldots + D_{ji}^N\,(1 - \alpha_1\beta)\,(1 - \alpha_2\beta)\ldots(1 - \alpha_{N-1}\beta)}{det\,(I - \beta P)}.$$

After rejection of denominators of both sides we compare factors which stand in front of the same powers β, $\left(\beta^0, \beta^1, \ldots, \beta^{N-1}\right)$ of the left and right side of numerators.
So each element of matrix

$$\frac{[D_{ji}\,(\beta)]}{det\,(I - \beta P)} = \left[\frac{D_{ji}\,(\beta)}{det\,(I - \beta P)}\right], \quad i, j = \overline{1, N}$$

was decomposed into N component forms:

$$\frac{[D_{ji}(\beta)]}{det\,(1 - \alpha_k\beta)}, \quad k = 1, 2, 3, \dots, N.$$

It can be shown, that $|D_{ij}|^k < 1$ and even values express probability and $\alpha_1 \equiv 1$. Next connecting so received N elements on condition equal factors $(1 - \alpha_K\beta)$, we create N separate $(N \times N)$ matrices. The first matrix is ergodic matrix producted by $1/(1 - \beta)$. For irreducible ergodic Markov Chain, this matrix will be constructed from the same rows.

Next $N - 1$ matrices will be difference matrices, each different and elements will be divided by factors $(I - \alpha_k\beta)$, $k = 2, 3, 4, \dots, N$.

Now the formula for total rewards can be written in the following form:

$$\nu_\infty(\beta) = \left(\frac{1}{1 - \beta}\left[D_{ji}^1\right] + \frac{1}{1 - \alpha_2\beta}\left[D_{ji}^2\right] + \dots + \frac{1}{1 - \alpha_N\beta}\left[D_{ji}^N\right]\right) \cdot q. \qquad (19)$$

Next we consider and solve two simple examples, which show application of presented method.

4   Examples

Example 1

Let P be one-step stochastic matrix of transition of irreducible Markov Chain with finite set of states N. Let $N = 2$ and let matrix P have the following form [25]:

$$P = \begin{bmatrix} 0,5 & 0,5 \\ 0,4 & 0,6 \end{bmatrix}.$$

Matrix of rewards R is also given,

$$R = \begin{bmatrix} 9 & 3 \\ 3 & -7 \end{bmatrix}.$$

Hence

$$q = \begin{bmatrix} 6 \\ -3 \end{bmatrix}.$$

After some easy transformation we receive:

$$(I - \beta P) = \begin{bmatrix} 1 - 0,5\beta & -0,5\beta \\ -0,4\beta & 1 - 0,6\beta \end{bmatrix},$$

$$(I - \beta P)^T = \begin{bmatrix} 1 - 0,5\beta & -0,4\beta \\ -0,5\beta & 1 - 0,6\beta \end{bmatrix}$$

and

$$(I - \beta P)^{-1} = \frac{1}{det\,(I - \beta P)} \begin{bmatrix} 1 - 0,6\beta & 0,5\beta \\ 0,4\beta & 1 - 0,5\beta \end{bmatrix},$$

$$det\,(I - \beta P) = (1 - 0,5\beta)\,(1 - 0,6\beta) = 1 - \frac{11}{10}\beta + \frac{\beta^2}{10} = (1 - \beta)\,(1 - 0,1\beta).$$

Finally

$$(I - \beta P)^{-1} = \begin{bmatrix} \frac{1-0,6\beta}{(1-\beta)(1-0,1\beta)} & \frac{0,5\beta}{(1-\beta)(1-0,1\beta)} \\ \frac{0,4\beta}{(1-\beta)(1-0,1\beta)} & \frac{1-0,5\beta}{(1-\beta)(1-0,1\beta)} \end{bmatrix}.$$

Now as an example, we decompose into partial fractions the first element of matrix $(1 - \beta P)^{-1}$.
We receive:

$$\frac{1 - 0,6\beta}{(1 - \beta)\,(1 - 0,1\beta)} = \frac{D_{11}^1}{(1 - \beta)} + \frac{D_{11}^2}{(1 - 0,1\beta)} = \frac{\frac{4}{9}}{(1 - \beta)} + \frac{\frac{5}{9}}{(1 - 0,1\beta)},$$

because

$$1 - 0,6\beta = D_{11}^1\,(1 - 0,1\beta) + (1 - \beta)\,D_{11}^2.$$

8

In this same way, we decompose the other three elements of matrix and we receive:

$$(I - \beta P)^{-1} = \begin{bmatrix} \frac{\frac{4}{9}}{(1-\beta)} + \frac{\frac{5}{9}}{(1-0,1\beta)} & \frac{\frac{5}{9}}{(1-\beta)} + \frac{-\frac{5}{9}}{(1-0,1\beta)} \\ \frac{\frac{4}{9}}{(1-\beta)} + \frac{-\frac{4}{9}}{(1-0,1\beta)} & \frac{\frac{5}{9}}{(1-\beta)} + \frac{\frac{4}{9}}{(1-0,1\beta)} \end{bmatrix}.$$

Finally formula (15) becomes the following form:

$$(I - \beta P)^{-1} = \frac{1}{(1-\beta)} \begin{bmatrix} \frac{4}{9} & \frac{5}{9} \\ \frac{4}{9} & \frac{5}{9} \end{bmatrix} + \frac{1}{1-0,1\beta} \begin{bmatrix} \frac{5}{9} & -\frac{5}{9} \\ -\frac{4}{9} & \frac{4}{9} \end{bmatrix}.$$

We can check that the first matrix with factor $1/(1-\beta)$ is ergodic matrix of Markov Process for given stochastic matrix of transition P. The second matrix is so named difference matrix. The sum of elements is equal zero in rows of this matrix. Taking formula (19) into consideration, we receive total finite expected reward:

$$\nu_\infty (\beta) = \left(\frac{1}{(1-\beta)} \begin{bmatrix} \frac{4}{9} & \frac{5}{9} \\ \frac{4}{9} & \frac{5}{9} \end{bmatrix} + \frac{1}{1-0,1\beta} \begin{bmatrix} \frac{5}{9} & -\frac{5}{9} \\ -\frac{4}{9} & \frac{4}{9} \end{bmatrix} \right) \cdot q.$$

Now we can find value $\nu_\infty (\beta)$ for two different β, $\beta_1 = 0,5$ and $\beta_2 = 0,99$. After providing of values and simple calculations we receive:

$$\nu_\infty (0,5) = \left(2 \begin{bmatrix} \frac{4}{9} & \frac{5}{9} \\ \frac{4}{9} & \frac{5}{9} \end{bmatrix} + 1,052 \begin{bmatrix} \frac{5}{9} & -\frac{5}{9} \\ -\frac{4}{9} & \frac{4}{9} \end{bmatrix} \right) \cdot \begin{bmatrix} 6 \\ -3 \end{bmatrix}.$$

Hence we obtain for the starting state and $n \to \infty$

$$\nu_{1,\infty} (0,5) = 2 \cdot 1 + 1,052 \cdot 5 = 7,260,$$

and for the second state

$$\nu_{2,\infty}\left(0,5\right) = 2 \cdot 1 - 1,052 \cdot 4 = -2,208.$$

For $\beta = 0,99$ we obtain

$$\nu_{\infty}\left(0,99\right) = 100 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 5 \\ -4 \end{bmatrix},$$

and

$$\nu_{1,\infty}\left(0,99\right) = 100 + 5 = 105,$$

$$\nu_{2,\infty}\left(0,99\right) = 100 - 4 = 96.$$

Example 2
Let $N = 3$ and

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0,5 & 0,5 \end{bmatrix}, \quad R = \begin{bmatrix} 9 & 8 & 3 \\ 3 & -7 & -2 \\ 5 & -9 & 3 \end{bmatrix}.$$

Hence

$$q = \begin{bmatrix} -8 \\ 3 \\ -3 \end{bmatrix}.$$

Further calculations give us the following results:

$$(I - \beta P) = \begin{bmatrix} 1 & \beta & 0 \\ -\beta & 1 & 0 \\ 0 & 0,5\beta & 1 - 0,5\beta \end{bmatrix},$$

10

$$det\,(I - \beta P) = (1 - 0,5\beta)\left(1 - \beta^2\right) = (1 - \beta)\,(1 + \beta)\,(1 - 0,5\beta)$$

$$(I - \beta P)^{-1} = \begin{bmatrix} \frac{1}{1-\beta} & \frac{\beta}{1-\beta^2} & 0 \\ \frac{\beta}{1-\beta^2} & \frac{1}{1-\beta^2} & 0 \\ \frac{0,5\beta^2}{(1-\beta^2)(1-0,5\beta)} & \frac{0,5\beta^2}{(1-\beta^2)(1-0,5\beta)} & \frac{1}{1-0,5\beta} \end{bmatrix} = \frac{1}{1-\beta}\begin{bmatrix} 0,5 & 0,5 & 0 \\ 0,5 & 0,5 & 0 \\ 0,5 & 0,5 & 0 \end{bmatrix} +$$

$$+ \frac{1}{1+\beta}\begin{bmatrix} 0,5 & -0,5 & 0 \\ -0,5 & 0,5 & 0 \\ \frac{1}{6} & -\frac{1}{6} & 0 \end{bmatrix} + \frac{1}{1-0,5\beta}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{2}{3} & -\frac{1}{3} & 1 \end{bmatrix}.$$

Hence

$$\nu_\infty\,(\beta) = (1 - \beta \cdot P)^{-1}\,q = \left(\frac{1}{1-\beta}\,[.] + \frac{1}{1+\beta}\,[\ldots] + \frac{1}{1-0,5\beta}\,[\ldots]\right)\cdot\begin{bmatrix} 8 \\ 3 \\ -3 \end{bmatrix} =$$

$$= \frac{1}{1-\beta}\begin{bmatrix} 5,5 \\ 5,5 \\ 5,5 \end{bmatrix} + \frac{1}{1+\beta}\begin{bmatrix} 2,5 \\ -2,5 \\ \frac{5}{6} \end{bmatrix} + \frac{1}{1-0,5\beta}\begin{bmatrix} 0 \\ 0 \\ -\frac{28}{3} \end{bmatrix}.$$

Now we can calculate total finite expected rewards for given values β, $\beta_1 = 0,5$ and $\beta_2 = 0,99$. For $\beta_1 = 0,5$ we obtain

$$\nu_\infty\,(0,5) = \frac{1}{1-0,5}\begin{bmatrix} 5,5 \\ 5,5 \\ 5,5 \end{bmatrix} + \frac{1}{1+0,5}\begin{bmatrix} 2,5 \\ -2,5 \\ \frac{5}{6} \end{bmatrix} + \frac{1}{1-0,5\cdot 0,5}\begin{bmatrix} 0 \\ 0 \\ -\frac{28}{3} \end{bmatrix},$$

and next

$$\nu_{1,\infty}\,(0,5) = 2 \cdot 5,5 + 0,666 \cdot 2,5 + 1,333 \cdot 0 = 11 + 1,666 + 0 = 12,666,$$
$$\nu_{2,\infty}\,(0,5) = 2 \cdot 5,5 - 0,666 + 0 = 9,334,$$

$$\nu_{3,\infty}\left(0,5\right) = 11 + 0,666 \cdot \frac{5}{6} + 1,333 \cdot \left(-\frac{28}{3}\right) = 11 + 1,388 - 12,441 = -0,053.$$

Finally

$$\nu_{\infty}\left(0,5\right) = \begin{bmatrix} 12,666 \\ 9,334 \\ -0,053 \end{bmatrix}.$$

For $\beta_2 = 0,99$ we obtain

$$\nu_{\infty}\left(0,99\right) = \frac{1}{1-0,99}\begin{bmatrix} 5,5 \\ 5,5 \\ 5,5 \end{bmatrix} + \frac{1}{1+0,99}\begin{bmatrix} 2,5 \\ -2,5 \\ \frac{5}{6} \end{bmatrix} + \frac{1}{1-0,5\cdot 0,99}\begin{bmatrix} 0 \\ 0 \\ -\frac{28}{3} \end{bmatrix} =$$

$$= 100\begin{bmatrix} 5,5 \\ 5,5 \\ 5,5 \end{bmatrix} + 0,502\begin{bmatrix} 2,5 \\ -2,5 \\ \frac{5}{6} \end{bmatrix} + 1,980\begin{bmatrix} 0 \\ 0 \\ -\frac{28}{3} \end{bmatrix}$$

and the result

$$\nu_{1,\infty}\left(0,99\right) = 100\cdot 5,5 + 0,502\cdot 2,5 + 1,980\cdot 0 = 550 + 1,255 + 0 = 551,255,$$

$$\nu_{2,\infty}\left(0,99\right) = 100\cdot 5,5 - 0,502\cdot 2,5 + 0 = 548,745,$$

$$\nu_{3,\infty}\left(0,99\right) = 550 + 0,502\cdot\frac{5}{6} + 1,980\cdot\frac{28}{3} = 531,938.$$

Finally

$$\nu_{\infty}\left(0,99\right) = \begin{bmatrix} 551,255 \\ 548,745 \\ 531,938 \end{bmatrix}.$$

5   Conclusions and comments

Identical results would be obtained for $\nu_\infty(\beta)$ if we use directly formula (10) (passing over difficulties connected with inverse matrix $(I - \beta P)$). Using formula (19) we can separate two components of total reward; a constant component connected with $1/(1 - \beta)$ factor and ergodic matrix and variable component which represents this part of $\nu_\infty(\beta)$ which arises under the influence of unsteady transient process. Effect of this process is especially visible during the initial phase of Markov Decision Process. Value of this part of component of quantity $\nu_\infty(\beta)$ rises along with decrease of discounted factor β and increase of disturbances which are generated by matrix P. Two presented examples show it.

We can create, relying on above observations, a performance index of tested Discounted Markov Decision Processes. Let $\nu_\infty^1(\beta)$ mean component which stand in front of $\frac{1}{(1-\beta)}$, and $\nu_\infty^k(\beta)$, $k = 2, 3, \ldots, N$ mean components which stand in front of $\frac{1}{1-\alpha_K\beta}$. Then performance index of mentioned Markov Chain can have the following form:

$$J(\beta) = [J_i(\beta)] = \left[\frac{\sum_{K=2,3,\ldots} \nu_{\infty,i}^K(\beta)}{\nu_{\infty,i}^1(\beta)} \right]. \tag{20}$$

From definition of coefficient $J(\beta)$ for given β results that when absolute value of this coefficient is more close zero, then better properties have tested Markov Process. For presented two examples, values of coefficients amount to:

$$J(0,5) = \begin{bmatrix} 0,151 \\ -0,151 \\ -1,005 \end{bmatrix}, \quad J(0,5) = \begin{bmatrix} 2,630 \\ -2,104 \end{bmatrix},$$

$$J(0,99) = \begin{bmatrix} 0,002 \\ -0,002 \\ -1,032 \end{bmatrix}, \quad J(0,99) = \begin{bmatrix} 0,05 \\ -0,04 \end{bmatrix}.$$

We observe that given stochastic matrix P always generates the same transient process for $n = 0, 1, 2$ (it means P^n). It results from calculation that effect of the process depends on value of β. Hence optimisation of Discounted Markov Decision Process can rely on selection of adequately large factor β for given quality coefficient. But usually β is given and depends on different economic–technical conditions. Then optimisation MDP can rely on selection of adequately matrix

P, it means, the control strategy should minimize quality coefficient if the factor β is given.

These aspects mentioned above will be a subject area of next papers. It results from analyses of MDP some conclusions:

1. Formula (19) allows to calculate in analytical way value of total reward $\nu_\infty^k(\beta)$ without difficult process of reverse of the matrix $(I - \beta P)$. Reverse of matrices using computer technology goes on in iterative way. The number of iteration rises along with the size N of matrix rapidly. It leads to loss of calculation's accuracy.

2. Proposed analytical method of calculation of ergodic and difference matrices gives us the possibility of selection of two components of total reward. It increases the possibility of analysis of Discounted Markov Decision Process.

REFERENCES:

1. Altman E., Shwartz A.: Markov Decision Problems and State – Actions Frequencies,SIAM Journal on Control and Optimization, 29, 1991.

2. Arapostathis A., Borkar V.S., Fernandez-Gaucherand E., Ghosh M.K. and Marcus S.I.: Discrete -Time Controlled Markov Processes with an Average Costs Criterion: A Servey. SIAM Journal on Control and Optimization. 31, 1993.

3. Bertsekas P.: Dynamic Programming and Stochastic Control. Acad. Press, New York,1976.

4. Bewley T., Kohlberg E.: The Asymptotic Theory of Stochasic Games. Mathematics of Oper. Res., 1, 1976.

5. Bewley T., Kohlberg E.: On Stochastic Games with Stationary Optimal Strategies. Mathematics of Oper. Res., 3, 1978.

6. Breton M., Filar J.A., Haurie A. and Shultz T.A.: On the Computation of Equilibria in Discounted Stochastic Games. In T. Basar, editor. Dynamics Games and Applications in Economics. Springer – Verlag, Berlin,1985. Lecture Notes in Economics and Mathematical Systems, 265.

7. Chan L.M.: A Markovian approach to the study of the canadian cattle industry. Reprint, ser. 162, University Toronto, Canada, 1981.

8. Chan L.M., Gal S.: A Markovin model for a perishable product invertory. Man. Science,23, 1985.

9. Chrzan P.: O pewnym algorytmie rozwiązania zadania optymalizacji niejednorodnego okresowego łańcucha Markowa. PN.351, AE-Wrocław, 1986.

10. Chrzan P.: Iteracyjny algorytm wyznaczania macierzy ergodycznej i fundamentalnej dla skończonego łańcucha Markowa. Problemy budowy i zastosowań modeli ekonomicznych. SGPiS-Warszawa, 1987.

11. Chrzan P.: Model Howard'a dla niejednorodnego okresowego łańcucha Markowa. Część I. Przegląd Statystyczny, 35/2, PWN, Poznań, 1988.

12. Chrzan P.: Model Howard'a dla niejednorodnego okresowego łańcucha Markowa. Część II. Przegląd Statystyczny, 35/3, PWN, Poznań, 1988.

13. Chrzan P.: Stacjonarny łańcuch Markowa z czasem dyskretnym. Zadanie optymalizacji z dyskontem. Program w języku FORTRAN dla IBM-PC, metoda iteracji w przestrzeni polityk. Working Paper, Instytut Ekonomii, AE- Katowice, 1990.

14. Chrzan P.: Markov Decisin Chains and their Applications in Economics. AE-Katowice , 1990 (In Polish).

15. Dirven C.A.J.M., Vrieze O. J.: Advertising Models, Stochastic Games and Myopic Strategies. Oper. Res., 34, 1986.

16. Federgruen A.: Markovia Control Problems. Mathematical Centre Treats 97, Amsterdam, 1983.

17. Filar A.J., Shiltz T.A.: Nonlinear Programming and Stationary Strategies in Stochastic Games. Mathematical Programming, 35, 1986.

18. Filar J.A., Vrieze K.: Competitive Markov Decision Processes. Springer - Verlag, New York , Inc, 1997.

19. Foguel S.R.: The ergodic theory of Markov processes. Van Nostrand Renhold Company, New York, 1969.

20. Ghellinck G.T., Eppem G.D.: Linear programming solution for separable Markovian decision problems. Man Science, 13, 1967.

21. Gillette D.: Stochastic Games with Zero Stop Probabilities. In A.W. Tucker, M. Dresher and P. Wolfe, editors: Contributions to the Theory of Games. Princeton University Press, Princeton, New Jersey, 1957. Annals of Mathematics Studies. 39.

22. Goldwerger J.: Dynamic programming for a stochastic Markovian processes with an application to the mean variance models. Man Science 23, 1977.

23. Hauser J.R., Wiśniewski K.J.: Dynamic analysis of consumer response to marketing strategies. Man. Science 28, 1982.

24. Hordijk A.: Dynamic programming and Markov potential theory. MCT 51, Amsterdam, 1974.

25. Howard R.: Dynamic programming and Markov processes. John Wiley, New York, 1960.

26. Jędrzejowicz P.: Wybrane modele decyzyjne w produkcji i eksploatacji . WkiŁ, Warszawa, 1981.

27. Kałuski J.: Markov Processes in the Control of Discrete Events Industrial Processes. ZN. Politechniki Śląskiej. Seria: AUTOMATYKA,z.108, Gliwice, 1994 (In Polish).

28. Kałuski J.: An n-person stochastic games with coalitions. ZN. Politechniki Śląskiej. Seria: AUTOMATYKA, z.123, Gliwice, 1998. (In polish).

29. Kao E.P.: A preference order dynamic prgram for a stochastic traveling solesman problem. Oper. Res., 26, 1978.

30. Lee T.C., Judge G.G., Zellner A.: Estimating the parameters of the Markov probability model from aggregate time series data. North Holland , Amsretdam, 1977.

31. Liggett T., Lipman S.: Stochastic Games with Perfect Information and Time Average Payoff. SIAM Review, 11, 1969.

32. Meridith J.: A Markovian analysis of a geriatric wards. Man. Science 19, 1973.

33. Mertens J.F.: Stochastic Games . In R.J. Aumann and S.Hart, editors. Handbook of Game Theory with Economic Application. North-Holland, Amsterdam, 1992.

34. Puterman M., Shin M.: Modified policy iterration algorithm for discounted Markov decision problems. Man. Science, 24, 1978.

35. Puterman M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley, New York, 1994.

36. Shapley L.S.: Stochastic Games.Proceedings National Acad. Of Science, USA, 39, 1953.

37. Van Der Wal J.: Discounted Markov Games: Successive Approximations and Stopping Times. Internationale Journal of Game Theory. 6, 1977.

38. Van Der Wal J.: Stochastic Dynamic Programming, Successive Approximation and Nearly Optimal Solutions for Markov Decision Processes and Markov Games. Technical Report. 1981.

39. Vorob'iov N.N.: Game Theory. Lectures for Economists and Systems Scientists. Springer-Verlag, Berlin, 1977.

40. Vrize O.J.: Linear Programming and Undiscounted Stochastic Games. OR Spectrum, 3, 1981.

41. White D.J.: Real Applications of Markov Decision Processes. Interfaces, 15 (6), 1985.

42. White D.J.: Markov Decision Processes. John Wiley, Chichester, 1993.