Tomáš Kojecký; Petr Mayer
A modification of a class of IAD methods

In: Jan Chleboun and Petr Přikryl and Karel Segeth (eds.): Programs and Algorithms of Numerical Mathematics, Proceedings of Seminar. Dolní Maxov, June 6-11, 2004. Institute of Mathematics AS CR, Prague, 2004.
pp. 116–130.

Persistent URL: http://dml.cz/dmlcz/702784

A MODIFICATION OF A CLASS OF IAD METHODS *

Tomáš Kojecký, Petr Mayer

Abstract

We provide a short overview of algorithms useful for computing of stationary probability vectors of stochastic matrix. Some care is devoted to the problem of computing of all extremal stationary probability vectors for the reducible stochastic matrices. We present some modifications of standard Iterative Aggregation/Disaggregation algorithm.

1. Introduction

Our aim is to provide (not complete) some overview of methods available for computing stationary probability vectors of stochastic matrices. We do not suppose any information about block structure of these matrices.

Definition 1 *Matrix $T \in \Re^{N\times N}$ with nonnegative elements such that the sum of all elements in each column is equal to one is* column stochastic matrix.

We solve the following problem: Find all solutions $x \in \Re^N$ of

$$Tx = x \quad \text{with conditions} \quad x \geq 0 \quad \text{and} \quad \sum_{i=1}^{N} x^{(i)} = 1\,. \tag{1}$$

For simplicity let us introduce the vector of all ones $e(n) \in \Re^n$. Similarly by $e_i(n) \in \Re^n$ we denote i-th vector of canonical base, i.e.

$$e_i(n) = \left(\begin{array}{ccccccc} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ & & & i-\text{th element} & & & \end{array} \right).$$

Furthermore we denote by $I(n)$ the identity matrix and by $E(n,m) \in \Re^{n\times m}$ the matrix of all ones. When the dimension is clear from the context we omit the indication of the dimension. By $r(A)$ we denote spectral radius of the matrix A , by $\sigma(A)$ we denote its spectrum and elements of vectors we denote by upper indices in brackets.

Definition 2 *Any solution of the problem (1) is called* stationary probability vector (SPV).

*This work was supported by the grants GAČR 201/02/0684, GAČR 201/02/0595 and MŠM J13/98:113200007.

From the Perron-Frobenius theorem (see [5]) we know that such solution always exists, and this solution is unique when the matrix T is irreducible . General case is solved by the following lemmas.

Lemma 1 *For any stochastic matrix T, there exists a permutation matrix P such that*

$$P^T T P = \begin{pmatrix} F_0 & 0 & 0 & \dots & 0 \\ F_1 & G_1 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ F_p & 0 & \dots & 0 & G_p \end{pmatrix} \tag{2}$$

where F_0 is zero convergent matrix, $F_i, i > 0$ are nonnegative matrices and $G_1, \dots, G_p$ are irreducible stochastic matrices.

Definition 3 *The matrix of the form (2) is called* Romanovski canonical form *of the stochastic matrix T.*

Let us note, that the number of the irreducible blocks in the Romanovski form of a stochastic matrix is the number of linearly independent stationary probability vectors. Moreover, if we wish to compute other significant characteristics of Markov chain which is described by the stochastic matrix T, such as the probability of reaching one state from another state, the mean time (in transition steps of Markov chain) or the variance of the mean time, the knowledge of the Romanovski form gives us the information of zero/nonzero structure of matrices we are looking for.

For computing such an information, we can use the Tarjan algorithm (see [1]). Drawback of this method is that one needs a direct access to elements of matrix studied. In many situations, for example if a matrix is given in a form of sum of tensor products, it is impossible to find an effective way to obtain elements of T. The other possibility is to exploit knowledge of existence of the Romanovski canonical form, to compute the special base of stationary probability vectors and to decipher the structure of the Romanovski form from it.

We show a simple lemma which gives a connection between the Romanovski canonical form and the stationary probability vectors.

Lemma 2 *Let T be stochastic matrix with the Romanovski form (2), then vector u_k of the form*

$$u_k = P\tilde{u}_k \tag{3}$$

where

$$\widetilde{u}_k = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \widetilde{\widetilde{u}}_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{and} \quad G_k\widetilde{\widetilde{u}}_k = \widetilde{\widetilde{u}}_k,\ e^T\widetilde{\widetilde{u}}_k = 1 \tag{4}$$

is a solution of the problem (1).

Definition 4 *The eigenvectors defined by the Lemma 2 are called* extremal stationary probability vectors.

Lemma 3 *Let T be stochastic matrix, then any stationary probability vector (solution of the problem (1)) is a convex combination of extremal stationary probability vectors.*

Proof From the Romanovski canonical form we know that the eigenvalue one has its algebraic multiplicity p and on the other hand we have p linearly independent eigenvectors for the eigenvalue one defined by (3) and (4). □

For convenience we define M-matrix.

Definition 5 *[6, Def. 3.22] A real $n \times n$ matrix $A = (a_{ij})$ with $a_{ij} \leqslant 0$ for all $i \neq j$ and $A^{-1} \geqslant 0$ is an M-*matrix.

The paper is organized as follows: The second section describes some possibility how to compute the Romanovski form utilizing SPV. Sections three to five give partial overview of methods useful for computing SPV's. Section three deals with standard iterative methods based on regular splittings, section four describes standard IAD methods and shows possible overcome of necessity of irreducibility of the transition matrix T. Section five is devoted to modification of IAD methods which authors hope is promising. The sixth section contains some numerical examples and the paper finish by conclusions.

2. The construction of Romanovski form from the SPV's

A substantial part of constructing the Romanovski form is to find such index sets in the original matrix which are mapped to separate blocks of the transformed matrix. To determine behavior of the graph of the matrix T, we will use the following lemma

Lemma 4 *Let $x_0 = e_i$, let $x_{k+1} = Tx_k$ for $k = 0, \ldots, \infty$. If $x_k^{(j)} > 0$ for some $k > 0$, then there is a path in the graph of matrix T connecting i and j.*

One possibility for the construction of the Romanovski form is the following approach: denote $I_k, k = 0, \ldots, p$ the indices which belong to the k-th block, then the steps 1.-14. follow:

1. set $p := 0$, $I_0 := \emptyset$, $I := \{1, \ldots, N\}$, $i = 1$
2. while $I \neq \emptyset$ do
3. set $x_0 = e_i$ and find $x = \lim_{k\to\infty} x_k$ where $x_{k+1} = \frac{1}{2}(I + T)x_k$
4. if $x^{(i)} \neq 0$ then

   (i is the element of some of the index set of irreducible matrices)
5. $p := p + 1$
6. $I_p := \{j \in I; x^{(j)} \neq 0\}$
7. $I := I - I_p$
8. $i := \min\{j \in I\}$
9. else

   (i is the element of the index set of transient states)
10. $I_0 := I_0 \cup \{i\}$
11. $I := I - \{i\}$
12. $i := \min\{j \in I; x^{(j)} \geqslant x^{(k)}, k \in I\}$
13. end if
14. end while

Note 1 We can increase the number of elements added in step 10., by taking in account, that any index l for which $x^{(l)}$ is zero and $x_k^{(l)}$ is nonzero in at least in one iteration of the third step of the algorithm, i.e. if $x_k^{(l)} > 0$ for some k, then l is element of I_0, too.

Deeper analysis of computing of the Romanovski canonical form of a stochastic matrix by use of stationary probability vectors can be found in Tanabe [2].

From previous observation we see, that quicker method which is able to compute SPV for a reducible matrix with the same behavior concerning the zero/nonzero structure as power method is needed.

3. Methods based on regular splitting

The problem (1) is to solve the singular system of linear equations

$$(I - T)x = 0 \quad \text{with the condition} \quad e^T x = 1\,.$$

Let us consider a regular splitting $M - W$ of the matrix $I - T$, where M is an M-matrix and W is a nonnegative matrix. Then we will investigate two types of iterative methods.

The first one computes directly the SPV:

Algorithm MW (input: $T, M, W, x_{\text{init}}, \varepsilon$; output: x)

1. set
$$A := M^{-1}W \quad \text{and} \quad \tilde{A} := \frac{1}{2}\,(I + A) \tag{5}$$
$$x_0 := x_{\text{init}} \quad k := 0$$
2. while $\|Tx_k - x_k\| > \varepsilon$ do
3. $x_{k+1} := \widetilde{A}x_k$
4. $k := k + 1$
5. end while
6. $x := x_k$

The second one computes transformed SPV and then transforms it back.

Algorithm WM (input: $T, M, W, x_{\text{init}}, \varepsilon$; output: x)

1. set
$$B := WM^{-1} \text{ and } \tilde{B} := \frac{1}{2}(I + B) \tag{6}$$
2. $y_0 := Mx_{\text{init}}$
3. $x_0 := x_{\text{init}}$
4. $k := 0$
5. while $\|Tx_k - x_x\| > \varepsilon$ do
6. $x_k := M^{-1}y_k$
7. $k := k + 1$
8. $y_k := \frac{1}{2}(Wx_{k-1} + y_{k-1})$ (i.e. $y_k := \widetilde{B}y_{k-1}$)

9. end while

10. $x := x_k$

Let us note, that the vectors x_k generated by Algorithm MW and WM are identical.

Theorem 1 *The iteration matrices A, B and $\tilde{A}, \tilde{B}$ defined by (5) and (6) used in algorithms MW and WM are similar. Moreover, matrices $B, \tilde{B}$ are stochastic and in the case of matrices $\tilde{A}, \widetilde{B}$, the eigenvalue one is the only eigenvalue with absolute value one.*

Proof Because M is M-matrix, M is invertible, and we have similarity

$$M(M^{-1}W)M^{-1} = WM^{-1}$$

and

$$M\left[\frac{1}{2}(I + M^{-1}W)\right]M^{-1} = \frac{1}{2}(I + WM^{-1}).$$

To show stochasticity we use the fact that matrices W, M^{-1} are both nonnegative and since $I - T = M - W$ we obtain

$$\begin{aligned} 0 &= e^T(I - T) \\ 0 &= e^T(M - W) \\ 0 &= e^T M - e^T W \\ e^T M &= e^T W \\ e^T &= e^T W M^{-1}. \end{aligned}$$

The last part of Theorem follows from (5) and (6). □

We note, that for any choice of M and a stopping residual ε sufficiently small, the nonzero structure of the computed solution is the same as nonzero structure of the vector x computed in step 3. of the algorithm for construction of Romanovski form of the stochastic matrix T.

4. IAD methods

For introducing Iterative Aggregation/Dissaggregation methods, we have to define aggregation mapping

$$g : \{1, \ldots, N\} \longrightarrow \{1, \ldots, n\}, n \ll N,$$

where n is the size of the coarse space.

The indices which are mapped to the same value of g define one aggregation group. The optimal choice of mapping g is difficult and often depends on further information about the solved problem. As we will see, the difference between two choices of g for the same transition matrix can be substantial.

By using the aggregation mapping we define restriction and prolongation matrices. The restriction matrix $R \in \Re^{n \times N}$ is defined by nonzero elements

$$R_{g(i),i} = 1,$$

i.e. $(Rx)^{(j)} = \sum_{i=1,g(i)=j}^{N} x^{(i)}$.

The prolongation matrix $S(x)$ is parametrised by the vector $x \in \Re^N$, the nonzero elements of this matrix are

$$(S(x))_{i,g(i)} = \frac{x^{(i)}}{(Rx)^{(g(i))}},$$

it means that $(S(x)z)^{(i)} = z^{(g(i))} \frac{x^{(i)}}{(Rx)^{(g(i))}}$.

Let us denote the aggregated matrix defined by the vector x and by the aggregation mapping g as $A(x) = RTS(x)$. Some properties of the matrix $A(x)$ follow.

Lemma 5 *Let T be a column stochastic matrix, let g be an aggregation mapping and $x \in \Re^N$ such that $x \geqslant 0$ and $Rx > 0$. Then the aggregated matrix $A(x)$ is column stochastic. If the matrix T is irreducible and the vector x is strictly positive, then $A(x)$ is irreducible.*

Proof To proof the stochasticity of $A(x)$ we have to show non-negativity of $A(x)$ and the stochastic property $e(n)^T A(x) = e(n)^T$.

Non-negativity follows directly from the fact that all matrices $R, S(x)$ and T are nonnegative and their product is nonnegative, too.

To prove the second part we use $e(n)^T R = e(N)^T$ and $e(N)^T S(x) = e(n)^T$ for any $x \geqslant 0$ which satisfies $Rx > 0$. Then simple computation gives $e^T(n)A(x) = e(n)^T RTS(x) = e(N)^T TS(x) = e(N)^T S(x) = e(n)^T$.

If T is irreducible, then there is no permutation matrix P, such that

$$P^T TP = \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix},$$

suppose that $A(x)$ is reducible, then $P_1^T A(x) P_1 = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$ for some permutation P_1. If $x > 0$ then any element of $A(x)$ is a positive combination of the corresponding elements of the matrix T. Then from the reducibility of $A(x)$, it follows the reducibility of T and this is the contradiction. □

Note 2 Let us note, that the strict positivity of x is essential. Let us consider the irreducible matrix T and a nonnegative vector x as follows

$$T = \begin{pmatrix} \frac{1}{3} & \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{2}{3} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} & \frac{1}{4} \end{pmatrix}, \quad x = \begin{pmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \\ 0 \end{pmatrix}, \quad g : \begin{matrix} 1 \mapsto 1 \\ 2 \mapsto 1 \\ 3 \mapsto 2 \\ 4 \mapsto 2 \end{matrix}.$$

We get matrix $A(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ which is reducible.

With the previous knowledge we can define IAD algorithm for the irreducible stochastic matrix T and for a positive initial approximation x_{init}.

Algorithm IAD (input: $T, M, W, x_{\text{init}}, \varepsilon, g, s$; output: x)

1. $k := 1, x_1 := x_{\text{init}}$
2. while $\|Tx_k - x_k\| > \varepsilon$ do
3. $\tilde{x} := (M^{-1}W)^s x_k$
4. $A(\tilde{x}) := RTS(\tilde{x})$
5. solve $A(\tilde{x})z = z$ and $e^T z = 1$
6. $k := k + 1$
7. $x_k = S(\tilde{x})z$
8. end while

The convergence theory for IAD can be found in [3]. We just state the main theorem:

Theorem 2 *Let T be a column stochastic matrix, let $\hat{x}$ be the solution of (1), then there exist s_0 and a neighborhood of $\hat{x}$ such that for any x_{init} from this neighborhood and any $s > s_0$ the Algorithm IAD is convergent.*

In some practical problems we find that there are special cases, when the IAD algorithm returns exact solution in one step.

Theorem 3 *Let for splitting M, W be* $\text{range}(M^{-1}W) \subseteq \text{range}(S(\hat{x}))$*, then Algorithm IAD terminates after the first iteration.*

Proof see [4]. □

The next example ilustrates Theorem 3:

Example 1 Let

$$T = \begin{pmatrix} 0.1 & 0.1 & 0.1 & 0.05 & 0.15 & 0.25 \\ 0.5 & 0.2 & 0 & 0.02 & 0.06 & 0.10 \\ 0.2 & 0.1 & 0.1 & 0.03 & 0.09 & 0.15 \\ 0.04 & 0.12 & 0.16 & 0.2 & 0.2 & 0.1 \\ 0.08 & 0.24 & 0.32 & 0.6 & 0.2 & 0.1 \\ 0.08 & 0.24 & 0.32 & 0.1 & 0.3 & 0.3 \end{pmatrix}.$$

The splitting is $I - T = M - W$, where

$$M = \begin{pmatrix} 0.9 & -0.1 & -0.1 & 0 & 0 & 0 \\ -0.5 & 0.8 & 0 & 0 & 0 & 0 \\ -0.2 & -0.1 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.8 & -0.2 & -0.1 \\ 0 & 0 & 0 & -0.6 & 0.8 & -0.1 \\ 0 & 0 & 0 & -0.1 & -0.3 & 0.7 \end{pmatrix}$$

and

$$W = \begin{pmatrix} 0 & 0 & 0 & 0.05 & 0.15 & 0.25 \\ 0 & 0 & 0 & 0.02 & 0.06 & 0.10 \\ 0 & 0 & 0 & 0.03 & 0.09 & 0.15 \\ 0.04 & 0.12 & 0.16 & 0 & 0 & 0 \\ 0.08 & 0.24 & 0.32 & 0 & 0 & 0 \\ 0.08 & 0.24 & 0.32 & 0 & 0 & 0 \end{pmatrix}.$$

Let the aggregation mapping is

$$g : \begin{array}{l} 1 \mapsto 1 \\ 2 \mapsto 1 \\ 3 \mapsto 1 \\ 4 \mapsto 2 \\ 5 \mapsto 2 \\ 6 \mapsto 2 \end{array}$$

and the initial approximation is $x_0 = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})^T$, we get

$$\tilde{x} = (0.1041237, 0.1025773, 0.0845361, 0.1829060, 0.3094017, 0.3111111)^T.$$

The aggregated problem is $A(\tilde{x}) = \begin{pmatrix} 0.4849558 & 0.3319149 \\ 0.5150442 & 0.6680851 \end{pmatrix}$ and the aggregated solution is $z = (0.3918901, 0.6081099)^T$. We finish with

$$x_1 = (0.1401094, 0.1380285, 0.1137522, 0.1384420, 0.2341870, 0.2354809)^T$$

which is the exact

solution of $Tx = x$.

The main drawback of Algorithm IAD is that we need irreducibility of the transition matrix T and strict positivity of iterations x_k. We can solve this problem by modificating the prolongation operator $S(x)$. We define the new prolongation operator $\tilde{S}(x) = \lim_{\varepsilon \to 0} S(x + \varepsilon e)$ i.e.

$$\left(\tilde{S}(x)z\right)^{(i)} = \begin{cases} (z^{(g(i))}x^{(i)})/((Rx)^{(g(i))}) & \text{for} \quad (Rx)^{(g(i))} \neq 0\,, \\ 1/(\text{card}\,\{j : g(j) = g(i)\}) & \text{for} \quad (Rx)^{(g(i))} = 0\,. \end{cases}$$

To make the solution of the aggregated problem unique, we change the definition of the solution of the aggregated problem to $z = \lim_{l\to\infty}\left[\frac{1}{2}\left(I + RT\tilde{S}(\tilde{x})\right)\right]^l R\tilde{x}$. We finish with a slightly modified algorithm

Algorithm IAD1 (input: $T, M, W, x_{\text{init}}, \varepsilon, g, s$ output: x)

1. $k := 1, x_1 := x_{\text{init}}$
2. while $\|Tx_k - x_k\| > \varepsilon$ do
3. $\tilde{x} := (M^{-1}W)^s x_k$
4. $\tilde{A}(\tilde{x}) := RT\tilde{S}(\tilde{x})$
5. $z = \lim_{l\to\infty}\left[\frac{1}{2}\left(I + RT\tilde{S}(\tilde{x})\right)\right]^l R\tilde{x}$
6. $k := k + 1$
7. $x_k = \tilde{S}(\tilde{x})z$
8. end while
9. $x := x_k$

The limit in step 5. is used just only for the definition of z to be unique. Algorithm IAD1 can be used for any column stochastic matrix T and for any initial approximation x_{init}. It can be used for computing stationary vectors for general stochastic matrices. But unfortunately, this method can fail on computing all extremal eigenvectors. The next example shows unability of IAD1 to keep the same zero/nonzero structure as the power method. It shows that IAD1 cannot be used for computing of all SPV's.

Example 2 Let us consider the following transition matrix

$$T = \begin{pmatrix} 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix}$$

and the initial vector $x_0 = (1, 0, 0, 0, 0, 0)^T$. For the splitting $M = I$, $W = T$, the aggregation mapping

$$g : \begin{array}{l} 1 \mapsto 1 \\ 2 \mapsto 1 \\ 3 \mapsto 2 \\ 4 \mapsto 2 \\ 5 \mapsto 3 \\ 6 \mapsto 3 \end{array}$$

after one smoothing iteration we have $\tilde{x} = (0.5, 0.5, 0, 0, 0, 0)^T$, the restriction matrix is

$$R = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

and the prolongation matrix is

$$\tilde{S}(\tilde{x}) = \begin{pmatrix} 0.5 & 0 & 0 \\ 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \\ 0 & 0 & 0.5 \end{pmatrix}.$$

Then we have the aggregated matrix

$$\tilde{A}(\tilde{x}) = \begin{pmatrix} 0.75 & 0.25 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.25 & 0.75 \end{pmatrix}$$

which is irreducible and in step 5 we have $z = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T$. Then $x_1 = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})^T$. And we are not able to compute the extremal stationary probability vector.

5. Modification of IAD

In this section we introduce a different modification of IAD algorithm. Our aim is to make possible to use such algorithm for computing extremal eigenvectors and to keep advantages of the acceleration of computations given by the aggregation correction.

Algorithm MIAD (input: $T, M, W, x_{\text{init}}, \varepsilon, g, s, q, p$ output: x)

1. $x_0 := x_{\text{init}}, k := 0$
2. while $\|Tx_k - x_k\| > \varepsilon$ do
3. $v_0 := x_k$
4. for $i = 1, \ldots, s$ do
5. $v_i := (M^{-1}Wv_{i-1} + v_{i-1})/2$
6. end for
7. for $i = 1, \ldots, q$ do
8. for $j = 1, \ldots, p$ do

9. define vector $y_{(i-1)p+j}$ by

$$(y_{(i-1)p+j})^{(l)} = (v_{s+1-i})^{(l)} \text{ if } g(l) = j$$
$$(y_{(i-1)p+j})^{(l)} = 0 \text{ if } g(l) \neq j$$

10. end for

11. end for

12. define a new iteration $x_{k+1} \in \text{span}(y_1, \ldots, y_{\text{pq}})$ and $e^T x_{k+1} = 1$ such that

$$\|Tx_{k+1} - x_{k+1}\| = \min_{\substack{y \in \text{span}(y_1, \ldots, y_{\text{pq}}) \\ e^T y = 1}} \|Ty - y\| \tag{7}$$

13. $k := k + 1$

14. end while

15. $x := x_k$

Lemma 6 *Let for nonzero vector y be $(I - T)y = 0$ and $e^T y = 0$. Then vectors y^+ and y^- defined by $y_i^+ = \max(y_i, 0)$ and $y^- = y^+ - y$, are both eigenvectors for the eigenvalue 1. Moreover the matrix T is reducible.*

Proof Since the transition matrix T is nonnegative and due to Perron-Frobenius theorem, the spectral radius is equal to the largest eigenvalue. Moreover, when T is irreducible, the corresponding eigenvector is strictly positive. From $(I - T)y = 0$ we have $Ty = y$, the vector y is the eigenvector corresponding to the spectral radius and as a consequence the matrix T could not be irreducible. Lemma 2 gives us the basis for all eigenvectors of the reducible stochastic matrix. Then any stationary probability vector is a linear combination of the extremal stationary vectors. From their definition, the zero/nonzero structure of them is disjoint. Then both positive and negative parts of y have to be a linear combination of the different extremal stationary vectors. □

Note 3 Let us say, that neither y^+ nor y^- has to be the extremal stationary probability vector.

By the use of previous lemma we can stop MIAD algorithm either when the final residual is small enough, and we finish with one stationary probability vector, or with two stationary probability vectors $y^+/(e^T y^+)$ and $y^-/(e^T y^-)$ and additional information on reducibility of the transition matrix T.

Theorem 4 *For any stochastic matrix T there exists s_0 such that for any $s > s_0$, any $q \leqslant s$, any number of blocks p and any initial approximation x_{init} the MIAD is convergent.*

Proof We know that the matrix WM^{-1} is stochastic with the same stationary probability vectors. The same is valid for $(WM^{-1} + I)/2$. Then we have

$$(WM^{-1} + I)/2 = Q + Z, QZ = ZQ = O, r(Z) < 1, Q^2 = Q.$$

For the smoothing iteration matrix

$$(M^{-1}W + I)/2 = M^{-1}[(WM^{-1} + I)/2]M$$

and for the smoothed residual

$$\begin{aligned}
\|Tv_s - v_s\| &= \left\|(T - I)[(M^{-1}W + I)/2]^s x_k\right\| \\
&= \left\|(T - I)[M^{-1}(WM^{-1} + I)M/2]^s x_k\right\| \\
&= \left\|(T - I)M^{-1}[(WM^{-1} + I)/2]^s M x_k\right\| \\
&= \left\|(W - M)M^{-1}\left(Q + Z^s\right) M x_k\right\| \\
&= \left\|(WM^{-1} - I)\left(Q + Z^s\right) M x_k\right\| \\
&= \|2(Q + Z - I)(Q + Z^s)Mx_k\| \\
&= \|2Z^s(Z - I)Mx_k\| \\
&= \|Z^s 2(Q + Z - I)Mx_k\|.
\end{aligned}$$

For the original residual we have

$$\begin{aligned}
\|Tx_k - x_k\| &= \|(T - I)x_k\| \\
&= \|(W - M)x_k\| \\
&= \left\|(WM^{-1} - I)Mx_k\right\| \\
&= \|2(Q + Z - I)Mx_k\|.
\end{aligned}$$

Then we finish with

$$\|Tv_s - v_s\| \leqslant \|Z^s\| \|Tx_k - x_k\|.$$

Since x_{k+1} is the solution of the minimalization problem (7), we have $\|Tx_{k+1} - x_{k+1}\| \leqslant \|Tv_s - v_s\| \leqslant \|Z^s\| \|Tx_k - x_k\|$ and $r(Z) < 1$. Then we can choose s_0 such, that for any $s > s_0$ is $\|Z^s\| < 1 - \delta$ for some positive δ. Then we have

$$\|Tx_k - x_k\| < (1 - \delta)^k \|Tx_0 - x_0\|.$$

□

We finish with a generalization of Theorem 3.

Definition 6 *The matrix of the block form*

$$A = \begin{pmatrix} \sum_{i=1}^{k} u_{1i} v_{1i}^T \\ \vdots \\ \sum_{i=1}^{k} u_{pi} v_{pi}^T \end{pmatrix}$$

is called k-dyadic.

Theorem 5 *Let for splitting $M - W$ of the matrix $I - T$ be the matrix $M^{-1}W$ k-dyadic. If g conserves the same block structure as $M^{-1}W$ and $q \geqslant pk + p$ then MIAD algorithm is convergent in finite number of iterations.*

Proof Since the SPV belongs to $\mathrm{Rng}(A)$, the comdition

$$\mathrm{Rng}(A) \subset\subset \mathrm{Span}(y_1, \ldots, y_{\mathrm{pq}}) \tag{8}$$

is sufficient for the aggregation correction to compute the exact solution. The condition (8) is fulfilled if the condition $\mathrm{Rng}(A) \subset\subset \mathrm{Span}(v_{s+1-q}, \ldots, v_s)$ is valid. This condition is a consequence of the fact that the vectors $v_{s+1-i}, \ldots, v_s$ generate the Krylov space $K(M^{-1}W, v_{s+1-i})$. Since the rank of the matrix $M^{-1}W$ is less or equal to kp, the dimension of the Krylov space $K(M^{-1}W, v_{s+1-i})$ cannot be larger.

□

Let us note, that obviously the minimal q is just $k + 1$.

6. Numerical example

Consider matrix

$$T = \begin{pmatrix} 0.0094 & 0.0268 & 0.0288 & 0.0593 & 0.0954 & 0.7434 & 0.2903 & 0.1678 \\ 0.0280 & 0.0357 & 0.0504 & 0.0763 & 0.1908 & 0.0930 & 0.0683 & 0.1420 \\ 0.0280 & 0.0357 & 0.0360 & 0.0678 & 0.4770 & 0.0465 & 0.1025 & 0.3292 \\ 0 & 0.0089 & 0.1655 & 0.0339 & 0.1908 & 0.0465 & 0.0512 & 0.1356 \\ 0.7477 & 0 & 0 & 0 & 0.0065 & 0.0106 & 0.1055 & 0.0364 \\ 0.0935 & 0 & 0.3597 & 0.1695 & 0.0103 & 0.0173 & 0.1447 & 0.0578 \\ 0.0467 & 0 & 0 & 0 & 0.0086 & 0.0148 & 0.1276 & 0.0500 \\ 0.0467 & 0.8929 & 0.3597 & 0.5932 & 0.0206 & 0.0282 & 0.1099 & 0.0812 \end{pmatrix},$$

and

$$g : \begin{matrix} 1 \to 1 \\ 2 \to 1 \\ 3 \to 1 \\ 4 \to 1 \\ 5 \to 2 \\ 6 \to 2 \\ 7 \to 2 \\ 8 \to 2 \end{matrix}.$$

We use

$$M = \begin{pmatrix} 0.9906 & -0.0268 & -0.0288 & -0.0593 & 0 & 0 & 0 & 0 \\ -0.0280 & 0.9643 & -0.0504 & -0.0763 & 0 & 0 & 0 & 0 \\ -0.0280 & -0.0357 & 0.9640 & -0.0678 & 0 & 0 & 0 & 0 \\ 0 & -0.0089 & -0.1655 & 0.9661 & 0 & 0 & 0 & 0 \\ -0.7477 & 0 & 0 & 0 & 0.9935 & -0.0106 & -0.1055 & -0.0364 \\ -0.0935 & 0 & -0.3597 & -0.1695 & -0.0103 & 0.9827 & -0.1447 & -0.0578 \\ -0.0467 & 0 & 0 & 0 & -0.0086 & -0.0148 & 0.8724 & -0.0500 \\ -0.0467 & -0.8929 & -0.3597 & -0.5932 & -0.0206 & -0.0282 & -0.1099 & 0.9188 \end{pmatrix},$$

and initial approximation $x_{\text{init}} = (\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$. With such setting IAD method needs 20 iterations to reach the residual less then 10^{-14} .

But the matrix $M^{-1}W$ is 2-dyadic and Modified IAD method gives for $q = 3$ the exact solution after the first iteration.

7. Conclusions

We can use a wide range of methods for computing SPV. From the efficiency point of view we prefer using iterative aggregation/disaggregation methods. For standard IAD methods it is necessary to ensure irreducibility of the transition matrix T. It can be achieved by using of Tarjan's algorithm, but such approach needs access to elements of the matrix T. In many practical situations we can only compute product of a matrix and a vector. On the other side we can use knowledge of extremal SPV's to construct Romanovski canonical form. For such computations we can effectively use the presented modification of IAD methods.

References

[1] J.S. Duff, J.K. Reid: *An implementation of Tarjan's alghorithm for the block triangularization of a matrix.* ACM Transaction on Mathematical Software, 4(2), 1978, 137–147.

[2] K. Tanabe: *The conjugate gradient method for computing all the extremal stationary probability vectors of a stochastic matrix.* Ann. Ins. Statist. Math., 37 (Part B), 1985, 173–187.

[3] I. Marek, P. Mayer: *Iterative aggregation/disaggregation methods for computing some characteristic of of Markov chains.* In: Large Scale Scientific Computing, Third International Conference, LSSC 2001, Sozopol, Bulgaria, 2001, 68–82.

[4] I. Marek, P. Mayer: *Iterative aggregation/disaggregation methods for computing stationary probability vectors of a stochastic matrices can be finitely terminating.* International Journal of Differential Equations, 3, 2001, 301–313.

[5] W.J. Stewart: *Introduction to the numerical solution of Markov chains.* Princeton University Press, 1994.

[6] R.S. Varga: *Matrix iterative analysis.* Springer-Verlag Berlin Heidelberg, 2000.