Convergence of iterative aggregation/disaggregation methods based on splittings with cyclic iteration matrices

Ivo Marek[1,2], Petr Mayer[1,3], Ivana Pultarova[1,4],

[1] Department of Mathematics,
Faculty of Civil Engineering, Czech Technical University in Prague,
Thakurova 7, 166 29, Praha 6, Czech Republic
[2] marek@ms.mff.cuni.cz
[3] pmayer@mat.fsv.cvut.cz
[4] ivana@mat.fsv.cvut.cz

Abstract. Iterative aggregation/disaggregation methods (IAD) belong to competitive tools for computation the characteristics of Markov chains as shown in some publications devoted to testing and comparing various methods designed to this purpose. According to Dayar T., Stewart W.J. *Comparison of partitioning techniques for two-level iterative solvers on large, sparse Markov chains*. SIAM J. Sci. Comput.Vol **21**, No. 5, 1691-1705 (2000), the IAD methods are effective in particular when applied to large ill posed problems. One of the purposes of this paper is to contribute to a possible explanation of this fact. The novelty may consist of the fact that the IAD algorithms do converge independently of whether the iteration matrix of the corresponding process is primitive or not. Some numerical tests are presented and possible applications mentioned; e.g. computing the PageRank.

Keywords. Iterative aggregation methods, stochastic matrix, stationary probability vector, Markov chains, cyclic iteration matrix, Google matrix, PageRank.

1 Introduction

The paper is organized as follows. Section 2 contains notation and some faithful concepts as well as some new concepts useful to realize our goals and namely to support via theory the statements cited in the introductory section. The aggregation methods are treated in connection with either the Leontev models or stochastic modeling. Some possible generalizations of the notion stochastic matrix is presented in Section 3. After a brief introduction of a class of IAD methods in Section 4 and Section 5 some of their basic properties are surveyed in Section 6. A proof of our main result is presented in Section 7 and its subsections. Some experiments documenting the theory are collected in Section 10. There are two categories of experiments there. The first one concerns convergence properties of IAD methods for randomly chosen stochastic matrices in order to show the

Dagstuhl Seminar Proceedings 07071
Web Information Retrieval and Linear Algebra Algorithms
http://drops.dagstuhl.de/opus/volltexte/2007/1067

influence of cyclicity. The experiments of the second category are devoted to applications of IAD methods to PageRank computations with a sample of a part of the Google matrix. A special technique of reordering the blocks is utilized in order to accelerate the convergence. Section 11 contains some conclusions.

2 Definitions and notation

As standard, we denote by $\rho(C)$ the spectral radius of matrix C, i.e.

$$\rho(C) = \mathrm{Max}\,\{|\lambda| : \lambda \in \sigma(C)\}\,,$$

where $\sigma(C)$ denotes the spectrum of C. Further we define the quantity

$$\gamma(C) = \sup\,\{|\lambda| : \lambda \in \sigma(C), \lambda \neq \rho(C) = 1\}\,.$$

We are going to call $\gamma(C)$ *the convergence factor* of C. We also need another more general characteristic of convergence, therefore we introduce

1 Definition For any $N \times N$ matrix $C = (c_{jk})$, where $c_{jk}, j, k = 1, ..., N$, are complex numbers, let us define quantity

$$\tau(C) = \mathrm{Max}\{|\lambda| : \lambda \in \sigma(C), |\lambda| \neq \rho(C)\}.$$

This quantity is called *spectral subradius* of C.

2 Remark Let C be any $N \times N$ matrix. Then obviously,

$$\rho(C) \geq \gamma(C) \geq \tau(C)$$

with possible strict inequalities in place of the nonstrict ones.

3 g-stochastic matrices

We are going to generalize slightly the concept of stochastic matrix. Obviously, every stochastic matrix B is g-stochastic, where $g = e = (1, ..., 1)^T$.

3 Definition *Suppose $g \in \mathcal{R}^N$ is such that $g^T = (g_1, ..., g_N)$ and that $g_j > 0$ for all $j = 1, ..., N$. Matrix $C = (c_{jk})$, where $c_{jk} \geq 0, j, k = 1, ..., N$, is called g-stochastic if*

$$C^T g = g.$$

It is easy to see that the class of all g-stochastic matrices shares most of the properties typical for stochastic matrices. In particular,

4 Proposition *Suppose C is a g-stochastic matrix. The following statements hold true*

a) Spectral radius $\rho(C) = 1$.

b) Possible eigenvalues of C, let us denote them by $\lambda_1, ..., \lambda_p$, for which $|\lambda_j| = 1$, are simple poles of the resolvent matrix $(\lambda I - C)^{-1}$.
c) The Perron projection Q can be expressed as

$$Q = \lim_{M\to\infty} \frac{1}{M} \sum_{k=1}^{M} C^k \tag{3.1}$$

Furthermore,

$$Q = CQ = QC, \tag{3.2}$$

and Q is componentwise nonnegative. Moreover, if C is irreducible, then Q is componentwise strictly positive. In the latter case $Q = xg^T$, where x is a unique eigenvector of C corresponding to eigenvalue 1 and g comes from the definition.

5 Remark The concept just introduced can be met quite frequently. We will see in the following sections, in particular in Section 9 that iteration matrices governing most of the iterative aggregation/disaggregation processes to computing stationary probability vectors of stochastic matrices are g-stochastic not necessarily stochastic matrices.

4 Iterative aggregation/disaggregation methods

Let $\mathcal{E} = \mathcal{R}^N, \mathcal{F} = \mathcal{R}^n, n < N$, $e^T = e(N)^T = (1, ..., 1) \in \mathcal{R}^N$. Let $\mathcal{G}$ be a map defined on the index sets:

$$\mathcal{G} : \{1, ..., N\} \overset{onto}{\rightarrow} \{1, ..., n\}$$

By means of $\mathcal{G}$ IAD communication operators are defined as

$$(Rx)_j = \sum_{\mathcal{G}(j)=j} x_j$$

$$S = S(u),\ (S(u)z)_j = \frac{u_j}{(Ru)_j}(Rx)_j.$$

We obviously have

$$RS(u) = I_{\mathcal{F}}$$

For the *aggregation projection* $P(x) = S(x)R$

$$P(x)^T e = e \quad \forall x \in \mathcal{R}^N, x_j > 0, j = 1, ..., N$$

and

$$P(x)x = x \quad \forall x \in \mathcal{R}^N, x_j > 0, j = 1, ..., N \tag{4.3}$$

Define the *aggregated matrix* as

$$\mathcal{B}(x) = RBS(x)$$

How to choose the map $\mathcal{G}$?

Naturally, if B is in "suitable" block form

$$B = \begin{pmatrix} B_{11} & B_{12} & \dots & B_{1n} \\ B_{21} & B_{22} & \dots & B_{2n} \\ . & . & \dots & . \\ B_{n1} & B_{n2}, & \dots & B_{nn} \end{pmatrix}, \text{ with diagonal } n_j \times n_j \text{ block } B_{jj},\ j = 1, ..., n.$$

Then one lets usually,

$$\overline{j} = \mathcal{G}(j) \text{ for } n_0 + n_1 + ... + n_{j-1} + 1 \leq j \leq n_0 + n_1 + ... + n_{j-1} + n_j,\ n_0 = 0.$$

and this means that each of the blocks B_{jk} is aggregated to a 1×1 matrix.

Conversely, the map $\mathcal{G}$ gives a rise to appropriate block form of B (up to a permutation).

6 Example *Let B be a stochastic matrix written in a block form*

$$B = \begin{pmatrix} B_{11} & B_{12} & \dots & B_{1n} \\ B_{21} & B_{22} & \dots & B_{2n} \\ . & . & \dots & . \\ B_{n1} & B_{22} & \dots & B_{nn} \end{pmatrix}$$

Take $N = 4$ and $n = 2$ and each block B_{jk} to be 2×2. Then, choosing

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

and

$$S(x) = \begin{pmatrix} \frac{1}{x_1+x_2}x_1 & 0 \\ \frac{1}{x_1+x_2}x_2 & 0 \\ 0 & \frac{1}{x_3+x_4}x_3 \\ 0 & \frac{1}{x_3+x_4}x_4 \end{pmatrix}$$

the aggregated matrix becomes

$$\mathcal{B}(x) = \begin{pmatrix} \beta_{\overline{11}}(x) & \beta_{\overline{12}}(x) \\ \beta_{\overline{21}}(x) & \beta_{\overline{22}}(x) \end{pmatrix}$$

$$\beta_{\overline{11}}(x) = \frac{1}{x_1 + x_2}[(b_{11} + b_{21})x_1 + (b_{12} + b_{22})x_2]$$

$$\beta_{\overline{12}}(x) = \frac{1}{x_3 + x_4}[(b_{13} + b_{23})x_3 + (b_{14} + b_{24})x_4]$$

$$\beta_{\overline{21}}(x) = \frac{1}{x_1 + x_2}[(b_{31} + b_{41})x_1 + (b_{32} + b_{42})x_2]$$

$$\beta_{\overline{22}}(x) = \frac{1}{x_3 + x_4}[(b_{33} + b_{43})x_3 + (b_{34} + b_{44})x_4]$$

To choose $\mathcal{G}$ is a very difficult task and any advice is welcome. In practice external information frequently coming from sources far of mathematics is utilized.

The following three quotations excerpted from [1] showing competence of iterative aggregation/disaggregation methods (shortly IAD) are based on practical experiments and experience of the authors. Our study can be considered as an attempt to support the statements by some theoretical results that typically are not valid for all standard methods frequently applied in practice.

"Results of experiments on a test suite of 13 Markov chains show that the particular two-level iterative solvers BSOR and IAD are in general very competitive with ILU preconditioned Krylov subspace solvers BSGStab, CGS, and GMRES."

"When the Markov chain is extremely ill-conditioned (leaky), incomplete ILU factorization may even fail. For NCD matrices, we recommend IAD and BSOR with *newncd* partitioning and relaxation parameter 1.0."

"However, higher ill-conditioning not always imply poorer performance. It is noticed in some cases that it may even help a solver, especially IAD (compare results of *ncd-alt2* and *ncd*), to converge faster. "

5 IAD Algorithms

5.1 Algorithm SPV($B; T; t, s; x^{(0)}; \varepsilon$)

Let B be an $N \times N$ irreducible stochastic matrix and $\hat{x}$ its unique stationary probability vector. Further, let $I - B = M - W$ be a splitting of $I - B$ such that T is an elementwise nonnegative matrix. Finally, let t, s be positive integers, $x^{(0)} \in \mathcal{R}^N$ an elementwise positive vector and $\varepsilon > 0$ a tolerance.

Step 1. Set $k = 0$.

Step 2. Construct the *aggregated matrix* (in case $s = 1$ irreducibility of B implies that of $\mathcal{B}(x^{(k)})$)

$$B(x^{(k)}) = RB^s S(x^{(k)})$$

Step 3. Find the unique stationary probability vector $z^{(k)}$ from

$$\mathcal{B}(x^{(k)})z^{(k)} = z^{(k)},\ e(p)^T z^{(k)} = 1,\ e(p) = (1, ..., 1)^T \in \mathcal{R}^p.$$

Step 4. Let

$$Mx^{(k+1,m)} = Wx^{(k+1,m-1)},\ x^{(k+1,0)} = x^{(k)},\ m = 1, ..., t,$$

$$x^{(k+1)} = x^{(k+1,t)},\ e(N)^T x^{(k+1)} = 1.$$

Step 5. Test whether
$$\|x^{(k+1)} - x^{(k)}\| < \epsilon.$$
Step 6. If NO in Step 6, then let
$$k + 1 \to k$$
and GO TO Step 2.
Step 7. If YES in Step 6, then set
$$\hat{x} := x^{(k+1)}$$
and STOP.

5.2 Algoritm LM$(C; M, W; t; y^{(0)})$

Let C be an $N \times N$ aggregation convergent matrix with nonnegative real elements, and let $\{M, W\}$ be a splitting of of $A = I - C$ such that the iteration matrix $T = M^{-1}W$ is elementwise nonnegative.
Step 1. Set $0 \to k$.
Step 2. Construct the matrix
$$\mathcal{C}(y^{(k)}) = RCS(y^{(k)}).$$
Step 3. Find a unique the solution $\tilde{z}^{(k)}$ to the problem
$$\tilde{z}^{(k)} - \mathcal{C}(y^{(k)})\tilde{z}^{(k)} = Rb. \tag{5.4}$$
Step 4. Disaggregate by setting
$$v^{(k+1)} = S(y^{(k)})\tilde{z}^{(k)}.$$
Step 5. Let
$$My^{(k+1,m)} = Ny^{(k+1,m-1)} + b,\ y^{(k+1,0)} = v^{(k+1)},\ m = 1, ..., t,$$
$$y^{(k+1)} = y^{(k+1,t)}.$$
Step 6. Test whether
$$\|y^{(k+1)} - y^{(k)}\| < \epsilon.$$
Step 7. If NO in Step 6, then let
$$k + 1 \to k$$
and GO TO Step 2.
Step 8. If YES in Step 6, then set
$$x^* := y^{(k+1)}$$
and STOP.

7 Remark The algorithms of the type introduced in this section are known as Leontev procedures invented by Leontev in the thirties of the twentieth century in his famous sectorial economy theory. Actually, his sectorial variables are just the aggregates of the initial variables and sectorial production matrix is our aggregation matrix etc.

Since both algorithms SPV and LM possess the property that the corresponding error-vector formulas are identical and the corresponding theories are very similar we will investigate the case of SPV algorithms only.

6 Some properties of IAD methods

According to definition of bf SPV algorithm the error-vector formula for the sequence of approximants reads

$$x^{(k+1)} - \hat{x} = J_t(x^{(k)}) \left(x^{(k)} - \hat{x} \right), \tag{6.1}$$

where [2]

$$J_t(x) = J(B; T^t; x) = T^t \left[I - P(x)Z \right]^{-1} \left(I - P(x) \right), \tag{6.2}$$

and where Z comes from the spectral decomposition of $B = Q+Z, Q^2 = Q, QZ = ZQ = 0, 1 \notin \sigma(Z)$. Furthermore, $J_t(x) = T^{t-1} J_1(x),\ t \geq 1$, holds for any x with all components positive.

We want to analyze convergence properties of IAD methods with no explicit requirement that the basic iteration matrix is convergent, i.e.

$$\lim_{k \to \infty} Z^k = 0.$$

8 Remark One of the most delicate questions concerning Theorem 11 reads: How to choose the number of smoothings $\hat{t}$? The answer to this question is not a simple matter as does the following example show. It turns us back to another basic question and namely, how to aggregate. Some results concerned with convergence issues of the SPV algorithm with small number of smoothings t can be found in [3].

6.1 An example

9 Example *Assume $p > 1$ is a positive integer and B is the transition matrix of a Markov chain such that it can be written in a block form as*

$$\begin{pmatrix} B_{11} & 0 & \ldots & 0 & B_{1p} \\ B_{21} & B_{22} & \ldots & 0 & 0 \\ . & . & \ldots & . & . \\ 0 & 0 & \ldots & B_{pp-1} & B_{pp} \end{pmatrix}.$$

The iteration matrix $T = M^{-1}W$ is defined via splitting $I - B = M - W$ with

$$M = \text{diag}\{B_{11}, ..., B_{pp}\},\ W = I - B - M.$$

We see that the iteration matrix T is block p-cyclic.

The aggregation communication operators are chosen such that

$$R = (1, ..., 1)^T$$

is $1 \times N$ *matrix and*

$$S(x)z = \frac{z}{Rx}x, \ x \in \mathrm{Int}\mathcal{R}^N, \ z \in \mathcal{R}^1.$$

This means that the SPV algorithm reduces to the simple power method with the iteration matrix T^t. Assume the off-diagonal blocks are elementwise positive. Obviously, the SPV process possesses the following properties: It does not converge for $t < p$ and does converge for $t = kp, k = 1, 2, ...$ We see that our IAD method does preserve the nonconvergence property of the original power method.

On the other hand, if the aggregation operators are chosen as shown in Section 6 i.e. each single block of matrix B is aggregated to 1×1 *matrix, the situation may change dramatically. As example let us take transition matrix whose off-diagonal row blocks satisfy $B_{jk} = v_j u_{jk}^T, j \neq k$ where v_j and $u_{jk}, j, k = 1, ..., n$, are some vectors. Then taking the same splitting as in the example discussed in this section the exact stationary probability vector is obtained after at most two iteration sweeps* [4].

These examples show that some of the aggregation/disaggregation procedures may be inefficient whilst some other ones can be extremely efficient. The simplicity of these examples should not lead to conclusion that inefficiency is due to our "wish" to demonstrate existence of a poor situation. Divergence may appear whenever one aggregates inappropriately within some blocks of a given transition matrix. Dangerous may be aggregations leading to mixing the cycles. Thus, the situation does not seem to be trivial, but anyhow, inefficiency and even divergence may always be expected. A way out leads to some "order": We propose a suitable concept $-$ *aggregation-convergence*.

6.2 Aggregation-convergence

Let us remind a definition relevant in this context of IAD methods [5].

10 Definition Assume B is $N \times N$ irreducible stochastic matrix with stationary probability vector $\hat{x}$ and R and $S(x)$ IAD communication operators. A splitting of $I - B$, where

$$I - B = M - W = M(I - T), \ T \geq 0,$$

is called *aggregation-convergent* if

$$\lim_{k\to\infty} (I - P(\hat{x}))\, T^k = 0$$

An interesting question is how to recognize that a splitting is aggregation-convergent.

If looking at the error-vector formula valid for any IAD constructed utilizing splitting of

$$A = I - B = M(I - T),\ T \geq 0 \tag{6.3}$$

we can summarize our knowledge concerning the class of IAD algorithms as

11 Theorem [5] *Consider algorithm* $SPV(B; M, W, T; t, s = 1; x^{(0)}; \varepsilon)$ *with an irreducible stochastic matrix* B*, aggregation-convergent splitting (6.3) and initial guess taken such that* $x^{(0)} \in \mathrm{Int}\mathcal{R}^N_+$.

Then there exist generally two positive integers a, b *and two, generally different, neighborhoods* $\Omega_a(\hat{x})$ *and* $\Omega_b(\hat{x})$ *such that Algorithm* $SPV(B; M, W, T; t, s = 1; x^{(0)}; \varepsilon)$ *returns a sequence of iterants* $\{x^{(k)}\}$ *for which*

$$\lim_{k\to\infty} x^{(k)} = \hat{x} = B\hat{x},\ e^T\hat{x} = 1,$$

$$\text{for } t = a \text{ and } x^{(0)} \in \Omega_a(\hat{x}), \tag{6.4}$$

$$\text{for } t \geq b \text{ and } x^{(0)} \in \Omega_b(\hat{x}). \tag{6.5}$$

12 Remark Theorem 11 deserves some comments.

a) First of all, generally, a in (6.4) may be large.
b) There are examples [3] showing that $SPV(B; B; t = 1, s = 1; x^{(0)}; \varepsilon)$ does converge and $SPV(B; B; t = 2, s = 1; x^{(0)}; \varepsilon)$ does not.

13 Example *(I. Pultarová [3]) Let us consider*

$$B = \left(\begin{array}{cc|ccc} 0 & 0 & 0 & 1/2 & 0 \\ 1 & 1/2 & 1/100 & 1/2 & 1/100 \\ \hline 0 & 0 & 0 & 0 & 99/100 \\ 0 & 0 & 99/100 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \end{array}\right).$$

It can be shown that

$$\rho(J(\hat{x})) = 0.9855 < 1 \text{ for } SPV(B; B; t = 1; s = 1; x^{(0)}; \varepsilon = 1.10^{-5})$$

and

$$\rho(J(\hat{x})) = 1.1271 > 1 \text{ for } SPV(B; B; t = 2; s = 1; x^{(0)}; \varepsilon = 1.10^{-5}).$$

The effect just shown is caused by nonnormality of the iteration matrix. In this context let us recall a popular problem of shuffling the cards (see A. Greenbaum [6]).

If looking at the error-vector formula one recognizes immediately that convergence will take place if the spectral radii $\rho(J(b,T,x^{(k)})) < 1, k \geq \hat{k}$ for some $\hat{k}$. On the first look, there seems to be no reason for validity of such properties. The only factor in the product forming matrix $J(B,T^k,x)$ that changes with k is T^k. However, $\{T^k\}$ does not converge if T is cyclic. On the other hand, we did have massive numerical evidence that the IAD processes with iteration matrices $T_m, m = 1, 2, ...,$ where

$$M_m = \left(1 + \frac{1}{m}\right) I$$

implying that

$$T_m = \left(\frac{1+m}{m}\right)^{-1} \left(\frac{1}{m} I + B\right) = \frac{1}{1+m} I + \frac{m}{1+m} B,$$

showed a monotonically increasing rate of convergence for incr变rasing index m. This observation led us to a conclusion that cyclicity of the iteration matrix is harmless. Our theory confirms this claim.

7 Convergence of IAD within the class of irreducible stochastic matrices

Let us consider a subclass of the class of all irreducible Markov chains whose transition matrices are block cyclic. Let B be such a matrix. Then

$$\begin{aligned} B &= \begin{pmatrix} B_{11} & \dots & B_{1p} \\ . & \dots & . \\ B_{p1} & \dots & B_{pp} \end{pmatrix} \\ &= H \begin{pmatrix} 0 & \dots & 0 & \tilde{B}_{1p} \\ \tilde{B}_{21} & \dots & . & 0 \\ . & \dots & . & . \\ 0 & \dots & \tilde{B}_{pp-1} & 0 \end{pmatrix} H^T, \end{aligned} \tag{7.1}$$

where H is some permutation matrix.

14 Agreement *In our analysis we will always assume that the examined stochastic matrix is in a block form obtained by applying an aggregation map $\mathcal{G}$. This concerns in particular the case of cyclic matrices where we assume the block form shown in (7.1).*

Now we consider Algorithm 1 and assume that our transition matrix B has the form

$$B = Q + Z(B),\ \rho(Z(B)) \leq 1,\ 1 \notin \sigma(Z(B)),$$

and

$$Q^2 = Q,\ QZ(B) = Z(B)Q = 0,$$

B as well as T have the blocks of identical sizes and T is block p-cyclic, i.e.

$$T = M^{-1}W = \sum_{j=1}^{p} \lambda^{j-1} Q_j + Z(T),\ \lambda = \exp\left\{\frac{2\pi i}{p}\right\},$$

where

$$Q_j^2 = Q_j, Q_j Q_k = Q_k Q_j = 0,\ j \neq k,$$

$$Q_j Z(T) = Z(T) Q_j = 0$$

$$\rho(Z(T)) < 1.$$

Defining

$$U = \sum_{j=2}^{p} \lambda^{j-1} Q_j + Z(T),$$

we see that 1 is not an eigenvalue of $P(\hat{x})Z(B)$, $I - P(\hat{x})Z(B)$ is invertible and

$$J(x) = T^t[I - P(\hat{x})Z(B)]^{-1}(I - P(\hat{x})).$$

Suppose y is an eigenvector of T corresponding to an eigenvalue λ such that $|\lambda| = 1$ and $\hat{x}$ is the unique stationary probability vector of B. Then, according to [7], the multi-components of vectors $\hat{x}$ and y satisfy

$$y_{(j)} = \alpha_j \hat{x}_{(j)}, \quad y^T = (y_{(1)}^T, ..., y_{(p)}^T), \tag{7.2}$$

with some $\alpha_j \neq 0, j = 1, ..., p$. It follows that

$$\begin{aligned} (P(\hat{x})y)_{(j)} &= \hat{x}_{(j)} \left(\tfrac{1}{(R\hat{x})_j}\right) (Ry)_j \\ &= \alpha_j \hat{x}_{(j)} \tfrac{1}{(R\hat{x})_j} (R\hat{x})_j \\ &= y_{(j)} \end{aligned}$$

and thus,

$$(I - P(\hat{x}))y = 0. \tag{7.3}$$

Let w be an eigenvector of $J(\hat{x})$, i.e.

$$J(\hat{x})w = \lambda w.$$

Since

$$J(\hat{x}) = J(\hat{x})(I - P(\hat{x})),$$

we also have that

$$\lambda(I - P(\hat{x}))w = (I - P(\hat{x}))J(\hat{x})(I - P(\hat{x}))w.$$

Thus, together with w vector $(I - P(\hat{x}))w$ is an eigenvector of $J(\hat{x})$ corresponding to the same λ.

Since, according to (7.3),

$$(I - P(\hat{x}))Q_j = 0,$$

we have

$$(I - P(\hat{x}))U = (I - P(\hat{x}))Z(T)$$

and thus, there is a $\tilde{t} \geq 1$ such that

$$\tau(T^t) = \rho((I - P(\hat{x}))(Z(T))^t) < 1, \text{ for } t \geq \tilde{t}.$$

It follows that there is a $\hat{t} \geq \tilde{t}$ such that

$$\rho\left(J(\hat{x})\right) = \tau\left(T^t[I - P(\hat{x})Z(B)]^{-1}(I - P(\hat{x}))\right) < 1 \text{ for } t \geq \hat{t}.$$

Thus, we have convergence.

Summarizing we can state the following

15 Theorem *Let B be an irreducible stochastic matrix and $I - B = M - W$ its splitting such that the iteration matrix $T = M^{-1}W$ is block p-cyclic.*

Then there exists a positive integer $\hat{t}$ and a neighborhood $\Omega(\hat{x})$ such that the SPV Algorithm returns a sequence of iterants $\{x^{(k)}\}$ such that

$$\lim_{k\to\infty} x^{(k)} = \hat{x} = B\hat{x} = T\hat{x},$$

whenever $x^{(0)} \in \Omega(\hat{x})$.

16 Remark Because of the counterexamples shown one cannot prove more. There are some results on the local convergence property of some special type of the aggreagtion algorithm [3].

8 A class of stochastic matrices for which IAD is fast convergent

In this section we introduce new results concerning convergence of the IAD algorithm introduced in Section 4 and 5. We show under what circumstances the exact solution is obtained within n IAD sweeps, where n is the number of the aggregation groups. The phenomenon when the exact solution is reached within an a priori known finite number of the IAD steps is called *fast convergence* of the IAD method [8,3]. Let us stress that we follow and generalize the basic theorem on this interesting property of the aggregation methods. [8].

17 Definition Let us consider a class of special stochastic matrices of the block form

$$B_{\text{dyad}} = \begin{pmatrix} B_{11} & B_{12} & \ldots & B_{1p} \\ B_{21} & B_{22} & \ldots & B_{2p} \\ . & . & \ldots & . \\ B_{p1} & B_{p2} & \ldots & B_{pp} \end{pmatrix}$$

where

$$B_{jj}, j = 1, ..., p, \ \text{arbitrary substochastic} \tag{8.1}$$

and

$$B_{jk} = v_j \ u_{jk}^T, j \neq k, \ \text{rank one matrices} \tag{8.2}$$

$$v = \begin{pmatrix} v_1 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ v_2 \\ . \\ . \\ . \\ 0 \end{pmatrix} + ... + \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ . \\ v_p \end{pmatrix}, \ \ v_j > 0, \ j = 1, ..., p.$$

Such matrices B_{dyad} will be called *dyadic matrices*.

Let B_{dyad} be irreducible. Then the IAD algorithm $SPV(B; I - B_{\text{diag}}, B_{\text{off}}; s, t; x^{(0)}; \varepsilon)$ returns as $\hat{x}^{(2)}$ the exact stationary probability vector $\hat{x}$ [8] having the form

$$\hat{x} = v = x^{(2)}.$$

Let v be an arbitrary positive vector such that, $e^T v = 1$. Let us denote

$$P = ve^T.$$

Let us consider a stochastic cyclic matrix C,

$$C_{j,k} = \frac{v_{(j)}}{||v_{(j)}||} e^T, \tag{8.3}$$

if $k = j + 1$ or $j = n$ and $k = 1$, and

$$C_{j,k} = 0,$$

otherwise. The example of the block structure of C for $n = 4$ is

$$C = \begin{pmatrix} 0 & \times & 0 & 0 \\ 0 & 0 & \times & 0 \\ 0 & 0 & 0 & \times \\ \times & 0 & 0 & 0 \end{pmatrix},$$

where all of the nonzero blocks are rank-one positive stochastic matrices. The ordering of the blocks influences the convergence, as it will be seen in the next sections.

Let us choose $\frac{1}{2} < \alpha < 1$. According to the previous theory, we will observe the convergence of the IAD algorithm for the matrices

$$B_P = \alpha B + (1 - \alpha) P, \tag{8.4}$$

$$B_C = \alpha B + (1 - \alpha) C \tag{8.5}$$

and

$$B_S = B_{\mathrm{diag}} + \alpha B_{\mathrm{off}} + (1-\alpha) C_w, \tag{8.6}$$

where B_{diag} is a block diagonal part of B, B_{off} is a block off-diagonal part of B and C_w is of the same type as C, but the columns are weighted in order to obtain B_S stochastic.

18 Remark *Let us note that matrix B_P coincides with the* Google matrix *utilized in computations of the PageRank by the Google search engine while matrices B_C and B_S represent its analogs utilizing cyclic perturbations in place of rank-one matrix.*

19 Lemma *Let $x^{(k)}_{(j)} = c_j \hat{x}_{(j)}$ for some positive constants $c_1, \dots c_n$ in step k of the algorithm SPV$(B; T; t, 1; x^{(0)}; \varepsilon)$. Then $x^{(k+1)} = \hat{x}$.*

The proof [8] exploits that relation $(I - P(x^{(k)}))(x^{(k)} - \hat{x}) = 0$ holds for such $x^{(k)}$.

20 Theorem [8] *Let $I - B = M - W$ be a splitting such that the iteration matrix $= M^{-1}W$ is elementwise nonnegative. Let M be a block diagonal part of B or a block triangular part of B. Let the off-diagonal blocks composed from the block-rows of W corresponding to each particular aggregation group be rank-one matrices possessing the same range, i.e. they have the properties described in (8.1) and (8.2). Then the algorithm SPV$(B; T; 1, 1; x^{(0)}; \varepsilon)$ with $T = M^{-1}W$ reaches the exact solution within two iteration sweeps.*

Proof [8] The statement directly follows from Lemma 19. It is sufficient to realize that the assumptions of Lemma 19 are fulfilled after the first sweep, $x^{(2)} = M^{-1}Wy^1$ with appropriate y^1.

In the following we introduce some modifications and generalizations of Theorem 20 for some other special structures of matrix B. In the first of them, we consider the case of finding the stationary probability vector of B_P.

21 Theorem [3] *Let B be a block triangular stochastic matrix and let P be a rank-one stochastic matrix. Let $I - B_P = I - \alpha B - (1-\alpha)P = M - W$ is a splitting of nonnegative type where*

$$M = I - \alpha B_{\mathrm{diag}}, \quad W = \alpha B_{\mathrm{off}} + (1-\alpha)P,$$

where B_{diag} is the block diagonal part of B and B_{off} is the block off-diagonal part of B. Then the IAD method SPV$(B; M^{-1}W; 1, 1; x^{(0)}; \varepsilon)$ yields the exact solution within at most $n+1$ sweeps where n is the number of the aggregation groups and where the blocks correspond to the aggregation groups.

Proof [3] Let us recall the error formula [2,8] for the aggregation algorithm considered in Section 4,

$$x^{(k+1)} - \hat{x} = T(I - P(x^{(k)})Z)^{-1}(I - P(x^{(k)}))(x^{(k)} - \hat{x}), \tag{8.7}$$

where Z comes from the spectral decomposition of $\alpha B + (1 - \alpha P) = Q + Z$, where Q is the Perron projection corresponding to $\alpha B + (1 - \alpha P)$. It can be shown [9] that (8.7) can be substituted by

$$x^{(k+1)} - \hat{x} = T(I - \alpha P(x^(k))B)^{-1}(I - P(x^(k)))(x^(k) - \hat{x}).$$

Since all of the vectors of the last set of rows of B corresponding to the last aggregation group G_n are equal, then $x^{(1)}_{(n)}$ is parallel to $\hat{x}_{(n)}$. Then the last aggregate of

$$(I - P(x^{(1)}))(x^{(1)} - \hat{x})$$

is the zero-vector and due to the upper triangular shape of $(I - \alpha P(x^{(1)})Z)^{-1}$, the last block-component of

$$(I - \alpha P(x^{(1)}Z)^{-1}(I - P(x^{(1)}))(x^{(1)} - \hat{x})$$

is the zero-vector too. Thus $x^{(2)}_{(n-1)}$ is parallel to $\hat{x}_{(n-1)}$. Once again, the last and the last but one aggregates of

$$(I - P(x^{(2)}))(x^{(2)} - \hat{x})$$

are the zero-vectors. Due to the upper triangular shape of $(I - \alpha P(x^2)Z)^{-1}$, the last and the last but one aggregates of

$$(I - \alpha P(x^{(2)})B)^{-1}(I - P(x^{(2)}))(x^{(2)} - \hat{x})$$

are the zero-vectors. If we repeat these considerations, we come to the conclusion that all of the aggregates of $x^{(n)}$ are pairwise parallel to the corresponding aggregates of $\hat{x}$. Then according to Lemma 19, $x^{(k+1)} = \hat{x}$.

Let us come to the next result of this section. We consider the stationary probability vector of matrix B_C.

22 Theorem *Let B be a block upper triangular stochastic matrix and let C be a block-cyclic stochastic matrix constructed according to (8.3). Let $I - B_C = I - \alpha B - (1 - \alpha)C = M - W$ be a splitting of nonnegative type where*

$$M = I - \alpha B_{\text{diag}}, \quad W = \alpha B_{\text{off}} + (1 - \alpha)C,$$

where B_{diag} is the block diagonal part of B and B_{off} is the block off-diagonal part of B. Then the IAD method algorithm $SPV(B; M^{-1}W; 1, 1; x^{(0)}; \varepsilon)$ gives the exact solution within $n + 1$ sweeps.

Proof The structure of B_C is block triangular with a single rank-one nonzero block in the left lower corner in the matrix. The structure of the aggregated matrix $RBS(x^{(k)})$ is the same but considering elements instead of blocks. Then for example in the case $n = 4$,

$$B = \begin{pmatrix} \alpha B_{11} & \alpha B_{12} + (1-\alpha)C_{12} & \alpha B_{13} & \alpha B_{14} \\ 0 & \alpha B_{22} & \alpha B_{23} + (1-\alpha)C_{23} & \alpha B_{24} \\ 0 & 0 & \alpha B_{33} & \alpha B_{34} + (1-\alpha)C_{34} \\ (1-\alpha)C_{41} & 0 & 0 & \alpha B_{44} \end{pmatrix},$$

and the structure of $RBS(x^{(k)})$ is

$$RBS(x^{(k)}) = R \begin{pmatrix} \alpha & \times \times \times \\ 0 & \times \times \times \\ 0 & 0 \times \times \\ (1-\alpha) & 0 \ 0 \times \end{pmatrix} S(x^{(k)}).$$

The block sparsity structure of T is

$$T = \begin{pmatrix} 0 & \times & \times & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & 0 & \times \\ \times & 0 & 0 & 0 \end{pmatrix}.$$

Since $M = I - \alpha B_{\text{diag}}$ and $W_{n,1} =$ is a rank-one matrix then

$$\hat{x}_{(n)} = \beta_n^1 x_{(n)}^{(1)}$$

for some real β_n^1, and the last columns of $S(x^{(1)})$ and that of $S(\hat{x})$ are equal. Then in the next step, the last row and the last column of the aggregated matrix $RBS(x^{(1)})$ are equal to the last row and to the last column of the aggregated matrix $RBS(\hat{x})$, respectively. From the last row (only the first and the last elements are nonzero in it) we see that the ratio of the first and the last elements of the vector z^2 is equal to the ratio of the 1-norms of the parts of the exact solution $\hat{x}_{(1)}$ and $\hat{x}_{(n)}$. Further, the same result is obtained for the parts y_1^2 and $y_{(n)}^2$ of vector $y^2 = S(x^{(1)})z^2$. Let us recall once more,

$$\hat{x}_{(n)} = \beta_n^1 x_{(n)}^{(1)} = \gamma_n^2 y_{(n)}^2$$

for some real scalars β_n^1, γ_n^2. Thus from the last block row and from the last but one block row of the matrix $T = M^{-1}W$ we get

$$\hat{x}_{(n-1)} = \beta_{n-1}^2 x_{(n-1)}^{(2)}, \quad \hat{x}_{(n)} = \beta_{(n)}^{(2)} x_{(n)}^2,$$

for some scalars β_{n-1}^2, β_n^2 and that the ratio of the 1-norms of $x_{(n-1)}^{(2)}$ and $x_{(n)}^{(2)}$ is equal to the ratio of the norms of $\hat{x}_{(n-1)}$ and $\hat{x}_{(n)}$.

In the third step of the IAD algorithm, the block columns $n-1$ and n in the aggregated matrix $RBS(x^{(2)})$ are the same as in the matrix $RBS(\hat{x})$. Thus the ratios among the elements z_1, z_{n-1}, z_n are equal to the ratios among the norms of $\hat{x}_{(1)}, \hat{x}_{(n-1)}, \hat{x}_{(n)}$. The same ratios are valid for the norms of $y_{(1)}, y_{(n-1)}, y_{(n)}$. Thus the multiplication $Ty = M^{-1}Wy$ yields due to the nonzero block structure of T,

$$\hat{x}_{(n-2)} = \beta_{n-2}^3 x_{(n-2)}^{(3)}, \quad \hat{x}_{(n-1)} = \beta_{n-1}^3 x_{(n-1)}^{(3)}, \quad \hat{x}_{(n)} = \beta_n^3 x_{(n)}^{(3)}.$$

Then we get similarly in the fourth step the proper ratios among the norms of the parts $n-3, n-2, n-1, n$ of y from the solution of principal eigenvector of aggregated matrix, and further,

$$\hat{x}_{(n-3)} = \beta_{(n-3)}^4 x_{(n-3)}^{(4)}, \quad \ldots, \quad \hat{x}_{(n)} = \beta_{(n)}^4 x_{(n)}^{(4)}$$

from the correcting step.

Following these considerations, we can see that after k, $k \le n$ sweeps of the IAD algorithm, we have

$$\hat{x}_{(n-(k-1))} = \beta^k_{n-(k-1)} x^{(k_{(n-(k-1))})}, \quad \ldots, \quad \hat{x}_{(n)} = \beta^k_n x^{(k)}_{(n)}.$$

Especially, after n steps, all of the aggregates of $x^{(n)}$ are parallel to the corresponding parts of the exact solution $\hat{x}$. Now using Lemma 19 we get $x^{(n+1)} = \hat{x}$.

23 Remark *The ordering the blocks in the block-cyclic matrix C is the important assumption in Theorem 22. The fast convergence couldn't be obtained, without some additional requirements, for any other ordering of the blocks, for example, for such of the type*

$$C = \begin{pmatrix} 0 & 0 & 0 & \times \\ \times & 0 & 0 & 0 \\ 0 & \times & 0 & 0 \\ 0 & 0 & \times & 0 \end{pmatrix}.$$

24 Remark *If the examined matrix B is block diagonal (instead of block upper triangular), we get the exact solution after at most two sweeps of the IAD method algorithm $SPV(B; M^{-1}W; 1, 1; x^0; \varepsilon)$. It is the consequence of Theorem 20.*

The introduced theorems give rise to guessing that a special reordering may speed up the convergence of the IAD methods. After performing such a procedure it can be seen that a block diagonal or a block triangular reordered matrix B is obtained . Even if an appropriate structure is not obtained exactly, the convergence is faster than in case of the originally ordered matrix in many practical examples. Some numerical experiments shown in the next subsection confirm these claims.

8.1 Numerical experiments

The main goal of this section is to compare the convergence of some of the IAD methods applied for different types of perturbing the initial stochastic matrix B and for different types of reorderings of the matrix B. We have in mind the perturbation a) by a rank-one stochastic matrix , b) by a block cyclic stochastic matrix and c) by a special combination of a block diagonal and block cyclic matrix. These three types of stochastic matrices correspond to the matrices B_P, B_C nad B_S defined by (8.4), (8.5) and (8.6), respectively.

As it turned out in our experiments, a reordering of the matrix may influence the speed of the convergence significantly. That is also a reason why we try to use two special reordering methods, in order to obtain the nonzero structure of B more appropriate for the faster convergence.

One of the ways how to perform reordering is the Tarjan's algorithm [10] which finds all the strongly connected parts of the incidence graph structure of the stochastic matrix. In other words, the algorithm leads to a symmetric reordering of the matrix which yields the block triangular structure with irreducible diagonal blocks. We can also use a threshold adaptation of this algorithm, where only the indices j, k of elements b_{jk} greater than some positive constant θ are treated as the edges of the graph. The complexity of the method corresponds to the number of nonzero elements of the matrix. Let us stress, that in practical examples, the speed of convergence is better for the reordered matrix even if the irreducible diagonal blocks do not correspond to the aggregation blocks.

The second algorithm used in our experiments is a special symmetric reordering which tries to move great elements in each column to the first "subdiagonal". The algorithm finds the greatest value in the column among the elements with row indices which were not previously checked. According to this, a path is found in the graph, and the corresponding reordering is fixed. An interruption of the graph path, when all elements in a column have been tested, can be treated in several different manner. When no further nonzero element can be found in the actuell column, a new starting vertex is chosen. For example, it may correspond to the column with the lowest index still not visited. We can call this method as "following the maximal column elements". Such reordering is not unique and may have no positive influence on the speed of convergence. However, both these types of reordering end up with almost identical conclusions when very sparse matrices are considered.

In the first test, matrix B is a sparse block triangular column stochastic matrix of the size 1000. We compute the stationary probability vector of B_C. Before testing, the columns and the rows are randomly symmetrically permuted. Then several reorderings are considered: Tarjan's threshold reordering, ordering following the maximal columns elements, and the original state without reordering. In the tests we observe the norms of the residual vectors in each step of the IAD methods, see Fig. 1. The line without markers corresponds to the power method, the circled line correspods to the algorithm $SPV(B; M^{-1}W; 1, 1; x^0; \varepsilon)$ used for the matrix without any ordering rows and columns. The lines marked by asterisks and crosses belong to this algorithm used after Tarjan's reordering and "maximal subdiagonal" reordering, respectively.

The second test is performed for the 5000×5000 part of the Stanford matrix [11], see Fig. 2. The residuals are displayed for power method, for the method $SPV(B; M^{-1}W; 1, 1; x^0; \varepsilon)$ and for this method with reordering by Tarjan's algorithm and by "maximal subdiagonal" reordering, respectively. The corresponding lines are the black line without markers, the black line with circles, the blue line with stars and the red line with pluses, respectively.

9 Google-like applications

9.1 A motivating example

We are going to examine the following system of problems parametrized by parameter $\alpha \in (\frac{1}{2}, 1)$:

$$G(\alpha) = \alpha G^{(1)} + (1 - \alpha) G^{(2)},$$

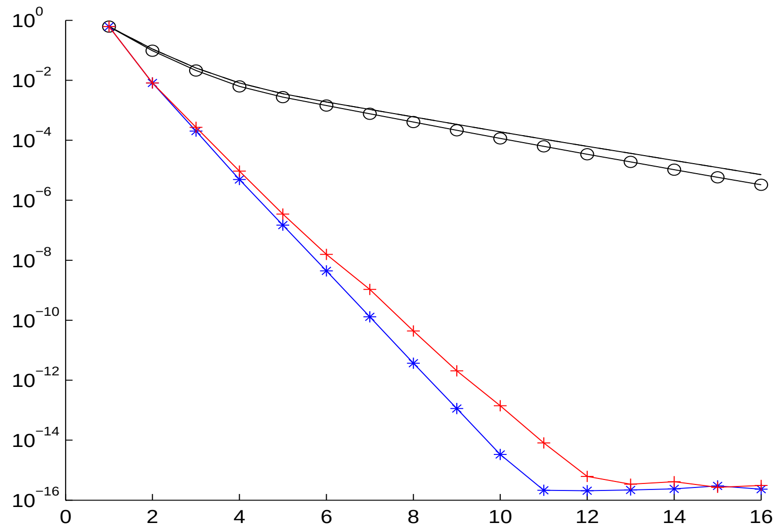

Fig. 1. *The graphical plot of the norms of the residuals of the algorithm* $SPV(B; M^{-1}W; 1, 1; x^0; \varepsilon)$ *for matrix* B_C*, where* $N = 1000$*,* $n = 10$ *and* B *is block upper triangular.*

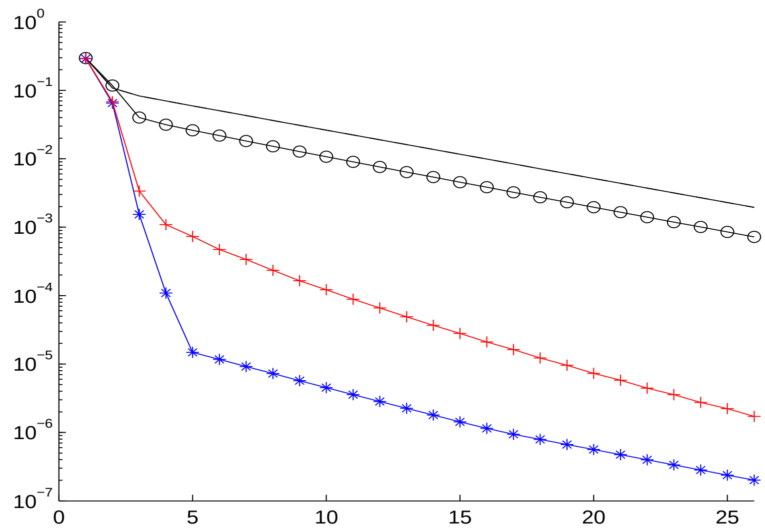

Fig. 2. *The graphical plot of the norms of the residuals of the IAD algorithm* $SPV(B; M^{-1}W; 1, 1; x^0; \varepsilon)$ *for matrix* B_P*, where* $N = 5000$*,* $n = 10$ *and* B *is a part of the Stanford matrix.*

where $G^{(1)}$ is a (column) stochastic matrix and $G^{(2)}$ a suitable (low rank) irreducible stochastic matrix.

A prototype of results we would like to establish is the following one presented as

25 Theorem *Suppose $G^{(2)} = ve^T$, where v is a vector with all its components strictly positive and $e^T = (1, ..., 1)$, $e^T v = 1$, i.e. $G^{(1)}$ represents a rank-one primitive stochastic matrix.*

Then

$$\gamma(G(\alpha)) \leq \alpha.$$

26 Remark To prove the above theorem is very easy and there is a lot of proofs using very different means. Among the existing proofs there are some quite elementary ones. Our proof does not belong to such category. The reason is that our intention is to prove a more general result and we did not have success with elementary methods. A direction of our generalization is led by the fact that the dyadic matrices studied in Section 8 are fast convergent similar just as do rank-one matrices applied in the Google search engine. One of the most compact proofs is presented in [12], see also [13, p.46].

A new approach to computing the PageRank is proposed by applying some variant from a class of aggregation/disaggregation iterative methods. In contrast to frequently used rank-one perturbations in constructing the appropriate Markov chain transition matrix our novelty allows perturbations of arbitrary rank. Similarly, the iteration matrix of the aggregation/disaggregation iterative process is a rank p perturbation of an analog of the original iteration matrix used for power iterations as standard.

Great interest to investigations of modern techniques in communication and information retrieval led to a complex and massive research of appropriate computer systems such as web search engines, their parts and combinations. A nice and nearly complete source of information concerning the mathematics behind the research of the Google type search engine is available in the monograph of A. Langville and C.D. Meyer [13].

The PageRank computations are of particular interest. PageRank became a frequent problem discussed in specialized scientific literature. This problem influenced quite many areas of research mainly within Mathematics and Computer Sciences in general and Numerical Algebra in particular. Our interest in PageRank computations will be reflected in analysis of certain type of iterative processes in the spirit of the title of this contribution.

In connection with computing the PageRank utilizing the power method a problem arises to estimate the spectral radius of a convex combination of two stochastic matrices. Our approach is based on application of some suitable variant from a class of iterative aggregation/diaggregation methods (IAD) [14, pp.307-342], [5]. A similar problem of estimating the rate of convergence arises as well. Our goal will be to show that though the IAD approach allows more general models of computing the PageRank than those exploiting the power method the rate of convergence remains the same. Actually, it is identical with the $\alpha > 1/2$

in the convex combination of the initial Markov chain matrix and the p-rank perturbation.

Proof of Theorem 25 First we determine the Perron eigenvector of $G(\alpha)$.

Let $\hat{x}(\alpha)$ denote the Perron eigenvector. It is easy to see that $\hat{x}(\alpha)$ is strictly positive and it can be normalized as follows

$$e^T \hat{x}(\alpha) = 1.$$

It follows that

$$\hat{x}(\alpha) = G(\alpha)\hat{x}(\alpha) = \alpha G^{(1)}\hat{x}(\alpha) + (1-\alpha)v$$

and

$$\hat{x}(\alpha) = \left[\frac{1}{1-\alpha}(I - \alpha B^{(1)})\right]^{-1} v.$$

Thus, the Perron projection of $G(\alpha)$ reads

$$Q(\alpha) = \hat{x}(\alpha)e^T.$$

We check easily that

$$Q(\alpha)G(\alpha)Q(\alpha) = G(\alpha)Q(\alpha) = Q(\alpha)$$

and

$$(I - Q(\alpha))\, G^{(2)}\, (I - Q(\alpha)) = \left(G^{(2)} - Q(\alpha)\right)(I - Q(\alpha)) = G^{(2)}\,(I - G(\alpha)) = 0. \tag{9.1}$$

The validity of the statement of the Theorem follows from the relation

$$G(\alpha) = Q(\alpha) + (I - Q(\alpha))\, \alpha G^{(1)}\, (I - Q(\alpha))\,.$$

The above proof opens a way to generalizations. A crucial point in the above proof is a special kind of relationship between the original transition matrix $G^{(1)}$ and the perturbation $G^{(2)}$ consisting of equality (9.1).

9.2 Rank-p perturbations

Since aggregation/disaggregation iterative methods behave very friendly with respect to cyclic matrices [1], [5] it is natural to manage computing the PageRank via IAD methods. An appropriate analysis is to be provided in this section.

Similarly as in Subsection 9.1 we are going to consider a convex combination of two stochastic matrices $B(\alpha)$ where $B^{(1)}$ is arbitrary and $B^{(2)}$ possesses a particular form. The only difference can be seen in the fact that the iteration matrix is based on a more general splitting. The iteration matrix coincided with the transition matrix $B(\alpha)$. It should be mentioned that the just mentioned fact has only theoretical value, in practice an implementation of the proposed approach might be very costly and would require computer technique which is still too advanced in standard measures. Some attempts have successfully been made with rather simple splittings, e.g. the IAD iteration matrix coincided with the transition matrix $T(\alpha) = B(\alpha)$.

To construct the iteration matrix we assume the matrices $B^{(1)}$ and $B^{(2)}$ to be written in a block form and then let

$$I - B(\alpha) = I - B^{(1)}_{\text{diag}} - \alpha B^{(1)}_{\text{off}} - (1-\alpha)B^{(2)}_{\text{off}}, \tag{9.2}$$

where

$$B^{(t)} = B^{(1)}_{\text{diag}} + B^{(t)}_{\text{off}},\ t = 1, 2,$$

$$B^{(2)}_{\text{off}} = \begin{pmatrix} 0 & 0 \dots & 0 & B^{(2)}_{1p} \\ B^{(2)}_{1p} & 0 \dots & 0 & 0 \\ . & \dots\dots & . & . \\ 0 & 0 \dots & B^{(2)}_{1p} & 0 \end{pmatrix} \tag{9.3}$$

and

$$B^{(2)}_{1p} = v_1 c(n_p)^T,\ B^{(2)}_{jj-1} = v_j c(n_{j-1}), j = 2, ..., p$$

with

$$v^T = (v_1^T, ..., v_p^T),\ c^T = (c(n_1)^T, ..., c(n_p))^T,\ e^T = (1, ..., 1) = (e(n_1)^T, .., e(n_p)^T)$$

assuming that all the components of v are positive real numbers and vector c is such

$$\left(I - (B^{(1)}_{\text{diag}})^T\right) e = f,\ f^T = \left(f^T_{(1)}, ..., f^T_{(p)}\right)$$

and

$$f^T_{(j)} v^{(j)} = 1,\ j = 1, ..., p.$$

Splitting (9.2) defines iteration matrix

$$T(\alpha) = \left(I - B^{(1)}_{\text{diag}}\right)^{-1} \left[\alpha B^{(1)}_{\text{off}} + (1-\alpha)B^{(2)}_{\text{off}}\right]$$

$$= \alpha T^{(1)} + (1-\alpha)T^{(2)}$$

and one can check easily that $T^{(2)}$ is block p-cyclic, irreducible and

$$(T(\alpha))^T f = f.$$

It follows that $T(\alpha)$ is irreducible and hence, it possesses a unique Perron eigenvector $x(\alpha)$. If we normalize this vector by setting

$$f^T x(\alpha) = 1,$$

we obtain the Perron projection $Q_1(\alpha)$

$$Q_1(\alpha) = x(\alpha) f^T = \left(x(\alpha) f^T\right)^2 = [Q_1(\alpha)]^2.$$

Our goal will be

27 Theorem *Irreducibility of matrix $B^{(2)}$ implies that $T(\alpha)$ can be expressed in the form*

$$T(\alpha) = \sum_{t=1}^{p} \lambda^{t-1} Q_t(\alpha) + \alpha(I - Q_1(\alpha))Z^{(1)}(T)(I - Q_1(\alpha)),\ \lambda = \exp\frac{2\pi i}{p},$$

where

$$\begin{aligned} Q_t(\alpha) &= y^{(t)} \left(f^{(t)}\right)^T \\ &= (I - Q_1(\alpha)) \left[\alpha Q_t^{(1)} + (1-\alpha) Q_t^{(2)}\right] (I - Q_1(\alpha))\ , t > 1, \\ T^{(1)} &= \sum_{t=1}^{p} \lambda^{t-1} Q_t^{(1)} + Z^{(1)}(T),\ T^{(2)} = \sum_{t=1}^{p} \lambda^{t-1} Q_t^{(2)} \end{aligned}$$

and

$$y^{(1)}(\alpha) = x(\alpha),\ \left(y^{(t)}(\alpha)\right)^T = \left(\lambda^t \left(x(\alpha)_{(1)}\right)^T, ..., \lambda^{pt} \left(x(\alpha)_{(p)}\right)^T\right),\ t > 1.$$

Furthermore, the spectrum $\sigma(T) = \sigma(\alpha T^{(1)}) \bigcup_{t=1}^{p} \{\lambda^t\}$ and

$$\tau(T(\alpha)) = \max\{|\,\mu| : \mu \in \sigma(T(\alpha)), \mu \neq \lambda^t, t = 1, ..., p\} \leq \alpha.$$

Let us mention two obvious facts

28 Proposition *The assumption that $B^{(2)}$ as well as $T(\alpha)$ are both p-cyclic implies that $T^{(1)}$ is p-cyclic too.*

29 Proposition *Perron projection of $T(\alpha)$ and matrix $T^{(2)}$ satisfy*

$$Q_1(\alpha) T^{(2)} = Q_1(\alpha) \tag{9.4}$$

and

$$Q_1(\alpha) \left[T^{(2)} - Q_1(\alpha)\right] = 0. \tag{9.5}$$

Proof Both relations (9.4) and (9.5) are immediate consequences of the following relations

$$(T(\alpha))^T f = f = \left(T^{(1)}\right)^T f.$$

Proof of Theorem 27 Relations (9.4) and (9.5) are direct consequences of the hypothesis concerning cyclicity of $T(\alpha)$ [7].

According to Proposition 29 we have

$$\begin{aligned} T(\alpha) &= Q_1(\alpha)Q_1(\alpha) + (I - Q_1(\alpha))T(\alpha)(I - Q_1(\alpha))) \\ &= Q_1(\alpha) + (I - Q_1(\alpha))\alpha T^{(1)}(I - Q_1(\alpha)) \\ &= \textstyle\sum_{t=1}^{p} \lambda^{t-1} Q_t(\alpha)) + (I - Q_1(\alpha))\alpha Z^{(1)}(T)(I - Q_1(\alpha)) \end{aligned}$$

We see that

$$\sigma\left(T(\alpha)\right) = \bigcup_{j=1}^{p}\left\{\lambda^{j}\right\} \cup \sigma\left(\alpha Z^{(1)}(T)\right)$$

and thus,

$$\tau(T(\alpha)) = \alpha\rho\left(Z^{(1)}(T)\right) \leq \alpha.$$

The proof is complete.

30 Remark It is easy to see that all the statements of Theorem 27 hold in a more general situation in which the iteration matrices are not formed on basis of splittings of some Markov chain.

Actually, we have.

31 Theorem *Assume $g \in \mathcal{R}^N$ possesses all coordinates strictly positive and $T^{(j)}, j = 1, 2$ are g-stochastic matrices and $T^{(2)}$ is irreducible and p-cyclic. If $T(\alpha) = \alpha T^{(1)} + (1-\alpha)T^{(2)}$ is p-cyclic then*

$$\tau(T(\alpha)) \leq \alpha\rho(T^{(1)}).$$

10 Some further experiments

Let $X_{jk}, \sum_{j=1}^{p} = N_j$ denote a randomly chosen $n_j \times n_k$ matrix with nonnegative reals. A transition matrix B is obtained by "normalizing" each of the blocks X in order to obtain a column stochastic matrix of the form

$$B(\phi) = B_{\text{diag}} + \phi B_{\text{off}}$$

where

$$B_{\text{diag}} = \begin{pmatrix} \beta_{11}X_{11} & . & \dots & 0 \\ 0 & \beta_{22}X_{22} & . . 0 & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & \beta_{pp}X_{pp} \end{pmatrix},$$

$$B_{\text{off}} = \begin{pmatrix} 0 & 0 . . & 0 & \beta_{1p}X_{1p}0 \\ \beta_{21}X_{21} & \dots & 0 & 0 \\ . & \dots & . & . \\ 0 & 0 . . & \beta_{pp-1}X_{pp-1} & 0 \end{pmatrix}$$

Parameter $\phi > 0$ is utilized in order to test influence of cyclicity. Actually, all the examples considered possess the same kind of cyclic character, i.e. concerning the incidence graf of the corresponding matrices.

A solution vector is defined as $x^{(f)}$, where f is the smallest index for which $\|x^{(f-1)} - f^{(f)}\| < \varepsilon = 10^{-15}$ and the norm used is the l_2 norm.

We let

$$A(\phi) = I - B(\phi) = M - W$$

and compare the following block methods

MM [4] $M = I$

KMS [15] $M = I - B_{\text{diag}} - L(B_{\text{off}})$, where $L(B_{\text{off}})$ denotes the lower block triangle of B_{off}

Vant [16] $M = I - B_{\text{diag}}$

Block Jacobi, Block Gauss-Seidel methods diverge.

$N = 1000, \phi = 1$

index of cyclicity	number of iteration sweeps			
p	power method	MM	KMS	Vant
10	624	12	10	11
20	2526	13	11	12
30	5877	14	12	13
40	9935	15	13	13
50	15570	16	14	14
60	23730	17	14	14
70	31960	17	14	15
80	42060	18	15	15
90	51430	18	15	15
100	60810	18	15	15

$N = 1000, \phi = 1/10$

index of cyclicity	number of iteration sweeps			
p	power	MM	KMS	Vant
10	1874	13	7	7
20	7330	15	8	8
30	16730	16	8	8
40	29160	18	9	9
50	45190	19	9	9
60	67210	20	9	9
70	89210	21	9	9
80	118700	22	9	9
90	144300	22	9	9
100	175900	24	9	9

$N = 1000, \phi = 10$

index of cyclicity	number of iteration sweeps			
p	power	MM	KMS	Vant
10	1877	13	12	13
20	7348	14	14	14
30	18110	16	16	16
40	28720	17	17	17
50	44800	18	17	17
60	71200	18	18	18
70	96900	19	19	19
80	127700	20	19	20
90	155900	20	20	20
100	171200	21	20	21

The results obtained lead to conclusion that the convergence effects show a similar behavior as do the appropriate examples in absence of cyclicity. In our experiments, since the power method with matrix B is convergent the control case MM is acyclic we conclude that cyclic character of the studied examples is preferable in comparison with the standard procedures. Even more, the difficulties concerned with treating problems with cyclic transition matrices reported in some respected sources such as [14] are absent in our computations.

11 Concluding remarks

- The IAD algorithms are shown to be suitable for computing stationary probability vectors of general MC's. Roughly speaking, the more cyclicity the better; cyclicity helps more than primitivity of the iteration matrix.
- Most of the IAD algorithms are very suitable for computations on parallel architectures. This remark concerns in particular the IAD algorithms exploiting inverses of the diagonal blocks as preconditioner, i.e. $M = I - B_{\mathrm{diag}}$ as utilized in our scheme for PageRank computations.
- The above mentioned properties will find applications in modelling real world problems (railway reliability, Google search engine etc.)

Acknowledgements

The work on which this contribution is based was supported by the Program Information Society under Project 1ET400300415, Grant No. 201/05/0453 of the Grant Agency of the Czech Republic and by the Grants No. MSM 210000010 and No. MSM 6840770001 of the Ministry of Education, Youth and Sports of the Czech Republic.

References

1. Dayar, T., Stewart, W.J.: Comparison of partitioning techniques for two-level iterative solvers on large, sparse markov chains. In: SIAM J. Sci. Comput. Number 5 (2000)
2. Marek, I., Mayer, P.: Convergence analysis of an iterartive aggregation/disaggregation method for computing stationary probability vectors of stochastic matrices. In: Numerical Linear Algebra with Applications. Number 5 (1998)
3. Pultarova, I.: Numerical solution of principal eigenvectors of stochastic matrices. In: Submitted for publication. (2007)
4. Marek, I., Mayer, P.: Iterative aggregation/disaggregation methods for computing stationary probability vectors of stochastic matrices can be finitely terminating. In: International Journal of Differential Equations. (2001)
5. Marek, I., Mayer, P.: Convergence theory of a class of aggregation/ disaggregation iterative methods for computing stationary probability vectors of stochastic matrices. In: Linear Al;gebra Appl. (2002)
6. Greenbaum, A.: Card shuffling and the polynomial numerical hull of degree k. In: Manuscript, University of Washington, Seattle. (2002)
7. Courtois, P.J., Semal, P.: Block iterative algorithms for stochastic matrices. In: Linear Algebra Appl. (1986)
8. Marek, I., Mayer, P.: Convergence theory of some classes of iterative aggregation-disaggregation methods for computing stationary probability vectors of stochastic matrices. In: Linear Algebra and its Applications. Number 363 (1998)
9. Marek, I., Pultarova, I.: A note on local and global convergence analysis of iterative aggregation-disaggregation methods. In: Linear Algebra and Its Applications. Number 413 (2006)
10. Duff, I., Reid, J.K.: An implementation of tarjan's algorithm for the block triangularization of a matrix. In: ACM Transactions on Mathematical Software. Number 4 (1978)
11. Kamvar, S.: (Stanford web matrix and stanford-berkeley web matrix, data sets)
12. Langville, A.N., Meyer, C.D.: Deeper inside pagerank. In: Internet Journal. Number 3 (2005)
13. Langville, A.N., Meyer, C.D.: Google's pagerank and beyond. the science of search engine rankings, Princeton University Press (2006)
14. Stewart, W.J. In: Introduction to the Numerical Solution of Markov Chains, Princeton University Press, Princeton, NJ (1994)
15. Koury, R., McAllister, D.F., Stewart, W.J.: Methods for computation stationary distributions of nearly completely decomposable markov chains. In: SIAM J. Alg. Disc. Methods. Number 2 (1984)
16. Vantilborg, H.: The error aggregation. a contribution to the theory of decomposable system and applications, PhD Thesis, Faculty of Applied Sciences, Louvain Catholic University, Louvain-la Neuve, Belgium (1981)