CONVERGENCE ISSUES IN THE THEORY AND PRACTICE OF ITERATIVE AGGREGATION/DISAGRREGATION METHODS

IVO MAREK *, PETR MAYER †, AND IVANA PULTAROVÁ ‡

Abstract. Iterative aggregation/disaggregation methods (IAD) for computation stationary probability vectors of large scale Markov chains is an efficient tool in practice. However, its convergence theory is still not developed appropriately. Similarly as in other multi-level methods such as multigrid methods the number of relaxations on the fine level of the IAD algorithms that is to be executed plays a very important role Some new concepts to better understanding the methods as well as behavior and dependence of various parameterrs involved in the aggregation algorithms are studied and some necessary and/or sufficient conditions for convergence are established. The theory developed offers as one of the main results a proof of convergence of IAD algorithms independently of the fact whether the governing iteration matrix is primitive or cyclic. Another important result concerns comparison of the rates of convergence two IAD processes. Some examples document the diversity of behavior of IAD methods.

Key words. stationary probability vector of Markov chain, iterative aggregation/disaggregation

AMS subject classifications. 15A15, 15A09, 15A23, 65F05

1. Introduction. As documented in the literature, e.g. [2] and [13, Section 6.3], aggregation/disaggregation iterative methods (IAD) belong to competitive classes of methods for computation stationary probability vectors of Markov chains and Leontev systems. The appropriate convergence theory still is far of being complete in the sense of understanding the dependence of these methods upon all parameters influencing convergence and its rate. The aim of this contribution is to decode some of the subtle parts of the theory. In particular, some new concepts such as convergence indices are introduced and their properties studied. These convergence indices allow to establish some relations between the number of relaxations on the fine level in each iteration sweep and the convergence/divergence behavior of the IAD. Some results in this direction appear as decisive in order to understand the interplay between the basic iteration matrix and the IAD algorithms. Consequently, a new important result is established: The IAD processes do return convergent sequences of iterants without requiring the splittings which the IAD algorithms are based on to be convergent; e.g. splittings leading to cyclic iteration matrices are not only allowed but they may appear as preferable in comparison with primitive iteration matrices obtained using shifts of the originally cyclic iteration matrices.

2. Definitions and notation.

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

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

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

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

*Department of Mathematics, Czech University of Technology, Thákurova 7, 166 29 Praha 6, Czech Republic (marek@ms.mff.cuni.cz).

†Department of Mathematics, Czech University of Technology, Thákurova 7, 166 29 Praha 6, Czech Republic , (pmaye@ms.mff.cuni.cz).

‡Department of Mathematics, Czech University of Technology, Thákurova 7, 166 29 Praha 6, Czech Republic , (ivana@mat.fsv.cvut.cz).

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

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

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

This quantity is called *spectral subradius* of C.

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.

Let p be a positive integer B is an irreducible column stochastic matrix with spectral decomposition

$$B = Q + Z,\ Q^2 = Q, QZ = ZQ = 0,\ \rho(Z) < 1,$$

where

$$Q = \sum_{j=1}^{p} \lambda^{j-1} Q_j,\ Q_j Q_k = Q_k Q_j = \delta_{jk} Q_j, j, k = 1, ..., p.$$

Let us note that the above formulas describe two essentially different situations: The *primitive case* appears if $p = 1$ and the *cyclic case* if $p > 1$ respectively.

2.2. Aggregation communication. 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:

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

With this notation we can write $e^T = (e(r_1)^T, ..., e(r_n)^T)$, where

$$r_j = \text{card}\left(\left\{\overline{j} \in \{1, ..., N\} : \mathcal{G}(\overline{j}) = j\right\}\right).$$

Iterative Aggregation/disaggregation 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

$$P(x)x = x \quad \forall x \in \mathcal{R}^N, x_j > 0, j = 1, ..., N. \tag{2.1}$$

Define the *aggregated matrix* as

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

3. IAD Algorithms.

3.1. Algorithm SPV$(B;T;t,s;x^{(0)};\mathcal{G};\varepsilon)$ **(**Stationary Probability Vector Algorithm**).** 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 $T=M^{-1}W$ 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 $\mathcal{B}(x^{(k)})$)

$$\mathcal{B}(x^{(k)})=RB^sS(x^{(k)})$$

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

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

Step 4. Let

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

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

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.

The error matrix is defined as

$$J(B;T;t,s;\mathcal{G};x^{(0)};\varepsilon)=T^t\left[I-P(x)\left(B-Q_1\right)\right]^{-1}\left(I-P(x)\right),$$

where we set $Q_1=Q$ if $p=1$ and

$$x^T=(x_1,...,x_N),\ x_j>0,j=1,...,N.$$

If the dependence of the iteration sequence on the iteration matrix should be made explicit we write $x^{(k)}(T)$ in place of $x^{(k)}$, where $T=M^{-1}W$.

3.2. Algoritm LM$(C; M, W; t; \mathcal{G}; y^{(0)})$ **(**Leontev Model Algorithm**).** 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

$$\tilde{z}^{(k)} - \mathcal{C}(y^{(k)})\tilde{z}^{(k)} = Rb. \tag{3.1}$$

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.

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.

4. Some properties of IAD methods. According to definition of **SPV** algorithm the error-vector formula for the sequence of approximants reads

$$x^{(k+1)} - \hat{x} = J_t(x^{(k)}) \left(x^{(k)} - \hat{x} \right), \tag{4.1}$$

where [5]

$$J_t(x) = J(B; T^t; x) = T^t \left[I - P(x)Z\right]^{-1} \left(I - P(x)\right), \tag{4.2}$$

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. the following limit

$$\lim_{k \to \infty} T^k$$

exists.

REMARK *One of the most delicate questions concerning Theorem 4.2 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 [11].*

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

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

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

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

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 4 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 *[7]*.

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*.

4.1. Aggregation-convergence. Let us remind a definition relevant in the context of IAD methods [6].

DEFINITION 4.1. 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

$$A = I - B = M(I - T),\ T \geq 0 \tag{4.3}$$

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

THEOREM 4.2. [6] *Consider algorithm SPV*$(B; M, W, T; t, s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$ *with an irreducible stochastic matrix* B*, aggregation-convergent splitting (4.3) and initial guess taken such that* $x^{(0)} \in \mathrm{Int}\mathcal{R}_+^N$.

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

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

for $t = \tilde{t}$ and $x^{(0)} \in \Omega_{\tilde{t}}(\hat{x})$,
(4.4)

for $t \geq \hat{t}$ and $x^{(0)} \in \Omega_{\hat{t}}(\hat{x})$.
(4.5)

REMARK *Theorem 4.2 deserves some comments.*

a) First of all, generally, $\tilde{t}$ *in (4.4) may be large.*

b) There are examples [12] 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.*

EXAMPLE [12] 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 [3]).

5. Necessary and/or sufficient conditions for local and global convergence. Let $\mathcal{G}$ be a mapping of the index sets $\{1, ..., N\}$ onto $\{1, ..., n\}$ and R and $S(x)$ the corresponding communication maps determining the aggregation projection $P(x) = S(x)R$ [6]. Let B denote a fixed irreducible column stochastic matrix and $I - B = M - W$ its splitting such that the iteration matrix $T = M^{-1}W$ is elementwise nonnegative.

DEFINITION 5.1. *A nonnegative integer r is called a-*index *of algorithm $SPV(B; T = M^{-1}W; t, s; \mathcal{G}; x^{(0)}; \varepsilon)$ if this algorithm returns convergent sequences of iterants for $t = r$ and divergent ones for $r = r + 1$.*

DEFINITION 5.2. *A positive integer t_b is called (convergence) b-*index *of algorithm $SPV(B; T = M^{-1}W; t, s; \mathcal{G}; x^{(0)}; \varepsilon)$ if t_b is the smallest positive integer such that SPV algorithm returns convergent sequences of iterants for all $t \geq t_b$. If a considered SPV process is not convergent we also say that its b-index is zero.*

In this section we are going to examine convergence issues concerned with Algorithm SPV$(B; T = M^{-1}W; t, s; \mathcal{G}; x^{(0)}; \varepsilon)$ formulated in terms of the a and b indices of this algorithm .

Prior we formulate the appropriate statements we comment on Theorem 4.2 utilizing the above introduced convergence indices.

REMARK Parts of Theorem 4.2 concerned with relations (4.4) and (4.5) can be reformulated as follows: The a-index of SPV$(B; T; t, s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$, let us denote it by t_a, satisfies $t_a \geq \tilde{t}$ and for the b-index, denoted by t_b, we have relation $t_b \leq \hat{t}$.

The convergence indices just introduced allow us to formulate natural conditions necessary and sufficient for convergence and/or divergence of SPV algorithms. They are presented in the form of Propositions.

PROPOSITION 5.3. *(Sufficient conditions for global convergence) Assume $B = C + \gamma I$, where $\gamma > 0$ is an irreducible stochastic matrix.*

Then, there exist positive integers $\tilde{t}$ and $\tilde{s}$ such that for any $x^{(0)} \in \mathcal{R}^N$ with strictly positive components algorithm $SPV(B; B; t, s; \mathcal{G}; x^{(0)}; \varepsilon)$ for $t \geq \tilde{t}$ and $s \geq \tilde{s}$ returns convergent sequence of iterants $\{x^{(k)}\}$, i.e.

$$\lim_{k \to \infty} x^{(k)} = \hat{x} = B\hat{x}. \tag{5.1}$$

In other words, the b-index of algorithm $SPV(B; B; t \geq \tilde{t}, s \geq \tilde{s}; \mathcal{G}; x^{(0)}; \varepsilon)$, say t_b, is bounded above by $\tilde{t}$: $t_b \leq \tilde{t}$.

Proof. Let $\|.\|_{MS}$ denote the norm defined in (10.1) (see also [4]). It follows that

$$\|P(\hat{x})\|_{MS} = 1$$

implying that

$$\|I - P(\hat{x})\|_{MS} \leq 2,$$

and

$$\left\| [I - P(\hat{x})C^s]^{-1} \right\|_{MS} \leq \frac{1}{1 - \|C\|_{MS}}.$$

In a manner similar to that in [4] we derive that

$$\begin{aligned} \|x^{k+1} - \hat{x}\|_{MS} &= \|C^t[I - P(\hat{x})C^s]^{-1}(I - P(\hat{x}))\|_{MS}\|x^{k+1} - \hat{x}\|_{MS} \\ &\leq \|C^t\|_{MS}\|[I - P(\hat{x})C^s]^{-1}\|_{MS}\|I - P(\hat{x})\|_{MS}\|x^{k+1} - \hat{x}\|_{MS} \\ &\leq \tfrac{2\alpha}{1-\beta}\|x^{k+1} - \hat{x}\|_{MS}. \end{aligned}$$

To complete the proof it is enough to take $\tilde{t}$ and $\tilde{s}$ such that $\|C^{\tilde{t}}\|_{MS} = \alpha$ and $C^{\tilde{s}}\|_{MS} = \beta$ satisfy

$$2\alpha + \beta < 1.$$

□

REMARK It may seem strange that such a simple statement with a quite trivial proof may appear as new after a rather long period of investigating IAD methods. The reason might be the absence of well understanding the role of the indices t and s in the SPV Algorithm. The examples of Section 8 showing e.g. that SPV$(B; B; t = 1, s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$ may not converge even locally while SPV$(B; B; t = 1, s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$ does led us to defining the *a-* and *b-* convergence indices and finding a way to local and global convergence of IAD algorithms.

PROPOSITION 5.4. *(Necessary condition for local convergence) Suppose, all free variables of Algorithm SPV$(B; T = M^{-1}W; t = s = 1; x^{(0)}; \varepsilon)$ are fixed and*

$$\rho\left(J(B; B; t, s = 1; x^{(0)}; \varepsilon)\right) < 1 \quad \forall t \geq \hat{t}$$

.

Then

$$\rho\left((I - P(\hat{x}))T^t\right) < 1 \; \forall t \geq \hat{t}. \tag{5.2}$$

Proof. Let us assume that the conclusion of Proposition 5.4 is false. Then the spectral resolution of the of matrix $(I - P(\hat{x}))\,T^t$ reads

$$(I - P(\hat{x}))\,T^t = \sum_{j=1}^{p} \lambda_j P_j + F,$$

where

$$\rho(F) < \rho\left((I - P(\hat{x}))T^t\right) = |\lambda_j| \geq 1.$$

Let $y \in \mathcal{R}^N$ be such that

$$P_{j_0} y \neq 0,\ j_0 \in \{1, ..., p\}.$$

It follows that

$$\liminf_{k\to\infty} \left\| \left[(I - P(\hat{x}))\, T^t \right]^k y \right\| > 0.$$

Thus, the implication

$$\left\| J(B; B; t, s = 1; x^{(0)}; \varepsilon) \right\| \to 0 \text{ as } t \to \infty$$

is contradictory. □

PROPOSITION 5.5. *(Necessary and sufficient conditions for local convergence) Let $t_b \geq 0$ be the b-index of algorithm SPV$(B; T; t \geq t_b, s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$. Then the following conditions (i) and (ii) are equivalent.*

(i) The SPV iterative process returns convergent sequences of iterants.

(ii) Relation

$$\rho\left((I - P(\hat{x}))T^t\right) < 1$$

holds for all $t \geq t_b$.

As a consequence of Propositions 5.5 and 5.3 we deduce

THEOREM 5.6. *To every SPV algorithm SPV$(B; T; t, s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$ there belongs a finite b-index of convergence.*

REMARK Logically, the negation of the condition necessary for convergence described in the above proposition is sufficient for divergence of the SPV iterative process. However, according to the next proposition, possible divergence may appear only if the extreme case of the sufficient conditions takes place, i.e. if $\rho((I - P(\hat{x}))T) = 1$.

THEOREM 5.7. *Let B be a column stochastic matrix and $I - B = M - W = M(I - T)$ its splitting of nonnegative type with iteration matrix $T \geq 0$ and $e^T M = \tilde{e}^T = (\eta_1 e(r_1)^T, ..., \eta_n e(r_n)^T)$.*

Then

$$\rho((I - P(\hat{x}))T) \leq 1.$$

Proof. Instead of $(I - P(\hat{x}))T$ we will explore the spectral radius of

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

where

$$D = \operatorname{diag}\left\{ \sqrt{\frac{\hat{x}_1}{\eta_1}}, ..., \sqrt{\frac{\hat{x}_{r_1}}{\eta_1}}, ..., \sqrt{\frac{\hat{x}_{N-r_{n-1}+1}}{\eta_n}}, ..., \sqrt{\frac{\hat{x}_N}{\eta_n}} \right\}.$$

Denoting $T_s = D^{-1}TD$ we will utilize norm (10.1) with $z = DM^T e = D\tilde{e}$. An estimate of $||T_s||_{(z)}$ is obtained using relations

$$z^T T_s = e^T M D T_s = e^T M D D^{-1} M^{-1} W D = e^T W D = e^T M D = z^T.$$

Thus $||T_s||_{(z)} \leq 1$. An estimate of $||T_s^T||_{(z)}$ can be obtained from

$$T_s z = D^{-1}M^{-1}WDDM^T e.$$

Further, since

$$D^2 M^T e = \hat{x},$$

we have

$$T_s z = D^{-1}M^{-1}W\hat{x} = D^{-1}\hat{x} = DM^T e = z.$$

Then $||T_s||_{(z)} \leq 1$ and $||T_s^T||_{(z)} \leq 1$. For 2-norm of $D^{-1}(I - P(\hat{x}))TD$ we have

$$||D^{-1}(I-P(\hat{x}))TD||_2 \leq ||D^{-1}(I-P(\hat{x}))D||_2\,||T_s||_2 = ||D^{-1}(I-P(\hat{x}))D||_2\,\rho(T_s^T T_s) \leq$$

$$\leq ||D^{-1}(I-P(\hat{x}))D||_2\,||T_s^T T_s||_{(z)} \leq ||D^{-1}(I-P(\hat{x}))D||_2\,||T_s||_{(z)}\,||T_s^T||_{(z)} \leq$$

$$\leq ||D^{-1}(I-P(\hat{x}))D||_2.$$

Since $D^{-1}(I - P(\hat{x}))D$ is a symmetric projection, we get

$$||D^{-1}(I-P(\hat{x}))TD||_2 \leq 1$$

then $\rho((I - P(\hat{x}))T) \leq 1$. □

REMARK

The assumptions of Theorem 5.7 were restricted to the case where M is chosen in such a way that

$$e^T M = \tilde{e}^T = (\eta_1 e(r_1)^T, ..., \eta_n e(r_n)^T) \tag{5.3}$$

The reason was that in the proof we use a matrix D which has to fulfill the following three properties

a) D is symmetric,
b) $D^2 M^T e = \hat{x}$,
c) $D^{-1}P(\hat{x})D$ is symmetric.

In case of the special choice of M given by the relation (5.3), matrix D can be diagonal with

$$D_{ii} = \sqrt{\frac{\hat{x}_i}{c_k}}$$

for i with $\mathcal{G}(i) = k$ as introduced in the proof of Theorem 5.7. But when we do not assume (5.3), matrix D in a diagonal form does not fulfill these three conditions any more. When the size of B is $N \times N$, the number of equations corresponding to these three conditions a), b) and c) are

$$\frac{N^2 - N}{2}, \quad N, \quad \frac{N^2 - N}{2},$$

respectively, and the sum of them is N^2. Thus it seems that such a matrix D can be found and that the statement of Theorem 5.7 is valid without the assumption (5.3). Still we do not provide the exact construction of D yet.

6. Convergence of IAD within the class of irreducible stochastic matrices. 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 guaranteeing convergence . 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 increrasing index m. This observation led us to a conclusion that cyclicity of the iteration matrix is harmless. Our theory confirms this claim.

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

$$\begin{aligned} B &= \begin{pmatrix} B_{11} & . & . & . & B_{1p} \\ . & . & . & . & . \\ B_{p1} & . & . & . & B_{pp} \end{pmatrix} \\ &= H \begin{pmatrix} 0 & . & . & . & 0 & \tilde{B}_{1p} \\ \tilde{B}_{21} & . & . & . & . & 0 \\ . & . & . & . & . & . \\ 0 & . & . & . & \tilde{B}_{pp-1} & 0 \end{pmatrix} H^T, \end{aligned} \tag{6.1}$$

where H is some permutation matrix.

6.1. **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 (6.1).*

Now we consider Algorithm 3.1 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 [1], the multi-components of vectors $\hat{x}$ and y satisfy

$$y_{(j)} = \alpha_j \hat{x}_{(j)}, \quad y^T = (y_{(1)}^T, ..., y_{(p)}^T), \tag{6.2}$$

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,

$$(I - P(\hat{x}))y = 0. \tag{6.3}$$

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 (6.3),

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

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

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

6.3. **Remark** Because of the counterexamples shown generally one cannot prove more. There are some results on the local convergence properties of some special types of the aggregation algorithm [11].

7. A comparison result. Our numerous experiments with applying Algorithm **SPV**$(B; I, B; t = 1, s = 1; x^{(0)}; \varepsilon)$ never failed to converge if applied to practical problems. A possible explanation might consist of the fact that in any neighborhood of an irreducible stochastic for which a given SPV algorithm returns divergent sequences of iterants there is another stochastic matrix for which the same algorithm does return convergent sequences of iterants.

The next result enlightens a bit the role of b-index of the basic algorithm SPV. In fact, a very natural fact is shown and namely that the smaller is the spectral radius of the variable part of the error matrix the faster convergence of the corresponding SPV algorithm is guaranteed.

THEOREM 7.1. *Let $\|.\|$ denote any norm on $\mathcal{R}^N$ and also the corresponding operator norm. Further, let B is an irreducible stochastic matrix and $I - B = M_j^{-1}W_j, j = 1, 2$ are two splittings such that T_j is elementwise nonnegative.*

Assume that relations

$$\left\|P(T_2^t x) - P(\hat{x})\right\| \leq \left\|P(T_1^t x) - P(\hat{x})\right\|, \; x \in \Omega_1(\hat{x}) \tag{7.1}$$

hold for $t \geq t_1$ where t_1 is the b-convergence index of SPV$(B; T_1; t, s = 1; \mathcal{G}, x^{(0)}; \varepsilon)$ and where $\Omega_1(\hat{x})$ is a corresponding neigborhood of local covergence.

Then algorithm SPV$(B; T_2; t, s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$ is locally convergent too and its B-convergence index $t_2 \leq t_1$.

Proof. Let $x^{(k)}(T_j^t)$ denote the iterant returned by Algorithm SPV$(B; T_2; t, s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$. Our goal is to show convergence of the sequence $\{x^{(k)}(T_2^t)\}$ with $t \geq t_1$. By construction of the sequence we know that it is componentwise uniformly bounded and hence it is pre compact as a bounded set. Let $\tilde{y}$ be any of its points of condensation and let us assume, without loss of generality, that the corresponding convergent subsequence coincides with that of $\{x^{(k)}(T_2^t)\}$. Relations (7.1) imply that

$$\|P(x^{(k)})(T_2) - P(\hat{x})\| \leq \|P(x^{(k)})(T_1) - P(\hat{x})\|$$

hold for $k = 1, 2, ...$ and consequently,

$$\lim_{k\to\infty} P(x^{(k)}(T_2)) = P(\hat{x}).$$

Local convergence of Algorithm SPV$(B; T_2; t, s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$ then follows because of validity of the next proposition proven in [6], [7].

PROPOSITION 7.2. *Let* $x \in \mathcal{R}^N \oplus i\mathcal{R}^N$, $i^2 = 1$, *satisfy*

$$P(x) = P(\hat{x}).$$

Then $\hat{x} = x^{(1)}$, *where* $x^{(1)}$ *is the vector returned after one iteration sweep of Algorithm SPV*$(B; T_2; t, s = 1; \mathcal{G}; x^{(0)} = x; \varepsilon)$.

The proof of Theorem 7.1 is complete. □

8. Examples. Example 1. We compute spectral radii of error matrices

$$J_t = B^t(I - P(\hat{x})Z)^{-1}(I - P(\hat{x})),$$

for a trivial example, namely for a primitive 3×3 matrix

$$B = \begin{pmatrix} a & 0 & b \\ 1-a & 0 & 1-b \\ 0 & 1 & 0 \end{pmatrix}$$

for $t = 1, \ldots, 15$ and for various values of a and b, respectively. The values of a and b represent situations of nearly cyclic or nearly reducible matrix B,

i) $a = b = 0.9$, it means nearly reducible matrix B,
ii) $a = b = 0.1$, nearly reducible matrix B,
iii) $a = 0.9$ and $b = 0.1$, nearly cyclic matrix B,
iv) $a = 0.1$ and $b = 0.9$, nearly cyclic matrix B.

All of these four cases lead to local divergenece for $t = 1$. The nearly reducible cases (i) and (ii) differ significantly for increasing t. While in (i) the spectral radii decrease rapidly, the spectral radii in (ii) decrease very slowly. The effect is more remarkable for $a = b \to 0$ in (ii). The behavior of nearly cyclic cases (iii) and (iv) also differ for changing t. Thus one can see that even in such a trivial example, the choice of the aggregation groups is crucial.

Example 2. While in Example 1 all choices of a and b lead to local convergent IAD process for any $t = 1, 2, \ldots, 15$, the situation is different for matrix

$$B = \begin{pmatrix} 0 & 0 & 0 & 0.1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \end{pmatrix}. \tag{8.1}$$

In Figure 8.2, we show the spectral radii of error matrices $J_t = B^t(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))$ (solid line) and for $(I - P(\hat{x}))Z^t$ (dashed line) for $t = 1, 2, \ldots, 12$. We can observe that smoothings with some of the powers of B lead to processes divergent even locally.

9. Concluding remarks. When summarizing our observations we can say that the theory, computer experiments and practical computations confirm a view accepted in the literature and namely that iterative aggregation/disaggregation methods are competitive means for computations the characteristics of Markov chains, in particular stationary probability vectors. Let us recall that any IAD method possesses a finite b-index of convergence independently of whether the governing iteration matrix is

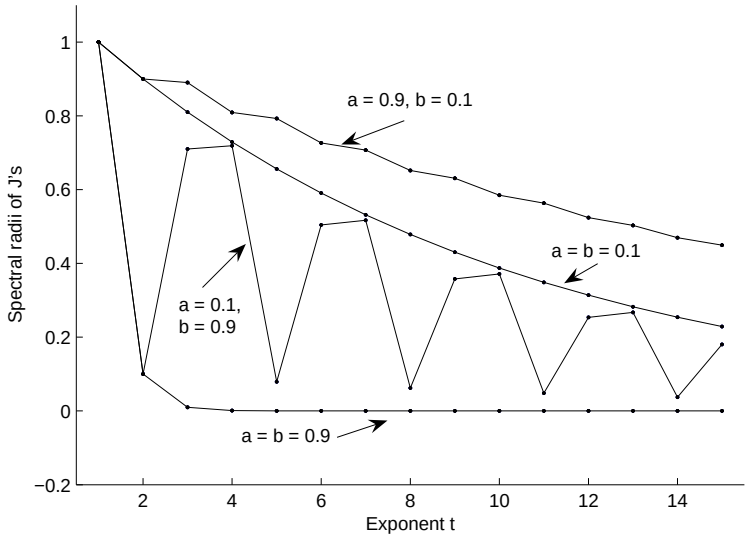

FIG. 8.1. *Spectral radii of matrices J_t for four different choices of a and b in Example 1.*

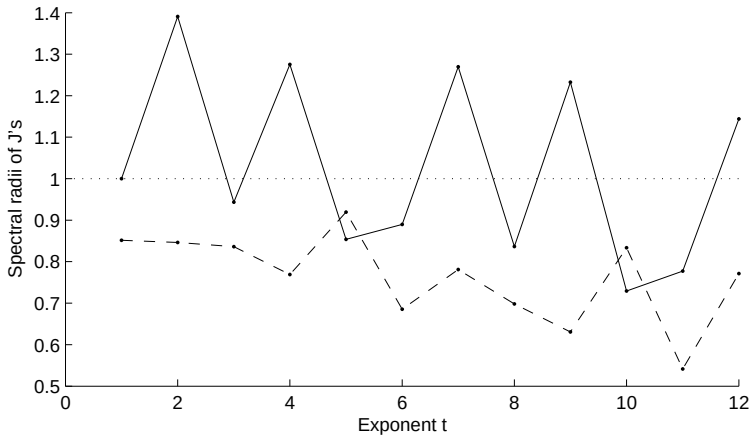

FIG. 8.2. *Spectral radii of matrices J_t (solid line) and $(I - P(\hat{x}))Z^t$ (dashed line) in Example 2.*

primitive or cyclic. This property significantly distinguishes the IAD methods of other methods. Another of our results we want to mention explicitly is concerned with the theory of convergence indices including a new type of comparison of rates of convergence. We observed examples showing divergence of some SPV algorithms too. We show however that the divergence can take place only for indices $t < t_b$, where t_b is the convergence b-index of the appropriate SPV algorithm.

10. Appendix: A special norm. Let C be an $N \times N$ matrix whose elements are reals and let $z^T = (\zeta_1, ..., \zeta_N), \zeta_j > 0, j = 1, ..., N$. Define

$$\|C\|_{(z)} = \nu_z(C) = \min \left\{ \alpha \in \mathcal{R}_+ : |C^T| z \leq \alpha z \right\} \tag{10.1}$$

and $|C|$ denotes the matrix of absolute values of elements of matrix C.

PROPOSITION 10.1. [4] *The expression (10.1) is a norm on the space of $N \times N$ matrices over R.*

Acknowledgments. 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] Courtois P.J., Semal P. *Block iterative algorithms for stochastic matrices*, Linear Algebra Appl. **76**, 59-80 (1986).

[2] Dayar T., Stewart W.J. *Comparison of partitioning techniques for two-level iterative solvwers on large, sparse Markov chains*. SIAM J. Sci. Comput.Vol **21**, No. 5, 1691-1705 (20

[3] Greenbaum A. *Card shuffling and the polynomial numerical hull of degree k*. Manuscript, Uniersity of Washington, Seattle (2002.

[4] J. Mandel, B. Sekerka, A local convergence proof for the iterative aggregation method. *Linear Algebra and Its Applications* 51, 163 - 172, 1983.

[5] Marek I., Mayer P. *Convergence analysis of an aggregation/disaggregation iterative method for computation stationary probability vectors of stochastic matrices*. Numerical Linear Algebra With Applications, **5** 253-274 (1998).

[6] Marek I., Mayer P. *Convergence theory of a class of aggregation/ disaggregation iterative methods for computing stationary probability vectors of stochastic matrices*. Linear Algebra Appl. **363** 177- 200, 2002).

[7] Marek I., Mayer P. *Iterative aggregation/disaggregation methods for computing stationary probability vectors of stochastic matrices can be finitely terminating*, International Journal of Differential Equations Vol. **3**, 301-313 (2001).

[8] Marek I., Mayer P., Pultarová I. *IAD methods based on splittings with cyclic iteration matrices*. Dagstuhl Seminar Proceedings 07071, 2007, 27 pp. 2007.

[9] Marek I., Pultarová I. *A note on local and global convergence analysis of iterative aggregation/disaggregation methods*. Linear Algebra Appl. **413** 327-341 (2006).

[10] Marek I., Žitný K. *Analytic Theory of Matrices for Applied Sciences Vol 1*. Teubner Texte zur Mathematik Band 60, Leipzig, 1983.

[11] Pultarová I. *Necessary and sufficient local convergence condition of one class of iterative aggregation - disaggregation methods*. To appear in Numer. Linear Algebr. Appl.

[12] Pultarová I. *Local convergence analysis of iterative aggregation-disaggregation methods with polynomial correction*. Linear Algebra Appl. **421** 122-137 (2007).

[13] Stewart W.J. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, Princeton, NJ., 1994.