Electronic Transactions on Numerical Analysis.
Volume 35, pp. 185-200, 2009.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

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

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

Abstract. Iterative aggregation/disaggregation (IAD) methods for the computation of stationary probability vectors of large scale Markov chains form efficient practical analysis tools. However, their convergence theory is still not developed appropriately. Furthermore, as in other multilevel methods such as multigrid methods, the number of relaxations on the fine level of the IAD algorithms which is to be executed plays a very important role. To better understand these methods, in this paper we study some new concepts as well as their behavior and dependence on the parameters involved in aggregation algorithms, and establish some necessary and/or sufficient conditions for convergence. The theory developed offers a proof of convergence of IAD algorithms independent of whether the governing iteration matrix is primitive or cyclic as one of its main results. Another important result concerns a comparison of the rates of convergence of two IAD processes. Some examples documenting the diversity of behavior of IAD methods are given.

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 [14, Section 6.3]), aggregation/disaggregation iterative (IAD) methods belong to competitive classes of methods for the computation of stationary probability vectors of Markov chains and Leontief systems. An appropriate convergence theory is still far from being complete—in the sense of a full understanding of the dependence of these methods upon all the parameters influencing convergence and its rate. The aim of this contribution is to explain some of the more subtle parts of the theory. In particular, some new concepts, such as convergence indices, are introduced and their properties are studied. These convergence indices allow one 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 methods. Some results in this direction appear to be decisive in understanding the interplay between the basic iteration matrix and the IAD algorithms. Consequently, a new important result is established: IAD processes return convergent sequences of iterates without requiring the splittings, upon which the IAD algorithms are based, to be convergent; for example, splittings leading to cyclic iteration matrices are not only allowed but they may even be preferable in comparison with primitive iteration matrices obtained using shifts of the originally cyclic iteration matrices.

2. Definitions and notation.

2.1. Generalities.

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

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

where $\sigma(C)$ denotes the spectrum of C. When we suppose C to be stochastic, we can assume that $\rho(C) = 1$. Further, we define the quantity

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

*Received December 27, 2007. Accepted November 11, 2008. Published online on November 5, 2009. Recommended by Anne Greenbaum.

†Department of Mathematics, Czech University of Technology, Thákurova 7, 16629 Praha 6, Czech Republic ({marek,pmaye}@ms.mff.cuni.cz, ivana@mat.fsv.cvut.cz).

ETNA
Kent State University
http://etna.math.kent.edu

We are going to call $\gamma(C)$ *the convergence factor* of C. We also need another more general characteristic of convergence. Therefore we give the following definition.

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

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

This quantity is called the spectral subradius *of C.*

REMARK 2.2. 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 and B an irreducible column stochastic matrix with spectral decomposition

$$B = Q + Z,$$

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

$$Q = \sum_{j=1}^{p} \lambda^{j-1} Q_j,$$

with $Q_j Q_k = Q_k Q_j = \delta_{jk} Q_j$, $j, k = 1, \ldots, p$.

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

$$\mathcal{G} : \{1, \ldots, N\} \xrightarrow{\text{onto}} \{1, \ldots, n\}.$$

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

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

Iterative aggregation/disaggregation communication operators are defined as

$$(Rx)_j = \sum_{\bar{j} : \mathcal{G}(\bar{j}) = j} x_{\bar{j}},$$

and

$$S = S(u), \quad \big(S(u)x\big)_j = \frac{u_j}{(Ru)_{\bar{j}}}(Rx)_{\bar{j}}.$$

Obviously for any $u \in \mathbb{R}^N$, $u^T = (u_1, \ldots, u_N)$, $u_j > 0$, $j = 1, \ldots, N$, we have

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

and

$$S(u)Ru = u,$$

ETNA
Kent State University
http://etna.math.kent.edu

but, in general,

$$S(u)Rx \neq x.$$

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

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

and

$$P(x)x = x, \quad \forall x \in \mathbb{R}^N,\; x_j > 0,\; j = 1, \ldots, N.$$

We define the *aggregated matrix* as

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

3. IAD algorithms.

3.1. Algorithm $\mathrm{SPV}(B; T; t, s; x^{(0)}; \mathcal{G}; \varepsilon)$ **(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, $x^{(0)} \in \mathbb{R}^N$ an elementwise positive vector, and $\varepsilon > 0$ a tolerance.

Step 1. Set $k = 0$.

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

$$\mathcal{B}\big(x^{(k)}\big) = RB^s S\big(x^{(k)}\big).$$

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

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

Step 4. Let

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

and

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

Step 5. Test whether

$$\big\|x^{(k+1)} - x^{(k)}\big\| < \epsilon.$$

Step 6. If NO in Step 5, then let

$$k + 1 \longrightarrow k$$

and GOTO Step 2.

Step 7. If YES in Step 5, then set

$$\hat{x} := x^{(k+1)}$$

and STOP.

The error matrix is defined as

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

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

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

When the dependence of the iteration sequence on the iteration matrix must be specified explicitly, we write $x^{(k)}(T)$ in place of $x^{(k)}$, where $T = M^{-1}W$.

REMARK 3.1. Algorithm SPV can be easily modified in order to construct solutions for Leontief systems, i.e., $x = Cx + b$, with $b \in \mathbb{R}^N, C \in \mathbb{R}^{N\times N}$ elementwise nonnegative, and $\lim_{k\to\infty} C^k = 0$. Let us denote the modified algorithm by Algorithm LM. Such algorithms have been studied in [4]. It is shown in [7] that the error matrices of both classes of algorithms are essentially the same. This is the reason why we are only going to study algorithms of class SPV.

Algorithms of the type introduced in this section are known as Leontief procedures; they were invented by Leontief around 1930 in his famous sectorial economy theory. Actually, his sectorial variables are just aggregates of the initial variables, and the sectorial production matrix is our aggregation matrix.

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

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

where [6]

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

The matrix Z comes from the spectral decomposition of $B = Q + Z$, where $Q^2 = Q$, $QZ = ZQ = 0$, and $1 \notin \sigma(Z)$. 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 the convergence properties of IAD methods, without the explicit requirement that the basic iteration matrix is convergent, i.e., we do not assume that the limit

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

exists.

REMARK 4.1. One of the most delicate questions concerning IAD methods is the following: how to choose the number of smoothing t? The answer to this question is not a simple matter, as illustrated by the following example. This takes us back to another basic question, namely, how to aggregate? Some results concerning the convergence issues of the SPV algorithm with a small number of smoothing t can be found in [12].

EXAMPLE 4.2. Assume that $p > 1$ is a positive integer and that B is the transition matrix of a Markov chain, which can be written in a block form as

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

ETNA
Kent State University
http://etna.math.kent.edu

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

$$M = \operatorname{diag}\left\{B_{11}, \dots, B_{pp}\right\}, \quad W = B + M - I.$$

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

The aggregation communication operators are chosen, such that

$$R = (1, \dots, 1)^T$$

is $1 \times N$ matrix and

$$S(x)z = \frac{z}{Rx}x, \quad x \in \mathbb{R}^N, \quad x_j > 0, \quad j = 1, \dots, N, \quad z \in \mathbb{R}.$$

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

On the other hand, if the aggregation operators are chosen as shown above, i.e., each single block of matrix B is aggregated to a 1×1 matrix, the situation may change dramatically. As an example, let us take the transition matrix whose off-diagonal row blocks satisfy $B_{jk} = v_j u_{jk}^T$, $j \neq k$, where v_j and u_{jk}, $j, k = 1, \dots, n$, 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 [8].

The theoretical knowledge of the IAD methods expressed e.g. in Theorem 4.4 and our computational experience with such methods leads to the following conclusions: Even a quite deep theoretical knowledge and a massive computational experience do not offer sufficient information how to aggregate successfully. This is because the quantities relevant for determining suitable parameters are in practice very difficult to get. Our recommendation how to obtain the parameters just mentioned is to exploit any knowledge and the knowledge coming from the outside of mathematics in particular. By these statements we do not say that the IAD methods should not be considered as suitable means for computations. Just the opposite is true and in particular to determining the characteristics of Markov chains such as the stationary probability vectors, the mean first visit times matrices, etc. and solutions to problems whose mathematical model states are probabilities.

4.1. Aggregation-convergence. Let us remind ourselves of a definition that is relevant within in the context of IAD methods [7].

DEFINITION 4.3. *Assume that B is an $N \times N$ irreducible stochastic matrix with stationary probability vector $\hat{x}$ and R and $S(x)$ IAD communication operators. A splitting of $I - B$, where*

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

is called aggregation-convergent *if*

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

An interesting question is how to recognize that a splitting is aggregation-convergent. When looking at the error-vector formula valid for any IAD constructed using the splitting

$$A = I - B = M(I - T), \quad T \geq 0, \tag{4.1}$$

we can summarize our knowledge concerning the class of IAD algorithms by the following theorem.

THEOREM 4.4 (see [7]). *Consider Algorithm* $\mathrm{SPV}(B; M, W, T;\ t, s = 1;\ \mathcal{G}; x^{(0)}; \varepsilon)$ *with an irreducible stochastic matrix* B*, an aggregation-convergent splitting (4.1) and an initial guess taken such that* $x^{(0)} \in \mathrm{Int}\,\mathbb{R}^N_+$*. Then, there exist 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* $\mathrm{SPV}(B; M, W, T; t, s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$ *returns a sequence of iterates* $\{x^{(k)}\}$ *for which*

$$\lim_{k\to\infty} x^{(k)} = \hat{x} = B\hat{x}, \quad e^T\hat{x} = 1, \tag{4.2}$$

for $t = \tilde{t}$ *and* $x^{(0)} \in \Omega_{\tilde{t}}(\hat{x})$*, or for* $t \geq \hat{t}$ *and* $x^{(0)} \in \Omega_{\hat{t}}(\hat{x})$*.*

REMARK 4.5. Theorem 4.4 deserves some comments. Let $\underline{t}$ and $\overline{t}$ be the minimum of those $\tilde{t} \geq 1$ and $\hat{t}$ for which (4.2) holds, respectively.

(a) First of all, $\underline{t}$ in (4.2) may be large, in particular if $\underline{t} = \overline{t}$. This effect is caused by the "interaction" of the matrices $P(\hat{x})B$ and $P(x)T$; in the case of $T = B$ it may also be caused by the possible nonnormality of $(I - P(x))B[I - P(x)Z]^{-1}$. In this context let us recall a popular problem of shuffling cards (see Greenbaum [3]).

(b) There may be lots of integers m and r such that $m < r < \overline{t}$ and Algorithm $\mathrm{SPV}(B; T;\ t, s = 1; x^{(0)}; \varepsilon)$ is divergent for $t = m$ and convergent for $t = r$ (see Example 8.1).

(c) There are examples [13] showing that $\mathrm{SPV}(B; B;\ t = 1,\ s = 1; x^{(0)}; \varepsilon)$ converges and $\mathrm{SPV}(B; B;\ t = 2,\ s = 1; x^{(0)}; \varepsilon)$ does not.

EXAMPLE 4.6 (see [13]). 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

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

and

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

5. Necessary and/or sufficient conditions for local and global convergence. Let $\mathcal{G}$ be a mapping of the index set $\{1, \ldots, N\}$ onto $\{1, \ldots, n\}$, and R and $S(x)$ the corresponding communication maps determining the aggregation projection $P(x) = S(x)R$ [7]. 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* $\mathrm{SPV}(B; T = M^{-1}W; t, s; \mathcal{G}; x^{(0)}; \varepsilon)$ *if this algorithm returns convergent sequences of iterates for* $t = r$*, and divergent ones for* $t = r + 1$*, where the integer* r *is the smallest among all such numbers.*

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

In this section we are going to examine convergence issues concerning Algorithm $\mathrm{SPV}(B;\ T = M^{-1}W; t, s; \mathcal{G}; x^{(0)}; \varepsilon)$, formulated in terms of its a and b indices. Before we formulate the appropriate statements, we comment on Theorem 4.4 utilizing the above-introduced convergence indices.

REMARK 5.3. Parts of Theorem 4.4 concerning the relation (4.2) can be reformulated as follows: the a-index of $\mathrm{SPV}(B;T;t,\ s=1;\mathcal{G};x^{(0)};\varepsilon)$, which we denote by t_a, satisfies $t_a \geq \tilde{t}$, and for the b-index, denoted by t_b, we have $t_b \leq \hat{t}$.

There are classes of stochastic matrices for which the SPV Algorithms possess small a-indices, actually the smallest possible, i.e., $\tilde{t} = 1$; see [13]. To this class there belong stochastic matrices possessing positive diagonals. The rate of convergence of SPV iterative procedures with just one smoothing on the fine level for such matrices may be slow however. To speed up the computational process is not an easy task because the process with number of smoothing $t = \tilde{t} + m, m \geq 1$ may diverge for rather quite a lot of values m's! We see that though the relation $B = C + \gamma I$, where $\gamma > 0$ and C is elementwise nonnegative, implies that $B^k \geq \delta_k I$ for some $\delta_k > 0$, index 1 need not be the b-index of B^k.

The convergence indices just introduced allow us to formulate adequate conditions necessary and sufficient for convergence and/or divergence of SPV algorithms. At this point a note is appropriate. The counterexamples shown, as well as the appearance of the convergence indices, are needed if one is to fully understand the convergence of IAD methods. On the one hand, there are some irregularities in even the local convergence behavior for cases with a small number of smoothings and, on the other hand, the necessity of requiring a large number of smoothings for guaranteeing global convergence. Only a good understanding of as much as possible of all these convergence issues may give correct recommendations for practical computations.

PROPOSITION 5.4 (sufficient conditions for global convergence). *Assume that* $A = I - B = M - W$, *where* $B = P + Z$, $P^2 = P$, $PZ = ZP = 0$, $T = M^{-1}W$, *where* B *is an* $N \times N$ *irreducible stochastic matrix, and* $T = C + \gamma I$, *where* $\gamma > 0$, *is a nonnegative matrix.*

Then there exist two positive integers $\tilde{r}$ *and* $\tilde{t}$, *such that for any* $x^{(0)} \in \mathbb{R}^N_+ \cap \mathrm{range}(T^t)$, $t \geq \tilde{r}$, *Algorithm* $\mathrm{SPV}(B;T;t,s;\mathcal{G};x^{(0)};\varepsilon)$ *for* $t \geq \tilde{t}$ *and* $s \geq 1$ *returns a convergent sequence of iterates* $\{x^{(k)}\}$, *i.e.,*

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

Proof. Let $\mathcal{K} = \{x \in \mathbb{R}^N_+ : [e(N)]^T x = 1\}$. Since for every $x \in \mathcal{K}$

$$\lim_{k\to\infty} T^k x = \hat{x}e^T x, \quad e^T x = 1, \tag{5.1}$$

it follows that there is a positive integer $\overline{r}$, such that

$$\left(T^t x\right)_j \geq \frac{1}{2}(\hat{x})_j, \quad j = 1, 2, \ldots, N \text{ and } t \geq \overline{r}.$$

We want to prove that there is a positive integer $\tilde{t}$, such that

$$\left\| T^t \left[I - P(x)Z\right]^{-1} \left(I - P(x)\right) \right\| < 1 \quad \forall t \geq \tilde{t}, \tag{5.2}$$

whenever $x \in \mathbb{R}^N_+ \cap \mathrm{range}(T^{\overline{r}})$. For this purpose we utilize the relations

$$T = Q + U, \quad Q = \hat{x}e^T, \quad e^T\hat{x} = 1, \quad QU = UQ = 0, \quad \rho(U) < 1,$$
$$w = \left[I - P(x)\right]w, \quad \lambda w = T^t\left[I - P(x)Z\right]^{-1}\left(I - P(x)\right)w, \quad w \neq 0,$$

valid for any eigenpair λ and w, implying by (5.1) that

$$\lim_{k\to\infty} \left[I - P(x)\right]\left\{T^t\right\}^k = \lim_{k\to\infty} \left(I - P(x)\right)\left[\hat{x}e^T + U\right]^k = 0, \quad x \in \mathcal{K},\ t \geq \overline{t}.$$

ETNA
Kent State University
http://etna.math.kent.edu

In other words, there is a positive integer $\check{t}$, such that

$$\rho\Big(\big[I - P(x)\big]T^t\Big) < 1, \quad t \geq \check{t}. \tag{5.3}$$

The compactness of the set $\mathcal{K}$ implies the existence of a constant τ independent of $x \in \mathcal{K}$, such that relation (5.3) can be stated in a stronger form as

$$\rho\Big(\big[I - P(x)\big]T^t\Big) \leq \tau < 1, \quad t \geq \bar{t} \text{ uniformly with respect to } x \in \mathcal{K}.$$

Set

$$\|u\|_{P(x)} = \big\|P(x)u\big\|_1 + \big\|\big(I - P(x)\big)u\big\|_1, \quad u \in \mathbb{R},\ x \in \mathcal{K}.$$

It follows that

$$\big\|\big[I - P(x)\big]T^t\big\|_{P(x)} < 1, \quad t \geq \bar{t}.$$

Consequently, the required positive integer in (5.2) is obtained as

$$\tilde{t} = \min\Big\{t : \kappa\Big\|\big[I - P(x)\big]\big\{T^{\bar{t}}\big\}^t\Big\|_{P(x)} < 1\Big\},$$

where (since $\|(I - P(x))\|_{P(x)} = 1$)

$$\kappa = \max\Big\{\Big\|\big[I - P(x)Z\big]^{-1}\Big\|_{P(x)} : x \in \text{range}\,\big(T^{\overline{r}}\big) \cap \mathbb{R}^N_+,\ x\big[e(N)\big]^T x = 1\Big\}.$$

The proof is complete. □

REMARK 5.5. The above examples and Proposition 5.4 show several reasons for having difficulties in proving the global convergence of IAD methods. We must, therefore, consider sufficient conditions for global convergence in light of this. The conditions are rather complicated from the viewpoint of computer implementation and the conditions guaranteeing global convergence may seem impractical. Thus, if one is interested in computing the stationary probability vector in practice, then one should better choose $\text{SPV}(B; T;\ s = 1,\ t = 1; x^{(0)}; \varepsilon)$ yielding the approximation sequence $\{x^{(0)}\}$ locally convergent [10, 12] than $\text{SPV}(B; T;\ s = 1,\ t = 1; x^{(0)}; \varepsilon)$ with sufficiently large s and t in order to guarantee global convergence.

REMARK 5.6. It may seem strange that such a simple statement with a quite trivial proof may appear as new after a relatively long period of investigating IAD methods. The reason might be the absence of a good understanding of the role of the indices t and s in the SPV algorithm. The examples of Section 8, showing, for example, that $\text{SPV}(B; B;\ t = 1,\ s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$ may not converge even locally while $\text{SPV}(B; B;\ t = 2,\ s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$ does, have led us to define the a- and b- convergence indices and to find a way towards local and global convergence of IAD algorithms.

PROPOSITION 5.7 (necessary condition for local convergence). *Suppose that all free variables of Algorithm $SPV(B;\ T = M^{-1}W;\ t = s = 1; x^{(0)}; \varepsilon)$ are fixed and*

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

Then

$$\rho\big(\big(I - P(\hat{x})\big)T^t\big) < 1 \quad \forall t \geq \hat{t}.$$

ETNA
Kent State University
http://etna.math.kent.edu

Proof. Let us assume that the conclusion of Proposition 5.7 is false. Then the spectral resolution of the of matrix $(I - P(\hat{x}))T^t$ reads

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

where

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

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

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

It follows that

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

Thus, the implication

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

is contradictory. □

PROPOSITION 5.8 (necessary and sufficient conditions for local convergence). *Let $t_b \geq 0$ be the b-index of Algorithm* $\mathrm{SPV}(B; T;\ t \geq t_b,\ s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$. *Then the following conditions (i) and (ii) are equivalent.*

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

(ii) The relation

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

holds for all $t \geq t_b$.

As a consequence of Propositions 5.8 and 5.4 we deduce the following theorem.

THEOREM 5.9. *To every SPV Algorithm* $\mathrm{SPV}(B; T; t,\ s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$ *there belongs a finite b-index of convergence.*

REMARK 5.10. 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, divergence may appear only in the extreme case of the sufficient conditions taking place, i.e., if $\rho((I - P(\hat{x}))T) = 1$.

THEOREM 5.11. *Let B be a column stochastic matrix and $I - B = M - W = M(I - T)$ its splitting of a nonnegative type with iteration matrix $T \geq 0$ and $e^T M = \tilde{e}^T = (\eta_1 e(r_1)^T, \ldots, \eta_n e(r_n)^T)$.*

Then

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

Proof. In order to estimate the spectral radius of $V = (I - P(\hat{x}))T$, we will consider the matrix

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

ETNA
Kent State University
http://etna.math.kent.edu

similar to V, where

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

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

$$z^T T_s = e^T MDT_s = e^T MDD^{-1}M^{-1}WD = e^T WD = e^T MD = z^T.$$

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

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

Further, since

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

we have

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

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

$$\begin{aligned}\big\|D^{-1}\big(I - P(\hat{x})\big)TD\big\|_2 &\leq \big\|D^{-1}\big(I - P(\hat{x})\big)D\big\|_2\big\|T_s\big\|_2\\ &= \big\|D^{-1}\big(I - P(\hat{x})\big)D\big\|_2\rho\big(T_s^TT_s\big)\\ &\leq \big\|D^{-1}\big(I - P(\hat{x})\big)D\big\|_2\big\|T_s^TT_s\big\|_{(z)}\\ &\leq \big\|D^{-1}\big(I - P(\hat{x})\big)D\big\|_2\big\|T_s\big\|_{(z)}\big\|T_s^T\big\|_{(z)}\\ &\leq \big\|D^{-1}\big(I - P(\hat{x})\big)D\big\|_2.\end{aligned}$$

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

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

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

REMARK 5.12. The assumptions of Theorem 5.11 were restricted to the case where M is chosen in such a way that

$$e^TM = \tilde{e}^T = \Big(\eta_1 e\big(r_1\big)^T,\dots,\eta_n e\big(r_n\big)^T\Big). \tag{5.4}$$

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

In the case of the special choice of M given by relation (5.4), the matrix D can be diagonal with

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

ETNA
Kent State University
http://etna.math.kent.edu

for i with $\mathcal{G}(i) = k$ as introduced in the proof of Theorem 5.11. But when we do not assume (5.4), the matrix D in a diagonal form does not fulfill these three conditions. 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 5.11 is valid without assumption (5.4). Still we cannot provide the exact construction of D yet.

6. Convergence of IAD within the class of irreducible stochastic matrices. If one looks 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 $\hat{k}$. At a first glance, 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 have built up massive numerical evidence that the IAD processes with iteration matrices T_m, $m = 1, 2, \ldots$, 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 increasing index m. This observation has led us to the 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

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

where H is a permutation matrix.

AGREEMENT 6.1. *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, for which 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

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

and

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

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

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

ETNA
Kent State University
http://etna.math.kent.edu

where

$$\begin{aligned}
&Q_j^2 = Q_j, \quad Q_j Q_k = Q_k Q_j = 0, \quad j \neq k,\\
&Q_j Z(T) = Z(T) Q_j = 0,\\
&\rho\big(Z(T)\big) < 1.
\end{aligned}$$

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

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

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

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

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

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

and, thus,

$$\big(I - P(\hat{x})\big) y = 0. \tag{6.2}$$

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

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

Since

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

we also have that

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

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

Since, according to (6.2),

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

ETNA
Kent State University
http://etna.math.kent.edu

we have

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

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

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

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

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

Thus, we have convergence.

Summarizing we obtain the following theorem.

THEOREM 6.2. *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 neighbourhood $\Omega(\hat{x})$, such that the SPV Algorithm returns a sequence of iterates $\{x^{(k)}\}$, such that

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

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

REMARK 6.3. Because of the counterexamples shown, generally one cannot prove more. There are some results on the local convergence properties for some special types of the aggregation algorithm [12].

7. A comparison result. Our numerous numerical experiments concerning the application of Algorithm $\mathrm{SPV}(B; I, B;\ t = 1,\ s = 1; x^{(0)}; \varepsilon)$ never failed to converge when applied to practical problems. One possible explanation might be that in any neighbourhood of an irreducible stochastic matrix for which a given SPV algorithm returns divergent sequences of iterates, there is another stochastic matrix for which the same algorithm returns convergent sequences of iterates.

The next result enlightens to some extent the role of the b-index of the basic algorithm SPV. An obvious fact is shown, namely that the smaller the spectral radius of the variable part of the error matrix, the faster convergence of the corresponding SPV algorithm.

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

Assume that the inequality

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

holds for $t \geq t_1$, where t_1 is the b-convergence index of $\mathrm{SPV}(B; T_1;\ t, s = 1; \mathcal{G}, x^{(0)}; \varepsilon)$ *and where $\Omega_1(\hat{x})$ is a corresponding neighbourhood of local convergence. Then Algorithm* $\mathrm{SPV}(B; T_2;\ t, s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$ *is locally convergent too with b-convergence index $t_2 \leq t_1$.*

Proof. Let $x^{(k)}(T_j^t)$ denote the iterate returned by Algorithm $\mathrm{SPV}(B; T_2;\ t, s = 1; \mathcal{G}; x^{(0)}; \varepsilon)$. Our goal is to show convergence of the sequence $\{x^{(k)}(T_2^t)\}$ for $t \geq t_1$. By constructing the sequence we know that it is componentwise uniformly bounded and hence it is precompact as a bounded set. Let $\tilde{y}$ be any of its points of condensation and let us assume,

ETNA
Kent State University
http://etna.math.kent.edu

without loss of generality, that the corresponding convergent subsequence coincides with that of $\{x^{(k)}(T_2^t)\}$. Inequality (7.1) implies that

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

holds for $k = 1, 2, \ldots$. Thus,

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

Local convergence of Algorithm $\mathrm{SPV}(B; T;\ t, s = 1\mathcal{G}; x^{(0)}; \varepsilon)$ then follows according to the next Proposition 7.2. The proof of Theorem 7.1 is complete. □

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

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

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

8. Examples.

EXAMPLE 8.1. We compute the spectral radii of error matrices

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

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

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

and for $\mathcal{G}(1) = \mathcal{G}(2) = 1$, $\mathcal{G}(3) = 2$. We assume that $t = 1, \ldots, 15$ and vary values of a and b. The values of a and b represent situations of a nearly cyclic or a nearly reducible matrix B:

(i) $a = b = 0.9$, this means a 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 divergence 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 8.2. While in Example 1 all choices of a and b lead to a local convergent IAD process for all $t = 1, 2, \ldots, 15$, the situation is different for the matrix

$$B = \begin{bmatrix} 0 & 0 & 0 & 0.1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \end{bmatrix}.$$

We consider $\mathcal{G}(1) = \mathcal{G}(2) = 1$, $\mathcal{G}(3) = \mathcal{G}(4) = \mathcal{G}(5) = 2$. In Figure 8.2, we show the spectral radii of the 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 that diverge locally.

ETNA
Kent State University
http://etna.math.kent.edu

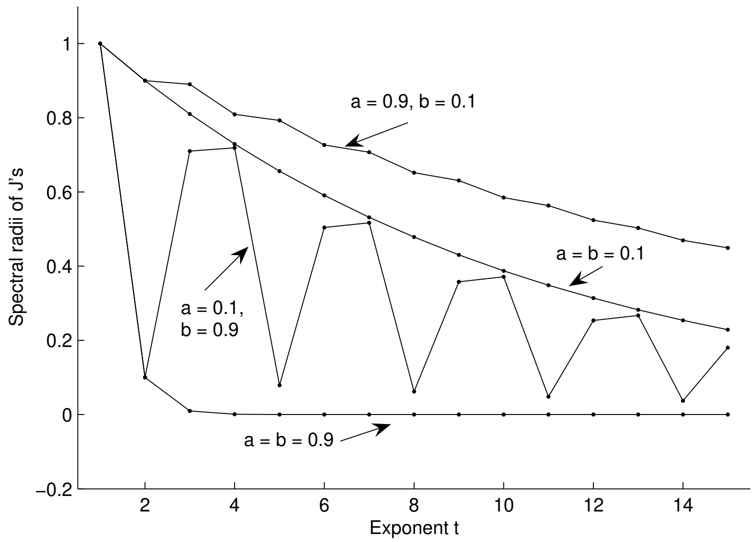

FIG. 8.1. *Spectral radii of the matrices J_t for four different choices of a and b in Example 1.*

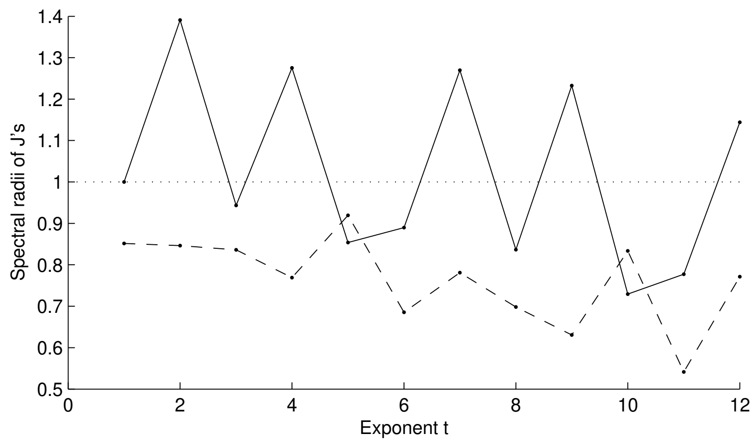

FIG. 8.2. *Spectral radii of the matrices J_t (solid line) and $(I - P(\hat{x}))Z^t$ (dashed line) in Example 2.*

9. Concluding remarks. Summarizing our observations, we can say that the theory, computer experiments, and practical computations confirm a view accepted in the literature, namely that iterative aggregation/disaggregation methods are competitive means for computing the characteristics of Markov chains, in particular stationary probability vectors. Let us recall that any IAD method possesses a finite b-index of convergence independent of whether the governing iteration matrix is primitive or cyclic. This property significantly distinguishes the IAD methods from other methods. Other results we want to mention explicitly concerns the theory of convergence indices including a new type of comparison of rates of convergence. We have also observed examples showing the divergence of some SPV algorithms. 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.

ETNA
Kent State University
http://etna.math.kent.edu

Acknowledgments. The authors of this paper express their gratitude to anonymous referees and to the Editor of the special volume devoted to contributions delivered at the Applied Linear Algebra Conference held in Harrachov in August 2007 and the Chief Editor for their help in preparing the final version of the paper for publication, in particular for their help in managing a linguistic expertise. The work on which this contribution is based was supported by the Program Information Society under Project 1ET400300415, Grant No. 201/09/1544 of the Grant Agency of the Czech Republic and by the Grants No. MSM 6840770010 and No. MSM 6840770001 of the Ministry of Education, Youth and Sports of the Czech Republic.

Appendix A. A special norm.

Let C be a real $N \times N$ matrix and let $z^T = (\zeta_1, \ldots, \zeta_N)$, $\zeta_j > 0$, $j = 1, \ldots, N$. Define

$$\|C\|_{(z)} = \nu_z(C) = \min\left\{\alpha \in \mathbb{R}_+ : \left|C^T\right| z \leq \alpha z\right\}, \tag{A.1}$$

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

PROPOSITION A.1 (see [4]). *Expression (A.1) is a norm on the space of $N \times N$ matrices over R.*

REFERENCES

[1] P. J. COURTOIS AND P. SEMAL, *Block iterative algorithms for stochastic matrices*, Linear Algebra Appl., 76 (1986), pp. 59–80.

[2] T. DAYAR AND W. J. STEWART, *Comparison of partitioning techniques for two-level iterative solvers on large, sparse Markov chains*, SIAM J. Sci. Comput., 21 (2000), no. 5, pp. 1691–1705.

[3] A. GREENBAUM, *Card shuffling and the polynomial numerical hull of degree k*, SIAM J. Sci. Comput., 25 (2003), no. 2, pp. 408–416.

[4] J. MANDEL AND B. SEKERKA, *A local convergence proof for the iterative aggregation method*, Linear Algebra Appl., 51 (1983), pp. 163–172.

[5] I. MAREK AND P. MAYER, *Aggregation/disaggregation iterative methods applied to Leontev systems and Markov chains*, Appl. Math., 47 (2002), no. 2, pp. 139–156.

[6] I. MAREK AND P. MAYER, *Convergence analysis of an aggregation/disaggregation iterative method for computation stationary probability vectors of stochastic matrices*, Numer. Linear Algebra Appl., 5 (1998), pp. 253–274.

[7] I. MAREK AND P. MAYER, *Convergence theory of a class of aggregation/disaggregation iterative methods for computing stationary probability vectors of stochastic matrices*, Linear Algebra Appl., 363 (2002), pp. 177–200.

[8] I. MAREK AND P. MAYER, *Iterative aggregation/disaggregation methods for computing stationary probability vectors of stochastic matrices can be finitely terminating*, Int. J. Differ. Equ. Appl., 3 (2001), pp. 301–313.

[9] I. MAREK, P. MAYER AND I. PULTAROVÁ, *IAD methods based on splittings with cyclic iteration matrices*, Dagstuhl Seminar Proceedings 07071, 2007.

[10] I. MAREK AND I. PULTAROVÁ, *A note on local and global convergence analysis of iterative aggregation/disaggregation methods*, Linear Algebra Appl., 413 (2006), pp. 327–341.

[11] I. MAREK AND K. ŽITNÝ, *Analytic Theory of Matrices for Applied Sciences Vol 1*, Teubner Texte zur Mathematik Band 60, Leipzig, 1983.

[12] I. PULTAROVÁ, *Necessary and sufficient local convergence condition of one class of iterative aggregation – disaggregation methods*, Numer. Linear Algebra Appl., 15 (2008), pp. 339–354.

[13] I. PULTAROVÁ, *Local convergence analysis of iterative aggregation-disaggregation methods with polynomial correction*, Linear Algebra Appl., 421 (2007), pp. 122–137.

[14] W. J. STEWART, *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, Princeton, New Jersey, 1994.

[15] R. S. VARGA, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey, 1962. Second Edition, revised and expanded, Springer, Berlin, Heidelberg, New York, 2000.