
ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

LINEAR ALGEBRA AND ITS APPLICATIONS

Linear Algebra and its Applications 363 (2003) 177–200

www.elsevier.com/locate/laa

Convergence theory of some classes of iterative aggregation/disaggregation methods for computing stationary probability vectors of stochastic matrices

Ivo Marek [a,*], Petr Mayer [b]

[a] *Katedra Matematiky, Stavební Fakulta Českého Vysokého učení Technického, Thákurova 7, 160 00 Praha 1, Czech Republic*

[b] *Katedra Numerické Matematiky, na Matematicko-Fyzikální Fakultě University Karlovy, Sokolovská 83, Czech Republic*

Received 8 April 2001; accepted 12 February 2002

Submitted by R. Nabben

Abstract

This contribution is a natural follow-up of the paper of the same authors entitled *Convergence theory of an aggregation/disaggregation methods for computing stationary probability vectors of stochastic matrices* published in [Numer. Linear Algebra Appl. 5 (1998) 253]. In contrast to that paper in which the algorithm studied was based on the splitting whose iteration matrix was identical with the matrix whose stationary probability vectors are computed, the present paper presents a convergence analysis of algorithms based on fully general splittings of nonnegative type. Together with this generalization another feature of the older paper and namely the independence of the convergence results on the size of the elements of the examined stochastic matrix is shown to remain valid for the new algorithms as well. This concerns in particular the possibility of computing stationary probability vectors of Markov chains containing rare events, i.e. events whose stationary probabilities are substantially smaller than some of elements of the transition matrix of the chain.

Keywords: Markov chain; Stochastic matrix; Aggregation/disaggregation procedure; Stationary distribution

* Corresponding author.

0024-3795/03/$ - see front matter
PII: S0024-3795(02)00333-6

1. Introduction

One of the aims of mathematical modeling of reliable safety railway systems is to compute the probabilites of rare events, i.e. such events whose appearance is as a rule unwanted, e.g. crashes on the railways etc. [11,13]. Rare events are such events where probability is much smaller than some of the input data, i.e. much smaller than some elements of the transition matrix B. Usually, these probabilites range around 10^{-12} in our nonacademic examples. Thus, some fast convergent methods are needed. Though, row block rank one matrices may look too academic, they often appear in practice and their tensor products serve as tools for constructing suitable Markov chains that model the appropriate effects. Stationary probability vectors (SPVs) of weakly perturbed row block rank one matrices can be efficiently computed by adequate IAD methods (see Theorem 6.7).

In various applications it is desirable to be able to obtain the corresponding probabilites that are essentially smaller compared with the probabilites of other events. If the mathematical models considered, e.g. models of safety systems, are based on Markov chains then obviously the corresponding transition matrices will become unstable and consequently, computing the stationary probability (distribution) vectors represents a difficult task.

For computing SPVs many algorithms have been developed in the past and the problem of finding all SPVs has been attacked both from theoretical as well as experimental point of view. To the knowledge of the authors the works devoted to comparing various methods how to compute SPVs conclude that the winners in the competition are some versions of iterative aggregation/disaggregation (IAD) methods [7,15]. However, there is no adequate convergence theory. Essentially, there are very few convergence proofs of IAD methods available in the literature [3,18,21,25] and those existing cover a rather limited area of problems. The monograph [25] presents the most complete description of the situation in the area of computation of Markov chain modeling in the middle of the nineties. The convergence proof presented there in fact reproduces the original proof from [3]. It is based on the assumption that the examined Markov chains are nearly completely decomposable (NCD). Thus, the smallness of the elements of the off-diagonal blocks of the corresponding transition matrices is a tool for achieving the convergence. A weakness of this approach is first that convergence is achieved only when the elements of the off-diagonal blocks are sufficiently small and second, the hypothesis that the transition matrices must satisfy certain *regularity conditions* [25, p. 335] excludes explicitly a possibility of computing probabilities of rare events.

In this contribution a general convergence theory is presented. This paper is a natural follow-up of [21], where convergence theory is presented for one particular splitting of matrix $A = I - B$, where B is a transition matrix of a Markov chain. We extend the theory by admitting any splitting of nonnegative type. The generalization is essential in the sense that the corresponding iteration matrix may not commute with B. We show that any IAD algorithm based on a splitting $\{M, W\}$ of $A = I - B$,

is convergent whenever $T = M^{-1}W$ is nonnegative and such that a possible cyclicity of T is preserved in an appropriate way by the aggregation communication. As analytical tools for proving that some new concepts such as $\mathcal{G}$-convergence and $\mathcal{G}$-consistence of splittings of nonnegative type are invented and a theory developed. A sufficient condition for $\mathcal{G}$-convergence (see Definition 5.4) is a consequence of an interplay between the corresponding iteration matrix T, possibly cyclic, and the IAD map $\mathcal{G}$. Thus, information that a given IAD is convergent is available without the aggregation projection to be known explicitely (see Theorem 5.8). The well-known IAD methods such as [17,21,27] can serve as examples of this fact.

We also identify classes of problems and suitable splittings for which the IAD methods terminate after two iteration sweeps. The finite termination results are then extended via continuity of the IAD process to fast convergence for a wide class of problems. Furthermore, we specify some results to a class of p-cyclic Markov chains.

The first four sections are devoted to introduction, definitions, notation and presentation of the algorithms. Our theory is based on splittings which need not be even convergent, nevertheless, this theory applies to any stochastic matrix and we emphasize that the SPVs are allowed to possess small componets, e.g. corresponding to rare events as well (Section 5). We also identify some classes of Markov chains for which some of the IAD algorithms are fast convergent, in some cases exact SPV are returned after a finite number of iteration sweeps (Section 6). We elucidate the situation with convergence characteristics by introducing a new concept of *aggregation-convergent matrix*. Actually, we discovered the fast convergence effect just by analyzing splittings based on divergent splittings that are aggregation-convergent. Continuing in this way IAD appear as very efficient for p-cyclic Markov chains (Section 7). In this context we show a rather surprising result: Utilizing as bases for IAD algorithms the SOR methods, both within standard as well as extended convergence range of relaxation parameters, the Gauss–Seidel procedure appears as optimal. Moreover, if starting with a vector whose subvectors are limits obtained by the Gauss–Seidel procedure, just one IAD step is sufficient in order to get the required SPV. It is worth mentioning that in our theory concerning the p-cyclic Markov chains the frequently used hypotheses that the transition matrix B is consistently ordered and the spectrum of B^p is real can be omitted. From the point of view of computations a relatioship between the error on the fine and aggregated levels is of high importance (Section 8).

2. Definitions and notation

In this section we present notation and definitions which we are going to use. Some very standard concepts not explicitely defined here can be found in the monograph [1] a standard reference of ours.

Our analyses are provided in finite dimensional Banach spaces. Because of equivalence of all norms on such spaces we can in principle use any norm. In the context of

stochastic matrices the l_1-norm is the most adequate, however. In the whole paper the symbol $\|\cdot\|$ denotes the l_1-norm on the appropriate space if any other specification is not declared.

Let N be a positive integer. Objects of our investigation are matrices whose elements are real numbers. An $N \times N$ matrix $C = (c_{jk})$ with $c_{jk} \in \mathbb{R}^1$, is called nonnegative if $c_{jk} \geqslant 0,\ j, k = 1, \ldots, N$. In particular let I denote the $N \times N$ identity matrix. We let $\mathbb{R}^N$ to denote the standard arithmetic space of N-tuples of real numbers. Let $[\cdot, \cdot]$ denote the standard inner product on $\mathbb{R}^N$

$$[x, y] = \sum_{j=1}^{N} x_j y_j, \quad x = (x_1, \ldots, x_N)^{\mathrm{T}}, \quad y = (y_1, \ldots, y_N)^{\mathrm{T}} \in \mathbb{R}^N.$$

We denote

$$\|x\| = \sum_{j=1}^{N} |x_j|, \quad x = (x_1, \ldots, x_N)^{\mathrm{T}} \in \mathbb{R}^N,$$

and

$$\|C\| = \max \left\{ \frac{\|Cx\|}{\|x\|} \ : \ x \in \mathbb{R}^N,\ x \neq 0 \right\},$$

for $C = (c_{jk}),\ c_{jk} \in \mathbb{R}^1,\ j, k = 1, \ldots, N$.

Definition 2.1. Let A be an $N \times N$ matrix. A pair of matrices $\{M, W\}$ is called a splitting of A if $A = M - W$ and M^{-1} exists. A splitting of matrix A is called of nonnegative type [20] if the matrix $T = M^{-1}W$ is nonnegative. If, in particular, the matrices M^{-1} and W are nonnegative, the splitting is called *regular* [28, p. 88]. If M^{-1} and $T = M^{-1}W$ are nonnegative, the splitting is called *weak regular* [24, p. 56]. A splitting $\{M, W\}$ is called *convergent* if $\lim_{k\to\infty} T^k$ exists and *zero-convergent*, if moreover $\lim_{k\to\infty} T^k = 0$.

Let Y denote an $N \times N$ matrix. We call a splitting $\{M, W\}$ *Y-convergent* or *Y-zero-convergent*, if the sequence $\{YT^k\}$ is convergent or zero-convergent, respectively.

Remark 2.2. The concept Y-convergent splitting reminds another concept *quotient convergence* introduced in [19]. Y-Convergence is very suitable in the context of IAD methods. The following example shows some flavour of the concept introduced.

Example 2.3. Let α, β be complex numbers, $|\alpha| > 1$.

$$T = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix},$$

and

$$Y = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

We see that T is Y-zero-convergent if $|\beta| < 1$ and Y-convergent if $\beta = 1$. Obviously, $\lim_{k\to\infty} T^k$ does not exist.

A collection of all distinct eigenvalues of a square matrix A is called the *spectrum* of A and it is denoted by $\sigma(A)$. We let

$$r(A) = \mathrm{Max}\{|\lambda| : \lambda \in \sigma(A)\}$$

and call it *spectral radius* of A.

3. Stationary probability vectors of stochastic matrices

We are going to consider the following:

Problem (P). Find a vector $x \in \mathbb{R}^N$ satisfying

$$x = Bx, \quad [x, e(N)] = 1, \tag{3.1}$$

under the restriction

$$\sum_{j=1}^{N} b_{jk} = 1, \quad \text{i.e. } B^{\mathrm{T}} e(N) = e(N), \tag{3.2}$$

where $e(N) = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^N$.

It follows from (3.1) that

$$r(B) = 1, \quad \deg_B(1) = 1,$$

where $\deg_B(\lambda)$ denotes the maximal size of the Jordan blocks corresponding to value λ. Any solution to (3.1) is called *stationary probability vector of* B.

Remark 3.1. It is well known that every irreducible stochastic $N \times N$ matrix B possesses a unique stationary probability vector. We are going to denote it by the symbol $\hat{x}$.

A matrix B satisfying (3.2) is called (*column*) *stochastic*. There is a permutation matrix H such that [12, p. 341]

$$HBH^{\mathrm{T}} = \begin{pmatrix} G_0 & 0 & . & . & . & 0 \\ G_1 & F_1 & . & . & . & 0 \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ G_p & 0 & . & . & . & F_p \end{pmatrix}, \tag{3.3}$$

where

$$\lim_{k\to\infty} G_0^k = 0$$

and F_j is an irreducible and stochastic matrix.

It is known [10] that matrix H can be obtained by the so-called *Tarjan algorithm* and that its complexity is almost linear.

If the permutation matrix H is available then one can compute the stationary probability vectors of each of matrices F_j, $j = 1, \ldots, p$, separately just executing the computations in parallel each F_j on its processor.

Proposition 3.2. *Let B be an irreducible $N \times N$ stochastic matrix, $\hat{x}$ its SPV, and $\{M, W\}$ a splitting of $A = I - B$ of nonnegative type. Then*

$$r(T) = 1, \quad \deg_T(1) = 1. \tag{3.4}$$

Proof. We see that

$$0 = M(I - T)\hat{x}$$

and thus,

$$\hat{x} = T\hat{x}$$

[2, Lemma 2.2] applies and this concludes the proof. □

4. Aggregation/disagregation algorithms

Let $\mathcal{G}$ be a map of $\{1, \ldots, N\}$ onto $\{1, \ldots, n\}$. Whenever it is necessary we distinguish the indices from the set $\{1, \ldots, n\}$ by bars from those belonging to the set $\{1, \ldots, N\}$; otherwise, we use a simplified notation. According to this aggreement we write, e.g. $x_{\mathrm{sub}(j)}$ instead of the more precise writing $x_{\mathrm{sub}(\overline{j})}$ if it is clear that $\overline{j}$ is fully determined by the associated j.

Let $\mathcal{K}$ denote the permutation of the set $\{1, \ldots, N\}$ given by the relations

$$\mathcal{K}(1) = j_1, \ldots, \mathcal{K}(N) = j_N$$

and denote the associated permutation matrix by K. We then have

$$(Kx)^{\mathrm{T}} = (x_{j_1}, \ldots, x_{j_{n_1}}, x_{j_{n_1+1}}, \ldots, x_{j_{n_1+n_2}}, \ldots, x_{j_N}).$$

It is then easy to see that

$$\{j \in \{1, \ldots, N\} : \mathcal{G}(j) = \overline{1}\} = \{j_1, \ldots, j_{n_1}\},$$

$$\{j \in \{1, \ldots, N\} : \mathcal{G}(j) = \overline{2}\} = \{j_{n_1+1}, \ldots, j_{n_1+n_2}\},$$

$$\vdots$$

$$\{j \in \{1, \ldots, N\} : \mathcal{G}(j) = \overline{n}\} = \{j_{n_1+\cdots+n_{p-1}+1}, \ldots, j_N\},$$

and

$$\{j \in \{1,\dots,N\} : \mathscr{G}_{\mathscr{K}}(j) = \overline{1}\} = \{1,\dots,n_1\},$$

$$\{j \in \{1,\dots,N\} : \mathscr{G}_{\mathscr{K}}(j) = \overline{2}\} = \{n_1+1,\dots,n_1+n_2\},$$

$$\vdots$$

$$\{j \in \{1,\dots,N\} : \mathscr{G}_{\mathscr{K}}(j) = \overline{n}\} = \{n_1+\cdots+n_{n-1}+1,\dots,N\}.$$

Denoting $\tilde{B} = K^{-1}BK$ we see that $\mathscr{G}$ yields an equivalent aggregation scheme for B as does $\mathscr{G}_{\mathscr{K}}$ for the permutation-similar matrix $\tilde{B}$.

We define communication operators R mapping space $\mathscr{E} = \mathbb{R}^N$ into $\mathscr{F} = \mathbb{R}^n$ and $S(x)$ mapping $\mathscr{F}$ into $\mathscr{E}$, respectively, by setting

$$\begin{aligned} &(Ru)_{\overline{j}} = \textstyle\sum_{\mathscr{G}(j)=\overline{j}} u_j, \quad u \in \mathbb{R}^N, \\ &Ku^{\mathrm{T}} = (u^{\mathrm{T}}_{\mathrm{sub}(1)},\dots,u^{\mathrm{T}}_{\mathrm{sub}(n)}), \quad u_{\mathrm{sub}(j)} \in \mathbb{R}^{n_j}, \end{aligned} \tag{4.1}$$

where

$$u_{\mathrm{sub}(1)} = (x_{j_1},\dots,x_{j_{n_1}})^{\mathrm{T}},\dots,u_{\mathrm{sub}(n)} = (x_{j_{n_1+\cdots+n_{n-1}}},\dots,x_{j_N})^{\mathrm{T}}$$

and

$$(S(x)z)_j = \frac{x_j}{(Rx)_{\overline{j}}} z_{\overline{j}}, \quad z \in \mathbb{R}^p, \quad z_{\overline{j}} \in \mathbb{R}^1, \quad \mathscr{G}(j) = \overline{j}, \quad j = 1,\dots,N. \tag{4.2}$$

for $x \in \mathscr{D}$, where

$$\mathscr{D} = \left\{ x \in \mathbb{R}^N \,:\, x^{\mathrm{T}} = (x_1,\dots,x_N),\ x_j > 0,\ j = 1,\dots,N \right\}.$$

We check immediately that

$$RS(x)z = z \quad \forall x \in \mathscr{D}, \quad z \in \mathbb{R}^p.$$

Therefore,

$$P(x) = S(x)R \tag{4.3}$$

is a projection called *aggregation projection*,

$$[P(x)]^2 = P(x).$$

Moreover,

$$S(x)^{\mathrm{T}} e(N) = e(p), \quad R^{\mathrm{T}} e(p) = e(N) \quad \forall x \in \mathscr{D}, \tag{4.4}$$

$$P(x)x = x \quad \forall x \in \mathscr{D} \tag{4.5}$$

and

$$P(x)^{\mathrm{T}}e = e \quad \forall x \in \mathscr{D}.$$

We define the matrix $\mathscr{B}(x) = RBS(x),\ x \in \mathscr{D}$, and call it *aggregated matrix* of B.

To guarantee that the proposed two-level algorithms can be unlimitedly realized we need the following two statements proven in [21].

Proposition 4.1 [21]. *Let matrix B be stochastic. Then its aggregated matrix $\mathscr{B}(x)$, $x \in \mathscr{D}$, is stochastic too.*

Proposition 4.2 [21]. *Let stochastic matrix B be irreducible. Then its aggregated matrix $\mathscr{B}(x),\ x \in \mathscr{D}$, is irreducible too.*

Remark 4.3. To choose a suitable map $\mathscr{G}$ defining the aggregation process is a rather difficult task. The reason is that one wants to obtain convergent and stable well conditioned IAD schemes. Some attempts to make construct of $\mathscr{G}$ determining an IAD processes possessing these properties for blockwise written matrices are presented in [4] as algorithms PABLO and TPABLO.

Algorithm 4.4 (**SPV**$(B; M, W; t, s; x^{(0)})$). Let B be an $N \times N$ irreducible stochastic matrix, let $\{M, W\}$ be a splitting of nonnegative type of the matrix $A = I - B$, $T = M^{-1}W$ and $s, t \geqslant 1$ positive integers.

Let $\epsilon > 0$ be a given tolerance and let $x^{(0)}$ with $(x^{(0)})_j > 0,\ j = 1, \ldots, N$, $[x^{(0)}, e(N)] = 1$, be an otherwise arbitrary vector.

Step 1. Set $0 \to k$.

Step 2. Construct the matrix

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

Step 3. Find the unique solution vector $z^{(k)}$ to the Problem (P) with $\mathscr{B}(x^{(k)})$, i.e. the unique solution to the problem

$$\mathscr{B}(x^{(k)})z^{(k)} = z^{(k)}, \tag{4.6}$$

$$[z^{(k)}, e(n)]_n = 1. \tag{4.7}$$

Step 4. Disaggregate by setting

$$v^{(k+1)} = S(x^{(k)})z^{(k)}.$$

Step 5. Let

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

and after these t smoothings a next approximant is defined as

$$x^{(k+1)} := x^{(k+1,t)}, \quad [x^{(k+1,t)}, e(N)]_N = 1.$$

Step 6. Test whether

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

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

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

and STOP.

Remark 4.5. Guaranteeing Algorithm 4.4 to be well defined requires the systems shown in Step 3 to be uniquely solvable. In connection with this requirement positive integer s must be chosen appropriately. The choice $s = 1$ is appropriate not only because it guarantees the irreducibility of the aggregated matrix $\mathscr{B}(x)$ (Proposition 4) but choosing $s = 1$ implies a consequent sparsity of $\mathscr{B}(x)$ inherited from B. On the other hand, as mentioned already, theoretically, any $s \geqslant 1$ is possible if the stochastic matrix examined is primitive. Recall that the latter is easy to achieve by considering $\tilde{B} = \frac{1}{2}(I + B)$ in place of B.

Remark 4.6. The **SPV** algorithm contains besides the transition matrix B five more variables offering a broad variety of realizations. We are going to show that at least theoretically, there are no restrictions for their choice. There are, however, some natural recommendations such as the choice $s = 1$ in case of sparse B and strict positivity of the zero-approximation $x^{(0)}$. On the other hand, the choice of t is a rather difficult task. In all our numerical tests the choice $t = 1$ never failed concerning convergence of the appropriate IAD scheme. Theoretically, both local as well as global convergence to be garanteed require t and alternatively s to be large enough.

The splitting $\{M, W\}$ is required to be of nonnegative type, otherwise arbitrary. Of course, in general one cannot expect fast convergence. On the other hand, $\{M, W\}$ need not be convergent and, as we show, the speed of convergence of the corresponding iteration matrix based on nonconvergent splitting may be very fast and for some classes of matrices even optimal.

In the literature various splittings $\{M, W\}$ of $I - B$ are recommended as bases for IAD methods such as Jacobi and Gauss–Seidel splittings. We mention them here in order to give a flavour of a form of some most frequently used splittings.

Let

$$B = D + L + U,$$

where D is the diagonal part of B, L the lower triangular part and U the upper triangular part, respectively. Set

$$M_V = I - D, \quad M_{KMS} = I - D - L.$$

Then the iteration matrices

$$T_V = (I - D)^{-1}[L + U], \quad T_{KMS} = (I - D - L)^{-1}U.$$

are well-known Jacobi and Gauss–Seidel iteration matrices. These two splittings form a basis for two IAD schemes. The first splitting was used for transition matrices in the blockwise written form by Vantilborgh [27] and the second similarly by Koury et al. [16]. A convergence theory of these methods together with the Takahashi method [26] are treated in the monograph [25]. Their convergence is proven there under the hypothesis that the transition matrix B is irreducible, it corresponds to a NCD Markov chain (for definition see [25, p. 285])) and furthermore as mentioned already, it satisfies four so called regularity conditions [25, p. 335]. The regularity conditions just mentioned explicitly exclude the case when the probability vectors possess components substantially smaller than one, i.e. the case of rare events is not allowed to happen in [25, p. 335]. More generally, it is not difficult to construct Markov chains with rare events without being NCD. We want to emphasize the fact that our theory is valid without any restrictions.

5. Convergence results

In the following sections we assume that the hypotheses made in introducing Algorithm 4.4 are satisfied. This concerns all the parameters of that algorithm, i.e. matrix B is irreducible and stochastic, splitting $\{M, W\}$ is of nonnegative type, $t \geqslant 1$ the number of relaxations of T is large enough, $s \geqslant 1$ the number of iterations of matrix B is arbitrary and the initial approximation vector $x^{(0)}$ is strictly positive.

For practical computations an important question is the choice of the parameteres t (the number of relaxations on the fine level) and s (the power of the investigated matrix). The choice $t = 1$ is interesting from the viewpoint of theory. However, this choice need not be preferable. This is explained by the fact that computing the aggregated matrix is costly and for $t = 1$ one must compute it after every relaxation on the fine level. The experiments reported in [21] show an "optimal choice" for t is $30 \leqslant t \leqslant 33$ if one requires to diminish the number of full IAD sweeps.

The very basic hypothesis required upon the splitting $\{M, W\}$ of the matrix $A = I - B$ in order to achieve convergence of the corresponding IAD process is its nonnegativity. Not surprisingly, the speed of convergence is generally slow. We show some splittings for which the speed of convergence of the corresponding IAD is fast. It is interesting to point out that iteration matrices deduced from splittings offering high speed of convergence may be divergent as solvers. In contrast, the speed of convergence of the corresponding IAD methods may be very fast and even

the exact solutions may be returned after a finite number of IAD iteration sweeps (Theorem 6.3).

By the symbol $I_{\mathscr{F}}$ we denote the identity map of the space $\mathscr{F} = \mathbb{R}^n$.

Lemma 5.1. *Let the hypotheses declared in the definition of Algorithm* 4.4 *be satisfied. Then the matrix* $I_{\mathscr{F}} - RZS(x^{(k)})$ *is invertible and*

$$\begin{aligned} S(x^{(k)})z^{(k)} &= S(x^{(k)})\left(I_{\mathscr{F}} - RZS(x^{(k)})\right)^{-1} R\hat{x} \\ &= \left(I - P(x^{(k)})Z\right)^{-1} P(x^{(k)})\hat{x}, \quad k = 0, 1, \ldots, \end{aligned} \tag{5.1}$$

with Z *coming from the spectral decomposition of* $B = Q + Z,\ Q^2 = Q,\ QZ = ZQ = 0, 1 \notin \sigma(Z)$.

Proof. Let λ be an eigenvalue of $RZS(x^{(k)})$ and z_λ a corresponding eigenvector. Writing $z_\lambda = Ry_\lambda \in \mathbb{R}^n$ with some $y_\lambda \in \mathbb{R}^N$ we see that

$$P(x^{(k)})ZP(x^{(k)})y_\lambda = \lambda P(x^{(k)})y_\lambda.$$

Since $\sigma\left(Z|_{P(x^{(k)})}\right) \subset \sigma(Z)$, where $Z|_{P(x^{(k)})}$ denotes the restriction of Z to the range of $P(x^{(k)})$, we conclude that $\lambda \in \sigma(Z|_{P(x^{(k)})})$ implies $\lambda \neq 1$ and invertibility of $I_{\mathscr{F}} - RZS(x^{(k)})$ follows.

From

$$0 = \left(I_{\mathscr{F}} - RBS(x^{(k)})\right) z^{(k)} = \left(I_{\mathscr{F}} - RZS(x^{(k)})\right) z^{(k)} - RQS(x^{(k)})z^{(k)},$$

it follows

$$z^{(k)} = \left(I_{\mathscr{F}} - RZS(x^{(k)})\right)^{-1} RQS(x^{(k)})z^{(k)}.$$

Since in our case $Qx = [x, e(N)]\hat{x},\ x \in \mathbb{R}^N$, the first formula in (4.4) implies

$$QS(x^{(k)})z^{(k)} = [S(x^{(k)})z^{(k)}, e(N)]\hat{x} = \hat{x}$$

and thus

$$z^{(k)} = \left(I_{\mathscr{F}} - RZS(x^{(k)})\right)^{-1} R\hat{x}.$$

Recalling that $I_{\mathscr{F}} = RS(x)$ and $P(x) = S(x)R$ for any $x \in \mathscr{D}$ we derive that

$$S(x)\left(I_{\mathscr{F}} - RZS(x)\right)^{-1} R\left(I - P(x)Z\right)P(x) = P(x)$$

and also

$$\left(I - P(x)Z\right)S(x)\left(I_{\mathscr{F}} - RZS(x)\right)^{-1} R = P(x).$$

Then it follows that the relation

$$S(x)\left(I_{\mathscr{F}} - RZS(x)\right)^{-1} RP(x)y = \left(I - P(x)Z\right)^{-1} P(x)y,$$

hold for all $y \in \text{range}(P(x))$, $x \in \mathscr{D}$. Since, in particular, $P(x)\hat{x} \in \text{range}(P(x))$, the proof is complete. □

Proposition 5.2. *Let B be any (column) stochastic matrix, $\{M, W\}$ a splitting of nonnegative type of $A = I - B$ and t, s any positive integers.*

Then the error-vector formula for the sequence of approximants $\{x^{(k)}\}$ returned by Algorithm 4.4 *reads*

$$x^{(k+1)} - \hat{x} = J_t(x^{(k)})(x^{(k)} - \hat{x}), \tag{5.2}$$

where

$$J_t(x) = T^t \left[I - P(x)Z\right]^{-1} \left(I - P(x)\right), \tag{5.3}$$

where Z comes, similarly as in Lemma 5.1 *from the spectral decomposition of $B = Q + Z$, $Q^2 = Q$, $QZ = ZQ = 0, 1 \notin \sigma(Z)$. Furthermore, $J_t(x) = T^{t-1} J_1(x)$, $t \geqslant 1$, holds for any x with all components positive.*

Proof. By definition of the **SPV**$(B; M, N; t, s; x^{(0)})$-algorithm,

$$Mx^{(k+1,1)} = WS(x^{(k)})z^{(k)}. \tag{5.4}$$

Lemma 5.1 together with the fact $Z\hat{x} = 0$ implies

$$\begin{aligned} Mx^{(k+1,1)} &= W(I - P(x^{(k)})Z)^{-1} P(x^{(k)})\hat{x} \\ &= W(I - P(x^{(k)})Z)^{-1} P(x^{(k)})(I - Z)\hat{x} \\ &= W(I - P(x^{(k)})Z)^{-1} (P(x^{(k)}) - I + I - P(x^{(k)})Z)\hat{x} \\ &= W\hat{x} - W(I - P(x^{(k)})Z)^{-1} (I - P(x^{(k)})\hat{x} \end{aligned}$$

and, since $M\hat{x} = W\hat{x}$ and $(I - P(x^{(k)}))x^{(k)} = 0$,

$$\begin{aligned} Mx^{(k+1,1)} - W\hat{x} &= M(x^{(k+1,1)} - \hat{x}) \\ &= W(I - P(x^{(k)})Z)^{-1} (I - P(x^{(k)}))(x^{(k)} - \hat{x}). \end{aligned}$$

Finally,

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

This is just formula (5.3) for $t = 1$. To obtain (5.3) for arbitrary $t \geqslant 1$ one needs to apply T^{t-1} to $J_1(x^{(k)})$. It is obvious that that algorithm **SPV**$(B; M, W : t, s; x^{(0)})$ achieves this purpose by applying the iteration procedure determined by the splitting $\{M, W\}$. □

Proposition 5.3. *Under the assumptions of Proposition* 5.2 *the spectra of $J_t(\hat{x})$ and $(I - P(\hat{x}))J_t(\hat{x})$ are related as follows*:

$$\sigma\left(J_t(\hat{x})\right) \subset \sigma\left(\left(I - P(\hat{x})\right) J_t(\hat{x})\right) \cup \{0\}.$$

Consequently, $r(J_t(\hat{x})) = r((I - P(\hat{x}))J_t(\hat{x}))$.

Proof. Let $0 \neq \lambda \in \sigma(J_t(\hat{x}))$ and w a corresponding eigenvector $J_t(\hat{x})w = \lambda w,\ w \neq 0$. According to the definition of $J_t(x)$ we see that

$$\left(I - P(\hat{x})\right) J_t(\hat{x}) \left(I - P(\hat{x})\right) w = \lambda \left(I - P(\hat{x})\right) w. \qquad \square$$

The following definition is a crucial concept in our convergence theory and Theorem 5.5 is one of our main results. Note that convergence of the well-known algorithms of Koury et al. [16] and Vantilborgh [27] is proven so far only for the case of NCD Markov chains under the assumption of validity of the regularity conditions [25, p. 335].

Definition 5.4. A splitting $\{M, W\}$ of $I - B = M - W$ is called *aggregation-convergent* if $T = M^{-1}W$ is Y-zero-convergent, where $Y = I - P(\hat{x})$ and $P(\hat{x})$ is the aggregation projection (4.3).

Theorem 5.5. *Let B be an $N \times N$ irreducible column stochastic matrix. Let $\{M, W\}$ be an aggregation-convergent splitting of nonnegative type of matrix $I - B$ and let $s = 1$. Denote the iteration matrix corresponding to this splitting by T, i.e. $T = M^{-1}W$.*

Then there is a neighborhood $\Omega(\hat{x})$ and a positive integer $t \geqslant 1$ such that **SPV** *$(B; M, W; t, s; x^{(0)})$ Algoritm* 4.4 *is convergent whenever $x^{(0)} \in \Omega(\hat{x})$. Moreover, the following error estimate holds*

$$\left\|x^{(k)} - \hat{x}\right\| \leqslant \kappa \rho^k \left\|x^{(0)} - \hat{x}\right\|,$$

where $\|\cdot\|$ denotes any norm on $\mathbb{R}^N$ and κ and $\rho < 1$ are positive real numbers independent of $k = 0, 1, \ldots$

Proof. The hypotheses of Theorem 5.5 imply that for every $y \in \mathbb{R}^N$

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

Let $\delta > 0$ be small enough. According to (5.3) there exist a $\hat{t} \geqslant 1$ and positive real $\tilde{\rho}$ such that

$$\|J_t(\hat{x})\| \leqslant \tilde{\rho} + \delta = \rho < 1, \quad t \geqslant \hat{t}.$$

Continuity of the projection-function $P = P(x)$ with respect to $x \in \mathrm{Int}(\mathbb{R}^N_+)$ allows us to extend the above estimate to a neighborhood $\Omega(\hat{x})$ and this completes the proof. $\square$

Definition 5.6. Let $\{M, W\}$ be a splitting of $N \times N$ matrix $A = I - B$ and $T = M^{-1}W$ be the corresponding iteration matrix. Let T be cyclic with index of cyclicity n and let

$$HTH^{\mathrm{T}} = \begin{pmatrix} 0 & T_{12} & 0 & . & . & . & 0 & 0 \\ 0 & 0 & T_{23} & . & . & . & 0 & 0 \\ . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & 0 & T_{n-1n} \\ T_{n1} & 0 & 0 & . & . & . & 0 & 0 \end{pmatrix} \tag{5.5}$$

with square diagonal blocks of sizes $d_j \times d_j$, $j = 1, \ldots, n$ and with a suitable permutation matrix H be a canonical form of T (see [12, p. 324]). Let $\mathscr{H}$ denote the permutation of $\{1, \ldots, N\}$ corresponding to the permutation H in the block form (5.5), i.e.

$$Hx = \left(x_{\mathscr{H}(1)}, \ldots, x_{\mathscr{H}(n_1)}, \ldots, x_{\mathscr{H}(l_N)}\right)^{\mathrm{T}}$$

and

$$\mathscr{H}(n_1 + \cdots + n_{k-1} + j) = l_{n_1+\cdots+n_{k-1}+j},$$
$$n_1 + \cdots + n_{k-1} + 1 \leqslant j \leqslant n_1 + \cdots + n_k, \quad k = 1, \ldots, n, \quad n_0 = 0.$$

We call a splitting $\{M, W\}\mathscr{G}$-*consistent* if

$$d_k = n_k, \quad k = 1, \ldots, n,$$

where $n_k = \mathrm{card}\{j : \mathscr{G}(j) = \overline{k}\}$ and

$$(Rx)_{\overline{k}} = \sum_{j=1}^{n_k} (Hx)_{n_1+\cdots+n_{k-1}+j}, \quad k = 1, \ldots, n, \quad n_0 = 0.$$

Remark 5.7. Definition of $\mathscr{G}$-consistency of a splitting $\{M, W\}$ implies that the iteration matrix $T = M^{-1}W$ is such that every state belonging to the same $\mathscr{G}$-aggregate of B, belongs to the same cyclic class of T. This is a condition formulated in [5] guaranteeing convergence of the subvectors $x^{(k)}_{\mathrm{sub}(\overline{j})}$ and $y^{(k)}_{\mathrm{sub}(\overline{j})}$ of the iteration sequences associated with the block form of the investigated stochastic matrix B determined by $\mathscr{G}$.

Theorem 5.8. *Any $\mathscr{G}$-consistent splitting of nonegative type $\{M, W\}$ of $A = I - B$, where B is an irreducible column stochastic matrix, is aggregation-convergent.*

Proof. Let $T = M^{-1}W$ be the iteration matrix of the splitting under consideration. Further, let y be an eigenvector of T corresponding to an eigenvalue λ, $|\lambda| = 1$, belonging to the spectrum of T. Our hypotheses allow us to show that

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

The validity of this claim follows from a known result of Courtois and Semal [5] according to which the subvectors of $\hat{x}$ and y corresponding to the block form of T implied by $\mathcal{G}$ are parallel. It follows from Proposition 5.3 that

$$J_t(\hat{x})y = 0.$$

Since the above property concerns eigenvectors corresponding to any eigenvalue of absolute value one, we conclude $J_t(\hat{x})$ is zero-convergent. □

Remark 5.9. Let B be written in a block form. The hypotheses of Theorem 5.8 apply in case of the splittings leading to the IAD methods KMS ([17], [25, p. 308]), V ([25, p. 316], [27]) and MM [21] whenever the aggregation map $\mathcal{G}$ is chosen such that each block of B is aggregated to a 1×1 matrix.

Remark 5.10. The above results concern the local convergence. The global convergence is obtained by taking either of the parameters t and s sufficiently large in a manner described in [21, Section 4.1]. Thus under the additional hypothesis requiring B to be primitive and the same hypotheses of Theorem 5.5 concerning matrix B, and splitting $\{M, W\}$ Algorithm 4.4 converges globally to the unique stationary probability vector of B and the speed of convergence is identical with the speed shown in theorems of this section.

6. Fast IAD methods

Proposition 6.1. *Let $\alpha_{\overline{1}}, \ldots, \alpha_{\overline{n}}$ be positive real numbers. Let*

$$x_{\mathrm{sub}(\overline{j})} = \alpha_{\overline{j}} y_{\mathrm{sub}(\overline{j})}, \quad \overline{j} = 1, \ldots, n, \tag{6.1}$$

where

$$x^{\mathrm{T}} = \left(x^{\mathrm{T}}_{\mathrm{sub}(\overline{1})}, \ldots, x^{\mathrm{T}}_{\mathrm{sub}(\overline{n})}\right), \quad y = \left(y^{\mathrm{T}}_{\mathrm{sub}(\overline{1})}, \ldots, y^{\mathrm{T}}_{\mathrm{sub}(\overline{n})}\right)$$

are blockwise written vectors x, y accordingly to the block form of B given by the map $\mathcal{G}$.

Then

$$P(x)y = y. \tag{6.2}$$

Proof. We see that

$$(P(x)y)_j = (S(x)Ry)_j = \frac{x_j}{(Rx)_{\overline{j}}}\,(Ry)_{\overline{j}}$$

and further, because of (6.1),

$$(P(x)y)_j = y_j, \quad j \in \left\{j : \mathcal{G}(j) = \overline{j}\right\}, \quad \overline{j} = 1, \ldots, n.$$

Consequently, (6.2). □

Corollary 6.2. *If* (6.1) *holds for some* $x^{(k)}$*—an element returned by the Algorithm* **SPV***—in place of* x *and* $\hat{x}$*—the unique stationary probability vector of* B*—in place of* y*, then*

$$\big(I - P(x^{(k)})\big)(x^{(k)} - \hat{x}) = 0$$

and, by (5.2)

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

Proof. The error vector formula implies

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

according to Proposition 6.1, we have

$$\begin{aligned}\big(I - P(x^{(k)})\big)(x^{(k)} - \hat{x}) &= (x^{(k)} - \hat{x}) - P(x^{(k)})(x^{(k)} - \hat{x}) \\ &= -\hat{x} + P(x^{(k)})\hat{x} = 0.\end{aligned}$$

Thus, by (5.2) and (5.3), $x^{(k+1)} - \hat{x} = 0$. □

Theorem 6.3. *Let* B *be an* $N \times N$ *irreducible stochastic matrix such that in the block form*

$$B = \begin{pmatrix} B_{11} & B_{12} & . & . & . & B_{1n} \\ B_{21} & B_{22} & . & . & . & B_{2n} \\ . & . & . & . & . & . \\ B_{n1} & B_{n2} & . & . & . & B_{nn} \end{pmatrix} \tag{6.3}$$

the off-diagonal blocks satisfy relations $R_{\overline{j}} = \mathrm{range}(B_{jk}) = \{\alpha f_{\mathrm{sub}(\overline{j})} : \alpha \in \mathscr{R}^1\}$, $\mathscr{G}(j) = \overline{j}$*, where* $f_{\mathrm{sub}(\overline{j})}$ *is a vector in* $\mathbb{R}^{n_j}$*,* n_j *being the size of the block* B_{jj}*. Let* $\{M, W\}$ *be a splitting of* $A = I - B$ *of nonegative type such that* $W = M - A = (W_{jk})$, $W_{jj} = 0$.

Then for $s = 1$*, any* $t \geqslant 1$ *and any initial approximation* $x^{(0)} \in \{x \in \mathbb{R}^N : (x)_j > 0,\ j = 1, 2, \ldots, N\}$*, we have* $x^{(2)} = \hat{x},\ B\hat{x} = \hat{x},\ [\hat{x}, e(N)] = 1$*, i.e. the exact solution is returned after two SPV sweeps.*

Proof. Let $x^{(0)}$ be an initial approximation of $\hat{x}$. By Proposition 6.1 we have

$$(x^{(1)})_{\mathrm{sub}(\overline{j})} \in R_{\overline{j}}, \quad \overline{j} = 1, \ldots, n,$$

and followingly,

$$(x^{(2)})_{\mathrm{sub}(\overline{j})} \in R_{\overline{j}}, \quad \overline{j} = 1, \ldots, n.$$

Then, by hypothesis,

$$(\hat{x})_{\mathrm{sub}(\overline{j})} = \alpha_{\overline{j}}(x^{(1)})_{\mathrm{sub}(\overline{j})},$$

with some positive reals $\alpha_{\overline{1}}, \ldots, \alpha_{\overline{n}}$. Utilizing Corollary 6.2 we deduce $x^{(2)} - \hat{x} = 0$ and this completes the proof. □

Remark 6.4. Though stochastic matrices belonging to the class of matrices (let us denote it $\mathscr{DYAD}$) with the properties described in the previous theorem may look very academic they find applications in practice of designing equipments in the railway safety systems. Such matrices possess some interesting spectral properties too. It is easy to see that the spectrum of some elements of the class just discussed may be quite complex. To get a feeling of such complexity, let us consider a class of dyadic stochastic $N \times N$ matrices of the form

$$B(\varepsilon) = \begin{pmatrix} (1-\varepsilon)B_{11} & & & & \\ & (1-\varepsilon)B_{22} & & & \\ & & \ddots & & \\ & & & & (1-\varepsilon)B_{nn} \end{pmatrix} + \begin{pmatrix} 0 & \varepsilon B_{12} & \cdots & \varepsilon B_{1n} \\ \varepsilon B_{21} & 0 & \cdots & \varepsilon B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon B_{n1} & \varepsilon B_{n2} & \cdots & 0 \end{pmatrix},$$

where B belongs to $\mathscr{DYAD}$ and $\varepsilon \in [0, 1]$ is otherwise arbitrary.

It is well known that the spectra of the diagonal blocks $B_{11}, \ldots, B_{nn}$ belong to the so called Karpelevich sets [14], [12, p. 350]. Since these matrices are arbitrary substochastic, i.e. their column sums do not exceed value one, we conclude that the spectra of $B(\varepsilon)$ may be complex, in particular if the sizes of some of their blocks become large and ε small.

Remark 6.5. Note that the iteration processes defined by the splittings described in Theorem 6.3 are in general divergent, e.g. the Jacobi method in case of a p-cyclic matrix. Nevertheless, the convergence of the **SPV** algorithm based on such splittings is fast and terminates after at most two IAD sweeps.

Remark 6.6. To check that a given block matrix satisfies the conditions of Theorem 6.3 might be difficult. This is because the problem to determine the rank of a matrix is not well posed. Fortunately, we have the following modification of Theorem 6.3.

Theorem 6.7. *Let B be such that every column of each of the blocks $B_{jk},\ j \neq k$, have the form $(1-\tau)\alpha_{\bar{j}} f_{\mathrm{sub}(\bar{j})} + \tau g_{\mathrm{sub}(\bar{j})}$, where $\|g_{\mathrm{sub}(\bar{j})}\| < \tau\alpha$, where α and τ are positive real numbers, $\tau \leqslant 1$. Furthermore, let $\{M, W\}$ be a splitting of $A = I - B$ of nonnegative type such that $W = M - A = (W_{jk})$, $W_{jj} = 0$.*

Then there is a $\tau_0 > 0$ such that the rate of convergence of the IAD algorithm **SPV** *is bounded above by the product $\tau \mathrm{Max}\{\kappa_j : j = 1, \ldots, n,\}$ for all $\tau \leqslant \tau_0$, where $\kappa_j = \mathrm{cond}(I_{n_j} - B_{jj})^{-1}$.*

In the literature there are known some algorithms such as PABLO and TPABLO [4] that transform the original matrix into its block-equivalent permutation-similar form with the goal to condition the diagonal blocks optimally. Our analysis shows that it is desirable simultaneously to force the off-diagonal blocks to be as close as possible to rank-one blocks having ranges of the row blocks the same. Obviously, these two requirements are in conflict with each other. Hence, a reasonable compromise criterion needs to be found.

7. p-Cyclic stochastic matrices

Let N, p be positive integers, $n_j,\ j = 1, \ldots, p$, such that B is a (column) stochastic matrix. Furthermore, let

$$HBH^{\mathrm{T}} = \begin{pmatrix} B_{11} & 0 & . & . & . & 0 & B_{1p} \\ B_{21} & B_{22} & . & . & . & 0 & 0 \\ & & . & & & & \\ & & & . & & & \\ & & & & . & & \\ 0 & 0 & . & . & . & B_{p-1p-1} & 0 \\ 0 & 0 & . & . & . & B_{pp-1} & B_{pp} \end{pmatrix}, \tag{7.1}$$

where H is a permutation matrix and the blocks B_{jk} are $n_j \times n_k$ matrices such that $I - B_{jj},\ j, = 1, \ldots, p$, are invertible.

Our permanent hypothesis $B^{\mathrm{T}} e(N) = e(N)$ implies that $[HBH^{\mathrm{T}}]^{\mathrm{T}} e(N) = e(N)$ and also

$$\begin{aligned} &\left(I_{n_j} - B_{jj}^{\mathrm{T}}\right)^{-1} B_{j+1\,j}^{\mathrm{T}} e(n_{j+1}) = e(n_j), \quad j = 1, \ldots, p-1, \\ &\left(I_{n_p} - B_{pp}^{\mathrm{T}}\right)^{-1} B_{1p}^{\mathrm{T}} e(n_1) = e(n_p), \end{aligned} \tag{7.2}$$

where $e(n_j)^{\mathrm{T}} = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^{n_j}$. Evidently,

$$e = e(N) = (1, \ldots, 1)^{\mathrm{T}} = \left(e(n_1)^{\mathrm{T}}, \ldots, e(n_p)^{\mathrm{T}}\right)^{\mathrm{T}} \in \mathbb{R}^N.$$

A block matrix $A = I - B$ satisfying relations (7.2) is called weakly p-cyclic stochastic.

Definition 7.1. Let N and p be positive integers. Let $\{M, W\}$ be a splitting of nonnegative type of $A = I - B$, where B is a p-cyclic stochastic matrix. If

$$M^{-1}W = T = \sum_{l=1}^{p} \lambda_l Q_l + Z,\ Q_l Q_k = Q_k Q_l = \delta_{lk} Q_l,$$

$$Q_l Z = Z Q_l = 0, \quad r(Z) < r(T) = |\lambda_l|, \quad l, k = 1, \ldots, p,$$

where δ_{lk} denotes the Kronecker "delta" and the limits denoted by symbols $w^{\infty}_{\mathrm{sub}(j)}$,

$$\lim_{k\to\infty} \left(\left[\frac{1}{\lambda_l} T \right]^k w \right)_{\mathrm{sub}(\overline{j})} = w^{\infty}_{\mathrm{sub}(\overline{j})}, \quad l, \overline{j} = 1, \ldots, p, \tag{7.3}$$

exist whenever $\sum_{l=1}^{p} Q_l w \neq 0$, where the blocks in $w^{\mathrm{T}} = (w^{\mathrm{T}}_{\mathrm{sub}(\overline{1})}, \ldots, w^{\mathrm{T}}_{\mathrm{sub}(\overline{p})})$ are consistent with the form of B in (7.1), we call such a splitting *p-shape convergent*.

Remark 7.2. In the context of IAD methods the subvectors of a given vector determined by the aggregation map $\mathscr{G}$ characterize the "shapes" of the desired solution up to constant multiples. Thus, we accepted this property as decisive for the name of the concept introduced in Definition 7.1.

Remark 7.3. It is obvious that every convergent splitting is automatically p-shape convergent.

Example 7.4. According to [5] as examples of p-shape convergent splittings one can take the block Jacobi and block Gauss–Seidel iteration schemes with the blocks appropriately associated.

As a consequence of the previous results we can formulate our next result.

Theorem 7.5. *Let $\{M, W\}$ be a p-shape convergent splitting of nonnegative type of $A = I - B$, where B is an irreducible block matrix such that $(B - \mathrm{diag}B)$ is blockwise p-cyclic.*

Let $\lambda_1 = r(T)$, $Q_1 v^{(0)} \neq 0$ and

$$Mv^{(k+1)} = Wv^{(k)}, \quad k = 0, 1, \ldots \tag{7.4}$$

To apply **SPV** *algorithm let us choose $t = 1$ and $s = 1$ and let*

$$\overline{x}^{\mathrm{T}} = \left(\left[\lim_{k\to\infty} v^{(k)}_{\mathrm{sub}(\overline{1})} \right]^{\mathrm{T}}, \ldots, \left[\lim_{k\to\infty} v^{(k)}_{\mathrm{sub}(\overline{p})} \right]^{\mathrm{T}} \right).$$

Then **SPV**$(B; M, W; t = 1, s = 1; \overline{x})$*-algorithm returns the exact solution after at most two IAD sweeps. In other words,*

$$x^{(2)}(\overline{x}) = \hat{x},$$

where $x^{(1)}(\overline{x})$ *and* $x^{(2)}(\overline{x})$ *denote the first and second returns of the* **SPV**$(B; M, W; t = 1, s = 1; \overline{x})$*-algorithm, respectively.*

Proof. The p-shape convergence of T implies that the appropriate subvectors of (7.3) converge to the subvectors of the exact stationary probability vector $\hat{x}$. Therefore, at most two IAD iteration sweeps described in the theorem applied to the limiting subvectors return the exact $\hat{x}$. □

Remark 7.6. Theorem 7.5 requires the knowledge of the blocks in the canonical form 7.1. This is a standard requirement in working with p-cyclic matrices however.

Corollary 7.7. *Let in addition to hypotheses of Theorem* 7.5 *at least one of the off-diagonal blocks be formed by a tensor product of two vectors. Moreover, let the structure of the splitting* $\{M, W\}$ *be as shown in Theorem* 6.7.

Then the iterative process described in this theorem terminates after at most two iterations (7.4) *returning the exact solution* $\hat{x}$ *after one* **SVP** *sweep.*

Proof. The hypotheses imply that all the blocks of T^p are rank one matrices with the structure described in Theorem 6.3, the conclusions follow. □

8. A remark concerning the error analysis

As mentioned already our aim is to compute SPV with a high degree of precision. Since IAD methods are two-level procedures there are two sources of error: On the fine as well on the aggregated level. In this section we explain why the computations on the aggregated level have to be performed very precisely in order to achieve a final goal on the fine level.

To this purpose a rather exotic measure of error appears as appropriate. Actually, the measures $\rho_{\mathscr{E}}$ and $\rho_{\mathscr{F}}$ introduced below are distance functions on the sets $\mathrm{Int}\left(\mathbb{R}^N_+\right)$ and $\mathrm{Int}\left(\mathbb{R}^p_+\right)$, respectively.

A natural question arises when analyzing Algorithm 4.4 similarly as for a particular case [21]. How is the error on the fine level influenced by the error on the aggregatedr level? The solution is identical with that in [21].

To answer the above question, let $z^* = R\hat{x},\ \hat{x} = T\hat{x}$ and let us define

$$\rho_{\mathscr{F}}(z^{(k)}, z^*) = \min\left\{\sum_{\overline{j}=1}^{n}\left|\frac{(z^{(k)})_{\overline{j}}}{(z^*)_{\overline{j}}} - \mu\right| : \mu \in \mathbb{R}^1_+\right\}, \tag{8.1}$$

and

$$\rho_{\mathscr{E}}(S(\hat{x})z^{(k)}, \hat{x}) = \min\left\{\sum_{j=1}^{N}\left|\frac{(S(\hat{x})z^{(k)})_j}{(\hat{x})_j} - \mu\right| : \mu \in \mathbb{R}^1\right\}. \tag{8.2}$$

Then obviously,

$$\frac{(S(\hat{x})z^{(k)})_j}{(\hat{x})_j} = \frac{\left(S(\hat{x})z^{(k)}\right)_j}{\left(S(\hat{x})z^{*}\right)_j}, \quad z^* = R\hat{x}.$$

However,

$$\left(1/(\hat{x})_j\right)\left(S(\hat{x})z^{(k)}\right)_j = \left(1/(\hat{x})_{j_0}\right)\left(S(\hat{x})z^{(k)}\right)_{j_0}$$
$$\text{for all } j : G(j) = \overline{j}, \quad \overline{j} = 1, \ldots, n.$$

Therefore,

$$\begin{aligned}\rho_{\mathscr{E}}(S(\hat{x})z^{(k)}, \hat{x}) &= \sum\nolimits_{\overline{j}=1}^{n}\sum\nolimits_{G(j)=\overline{j}}\rho_{\mathscr{F}}(z^{(k)}, z^*)\\ &= \sum\nolimits_{\overline{j}=1}^{n} d_{\overline{j}}\rho_{\mathscr{F}}(z^{(k)}, z^*),\end{aligned}$$

where

$$d_{\overline{j}} = \operatorname{card}\left\{j : G(j) = \overline{j}\right\}. \tag{8.3}$$

Finally,

$$\underline{d}\rho_{\mathscr{F}}(z^{(k)}, z^*) \leqslant \rho_{\mathscr{E}}((S(\hat{x}))z^{(k)}, \hat{x}) \leqslant \overline{d}\rho_{\mathscr{F}}(z^{(k)}, z^*), \tag{8.4}$$

where

$$\underline{d} = \min d_{\overline{j}}, \overline{d} = \max d_{\overline{j}}. \tag{8.5}$$

We collect our deductions in the following statement.

Proposition 8.1. *Let sequence* $\{x^{(k)}\}$ *be computed according to Algorithm* 4.4.
Then the characteristics of the error estimates (8.2) *and* (8.1) *are related by* (8.4) *with* $d_{\overline{j}}$, $\overline{j} = 1, \ldots, n$, $\underline{d}$ *and* $\overline{d}$ *given by* (8.3) *and* (8.5), *respectively*.

A special feature of IAD methods is a relationship between the error of the kth approximation of the exact solution $\hat{x}$ and the corresponding error of its aggregeted version $z^{(k)} - R\hat{x}$ on the aggregated level is described in Proposition 8.1. It says that if one chooses all the segments aggregated into one component to have the same dimension d then $\|x^{(k)} - \hat{x}\| = d\|z^{(k)} - R\hat{x}\|$. This fact has a serious impact on the precision of computations on the aggregated level. One must compute on the aggregated level very precisely in order to achieve a required accuracy on the fine level. This concerns in particular the cases when d is large. If, say, $d = 10^4$, then

$$E_f(x^{(k)}) = \|x^{(k)} - \hat{x}\| = 10^4 E_c, \quad E_c(z^{(k)}) = \|z^{(k)} - R\hat{x}\|$$

and thus, in order to achieve $E_f(x^{(k)}) < 10^{-q}$, $E_c(z^{(k)})$ must be smaller than $10^{-(q+4)}$. In our reliable safety modeling computations one requires $q = 12$ and thus, one must use a multiple precision on both levels.

Obviously, if the aggregates are general positive integer d is replaced by a pair of positive integers $\underline{d}$ and $\overline{d}$.

9. Concluding remarks

As mentioned already our work is motivated by practical needs of railway transportation systems. Project of the European Commission [11] requires an interdisciplinary cooperation of many specialists in most diverse areas of research in order to built up a reliable and safe transportation system [13]. The requirements are indeed major: e.g. the expected amount of input data is of order 10^7 or even larger. Some of the requirements are such that checking whether they are satisfied is not possible using standard means of testing. This is because any existing device is not able to outlive as many tests as needed in order to accomplish a sufficient number of experiments that would guarantee a desired smallness of the probability of their appearance. Some samples of our academic tests have been published in [21,22]. The results of our railway safety applications will be published elsewhere.

Acknowledgements

Research has been partly supported by the Grant Agency of the Czech Republic under the contract No. 201/02/0595 and by Grants CEZ J04/98:210000010 (IM) and grant No. MSM 113200007 (PM).

The paper has been prepared as a contribution to the workshop devoted to "M-matrices and their generalizations" held in Oberwolfach within the period of November 26–December 3, 2000. A kind invitation to participate, support of the stay, excellent facilities available and a nice working atmosphere are sincerely acknowledged.

The authors are indebted to two anonymous referees for their valuable comments and recommendations. Their suggestions and help substantially exceeding standard refereeing procedures are gratefully appreciated.

References

[1] A. Berman, R.J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, Academic Press, New York, San Francisco, London, 1979.

[2] E. Bohl, I. Marek, A model of amplification, J. Comput. Appl. Math. 63 (1995) 27–47.

[3] W.L. Cao, W.J. Stewart, Iterative asggregation/disaggregation techniques for nearly uncoupled Markov chains, J. Assoc. Comput. Mach. 32 (3) (1985) 702–719.

[4] H. Choi, D. Szyld, Application of threshold partitioning of sparse matrices to Markov chains, in: Proceedings of the IEEE International Computer Performance and Dependability Symposium IPDS'96, Urbana, IL, 1996, pp. 158–165.
[5] P.J. Courtois, P. Semal, Block iterative algorithms for stochastic matrices, Linear Algebra Appl. 76 (1986) 59–80.
[6] T. Dayar, Permuting Markov chains to nearly completely decomposable form, Technical Report BU-CEIS-9808, Department of Computer Engineering and Information Science, Bilkent University, Ankara, Turkey, August 1998. Available from: ftp://ftp.cs.bilkent.edu.tr/pub/tech-reports/1998/BU-CEIS-9808.ps.z.
[7] T. Dayar, W.J. Stewart, Comparison of partitioning techniques for two-level iterative solvers on large, sparse Markov chains, Technical Report BU-CEIS-9805, Department of Computer Engineering and Information Science, Bilkent University, Ankara, Turkey, April 1998. Available from: ftp://ftp.cs.bilkent.edu.tr/pub/tech-reports/1998/BU-CEIS- 9805.ps.z.
[8] T. Dayar, W.J. Stewart, Comparison of partitioning techniques for two-level iterative solvers on large, sparse Markov chains, SIAM J. Sci. Comput. 21 (2000) 1691–1705.
[9] T. Dayar, W.J. Stewart, On the effects of using the Grassman–Taksar–Heyman method in iterative aggregation–disaggregation, SIAM J. Sci. Comput. 19 (1996) 287–303.
[10] I.S. Duff, J.K. Reid, An implementation of Tarjan's algorithm for the block triangularization of a matrix, ACM Trans. Math. Software 4 (1978) 137–147.
[11] *European Standards:*
EN 50 126: Railway application: The specification and demonstration of dependability, reliability, availability, maintainability and safety (RAMS).
EN 50 128: Railway application: Software for railway control systems and protection systems.
EN 50 129: Railway application: Safety related electronic systems.
EN 50 159-1: Railway application: Communication, signalling and processing systems.
Part 1: Safety-related communication in closed transmission systems.
EN 50 159-2: Railway application: Communication, signalling and processing systems.
Part 2: Safety-related communication in open transmission systems, European Commission CENELEC, Bruxelles 1993.
[12] F.R. Gantmacher, The Theory of Matrices, Gos. Izd. Lit. Moscow, 1954 (in Russian); English translation; Applications of the Theory of Matrices, Interscience, New York, 1959.
[13] T. Hlavatý, Š. Klapka, L. Přeučil, P. Štěpán, Formal methods in development and testing of railway interlocking systems, in: Proceedings of the International Conference Railway Traction Systems, vol. 1, Session 2: Modelling, Capri, Italy, May 2001, pp. 15–17.
[14] F.I. Karpelevich, On characteristic roots of matrices with nonnegative elements, Izvestia Matem. Nauk Ser. Matem. 15 (1951) 361–383 (in Russian).
[15] K. Kontovasilis, R.J. Plemmons, W.J. Stewart, Block cyclic SOR for Markov chains with p-cyclic infinitesimal generator, Linear Algebra Appl. 154–156 (1991) 145–223.
[16] R. Koury, D.F. McAllister, W.J. Stewart, Methods for computing stationary distributions of nearly completely decomposable Markov chains, SIAM J. Alg. Disc. Meth. 5 (2) (1984) 164–186.
[17] R. Koury, D.F. McAllister, W.J. Stewart, On the Raleigh–Ritz refinement technique for nearly uncoupled stochastic matrices, Linear Algebra Appl. 60 (1984) 1–25.
[18] U. Krieger, On a two-level multigrid solution method for finite Markov chains, Linear Algebra Appl. 223/224 (1995) 415–438.
[19] G.I. Marchuk, Yu.A. Kuznetsov, Iterative Methods and Quadratic Functionals, Nauka, Moscow, 1972 (in Russian).
[20] I. Marek, Frobenius theory of positive operators. Comparison theorems and applications, SIAM J. Appl. Math. 19 (1970) 607–628.
[21] I. Marek, P. Mayer, Convergence analysis of an aggregation/disaggregation iterative method for computation stationary probability vectors of stochastic matrices, Numer. Linear Algebra Appl. 5 (1998) 253–274.

[22] I. Marek, P. Mayer, Aggregation/disaggregation methods for p-cyclic Markov chains, Computing 15 (Suppl.) (2001) 133–150.
[23] I. Marek, D. Szyld, Splittings of M-operators, irreducibility and the index of the iteration operators, Numer. Functional Anal. Optim. 11 (1990) 529–553.
[24] J. Ortega, W. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.
[25] W.J. Stewart, Introduction to the Numerical Solution of Markov Chains, Princeton University Press, Princeton, NJ, 1994.
[26] Y. Takahashi, A lumping method for numerical calculations of stationary distributions of Markov chains. Res. Rep. B-18, Department of Information Science, Tokyo Institute of Technology, Tokyo, Japan, June 1975.
[27] H. Vantilborgh, The error aggregation. A contribution to the theory of decomposable systems and applications. PhD Thesis, Faculty of Applied Sciences, Louvain Catholic University, Louvain-la Neuve, Belgium, 1981.
[28] R.S. Varga, Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1962.