Local convergence analysis of iterative aggregation–disaggregation methods with polynomial correction

Ivana Pultarová *

Department of Mathematics, Faculty of Civil Engineering, Czech Technical University in Prague, Thákurova 7, 166 29 Prague, Czech Republic

Received 28 October 2005; accepted 4 September 2006
Available online 25 October 2006
Submitted by S. Kirkland

Abstract

The paper introduces some new results on local convergence analysis of one class of iterative aggregation–disaggregation methods for computing a stationary probability distribution vector of an irreducible stochastic matrix. We focus on methods, where the basic iteration on the fine level corresponds to a multiplication by a polynomial of order one with nonnegative coefficients in the original matrix. We show that this process is locally convergent for matrices with positive diagonals or when the coefficients of the polynomial are positive. On the other hand there are examples for which the process may diverge in a local sense for higher degree polynomials even if it converges for a polynomial of a lower degree for the same matrix.

AMS classification: 65F10; 15A51; 60J10

Keywords: Stochastic matrix; Markov chains; Stationary probability distribution vector; Iterative aggregation–disaggregation method

1. Introduction

This paper can be viewed as a continuation of the work which can be found in the recently published paper [8]. We present several new conclusions on convergence analysis of a two-
level iterative algorithm, iterative aggregation–disaggregation (IAD) method, for computing a stationary probability distribution vector of a finite discrete Markov chain which is represented by an irreducible stochastic matrix $B$.

Various types of two-level iterative algorithms for computing a stationary probability distribution vector of a primitive stochastic matrix have been proposed in the literature. Depending on the matrix and the aggregation they may converge faster than classical methods like the power method or other matrix iterative methods. Basic analysis of the Koury–McAllister–Stewart algorithm and Takahashi's algorithm together with sufficient convergence conditions can be found in [9]. A local convergence analysis based on properties of the error operator is presented e.g. in [4–7]. Some special choice of aggregation groups enables more detailed results. For example the algorithm presented in paper [3] can be viewed as a special case of the IAD method considered in our paper, when only one of the aggregation groups contains more than one element. Then the authors of [3] show that we can decide about the convergence of the algorithm from the sparsity structure of the diagonal block of the stochastic matrix which corresponds to that nontrivial aggregation group. Up to the mentioned special case of the IAD methods considered in [3], the analysis introduced in [8] and in this paper is the first attempt to prove the local convergence of some classes of the IAD method with general aggregation.

In our work, a two-level IAD process with a special type of basic iteration is studied. The correcting step on the fine level is performed via multiplying by a polynomial in the original stochastic matrix $B$, let us call it $p$, such that $p(1) = 1$. The aim is to specify classes of matrices $B$ for which the IAD process is locally convergent, i.e. there exists a neighborhood of the exact stationary probability distribution vector such that for any initial approximation to it the sequence obtained converges to the exact solution. Though the basic iteration matrix is chosen in such a simple form, we do not provide a complete convergence analysis for all polynomials $p$. We extend the local convergence theory to two new cases, namely for the IAD methods with basic iteration represented by multiplication by polynomials of degree one with nonnegative coefficients in matrix $B$ and for matrices with positive diagonals.

This paper continues the research started in [8], where some results for matrices with a positive row were presented. It was proved that the existence of at least one positive row of a stochastic matrix $B$ is sufficient for local convergence of the IAD method when $p(t) = t$. Then also the asymptotic convergence factor can be estimated; it is not greater than $\sqrt{1 - \alpha}$, where $\alpha$ is a minimal element in this row. In addition, for an elementwise positive matrix $B$, the asymptotic convergence factor is not greater than $1 - \alpha$, where $\alpha$ is a minimal element in $B$. On the other hand, it is shown in [8] that there exist primitive matrices for which the IAD process with $p(t) = t$ is divergent even in a local sense. It is proved with help of some ideas found in [3] that the divergence is derived from the relation of the sparsity structure of the matrix $B$ and the aggregation. Based on the facts introduced above, some convergence properties could be studied dependent on the sparsity of $B$.

Let us note that for higher degree polynomials, the study becomes more complicated, as shown in two examples introduced in this paper. The situations correspond e.g. to the cases when several steps of the power method are done within the basic iteration. In this paper, we show that there exist matrices for which the IAD method converges locally for $p(t) = t$, but it is locally divergent for $p(t) = t^2$. This surprising fact coincides with the difficulties which are met when trying to prove the local convergence for $p$ of a higher degree in the basic iteration. However, we know from the error formula, that for any matrix $B$ there exists an integer $k_0$ such that the IAD method converges when basic iteration is done via $k$ multiplications by $B$, $k \geq k_0$. Understanding the behavior of
the spectra of the asymptotic error matrix when increasing the degree of the polynomial is the main goal of our work. We consider the present paper as an initial setting for the study of higher degree polynomials in the basic iteration, which is a quite new approach in the analysis of IAD methods.

The paper is organized as follows. In Section 2 we describe the IAD method including its error operator. In Section 3, the proof of the local convergence of the IAD method with one step of the power method as the basic iteration for irreducible matrices with positive diagonals is introduced. Then in Section 4, the smoothing with a polynomial of degree one is studied. The proof of the local convergence of IAD algorithm with the smoothing by matrix $\alpha B + (1 - \alpha)I$, $\alpha \in (0, 1)$, for any irreducible matrix $B$ is presented and it is proved that the optimal value of $\alpha$ is not less than $1/2$. A modification of the IAD algorithm is introduced in Section 5 and the local convergence is proved for irreducible matrices. Finally, examples of the absence of convergence when smoothing is done with some power of $B$ greater than one are presented in Section 6. Some discussion are introduced in the last section.

2. The IAD methods

In this part we describe the IAD methods which are considered in this paper and the corresponding error operator.

Throughout this paper, let $B$ be an $N \times N$ irreducible column stochastic matrix [9,11], i.e. $B \geq 0$ and

$$\sum_{i=1}^{N} B_{ij} = 1$$

for $j = 1, 2, \ldots, N$. Then there exists a unique Perron–Frobenius eigenvector $\hat{x}$ of $B$ [1,9,11] such that $\|\hat{x}\|_1 = 1$, $\hat{x} > 0$ and

$$B\hat{x} = \hat{x}.$$  

Here $\|\cdot\|_1$ denotes the 1-norm, and $\|M\|_1 = \max_j \sum_i |M_{ij}|$. The spectral norm will be denoted by $\|\cdot\|_2$, $\|M\|_2 = \sqrt{r(M^T M)}$. We will use $\sigma(M)$ and $r(M)$ for denoting the spectrum of $M$ and the spectral radius of $M$, respectively. The identity matrix will be denoted by $I$ and $e$ will be a column vector of ones of the length appropriate to the context.

Let $G_1, \ldots, G_n$, $n \leq N$, be aggregation groups of elements (events). Any pair of sets $G_i$, $i = 1, \ldots, n$, is assumed to be disjoint and the union of them covers the all set of elements,

$$\bigcup_{i=1}^{n} G_i = \{1, 2, \ldots, N\}.$$  

The $n \times N$ restriction (aggregation) matrix $R$ is defined by $R_{ij} = 1$ if $j \in G_i$, while $R_{ij} = 0$ otherwise. For any positive $x$ we define the prolongation (disaggregation) $N \times n$ matrix $S(x)$ by

$$S(x)_{ij} = \frac{x_i}{\sum_{k \in G_j} x_k}$$

if $i \in G_j$ and $S(x)_{ij} = 0$ otherwise. Let $P(x)$ denote the projection matrix $P(x) = S(x)R$. Note that $RS(x) = I$. See [8] for other properties of the introduced items. In this part a superscript of a vector denotes an order of the vector in a sequence computed by the IAD method, while a superscript of a matrix is an exponent.
Now we describe the IAD algorithm considered in this paper. We will use a polynomial \( p \) within the IAD algorithm such that \( p(1) = 1 \). The basic correcting iteration is then carried out via multiplication by \( p(B) \).

1. An elementwise positive initial approximation \( x^0, \|x^0\|_1 = 1 \), is chosen. The value \( k \) is set to 0. A polynomial \( p \) is chosen such that \( p(1) = 1 \).

2. An \( n \times n \) aggregated matrix \( RBS(x^k) \) is composed. The associated problem is solved, i.e. a positive vector \( z \) is found, which fulfills

\[
RBS(x^k)z = z, \quad \|z\|_1 = 1.
\]

3. The prolonged vector \( x^{k+1,1} \) of the original size \( N \) is computed by

\[
x^{k+1,1} = S(x^k)z.
\]

4. A new approximation \( x^{k+1} \) of \( \hat{x} \) is computed by

\[
x^{k+1} = p(B)x^{k+1,1}.
\]

This step can be called the smoothing step or the correction on the fine level or we can refer to it as the basic iteration of the IAD method.

5. The test for convergence is evaluated and then the algorithm finishes with the approximate solution \( x^{k+1} \) or continues with Step 2 and with \( k \) increased by 1.

Let us stress that in this paper we consider the smoothing by a polynomial \( p \) in \( B \) in Step 4 of the IAD algorithm. It seems to be one of the simplest approaches to the convergence analysis of the IAD methods where the smoothing matrix may be more general than \( B \). For example, the matrix \( M^{-1}W \) can be used instead of \( p(B) \) in Step 4, where \( M = I - B \) is a splitting of nonnegative type [11].

We now introduce matrices \( Q \) and \( Z \), the spectral decomposition of the matrix \( B, B = Q + Z \), such that \( Q^2 = Q \) and \( QZ = ZQ = 0 \). The matrix \( Q \) is the Perron–Frobenius projection

\[
Q = \hat{x}e^T.
\]

When \( B \) is primitive [11], then \( Z^k \to 0 \) for \( k \to \infty \).

Since we study the convergence of the IAD methods, we recall the error operator. We denote by \( J(p, x) \) a matrix given by

\[
J(p, x) = J(p, x^k)(x^k - \hat{x})
\]

for \( k = 1, 2, \ldots \). It depends on \( k \), on the polynomial \( p \) and of course on \( B \).

**Lemma 2.1.** The approximations \( x^k, k = 0, 1, 2, \ldots \), computed by the IAD method with the correcting step corresponding to multiplication by \( p(B) \), where \( p \) is a polynomial, follow the formula

\[
x^{k+1} - \hat{x} = J(p, x^k)(x^k - \hat{x}),
\]

where

\[
J(p, x) = p(B)(I - P(x)Z)^{-1}(I - P(x))
\]

or equivalently

\[
J(p, x) = p(Z)(I - P(x)Z)^{-1}(I - P(x)).
\]
Proof. The proof can be obtained directly from the theory in papers [8] or [4,6,7] when using $p(B)$ for the correcting matrix. □

Remark 2.2. Let us show that the matrix $J(p,x)$ is not a nonnegative matrix for any positive vector $x$ unless it is a zero matrix. We have

$$e^T p(B)(I - P(x)Z)^{-1}(I - P(x)) = e^T(I - P(x)Z)^{-1}(I - P(x)) = e^T(I + P(x)Z)(I - P(x)Z)^{-1}(I - P(x)) = e^T(I - P(x)) + e^TP(x)Z(I - P(x))Z^{-1}(I - P(x)) = 0,$$

because $e^T Z = e^T (B - Q) = 0$. This yields that $J(p,x)$ may not have any strictly positive column sum.

In the major part of this work we exploit the basic idea, which is outlined in the following remark.

Remark 2.3. Let $\hat{x}$ be a stationary probability distribution vector of an irreducible matrix $B$. Let the spectral radius of $J(p,\hat{x})$ be less than one,

$$r(J(p,\hat{x})) = 1 - \delta < 1,$$

for some $\delta \in (0, 1)$. For any $\epsilon > 0$ we can find a vector norm, let us denote it by $\|\cdot\|_*$, such that

$$\|J(p,\hat{x})\|_* = \sup_{v \neq 0} \frac{\|J(p,\hat{x})v\|_*}{\|v\|_*} < 1 - \delta + \epsilon.$$

The construction is well known and it is used e.g. in [6]. Let us choose $\epsilon \in (0, \delta/3)$, then the matrix norm

$$\|J(p,\hat{x})\|_* < 1 - 2\delta/3.$$

Now due to the continuous dependency of a norm of a matrix on matrix elements, there exists a neighborhood $d(\hat{x})$ of the exact solution $\hat{x}$ (measured in $\|\cdot\|_*$) such that

$$\|J(p,x)\|_* < 1 - \delta/3$$

for any $x \in d(\hat{x})$. Choosing $x^0 \in d(\hat{x})$ we have

$$\|x^1 - \hat{x}\|_* \leq \|J(p,x^0)\|_* \|x^0 - \hat{x}\|_*.$$

Thus $x^1 \in d(\hat{x})$ as well as any other vector of the computed sequence $\{x^k\}_{k=0}^\infty$. Moreover,

$$\lim_{k \to \infty} x^k = \hat{x}$$

and the asymptotic convergence factor is not greater than $1 - \delta$. Summarizing this remark, in the case that $r(J(p,\hat{x})) < 1$ the local convergence of the IAD method is obtained.

3. Matrices with positive diagonals

In this part we show that the IAD method is locally convergent for any irreducible matrix $B$ with a positive diagonal when $p(t) = t$, in other words, when we use one step of power method
for smoothing. Let us note that an irreducible matrix with at least one positive diagonal element is primitive. We will use a tool introduced in [5] and developed in [8] for these methods. But before that let us introduce a useful lemma.

**Lemma 3.1.** Let $B$ be an irreducible stochastic matrix with the spectral decomposition $B = Q + Z$. Then for any $x > 0$, $r(P(x)Z) \leq 1$ and $1 \notin \sigma(P(x)Z)$. If $B$ is primitive then $r(P(x)Z) < 1$.

**Proof.** Let 

$$P(x)Zu = \lambda u$$

for some $u \neq 0$ and let us assume $\lambda \neq 0$. Then from $QP(x) = Q$ we have 

$$\lambda Qu = QP(x)Zu = QZu = 0,$$

which gives $Qu = 0$. Then 

$$\lambda u = P(x)Zu = P(x)Zu + P(x)Qu = P(x)Bu.$$

Then $u$ is an eigenvector of $P(x)B$ corresponding to the eigenvalue $\lambda$. Then $|\lambda| < 1$. From $e^T P(x)B = e^T$ and $P(x)B \geq 0$, the matrix $P(x)B$ is stochastic. The matrix $P(x)B$ is also irreducible. Since $Qu = 0$, the vector $u$ is not nonnegative. Then it cannot be a stationary probability distribution vector of $P(x)B$ and thus $\lambda \neq 1$. When $B$ is primitive, then $P(x)B$ is primitive and thus $|\lambda| < 1$. The proof is complete. □

In the following theorem we show that the spectral radius of $J(p, \hat{x})$ is less than one, when $p(t) = t$ and $B$ has a positive diagonal. Then from the continuous dependency of $r(J(p, x))$ on $x$, the IAD algorithm is locally convergent.

**Theorem 3.2.** The spectral radius of $J(p, \hat{x})$ is less than one for $p(t) = t$ and for any irreducible stochastic matrix $B$ with a positive diagonal.

**Proof.** Let us suppose $B \geq \alpha I$ for some $\alpha \in (0, 1)$. From Lemma 2.1 we have 

$$J(p, \hat{x}) = (B - \alpha I)(I - P(\hat{x})Z)^{-1}(I - P(\hat{x})) + \alpha(I - P(\hat{x})Z)^{-1}(I - P(\hat{x})).$$

We denote by $\tilde{J}(p, \hat{x})$ the first term on the right hand side, 

$$\tilde{J}(p, \hat{x}) = (B - \alpha I)(I - P(\hat{x})Z)^{-1}(I - P(\hat{x})).$$

In the first part of the proof we show that $r(\tilde{J}(p, \hat{x})) < 1 - \alpha$. Recalling that any irreducible matrix with at least one positive diagonal entry is primitive [11] and recalling Lemma 3.1 we can write 

$$(I - P(\hat{x})Z)^{-1} = \sum_{k=0}^{\infty} (P(\hat{x})Z)^k.$$

Then 

$$\tilde{J}(p, \hat{x}) = (B - \alpha I)(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))$$

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

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

$$+ \alpha(B - \alpha I)P(\hat{x})(I - P(\hat{x})Z)^{-1}(I - P(\hat{x})).$$
where the last equality follows from \(Q(I - P(\hat{x})Z)^{-1}(I - P(\hat{x})) = 0\). Then
\[
\tilde{J}(p, \hat{x}) = (B - \alpha I)(I - P(\hat{x})) + (B - \alpha I)P(\hat{x})\tilde{J}(p, \hat{x})
+ \alpha(B - \alpha I)P(\hat{x})(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))
+ \alpha(B - \alpha I)P(\hat{x})\left(1 + P(\hat{x})Z(I - P(\hat{x})Z)^{-1}\right)(I - P(\hat{x}))
= (B - \alpha I)(I - P(\hat{x})) + (B - \alpha I)P(\hat{x})\tilde{J}(p, \hat{x})
+ \alpha(B - \alpha I)P(\hat{x})Z(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))
\]
from \(P(\hat{x})(I - P(\hat{x})) = 0\). We can proceed further,
\[
\tilde{J}(p, \hat{x}) = (B - \alpha I)(I - P(\hat{x})) + (B - \alpha I)P(\hat{x})\tilde{J}(p, \hat{x})
+ \alpha(B - \alpha I)P(\hat{x})(B - \alpha I)(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))
+ \alpha^2(B - \alpha I)P(\hat{x})Z(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))
= (B - \alpha I)(I - P(\hat{x})) + (B - \alpha I)P(\hat{x})\tilde{J}(p, \hat{x})
+ \alpha(B - \alpha I)P(\hat{x})\tilde{J}(p, \hat{x})
+ \alpha^2(B - \alpha I)P(\hat{x})Z(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))
= \cdots
= (B - \alpha I)(I - P(\hat{x})) + \sum_{k=0}^{\infty} \alpha^k(B - \alpha I)P(\hat{x})\tilde{J}(p, \hat{x}).
\]
Since \(\alpha \in (0, 1)\), then
\[
\tilde{J}(p, \hat{x}) = (B - \alpha I)(I - P(\hat{x})) + \frac{1}{1 - \alpha}(B - \alpha I)P(\hat{x})\tilde{J}(p, \hat{x}). \tag{2}
\]
Applying a similar idea as in the proof of Proposition 2 in [8], we now introduce the corresponding “symmetrized” operators; we denote them with subscript “s” here. Thus let us denote
\[
M_s = D(\hat{x})^{-1}MD(\hat{x})
\]
for any matrix \(M\). The matrix \(D(\hat{x})\) is a diagonal matrix in which \(D(\hat{x})_{ii} = \sqrt{\hat{x}_i}\). Then we can see that \(M_s(\hat{x})\) is a symmetric projection [5,8] with its spectral norm equal to one [10]. Let us also note that \(I_s = I\) and
\[
D(\hat{x})^{-1}\hat{x} = D(\hat{x})e.
\]
Let us now introduce a matrix norm induced by a positive vector \(v\). We denote it by \(\|M\|_v\) and define it by
\[
\|M\|_v = \min\{c; v^TM \leq cv^T\}.
\]
Note that \(r(M) \leq \|M\|_v\) for any nonnegative matrix \(M\) [11]. Now we can estimate the spectral norm of \(B - \alpha I\),
\[
\|B_s - \alpha I\|_2^2 = r((B_s - \alpha I)^T(B_s - \alpha I)) \leq \|(B_s - \alpha I)^T(B_s - \alpha I)\|_v
\leq \|(B_s - \alpha I)^T\|_v \|B_s - \alpha I\|_v.
\]
We also have
\[(D(\hat{x})^{-1}\hat{x})^T(B_s - \alpha I) = (D(\hat{x})e)^T(B_s - \alpha I) = e^T(B - \alpha I)D(\hat{x}) = (1 - \alpha)e^T D(\hat{x}) = (1 - \alpha)(D(\hat{x})^{-1}\hat{x})^T\]
and
\[(B_s - \alpha I)D(\hat{x})^{-1}\hat{x} = D(\hat{x})^{-1}(B - \alpha I)D(\hat{x})D(\hat{x})^{-1}\hat{x} = D(\hat{x})^{-1}(B - \alpha I)\hat{x}
= (1 - \alpha)D(\hat{x})^{-1}\hat{x}.

Then
\[\|B_s - \alpha I\|_2^2 \leq (1 - \alpha)^2.\]  

Now coming back to (2), we have
\[\tilde{J}_s(p, \hat{x}) = (B_s - \alpha I)(I - P_s(\hat{x})) + \frac{1}{1 - \alpha} (B_s - \alpha I)P_s(\hat{x})\tilde{J}_s(p, \hat{x}),\]  

and due to symmetry of \(P_s(\hat{x})\) [10]
\[\|(I - P_s(\hat{x}))\tilde{J}_s(p, \hat{x})u\|_2^2 + \|P_s(\hat{x})\tilde{J}_s(p, \hat{x})u\|_2^2 = \|\tilde{J}_s(p, \hat{x})u\|_2^2 \]  

for any vector \(u\). Then from (4) and (5) we have
\[\|(I - P_s(\hat{x}))(I - P_s(\hat{x}))u\|_2^2 + \|P_s(\hat{x})\tilde{J}_s(p, \hat{x})u\|_2^2 = \|\tilde{J}_s(p, \hat{x})u\|_2^2 \]
and together with (3)
\[\|(I - P_s(\hat{x}))(I - P_s(\hat{x}))\|_2^2 \leq \|(B_s - \alpha I)\|_2^2 \|(I - P_s(\hat{x}))u\|_2^2 \leq (1 - \alpha)^2 \|u\|_2^2.\]

From
\[\|(I - P_s(\hat{x}))\|_2^2 \leq 1 - \alpha\]
it follows that
\[r((I - P_s(\hat{x}))\tilde{J}_s(p, \hat{x})) \leq 1 - \alpha\]
and also that
\[r(\tilde{J}_s(p, \hat{x})) \leq 1 - \alpha\]

[8]. Since
\[\tilde{J}_s(p, \hat{x}) = D(\hat{x})^{-1}\tilde{J}(p, \hat{x})D(\hat{x})\]
then

\[ r(\tilde{J}(p, \hat{x})) \leq 1 - \alpha. \]

As a final part of the proof, let us suppose that \( J(p, \hat{x})u = \lambda u. \) Then we have

\[ (B - \alpha I)(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))u + \alpha(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))u = \lambda u \]

for some \( u \neq 0. \) We can suppose \( v = (I - P(\hat{x}))u \neq 0. \) Note that the matrix

\[ I - P(\hat{x}) \]

is a projection, so \((I - P(\hat{x}))v = v.\) Let us multiply Eq. (6) by \((I - P(\hat{x})).\) We get

\[ (I - P(\hat{x}))(B - \alpha I)(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))v = (\lambda - \alpha)v. \]

From \( r((I - P(\hat{x}))\tilde{J}(p, \hat{x})) \leq 1 - \alpha \) we have

\[ |\lambda - \alpha| \leq 1 - \alpha. \]

Then for getting \( r(J(p, \hat{x})) < 1 \) we need only to prove that \( 1 \notin \sigma(J(p, \hat{x})). \) Let \( J(p, \hat{x})u = u \) for some \( u \neq 0. \) Since \( \alpha \neq 1 \) then \( v = (I - P(\hat{x}))u \neq 0 \) and

\[ (I - P(\hat{x}))(B - \alpha I)(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))v = (1 - \alpha)v, \]

or from \( QZ = 0 \) and \( Q(I - P(\hat{x})) = 0 \)

\[ (I - P(\hat{x}))(Z - \alpha I)(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))v = (1 - \alpha)v, \]

which gives

\[ (I - P(\hat{x}))(Z - \alpha I)y = (1 - \alpha)(I - P(\hat{x})Z)y, \]

where \( y = (I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))v \neq 0. \) After some rearrangement we get

\[ (Z - I)y = \alpha P(\hat{x})(Z - I)y. \]

Since \( 1 \notin \sigma(Z) \) we have \((Z - I)y \neq 0. \) But then

\[ P(\hat{x})(Z - I)y = \alpha P(\hat{x})(Z - I)y \neq 0, \]

which contradicts \( \alpha \in (0, 1). \) Then \( \sigma(J(p, \hat{x})) \subseteq [\lambda; |\lambda - \alpha| \leq 1 - \alpha] \setminus \{1\}. \) The proof is complete. \( \square \)

4. Smoothing with a convex combination of \( B \) and \( I \)

In this part we show that for any irreducible stochastic matrix \( B, \) the IAD process is locally convergent when smoothing is carried out via multiplying by \( \alpha B + (1 - \alpha)I, \alpha \in (0, 1). \) We will need the following proposition.

Lemma 4.1. Spectral radius of \( J(p, \hat{x}) \) is less than or equal to one for any irreducible stochastic matrix \( B \) when \( p(t) = t \) and

\[ \sigma(J(p, \hat{x})) \subseteq \sigma((I - P(\hat{x}))J(p, \hat{x})) \cup \{0\} \subseteq \{x; x^Tx \leq 1\} \setminus \{1\}. \]

Proof. Both relations can be easily proved using the tools of the proof of Theorem 3.2. \( \square \)
Theorem 4.2. Spectral radius of $J(p, \hat{x})$ is less than one for any irreducible stochastic matrix $B$ when $p(t) = \alpha t + 1 - \alpha$ and $\alpha \in (0, 1)$. Moreover

$$\sigma(J(p, \hat{x})) \setminus \{0\} \subseteq 1 - \alpha + \alpha \sigma((I - P(\hat{x}))J(p, \hat{x})).$$

Proof. Let $\lambda \in \sigma(J(p, \hat{x}))$. We have

$$J(p, \hat{x})u = \lambda u$$

for some $u \neq 0$. It means that

$$\alpha Z(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))u + (1 - \alpha)(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))u = \lambda u.$$

Multiplying by $(I - P(\hat{x}))$ gives rise to

$$\alpha(I - P(\hat{x}))Z(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))u = (\lambda - (1 - \alpha))(I - P(\hat{x}))u.$$

If $(I - P(\hat{x}))u = 0$ then $\lambda = 0$. So it suffices to suppose $(I - P(\hat{x}))u = v \neq 0$. Then

$$\alpha(I - P(\hat{x}))Z(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))v = (\lambda - (1 - \alpha))v,$$

$$\alpha(I - P(\hat{x}))B(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))v = (\lambda - (1 - \alpha))v,$$

$v \neq 0$. From Lemma 4.1, the spectral radius of $(I - P(\hat{x}))B(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))$ is less than or equal to one, then

$$|\lambda - (1 - \alpha)| \leq \alpha.$$

It is now sufficient to show that $1 \notin \sigma(J(p, \hat{x}))$. Since $\alpha \neq 0$ and considering $\lambda = 1$ in (7), we get the desired property from Lemma 4.1. Then

$$\sigma(J(p, \hat{x})) \subseteq \{1 - \alpha + \alpha v; v \in \sigma((I - P(\hat{x}))B(I - P(\hat{x})Z)^{-1}(I - P(\hat{x})))\} \cup \{0\},$$

thus $|\lambda| < 1$ for any eigenvalue $\lambda$ of $J(p, \hat{x})$. The proof is complete. □

Let us stress that in Theorem 4.2 we do not allow $\alpha$ being equal to 0 or 1. Based on the proofs of Theorems 3.2 or 4.2, one can easily derive that $r(J(p, \hat{x})) \leq 1$ when $p(t) = \alpha t + (1 - \alpha)$ and either $\alpha = 0$ or $\alpha = 1$. However, in such cases the local convergence is not ensured [8].

A natural question arises: what value of $\alpha$ yields the fastest asymptotic convergence for a given matrix? A partial answer is given in the following theorem.

Theorem 4.3. Let $\alpha_0$ be the value of the parameter $\alpha$, $0 \leq \alpha \leq 1$, for which the asymptotic convergence rate of the IAD method with $p(t) = \alpha t + 1 - \alpha$ is minimal. Then $\alpha_0 \geq 1/2$.

Proof. The statement follows directly from Theorem 4.2 and Lemma 4.1. The elements of the spectrum of $J(p, \hat{x})$ are convex combinations of one and of some points of a unit disc excluding one. □

Example 4.4. The piecewise linear dependence of the spectral radii of $J(p, \hat{x})$ on $\alpha$ can be seen in Fig. 1. The graphs (a)–(c) display the asymptotical spectral radii for matrices
Fig. 1. Spectral radii of $J(p, \hat{x})$ for three different matrices and $0 \leq \alpha \leq 1$.

$B_a = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}$, $B_b = \begin{pmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix}$ and $B_c = \begin{pmatrix} 1/4 & 2/5 & 1/4 \\ 1/4 & 2/5 & 1/4 \\ 1/2 & 1/5 & 1/2 \end{pmatrix}$, respectively. The aggregation groups are $G_1 = \{1\}$ and $G_2 = \{2, 3\}$ and the parameter $\alpha$ changes along the horizontal axis, $0 \leq \alpha \leq 1$. It can be noticed that the optimal values of $\alpha$ can be found between $1/2$ and $1$.

5. Relaxed approximations

In this part a simple modification will be introduced in Step 4 of the IAD algorithm. Let us consider $\alpha \in (0, 1)$ and formulas

$x^{k+1.2} = p(B)x^{k+1.1}$, \hspace{1cm} (8)

$x^{k+1} = \alpha x^{k+1.2} + (1-\alpha)x^k$ \hspace{1cm} (9)

being valid instead of (1) in Step 4 in the IAD method. Based on techniques of the proofs of previous theorems, we show that this IAD process is locally convergent for $p(t) = t$ when $B$ is an arbitrary irreducible matrix.

**Theorem 5.1.** The IAD method converges locally when Step 4 is modified according to formulas (8) and (9) with $\alpha \in (0, 1)$ and $p(t) = t$ for any irreducible matrix $B$.

**Proof.** It can be derived that for the adapted IAD method we have

$x^{k+1} - \hat{x} = \bar{J}(p, x)(x^k - \hat{x})$.

where

$\bar{J}(p, x) = \alpha B(I - P(x^k)Z)^{-1}(I - P(x^k)) + (1-\alpha)I$.

If for some $u \neq 0$

$\alpha B(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))u + (1-\alpha)u = \lambda u$, 

$I. Pultarová / Linear Algebra and its Applications 421 (2007) 122–137
then
\[ \alpha B(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))u = (\lambda - (1 - \alpha))u. \]

We know from Lemma 4.1 that the spectral radius of \( B(I - P(\hat{x})Z)^{-1}(I - P(\hat{x})) \) is less than or equal to one and one is not an eigenvalue of this matrix. Then we have
\[ \lambda \in 1 - \alpha + \alpha \sigma \left( B(I - P(\hat{x})Z)^{-1}(I - P(\hat{x})) \right). \]

The proof is complete. \( \square \)

**Remark 5.2.** The spectrum of the asymptotic error operator considered in Theorem 5.1 is almost identical to that studied in Theorem 4.2. Once again, the optimal value of parameter \( \alpha \) is not less than \( 1/2 \).

### 6. Examples

In [8] it was shown that we can find primitive matrices for which the IAD process with one step of the power method as smoothing does not converge even locally. There is a question whether two steps of power method as smoothing are sufficient for local convergence already. The answer may seem to be affirmative because the spectral radius of \( J(p_1, \hat{x}) \) cannot be greater than one when \( p_1(t) = t \) (see [8] or Lemma 4.1 in this text). Then we could think that one additional step of power method in smoothing ensures the local convergence, i.e. \( J(p_2, \hat{x}) < 1 \), where \( p_2(t) = t^2 \). However the answer to that question is in the negative. A simple example of such a behavior follows.

**Example 6.1.** A primitive stochastic matrix \( B_1 \) is given by
\[
B_1 = \begin{pmatrix}
0 & 0 & 0 & 1/2 & 0 \\
1 & 1/2 & 0 & 1/2 & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 \\
0 & 1/2 & 0 & 0 & 0
\end{pmatrix}.
\]

The aggregation groups are \( G_1 = \{1, 2\} \) and \( G_2 = \{3, 4, 5\} \). Then we obtain
\[
\hat{x}_1 = (1, 4, 2, 2, 2)^T / 11.
\]

The spectral radii of \( J(\cdot, \hat{x}) \) are
\[
r(J(p_1, \hat{x}_1)) = 1
\]
and
\[
r(J(p_2, \hat{x}_1)) = 1.1570 > 1,
\]
where \( p_1(t) = t \) and \( p_2(t) = t^2 \). Note that the IAD process is locally convergent neither for exponent equal to one nor for exponent two in the smoothing step.

One could argue that when the spectral radius of \( J(p, \hat{x}) \) is greater than or equal to one, it may not necessarily follow that the IAD process diverges. We refer to [8] for a partial explanation of the divergence for some types of sparsity patterns of \( B \). We introduce there some general examples where \( B \) is primitive but the IAD processes yields divergent sequences for almost all
initial vectors from some neighborhood of $\hat{x}$. The more detailed convergence analysis of the IAD methods can be found in [3], where a special partitioning of the set of elements is considered. One of the aggregation groups contains more than one element and any of the other aggregation groups has only one element. In this case only the sparsity pattern of the diagonal block of the matrix $B$ corresponding to the first aggregation group determines the convergence behavior of this IAD process even in a global sense.

It is also interesting to mention within this example, that in case of $p_1$ in smoothing, when $x^0$ is chosen in such a way that the last three elements of $x^0$ are not equal, then a sequence obtained asymptotically oscillates among three limiting vectors, all of which are positive. When $p_2$ is used in the smoothing step, then the sequence of vectors obtained asymptotically oscillates between two vectors some components of which are equal to zero. These considerations are understandable when using an appropriate stochastic complement formulation of the IAD method, see Lemma 2 in [8].

Now we could hope that in case of local convergence, increasing the exponent in smoothing does not destroy the local convergence. Unfortunately it is not the case again, as shown in the following example, where the original matrix $B_2$ even contains a positive row.

**Example 6.2.** We are given a primitive stochastic matrix $B_2$,

\[
B_2 = \begin{pmatrix}
0 & 0 & 0 & 1/2 & 0 \\
1 & 1/2 & 1/100 & 1/2 & 1/100 \\
0 & 0 & 0 & 0 & 99/100 \\
0 & 0 & 99/100 & 0 & 0 \\
0 & 1/2 & 0 & 0 & 0
\end{pmatrix}.
\]

The aggregation groups are $G_1 = \{1, 2\}$ and $G_2 = \{3, 4, 5\}$. Then for the Perron–Frobenius eigenvector $\hat{x}_2$ of $B_2$ we obtain

\[
r(J(p_1, \hat{x}_2)) = 0.9855 < 1
\]

and

\[
r(J(p_2, \hat{x}_2)) = 1.1271 > 1
\]

where again $p_1(t) = t$ and $p_2(t) = t^2$. Note that the IAD process is locally convergent for the exponent one of $B$ in the smoothing step, but not for two.

The following example shows that the spectral radii of $J(p_k, \hat{x})$ may not decrease monotonically when $k$ increases and $p_k(t) = t^k$.

**Example 6.3.** The matrix $B_2$ is the same as in the Example 6.2. Fig. 2 shows the spectral radii of $J(p_k, \hat{x})$, when polynomials $p_k$ are defined by $p_k(t) = t^k$, $k = 0, 1, 2, \ldots, 14$. The aggregation groups are again $G_1 = \{1, 2\}$ and $G_2 = \{3, 4, 5\}$. Let us notice

\[
J(p_0, \hat{x}) \leq 1
\]

and

\[
J(p_1, \hat{x}) \leq 1
\]

according to Remark 5.2 in the end of Section 5.
Example 6.4. This example shows a primitive stochastic matrix $B$ such that the spectral radius of $J(p_k, \hat{x})$ can be greater than one for arbitrarily large exponent $k$ when $p_k(t) = t^k$. Let us consider a primitive stochastic matrix $B$,

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

and the aggregation groups $G_1 = \{1, 2\}$ and $G_2 = \{3, 4, 5\}$. The behavior of the spectral radii of $J(p_k, \hat{x})$ for $k = 1, 2, \ldots$ (Fig. 3) can be explained using the stochastic complement formulation similar to the approach adopted in [8]. A slower descent of the extremal values of the curve is obtained when $\epsilon$ approaches zero.

On the other hand when the matrix $B$ is normal [11] and $p_k(t) = t^k$, then the spectral radii of matrices $J(p_k, \hat{x})$ decay monotonically.

7. Discussion

The purpose of this paper is to extend the local convergence analysis of one class of the IAD methods to some additional cases not treated before. We want to specify the types of the IAD methods such that we could predict the local convergence for any given stochastic matrix $B$. 
We deal only with methods for which the smoothing step is represented by multiplication by a polynomial $p$ in $B$.

As a continuation of the analysis presented in [8], we show here three new results. While in [8] we show that the existence of at least one positive row of $B$ is a sufficient condition for the local convergence of the IAD scheme when having $p(t) = t$, in this paper we show that a positive diagonal of $B$ is another sufficient condition for the local convergence for the same $p$. However we are not able to estimate the asymptotical convergence factor in this case in contrast to the case of a positive row. Such an estimate would be difficult to get, because the diagonal itself says nothing about the convergence properties of $B$, less so about the spectrum of $J(p, \hat{x})$.

As a second main result we show that a nontrivial convex combination of matrix $B$ and an identity matrix, $\alpha B + (1 - \alpha)I$, as a correcting matrix leads to the local convergence of the IAD process for any irreducible matrix $B$. The convex combination the approximations obtained in the last and next to last iterations, $x^{k+1} := \alpha x^k + (1 - \alpha)x^{k-1}$ (also called semiiteration), also leads to a locally convergent process, which forms the third theorem of this paper. Let us stress that at present we are not able to estimate the optimal value $\alpha$ in any of these two methods in order to obtain the most effective method. But we show that the optimal value of $\alpha$ is not less than $1/2$. We also do not provide any method for verifying that the initial approximation is close enough to the exact solution in these introduced cases.

Many other questions remain unsolved as well. For some matrices the IAD method converges faster than the power method; furthermore there are matrices for which the IAD method reaches the exact solution in finite number of steps [7]. In spite of that there are matrices, for which the IAD method converges slower or even diverges for some aggregation. Let us mention for example a simple $3 \times 3$ matrix from [8],

$$ B = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}, $$

where one choice of the aggregation groups, $G_1 = \{1, 2\}$ and $G_2 = \{3\}$, leads to the local convergence and moreover, the exact solution is reached in two steps. In contrast, a different choice of the aggregation groups, $G_1 = \{1\}$ and $G_2 = \{2, 3\}$, yields a divergent (oscillating) sequence $x^k$. In both cases the smoother is $B$, i.e. $p(t) = t$.

Moreover, the examples introduced in this paper in Section 6 show that the local convergence analysis for higher degree polynomials $p$ is much more difficult than that for those of degree one. We can construct matrices $B$ such that some powers of $B$ taken for smoothing yield locally divergent processes. Moreover, there are examples such that the IAD method converges for $p(t) = t$ but it diverges for $p(t) = t^2$. This means that using $B^k, k > 1$, instead of $B$ as the basic iteration matrix may lead to slower convergence or the convergence may be lost even in a local sense. On the other hand there is another property of the IAD methods: for each primitive stochastic matrix there exists an exponent $k_0$ such that for any $k \geq k_0$ the IAD process is convergent for $p(t) = t^k$ in local sense. This follows from the estimate of a norm of $J(p, \hat{x})$. For any norm we have

$$ \|Z(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))\| \leq C $$

for some real $C$. Then for $k > 1$ we have from Lemma 2.1

$$ \|J(p, \hat{x})\| \leq \|Z^{k-1}\|\|Z(I - P(\hat{x})Z)^{-1}(I - P(\hat{x}))\| \to 0, $$

because $Z^k \to 0$ for $k \to \infty$. These observations are based on the estimate of the spectrum of $Z$. We would like to obtain some conditions on $k$ under which the process is more effective than at least with $k = 1$. Unfortunately, at present we do not have any tools for such analysis.
difficulties with the choice of the number of block Gauss–Seidel or block Jacobi iteration steps between the aggregation steps were met already in [2] and still remain unsolved. Certainly, the choice of the aggregation plays a crucial role. We can observe that the local convergence of the IAD methods possesses the same effect as do some other approximate methods e.g. Newton method. The convergence conditions are either difficult or expensive to check. Further variants of the IAD methods are obtained when the aggregation matrix is built from $B^m$, $m > 1$, instead of $B$, i.e. we solve

$$RB^mS(x^k)z = z$$

in the second step of the IAD method. But we stay at $m = 1$ in our considerations in order to minimize the cost of the aggregation step.

In this paper we concentrate on the local convergence issues. But probably the most interesting question concerns the relation between the local and global convergence.

Acknowledgments

The author thanks professor I. Marek for many helpful discussions. The author is grateful to the anonymous referee for his comments and suggestions, which improved the paper. The research was supported by the project CEZ MSM 6840770001, by the Grant Agency of Czech Republic under the contract No. 201/05/0453 and by the Information Society project No. 1ET400300415.

References