Sparsity pattern for a divergent IAD method

Ivana Pultarová

Czech Technical University in Prague
Czech Republic, Europe
http://mat.fsv.cvut.cz/ivana

ILAS'05 - Regina - Saskatchewan - Canada

Outline

We assume an $N \times N$ *irreducible acyclic column stochastic* matrix B.

We compute *the stationary probability distribution vector* $\hat{x}$,

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

We use *a two - level iterative method*:

the iterative aggregation - disaggregation (IAD) method

based on the aggregation groups of events, solving a smaller problem amost exactly, prolongation the solution to the original size and correction on the original level.

Basic notations

n is number of aggregation groups (G_i).

G_1, ..., G_n are sets of indexes of events, disjoined, $\cup_{i=1}^{n} G_i = \{1, \ldots, N\}$.

I is an identity matrix and e is a vector of ones.

R is a reduction matrix, $R_{ij} = 1$ if $j \in G_i$ and 0 otherwise.

$S(x)$ is a prolongation matrix, $S_{ji}(x) = x_j / \sum_{k \in G_i} x_k$ if $j \in G_i$ and 0 otherwise.

$P(x)$ is projection $P(x) = S(x)R$.

Five steps of IAD method, steps 2 - 5 are repeated:

1. Initial approximation $x^0 > 0$ is selected, k is set to 0.
2. Solving $RBS(x^k)z = z$.
3. Prolongation z to $y^k = S(x^k)z$.
4. Correction by some appropriate matrix T, T comes from weakly regular splitting, $I - B = M - W = M(I - T)$,

$$x^{k+1} = Ty^k.$$

5. Test for convergence.

Note.

In step 4, T can be taken as B which gives the power method, or it can correspond to block Jacobi or block Gauss-Seidel methods, etc.

We will consider $T = B$ in the following.

Then it holds

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

where

$$J(x) = B(I - P(x)Z)^{-1}(I - P(x)),$$

where Z comes from the spectral decomposition of B,

$$B = P + Z, \quad P^2 = P, \quad PZ = ZP = 0,$$

P is the Perron projection of B.

Observing $J(x)$ we can find the sufficient local convergence condition.

Based on Mandel, Sekerka (83) and Marek, Mayer (98) we get

Theorem. *When* B *contains at least one positive row then the IAD method is locally convergent and the asymptotic rate of convergence is* $\sqrt{1-\alpha}$*, where* α *is the minimum in this row.*

When $B > \alpha P$ *then the IAD method is locally convergent and the asymptotic rate of convergence is* $1-\alpha$*.*

Analysing each step in the proof, we want to find some weaker conditions for local convergence. And we can come to the condition

$$P(\hat{x})Bv = 0 \qquad \Rightarrow \qquad (I - P(\hat{x}))Bv = 0$$

or equivalently

$$P(\hat{x})Bv = 0 \qquad \Rightarrow \qquad Bv = 0.$$

This is fulfilled for example when the assumptions of the following theorem hold.

Theorem. *Let in some IAD iteration there exist* $b_i > 0$ *such that*

$$x^k_{G_i} = b_i \hat{x}^k_{G_i}$$

for $i = 1, \ldots, n$. *Then* $x^{k+1} = \hat{x}$.

We can call it the "rapid convergence", the exact solution is obtained after very small number of steps, which can be estimated in advance.

Example.

a) Each set of rows of B corresponding to a particular aggregation group is a rank one matrix $\Rightarrow$ IAD method yields the exact solution after at most 2 steps.

b) $I - B = M - W$, M block diagonal and each set if rows of W corresponding to a particular aggregation group is a rank one matrix $\Rightarrow$ IAD method results in the exact solution after at most 2 steps when using $T = M^{-1}W$.

Now we aim at finding the ”boundary-line” between divergence and convergence (and rapid convergence) and we present here a small contribution to it.

We have discovered the sparsity pattern of B that leads to locally divergent IAD process when setting $T = B$ in step 2.

Theorem. *The approximations* x^k, $k = 1, 2, \ldots$ *obtained by the IAD method fulfill*

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

Let us sellect an arbitrary aggregation group, say G_i. Then after a proper symmetric permutation, B can be partitioned into

$$B = \begin{pmatrix} B_{11} & B^C_{G_i} \\ B^R_{G_i} & B_{G_i} \end{pmatrix}.$$

Similarly, let us permute and partition $P(x^k)$ into the block diagonal matrix

$$P(x^k) = \begin{pmatrix} P(x^k)_1 & 0 \\ 0 & P(x^k)_{G_i} \end{pmatrix}.$$

Then the equation $BP(x^k)x^{k+1} = x^{k+1}$ can be rewritten to

$$x^{k+1} = \rho_{k+1}\tilde{B}_{G_i}(x^k)e,$$

where ρ_{k+1} is a scaling factor and

$$\tilde{B}_{G_i}(x^k) = \begin{pmatrix} 0 & (I - B_{11}P(x^k)_1)^{-1}B^C_{G_i} \\ 0 & S_{G_i}(x^k) \end{pmatrix} \begin{pmatrix} 0 \\ P(x^k)_{G_i} \end{pmatrix},$$

where S_{G_i} is a stochastic matrix,

$$S_{G_i}(x^k) = B_{G_i} + B^R_{G_i}P(x^k)_1(I - B_{11}P(x^k)_1)^{-1}B^C_{G_i}.$$

Theorem. *When $P(x^k)_1(I - B_{11}P(x^k)_1)^{-1}$ is positive and $S_{G_i}(x)$ is cyclic for some x then it does not depend on x up to one block of cyclicity.*

Now we come to the two final theorems of this talk.

Theorem. *When* $S_{G_i}(x)$ *is cyclic of index* m *for some* x *then the sequence* $\{x^k\}_{k=0}^{\infty}$ *obtained by the IAD method can converg to the exact solution* $\hat{x}$, *only if the partial sums of part* $x^0_{G_i}$ *of the initial approximation* x^0 *corresponding to the block of cyclicity of* $S_{G_i}(x)$ *are equal.*

This immediatelly yields

Theorem. *When* $S_{G_i}(x)$ *is cyclic for some* x *then in each neighbourhood* $U(\hat{x})$ *of the exact solution* $\hat{x}$ *there exist a zero measured set* $\tilde{U}(\hat{x})$ *such that* $\{x^k\}_{k=0}^{\infty}$ *obtained by the IAD algorithm doesn't converge to the exact solution* $\hat{x}$ *for any* $x^0 \in U(\hat{x}) \setminus \tilde{U}(\hat{x})$.

Example 1.

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

and $G_1 = \{1\}$ and $G_2 = \{2, 3\}$.

Example 2.

$$B = \left(\begin{array}{ccc|cccc} \times & \ldots & \times & \times & 0 & \ldots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \times & \ldots & \times & \times & 0 & \ldots & 0 \\ \hline 0 & \ldots & 0 & 0 & 0 & \ldots & \times \\ \times & \ldots & \times & \times & 0 & \ldots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & \times & 0 \end{array} \right).$$

Conclusion.

- The local convergence of IAD method is not ensured for B irreducible a acyclic.

- The above results are valid only for $T = B$.

- Comparing to the two - groups case of Ipsen, Kirkland (2004).

- Open questions. Is $T = B^2$ sufficient for local convergence? Is $T = M^{-1}W$, where $I - B = M - W$ and M is block diagonal of $I - B$ corresponding to the aggregation groups, sufficient for local convergence?

References.

- P. J. Courtois, P. Semal, *Block iterative algorithms for stochastic matrices*, 1986
- I. C. F. Ipsen, S. Kirkland, *Convergence analysis of an improved PageRank algorithm*, 2004
- I. Marek, P. Mayer, *Convergence theory of some classes of iterative aggregation - disaggregation methods for computing stationary probability vectors of stochastic matrices*, 2003
- I. Marek, I. Pultarová, *A note on local and global convergence analysis of iterative aggregation-disaggregation methods*, 2004
- W. J. Stewart, *Introduction to the numerical solutions of Markov chains*, 1994