Iterative Aggregation - Disaggregation Methods and Ordering Algorithms

Ivana Pultarová
Czech Technical University in Prague
Faculty of Civil Engineering, Department of Mathematics
Thákurova 7, Prague 6, Czech Republic
ivana@mat.fsv.cvut.cz

ABSTRACT

The paper presents some modifications of two-level methods for computing stationary probability distribution vector of large discrete time Markov chains. The approach considered is an iterative aggregation - disaggregation method. Two types of permuting the events are considered in order to improve the convergence.

Categories and Subject Descriptors

G.1 [**Numerical Analysis**]: Mathematics of Computing—*Eigenvalues and Eigenvectors (Direct and Iterative Metods)*

General Terms

Algorithms

Keywords

Markov chains, iterative aggregation - disaggregation methods, Tarjan's algorithm

1. INTRODUCTION

We assume a column stochastic matrix B which is irreducible but not necessarily primitive. Then there exists a unique eigenvector corresponding to the eigenvalue one of B [18].

In our considerations matrix B is a convex combination of some large and sparse stochastic matrix $\tilde{B}$ and a rank-one stochastic matrix qe^T, where e is the all ones vector of the appropriate size, and $e^T q = 1$,

$$B = \alpha\tilde{B} + (1-\alpha)qe^T, \qquad \alpha \in (0,1).$$

We want to find the eigenvector of B corresponding to the eigenvalue one, i.e. the stationary probability distribution vector of a Markov chain represented by the stochastic matrix B. Let us denote it by $\hat{x}$. Then

$$B\hat{x} = \hat{x}, \qquad e^T\hat{x} = 1.$$

SMCTools 2008, October 20, 2008, Athens, GREECE

Our aim is to find an effective algorithm for computing $\hat{x}$ based on a multilevel approach. As promising seem to be multi- or two-level iterative aggregation-disaggregation (IAD) methods. Though the general theory of their convergence has not been fully developed yet, some results are known for matrices with special structures or for special adaptations of the methods or of the aggregation. Furthermore, these methods may benefit from the possibility to be performed in parallel.

In the next section we describe the algorithm of the IAD methods which will be considered throughout the paper. Some basic known results on the convergence of the algorithm are recalled in section 3. Among the ordering algorithms suitable for Markov chains, we remind the Tarjan's algorithm and a simple method which can be called "following the maximal column value" in section 4. In section 5 we present two theorems which form the background for the proposed variants of the IAD methods. These are tested in some numerical examples shown in section 5.

2. IAD METHODS

The following notation will be used. The column stochastic matrix B is of dimension N and the set of events corresponding to numbers $1, 2, \ldots, N$ is divided into $n << N$ subgroups $G_1, G_2, \ldots, G_n$ such that

$$\cup_{k=1}^{n} G_k = \{1, 2, \ldots, N\}$$

and

$$G_k \cap G_j = \emptyset \qquad \text{if} \quad k \neq j.$$

Communication matrices R (of type $n \times N$, from fine to coarse level) and $S(x)$ (of type $N \times n$, from coarse to fine level) are defined by

$$\begin{aligned} R_{rc} &= 1 \quad \text{if} \quad c \in G_r, \\ &= 0 \quad \text{otherwise}, \end{aligned}$$

$$\begin{aligned} S(x)_{rc} &= \frac{x_r}{\sum_{k \in G_c} x_k} \quad \text{if} \quad r \in G_c, \\ &= 0 \quad \text{otherwise}. \end{aligned}$$

For example, when $G_1 = \{1, 2\}$, $G_2 = \{3, 4, 5\}$ then

$$R = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix},$$

and

$$S(x) = \begin{pmatrix} 1/3 & 0 \\ 2/3 & 0 \\ 0 & 2/6 \\ 0 & 3/6 \\ 0 & 1/6 \end{pmatrix}, \qquad \text{if} \quad x = \frac{1}{12} \begin{pmatrix} 2 \\ 4 \\ 2 \\ 3 \\ 1 \end{pmatrix}.$$

The basic iteration matrix will be

$$T = M^{-1}W,$$

where

$$I - B = M - W$$

is a nonnegative splitting [18], $M^{-1} \geq 0$, $W \geq 0$. For example, it can be chosen $M = I$, M can be a block-diagonal of $I-B$, M can be a block-upper-triangle of $I-B$, etc. For the purpose of the abbreviation let us denote T_J the iteration matrix corresponding to the block Jacobi iteration, T_{GS} the iteration matrix corresponding to the block Gauss-Seidel iteration and let $T_P = B$ the iteration matrix corresponding to the power method.

The algorithm of the IAD method is then the following.

1. Choose an initial approximation x^0 and a tolerance ϵ. Set $k = 0$.
2. Solve the $n \times n$ problem
$$RBS(x^k)z = z.$$
3. Prolong z to the original size N,
$$y = S(x^k)z.$$
4. Apply $t \geq 1$ steps of the basic iteration,
$$x^{k+1} = T^t y.$$
5. Check whether $||x^k - x^{k-1}|| \leq \epsilon$. If it is not satisfied, then set $k := k + 1$ and continue in step 2.

The matrix $RBS(x^k)$ is an irreducible stochastic matrix, then the solution z is unique [13]. The eigenvector z is supposed to be carried out exactly in step 2. The number of basic iteration steps t is chosen a priori, in contrast to other forms of IAD where the convergence of the algorithm is ensured by controlling the number of basic iteration steps and checking the improvements of the error [4, 10].

3. CONVERGENCE

There are only few general results on the convergence of the IAD methods. The existence of at most one fixed point of the sequence $\{x^k\}_{k=0}^{\infty}$ computed by this method was proved in [17] for $T = T_J$ and $T = T_{GS}$. The fixed point is then equal to $\hat{x}$.

For a general column stochastic matrix B, a special choice of the aggregation groups and $T = T_P = B$ allows us to recognize whether the process converges in global sense only from the sparsity pattern of B [7]. The configuration of the aggregation groups is such that only one of the groups contains more than one event. In addition, the estimate of the asymptotic rate of convergence can be obtained. It is based on the relation [16]

$$\lambda_2(B) \leq \tau(B),$$

where the spectral sub-radius $\lambda_2(B)$ is

$$\lambda_2(B) = \max\{|\lambda|;\, \lambda \in \sigma(B),\, \lambda \neq 1\},$$

$\sigma(B)$ is a spectrum of B, and

$$\tau(B) = \frac{1}{2} \max_{i,j} |e^T B(e_i - e_j)|.$$

Here e_i is the all zero vector with the value one in the coordinate i.

Motivated by [7] the criteria for local convergence and the asymptotic convergence rate for general choice of the aggregation were proved for $T = T_P = B$ [15]. The local convergence depends only on the nonzero pattern of B. Examples of local divergence of IAD methods for primitive matrix B are also available in [15].

In [14] it was derived that when

$$B \geq qe^T$$

$q \geq 0$ and $T = T_P = B$ in the IAD method then the asymptotic rate of convergence is less than or equal to

$$\sqrt{1 - e^T q}.$$

When $q \geq \beta\hat{x}$, $\beta > 0$, the asymptotic convergence rate is less than or equal to

$$1 - \beta.$$

In [13] one can find the basic concept of the so called fast convergence. Under certain conditions the IAD method yields the exact solution $\hat{x}$ within the finite number of steps which is known a priori. In [13] it is shown that if each of the off-diagonal block rows has the range of dimension one then the IAD method with $T = B_J$ or $T = B_{GS}$ yields the exact solution after at most two IAD steps.

The explanation of this effect (not presented in [13]) is closely connected with the concept of lumpability of the iteration matrix. In fact, the iteration matrices B_J and B_{GS} are exactly lumpable. The definition of lumpability and the associated properties can be found e.g. in [3].

The mentioned result of [13] can be generalized in the following sense. Since this property will be exploited in the numerical experiments in section 5 we introduce it as a theorem [15].

THEOREM 1. *[15] Let B be an irreducible column stochastic matrix. Let B have a block structure and let the blocks be denoted by B_{rc}, $r, c = 1, 2, \ldots, n$. Let for every $r = 1, 2, \ldots, n$ the block row*

$$\begin{pmatrix} B_{r1} & \ldots & B_{r,r-1} \end{pmatrix}$$

has the rank equal to one. If in the IAD algorithm the block Jacobi basic iteration is used, $T = T_J$, then the exact solution is obtained within at most $n + 1$ IAD steps.

A special situation which fulfils the assumptions of Theorem 1 is when B is a convex combination of a block triangular stochastic matrix $\tilde{B}$ and a rank one stochastic matrix qe^T.

Now let us introduce a comparison theorem which is a consequence of the theory of nonnegative matrices and of M-matrices [1, 18].

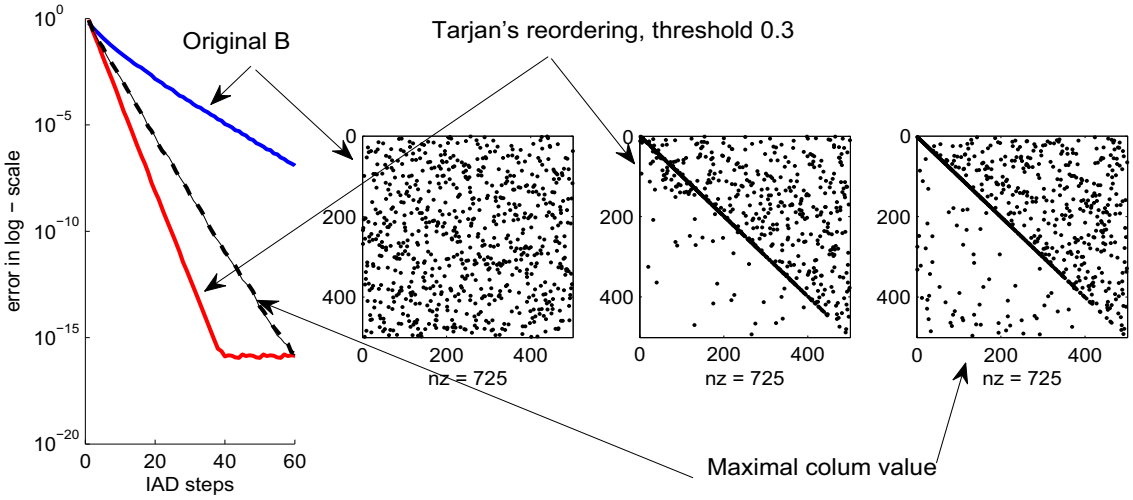

Figure 1: A test for a randomly distributed elements, 1.5 element per row.

THEOREM 2. *Let $\tilde{B}$ be a stochastic matrix, q a nonnegative vector, $e^T q = 1$, and $\alpha \in (0,1)$. Let $\tilde{B}_1$ and $\tilde{B}_2$ be two nonnegative matrices such that*

$$\tilde{B}_1 \leq \tilde{B}, \quad \tilde{B}_2 \leq \tilde{B}$$

and

$$e^T \tilde{B}_1 = c_1 e^T \geq c_2 e^T = e^T \tilde{B}_2.$$

Then for the spectral radii

$$\rho_1 = \rho\left(\alpha(I - \alpha\tilde{B}_1)^{-1}(\tilde{B} - \tilde{B}_1)\right)$$

and

$$\rho_2 = \rho\left(\alpha(I - \alpha\tilde{B}_2)^{-1}(\tilde{B} - \tilde{B}_2)\right),$$

it is

$$\rho_1 \leq \rho_2.$$

PROOF. Since $(I - \alpha\tilde{B}_i)^{-1} \geq 0$ then also

$$\alpha(I - \alpha\tilde{B}_i)^{-1}(\tilde{B} - \tilde{B}_i) \geq 0,$$

$i = 1, 2$. We have

$$\alpha e^T (I - \alpha\tilde{B}_i)^{-1}(\tilde{B} - \tilde{B}_i) =$$

$$= \frac{1 - c_i}{1/\alpha - c_i} e^T \geq 0.$$

Thus

$$\rho_i = \frac{1 - c_i}{1/\alpha - c_i}$$

and

$$\rho_2 - \rho_1 = \frac{1 - c_2}{1/\alpha - c_2} - \frac{1 - c_1}{1/\alpha - c_1} =$$

$$= \frac{(c_1 - c_2)(1/\alpha - 1)}{(1/\alpha - c_1)(1/\alpha - c_2)} \geq 0$$

□

In fact, Theorem 2 states that the iterative process

$$x^{k+1} = \alpha(I - \alpha\tilde{B}_i)^{-1}(\tilde{B} - \tilde{B}_i)x^k + (1 - \alpha)(I - \alpha\tilde{B}_i)^{-1}q$$

for solving the eigenvector u of

$$(\alpha\tilde{B} + (1 - \alpha)qe^T)u = u$$

converges faster when the l_1-norm of $\tilde{B}_i$ is larger.

Now the question may arise how to achieve the symmetric permutation of $\tilde{B}$ which leads to the desired form according to the assumption of Theorems 1 or how to obtain the structure of $\tilde{B}$ where according to Theorem 2 matrix $\tilde{B}_1$ contains large values and the solution of the system with matrix $I - \alpha\tilde{B}_1$ not expensive. In addition to it, the iteration should be performed in parallel.

These requirements yield that the desired form is when $\tilde{B}$ is block diagonal and we use the block Jacobi basic iteration and obtain the exact solution within $n+1$ steps (Theorem 1) or when the diagonal blocks involve large elements of $\tilde{B}$ and the basic iteration corresponds to block Jacobi iteration or some simpler iteration with a similar effect (Theorem 2). Some possibilities how to find the desired permutations of the events are shown in the next section.

4. ORDERING METHODS

Ordering the columns and rows of B influences the convergence of iterative algorithms. A famous example of [9] and some suggestions for checking the ordering satisfying the Property-R [5] illustrate the complexity of the problem. Some other results are oriented to the practical large scale PageRank problems. A reordering method which successively separates zero columns from diagonal subblocks applied in [12] results in a significant increase of the speed of the iterative algorithm of solving $\hat{x}$. We would like to contribute to this kind of investigation and introduce some further reordering methods. Our aim is to compare two of them when they are used as a preprocessing befor the two-level iterative algorithm described in section 2.

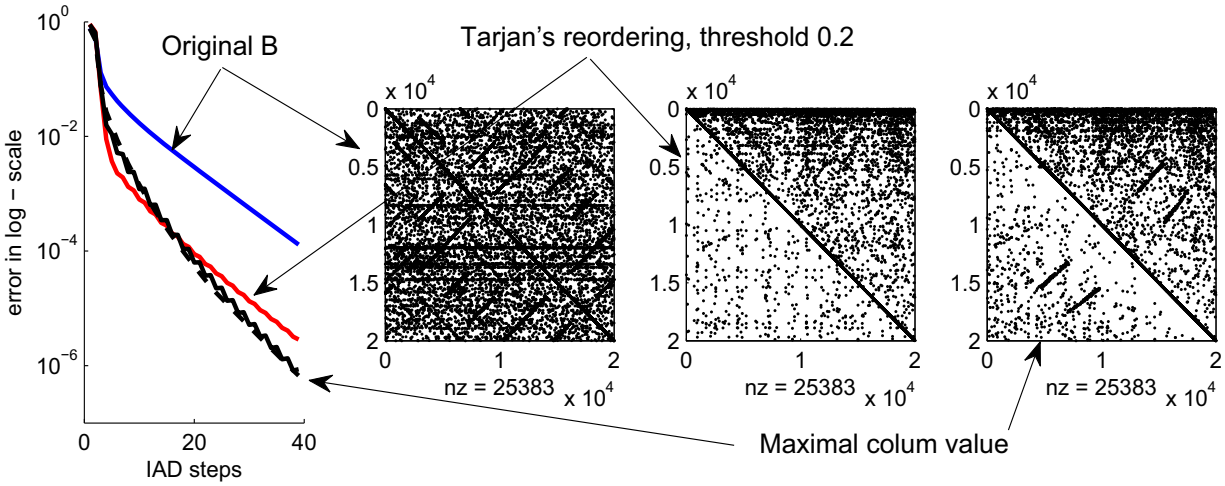

Figure 2: A test for a diagonal block of the Stanford matrix of the size 20 000.

4.1 Tarjan's Algorithm

A structure of a column stochastic matrix can be represented by an directed graph where each nonzero element B_{rc} corresponds to an edge leading from vertex c to vertex r.

A strongly connected component is a part of the graph where every vertex is reachable from another. The method for finding all of the strongly connected components is called Tarjan's algorithm and is described in [6]. The result of this algorithm is the symmetric permutation of the rows and columns of matrix $\tilde{B}$ such that it has a block upper triangular form with irreducible square diagonal blocks. Decomposition of the set of vertices into the subsets belonging to the individual strongly connected components is unique up to their order.

Complexity is approximately equal to the number of nonzero elements in B. In practise maintaining two stacks and searching in the graph may by costly. Moreover, in practical examples one of the strongly connected components is usually huge relatively to the others [2]. This is why various threshold adaptations of Tarjan's algorithm are worth being adopted.

4.2 Following the Maximal Column Value

In order to develop some less sophisticated and cheaper algorithm for accumulating large entries in diagonal blocks of $\tilde{B}$, we consider the following naive method. It is started from an arbitrary column (vertex) and then the largest value in it is found. The corresponding row index (vertex) is then taken as for the next column (next vertex). Then in this column the maximal element is found and its row index is taken as for the next column (next vertex) and so on. When any column is gone through, the row of $\tilde{B}$ with the same index is annihilated. When it happens that there is no nonzero element in some column, the path then continues in an arbitrary column (vertex). We can call this process "following the maximal column value" (FMCV).

In the formal description of the FMCV algorithm we use a vector $Perm$ in which we store the permutation vector, a set C which contains the indices of columns which have not been checked yet and a variable F which points to the last number (position) stored in $Perm$. The FMVC algorithm then can be performed in the following two steps.

1. Set
   $C := \{2, \ldots, N\}$,
   $F := 1$,
   $Perm(F) := 1$.

2. Repeat until C is empty (or equivalently F is equal to N):

   find m that $\tilde{B}_{m,Perm(F)} = \max_{j \in C} \tilde{B}_{j,Perm(F)}$,
   if $\tilde{B}_{m,Perm(F)} = 0$ then $m := \min C$,
   $F := F + 1$,
   $Perm(F) := m$,
   $C := C \setminus \{m\}$.

The desired nonzero pattern of $\tilde{B}$ is then obtained by $\tilde{B} := \tilde{B}(Perm, Perm)$.

The resulting ordering is not unique. It depends on the choices of the initial column and of the columns when the quantity $\max_{j \in C} \tilde{B}_{j,Perm(F)}$ is zero. Under fortunate circumstances the resulting matrix has zero values under the second diagonal or at least the maximal value in each column is in the second diagonal or above it. Complexity of the algorithm is equal to N. We do not provide any theory of this algorithm but we use it in the experiments in the next section.

5. NUMERICAL EXPERIMENTS

For the purpose of testing we use either randomly generated matrices or test matrices simulating the Web links which are available in [8], Stanford and Stanford-Berkeley matrices, respectively.

In each test we compute the sequence of approximations to the vector $\hat{x}$ which solves

$$(\alpha\tilde{B} + (1-\alpha)qe^T)\hat{x} = \hat{x}$$

where $\alpha = 0.85$, q is a random vector, $e^T q = 1$, $\tilde{B}$ is either a random matrix or a part of one of two mentioned test Web-matrices. The diagonal blocks of Web-matrices are modified in the following sense. In the columns with no

nonzero elements the diagonal entry is substituted by 1. In our examples, there are only about 1% of such columns. This modification should ensure the matrix will be stochastic. In many applications of PageRank type, a lot of dangling nodes are presented in the matrix B. The corresponding columns are thus empty and then they should be substituted by some appropriate unit vectors. The whole problem can be then reduced (see e.g. [12]) and only the "non-dangling block" can be handled. The experiments presented in this paper can be thus viewed as simulations for computing the "non-dangling" problems.

We measure the l_1-distance of two consecutive approximations

$$||x^k - x^{k-1}||_{l_1}$$

for two modifications of the IAD method used for different permutations of $\tilde{B}$.

The first series is computed for the original ordering of $\tilde{B}$. Then the Tarjan's reordering is used with some threshold, i.e. the values bellow the threshold are considered as for zero. The second series is thus computed for the permuted events according to Tarjan's algorithm. Third and fourth series are computed for the FMCV permutation described in paragraph 4.2.

In every test the block Jacobi basic iteration is used except of the fourth. When computing the fourth sequence, the block Jacobi basic iteration is substituted by solving the systems where the matrix is composed only of the diagonal and of the second diagonal of the diagonal blocks $I - \alpha\tilde{B}_{rr}$, $r = 1, 2, \ldots, n$. It means that this basic iteration is much faster than the block Jacobi iteration.

In each test the number of basic iteration steps t is one.

The mentioned computations were performed for a randomly generated matrix (Figure 1) and for a part of the Stanford matrix (Figure 2), respectively. The dimensions of these matrices are 500 and 20 000, respectively.

In each figure, the magnitudes of the errors are displayed in log-scale with respect to the number of steps of the IAD method. There are only three different visible curves corresponding to four different series computed. For the FMCV permutation two types of the basic iteration were used: block Jacobi basic iteration and the iteration with the diagonal and with the first sub-diagonal in each block of the reordered $\tilde{B}$. But in fact the difference between the two corresponding error curves almost cannot be distinguished.

The graphical plot of the nonzero patterns of the original matrix $\tilde{B}$, of $\tilde{B}$ after Tarjan's permutation and after FMCV permutation are displayed in the figures. The graphical representation is not too much realistic, in fact there are only about 0.3% of nonzero entries in the matrix in Figure 1 and about 0.025% of nonzero entries in the matrix on Figure 2.

6. CONCLUSIONS

We have stressed that the convergence proofs for the general IAD methods are not simple tasks even in the sense of the local behavior. Some partial results have been obtained for several particular matrix structures and for simple types of the IAD methods. One of the circumstances under which the IAD process is globally convergent was presented in Theorem 1. It says that when the block Jacobi basic iteration is used in the IAD method then the exact solution is obtained not later than after $n + 1$ IAD steps when matrix B is a convex combination of a block triangular stochastic matrix $\tilde{B}$ and a rank one matrix qe^T. The second estimate important for the practical computation is based on Theorem 2. The larger values are involved in the diagonal blocks of $\tilde{B}$ the faster is the convergence of the block Jacobi iteration.

Now we must solve how to obtain the matrix $\tilde{B}$ is the desired form. The Trajan's algorithm or the FMCV strategy may give the answer to this problem. In the latter case, the block Jacobi basic iteration can be substituted by solving much more simple system with almost the same effect: the matrix of the system contains only the diagonal and the first sub-diagonal. Once again we stress that we prefer the block method which can be parallelized, i.e. the block Jacobi before the block Gauss-Seidel.

In real situations like the PageRank solution [11], the block triangular system is practically unreachable due to the size of the problem and to the complexity of the Tarjan's algorithm. Also the permutation based on FMCV may not give the proper ordering the matrix B. Still the experiments presented in this paper show that when the data are close to that mentioned desirable structures the convergence improves. This theory thus may motivate the ideas in which order to store and how to maintain the data.

7. ACKNOWLEDGMENTS

The author thanks the anonymous referees for their stimulating suggestions and comments. This research has been supported by the project CEZ MSM 6840770001 and by the Information Society project No. 1ET400300415.

8. REFERENCES

[1] R. J. P. A Berman. *Nonnegative Matrices in the Mathematical Sciences*. Classics in Applied Mathematics, 9, 1994.

[2] K. Avrachenkov, N. Litvak, and K. S. Pham. Distribution of PageRank mass among principle components of the Web. *Electronic journal arXiv.org*, (arXiv:0709.2016v1), 2007.

[3] P. Buchholz. Exact and ordinary lumpability in finite Markov chains. *Journal of Applied Probability*, 31:59–75, 1994.

[4] P. Buchholz and T. Dayar. On the convergence of a class of multilevel methods for large sparse Markov chains. *SIAM Journal on Matrix Analysis and Applications*, 29:1025–1049, 2007.

[5] T. Dayar. State space orderings for Gauss-Seidel in Markov chains revisited. *SIAM Journal on Scientific Computing*, 19(1):148–154, 1998.

[6] I. Duff and J. K. Reid. An implementation of Tarjan's algorithm for the block triangularization of a matrix. *ACM Transactions on Mathematical Software*, (4), 1978.

[7] I. Ipsen and S. Kirkland. Convergence analysis of a PageRank updating algorithm by Langville and Meyer. *SIAM Journal on Matrix Analysis and Applications*, 27:952–967, 2006.

[8] S. Kamvar. Stanford Web Matrix and Stanford-Berkeley Web Matrix. *Data sets, available online, http://www.kamvar.org/personalization*.

[9] L. Kaufman. Matrix methods for queuing problems. *SIAM Journal on Scientific Computing*, 4(3):525–552, 1983.

[10] U. R. Krieger. Numerical solution of large finite Markov chains by algebraic multigrid techniques. *Computations with Markov Chains, Kluwer Academic Publisher, Boston*, pages 403–424, 1995.

[11] A. N. Langville and C. D. Meyer. A survey of eigenvector methods of Web information retrieval. *SIAM Review*, 47:135–161, 2005.

[12] A. N. Langville and C. D. Meyer. A reordering for the PageRank problem. *SIAM Journal on Scientific Computing*, 27(6):2112–2120, 2006.

[13] I. Marek and P. Mayer. Convergence analysis of an iterartive aggregation/disaggregation method for computing stationary probability vectors of stochastic matricess. *Numerical Linear Algebra with Applications*, 5:253–274, 1998.

[14] I. Marek and I. Pultarova. A note on local and global convergence analysis of iterative aggregation - disaggregation methods. *Linear Algebra and Its Applications*, (413):327–341, 2006.

[15] I. Pultarova. Necessary and sufficient local convergence condition of one class of iterative aggregation - disaggregation methods. *Numerical Linear Algebra with Applications*, 15(4):339–354, 2008.

[16] E. Seneta. Explicit forms for ergodicity coefficients and spectrum localization. *Linear Algebra and its Applications*, 60(421):187–197, 1984.

[17] W. J. Stewart. *Introduction to the Numerical Solutions of Markov Chains*. Princeton University Press, Princeton, 1994.

[18] R. S. Varga. *Matrix Iterative Analysis*. Springer-Verlag, New York, 2000.