Proceedings of the International Multiconference on
Computer Science and Information Technology pp. 281–288
ISBN 978-83-60810-27-9
ISSN 1896-7094

The experimental analysis of GMRES convergence for solution of Markov chains

Beata Bylina
Institute of Mathematics
Marie Curie-Sklodowska University
plac Marii Curie-Skłodowskiej 5, 20-031 Lublin, Poland
Email: beatas@hektor.umcs.lublin.pl

Jarosław Bylina
Institute of Mathematics
Marie Curie-Sklodowska University
plac Marii Curie-Skłodowskiej 5, 20-031 Lublin, Poland
Email: jmbylina@hektor.umcs.lublin.pl

***Abstract*—The authors consider the impact of the structure of the matrix on the convergence behavior for the GMRES projection method for solving large sparse linear equation systems resulting from Markov chains modeling. Studying experimental results we investigate the number of steps and the rate of convergence of GMRES method and the IWZ preconditioning for the GMRES method. The motivation is to better understand the convergence characteristics of Krylov subspace method and the relationship between the Markov model, the nonzero structure of the coefficient matrix associated with this model and the convergence of the preconditioned GMRES method.**

I. Introduction and Motivation

MARKOV chains are a particularly robust and wide used tool for analyzing a variety of stochastic (probabilistic) systems over time.

A CTMC (Continuous-Time Markov Chain) may be represented by a set of states and a transition rate matrix $\mathbf{Q}$ containing state transition rates as coefficients. To compute the steady-state probabilities we must solve a (homogeneous) sparse system of linear equations, of the form $\mathbf{Q^T x} = \mathbf{0}$, of size equal to the number of states in the CTMC. $\mathbf{Q}$ is a singular matrix demanding adequate methods to solve the equation. Solving the equation system generally requires applying iterative methods, projection methods or decomposition methods but occasionally (for the need of an accurate solution) direct methods are used as well. The rich material concerning the methods mentioned above can be found in [13].

In this article we consider one of the Krylov subspace methods, namely the GMRES method. This method was first introduced by Y. Saad in the article [12] as the method to solve linear systems of equations. GMRES for Markov chains was studied in the article [11]. The full GMRES algorithm is guaranteed to converge in at most n steps, but it is not useful for large systems of equations, because a good approximate solution is often computed quite early, after very few iterations. In the literature, we find results, which would provide an upper bound on the convergence rate of the GMRES [9]. The traditional bounds of the residual are expressed in terms of eigenvalues of $\mathbf{Q}$ and the condition number of the eigenvector matrix. It is of limited practical interest because we need the condition number, which is typically not known. For any matrix determination of its condition number is a task of the complexity $O(n^3)$. In practice, it is difficult to use the theoretical knowledge about the convergence of the GMRES method.

One of the tools used in the convergence analysis of GMRES are numerical experiments. We perform numerical experiments to help us understand the effect of nonzero structure of the matrix on the convergence characteristics of preconditioned Krylov subspace methods. We try to provide some properties of the coefficients of the matrix $\mathbf{Q}$, which affect the convergence of the method GMRES and the preconditioned GMRES.

One of the famous preconditioning techniques is incomplete factorization, for example IWZ factorization. The incomplete WZ factorization is originally described in our previous works [3], here we discuss its performance for Krylov subspace methods like GMRES.

Basing on our previous investigation we consider impact of the incomplete WZ factorization on the GMRES method for the numerical solution of Markov chains. We study relationship between the number of iterations, the convergence rate of the GMRES method and properties of the matrix $\mathbf{Q}$. Research was carried out for two models. The first model concern matrices associated with some abstract model. These matrices have not got any structure. The second model concern matrices known from the literature as the epidemic model and this matrix has got a structure.

The rest of the paper is organized as follows. Section II presents the problem. In Section III Krylov subspace are reminded. Section IV recall briefly the IWZ preconditioning. Section V presents two test models. Section VI describes conducted numerical experiments. Section VII contains some conclusions.

II. CTMCs and the Steady-State Solution

While modeling with Markov chains, in a steady state (independent of time), we obtain a linear equation system like following;

$$\mathbf{Q}^T\mathbf{x} = \mathbf{0}, \qquad \mathbf{x} \geq \mathbf{0}, \qquad \mathbf{x}^T\mathbf{e} = 1 \tag{1}$$

where $\mathbf{Q}$ is a transition rate matrix, $\mathbf{x}$ is an unknown vector of states probabilities and $\mathbf{e} = (1, 1, ..., 1)^T$. The matrix $\mathbf{Q}$ is a square one of size $n \times n$, usually a big one, of rank $n - 1$, sparse, with dominant diagonal.

1) choose:
   - an initial approximation $\mathbf{x}^{(0)}$
   - a subspace $\mathcal{K}$ spanned by $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_m]$
   - a subspace $\mathcal{L}$ spanned by $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_m]$
2) $\mathbf{r}^{(0)} \leftarrow -\mathbf{Q}^T\mathbf{x}^{(0)}$
3) $\mathbf{y} \leftarrow (\mathbf{W}^T\mathbf{Q}^\mathbf{T}\mathbf{V})^{-1}\mathbf{W}^T\mathbf{r}^{(0)}$
4) $\mathbf{x}^{(0)} \leftarrow \mathbf{x}^{(0)} + \mathbf{V}\mathbf{y}$

Fig. 1. A basic projection step for the equation $\mathbf{Q}^T\mathbf{x} = \mathbf{0}$

1) $\mathbf{v}_1 \leftarrow \mathbf{v}/||\mathbf{v}||_2$
2) for $j = 1, 2, \ldots, m$:
   a) $\mathbf{w} \leftarrow \mathbf{A}\mathbf{v}_j$
   b) for $i = 1, 2, \ldots, j$:
      i) $h_{ij} \leftarrow \mathbf{v}_i^T\mathbf{w}$
      ii) $\mathbf{w} \leftarrow \mathbf{w} - h_{ij}\mathbf{v}_i$
   c) $h_{j+1,j} \leftarrow ||\mathbf{w}||_2$
   d) $\mathbf{v}_{j+1} \leftarrow \mathbf{w}/h_{j+1,j}$

Fig. 2. The basic Arnoldi process for a subspace $\mathcal{K}_m(\mathbf{A}, \mathbf{v})$

III. Krylov subspace methods — GMRES

In this section we recap the basics of projection methods.

The projection methods consist in approximating the solution vector with a vector from a small-dimension subspace. Such approximations are repeated until our approximation is sufficiently close to the solution — in some sense the projection methods are iterative methods.

The projection methods need more space than iterative methods (because they have to store huge basis vectors of subspaces), but can converge faster than classical iterative methods — although the convergence rate is much better for the matrices 'more beautiful' in their structure than the ones arising in solving Markov chains.

A. The Projection Step

To solve a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ by a projection method first we have to choose two subspaces of dimension m from the n-dimensional space:

- $\mathcal{K}$ which is a subspace containing the approximation;
- $\mathcal{L}$ which is a subspace defining constraints for selection of approximation from $\mathcal{K}$.

Let the subspace $\mathcal{K}$ be spanned by $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$. The approximated solution is in $\mathcal{K}$ so it can be written

$$\mathbf{x} = \mathbf{V}\mathbf{y}$$

where $\mathbf{y}$ is an m-dimensional unknown vector. To find $\mathbf{y}$ we require that the residual vector $\mathbf{b} - \mathbf{A}\mathbf{x} = \mathbf{b} - \mathbf{A}\mathbf{V}\mathbf{y}$ be orthogonal to the subspace $\mathcal{L}$ spanned by $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$, that is:

$$\mathbf{W}^T(\mathbf{b} - \mathbf{A}\mathbf{V}\mathbf{y}) = 0,$$

and then (if the matrix $\mathbf{W}^T\mathbf{A}\mathbf{V}$ is nonsingular):

$$\mathbf{y} = (\mathbf{W}^T\mathbf{A}\mathbf{V})^{-1}\mathbf{W}^T\mathbf{b}.$$

If we know an initial approximation $\mathbf{x}^{(0)}$ we will rather seek a difference $\mathbf{d}$ between the exact solution $\mathbf{x}$ and $\mathbf{x}^{(0)}$: $\mathbf{x} = \mathbf{x}^{(0)} + \mathbf{d}$. Setting $\mathbf{r}^{(0)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(0)}$ we are to solve the equation

$$\mathbf{A}\mathbf{d} = \mathbf{r}^{(0)},$$

what can be done with the described above projection step.

A basic projection step for our equation 1 (where $\mathbf{A} = \mathbf{Q}^T$ and $\mathbf{b} = \mathbf{0}$) is shown in Figure 1.

The most efficient method for general, non-symmetric coefficient matrices (like $\mathbf{Q}^T$) are methods based on Krylov subspaces. A Krylov subspace is defined by its dimension m, a matrix $\mathbf{A}$ and a vector $\mathbf{v}$:

$$\mathcal{K}_m(\mathbf{A}, \mathbf{v}) = \mathrm{span}\{\mathbf{v}, \mathbf{A}\mathbf{v}, \mathbf{A}^2\mathbf{v}, \ldots, \mathbf{A}^{m-1}\mathbf{v}\}.$$

Many of such methods require that an orthonormal basis be found for the Krylov subspace. Unfortunately, classical Gram-Schmidt orthogonalization is numerically poor. To deal with it there are two main kinds of methods: Arnoldi process (which is a modified Gram-Schmidt orthogonalization) and Lanczos methods (originally for symmetric coefficient matrices but generalized in some ways).

B. The Arnoldi Process

The Arnoldi process [1] on its own (see Figure 2) generates the orthonormal basis $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$ for the subspace $\mathcal{K}_m(\mathbf{A}, \mathbf{v})$ and an upper Hessenberg matrix $\mathbf{H} = (h_{ij})$:

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} & h_{13} & \cdots & h_{1,m-1} & h_{1m} \\ h_{21} & h_{22} & h_{23} & \cdots & h_{2,m-1} & h_{2m} \\ 0 & h_{32} & h_{33} & \cdots & h_{3,m-1} & h_{3m} \\ 0 & 0 & h_{43} & \cdots & h_{4,m-1} & h_{4m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & h_{m,m-1} & h_{mm} \end{pmatrix},$$

which represents the linear transformation $\mathbf{A}$ restricted to $\mathcal{K}_m(\mathbf{A}, \mathbf{v})$ with respect to the basis $\mathbf{V}$, that is $\mathbf{H} = \mathbf{V}^T\mathbf{A}\mathbf{V}$.

The original Arnoldi process applied to a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is called *the full orthogonalization method* (FOM) [10] but a better approach is *the generalized minimum residual* algorithm (GMRES) [12]. Both the methods are shown in Figure 3. They differ only in one step — how to find the vector $\mathbf{y}$ (but both the procedures are projections [12]).

The GMRES algorithm is very popular in its iterative form. In the iterative GMRES after computing the new vector $\mathbf{x}^{(0)}$, the new residual $-\mathbf{Q}^T\mathbf{x}^{(0)}$ is checked if it is sufficiently close to $\mathbf{0}$. If not, the whole algorithm is repeated with the new $\mathbf{x}^{(0)}$ as the initial guess.

One of the advantages of this method is no fill-in generation (because the matrix $\mathbf{Q}$ is only used in the matrix-vector multiplication), the other is the fast convergence rate. The

1) choose $\mathbf{x}^{(0)}$ and m
2) $\mathbf{r}^{(0)} \leftarrow -\mathbf{Q}^T\mathbf{x}^{(0)}$
3) $\beta \leftarrow ||\mathbf{r}^{(0)}||_2$
4) $\mathbf{v}_1 \leftarrow \mathbf{r}^{(0)}/\beta$
5) for $j = 1, \ldots, m$:
   a) $\mathbf{w} \leftarrow \mathbf{Q}^T\mathbf{v}_j$
   b) for $i = 1, \ldots, j$:
      i) $h_{ij} \leftarrow \mathbf{v}_i^T\mathbf{w}$
      ii) $\mathbf{w} \leftarrow \mathbf{w} - h_{ij}\mathbf{v}_i$
   c) $h_{j+1,j} \leftarrow ||\mathbf{w}||_2$
   d) $\mathbf{v}_{j+1} \leftarrow \mathbf{w}/h_{j+1,j}$
6) FOM only:
   find $\mathbf{y} = (y_1, \ldots, y_m)$ from the $m \times m$ Hessenberg system $\mathbf{H}\mathbf{y} = \beta\mathbf{e}_1$
7) GMRES only:
   find $\mathbf{y} = (y_1, \ldots, y_m)$ minimizing $||\beta\mathbf{e}_1 - \bar{\mathbf{H}}\mathbf{y}||_2$ where $\bar{\mathbf{H}} = (h_{ij})$ is an $(m+1) \times m$ upper Hessenberg matrix
8) $\mathbf{x}^{(0)} \leftarrow \mathbf{x}^{(0)} + \sum_{i=1}^{m} \mathbf{v}_i y_i$

Fig. 3. The FOM and GMRES methods for the equation $\mathbf{Q}^T\mathbf{x} = \mathbf{0}$

iterative GMRES algorithm is also convenient to vectorize [5] and parallelize [4]).

One of the problems which materialize when using the GMRES method is to select the optimal parameter m. The lower the value of the parameter m, the shorter loop and thus the less calculation time and space. Additionally, the vector $\mathbf{v}$ is also shorter.

IV. IWZ Preconditioning

The convergence rate of iterative methods depends on properties of the coefficient matrix of the linear system. If matrix $\mathbf{Q}$ is ill-conditioned, this can make the convergence of iterative methods slow. One way to prevent such problems is to transform the system (1) into an equivalent system (having the same solution), but with better numerical properties. Such a transformation can be done by preconditioning, that is by converting the system (1) into:

$$\mathbf{M}^{-1}\mathbf{Q}^T\mathbf{x} = \mathbf{0}, \qquad \sum_{i=1}^{n} x_i = 1, \qquad \mathbf{x} \geq \mathbf{0}, \tag{2}$$

where the nonsingular matrix $\mathbf{M}$ (known as a preconditioner) approximates the matrix $\mathbf{Q}^T$ in a manner. The system (2) has the same solutions as (1) but it is (hopefully) better conditioned.

The matrix $\mathbf{M}$ should have the following properties:

- its use should entail low memory requirements;
- its inverse should be cheaply applicable;
- the transformed problem (2) should converge faster (in computational less time) than the original problem.

There is a clear conflict among these three requirements, especially for the construction of general purpose preconditioners.

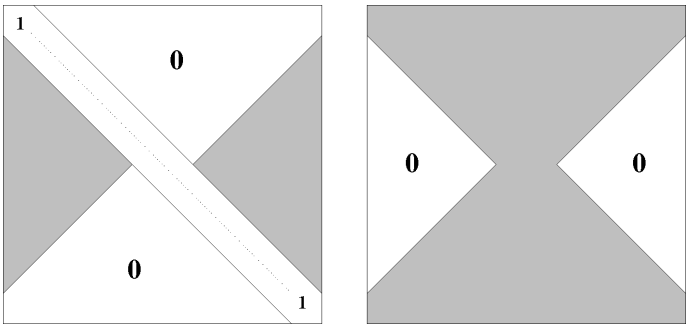

Fig. 4. The form of the output matrices in the WZ factorization (left: $\mathbf{W}$; right: $\mathbf{Z}$)

Generally, computing and using a good preconditioner is an expensive task consisting of finding the matrix $\mathbf{M}$ and its inverse. If the preconditioning is to be used, that cost should be refunded by reduced number of iterations needed to acquire required accuracy — or by using the same preconditioner for various linear systems.

The preconditioner matrix is usually built on the basis of the original coefficients of the matrix $\mathbf{Q}$. In [2] preconditioners for Krylov subspace methods for solving large singular linear systems arising from Markov modeling are considered. The incomplete WZ factorization is originally described in our previous works [3]; here we only recall it. The WZ factorization consists in decomposition of the given matrix ($\mathbf{Q}^T$ in the paper) into a product of two matrices: $\mathbf{W}$ and $\mathbf{Z}$ (Figure 4).

Incomplete WZ factorization (denoted IWZ) is based on the described above WZ factorization, where we find matrices $\widetilde{\mathbf{W}}$ and $\widetilde{\mathbf{Z}}$ (of the form of matrices $\mathbf{W}$ and $\mathbf{Z}$ shown in Figure 4) and the product $\widetilde{\mathbf{W}}\widetilde{\mathbf{Z}}$ is a kind of approximation for the matrix $\mathbf{Q}^T$.

In IWZ computations are conducted as in complete WZ factorization, but new non-zero elements (w_{ij} and z_{ij}) arising in the process are dropped if they appear in the place of a zero element in the original matrix $\mathbf{Q}^T$. Hence, the factors together have the same number of non-zeros as the original matrix $\mathbf{Q}^T$. It is worth noting that we got the inverse of $\widetilde{\mathbf{W}}$, because [14]:

$$\widetilde{\mathbf{W}}^{-1} = (-1) \cdot (\widetilde{\mathbf{W}} - \mathbf{I}) + \mathbf{I} \tag{3}$$

$$\text{(just like} \quad \mathbf{W}^{-1} = (-1) \cdot (\mathbf{W} - \mathbf{I}) + \mathbf{I}). \tag{4}$$

After IWZ we have:

$$\mathbf{Q}^T = \widetilde{\mathbf{W}}\widetilde{\mathbf{Z}} + \mathbf{R}_{WZ}, \tag{5}$$

where $\widetilde{\mathbf{W}}$ and $\widetilde{\mathbf{Z}}$ are (respectively) matrices of the form of $\mathbf{W}$ and $\mathbf{Z}$ from Figure 4 and the remainder matrix $\mathbf{R}_{WZ}$ is supposed to be small in a sense.

V. The test models

Two models are chosen to test: an abstract model (Model I) and a model of epidemics known from literature (Model II). Parameter d was introduced for the characterization of the matrixes. So, d is an average number of non-zeros in a row/column of the matrix.

TABLE I
ESSENTIAL CHARACTERISTICS OF THE MATRICES USED IN TESTS

Group	Matrix ID	n	nz	d
I	1	100	1190	11.9
II	2	100	388	3.88
I	3	1500	37955	25.3
II	4	1500	5873	3.9
I	5	3000	120590	40.2
II	6	3000	11636	3.9

TABLE II
THE TEST MATRIX ATTRIBUTES FOR A 2D MARKOV MODEL.

ID	7
N_x	64
N_y	16
n	1105
nz	9457
d	8.56

A. Model I

In this section we describe matrices corresponding to the model I. Matrices (with IDs from 1 to 6) used in tests were generated by the paper authors on the basis of some abstract queuing models — the matrices are infinitesimal generators of Markov chains describing these models — and they are neither symmetric nor anyway structural. In Table I the essential characteristics of the matrices are presented (n is the number of rows/columns of the matrix, nz is the number of non-zeros in the matrix, $d = nz/n$).

For these matrices we can observe, that the matrices might have the same size and a different value of the parameter d. The matrices were divided into two groups, the first group include matrices with $d > 8$, the second group include matrices with $d \leq 8$.

B. Model II

The matrix from model II was generated from a standard two-dimensional model [6], [7]. Table II shows the test matrix attributes for a 2D Markov model. This particular example has been taken from [8], [7]. The model is discussed there and it has been used to compare different solution methods in [13].

The state of the chain is described as a two-dimensional vector. In the first dimension, the state variable assumes all values from 0 through N_x; in the second dimension the state variable takes on values from 0 through N_y. The states of such a chain are described with two numbers (u, v), $u = 0, \ldots, N_x$, $y = 0, \ldots, N_y$ (here $N_x = 64$, $N_y = 16$) and transitions are only allowed from (u, v) to (u', v') if $|u' - u| \leq 1$ and $|v' - v| \leq 1$. There was assumed — as in [6] — that only some transition from each state are permitted. This two-dimensional Markov chain model allows for transitions from any non-boundary state to adjacent states in fixed directions (chosen from North, South, East, West, North-East, North-West, South-East, South-West). A sample scheme of the model (with allowed directions: South, East and North-West) is shown in Figure 5.

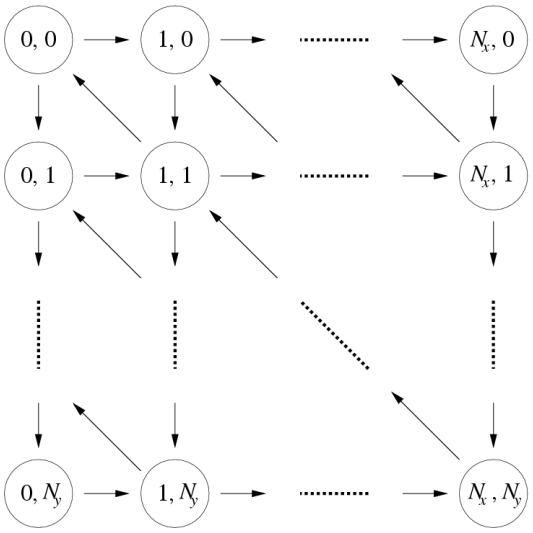

Fig. 5. A sample scheme of a two-dimensional Markov chain.

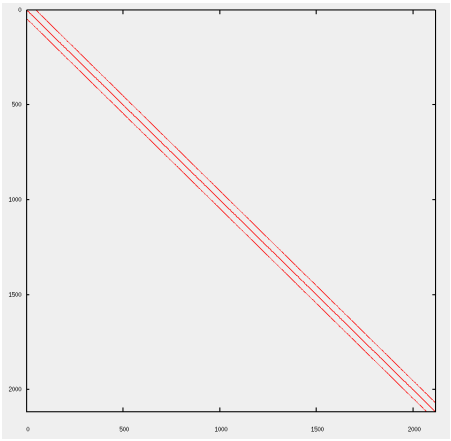

Fig. 6. The structure for the model II matrix

The matrix describing the two-dimensional Markov chain has a structure shown in Figure 6.

VI. EXPERIMENTAL RESULTS

The experiment was performed on a Pentium IV 2.8GHz computer, 1GB RAM, with Debian GNU/Linux operating system. We used high-level programming language, namely Octave.

A vector $\mathbf{x}^{(0)} = (x_i^{(0)})$ with $x_i^{(0)} = \frac{1}{i}$ was chosen as an initial vector. (We chose $x_i^{(0)} = \frac{1}{i}$ because the starting vector can be selected almost freely, but its elements should not be equal — $x_i^{(0)} = \frac{1}{i}$ fulfils this condition.) As a measure of accuracy of the solution we chose:

$$\varepsilon^{(k)} = ||\mathbf{0} - \mathbf{Q}^T\mathbf{x}^{(k)}||_2. \tag{6}$$

Accuracy has been studied experimentally for the matrix of model I and model II. We study both the number of iterations needed to achieve a given accuracy, and the rate of convergence. The stopping criterion used is that the 2-norm of the residual $||\mathbf{Q}^T\mathbf{x}^k||_2$ is less than e^{-15}

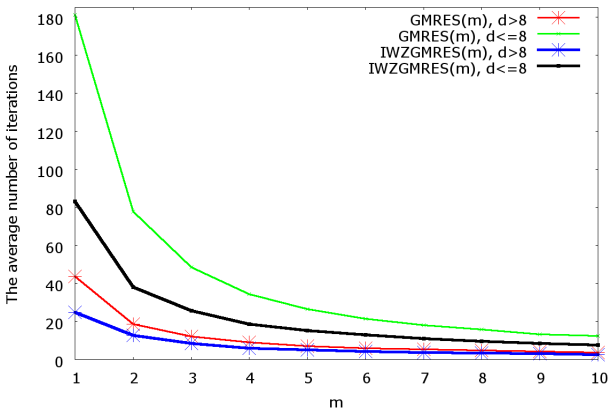

Fig. 7. Relationship between the parameter m and the average number of iterations needed to achieve accuracy e^{-16} for GMRES method and IWZGMRES method, for matrices of Group I ($d > 8$) and II ($d <= 8$).

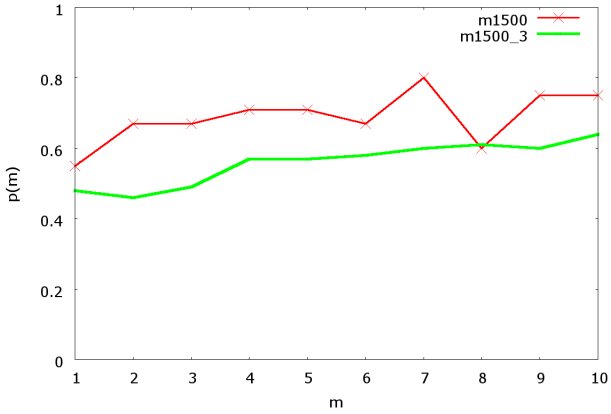

Fig. 8. Relationship between the parameter m and the value of the coefficient $p(m)$ for matrices 3 (m1500) and 4 (m1500_3)

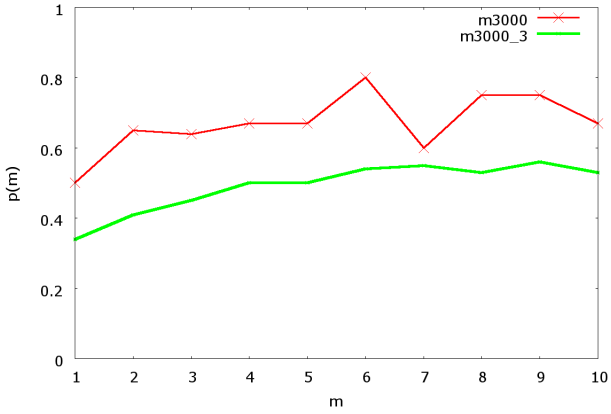

Fig. 9. Relationship between the parameter m and the value of the coefficient $p(m)$ for matrices 5(m3000) and 6 (m3000_3)

A. Number of iterations

Table III shows numbers of iteration used to achieve a given accuracy for selected parameters m for two methods: GMRES(m) alone (denoted GMRES(m)) and GMRES(m) preconditioned with IWZ (denoted IWZGMRES(m)).

Figure 7 shows the relationship between the parameter m and the average number of iterations needed to achieve a given accuracy. The average number of iterations is counted for two groups of matrices. First for the group with $d > 8$ (matrix number 1, 3, 5) and the second group of matrices with $d \leq 8$ (matrix number 2, 4, 6).

Applications that may be present at the table III and Figure 7 are as follows:

- With the increase of parameter m the average number of iterations needed to achieve the assumed accuracy of the method GMRES(m) and IWZGMRES(m) decreases inversely.
- For each matrix from the first group number of iterations needed to achieve the assumed precision for the selected value of parameter m is almost the same. Analogous relationship can be seen for matrices of the second group. It means that regardless of the size of the matrix the number of iterations needed to achieve a given accuracy is the same and depends on the parameter d.
- In the method IWZGMRES(m) the number of iterations needed to achieve a given convergence is less than the GMRES(m) method, regardless of the parameter m and the parameter d.

Let us define the coefficient $p(m)$, which shows the relationship between the number of iterations needed to achieve (given the convergence e^{-16}) in the method of GMRES(m) and the method IWZGMRES(m).

Let $I_{IWZGMRES(m)}(m)$ mean the number of iterations needed to achieve the fixed accuracy of the method IWZGMRES(m) depending on the parameter m

Let $I_{GMRES(m)}(m)$ mean the number of iterations needed to achieve the fixed accuracy of the method GMRES(m) depending on the parameter m.

Let

$$p(m) = \frac{I_{IWZGMRES(m)}(m)}{I_{GMRES(m)}(m)}.$$

Figure 8 shows the relationship between the parameter m and the value of the coefficient $p(m)$ for the matrices 3 and 4 in Table II. Matrices have size 1500 and vary in the value of the parameter d. Figure 9 shows the relationship between the parameter m and the value of the coefficient $p(m)$ for the matrices with the numbers 5 and 6 in Table II. Matrices have size 3000 and vary in the value of the parameter d.

Figures 8 and 9 show how the value of the parameter m influence the convergence. The conclusions are:

- With the increase of parameter m (where m changes from 1 to 10) the value of the coefficient of $p(m)$ grows.
- Value of the parameter $p(m)$ for the matrices of group I is higher than for the matrices in group II.
- For the matrices of the group II growth factor $p(m)$ is more uniform than for the matrix of the group I.

Let

$$p = \max_{1<=m<=10} |p(m)|.$$

TABLE III
NUMBER OF ITERATION k NEED TO ACHIEVE A GIVEN ACCURACY $\varepsilon^{(k)} = e^{-16}$ FOR THE SELECTED OF THE VALUE PARAMETER m.

	$m=1$		$m=5$		$m=10$	
ID	GMRES(m)	IWZGMRES(m)	GMRES(m)	IWZGMRES(m)	GMRES(m)	IWZGMRES(m)
1	45	29	8	6	4	3
2	87	69	17	13	8	7
3	44	24	7	5	4	3
4	185	88	30	17	14	9
5	42	21	6	4	3	2
6	271	92	32	16	15	8

TABLE IV
VALUE OF THE PARAMETER p FOR MATRICES OF THE MODEL I

ID	*p*
1	0.8
2	0.8
3	0.8
4	0.64
5	0.8
6	0.56

TABLE V
VALUE OF THE PARAMETER $p(m)$ FOR THE MATRIX OF MODEL II FOR DIFFERENT VALUES m

m	*p(m)*
5	0.86
14	1.0
25	1.33
29	1.14
33	1.0
41	1.0
49	1.0
61	0.67

Now, p can be interpreted as a number that indicates how many times GMRES(m) method can be faster, if we use IWZ as a preconditioning method.

Table IV shows the value of p for the matrices of model I.

From table IV it can be deduced that the matrices of group I, the coefficient p is the same regardless of the size of the matrix and the rate is 0.8. While for the matrices of group II ratio p decreases with increasing size of the matrix.

Table V provides a value for $p(m)$ for model II. The value of $p(m)$ is the highest for $m = 25$ and for $m = 33$, $m = 41$, $m = 49$ is a constant which means that preconditioning has no effect on the rate of convergence, and for example $m = 61$ this ratio is less than 1 which means that the method IWZGMRES(61) is faster convergent than the GMRES(61).

B. The convergence rate of the GMRES

Figures 10 and 11 present relationship between the number of iterations and the convergence $\log(||Q^T x_i||_2)$ for the matrices with the number 3 and 4 for methods GMRES(m), IWZGMRES(m) for a few selected values of parameter m. The plot shows that the higher value of the parameter m, the more rapidly convergent is the method GMRES(m). Analogously, the higher value of the parameter m means that IWZGMRES(m) method is faster convergent.

The convergence curve $\log(||Q^T x_i||_2)$ as a function i is almost of the same shape for a particular parameter m for the

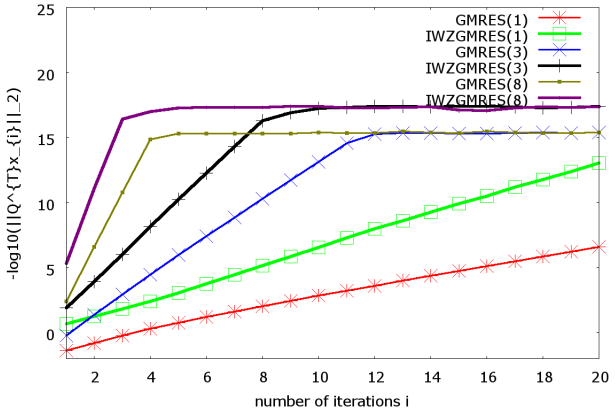

Fig. 10. Plot of the convergence curve $\log(||Q^T x_i||_2)$ as a function of i for the matrix 3

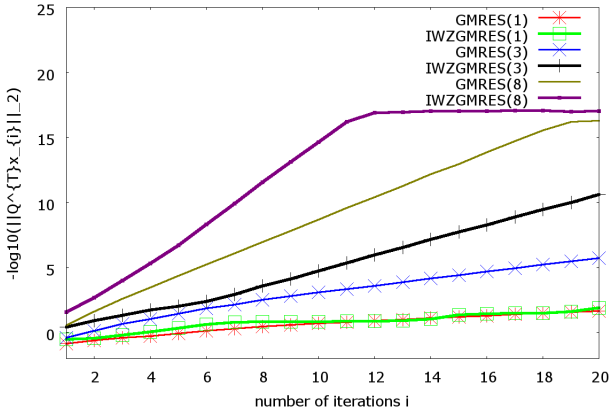

Fig. 11. Plot of the convergence curve $\log(||Q^T x_i||_2)$ as a function of i for matrix 4

GMRES(m) method and the IWZGMRES(m) method, only for the IWZGMRES(m) method the curve is shifted upwards. It means that the IWZGMRES(m) method is faster convergent than the GMRES(m) method.

Figures 12 and 13 show relationship between the number of the iterations and the convergence $\log(||Q^T x_i||_2)$ for matrices size, respectively 1500 and 3000 with different parameter d for the GMRES(m) methods and the IWZGMRES(m) methods for the parameter $m = 8$. The plots show that the GMRES(m) method and the IWZGMRES(m) method are faster convergent for matrices 3 and 5 than for matrices 4 and 6. The rate of the convergence depends on the d of the matrix and not on its

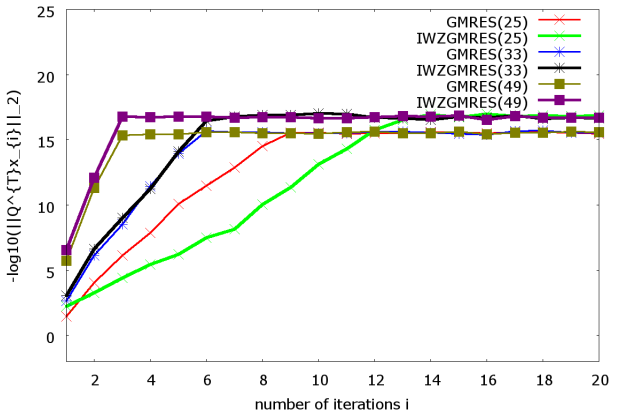

Fig. 14. Plot of the convergence curve $\log(||Q^T x_i||_2)$ as a function of i for the matrix of the model II

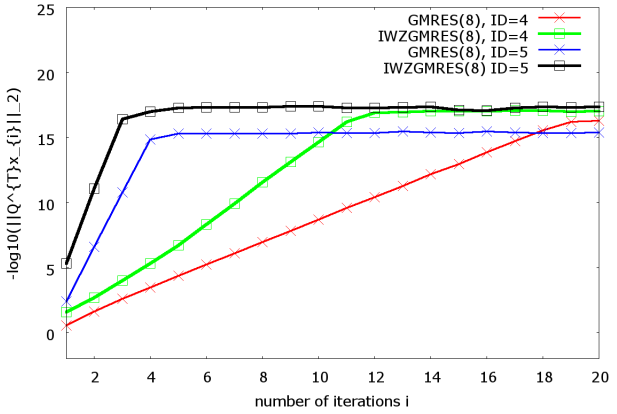

Fig. 12. Plot of the convergence curve $\log(||Q^T x_i||_2)$ as a function of i for the matrix 3 and 4

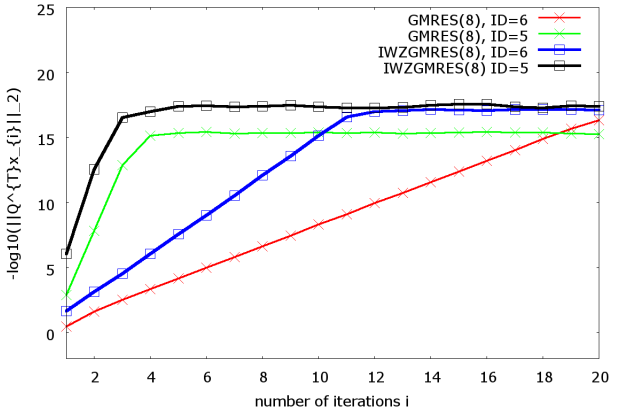

Fig. 13. Plot of the convergence curve $\log(||Q^T x_i||_2)$ as a function of i for the matrix 5 and 6

size.

Figure 14 shows relationship between number of iterations and the convergence $\log(||Q^T x_i||_2)$ for the GMRES(m) method and the IWZGMRES(m) method for parameter m with values $m = 25$, $m = 33$ i $m = 49$. Figure 14 and Table V present, that for certain values of m the method GMRES(m) had got a faster rate of convergence and for some the IWZGMRES(m) method had.

VII. Conclusion

Those numerical experiments helped us understand the effect of the nonzero structure and the size of the matrix on the convergence characteristics of preconditioned Krylov subspace methods like GMRES. The rate of convergence of the projection methods GMRES does not depend on the size of the matrix. Speed of convergence in terms of numbers of iterations of GMRES depends on the structure of the matrix. Tested matrices from two different models were characterized by the fact that the matrices of the model I have no structure, the matrix form the model II has got some structure.

Matrices from the first model was characterized by a parameter d. Namely, the set of matrices, which have a low value of the parameter d, ($d < 8$) are slowly convergent and require additional techniques to improve the rate of convergence. This technique was preconditioning.

The convergence, expressed in terms of p showed that we can identify the most optimal value of the parameter m, for which instead of the use of the GMRES(m) method we use the preconditioned GMRES(m) method, namely IWZGMRES(m).

On the basis of additional studies it may be concluded that irrespectively of the size and structure of the matrix GMRES($m + 1$) is faster convergent than the GMRES(m) for any matrix: similarly for the IWZGMRES(m) method the same dependence holds. With the increase parameter m, the rate of convergence of the method GMRES(m) increases, of course, up to some m_0, for which the rate is the largest, for all $m > m_0$ the rate is already the same.

The numerical example shows, that it is good to examine whether the matrix associated with a Markov chain is structured and has some properties, for example, the parameter d. If there is no structure you can use the preconditioning technique.

For the matrix of the model II it is not always the IWZGMRES(m) improves the convergence rate, because the matrices a structure. For this model we need to develop a separate algorithm to determine the vector of probabilities. These algorithms should take advantage of some properties of matrices associated with models.

References

[1] W. E. Arnoldi: The principle of minimized iteration in the solution of the matrix eigenvalue problem, *Quarterly for Applied Mathematics* 9, 1951, p. 17–29.

[2] M. Benzi, B. Ucar: Block Triangular preconditioners for M-matrices and Markov chains. [To appear in *Electronic Transactions on Numerical Analysis*.]

[3] B. Bylina, J. Bylina: Incomplete WZ decomposition algorithm for solving Markov chains, *Journal of Applied Mathematics*, vol. 1 (2008), n. 2, p. 147–156.

[4] J. Bylina: Distributed solving of Markov chains for computer network models, *Annales UMCS Informatica* 1 (2003), Lublin 2003, p. 15–20.

[5] J. Bylina, B. Bylina: GMRES dla rozwiazywania łańcuchów Markowa na komputerze wektorowym CRAY SV1, *Algorytmy, metody i programy naukowe*, Polskie Towarzystwo Informatyczne, Lublin 2004, p. 19–24 [in Polish].

[6] T. Dayar, W. J. Stewart.: Comparison of partitioning techniques for two-level iterative solvers on Large, Sparse Markov chins, *SIAM Journal on Scientific Computing* 21, 1691 (2000)

[7] P. K. Pollett, D. E. Stewart: An Efficient Procedure for Computing Quasi-Stationary Distributions of Markov Chains with Sparse Transition Structure, *Advances in Applied Probability* 26 (1994), p. 68.
[8] C. J. Ridler-Rowe: On a Stochastic Model of an Epidemic, *Advances in Applied Probability*, vol. 4, 1967, p. 19–33.
[9] Y. Saad: *Iterative methods for sparse linear systems*, SIAM, Philadelphia, 2003.
[10] Y. Saad: Krylov subspace methods for solving unsymmetric linear systems, *Mathematics of Computation* 37, 1981, p. 105-126.
[11] Y. Saad: Preconditioned Krylov subspace methods for the numerical solution of Markov chains, in: W. J. Stewart (Ed.), *Computations with Markov Chains*, Kluwer Academic, Dordrecht, 1995, p. 49–64.
[12] Y. Saad, M. H. Schultz: GMRES: A generalized minimal residual algorithm for solving non-symmetric linear systems, *SIAM Journal of Scientific and Statistical Computing*, 7, 1986, p. 856–869.
[13] W. Stewart: *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, Chichester, West Sussex 1994.
[14] P. Yalamov, D. J. Evans: The WZ matrix factorization method, *Parallel Computing* 21 (1995), p. 1111.