

Krylov Methods for Large Sparse Systems: A Comprehensive Overview

Amanda Zeqiri

Arben Malko

University of Tirana, Faculty of Natural Sciences,
Department of Applied Mathematics, Albania

[Doi:10.19044/esj.2021.v17n17p39](https://doi.org/10.19044/esj.2021.v17n17p39)

Submitted: 12 April 2021

Accepted: 11 May 2021

Published: 31 May 2021

Copyright 2021 Author(s)

Under Creative Commons BY-NC-ND

4.0 OPEN ACCESS

Cite As:

Zeqiri A. & Malko A. (2021). *Krylov Methods for Large Sparse Systems: A Comprehensive Overview*. European Scientific Journal, ESJ, 17(17), 39.

<https://doi.org/10.19044/esj.2021.v17n17p39>

Abstract

In this paper are analyzed behavior and properties for different Krylov methods applied in different categories of problems. These categories often include PDEs, econometrics and network models, which are represented by large sparse systems. For our empirical analysis are taken into consideration size, the density of non-zero elements, symmetry/un-symmetry, eigenvalue distribution, also well/ill-conditioned and random systems. Convergence, approximation error and residuals are compared for the full version of methods, some restarted methods and preconditioned methods. Two preconditioners are considered respectively, ILU(0) and IC(0) by using at least five preconditioning techniques. In each case, empirical results show which technique is best to use based on properties of the system and are backed up by general theoretical information already found on Krylov space methods.

Keywords: Krylov methods, numerical experiments, sparse linear systems, preconditioning

Introduction

Krylov methods are a large class of iterative algorithms that work with finite matrices and vectors in real and complex arithmetic. Many applications, which include solving linear systems, eigenvalue problems, and singular value

problems. Mathematically, these methods are based on projection methods. The main idea behind the Krylov subspace and of a projection process in general is to find the approximated solution of a potentially very large and/or sparse $Ax = b$ system by solving a much smaller system. Also, this idea has been used to design several effective algorithms for large eigenvalue problems. Such methods can have several different applications creating different algorithms, which are mathematically equivalent.

Let $Ax = b$ be a linear system where the general system matrix $A \in \mathbb{C}^{N \times N}$, $b \in \mathbb{C}^N$ and x_0 the initial approximation of the solution with its initial residual approximation $r_0 = b - Ax_0$. The Krylov subspace approach generates solutions x_m by the recursion $x_m \in x_0 + K_m(A, r_0)$ where $K_m(A, r_0) = \text{span}\{r_0, Ar_0, A^2r_0, \dots, A^{m-1}r_0\}$ is the m -th Krylov subspace generated by A from r_0 .

Since $K_m \subseteq \mathbb{C}^N$ or $K_m \subseteq \mathbb{R}^N$ (if the data are real) and based on the same recursion as x_m for the m -th residual $r_m = b - Ax_m$ the implication $r_m \in r_0 + AK_m(A, r_0) \subset K_{m+1}(A, r_0)$ occurs. The purpose is to find the approximation with short recursions very close the real (direct) solution x^{real} in few iterations (based on the limited computer memory). Also, to approximate r_m as small as possible, by elements of AK_m .

Different versions of Krylov space methods come from different choices of the m dimensional subspace and by the way the system is preconditioned (meaning to condition a given problem into a form that is more suitable for numerical solution methods by reducing its conditional number and applying a transformation of the original problem into another similar problem). Although all techniques offer the same type of polynomial approximations, the constraints used to construct these approximations will have a significant effect on the iterative techniques. The determination of some types of preconditions is related to the construction of an invariant subspace of coefficients matrix A corresponding to the eigenvalues closest to zero.

The priority of the implemented algorithms addressed in the paper consists in the ability for all of them to use multi preconditioning techniques (explained further in the article). This makes the decision of choosing an appropriate technique based on properties of A far easier compared to methods that only use one specific technique such as just right or just left preconditioning (Aditi Ghai et al., 2018).

Methodology

First, are considered two different types of sparsity for large coefficient matrix A , respectively less than 1% and 30% of non-zero elements. The most relevant Krylov methods are implemented in Matlab software and applied in both cases. Experiments have been performed by using basic methods, their

restarted version, and then by applying different types of preconditioning. Since the matrix properties variate, different behaviors in each case are observed.

Methods used are the following:

```
gmres=modified_GMRES(x0,MaxIt,tol,A,b,method,K1,K2,restart);
```

The GMRES algorithm is modified to be implemented with or without the ‘restart’ parameter using *restart* = 1 for full GMRES, or to consider preconditioning or not by using inputs ‘K1’ & ‘K2’. Also, the ‘method’ input is used to choose between two versions of GMRES as described by (Henk A. vander Vorst, 2003), with *method* = 1 corresponding to GMRES with modified Gram-Schmidt orthogonalization and *method* = 2 corresponding GMRES algorithm for complex systems.

```
[x,~,~,iter]=minres(A,b,tol,MaxIt,K1,K2,x0);
```

The MINRES algorithm is modified to consider preconditioning or not by using inputs ‘K1’ & ‘K2’, since its vulnerability to rounding errors and especially in the case of ill-conditioned matrices (Henk A. vander Vorst, 2003).

```
[x,R,H,Q]=FOM(A,b,x0,K1,K2,restart);
```

The un-preconditioned FOM algorithm uses the modified Gram-Schmidt procedure (L. Giraud & S. Gratton, 2006-2007) and a maximum number of iterations same as *N*. Then is considered preconditioning or not by using inputs ‘K1’ & ‘K2’. FOM is modified to be implemented with or without the ‘restart’ parameter.

```
bicg=BiCG(x0,MaxIt,tol,A,b,K);
```

The BiCG algorithm (Yousef Saad, 2003) is modified to consider preconditioning or not by using input ‘K’.

```
[x,~,~,iter]=QMR(A,b,tol,MaxIt,K1,K1,x0);
```

The un-preconditioned QMR algorithm uses the two-sided Lanczos iteration (Yousef Saad, 2003) and it has been modified to consider preconditioning or not by using inputs ‘K1’ & ‘K2’.

```
[x,~,~,iter]=BiCGstab(A,b,tol,MaxIt,K1,K2,x0);
```

The un-preconditioned BiCGStab algorithm described by (Henk. A. van der Vorst, 1992) and it has been modified to consider preconditioning or not by using inputs 'K1' & 'K2'.

Preconditioning

Preconditioner K is an operator close to A and if properly selected, when applied Krylov methods, it would need only a few iterations that lead to a good enough approximation of the system $Ax = b$. The difficulty to find a good preconditioner consists in the approximation, missing information about the behavior of the solution or the spectral properties of A .

Precondition techniques used are Incomplete Cholesky Factorization (IC) and Incomplete LU Factorization (ILU), both with zero fill-ins. To see which preconditioner works more efficiently are used different implementations: Left-preconditioning, Right-preconditioning and Two-sized preconditioning.

Numerical Experiments

The first two cases use $\varepsilon = 0.0001$ tolerance and 2500 maximal number of iterations. As vector b is generated a random vector of length N and as an initial approximation of the solution x_0 is used vector $\mathbf{1}$ of the same length. The difference between direct and approximate solutions is evaluated to compare iterative methods.

Case I:

At first it is used a large symmetric positive definite matrix, that its coefficients represent variables of a neural network problem, `A=load('494_bus.mat').Problem.A`. It has a density of 0.6827% non-zero elements and it is ill-conditioned. In figure 1 it is shown a visual representation of the 494×494 matrix with 1666 non-zero elements.

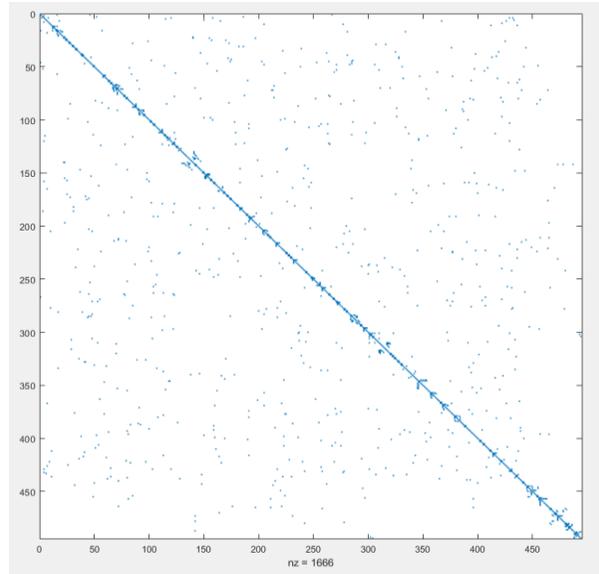


Figure 1. 0.6827% density matrix.

The full version of all six Krylov methods is applied, with no restart parameter and no preconditioning. Overall results are, GMRES ends with 250 iterations and an approximation error of 0.0012. MINRES converges after 908 iterations and an approximation error of 0.0013. FOM uses $N = 494$ maximum number of iterations which gives an approximation error of $0.3888e-06$. BiCG converges after 880 iterations with an approximation error of 0.0088. Followed by QMR and BiCGStab with approximation errors 0.0012 and 0.00694, which converge respectively after 1026 and 1034 iterations.

It is FOM method that makes a noticeable difference in the approximation error in this case, but it is pointed out that FOM is not very stable depending on the properties of the system $Ax = b$. BiCGStab is a very more stable alternate with a good enough approximation of the solution. Although QMR, GMRES and MINRES do not differ a lot from one another, GMRES makes the slightest difference in converging faster based on the number of iterations.

Table 1. Restarted methods with parameter 1 to 10.

Method Restart	GMRES		MINRES		BiCG		QMR		BiCGStab		FOM
	Error	It.	Error	It.	Error	It.	Error	It.	Error	It.	Error
1	0.0012	250	0.0013	908	0.0088	880	0.0012	1026	6.94e-04	1034	3.888e-07
2	0.0011	1	0.0013	1	0.0099	7	0.0012	1	6.94e-04	1	2.00e-10
3	0.0011	1	0.0013	1	0.0083	3	0.0012	1	6.94e-04	1	2.00e-10
4	0.0011	1	0.0013	1	0.0078	3	0.0012	1	6.94e-04	1	2.00e-10
5	0.0011	1	0.0013	1	0.0099	14	0.0012	1	6.94e-04	1	2.00e-10
6	0.0011	1	0.0013	1	0.0072	3	0.0012	1	6.94e-04	1	2.00e-10
7	0.0011	1	0.0013	1	0.0068	3	0.0012	1	6.94e-04	1	2.00e-10
8	0.0011	1	0.0013	1	0.0053	3	0.0012	1	6.94e-04	1	2.00e-10
9	0.0011	1	0.0013	1	0.0063	3	0.0012	1	6.94e-04	1	2.00e-10
10	0.0011	1	0.0013	1	0.0084	2	0.0012	1	6.94e-04	1	2.00e-10

Furthermore, the restarting process is applied to the same methods. In Table 1 it is noted as expected that increasing the restarting parameter, it does not affect the approximation accuracy and most methods converge after the first iteration, except for BiCG method as it is pretty sensitive to the initial approximation. It is proceeded by applying different preconditioning techniques using ILU and IC factorizations for each method (M. Benzi, 2002) & (Henk A. vander Vorst 2003). It is used parameter K1 to set Right ILU Preconditioning, parameter K2 to set Left ILU Preconditioning, parameter $K = L'$ to set the Right IC Preconditioning (where $A = LL' + E$) and parameter K' to set Left IC Preconditioning. Results are shown below in Table 2:

Table 2. Preconditioning Techniques Case 1.

Method	Two-sized ILU	Right ILU	Left ILU	Right IC	Left IC
GMRES	-	-	-	-	-
MINRES	0.0011, 80	17.8680	18.1915	-	-
BiCG	-	-	-	-	-
QMR	0.0011, 80	0.0012, 865	0.0012, 113	9.6295e-04, 254	0.0010, 231
BiCGStab	8.5883e-04, 59	8.0051e-04, 823	3.5014e-04, 72	0.0011, 269	0.0011, 269

GMRES and BiCG convergence is not improved at all with both ILU and IC factorization, error increases greatly. This is an exact case where preconditioning makes things worst, due to specific properties of matrix A (a very sparse one). MINRES convergence on the other hand is improved with two-sized ILU preconditioning by reducing the number of iterations at 80 from 908. K1 (right) and K2 (left) preconditioning not so successful. Both successful improvements in convergence speed and error are not always possible. QMR method has the same accuracy as no-preconditioning with a noticeable reduction of the number of iterations using two-sized ILU preconditioning although right IC preconditioning can cause a more drastic decrease in error approximation. When preconditioning is applied for

BiCGStab method, the most effective technique is left-preconditioning with ILU factorization. It affects both the accuracy and speed of the method.

Case II:

In the next experiment a random matrix by increasing size to $N = 1000$ and density up to 30%, is considered:

```
A = sprand(n, n, density);  
A = 100*A - 50*spones(A);
```

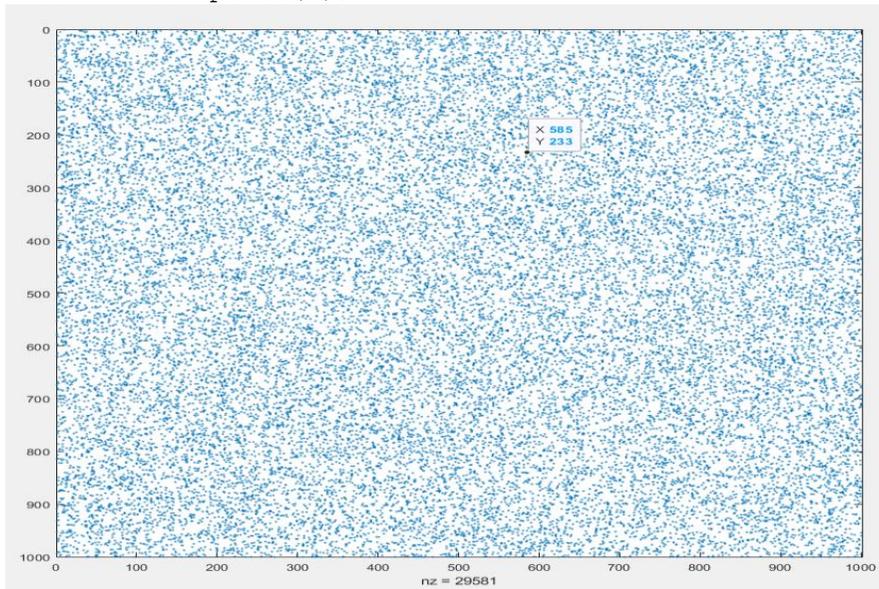


Figure 2. 30% density matrix

As expected, since the matrix is neither symmetric nor positive definite, random and has a very high condition number $\kappa(A) = 6.1951e + 04$, different behavior from the same methods is shown. GMRES is the only method that shows stability, with a satisfactory approximation error of $2.3017e - 13$ and 1000 number of iterations. No restart and no precondition FOM had an unexpected result considering sensibility to ill-condition problems, in this case, after GMRES it has an error of $2.9153e - 13$. Applying the restart process once more it does not improve the approximation accuracy and convergence. Preconditioning with ILU and IC Factorization, using the same techniques as in Table 2 does not improve the results, leading to search for other successful preconditioning techniques.

Case III:

In this case is considered a modification of a real un-symmetric matrix from transient stability analysis of Navier-Stokes solvers of order $N = 23560$. The source of the matrix is The Matrix Market from National Institute of

Standards & Technology. A small perturbation analysis of a FD (finite difference) approximation of the Navier-Stokes equations is associated with the eigenvalues and corresponding eigenvectors of A . These equations represent flows over airfoils. Conditioning is applied by increasing the density of non-zero elements around the main diagonal as following:

```
load airfoil
x = pow2(x,-32); y = pow2(y,-32); n = max(max(i),max(j));
A = sparse(i,j,-1,n,n); A = A + A';
d = abs(sum(A)) + 1;
A = A + diag(sparse(d));
```

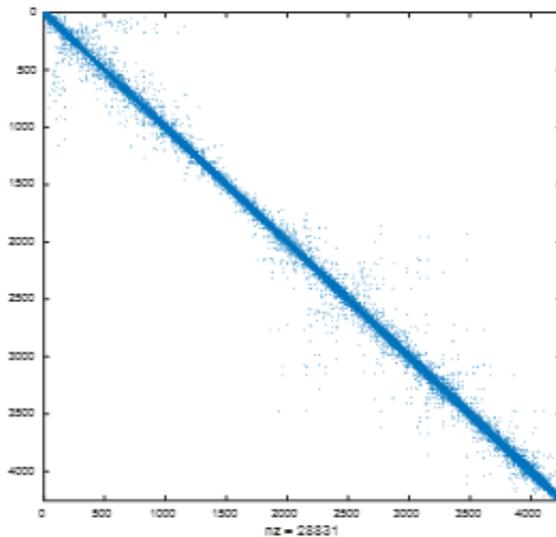


Figure 3. Matrix of order $N = 23560$, with 28831 non-zero elements

Tolerance $\varepsilon = 0.001$ is used and 1000 maximal number of iterations. Vector b equals vector y and the condition number, in this case, is 19. GMRES error is stable with restarting after 7 iterations with an average 3% error and an average relative residual norm of 0.08% as shown in Figure 4. The best approximation is achieved with restarting after 4 iterations with a 1.6 % error and 0.05% relative residual norm, considering the maximum number of iterations or not. Decreasing error tolerance by 10% will lead to reducing relative residual norm from $e-3$ to $e-5$, but will maintain the same proportions concerning tolerance in each case, whereas the approximation error does not change.

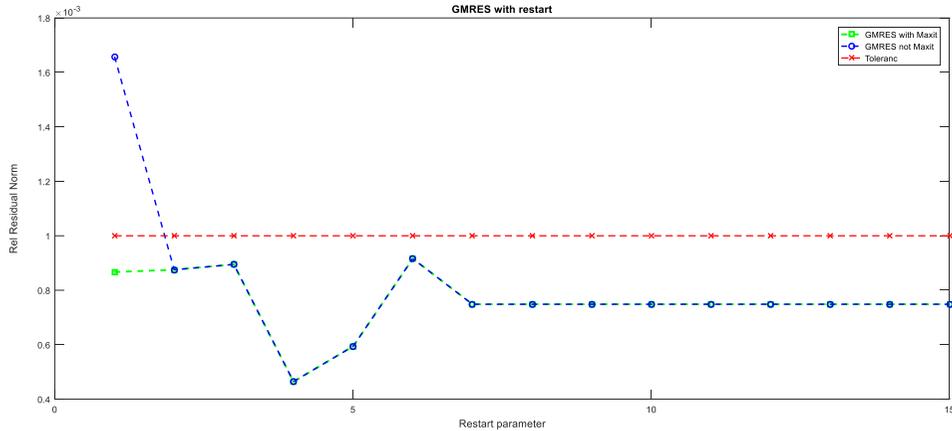


Figure 4. GMRES restarting process

QMR and MINRES have respectively error and relative residual norm of 2.58% and 0.07% for both methods. BiCG method compared to other methods has the highest error 3.09% and the highest relative residual norm 0.09%. Meanwhile, BiCGStab has the smallest error of 1.95% and the smallest relative residual norm of 0.06%.

By applying five different preconditioning techniques for each method, the following results were found:

- GMRES with two-sized ILU reduces the error to 0.02% and relative residual norm to 1.49%.
- Preconditioned QMR with Left-U technique reduces the error to 0.03% and relative residual norm to 1.20%.
- Preconditioned MINRES does not show improvement with any of the five techniques.
- Both Left-U & Right-U preconditioning for BiCG result in the reduction of the error to 0.03% and relative residual norm to 1.03%.
- Preconditioned BiCBStab gives the smallest error of 0.03% and the smallest relative residual norm 1.15% with the IC decomposition.

Figure 5 and Figure 6 show the error and the relative residual norm of Krylov methods for different preconditioning techniques.

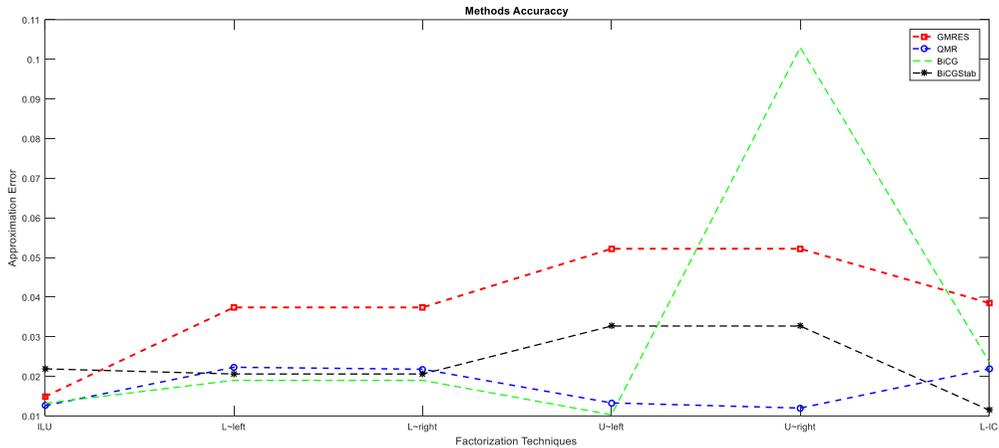


Figure 5. Methods accuracy with different preconditioning techniques

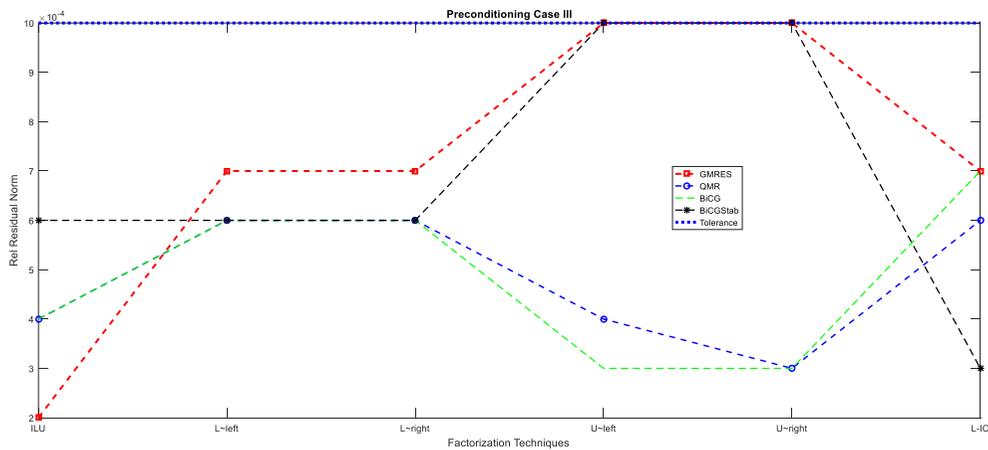


Figure 6. Relative residual norm with different preconditioning techniques

Conclusion

In the first case, where matrix A is symmetric and positive definite of order 494, the most robust Krylov method is BiCGStab, which improves noticeably after preconditioning with Left-ILU (in most numerical experiments Right preconditioning is more effective). BiCGStab has smoother convergence compared to other methods used on symmetric systems such as CG or BiCG and with a good preconditioning technique BiCGStab is not likely to breakdown due to its Bi-Lanczos iteration. Full FOM in the first case also gives a convenient result.

In the second case where A is neither symmetric nor positive definite and random of order 1000, the effective method was full GMRES. All other methods stagnated. The preconditioning techniques taken into consideration do not improve the convergence of methods, leading us to try other preconditioning schemes such as ILU(k) with fill-in since A has a random

distribution. In both first two cases increasing the restart parameter, slowed down the convergence drastically.

The last experiment where A is a matrix of order 23560 and unsymmetric, all methods except MINRES give approximation errors below the given tolerance. BiCGStab has the smallest error and relative residual norm with no precondition. Preconditioning on the other hand by using different schemes gives BiCGStab and other methods, except QMR, similar performance and apparent error or residual oscillations. QMR is the most stable method after all preconditions. Overall, GMRES is the most robust method that shows improvement with restart and two-sized ILU preconditioning with the smallest average error and smallest average relative residual.

The convergence of the most robust methods based on the experiments, GMRES and BiCGStab, depends on the distribution of the eigenvalues of A and its corresponding eigenvectors (O. Axelsson & G. Lindskog, 1986).

References:

1. Henk A. vander Vorst (2003). *Iterative Krylov Methods for Large Linear Systems*. Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press.
2. L. Giraud & S. Gratton (2006-2007). *Introduction to Krylov subspace methods for the solution of linear systems*.
3. Yousef Saad (2003). *Iterative Methods for Sparse Linear Systems, Second Edition*. Society for Industrial and Applied Mathematics.
4. Aditi Ghai, Cao Lu and Xiangmin Jiao (2018). *A Comparison of Preconditioned Krylov Subspace Methods for Large-Scale Nonsymmetric Linear Systems*. Cornell University.
5. Zhong-Zhi Bai (2015). *Motivations and realizations of Krylov subspace methods for large sparse linear systems*. Journal of Computational and Applied Mathematics.
6. David Chin-Lung Fong and Michael Saunders (2012). *CG Versus MINRES: An Empirical Comparison*. SQU Journal for Science.
7. O. Axelsson and G. Lindskog (1986). On the eigenvalue distribution of a class of preconditioning methods. Numer. Math.
8. Henk. A. van der Vorst (1992). *Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*. SIAM J. Sci. Stat. Comput.
9. M. Benzi (2002). *Preconditioning Techniques for Large Linear Systems: A Survey*. Journal of Computational Physics, 182.
10. P. N. Brown (1991). *A theoretical comparison of the Arnoldi and GMRES algorithms*. SIAM Journal on Scientific and Statistical Computing, 12.

11. M. H. Gutknecht (1992). *A completed theory of the unsymmetric Lanczos process and related algorithms, Part I*. SIAM J. Matrix Anal. Applic.