Grid computing: a case study in hybrid GMRES method

Ye Zhang[1], Guy Bergere[1], Serge Petiton[1]

[1] Laboratoire d'Informatique Fondamentale de Lille, USTL
59650 Villeneuve D'Ascq, France
{Ye.Zhang, Guy.Bergere, Serge.Petiton}@lifl.fr

Abstract. Grid computing in general is a special type of parallel computing. It intends to deliver high-performance computing over distributed platforms for computation and data-intensive applications by making use of a very large amount of resources. The GMRES method is used widely to solve the large sparse linear systems. In this paper, we present an effective parallel hybrid asynchronous method, which combines the typical parallel GMRES method with the Least Square method that needs some eigenvalues obtained from a parallel Arnoldi process. And we apply it on a Grid Computing platform Grid5000. From the numeric results, we will present that this hybrid method has some advantage for some real or complex systems compared to the general method GMRES.

Keywords: Grid, hybrid, GMRES, complex

1 Introduction

Iterative methods are a common choice for solving the large linear sparse system of the form Ax=b. A popular class of iterative methods is Krylov subspace methods. The generalized minimum residual algorithm (GMRES) [2] is used widely and it is often referred to as an "optimal" method because it finds the approximate solution in the Krylov subspace that minimaized the 2-norm of the residual. In order to limit both computation and memory requirements, a restarted version is often used. It has been implemented on parallel systems [1], but this method does not always converge very fast. There are some existing modifications to the standard GMRES algorithm. We study a hybrid method [9] which calculates in parallel some eigenvalues by the Arnoldi method [3], [4]. As soon as they are approximated with a sufficient accuracy, the eigenvalues are used to perform some iterations of the Least Squares method [6] for getting a new initial vector for the next GMRES restarts. We have applied it on the supercomputer to solve some small real linear system [9], [15].

We perform our experiments on a Grid system. As known to all, the Grid is well established as a research domain and proposes technologies that are mature enough to be used for real-life applications. It is dedicated to achieve a high performance of large scale computing by using a large amount of unoccupied computing resources.

The Grid5000 project has been launched to provide the community of Grid researchers with an unprecedented large-scale infrastructure to study Grid issues under real experimental conditions. Grid'5000 is a large scale computing tool composed of many clusters distributed in several computing centers in France.

In this paper, we present the distributed hybrid method GMRES(m)/LS-Arnoldi which is well implemented on the GRID system Grid'5000. And we try to apply it on the large linear systems and the complex systems.

This paper is organized as follows. The numerical methods used in our hybrid method will be present in section 2. In section 3, we introduce the implementation on Grid'5000. In section 4, we present the results obtained on the platform Grid'5000. At the same time we sum up the advantages and characteristic and see the effect on complex problem. Finally, in section 5, we present a summary and discuss directions for future research.

2 GMRES(m)/LS-Arnoldi Hybrid Parallel Method

This method aims to accelerate the convergence with the benefit of Arnoldi and Least Square methods. Arnoldi method uses the Gram-Schmidt method to compute the orthonormal basis of the Krylov subspace. It is well-known for approximating eigenvalues of large sparse matrices. Least Square is a polynomial iteration method that can offers us a new initial vector by using the eigenvalues information.

The idea is that during the GMRES iterations, if we can offer more information about the matrix, the efficiency of convergence can be increased. So we use Least Square method to obtain the new initial vector for the next GMRES iterations. And the Arnoldi process is performed in parallel to calculate some eigenvalues with a sufficient accuracy for the Least Square computation.

2.1 GMRES Method

GMRES (Generalized Minimum RESidual) method is one of the iterative methods based on Krylov subspace. Such methods find an approximate solution $x_i \in x_0 + K_i(A, r_0)$, where $K_i(A, r_0) \equiv span\{r_0, Ar_0, ..., A^{i-1}r_0\}$ denotes an i-demensional Krylov subspace, x_0 is the initial guess, and r_0 is the initial residual.

The GMRES method was proposed by Saad and Schultz[2] in 1986. It is used widely to solve non-symmetric linear systems. The m^{th} iterate x_m of GMRES is the solution of the least squares problem: $minimize_{x \in x_0 + K_m(A, r_0)} ||b - Ax||_2$, where $r_0 = b - Ax_0$ is the residual of the initial solution. The Arnoldi process applied to $K_m(A, r_0)$ builds $V_{m+1} = [V_m, v_{m+1}]$, an orthonormal basis of $K_m(A, r_0)$, the m+1 by m matrix $\overline{H}_m$ and $\beta = \|r_0\|_2$. These matrices satisfy the relation $AV_m = V_{m+1}\overline{H}_m$. The

iterate x_m can be written as $x_m = x_0 + V_m y_m$, where $y_m \in \Re^m$ is the solution of the least squares problem: $minimize_{y \in \Re^m} \left\| \beta e_1 - \overline{H}_m y \right\|_2$.

In the GMRES algorithm the number of vectors requiring storage increases with m. In order to limit both computation and memory requirements, a restarted version is often used.

In the algorithm, x_0 denotes an initial guess of the solution, m denotes the size of Krylov subspaces, and ε denotes the tolerance.

Algorithm restarted GMRES(m):

1. Initialization
2. $r_0 = b - Ax_0$
3. Apply Arnoldi process to $K_m(A, r_0)$
4. $y_m = \underset{y \in \Re^m}{arg\,min} \| \beta e_1 - \overline{H}_m y \|_2$

   $x_m = x_0 + V_m y_m$, $r_m = b - Ax_m$
5. if $\| r_m \|_2 \leq \varepsilon$ then stop

   else

   $x_0 = x_m$, $r_0 = r_m$

   goto step3

   end if

2.2 The hybrid Algorithm GMRES(m)/LS(k,l)

The whole process is that we calculate in parallel some eigenvalues by the Arnoldi method [5]. As they will be approximated with a sufficient accuracy, eigenvalues are used to perform some iterations of the Least Squares method [6] in order to obtain a new initial vector for the next GMRES iterations.

The hybrid algorithm GMRES(m)/LS(k,l) can be given as follows. There are some important parameters, m' denotes the size of Krylov subspace for Arnoldi method, k denotes the degree of the least squares polynomial, and l denotes the number of the successive applications of the Least Squares method.

Algorithm: GMRES(m)/LS(k,l)

1. Initialization
2. Compute x_m, the m^{th} iterate of GMRES starting with x_0

   if $\| b - Ax_m \|_2 < \varepsilon$ then Stop

   else

   $x_0 = x_m$, $r_0 = b - Ax_0$

   end if

2′. Perform m' iterations of the Arnoldi process starting with r_0, to $K_{m'}(A, v)$, and compute the eigenvalues of $H_{m'}$

2''. Compute the least squares polynomial P_k on the boundary of H, the hull convex enclosing all computed eigenvalues.

3. Do $j = 1,\cdots,l$

$\tilde{x} = x_0 + P_k(A)r_0$

$x_0 = \tilde{x}$, $r_0 = b - Ax_0$

end do

4. if $||r_0||_2 < \varepsilon$ then Stop

else goto step 2

end if

In the algorithm, step 2' and 2'' means that these two steps are performed independently of the GMRES iterations, and the step 2'' is performed following the step 2'.

Step 2' is the Arnoldi process. At first we apply Arnoldi process to the krylov subspaces of Arnldi $K_m(A,v)$. Then we calculate the eigenvalues(λ_i, $1 \le i \le d$) and the associate eigenvectors (y_i, $1 \le i \le d$) of H_m. After that we compute the Ritz vectors $u_i = V_m y_i$, for i=1, ···, d. Set $v = \sum_{i=1}^{d} Re(u_i)$, and repeat the process above until $\max_{i=1}^{d} |\rho_i| < \varepsilon$, where $\rho_i = \|\lambda_i u_i - Au_i\|_2$, $1 \le i \le d$.

Step 2'' is the sequential part of Least Square method. For the Least Square method, it can be written as follows: $\tilde{x} = x_0 + P_k(A)r_0$ where x_0 an initial approximation, r_0 its residual, and P_k is a polynomial of degree k-1. Let P_k^1 be the set of the real polynomials p of degree k, such that p(0)=1, and define the polynomial $R_k \in P_k^1$ by $R_k(z) = 1 - zP_k(z)$. Then the residual of the iterate $\tilde{x}$ is $\tilde{r} = R_k(A)r_0$.

In general, we do not have the whole spectrum of A, but only some eigenvalue estimates contained in a convex hull H because all eigenvalue calculation will spend a very long time. H is constructed such as it does not contain the origin. Smolarski and Saylor [11] proposed to find R_k minimizing a weighted L_2-norm on the space of real polynomials, with a suitable weight function w, defined on the boundary of H. We obtain the following least squares problem $\min_{R_k \in P_k^1} \|R_k\|_w$.

The obtained polynomial $P_k = \sum_{i=0}^{k-1} \eta_i t_i$ is expressed in the scaled and shifted Chebyshev basis defined by $t_j(\lambda) = T_j\left(\frac{\lambda - c}{d}\right) / T_j\left(\frac{a}{d}\right)$ j=0,1,… This is the best basis of polynomials on the ellipse $\varepsilon(c,d,a)$ of smallest area enclosing H (see [10] and [5] for an algorithm computing this optimal ellipse). For more details, see [6].

2.3 The hybrid Method for Complex Problem

In fact, we realize the solution by the prior treatment for the complex matrix. The complex problem $(Ar + Ai) \times (Xr + Xi) = (br + bi)$ can be split into real part and image part:

$$Ar \times Xr - Ai \times Xi = br$$
$$Ar \times Xi + Ai \times Xr = bi$$

So we extend the complex matrix (size N*N) into a real counterpart (size 2N*2N) [7].

Ar	-Ai		Xr		br
		×		=	
Ai	Ar		Xi		bi
A’			X’		b’

We apply the hybrid method to this new system $A' \times X' = b'$.

3 Implementation on GRID System

Grid’5000 is a Nation Wide Grid environment that is composed of many clusters distributed in 9 computing centers in France. A fast dedicated network interconnects those clusters. It is a highly configurable, controllable and monitorable instrument that can be configured to work as a real Grid. We implement our experiments on Grid5000 because it isolates the perturbations from outside, par example the communication over the Internet and the load of the computing devices. We can devote ourselves to research the algorithm itself and it would help us improve our analysis of more tests on the worldwide platforms in future.

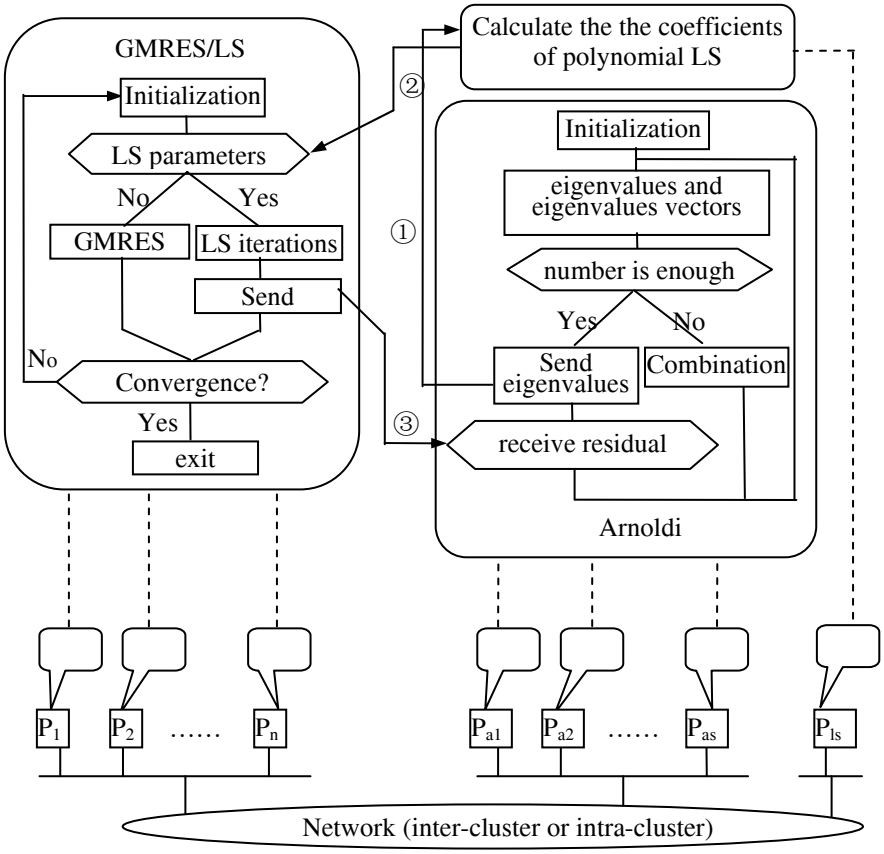
The Grid5000 usage is based on a reservation policy and a deployment mechanism allowing people configuring their own environment. Details can be found on the Grid5000 website [16]. We distribute our application on one or several sites of Grid5000 with the environment MPI.

We reserve most processors to run the algorithm GMRES(m) by the way of the SPMD model, where one act as an administrative process and the other p identical calculation processes play the role of workers. The calculation processors read directly their own data and execute the method GMRES(m), communicating with their brother processes.

The processors dedicated to the parallel package “PARPACK” are in charge of the residuals reception, the Arnoldi projection and the eigenvalues calculation, performing independently of the processes GMRES.

Only one processor is in charge of the sequential part because of the small set of data for calculation. The parameters “Least Square” obtained are then sent to the processors executing the parallel part of LS method and algorithm GMRES(m) .

The whole process and the relationship of the communication between the three parts are presented in Fig. 1.



① – eigenvalues, ② – parameter c,a,d, P_{LS} ,③ – residual

Fig.1. General scheme of asynchronous hybrid GMRES/LS-Arnoldi process

There are two threads for the whole calculation. The first is the GMRES iteration or Least Square iteration. After each iteration, the GMRES(m) process always checks if the LS parameters arrive. In this case, the GMRES algorithm is then suspend, and the processes perform the parallel part of the LS hybridizations. Then GMRES(m) restart with the obtained initial vector.

Another thread is the eigenvalues calculation by Arnoldi method and the coefficients computation by sequential part of Least Square method. These two processes are performed in serial.

All the processors are interconnected inter-cluster or intra-cluster. Intra-cluster means that the algorithm is performed in one cluster. In other word, the GMRES iteration, the Arnoldi process and LS method are distributed in the nodes of the same cluster. Inter-cluster means that these three components are distributed in different

clusters, and each component can be assigned completely in one cluster or be distributed in different clusters.

4 Numeric Results and Analysis

Table 1 shows the excellent network configuration of our experimental platform. The detailed information can be referred to the web site of Grid'5000 [16].

Table 1. Bandwidth and average latency between the clusters of the sites.

	Bandwidth	Latency
Inter-Cluster		
Orsay	48.4MB/s	0.11ms
nancy	42.8MB/s	0.09ms
Bordeaux	53.7MB/s	0.086ms
Intra-Cluster		
Orsay – nancy	9.7MB/s	5.7ms
Orsay – Bordeaux	8.1MB/s	7.9ms
Nancy – Bordeaux	4.0MB/s	17ms

All the sparse matrices are stocked in the compressed format CSR (Compress Sparse Row) for saving the memory and reducing the communication on the network. Moreover, in order to be able to verify the results accuracy, we have chosen in all examples the right-hand side so that the solution of the system is $x = (1,1,...,1)^T$. The iteration starts with $x_0 = (0,0,...,0)^T$.

First example (af23560): We experiment some industrial matrices from the site MatrixMarket. In this paper, we present the results obtained with the matrix af23560 (size 23560*23560, 484256 nonzero elements)

Second example: are created by a generator and are block diagonal matrices. (size 17000*17000, 426260 nonzero elements)

Third example: symmetric complex matrix young1c (size 841*841, 4089 nonzero elements). It is from the site MatrixMarket.

Forth example: symmetric complex matrix dwg961b (size 961*961, 10591 nonzero elements).

4.1 Advantages of the Hybrid Method

We can see the first advantage from Fig 2: the high degree of parallelism. GMRES method is a compute-intensive and data parallelism application. During the parallel GMRES processes, there are intensive communication and multiple synchronizations. So the parallelism degree is limited. It can't be increased easily. The more processors involve, the more communication spend and the slower the convergence is. In the hybrid method, we add the task parallelism by the participation of Arnoldi process and Least Square method. We use more processors and we accelerate the convergence. In this example, the classic gmres method has an optimal

number of processors, 26, and the optimal number for the hybrid method is 34. And we can also remark that the hybrid method spent less time.

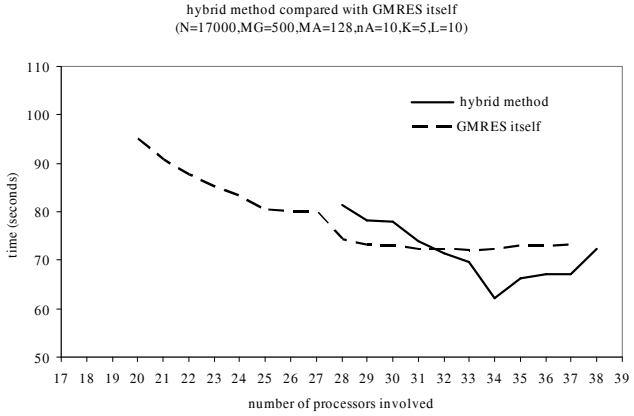


Fig. 2. The comparison of hybrid method and GMRES itself of the matrix (N=17000) with the number of processors

The second advantage is the obvious speed up of convergence. The convergence with the hybrid method can be faster, even when it is difficult by using the classic GMRES method (Fig. 3). In this example, we choose a relatively small size of krylov subspace. We can notice a stagnation of convergence for the classic restarted GMRES method. However, for the hybrid method, despite the appearance of many peaks, it converges.

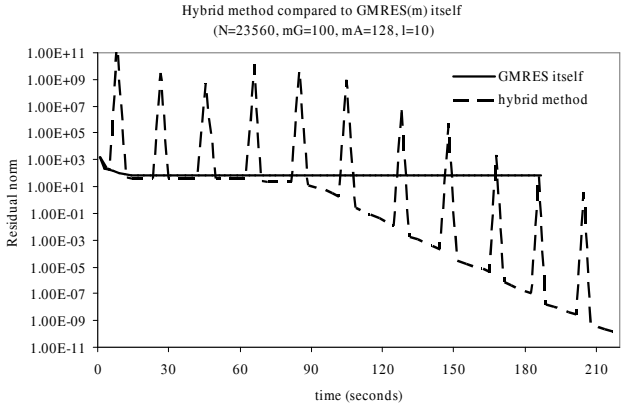


Fig. 3. The comparison of hybrid method and GMRES itself of the matrix (N=23560) in condition of the difficult convergence

The third advantage accompanies the second advantage: low requirement of memory. The use of the restarted GMRES version is because it can limit both computation and memory requirements. The bigger the size of Krylov subspace is, the better the convergence is. However, the bigger size means the more memory requirement. Using the hybrid method, we can realize the convergence by the smaller size of Krylov subspace.

4.2 Characteristics of the Hybrid Method

As the mention in section 4.1, we can obverse some peaks during the convergence of hybrid method (Fig. 3). These peaks appear when the process of hybridization occurs. The residual increases sharply temporary, however a sharper deduce follows. Overall, the convergence is achieved. Thereby, too many peaks will damage the efficiency, as each hybridization influences many GMRES iterations. When the peaks are high and nearby, divergence may even occur.

Additional, we combine the Arnoldi and Least Square method with the restarted GMRES in order to accelerate the convergence. Although the computation and communication increases, in fact there is almost no influence for the whole performance because these two methods are performed in parallel and the communication between different components is relatively little. Most of their computation time and their communication time can be overlapped. Table 2 illustrates this characteristic.

The symbols (1), (2), (3) correspond to the condition 1, 2, 3.

Condition1: GMRES on Nancy, Anoldi on Nancy, and Least Square on Bordeaux.

Condition2: GMRES on Nancy, Arnoldi on Orsay, and Least Square on Bordeaux

Condition3: GMRES on Nancy and Orsay, Arnoldi on Nancy and Orsay, and Least Square on Bordeaux

The term com1 denotes the communication for GMRES computing like exchange the data with their brother processors. The term com2 denotes the communication between the components, like eigenvalues, LS parameters.

Table 2. Bandwidth and average latency between the clusters of the sites.

Time(s)	Intra-cluster (Nancy)	Inter-cluster (Nancy+Orsay)		
		Distribution Inter-component		Distribution Intra-component
		(1)	(2)	(3)
Total time	70.73	71.57	72.69	120.8
Computing	58	58.224	58.358	65.54
Com1	12.42	13.319	14.132	54.42
Com2	0.009	0.024	0.055	0.084
Iteration	10	10	10	10

(2) represents the condition that three different components of the algorithm are distributed respectively in three different clusters. It is obviously that this distribution strategy can't bring much more burden for the communication.

4.3 Complex Problems

Due to the incompatible of the part of hybridization algorithm for the complex elements, the hybrid method is always used to solve the real linear systems. Now we apply the hybrid method on the prior transformed system.

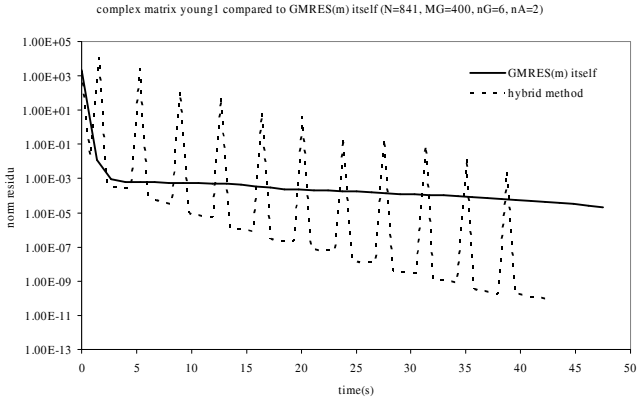


Fig. 4. The comparison of hybrid method and GMRES itself for the complex matrix (N=841)

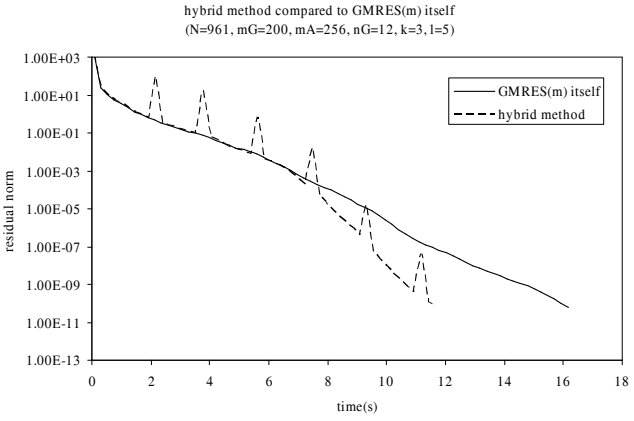


Fig. 5. The comparison of hybrid method and GMRES itself for the complex matrix (N=961)

From the experimental results (Fig.4, Fig.5), we can notice that the hybrid method also shows its advantages and characteristic for the solution of complex problem.

5 Conclusion

We implemented our algorithm for the hybrid method GMRES(m)/LS-Arnoldi on Grid computing platforms: Grid'5000 with the environment MPI, and applied it to the real problems and the complex problems.

From the experimental results, we sum up the advantages and characteristics of the hybrid method. We obtain very important convergence accelerations, and increase the degree of parallelism.

In future, we will try more complex problems, and extend our method to the scientific problems of larger size.

Moreover, the hybrid method can be improved in many places. We think that it's better to change the parameters dynamically during the solution of the problem. For example, we can decide whether or not to proceed the hybridization of LS according to the speed of the convergence. And we can change the size of Krylov subspaces of Arnoldi after each LS hybridization to obtain the more important eigenvalues for the next hybridization.

In addition, we will do some tests on the other supercomputers or cluster (i.e. Tsubame in Japon, IBM cell in France) to see the performances. And we will analyze the energy consumption of every component to optimize the implementation in the grid environment.

References

1. Cunha, R.D.Da, Hopkins, T.: A Parallel Implementation of the Restarted GMRES Iterative Algorithm for Nonsymmetric Systems of Linear Equations. In: Advances in Computational Mathematics, vol. 2, pp. 261--277. Springer (1994)
2. Saad, Y., Schultz, M.H.: GMRES: A Generalized GMRES Algorithm for Solving Nonsymmetric Linear Systems. J. SIAM. Sci. Statist. Compt. 7, 856--869 (1986)
3. Edjlali, G., Petiton, S., Emad, N.: Interleaved Parallel Hybrid Arnoldi Method for a Parallel Machine and a Network of Workstations. In: Conference on Information, Systems, Analysis and Synthesis (ISAS'96), pp. 22--26. Orlando (1994)
4. Saad, Y.: Variations on Arnoldi's Method for Computing Eigenelements of Large Unsymmetric Matrices. Linear Algebra Appl. 34, 269--295 (1980)
5. Saad, Y.: Numerical Methods for Large Eigenvalue Problems. Manchester University Press, Manchester (1992)
6. Saad, Y.: Least Squares Polynomials in the Complex Plane and their Use for Solving Nonsymmetric Linear Systems. J. SIAM. Sci. Statist. Compt. 7, 155--169 (1987)
7. YinTsung, H., WeiDa, C.: A Low Complexity Complex QR Factorization Design for Signal Detection in MIMO OFDM Systems. In: International Symposium on Circuits and Systems, pp. 932--935. IEEE Press, Seattle (2008)
8. Baker, A.H., Jessup, E.R., Manteuffel, T.: A Technique for Accelerating the Convergence of Restarted GMRES. J. SIAM. MATRIX. Anal. Appl. 26, 962--984 (2005)
9. Essai, A., Bergere, G., Petiton, S.: Heterogeneous Parallel Hybrid GMRES/LS-Arnold Method. In: Ninth SIAM Conference on Parallel Processing for Scientific Computing, Texas (1999)
10. Manteuffel, T.A.: The Tchebychev Iteration for Nonsymmetric Linear Systems. J.Numer. Math. 28, 307--327 (1997)

11. Smolarski, D.C., Saylor, P.E.: An Optimum Iterative Method for Solving any Linear System with a Square Matrix. J. BIT. 28, 163--178 (1988)
12. Georgiou, Y., Richard, O., Neyron, P., Huard, G., Martin, C.: A Batch Scheduler with High Level Components. In: 5th International Symposium on Cluster Computing and the Grid, pp. 2, 776--783. IEEE Press, Cardiff (2005)
13. Franck, C., Eddy, C., Michel, D., Frederic, D., Yvon, J.: Grid'5000: a Large Scale and Highly Reconfigurable Grid Experimental Testbed. International Journal of High Performance Computing Applications. 20, 481--494 (2006)
14. Benjamin, Q., Franck, C.: A Survey of Grid Research Tools: Simulators, Emulators and Real Life Platform. In: 17th IMACS World Congress Scientific Computation, Applied Mathematics and Simulation, Paris, France (2005)
15. Ye, Z., Bergere, G., Petiton, S.: A Parallel Hybrid Method of GMRES on GRID System, IEEE International Parallel and Distributed Processing Symposium, pp. 356. IEEE Press, California (2007)
16. Grid'5000, http://www.grid5000.fr