Performance of Eigenvalues and Eigenvectors Solutions of Complex Hermitian Matrices on GPU Accelerator

Edita Gichunts

Institute for Informatics and
Automation Problems of NAS
RA
e-mail: editagich@ipia.sci.am

ABSTRACT

Solutions of eigenvalues and eigenvectors of complex Hermitian matrices are widespread and have a very important role in scientific calculations. These solutions can be obtained from linear algebra libraries by the functions of Lapack and by its parallel version ScaLapack in systems with general and distributed memory, respectively. However, it's more beneficial to get these solutions in hybrid architectures which require a new development of algorithms to efficiently organize non-uniformity and massive parallelization in a graphical processor. The main objective of this paper is to present the performance of standard, as well as generalized case of solutions of eigenvalues and eigenvectors of complex and double complex Hermitian matrices on Tesla C1060 accelerator.

Keywords

GPU accelerator; symmetric tridiagonal reduction; eigenvalue problems; hybrid architectures; Householder transformations, one-stage, two-stage, tridiagonal

1. INTRODUCTION

Calculations of eigenvalue problems are usually in great demand. Scientific calculation expressions, which are transmitted through the environment, such as earthquakes, a building vibration, the energy levels of electrons in nanostructured compounds, require solutions of eigenvalue problems. Solutions of eigenvalue problems are also required in quantum chemistry and atomic physics.

Subprograms of the package Lapack of linear algebra are intended for systems with general memory. MAGMA package expands Lapack for heterogeneous architecture, and its performance for modeling is an important value using the computing capacity of such architecture. To use the heterogeneous architecture, MAGMA is based on a hybrid programming paradigm and static arrangement of the scheme, in other words the algorithm first is divided into smaller computational problems that afterwards are statically arranged in a graphical processor.

This work focuses on one node of GPU-CPU hybrid architecture. In this work the following form of Hermitian matrix eigenproblem is studied: $Az = \lambda z$ is standard and $Az = \lambda Bz$ is generalized, where A is a Hermitian matrix and B is Hermitian positively defined.

The sizes of matrices here are several thousands and there is no need to solve this problem in large distributed memory systems. That's why we develop algorithms that will effectively scale the massive parallelization in general memory systems, particularly in GPU-CPU hybrid multiprocessor systems.

The eigensolutions obtained in this work are important results in high performance calculations which provide protection for hybrid architectures and sufficiently increase the productivity.

2. RELATED WORK

Linear algebra libraries Lapack [1] and ScaLapack [2] contain a collection of eigensolutions procedures for general and distributed memory systems, respectively.

Hermitian eigenproblem solution is implemented through the following main steps. Using the Householder transformation, the matrix is reduced to a tridiagonal form. The acceleration of the eigenproblem solution depends on the acceleration of reduction to a tridiagonal form.

The standard step of reducing to a tridiagonal form in Lapack is a "single phase" (also known as "one-stage") [3]. The two-stage approach is actual where the matrices are first transformed into a band form and second they are reduced to a final tridiagonal form.

For two-sided analysis Tomov et al. [4] have presented algorithms of a new hybrid transformation which complete the graphic processor high performance taking advantage of GPU high throughput capacity.

Bientinesi et al. [5] have accelerated the two-stage approach through Successive Band Reduction (SBR) tools. SBR [6] tools used two-sided orthogonal transformations based on Householder reflectors and successively reduced the matrix bandwidth size until a suitable width was reached.

Making use of subprograms of MAGMA set, the task of eigenvalue problem was run for the complex Hermitian matrices on GPU graphic processor. Programs have been run to find complex and double complex eigenvalues and eigenvectors for $Az = \lambda z$ and $Az = \lambda Bz$ eigensolutions, as well as performance was obtained when the reduction of matrices to tridiagonal form was realized in case of one-stage and two-stage analyses.

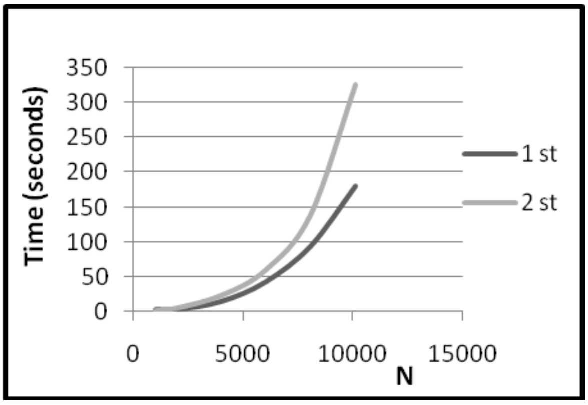
3. HYBRID STANDARD EIGENSOLUTIONS

To solve the eigensolutions of the Hermitian problem of the following form $Az = \lambda z$, Λ eigenvalues and Z eigenvectors should be found so, that $A=Z \Lambda Z^H$, where ^{H} is the conjugate-transpose. The standard algorithm consists of three steps [7], [8]. First, by the Q orthogonal transformation the matrix is reduced to a tridiagonal form called "a reduction phase", so that $A=QTQ^H$, where T is a tridiagonal matrix. Second, the eigencouples of tridiagonal matrix are calculated (Λ, E), called "a solution phase". Third, the eigenvectors of the tridiagonal matrix should be brought back to the eigenvectors of the initial matrix Z=QE called "a back transformation phase". It is well known that the first phase - "the reduction phase", depending on some processes, takes much more time than the other two stages. Therefore, there are several approaches for the reduction of Hermitian matrix to tridiagonal form. Among the existing approaches are one-stage and two-stage analyses. The standard one-stage approach of reducing to a tridiagonal form is used in eigensolutions of LAPACK library [9]. Now it has been developed for hybrid architecture and is also used in MAGMA package [10]. It performs the function xHEEVD, where x can be as c and z which are for complex and double complex values, correspondingly. MAGMA also contains the function xHEEVDX which finds not only all eigensolutions but also the eigensolutions in the mentioned range. For the generalized form of eigensolutions they are the functions xHEGVD and xHEGVDX, correspondingly. In these functions the reduction of the symmetric matrix to the tridiagonal form is realized through xHETRD subprogram.

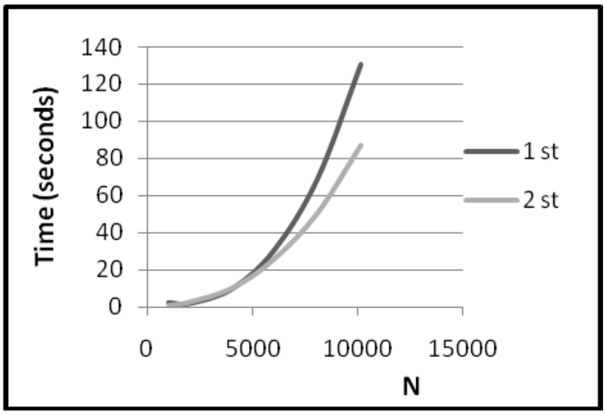
In case of two-stage approach of eigensolutions the matrix is reduced to a band matrix in the first step, and in the second step the band matrix is reduced to a tridiagonal form using a bulge chasing technique [11]. In case of two-stage approach xheevdx_2stage and xhegvdx_2stage subprograms of MAGMA package are used correspondingly for Az = λz standard and Az = λBz generalized solutions of both forms. During this hybrid development the xHEGST subprogram transforms the general problem into a standard one.

Divide and Conquer algorithm is used in all programs to find the eigensolutions. Standard Divide and Conquer algorithm calculates all the eigenvalues and eigenvectors of the tridiagonal matrix [12].

4. RESULTS OF STANDARD EIGENSOLUTIONS

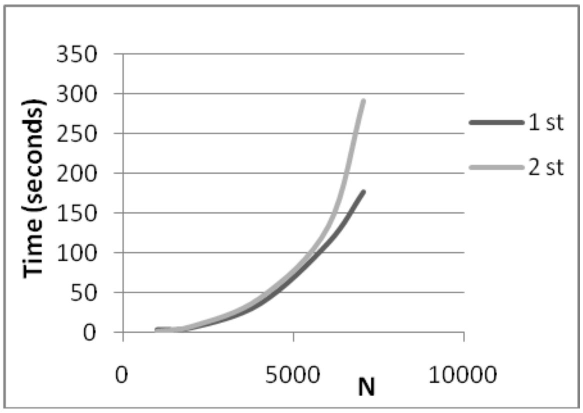


(a) all eigenvectors

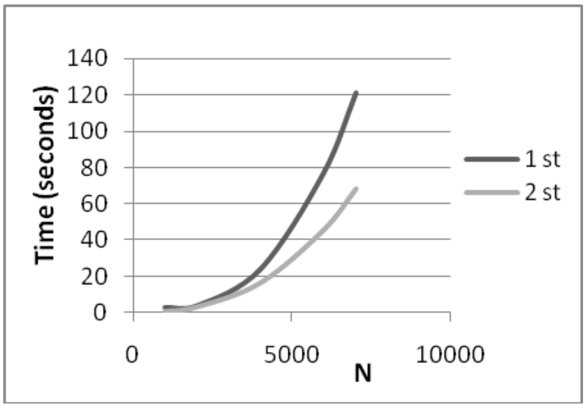


(b) 10% eigenvectors

Fig. 1. Complex precision



(c) all eigenvectors



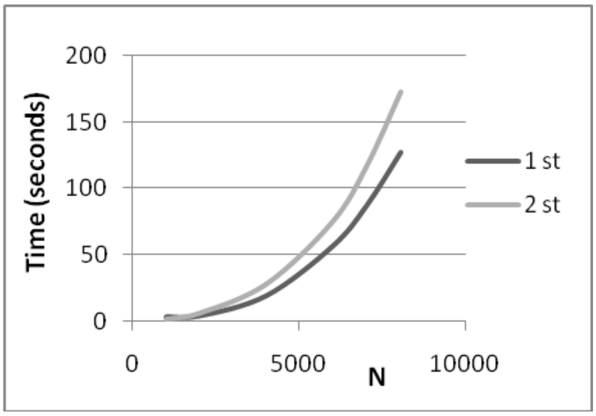
(d) 10% eigenvectors

Fig. 2. Double complex precision

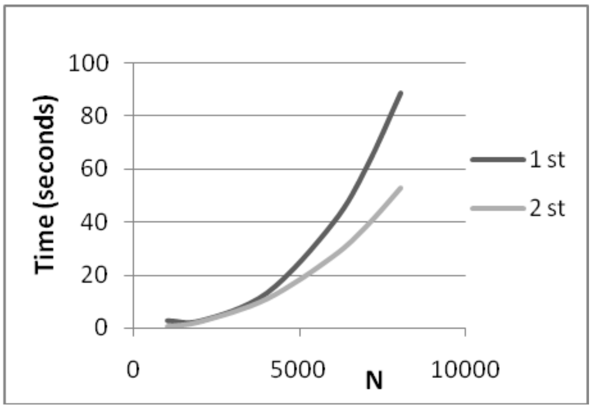
Figures 1 and 2 show the results of Az = λz standard eigensolutions for various sizes of complex precision and double complex precision matrices, correspondingly. Given the graphs of one-stage and two-stage approaches of the problem solution. During the problem solution much more time takes the reduction of the matrix to a tridiagonal form. Note that in case of one-stage the matrix is immediately reduced to a tridiagonal form through xHETRD subprogram, then it finds the solutions, while in case of two-stage the matrix first is reduced to a banded form using BLAS 3, then the banded matrix is reduced to a tridiagonal form using the so-called bulge chasing technique; and after finding the solutions a back transformation of eigenvectors is carried out. The reduction of the matrix to a tridiagonal form in case of two-stage is twice faster than in one-stage case.

However, if all the vectors are required to be found during the process of finding the eigenvectors of two-stage, then the problem operation slows down by virtue of back transformation of vectors. Since the global memory of Tesla C1060 is 4Gb, and the matrices are completely transferred to GPU, and as in case of standard solution one matrix is entered, hence, in case of complex values a 10112 size-matrix is entered, while in case of double complex – 7040 size matrix.

5. RESULTS OF GENERALIZED EIGENSOLUTIONS

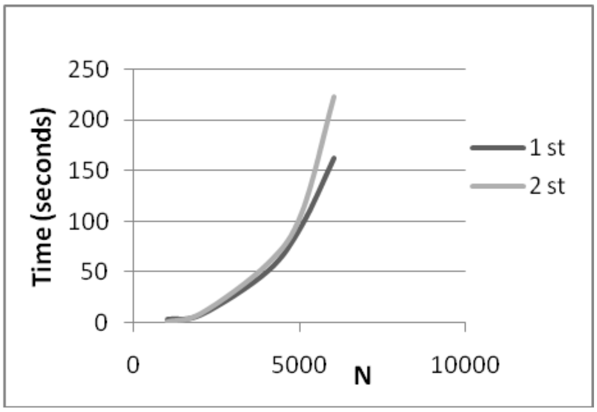


(a) all eigenvectors

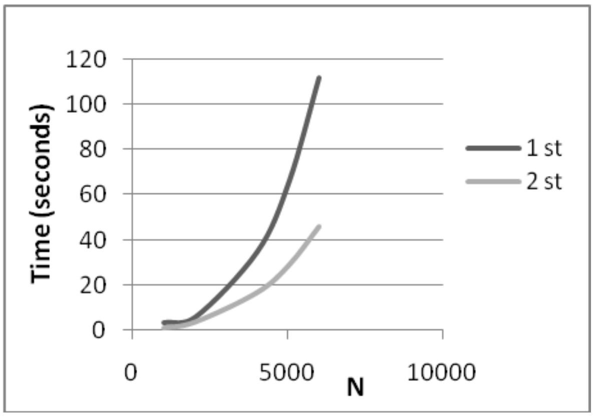


(b) 10% eigenvectors

Fig. 3. complex precision



(c) all eigenvectors



(d) 10% eigenvectors

Fig. 4 double complex precision

Figures 3 and 4 show the results of Az = λBz generalized Hermitian eigensolutions correspondingly for complex precision and double complex precision matrices of various sizes. There have also been observed one-stage and two-stage approaches of reducing to a tridiagonal form. In case of Az = λBz generalized solutions first of all xPOTRF analysis of Kholetski is made, afterwards in order to transform the generalized problem into a standard one, xHEGST and xTRSMM transformations are applied. In case of Az = λ Bz generalized solutions the same result is obtained as in case of standard solutions, but as two matrices A and B are to be entered, therefore, in case of complex the sizes of the entered matrices should be 8064-dimensional and in case of double complex - 6016-dimensional based on GPU global memory.

6. CONCLUSION

Performance of eigenvalues and eigenvectors solutions of Hermitian matrices on Tesla C1060 accelerator is presented in this article. The problem was observed for both forms of Az = λz standard and Az = λBz generalized solutions. One-stage and two-stage approaches of reduction to tridiagonal form of a matrix have been presented for them. The results show that if one needs to find all the eigenvectors despite the fact that the approach of reducing to a tridiagonal form is twice faster, nevertheless, the two-stage approach of problem solution concedes to the one-stage approach as the back transformation of eigenvectors takes considerable time. In this case the problem solution time, depending on the increase of the matrix sizes, grow at 10%-20%. And if less eigenvectors are required to be found in the problem of eigensolutions, e.g., 10% - 50%, then the two-stage approach is more effective than the one-stage approach as in case of double complex it is twice faster and in case of complex – 1,5 times, therefore, much more time is saved for solving eigenproblems.

REFERENCES

[1] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, LAPACK Users' Guide. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1999.

[2] L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley,

ScaLAPACK Users' Guide. PA: Society for Industrial and Applied Mathematics, 1997.
[3] E. Anderson, Z. Bai, C. Bischof, S. L. Blackford, J. W. Demmel, J. J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. C. Sorensen, LAPACK User's Guide, 3rd ed. Philadelphia: Society for Industrial and Applied Mathematics,1999.
[4] S. Tomov, R. Nath, and J. Dongarra, "Accelerating the reduction to upper Hessenberg, tridiagonal, and bidiagonal forms through hybrid GPU-based computing," Parallel Comput., vol. 36, no. 12, pp. 645–654, 2010.
[5] P. Bientinesi, F. D. Igual, D. Kressner, and E. S. Quintana-Ortı, "Reduction to condensed forms for symmetric eigenvalue problems on multi-core architectures," ser. PPAM'09. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 387–395.
[6] C. H. Bischof, B. Lang, and X. Sun, "Algorithm 807: The SBR Toolbox—software for successive band reduction," ACM Transactions on Mathematical Software, vol. 26, no. 4, pp. 602–616, 2000.
[7] J. O. Aasen, "On the reduction of a symmetric matrix to tridiagonal form," BIT, vol. 11, pp. 233–242, 1971.
[8] B. N. Parlett, The Symmetric Eigenvalue Problem. Englewood Cliffs, NJ, USA: Prentice-Hall, 1980.
[9] E. Anderson, Z. Bai, C. Bischof, L. S. Blackford, J. W. Demmel, J. J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, LAPACK Users' Guide. Philadelphia, PA: SIAM, 1992, http://www.netlib.org/lapack/lug/.
[10] C. Vomel, S. Tomov, and J. Dongarra, "Divide and conquer on hybrid gpu-accelerated multicore systems," SIAM Journal on Scientific Computing, vol. 34, no. 2, pp. C70–C82, 2012.
[11] A. Haidar, H. Ltaief, and J. Dongarra, "Parallel reduction to condensed forms for symmetric eigenvalue problems using aggregated fine-grained and memory-aware kernels," in SC11: International Conference for High Performance Computing, Networking, Storage and Analysis, Seattle, WA, USA, November 12-18 2011.
[12] J. Demmel, Applied Numerical Linear Algebra. Society for Industrial and Applied Mathematics, 1997.