Vol.2017-MPS-114 No.9
2017/7/17

Performance Evaluation of a Combination of the Parallel Bisection Method and the Block Inverse Iteration Method with Reorthogonalization for Eigenvalue Problems on MIC processor

Sho Araki[1], **Hiroyuki Ishigami**[2], **Masayuki Osawa**[1], **Kinji Kimura**[1] **and Yoshimasa Nakamura**[1]
[1]Graduate School of Informatics, Kyoto University, Kyoto, Japan
[2]Yahoo Japan Corporation, Tokyo, Japan

Abstract— *We discuss the implementation, performance tuning, and evaluation of an eigensolver of real symmetric tridiagonal matrices using the bisection method and the block inverse iteration method with reorthogonalization on Intel Xeon Phi (Xeon Phi) many integrated core (MIC) processor. We develop an OpenMP thread parallel program for the eigensolver for Xeon Phi and experimentally determine the optimal block size parameter for both the MIC and CPU environments. Moreover, we perform experiments for evaluating the performance of the algorithm with the optimal block size. The eigensolver exhibits higher computation speed and accuracy in the MIC environment than the MRRR algorithm, which is known as the conventional high-speed eigensolver.*

Keywords: eigenpair problem; bisection method; block inverse iteration method with reorthogonalization; parallel computing; MIC processor;

1. Introduction

This study is mainly concerned with the standard eigenvalue problem for an $n \times n$ dense symmetric matrix A, as follows

$$Av_i = \lambda_i v_i, i = 1, \ldots, n,$$

where $\{\lambda_i\}_{i=1}^n (\lambda_1 \leq \cdots \leq \lambda_n)$ is the sequence of real eigenvalues of A. We refer to the pair of eigenvalue and eigenvector as "eigenpair" below.

There are many applications of the eigenvalue problem, since problems in computational science are often reduced to primal linear programming. Sometimes, only a small number of eigenpairs are required. Target matrices of eigenpair problems have become increasingly large, and the solvers need to be accelerated by means of parallel computing. Intel Xeon Phi is a parallel processor exhibiting higher watt performance than ordinary CPU (Central Processing Units) , and was very similar to GPU (Graphics Processing Units) until the previous "Knights Corner" generation. The largest difference between MIC (Many Integrated Core) architecture and GPU is the remarkably high compatibility of the former with CPU architecture, which is capable of compiling existing source code of C and Fortran languages without major modifications. In addition, "Knights Landing" generation processors, released in 2016, are mounted directly on CPU sockets. Therefore, the bandwidth limitation of PCI-Express is no longer an issue. Moreover, the higher compatibility of the updated Xeon Phi with ordinary CPU widens the range of application of MIC architecture in numerical calculations. The solver of eigenpair problems considered in this paper is an example.

The computation of the eigenpairs of a real symmetric matrix A generally employs the transformation of A to a symmetric tridiagonal or band matrix. Then, the eigenpairs of the tridiagonal or band matrix are computed. The eigenvalues of the transformed tridiagonal or band matrix are equal to the eigenvalues of the given matrix A, and the desired eigenvectors are obtained by the reverse transformation of the previous tridiagonal or band transformation. Many types of highly efficient parallel computing methods have been proposed for pre-processing and post-processing. Subsequently, one may concentrate on the eigenvalue computation part.

In this paper, we consider the bisection method [1] for obtaining the set (or a subset) of the eigenvalues of a pre-processed real symmetric tridiagonal matrix. It should be noted that there are several well-known methods for obtaining the eigenvectors of a real symmetric tridiagonal matrix, such as reverse iteration, MRRR (Multiple Relatively Robust Representation) [2], QR, and Divide and Conquer methods. MRRR can be used to compute all or a part of the eigenpairs of a given matrix. By contrast, the QR and Divide and Conquer methods can be used to compute all eigenpairs. The advantage of the QR method is accuracy in terms of absolute error measurement, whereas the advantage of the Divide and Conquer method is high computation speed in parallel environments. Even though the bisection method is suitable for the partial eigenvalue problem as well, we evaluate its performance in obtaining all eigenvalues, in order to make a comparison with the QR and Divide and Conquer algorithms. For parallel computing, Intel Math Kernel Library [3] (MKL) provides the routines for the methods.

2. Implementation of the Target Eigensolver

The computation of the eigenpairs of a real symmetric matrix is generally performed through the transformation of the target matrix A to a symmetric band matrix, and subsequent computation of the eigenpairs of the band matrix. The transformed symmetric band matrix, often a tridiagonal matrix, has the same eigenvalues as the original matrix. Then, the eigenvectors of the original matrix are obtained by the reverse

transformation of the previous band transformation. In this section, we briefly present the latter procedure of obtaining eigenpairs of symmetric tridiagonal matrix.

2.1 Implementation of the Bisection Method

Herein, we discuss the implementation of the bisection method for real symmetric tridiagonal matrices. The bisection method is an algorithm for computing the eigenvalues of a real symmetric matrix using binary search. It is proposed in [1], and its implementation for ordinary CPU is provided by the DSTEBZ routine of LAPACK (Linear Algebra PACKage) [4].

We adopt the implementation introduced in [8] as a thread parallel bisection method suitable for shared memory systems such as MIC environments.

2.2 Implementation of the Inverse Iteration Method

Herein, we briefly present the block inverse iteration method with reorthogonalization [5]–[8] for the eigenvector problem.

For an $n \times n$ real symmetric tridiagonal matrix T, let $\lambda_i \in \mathbb{R}(\lambda_1 < \cdots < \lambda_n)$ be its eigenvalues and $\boldsymbol{q}_i \in \mathbb{R}$ be the eigenvectors corresponding to λ_i. If $\tilde{\lambda}_i$ is the approximate value of λ_i, and the initial vector $\boldsymbol{v}_i^{(0)}$ is randomly (uniformly) generated, then the vector $\boldsymbol{v}_i^{(j)}$ converges to the eigenvector $\boldsymbol{q}_i$ as $j \to \infty$ in the following linear iterative equation

$$(T - \tilde{\lambda}_i I)\boldsymbol{v}_i^{(j)} = \boldsymbol{v}_i^{(j-1)}, j = 1, 2, \ldots, \quad (1)$$

where I is the $n \times n$ identity matrix. The inverse iteration method [5]–[7] is based on the above procedure. The complexity of this method for obtaining $m(< n)$ eigenvectors is $O(mn)$. Practically, the vector $\boldsymbol{v}_i^{(j)}$ must be normalized at each iteration to avoid overflow and underflow. The obtained eigenvectors are orthogonal if the eigenvalues of T are sufficiently separated. By contrast, it is known that the eigenvectors may fail to be orthogonal if the eigenvalues of T are clustered. In this case, it is proposed that eigenvectors corresponding to clustered eigenvalues should be reorthogonalized.

We adopt the block inverse iteration method with reorthogonalization proposed in [8] as the implementation of the inverse iteration method for MIC environment. This is a modification of the simultaneous inverse iteration method. The dominant part of the block inverse iteration method could be implemented, with efficient execution of matrix multiplications, on SSE and AVX enabled processors.

3. Performance Evaluation

We present the results of the numerical experiments that were conducted to evaluate the performance of the eigensolver using the parallel bisection method (PBi) and the block inverse iteration method with reorthogonalization (BIR). Specification of CPU and MIC environments are shown in Tables 1 and 2, respectively.

Test matrix for the evaluation is $n \times n$ symmetric tridiagonal matrices T, random matrix with uniformly distributed random numbers $d_i, e_i \in [0, 1]$.

Table 1: Specification of the experimental environment (CPU)

CPU	Intel Xeon E5-2695 v4 x2 (2.10GHz, 18 cores x2)
RAM	DDR4-2400 128GB
Compiler	icc 16.0.4, ifort 16.0.4
Options	-O3 -ipo -xCORE-AVX2 -fp-model precise -qopenmp -mkl
Software	Intel Math Kernel Library 11.3.4

Table 2: Specification of the experimental environment (MIC)

CPU	Intel Xeon Phi 7250 (1.4GHz, 68 cores)
RAM	DDR4-2133 96GB + MCDRAM 16GB (Cache mode)
Compiler	icc 16.0.4, ifort 16.0.4
Options	-O3 -ipo -xMIC-AVX512 -fp-model precise -qopenmp -mkl
Software	Intel Math Kernel Library 11.3.4

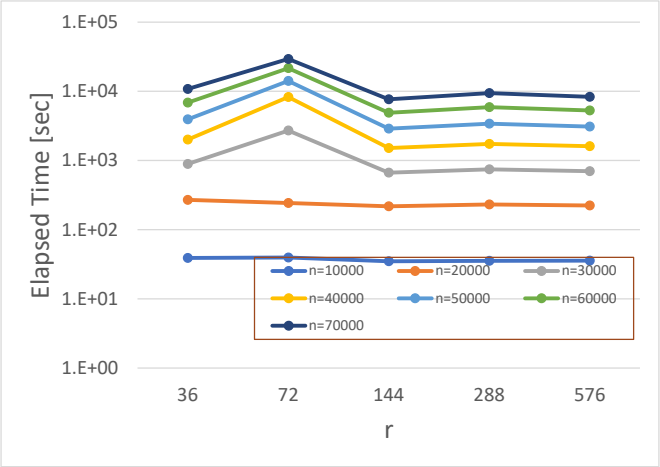

Fig. 1: Computation time for each block size (CPU, T)

3.1 Determining the Optimal Block Size for the Eigensolver

We first determine the optimal block size for the BIR algorithm. To this end, we use the following search method.

1) Set the initial block size r to the number of processor cores of each experimental environment.
2) Measure the computation time for obtaining all eigenvalues and eigenvectors of T.
3) If the computation time is longer than that of previous two trials, stop searching and let the block size giving minimum computation time be optimal.
4) Otherwise, set the block size $r := 2 \times r$ and continue searching.

Figures 1 and 2 show the computation time for T using the block inverse iteration method with reorthogonalization for different block sizes r on CPU and MIC environments for the matrix T of size n. It should be noted that all graphs are logarithmic.

We optimize the parameter for the BIR algorithm, since it was more than 100 times slower than the PBi algorithm. In addition, if we use hyper-threading technology, the PBi algorithm becomes faster, whereas the BIR algorithm becomes slower. Thus, hyper-threading technology is not adopted.

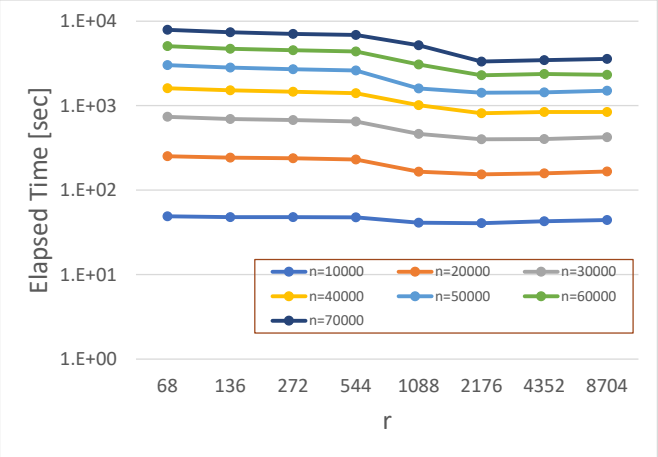

Fig. 2: Computation time for each block size (MIC, T)

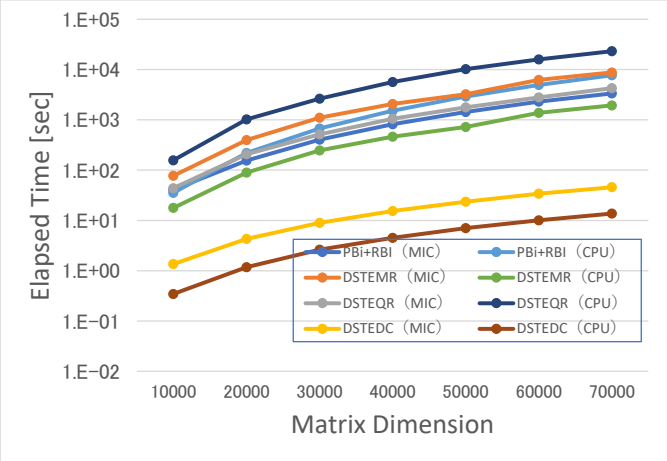

Fig. 3: Computation time for T

These graphs show that block size $r = 144$ for CPU environment and $r = 2176$ for MIC yield the best performance.

3.2 Comparison with Other Algorithms

We compare the execution time for obtaining all eigenvalues and eigenvectors of the matrix T using the PBi+BIR algorithm with the corresponding time for the MRRR algorithm. For further comparison, we add the QR [9] and Divide and Conquer algorithms [10], which are well-known methods. Because QR algorithm [9] and Divide and Conquer algorithm [10] are not able to compute partial eigenvalue and eigenvectors of target matrix. The implementations of these three algorithms are provided in Intel Math Kernel Library (MKL) [3]. In these experiments, we use DSTEMR, DSTEQR, and DSTEDC LAPACK routines provided by Intel MKL as the parallel implementations of the MRRR, QR, and Divide and Conquer algorithms, respectively. We note that the number of threads in all numerical experiments is set to be the number of processor cores.

Figure 3 shows the execution time for obtaining all eigenvalues and eigenvectors using each algorithm for $T(n = 10000, \ldots, 70000)$. It should be noted that the graph is semilogarithmic. The Divide and Conquer algorithm is the fastest. However, it cannot be adopted for partial eigenvalue and eigenvectors of the target matrix. PBi+BIR achieves higher performance than MRRR in MIC environment, as the graph shows.

Moreover, evaluations for accuracy are performed. Figure 4 shows the orthogonality $\|Q^{\mathsf{T}}Q - I\|_{\mathrm{F}}$ of eigenvectors obtained by each eigensolver for T, respectively (with $n = 10000, \ldots, 70000$). n denotes the dimension of the target matrix, $D = \mathrm{diag}\{\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n\}$, and $Q = [\boldsymbol{q}_1 \cdots \boldsymbol{q}_n]$. The graphs are semilogarithmic. The lines of PBi+BIR (MIC), DSTEMR(MIC), DSTEQR(MIC) and DSTEDC (MIC) are hidden behind the corresponding lines in CPU environment, since the results are nearly equal. The orthogonality among the eigenvectors obtained by the PBi+BIR algorithm is significantly smaller than that obtained by the other algorithms.

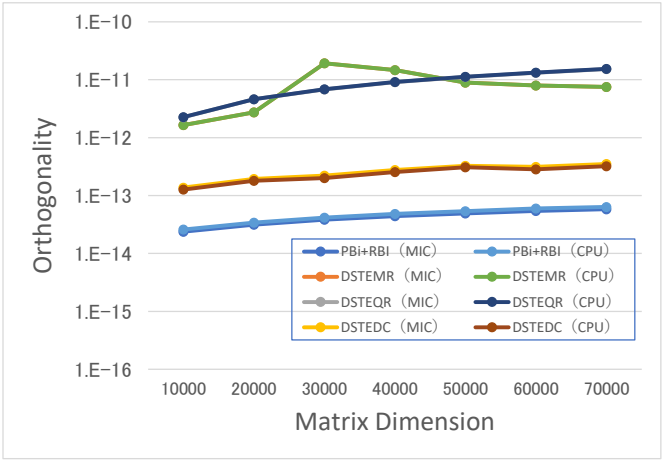

Fig. 4: Orthogonality of obtained eigenvectors for T_1

Figure 5 shows the results of the evaluation of decomposition accuracy, namely, the residuals $\|TQ - QD\|_{\mathrm{F}}$ of the decomposed eigenvalues and eigenvectors for T, respectively. PBi+BIR achieves higher accuracy than MRRR in terms of decomposition. In conclusion, in terms of both speed and accuracy, the PBi+BIR algorithm is superior to the MRRR algorithm in MIC environment. The high accuracy of PBi+BIR may be attributed to the high relative accuracy of the PBi algorithm.

4. Conclusion

We evaluated the performance of an eigensolver using a combination of the parallel bisection method and the block inverse iteration method with reorthogonalization (PBi+BIRalgorithm) for computing all eigenpairs of a real symmetric tridiagonal matrix. In particular, we compared the performance of PBi+BIR with that of MRRR, which can be adopted for obtaining a subset of eigenpairs. From the preliminary experiments, we determined the optimal block size for the block inverse iteration method with reorthogonalization. We measured the computation time of PBi+BIR, for each block

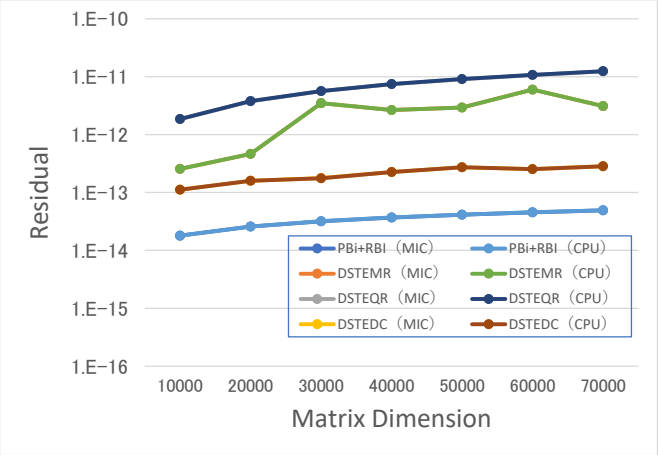

Fig. 5: Residuals of decomposition for T

of size $r = 2^m \times C$, where C is the number of processor cores, and let the block size yielding the minimum computation time be the optimal.

For T_1, PBi+BIR with the optimal block size parameter obtains the eigenpairs faster than MRRR in MIC environment. This is true for T_2 as well, with the condition that size of the target matrix is $n \geq 60000$. In addition, PBi+BIR is the only algorithm computing eigenpairs faster in MIC environment than in CPU environment. It is conceivable the AVX512 instruction set contributes to the superior performance of PBi+BIR in MIC environment, as this algorithm involves a large number of matrix multiplications. This suggests that PBi+BIR is suitable for MIC environment. Moreover, in a comparison of orthogonality of eigenvectors and residuals of eigenpairs, PBi+BIR achieves higher accuracy than MRRR, and there is no trade-off between computation speed and accuracy in MIC environment.

There are many applications of eigensolvers for real symmetric matrices such as kernel principal component analysis, which frequently appears in the industrial field. Fast and accurate eigensolvers are in great demand in this field, and PBi+BIR in MIC environment may be the eigensolver of choice. In future work, we would perform a more detailed examination of the relation between computation time and block size, and establish a method for auto-tuning of the algorithm in MIC environment.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 17H02858.

References

[1] J. Wilkinson, “Calculation of the eigenvalues of a symmetric tridiagonal matrix by the method of bisection”, *Numer. Math.*, vol. 4, no. 1, pp. 362 367, 1962.
[2] I. S. Dhillon, B. N. Parlett, and C. Vömel, “The design and implementation of the MRRR algorithm”, *ACM Trans. Math. Softw.*, vol. 32, no. 4, pp. 533–560, 2006.
[3] Intel Math Kernel Library, “Available electronically at https://software.intel.com/en-us/intel-mkl/.”
[4] E. Anderson, Z. Bai, C. Bischof, L. S. Blackford, J. W. Demmel, J. Dongarra, J. Du Croz, S. Hammarling, A. Greenbaum, A. McKenney, and D. Sorensen, *LAPACK Users Guide (Third ed.)*. Philadelphia, PA, USA: SIAM, 1999.
[5] G. H. Golub and C. F. van Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.
[6] J. W. Demmel, *Applied Numerical Linear Algebra*. Philadelphia, PA, USA: SIAM, 1997.
[7] B. N. Parlett, *The Symmetric Eigenvalue Problem*. Philadelphia, PA, USA: SIAM, 1998.
[8] H. Ishigami, K. Kimura, Y. Nakamura, “A New Parallel Symmetric Tridiagonal Eigensolver Based on Bisection and Inverse Iteration Algorithms for Shared-memory Multi-core Processors”, *2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)*, pp. 216–213, 2015.
[9] W. Kahan, “Accurate eigenvalues of a symmetric tridiagonal matrix”, *Technical Report*, Computer Science Dept. Stanford University, no. CS41, 1966.
[10] M. Gu and S. C. Eisenstat, “A divide-and-conquer algorithm for the symmetric tridiagonal eigenproblem”, *SIAM J. Matrix Anal. Appl.*, vol. 16, no. 1, pp. 172–191, 1995.