Krylov Subspace Method with Communication Avoiding Technique for Linear System Obtained from Electromagnetic Analysis*)

Soichiro IKUNO, Gong CHEN, Susumu YAMAMOTO, Taku ITOH[1)], Kuniyoshi ABE[2)] and Hiroaki NAKAMURA[3,4)]

School of Computer Science, Tokyo University of Technology, Tokyo 192-0982, Japan
[1)]*College of Industrial Technology, Nihon University, Narashino 275-8575, Japan*
[2)]*Faculty of Economics and Information, Gifu Shotoku Gakuen University, Gifu 500-8288, Japan*
[3)]*Department of Energy Engineering and Science, Nagoya University, Toki 509-5292, Japan*
[4)]*National Institute for Fusion Science, Toki 509-5292, Japan*

(Received 9 December 2015 / Accepted 20 January 2016)

Krylov subspace method and the variable preconditioned Krylov subspace method with communication avoiding technique for a linear system obtained from electromagnetic analysis are numerically investigated. In the $k-$skip Krylov method, the inner product calculations are expanded by Krylov basis, and the inner product calculations are transformed to the scholar operations. $k-$skip CG method is applied for the inner-loop solver of Variable Preconditioned Krylov subspace methods, and the converged solution of electromagnetic problem is obtained using the method.

Keywords: Krylov subspace method, communication avoiding technique, electromagnetic analysis, pralleliza-tion, GPGPU

DOI: 10.1585/pfr.11.2406021

1. Introduction

Recently, the performance of the Central Processing Unit (CPU), the Graphics Processing Unit (GPU) and Many Integrated Core (MIC) have increased in each of the past ten years. In addition, these devices are constituted by multitudes of processing units. Therefore, a parallelization scheme must be implemented on the simulation code in order to educe the performance of the devices.

As is well known that the speedup of the parallelization technique is governed by Amdahl's law. The simulation code can be divided into two parts. One is the parallelizable part, the other part consist of preprocessing, sequential calculation part. Thus, the value of speedup S is defined by following equation.

$$S = \frac{1}{1 - r + \frac{r}{n}}. \tag{1}$$

Here, r denotes a ratio of parallelizable part in the code, and n denotes a number of process. Equation (1) indicates that although the calculation cost decreases as the number of processing units increases, the lower limit inevitably exists i.e. $1 - r$. According to the idea of Amdahl's law, the calculation time decreases as the number of processing unit increases. However, in real calculations, the situation is even worse than predicted by Amdahl's law. In the parallelization calculation of the vector inner product, elements should be scatter to processing units (PUs). After the calculation, the results must be gather from PUs. Thus, the communication time increases as the parallelization units increase, and the communication time cannot be ignored.

The conjugate gradient (CG) method is one of solvers for a large-sparse linear system. In addition, the algorithm of CG method is very simple, and the most of all the procedure of the method is constituted by addition of vectors, inner products and multiplication of matrices and vectors as shown in Fig. 1. These operations are very easy to derive a parallelization efficiency. However, the communication between PUs must be necessary for parallelized inner products calculation, and the amount of the communication time increases as the size of the system increases. That is to say, the communication time is bottle neck for the effective parallelization of CG method [1, 2]. In order to enhance the performance of Krylov subspace method, new strategies for the communication bound kernels should be explored to minimize communication and data movement [3–6].

The purpose of the present study is to implement Krylov subspace method with communication avoiding techniques, and to evaluate the numerical features of the method.

author's e-mail: ikuno@nal.ikulab.org

*) This article is based on the presentation at the 25th International Toki Conference (ITC25).

Let $\boldsymbol{x}_0$ be an initial guess.
Set $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$
Set $\boldsymbol{p}_0 = \boldsymbol{r}_0$
for $k = 0, 1, \cdots$, until $\|\boldsymbol{r}_k\|_2/\|\boldsymbol{b}\|_2 \le \varepsilon$ **do**
$\alpha_k = \dfrac{(\boldsymbol{r}_k, \boldsymbol{r}_k)}{(\boldsymbol{p}_k, A\boldsymbol{p}_k)}$
$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$
$\boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \alpha_k A\boldsymbol{p}_k$
$\beta_k = \dfrac{(\boldsymbol{r}_{k+1}, \boldsymbol{r}_{k+1})}{(\boldsymbol{r}_k, \boldsymbol{r}_k)}$
$\boldsymbol{p}_{k+1} = \boldsymbol{z}_{k+1} + \beta_k \boldsymbol{p}_k$
end for

Fig. 1 The algorithm of the conjugate gradient (CG) method. Here, A denotes a coefficient matrix of a linear system and b denotes a known vector.

2. $k-$skip CG and CR Method

The communication avoiding technique is one of settlements for communication bottle neck issue for the parallelization CG. Although the algorithm of Krylov subspace method with the communication avoiding technique is different from original method, the theoretical methodology is equivalent. In the technique, inner products is rewritten into the recurrence formula using Krylov basis, and the inner products are transcribed into scalar calculation. As the results, collective communications between PUs can be gathered in only one time. As is obvious the communication time for the parallelization CG can be reduced.

Let us introduce the $k-$skip conjugate gradient method [4]. The inner product $(\boldsymbol{r}_{k+1}, \boldsymbol{r}_{k+1})$ can be rewritten by using $\alpha_k = (\boldsymbol{r}_k, \boldsymbol{r}_k)/(\boldsymbol{r}_k, A\boldsymbol{p}_k)$ as follows [1].

$$(\boldsymbol{r}_{k+1}, \boldsymbol{r}_{k+1}) = \alpha_k^2 (A\boldsymbol{p}_k, A\boldsymbol{p}_k) - (\boldsymbol{r}_k, \boldsymbol{r}_k). \tag{2}$$

In the $k-$skip strategy, a number of bases are necessary as a function of skip amount. Thus, other inner products can be also rewritten as follows.

$$\begin{aligned}(\boldsymbol{p}_{k+1}, A^j\boldsymbol{p}_{k+1}) &= (\boldsymbol{r}_{k+1}, A^j\boldsymbol{p}_{k+1}) + \beta_k(\boldsymbol{p}_k, A^j\boldsymbol{r}_k) \\ &\quad + \beta_k^2(\boldsymbol{p}_k, A^j\boldsymbol{p}_k) \\ &\quad - \alpha_k\beta_k(\boldsymbol{p}_k, A^{j+1}\boldsymbol{p}_k), \end{aligned} \tag{3}$$

$$\begin{aligned}(\boldsymbol{r}_{k+1}, A^j\boldsymbol{p}_{k+1}) &= (\boldsymbol{r}_{k+1}, A^j\boldsymbol{r}_{k+1}) \\ &\quad + (\boldsymbol{r}_k, A^j\boldsymbol{p}_k) - \alpha_k\beta_k(\boldsymbol{p}_k, A^{j+1}\boldsymbol{p}_k), \end{aligned} \tag{4}$$

$$\begin{aligned}(\boldsymbol{r}_{k+1}, A^j\boldsymbol{r}_{k+1}) &= (\boldsymbol{r}_k, A^j\boldsymbol{r}_k) - 2\alpha_k(\boldsymbol{r}_k, A^{j+1}\boldsymbol{p}_k) \\ &\quad + \alpha_k^2(\boldsymbol{p}_k, A^{j+2}\boldsymbol{p}_k). \end{aligned} \tag{5}$$

Here, the superscript j denotes a dimension size of Krylov subspace. The values of left hand side of (2), (3), (4) and (5) can be calculated by using value of previous step α_k, β_k, $(\boldsymbol{p}_k, A^j\boldsymbol{p}_k)$, $(\boldsymbol{p}_k, A^{j+1}\boldsymbol{p}_k)$, $(\boldsymbol{p}_k, A^{j+2}\boldsymbol{p}_k)$, $(\boldsymbol{r}_k, A^j\boldsymbol{p}_k)$, $(\boldsymbol{r}_k, A^{j+1}\boldsymbol{p}_k)$ and $(\boldsymbol{r}_k, A^j\boldsymbol{r}_k)$. Especially, in CG, $(k+1)-$th step values $(\boldsymbol{r}_{k+1}, \boldsymbol{r}_{k+1})$, α_{k+1} and β_{k+1} can be calculated using $(\boldsymbol{p}_k, A\boldsymbol{p}_k)$ and $(\boldsymbol{p}_k, A^2\boldsymbol{p}_k)$, and the step can be forwarded one step without inner product calculation.

Let $\boldsymbol{x}_0$ be an initial guess.
Set $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$
Set $\boldsymbol{p}_0 = \boldsymbol{r}_0$
for while $|\gamma_k|/\|\boldsymbol{b}\|_2 \le \varepsilon$ **do**
calculate $\boldsymbol{r}_n, A\boldsymbol{r}_n, A^2\boldsymbol{r}_n, \cdots, A^{2k+1}\boldsymbol{r}_n$
calculate $\boldsymbol{p}_n, A\boldsymbol{p}_n, A^2\boldsymbol{p}_n, \cdots, A^k\boldsymbol{p}_n, A^{2k+1}\boldsymbol{p}_n$
calculate γ_n
calculate $\delta_{n,1}, \cdots, \delta_{n,2k+1}$
calculate $\eta_{n,1}, \cdots, \eta_{n,2k+1}$
calculate $\zeta_{n,1}, \cdots, \zeta_{n,2k+1}$
for $i = n, 1, \cdots, n+k$ **do**
$\alpha_i = \delta_{i,1}/\eta_{i,1}$
$\beta_i = (\delta_{i,1} - 2\alpha_i\zeta_{i,2} + \alpha_i^2\eta_{i,2})/\delta_{i,1}$
for $j = 1, \cdots, 2k - 2(i-n)$ **do**
$\delta_{i+1,j} = \delta_{i,j} - 2\alpha_i\zeta_{i,j+1} + \alpha_i^2\eta_{i,j+1}$
$\eta_{i+1,j} = \delta_{i+1,j+1} + 2\beta_i(\zeta_{i,j+1} - \alpha_k\eta_{i,j+1}) + \beta_i^2\eta_{i,j}$
$\zeta_{i+1,j} = \delta_{i,j} + \alpha_i\zeta_{i,j+1} - \alpha_i(\zeta_{i,j+1} - \alpha_i\eta_{i,j+1}) + \beta_i\zeta_{i,j} - \alpha_i\beta_i\eta_{i,j}$
end for
$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i\boldsymbol{p}_i$
$\boldsymbol{r}_{i+1} = \boldsymbol{r}_i - \alpha_i A\boldsymbol{p}_i$
$\boldsymbol{p}_{i+1} = \boldsymbol{r}_{i+1} + \beta_i\boldsymbol{p}_i$
end for
$n = n + k + 1$
end for

Fig. 2 The algorithm of $k-$skip conjugate residual method.

As is well known that the Conjugate Residual (CR) method has much stable convergence property than that of CG. Additionally, it is difficult to obtain the converged solution of the ill-posed linear system using CG. From this reason, we derive the $k-$skip CR method by using above strategy. Substituting the recurrence equations

$$\begin{aligned}\boldsymbol{r}_{i+1} &= \boldsymbol{r}_i - \alpha_i A\boldsymbol{p}_i, \\ \boldsymbol{p}_{i+1} &= \boldsymbol{r}_{i+1} + \beta_i\boldsymbol{p}_i,\end{aligned}$$

to inner product operation, we can obtain following equations.

$$\begin{aligned}(A^j\boldsymbol{r}_{k+1}, \boldsymbol{r}_{k+1}) &= (A^j\boldsymbol{r}_k, \boldsymbol{r}_k) - 2\alpha_k(A^{j+1}\boldsymbol{p}_k, \boldsymbol{r}_k) \\ &\quad + \alpha_k^2(A^{j+1}\boldsymbol{p}_k, A\boldsymbol{p}_k), \\ (A^j\boldsymbol{p}_{k+1}, A\boldsymbol{p}_{k+1}) &= (A^{j+1}\boldsymbol{r}_{k+1}, \boldsymbol{r}_{k+1}) \\ &\quad + 2\beta_k(A^{j+1}\boldsymbol{p}_k, \boldsymbol{r}_{k+1}) \\ &\quad + \beta_k^2(A^j\boldsymbol{p}_k, A\boldsymbol{p}_k), \\ (A^j\boldsymbol{p}_k, \boldsymbol{r}_{k+1}) &= (A^j\boldsymbol{p}_k, \boldsymbol{r}_k) - \alpha_k(A^j\boldsymbol{p}_k, A\boldsymbol{p}_k), \\ (A^j\boldsymbol{p}_{k+1}, \boldsymbol{r}_{k+1}) &= (A^j\boldsymbol{r}_{k+1}, \boldsymbol{r}_k) - \alpha_k(A^{j+1}\boldsymbol{p}_k, \boldsymbol{r}_{k+1}) \\ &\quad + \beta_k(A^j\boldsymbol{p}_k, \boldsymbol{r}_k) - \alpha_k\beta_k(A^j\boldsymbol{p}_k, A\boldsymbol{p}_k), \\ (A^j\boldsymbol{r}_{k+1}, \boldsymbol{r}_k) &= (A^j\boldsymbol{r}_k, \boldsymbol{r}_k) - \alpha_k(A^{j+1}\boldsymbol{p}_k, \boldsymbol{r}_k).\end{aligned}$$

By using following notation, we can derive the $k-$skip CR, and the algorithm of the method is shown in Fig. 2.

$$\begin{aligned}\delta_{k,j} &\equiv (A^j\boldsymbol{r}_k, \boldsymbol{r}_k), \\ \eta_{k,j} &\equiv (A^j\boldsymbol{p}_k, A^j\boldsymbol{p}_k), \\ \zeta_{k,j} &\equiv (A^j\boldsymbol{p}_k, \boldsymbol{r}_k).\end{aligned}$$

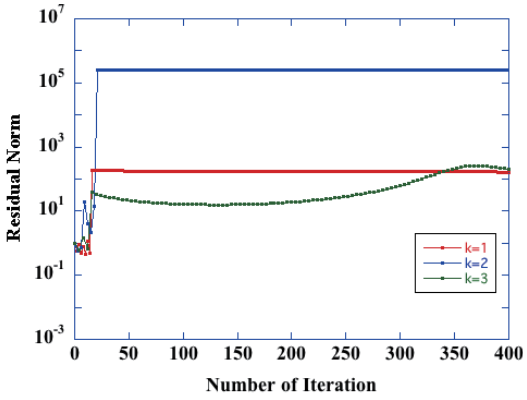

Fig. 3 The residual histories of $k-$skip CG method.

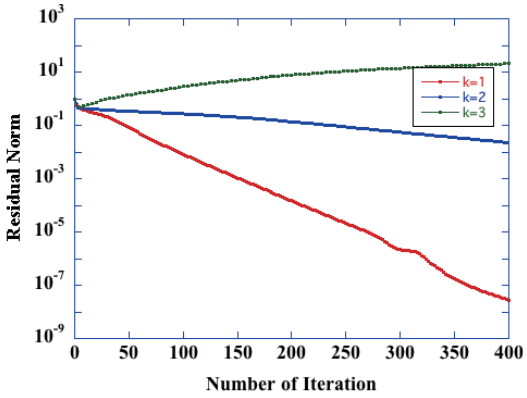

Fig. 4 The residual histories of $k-$skip CR method.

$k-$skip CG and $k-$skip CR are numerically evaluated by following tridiagonal matrix. The values of subdiagonals of the coefficient matrix are set to -1, and main diagonals are set to 2.05, and the dimension size of the matrix is fixed as 100. Note that the condition number of the matrix is 79.4. Moreover, elements of right hand side vector and the maximun iteration number are fixed as 1 and 400, respectively. The residual histories of $k-$skip CG and CR for various k are shown in Fig. 3 and Fig. 4. We can see from figures that the converged solution cannot obtain by $k-$skip CG. Additionally, although the converged solution is derived by $k-$skip CR with $k = 1$, the residual diverges as increase of value k. Furthermore, the residual property is stabilized by CR. However, given the convergence property of Krylov subspace method, $k-$skip technique takes much iterations. This is because the orthogonality of Krylov basis destroyed by expansion of the inner products and the truncation error of the power of the coefficient matrix for Krylov bases. On the other hand, during parallelization, collective communication can be avoided. From this reason, even amount of the operation is increased, it can be covered with parallel efficiency.

In the next section, $k-$skip CG, $k-$skip CR and the variable preconditioned (VP) CG with $k-$skip technique are adopted for realistic electromagnetic problem.

Let $\boldsymbol{x}_0$ be an initial guess.
Set $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$
Roughly solve $Az_0 = \boldsymbol{r}_0$ using some iterative method
Set $\boldsymbol{p}_0 = z_0$
for $k = 0, 1, \cdots$, until $\|\boldsymbol{r}_k\|_2 / \|\boldsymbol{b}\|_2 \leq \varepsilon$ **do**
$\alpha_k = \dfrac{(\boldsymbol{r}_k, z_k)}{(\boldsymbol{p}_k, A\boldsymbol{p}_k)}$
$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$
$\boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \alpha_k A\boldsymbol{p}_k$
Roughly solve $Az_{k+1} = \boldsymbol{r}_{k+1}$ using some iterative method
$\beta_k = \dfrac{(\boldsymbol{r}_{k+1}, z_{k+1})}{(\boldsymbol{r}_k, z_k)}$
$\boldsymbol{p}_{k+1} = z_{k+1} + \beta_k \boldsymbol{p}_k$
end for

Fig. 5 The algorithm of variable preconditioned conjugate gradient (VPCG) method.

3. Variable Preconditioned Krylov Subspace Method with $k-$skip Algorithm

A variable preconditioning method has been developed as a new preconditioning strategy for Krylov subspace methods [7]. VP Krylov subspace method has two nested iterations for main Krylov subspace method and variable preconditioning for Krylov subspace method are called as outer-loop and inner-loop. The VP Krylov subspace method has the sufficient condition for convergence. The residual of the problem converges if the relative residual norm of inner-loop satisfies the following inequality in each steps.

$$\frac{\|\boldsymbol{r}_{k+1} - Az_{k+1}\|_2}{\|\boldsymbol{r}_{k+1}\|_2} < 1. \tag{6}$$

That is to say, in the preconditioning procedure, the residual equation is solved roughly by using some iterative method with only a few iteration. Therefore, the algorithm of VP Krylov subspace method is very simple, so that the method have a good chemistry with parallelization techniques [8]. The algorithm of VPCG is shown in Fig. 5.

In this section, the Problem 20 in testing electromagnetic analysis methods (T.E.A.M.) Workshop for the benchmark [9]. Though the original Problem 20 is a nonlinear problem, the value of relative magnetic permeability is fixed as 200 so that the problem becomes a linear problem. The number of edge element of the problem is 27,549,822 and the dimension size of the coefficient matrix is 1,709,028. Note that the coefficient matrix becomes very sparse matrix, and only 42 nonzero elements include in unit column. In order to avoid the zero calculation, Compressed Row Storage (CRS) are employed for the calculation.

The residual histories of $k-$skip CG and CG are shown

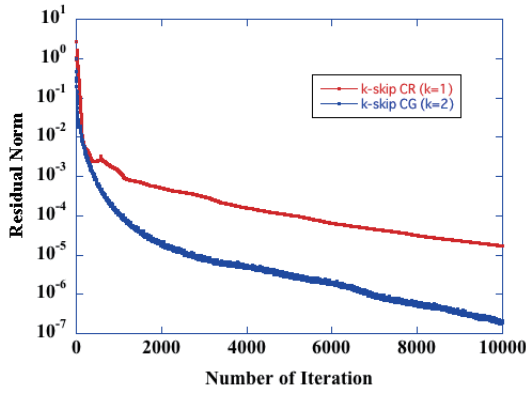

Fig. 6 The residual histories of $k-$skip CG and CR method in case of electromagnetic problem.

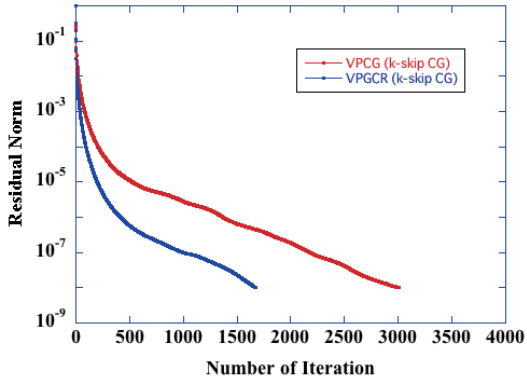

Fig. 7 The residual histories of VPCG and VPGCR with $k-$skip CG method. Note that the value of k is fixed as 2.

in Fig. 6. This figure indicates that the residual norm decreases less than one in a few iterations in both cases. Although, the residual of $k-$skip CG converges in case of $k = 2$, the converged solution cannot be obtained in case of $k-$skip CG method with $k = 1$. Essentially, the convergence property deteriorates as the skip size increase because of the truncation error of the power of the coefficient matrix for Krylov bases. Thus, this result remaining as our future work. In either case, above result leads us that $k-$skip CG and CR method can be adopted for the inner-loop solver for VP Krylov subspace method.

To take above fact into account, we adopt the $k-$skip CG and CR on the inner-loop solver for VP Krylov subspace method. In this study, VPCG and VPGCR is employed for the numerical evaluation, and the residual histories of VPCG and VPGCR with $k-$skip CG are shown in Fig. 7. Note that the termination condition is set as 10^{-8} in this calculation. We can see from this figure that the converged solutions are derived by both methods. In addition, VPGCR is converges faster than that of VPCG. However, the converged solution does not derived if $k-$skip CR method is adopted for inner-loop solver.

From these results, we can conclude that $k-$skip CG method is effective inner-loop solver for VP Krylov subspace method. Implementation on graphics processing unit (GPU) and evaluation of the parallelization efficiency are our future work.

4. Conclusion

In the present study, we have derived $k-$skip CR method by using Krylov basis. Furthermore, numerical features of $k-$skip CG and CR method have been evaluated. In addition, VPCG and VPGCR with $k-$skip CG have been developed, and the methods have been adopted for the electromagnetic problem.

Conclusions obtained in this study are summarized as follows.

- The converged solution cannot be obtained by $k-$skip CG in case of the test problem. Although, the converged solution is derived by $k-$skip CR method with $k = 1$, the residual diverges as increase of value k.
- The residual property is stabilized by CR method. However, given the convergence property of Krylov subspace method, $k-$skip technique takes much iterations. This is because the orthogonality of Krylov basis destroyed by expansion of the inner products. The truncation error of the power of the coefficient matrix for Krylov bases is also the cause of the deterioration.
- The converged solutions are derived by VPCG and VPGCR with $k-$skip CG. In addition, VPGCR is converges faster than that of VPCG. However, the converged solution does not derived if $k-$skip CR method is adopted for inner-loop solver.

In the future work, VPCG and VPGCR with $k-$skip CG method will be implemented on graphics processing unit (GPU) to derive high-performance parallelization efficiency.

Acknowledgment

This work was supported in part by Japan Society for the Promotion of Science under a Grant-in-Aid for Scientific Research (C) No. 26390135.

[1] Y. Saad, SIAM J. Sci. Stat. Comput. **6(4)**, 865 (1985).
[2] G. Meurant, Parallel Comput. **5(3)**, 267 (1987).
[3] M. Hoemmen, The doctoral dissertation of University of California at Berkeley, Berkeley, CA, USA, 2010.
[4] T. Motoya and R. Suda, IPSJ SIG Tech. Rep., Vol.2012-HPC-133, No.30 (2012) (Japanese).
[5] E.F. D'Azevedo, V. Eijkhout and C.H. Romine, Technical Report, Mathematical Sciences Section, Oak Ridge National Laboratory, 1999.
[6] A.T. Chronopoulos and C.W. Gear, J. Comput. Appl. Math. **25(2)**, 153 (1989).
[7] K. Abe and S.L. Zhang, Int. J. Numer. Anal. Model. **2**, No.2, 137 (2005).
[8] S. Ikuno, Y. Kawaguchi, T. Itoh, S. Nakata and K. Watanabe, IEEE Trans. Magn. **48**, No.2, 467 (2012).
[9] http://www.compumag.org/jsite/team.html