

Parallel Preconditioning Methods with Selective Fill-Ins and Selective Overlapping for Ill-Conditioned Problems in Finite-Element Methods

Kengo Nakajima

Department of Earth and Planetary Science, The University of Tokyo
 7-3-1 Hongo, Bunkyo-ku, Tokyo 112-0002, Japan
 nakajima@eps.s.u-tokyo.ac.jp
<http://www-solid.eps.s.u-tokyo.ac.jp/~nakajima>

Abstract. In the present paper, parallel preconditioning methods with selective fill-ins and selective overlapping for contact problems have been developed and evaluated on PC clusters with 64 cores. The proposed methods provide robust convergence with efficiency. Finally, the proposed methods have been applied to ill-conditioned problems with heterogeneity and robustness, and their efficiency has been evaluated on TSUBAME super-cluster with up to 512 cores.

1 Introduction

Contact phenomenon is one of the most important and critical issues in various types of science and engineering applications. The author has been developing preconditioning methods for contact problems in geophysics, in which stress accumulation on plate boundaries (faults) is computed for estimating the earthquake generation cycle [1,2]. In [1,2], the augmented Lagrangian method (ALM) and the penalty method are implemented, and a large penalty number λ is introduced for constraint conditions around faults. The nonlinear process is solved iteratively by the Newton-Raphson (NR) method. A large λ ($\sim 10^4$) can provide an accurate solution and fast nonlinear convergence for NR processes, but the condition number of the coefficient matrices for linear equations is large, and several iterations are required for the convergence of iterative solvers. Therefore, a robust preconditioning method is essential for such ill-conditioned problems.

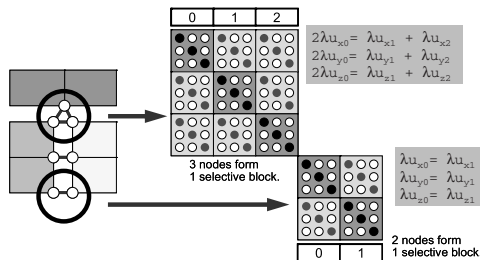


Fig. 1. Matrix operation of nodes in contact group for *selective blocking* preconditioning

Selective blocking is a special preconditioning method developed for this type of application by the author [1]. In this method, finite element nodes in the same contact group coupled through penalty constraints are placed into a large block (selective block or super node) (Fig. 1). For symmetric positive definite matrices, preconditioning with *block incomplete Cholesky factorization using selective blocking* (SB-BIC) shows excellent performance and robustness [1, 2]. In previous studies [1, 2], the number and location of nodes on both sides of the contact surface are identical. This is not flexible and cannot be applied to fault contact simulations with large slip/deformation, in which the number and location of nodes in contact groups might be inconsistent.

In the present study, new parallel preconditioning methods for this type of general contact problem have been developed. One remedy is a preconditioning method with *selective fill-ins*, in which *fill-ins* of higher order are introduced only for nodes connected to special elements for contact conditions [1, 2]. Another remedy is extension of overlapped elements between domains. In the present study, the *selective overlapping* method has been proposed, which extends the layers of overlapped elements according to the information of the special elements for contact conditions. Both methods are based on the idea of *selective blocking*, but are more general and flexible.

In the remainder of this paper, algorithms of parallel preconditioning methods with *selective fill-ins* and *selective overlapping* will be described, and the results of example problems with contact conditions are shown. Finally, the developed methods are applied to general ill-conditioned problems with the heterogeneous material property, and parallel performance up to 512 cores is evaluated.

2 Parallel Preconditioning Methods

2.1 Selective Fill-Ins

The *selective blocking* preconditioning method [1, 2] is a very robust and efficient preconditioning method for contact problems. However, it can be applied to a very limited number of situations. **ILU(p)** (Incomplete LU factorization with p-th-order fill-ins) preconditioning methods are widely used for various types of applications [3]. If the order of fill-ins (**p**) is higher, the preconditioner is more robust, but is usually more expensive. The required memory for coefficient matrices increases by a factor of from 2 to 5, if the order of fill-ins (**p**) increases from 0 to 1 or from 1 to 2 [1, 2].

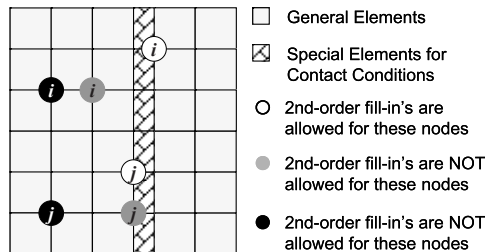


Fig. 2. Example of ILU(1+)

In the present study, new preconditioning methods for general contact problems have been developed. The first approach is a preconditioning method with *selective fill-ins*, called $\mathbf{ILU}(\mathbf{p}+)$. Figure 2 describes the idea of $\mathbf{ILU}(\mathbf{p}+)$. In $\mathbf{ILU}(\mathbf{p}+)$, $(p+1)$ -th order fill-ins are allowed for m_{ij} , which represents the component of preconditioned matrices, if both the i -th and j -th nodes are connected to special elements for contact conditions, such as *master-slave* type elements [2]. In Fig. 2, second-order fill-ins can be allowed for all three i - j pairs, according to graphical connectivity information. However, only the white circles are allowed to generate second-order fill-ins.

This approach is very similar to that of *selective blocking*, in which full LU factorization is applied to nodes in contact groups, but is much more general and flexible. Since the constraint conditions through penalty terms are applied to the nodes that are connected to special elements, *selective ILU* factorization with higher order fill-ins for these nodes is expected to provide robust convergence with efficiency. In [4], a preconditioning method with block ILU factorization is proposed for coupled equations of incompressible fluid flow and solid structure. Different orders of fill-ins are applied to velocity and pressure components for generating block ILU factorization of coefficient matrices. $\mathbf{ILU}(\mathbf{p}+)$ is very similar to this idea.

2.2 Selective Overlapping

Another approach is the extension of overlapped zones between domains for parallel computing. In previous studies [1, 2], the *GeoFEM* local data structure has been applied. This data structure is node-based with a single layer of overlapped elements (the depth of overlapping is 1) and is appropriate for parallel iterative solvers with block Jacobi-type localized preconditioning methods [1, 2]. Figure 3 shows an example of the local data for contact problems, in which the depth of overlapping is 1.

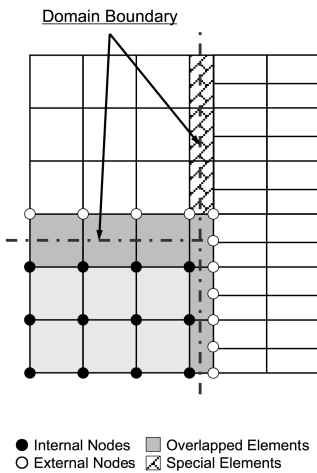


Fig. 3. Example of GeoFEM local data structure for contact problems

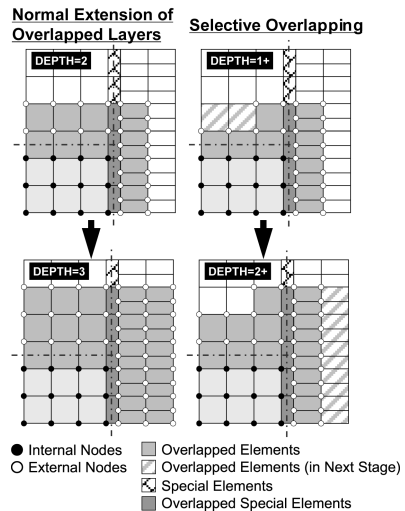


Fig. 4. Example of *selective overlapping*. Precedence for extensions of overlapped layers is taken over nodes connected to special elements for contact conditions

In the present study, a greater number of layers of overlapped elements are considered for the robustness of parallel preconditioners. Generally speaking, a larger depth of overlapped layers provides faster convergence in block Jacobi-type localized preconditioning methods, but the cost for computation and communications is more expensive [1,2,5].

In the present study, the *selective overlapping* method is proposed. This method provides priority over nodes connected to special elements for contact conditions, when extensions of overlapped layers are applied, as shown in Fig. 4. In *selective overlapping*, extension of overlapping for nodes that are *not* connected to special elements for contact conditions is *delayed*. *Hatched* elements in Fig. 4 are to be included as extended overlapped elements in normal extension cases. However, in selective overlapping, extension of overlapping for these elements is performed in the next stage of overlapping. Thus, the increase in cost for computation and communication by extension of overlapped elements is suppressed.

This idea is also an extension of the idea of *selective blocking*, and is also based on the idea of special partitioning strategy for contact problems, developed in [1, 2]. The convergence rate of parallel iterative solvers with block Jacobi-type localized preconditioning is generally bad, because the *edge-cut* may occur at inter-domain boundary edges that are included in contact groups [1, 2]. All nodes in the same contact group should be in the same domain in order to avoid such edge-cuts [1, 2]. Because the constraint conditions through penalty terms are applied to those nodes that are connected to special elements, *selective* extension of overlapping for these nodes is expected to provide robust convergence with efficiency.

3 Examples

3.1 Problem Description

Figure 5 describes the model for validation of the developed preconditioning methods. This problem simulates general contact conditions, in which the positions and number of nodes on contact surfaces are inconsistent. There are four blocks with elastic material that are discretized into cubic tri-linear type finite-elements. Each block is connected through elastic truss elements generated at each node on contact surfaces. Truss elements are in the form of a cross, as shown in Fig. 5. In this case, the elastic coefficient of the truss elements is set to 10^4 times that of the solid elements, which corresponds to the coefficient $\lambda (=10^4)$ for constraint conditions of the augmented Lagrangian method (ALM). Poisson's ratio for cubic elements is set to 0.25.

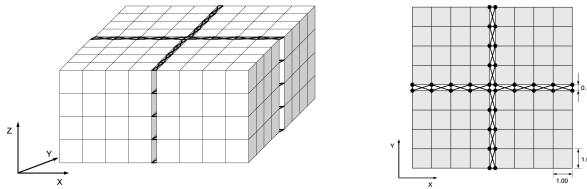


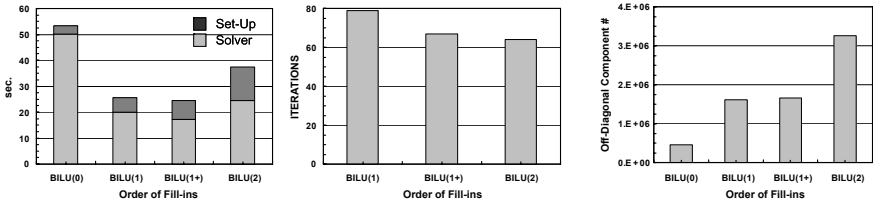
Fig. 5. Elastic blocks connected through truss elements

Symmetric boundary conditions are applied to $x = 0$ and $y = 0$ surfaces, and the Dirichlet fixed condition for deformation in the direction of the z -axis is applied to $z = 0$ surfaces. Finally, a uniform distributed load in the direction of the z -axis is applied to $z = Z_{\max}$ surfaces. This problem is linear-elastic, but the coefficient matrices are very ill-conditioned and simulate those for nonlinear contact problems very well [2].

3.2 Serial Cases with a Single CPU

Each chart in Fig. 6 shows the results of linear-elastic computation on the model in Fig. 5 using a single core of AMD Opteron 275 (2.2 GHz) with PGI compiler. Each block in Fig. 5 has 8,192 ($=16 \times 16 \times 32$) cubes, where the total problem size is 117,708 DOF ($=$ degrees of freedom) (38,148 cubes, 39,236 nodes). GPBi-CG (Generalized Product-type methods based on Bi-CG) [6] for general coefficient matrices have been applied as an iterative method, although the coefficient matrices of this problems are positive indefinite. Each node has three DOF in each axis in 3D solid mechanics; therefore, **block ILU (BILU)** type preconditioning [1,2] has been applied.

BILU(1+), in which additional *selective fill-ins* have been applied for nodes connected to special elements (elastic truss elements in Fig. 5) to **BILU(1)**, provides the most robust and efficient convergence. **BILU(p)** provides faster convergence if p is larger, as shown in Fig. 6, but the cost for computation is more expensive. **BILU(1)** and **BILU(1+)** are competitive, but **BILU(1+)** provides a better convergence rate.



(a) Computation time (b) Iterations for convergence (c) Off-diagonal component # of preconditioned matrix (320 for BILU(0))

Fig. 6. Results of linear-elastic problem in Fig.5 with a single core of AMD Opteron 275 (2.2 GHz) with PGI compiler. 117,708 DOF (38,148 cubes, 39,236 nodes)

3.3 Parallel Cases with Multiple CPU's

Each chart in Fig. 7 shows the results of linear-elastic computation on the model shown in Fig. 5 using 64 cores of AMD Opteron 275 cluster with PGI compiler and MPICH connected through Infiniband network. Each block in Fig. 5 has 250,000 ($=50 \times 50 \times 100$) cubes, where the total problem size is 3,152,412 DOF (1,000,000 cubes, 1,050,804 nodes). The effect of the extension of overlapping is evaluated for **BILU(1)**, **BILU(1+)**, and **BILU(2)**, respectively. **BILU(p)-(d)** means **BILU(p)** preconditioning, where the depth of overlapping is equal to d . Partitioning was applied in an RCB (recursive coordinate bisection) manner [2].

Generally speaking, the convergence rate is improved by the extension of overlapping (Fig. 7(a)). This is significant, when the depth of overlapping (d) is increased from ($d = 1$) to ($d = 1+$) because *edge-cuts* may occur at truss elements for contact

conditions, if the depth of overlapping is 1. However, the decrease in the number of iterations for convergence is very slight if the depth of overlapping is greater than 2.

The number of off-diagonal components of preconditioned matrices increases, as the depth of overlapping is larger (Fig. 7(c)). Finally, computation with a larger depth of overlapping is more expensive. Therefore, the computation time increases, because the depth of overlapping is greater than 2 (Fig. 7(a)). **BILU(1)-(1+)** and **BILU(1+)-(1+)** are the best cases and they are competitive.

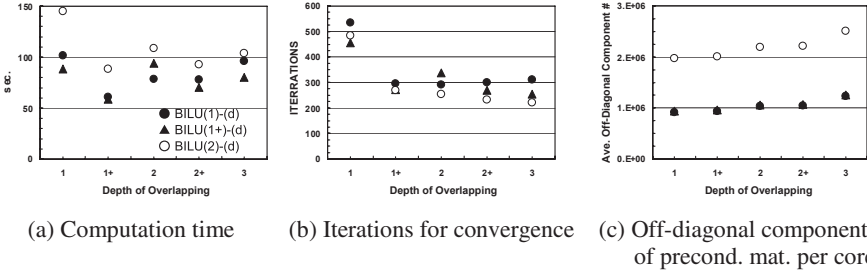


Fig. 7. Results of linear-elastic problem in Fig.5 with 64 cores of AMD Opteron 275 (2.2 GHz) cluster. 3,152,412 DOF (1,000,000 cubes, 1,050,804 nodes)

4 Applications to Problems with Heterogeneous Material Property

Parallel preconditioning methods with selective fill-ins and selective overlapping (**BILU(p)-(d)** approach), which are developed in the present study, provide robust and efficient convergence in ill-conditioned problems with contact conditions. In this section, this **BILU(p)-(d)** is extended to **BILU(p+,ω)-(d+,α)** for ill-conditioned problems with heterogeneous material property, where ω and α are threshold parameters for the extension of fill-ins and overlapping. In applications for a heterogeneous distribution of material property, coefficient matrices for linear solvers are generally ill-conditioned and the rate of convergence is poor.

In this section, linear-elastic problems for simple cube geometries with heterogeneity are solved. Poisson’s ratio is set to 0.25 for all elements, and the heterogeneous distribution of Young’s modulus in each tri-linear element is calculated by a sequential Gauss algorithm, which is widely used in the area of geo-statistics [7]. The minimum and maximum values of Young’s modulus are 10^{-2} and 10^2 , respectively, where the average value is 1.0. Boundary conditions in Section 3.1 have been applied. Each chart of Fig. 8 shows the results of linear-elastic computation on the model with heterogeneity using 64 cores of AMD Opteron 275 cluster. The total problem size is 3,090,903 DOF (1,000,000 cubes). In **BILU(p+,ω)-(d+,α)**, (p+1)-th-order fill-ins are allowed for pairs of nodes if both nodes are connected to elements for which the Young’s modulus is greater than ω , and *selective overlapping* is applied to nodes if the nodes are connected to elements for which the Young’s modulus is greater than α . In this case, **BILU(1)** generally requires more iterations for convergence than **BILU(1+,ω)** and **BILU(2)**. **BILU(1+,5)** and **BILU(1+,10)** are competitive, but **BILU(1+,5)-(1+,10)** ($\omega = 5, \alpha = 10$) provides the best performance. In **BILU(2)** and

BILU(1+, ω), the effect of the convergence rate by selective overlapping is similar to that in the previous cases for contact problems. Finally, the parallel performance of **BILU(1+,5)-(1+,10)** has been evaluated using between 32 and 512 cores of TSUBAME super-cluster [8] at the Tokyo Institute of Technology. In this evaluation, a *strong scaling test* has been applied, where the entire problem size is fixed as 3,090,903 DOF (1,000,000 cubes). Figure 9 shows parallel performance. Usually, the convergence rate for block Jacobi-type localized parallel preconditioning in strong scaling cases becomes worse as the number of domains increases [1,2]. However, selective overlapping provides an almost constant number of iterations up to 512 cores, as shown in Fig. 9(a), and excellent parallel performance, as shown in Fig. 9(b).

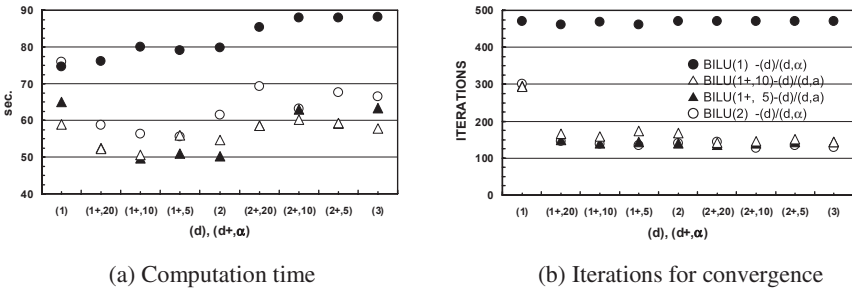


Fig. 8. Results of linear-elastic problem with heterogeneity using 64 cores of AMD Opteron 275 (2.2 GHz) cluster. 3,090,903 DOF (1,000,000 cubes, 1,030,301 nodes)

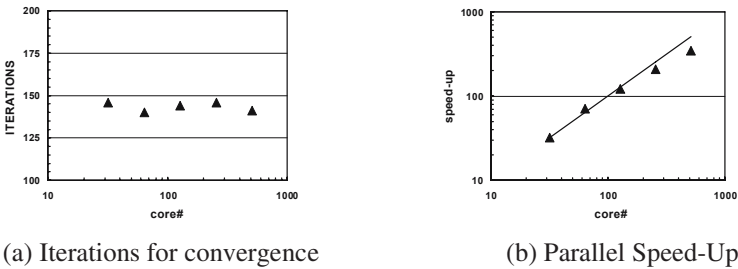


Fig. 9. Parallel performance of BILU(1+,5)-(1+,10) for linear-elastic problem with heterogeneity using TSUBAME super-cluster using up to 512 cores (3,152,412 DOF)

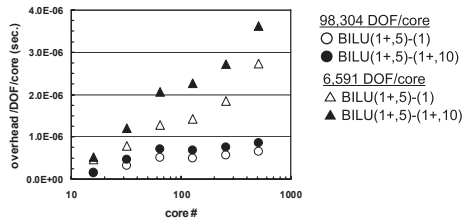


Fig. 10. Overhead for weak scaling test on TSBUME cluster using **BILU(1+,5)**

Figure 10 shows the results of the *weak scaling test*, in which the problem size per core is fixed. The obtained results correspond to the overhead by communications and additional computations per iteration divided by the problem size per core. If the problem size per each core is sufficiently large, the additional overhead by *selective overlapping* ($(\mathbf{d}=1)$ and $(\mathbf{d},\boldsymbol{\alpha})=(\mathbf{1},\mathbf{10})$) is negligible.

5 Concluding Remarks

In the present paper, parallel preconditioning methods with *selective fill-ins* and *selective overlapping* for contact problems have been developed and evaluated on PC clusters. The proposed methods are based on the concept of *selective blocking* in previous studies, but are much more flexible and provide robust convergence with efficiency. The proposed methods have also been applied to ill-conditioned problems with heterogeneity and were found to be robust and efficient on TSUBAME super-cluster with up to 512 cores. Generally speaking, **BILU(1+)-(1+)** or **BILU(1+, ω)-(1+, $\boldsymbol{\alpha}$)**, with selective fill-in ($\mathbf{p} = \mathbf{1+}$) and selective overlapping ($\mathbf{d} = \mathbf{1+}$), provides the best performance with robustness. The effect of *selective overlapping* is very significant, if the depth of overlapping increases from ($\mathbf{d} = \mathbf{1}$) to ($\mathbf{d} = \mathbf{1+}$). In future studies, the proposed methods will be evaluated in various types of real applications with different parameters on massively parallel computers.

Acknowledgements

This work is supported by the 21st Century Earth Science COE Program at the University of Tokyo, and CREST/Japan Science and Technology Agency. The author would like to thank Professor Satoshi Matsuoka and members of the Global Scientific Information and Computing Center, Tokyo Institute of Technology, for use of TSUBAME super-cluster.

References

1. Nakajima, K. (2003), Parallel Iterative Solvers of GeoFEM with Selective Blocking Preconditioning for Nonlinear Contact Problems on the Earth Simulator, ACM/IEEE Proceedings of SC2003.
2. <http://geofem.tokyo.rist.or.jp/>
3. Saad, Y. (2003), Iterative Methods for Sparse Linear Systems (2nd Edition), SIAM.
4. Washio, T., Hisada, T., Watanabe, H., and Tezduyar, T.E. (2005), A Robust and Efficient Iterative Linear Solver for Strongly Coupled Fluid-Structure Interaction Problems, Computer Methods in Applied Mechanics and Engineering, 194, 4027-4047.
5. Smith, B., Björstad, P. and Gropp, W. (1996), Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations, Cambridge University Press.
6. Zhang, S.L. (1997), GPBi-CG: Generalized Product-type methods based on Bi-CG for solving nonsymmetric linear systems, SIAM Journal of Scientific Computing, 18, 537-551.
7. Deutsch, C.V. and Journal, A.G. (1998), GSLIB Geostatistical Software Library and User's Guide, Second Edition, Oxford University Press.
8. <http://www.gsic.titech.ac.jp/>