Research activities of Hokkaido University group on next-generation linear solvers -mixed precision computing, accelerators, subspace correction techniques-

Takeshi Iwashita
(Kyoto Univ. / Hokkaido Univ.)

Research results of our group

- Mixed precision computing
  1. Development of mixed precision GMRES(m) solver
     - The solver is based on the iterative refinement method. An *m*-step FP32 GMRES solver is used for the inner iteration. In the outer iteration, the residual vector is calculated using FP64 and the convergence of the relative residual norm is checked. Thus the accuracy of the solution is the same as that of a conventional FP64 GMRES solver.
       - ***Yingqi Zhao***, ***Takeshi Fukaya***, *Linjie Zhang*, ***Takeshi Iwashita***: *Numerical Investigation into the Mixed Precision GMRES(m) Method Using FP64 and FP32. J. Inf. Process. 30: 525-537 (2022)*
  2. Investigation of mixed precision IR-based Bi-CGSTAB solver
     - Methods based on the Lanczos principle look unsuitable for the inner solver of the iterative refinement method. However, we implemented it and got some good results.
       - *To be published*

Research results of our group

3. Introduction of mixed precision computing to H-matrices
   - A part of low-rank submatrices are expressed using lower precision data. The proposed method was evaluated in a BEM analysis for electric field.
     - ***Rise Ooi***, ***Takeshi Iwashita***, ***Takeshi Fukaya***, *Akihiro Ida, Rio Yokota: Effect of Mixed Precision Computing on H-Matrix Vector Multiplication in BEM Analysis. HPC Asia 2020: 92-101.*
4. Integer arithmetic based GMRES solver
   - An integer arithmetic based preconditioned GMRES solver was developed. The solver is based on the iterative refinement method and only integer arithmetic are used in the inner iteration.
     - ***Takeshi Iwashita***, ***Kengo Suzuki***, ***Takeshi Fukaya***: *An Integer Arithmetic-Based Sparse Linear Solver Using a GMRES Method and Iterative Refinement. ScalA@SC 2020: 1-8.*

Research results of our group

- ILU-type preconditioning techniques effectively using SIMD vectorization
  1. Proposal of hierarchical block multi-color ordering
     - The proposed parallel ordering method attains the same convergence rate as the block multi-color ordering method, while the forward and backward substitutions are efficiently parallelized and vectorized. (The preconditioning matrix involves diagonal matrices.)
       - ***Takeshi Iwashita***, ***Senxi Li***, ***Takeshi Fukaya***: *Hierarchical block multi-color ordering: a new parallel ordering method for vectorization and parallelization of the sparse triangular solver in the ICCG method. CCF Trans. High Perform. Comput. 2(2): 84-97 (2020)*

Research results of our group

2. Proposal of ILUB preconditioning

- A new fill-in strategy was proposed for ILU preconditioning. Nonzero blocks are defined and all fill-ins in a nonzero block are permitted in ILU factorization. Then, the forward and backward substitutions consist of small dense matrix operations which is efficiently vectorized.
  - ***Kengo Suzuki***, ***Takeshi Fukaya***, ***Takeshi Iwashita***: *A novel ILU preconditioning method with a block structure suitable for SIMD vectorization. J. Comput. Appl. Math. 419: 114687 (2023)*

Research results of our group

- GPU oriented preconditioners
  1. Proposal of a new AINV preconditioner
     - A modification of the AIV algorithm was presented. A new approximation method for the preconditioner matrix greatly reduces the preconditioner setup time with preserving the preconditioning effect.
       - ***Kengo Suzuki***, ***Takeshi Fukaya***, ***Takeshi Iwashita****: A New AINV Preconditioner for the CG Method in Hybrid CPU-GPU Computing Environment. J. Inf. Process. 30: 755-765 (2022)*

Research result: convergence acceleration of an iterative solver

- Aim of the research: Speedup of the convergence of a sparse linear iterative solver
- Subspace correction method
  - J. Xu, "Iterative Methods by Space Decomposition and Subspace Correction", SIAM Rev., Vol. 34, 1992.
- Deflation method
  - R. A. Nicolaides, "Deflation of conjugate gradients with application to boundary value problems", SIAM J. Numer. Anal., Vol. 24, 1987.
  - Y. Saad et al., "A deflated version of the conjugate gradient algorithm", SIAM J. Sci. Comput., Vol. 21, 2000.
- These techniques can collaborate with standard preconditioning techniques and use a specific subspace.
  - It is important to set an appropriate subspace.
  - Here we consider a linear system of equations having a S.P.D. matrix. In many practical simulations, a few isolated eigenvalues with very small magnitudes cause a convergence problem. If we identify the eigenspaces corresponding to these eigenvalues, we can accelerate the convergence of the iterative (CG) solver.

Basic idea

- The way to obtain (approximate) eigenvectors with small eigenvalues
  - Based on knowledge of problems
    - Multigrid method, finite elements with high aspect ratio
- In this research, we focus on algebraic auxiliary matrix construction.
  - Reduction of programming cost
  - (Hidden) unapparent characteristics of a problem may be exploited.
- We focus on a situation in which a series of linear systems with an identical or similar coefficient matrix are solved.
  - In practical simulations, these linear systems often arise.
    - Time dependent calculations, non-linear analysis (Newton method)

Proposed algebraic auxiliary matrix construction method

- Problem:

$$A_k x_k = b_k, (k = 1, 2, \dots m)$$

$$A_k = A$$

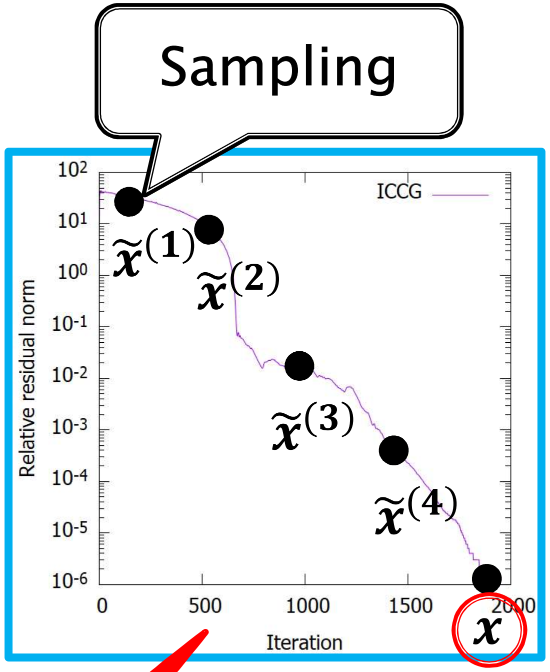
The right hand side vector b_k are functions of x_1, x_2,⋯,x_{k-1}.

- Proposed method
- In the first solution step for $Ax_0 = b_0$, the following 4 steps are performed. For simple explanation, let x_0 and b_0 denoted by x and b.
  1. Sampling of approximation vectors
  2. Calculation of error vectors
  3. Rayleigh-Ritz method
  4. Construction of auxiliary matrix

Proposed algebraic auxiliary matrix construction method (2)

❶ Sampling of approximation vectors

- Number of samples: m
- Sampling is done with a certain interval
- Obtain $\tilde{\boldsymbol{x}}^{(s)} (s = 1, \dots, m)$



❷ Calculation of error vectors

- After the solution process completes, the error vectors corresponding to the sampled approximation vectors are calculated: $\boldsymbol{e}^{(1)}, \boldsymbol{e}^{(2)}, \dots, \boldsymbol{e}^{(m)}$.
- Apply the Gram–Schmidt process to the vectors and get orthonormal basis vectors: $(\bar{\boldsymbol{e}}^{(1)}, \bar{\boldsymbol{e}}^{(2)}, \dots, \bar{\boldsymbol{e}}^{(\bar{m})}) = \boldsymbol{E}\ (\bar{m} \leq m)$

$$\boldsymbol{e}^{(s)} = \boldsymbol{x} - \tilde{\boldsymbol{x}}^{(s)}$$
$$(s = 1, \dots, m)$$

Proposed algebraic auxiliary matrix construction method (3)

❸ Rayleigh-Ritz method

- Solve an eigenvalue problem: $\boldsymbol{E}^T\boldsymbol{A}\boldsymbol{E}\boldsymbol{t} = \boldsymbol{\lambda}\boldsymbol{t}$
- Select Ritz vectors with small Ritz values less than θ
  - θ should be much less than 1 when the coefficient matrix is diagonally (or properly) scaled.
  - $\tilde{m}$: the number of selected Ritz vectors
  - Selected Ritz vectors: $\boldsymbol{p}_i$ $(i = 1, 2, \ldots, \tilde{m})$

❹ Auxiliary matrix

- $W = (\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_{\tilde{m}})$

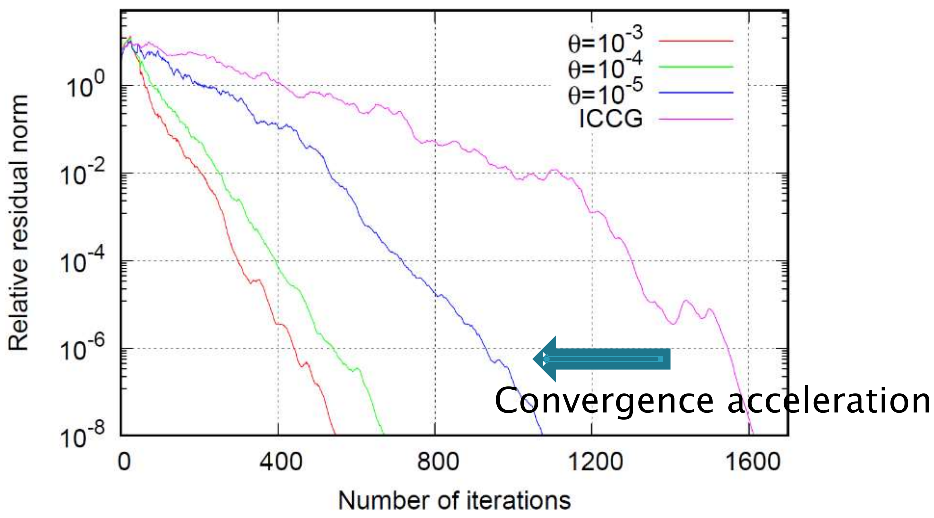
Numerical tests

- We used a node of Fujitsu CX2550 (2 Xeon Skylake processors)
- We tested three solvers
  - ICCG
  - ES-SC-ICCG: additive Schwarz SC and IC preconditioned CG solver
  - ES-D-ICCG: deflated ICCG solver

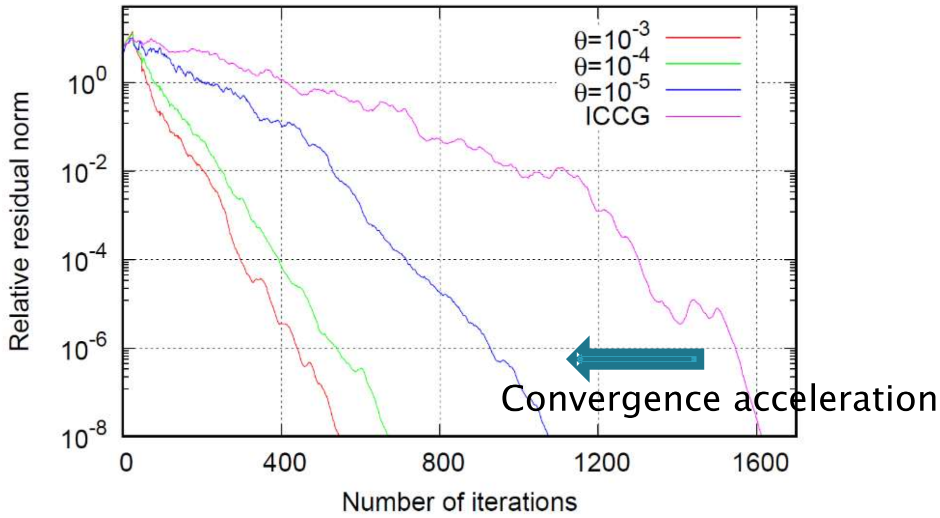
  (ES-SC-ICCG and ES-D-ICCG: Using the proposed technique)
- We also tested parallelized versions of three solvers
  - Block Jacobi IC method is used for the parallelization of IC preconditioning
- The convergence criterion: relative residual norm less than 10^{-8}
- Test problems: 30 matrices from SuiteSparse Matrix collection
  - We selected 30 largest datasets of symmetric positive-definite matrices.
- We solved a linear system 6 times. In / after the first solution step, vector sampling and the auxiliary matrix construction are performed. Two types of right-hand side vectors are used: (1) a vector of ones and (2) a random vector.
- The number of sampling vectors, m : 20
- The threshold for Ritz vectors, θ: 10^{-3}, 10^{-4}, 10^{-5}
- The number of threads in tests for parallel solvers: 40

Numerical results (evaluation in convergence)

Results of tests using random vectors (dataset: Hook_1498)



ES-SC-ICCG

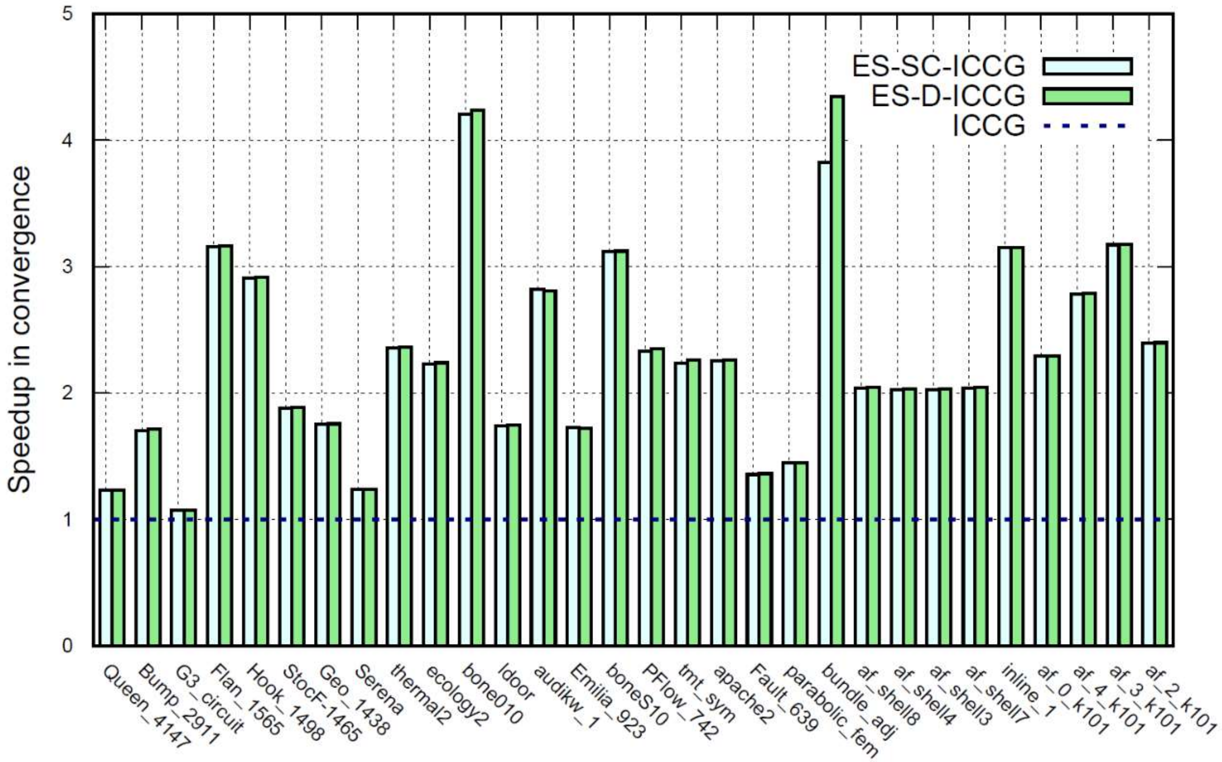


ES-D-ICCG

- Significant convergence acceleration was achieved.
- When we use a larger θ, which typically increases the number of selected Ritz vector, $\tilde{m}$, reduction in the number of iterations is enlarged.
- Although the mechanism of convergence acceleration is different between subspace correction and deflation, the convergence behaviors are mostly identical. (This phenomenon was observed in all datasets.) It implies that the difference of both methods are marginally when the used subspace is identical.

Numerical results (evaluation in convergence)

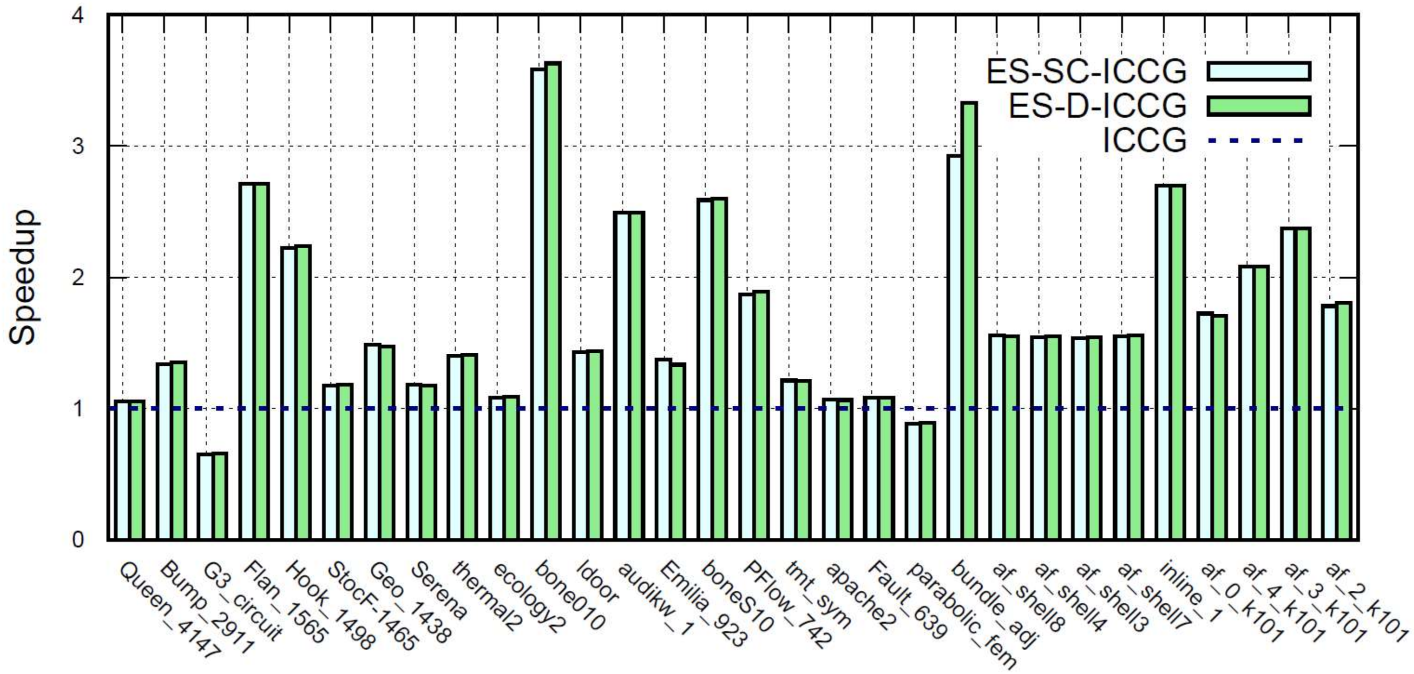
Results of tests using random vectors



- The convergence rate was doubled for 20 out of 30 datasets.
  - When a vector of ones was used, the improvement in convergence rate was more significant.
- For 12 datasets, the best result was obtained when $\tilde{m} = m$ (# sampled vectors). For these datasets, the increase in m possibly improve the performance.

Numerical results (evaluation in time to solution)

Results of tests using random vectors



- The time to solution was reduced for 28 out of 30 datasets using the proposed technique.
- For 12 datasets, the best result was obtained when $\tilde{m} = m$ (# sampled vectors). For these datasets, the increase in m possibly improve the performance.

Iwashita lab.

Summary

- We newly introduce an algebraic auxiliary matrix (subspace) construction method for subspace correction and deflation based on error vector sampling in a series of linear systems.
- In the method, limited number of approximation vectors are preserved. The error vectors are calculated after the solution step finishes. The auxiliary matrix is constructed using these error vectors.
- Numerical tests using matrices from SuiteSparse confirm that the subspace correction preconditioning and deflation using the proposed technique accelerate the convergence and the solution of a preconditioned CG solver.
- Numerical test also indicates that the proposed method can be used for the condition number estimation.

Publication

- Basic concept
  - *T. Iwashita, S. Kawaguchi, T. Mifune, and T. Matsuo, Automatic mapping operator construction for subspace correction method to solve a series of linear systems, JSIAM Letters, vol. 9, pp. 25-28, 2017.*
- Preliminary analysis on computational electromagnetics
  - *T. Iwashita, S. Kawaguchi, T. Mifune, and T. Matsuo, Acceleration of Transient Non-Linear Electromagnetic Field Analyses Using an Automated Subspace Correction Method, IEEE Trans. Magn., vol. 55, pp. 1-4, 2019.*
- Regular paper
  - *T. Iwashita, K. Ikehara, T. Fukaya, and T. Mifune, Convergence acceleration of preconditioned conjugate gradient solver based on error vector sampling for a sequence of linear systems, Numer. Linear Algebra Appl. (2023), e2512.*

Thank you

We will welcome your visit to Hokkaido & Kyoto Univ.