High Performance Numerical Algorithm

- Development of stable and high accuracy solvers for linear systems with multiple right-hand sides -

Hiroto Tadano
Division of High Performance Computing Systems
Center for Computational Sciences,
University of Tsukuba

External Review on CCS
February 19, 2014

Outline

- § 1 Introduction
- § 2 Development of a high accuracy linear solver (BiCGGR)
- § 3 Stablization of Block BiCGGR
- § 4 Numerical experiments
- § 5 Summary

§ 1 Introduction

Linear systems with multiple right-hand sides

Linear systems with L right-hand sides

$$AX = B$$

Here, $A \in \mathbb{C}^{n \times n}$: $n \times n$ non-Hermitian matrix,

$$X = \left[x^{(1)}, x^{(2)}, \ldots, x^{(L)}\right], \quad B = \left[b^{(1)}, b^{(2)}, \ldots, b^{(L)}\right]$$

This linear system appears in ...

- Eigensolver using contour integral（SS method）
- Physical value calculation in Lattice QCD
  - ➡ Linear system with 12 〜 100 multiple right-hand sides need to be solved.

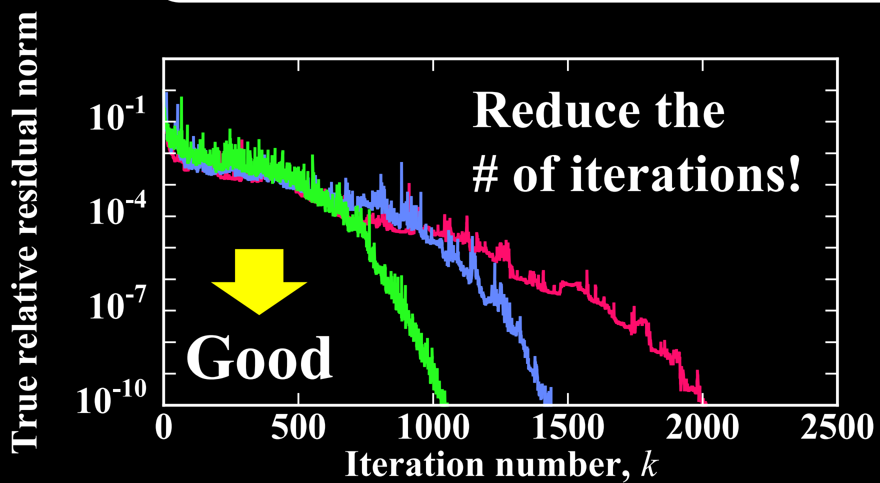
Block Krylov subspace methods

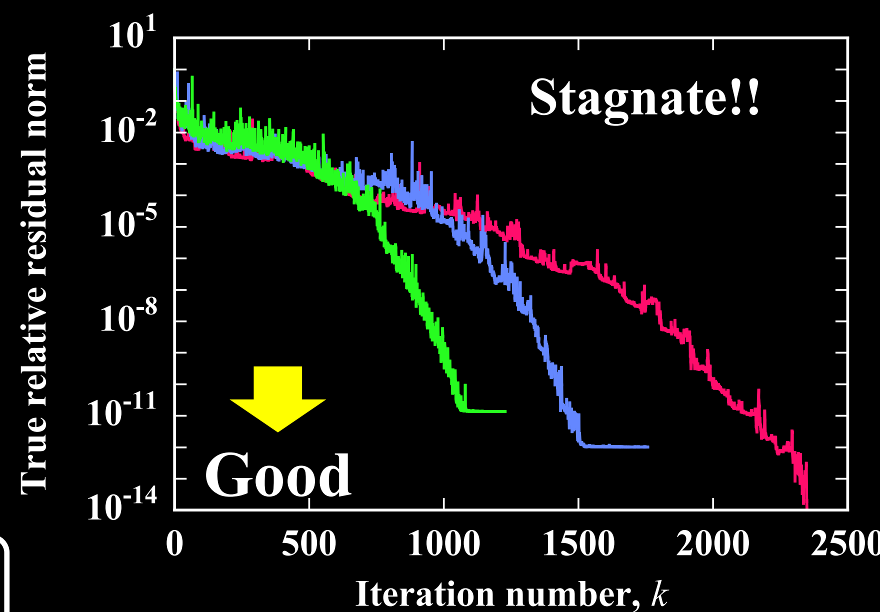
▪ Block BiCG	O'Leary (1980)
▪ Block GMRES	Vital (1990)
▪ Block QMR	Freund (1997)
▪ Block BiCGSTAB	El Guennouni (2003)

Linear system with multiple right-hand sides can be efficiently solved by using Block Krylov methods

What is "efficient?"

Residual norm of Block Krylov methods may converge in smaller number of iterations than that of Krylov methods





True relative residual : $\|B - AX_k\|_F / \|B\|_F$

Fig. 1. True relative residual norm histories of Block BiCGSTAB.
■ : $L = 1$, ■ : $L = 2$, ■ : $L = 4$.

Pros and cons of Block Krylov subspace methods

Pros

- Linear system with L RHSs can be solved simultaneously.
- The number of iterations of Block Krylov subspace methods may smaller than that of Krylov subspace methods.

Cons

- The accuracy of the obtained approximate solution may not good if the stopping condition is satisfied!
- The relative residual norm may not converge due to the influence of numerical instability when the number of right-hand sides L is large.

Objectives of this research

Objectives

1. We develop a Block Krylov subspace method for computing high accuracy solutions.

2. We improve the numerical instability of Block Krylov subspace methods when the number of right-hand sides is large.

§ 2 Development of a high accuracy linear solver

Definition of a residual and an operation

Linear systems

$$AX = B,\ A \in \mathbb{C}^{n\times n},\ X, B \in \mathbb{C}^{n\times L}$$

The (k+1)th residual

$$\begin{aligned} R_{k+1} &= B - AX_{k+1} \\ &\equiv (\mathcal{H}_{k+1}\mathcal{R}_{k+1})(A) \circ R_0 \end{aligned}$$

Def. of an operation $\circ$

$$\mathcal{M}_k(A) \circ V \equiv \sum_{j=0}^{k} A^j V M_j$$

Here, $\mathcal{M}_k(z) \equiv \sum_{j=0}^{k} z^j M_j$,

$M_j \in \mathbb{C}^{L\times L},\ V \in \mathbb{C}^{n\times L}$.

Recursions of polynomials

$$\mathcal{R}_0(z) = \mathcal{P}_0(z) = I_L, \qquad \mathcal{H}_0(z) = 1,$$

$$\mathcal{R}_{k+1}(z) = \mathcal{R}_k(z) - z\mathcal{P}_k(z)\alpha_k, \qquad \mathcal{H}_{k+1}(z) = (1 - \zeta_k z)\mathcal{H}_k(z)$$

$$\mathcal{P}_{k+1}(z) = \mathcal{R}_{k+1}(z) + \mathcal{P}_k(z)\beta_k$$

Here, $\alpha_k, \beta_k \in \mathbb{C}^{L\times L},\ \zeta_k \in \mathbb{C}$.

Derivation of recurrence formulas

There are two ways of derivation of recurrence formulas

The $(k+1)$th residual

$$
\begin{aligned}
R_{k+1} &= B - AX_{k+1} \\
&\equiv (\mathcal{H}_{k+1}\mathcal{R}_{k+1})(A) \circ R_0
\end{aligned}
$$

Expand from $\mathcal{H}_{k+1}$

Block BiCGSTAB

Algorithm of the Block BiCGSTAB method

$X_0 \in \mathbb{C}^{n\times L}$ is an initial guess,
Compute $R_0 = B - AX_0$,
Set $P_0 = R_0$,
Choose $\tilde{R}_0 \in \mathbb{C}^{n\times L}$,
For $k = 0, 1, \ldots,$ until $\|R_k\|_F \le \varepsilon \|B\|_F$ do:
Solve $(\tilde{R}_0^H A P_k)\alpha_k = \tilde{R}_0^H R_k$ for α_k,

$$
\begin{aligned}
T_k &= R_k - AP_k\alpha_k, \\
\zeta_k &= \frac{\mathrm{Tr}[(AT_k)^H T_k]}{\mathrm{Tr}[(AT_k)^H AT_k]}, \\
X_{k+1} &= X_k + P_k\alpha_k + \zeta_k T_k, \\
R_{k+1} &= T_k - \zeta_k AT_k,
\end{aligned}
$$

Solve $(\tilde{R}_0^H V_k)\beta_k = -\tilde{R}_0^H Z_k$ for β_k,

$$
P_{k+1} = R_{k+1} + (P_k - \zeta_k V_k)\beta_k,
$$

End

Relationship between the true residual and the recursive residual

- **Theoretically, the true residual $B - AX_k$ is equal to the recursive residual R_k.**

$$B - AX_k = R_k$$

- **If the recursive residual R_k becomes zero matrix, then the true residual $B - AX_k$ also becomes zero matrix.**
- **Hence, X_k is the exact solution.**

However, the equation $B - AX_k = R_k$ is not satisfied in the numerical computation.

The error matrix in Block BiCGSTAB

Recursions of X_{k+1} and R_{k+1}

$$X_{k+1} = X_k + P_k\alpha_k + \zeta_k T_k$$

$$R_{k+1} = R_k - AP_k\alpha_k - \zeta_k AT_k$$

Here,

$X_k, R_k, P_k, T_k \in \mathbb{C}^{n\times L}$,
$\alpha_k \in \mathbb{C}^{L\times L}$, $\zeta_k \in \mathbb{C}$.

Expansion of recursions

$$X_{k+1} = X_0 + \sum_{j=0}^{k} P_j\alpha_j + \sum_{j=0}^{k} \zeta_j T_j$$

$$R_{k+1} = R_0 - \sum_{j=0}^{k} (AP_j)\alpha_j - \sum_{j=0}^{k} \zeta_j (AT_j)$$

The relationship between the true res. and the recursive res.

$$B - AX_{k+1} = R_{k+1} + \sum_{j=0}^{k} \left[(AP_j)\alpha_j - A(P_j\alpha_j)\right] + \sum_{j=0}^{k} \left[\zeta_j(AT_j) - A(\zeta_j T_j)\right]$$

Example of effect of the error matrix

Matrix : JPWH991 (from Matrix Market)

#RHS : $L = 4$

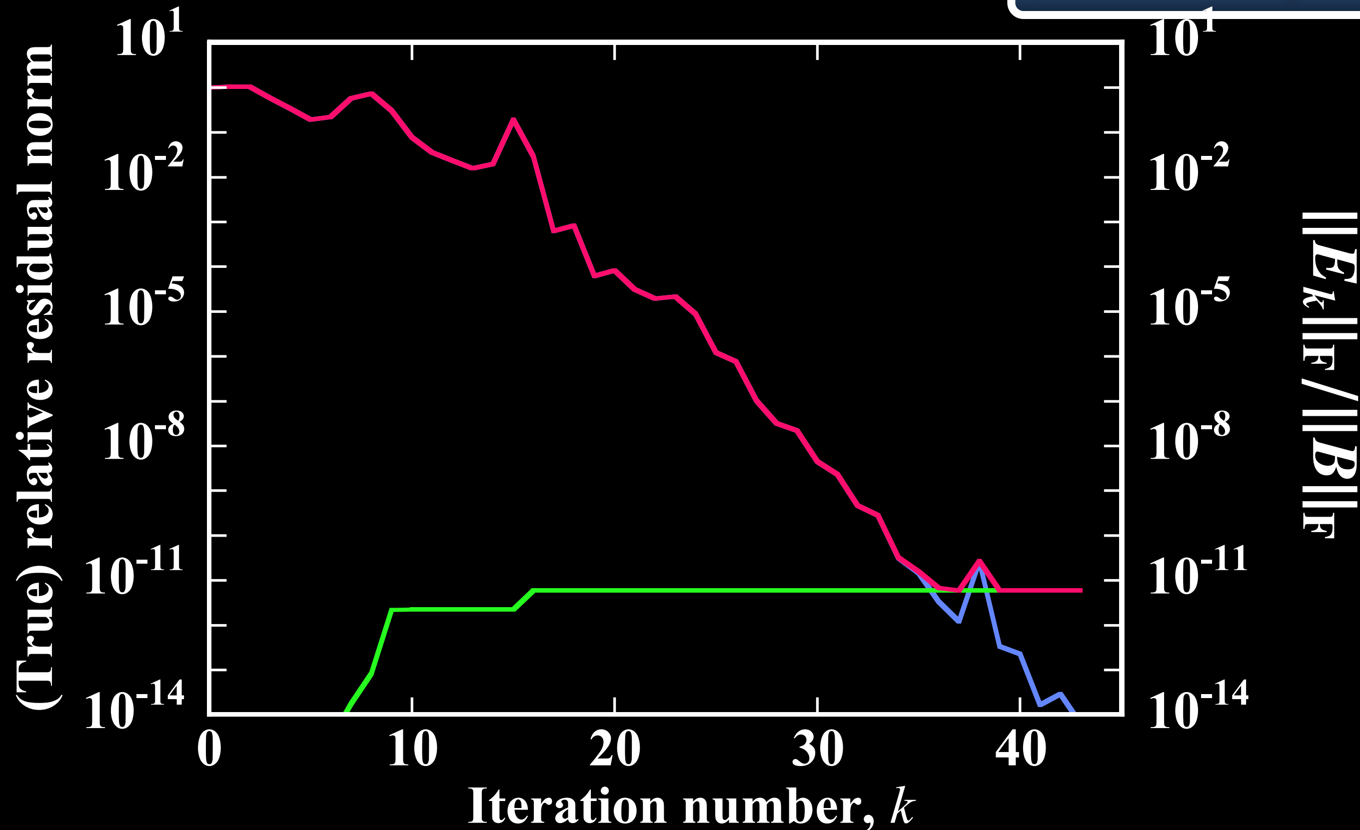


Fig. 2. Relation between the true rel. res. and the error matrix norm.

■ : $\|B - AX_k\|_F / \|B\|_F$, ■ : $\|R_k\|_F / \|B\|_F$, ■ : $\|E_k\|_F / \|B\|_F$.

Here, $E_k = \sum_{j=0}^{k-1} \left[(AP_j)\alpha_j - A(P_j\alpha_j) \right] + \sum_{j=0}^{k-1} \left[\zeta_j(AT_j) - A(\zeta_j T_j) \right].$

There are two ways of derivation of recurrence formulas

The (k+1)th residual

$$\begin{aligned} R_{k+1} &= B - AX_{k+1} \\ &\equiv (\mathcal{H}_{k+1}\mathcal{R}_{k+1})(A) \circ R_0 \end{aligned}$$

Expand from $\mathcal{H}_{k+1}$ → **Block BiCGSTAB**

Expand from $\mathcal{R}_{k+1}$ → **Block BiCGGR**

Algorithm of the Block BiCGGR method

$X_0 \in \mathbb{C}^{n\times L}$ is an initial guess,
Compute $R_0 = B - AX_0$,
Set $P_0 = R_0$,
Choose $\tilde{R}_0 \in \mathbb{C}^{n\times L}$,
For $k = 0, 1, \ldots,$ until $\|R\|_{\mathrm{F}} \le \varepsilon \|B\|_{\mathrm{F}}$ do:
  Solve $(\tilde{R}_0^{\mathrm{H}} A P_k)\alpha_k = \tilde{R}_0^{\mathrm{H}} R_k$ for α_k,

$$\zeta_k = \frac{\mathrm{tr}[(AR_k)^{\mathrm{H}} R_k]}{\mathrm{tr}[(AR_k)^{\mathrm{H}} AR_k]},$$

$$U_k = (P_k - \zeta_k A P_k)\alpha_k,$$

$$X_{k+1} = X_k + \zeta_k R_k + U_k,$$

$$R_{k+1} = R_k - \zeta_k A R_k - AU_k,$$

  Solve $(\tilde{R}_0^{\mathrm{H}} R_k)\gamma_k = \tilde{R}_0^{\mathrm{H}} R_{k+1}/\zeta_k$ for γ_k,

$$P_{k+1} = R_{k+1} + U_k \gamma_k,$$

$$AP_{k+1} = AR_{k+1} + AU_k \gamma_k,$$

End For

The error matrix in Block BiCGGR

Recursions of X_{k+1} and R_{k+1}

$$X_{k+1} = X_k + \zeta_k R_k + U_k$$
$$R_{k+1} = R_k - \zeta_k A R_k - A U_k$$

Expansion of recursions

$$X_{k+1} = X_0 + \sum_{j=0}^{k} \zeta_j R_j + \sum_{j=0}^{k} U_j$$
$$R_{k+1} = R_0 - \sum_{j=0}^{k} \zeta_j (A R_j) - \sum_{j=0}^{k} A U_j$$

Here,
$X_k, R_k, U_k \in \mathbb{C}^{n \times L}$, $\zeta_k \in \mathbb{C}$.

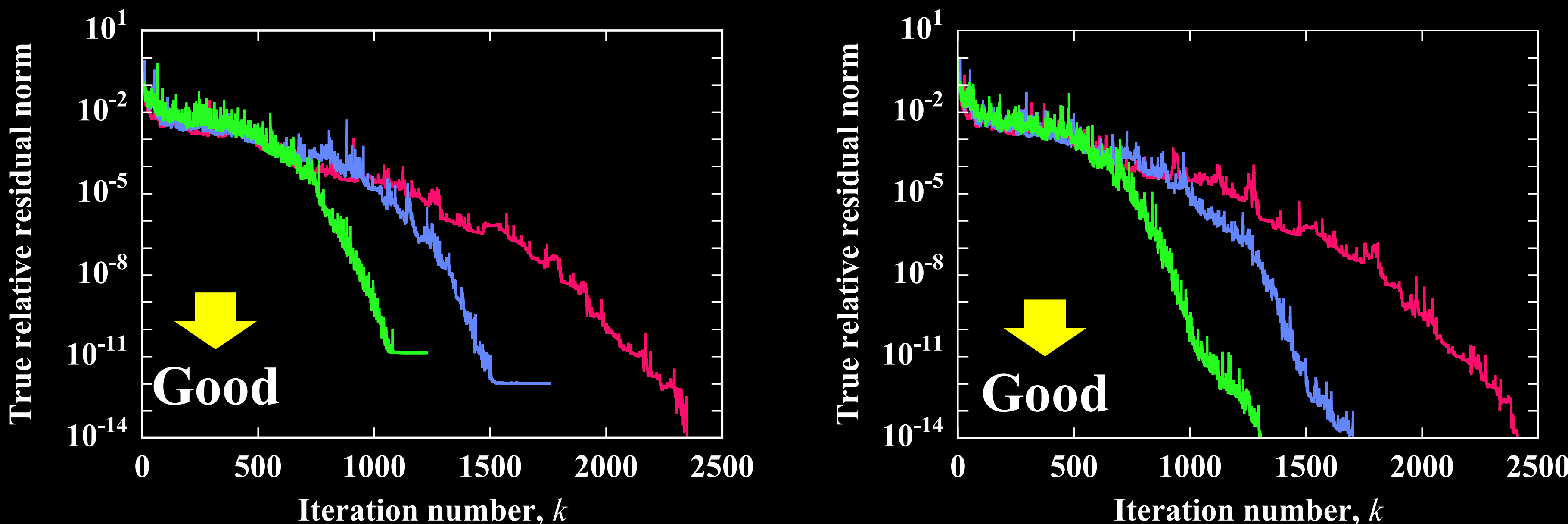
The relation between the true res. and the recursive residual

$$B - A X_{k+1} = R_{k+1} + \sum_{j=0}^{k} \left[\zeta_j (A R_j) - A(\zeta_j R_j) \right]$$

Comparison of two methods

§ 2 Development of a high accuracy linear solver

Stagnate! **Converge!**



(a) Block BiCGSTAB. (b) Block BiCGGR.

Fig. 3. True relative residual histories of two methods.
: $L = 1$, : $L = 2$, : $L = 4$.

§ 3 Stabilization of Block BiCGGR

Numerical instability when #RHSs is large

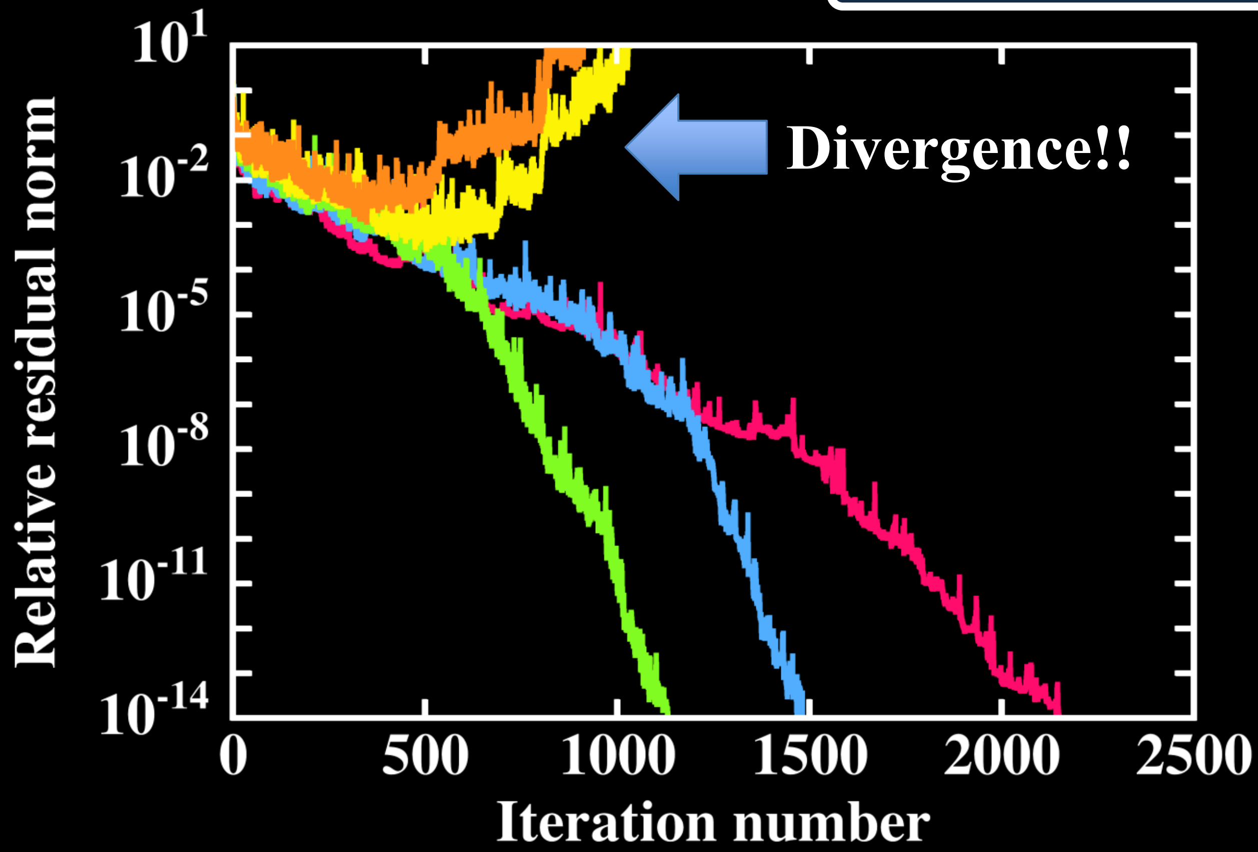


Fig. 4. Relative residual histories of the Block BiCGGR method.

■ : $L = 1$, ■ : $L = 2$, ■ : $L = 4$. , ■ : $L = 8$, ■ : $L =$ 12.

Pros and cons of Block Krylov subspace methods

Pros

- Linear system with *L* RHSs can be solved simultaneously.
- The number of iterations of Block Krylov methods is may smaller than that of Krylov subspace methods.

Cons

- The accuracy of the obtained approximate solution may not good even if the stopping condition is satisfied!
- The relative residual norm may not converge due to the influence of numerical instability when the number of right-hand sides *L* is large.

Cause of numerical instability of Block BiCGGR

$X_0 \in \mathbb{C}^{n\times L}$ is an initial guess,
Compute $R_0 = B - AX_0$,
Set $P_0 = R_0$,
Choose $\tilde{R}_0 \in \mathbb{C}^{n\times L}$,
For $k = 0, 1, \ldots,$ until $\|R\|_F \le \varepsilon \|B\|_F$ do:
  Solve $(\tilde{R}_0^H A P_k)\alpha_k = \tilde{R}_0^H R_k$ for α_k,

$$\zeta_k = \frac{\mathrm{tr}[(AR_k)^H R_k]}{\mathrm{tr}[(AR_k)^H AR_k]},$$

$$U_k = (P_k - \zeta_k AP_k)\alpha_k,$$

$$X_{k+1} = X_k + \zeta_k R_k + U_k,$$

$$R_{k+1} = R_k - \zeta_k AR_k - AU_k,$$

  Solve $(\tilde{R}_0^H R_k)\gamma_k = \tilde{R}_0^H R_{k+1}/\zeta_k$ for γ_k,

$$P_{k+1} = R_{k+1} + U_k\gamma_k,$$

$$AP_{k+1} = AR_{k+1} + AU_k\gamma_k,$$

End For

Small linear systems need to be solved to obtain $L \times L$ matrices α_k, γ_k.

Cause of numerical instability

If the linear independence of R_k and P_k is lost, the small coefficient matrices become ill-condition.

Stabilization of Block BiCGGR by residual orthonormalization

In order to improve the numerical instability ⋯

We consider to improve linear independence of the vectors.

→ Perform the orthonormalization of vectors.

In this stydy ⋯

We develop the Block BiCGGRRO method. The residual matrix R_k of this method is orthonormalized as follows.

$$R_k = Q_k \xi_k, \;\; Q_k^{\mathrm{H}} Q_k = I_L, \;\; \xi_k \in \mathbb{C}^{L \times L}$$

Algorithm of the Block BiCGGRRO method

§ 3 Stabilization of Block BiCGGR

Orthonormalization

$X_0 \in \mathbb{C}^{n\times L}$ is an initial guess,

Compute $Q_0\xi_0 = B - AX_0$,

Set $S_0 = Q_0$,

Choose $\tilde{R}_0 \in \mathbb{C}^{n\times L}$,

For $k = 0, 1, \ldots,$ until $\|\xi_k\|_{\mathrm{F}} \le \varepsilon\|B\|_{\mathrm{F}}$ do:

Solve $(\tilde{R}_0^{\mathrm{H}} AS_k)\alpha_k = \tilde{R}_0^{\mathrm{H}} Q_k$ for α_k,

$$\zeta_k = \arg\min_{\zeta} \|Q_k\xi_k - \zeta AQ_k\xi_k\|_{\mathrm{F}},$$

$$V_k = (S_k - \zeta_k AS_k)\alpha_k,$$

$$X_{k+1} = X_k + [\zeta_k Q_k + V_k]\,\xi_k,$$

$$Q_{k+1}\tau_{k+1} = Q_k - \zeta_k AQ_k - AV_k,$$

$$\xi_{k+1} = \tau_{k+1}\xi_k,$$

Solve $(\tilde{R}_0^{\mathrm{H}} Q_k)\gamma_k = \tilde{R}_0^{\mathrm{H}} Q_{k+1}/\zeta_k$ for γ_k,

$$S_{k+1} = Q_{k+1} + V_k\gamma_k,$$

$$AS_{k+1} = AQ_{k+1} + AV_k\gamma_k,$$

End For

§ 4 Numerical Experiments

Test problem

Test problem

Linear system with multiple right-hand sides derived from Lattice QCD.

$$AX = B$$

n = 1, 572, 864,   nnz(A) = 80, 216, 064,

the number of nnz(A) per row is 51.

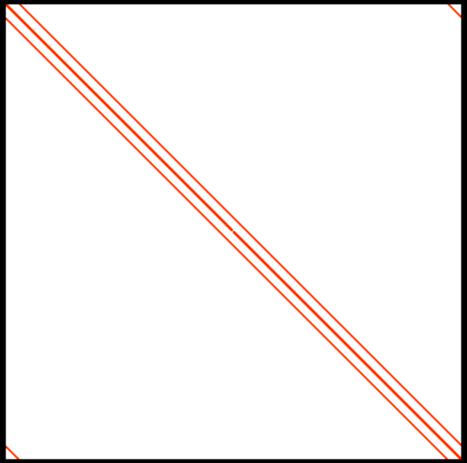


Fig. 5. Nonzero structure.

Experimental environment and conditions

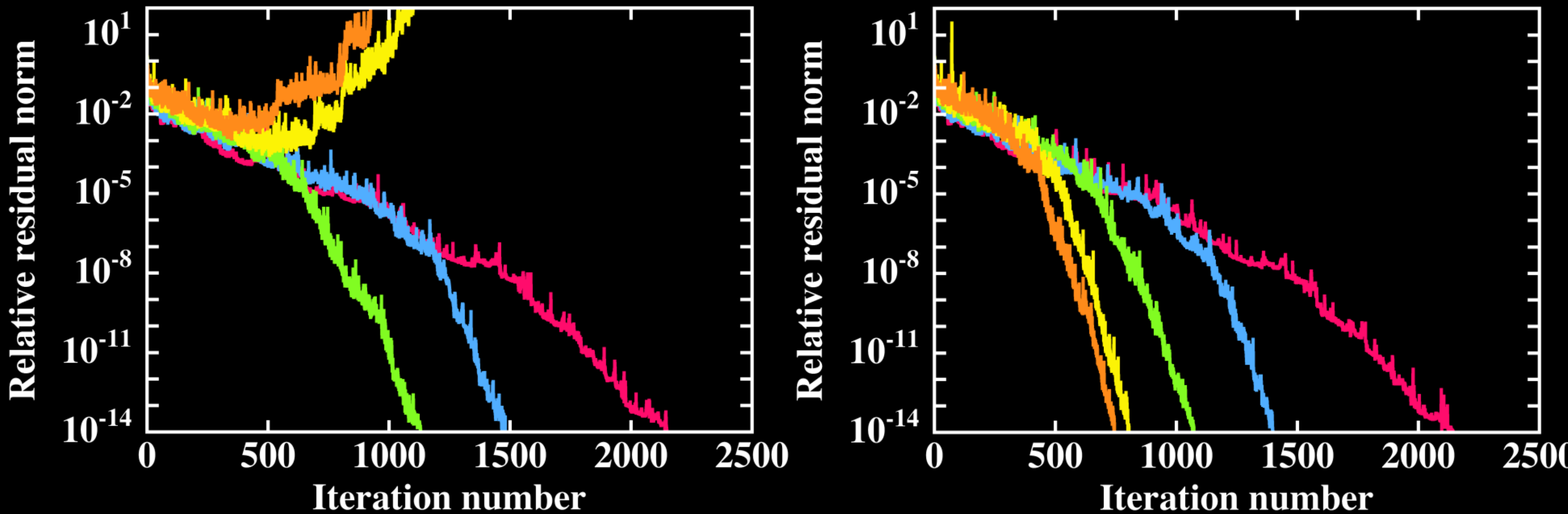
Table 1. Experimental environment.

CPU	AMD Opteron 6180 SE 2.5GHz × 4
Memory	256.0GBytes
Compiler	PGI Fortran ver. 11.5
Compile option	`-O3 -tp=x64 -mp`

Table 2. Experimental conditions.

Initial solution X_0	$[0, 0, \ldots, 0]$
Right hand side B	$[e_1, e_2, \ldots, e_L]$
Shadow residual $\tilde{R}_0$	Random number
Stopping criterion	$\|\|R_k\|\|_F / \|\|B\|\|_F \leq 1.0 \times 10^{-14}$ or $\|\|R_k\|\|_F / \|\|B\|\|_F \geq 1.0 \times 10^{6}$

Comparison of Block BiCGGR and Block BiCGGRO



(a) Block BiCGGR. (b) Block BiCGGRRO.

Fig. 6. Relative residual histories of BiCGGR and BiCGGRRO.

■ : $L = 1$, ■ : $L = 2$, ■ : $L = 4$, ■ : $L = 8$, ■ : $L = 12$.

Comparison of Block BiCGGR and Block BiCGGRRO

Table 3. Results of Block BiCGGR.

	$L = 1$	$L = 2$	$L = 4$	$L = 8$	$L = 12$
Iter.	2148	1481	1131	Divergence	Divergence
TRR	9.9×10^{-15}	6.2×10^{-15}	9.3×10^{-15}		
Time	107.7	106.6	152.5		

Table 4. Results of Block BiCGGRRO.

	$L = 1$	$L = 2$	$L = 4$	$L = 8$	$L = 12$
Iter.	2139	1421	1006	894	800
TRR	8.2×10^{-15}	8.9×10^{-15}	1.1×10^{-14}	1.1×10^{-14}	1.2×10^{-14}
Time	111.3	113.1	161.5	341.7	521.3

Iter. : Number of iterations, TRR : True relative residual norm,

Time : Computational time in seconds.

Block BiCGGRRO can also generate high accuracy solutions!

Summary

1. We developed the Block BiCGGR method. This method can generate high accuracy solutions compared to the conventional method. This method was developed through the collaborative research with Division of Particle Physics of CCS.

2. We improved the numerical instability of the Block BiCGGR method by performing the residual orthonormalization when the number of right-hand sides is large.