Modified Block BiCGSTAB for Lattice QCD

Yoshifumi Nakamura

RIKEN AICS

10 Jun 2011

計算科学研究機構
Advanced Institute for Computational Science

Introduction

Motivation: Understanding particle physics

Introduction

Motivation: Understanding particle physics

Non-perturbative approach of Quantum Chromo Dynamics(QCD)

Introduction

Motivation: Understanding particle physics

Non-perturbative approach of Quantum Chromo Dynamics(QCD)

Lattice QCD (LQCD)

Introduction

Motivation: Understanding particle physics

Non-perturbative approach of Quantum Chromo Dynamics(QCD)

Lattice QCD (LQCD)

But huge CPU resources require for LQCD simulations

Introduction

Motivation: Understanding particle physics

Non-perturbative approach of Quantum Chromo Dynamics(QCD)

Lattice QCD (LQCD)

But huge CPU resources require for LQCD simulations

It is important to reduce cost for LQCD simulations

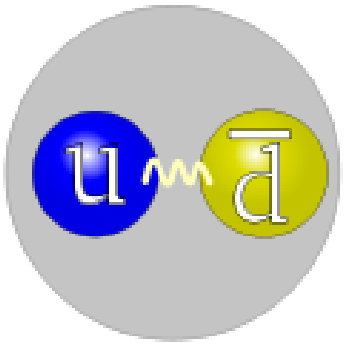
Contents

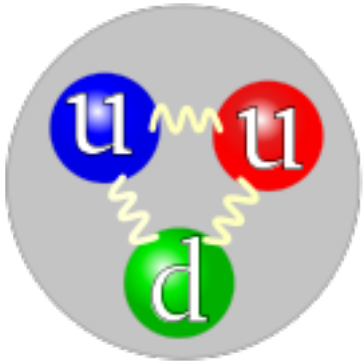
- QCD and LQCD
- Modified Block BiCGSTAB for Lattice QCD

Quantum Chromo Dynamics(QCD)

- basic theorem of hadron physics
- describing interaction for quark and gluon
- typical scale is **0.000 000 000 000 001** m = 1 fm

hadron(color-neutral): meson(π,K...), baryon(p,n...)





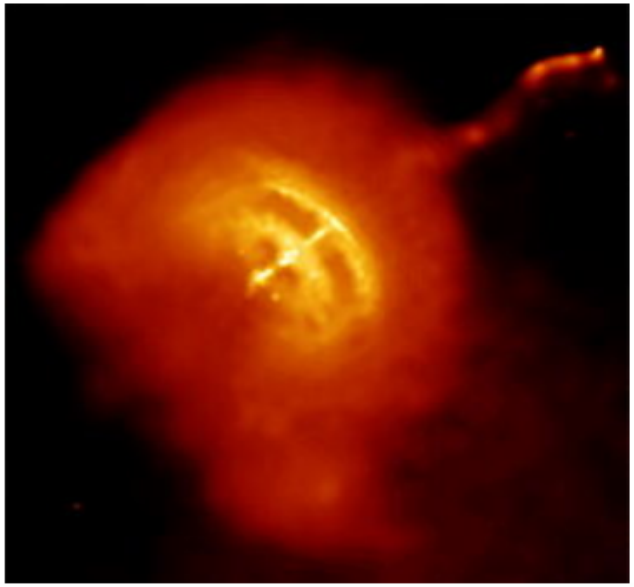
quark: 3 (R, G, B)

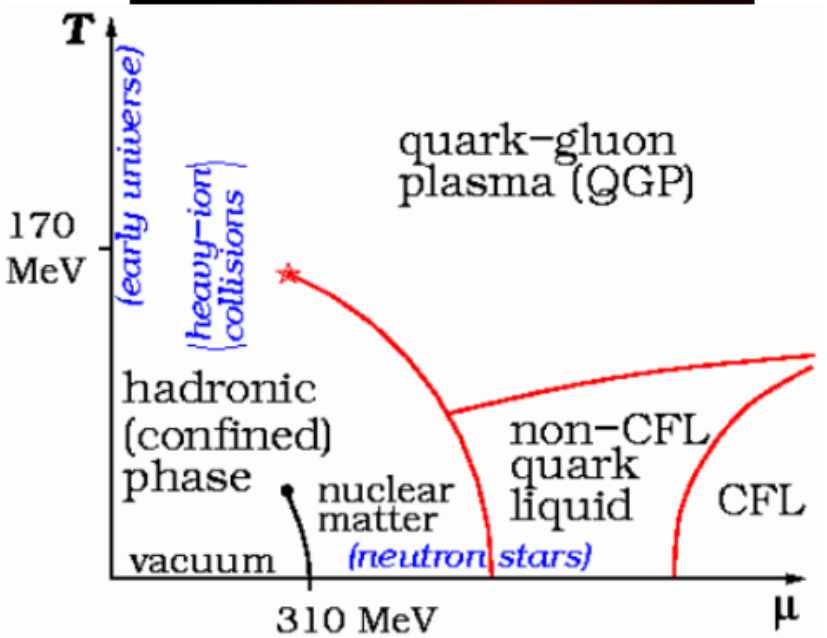
gluon: 8 (mass 0, charge 0, spin 1)

quark: 6 flavours

	u(up)	c(charm)	t(top)
mass	1.7-3.3 MeV	$1.27^{+0.07}_{-0.09}$ GeV	172.0(22) GeV
charge	2/3	2/3	2/3
spin	1/2	1/2	1/2

	d(down)	s(strange)	b(bottm)
mass	4.1-5.8 MeV	101^{+29}_{-21} MeV	$4.19^{+0.18}_{-0.06}$ GeV
charge	-1/3	-1/3	-1/3
spin	1/2	1/2	1/2





images from NASA and wikipedia

QCD action

$$L = \sum_i \bar{\psi}_i D(m_i) \psi_i - \frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu}$$

$$F^a_{\mu\nu}: \quad \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + i g f^{abc} A^b_\mu A^c_\nu$$

$$D(m_i): \quad \gamma_\mu (i\partial_\mu + g A^a_\mu T^a) - m_i$$

Path integral

Expectation value of observable $\mathcal{O}$:

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int dA d\bar{\psi} d\psi \; e^{i \int d^4x L(x,t)} \; \mathcal{O}(A, \bar{\psi}, \psi)$$

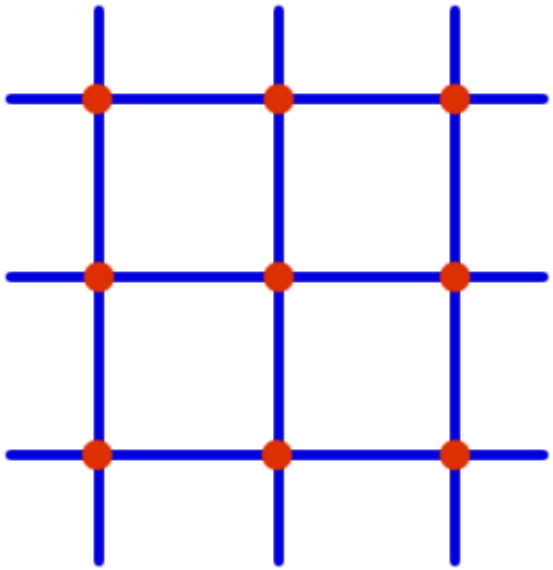
QED: perturbation + renormalization

QCD: perturbation does not work at low energy since coupling constant is big

Lattice QCD

Non-perturbative approach to solving QCD

- space - time discretization
  - quark field: color × spinor / site → 12 complex numbers
  - gluon field: SU(3) matrix / link → 9 complex numbers



Lattice QCD simulation

fermion field(Grassmann number)

→ pseudo-fermion field(usual number)

$$\int d\bar{\psi}_i d\psi_i \quad \exp(-\sum_{i=1}^{2} \bar{\psi}_i D(m_i)\psi_i) = \det D(m_1) \det D(m_2)$$

when $m = m_1 = m_2$

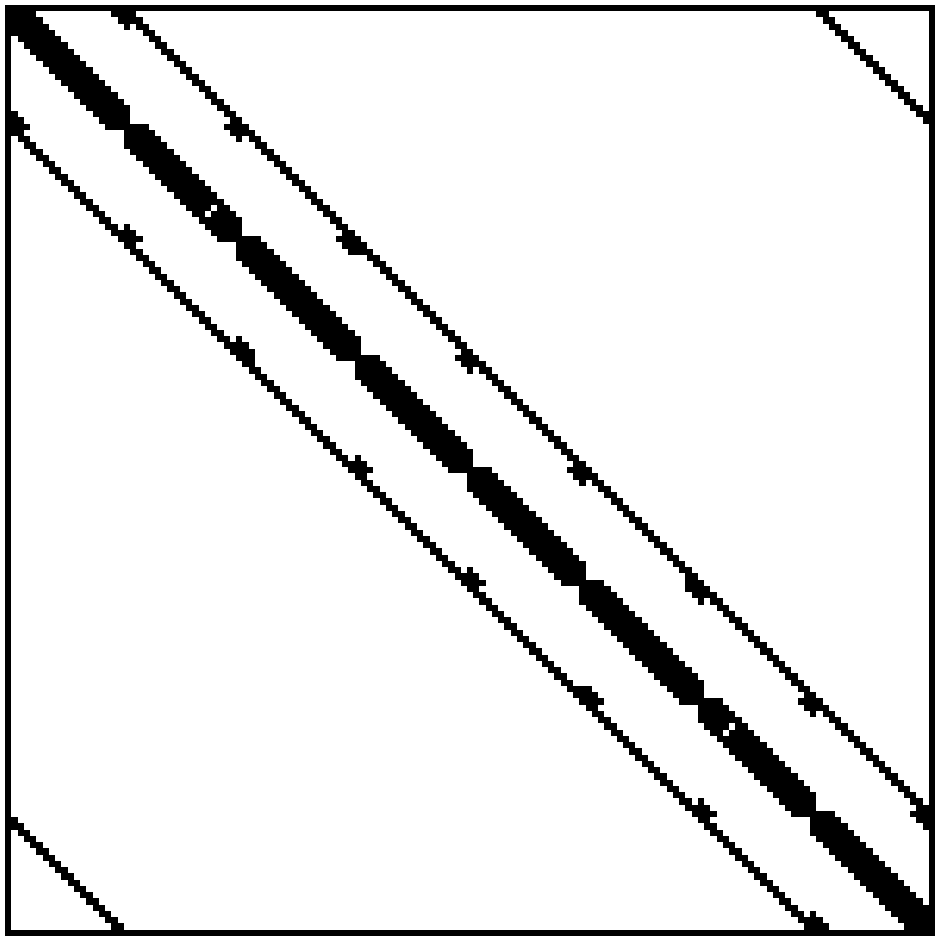
$$\det D(m)^2 = \int d\phi_i^\dagger d\phi_i \quad \exp(-\sum_i \phi_i^\dagger \left[D^\dagger(m) D(m) \right]^{-1} \phi_i)$$

condition number increases as m decreasing

D is $12V \times 12V$ complex sparse matrix

e.g. $V = 32^3 \times 64 \Rightarrow O(10^7)$

Sparsity pattern of Wilson-Dirac matrix



$$n = 8^4 \times 12$$

Modified Block BiCGSTAB for Lattice QCD

Y. Nakamura, K. -I. Ishikawa, Y. Kuramashi, T. Sakurai, H. Tadano (2011)

Outline

- Krylov subspace method
- Block Krylov subspace method
- Algorithm of Modified Block BiCGSTAB for Lattice QCD
- Numerical test results
- Summary

Krylov subspace method

iterative method to solve system of linear equations

$$Ax = b$$

$$x = A^{-1}b$$

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

by using matrix-vector multiplication

Krylov subspace

$$\mathcal{K}_k \equiv \mathbf{span}(v, Av, A^2v, ..., A^{k-1}v)$$

1. guess initial approx. solution vector x_0 for $Ax = b$
2. renew approx. solutions $x_1, x_2, x_3, ...$ with keeking condition of $x_k - x_0 \in \mathcal{K}_k(A, r_0) = \mathbf{span}(r_0, Ar_0, ...A^{k-1}r_0)$

First residual vector:

$$r_0 = b - Ax_0$$

Approximate solution:

$$x_k = x_0 + \sum_{i=0}^{k-1} c_i A^i r_0$$

Conjugate gradient (CG) method

Hestenes, Stiefel (1952)

Krylov subspace method for symmetric positive definite

- minimize $f(x) = (x, Ax) - 2(x, b)$
- (k+1)-th approx. solution vector: $x_{k+1} = x_k + \alpha p_k$
- (k+1)-th research vector: $p_{k+1} = r_{k+1} + \beta p_k$
- $(p_{k+1}, Ap_k) = 0$

(k+1)-th residual vector:

$$r_{k+1} = r_k - \alpha A p_k$$

Properties of CG method

- conjugate property

$$(p_i, Ap_j) = 0, \qquad i \neq j$$

- orthogonality

$$(r_i, r_j) = 0, \qquad i \neq j$$

System converges theoretically by ’n’ iteration at most

Block Krylov subspace method

$$AX = B$$

$$X = [x^{(1)}, x^{(2)}, ..., x^{(L)}]$$

$$B = [b^{(1)}, b^{(2)}, ..., b^{(L)}]$$

$$X_k - X_0 \in \mathcal{K}_k(A, R_0) = \mathrm{span}(R_0, AR_0, ...A^{k-1}R_0)$$

Approx. solutions X_k:

$$X_k = X_0 + \sum_{i=0}^{k-1} A^i R_0 \gamma_i$$

γ_i is $L \times L$ matrix

To solve $x^{(i)}$, one can use information of $\mathcal{K}_k(A, r_0^{(j)})$

Better cache usage

at Matrix Vector multiplication (MVM)

$$\begin{bmatrix} w_1^{(1)} & w_1^{(2)} \\ \vdots & \vdots \\ w_n^{(1)} & w_n^{(2)} \end{bmatrix} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix} \begin{bmatrix} v_1^{(1)} & v_1^{(2)} \\ \vdots & \vdots \\ v_n^{(1)} & v_n^{(2)} \end{bmatrix}$$

One can calculate $a_{1,1} \times v_1^{(2)}$ (not $a_{1,2} \times v_2^{(1)}$) right after $a_{1,1} \times v_1^{(1)}$

Wilson-Dirac operator

$$A\phi = \sum_{x=1}^{L_x \times L_y \times L_z \times L_t} (\phi_x - \kappa \eta_x), \qquad \eta_x = \sum_{\mu=1}^{4} \left[(1-\gamma_\mu) U_{x,\hat{\mu}} \phi_{x+\hat{\mu}} + (1+\gamma_\mu) U^{\dagger}_{x-\mu,\hat{\mu}} \phi_{x-\hat{\mu}} \right]$$

To computer η_x (hopping term multiplication)

Flops : 1320
store : 12 complex numbers
load : 72 + 96 complex numbers, for U and ϕ

Wilson-Dirac operator

$$A\phi = \sum_{x=1}^{L_x \times L_y \times L_z \times L_t} (\phi_x - \kappa \eta_x), \qquad \eta_x = \sum_{\mu=1}^{4} \left[(1-\gamma_\mu) U_{x,\hat{\mu}} \phi_{x+\hat{\mu}} + (1+\gamma_\mu) U^\dagger_{x-\mu,\hat{\mu}} \phi_{x-\hat{\mu}} \right]$$

To computer η_x (hopping term multiplication)

Flops : 1320
store : 12 complex numbers
load : 72 + 96 complex numbers, for U and ϕ

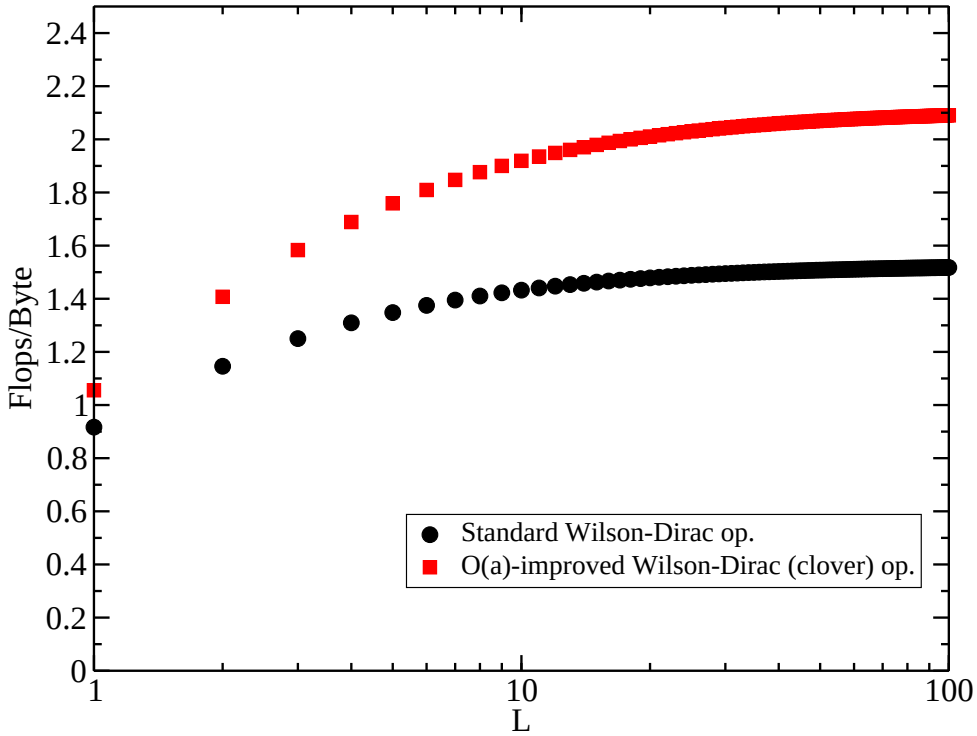
Hopping term mult. for multiple right hand sides

$$\eta_x{}^{(1,\ldots,L)} = \sum_{\mu=1}^{4} \left[(1-\gamma_\mu) U_{x,\hat{\mu}} \phi_{x+\hat{\mu}}{}^{(1,\ldots,L)} + (1+\gamma_\mu) U^\dagger_{x-\mu,\hat{\mu}} \phi_{x-\hat{\mu}}{}^{(1,\ldots,L)} \right]$$

Size of 8 U is 576 (1152) bytes in the single (double) precision
Able to keep in low level cache and use L times

Dirac op. Flops/Byte

with the single precision for multiple right hand sides



Block Krylov subspace method is suitable for recent high performance computer architecture

Numerical difficulty

Dirac matrix in lattice QCD is non-Hermitian

Numerical difficulty

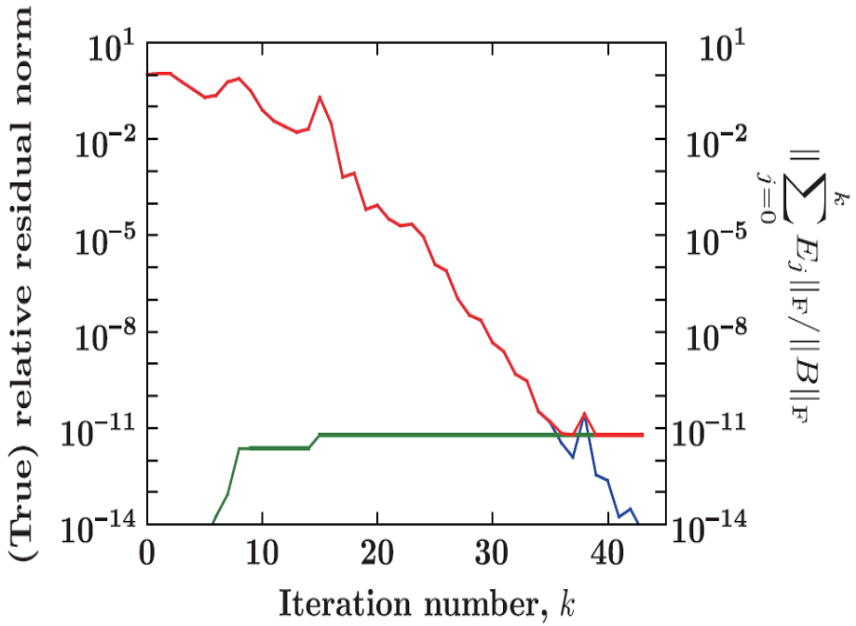
Dirac matrix in lattice QCD is non-Hermitian

→ Block BiCGSTAB (A. El Guennouni, K. Jbilou, H. Sadok (2003))

Numerical difficulty

Dirac matrix in lattice QCD is non-Hermitian

→ Block BiCGSTAB (A. El Guennouni, K. Jbilou, H. Sadok (2003))

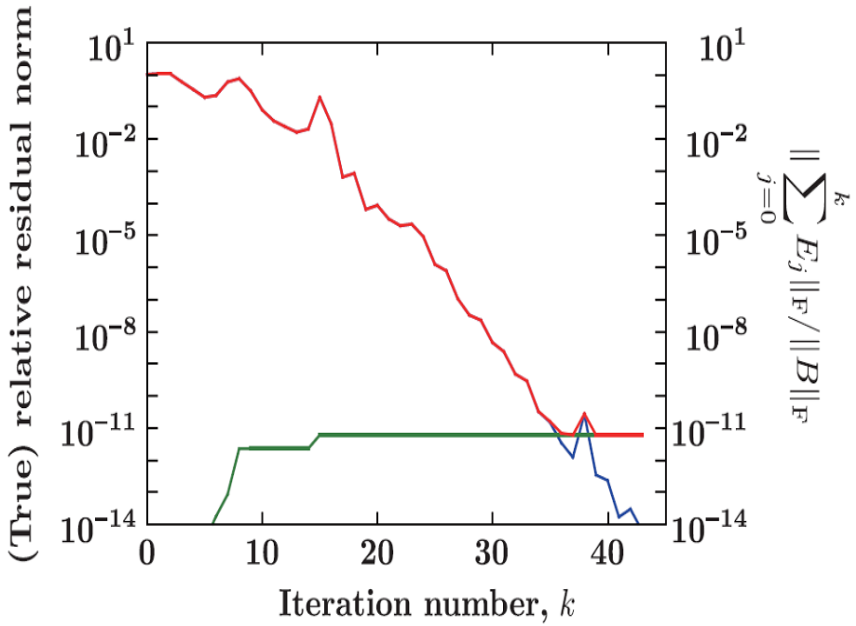


- Block BiCGSTAB has numerical error due to multiple right hand sides

Numerical difficulty

Dirac matrix in lattice QCD is non-Hermitian

→ Block BiCGSTAB (A. El Guennouni, K. Jbilou, H. Sadok (2003))



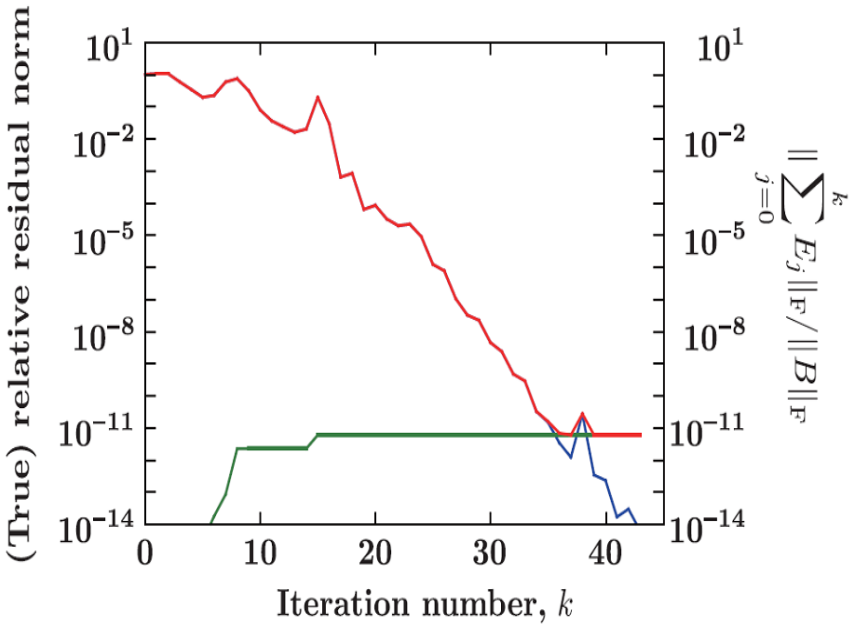
- Block BiCGSTAB has numerical error due to multiple right hand sides
- Block BiCGGR improved this problem significantly

Tadano, Sakurai, Kuramashi (2009)

Numerical difficulty

Dirac matrix in lattice QCD is non-Hermitian

→ Block BiCGSTAB (A. El Guennouni, K. Jbilou, H. Sadok (2003))



- Block BiCGSTAB has numerical error due to multiple right hand sides
- Block BiCGGR improved this problem significantly

Tadano, Sakurai, Kuramashi (2009)

Block BiCGGR sometimes fails to converge

Further robustness and convergence are needed!!

Modified Block BiCGSTAB algorithm

1 init. $X \in \mathbb{C}^{N\times L}$
2 $R = B - AX$
3 $P = R$
4 choose. $\tilde{R} \in \mathbb{C}^{N\times L}$
WHILE $\max_i(|r^{(i)}|/|b^{(i)}|) \leq \epsilon$
4.1 QR decomp $P = Q\gamma$, $P <= Q$
4.2 $U = MP$
4.3 $V = AU$
4.4 solve $(\tilde{R}^H V)\alpha = \tilde{R}^H R$ for α
4.5 $T = R - V\alpha$
4.6 $S = MT$
4.7 $Z = AS$
4.8 $\zeta = \mathrm{Tr}(Z_k^H T_k)/\mathrm{Tr}(Z_k^H Z_k)$
4.9 $X = X + U\alpha + \zeta S$
4.10 $R = T - \zeta Z$
4.11 solve $(\tilde{R}^H V)\beta = -\tilde{R}^H Z$ for β
4.12 $P = R + (P - \zeta V)\beta$
END

- by QR decomposition, numerical error ↘ convergence ↗
- minimize comm. overhead by domain decomposition preconditioning with single precision acceleration

Preconditioning

Original linear system:

$$\boldsymbol{Ax = b}$$

Preconditioned system:

$$\boldsymbol{x = My}$$

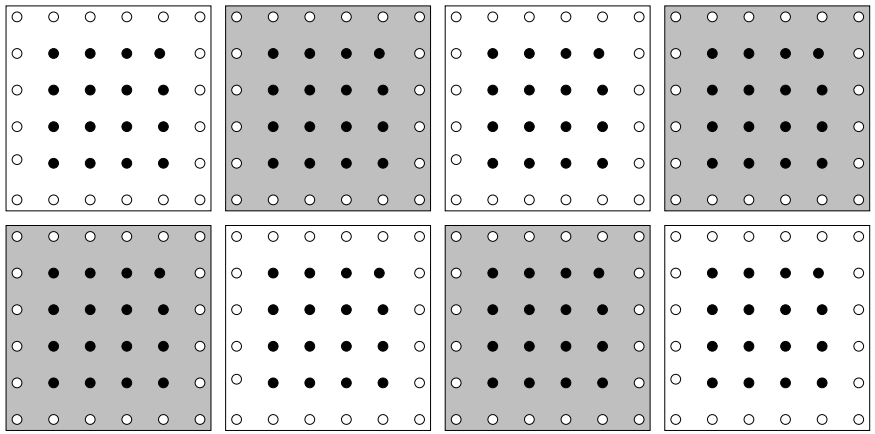
$$\boldsymbol{AMy = b}$$

Preconditioner

$$\boldsymbol{M \approx A^{-1}}$$

Condition number: $\boldsymbol{AM < A}$

Domain decomposition



$$M_{SAP} = K \sum_{j=0}^{N_{SAP}} (1 - AK)^j ,$$

$$K = \begin{pmatrix} B_{EE} & 0 \\ -B_{OO} A_{OE} B_{EE} & B_{OO} \end{pmatrix}$$

B_{EE} (B_{OO}) is an approximation for A_{EE}^{-1} (A_{OO}^{-1})

Single precision acceleration

"sloppy" precision can be used in right preconditioning

Suppose: calculation of $S = MT$ at line 4.6 in Algorithm is performed with "sloppy" precision in k-th iteration

$$
\begin{aligned}
S_k &\rightarrow S'_k = S_k + \delta S_k \\
Z_k &\rightarrow Z'_k = AS'_k \\
\zeta_k &\rightarrow \zeta'_k = \zeta_k + \delta\zeta_k \\
X_{k+1} &\rightarrow X'_{k+1} = X_k + U_k\alpha_k + \zeta'_k S'_k
\end{aligned}
$$

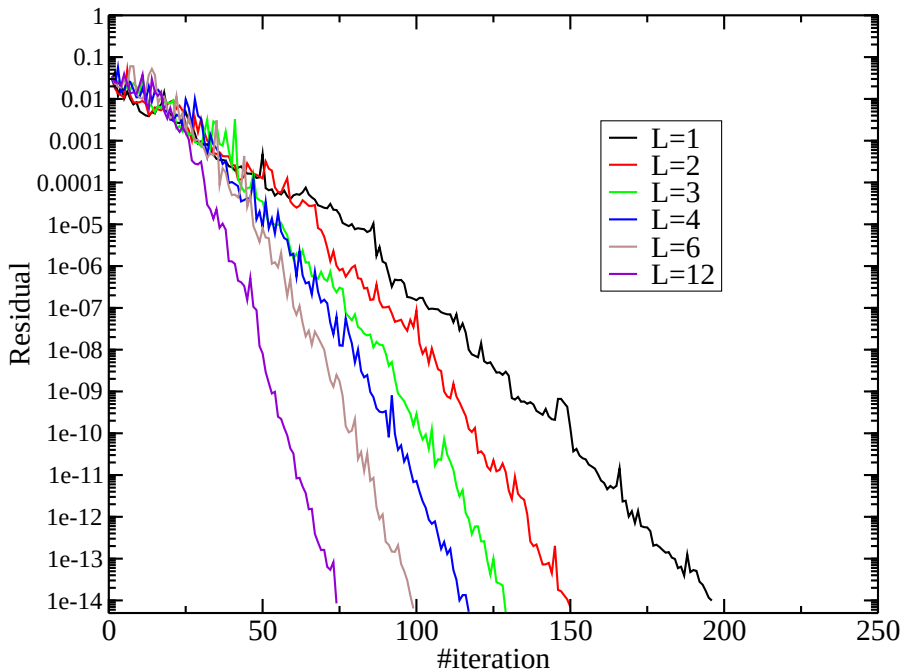
These yield

$$
\begin{aligned}
R'_{k+1} &= R_k - V_k\alpha_k - \zeta'_k Z'_k \\
&= R_k - AU_k\alpha_k - \zeta'_k AS'_k \\
&= B - AX_k - A(U_k\alpha_k + \zeta'_k S'_k) \\
&= B - AX'_{k+1}
\end{aligned}
$$

Numerical test

lattice size	: $32^3 \times 64$
quark masses	: almost physical
statistics	: 10 independent configurations
platform	: T2K-Tsukuba 16 nodes

T2K-Tsukuba	: quad-socket, 2.3GHz Quad-core AMD Opteron
	: 64KB/core L1$, 512KB/core L2$, 2MG/chip L3$
	: 8GB DDR2-667 /socket



計算科学研究機構 Advanced Institute for Computational Science

Results

$L \times 12/L$	time[s]	T(gain)	NMVM	NM(gain)
1×12	3827(755)	1	17146(3326)	1
2×6	2066(224)	1.9	12942(1379)	1.3
3×4	1619(129)	2.4	10652(832)	1.6
4×3	1145(99)	3.3	9343(835)	1.8
6×2	1040(87)	3.7	7888(663)	2.2
12×1	705(70)	5.4	6106(633)	2.8

- all tested case are converged
- better cache usage (gain ~ 2)
- less iteraion (gain ~ 3)

Summary

- introduced QCD, LQCD and Krylov subspace methods briefly
- Modified Block BiCGSTAB showed remarkable cost reduction

Summary

- introduced QCD, LQCD and Krylov subspace methods briefly
- Modified Block BiCGSTAB showed remarkable cost reduction
- and should accelerate LQCD simulations on

計算科学研究機構
Advanced Institute for Computational Science