Modified Block BiCGSTAB for Lattice QCD

Yoshifumi Nakamura

RIKEN AICS

10 Jun 2011
Motivation: Understanding particle physics
Motivation: Understanding particle physics

Non-perturbative approach of Quantum Chromo Dynamics (QCD)
Motivation: Understanding particle physics

Non-perturbative approach of Quantum Chromo Dynamics (QCD)

Lattice QCD (LQCD)
Motivation: Understanding particle physics

Non-perturbative approach of Quantum Chromo Dynamics (QCD)

Lattice QCD (LQCD)

But huge CPU resources require for LQCD simulations
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
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 $0.000\,000\,000\,000\,001\,\text{m} = 1\,\text{fm}$

hadron(color-neutral): meson($\pi, K...$), baryon($p, n...$)

quark: 3 (R, G, B)

 gluon: 8 (mass 0, charge 0, spin 1)
### quark: 6 flavours

<table>
<thead>
<tr>
<th></th>
<th>u(up)</th>
<th>c(charm)</th>
<th>t(top)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mass</td>
<td>1.7-3.3 MeV</td>
<td>1.27$^{+0.07}_{-0.09}$ GeV</td>
<td>172.0(22) GeV</td>
</tr>
<tr>
<td>charge</td>
<td>2/3</td>
<td>2/3</td>
<td>2/3</td>
</tr>
<tr>
<td>spin</td>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>d(down)</th>
<th>s(strange)</th>
<th>b(bottm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mass</td>
<td>4.1-5.8 MeV</td>
<td>101$^{+29}_{-21}$ MeV</td>
<td>4.19$^{+0.18}_{-0.06}$ GeV</td>
</tr>
<tr>
<td>charge</td>
<td>-1/3</td>
<td>-1/3</td>
<td>-1/3</td>
</tr>
<tr>
<td>spin</td>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
</tr>
</tbody>
</table>
QCD action

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

\[ F_{\mu\nu}^a : \quad \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + ig f^{abc} A^b_\mu A^c_\nu \]

\[ D(m_i) : \quad \gamma_\mu (i\partial_\mu + g A^a_\mu T^a) - m_i \]
Path integral

Expectation value of observable $O$:

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

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 $\times$ spinor / site $\rightarrow$ 12 complex numbers
  - gluon field: SU(3) matrix / link $\rightarrow$ 9 complex numbers
fermion field (Grassmann number)  
→ pseudo-fermion field (usual number)  

\[
\int d\bar{\psi}_id\psi_i \exp\left(-\sum_{i=1}^{2}\bar{\psi}_iD(m_i)\psi_i\right) = \det D(m_1) \det D(m_2)
\]

when \( m = m_1 = m_2 \)

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

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


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

\[ Ax = b \]
\[ x = A^{-1} b \]

by using matrix-vector multiplication

\[
\begin{bmatrix}
  a_{1,1} & \cdots & a_{1,n} \\
  \vdots & \ddots & \vdots \\
  a_{n,1} & \cdots & a_{n,n}
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
  \vdots \\
  x_n
\end{bmatrix}
= 
\begin{bmatrix}
  b_1 \\
  \vdots \\
  b_n
\end{bmatrix}
\]
Krylov subspace

\[ K_k \equiv \text{span}(v, Av, A^2v, \ldots, A^{k-1}v) \]

1. guess initial approx. solution vector \( x_0 \) for \( Ax = b \)

2. renew approx. solutions \( x_1, x_2, x_3, \ldots \) with keeping condition of \( x_k - x_0 \in K_k(A, r_0) = \text{span}(r_0, Ar_0, \ldots A^{k-1}r_0) \)

First residual vector:

\[ r_0 = b - Ax_0 \]

Approximate solution:

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

$$r_{k+1} = r_k - \alpha Ap_k$$
Properties of CG method

- conjugate property

\[(p_i, A p_j) = 0, \quad i \neq j\]

- orthogonality

\[(r_i, r_j) = 0, \quad i \neq j\]

System converges theoretically by ’n’ iteration at most
Block Krylov subspace method

\[ AX = B \]
\[ X = [x^{(1)}, x^{(2)}, \ldots, x^{(L)}] \]
\[ B = [b^{(1)}, b^{(2)}, \ldots, b^{(L)}] \]

\[ X_k - X_0 \in K_k(A, R_0) = \text{span}(R_0, AR_0, \ldots A^{k-1}R_0) \]

Approx. solutions \( X_k \):

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

\( \gamma_i \) is \( L \times L \) matrix

To solve \( x^{(i)} \), one can use information of \( K_k(A, r_0^{(j)}) \)
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

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

To compute \( \eta_x \) (hopping term multiplication)

- Flops: 1320
- Store: 12 complex numbers
- Load: 72 + 96 complex numbers, for \( U \) and \( \phi \)
Wilson-Dirac operator

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

To compute \( \eta_x \) (hopping term multiplication)

- Flops: 1320
- store: 12 complex numbers
- load: 72 + 96 complex numbers, for \( U \) and \( \phi \)

Hopping term mult. for multiple right hand sides

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

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 $P = Q\gamma$, $P \leq Q$
4.2 $U = MP$
4.3 $V = AU$
4.4 solve $(\tilde{R}^H V)\alpha = \tilde{R}^H R$ for $\alpha$
4.5 $T = R - V\alpha$
4.6 $S = MT$
4.7 $Z = AS$
4.8 $\zeta = \text{Tr}(Z^H T_k)/\text{Tr}(Z^H Z_k)$
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 $\beta$
4.12 $P = R + (P - \zeta V)\beta$

END

- by QR decomposition, numerical error $\downarrow$
  convergence $\uparrow$
- minimize comm. overhead by domain decomposition
preconditioning with single precision acceleration
Preconditioning

Original linear system:

\[ Ax = b \]

Preconditioned system:

\[ x = My \]
\[ AMy = b \]

Preconditioner

\[ M \approx A^{-1} \]

Condition number: \( AM < A \)
\[ M_{SAP} = K \sum_{j=0}^{N_{SAP}} (1 - AK)^j, \]
\[ K = \begin{pmatrix} B_{EE} & 0 \\ -B_{OO}A_{OE}B_{EE} & B_{OO} \end{pmatrix} \]

\( B_{EE} (B_{OO}) \) is an approximation for \( A^{-1}_{EE} (A^{-1}_{OO}) \)
“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

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

These yield

$$
R'_{k+1} = R_k - V_k \alpha_k - \zeta'_k Z'_k \\
= R_k - A U_k \alpha_k - \zeta'_k A S'_k \\
= B - A X_k - A(U_k \alpha_k + \zeta'_k S'_k) \\
= B - A X'_{k+1}
$$
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
## Results

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

- all tested case are converged
- better cache usage (gain $\sim 2$)
- less iteration (gain $\sim 3$)
introduced QCD, LQCD and Krylov subspace methods briefly

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

Modified Block BiCGSTAB showed remarkable cost reduction

and should accelerate LQCD simulations on