Bridge++のための

Multishift solver for shifted linear equations

格子 QCD コード Bridge++ 開発プロジェクト

松古栄夫 (KEK, hideo.matsufuru@kek.jp)

Version 1.1, 2014 年 12 月 17 日

Abstract

マルチシフトソルバーは、シフトされた線形方程式 $(A+\sigma_i)x_i = b$ を $(i = 0, 1, \ldots, k)$ に対し同時に解くためのアルゴリズムである。格子 QCD シミュレーションでは Rational HMC やオーバーラップ演算子において用いられ、計算コストを抑制するために重要な役割を果たす。このノートでは、マルチシフト CG 法を実装するにあたって必要となる表式を導出する。

Contents

1 Introduction

マルチシフトソルバーは、シフトされた線形方程式

$$(A+\sigma_i)x_i = b, \qquad i = 0, 1, \ldots, n \tag{1}$$

を同時に解くためのアルゴリズムである [1, 2, 3]。格子 QCD シミュレーションでは Rational HMC やオーバーラップ演算子の rational 近似において、計算コストを抑制するために重要な役割を果たす [4]。Eq. (1) を通常の反復法を用いて解くには各 i についてこれを適用する必要がある。マルチシフトソルバーでは一度の反復法の適用で、すべての i に対する解を同時に求めることができる。このため計算コストの大きい行列ベクトル積を、別々に解いた場合に比べて $O(1/n)$ の回数に抑えることができる。近似解と修正ベクトルを求めるためのベクトル演算は各 i に対しそれぞれ必要となるため、全体の計算コストは $1/n$ とはならないが、行列ベクトル積の計算コストが大きい場合には大きな効果がある。

今行列 A が正定値で、σ_i は小さいものから並んでいるものとする。このとき、$A+\sigma_i$ で最も条件数が大きいものは $A+\sigma_0$ となるため、この行列に対する線形方程式が最も大きい反復回数を必要とする。$i \geq 1$ に対しては、$i = 0$ に比べて収束が速いため、適当な収束条件で反復を切り上げることができる。

Krylov 部分空間法では、

$$\mathcal{K}_k = \mathrm{span}\{r_0, Ar_0, \ldots, A^{k-1}r_0\} \tag{2}$$

という部分空間内で近似解を構成する。シフト行列 $A+\sigma_i$ に対し、同一の初期残差ベクトル r_0 から Krylov 部分空間 $\mathcal{K}_k^{(i)}$ を構成したとすると、これらはすべて同一の部分空間を表していることがわかる。次のステップの残差ベクトルを、それまでの Krylov 部分空間に直交するように取れば[1]、各 σ_i に対する残差ベクトル $r_k^{(i)}$ は全て互いに並行となる。従って、σ_0 に対する $r_k^{(0)}$ と定数倍の違いしかない: $r_k^{(i)} = z_k^{(i)} r_k^{(0)}$. 一般に r_k を求めることができれば近似解ベクトル x_k も容易に求まるので、$z_k^{(i)}$ を求めるための表式が得られれば、シフトソルバーを構成できることになる。

このノートでは、マルチシフトソルバーのアルゴリズムを CG 法に対し実際に導出する。得られた表式は、Bridge++[5] の実装に用いられている。

2 Multishift CG algorithm

CG 法のアルゴリズムは、次のように表される[2]。線形方程式 $Ax = b$ に対し、

CG algorithm:

(i) initial step

$x_0 = 0$

[1] この残差ベクトルの選び方は Ritz-Galerkin アプローチと呼ばれる。

[2] 教科書でよく目にする表式では α と β の表記が逆で β の符号が逆の場合が多い。

$p_0 = r_0 = b$

(ii) repeat until convergence:

for $i = 0, 1, 2, \ldots$

$\beta_i = -(r_i, r_i)/(p_i, Ap_i)$

$x_{i+1} = x_i - \beta_i p_i$

$r_{i+1} = r_i + \beta_i A p_i$

$\alpha_i = (r_{i+1}, r_{i+1})/(r_i, r_i)$

$p_{i+1} = r_{i+1} + \alpha_i p_i$

ここで $x_0 = 0$ としたことにより、初期残差ベクトルは $r_0 = b$ となるため、後で全てのシフト値 σ に対して $r_0^\sigma = b$ を共通にでき、同一の Krylov 部分空間を取ることができる。

上記のようにシフトソルバーの構成には r_i の構成が本質的なので、r_i のみの漸化式の形に表すことを考える。以下のように、r_{i+1} は 3 項漸化式で表すことができる[3]。

$$\begin{aligned} p_i &= r_i + \alpha_{i-1} p_{i-1} \\ r_{i+1} &= r_i + \beta_i A p_i \end{aligned} \tag{3}$$

から p_i を消去すると、

$$\begin{aligned} r_{i+1} &= \left(1 + \frac{\beta_i \alpha_{i-1}}{\beta_{i-1}}\right) r_i + \beta_i A r_i - \frac{\beta_i \alpha_{i-1}}{\beta_{i-1}} r_{i-1} \\ &\equiv \hat{\alpha}_i r_i + \hat{\beta}_i A r_i + \hat{\gamma}_i r_{i-1}, \end{aligned} \tag{4}$$

と書ける。ここで

$$\hat{\alpha}_i = 1 + \frac{\beta_i \alpha_{i-1}}{\beta_{i-1}}, \qquad \hat{\beta}_i = \beta_i, \qquad \hat{\gamma}_i = -\frac{\beta_i \alpha_{i-1}}{\beta_{i-1}} = 1 - \hat{\alpha}_i. \tag{5}$$

Krylov 部分空間 $\mathcal{K}_k(A, r_0) = \mathrm{span}\{r_0, Ar_0, \ldots, A^{k-1} r_0\}$ を考える時、$A \to A + \sigma$ のシフトで、$\mathcal{K}_k$ は不変: $\mathcal{K}_k(A, r_0) = \mathcal{K}_k(A + \sigma, r_0)$ 。CG 法では、残差ベクトルに対し、

$$(r_i, r_j) = 0 \qquad \text{for } i \neq j \tag{6}$$

が成り立つ (Ritz-Galerkin アプローチに対応)。即ち、

$$r_k \perp \mathcal{K}_k(A, r_0) = \mathcal{K}_k(A + \sigma, r_0). \tag{7}$$

一方、r_k はある k-次多項式 $Z_k(x)$ によって、$r_k = Z_k(A) r_0$ と表される。従って、r_k は $A^k r_0$ の $\mathcal{K}_k(A, r_0)$ に垂直な成分に比例する。同じことは、$A \to A + \sigma$ とシフトした r_k^σ についても言える。即ち、$r_k^\sigma \propto r_k$ であるので、$Z_k(x)$ を k-次多項式、ただし $Z(0) = 1$ とすると、

$$\begin{aligned} r_k &= Z_k(A) r_0 \\ r_k^\sigma &= Z_k^\sigma(A + \sigma) r_0 = \zeta_k^\sigma Z_k(A) r_0, \end{aligned} \tag{8}$$

[3] CG 法の基底ベクトルは Lanczos 算法で構成されるため、基底ベクトルの漸化式は 3 項漸化式で表される。r_i は基底ベクトルに比例するため、r_i も 3 項漸化式で表される。

従って

$$Z_k^\sigma(A+\sigma) = \zeta_k^\sigma Z_k(A) \tag{9}$$

が成り立つ。3 項漸化式に対しては、

$$\begin{aligned} r_{i+1}^\sigma &= \zeta_{i+1}^\sigma r_{i+1} \\ &= \zeta_{i+1}^\sigma(\hat{\alpha}_i r_i + \hat{\beta}_i A r_i + \hat{\gamma}_i r_{i-1}) \end{aligned} \tag{10}$$

一方、

$$\begin{aligned} r_{i+1}^\sigma &= \hat{\alpha}_i^\sigma r_i^\sigma + \hat{\beta}_i^\sigma (A+\sigma) r_i^\sigma + \hat{\gamma}_i^\sigma r_{i-1}^\sigma \\ &= (\hat{\alpha}_i^\sigma + \hat{\beta}_i^\sigma \sigma)\zeta_i^\sigma r_i + \hat{\beta}_i^\sigma \zeta_i^\sigma A r_i + \hat{\gamma}_i^\sigma \zeta_{i-1}^\sigma r_{i-1} \end{aligned} \tag{11}$$

であるので、これらが等しいことから、

$$\begin{aligned} \zeta_{i+1}^\sigma \hat{\alpha}_i &= (\hat{\alpha}_i^\sigma + \hat{\beta}_i^\sigma \sigma)\zeta_i^\sigma, \\ \zeta_{i+1}^\sigma \hat{\beta}_i &= \hat{\beta}_i^\sigma \zeta_i^\sigma, \\ \zeta_{i+1}^\sigma \hat{\gamma}_i &= \hat{\gamma}_i^\sigma \zeta_{i-1}^\sigma, \end{aligned} \tag{12}$$

これを $\hat{\alpha}_i^\sigma$, $\hat{\beta}_i^\sigma$, $\hat{\gamma}_i^\sigma$ についての式に表せば、

$$\begin{aligned} \hat{\alpha}_i^\sigma &= (\hat{\alpha}_i - \sigma\hat{\beta}_i)\zeta_{i+1}^\sigma/\zeta_i^\sigma, \\ \hat{\beta}_i^\sigma &= \hat{\beta}_i \zeta_{i+1}^\sigma/\zeta_i^\sigma, \\ \hat{\gamma}_i^\sigma &= \hat{\gamma}_i \zeta_{i+1}^\sigma/\zeta_{i-1}^\sigma = (1-\hat{\alpha}_i)\zeta_{i+1}^\sigma/\zeta_{i-1}^\sigma. \end{aligned} \tag{13}$$

Eq. (5) と同様 $\hat{\alpha}_i^\sigma + \hat{\gamma}_i^\sigma = 1$ が成り立つので、この条件から ζ_{i+1}^σ に対する漸化式が次のように求まる。

$$\zeta_{i+1}^\sigma = \left[(\hat{\alpha}_i - \sigma\hat{\beta}_i)/\zeta_i^\sigma + (1-\hat{\alpha}_i)/\zeta_{i-1}^\sigma\right]^{-1} \tag{14}$$

初期条件については、$i=0$ に対しての 3 項漸化式を実際の $r_1 = b - Ax_1 = r_0 + \beta_0 A r_0$, $\beta_0 = -(b,b)/(b,Ab)$ と比較して、

$$\hat{\alpha}_0 = 1 + \frac{\beta_0 \alpha_{-1}}{\beta_{-1}} = 1, \qquad \hat{\beta}_0 = \beta_0, \qquad \hat{\gamma}_0 = 1 - \hat{\alpha}_0 = 0, \tag{15}$$

従って $\alpha_{-1} = 0$, $\beta_{-1} = 1$ と取っておけば矛盾しない。シフトされた線形方程式 $(A+\sigma)x^\sigma = b$ に CG 法を 1 ステップ適用すると

$$r_1^\sigma = (1-\sigma\beta_0)^{-1} r_1 \equiv= \zeta_1^\sigma r_1 \tag{16}$$

となることがわかる。これと ζ^σ に関する漸化式 (14) を比較すると、

$$\zeta_{-1}^\sigma = 1, \qquad \zeta_0^\sigma = 1 \tag{17}$$

と取ればよいことがわかる。

実際に反復法を数値的に解く際には x_i^σ と p_i^σ を更新してゆく。これには α_i^σ と β_i^σ が必要である。Eqs. (5), (13), 及び (14) を使うと、

$$\beta_i^\sigma = \beta_i \frac{\zeta_{i+1}^\sigma}{\zeta_i^\sigma}, \qquad \alpha_i^\sigma = \alpha_i \left(\frac{\zeta_{i+1}^\sigma}{\zeta_i^\sigma}\right)^2 \tag{18}$$

と書ける。これらを使って、上の CG 法アルゴリズムと同様、

$$\begin{aligned} x_{i+1}^\sigma &= x_i^\sigma - \beta_i^\sigma p_i^\sigma, \\ p_{i+1}^\sigma &= r_{i+1}^\sigma + \alpha_i^\sigma p_i^\sigma \ = \ \zeta_{i+1}^\sigma r_{i+1} + \alpha_i^\sigma p_i^\sigma \end{aligned} \tag{19}$$

のように更新してゆく。

まとめると、

Multi-shift CG algorithm:

(i) initial step

$$\begin{array}{ll} x_0 = 0 & x_0^\sigma = 0 \\ r_0 = p_0 = b & p_0^\sigma = b \\ & \zeta_0^\sigma = \zeta_{-1}^\sigma = 1 \end{array}$$

(ii) repeat until convergence

for $i = 0, 1, 2, \ldots$

$\beta_i = -(r_i, r_i)/(p_i, Ap_i)$

$x_{i+1} = x_i - \beta_i p_i$

$r_{i+1} = r_i + \beta_i A p_i$

$\alpha_i = (r_{i+1}, r_{i+1})/(r_i, r_i)$

$p_{i+1} = r_{i+1} + \alpha_i p_i$

$$\begin{aligned} &\hat{\alpha}_i = 1 + (\alpha_{i-1}\beta_i/\beta_{i-1}) \\ &\zeta_{i+1}^\sigma = \left[(\hat{\alpha}_i - \sigma\beta_i)/\zeta_i^\sigma \ + \ (1-\hat{\alpha}_i)/\zeta_{i-1}^\sigma\right]^{-1} \\ &\beta_i^\sigma = \beta_i\,(\zeta_{i+1}^\sigma/\zeta_i^\sigma) \\ &\alpha_i^\sigma = \alpha_i\,(\zeta_{i+1}^\sigma/\zeta_i^\sigma)^2 \\ &x_{i+1}^\sigma = x_i^\sigma - \beta_i^\sigma p_i^\sigma \\ &p_{i+1}^\sigma = \zeta_{i+1}^\sigma r_{i+1} + \alpha_i^\sigma p_i^\sigma \end{aligned}$$

左列が $\sigma = 0$ に対する演算、右列が σ でシフトされた線形方程式に対する演算である。収束条件としては、$|r_i^\sigma| = |\zeta_i^\sigma r_i| < \epsilon |b|$ とすればよい。Bridge++ では代わりに修正ベクトル $|p_i^\sigma|$ が十分小さくなるという条件で収束を判定している。

3 Miscellaneous

- CG 法以外にも、シフトソルバーが構成できるアルゴリズムは存在する。例えば、[3] とその参考文献を参照のこと。

- 反復法を 2 度適用することにより、ベクトルを保持するためのメモリ量を抑制する方法が提案されている [6]。アーキテクチャや問題サイズに依存して、キャッシュ効果などのため、元の方法より高速に実行できる可能性がある。

Acknowledgment

本文書を作成するにあたって、2005 年 2 月 23 日に KEK で行われた格子 QCD 勉強会での新谷栄悟氏のノートが参考になった。Bridge++プロジェクトは、HPCI 戦略プログラム分野 5「物質と宇宙の起源と構造」からサポートを受けている。

改訂履歴

- 2014.12.17 ver.1.1: Eq. (3) の typo 修正 (thanks to 上田悟氏)
- 2014.11.20 ver.1.0: 初版

References

[1] A. Frommer, B. Nockel, S. Gusken, T. Lippert and K. Schilling, "Many masses on one stroke: Economic computation of quark propagators," Int. J. Mod. Phys. C **6** (1995) 627 [hep-lat/9504020].

[2] B. Jegerlehner, "Krylov space solvers for shifted linear systems," hep-lat/9612014.

[3] 日本語のレビューとしては例えば、曽我部 知広、張 紹良「大規模シフト線形方程式の数値解法: クリロフ部分空間の性質に着目して」, 応用数理 19 (2009) 163.

[4] For a review, *e.g.*, A. D. Kennedy, "Algorithms for dynamical fermions," hep-lat/0607038.

[5] Bridge++ project, `http://bridge.kek.jp/Lattice-code/`

[6] H. Neuberger, "Minimizing storage in implementations of the overlap lattice Dirac operator," Int. J. Mod. Phys. C **10** (1999) 1051 [hep-lat/9811019].