双共役勾配安定化法,BiCGSTAB 法*

梅谷　信行

平成 22 年 11 月 13 日

目 次

1　概要

Bi-CGSTAB 法 [1] は H.A. van der Vorst によって提案された非対称行列用の反復法である。BiCG 法 (Biconjugate Gradient,BiCG) における j 回の反復における残差

*これは忘れっぽい著者が昔勉強したことを書き留めておく備忘録です．きっと間違えを多く含んでいます．すいません．ご指摘ご意見などありましたら教えてもらえると有り難いです．

$\mathbf{r}'_j$とすると、BiCGSTAB法とはこのBiCG法の残差$\mathbf{r}'_j$に以下のような多項式ψ_jを用いることで残差$\mathbf{r}_j$を減少させ、BiCG法を安定化した方法である。

$$\mathbf{r}_j = \chi_j(\mathbf{A})\mathbf{r}'_j \tag{1}$$

但し、χ_jは次のような漸化式で表される多項式である。

$$\chi_{j+1}(t) = (1 - \omega_j t)\chi_j(t) \tag{2}$$

つまり、j回目の反復では残差

$$\|\mathbf{r}_{j+1}\|_2 = \|(\mathbf{I} - \omega_j \mathbf{A})\chi_{j-1}(\mathbf{A})\mathbf{r}'_j\|_2 \tag{3}$$

を最小化するようにω_jを選ぶ。

2 基本的な手続き

BiCGSTAB法の手続きで難しいのは、いかにしてBiCG法(Biconjugate Gradient,BiCG)で用いる残差ベクトルや解ベクトルなどを明示的に持たずにBiCG法に残差の最小化の機能を追加するのかということである。クリロフ部分空間法であるBiCG法では残差などのベクトルは$\mathbf{A}$の多項式と初期残差を使って表現できる。これを利用してBiCGSTAB法では初期残差からBiCG法における係数を計算する。

2.1 BiCGSTAB法における残差ベクトルr、解の更新ベクトルp

BiCG法におけるj回反復における残差を$\mathbf{r}'_j$、解の更新ベクトルを$\mathbf{p}'_j$とする。次のように多項式ϕ_jとψ_jを用いて初期残差$\mathbf{r}_0$でこれらを表すとする。つまり、

$$\mathbf{r}'_j = \phi_j(\mathbf{A})\mathbf{r}_0 \tag{4}$$

$$\mathbf{p}'_j = \psi_j(\mathbf{A})\mathbf{r}_0 \tag{5}$$

この多項式ψとϕは次の漸化式を満たす。

$$\phi_{j+1}(t) = \phi_j(t) - \alpha_j t\psi_j(t) \tag{6}$$

$$\psi_{j+1}(t) = \phi_{j+1}(t) + \beta_j\psi_j(t) \tag{7}$$

多項式 ψ と χ を用いると残差は次のように書ける。

$$\mathbf{r}_j = \chi_j(\mathbf{A})\phi_j(\mathbf{A})\mathbf{r}_0 \tag{8}$$

ここで、$\chi_j\phi_j$ が出てきたので、これがどのように更新されるのか詳しく調べる。

$$\begin{aligned} \chi_{j+1}\phi_{j+1} &= (1-\omega_j t)\chi_j(t)\phi_{j+1} &(9)\\ &= (1-\omega_j t)(\chi_j\phi_j - \alpha_j t\chi_j\psi_j) &(10)\end{aligned}$$

$\chi_j\phi_j$ を更新するためには $\chi_j\psi_j$ が必要であることがわかる。そこで $\chi_j\psi_j$ についても、どのように更新されるのかを調べてみる。

$$\begin{aligned} \chi_{j+1}\psi_{j+1} &= \chi_{j+1}(\phi_{j+1} + \beta_j\psi) &(11)\\ &= \chi_{j+1}\phi_{j+1} + \beta_j(1-\omega_j t)\chi_j\psi_j &(12)\end{aligned}$$

このことから、$\chi\phi$ と $\chi\psi$ があれば多項式を更新できるということがわかる。そこで BiCGSTAB 法におけるベクトル $\mathbf{p}$ を次のように定義する。

$$\mathbf{p}_j = \chi_j(\mathbf{A})\psi_j(\mathbf{A})\mathbf{r}_0 \tag{13}$$

これらを用いて上の漸化式をまとめると、

$$\mathbf{r}_{j+1} = (\mathbf{I} - \omega_j\mathbf{A})(\mathbf{r}_j - \alpha_j\mathbf{A}\mathbf{p}_j) \tag{14}$$

$$\mathbf{p}_{j+1} = \mathbf{r}_{j+1} + \beta_j(\mathbf{I} - \omega_j\mathbf{A})\mathbf{p}_j \tag{15}$$

となる。次はこのベクトル $\mathbf{r}$、$\mathbf{p}$ から、BiCG 法 (Biconjugate Gradient,BiCG) の係数 α、β を作る方法を考える。

2.2 β の導出

まず始めに β を導入する。ρ_j を次のように定めると、

$$\rho_j = (\mathbf{r}'_j, \mathbf{r}^{*'}_j) = \left(\phi_j(\mathbf{A})\mathbf{r}_0, \phi_j(\mathbf{A}^T)\mathbf{r}_0\right) \tag{16}$$

BiCG 法より、β は ρ を使って $\beta = \rho_{j+1}/\rho_j$ のように定義された。

しかし、ρ_j を作る時に必要な BiCG 法 (Biconjugate Gradient,BiCG) のベクトル $\mathbf{r}'_j$、$\mathbf{r}^{*\prime}_j$ は用いることができない。そこで、次のような値 $\tilde{\rho}_j$ を考える。

$$\tilde{\rho}_j = (\mathbf{r}_j, \mathbf{r}^*_0) \tag{17}$$

この $\tilde{\rho}_j$ は次のように変形される。

$$\begin{aligned}
\tilde{\rho}_j &= \left(\chi_j(\mathbf{A})\phi_j(\mathbf{A})\mathbf{r}_0, \mathbf{r}^*_0\right) && (18)\\
&= \left(\phi_j(\mathbf{A})\mathbf{r}_0, \chi_j(\mathbf{A}^T)\mathbf{r}^*_0\right) && (19)\\
&= \left(\mathbf{r}'_j, \chi_j(\mathbf{A}^T)\mathbf{r}^*_0\right) && (20)
\end{aligned}$$

ここで、BiCG 法は残差 $\mathbf{r}'_j$ が $\mathbf{r}^*_0$ と $\mathbf{A}^T$ で作られるクリロフ部分空間 $\mathcal{K}_j(\mathbf{A}^T, \mathbf{r}^*_0) = span\{\mathbf{r}^*_0, \mathbf{A}^T\mathbf{r}^*_0, \ldots, (\mathbf{A}^T)^{j-1}\mathbf{r}^*_0\}$ が直交するという条件の中で解を探すものであった。$\mathbf{a}$ についての多項式 χ_j の中で次数がもっとも高い j 次の項以外の項は $\mathbf{r}'_j$ と内積を計算すると 0 になることがわかる。よってこの性質を利用すると $\tilde{\rho}$ は χ_j の最高次数の係数を η_j とすると次のように書くことができる。

$$\tilde{\rho}_j = \left(\mathbf{r}'_j, (\mathbf{A}^T)^j\mathbf{r}^*_0\right)\eta_j \tag{21}$$

同じように、ρ_j についてもこの性質を利用して表示してみる。

$$\rho_j = (\mathbf{r}'_j, \phi_j(\mathbf{A}^T)\mathbf{r}_0) = \left(\mathbf{r}'_j, (\mathbf{A}^T)^j\mathbf{r}^*_0\right)\gamma_j \tag{22}$$

但し、γ_j は多項式 ϕ_j の最高次数の係数である。上の２式を見比べると ρ_j は次のように作ることができることがわかる。

$$\rho_j = \frac{\gamma_j}{\eta_j}\tilde{\rho}_j \tag{23}$$

η_j と γ_j はそれぞれ χ_j と ϕ_j の最高次数の係数であったことを考えると、η_j と γ_j は明らかに次の漸化式を満たす

$$\eta_{j+1} = -\omega_j\eta_j \qquad \gamma_{j+1} = -\alpha_j\gamma_j \tag{24}$$

よって、β_j は次のようになる。

$$\beta_j = \frac{\rho_{j+1}}{\rho_j} \tag{25}$$

$$= \frac{\eta_j}{\eta_{j+1}} \times \frac{\gamma_{j+1}}{\gamma_j} \times \frac{\tilde{\rho}_{j+1}}{\tilde{\rho}_j} \tag{26}$$

$$= \frac{\alpha_j}{\omega_j} \frac{\tilde{\rho}_{j+1}}{\tilde{\rho}_j} \tag{27}$$

2.3 α_j の導出

つぎに BiCG 法 (Biconjugate Gradient,BiCG) の係数 α_j について調べる。BiCG 法によると α_j は次のように定義された。

$$\alpha_j = \frac{(\mathbf{r}'_j, \mathbf{r}^{*'}_j)}{(\mathbf{A}\mathbf{p}'_j, \mathbf{p}^{*'}_j)} \tag{28}$$

分子は前の議論によると、$(\mathbf{r}'_j, \mathbf{r}^{*'}_j) = \rho_j = \frac{\gamma_j}{\eta_j}\tilde{\rho}$ であり計算できる値である。そこで分母の $(\mathbf{A}\mathbf{p}'_j, \mathbf{p}^{*'}_j)$ を取り出して詳しく調べる。

$$\tag{29}$$

また、BiCG 法のアルゴリズムから

$$\mathbf{A}\mathbf{p}'_j = -\frac{1}{\alpha_j}(\mathbf{r}'_{j+1} - \mathbf{r}'_j) \tag{30}$$

が成り立つ。

ここでも β の導出の際と同様に、BiCG 法は残差 $\mathbf{r}'_j$ とクリロフ部分空間 $\mathcal{K}_j(\mathbf{A}^T, \mathbf{r}^*_0) = span\{\mathbf{r}^*_0, \mathbf{A}^T\mathbf{r}^*_0, \ldots, (\mathbf{A}^T)^{j-1}\mathbf{r}^*_0\}$ を直交させるという条件を用いる。

$$\mathbf{r}'_j \perp \mathcal{K}_j(\mathbf{A}^T, \mathbf{r}^*_0) \tag{31}$$

$$\mathbf{r}'_{j+1} \perp \mathcal{K}_{j+1}(\mathbf{A}^T, \mathbf{r}^*_0) \ni \mathcal{K}_j(\mathbf{A}^T, \mathbf{r}^*_0) \tag{32}$$

よって次が成り立つ。

$$\mathbf{Ap}_j^* = -\frac{1}{\alpha_j}(\mathbf{r}'_{j+1} - \mathbf{r}'_j) \perp \mathcal{K}_j(\mathbf{A}^T, \mathbf{r}_0^*) \tag{33}$$

以上から $\phi_j(\mathbf{A}^T)\mathbf{r}_0^*$ は次数が j であるため、最高次の項以外は $\mathbf{Ap}_j^*$ との内積を計算すると 0 になる。$\phi_j(\mathbf{A}^T)\mathbf{r}_0^*$ の最高次の係数を γ_j とおいていたので、これを用いると

$$(\mathbf{Ap}'_j, \mathbf{p}_j^{*'}) = \left(\mathbf{Ap}'_j, (\mathbf{A}^T)^j \mathbf{r}_0^*\right)\gamma_j \tag{34}$$

同様の議論から次がいえる。

$$\begin{aligned}(\mathbf{Ap}_j, \mathbf{r}_0^*) &= \left(\mathbf{A}\chi_j(\mathbf{A})\psi_j(\mathbf{A})\mathbf{r}_0, \mathbf{r}_0^*\right) && (35)\\ &= \left(\mathbf{A}\psi_j(\mathbf{A})\mathbf{r}_0, \chi_j(\mathbf{A}^T)\mathbf{r}_0^*\right) && (36)\\ &= \left(\mathbf{Ap}'_j, (\mathbf{A}^T)^j \mathbf{r}_0^*\right)\eta_j && (37)\end{aligned}$$

上の２式から次のようになる。

$$(\mathbf{Ap}'_j, \mathbf{p}_j^{*'}) = \frac{\gamma_j}{\eta_j}(\mathbf{Ap}_j, \mathbf{r}_0^*) \tag{38}$$

よって α_j を求めることができる。

$$\begin{aligned}\alpha_j &= \frac{\frac{\gamma_j}{\eta_j}\tilde{\rho}}{\frac{\gamma_j}{\eta_j}(\mathbf{Ap}_j, \mathbf{r}_0^*)} && (39)\\ &= \frac{\tilde{\rho}}{(\mathbf{Ap}_j, \mathbf{r}_0^*)} && (40)\end{aligned}$$

2.4　ω の導出

続いて ω の導出する。これは、残差の２乗ノルムを最小化するように選ばれる。j 回目の反復における残差の２乗ノルムは以下のように表された。

$$\|\mathbf{r}_{j+1}\|_2 = \|(\mathbf{I} - \omega_j\mathbf{A})\chi_{j-1}(\mathbf{A})\mathbf{r}'_j\|_2 \tag{41}$$

簡単のために以下のようにベクトル $\mathbf{s}_j$ を定義する。

$$\mathbf{s}_j = \chi_{j-1}(\mathbf{A})\mathbf{r}'_j = \mathbf{r}_j - \alpha_j \mathbf{A}\mathbf{p}_j \tag{42}$$

これを用いると、残差の２乗ノルムの２乗は次のようにかける

$$\begin{aligned}
\|\mathbf{r}_{j+1}\|_2^2 &= \|(\mathbf{I} - \omega_j \mathbf{A})\mathbf{s}_j\|_2^2 & (43)\\
&= \Big((\mathbf{I} - \omega_j \mathbf{A})\mathbf{s}_j, (\mathbf{I} - \omega_j \mathbf{A})\mathbf{s}_j\Big) & (44)\\
&= (\mathbf{A}\mathbf{s}_j, \mathbf{A}\mathbf{s}_j) & (45)
\end{aligned}$$

$$\omega_j^2 - 2(\mathbf{s}_j, \mathbf{A}\mathbf{s}_j)\omega_j + (\mathbf{s}_j, \mathbf{s}_j) \tag{46}$$

残差の２乗ノルムが最小であるという条件から

$$\frac{\partial}{\partial \omega_j}\|\mathbf{r}_{j+1}\|_2^2 = 2(\mathbf{A}\mathbf{s}_j, \mathbf{A}\mathbf{s}_j)\omega_j - 2(\mathbf{s}_j, \mathbf{A}\mathbf{s}_j) = 0 \tag{47}$$

となる。よって ω_j は次のようになる。

$$\omega_j = \frac{(\mathbf{s}_j, \mathbf{A}\mathbf{s}_j)}{(\mathbf{A}\mathbf{s}_j, \mathbf{A}\mathbf{s}_j)} \tag{48}$$

2.5 解の更新

残差 $\mathbf{r}$ は次のように更新された

$$\begin{aligned}
\mathbf{r}_{j+1} &= (\mathbf{I} - \omega_j \mathbf{A})(\mathbf{r}_j - \alpha_j \mathbf{A}\mathbf{p}_j) & (49)\\
&= \mathbf{s}_j - \omega_j \mathbf{A}\mathbf{s}_j & (50)\\
&= \mathbf{r}_j - \alpha_j \mathbf{A}\mathbf{p}_j - \omega_j \mathbf{A}\mathbf{s}_j & (51)
\end{aligned}$$

つまり、残差の増分は以下のように表される。

$$\mathbf{r}_{j+1} - \mathbf{r}_j = -\mathbf{A}(\alpha \mathbf{p}_j + \omega_j \mathbf{s}_j) \tag{52}$$

また解の増分と残差の増分の関係から以下がいえる。

$$\mathbf{r}_{j+1} - \mathbf{r}_j = -\mathbf{A}(\mathbf{x}_{j+1} - \mathbf{x}_j) \tag{53}$$

上の２式を比較すると解の更新の関係が得られる。

$$\mathbf{x}_{j+1} = \mathbf{x}_j + \alpha \mathbf{p}_j + \omega_j \mathbf{s}_j \tag{54}$$

2.6　アルゴリズム（BiCGSTAB 法)

1. Compute $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$,
2. Choose $\mathbf{r}_0^*$ arbitrary
3. $\mathbf{p}_0 = \mathbf{r}_0$
4. For $j = 1, 2, \ldots, m$ Do:
   (a) $\alpha_i = \frac{(\mathbf{r}_j, \mathbf{r}_0^*)}{(\mathbf{A}\mathbf{p}_j, \mathbf{r}_0^*)}$
   (b) $\mathbf{s}_j = \mathbf{r}_j - \alpha_j \mathbf{A}\mathbf{p}_j$
   (c) $\omega_j = \frac{(\mathbf{s}_j, \mathbf{A}\mathbf{s}_j)}{(\mathbf{A}\mathbf{s}_j, \mathbf{A}\mathbf{s}_j)}$
   (d) $\mathbf{x}_{j+1} = \mathbf{x}_j + \alpha_j \mathbf{p}_j + \omega_j \mathbf{s}_j$
   (e) $\mathbf{r}_{j+1} = \mathbf{s}_j - \omega_j \mathbf{A}\mathbf{s}_j$
   (f) $\beta_j = \frac{(\mathbf{r}_{j+1}, \mathbf{r}_0^*)}{(\mathbf{r}_j, \mathbf{r}_0^*)} \times \frac{\alpha_j}{\omega_j}$
   (g) $\mathbf{p}_{j+1} = \mathbf{r}_{j+1} + \beta_j(\mathbf{p}_j - \omega_j \mathbf{A}\mathbf{p}_j)$
5. End Do

以上が BICGSTAB 法の基本的な手続きである。(c),(d),(e) 行目が残差最小化のためのステップであり、(a),(b),(f),(g) 行目が BiCG 法に相当するステップであると解釈できる。

3　前処理つき BiCGSTAB 法

解くべき方程式 $\mathbf{A}\mathbf{x} = \mathbf{b}$ を前処理行列 M を使って次のように変形する。

$$\mathbf{A}\mathbf{M}^{-1}\mathbf{u} = \mathbf{b}, \qquad \mathbf{x} = \mathbf{M}^{-1}\mathbf{u} \tag{55}$$

これは $\mathbf{M}^{-1}$ が $\mathbf{A}$ の右から掛けられているので右前処理と呼ばれる。係数行列が $\mathbf{A}$ から $\mathbf{A}\mathbf{M}^{-1}$ に変わっている。この $\mathbf{A}\mathbf{M}^{-1}$ の性質が単位行列に近ければ元の方程式よりも早い収束が期待される。この前処理を使って BiCGSTAB 法で解を求める方法は基本的に BiCGSTAB 法で $\mathbf{A}$ の部分を $\mathbf{A}\mathbf{M}^{-1}$ に変えて $\mathbf{u}$ を求めた後に最後に 1 回解く事で $\mathbf{u}$ を求めるということだが、少し変形すると $\mathbf{u}$ を明示的に持たずに直接 $\mathbf{x}$ の増分を求めることが可能である。$\hat{\mathbf{p}} = \mathbf{M}^{-1}\mathbf{p}$、

$$\hat{\mathbf{s}} = \mathbf{M}^{-1}\mathbf{s} \tag{56}$$

とおくと、$\Delta\mathbf{x} = \mathbf{M}^{-1}\Delta\mathbf{u} = \mathbf{M}^{-1}(\alpha\mathbf{p} + \omega\mathbf{s}) = \alpha\hat{\mathbf{p}} + \omega\hat{\mathbf{s}}$ というように解の増分を求めることができる。$\hat{\mathbf{p}}$、$\hat{\mathbf{s}}$ を用いることで $\mathbf{u}$ を明示的に持たなくてもよいことがわかる。また明らかに $\mathbf{AM}^{-1}\mathbf{p} = \mathbf{A}\hat{\mathbf{p}}$、

$$\mathbf{AM}^{-1}\mathbf{s} = \mathbf{A}\hat{\mathbf{s}} \tag{57}$$

である。残差については

$$\mathbf{r} = \mathbf{b} - \mathbf{Ax} = \mathbf{b} - \mathbf{AM}^{-1}\mathbf{u} \tag{58}$$

であるから $\mathbf{u}$ に対する残差はそのまま $\mathbf{x}$ に対する残差でもある。よって反復計算しながら $\mathbf{u}$ を求める際に得られる残差の大きさは $\mathbf{x}$ に対する残差の大きさでもある。右前処理は反復しながら解の残差の大きさがわかるので残差の大きさを判断して計算を打ち切るかどうか決める際に便利である。

3.1 アルゴリズム（右前処理つき BiCGSTAB 法）

1. Compute $\mathbf{r}_0 = \mathbf{b} - \mathbf{Ax}_0$
2. Choose $\mathbf{r}_0^*$ arbitrary
3. $\mathbf{p}_0 = \mathbf{r}_0$
4. For $j = 1, 2, \ldots, m$ Do:
   (a) $\hat{\mathbf{p}}_i = \mathbf{M}^{-1}\mathbf{p}$
   (b) $\alpha_i = \frac{(\mathbf{r}_j, \mathbf{r}_0^*)}{(\mathbf{A}\hat{\mathbf{p}}_j, \mathbf{r}_0^*)}$
   (c) $\mathbf{s}_j = \mathbf{r}_j - \alpha_j \mathbf{A}\hat{\mathbf{p}}_j$
   (d) $\hat{\mathbf{s}}_j = \mathbf{M}^{-1}\mathbf{s}_j$
   (e) $\omega_j = \frac{(\mathbf{s}_j, \mathbf{A}\hat{\mathbf{s}}_j)}{(\mathbf{A}\hat{\mathbf{s}}_j, \mathbf{A}\hat{\mathbf{s}}_j)}$
   (f) $\mathbf{x}_{j+1} = \mathbf{x}_j + \alpha_j\hat{\mathbf{p}}_j + \omega_j\hat{\mathbf{s}}_j$
   (g) $\mathbf{r}_{j+1} = \mathbf{s}_j - \omega_j\mathbf{A}\hat{\mathbf{s}}_j$
   (h) $\beta_j = \frac{(\mathbf{r}_{j+1}, \mathbf{r}_0^*)}{(\mathbf{r}_j, \mathbf{r}_0^*)} \times \frac{\alpha_j}{\omega_j}$
   (i) $\mathbf{p}_{j+1} = \mathbf{r}_{j+1} + \beta_j(\mathbf{p}_j - \omega_j\mathbf{A}\hat{\mathbf{p}}_j)$

5. End Do

以上が右前処理つき BICGSTAB 法の基本的な手続きである。

4　参考にしたもの

[1],[2],[3]

参考文献

[1] Vorst, Van der H.: BiCGStab, A fast and smoothly conueqing variant of BiCG for the solution of nonsymmetn'c linear systems SIAM J, *Sci. Stat. Comput*, Vol. 13, pp. 631–644 (1992).

[2] Barrett, R., Berry, M., Chan, T. F., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C. and Vorst, van der H.: *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods (Miscellaneous Titles in Applied Mathematics Series No 43)*, Society for Industrial Mathematics (1987).

[3] Saad, Y.: *Iterative Methods for Sparse Linear Systems, Second Edition*, Society for Industrial and Applied Mathematics, 2 edition (2003).