前処理付き CG 法系解法における行列の条件数と収束判定

石黒 美佐子* 陽遊美由紀** 平津 忍*
* 茨城大学工学部 ** ソフトバンク

線形方程式の反復解法の収束状況は，係数行列の性質，特に条件数に大いに影響される．ここでは，並列，ベクトル処理で有効とされる前処理付き CG 法系解法の条件数，収束判定，数値誤差について論じる．差分法による離散化により生じる行列の条件数を近似計算したときの有効性，前処理後の行列の最大・最小固有値と条件数との関係，正規化や前処理と収束判定や解の精度との関係，などについて数値実験結果を示す．

Condition Number and Convergence Criterion in the Preconditioned CG and Its Variant Algorithms

Misako Ishiguro*, Miyuki Yohyu**, Shinobu Hiratsu*
* Faculty of Engineering, Ibaraki University ** Soft Bank

Convergence behavior of iterative methods of a linear system is affected by the property of the coefficient matrix, especially, by the "condition number" of the matrix. In this report, we deal with preconditioned CG and its variants which are useful for vector and parallel processors and discuss on the condition number, convergence criterion, and resulting numerical error. The effectiveness of the approximate calculation of the condition number is first validated. Numerical experiments are made to investigate the relationship among the maximum and minimum eigenvalues, condition number, convergence criterion, and numerical error, for normalized and/or preconditioned matrices arisen from finite difference discretization.

1. はじめに

気象予報，航空，宇宙工学，半導体設計，核融合，原子力などの工学分野で起こる非線型で時間依存の大規模数値計算は，通常，連続体モデルに基づいて計算されている．これらの課題は，多くの場合において大次元連立方程式解法を繰り返し解くことに帰着する．連立方程式を高速に精度よく解くことは重要な課題で在り続けると言える．ここでは，ベクトル処理や並列処理に有効とされる対称行列のためのＣＧ法，非対称行列のための Bi-CGSTAB 法[7]で連立方程式 $Ax=b$ を解くことを取上げる．行列の条件数が小さいほど反復回数が少ないことが知られている．大規模数値計算の現場でしばしば遭遇する正規化や前処理による反復回数の減少を行列の条件数を介して考察し，数値実験で検証する．

真の解 x，近似解 $\hat{x}$ としたとき，誤差 $e=x-\hat{x}$ の誤差のノルムは次式のように評価される[2].

(1.1) $\|e\|/\|\hat{x}\| \leq \varepsilon_M \|A\|\cdot\|A^{-1}\|$.

ここで，ε_M は使用する計算機の丸め誤差で，$\|A\|\cdot\|A^{-1}\|$ は行列 A の条件数（$=cond(A)$ と記す）と呼ばれるもので数値誤差に対する性質の悪さを示す尺度となる．$\|A\|$ は A のノルムを表し，$\|A\| = \max_{x\neq 0}(\|Ax\|/\|x\|)$ の定義から A

の絶対値最大固有値 λ_{max} と絶対値最小固有値 λ_{min} に対し，次式が成り立つ．

(1.2) $\quad \| A \| \geq |\lambda_{max}|, \quad \| A^{-1} \| \geq 1 / |\lambda_{min}|.$

対称行列でユークリッドノルム $\| \ \|_2$ を用いたときには，式(1.2) は等号が成立する．故に，対称行列の条件数は最大固有値と最小固有値の絶対値の比 $|\lambda_{max} / \lambda_{min}|$ で表せる．

非対称行列では条件数を固有値から求めることはできない．逆行列のノルム $\| A^{-1} \|_1$ を近似的に求める方法がいくつか提案され，評価されている[3]．これらによって非対称行列の条件数の近似値を求めることができる．

行列に前処理を施すことによって条件数を下げられれば反復回数を減らすことができる．前処理付き CG 法系の解法では，例えば，A が非対称行列の場合は，$M^{-1}A \fallingdotseq I$ となる前処理行列 M を選び，固有値を 1 に集めようとするものである．$Ax = b$ を解く代わりに前処理後の行列 $\tilde{A} = M^{-1}A$ に対して連立方程式を解くことになる．

ε_M が 10^{-m} 程度の計算機で条件数が 10^c 程度の大きさを持つ連立方程式を解くと，得られる解の相対誤差は 式(1.1)からおよそ 10^{c-m} 程度になる[2]．解の有効桁数が条件数と計算機の丸め誤差によって決まるということは，前処理によって条件数を小さくすれば，有効桁数を増やすことができると考えられる．前処理後の条件数を $10^{c'}$ 程度であるとする．計算機のまるめ誤差を前処理前後で一定の 10^{-m} 程度とすれば，有効桁は (c-c') 桁分増加しているといえる．反復回数を論じるには，このような有効桁数の違いに注意して収束判定を行う必要がある．本論文では，数値実験によってこのことを検証し，結果の考察を行う．

CG 法系解法の丸め誤差については，解法本来の反復過程からくる残差から誤差を分離することによって解の収束に与える影響という視点で考察され，非対称行列用の CGS 法[5]の丸め誤差が累積する傾向が指摘されている[6]．本報告では，丸め誤差の振る舞いよりはむしろ，収束条件，前処理との関係から解の数値誤差について論じる．

2. 前処理付きＣＧ法系解法の誤差と残差

ＣＧ法系解法で得られた近似解を $\hat{x}$ とするとき，$r = b - A\hat{x}$ を残差という．CGS 法では，アルゴリズムの中で更新される残差ノルムが小さくても，真の残差ノルム $\| b - A\hat{x} \|$ は大きな値となる場合があることが指摘されているが，CG 法や Bi-CGSTAB 法では両者はほぼ一致する[7]．

誤差 e と残差との間に次の関係がある．

(2.1) $\quad e = A^{-1} r.$

相対誤差ノルムは，式(1.1)で示したように計算機のまるめ誤差と行列 A の条件数によって評価される．以下の手順に従って，CG 法系解法の収束判定基準と計算精度との関係を論じよう．

2.1 CG 法等解法の収束判定

CG 法，Bi-CGSTAB 法の収束判定は，通常，次式に従う．

(2.2) $\quad \| r \| \leq \varepsilon \| b \|.$

ここで ε はユーザが指定する計算精度である．

前処理つき CG 法系解法の場合には，不完全 LU 分解などによって $K = K_1 K_2$ と分解したものを前処理行列とし，

(2.3) $\quad \tilde{A} = K_1^{-1} A K_2^{-1}, \quad \tilde{b} = K_1^{-1} b, \quad x = K_2^{-1} \tilde{x}$

とおいて $\tilde{A}\tilde{x} = \tilde{b}$ を解くことになる．

$K_1 = M$，$K_2 = I$ とおいて前処理行列を M とし，

(2.4) $\quad \tilde{A} = M^{-1} A, \quad \tilde{b} = M^{-1} b$

で前処理する方法もよく用いらる．いずれも $\tilde{x}$ を近似解とおけば，$\tilde{r} = \tilde{b} - \tilde{A}\tilde{x}$，収束判定は次式による．

(2.5) $\quad \| \tilde{r} \| \leq \varepsilon \| \tilde{b} \|$

$\tilde{r}$ は，前処理した方が早く小さくなるが反復回数は右辺ベクトルのノルムに関係する．

2.2 収束判定条件と解の有効桁数

$\| r \| \leq \varepsilon \| b \|$ で収束判定するとすれば，

$\| r \| \leq \varepsilon \| b \| \leq \varepsilon \| A \| \| x \|,$

式(2.1)より誤差ノルムは，

$$\| e \| \leq \| A^{-1} \| \| r \| \\ \leq \varepsilon \| A^{-1} \| \| A \| \| x \|$$

となる．解の相対誤差ノルムは

(2.6) $\| e \| / \| x \| = \varepsilon\, cond(A)$.

式(1.1)と類似の式が得られた．CG 法系解法での収束判定で使われる ε を計算機の丸め誤差 ε_M と同様に捕らえることができる．ただし，$\varepsilon_M \leqq \varepsilon$ である．$\varepsilon = 10^{-m}$，$cond(A)$ の桁数を c 程度とすると解の有効桁数は m-c 以上と期待できる。

2.3 前処理後の解の有効桁数

前処理後の連立方程式に対して式(2.4)に示す表記法を用いると

(2.7) $\tilde{A}^{-1}\tilde{r} = \tilde{A}^{-1}\tilde{b} - \bar{x} = x - \bar{x} = \tilde{e}$.

前処理後の誤差 $\tilde{e}$ は，前処理なしのときの e と同様の定義となる．行列 $\tilde{A}$ の条件数を c' とおくと，条件数は小さくなっているはずだから，c'≦c.

$\| \tilde{r} \| \leqq \varepsilon \| \tilde{b} \|$ で収束判定した場合の解の有効桁数は，

(2.8) $\| \tilde{e} \| / \| x \| = \varepsilon\, cond(\tilde{A})$.

により m-c' 以上となる．有効桁数は前処理したものの方が大きいと期待される．

2.4 前処理後の収束判定の変更

前処理なしで $\| r \| \leqq \varepsilon \| b \|$ で収束判定したときの反復回数を n，前処理後に $\| \tilde{r} \| \leqq \varepsilon \| \tilde{b} \|$ で収束判定したときの反復回数を n' とおく．解の有効桁数はそれぞれ m-c と m-c' で m-c ≦m-c'，かつ n'≦n．前処理なしの場合と同じ精度を保証するためにはもっと大きな ε で良いはずである，e.g.，$\varepsilon' = 10^{(c\text{-}c')\text{-}m}$．$\| \tilde{r} \| \leqq \varepsilon' \| \tilde{b} \|$ で収束判定したとして，その反復回数を n" とすると，収束条件を緩めたことになるので n"≦n' となるはずである．結局，n"≦n'≦n が期待できる．

3．正規化と前処理

3.1 正規化と条件数

前処理行列 M の代わりに元の行列 A の対角要素 $d_1, d_2, \ldots, d_n$ からなる行列 D でスケーリングする．非対称行列 A の正規化後の行列を $\bar{A}$ とおく．

(3.1) $\bar{A} = D^{-1}A$

A の任意の固有値を λ，その固有ベクトルを u とすれば，次の式が成り立つ．

(3.2) $D\bar{A}u = \lambda u$,　$D^{-1}\bar{A}^{-1} v = (1/\lambda) v$

ここで $v = D^{-1}u$ とおく．両辺ノルムをとってそれぞれ $\| u \|$，$\| v \|$ で割れば，

(3.3) $\| \bar{A} u \| / \| u \| \geqq |\lambda| \, \| D \|^{-1}$,

$\| \bar{A}^{-1} v \| / \| v \| \geqq |1/\lambda| \, \| D^{-1} \|^{-1}$.

行列のノルムの定義から、

(3.4) $\| \bar{A} \| \geqq |\lambda_{max}| \, \| D \|^{-1}$,

$\| \bar{A}^{-1} \| \geqq 1 / |\lambda_{min}| \, \| D^{-1} \|^{-1}$.

故に

(3.5) $cond(\bar{A}) \geqq |\lambda_{max} / \lambda_{min}| \, |d_{min} / d_{max}|$.

ここで，λ_{max} と λ_{min} は元の行列の絶対値最大と最小の固有値，d_{max} と d_{min} は元の行列の絶対値最大と最小の対角要素を表す．

一般行列では，条件数は対角要素の値のばらつきが大きければ正規化によって $|d_{min} / d_{max}|$ の比率で減少することを示唆している．

3.2 不完全 LU 分解による前処理

正規化後の行列を，対角部分 I，下三角行列部分 L，上三角部分 R に 3 区分し，$\bar{A} = I + L + R$ で表し，次の不完全 LU 分解後の行列を前処理行列として用いる．

(3.6) $K = K_1 K_2$,　$K_1 = (I + L)$,
$K_2 = (I + R)$.

3.3 行列多項式前処理

$\bar{A}^{-1}$ を固有値を 1 に集めるために展開した行列多項式[1]を 2 次で打切ったものを前処理行列 M の逆数として使う．

(3.7) $\bar{A}^{-1} = (I + L + R)^{-1}$
$\fallingdotseq 35/32I - 50/32(L + R) + 35/32(L + R)^2$

4．数値実験

4.1 テスト行列

内部発熱がある２次元熱伝導問題の定常解を$x=\pm 1$, $y=\pm 1$ の形状で求める[4]. $y=1$ で $\partial u/\partial y=-u$ という境界条件を仮定し，差分法で離散化すると非対称のブロック行列が生じる．これをテスト行列として用いる．対称行列のテストには，この行列要素を一部置換え対称化する．

4.2 条件数の近似計算の有効性

まず正規化後の行列に対して塚本らの近似的方法[3]を用いて条件数を計算し，$|\lambda_{max}/\lambda_{min}|$ の値と比較する（Fig.1）．対称行列では，下方の２つのラインに示すとおり，近似計算は，$|\lambda_{max}/\lambda_{min}|$から計算した値の 7/10 でかなりよく一致する．両者のノルムの取り方が違うことを考慮に入れれば，この程度の差は止むおえない．非対称行列では式(1.1)と(1.2)の関係から条件数は固有値からは計算できない．上方の２つのラインで示すとおり，近似値的方法では$|\lambda_{max}/\lambda_{min}|$よりもかなり大きな値となっている．右端の数字は 420×420 の行列の条件数を示している．なお，Hager の近似法[3]でもほぼ同じ条件数を得る．

4.3 前処理後の行列の最大･最小固有値と条件数

前処理によって条件数がどの程度減少するの

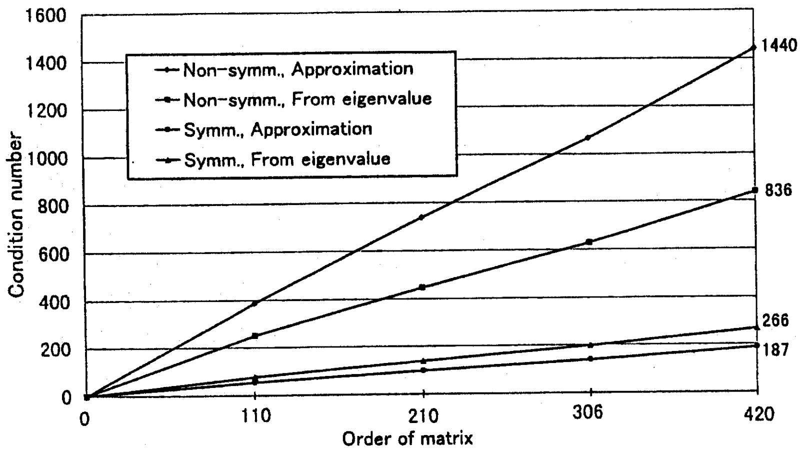

Fig.1 Comparison of condition number of matrix between two methods

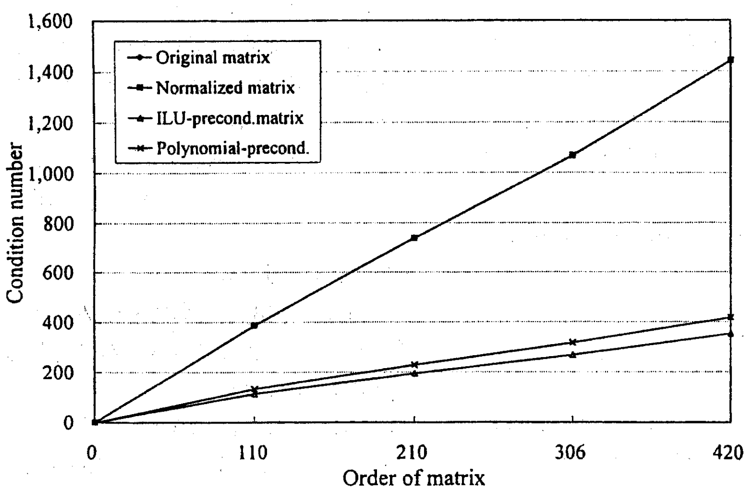

Fig.2 Condition number of four non-symmetric matrices

かを非対称行列で比較する．Fig.2 には，(1)元の行列，(2)正規化後の行列（式(3.1)による），(3) ILU 前処理（式(3.6)による），(4)行列多項式前処理（式(3.7)による）の行列 の条件数を近似法で求め図示した．ただし，ケース(1)，(2)はプロットが重なっている．差分法によるこの種の行列の条件数は行列の次数に対して線形に増加すること，このように対角要素のばらつきが少ない行列では正規化によって条件数は変わらないこと，行列多項式前処理によって条件数は対称行列で約 1/7.5 に，非対称行列では 1/4 に減少すること，行列多項式前処理よりも ILU 前処理の方が条件数は小さくなることなどが数値実験によって確認できた．一方，行列多項式前処理の方がベクトル処理や並列処理向きであるため，ILU 前処理で超平面法を用いるよりも計算の高速化になり有利な面もある [1]．

Table 1 には，420×420 の行列の各種行列の $|\lambda_{max}|$ と $|\lambda_{min}|$（その逆数），近似計算による行列のノルムと逆行列のノルムを示した．また，対称行列では固有値から求めた条件数と CG 法での反復回数を，非対称行列では近似法で求めた条件数と Bi-CGSTAB 法での反復回数を参考までに示した．

4.4 収束判定と解の有効桁数

式(2.8)は，条件数が小さい行列では相対誤差が小さくなることを示している．Fig.3 は次数 1190 の非対称行列に Bi-CGSTAB 法を適用したときの 4 つのケースでの計算格子上での解の相対誤差を示している：(A)元の行列を $\varepsilon=10^{-8}$ で計算，(B)正規化後の行列を $\varepsilon=10^{-8}$ で計算，(C) ILU 前処理後 $\varepsilon=10^{-8}$ で計算，(D) ILU 前処理後 $\varepsilon=10^{-7}$ で計算．なお，真の解として $\varepsilon=10^{-12}$ で計

Table 1 $|\lambda_{max}|$, $|\lambda_{min}|$, and $\|A\|$, $\|A^{-1}\|$, etc. for various matrices (420 × 420)

行列の種類		$\lvert\lambda_{max}\rvert$	$\|A\|$ の近似値	$\lvert\lambda_{min}\rvert$	$\|A^{-1}\|$ の近似値	条件数	反復回数 $\varepsilon=10^{-8}$
対称行列	正規化前	16.9	16.0	0.0718 (13.9)	16.6	235.0	60
	正規化後	1.99	2.00	0.0107 (93.5)	133	187.0	60
	ILU 前処理	1.00	1.61	0.0417 (24.0)	34.1	24.0	27
	行列多項式前処理	1.23	1.89	0.0396 (25.3)	35.1	31.1	24
非対称行列	正規化前	16.2	20.0	0.0191 (52.36)	72.0	1440	72
	正規化後	2.00	2.50	0.00239 (418)	576	1440	71
	ILU 分解	1.00	2.01	0.00922 (108)	175	351	27
	行列多項式前処理	1.25	2.39	0.00893 (112)	175	418	31

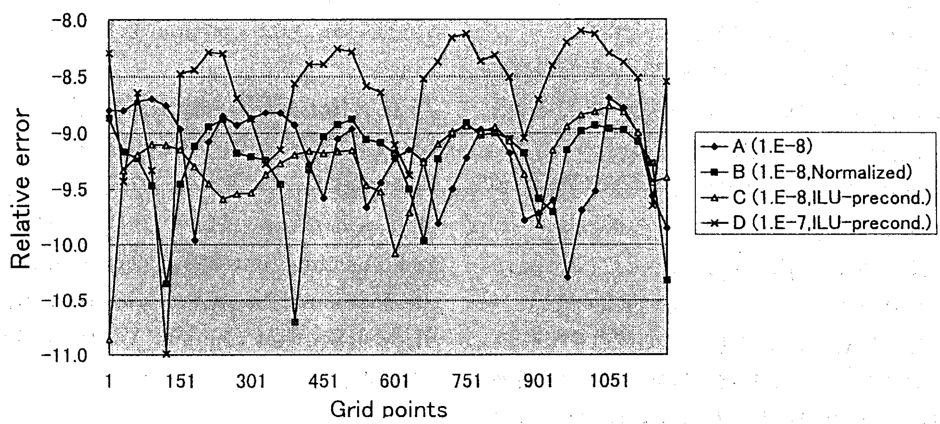

Fig.3 Comparison of accuracy between convergence criteria by Bi-CGSTAB

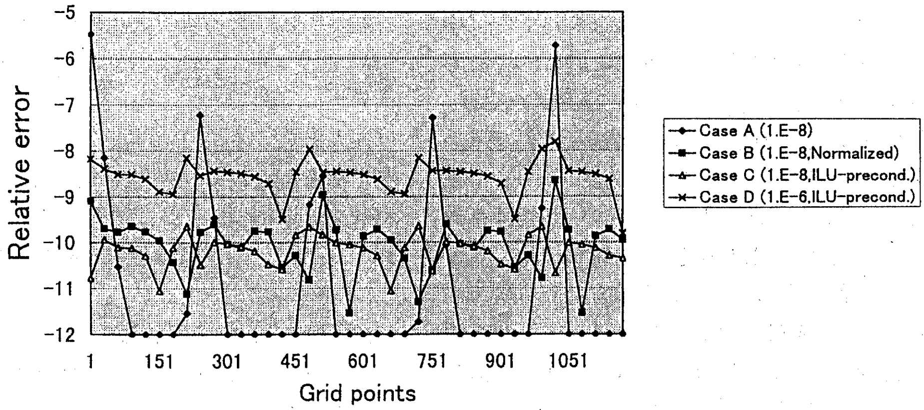

Fig.4 Comparison of accuracy for a diagonal dominant matrix

算したものを用いた．次数 1190 のときの条件数は元の行列では 3895，ILU 前処理後は 1025 で 1 桁以下の差である．ILU で収束判定を緩くした条件(D)でも条件(A)と同じ程度の精度が得られることが検証できた．このときの反復回数は 100，100，43，38 である．Fig.4 では，式(3.5)を検証する目的で優対角行列の正規化による精度向上を調査した．2 次元 (i,j)-格子に対応する対角要素を j 倍して，(A)〜(D)の 4 ケースで誤差を調べた．この場合，次数 1190 で $d_{min} / d_{max} = 1/34$ である．正規化なしの場合は部分的に誤差の大きい解が出るが，正規化後の誤差は一様に縮小し ε を 2 桁上げても所定の精度が得られた．

5．おわりに

線形方程式の CG 法系解法における行列の条件数と収束判定との関係を数値的に考察した．

(1) 塚本らの方法による条件数の近似計算の有効性を対称行列で評価した．他の近似方法でも対称行列では精度よく条件数を求めることができる．

(2) 非対称行列に対し 2 通りの方法で前処理された行列の条件数を近似計算した．条件数 κ からの反復回数の推定 $\log\varepsilon / \log((\sqrt{\kappa}-1)/(\sqrt{\kappa}+1))$ [5]との係わりについても，今後，考察していく．

(3) Bi-CGSTAB 法による解の相対誤差を定性的，定量的に調査し，前処理後の行列に対して，条件数の減少度を考慮した収束判定条件が適用できることを示した．また，対角要素の値にばらつきが大きいときは，正規化によって条件数が減少し，計算精度が上がること示した．

参考文献

[1] 石黒美佐子，鈴木信太郎，浅田 潤：共役勾配法系アルゴリズムにおけるベクトル、並列処理向き前処理法，日本応用数理学会論文誌，Vol.7, No.2 (1997)，109-119.

[2] 森 正武，数値解析法：朝倉書店, 1984.

[3] 松尾，杉原，森：行列の条件数の推定方法と数値的評価，日本応用数理学会論文誌，Vol.7, No.3, (1997)，307-319.

[4] Smith,G.D.（藤川洋一郎訳）：偏微分方程式の解法，サイエンス社，1989.

[5] Sonneveld,P : CGS, A Fast Lanczos-type Solver for Nonsymmetric Linear Systems, SIAM J. Sci. Stat. Comput., Vol.10, No.1 (1989), 36-52.

[6] 張 紹良，藤野清：丸め誤差の分離に基づく共役勾配系統の解法の収束特性の考察，日本応用数理学会論文誌，Vol.3, No.3 (1993)，135-146.

[7] Van der Vorst, H.A. : Bi-CGSTAB; A Fast and Smoothly Converging Variant of Bi-CG for the Solution of Nonsymmetric Linear Systems, SIAM J. Sci. Stat. Comput., Vol.13, No.2 (1992), 631- 644.