数理解析研究所講究録
第422巻 1981年 1-11

方程式系の近似解に対する誤差評価

愛媛大 理　山本 哲朗

$x^{(0)}$ を $\mathbb{R}^n$ または $\mathbb{C}^n$ のある領域 D で定義された方程式

(1)　$f(x)=(f_1(x),\cdots,f_n(x))^t=0$

の近似解とするとき，その精度を判定することは，数値解析における主要な課題の一つであって，多くの文献があるが，それらは通常適当なノルムを用いてなされている。しかし，$x^{(0)}$ の成分の絶対値に大きな差異がある場合，通常のノルム評価は好ましくない。本稿では，筆者が最近発表した一連の結果 [11], [12], [13] を更に一般化し，代数的問題に対する浮動小数点誤差解析の結果を数値例と共に報告したい。

§1. 解の存在定理

次の記号を用いる。$x=(x_i)$, $y=(y_i)\in\mathbb{R}^n$ or $\mathbb{C}^n$ に対し

$$|x|=(|x_i|),\quad \rho(x,y)=|x-y|.$$

また，$x,y\in\mathbb{R}^n$ かつ $x_i\geq y_i\ (\forall i)$ のとき $x\geq y$ または $y\leq x$ とおく。行列，テンソルについても同じ記号を用いる。さらに

2

$v \geqq 0$ に対し

$$U(x^{(0)}, v) = \{x \in \mathbb{R}^n (\text{or } \mathbb{C}^n) \mid \rho(x, x^{(0)}) \leqq v\}$$

とおく．以下において次のことを仮定する。

(i) f のヤコビ行列 $J(x) = \left(\frac{\partial f_i}{\partial x_j}\right)$ は D 上存在する．

(ii) n 次行列 $L = (l_{ij})$ で

$$K = V[I_n - LJ(x^{(0)})] = (\kappa_{ij})$$

のスペクトル半径 $\rho(K) < 1$ となるものが存在する。

(iii) 3次の対称テンソル $H = (h_{ijk}(D))$ で

(2) $$\rho(LJ(x), LJ(y)) \leqq H\rho(x, y), \quad x, y \in D$$

となるものが存在する。

このとき，次の定理が成り立つ．

定理1. $\|\cdot\|$ を任意のノルムとし，$\|x\| = \|y\| = 1$ なるすべての x, y につき

(3) $$Kx \leqq \kappa, \quad Hxy \leqq h.$$

かつ $\|\kappa\| < 1$ とする。さらに

$$\varepsilon = V[Lf(x^{(0)})], \quad t = (1 - \|\kappa\|)^2 - 2\|h\| \cdot \|\varepsilon\| \geqq 0$$

として

$$a = \frac{2\|\varepsilon\|}{1 - \|\kappa\| + \sqrt{t}}, \quad \beta = \varepsilon + a\kappa + \frac{1}{2}a^2 h = (\beta_1, \cdots, \beta_m)^t$$

とおけば

(i) $U(x^{(0)}, \beta)$ 内に (1) の解 x^* が存在する

(ii) $\{\beta^{(k)}\}$ を

$$\beta^{(0)}=\beta,\ \beta^{(k+1)}=\varepsilon+K\beta^{(k)}+\frac{1}{2}H\beta^{(k)2},\quad k\geq 0$$

により定義すれば，$\beta^{(0)}\geqq\beta^{(1)}\geqq\cdots\to\beta^*\geqq 0$ かつ $x^*\in U(x^{(0)},\beta^*)$.

(iii) x^* は $U(x^{(0)},\beta^*)$ 内で一意的.

(iv) 特に $t>0$ ならば x^* は単純である.

(証明). Schröder, Collatz による majorant principle による.

詳細は 山本[14].

系. 定理1の仮定の下で，$x^{(1)}=x^{(0)}-Lf(x^{(0)})$ とおけば

$$x^*\in U(x^{(1)},\beta^*-\varepsilon).$$

注意1. ノルムとして l_1 ノルム $\|\cdot\|_1$ と最大ノルム $\|\cdot\|_\infty$ をとれば $\kappa=(\kappa_i)$, $h=(h_i)$ はそれぞれ

$$\|\cdot\|_1:\quad \kappa_i=\max_j \kappa_{ij},\quad h_i=\max_{j,k} h_{ijk}(D)$$

$$\|\cdot\|_\infty:\quad \kappa_i=\sum_{j=1}^{n}\kappa_{ij},\quad h_i=\sum_{j,k=1}^{n}h_{ijk}(D)$$

により与えられる.

注意2. 定理1は最大ノルムにつき，Newton-Kantorovich の定理を改良している. 実際，f が D において C^2 級とし

$$\left|\frac{\partial^2 f_i}{\partial x_j\partial x_k}\right|\leqq m_{ijk}\quad(\forall i,j,k),\quad M=(m_{ijk})$$

とすれば，注意1により

$$\|\kappa\|_\infty=\|K\|_\infty,\quad \|h\|_\infty=\|H\|_\infty\leqq\|L\|_\infty\cdot\|M\|_\infty.$$

故に $\tilde{t}=(1-\|\kappa\|_\infty)^2-2\|L\|_\infty\cdot\|M\|_\infty\geq 0$ ならば

$$t_\infty=(1-\|K\|_\infty)^2-2\|h\|_\infty\cdot\|\varepsilon\|_\infty\geq 0$$

で

$$\|x^{(0)}-x^*\|_\infty \leqq \|\beta\|_\infty \leqq a_\infty = \frac{2\|\varepsilon\|_\infty}{1-\|K\|_\infty+\sqrt{t_\infty}} \leqq \frac{2\|\varepsilon\|_\infty}{1-\|K\|_\infty+\sqrt{t}}$$

ここで特に $L=J(x^{(0)})^{-1}$ とと小は $K=0$ で，上式最後の値は Newton－Kantorovich の評価と一致する。

さて，定理1の誤差限界ベクトル β^* は L に依存している。では，どのような L に対し最良の β^* が与られるであろうか。当然予想される結果は

$$\beta^*_{J(x^{(0)})^{-1}} \leqq \beta^*_L \quad (\forall_L)$$

である。実際，この予想は，ある条件の下で正しい。これをみるために，K，ε，t，β^* 等を K_L，ε_L，t_L，β^*_L 等とかき，特に $L=J(x^{(0)})^{-1}$ に対するそれらを $\hat{K}$，$\hat{\varepsilon}$，$\hat{t}$，$\hat{\beta}^*$ 等により表わす。

定理2. H_L は次の意味で最良とする。すなわち $H'_L \geqslant 0$ を

$$\rho(LJ(x),\ LJ(y)) \leqq H'_L \rho(x,y), \quad x,y \in D$$

をみたす3次の対称テンソルとするとき，$H_L \leqq H'_L$。もし $\|\cdot\|$ が単調で $\|K_L\|<1$ かつ $t_L=(1-\|K_L\|)^2-2\|h_L^2\|\cdot\|\varepsilon_L\|>0$ ならば

$$\hat{t}=1-2\|\hat{h}_L^2\|\cdot\|\hat{\varepsilon}\|>0 \quad かつ \quad \hat{\beta}^* \leqq \beta^*_L.$$

(証明) 山本[14].

§2. 応用

2.1. 連立一次方程式

定理3. $x^{(0)}$ を $Ax=b$ の近似解，L を A^{-1} の近似とし

$$\|K\|_\infty<1 \quad (但し\ K=\nu[I_n-LA]=(K_{ij}))$$

と仮定する。

$$\varepsilon = \nu[L(Ax^{(0)}-b)],\quad \kappa_i = \sum_{j=1}^{m}\kappa_{ij},\quad \kappa = (\kappa_1,\cdots,\kappa_m)^t$$

$$\beta^{(0)} = \varepsilon + \frac{\|\varepsilon\|_\infty}{1-\|\kappa\|_\infty}\kappa,\qquad \beta^{(k+1)} = \varepsilon + K\beta^{(k)},\quad k\geq 0$$

とすれば $\beta^{(0)} \geqq \beta^{(1)} \geqq \cdots \to \beta^* = (I_m - K)^{-1}\varepsilon$ かつ

$$|x_i^* - x_i^{(0)}| \leqq \beta_i^* \leqq \beta_i^{(k)},\quad i=1,2,\cdots,n\quad (k\geq 0).$$

2.2. 固有値問題

実行列 A の近似実固有値 $\lambda^{(0)}$ と近似実固有ベクトル $x^{(0)}$ が与えられたとする。$z^{(0)} = (x_1^{(0)},\cdots,x_n^{(0)},\lambda^{(0)})^t$ は

$$f(z) = \begin{pmatrix} f_1(z) \\ \vdots \\ f_{n+1}(z) \end{pmatrix} = \begin{pmatrix} Ax-\lambda x \\ \frac{1}{2}(1-\|x\|_2^2) \end{pmatrix} = 0$$

の近似解とみなせる。この場合，$m = n+1$ として

$$J(z^{(0)}) = \begin{pmatrix} A-\lambda^{(0)}I_n & -x^{(0)} \\ -x^{(0)t} & 0 \end{pmatrix}$$

$$H = \nu\left[\left(\begin{array}{cc|cc|c|cc} l_{1m} & 0\cdots 0\, l_{11} & 0 & l_{1m}\ 0\cdots 0\ l_{12} & \cdots & 0\cdots 0\, l_{1m}\, l_{1n} & l_{11}\cdots l_{1n}\ 0 \\ l_{2m} & 0\cdots 0\, l_{21} & 0 & l_{2m}\ 0\cdots 0\ l_{22} & & 0\cdots 0\, l_{2m}\, l_{2n} & l_{21}\cdots l_{2n}\ 0 \\ \cdots & & \cdots & & & & \\ l_{mm} & 0\cdots 0\, l_{m1} & 0 & l_{mm}\ 0\cdots 0\ l_{m2} & & 0\cdots 0\, l_{mm}\, l_{mn} & l_{m1}\cdots l_{mn}\ 0 \end{array}\right)\right]$$

に注意すれば次の結果を得る。

定理4. $\|\cdot\|$ を任意のノルムとし，L は $J(z^{(0)})^{-1}$ の近似とする。

$$K = \nu[I_{n+1} - LJ(z^{(0)})],\quad \varepsilon = \nu[Lf(z^{(0)})]$$

とし，κ, h を定理1のように定義する。$\|\kappa\|<1$ かつ $t = (1-\|\kappa\|)^2 - 2\|h\|\cdot\|\varepsilon\| \geq 0$ と仮定して

$$a = \frac{2\|\varepsilon\|}{1-\|\kappa\|+\sqrt{t}},\quad \beta = \varepsilon + a\kappa + \frac{1}{2}a^2 h = (\beta_1,\cdots,\beta_{n+1})^t,$$

$$\beta^{(0)} = \beta,\quad \beta^{(k+1)} = \varepsilon + K\beta^{(k)} + \frac{1}{2}H\beta^{(k)2},\quad k\geq 0$$

とおけば 実固有値λ^*と実固有ベクトルx^*（$\|x^*\|_2=1$）が存在して

$$\begin{pmatrix} x^* \\ \lambda^* \end{pmatrix} \in U(z^{(0)}, \beta^*).$$

すなわち

$$|x_i^* - x_i^{(0)}| \leqq \beta_i^* \leqq \beta_i^{(k)}, \quad i=1,2,\cdots,n$$

$$|\lambda^* - \lambda^{(0)}| \leqq \beta_{n+1}^* \leqq \beta_{n+1}^{(k)}, \qquad k \geq 0.$$

この結果は，すべての固有値，固有ベクトルを必要としないこと，Aの対称性を仮定しないこと等，従来のものより実用的と考えられる。

§3. 誤差解析

実計算では，$\beta^{(k)}$は正しく計算されないから，定理3,4を直接応用するには幾分の配慮が必要となる。いま演算結果をN進β桁の仮数部に切捨てる計算機を考え，Wilkinson[9]，Paige[4]に従がって，$n \geq 2$，かつ

(4) $fl(\hat{a} \circ \hat{b}) = \hat{a} \circ \hat{b}(1+\xi)$, $|\xi| < N^{1-\beta}$

(5) $fl(\sqrt{\hat{a}}) = \sqrt{\hat{a}}(1+\eta)$, $|\eta| < 1.01\, N^{1-\beta}$

(6) $1.006(n+1)N^{1-\beta} < 0.01$

を仮定する。ただし$\hat{a}, \hat{b}$はa, bのN進β桁浮動小数点表現とし，$\circ$は四則演算の一つをあらわすものとする。式c, d, ...についても，それらの演算結果を$\hat{c}, \hat{d}, \cdots$であらわす。

3.1. 定理3に対する誤差解析（山本[16]）

簡単のため $\hat{A}=A$, $\hat{b}=b$, $\hat{x}^{(0)}=x^{(0)}$ とする。また L は A^{-1} に対する計算結果とし，$\hat{L}=L$ とする。このとき，(4)と(6)を用いて定理3にあらわれる諸量 $\hat{K}$, $\hat{\varepsilon}$, $\hat{\alpha}$, … を評価する。結果は次の通り。

定理5. $A=(a_{ij})(=\hat{A})$, $\theta_n = n N^{1-\Lambda}$

$$\hat{A}_\infty = \max_i \left(\sum_{j=1}^{n} |a_{ij}| \right), \quad \hat{L}_\infty = \max_i \left(\sum_{j=1}^{n} |l_{ij}| \right),$$

$$\hat{\kappa}_\infty = \max_i \left(\sum_{j=1}^{n} \hat{\kappa}_{ij} \right) = \|\hat{\kappa}\|_\infty,$$

$$\hat{c}_\infty = 1.02\, \hat{A}_\infty \|x^{(0)}\|_\infty + 0.502 \|\widehat{Ax^{(0)}-b}\|_\infty$$

$$\hat{d}_\infty = 1.02\, \hat{L}_\infty \left(\|\widehat{Ax^{(0)}-b}\|_\infty + \hat{c}_\infty \right)$$

$$\hat{e}_\infty = 1.03 \left(\hat{L}_\infty \hat{A}_\infty + \hat{\kappa}_\infty \right) + 0.502 \max_i \hat{\kappa}_{ii}$$

とおく。もし

$$\hat{e}_\infty \theta_n < 1 \quad \text{すなわち} \quad \hat{\kappa}_\infty < 1 - \frac{1}{m} - \hat{e}_\infty \theta_n \quad (m>0)$$

ならば

$$\hat{f}_\infty = 1.004(m-1)\left\{ \left(1 + \frac{1}{n} + m\hat{e}_\infty \theta_n\right) \hat{d}_\infty + \left(2n^{-1} + m\hat{e}_\infty\right) \|\hat{\varepsilon}\|_\infty \right\}$$

$$\Delta\hat{\beta}_i = 1.004\, N^{1-\Lambda} \hat{\beta}_i + \left\{ \hat{d}_\infty \left(1 + m\hat{e}_\infty \theta_n\right) + m\|\hat{\varepsilon}\|_\infty \hat{e}_\infty + \hat{f}_\infty + \frac{1}{n} \hat{\alpha} \hat{\kappa}_\infty \right\} \theta_n$$

とおいて

$$|x_i^* - x_i^{(0)}| \leqq \hat{\beta}_i + \Delta\hat{\beta}_i, \quad i = 1, 2, \cdots, n.$$

注意3. $\hat{L}_\infty \hat{A}_\infty$ は A の条件数を近似する量であり，これが大きくなれば $\Delta\hat{\beta}_i$ は大きくなく，定理3は有効に効く。また方程式が悪条件の場合，計算桁数 t をどのようにとるべきかを定理5は示唆している。

3.2. 定理4に対する誤差解析.(山本[15]).

定理5と同様に次の結果が成り立つ.

定理6. $\beta_i < (1+1.03\theta_2+1.02\theta_{n+1})\hat{\beta}_i + \theta_{n+1}\delta_i$, $i=1,2,\cdots,n+1$

但し $\hat{\varepsilon}>0$ を仮定し,

$$\delta_i = e_i + 1.004(ag_i+\alpha) + 2^{-1}(2a+\alpha\theta_{n+1})\alpha h_i,$$

$$a = 2d^{-1}\|\varepsilon\|_1, \quad d = 1-\|\kappa\|_1+\sqrt{\varepsilon},$$

$$\alpha = 2.02(d\hat{d})^{-1}\{(\|e\|+1.4\|\varepsilon\|_1)d + (p+\omega+1)\|\varepsilon\|_1\}$$

$$p = 1.02\,fl(g_1+\cdots+g_{n+1}) + 1.02\,fl(\hat{\kappa}_1+\cdots+\hat{\kappa}_{n+1}) + 1$$

$$g_i = \max_j g_{ij}$$

$$g_{ij} = \frac{1.004}{n+1}\{|l_{ij}|\cdot|\hat{u}_{jj}|(1-\delta_{j,n+1}) + |fl(1-fl(LJ)_{jj})|\}\delta_{ij} + d_{ij}$$

$$(\hat{l}_{ij}) = fl(A-\lambda^{(0)}I_n)$$

$$d_{ij} = 1.02\,fl(|l_{i1}|+\cdots+|l_{i,n+1}|)\|J^{(j)}\|_\infty \quad (J^{(j)}\text{は}J(\hat{x}^{(0)})\text{の第}j\text{列})$$

$$e=(e_1,\cdots,e_{n+1})^t,$$

$$e_i = 1.02\{1.006\|\hat{f}\|_\infty + \frac{n}{n+1}\|c\|_\infty\}\,fl(\sum_{j=1}^{n+1}|l_{ij}|)$$

$$f = (f_1(\hat{x}^{(0)}),\cdots,f_n(\hat{x}^{(0)}))^t,$$

$$c = (c_1,\cdots,c_{n+1})^t, \quad c_i = \begin{cases} fl(\sum_{j=1}^{n+1}|\hat{b}_{ij}|)(1.02)\|x^{(0)}\|_\infty & (i\le n) \\ \frac{1.004}{n}|\hat{f}_{n+1}| + 0.508\|x^{(0)}\|_2^2 & (i=n+1) \end{cases}$$

$$\omega = 1+\frac{\tau}{\sqrt{\hat{\varepsilon}}}, \quad \tau = g + 2.02\|h\|\cdot\|e\| + 2$$

$$g = 2p + p^2\theta_{n+1} + 1$$

§4. 数値例

例1. Wilkinsonのよった悪条件の方程式

$$0.876543x_1 + 0.617341x_2 + 0.589973x_3 = 0.863257$$

$$0.612314\,x_1 + 0.784461\,x_2 + 0.827742\,x_3 = 0.820647$$

$$0.317321\,x_1 + 0.446779\,x_2 + 0.476349\,x_3 = 0.450098$$

を FACOM 230-28（単精度 16進 6桁）で Gauss 消去（行交換）を用いて解くと

$$x^{(0)} = (0.6363233,\ -0.2946413E-1,\ 0.5486381)^t,$$

を得る．A^{-1} の計算値 $L(=\hat{L})$ は $\hat{L}_\infty = 0.657\cdots E5$（山本[C]）であった．この場合，倍精度（16進14桁）計算により $\hat{K}_\infty$, $\hat{e}_\infty$, … を求めて，定理5を適用すれば

$$\hat{\beta} = \begin{pmatrix} 0.573591\cdots E-5 \\ 0.427810\cdots E-4 \\ 0.362315\cdots E-4 \end{pmatrix}, \quad \Delta\beta_i = 0.150\cdots E-9 < 0.151E-9.$$

故に $\nu[x^* - x^{(0)}] \leqq \hat{\beta} + 0.151E-9\begin{pmatrix}1\\1\\1\end{pmatrix} \leqq \begin{pmatrix} 0.574E-5 \\ 0.428E-4 \\ 0.363E-4 \end{pmatrix}$.

例2. 行列

$$A = \begin{pmatrix} 14 & 9 & 6 & 4 & 2 \\ -9 & -4 & -3 & -2 & -1 \\ -2 & -2 & 0 & -1 & -1 \\ 3 & 3 & 3 & 5 & 3 \\ -9 & -9 & -9 & -9 & -4 \end{pmatrix}$$

の近似固有値 $\lambda^{(0)} = 4.990$ と近似固有ベクトル $x^{(0)} = (0.707,\ -0.707,\ 0.0002,\ -0.0001,\ 0.0000)^*$ が与えられたとき，それらの誤差をしらべる．

定理4により $\hat{\xi} = 0.799\cdots < 1$ かつ

$$\hat{\beta} = \begin{pmatrix} 0.1682048E-3 \\ 0.1686592E-3 \\ 0.2619984E-3 \\ 0.1619263E-3 \\ 0.6216292E-4 \\ 0.1005869E-1 \end{pmatrix}, \quad \hat{\beta}^{(1)} = \begin{pmatrix} 0.1084336E-3 \\ 0.1088856E-3 \\ 0.2015247E-3 \\ 0.1009613E-3 \\ 0.4388653E-5 \\ 0.1003479E-1 \end{pmatrix}$$

この場合，定理6により

$$\beta_i \leqq \hat{\beta}_i + 10^{-12}$$

を得る．（詳細は 山本[15]）．故に 定理4(6) は有効に働く．

References

[1] Collatz,L.: Functional analysis and numerical mathematics, Academic Press 1966.

[2] Kantorovich,L.V. and Akilov,G.P.: Functional analysis in normed spaces. Pergamon Press 1964.

[3] Ortega,J.M. and Rheinboldt,W.C.: Iterative solution of nonlinear equations in several variables, Academic Press 1970.

[4] Paige,C.C.: Error analysis of the symmetric Lanczos process for the eigenproblem, London Univ.Inst. of Computer Science,Tech.Note ICSI 209, 1969.

[5] Rall,L.B.: Computational solution of nonlinear operator equations, John Wiley & Sons.Inc. 1969.

[6] Schröder,J.: Nichtlineare Majoranten beim Verfahren der schrittweiseen Näherung, Arch.Math.7(1956),471-484.

[7] Schröder,J.: Über das Newtonsche Verfahren, Arch.Rat.Mech.Anal.1(1957), 154-180.

[8] Urabe,M.: A posteriori component-wise error estimation of approximate solutions to nonlinear equations, Lect.Notes in Computer Sci.29,, Springer, 1975.

[9] Wilkinson,J.H.: Rounding errors in algebraic processes, Prentice Hall 1963.

[10]Wilkinson,J.H.:The algebraic eigenvalue problem, Oxford Univ Press 1965.

[11]Yamamoto,T.: Componentwise error estimates for approximate solutions of nonlinear equations, JIP 2(1979), 121-126.

[12]Yamamoto,T.: An existence theorem for solutions of nonlinear systems and its applications to algebraic equations, In 3rd USA-JAPAN computer conference proceedings,300-304. AFIPS and IPSJ 1978.

[13] Yamamoto,T.: Error bounds for computed eigenvalues and eigenvectors, Numer.Math.34(1980), 189-199.

[14] Yamamoto,T.: Error bounds for approximate solutions of systems of equations, to appear.

[15] Yamamoto,T.: Componentwise error estimates for approximate solutions of systems of equations, to appear.

[16] Yamamoto,T.: A posteriori componentwise error estimate for a computed solution of a system of linear equations, to appear.