

浮動小数点型演算による反復法の component-wise な誤差解析

久大 理 占 部 実

まえがき

与えられた方程式を

$$\varphi_i(x_1, x_2, \dots, x_d) = 0 \quad (i=1, 2, \dots, d),$$

あるいはベクトル記法を用いて

$$(0.1) \quad \varphi(x) = 0$$

とする。この方程式を反復法を用いて解くには、上の方程式を

$$(0.2) \quad x = f(x)$$

の形に書き直し、適当な $x^{(0)}$ から出発して

$$(0.3) \quad x^{(n+1)} = f[x^{(n)}] \quad (n=0, 1, 2, \dots)$$

により, $x^{(1)}, x^{(2)}, \dots$ を逐次求め, 適当なところでこの反復計算を止めて, 最後に得られた $x^{(n)}$ をもって望む近似解とする.

$f[x^{(n)}]$ が固定小数点型演算で計算される場合, 上の反復法が実際計算ではどのようなになるかは, 一応「1」, 「2」で論じたので, ここでは $f[x^{(n)}]$ が浮動小数点型演算で計算される場合について考察する.

なお, 連立方程式の解の誤差を考察するとき, 誤差評価には普通, 誤差ベクトルのノルムが用いられているが, これはしばしば過大評価を与えることになって, 実際の誤差評価には必ずしも適切ではない. そこで, ここでは, 誤差評価を成分毎に行う方法をとった.

われわれの結果は, 以下「注意」で述べられているように, ノルムを用いる場合にも成り立つので, 必要があれば, ノルムを用いる場合にそのまま適用して差し支えない.

§1. 予備

定義1. "任意のベクトル $v = (v_1, v_2, \dots, v_n)^T$ (T は転置を表わす) に対して, ベクトル $(|v_1|, |v_2|, \dots, |v_n|)^T$ を v の 絶対ベクトル とよび, $\sigma(v)$ で表わす."

(1)

$\sigma(v)$ の第 i 成分を $\sigma_i(v)$ で表わすと, 明らかに

$$\sigma_i(v) = |v_i| \quad (i=1, 2, \dots, d)$$

である.

定義 2. "任意の行列 $A = (a_{ij})$ ($i, j = 1, 2, \dots, d$) に対して, i, j -要素が $|a_{ij}|$ である行列を A の 絶対行列 とし, $\sigma(A)$ で表わす."

$\sigma(A)$ の i, j -要素を $\sigma_{ij}(A)$ で表わすと, 明らかに

$$\sigma_{ij}(A) = |a_{ij}| \quad (i, j = 1, 2, \dots, d)$$

である.

定義 3. "ベクトル $v = (v_1, v_2, \dots, v_d)^T$, $w = (w_1, w_2, \dots, w_d)^T$ に対して $v_i \geq w_i$ ($i=1, 2, \dots, d$) が成り立つとき, $v \geq w$ と書く. また, 行列 $A = (a_{ij})$, $B = (b_{ij})$ に対して $a_{ij} \geq b_{ij}$ ($i, j = 1, 2, \dots, d$) が成り立つとき, $A \geq B$ と書く."

もちろん, $v \geq w$ のときには, $w \leq v$ とも書き, $A \geq B$ のときには $B \leq A$ とも書く. 上の定義から明らかのように,

$$v_i \geq 0 \quad (i=1, 2, \dots, d) \quad \text{と} \quad v \geq 0,$$

$$a_{ij} \geq 0 \quad (i, j = 1, 2, \dots, d) \quad \text{と} \quad A \geq 0,$$

はそれぞれ互に同値であり, また

$$v \geq w \quad \text{と} \quad v - w \geq 0,$$

$$A \geq B \quad \text{と} \quad A - B \geq 0,$$

はそれぞれ互に同値である.

補題 1. " v, w をベクトル, A, B を行列とすると,

$$\sigma(v+w) \leq \sigma(v) + \sigma(w),$$

$$\sigma(A+B) \leq \sigma(A) + \sigma(B),$$

$$\sigma(Av) \leq \sigma(A)\sigma(v),$$

$$\sigma(AB) \leq \sigma(A)\sigma(B).$$

また, $\sigma(v) \geq 0, \sigma(A) \geq 0$ であり

$$\sigma(v) = 0 \quad \text{と} \quad v = 0,$$

$$\sigma(A) = 0 \quad \text{と} \quad A = 0,$$

はそれぞれ互に同値である. "

この補題は, 定義 1, 2, 3 により明らかである.

つぎの補題はよく知られていることであるが, 以下でしばしば用いるので, ここに一応あげておく.

補題2. "任意の正定値行列 H に対して, その spectral radius を $\rho(H)$ で表わすと, $\rho(H) < 1$ ならば $\lim_{n \rightarrow \infty} H^n = 0$ であり, また逆に $\lim_{n \rightarrow \infty} H^n = 0$ ならば $\rho(H) < 1$ である. ただしここで n は正の整数である."

補題3. "任意の正定値行列 H に対して, $\rho(H) < 1$ ならば, $(E - H)^{-1}$ (E は単位行列) が存在し,

$$(E - H)^{-1} = \sum_{n=0}^{\infty} H^n$$

が成立する."

補題3から, 正定値行列 $H \geq 0$ に対して $\rho(H) < 1$ ならば, $(E - H)^{-1} \geq 0$ であることが, すぐわかる.

[注意] $\sigma(\cdot)$ をノルム, あるいは行列のノルムとしても, 補題1はそのまま成立する. このことは, 以下の議論において, $\sigma(\cdot)$ をノルム, あるいは行列のノルムとしても結果はそのまま成立することを示すことになる.

§2. 反復法の収束定理

定理1. " $f(x)$ は x -空間の集合 D で定義され, D に属する任意の x', x'' に対して

$$(2.1) \quad \sigma[f(x') - f(x'')] \leq K_0 \sigma(x' - x'')$$

をみたす, とする. ただし K_0 は正定行列で

$$(2.2) \quad K_0 \geq 0, \quad \rho(K_0) < 1$$

をみたすものとする.

D に属し, 次の条件をみたすような $x^{(0)}$ が存在するとする:

$$(2.3) \quad S = \{x \mid \sigma[x - x^{(0)}] \leq K_0 (E - K_0)^{-1} \sigma[x^{(0)} - x^{(0)}]\} \subset D.$$

ただしここで

$$(2.4) \quad x^{(1)} = f[x^{(0)}]$$

である.

このとき,

$$(2.5) \quad x^{(n+1)} = f[x^{(n)}] \quad (n=0, 1, 2, \dots)$$

によって定まる列 $\{x^{(n)}\}$ ($n=1, 2, \dots$) は S の中にある
 24収束し, その極限

$$(2.6) \quad \hat{x} = \lim_{n \rightarrow \infty} x^{(n)}$$

は, 方程式'

$$(2.7) \quad x = f(x)$$

の D における唯一の解である。"

[証明] まず, (2.5) により S の中に無限点列 $\{x^{(n)}\}$ が得られることを示そう。

$x^{(1)} \in S$ は明らかであるから, $x^{(1)}, x^{(2)}, \dots, x^{(n)}$ が S の中に得られた, とする。すると, $x^{(n)} \in S \subset D$ であるから, $x^{(n+1)} = f[x^{(n)}]$ によって $x^{(n+1)}$ が得られる。問題は、この $x^{(n+1)}$ が S に属するか, ということである。まず, (2.1) により

$$\sigma[x^{(2)} - x^{(1)}] = \sigma[f(x^{(1)}) - f(x^{(0)})] \leq K_0 \sigma[x^{(1)} - x^{(0)}]$$

を得, 同様にして

$$\sigma[x^{(3)} - x^{(2)}] \leq K_0 \sigma[x^{(2)} - x^{(1)}] \leq K_0^2 \sigma[x^{(1)} - x^{(0)}],$$

.....

$$\sigma[x^{(n+1)} - x^{(n)}] \leq K_0 \sigma[x^{(n)} - x^{(n-1)}] \leq K_0^n \sigma[x^{(1)} - x^{(0)}]$$

を得る。すると, これらの式より, σ の不等式を得る:

$$\begin{aligned} \sigma[x^{(n+1)} - x^{(1)}] &\leq \sigma[x^{(n+1)} - x^{(n)}] + \sigma[x^{(n)} - x^{(n-1)}] \\ &\quad + \dots + \sigma[x^{(2)} - x^{(1)}] \end{aligned}$$

$$\leq (K_0^n + K_0^{n-1} + \dots + K_0) \sigma[x^{(1)} - x^{(0)}]$$

$$\leq K_0 (E - K_0)^{-1} \sigma[x^{(1)} - x^{(0)}].$$

S の定義により, これは $x^{(n+1)} \in S$ を表わしている. かくて, 帰納法により, (2.5) によって S の中に無限点列 $\{x^{(n)}\}$ が得られることがわかる.

上の結果から, 任意の正の整数 r に対して

$$(2.8) \quad \sigma[x^{(n+r)} - x^{(n)}] \leq \sigma[x^{(n+r)} - x^{(n+r-1)}] + \dots + \sigma[x^{(n+1)} - x^{(n)}]$$

$$\leq (K_0^{n+r-1} + \dots + K_0^n) \sigma[x^{(1)} - x^{(0)}]$$

$$\leq K_0^n (E - K_0)^{-1} \sigma[x^{(1)} - x^{(0)}]$$

を得る. $\rho(K_0) < 1$ であるから, 補題 2 により,

$$\lim_{n \rightarrow \infty} K_0^n = 0$$

となり, 任意の正の整数 r に対して, $n \rightarrow \infty$ のとき

$$\sigma[x^{(n+r)} - x^{(n)}] \rightarrow 0$$

を得る. これは任意の正の整数 r に対して

$$\lim_{n \rightarrow \infty} |x_i^{(n+r)} - x_i^{(n)}| = 0 \quad (i=1, 2, \dots, d)$$

を意味するから,

$$\hat{x}_i = \lim_{n \rightarrow \infty} x_i^{(n)} \quad (i=1, 2, \dots, d)$$

が存在する. $\hat{x} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_d)^T$ とおけば, 上式は

$$(2.9) \quad \hat{x} = \lim_{n \rightarrow \infty} x^{(n)}$$

を意味する. $x^{(n)} \in S$ ($n=1, 2, \dots$) であるから,

$$(2.10) \quad \hat{x} \in S \subset D$$

であることは明らかである.

(2.1) により $f(x)$ は D において連続であるから,
(2.5) において $n \rightarrow \infty$ とすれば, (2.9) により

$$\hat{x} = f(\hat{x})$$

を得る. これは $\hat{x}$ が '方程式' (2.7) の解であることを示している.

最後に, '方程式' (2.7) は D の中では, 上に求められた $\hat{x}$ 以外には解をもたないことを示そう.

いま, $\hat{x}$ 以外に D の中に解 $\hat{x}'$ があったとすると, すると,

$$\hat{x}' = f(\hat{x}'), \quad \hat{x} = f(\hat{x})$$

より, (2.1) によって

$$\sigma(\hat{x}' - \hat{x}) = \sigma[f(\hat{x}') - f(\hat{x})] \leq K_0 \sigma(\hat{x}' - \hat{x})$$

を得る。したがって

$$(E - K_0) \sigma(\hat{x}' - \hat{x}) \leq 0$$

を得る。ここで、 $\rho(K_0) < 1$ であるから、補題3により、 $(E - K_0)^{-1} \geq 0$ が存在する。したがって、これを上の不等式に左から掛ければ

$$\sigma(\hat{x}' - \hat{x}) \leq 0$$

を得る。 $\sigma(\hat{x}' - \hat{x}) \geq 0$ であるから、

$$\sigma(\hat{x}' - \hat{x}) = 0, \text{ すなわち } \hat{x}' - \hat{x} = 0$$

を得る。これは $\hat{x}' = \hat{x}$ を意味し、解の一意性を示している。

これで定理の証明は完全にできたことになる。

[注意1] 行列 K_0 の要素を k_{ij} ($i, j = 1, 2, \dots, d$) とする。条件 (2.1) は

$$|f_i(x') - f_i(x'')| \leq \sum_{j=1}^d k_{ij} |x_j' - x_j''|$$

を表わしている。したがって、ベクトル $v = (v_1, v_2, \dots, v_d)^T$ のノルム $\|v\|$ として、 $\|v\| = \max_i |v_i|$ を用いると、行列 K_0 の対応するノルム $\|K_0\|$ は $\|K_0\| = \max_i \sum_{j=1}^d k_{ij}$ ($k_{ij} \geq 0$ であるから) となり、条件 (2.1) より

$$\|f(x') - f(x'')\| \leq \|K_0\| \cdot \|x' - x''\|$$

が成り立つことになる。

〔注意 2〕 $\sigma(v)$ をベクトル v のノルムとすると、 K_0 はスカラーとなり、(2.2) は

$$0 \leq K_0 < 1$$

を意味する。このとき、定理 1 の結果がそのまゝ成立することはい上の証明から明らかである。

〔注意 3〕 条件 (2.3) は、 $x^{(0)}$ が適当な方程式 (2.7) の近似解であることを要請しているものである。なぜなら、もし $x^{(0)}$ が D の内部にあり、相当な精度をもつ方程式 (2.7) の近似解であるならば、(2.4) により $x^{(1)}$ は $x^{(0)}$ に近く、したがって S は $x^{(1)}$ を中心とする小さな長方形領域になり、条件 (2.3) はあつちからみたされるからである。

§ 3. 浮動小数点型演算による反復法の誤差解析

3.1 反復法の振動定理

反復計算 (0.3) を計算機によって実行するとき, $f(x)$ の計算は浮動小数点型演算で行われるものとし, 得られる値を $\tilde{f}(x)$ で表わす. $\tilde{f}(x)$ は一般に $f(x)$ と異なっているが, いまの場合

$$(3.1) \quad \sigma[\tilde{f}(x) - f(x)] \leq \varepsilon \cdot \sigma[f(x)]$$

が成立する, と考える. ただしこのとき, ε は対角行列で, その対角線要素は正数 ε_i ($i=1, 2, \dots, d$) であるとする. この場合, (3.1) は

$$|\tilde{f}_i(x) - f_i(x)| \leq \varepsilon_i \cdot |f_i(x)| \quad (i=1, 2, \dots, d)$$

を意味し, これは $\tilde{f}_i(x)$ の相対誤差が ε_i を越えないことを示している. 便宜上, われわれは

$$(3.2) \quad \varepsilon_i \leq \hat{\varepsilon} \quad (i=1, 2, \dots, d)$$

で,

$$(3.3) \quad 0 < \hat{\varepsilon} < 1$$

であるとする. 行列 ε の定義, および (3.2) により明らかに

$$(3.4) \quad 0 \leq \varepsilon \leq \hat{\varepsilon} E$$

が成り立っている。

反復計算 (0.3) を計算機の上で実行すると、上に述べたことから、われわれは

$$(3.5) \quad \tilde{x}^{(n+1)} = \tilde{f}[\tilde{x}^{(n)}] \quad (\tilde{x}^{(0)} = x^{(0)}; n=0, 1, 2, \dots)$$

によって定まる列 $\{\tilde{x}^{(n)}\}$ を得、これは (0.3) によって定まる列 $\{x^{(n)}\}$ とは一般に異なっている。この $\{\tilde{x}^{(n)}\}$ の挙動については、つぎの定理が成り立つ。

定理2. " $f(x)$ は定理1の条件をみたしているとする。
また $\hat{\varepsilon}$ は小さくて

$$(3.6) \quad \hat{L}_0 = (1 + \hat{\varepsilon}) K_0$$

とあるとき、

$$(3.7) \quad \rho(\hat{L}_0) < 1$$

が成り立つとする。

D に属し、つぎの条件をみたすような $x^{(0)}$ が存在するとする：

$$(3.8) \quad \Sigma = \{x \mid \sigma[x - \tilde{x}^{(0)}] \leq \rho_1 + \rho_2\} \subset D.$$

ただしここで,

$$(3.9) \quad \tilde{x}^{(1)} = \tilde{f}[x^{(0)}],$$

$$(3.10) \quad \rho_1 = K_0(E - K_0)^{-1} \sigma[\tilde{x}^{(1)} - x^{(0)}] + (E - K_0)^{-1} \varepsilon (E - \varepsilon)^{-1} \sigma[\tilde{x}^{(0)}],$$

$$(3.11) \quad \rho_2 = (E - L_0)^{-1} \varepsilon \{ \rho_1 + \sigma[\tilde{x}^{(0)}] \}$$

で,

$$(3.12) \quad L_0 = (E + \varepsilon) K_0$$

である.

このとき, 方程式 (2.7) は D において一つただ一つの解 $\hat{x}$ をもち, $\sigma[\tilde{x}^{(n)}]$ は有限回の反復のうち有限個の値をとって振動し, 振動状態に達した $\tilde{x}^{(n)}$ に対しては

$$(3.13) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq (E - L_0)^{-1} \varepsilon \cdot \sigma(\hat{x})$$

が成り立つ."

〔証明〕 初めに, 次の不等式が成り立つことを示しておく:

$$(3.14) \quad \sigma[\tilde{x}^{(1)} - x^{(0)}] \leq \varepsilon \cdot \sigma[x^{(0)}] \leq \varepsilon (E - \varepsilon)^{-1} \sigma[\tilde{x}^{(0)}].$$

(3.1) により

$$(3.15) \quad \sigma[\tilde{x}^{(1)} - x^{(1)}] = \sigma[\tilde{f}(x^{(0)}) - f(x^{(0)})] \\ \leq \varepsilon \cdot \sigma[f(x^{(0)})] = \varepsilon \cdot \sigma[x^{(0)}]$$

であるから, (3.14) の最初の不等式が成立する. 次に $\sigma[x^{(1)}]$ を考えると,

$$\sigma[x^{(1)}] \leq \sigma[x^{(0)} - \tilde{x}^{(0)}] + \sigma[\tilde{x}^{(0)}]$$

であるから, (3.15) より

$$\sigma[x^{(1)}] \leq \varepsilon \cdot \sigma[x^{(0)}] + \sigma[\tilde{x}^{(0)}],$$

すなわち,

$$(E - \varepsilon) \sigma[x^{(1)}] \leq \sigma[\tilde{x}^{(0)}]$$

を得る. (3.3) により $\rho(\varepsilon) < 1$ であるから, 補題3を用いて

$$(3.16) \quad \sigma[x^{(1)}] \leq (E - \varepsilon)^{-1} \sigma[\tilde{x}^{(0)}]$$

を得る. すると, (3.15) より (3.14) の第2の不等式が成立する. これで (3.14) は証明された.

われわれは, まず, (3.14) を利用し, 定理1で定義され

に S に対して

$$(3.17) \quad S \subset \Sigma_0 = \{h \mid \sigma[h - \tilde{x}^{(1)}] \leq \rho_1\}$$

が成り立つことを示そう。 S に属する任意の h に対しては、

$$\sigma[h - x^{(1)}] \leq K_0 (E - K_0)^{-1} \sigma[x^{(1)} - x^{(0)}]$$

が成り立つから、(3.14), (3.10) により逐次つぎの不等式を得る：

$$\begin{aligned} \sigma[h - \tilde{x}^{(1)}] &\leq \sigma[h - x^{(1)}] + \sigma[x^{(1)} - \tilde{x}^{(1)}] \\ &\leq K_0 (E - K_0)^{-1} \sigma[x^{(1)} - x^{(0)}] + \sigma[x^{(1)} - \tilde{x}^{(1)}] \\ &\leq K_0 (E - K_0)^{-1} \sigma[x^{(1)} - \tilde{x}^{(1)}] + K_0 (E - K_0)^{-1} \sigma[\tilde{x}^{(1)} - x^{(0)}] \\ &\quad + \sigma[x^{(1)} - \tilde{x}^{(1)}] \\ &= K_0 (E - K_0)^{-1} \sigma[\tilde{x}^{(1)} - x^{(0)}] + (E - K_0)^{-1} \sigma[x^{(1)} - \tilde{x}^{(1)}] \\ &\leq K_0 (E - K_0)^{-1} \sigma[\tilde{x}^{(1)} - x^{(0)}] + (E - K_0)^{-1} \varepsilon (E - \varepsilon)^{-1} \sigma[\tilde{x}^{(1)}] \\ &= \rho_1. \end{aligned}$$

これは $h \in \Sigma_0$ を意味するから、われわれは (3.17) を得る。

(3.17) が成り立つのは、われわれの仮定 (3.8) により

$$(3.18) \quad S \subset \Sigma_0 \subset \Sigma \subset D$$

が成り立つから、定理 1 の条件はすべて満たされることになり、定理 1 の結果がすべて成り立ち、方程式 (2.7) は D において一つただ一つの解 x をもつことになる。

さて、数値計算による反復演算 (3.5) により、 Σ の中に無限点列 $\{\tilde{x}^{(n)}\}$ が得られることを示そう。

$\tilde{x}^{(1)} \in \Sigma$ は明らかであるから、 $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots, \tilde{x}^{(n)}$ が Σ の中に得られた、とする。すると、 $\tilde{x}^{(n)} \in \Sigma \subset D$ であるから、 $\tilde{x}^{(n+1)} = \tilde{f}[\tilde{x}^{(n)}]$ によって $\tilde{x}^{(n+1)}$ が得られる。問題は $\tilde{x}^{(n+1)}$ が Σ に属するか、というところである。

さて、(3.1), (2.1), (3.12), (3.14) により

$$\begin{aligned}
 (3.19) \quad \sigma[\tilde{x}^{(2)} - x^{(2)}] &= \sigma[\tilde{f}(\tilde{x}^{(1)}) - f(x^{(1)})] \\
 &\leq \sigma[\tilde{f}(\tilde{x}^{(1)}) - f(\tilde{x}^{(1)})] + \sigma[f(\tilde{x}^{(1)}) - f(x^{(1)})] \\
 &\leq \varepsilon \cdot \sigma[f(\tilde{x}^{(1)})] + K_0 \sigma[\tilde{x}^{(1)} - x^{(1)}] \\
 &\leq \varepsilon \{ \sigma[f(\tilde{x}^{(1)}) - f(x^{(1)})] + \sigma[f(x^{(1)})] \} \\
 &\quad + K_0 \sigma[\tilde{x}^{(1)} - x^{(1)}] \\
 &\leq (E + \varepsilon) K_0 \sigma[\tilde{x}^{(1)} - x^{(1)}] + \varepsilon \sigma[x^{(2)}] \\
 &\leq L_0 \varepsilon \sigma[x^{(1)}] + \varepsilon \sigma[x^{(2)}]
 \end{aligned}$$

を得る。同様に

$$\sigma[\tilde{x}^{(3)} - x^{(3)}] \leq L_0 \sigma[\tilde{x}^{(2)} - x^{(2)}] + \varepsilon \sigma[x^{(3)}]$$

を得, (3.19) より

$$\sigma[\tilde{x}^{(3)} - x^{(3)}] \leq L_0^2 \varepsilon \sigma[x^{(1)}] + L_0 \varepsilon \sigma[x^{(2)}] + \varepsilon \sigma[x^{(3)}]$$

を得る. これを続けてゆくと,

$$(3.20) \quad \sigma[\tilde{x}^{(n+1)} - x^{(n+1)}] \leq L_0^n \varepsilon \sigma[x^{(1)}] + L_0^{n-1} \varepsilon \sigma[x^{(2)}] \\ + \dots + \varepsilon \sigma[x^{(n+1)}]$$

が得られる. ここで, 定理 1 により $x^{(n)} \in S$ ($n=1, 2, \dots$)

であるから,

$$\sigma[x^{(n)} - x^{(1)}] \leq K_0 (E - K_0)^{-1} \sigma[x^{(1)} - x^{(0)}] \quad (n=1, 2, \dots)$$

が成り立つ. すると,

$$\begin{aligned} \sigma[x^{(n)}] &\leq \sigma[x^{(n)} - x^{(1)}] + \sigma[x^{(1)}] \\ &\leq K_0 (E - K_0)^{-1} \sigma[x^{(1)} - x^{(0)}] + \sigma[x^{(1)}] \\ &\leq K_0 (E - K_0)^{-1} \sigma[x^{(1)} - \tilde{x}^{(1)}] + K_0 (E - K_0)^{-1} \sigma[\tilde{x}^{(1)} - x^{(0)}] \\ &\quad + \sigma[x^{(1)} - \tilde{x}^{(1)}] + \sigma[\tilde{x}^{(1)}] \\ &= (E - K_0)^{-1} \sigma[x^{(1)} - \tilde{x}^{(1)}] + K_0 (E - K_0)^{-1} \sigma[\tilde{x}^{(1)} - x^{(0)}] \\ &\quad + \sigma[\tilde{x}^{(1)}] \end{aligned}$$

となるから, (3.14) により,

$$\begin{aligned}
\sigma[x^{(n)}] &\leq K_0(E-K_0)^{-1}\sigma[\tilde{x}^{(0)}-x^{(0)}] + (E-K_0)^{-1}\varepsilon(E-\varepsilon)^{-1}\sigma[\tilde{x}^{(0)}] \\
&\quad + \sigma[\tilde{x}^{(0)}] \\
&= \rho_1 + \sigma[\tilde{x}^{(0)}]
\end{aligned}$$

を得る。かく、(3.20) よりわれわれは次の不等式を得る:

$$(3.21) \quad \sigma[\tilde{x}^{(n+1)}-x^{(n+1)}] \leq (L_0^n + L_0^{n-1} + \dots + E)\varepsilon \{\rho_1 + \sigma[\tilde{x}^{(0)}]\}.$$

ところで、(3.4), (3.6), (3.12) により

$$(3.22) \quad L_0 = (E + \varepsilon)K_0 \leq (1 + \varepsilon^1)K_0 = \hat{L}_0$$

である。(3.7) から、補題2により、 $\lim_{n \rightarrow \infty} \hat{L}_0^n = 0$ である。すると、(3.22) により

$$0 \leq L_0 \leq \hat{L}_0$$

であるから、 $\lim_{n \rightarrow \infty} L_0^n = 0$ となり、補題2によって

$$(3.23) \quad \rho(L_0) < 1$$

となる。すると、補題3により

$$\begin{aligned}
(3.24) \quad E + L_0 + L_0^2 + \dots + L_0^n &\leq E + L_0 + L_0^2 + \dots \\
&= (E - L_0)^{-1}
\end{aligned}$$

となるので, (3.21) より (3.11) により

$$(3.25) \quad \sigma[\tilde{x}^{(n+1)} - x^{(n+1)}] \leq \rho_2$$

を得る. $x^{(n+1)} \in S \subset \Sigma_0$ であるから, (3.25) より

$$\sigma[\tilde{x}^{(n+1)} - \tilde{x}^{(1)}] \leq \rho_1 + \rho_2$$

を得る. これは Σ の定義により $\tilde{x}^{(n+1)} \in \Sigma$ であることを示している.

上の結果から, 帰納法により, (3.5) によって Σ の中に無限点列 $\{\tilde{x}^{(n)}\}$ が得られることがわかる.

さて, Σ は d -次元ユークリッド空間の有界な閉領域で, $\tilde{x}^{(n)}$ の成分 $\tilde{x}_i^{(n)}$ ($i=1, 2, \dots, d$) は計算機によって得られる有限桁, 有限指数の数である. したがって, $\tilde{x}^{(n)}$ がこゝとこく異なる値をとる, ということはできない. すなわち, 必ず, ある正の整数 p に対して

$$(3.26) \quad \tilde{x}^{(m+p)} = \tilde{x}^{(m)}$$

が成り立つような m が存在することになる. 一旦 (3.26) が起ると,

$$\tilde{x}^{(m+p+r)} = \tilde{x}^{(m+r)} \quad (r=1, 2, \dots)$$

となるので, $\{\tilde{x}^{(n)}\}$ は $\tilde{x}^{(m)}$ 以後は p 個の値

$$\tilde{x}^{(m)}, \tilde{x}^{(m+1)}, \dots, \tilde{x}^{(m+p-1)}$$

をとって振動するようになる.

いま, $\tilde{x}^{(m)}$ を振動状態にある任意の $\tilde{x}^{(n)}$ の値としよう.

(3.20) の証明からわかるように, 任意の $\tilde{x}^{(n)}$ に対しては

$$(3.27) \quad \sigma[\tilde{x}^{(n)} - x^{(n)}] \leq L_0^{n-1} \varepsilon \sigma[x^{(1)}] + L_0^{n-2} \varepsilon \sigma[x^{(2)}] + \dots + \varepsilon \sigma[x^{(n)}]$$

が成り立つ. ここで, 定理 1 の証明中の (2.8) において,
 $r \rightarrow \infty$ とすると, (2.9) により

$$(3.28) \quad \sigma[\hat{x} - x^{(n)}] \leq K_0^n (E - K_0)^{-1} \sigma[x^{(1)} - x^{(0)}]$$

を得る. したがって

$$\sigma[x^{(n)}] \leq \sigma[x^{(n)} - \hat{x}] + \sigma[\hat{x}]$$

より

$$(3.29) \quad \sigma[x^{(n)}] \leq K_0^n (E - K_0)^{-1} \sigma[x^{(1)} - x^{(0)}] + \sigma[\hat{x}]$$

を得る. すると, (3.27) より次の不等式が得られる:

$$\begin{aligned}
(3.30) \quad \sigma[\tilde{x}^{(n)} - x^{(n)}] &\leq L_o^{n-1} \varepsilon K_o (E - K_o)^{-1} \sigma[x^{(1)} - x^{(0)}] + L_o^{n-1} \varepsilon \sigma[\hat{x}] \\
&\quad + L_o^{n-2} \varepsilon K_o^2 (E - K_o)^{-1} \sigma[x^{(1)} - x^{(0)}] + L_o^{n-2} \varepsilon \sigma[\hat{x}] \\
&\quad \dots \dots \dots \\
&\quad + \varepsilon K_o^n (E - K_o)^{-1} \sigma[x^{(1)} - x^{(0)}] + \varepsilon \sigma[\hat{x}] \\
&= (L_o^{n-1} \varepsilon K_o + L_o^{n-2} \varepsilon K_o^2 + \dots + \varepsilon K_o^n) (E - K_o)^{-1} \sigma[x^{(1)} - x^{(0)}] \\
&\quad + (L_o^{n-1} + L_o^{n-2} + \dots + E) \varepsilon \sigma[\hat{x}].
\end{aligned}$$

さて, (3.4), (3.6), (3.12) により次の式が成り立つ:

$$\begin{aligned}
(3.31) \quad L_o^{n-1} \varepsilon K_o + L_o^{n-2} \varepsilon K_o^2 + \dots + \varepsilon K_o^n \\
\leq \hat{\varepsilon} (\hat{L}_o^{n-1} K_o + \hat{L}_o^{n-2} K_o^2 + \dots + K_o^n) \\
= \hat{\varepsilon} (\hat{L}_o^{n-1} + \hat{L}_o^{n-2} K_o + \dots + K_o^{n-1}) K_o.
\end{aligned}$$

ところで, (3.6) により $\hat{L}_o K_o = K_o \hat{L}_o$ であるから,

$$(\hat{L}_o - K_o)(\hat{L}_o^{n-1} + \hat{L}_o^{n-2} K_o + \dots + K_o^{n-1}) = \hat{L}_o^n - K_o^n,$$

すなわち,

$$\hat{\varepsilon} K_o (\hat{L}_o^{n-1} + \hat{L}_o^{n-2} K_o + \dots + K_o^{n-1}) = \hat{L}_o^n - K_o^n$$

を得る。すると, (3.31) より

$$(3.32) \quad 0 \leq L_o^{n-1} \varepsilon K_o + L_o^{n-2} \varepsilon K_o^2 + \dots + \varepsilon K_o^n \leq \hat{L}_o^n - K_o^n$$

を得る。ところで, (2.2), (3.7) により

$$\rho(K_0), \rho(\hat{L}_0) < 1$$

であるから, 補題2により $n \rightarrow \infty$ のとき $\hat{L}_0^n - K_0^n \rightarrow 0$ となる。
かくて (3.32) より

$$(3.33) \quad \lim_{n \rightarrow \infty} (L_0^{n-1} \varepsilon K_0 + L_0^{n-2} \varepsilon K_0^2 + \dots + \varepsilon K_0^n) = 0$$

が得られる。

さて, 振動状態にある任意の $\tilde{\chi}^{(n)}$ に対しては, ある正の整数 p に対して $\tilde{\chi}^{(n+p)} = \tilde{\chi}^{(n)}$ が成り立つから, いまの $\tilde{\chi}^{(m)}$ に対しては $\tilde{\chi}^{(m+p)} = \tilde{\chi}^{(m)}$ が成り立つとする。すると, 明らかに任意の正の整数 r に対して

$$(3.34) \quad \tilde{\chi}^{(m+rp)} = \tilde{\chi}^{(m)}$$

が成り立つことになる。そこで (3.30) において

$$n = m + rp$$

と置き, $r \rightarrow \infty$ とすると, (3.34), (2.9), (3.24), (3.33) により

$$(3.35) \quad \sigma[\tilde{\chi}^{(m)} - \hat{\chi}] \leq (E - L_0)^{-1} \varepsilon \sigma[\hat{\chi}]$$

を得る。これは望む不等式'(3.13)である。

これで定理は完全に証明されたことに至る。

〔注意1〕 方程式'(2.7)の解 $\hat{x}$ に対して、 $\hat{x}$ のある成分 $\hat{x}_i$ が0であるとせば、 $f_i(\hat{x}) = \hat{x}_i = 0$ となり、一般には $\tilde{f}_i(\hat{x}) \neq 0$ であるから、(3.1)は成立しない。このようなことが予知、あるいは予想されるような場合には、もとの方程式で x_i を適当な量だけずらしておいて解くようにすることが望ましい、と考えられる。

〔注意2〕 条件(3.8)は、定理1の条件(2.3)と同じく、 $x^{(0)}$ が適当な方程式'(2.7)の近似解であることを要請しているものである。なぜなら、 $x^{(0)}$ が D の内部にあり、相当の精度をもつ方程式'(2.7)の近似解であり、 $f(x)$ の計算は相当精密に行われて $\hat{\varepsilon}$ が小さくすると、(3.9)により $\tilde{x}^{(1)}$ は $x^{(0)}$ に近くなり、(3.8)、(3.10)、(3.11)により Σ は $\tilde{x}^{(1)}$ を中心とする小さな長方形領域になって、条件(3.8)はあつからみたされるようになるからである。

〔注意3〕 K_0 の各要素が小さくて $K_0 \approx 0$ である場合には、(3.12)により $L_0 \approx 0$ となり、(3.13)は

$$\sigma[\tilde{x}^{(n)} - \hat{x}] \leq \varepsilon \cdot \sigma(\hat{x}),$$

すなわち,

$$|\tilde{x}_i^{(n)} - \hat{x}_i| \leq \varepsilon_i |\hat{x}_i| \quad (i=1, 2, \dots, d)$$

となる。これは $\tilde{x}^{(n)}$ の各成分 $\tilde{x}_i^{(n)}$ の相対誤差がそれぞれ ε_i を越えないことを示している。このような誤差評価は実際問題ではきわめて便利であり、またノルムを用いての評価より一般にはるかに精密になっている。

〔注意4〕 $\sigma(v)$ をベクトルのノルムとすると、 K_0, ε はともにスカラーになり、前に述べたように (2.2) は

$$0 \leq K_0 < 1$$

を意味し、(3.3), (3.4) は

$$0 \leq \varepsilon \leq \hat{\varepsilon} < 1$$

を意味し、(3.7) は

$$(1 + \hat{\varepsilon}) K_0 < 1$$

を意味する。このとき、定理2の結果が、定理1の場合と同じく、そのまま成立することは、上の証明から明らかである。

3.2 反復計算の停止

反復計算 (0.3) を計算機⁹上で実行し, (3.5) により定まる $\tilde{x}^{(n)}$ ($n=1, 2, \dots$) を求めてゆくとき, $\tilde{x}^{(n)}$ が振動状態に達したかどうかを直接確かめるためには, 反復計算で得られる $\tilde{x}^{(n)}$ をすべて記憶装置にとっておかなければならない。これは実際計算にはあまり便利ではない。そこで, 次の不等式を考えて, これがみたされると, そこで反復計算を停止する, という方法がよく行われる:

$$(3.36) \quad \sigma[\tilde{x}^{(n+1)} - \tilde{x}^{(n)}] \leq \alpha \cdot \sigma[\tilde{x}^{(n)}].$$

ただしここで, α は $\alpha \geq 0$ である適当な行列である。

α の各要素を無暗に小さくすると, 上の方法は有効に働かない。いいかえると, 不等式 (3.36) は決してみたされず, したがって反復計算は停止をいなりことになる。上の方法が有効に働くためには, α をどのように選ぶべきか。次の定理はこれに答えるものである。

定理 3. "定理 2 にあって

$$(3.37) \quad \delta_0 = (E - L_0)^{-1} \varepsilon$$

と置く。 $\hat{\varepsilon}$ が小さくて

$$(3.38) \quad \rho(\delta_0) < 1$$

が成り立つときは, 行列 $\alpha \geq 0$ を

$$(3.39) \quad \alpha \geq 2\delta_0(E - \delta_0)^{-1}$$

が成り立つように選ぶならば, (3.36) は有限の n に達して必ず成り立つ, いいかえると, (3.36) がみたされるところで反復計算を停止するという方法は, 実際には有効に働く."

[証明] いま, $\tilde{x}^{(n)}$ が振動状態にあるとすると, $\tilde{x}^{(n+1)}$ も振動状態にあるから, 定理2の (3.13) により (3.37) を用いて

$$\sigma[\tilde{x}^{(n)} - \hat{x}], \sigma[\tilde{x}^{(n+1)} - \hat{x}] \leq \delta_0 \cdot \sigma(\hat{x})$$

を得る. したがって

$$(3.40) \quad \sigma[\tilde{x}^{(n+1)} - \tilde{x}^{(n)}] \leq 2\delta_0 \cdot \sigma(\hat{x})$$

を得る. ここで,

$$\begin{aligned} \sigma(\hat{x}) &\leq \sigma[\hat{x} - \tilde{x}^{(n)}] + \sigma[\tilde{x}^{(n)}] \\ &\leq \delta_0 \cdot \sigma(\hat{x}) + \sigma[\tilde{x}^{(n)}] \end{aligned}$$

であるから,

$$(E - \delta_0) \sigma(\chi^1) \leq \sigma[\tilde{\chi}^{(n)}]$$

を得, (3.38) により補題3を用いて

$$\sigma(\chi^1) \leq (E - \delta_0)^{-1} \sigma[\tilde{\chi}^{(n)}]$$

を得る. かくて (3.40) により, 振動状態にある $\tilde{\chi}^{(n)}$ に対しては

$$(3.41) \quad \sigma[\tilde{\chi}^{(n+1)} - \tilde{\chi}^{(n)}] \leq 2\delta_0 \cdot (E - \delta_0)^{-1} \sigma[\tilde{\chi}^{(n)}]$$

を得る.

いま, 行列 $\alpha \geq 0$ は (3.39) をみたすとして, (3.36) がすべての n に対して成り立たない, としてみよう. すなわち,

$$(3.42) \quad \sigma[\tilde{\chi}^{(n+1)} - \tilde{\chi}^{(n)}] > \alpha \cdot \sigma[\tilde{\chi}^{(n)}] \quad (n=0, 1, 2, \dots)$$

となる. ここで, 定理2により $\tilde{\chi}^{(n)}$ はある n に対して必ず振動状態にあるから, そのような n に対しては (3.41) が成り立つ. すなわち, (3.42) により

$$2\delta_0 (E - \delta_0)^{-1} \sigma[\tilde{\chi}^{(n)}] > \alpha \cdot \sigma[\tilde{\chi}^{(n)}]$$

を得る. これは (3.39) に矛盾する. これで定理は証明された.

反復計算を停止するために不等式'(3.36)を用いる場合、最後に得られる $\tilde{x}^{(n+1)}$ はどの程度の誤差をもっているであろうか。次の定理はこれに答えるものである。

定理4. "定理2において、ある n に対し (3.36) が成り立つとする。このとき、行列 $\alpha \geq 0$ の各要素が小さくて

$$(3.43) \quad \rho(L_0 + \alpha) < 1$$

が成り立つならば、 $\tilde{x}^{(n+1)}$ に対して次の不等式が成り立つ:

$$(3.44) \quad \sigma[\tilde{x}^{(n+1)} - \hat{x}] \leq \left\{ L_0 [E - (L_0 + \alpha)]^{-1} \alpha + (E - \alpha) [E - (L_0 + \alpha)]^{-1} \varepsilon \right\} \sigma(\hat{x}).$$

[証明] (3.36) より逐次つぎの不等式を得る:

$$\begin{aligned} \sigma[\tilde{x}^{(n)} - \hat{x}] &\leq \sigma[\tilde{x}^{(n)} - \tilde{x}^{(n+1)}] + \sigma[\tilde{f}(\tilde{x}^{(n)}) - f(\hat{x})] \\ &\leq \alpha \cdot \sigma[\tilde{x}^{(n)}] + \sigma[\tilde{f}(\tilde{x}^{(n)}) - f(\tilde{x}^{(n)})] + \sigma[f(\tilde{x}^{(n)}) - f(\hat{x})] \\ &\leq \alpha \cdot \sigma[\tilde{x}^{(n)} - \hat{x}] + \alpha \cdot \sigma(\hat{x}) + \varepsilon \cdot \sigma[f(\tilde{x}^{(n)})] + K_0 \sigma[\tilde{x}^{(n)} - \hat{x}] \\ &\leq \alpha \cdot \sigma[\tilde{x}^{(n)} - \hat{x}] + \alpha \cdot \sigma(\hat{x}) + \varepsilon \cdot \sigma[f(\tilde{x}^{(n)}) - f(\hat{x})] \\ &\quad + \varepsilon \cdot \sigma[f(\hat{x})] + K_0 \sigma[\tilde{x}^{(n)} - \hat{x}] \\ &\leq \alpha \cdot \sigma[\tilde{x}^{(n)} - \hat{x}] + \alpha \cdot \sigma(\hat{x}) + \varepsilon K_0 \sigma[\tilde{x}^{(n)} - \hat{x}] + \varepsilon \cdot \sigma(\hat{x}) \end{aligned}$$

$$\begin{aligned}
& + K_0 \sigma[\tilde{x}^{(n)} - \hat{x}] \\
& = (L_0 + \alpha) \sigma[\tilde{x}^{(n)} - \hat{x}] + (\alpha + \varepsilon) \sigma(\hat{x}).
\end{aligned}$$

これより

$$[E - (L_0 + \alpha)] \sigma[\tilde{x}^{(n)} - \hat{x}] \leq (\alpha + \varepsilon) \sigma(\hat{x})$$

と仮定し, (3.43) により補題3を用いて

$$(3.45) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq [E - (L_0 + \alpha)]^{-1} (\alpha + \varepsilon) \sigma(\hat{x})$$

を得る。すると,

$$\begin{aligned}
\sigma[\tilde{x}^{(n+1)} - \hat{x}] &= \sigma[\tilde{f}(\tilde{x}^{(n)}) - f(\hat{x})] \\
&\leq \sigma[\tilde{f}(\tilde{x}^{(n)}) - f(\tilde{x}^{(n)})] + \sigma[f(\tilde{x}^{(n)}) - f(\hat{x})] \\
&\leq \varepsilon \cdot \sigma[f(\tilde{x}^{(n)})] + K_0 \sigma[\tilde{x}^{(n)} - \hat{x}] \\
&\leq \varepsilon \cdot \sigma[f(\tilde{x}^{(n)}) - f(\hat{x})] + \varepsilon \cdot \sigma(\hat{x}) + K_0 \sigma[\tilde{x}^{(n)} - \hat{x}] \\
&\leq \varepsilon K_0 \sigma[\tilde{x}^{(n)} - \hat{x}] + \varepsilon \cdot \sigma(\hat{x}) + K_0 \sigma[\tilde{x}^{(n)} - \hat{x}] \\
&= L_0 \sigma[\tilde{x}^{(n)} - \hat{x}] + \varepsilon \cdot \sigma(\hat{x})
\end{aligned}$$

と仮定し, (3.45) により

$$(3.46) \quad \sigma[\tilde{x}^{(n+1)} - \hat{x}] \leq \left\{ \varepsilon + L_0 [E - (L_0 + \alpha)]^{-1} (\alpha + \varepsilon) \right\} \sigma(\hat{x})$$

を得る。これより (3.44) は容易にみちいかれる。これで

定理は証明された。

[注意 1] α の各要素が小さく, L_0 の要素と α の要素との積が ε の対角線要素に比べて無視できるときは, (3.44) は近似的に

$$(3.47) \quad \sigma[\tilde{x}^{(n+1)} - \hat{x}] \leq (E - L_0)^{-1} \varepsilon \cdot \sigma(\hat{x})$$

と書くことができる。(3.13) と比較すると, これは $\tilde{x}^{(n+1)}$ が振動状態にあるものと同程度の精度をもっていることを示している。

このことから, (3.39) を考慮に入れて, つぎの結論が得られる:

α としては, ε の要素が ε の対角線要素に比べては大きく, しかも L_0 の要素との積は ε の対角線要素に比べて小さくなるように選ぶことが望ましい。もし α がこのように選ばれらば, 最後に得られる $\tilde{x}^{(n+1)}$ は振動状態にあるもので, ほぼ同程度の精度をもつことになる。

[注意 2] $\sigma(\cdot)$ をベクトルのノルムとしても, 定理 3, 4 はそのまま成立する。ただしこのときは, $K_0, L_0, \varepsilon, \delta_0, \alpha$ はいずれもスカラーになる。

§ 4. Newton法の誤差解析

4.1 Newton法

与えられた方程式を

$$(4.1) \quad \varphi(x) = 0$$

とすると, Newton法とは,

$$(4.2) \quad f(x) = x - J^{-1}(x) \varphi(x)$$

として, 適当な $x^{(0)}$ から出発して

$$(4.3) \quad x^{(n+1)} = f[x^{(n)}] \quad (n=0, 1, 2, \dots)$$

により $x^{(n)}$ ($n=1, 2, \dots$) を逐次求め, 適当なところでこの反復計算を止めて, 最後に得られたものを与えられた方程式の解の近似とする方法である. ただし (4.2) で $J(x)$ は $\varphi(x)$ の x に関するヤコビ行列を表わしている. 実際計算では, しかし, 適当な行列 $H(x)$ を選んで,

$$(4.4) \quad f(x) = x - H(x) \varphi(x)$$

として (4.3) によって $x^{(n)}$ ($n=1, 2, \dots$) を計算する方法がよく用いられる. ここで, このような方法を Newton

法とよび、前者の方法を特に区別する必要がありときにはそれを本来の Newton法 とよぶことにする。

$\varphi(x)$, $H(x)$ については次のことを仮定する：

(a) $\varphi(x)$ は x -空間の凸コンパクトな集合 D において x に関し 2 回連続微分可能である，

(b) 行列 $H(x)$ は D に属する任意の x' , x'' に対して次の条件を満たしている：

$$(4.5) \quad \sigma[H(x') - H(x'')] \leq C \cdot \sigma(x' - x'').$$

ただしここで C はその成分がすべて非負である 3 次のテンソルである。

$H(x)$ の要素を $h_{ij}(x)$ ($i, j = 1, 2, \dots, d$), C の成分を c_{ijk} ($i, j, k = 1, 2, \dots, d$) とすると, (4.5) は次のことを表わしている：

$$|h_{ij}(x') - h_{ij}(x'')| \leq \sum_{k=1}^d c_{ijk} |x'_k - x''_k|.$$

ただしここで x'_k , x''_k はそれぞれベクトル x' , x'' の成分である。

D に属する任意の x' , x'' に対しては、われわれの仮定

から次の等式を得る:

$$(4.6) \quad \begin{cases} \varphi(x'') - \varphi(x') = \int_0^1 J[x' + \theta(x'' - x')] (x'' - x') d\theta, \\ J[x' + \theta(x'' - x')] - J(x') = \theta \int_0^1 J_x[x' + \theta, \theta(x'' - x')] (x'' - x') d\theta, \\ (0 \leq \theta \leq 1). \end{cases}$$

ただしここで, $J_x(x)$ は成分が $\partial^2 \varphi_i / \partial x_j \partial x_k$ ($i, j, k = 1, 2, \dots, d$) であるテンソルで, φ_i, x_j はそれぞれベクトル φ, x の成分である. (4.6) から次の式が得られる:

$$(4.7) \quad \varphi(x'') - \varphi(x') = J(x')(x'' - x') + \Psi(x', x'').$$

ただし

$$(4.8) \quad \Psi(x', x'') = \int_0^1 \int_0^1 J_x[x' + \theta_1 \theta (x'' - x')] \theta d\theta_1 d\theta (x'' - x')(x'' - x').$$

である. (4.8) から

$$(4.9) \quad \sigma[\Psi(x', x'')] \leq M_0 \sigma(x'' - x') \sigma(x'' - x')$$

をみたす定数テンソル M_0 の存在することがわかる. ただし M_0 の成分は σ_1 で非負である.

(4.7) を用いると, (4.4) で定義される $f(x)$ に対しては,

$$\begin{aligned}
f(x') - f(x'') &= x' - x'' - [H(x')\varphi(x') - H(x'')\varphi(x'')] \\
&= (x' - x'') - H(x')[\varphi(x') - \varphi(x'')] - [H(x') - H(x'')]\varphi(x'') \\
&= (x' - x'') - H(x')J(x')(x' - x'') + H(x')\bar{E}(x', x'') \\
&\quad - [H(x') - H(x'')]\varphi(x'')
\end{aligned}$$

を得る。そこで、次の不等式を得る：

$$\begin{aligned}
(4.10) \quad \sigma[f(x') - f(x'')] &\leq \sigma[E - H(x')J(x')] \sigma(x' - x'') \\
&\quad + \sigma[H(x')]M_0 \sigma(x' - x'') \sigma(x' - x'') \\
&\quad + C \sigma(x' - x'') \sigma[\varphi(x'')] \\
&= \{ \sigma[E - H(x')J(x')] + \sigma[H(x')]M_0 \sigma(x' - x'') \\
&\quad + C \sigma[\varphi(x'')] \} \sigma(x' - x'').
\end{aligned}$$

われわれは、 D に属する任意の x に対しては、

$$(4.11) \quad \begin{cases} \sigma[E - H(x)J(x)] \leq \kappa, \\ \sigma[H(x)]M_0 \leq M \end{cases}$$

が成り立つとする。ただし、 κ は $\kappa \geq 0$ である行列であり、 M は各要素が非負であるテンソルである。集合 D はコンパクトであるから、(4.11)はつねに成り立つ。(4.11)により(4.10)から次の不等式が得られる：

$$(4.12) \quad \sigma[f(x') - f(x'')] \leq K(x', x'') \sigma(x' - x'').$$

ただし

$$(4.13) \quad K(x', x'') = \kappa + M \sigma(x' - x'') + C \sigma[\varphi(x'')] \geq 0$$

である。

(4.13) で与えられる行列 $K(x', x'')$ に対して, われわれは次のことを仮定する:

(C) D に属する任意の x', x'' に対して

$$(4.14) \quad K(x', x'') \leq K_0$$

が成り立ち,

$$(4.15) \quad \rho(K_0) < 1$$

であるような行列 $K_0 \geq 0$ が存在する。

上の仮定は, D が与えられた方程式 (4.1) の解を含む小さな領域であると示は明らかにみだされている。なぜなら, このような場合には, D に属する任意の x', x'' に対して $\sigma(x' - x'') \approx 0$, $\sigma[\varphi(x'')] \approx 0$ であり, また

Newton法の場合には (4.11) の第一式をみたす K の要素は
すべて小さい, と考えてよいからである.

(c) がみたされていると, (4.12) から

$$\sigma[f(x') - f(x'')] \leq K_0 \sigma(x' - x'')$$

が, D に属する任意の x', x'' に対して成り立つことになる.
したがって定理 1, あるいは定理 2 に示された条件をみたす
 $x^{(0)}$ とすれば, それぞれ定理 1, あるいは定理 2 の結果が成
り立つことになる (ただし定理 2 の場合には ε は小さくて
(3.7) が成り立つとして).

[注意 1] (4.11) により

$$(4.16) \quad E - H(x)J(x) = v(x)$$

とおくと,

$$\sigma[v(x)] \leq K$$

を得る. ここで, (4.13), (4.14) により

$$K \leq K_0$$

であり, しかも (4.15) により $\rho(K_0) < 1$ である. す
ると,

$$\{\sigma[\nu(x)]\}^n \leq \kappa^n \leq K_0^n$$

で, 補題 2 により $\lim_{n \rightarrow \infty} K_0^n = 0$ であるから,

$$\lim_{n \rightarrow \infty} \{\sigma[\nu(x)]\}^n = 0$$

を得る. すると

$$\sigma[\nu^n(x)] \leq \{\sigma[\nu(x)]\}^n$$

であるから,

$$\lim_{n \rightarrow \infty} \sigma[\nu^n(x)] = 0$$

を得, したがって

$$\lim_{n \rightarrow \infty} \nu^n(x) = 0$$

を得る. すると, 補題 2 により

$$\rho[\nu(x)] < 1$$

を得る. すると, 補題 3 により $[E - \nu(x)]^{-1}$ が存在し, 明らかに

$$\det [E - \nu(x)] \neq 0$$

となる。これは (4.16) から

$$\det [H(x)J(x)] \neq 0,$$

すなわち,

$$\det H(x), \det J(x) \neq 0$$

を意味する。すると, $H^{-1}(x)$, $J^{-1}(x)$ が存在するから, 与えられた方程式 (4.1) と方程式

$$(4.17) \quad x = f(x) = x - H(x)\phi(x)$$

とは明らかに同値になる。

[注意2] $\sigma(\cdot)$ をベクトル, あるいは行列のノルムとしても上の議論は明らかに成り立つ。ただしこのときは, C , M_0 , κ , M , K_0 はすべてスカラーになる。

4.2 Newton法の誤差解析

(4.4) で定義された $f(x)$ に対して, その計算値を $\tilde{f}(x)$ とするとき (3.1) が成り立つとし, ε の対角線要素はすべて小さく (3.7) が成立するとする。 $x^{(0)}$ は定理2の条件 (3.8) をみたす D に属する点とする。すると, 定理2により, (3.5) により得られる列 $\{\tilde{x}^{(n)}\}$ はある

所から先きは振動状態になる。この場合、振動状態にある任意の $\tilde{x}^{(n)}$ に対しては次の定理を得る。

定理 5. "与えられた方程式 (4.1) の D の中にある解を $\hat{x}$ とするとき、Newton 法によって得られる近似解のうち振動状態にある任意のものを $\tilde{x}^{(n)}$ とすれば、

$$(4.18) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq \hat{\delta} \cdot \sigma[\hat{x}]$$

が成立する。ただしここで $\hat{\delta}$ は

$$(4.19) \quad \hat{\delta} = [E - (E + \varepsilon)\{\kappa + M \hat{\delta} \cdot \sigma(\hat{x})\}]^{-1} \varepsilon$$

によって定まる行列である。"

[証明] 定理 2 により、振動状態にある $\tilde{x}^{(n)}$ に対しては

$$(4.20) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq \delta_0 \cdot \sigma(\hat{x})$$

が成立する。ただしここで

$$(4.21) \quad \delta_0 = (E - L_0)^{-1} \varepsilon$$

である。

さて、定理 1 により $\hat{x} \in S$ であり、しかも定理 2 の証明で

示したように, (3.17) が成り立つから,

$$(4.22) \quad \sigma[\hat{x} - \tilde{x}^{(1)}] \leq \rho_1$$

が成立する。すると,

$$\sigma(\hat{x}) \leq \rho_1 + \sigma[\tilde{x}^{(1)}]$$

となるから, (4.21) により

$$\delta_0 \cdot \sigma(\hat{x}) \leq (E - L_0)^{-1} \varepsilon \{ \rho_1 + \sigma[\tilde{x}^{(1)}] \}$$

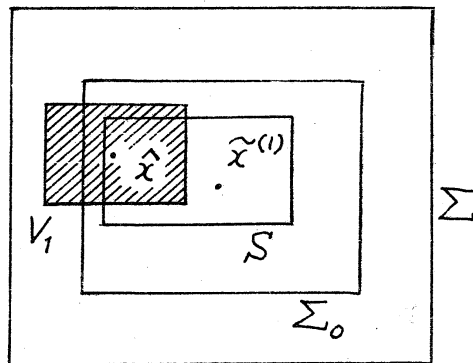
が成り立ち, (3.11) により

$$(4.23) \quad \delta_0 \cdot \sigma(\hat{x}) \leq \rho_2$$

を得る。すると, (4.22), (4.23), (3.8) により

$$(4.24) \quad V_1 = \{x \mid \sigma(x - \hat{x}) \leq \delta_0 \cdot \sigma(\hat{x})\} \subset \Sigma \subset D$$

を得る。



すると, (4.13), (4.14) により,

$$(4.25) \quad K_1 = \max_{x \in V_1} K(x, \hat{x}) = K + M \delta_0 \sigma(\hat{x}) \leq K_0$$

が成り立つ. さて, $\tilde{x}^{(n)}$ が振動状態にあれば, (4.20), (4.24) により $\tilde{x}^{(n)} \in V_1$ と仮定から, (3.1), (4.12) により逐次つぎの不等式を得る:

$$\begin{aligned} (4.26) \quad \sigma[\tilde{x}^{(n+1)} - \hat{x}] &= \sigma[\tilde{f}(\tilde{x}^{(n)}) - f(\hat{x})] \\ &\leq \sigma[\tilde{f}(\tilde{x}^{(n)}) - f(\tilde{x}^{(n)})] + \sigma[f(\tilde{x}^{(n)}) - f(\hat{x})] \\ &\leq \varepsilon \cdot \sigma[f(\tilde{x}^{(n)})] + K(\tilde{x}^{(n)}, \hat{x}) \sigma[\tilde{x}^{(n)} - \hat{x}] \\ &\leq \varepsilon \cdot \sigma[f(\tilde{x}^{(n)}) - f(\hat{x})] + \varepsilon \cdot \sigma(\hat{x}) + K(\tilde{x}^{(n)}, \hat{x}) \sigma[\tilde{x}^{(n)} - \hat{x}] \\ &\leq (E + \varepsilon) K(\tilde{x}^{(n)}, \hat{x}) \sigma[\tilde{x}^{(n)} - \hat{x}] + \varepsilon \cdot \sigma(\hat{x}) \\ &\leq (E + \varepsilon) K_1 \sigma[\tilde{x}^{(n)} - \hat{x}] + \varepsilon \cdot \sigma(\hat{x}) \\ &= L_1 \sigma[\tilde{x}^{(n)} - \hat{x}] + \varepsilon \cdot \sigma(\hat{x}). \end{aligned}$$

ただし

$$(4.27) \quad L_1 = (E + \varepsilon) K_1$$

である. (4.25) により明らかに

$$(4.28) \quad 0 \leq L_1 \leq L_0$$

であるから, (3.7) より補題 2 によって

$$(4.29) \quad \rho(L_1) < 1$$

である。さて、 $\tilde{x}^{(n)}$ が振動状態にあるとは、 $\tilde{x}^{(n+1)}$ も振動状態にあるから、(4.26) と同様にして

$$\sigma[\tilde{x}^{(n+2)} - \hat{x}] \leq L_1 \sigma[\tilde{x}^{(n+1)} - \hat{x}] + \varepsilon \cdot \sigma(\hat{x})$$

を得る。すると、(4.26) を用いて

$$\sigma[\tilde{x}^{(n+2)} - \hat{x}] \leq L_1^2 \sigma[\tilde{x}^{(n)} - \hat{x}] + (L_1 + E) \varepsilon \cdot \sigma(\hat{x})$$

を得る。この手続きを続けてゆくと、任意の正の整数 r に対して

$$(4.30) \quad \sigma[\tilde{x}^{(n+r)} - \hat{x}] \leq L_1^r \sigma[\tilde{x}^{(n)} - \hat{x}] + (E + L_1 + \dots + L_1^{r-1}) \varepsilon \cdot \sigma(\hat{x})$$

を得る。ところが、 $\tilde{x}^{(n)}$ は振動状態にあるから、ある正の整数 p に対して $\tilde{x}^{(n+p)} = \tilde{x}^{(n)}$ となる。そこで(4.30)において $r = p$ とすれば、

$$(4.31) \quad (E - L_1^p) \sigma[\tilde{x}^{(n)} - \hat{x}] \leq (E + L_1 + \dots + L_1^{p-1}) \varepsilon \cdot \sigma(\hat{x})$$

を得る。さて、

$$(E + L_1 + \dots + L_1^{p-1})(E - L_1) = E - L_1^p$$

で、(4.29) より補題3によって

$$(E - L_1)^{-1}, (E - L_1^p)^{-1} \geq 0$$

が存在するから, (4.31) より

$$(4.32) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq (E - L_1)^{-1} \varepsilon \cdot \sigma(\hat{x})$$

を得る.

$$(4.33) \quad \delta_1 = (E - L_1)^{-1} \varepsilon$$

とあけは, (4.28) より補題3によって

$$(4.34) \quad \delta_1 \leq \delta_0$$

が成立し, (4.32) より

$$(4.35) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq \delta_1 \cdot \sigma(\hat{x})$$

を得る.

上の手続きを繰り返してゆくと,

$$(4.36) \quad \begin{cases} K_r = K + M \delta_{r-1} \cdot \sigma(\hat{x}), & L_r = (E + \varepsilon) K_r, \\ \delta_r = (E - L_r)^{-1} \varepsilon \end{cases} \quad (r=1, 2, \dots)$$

で定まる行列の列 $\{K_r\}$, $\{\delta_r\}$ を得, これらは

$$(4.37) \quad \begin{cases} K_0 \geq K_1 \geq K_2 \geq \dots \geq 0, \\ \delta_0 \geq \delta_1 \geq \delta_2 \geq \dots \geq 0 \end{cases}$$

をみたし, しかもすべての r に対して

$$(4.38) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq \delta_r \cdot \sigma(\hat{x})$$

が成り立つようになる. (4.37) により

$$(4.39) \quad \hat{K} = \lim_{r \rightarrow \infty} K_r \geq 0, \quad \hat{\delta} = \lim_{r \rightarrow \infty} \delta_r \geq 0$$

が存在し, (4.38) により

$$(4.40) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq \hat{\delta} \cdot \sigma(\hat{x})$$

を得る. (4.36) より明らかに

$$(4.41) \quad \hat{K} = K + M \hat{\delta} \cdot \sigma(\hat{x}), \quad \hat{\delta} = [E - (E + \varepsilon) \hat{K}]^{-1} \varepsilon$$

が成り立つ. (4.41) の第1式を第2式に代入すれば, (4.19) の式が得られる. これで定理は証明された.

[注意 1] ε が小さいときは, (4.19) からわかるように, 行列 $\hat{\delta}$ の各要素の絶対値はすべて小さくなり, したがって,

$$(4.42) \quad \hat{\delta} \approx (E - K)^{-1} \varepsilon$$

となり, 本来の Newton 法の場合にはとくに

$$(4.43) \quad \hat{\delta} \approx \varepsilon$$

となる. (4.43) の場合には, (4.18) は

$$(4.44) \quad |\tilde{x}_i^{(n)} - \hat{x}_i| \leq \varepsilon_i |\hat{x}_i| \quad (i=1, 2, \dots, d)$$

を意味することになる. これは得られた近似解 $\tilde{x}^{(n)}$ の各成分 $\tilde{x}_i^{(n)}$ の相対誤差がそれぞれ ε_i を越えないうことを表わしている. これは実際問題ではきわめて重要なことである, と思われる. なお, 本来の Newton 法でなくても, 行列 K の各要素が十分小さいときは, (4.42) は (4.43) のようになるので, 上と同じ結論が得られる.

〔注意 2〕 $\Gamma(v)$ をベクトルのノルムとしても, 定理 5 はそのまま成り立つ. ただしこのときは, $\varepsilon, K, M, \hat{\delta}$, はすべてスカラーになる.

〔注意 3〕 ε が十分小さいときは, $\hat{\delta}$ は (4.19) によってつねに一意的に定まる.

これを示すために, (4.19) を

$$(4.45) \quad \delta = [E - (E + \varepsilon)(K + N\delta)]^{-1} \varepsilon$$

と書いて, この方程式が $\varepsilon \rightarrow 0$ のとき $\delta \cdot \varepsilon^{-1} \rightarrow (E - K)^{-1}$ となるような解を一つただ一つもつことを示そう.

(4.13), (4.14) により $K \leq K_0$ であるから, (4.15) により補題2を用いて

$$(4.46) \quad \rho(K) < 1$$

を得る. K の Jordan 標準形を K' とすると,

$$(4.47) \quad S^{-1} K S = K' = \sum \oplus \begin{bmatrix} \lambda & & 0 \\ \zeta & \lambda & \\ & \ddots & \ddots \\ 0 & & \zeta & \lambda \end{bmatrix}$$

の形の式が成り立ち, λ は K の固有値で ζ は任意の正数である. ここで ζ を十分小さくすると, $|\lambda| < 1$ であるから,

$$(4.48) \quad \|K'\| < 1$$

を得る. ただしここで, $\|\cdot\|$ は, 任意の行列 $A = (a_{ij})$

($i, j = 1, 2, \dots, d$) に対して

$$(4.49) \quad \|A\| = \max_i \sum_{j=1}^d |a_{ij}|$$

で定義される行列のノルムを表わしている. (4.47) に対応

して

$$(4.50) \quad S^{-1} \delta S = \delta', \quad S^{-1} N S = N', \quad S^{-1} \varepsilon S = \varepsilon'$$

とあくく, 方程式 (4.45) は次のようになる:

$$(4.51) \quad \delta' = [E - (E + \varepsilon')(K' + N'\delta')]^{-1} \varepsilon'.$$

さて, $\varepsilon \rightarrow 0$ のとき $\delta \cdot \varepsilon^{-1} \rightarrow (E - K)^{-1}$ となる δ に対しては,
(4.47), (4.50) より, $\varepsilon' \rightarrow 0$ のとき $\delta' \cdot \varepsilon'^{-1} \rightarrow (E - K')^{-1}$
となる. そこで

$$\delta' \cdot \varepsilon'^{-1} = (E - K')^{-1} + R'$$

とあけは, $\varepsilon' \rightarrow 0$ のとき $R' \rightarrow 0$ で, 上の式より

$$(4.52) \quad \|\delta'\| \leq [\|(E - K')^{-1}\| + \|R'\|] \|\varepsilon'\|$$

を得る. ここで, よく知られているように, (4.48) により

$$(E - K')^{-1} = \sum_{n=0}^{\infty} K'^n$$

となるから,

$$(4.53) \quad \|(E - K')^{-1}\| \leq (1 - \|K'\|)^{-1}$$

を得, (4.52) より

$$(4.54) \quad \|\delta'\| \leq [(1 - \|\kappa'\|)^{-1} + \|R'\|] \|\varepsilon'\|$$

を得る. $\|\varepsilon'\| \rightarrow 0$ のとき $\|R'\| \rightarrow 0$ であるから,

$$(4.55) \quad l_1 > (1 - \|\kappa'\|)^{-1}$$

とみたす正数 l_1 をとれば, $\|\varepsilon'\|$ が十分小さいとき, (4.54) をみたす δ' は次の不等式をみたすようになる:

$$(4.56) \quad \|\delta'\| \leq l_1 \|\varepsilon'\|.$$

以上のことから, われわれは, $\|\varepsilon'\|$ が十分小さいとき, (4.56), (4.51) をみたす δ' が一つただ一つ存在することを示せばよいことになる.

(4.51) を簡単に

$$(4.57) \quad \delta' = F(\delta')$$

と書こう. 明らかに

$$(4.58) \quad F(\delta') = [E - (E + \varepsilon')(\kappa' + N'\delta')]^{-1} \varepsilon'$$

である. (4.56) をみたす δ' の集合を Ω とすれば, Ω に属する任意の δ' に対しては

$$(4.59) \quad \|(E+\varepsilon')(K'+N'\delta')\| \leq (1+\|\varepsilon'\|)(\|K'\|+l_1\|N'\|\cdot\|\varepsilon'\|)$$

となる。したがって, (4.48) により

$$(4.60) \quad \|K'\| < l_2 < 1$$

をみたす正数 l_2 をとると, 十分小さな正数 ζ_1 をとれば $\|\varepsilon'\| < \zeta_1$ をみたす ε' に対してつねに

$$(4.61) \quad (1+\|\varepsilon'\|)(\|K'\|+l_1\|N'\|\cdot\|\varepsilon'\|) \leq l_2 < 1$$

が成立する。すると, (4.59), (4.61) により, (4.53) と同様にして

$$\begin{aligned} & \| [E - (E+\varepsilon')(K'+N'\delta')]^{-1} \| \\ & \leq [1 - (1+\|\varepsilon'\|)(\|K'\|+l_1\|N'\|\cdot\|\varepsilon'\|)]^{-1} \end{aligned}$$

が成立する。上の式の右辺は, $\|\varepsilon'\| \rightarrow 0$ のとき $(1-\|K'\|)^{-1}$ に収束するから, (4.55) により, 十分小さな正数 $\zeta_2 \leq \zeta_1$ をとれば $\|\varepsilon'\| < \zeta_2$ のときつねに

$$(4.62) \quad \| [E - (E+\varepsilon')(K'+N'\delta')]^{-1} \| \leq l_1$$

が成立する。このことは, (4.58) により, $\|\varepsilon'\| < \zeta_2$ であれば

$$(4.63) \quad F(\Omega) \subset \Omega$$

が成立することの意味している。

次に、 Ω に属する任意の δ'_1, δ'_2 をとると、(4.58) から次の等式が得られる：

$$\begin{aligned} & F(\delta'_1) - F(\delta'_2) \\ &= [E - (E + \varepsilon')(\kappa' + N'\delta'_1)]^{-1} \{ [E - (E + \varepsilon')(\kappa' + N'\delta'_2)] \\ &\quad - [E - (E + \varepsilon')(\kappa' + N'\delta'_1)] \} [E - (E + \varepsilon')(\kappa' + N'\delta'_2)]^{-1} \varepsilon' \\ &= [E - (E + \varepsilon')(\kappa' + N'\delta'_1)]^{-1} (E + \varepsilon') N' (\delta'_1 - \delta'_2) \times \\ &\quad \times [E - (E + \varepsilon')(\kappa' + N'\delta'_2)]^{-1} \varepsilon'. \end{aligned}$$

すると、 $\|\varepsilon'\| < \zeta_2$ ならば (4.62) により次の不等式が成立することになる：

$$(4.64) \quad \|F(\delta'_1) - F(\delta'_2)\| \leq \lambda_1^2 (1 + \|\varepsilon'\|) \|N'\| \cdot \|\varepsilon'\| \cdot \|\delta'_1 - \delta'_2\|.$$

さて、1 より小さい任意の正数 η をとるとき、十分小さい正数 $\zeta_3 \leq \zeta_2$ をとれば、 $\|\varepsilon'\| < \zeta_3$ のとき

$$\lambda_1^2 (1 + \|\varepsilon'\|) \|N'\| \cdot \|\varepsilon'\| \leq \eta$$

が成り立つようにすることができる。すると、(4.64) により、 $\|\varepsilon'\| < \zeta_3$ のとき、 Ω に属する任意の δ'_1, δ'_2 に

対して

$$(4.65) \quad \|F(\delta'_1) - F(\delta'_2)\| \leq k \|\delta'_1 - \delta'_2\|$$

が成立することになる。

(4.63), (4.65) から, よく知られている縮小写像の原理により, $\|\varepsilon'\| < \zeta_3$ ならば方程式 (4.57) は Ω にあって一つただ一つの解をもつことになる。これは (4.56), (4.51) をみたす δ' が, $\|\varepsilon'\|$ が十分小さければ, 一つただ一つ存在することの意味している。これによって, われわれの望む結果は証明されたことになる。

4.3 Newton 法における反復計算の停止

Newton 法の場合に, 反復計算 (3.5) を不等式 (3.36) を用いて停止することを考える。

Newton 法の場合, 振動状態にある $\tilde{x}^{(n)}$ に対しては, 定理 5 により (4.18) が成り立つので, 定理 3 と同様にして次の定理が得られる。

定理 6. "定理 5 において, $\hat{\varepsilon}$ は小さく,

$$L_0 = (E + \varepsilon)K_0, \quad \delta_0 = (E - L_0)^{-1}\varepsilon$$

とあるとき,

$$(4.66) \quad \rho(\delta_0) < 1$$

が成り立つとする. このとき, 行列 α を

$$(4.67) \quad \alpha \geq 2\hat{\delta}(E - \hat{\delta})^{-1}$$

が成り立つように選ぶならば, 不等式 (3.36) を用いて, たしかに反復計算を停止することが出来る."

定理5の証明に示されているように,

$$\delta_0 \geq \delta_1 \geq \delta_2 \geq \dots \geq 0$$

で $\lim_{r \rightarrow \infty} \delta_r = \hat{\delta}$ であるから, (4.66) により

$$\rho(\hat{\delta}) < 1$$

が成り立つ. したがって定理3と全く同様にして定理6は証明される.

Newton法において, 反復計算を不等式 (3.36) を用いて停止した場合, 最後に得られる $\tilde{x}^{(n+1)}$ はどの程度の誤差をもっているであろうか. 次の定理はこれに答えるものである.

定理 7. "定理 5 において, ある n に対して (3.36) が成り立つとする. ただしこのとき, 行列 $\alpha \geq 0$ の各要素は小さくして,

$$(4.68) \quad \rho(L_0 + \alpha) < 1$$

で,

$$(4.69) \quad \rho_3 = [E - (L_0 + \alpha)]^{-1} (\alpha + \varepsilon) \{ \rho_1 + \sigma[\tilde{x}^{(n)}] \}$$

とすると

$$(4.70) \quad \Sigma_\alpha = \{ k \mid \sigma[k - \tilde{x}^{(n)}] \leq \rho_1 + \rho_3 \} \subset D$$

が成り立つとする.

すると, $\tilde{x}^{(n+1)}$ に対して次の不等式が成り立つ:

$$(4.71) \quad \sigma[\tilde{x}^{(n+1)} - \hat{x}] \leq (\hat{\eta} - \alpha \hat{\eta} - \alpha) \sigma(\hat{x}).$$

ただしここで $\hat{\eta}$ は

$$(4.72) \quad \hat{\eta} = [E - \{ \alpha + (E + \varepsilon)(K + M \hat{\eta} \sigma(\hat{x})) \}]^{-1} (\alpha + \varepsilon)$$

によって定まる行列である. "

[証明] (3.36) が満たされているから, (3.1), (4.

12) により逐次つぎの不等式を得る:

$$\begin{aligned}
 (4.73) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] &\leq \sigma[\tilde{x}^{(n)} - \tilde{x}^{(n+1)}] + \sigma[\tilde{f}(\tilde{x}^{(n)}) - f(\tilde{x}^{(n)})] \\
 &\quad + \sigma[f(\tilde{x}^{(n)}) - f(\hat{x})] \\
 &\leq \alpha \cdot \sigma[\tilde{x}^{(n)}] + \varepsilon \cdot \sigma[f(\tilde{x}^{(n)})] + K(\tilde{x}^{(n)}, \hat{x}) \sigma[\tilde{x}^{(n)} - \hat{x}] \\
 &\leq \alpha \cdot \sigma[\tilde{x}^{(n)} - \hat{x}] + \alpha \cdot \sigma(\hat{x}) \\
 &\quad + \varepsilon \cdot \sigma[f(\tilde{x}^{(n)}) - f(\hat{x})] + \varepsilon \cdot \sigma(\hat{x}) + K(\tilde{x}^{(n)}, \hat{x}) \sigma[\tilde{x}^{(n)} - \hat{x}] \\
 &\leq [\alpha + (E + \varepsilon) K(\tilde{x}^{(n)}, \hat{x})] \sigma[\tilde{x}^{(n)} - \hat{x}] + (\alpha + \varepsilon) \sigma(\hat{x}).
 \end{aligned}$$

(4.74) により, 再び,

$$\sigma[\tilde{x}^{(n)} - \hat{x}] \leq [\alpha + (E + \varepsilon) K_0] \sigma[\tilde{x}^{(n)} - \hat{x}] + (\alpha + \varepsilon) \sigma(\hat{x}),$$

すなわち,

$$(4.74) \quad [E - (L_0 + \alpha)] \sigma[\tilde{x}^{(n)} - \hat{x}] \leq (\alpha + \varepsilon) \sigma(\hat{x})$$

を得る. (4.68) により $[E - (L_0 + \alpha)]^{-1} \geq 0$ が存在するから (補題3参照), (4.74) より

$$(4.75) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq \eta_0 \sigma(\hat{x})$$

が得られる. ただしここで

$$(4.76) \quad \eta_0 = [E - (L_0 + \alpha)]^{-1} (\alpha + \varepsilon)$$

である。

さて, $\hat{x}$ に対しては (4.22) が成り立つから, (4.76) より

$$(4.77) \quad \eta_0 \cdot \sigma(\hat{x}) \leq [E - (L_0 + \alpha)]^{-1} (\alpha + \varepsilon) \{ \rho_1 + \sigma[\tilde{x}^{(1)}] \} = \rho_3$$

が成り立つ。なお, (4.69) を (3.11) と比べると

$$\rho_2 \leq \rho_3$$

であるから, (3.17), (3.8), (4.70) より

$$(4.78) \quad S \subset \Sigma_0 \subset \Sigma \subset \Sigma_\alpha \subset D$$

が成り立つことに注意。

さて, (4.22) により

$$\sigma[\hat{x} - \tilde{x}^{(1)}] \leq \rho_1$$

であるから, (4.77), (4.70) により

$$(4.79) \quad W_1 = \{x \mid \sigma(x - \hat{x}) \leq \eta_0 \cdot \sigma(\hat{x})\} \subset \Sigma_\alpha \subset D$$

が成り立つ。すると, (4.13), (4.14) により次の式が得られる:

$$(4.80) \quad R_1 = \max_{x \in W_1} K(x, \hat{x}) = K + M \eta_0 \cdot \sigma(\hat{x}) \leq K_0.$$

すると, (4.75) により $\tilde{x}^{(n)} \in W_1$ であるから, (4.73) より

$$(4.81) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq [\alpha + (E + \varepsilon)R_1] \sigma[\tilde{x}^{(n)} - \hat{x}] + (\alpha + \varepsilon) \sigma(\hat{x})$$

を得る. このとき

$$(4.82) \quad \alpha + (E + \varepsilon)R_1 \leq \alpha + (E + \varepsilon)K_0 = \alpha + L_0$$

であるから, (4.68) により $[E - \{\alpha + (E + \varepsilon)R_1\}]^{-1} \geq 0$ が存在する (補題 3 参照). すると, (4.81) から次の不等式が得られる:

$$(4.83) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq \eta_1 \cdot \sigma(\hat{x}).$$

ただし

$$(4.84) \quad \eta_1 = [E - \{\alpha + (E + \varepsilon)R_1\}]^{-1} (\alpha + \varepsilon)$$

である. (4.82), (4.76) により

$$(4.85) \quad \eta_1 \leq \eta_0$$

であることは明らかである.

上の手続きを繰り返してゆくと,

$$(4.86) \quad \begin{cases} R_r = \kappa + M\eta_{r-1}\sigma(\hat{x}), \\ \eta_r = [E - \{\alpha + (E + \varepsilon)R_r\}]^{-1}(\alpha + \varepsilon) \end{cases} \\ (r=1, 2, \dots)$$

によって定まる行列の列 $\{R_r\}$, $\{\eta_r\}$ を得る。これらは明らかに

$$(4.87) \quad \begin{cases} K_0 \geq R_1 \geq R_2 \geq \dots \geq 0, \\ \eta_0 \geq \eta_1 \geq \eta_2 \geq \dots \geq 0 \end{cases}$$

をみたし, しかもすなわち r に対して

$$(4.88) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq \eta_r \sigma(\hat{x})$$

が成り立っている。

(4.87) により

$$(4.89) \quad \hat{R} = \lim_{r \rightarrow \infty} R_r \geq 0, \quad \hat{\eta} = \lim_{r \rightarrow \infty} \eta_r \geq 0$$

が存在し, (4.88) により

$$(4.90) \quad \sigma[\tilde{x}^{(n)} - \hat{x}] \leq \hat{\eta} \cdot \sigma(\hat{x})$$

が成り立つようになる。(4.86) より明らかに

$$(4.91) \quad \begin{cases} \hat{R} = \kappa + M\hat{\eta} \sigma(\hat{x}), \\ \hat{\eta} = [E - \{\alpha + (E+\varepsilon)\hat{R}\}]^{-1}(\alpha + \varepsilon) \end{cases}$$

が成り立つ。(4.91) の第1式を第2式'に代入すると,
(4.72) が得られる。

(4.90) が成り立つと, (4.12), (4.13) により次の不等式が得られる:

$$\begin{aligned} (4.92) \quad \sigma[\tilde{x}^{(n+1)} - \hat{x}] &= \sigma[\tilde{f}(\tilde{x}^{(n)}) - f(\hat{x})] \\ &\leq \sigma[\tilde{f}(\tilde{x}^{(n)}) - f(\tilde{x}^{(n)})] + \sigma[f(\tilde{x}^{(n)}) - f(\hat{x})] \\ &\leq \varepsilon \cdot \sigma[f(\tilde{x}^{(n)})] + K(\tilde{x}^{(n)}, \hat{x}) \sigma[\tilde{x}^{(n)} - \hat{x}] \\ &\leq \varepsilon \cdot \sigma[f(\tilde{x}^{(n)}) - f(\hat{x})] + \varepsilon \cdot \sigma(\hat{x}) \\ &\quad + K(\tilde{x}^{(n)}, \hat{x}) \sigma[\tilde{x}^{(n)} - \hat{x}] \\ &\leq (E+\varepsilon)[\kappa + M\hat{\eta} \sigma(\hat{x})] \hat{\eta} \sigma(\hat{x}) + \varepsilon \sigma(\hat{x}) \\ &= [\varepsilon + (E+\varepsilon)\{\kappa + M\hat{\eta} \sigma(\hat{x})\} \hat{\eta}] \sigma(\hat{x}). \end{aligned}$$

ところが, (4.72) より

$$[E - \{\alpha + (E+\varepsilon)(\kappa + M\hat{\eta} \sigma(\hat{x}))\}] \hat{\eta} = \alpha + \varepsilon,$$

すなわち,

$$\hat{\eta} - \alpha \hat{\eta} - (E+\varepsilon)\{\kappa + M\hat{\eta} \sigma(\hat{x})\} \hat{\eta} = \alpha + \varepsilon$$

が成り立つ。したがって、(4.92) より次の不等式が得られる：

$$\sigma[\tilde{x}^{(n+1)} - \hat{x}] \leq (\hat{\eta} - \alpha \hat{\eta} - \alpha) \sigma(\hat{x}).$$

これは望む(4.71)である。これで定理は証明された。

[注意1] (4.72) は補題3により次のように書くことができる：

$$(4.93) \quad \hat{\eta} = (\alpha + \varepsilon) + \left\{ \alpha + (E + \varepsilon)[\kappa + M \hat{\eta} \sigma(\hat{x})] \right\} (\alpha + \varepsilon) \\ + \left\{ \alpha + (E + \varepsilon)[\kappa + M \hat{\eta} \sigma(\hat{x})] \right\}^2 (\alpha + \varepsilon) + \dots$$

さて、(4.72) より、 $\kappa, \alpha, \varepsilon \rightarrow 0$ のとき

$$\sigma(\hat{\eta}) = O[\sigma(\alpha) + \sigma(\varepsilon)]$$

であるから、(4.93) より

$$(4.94) \quad \hat{\eta} = \alpha + \varepsilon + [\kappa, \alpha, \varepsilon]_2$$

を得る。ただしここで $[\kappa, \alpha, \varepsilon]_2$ は $\kappa, \alpha, \varepsilon$ の要素に関する2次以上の項の和を表わしている。(4.94) を (4.93) の右辺に代入すると次の式が得られる：

$$(4.95) \quad \hat{\eta} = \alpha + \varepsilon + \left\{ \alpha + \kappa + M(\alpha + \varepsilon)\sigma(\hat{x}) \right\} (\alpha + \varepsilon) + [\kappa, \alpha, \varepsilon]_3.$$

ただしここで $[\kappa, \alpha, \varepsilon]_3$ は $\kappa, \alpha, \varepsilon$ の要素に関する3次以上の項の和を表わしている。

対角行列 ε の対角線要素 ε_i ($i=1, 2, \dots, d$) は相当に小さいので、われわれは κ あるいは α の要素と ε_i との積は ε_i について無視できるものとしよう。すると、(4.95) からわれわれは次の式を得る：

$$(4.96) \quad \hat{\eta} \approx \alpha + \varepsilon + \alpha^2 + \kappa\alpha + M\alpha\sigma(\hat{x})\alpha.$$

すると、

$$(4.97) \quad \hat{\eta} - \alpha\hat{\eta} - \alpha \approx \varepsilon + \kappa\alpha + M\alpha\sigma(\hat{x})\alpha$$

となるので、(4.71) から近似的に次の式を得る：

$$(4.98) \quad \sigma[\hat{x}^{(n+1)} - \hat{x}] \leq [\varepsilon + \kappa\alpha + M\alpha\sigma(\hat{x})\alpha]\sigma(\hat{x}) \\ = [E + \kappa\alpha\varepsilon^{-1} + M\alpha\sigma(\hat{x})\alpha\varepsilon^{-1}]\varepsilon\sigma(\hat{x}).$$

もし、 κ の各要素の絶対値が小さくて

$$(4.99) \quad \kappa\alpha\varepsilon^{-1} \approx 0$$

であり, また α の各要素の絶対値が小さくて

$$(4.100) \quad M \alpha \sigma(\hat{x}) \alpha \varepsilon^{-1} \approx 0$$

である場合には, (4.98) より近似的に

$$(4.101) \quad \sigma[\tilde{x}^{(n+1)} - \hat{x}] \leq \varepsilon \cdot \sigma(\hat{x})$$

を得る. この場合には, (4.44) と比べてわかるように, $\tilde{x}^{(n+1)}$ は振動状態にあるものと同程度の精度をもつことになる. 本来の Newton 法の場合には, $\alpha = 0$ としてよりから, (4.99) はあつてからみたされており, 必要な条件は (4.100) のみとなる.

本来の Newton 法の場合, ある α は行列 $H(\alpha)$ が $J^{-1}(\alpha)$ に近くて (4.11) の第 1 式をみたす行列 α の各要素の絶対値が小さい場合には, 対角行列 ε の対角線要素 ε_i ($i=1, 2, \dots, d$) がすべて小さければ, 4.2 の注意 1 で述べたように (4.19) で定まる $\hat{\sigma}$ に対しては $\hat{\sigma} \approx \varepsilon$ となるので, 定理 6 の条件 (4.67) は

$$(4.102) \quad \alpha \geq 2\varepsilon$$

となる. またこの場合には条件 (4.99) は上に述べたように自動的にみたされている. かくてわれわれは次の結論を得

る.

本来の Newton 法の場合, ある ϵ は行列 $H(x)$ が $J^{-1}(x)$ に十分近い場合には, 対角行列 ϵ の対角線要素 ϵ_i ($i=1, 2, \dots, d$) がすべて小さければ, 行列 $\alpha = (\alpha_{ij})$ を

(i) $\alpha_{ii} \epsilon_i^{-1}$ ($i=1, 2, \dots, d$) が 1 に比べて大きく,

(ii) α の要素の二つの積の絶対値がすべて ϵ_i ($i=1, 2, \dots, d$) に比べて小さく,

なるように選ぶと, 不等式 (3.36) を用いて反復計算を停止することができ, しかも最後に得られる $\hat{x}^{(n+1)}$ は振動状態にあるものと同程度の精度をもち,

$$(4.103) \quad \sigma[\hat{x}^{(n+1)} - \hat{x}] \leq \epsilon \cdot \sigma(\hat{x})$$

がほぼ成立する.

[注意 2] $\sigma(v)$ をベクトルのノルムとしても本節の議論はそのまゝ成り立つ. ただしこのときは, $\epsilon, \alpha, K_0, L_0, \rho_1, \rho_2, \rho_3, R_r, \eta_r$ ($r=0, 1, 2, \dots$), $\hat{R}, \hat{\eta}$ はすべてスカラーとなり, 注意 1 の条件 (i), (ii) は次のようになる:

" α は ϵ に比べて大きく, $\epsilon^{1/2}$ に比べて小さい. "

[注意 3] K, α, ϵ の各要素の絶対値が十分小さいときは, $\hat{\eta}$ は方程式 (4.72) によって一意的に定まる.

実際, (4.72) を $\hat{\eta}$ に関する方程式'と考えるとき, $K, \alpha, \varepsilon \rightarrow 0$ ならば $\hat{\eta} = (E+R)(\alpha+\varepsilon)$ で $R \rightarrow 0$ となるような (4.72) の解 $\hat{\eta}$ はつねに一つただ一つ存在する. このことは, 方程式' (4.45) の場合と全く同じ方法で証明されるので, ここでは一切省略する.

§5. 数値例

次の方程式'を考える:

$$(5.1) \quad \begin{cases} \varphi(p, q) = -p^2 + ap + q - b = 0, \\ \psi(p, q) = pq - aq + c = 0. \end{cases}$$

この方程式'は, 3次の多項式'

$$P(x) = x^3 - ax^2 + bx - c$$

の2次因数 $x^2 - px + q$ を求めると変に現われる. すなわち,

$$P(x) = (x^2 - px + q)(x + l) - (Ax + B)$$

とあいて両辺の係数を比較すると

$$a = p - l, \quad b = q - pl - A, \quad c = B - ql$$

を得るので, これらから l を消去して

$$\begin{cases} A = -p^2 + ap + q - b, \\ B = pq - aq + c \end{cases}$$

を得る. $x^2 - px + q$ が $P(x)$ の因数になるための条件は

$$A = B = 0$$

であるから, $P(x)$ が与えられるとき, 2次因数 $x^2 - px + q$ を求めるためには, 方程式 (5.1) を解かばよいことになる.

いま,

$$P(x) = (x - \alpha)(x - \beta)(x - \gamma)$$

とすれば, 明らかに

$$(5.2) \quad a = \alpha + \beta + \gamma, \quad b = \alpha\beta + \beta\gamma + \gamma\alpha, \quad c = \alpha\beta\gamma$$

であり, また $x^2 - px + q$ が $P(x)$ の因数であるならば, それは

$$(x - \alpha)(x - \beta), \quad (x - \beta)(x - \gamma), \quad (x - \gamma)(x - \alpha)$$

のいずれかであるから, (p, q) は

$$(5.3) \quad (\alpha + \beta, \alpha\beta), \quad (\beta + \gamma, \beta\gamma), \quad (\gamma + \alpha, \gamma\alpha)$$

のいずれかである。すなわち、 a, b, c が(5.2)であることは、方程式(5.1)の解は(5.3)で与えられることになる。

いま、たとえば

$$(5.4) \quad \alpha=1, \quad \beta=10^{-3}, \quad \gamma=10$$

とすると、(5.2)により

$$(5.5) \quad a=11.001, \quad b=10.011, \quad c=0.01$$

となり、(5.3)は次のようになる：

$$(5.6) \quad (1.001, 0.001), (10.001, 0.01), (11, 10).$$

われわれは、(5.5)で与えられる a, b, c に対して、方程式(5.1)の解のうち $(1.001, 0.001)$ を本来のNewton法を用いて計算してみた。

(5.1)で与えられる $\varphi(p, q), \psi(p, q)$ の p, q に関するヤコビ行列 $J(p, q)$ は

$$(5.7) \quad J(p, q) = \begin{bmatrix} -(2p-a) & 1 \\ q & p-a \end{bmatrix}$$

となるので、 $J^{-1}(p, q)$ は次のようになる：

$$(5.8) \quad J^{-1}(p, q) = \frac{1}{D(p, q)} \begin{bmatrix} p-a & -1 \\ -q & -(2p-a) \end{bmatrix}.$$

ただし

$$(5.9) \quad \begin{aligned} D(p, q) &= \det J(p, q) \\ &= -(2p^2 - 3ap + q + a^2) \end{aligned}$$

である。したがって

$$(5.10) \quad \begin{cases} f(p, q) = p - \frac{1}{D(p, q)} [(p-a)\varphi(p, q) - \psi(p, q)], \\ g(p, q) = q - \frac{1}{D(p, q)} [-q\varphi(p, q) - (2p-a)\psi(p, q)] \end{cases}$$

とすれば、本来の Newton 法は

$$(5.11) \quad \begin{cases} p_n = f(p_{n-1}, q_{n-1}), \\ q_n = g(p_{n-1}, q_{n-1}) \end{cases}$$

によって逐次 (p_n, q_n) ($n=1, 2, \dots$) を計算することになる。

$p_0=2$, $q_0=0$ とし, 10進8桁の浮動小数点型演算で上の計算を実行した結果は次の通りである:

n	p_n	$\left \frac{p_n - p_{n-1}}{p_{n-1}} \right $	q_n	$\left \frac{q_n - q_{n-1}}{q_{n-1}} \right $
0	2.0		0.0	
1	0.85843290	0.5708	0.0011109877	
2	0.99879570	0.1635	0.0011396923	0.8974
3	1.0009995	2.206×10^{-3}	0.0010000307	0.1225
4	1.0010001	5.994×10^{-7}	0.00099999998	3.072×10^{-5}
5	1.0010000	9.990×10^{-8}	0.00099999998	0
6	1.0010000	0	0.00099999998	0

表 1

$n \geq 3$ では

$$(5.12) \quad 1 \leq p \leq 1.002, \quad 0.0009 \leq q \leq 0.0011$$

であるから、この範囲で10進8桁浮動小数点型演算で

$f(p, q)$, $g(p, q)$ を計算した場合の誤差を考える。 $f(p, q)$,

$g(p, q)$ の計算に含まれる丸め誤差を逐次評価してゆくと、

(5.12) の範囲においては最終的に次の結果が得られる:

$$(5.13) \quad \left\{ \begin{array}{l} |\tilde{f}(p, q) - f(p, q)| < 2.231 \times 10^{-7} |f(p, q)|, \\ \end{array} \right.$$

$$\left| \tilde{g}(p, q) - g(p, q) \right| < 2.477 \times 10^{-7} |g(p, q)|.$$

ただしここで $\hat{f}(p, q)$, $\hat{g}(p, q)$ はそれぞれ $f(p, q)$, $g(p, q)$ の計算によって得られる値を示している. (5.13) は (3.1) が

$$(5.14) \quad \varepsilon_1 = 2.231 \times 10^{-7}, \quad \varepsilon_2 = 2.477 \times 10^{-7}$$

に対して成り立っていることを示している.

さて, 表1からわかるように, 反復計算 (5.11) によって得られる (p_n, q_n) のうちで振動状態にあるものは

$$(5.15) \quad p_5 = 1.001, \quad q_5 = 0.000999999998$$

である. いまの場合, (p, q) の真の値 $(\hat{p}, \hat{q})$ は (5.6) により

$$(5.16) \quad \hat{p} = 1.001, \quad \hat{q} = 0.001$$

であるから, (5.15) の相対誤差は次のようになる:

$$(5.17) \quad \begin{cases} |p_5 - \hat{p}| / |\hat{p}| = 0, \\ |q_5 - \hat{q}| / |\hat{q}| = 2 \times 10^{-8}. \end{cases}$$

(5.14) により, これは定理5の (4.18), $\gamma = (4.44)$

が成り立っていることを示している。

なお, (5.14)より

$$\varepsilon_1, \varepsilon_2 < \hat{\varepsilon} = 2.5 \times 10^{-7}.$$

したがって, 不等式 (3.36) を用いて (5.11) の反復計算を停止するため, 4.3 の注意 1, 2 により,

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}$$

として, たとえば

$$(5.18) \quad \begin{cases} \alpha_{11} = \alpha_{22} = 10^{-5} < \sqrt{2.5 \times 10^{-7}} = 5 \times 10^{-4} \\ \alpha_{12} = \alpha_{21} = 0 \end{cases}$$

を選ぶと, 表 1 からわかるように反復計算は (p_5, q_5) を出して停止する. (p_5, q_5) は定常状態にある値であるから, これは 4.3 の結論が正しいことを示している。

αとして

$$(5.19) \quad \alpha_{11} = \alpha_{22} = 10^{-4}, \quad \alpha_{12} = \alpha_{21} = 0$$

を選ぶと, 表 1 からわかるように反復計算は (p_4, q_4) を出して停止する. (p_4, q_4) の相対誤差は

$$\begin{cases} |p_4 - \hat{p}|/|\hat{p}| = 10^{-7}/1.001 \approx 0.9990 \times 10^{-7}, \\ |q_4 - \hat{q}|/|\hat{q}| = 2 \times 10^{-8} \end{cases}$$

であるから、擾動状態にあるものとほとんど変わらない。

α として

$$(5.20) \quad \alpha_{11} = \alpha_{22} = 10^{-6}, 10^{-7}, \quad \alpha_{12} = \alpha_{21} = 0$$

を選ぶと、反復計算は(5.18)の場合と同じく (p_5, q_5) を出して停止する。

α として

$$\alpha_{11} = \alpha_{22} = 10^{-8}, \quad \alpha_{12} = \alpha_{21} = 0$$

を選ぶと、表1からわかるように反復計算は (p_6, q_6) を出して停止する。 (p_6, q_6) は (p_5, q_5) と同じであるから、この場合には計算を1回余計にしたことになる。

文 献

- [1] Urabe, M.: Convergence of numerical iteration in solution of equations, J. Sci. Hiroshima Univ., Ser. A, 19(1956), 479-489.
- [2] Urabe, M.: Error estimation in numerical solution of equations by iteration process, J. Sci. Hiroshima Univ., Ser. A-I, 26(1962), 77-91.