

丸め誤差を考慮したNewton法の減速について

統計数理研究所 土谷 隆 (Takashi Tsuchiya)

0. はじめに

Newton法は非線形方程式 $f(x)=0$ ($f_i(x)=0$ ($i=1, \dots, n$)) の解を $x_+ = x + \Delta x$ ($\Delta x = -J^{-1}f(x)$) という反復を繰り返して求めていく算法である。しばしば指摘されるように、この算法は解に十分近い近傍では収束が非常に速いという長所と、大域的な収束性が悪いという短所を合わせ持っている。大域的な収束性能を良くするためにさまざまな工夫が図られてきた。解 x の修正ベクトルを $\Delta x = -h \cdot J^{-1}f$ とし、適当に関数のノルム $\|f(x)\|$ を決めてこれを減速の基準とし、ノルムがある程度減るようにステップ幅 h ($0 < h \leq 1$; $h=1$ ならば普通のNewton法) を取るようにした減速Newton法はそのための工夫の一つである。しかしながら、この方法は減速に利用するノルムをどのようにとるかによって反復過程が異なったり (収束したりしなかったり)、関数計算時の丸め誤差のため実際の場面では減速がうまくいかない、などの問題点を持っている。また、ノルムによって減速を行うということは“関数値全体が減少しつつある”ということをも1つの数値によって代表させてしまっている不自然さももっている。一方、算法の挙動がノルムのとりかたに依存しないように全関数値を減らすように減速を行うと、ステップ幅は小さくなり、収束が著しく妨げられることも観察される。

ここでは、ノルムに頼る減速を離れて、関数値一つ一つがどのような挙動を示しているかをより意識した減速を考えるための一つの試みとして、方程式の解を一種の連立不等式系の解の極限とみなす減速法を提案し、これが方程式を数値的に解くという立場から見たときに自然で、しかもより細かい減速の制御を行いうることを示す。

1. 減速Newton法

非線形方程式系 $f(x)=0$ ($f_i(x)=0$ ($i=1, \dots, n$)) を解くための減速

Newton法の反復過程は、例えば以下のように表現される：

<減速Newton法の算法>

Step0) 初期値 $x^{(0)}$ を選ぶ.

Step1) 以下の反復を行う.

- 1) $\mu := 1$; $\Delta x^{(i)} := -J^{-1} f(x^{(i)})$. (J は f の Jacobi 行列.)
- 2) $y := x^{(i)} + \mu \Delta x^{(i)}$. もし, $f(y)$ が 減速条件 を満たすならば4)へ.
- 3) $\mu := \mu / 2$ として2)へ.
- 4) $x^{(i+1)} := y$ として, $x^{(i+1)}$ に対して 停止条件 のチェックを行う. $x^{(i+1)}$ が条件を満たしたら $x^{(i+1)}$ を解として終了.
 $x^{(i+1)}$ が条件を満たさなければ $i := i+1$ として1)へ戻る.

ここで, Step1)の2)の減速条件をどうするかによっていろいろなタイプの減速が考えられる.

2. ノルムによる減速

ここで反復を通じて一つのノルム ($\|f\|_w = \max_i |w_i f_i|$; $w = (w_1, \dots, w_n)^t$, w の全要素は正とする.) を固定しておき, そのノルムに関して

$$\|f(y)\|_w \leq (1 - \mu/2) \|f(x)\|_w \quad (2.1)$$

が成立することを減速条件とすると, 次の定理が成立する.

定理[1]

初期値および近似解の列が常にそのなかにあるような, 十分広い解のある近傍での $\max_{i,j,k} w_i |\partial^2 f_i / \partial x_j \partial x_k|$ の最大値を H , 上のノルムと consistent な Jacobi 行列の逆行列のノルムの最大値を K とおくと, 残差のノルム $\|f\|_w$ は少なくとも, 縮小率 $1 - \min[1/2, (4HK^2\|f\|_w)^{-1}]$ の一次収束をする.

しかしながらこのようにすることで, Newton法自身が持っていたスケーリング不変性は失われる. したがって, 方程式を解くという観点から, どのようなノルムを入れて減速を行うことが効果的か, という問題が派生するが, このような観点から (物理的あるいは数値計算的に) 意味のあるノルムを導入することは難し

く、問題のもともとのスケーリングをそのまま踏襲したごく普通のノルム

$$\|f(y)\| = \max_i |f_i(y)| \quad (2.2)$$

を使うことが多かった。以下、このノルムを<0D> (Ordinary norm の略) と呼び、添え字のないノルム記号はすべてこのノルムを表すものとする。

3. 丸め誤差による正規化ノルム

数値計算において広く受け入れられている考え方として、方程式を数値的に解くということを、 $\overline{\Delta f_i(x)}$ を x で f_i を計算したときの丸め誤差の絶対値のシャープな上界であるとして、

$$|f_i(\hat{x})| \leq \overline{\Delta f_i(\hat{x})} \quad (i=1, \dots, m) \quad (3.1)$$

を満たす $\hat{x}$ を探すことであると定義する立場がある。そのような $\hat{x}$ を求めるという観点からは、<0D> のようなノルムはその値が数値的に収束したように見えても条件(3.1)を満たす $\hat{x}$ が得られることが保証されないので好ましくない。また、減速においても丸め誤差の大きい関数の丸め誤差によって本質的な関数値の減少が見えなくなってしまう、減速が意味のないものになってしまう恐れもある。^(他の) 近似解が(3.1)を満たしていることを確実に保証するためには、計算機イプシロンを ε 、 $r = (r_1, \dots, r_m)^t$ を $r_i(x) \cdot \varepsilon \leq \overline{\Delta f_i(x)}$ となるような“ ε について正規化された $f_i(x)$ の丸め誤差評価”として、

$$\|f(\hat{x})\|_{r^{-1}(\hat{x})} = \max_i \left| \frac{f_i(\hat{x})}{r_i(\hat{x})} \right| \quad (i=1, \dots, m) \quad (3.2)$$

$$r^{-1}(\hat{x}) = (r_1^{-1}(\hat{x}), \dots, r_m^{-1}(\hat{x}))^t \quad (3.3)$$

をノルムとして採用し、このノルムが計算機イプシロンとほぼ同程度になったときに反復を停止すればよいことは明らかである。以下、このノルムを<NN> (Normalized Norm の略) と呼ぶ。近年提案されている高速微分法[2]によると非常にシャープな丸め誤差評価が得られるため、このようなノルムを実際に使うことが可能になる[3]。減速に際してもこのノルムを使えば、丸め誤差などのために無意味な減速をすることを避けられる。その際には減速の条件を

$$\|f(y)\|_{r^{-1}(\hat{x})} \leq (1 - \mu/2) \|f(x)\|_{r^{-1}(\hat{x})} \quad (3.4)$$

とすればよい。このようにすると各反復ごとのノルムの重み $r^{-1}(x)$ が異なるため、2で述べたような厳密な意味での一次収束性がいえなくなるが、経験的にはこのようにすることによって収束性能はそれほど変化しない。

いずれにせよ、ノルムが減ることはすべての関数値の絶対値の減少を意味しているわけではない。このことは“Newton法の方方向ベクトルは μ を小さく取れば必ず各関数の絶対値が減少するように選ばれている”という事実にあまりそぐわないように見える。

4. Pseudometricによる減速

一方、大域的収束性を向上させ、しかもスケーリング不変性を保つようにするための工夫として、Newton法を常微分方程式

$$\frac{dx_i}{dt} = - \sum_j J_{ij}^{-1}(x) f_j(x) \quad (4.1)$$

を離散化したものであると考えてむしろ(4.1)を積分して解を得ることもしばしば試みられている。これは減速Newton法の反復で μ を常に一定に保ち、非常に小さく取ることと相当する。しかしながら、このような算法は2次収束せず多量の計算を必要とするという欠点を持っている。そこで、スケーリング不変性を保ちつつ、しかも出来るだけステップサイズを大きくとるための方法としては“Newton法の与える方向はステップ幅を十分小さく取れば、すべての関数値の絶対値を減少させることができる”ことを利用して、次の条件

$$\text{すべての } i \text{ について, } |f_i(y)| \leq (1 - \mu/2) |f_i(x)|$$

$$\text{が成立する;} \quad (4.2)$$

が満たされるように μ を選ぶことが考えられる[4]。これは、関数のノルムを小さくする代わりに、“関数値の空間 (R^n) に自然と導入される pseudometric による関数値の半順序”を小さくするように減速することに相当する。ところが、このようにすると、特に“方程式の解ではないけれどもいくつかの関数値が非常に小さくなっているような初期値”から出発する場合、たとえ解に非常に近いような初期値から出発してもステップ幅 h が小さくなって収束が遅くなるということも観察される。

5. 領域による制御

Pseudometricによる減速で上記のようなことが起こるのは、以下の理由による。Newton法の方角 Δx へステップ幅 μ 進んだときの各関数 f_i の振舞は f のJacobi行列の逆行列 K_{ij} 、 f_i のHesse行列 $H_{ijk}(=\partial^2 f_i / \partial x_j \partial x_k)$ を用いてテーラー展開によって

$$f_i(x+\Delta x) \cong (1-\mu) f_i(x) + \frac{\mu^2}{2} \sum_{j,k,l,m} H_{ijk}(x) K_{jl}(x) K_{km}(x) f_l(x) f_m(x) \quad (5.1)$$

と近似できる。 f_i を減らすためにステップ幅をどの程度に選べばよいかは(5.1)の(第一項の大きさ) / (第二項の大きさ) で決まる。しかし、 f_i が非線形関数である以上は、超平面 $f_i(x)=0$ のごく近くでは(x が解のごく近くでない限りは) 第一項は0に近くなるのに対し、第二項は0にいくとは限らない。結局、Newton法の方角に進んで f_i を減少させようとする限りはステップ幅を非常に短くとらないといけなくなる。したがって、全関数のなかに上で述べたような意味で関数値が0に近いものが1つでもあるとpseudometricを使った減速ではその関数に合わせてステップ幅を選ばざるをえない。

上で述べたようなことが起こっているならば、関数値全てを減少させるということにあまりこだわらないでステップ幅を大きく取った方が、効率的に反復を収束させることができると考えられる。実際、ノルムを使った減速Newton法は、非常に0に近い値になっている関数値を一回大きくしたのちに収束することが多い (いくつかの関数の値は増加していてもノルム全体を減少させることはできることに注意)。このように、経験的にはNewton法をうまく働かせるためには関数値が増大することも許したほうがよい。ところがあまり自由に増大させることを許すと通常のNewton法のように発散することが多くなる。Newton法をうまく働かせるには、各関数値を減らすことよりも、むしろ近似解の列が飛び回る領域をある範囲内にとどめておくことが大事であると考えられる。そこで、“範囲を制御する” ということをより意識した減速法を以下では提案する。

6. 混合型の減速[4]

今、 f を解を含む適当な近傍で連続な関数、 $w = (w_1, \dots, w_n)^t$ を全要素が

正であるようなベクトルとし、 $S(f, w)$ を以下のような連立不等式の解集合とする：

$$|f_i(x)| \leq w_i \quad (i=1, \dots, m) \quad (6.1)$$

このときもし、 $w(t)$ が开区間 $(0, t_{\max})$ 上での実数パラメータ t から R^n の開集合への連続関数であり、

$$\lim_{t \rightarrow 0} \|w(t)\| = 0 \quad (6.2)$$

であるとする、方程式の解（一つとは限らない）の集合はすべての $S(f, w(t))$ の解集合の交わりとして表現できる。すなわち、

$$\{x \mid f(x) = 0\} = \bigcap_t S(f, w(t)) \quad (0 < t < t_{\max}) \quad (6.3)$$

与えられた初期値 $x^{(0)}$ に対して、

$$x^{(0)} \in S(f, w(t_{\max})) \quad (6.4)$$

となるように、適当に $w(t)$ を構成してやることは $x^{(0)}$ の性質がそれほど悪くなく、 f が性質の悪くない関数であれば（例えば、減速Newton法を使い得るような状況では）難しくないと考えられるので、 $x^{(0)}$ よりはじめて、少しずつ t を 0 に近づけながら $x(t) \in S(f, w)$ を求めていくという連続変形法を考えることができる。

このようなアプローチは、 $w(t)$ を問題に合わせていろいろに構成したり、反復の過程を観察しながら収束を速めるように適応させたりできる可能性があり、その分ノルムによる減速などよりもこまかい制御を行いうると思えられる。実際には適当に離散化を行って $(t_{\max} = t^{(1)} > t^{(2)} > t^{(3)} > t^{(4)} > \dots; \lim_{v \rightarrow \infty} t^{(v)} = 0)$ として、

問題 $P^{(v)}$ ：

“ $x(t^{(v)}) \in S(f, w(t^{(v)}))$ をもとにして、 $x(t^{(v+1)}) \in S(f, w(t^{(v+1)}))$ を求める”

を解く、という作業を繰り返して解に到達できると考えられる。ここで問題になるのは

- ① $w(t)$ として何を選ぶか,
- ② $t^{(v)}$ をどのように更新していくか,
- ③ $x(t^{(v)})$ から $x(t^{(v+1)})$ をどのように更新していくか
(問題 $P^{(v+1)}$ をどうやって解くか)

ということである. 以下, $x(t^{(v)})$ を $x^{(v)}$ と記すことにする. ここでは, 特に簡単な場合として

- ① $w(t) = w \cdot t$
- ② $t^{(0)} = \max \{ |f_i(x^{(0)})| / w_i \},$
 $t^{(v+1)} = \max \{ |f_i(x^{(v)})| / w_i \} \cdot \alpha \quad (0 < \alpha < 1)$

とし, $P^{(v+1)}$ を解くために以下の算法を行う場合について解析する.

<問題 $P^{(v+1)}$ の解法>

Step0) $z := x^{(v)}$ とする.

Step1) 以下の反復を行う.

- 1) $\mu := 1; \Delta z := -J^{-1}f(z).$ (J は f のJacobi行列.)
- 2) $y := z + \mu \Delta z.$

y に対して減速条件のチェックを行う.

$$\Omega^+ = \{ f_i \mid |f_i(z)| > w_i \cdot t^{(v+1)} \},$$

$$\Omega^- = \{ f_i \mid |f_i(z)| \leq w_i \cdot t^{(v+1)} \}$$

として,

(i) Ω^+ に属するすべての関数について

$$|f_i(y)| \leq (1 - \mu/2) |f_i(z)|$$

(ii) Ω^- に属するすべての関数について

$$|f_i(y)| \leq w_i \cdot t^{(v+1)}$$

が満たされるすならば4)へ.

3) $\mu := \mu/2$ として2)へ.

4) $z := y;$

z に対して停止条件のチェックを行う.

すべての関数について

$$|f_i(z)| \leq w_i \cdot t^{(v+1)}$$

が成立したら、すなわち $z \in S(f, w \cdot t^{(v)})$ が成り立てば $x^{(v+1)} = z$ として終了。 z が条件を満たさなければ1)に戻る。

この減速条件は、関数値 f_i が w_i よりも大きいものについては要素毎の減速を行い、それ以外のものはもともとの値を大きくしないようになっている。以下では、適当な仮定のもとで、上記の条件が有限回の反復で満たされることを示す。そのために関数 $g_i(x) = f_i(x)/w_i$ を考えると便利である。 g に関しては、 $t^{(v+1)} = \alpha \cdot \max |g_i(x^{(v)})| = \alpha \|g(x^{(v)})\|$ の関係があり、減速条件は

$$\Omega^+ = \{g_i \mid |g_i(z)| > t^{(v+1)}\},$$

$$\Omega^- = \{g_i \mid |g_i(z)| \leq t^{(v+1)}\}$$

として、

(i) Ω^+ に属するすべての関数について

$$|g_i(y)| \leq (1 - \mu/2) |g_i(z)|$$

(ii) Ω^- に属するすべての関数について

$$|g_i(y)| \leq t^{(v+1)}$$

と書き換えられる。また、停止条件は

$$\text{すべての関数について } |g_i(y)| \leq t^{(v+1)}$$

となる。このとき次の定理が成り立つ。

定理

$H_{ijk} = \partial^2 g_i / \partial x_j \partial x_k$, K_{ij} を g の Jacobi 行列の逆行列とする。 $S(f, w \cdot t^{(v)})$ での $\max_{i,j,k} |\partial^2 g_i / \partial x_j \partial x_k|$ の最大値を H , $\|g\|$ に consistent な K_{ij} のノルムの最大値を K とおくと、 H, K が存在するならば、問題 $P^{(v+1)}$ は必ず $(\log \alpha / \log(1 - (\alpha / 4 H K^2 \|g(x^{(v)})\|))) + 1$ 回以下の反復で解ける。

証明

μ があまり小さくならないうちに減速条件が満たされることをいえばよい。まず、 $y = z + \mu \Delta z$ とおくと、中間値の定理より適当な θ_i ($0 < \theta_i \leq 1$; θ_i は μ の関数。) に対して、

$$g_i(z + \mu \Delta z) = g_i(z) - \mu g_i(z) + \frac{\mu^2}{2} \sum_{j,k,l,m} H_{ijk}(z + \theta_i \mu \Delta z) K_{jl}(z) K_{km}(z) g_l(z) g_m(z) \quad (6.5)$$

が成り立つ。以下、 $H_{ijk}(z + \theta_i \mu \Delta z)$ のことを単に H_{ijk} と書くことにする。また、特に断わりのない限り、 $K_{ij}(z)$ 、 $g_i(z)$ をそれぞれ K_{ij} 、 g_i と表す。

(1) Ω^+ に属する関数について

$|g_i(y)| \leq (1 - \mu/2) |g_i(z)|$ が成立する μ の大きさを評価する。

(6.5)は以下のように変形できる：

$$|g_i(y)| = |(1 - \mu)g_i + \frac{\mu^2}{2} \sum_{j,k,l,m} H_{ijk} K_{jl} K_{km} g_l g_m|, \quad (6.6)$$

$$\left| \frac{g_i(y)}{g_i(z)} \right| \leq 1 - \mu + \frac{\mu^2}{2} \left| \frac{\sum_{j,k,l,m} H_{ijk} K_{jl} K_{km} g_l g_m}{g_i} \right|. \quad (6.7)$$

したがって、

$$\frac{\mu}{2} \geq \frac{\mu^2}{2} \left| \frac{\sum_{j,k,l,m} H_{ijk} K_{jl} K_{km} g_l g_m}{g_i} \right| \quad (6.8)$$

であれば、

$$\left| \frac{g_i(y)}{g_i(z)} \right| \leq \left(1 - \frac{\mu}{2}\right) \quad (6.9)$$

すなわち Ω^+ に関する条件が成立する。したがって、 μ を

$$\left| \frac{g_i}{\sum_{j,k,l,m} H_{ijk} K_{jl} K_{km} g_l g_m} \right| \quad (6.10)$$

よりも小さくとれば減速条件(i)が満たされる。今、各 g_i に対して $|g_i| > t^{(Q+1)}$

であること、そして、 $\left| \sum_{j,k,l,m} H_{ijk} K_{jl} K_{km} g_l g_m \right|$ が

$$\left| \sum_{j,k,l,m} H_{ijk} K_{jl} K_{km} g_l g_m \right| \leq HK^2 \|g\|^2 \quad (6.11)$$

と評価されることを使うと、

$$\mu \leq \frac{\alpha}{HK^2 \|g\|} \quad (6.12)$$

とすれば減速条件(i)が満たされることが分かる。

(2) Ω^- に属する関数について

すべての $g_i \in \Omega^-$ に関して、条件を満たすように μ をある程度大きくとれるこ

とを言えばよい。(6.5), (6.11)より,

$$\begin{aligned} |g_i(q)| &\leq (1-\mu)|g_i| + \frac{\mu^2}{2} HK^2 \|g\|^2 \\ &= (1-\mu)|g_i| + \frac{\mu^2}{2} HK^2 \alpha^{-2} t^{(\nu+1)^2} \quad (6.13) \end{aligned}$$

したがって, Ω^- に属する全関数について,

$$(1-\mu)|g_i| + \frac{\mu^2}{2} HK^2 \alpha^{-2} t^{(\nu+1)^2} \leq t^{(\nu+1)} \quad (6.14)$$

を満たす μ の値を評価してやる. 今, $\beta_i = |g_i| / t^{(\nu+1)} (\leq 1)$ とおき, $A = HK^2 \alpha^{-2} t^{(\nu+1)} / 2$ とおくと(6.14)は

$$(1-\mu)\beta_i t^{(\nu+1)} + A t^{(\nu+1)} \mu^2 \leq t^{(\nu+1)} \quad (6.15)$$

となる. $t^{(\nu+1)} > 0$ なので, 結局 β_i を 0 から 1 まで動かしたときに,

$$A\mu^2 - \beta_i\mu + \beta_i - 1 \leq 0 \quad (6.16)$$

を満たす μ の大きさを評価すればよい.(6.16)は

$$A\mu^2 \leq \beta_i(\mu-1) + 1 \quad (6.17)$$

と書き換えられる. A があたえられ, β_i が 0 から 1 までの間を動くときに(6.17)が常に満たされるためには, μ を $A^{-1} = 2\alpha^2 / (HK^2 t^{(\nu+1)})$ よりも小さく選べばよいことが, $0 < \mu \leq 1$, $A > 0$, μ を変数とする直線 $(\mu, \beta_i(\mu-1) + 1)$ が β_i の値にかかわらず $(1, 1)$ を通ること, などから分かる.

結局, (1), (2)で得られた結果をまとめると, 少なくとも

$$\mu \leq \frac{\alpha^2}{HK^2 t^{(\nu+1)}} = \frac{\alpha}{HK^2 \|g\|} \quad (6.18)$$

とすれば, 常に減速条件(i), (ii)が満たされることになり, Ω^+ に属する関数値が1回反復を行うごとに $(1 - \mu/2)$ の割合で減少することが分かる. 実際に反復を行うときには, μ の値は $1/2$ のべき乗の値をとることを考えても, 最悪の場合でも $\alpha / (2HK^2 \|g(z)\|)$ よりも大きい μ が選ばれ, $\|g(z)\|$ は 1

反復ごとに、 $(1 - \alpha / 4HK^2 \|g(z)\|)$ よりも大きい割合で減少する。このことと、問題 $P^{(v+1)}$ を解くためには g のノルムを $\alpha \|g(x^{(v)})\|$ まで減少させればよいことを考えて、反復の回数を評価すると、定理の結果を得る。

この算法の定性的な挙動はパラメータ α によって決定される。 α が 0 に近いと Ω^+ に入る関数が増えるため、要素ごとに減らされなくてはならない関数値の数が増える ($\alpha=0$ では少なくとも全関数が減らされるような μ が選ばれ、pseudo-metric による減速に一致する)。一方、 α が 1 に近いときには Ω^- に入る関数が多くなり、大きい値を持ついくつかの関数のみを小さくすればよくなり、ノルムによる減速に近づく。

7. 丸め誤差評価と混合型減速法

数値的に方程式を解くことを (3.1) のように定式化する立場は、以下のようにして 6 で述べた算法と自然に結び付けられる。すなわち、計算機イプシロンが ε^* の計算機で非線形方程式系 $f_i(x)=0$ ($i=1, \dots, n$) を数値的に解くことは、次の連立不等式を満たす x を求めることと同じである：

$$|f_i(x)| \leq r_i \cdot \varepsilon^* \quad (i=1, \dots, n) \quad (7.1)$$

(ここで、 r は 3 で定義されているものと同じである。) したがって、6 で述べた算法を、 $w(\varepsilon)=r \cdot \varepsilon$ とおいて適用し、 ε を徐々に小さくしていき ε^* での解を求めることが考えられる。 r としては解における丸め誤差評価を一貫して使うのが理想的であるがそれを知ることは不可能なので、実際に implement する際は各問題 $P^{(v+1)}$ ごとに変わるようにし、 $P^{(v+1)}$ に対する r としては $x^{(v)}$ での丸め誤差評価を採用する。そして $\varepsilon^{(v+1)} \geq \varepsilon^*$ であることも考慮すると、問題 $P^{(v+1)}$ は次のようになる。問題 $P^{(v+1)}$ 。

“ $x(t^{(v)}) \in S(f, r(x^{(v-1)}) \cdot \varepsilon^{(v)})$ をもとにして、 $x(t^{(v+1)}) \in S(f, r(x^{(v)}) \cdot \varepsilon^{(v+1)})$ を求める ($\varepsilon^{(v+1)} = \max_i [\max_j \{ |f_j(x^{(v)})| / r_j(x^{(v)}) \} \cdot \alpha, \varepsilon^*]$).”

8. 計算機実験[5]

種々の減速法の比較を行うため、簡単な計算機実験を行ってみた。2つの簡単

な電子回路の動作点を求める例題について、いくつかの初期値について収束の様子を観察し、減速法によって収束の様相がどのように変わるかを観察した。以下の4つの減速法および通常のNewton法<N>を比較した。

減速法<0D>:---通常ノルムによる減速

減速法<NN>:---丸め誤差による正規化ノルムによる減速

減速法<PM>:---pseudometricを減らす減速

減速法<HB>:---丸め誤差推定値を利用した混合型減速

実験は統計数理研究所のHITAC M-280H(OS:VOS3;言語:FORTRAN77;演算形式:16進切り捨て;計算機イプシロン $\varepsilon=16^{-5}$)で行った。なお、例題中のpnpトランジスタのモデルはすべて同じもので、以下のEbers-Mollモデルによってベース=エミッター電圧 V_{BE} 、ベース=コレクタ電圧 V_{CE} からベース電流 I_B 、コレクタ電流 I_C を計算している(図8.1):

$$I_B(V_{BE}, V_{CE}) = -(1 - \alpha_F) I_{ES} [\exp(-q V_{BE}/kT) - 1] \\ - (1 - \alpha_R) I_{CS} [\exp(q(V_{CE} - V_{BE})/kT) - 1],$$

$$I_C(V_{BE}, V_{CE}) = -\alpha_F I_{ES} [\exp(-q V_{BE}/kT) - 1] \\ + I_{CS} [\exp(q(V_{CE} - V_{BE})/kT) - 1],$$

$$\alpha_F = 0.98, \alpha_R = 0.5, I_{ES} = 1.0 \times 10^{-9} \text{ A}, I_{CS} = 2.0 \times 10^{-9} \text{ A}, \\ T = 300.0 \text{ K}, k = 1.380539 \times 10^{-23} \text{ J/K}, q = 1.6021 \times 10^{-19} \text{ C}.$$

<例題1> 直流増幅回路

トランジスタ1つの一番簡単な増幅回路である(図8.2)。 V_B , V_C を未知数として、<例題1>の回路の動作点を求める方程式は以下ようになる:

$$f_1 = (V_C - V_{CC})/R_L + I_C(V_B, V_C) = 0,$$

$$f_2 = (V_B - V_{CC})/R_L + I_B(V_B, V_C) = 0,$$

$$R_1 = 26 \text{ K}\Omega, R_L = 300 \Omega, V_{CC} = -3 \text{ V}.$$

領域 $(V_B, V_C) \in [-1.2, 0.4] \times [-4.0, 0.4]$ における f_1 , f_2 の等高線を図8.3に示す。この方程式の解は $(V_B, V_C) \approx (-0.39878 \text{ V}, -1.52929 \text{ V})$ となる。

<実験 1>

上記の領域の中の

$$(V_B, V_C) = (-1.2 + 0.4i, -4.0 + 1.1j) \quad (i=0, \dots, 4; j=0, \dots, 4)$$

を初期値として<N>と<OD>, <NN>, <PM>, <HB>による減速を行った時の解に至るまでの反復回数とその模様を比較した。特に, <HB>に関しては, パラメータ α によって, 収束の様子が変化すると思われるので, $\alpha = 0.01, 0.5, 0.9$ としてパラメータの違いによる収束の様子の变化を観察した。

<結果と考察>

結果は図8.4に示されている通りである。全体としては減速を行うことによって, 通常のNewton法に比べると収束域が広がっていることが観察される。ノルムによる減速では, 通常ノルム<OD>よりも, 丸め誤差による正規化ノルム<NN>が優れていることが分かる。Pseudometricによる減速を行うと, やはり, 特に $f_1=0$ の近辺で性能が悪くなることが観察される。また, 丸め誤差を用いた混合型減速法は予想通り α が小さくなるにつれてpseudometricによる減速に近づき, 性能が下がってくることも観察された。 α が1に近づくにつれて, 丸め誤差による正規化ノルム<NN>による減速に近づくことが予想されたが $\alpha=0.9$ での結果と<NN>での結果を比べてみると, その振舞いは, 少し異なっている。

<例題 2> Flip-Flop回路

トランジスタ2つのFlip-Flop回路である(図8.5)。 V_1 から V_4 を未知数として,

<例題 2>の回路の動作点を求める方程式は以下のようなになる:

$$\begin{aligned} f_1(V_1, V_2, V_3, V_4) &= (V_2 - V_{CC}) / R_L + (V_2 - V_3) / R_1 \\ &\quad + I_C(V_1, V_2) = 0, \\ f_2(V_1, V_2, V_3, V_4) &= (V_1 - V_4) / R_1' + I_B(V_1, V_2) = 0, \\ f_3(V_1, V_2, V_3, V_4) &= (V_4 - V_{CC}) / R_L' + (V_4 - V_1) / R_1' \\ &\quad + I_C(V_3, V_4) = 0, \\ f_4(V_1, V_2, V_3, V_4) &= (V_3 - V_2) / R_1 + I_B(V_3, V_4) = 0. \end{aligned}$$

$$V_{CC} = -3V, R_1 = R_1' = 10K\Omega, R_L = R_L' = 300\Omega.$$

この方程式の解は3つあり、各々

(S 1)	$V_1 = -0.4163V$	(S 2)	$V_1 = -0.1347V$	(S 3)	$V_1 = -0.3998V$
	$V_2 = -0.1347V$		$V_2 = -2.9247V$		$V_2 = -1.4398V$
	$V_3 = -0.1347V$		$V_3 = -0.4163V$		$V_3 = -0.3998V$
	$V_4 = -2.9247V$		$V_4 = -0.1347V$		$V_4 = -1.4399V$

となる.

<実験2>

例題2の方程式を解くために、(S1)を中心とする超立方体

$$\begin{aligned} (V_1, V_2, V_3, V_4) \in & [-0.4163 \times 2, 0] \\ & \times [-0.1347 \times 2, 0] \\ & \times [-0.1347 \times 2, 0] \\ & \times [-2.9247 \times 2, 0] \end{aligned}$$

の中の32個の点を乱数によって初期値として選び、各初期値に対して<N>と、<0D>、<NN>、<PN>、<HB>による減速法を行った時の解に至るまでの反復回数と収束の様子を比較した。特に<HB>に関しては、パラメータ α によって収束の様子が変化すると考えて、 $\alpha = 0.01, 0.5, 0.9$ として α の違いによる収束性の違いを観察した。

<結果と考察>

表8.1の通りである。通常のNewton法はやはり収束せず、発散しているものが多い。それに比べると、ノルムによる減速、混合型で α が0.5, 0.9のものは安定してどれかの解に収束している。また、Pseudometricによる減速や混合型で α が小さいもの($\alpha = 0.01$)は、やはり減速をかけすぎて解に到達できないものが多いということも分かる。また、<0D>による減速Newton法は、ほぼ解に到達しているが、丸め誤差のノルムによる減速よりも解の精度が若干落ちていることも観察される。実験1の場合と同様、混合型ではパラメータ α が0に近づくにつれて、pseudometricによる減速に振舞が似てくるということも確認された。

まとめ

Newton法の収束域を広げるための減速Newton法について考察、実験を試みた。領域によるNewton法の制御を考え、これが、方程式を数値的に解くという概念と自然と結び付けられることを指摘し、一番簡単な場合について収束性を示し、実

験を行った。今後の課題としては、 $w(t)$ の選び方や、 t の下げ方などについて様々な工夫を試みるということがあげられる。特に、反復の過程で $w(t)$ を解法の収束性を保証できるようなやり方でadaptiveに変えていくことで、ノルムによる減速とはまた異なるタイプの算法を構成できると考えられるので、そのような方向での研究を進めて行きたい。

謝辞

普段よりいろいろとご指導いただいている東京大学工学部伊理正夫教授、統計数理研究所田辺國士教授に感謝いたします。また、計算機実験の例題は笹山晋一氏（現 東京ガス）の東京大学工学部卒業論文より引用させていただきました。ここに感謝の意を表したいと思います。

参考文献

- [1]伊理 正夫：数値計算。朝倉書店，東京(1981)。
- [2]M. Iri: Simultaneous Computation of Functions, Partial Derivatives and Estimates of Rounding Errors --- Complexity and Practicality. Japan Journal of Applied Mathematics, Vol.1, No.2, pp.223-252(1984).
- [3]伊理 正夫，土谷 隆，星 守：偏導関数計算と丸め誤差推定の自動化の大規模非線形方程式系への応用。 情報処理, Vol.26, No.11, pp.1411-1420(1985)。
- [4]土谷 隆：高速微分法および丸め誤差推定法とその応用。東京大学大学院工学系研究科計数工学専門課程修士論文(1986)。
- [5]笹山 晋一：丸め誤差を考慮した非線形方程式の解法の研究。東京大学工学部計数工学科卒業論文(1986)。

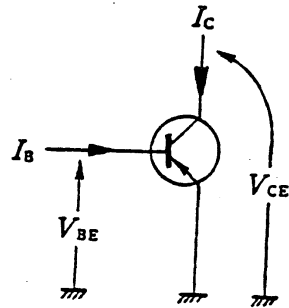


図8.1. pnp-トランジスタ.

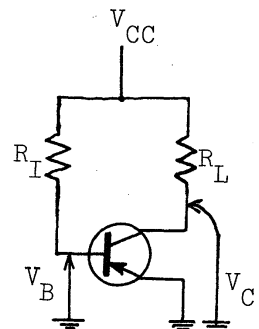
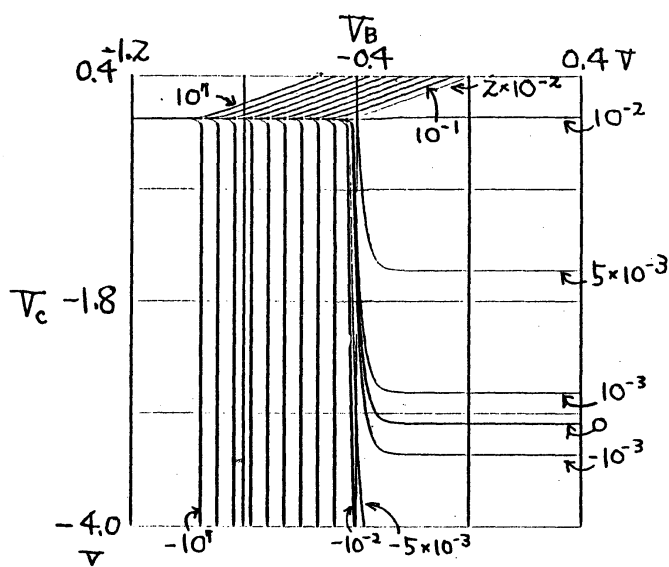
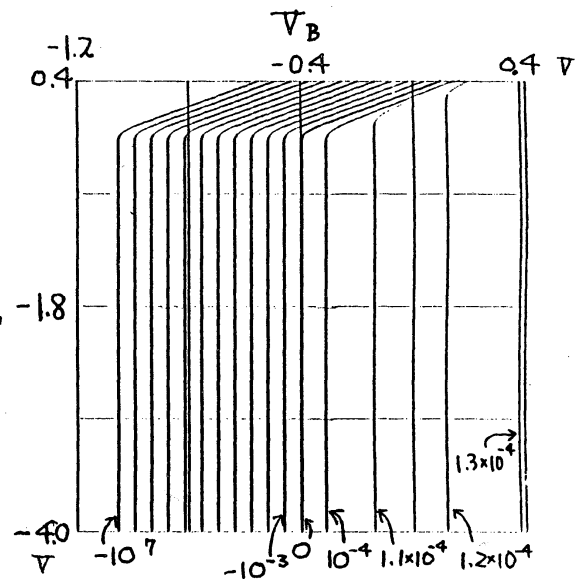


図8.2. <例題1>の回路.



(a) f_1 の等高線 (-10^{-2} から -10^7 , 10^{-1} から 10^1 までの等高線は10倍刻みで描かれている)



(b) f_2 の等高線 (-10^{-3} から -10^7 までの等高線は10倍刻みで描かれている)

図8.3 関数 f_1 と f_2 の等高線

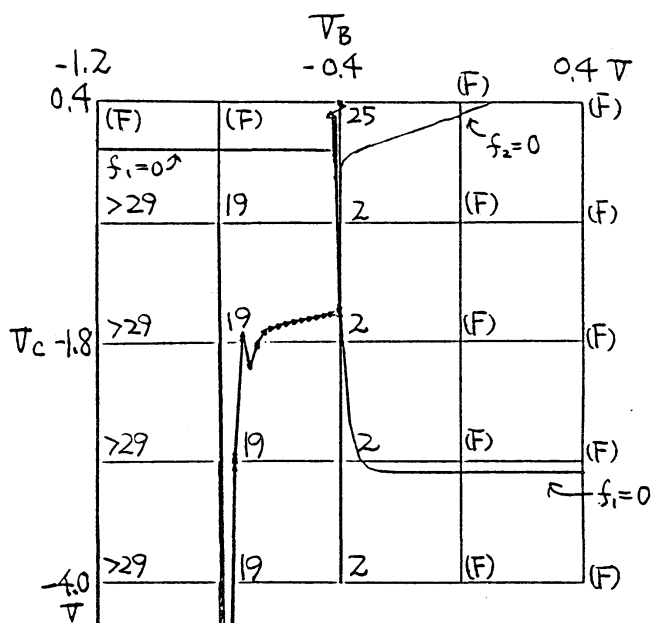
<図の見方>

- ・ 各格子点から出発した時の近似解の反復1回での動きを矢印で示してある。
- ・ 各格子点の数字は解に致るまでの反復回数である。

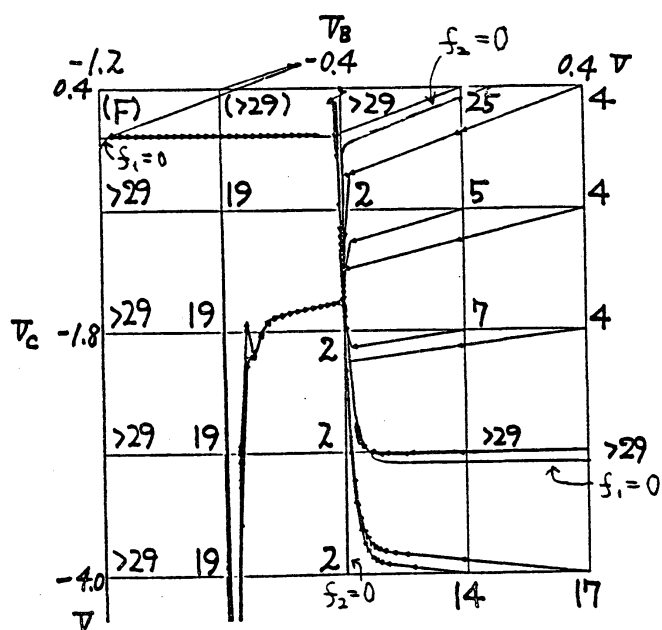
- ・ 記号: $>n$ は n 回反復しても収束しなかったことを,

(F) は関数値が大きくなりすぎて失敗したことを,

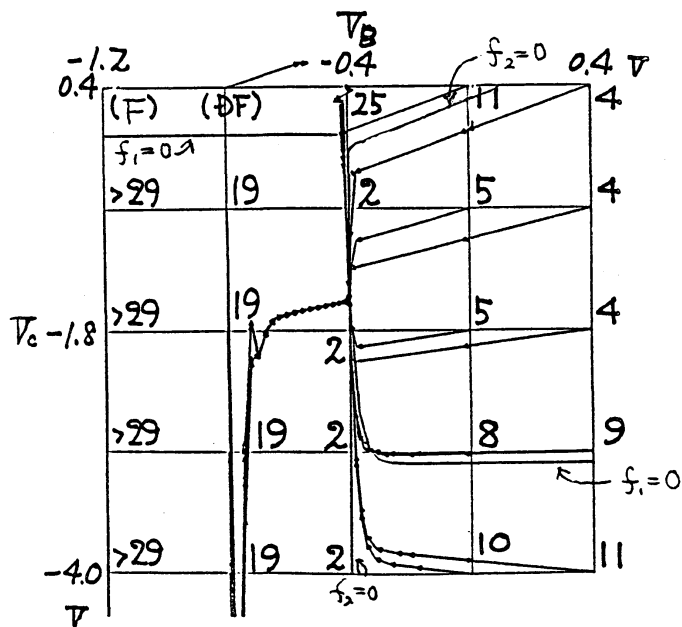
(DF) は減速に失敗したことを表す。



(a) 通常Newton法<N>

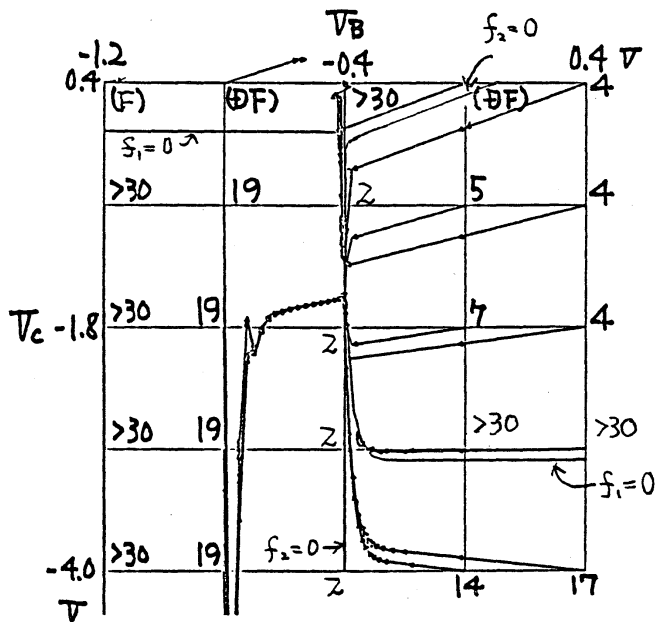


(b) 減速法<OD>



(c) 減速法<NN>

図8.4. 実験1における種々の減速法の収束の様子(続<).



(d) 減速法<PM>

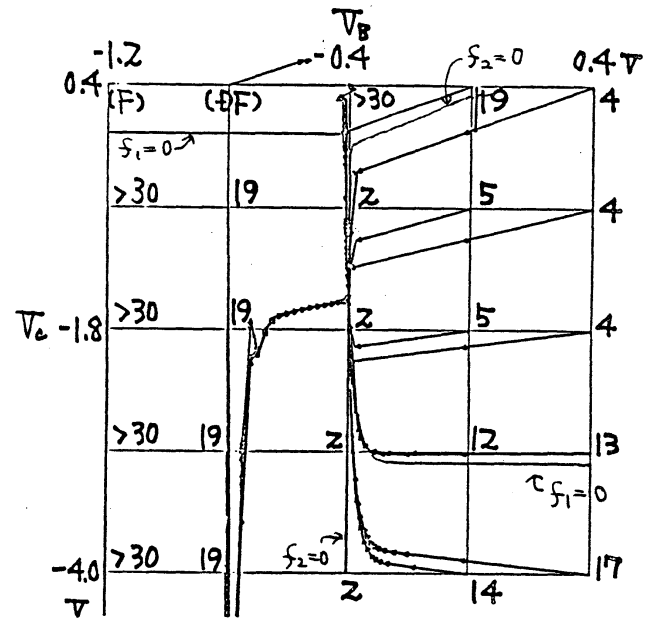
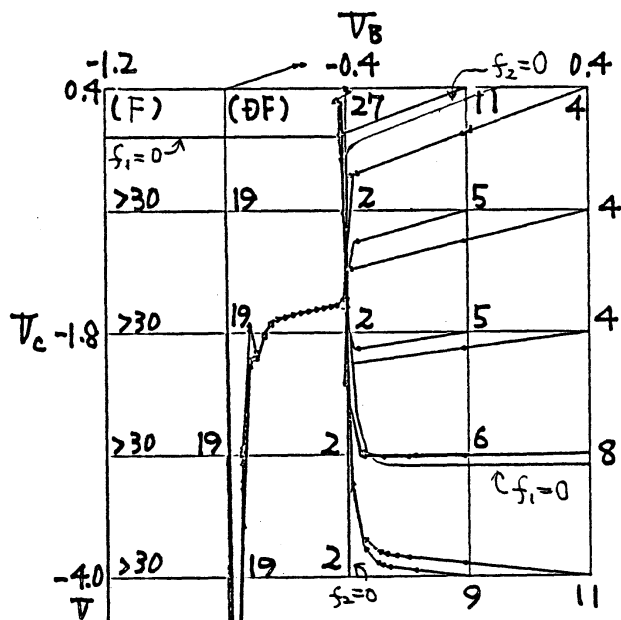
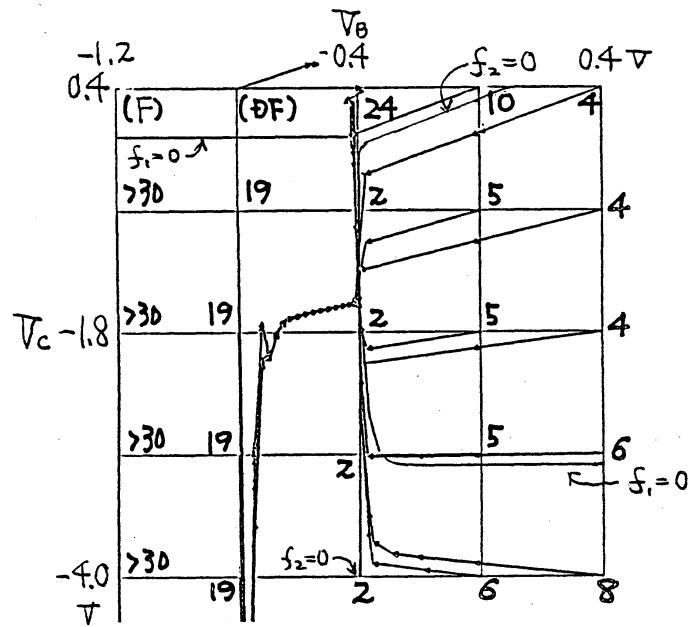
(e) 減速法<HB> ($\alpha=0.01$)(f) 減速法<HB> ($\alpha=0.5$)(g) 減速法<HB> ($\alpha=0.9$)

図8.4 実験1における種々の減速法の収束の様子(続き)

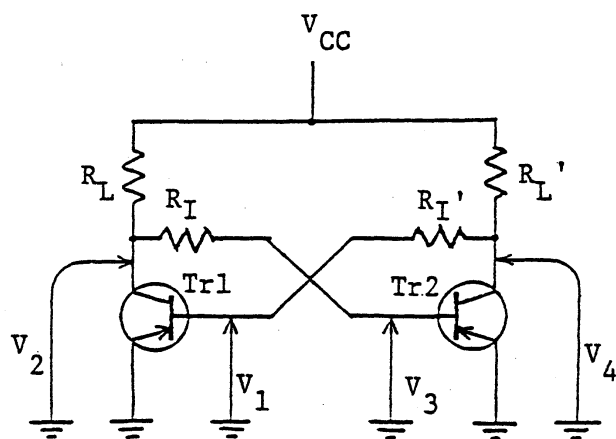


図 8.5. 例題 2 の回路.

表8.1. 実験値における種々の減速法の収束の様子(続)

表の見方

PROBLEM NO.: 問題番号;

NUM. OF ITE.: 反復回数;

NUM. OF DEC.: 減速回数 (反復のうち何回減速したか; 停止する直前の減速は数えられていない).

NUM. OF PRO. [= NUM. OF. UPDATED PROBLEM]:

減速法において作られた $P^{(n)}$ の数.

EFLG: どのようにして反復が終了したか.

0: (3.1) を満たす解が得られた.

1: 減速に失敗した (ステップ幅を小さくしても減速条件を満たせなかった.)

2: 反復を101回繰り返した収束しなかった.

NORMALIZED NORM: (3.2) で定義された NORM の最終的な値. この値が $\epsilon = 10^{-5}$ 与9.5 $\times 10^{-7}$ 以下であれば (3.1) の意味での数値解になっている.この値が 1.0×10^{-5} より大きいものに下線を引いてある. この欄の横に (S2), (S3) と書いてあるものは各々 (S2), (S3) に収束したことを表す. 残りのうち下線の引いてないものは (S1) に収束した.

(a) 通常のNewton法<N>

PROBLEM NO.	NUM. OF ITE.	NUM. OF DEC.	EFLG	NORMALIZED NORM (∞)
1	13	0	1	0.487472971E-05
2	5	0	1	0.490556067E-05
3	1	0	1	0.433588588E+36
4	16	0	1	0.500132410E-05
5	22	0	1	0.498055397E-05
6	9	0	1	0.508328139E-05
7	1	0	1	0.482310588E+14
8	16	0	1	0.498060490E-05
9	1	0	1	0.612688381E+17
10	16	0	1	0.498055397E-05
11	22	0	1	0.503227420E-05
12	14	0	1	0.496046141E-05
13	17	0	1	0.500132410E-05
14	11	0	1	0.498055397E-05
15	1	0	1	0.412245073E+16
16	1	0	1	0.412245073E+16
17	14	0	1	0.503227420E-05
18	1	0	1	0.503227420E-05
19	5	0	1	0.500132410E-05
20	5	0	1	0.496046141E-05
21	11	0	1	0.503227420E-05
22	10	0	1	0.503213232E-05
23	4	0	1	0.508378071E-05
24	4	0	1	0.213979954E+41
25	1	0	1	0.183505814E+45
26	18	0	1	0.496046324E-05
27	1	0	1	0.539430921E+44
28	1	0	1	0.330470805E+45
29	1	0	1	0.796597618E+45
30	20	0	1	0.496046141E-05
31	31	0	1	0.188314552E+15
32	1	0	1	0.255753735E+29

(b) 減速法<O>

PROBLEM NO.	NUM. OF ITE.	NUM. OF DEC.	EFLG	NORMALIZED NORM (∞)
1	12	0	1	0.130426809E-04
2	5	0	1	0.490556067E-05
3	3	0	1	0.493140817E-04 (S3)
4	16	0	1	0.500132410E-05
5	23	0	1	0.490556067E-05
6	9	0	1	0.508328139E-05
7	10	0	1	0.470713064E-06 (S3)
8	18	0	1	0.498060490E-05
9	5	2	1	0.128900401E-04
10	15	0	1	0.123718455E-04
11	14	1	1	0.503213232E-05
12	14	0	1	0.496046141E-05
13	11	0	1	0.129411028E-04
14	11	0	1	0.498055397E-05
15	16	11	0	0.304207674E-06 (S3)
16	8	0	1	0.129719019E-04
17	13	0	1	0.130673271E-04
18	8	2	1	0.492563777E-05
19	5	0	1	0.500132410E-05
20	7	2	1	0.498055397E-05
21	11	0	1	0.496046141E-05
22	10	0	1	0.503213232E-05
23	4	0	1	0.508378071E-05
24	5	2	1	0.496046141E-05
25	7	0	1	0.727611678E-07 (S3)
26	18	0	1	0.496048324E-05
27	7	3	1	0.843373255E-05
28	8	4	1	0.565976492E-07 (S3)
29	7	0	1	0.505290245E-05
30	19	0	1	0.129224025E-04
31	5	0	1	0.492563777E-05
32	5	2	1	0.508606720E-05

(c) 減速法<NN>

PROBLEM NO.	NUM. OF ITE.	NUM. OF DEC.	EFLG	NORMALIZED NORM (∞)
1	18	10	1	0.503213232E-05
2	5	0	1	0.490556067E-05
3	8	4	0	0.148124400E-06 (S3)
4	16	0	1	0.500132410E-05
5	25	9	1	0.503222600E-05
6	9	0	1	0.508328139E-05
7	9	0	1	0.299147302E-06 (S3)
8	23	9	1	0.492570325E-05
9	5	2	1	0.128900401E-04
10	14	0	1	0.492557410E-05
11	5	1	1	0.584577629E-05
12	14	1	1	0.282845394E-05
13	9	6	1	0.129416612E-04
14	10	0	1	0.493335677E-05
15	15	9	0	0.311538486E-06 (S3)
16	13	11	0	0.64829790E-07 (S3)
17	16	4	1	0.273488968E-05
18	12	7	1	0.492563777E-05
19	6	2	1	0.273488968E-05
20	6	0	1	0.496046141E-05
21	15	7	1	0.496046141E-05
22	16	0	1	0.503213232E-05
23	4	0	1	0.508378071E-05
24	6	3	0	0.408226151E-05
25	9	0	1	0.64676537E-07 (S3)
26	18	0	1	0.496046141E-05
27	7	3	1	0.843373255E-05
28	8	4	0	0.565976492E-07 (S3)
29	10	4	1	0.496046141E-05
30	21	3	1	0.345138324E-05
31	10	6	1	0.51078990E-05
32	8	4	1	0.496046141E-05

表 8.1. 実験2における種々の減速法の収束の様子 (続き)

(d) 減速法 <PM>

PROBLEM NO.	NUM. OF ITE.	NUM. OF DEC.	NUM. OF PRO.	EFLG	NORMALIZED NORM(INF.)
1	2	0	1	1	0.7435312E-02
2	6	2	1	1	0.4960633E-05
3	92	90	1	0	0.1486115E-06 (52)
4	22	16	1	1	0.8721750E-02
5	2	0	1	1	0.5152330E-02
6	15	13	1	1	0.7153386E-02
7	101	100	2	1	0.1974783E+00
8	3	1	1	2	0.5584717E-02
9	13	11	1	1	0.6740814E-05
10	19	14	1	1	0.9150678E-02
11	8	5	1	1	0.4925703E-05
12	12	0	1	1	0.7434640E-02
13	9	9	1	1	0.4980379E-05
14	53	49	1	1	0.2458800E-04
15	101	100	2	1	0.1351068E-02
16	18	16	1	1	0.4716670E-02
17	19	16	1	1	0.8778072E-02
18	6	5	1	1	0.5013132E-05
19	3	1	1	1	0.3451440E-05
20	5	4	1	1	0.6013166E-03
21	8	6	1	1	0.7310408E-02
22	11	9	1	1	0.9047361E-02
23	4	0	1	1	0.5083781E-05
24	7	4	1	1	0.4905338E-05
25	51	49	1	1	0.4905561E-05
26	11	11	1	1	0.8290198E-02
27	16	6	1	1	0.2487388E+00
28	101	100	2	1	0.1737295E+00
29	39	37	1	1	0.5255335E-05
30	17	9	1	1	0.8271281E-02
31	5	2	1	1	0.4905454E-05
32	7	4	1	1	0.4905538E-05

(e) 減速法 <HB> ($\alpha=0.01$)

PROBLEM NO.	NUM. OF ITE.	NUM. OF DEC.	NUM. OF PRO.	EFLG	NORMALIZED NORM(INF.)
1	90	85	2	1	0.4980672E-05
2	2	2	2	1	0.4960633E-05
3	56	54	2	0	0.1483754E-06 (52)
4	101	93	3	0	0.8619323E-02
5	101	89	4	2	0.3084615E-04
6	60	54	3	2	0.4980554E-05
7	101	100	3	2	0.1274783E+00
8	78	66	1	2	0.5001349E-05
9	6	4	2	1	0.3345624E-05
10	54	44	2	1	0.4833903E-05
11	15	1	3	1	0.5845776E-05
12	12	2	3	1	0.5001725E-05
13	9	9	2	1	0.4980579E-05
14	41	36	3	1	0.4925638E-05
15	101	100	3	2	0.1974783E+00
16	18	16	3	2	0.4874668E-05
17	101	96	2	2	0.8162107E-02
18	3	3	3	2	0.5032132E-05
19	6	1	3	1	0.3451440E-05
20	5	1	2	1	0.5107900E-05
21	3	0	2	1	0.3451440E-05
22	32	28	2	1	0.2383727E-02
23	4	0	2	1	0.2436211E-02
24	7	4	2	1	0.5083781E-05
25	49	49	2	1	0.2946325E-05
26	101	92	3	2	0.4108292E-05
27	6	4	3	2	0.4335314E-05
28	56	54	2	0	0.3103008E-04 (52)
29	31	28	2	0	0.1679305E-05
30	101	91	3	1	0.9161163E-02
31	5	2	4	2	0.4905454E-05
32	7	4	3	1	0.4905538E-05

(f) 減速法 <HB> ($\alpha=0.5$)

PROBLEM NO.	NUM. OF ITE.	NUM. OF DEC.	NUM. OF PRO.	EFLG	NORMALIZED NORM (CO)
1	15	7	9	1	0.5083898E-05
2	5	4	5	1	0.4905560E-05
3	18	4	4	0	0.1481251E-06 (52)
4	16	0	16	1	0.5001324E-05
5	22	5	17	1	0.7319571E-05
6	9	0	9	1	0.5083281E-05
7	14	9	3	0	0.148050674E-06 (52)
8	21	6	16	1	0.5083853E-05
9	5	2	4	1	0.128200401E-04
10	14	0	14	1	0.4923574E-05
11	14	1	13	1	0.5845776E-05
12	14	1	13	1	0.284813016E-05
13	5	5	4	1	0.4735377E-05
14	9	0	9	1	0.17353674E-02 (52)
15	23	18	2	0	0.5001324E-05
16	14	3	14	1	0.5001324E-05
17	16	3	14	1	0.496048324E-05
18	6	2	5	1	0.2734535E-05
19	4	0	4	1	0.496046141E-05
20	6	2	4	1	0.498055397E-05
21	14	8	12	1	0.5032132E-05
22	10	0	10	1	0.508378071E-05
23	4	3	4	1	0.408226151E-05
24	14	11	4	1	0.492554955E-05
25	6	0	18	1	0.496048324E-05
26	18	3	5	1	0.496048324E-05
27	7	3	3	1	0.843375255E-05
28	14	10	6	0	0.31255851E-06 (52)
29	14	10	4	1	0.498038935E-05
30	21	3	20	1	0.26797427E-05
31	10	6	7	1	0.530690977E-05
32	7	2	5	1	0.498055397E-05

(g) 減速法 <HB> ($\alpha=0.9$)

PROBLEM NO.	NUM. OF ITE.	NUM. OF DEC.	NUM. OF PRO.	EFLG	NORMALIZED NORM (CO)
1	16	7	16	1	0.490553793E-05
2	5	2	5	1	0.4905560E-05
3	11	3	10	0	0.148021039E-06 (52)
4	16	0	16	1	0.5001324E-05
5	23	6	23	1	0.7319571E-05
6	9	7	7	1	0.5083281E-05
7	12	7	3	0	0.1483754E-06 (52)
8	26	5	26	1	0.1711594E-05
9	5	1	5	1	0.498055397E-05
10	14	0	14	1	0.4923574E-05
11	4	1	14	1	0.5032132E-05
12	14	1	14	1	0.284843394E-05
13	5	1	6	1	0.503218234E-05
14	9	0	10	1	0.449339677E-05
15	12	9	3	0	0.148050617E-06 (52)
16	9	5	5	0	0.727620733E-07 (53)
17	14	0	14	1	0.503205138E-05
18	5	5	16	1	0.5001324E-05
19	5	0	5	1	0.5001324E-05
20	4	0	4	1	0.496046141E-05
21	13	4	13	1	0.496046141E-05
22	10	0	10	1	0.5032132E-05
23	4	0	4	1	0.5032132E-05
24	13	9	5	1	0.408226151E-05
25	6	0	18	1	0.496048324E-05
26	18	3	5	1	0.496048324E-05
27	7	3	3	1	0.843375255E-05
28	14	10	6	0	0.31255851E-06 (52)
29	14	10	4	1	0.498038935E-05
30	20	4	20	1	0.26797427E-05
31	9	1	9	1	0.530690977E-05
32	7	2	7	1	0.498055397E-05