数理解析研究所講究録
190巻 1973年 57-70

最急降下法に関する二三の問題点について

青山学院大・理工　馬渡鎮夫

§1．はじめに

n次元実Euclid空間R^nからR^mへの写像fに関する方程式

$$f(x)=0 \tag{1.1}$$

の数値解法について考察する。こゝで，

$$x=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad f(x)=\begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix} \tag{1.2}$$

であり，R^nにおいて写像fの全微分$Df(x)$は存在して連続であるとする。$m=n$のとき，(1.1)の反復解法の中で最も基本的な解法群は，Newton法系列のものである。これらはt_kを正数，H_kを$n\times n$行列とするとき，反復公式；

$$x^{[k+1]}=x^{[k]}-t_k\cdot H_k\cdot f(x^{[k]}) \tag{1.3}$$

によって第$k+1$番目の近似解を求めるものである。例えば$t_k=1$，$H_k=[Df(x^{[k]})]^{-1}$のときが通常のNewton法であり

$t_k>0$, $H_k=[Df(x^{[k]})]^T$のときが最急降下法であり，$t_k=1$ $\lambda_k\geqq 0$, $H_k=[[Df(x^{[k]})]^T[Df(x^{[k]})+\lambda_k I]^{-1}[Df(x^{[k]})]^T$のときが Levenberg-Marquardt 法等々である。これらの中で，最急降下法は，単独では特別効果的ではないものの，他の方法例えば通常の Newton 法と混用すれば非常に強力となり[a]，その解法は $m\neq n$ のとき，すなわち変数の数と条件等式の数とが一致しないときも容易に適用される上に，数理計画問題という重要な応用分野を持っているので，この方法について考察を深めることは有意義であると思われる。

最急降下法については，次の4つの問題点がある。

〔Ⅰ〕 反復公式(1.3)において，尺度因子 t_k をどのように定めるか[d]。——すべての $f_j(x)$ が線型であるときは適切な決め方が判明しているけれども[c]，そうでないときは従来から懸案とされて来た[e]。t_k の決め方は，(1.3)の収束性と収束速度に大きな影響をおよぼす。

〔Ⅱ〕 次式；

$$F(x)=\sum_{j=1}^{m}\{f_j(x)\}^2 \qquad (1.4)$$

の定める超曲面が，‘狭い谷’や‘細長い尾根’を持っていたり，$\max\{|D_iF(x)|;i=1,2,\cdots,n\}/\min\{|D_iF(x)|;i=1,\cdots,n\}$ が非常に大きかったりする $F(x)$ に対して，各 step におけ

探索方向に適切な工夫をすること。

〔Ⅲ〕　大域的な探索．すなわち，複数個の根または極小値をすべて求める問題．[f]

〔Ⅳ〕　全微分 $DF(x)$ を効率的に求める問題．[d]

本小論においては，〔Ⅰ〕，〔Ⅱ〕についてある程度の解決を与え，〔Ⅲ〕についての一つの試みを述べ，〔Ⅳ〕についての考察は省略した。また，§5において，非線型整数計画問題への応用について述べる。

§2.　尺度因子 t_k の定め方

(1.4) の全微分 $DF(x)$ は，変数 x_i に関する偏微分 $D_iF(x)$ を用いて次のように表現される．

$$DF(x)=(D_1F(x), D_2F(x), \cdots, D_nF(x)). \quad (2.1)$$

点 x における $F(x)$ の接平面；

$$H_x(y)=F(x)+DF(x)\cdot(y-x) \quad (2.2)$$

と，$DF(x)\neq 0$ であるときの vector；

$$p_x=\frac{1}{DF(x)\cdot[DF(x)]^T}[DF(x)]^T \quad (2.3)$$

とに対して，次の定理が成り立つ。

定理1

関数 $\alpha=F(x)$ は，R^n において連続的に微分可能であり，

点xにおいて$DF(x)\neq 0$とする。

(1) 実変数の関数$G(\gamma)=F(x-\gamma P_x)$は，$\gamma=0$の近傍において狭義の単調減少である。

(2) 関数$\alpha=H_x(y)$は，点xの近傍において，$-P_x$の方向で最も急激に減少する。

(3) vector $-P_x$は，点xを通る関数$\alpha=H_x(y)$の等高線と直交する。

(4) $$\forall \gamma\in R \quad H_x(x-\gamma P_x)-H_x(x)=-\gamma \qquad (2.4)$$

(5) 点$x^*=x-F(x)\cdot P_x$は，点xからn次元直線$H_x(y)=0$に下した垂線の足である。

(6) 反復公式；

$$x^{[k+1]}=x^{[k]}-F(x^{[k]})\cdot P_{x^{[k]}} \qquad (2.5)$$

は，$F(x)$に対する通常のNewton反復公式である。

証明

(1) $$\frac{d}{d\gamma}G(\gamma)\Big|_{\gamma=0}=DF(x-\gamma P)\cdot(-P)\Big|_{\gamma=0}=-DF(x)\cdot P=-1<0.$$

(2) δを正定数とし，$v\in R^n$を$\|v\|_2=\delta$となるように変動させ，Lagrangianを作り，v_iで偏微分して0とおく。

$$\varphi(v,\lambda)=H_x(x+v)-H_x(x)+\lambda\left(\sum_{i=1}^{n}v_i^2-\delta^2\right)$$

$$D_iF(x)+2\lambda v_i=0$$

$$\therefore\quad v_i=-\frac{1}{2\lambda}D_iF(x)$$

このことゝ，(1)をあわせて(2)を得る。

(3) $\alpha = H_x(y)$ の点xからの等高線の方向をwとすると,

$$0 = H_x(x+w) - H_x(x) = DF(x)\cdot w.$$

$$\therefore\ p^T\cdot w = 0.$$

(4) $$H_x(x - \gamma p_x) - H_x(x) = -\gamma\cdot DF(x)\cdot p = -\gamma.$$

(5) $H_x(y) = 0$ とする。

$$H_x(x^*) = F(x) + DF(x)\cdot(-F(x)\cdot p) = 0.$$

すなわち, 点x^*は直線$H_x(y)=0$ の上にある。そこで,

$x^* - x$ と $y - x^*$の内積$(x^* - x, y - x^*)$は,

$$-\frac{F(x)}{DF(x)\cdot[DF(x)]^T}\left\{([DF(x)]^T, y-x) - ([DF(x)]^T, x^*-x)\right\}$$

に等しくなり. これは0である。

(6) $$F(x+\delta) = F(x) + DF(x)\cdot\delta + o(\delta).$$

$F(x)+DF(x)\cdot\delta = 0$ の一つの特殊解は, $\delta = -F(x)\cdot p_x$.

Q.E.D.

定理1より, 尺度因子t_kを定めることができて, 反復公式は次のとおりである。

[1°] $x^{[0]}\in R^n$ を任意に与え, d_0を十分大きくとり (例えば $d_0 = 10^{35}$), $F(x^{[0]})$, $DF(x^{[0]})$を求める。

[2°] 次のvectorや関数値を求める。

$$p_k = \frac{1}{DF(x^{[k]})\cdot[DF(x^{[k]})]^T}[DF(x^{[k]})]^T \qquad (2.6)$$

$$y^{[k+1]} = x^{[k]} - \min\{d_k, F(x^{[k]})\} \cdot P_k \qquad (2.7)$$

$F(y^{[k+1]})$，$DF(y^{[k+1]})$.

［3°］$F(x^{[k]}) > F(y^{[k+1]})$ ならば，$x^{[k+1]} = y^{[k+1]}$ とし，k を一つ大きくして［2°］へもどる。

［4°］$F(x^{[k]}) \leqq F(y^{[k+1]})$ ならば，$d_k \leftarrow d_k/10$ とし，k の値を変えないで［2°］へもどる。

上の手続きを，$F(x^{[k]}) = 0$ となって根 $x^{[k]}$ が求まるか，または，$d_k = 0$ かつ $F(x^{[k]}) \neq 0$ となって極小点 $x^{[k]}$ が求まるまで繰り返す。

§3. 方程式の規格化

§2 で述べた問題［Ⅱ］は，(1.2) における写像 $f(x)$ の変数 x_i または成分の関数 $f_j(x)$ が *bad scale* であるときに起るものであるが，これは次に述べるような一見 *trivial* なしかし数値計算上は有効な命題と定義によって解決される。

命題1

$\alpha_j\ (j=1,2,\cdots,m)$，$\beta_i\ (i=1,2,\cdots,n)$ を任意の正数とするとき，次の (1)，(2)，(3) はすべて同値である。

(1) $\forall j; 1 \leqq j \leqq m \quad f_j(x) = 0$.

(2) $\forall j; 1 \leqq j \leqq m \quad \alpha_j f_j(x) = 0$.

(3) $\forall i; 1 \leqq i \leqq n,\ x_i = \beta_i y_i \Longrightarrow \forall j\ \ f_j(\beta_1 y_1, \beta_2 y_2, \cdots, \beta_n y_n) = 0$.

定義1

写像$f(x)$の点xにおける*Jacobian matrix*を

$$Df(x)=(D_i f_j(x)) \quad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m)$$

とし，xによって決る正数を$c(x)$とする。

$$\forall j;\ 1 \leqq j \leqq m \quad \sum_{i=1}^{n} |D_i f_j(x)| = c(x) \qquad (3.1)$$

であるとき，$f(x)$は点$x \in R^n$で規格化されているといい，$c(x)$を写像$f(x)$の点xにおける規格指数という。

$x^* \in R^n$を定点とすれば，(1.1) は，連立方程式；

$$\frac{1}{\sum_{i=1}^{n} |D_i f_j(x^*)|} f_j(x) = 0 \quad (j=1,2,\cdots,m) \quad (3.2)$$

と同値であり，左辺の写像は，点x^*において規格化されており，点x^*における規格指数は1である。

§2で述べた反復は，適当な回数ごとに（例えば点$x^{[0]}, x^{[10]}, \cdots, x^{[100]}, \cdots$において）規格指数1でもって規格化された$f(x)$に適用すればよい。(3.1)は，$\|Df(x)\|_\infty = c(x)$となるように規格化したが，$\|Df(x)\|_1 = c(x)$となるようにも規格化できる。どちらかの規格化を適宜使い分ければ，§1,[Ⅱ]で述べた問題は解決される。なおプログラムにおいては，命題1，(2) の極小点でしかも(1)の極小点でない点に収束して終りとならないような簡単な工夫を必要とする。

§4. 大域的な探索

Newton 法の反復公式が大域的に収束するための十分条件としてNewton-Baluev Theorem がある。[b] しかし，一般には(1.4)の$F(x)$の極小点は複数個存在する。こゝでは，$F(x)$は有限個の極小点を持ち，任意の2つの極小点の間には，ほぼ直線に近い‘谷線’が存在する場合について考察する。

(1.2)が点$x \in R^n$の十分小さい近傍において，

$$f(x+\delta)=f(x)+Df(x)\cdot\delta \qquad (4.1)$$

と書けたとする。内積$(f(x+\delta)-f(x), f(x+\delta)-f(x))$を変形して

$$(\delta, [Df(x)]^T\cdot Df(x)\cdot\delta)=F(x+\delta)-F(x)-DF(x)\cdot\delta$$

こゝで，$[Df(x)]^T Df(x)$は対称な$n\times n$行列であるから，その固有値λ，固有vector v_λは容易に求めることができて，

$$\lambda\|v_\lambda\|^2=F(x+v_\lambda)-F(x)-DF(x)\cdot v_\lambda \qquad (4.2)$$

この等式を幾何学的に解釈すれば，点xの近傍において，$[Df(x)]^T Df(x)$ の最小固有値λ_1に対応する固有vector v_{λ_1}の方向は，曲面$\alpha=F(x+\delta)$と接平面$H_x(x+\delta)$とが最も密着している方向に一致する。従って，v_{λ_1} の方向が谷線の方向と考えられる。このことに着目し，かつ，一つの極小点から谷線を登って行き，峠点を通過することができるならば，次のような手続きを経て他の極小点の探索を行うことができる。

$v(x)$ を $[Df(x)]^T Df(x)$ の最小固有値に対応する長さ1の固有vectorとし，ε を適当に小さい正数，$x^{[0]}$ を一つの極小点とする。

〔1°〕　$y^{[k]} = x^{[k-1]} + \varepsilon v_\lambda(x^{[k-1]})$ ，$F(y^{[k]})$ を求める．

〔2°〕(i) $F(y^{[k]}) \geqq F(x^{[k-1]})$ ならば，$x^{[k]} = y^{[k]}$ とし，k を一つ大きくして〔1°〕へもどる．

(ii) $F(y^{[k]}) < F(x^{[k-1]})$ のとき，内積 $(y^{[k]} - x^{[0]}, v_\lambda(x^{[k-1]}))$ を求め，これが非負ならば $x^{[k]} = y^{[k]}$ として〔5°〕へとぶ。

〔3°〕そうでないときは，次の値を求める．

$$z^{[k]} = x^{[k-1]} - \varepsilon v_\lambda(x^{[k-1]}) \text{ , } F(z^{[k]})$$

〔4°〕(i)　$F(z^{[k]}) \geqq F(x^{[k-1]})$　ならば，$x^{[k]} = z^{[k]}$ とし，k を一つ大きくして〔1°〕へもどる。

(ii) $F(z^{[k]}) < F(x^{[k-1]})$ ならば，$x^{[k]} = z^{[k]}$ とし〔5°〕へとぶ。

〔5°〕$x^{[k]}$ を初期値として，最急降下法を適用する。

§5.　非線型整数計画問題への応用

非線型の整数計画問題；

$$G(x_1, x_2, \cdots, x_n) \longrightarrow min \qquad (5.1)$$

(P)　*subject to*

$$\begin{cases} g_j(x_1, x_2, \cdots, x_n) \geqq 0 \ (j=1,2,\cdots,m) & (5.2) \\ x_i \geqq 0 \text{ , } x_i \text{ is integer.} \ (i=1,2,\cdots,n) & (5.3) \end{cases}$$

の解法は非常に困難とされている。しかし，応用上の問題においては，$G(x)\geqq 0$，$G(x)$，$g_j(x)$ $(j=1,2,\cdots,m)$ は連続的に微分可能と仮定しても一般性を失わないことに着目し，かつ問題（P）を次のように変形すれば連続型の問題となり，一つの極小値は比較的容易に求めることができる。

$$G(x_1, x_2, ---, x_n) \longrightarrow min$$

$$\text{(Q)} \quad \text{subject to} \quad \begin{cases} \{-g_j(x_1,x_2,---,x_n)\}_+^1 = 0 \quad (j=1,2,---,m) & (5.4) \\ K_i(x_i)\equiv\prod_{k=0}^{\infty}\left[1-\exp\{-20(x_i-k)^2\}\right] = 0 \quad (i=1,2,\cdots,n) & (5.5) \end{cases}$$

こゝで，l を任意の整数とするとき，

$$\{\alpha\}_+^l = \begin{cases} \alpha^l & \text{if } \alpha > 0. \\ 0 & \text{if } \alpha \leqq 0. \end{cases} \qquad (5.6)$$

関数 $K_i(x_i)$ の関数値および微分値は，$e^{-20} \fallingdotseq 0.2\times 10^{-8}$ 以下の数を0と見なすことにすれば，短時間で容易に計算することができる。問題（Q）の解法は次のとおりである。

〔 *Stage* 1 〕

連立方程式（5.4），（5.5）を解く。もし，解 $\bar{x}$ が存在しなければ，問題（P）は実行不可能である。

〔 *Stage* 2 〕

$\bar{x}$ を初期値とし，制約条件（5.4）の下で $G(x)$ の極小点 x^*

を求める。その方法は、前節までに述べた最急降下法を

$$F(x)=\{G(x)\}^2+\sum_{j=1}^{m}\{-g_j(x)\}_+^2 \qquad (5.8)$$

に適用するが，その際，$G(x)$は規格指数1でもって規格化し $g_j(x)<0$なる関数に対しては，$\{-g_j(x)\}_+^1$を規格指数4でもって規格化し，(2.7)の$y^{[k+1]}$は

$$y^{[k+1]}=x^{[k]}-d_k\cdot P_k \;.\; (d_0=10). \qquad (5.9)$$

と修正する。たゞし，

$$D_i\{-g_j(x)\}_+^1=-\{-g_j(x)\}_+^0\cdot D_i g_j(x) \qquad (5.10)$$

とする。

［Stage 3］

x^*を初期値とし，

$$F(x)=\{G(x)\}^2+\sum_{j=1}^{m}\{-g_j(x)\}_+^2+\sum_{i=1}^{n}\{K_i(x_i)\}^2 \qquad (5.11)$$

に最急降下法を適用して，極小点x^*を求める。各々の$K_i(x_i)$を規格指数1でもって規格化する外は，［Stage 2］と同様である。

最後に求まったx^*は，問題(P)の局所的な最適点である。

§6. 数値実験例

§2 ～ §4に述べた方法の数値実験例を2つだけあげる。

【実験例1】 [a], [d], [e]

$$\begin{cases} f_1(x_1, x_2) = 10(x_2 - x_1^2) = 0 \\ f_2(x_1, x_2) = 1 - x_1 = 0 \end{cases}$$

初期値 $= (-1.2, 1.0)$

精度 $\varepsilon = 10^{-6}$; $F(x) \equiv \{f_1(x)\}^2 + \{f_2(x)\}^2 \leqq \varepsilon$.

結果

関数評価回数	x_1	x_2	$F(x)$
1	−0.1200 0000E 1	0.1000 0000E 1	0.2420 0000E 2
6	0.9332 7951E 0	0.9207 6478E 0	0.2519 9917E 0
8 (*)	0.9833 0672E 0	0.9195 1014E 0	0.2247 8367E 0
14	0.9880 3321E 0	0.9217 5430E 0	0.2966 8128E 0
19 (*)	0.9915 1251E 0	0.9605 0528E 0	0.5111 0910E-1
21	0.9924 5645E 0	0.9611 6828E 0	0.5670 8142E-1
33	0.9981 9703E 0	0.9900 8717E 0	0.3985 0406E-2
35 (*)	0.9993 6250E 0	0.9946 6328E 0	0.1650 4912E-2
37	0.9988 7885E 0	0.9954 8050E 0	0.5203 9250E-3
48 (*)	0.9998 7647E 0	0.9989 2845E 0	0.6799 5877E-4
50	0.9998 5151E 0	0.9989 6667E 0	0.5424 5129E-4
51	0.9998 4048E 0	0.9990 2565E 0	0.4297 2257E-4
59	0.9999 1985E 0	0.9998 2380E 0	0.3168 0406E-7

(*) は，そこで d_k の修正が行われたことを示す。

〔実験例2〕

$$f_j(x_1, x_2, \cdots, x_n) = (x_j^2 - 1)\cdot \exp\left\{\frac{1}{4n}\sum_{i=1}^{n} x_i^2\right\}$$

$$(j = 1, 2, \cdots, n)$$

初期値；$x_i = -2$ for all i : $1 \leqq i \leqq n$.

精度；$\varepsilon = 10^{-6}$, $F(x) = \sum_{j=1}^{n} \{f_j(x)\}^2 \leqq \varepsilon$.

$n = 2$の場合の結果

(i) 56回の関数評価後

$x_1, x_2 = -0.10002261\mathrm{E}1$, $F(x) = 0.67444599\mathrm{E}{-6}$.

(ii) 峠点通過点

$x_1 = -0.97170623\mathrm{E}0$, $x_2 = 0.99271106\mathrm{E}{-1}$

110回の関数評価後, $F(x) = 0.98238353\mathrm{E}{-6}$.

$x_1 = -0.99999950\mathrm{E}0$, $x_2 = 0.10003859\mathrm{E}1$

(iii) 峠点通過点

$x_1 = 0.13056458\mathrm{E}0$, $x_2 = 0.34982609\mathrm{E}1$.

105回の関数評価後, $F(x) = 0.37315082\mathrm{E}{-6}$.

$x_1 = 0.10000416\mathrm{E}1$, $x_2 = 0.10002342\mathrm{E}1$.

(iv) 峠点通過点と誤認した点

$x_1 = 0.77661587\mathrm{E}0$, $x_2 = 0.18532237\mathrm{E}1$

57回の関数評価後, $F(x) = 0.47249107\mathrm{E}{-6}$.

$x_1 = 0.10000519\mathrm{E}1$, $x_2 = 0.10002625\mathrm{E}1$.

§7. おわりに

最急降下法に関する§1で述べた4つの問題点のうち，[I]と[II]については，それぞれ§2，§3でほぼ解決された。問題[III]に関する§4の試みは，未だ試験的段階の域を出なかったが，'谷線を登って峠点を通過する' Algorithm については，今後とも研究の価値があると思われる。

最後に，この小論をまとめるにあたっては，青山学院大学の山内二郎教授をはじめ多くの方々の御助力をいただいた。心から感謝の意を表します。

参考文献

[a] M.J.D. POWEL; *a hybrid method for nonlinear equations*, Goldon and Breach, 1970.

[b] J.M. ORTEGA; *Numerical analysis, second course*, Academic Press, 1972.

[c] 戸川隼人；数値計算技法，オーム社，1972.

[d] 大庭靖男；最適化，情報処理研修センターテキスト 1972

[e] 柳井浩；最適化手法のいろいろ，数理科学講究録62，p81～p138，1968

[f] 篠原能材；非線型連立方程式の大域における数値解法 数理科学講究録72，p69～p81，1969