The University of Osaka
Institutional Knowledge Archive

Title	非線型多変数関数の極小化
Author(s)	小谷, 恒之
Citation	大阪大学大型計算機センターニュース. 1979, 32, p. 27-48
Version Type	VoR
URL	https://hdl.handle.net/11094/65404
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

資　料

非線型多変数関数の極小化

大阪大学教養部　小　谷　恒　之

§ 1. 何を問題にするのか。

変数、あるいはパラメータとも呼ばれるXをN個含んでいる任意の関数Fがあり、Fを最小にするような変数Xの組を決定する問題を考える。このような要求は自然科学から社会科学までの広い分野に見出だされる。例えば、種々の条件を定量化したN変数の組をベクトルXで表したとしよう。このXを変化させて、最も効率的な生産計画を定める経済学の問題や装置を設計するパラメーターの最適値を探す問題から、彎曲した時空内の二点間の測地線を決定する微分幾何学の対象など、まったく異なった分野で多岐にわたる例をみつけることができる。関数$F(X)$は各分野特有の考えで決定されるのだが、ここでは$F(X)$を最小にする変数の組X_mを見出だす一般的な考え方の概略を述べたい。

なお最小化（Minimization）と述べたが、分野によっては、例えば経済活動で生産量の極大値（Maximization）を求める問題のこともある。後者は関数の符号を負にするか、ある十分に大きい数との差を考えるとかで最小化問題に帰着される。両者を併せて最適化(Optimization)と呼ばれることもあるし、強調点によっては最尤法（Maximizing likelihood）、回帰分析（Curve fitting)などいろいろに呼ばれている。ここでは、これらの場合をも含めて、単に最小化と呼ぶことにする。

最小化問題のなかで、よく知られた古典的な例は最小二乗法で、実験式に含まれたN個の変数をM個（一般に$M \gg N$）の実験値から決定する。この場合、関数$F(X)$は

$$F(X)=\sum_{l=1}^{M}\left\{f^{(l)}(X)\right\}^2 \qquad (1.1)$$

と表され、

$$f^{(l)}(X)=\frac{E_l-T_l(X)}{\Delta E_l} \qquad (1.2)$$

と定義される。ここで、E_lは測定値、ΔE_lはその測定の（統計）誤差、$T_l(X)$は変数Xによって決まる理論値である。lはl番目の測定のこともあれば、実験を行なった際のエネルギー等の外部条件を示すこともある。統計学でのχ^2検定も数学的には同種の問題である。

上式（1.2）の$T_l(X)$がXの1次式の場合は、狭義の最小二乗法として昔から知られており、Gaussの正規方程式と呼ばれる連立1次方程式を解くことにより、Xを決定する方法は解析的に

確立されている。一方、1920年代に、生物学のデータの取り扱い方に端を発し、$T_l(X)$ が X の任意の関数の場合に拡張され、Fisher により χ^2 − fit と名付けられた。[1] この非線型関数に関する最小二乗法の研究は、大型計算機が自由に使えるようになった1960年代になって急速に発展してきた。実用的要求とその手法が先行したためか、数学として統一された体系に完成されているとはいい難い。[2)〜5)]

どのような条件のもとに何を求めるかを整理してみよう。

(1) 関数 $F(X)$ はあらゆる X の組に対して、その値が与えられているとする。

(2) 関数の偏微分係数の存在を必ずしも要求しないが、その知識を利用した方が能率的なら用いる。

(3) 変数Xのとりうる範囲が限定されている制約条件付最小化問題も実用上要求されるが、ここでは制約条件は考えない。

(4) $F(X)$ が最小値をとる変数X の値を決定する。しかし、実際問題として、極小値が多数存在する場合、最小値を決定することは容易ではない。(§6で論ずる。)それで、§5まで、一つの極小点を求める問題に限定することにする。すなわち、極小点 X_m は、その近傍のすべての点 X で $F(X) > F(X_m)$ となるような点として定義することにする。

こうして、変数Xの張る空間(X −空間と以下略する)内のある点X_1 から出発して、$F(X)$ が減少する方向に、あるステップΔでXを変化させて、$F(X)$の値が小さくなる点をつぎつぎに探して行き、上記の X_m を見つけ出そうという極小化問題がこの小論の主目的である。

実用上は、$F(X)$の値を計算するのに一番多くの時間を必要とすると思われるので、関数 $F(X)$ の値を求めた回数が少ないことが特に重要である。もちろん、要求される記憶容量の大きさや全体の所要時間も問題になるであろうが、前者に重点を置くことにし、関数評価の回数と呼ぶことにする。

§2. 極小点の発見法

ある点X_1 から出発して、$F(X)$ の極小点を探す方法の基本的な考え方を説明するため、$F(X)$ の Taylor 展開を考える。Xはベクトルであるから、その一般形は

$$F(X) = F(X_1) + g_1^T \cdot (X - X_1) + \frac{1}{2}(X - X_1)^T \cdot G_1 \cdot (X - X_1) + \cdots\cdots \quad (21)$$

である。ここでTはベクトルの転置を意味する。勾配ベクトルg の点 X_j での成分は

$$g_{j\alpha} = \left(\frac{\partial F}{\partial X_\alpha}\right)_{X=X_j}, \quad (22)$$

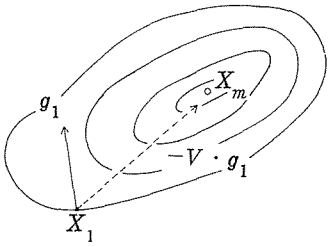
2階微分を成分とする2階テンソルはヘシアン行列と呼ばれ、やはり点 X_j でのそれらの要素は

$$G_{j\alpha\beta} = \left(\frac{\partial^2 F}{\partial X_\alpha \partial X_\beta}\right)_{X=X_j} \quad (23)$$

で定義される。Gは対称行列（$G_{\alpha\beta}=G_{\alpha\beta}$）である。注)

この級数の収束領域がわからなくても、（$X-X_1$)が小さくなれば、高次の項を無視してよいことは明らかである。さて、（2ノ）の第1項は定数項で、極小点を探すのには役に立たない。そこで、次の項に注目しよう。

第2項は勾配gに比例していて、関数が最も速く減少する方向を教えてくれる。この性質を利用した方法は一世紀以上前にCauchyによって用いられ、最大傾斜法（Method of Steepest Descent）として知られている。[4] これはあまり良い結果を与えてくれない。欠点の一つは第2項はXの1次のみで極小点までのステップの大きさを教えてくれないことと、より大きな欠点は、局所的には等高線に直角の方向で、最も速くFの値が減少する方向だが、大局的にFの極小点の方向を向いているとは限らないことである（第ノ図参照）。しかも、極小点の近傍では $g\to 0$ であるから、数値計算の誤差が大きく効き、計算値が収束しにくくなる。



第ノ図　最大傾斜法の短所と$-V\cdot g_1$，（24）：○印は極小点（X_m）を示す。

こうして、放物線型の特徴を与える第3項が問題になる。これは極小点X_mをきめる最低次の項で、X_mの近傍では、gと異なって、Gはほぼ一定値に近くなる。もし$F(X)$がXの2次関数であれば、Gは完全に一定値をとり、一度の操作でもって、極小点X_mをきめることができる。すなわち、$\nabla F=0$ より、

$$X_m=X_1-G_1^{-1}\cdot g_1=X_1-V_1\cdot g_1 \qquad (24)$$

ここでGの逆行列をVとし、共分散行列（Covariance Matrix または誤差（Error）行列）と呼ぶ。この式の利点は

(1) ステップの大きさが任意ではなく、$-V\cdot g$ として正確にきめられている、

注）　ギリシャ文字α、βはテンソルの成分を、ローマ字j、kはベクトルの種類を、lは最小二乗法でのl番目の$f^{(l)}$を表す。

(2) ステップの方向が g の方向ではなく、V を通して、変数間の相関がとりいれられている（第1図）、

からである。この方法は Newton 法とも呼ばれ、例えば、文献4）のプログラム CHIFIT などがある。しかし、この方法は実用上は不安定で、特に G または V が正値行列[注]でなければ発散してしまう。そのためになんらかの工夫が必要である。実際、現在知られている有効で強力な方法は、§4と§5で述べる具体的な例を含めて、大部分この（22）の考え方に基礎を置いているので、それらを一括して、傾斜法（Gradient Method、勾配法）と呼ぶことにする。

上述のように、傾斜法は（21）の第4項以上を無視した、すなわち、$F(X)$ が X の2次関数で近似できる場合に有効である。実際、極小点の近くから探索を始める場合には収束も速い。しかし、極小点から離れた領域では、高次の項の影響で G が X の関数として変化するので、場合によっては、とんでもない所へ暴走したり、進退きわまって同じ計算をくり返す可能性が十分考えられる。

そこで、極小点 X_m より離れた領域で有効な方法を考え出しておくことが必要になる。これらを一括して直接探索法（Direct Search Method）と呼んで、節を改めて§3で考察しよう。

以下の各方法では簡単な Rosenbrock の関数を実例として考えることにする。[6] それは $N=M=2$ の関数で、

$$F(x, y) = \sum_{l=1}^{2} \left[f^{(l)} \right]^2 , \qquad (25)$$

ここで

$$f^{(1)} = 10(y - x^2), \qquad (26)$$

$$f^{(2)} = (1 - x) \qquad (27)$$

である。出発点として点A（$x=-1.2$, $y=1.0$）と点B（$x=-1.2$, $y=5.0$）とをとり、最小点 $x=y=1.0$（精度 10^{-4}）に到達するまでに F の値を何回計算する必要があるかという関数評価の回数を、各方法の優劣の目安にしようというのである。

注） 正値行列の定義 ： A が正方対称行列で、次の条件を満たす場合に A を正値であるという。

(1) 任意のベクトル X に対して $X^T \cdot A \cdot X \geq 0$

(2) $X^T \cdot A \cdot X = 0$ であれば、必ず $X = 0$

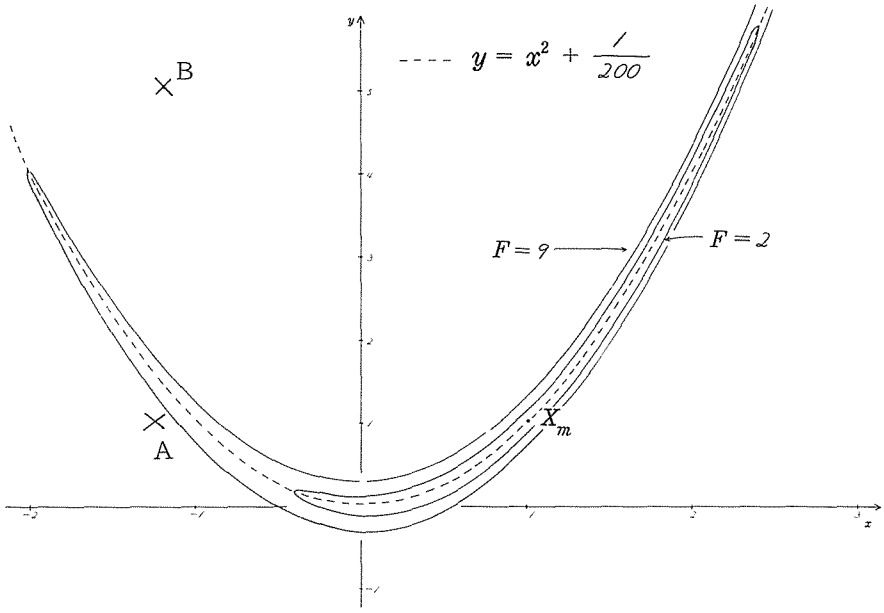
A が正値であることを数値的に確かめるのは簡単ではないが、必要条件は、A の行列要素に関し

(3) $A_{\alpha\alpha} > 0$

(4) $A_{\alpha\alpha} A_{\beta\beta} \geq A_{\alpha\beta}^2$

である。（2×2 行列についてはこれら二つが十分条件でもある。）

この関数の谷底の動きは、第2図に示されるように、ほぼ、$y=x^2+1/200$ の放物線に沿っている。この放物線より上の領域ではヘシアン行列Gは正値ではない（$G_{12}^2 > G_{11}G_{22}$）。出発点Bはわざとこの領域内に選んである。今後、出発点をA〔またはB〕に選んだものをRosenbrock(A)〔または(B)〕の例と呼ぶことにする。なお、計算に要した時間は、コンパイルに要した時間を含まないが、途中経過の書き出し量によって左右されるのであまり深刻にとってはならない。関数評価の回数も計算機の精度に敏感で、以下に述べる数値は日本電気のACOS 900 によるものである。



第2図　Rosenbrockの関数（25）の等高線　：　破線は渓谷の位置、X_mは極小点（$x=y=1$）、A（$x=-1.2$, $y=1$）とB（$x=-1.2$, $y=5$）は出発点を示す。

§3. 直線探索法

いくつかの点での関数値のみを知り、それらを簡単な方法で比較して、次の探索への指示を与え、Fの値がより小さくなる点を探して行く方法を考えよう。この方法なら、関数の性質によらず、どんな関数にも適用できる。さらに、可能なら、数値計算に伴う誤差に鈍感であり、関数Fの微細構造による浅い“くぼみ”に左右されないような方法を考えたいのである。

§3.1. 直 線 探 索

N変数の場合について述べる前に、試行点がある定まった直線上にあって、その直線上での局所的極小点を探す直線探索（Linear Search）について述べよう。これは一変数関数の極小点を求める問題に帰着され、N変数の場合でも、ここにでてくる考え方や処方をそのまま用いることが多いので、例題としても有用である。

最も初歩的な考えは、変数Xの変域内に数点を選び、各点での関数値を比較する方法である。それらのうち最小のFに対応するXの近くで、Xの変域と試行点間の間隔の両者を小さくして、同じ手法をくり返して行くのである。この方法の最大の難点は能率が極端に悪いことで、その原因はFの大きさの知識を次の段階での出発に際し考慮に入れないことである。

その点を改良したものとして、関数を2次の多項式で近似するという仮定を採用する。すなわち、

$$F(X) = a + gX + \frac{1}{2}GX^2 \qquad (3.1)$$

とおいて、3点X_1，X_2，X_3 における関数値をF_1，F_2，F_3 とする。これらの3点を通る放物線の極値は $X_4 = -g/G$ の所にある。すなわち、

$$X_4 = \frac{1}{2}(X_1 + X_2) - \frac{(F_1 - F_2)}{(X_1 - X_2)}\frac{1}{G} \qquad (3.2)$$

ここで

$$G = -\frac{2[F_1(X_2 - X_3) + F_2(X_3 - X_1) + F_3(X_1 - X_2)]}{(X_1 - X_2)(X_2 - X_3)(X_3 - X_1)} \qquad (3.3)$$

である。いうまでもなく、$G>0$ の場合にのみ、放物線の極小値が得られる。この考えを利用し、$F(X)$が非線型関数の場合にも同じく（3.2）、（3.3）を適用する。この場合に、次の諸注意が必要である。

(a) $G<0$ の場合、$X_4 = -g/G$ を採用すると極大値へ向い、発散してしまう。$G \approx 0$ では3点がほぼ一直線にあることを意味して、数値的な難点が発生する。これらの場合、

（a－1）例えば$G=1$とおいて、（3.2）の第2項の示す傾斜方向を用いて最大傾斜法に帰着させるか（ステップの方向は少くとも正しいが、大きさ不明）、（a－2）F_1とF_2から直線近似でステップを定めて次の試行点を決定するか、（a－3）Fがより小さい値になると期待される方向で適当に第4番目の点を定めるかの選択が必要になる。

(b) 次にX_1，X_2，X_3，X_4 の中から、改めて3点を選んで、新しい放物線近似を行なうのであるが、この選び方にも関数値Fとからめた工夫が含まれないと、収束しないことも起こりうる。

(c) 上述の難点が解決されても、この方法は極小値へ収束しないで、その近傍で行きつ戻りつ振動してしまう可能性は十分にある。

これらの諸点に注意し、変なことが発生すれば、再出発するようにすればうまく収束させることが可能である。極小点の近傍では、一般に放物線近似が成立するので、この方法は有効に作用すると期待される。

関数の微分の情報が得られる場合には、この方法のいくつかの修飾型が考えられる。しかし、一般にこれら修飾型はより少ない点からの情報を用いるので、原型よりも不安定で精度が悪くなるであろう。

この方法のN変数への拡張はgを勾配ベクトル、Gをヘシアン行列と読みかえれば、（2.2）になることから明らかであろう。§2でNewton法の難点と呼んだもの、例えば、$G_{\alpha\beta}$ が正値行列という要求はここでいう $G>0$ と同じ内容で、上記の三つの難点が多変数でも同じく問題になるのである。

もう一つ、始まりの3点はどう選ぶかという問題がある。よく用いられるのは[6]点X_1と最初のステップ巾Δとを任意に与え、X_1と$X_2=X_1+\Delta$で関数値を求める。$F(X_2)\leq F(X_1)$なら、$X_3=X_1+\alpha\Delta$（$\alpha>1.0$）とし、$F(X_2)>F(X_1)$なら、$X_3=X_1-\beta\Delta$（$\beta<1.0$）と選ぶ方法である。 $\alpha=2.0\sim3.0$ と $\beta\simeq0.5\sim0.4$ がよく用いられている。

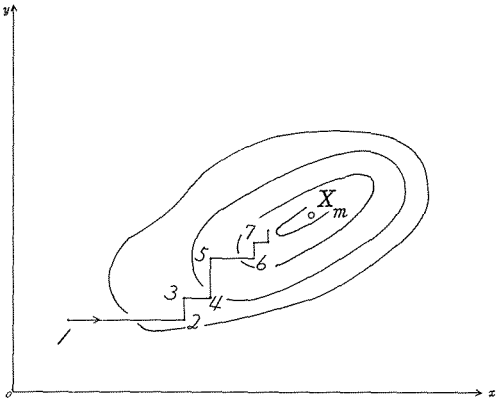
§3.2. 一変数変化法

数理計画法で、巡回緩和法（Cyclic Relaxation Method）と呼ばれている方法である。XのN成分中、（$N-1$）個を固定し、残りの一成分だけを変化させて$F(X)$を最小にする。これを各成分毎に順次何回もくり返そうというのである。原理が簡単なことや短いプログラムで書けるため古くから利用されてきた。

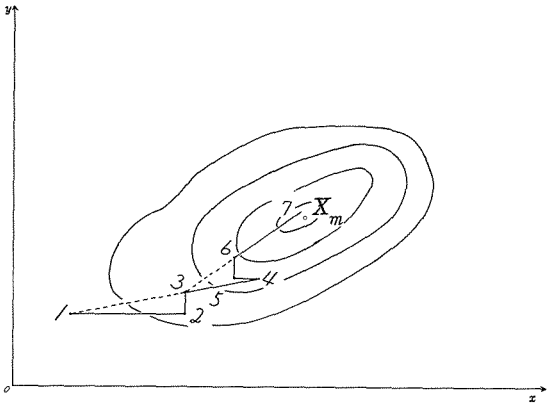
しかし、第3図に示されるように、あまり能率はよくなく、例えばRosenbrock（A）の例で、A点から出発して、答に到達するまでの関数評価の回数は実に2.2, 399回、時間にして31秒を要した。今後述べる方法では、対応する数字は大体300回以下、1秒以内であるから、この方法は簡単で見通しはよいが、実用の対象からは除外する。

§3.3. Hooke－Jeevesの方法[7), 5)]

一変数変化法の能率の悪いのは一方向ずつしか変化させないことである。それを改良するため、例えば、第3図で、点1と3を結ぶ線上に点4を定める。この点4の近くで、もとの点3でのF_3より小さい値（例えば第4図の点6でのF_6）を見出だせば、点3と6を結ぶ線上で点7をきめるという考えである。筆者が§3.1の直線探索の考えを入れて、点4を定める部分等を原著から修正したプログラム（ZIGZAG）では、Rosenbrock(A)の例で182回（0.402秒）、Rosenbrock(B)で283回（0.672秒）とぐっとよくなった。変数が多くなると効率がさがるようである。極小点附近での収束率は当然のことながらあまり速くない。



第3図　一変数変化法（$N=2$ の例）：曲線は$F(X)$の等高線、○印は極小点（X_m）を表す。



第4図　Hooke-Jeeves法（$N=2$ の例）。

§ 3.4. Rosenbrock の方法 [6),5)]

§ 3.3 の方法を更に工夫した方法ともいえる。第4図でいえば、点4で直線1－3と直交する方向を定めて、その線上で極小値を探し、次の探索方向を決定しようというのである。Rosenbrock(A) の例では150回で成功した。この方法は本質的に一つの主軸の方向に沿っての探索が主で、他の方向での探索は次の新しい主軸を決定するためにのみ使われる。変数Nが多くなると、この新しい主軸がX空間内の谷間の方向を必ずしも指向しないので、効率が低下する。

§ 3.5. シンプレックス法 [8),5)]

シンプレックス（単体）とは、N次元空間で（$N+1$）個の頂点を持つ図形のことで、2次元空間では三角形、3次元では四面体を意味する。極小化の問題として、（$N+1$）個の点での関数値についての知識を各段階毎に用いて、極小値に向かって、一定の法則で、図形を変形・移動させて行くので、Nelder と Mead[8)]によりこの名称がつけられた。2次元の場合を例にとれば、第5図に示されるように、初めに3点X_1, X_2, X_3を選ぶ。これらは任意でもよいが、筆者のプログラム（SIMPLX）では工夫がしてある。各点での関数値 F_j を求め、それらの最高値に対応する点をX_H、最低値の点をX_Lとし、X_H以外の点での平均値 $\overline{X}$ を求める。プログラムでは重みをつけて、

$$\overline{X} = \frac{1}{N}\left\{ \sum_{j=1}^{N+1} \left(\frac{X_j}{F_j}\right) - \frac{X_H}{F_H} \right\} \qquad (3.4)$$

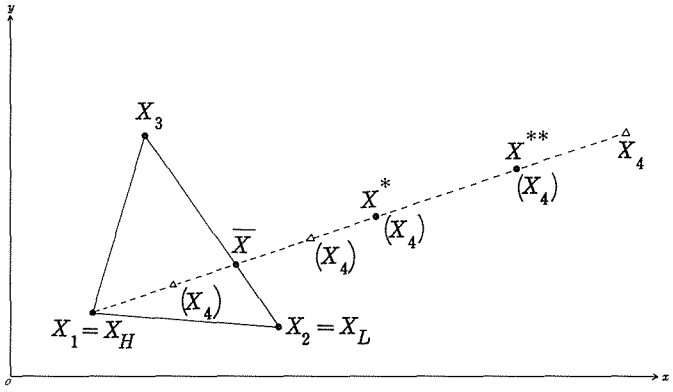
とした。始めのシンプレックスのX_Hをより良い点X_Bで置換して行くのであるが、X_Bをみつけるために、まず、$\overline{X}$についてX_Hの対称点 $X^* = \overline{X} + (\overline{X} - X_H)$ を決定する。$F(X^*) < F(X_L)$

なら、新しい点 $X^{**}=\overline{X}+2(\overline{X}-X_H)$ での $F(X^{**})$ を求める。$F(X^{**})<F(X^{*})$ なら、X_H、X^{*}、X^{**} で直線探索により X_4 を定め、X_H を X_4 で置き換えて新しいシンプレックスとする。ただし X_4 が X^{**} より X_H に近ければ、シンプレックスが小さくなるのを防ぐため、$X_4=X^{**}$ とする。$F(X^{**})>F(X^{*})$ なら、$X_4=X^{*}$ とし、$F(X^{*})>F(X_L)$ なら、X_4 を X_H と X^{*} の間で選ぶ。例えば $X_4=\overline{X}-0.5(\overline{X}-X_H)$ などが一例である。ただし、$\overline{X}$ に近い点を選ぶとシンプレックスがつぶれてしまって、多くの場合（$N-1$）次元の超曲面上の図形になって、回復しないので注意せねばならない。

各ステップが十分大きくなるように工夫されたシンプレックス法は、比較的浅い局所的くぼみや $F(X)$ の計算誤差に鈍感で、かなり収束の速いプログラムが作成できる。さらに、各段階毎に、関数値を1～2回求めるだけで、F の最高値から、F の小さい値の方向へ探索を続けるので、Rosenbrock の方法よりさらに有効に操作することが期待される。

筆者の作成したプログラム（SIMPLX）では Rosenbrock (A) の例で143回（0.054秒）、(B) の例で173回（0.062秒）であった。

上述のシンプレックスが、（$N-1$）次元曲面上に限定されるのを防ぐためにも、各段階で（$N+2$）個以上の点を用いることも考えられるが、筆者が試みた結果では、あまり効果的ではなかった。



第5図　シンプレックス法（$N=2$ の例）：

§ 3.6. その他の方法

過去の探索経験と無関係に次の段階を選ぶ方法として、コンプレックス法[5]などがあり、N が非常に大きく、$F(X)$ が不連続点を持っていたり、急激に変化する場合、また複雑な制約条件のあるときに有効である。この他にもいろいろな変型その他が提案されているが、ここでは割愛する。

§4. 傾斜法（I）— 共役方向を用いる方法

$F(X)$のヘシアン行列Gの特徴を利用するもので、（24）の考えを基礎としている。微分係数の値を直接要求することもあれば、探索の経過で適当にヘシアン行列のもつ特性のみを利用するものもある。

まず共役ベクトルの定義から始めよう。一般に、正値対称行列Gに関して、

$$d_j^T \cdot G \cdot d_k = 0 \qquad j \neq k \text{ に対し} \tag{41}$$

を満たす二つの方向を表すベクトル d_j と d_k は共役であるといわれる。Gが単位行列なら共役ベクトルは直交することから、共役性は直交性の一般化と考えられ、Gram−Schmidt の直交化と似た手続きで、d_1から始めて、N個の共役ベクトルの組を作ることができる。[5] すなわち、Gのn個の積に対し、

$$\langle j \mid G^n \mid j \rangle = d_j^T \cdot G^n \cdot d_j \tag{42}$$

を定義する。ここで $n=1$ または2、jはj番目の共役ベクトルを意味する（$N \geq j \geq 1$）。さらに、$d_0 = 0$、$\langle 0 \mid G \mid 0 \rangle = 1$ とすれば、$j+1$番目の共役ベクトルは次式で与えられる。

$$d_{j+1} = G \cdot d_j - \frac{\langle j \mid G^2 \mid j \rangle}{\langle j \mid G \mid j \rangle} d_j - \frac{\langle j \mid G \mid j \rangle}{\langle j-1 \mid G \mid j-1 \rangle} d_{j-1} \tag{43}$$

ところで、$F(X)$がXの2次関数で、（21）の第3項までの場合を考える。われわれの極小化問題では、ヘシアン行列Gは正値でなくてはならず、もちろん対称である。すると、次の定理がが有用である。

Fletcher−Reeves の定理[9]:

N変数の2次関数の場合、N個の共役方向のそれぞれで直線探索の極小化をすれば、関数そのものの極小化が実行できる。

すなわち、ベクトルの差 $X - X_1$ を d_j 方向の線型和として表す。

$$X = X_1 + \sum_{j=1}^{N} z_j d_j \tag{44}$$

これらを（21）の第3項までに代入し、整理すると

$$F(X) = F(X_1) + \sum_{j=1}^{N} b_j z_j + \sum_{j=1}^{N} c_{jj} z_j^2 \tag{45}$$

ここで、（41）を考慮すれば、$z_j z_k$ の項は現れないので、

$$b_j = g_1^T \cdot d_j, \tag{46}$$

$$c_{jj} = d_j^T \cdot G \cdot d_j \qquad (4.7)$$

この（4.5）の第3項の行列 c はヘシアン行列 G と異なり、対角要素しか存在しないので、N 個の共役ベクトル d_j の方向で、z_j に関して、それぞれ独立な直線探索により極小値を求めればよいことがわかる。こうして、この定理の正しいことが示された。（証明は文献 9）か 5）を参照されたい。）

この定理から、一変数変化法で収束が悪かったのは、共役方向ではなく、直交軸に沿って直線探索をやったのが原因であることが理解される。しかし、共役方向を決めるには、ヘシアン行列 G の知識が必要だし、G がわかっているなら、Newton 法を用いればよいので、N 回も直線探索をくり返すのは蛇足ということになる。

ところが、ヘシアン行列 G を最初に求めなくても、d の定義式（4.1）の特性から共役方向を決めることが可能なことがわかった。それらのうち代表的なものが、次に述べる二つの方法である。もちろん、N 個の共役方向がきまるまでに、行列 G に相当する情報が暗々裏に利用されている。さらに、共役方向を決定する過程で、極小化の手続きが含まれているので効率がよくなる。上記の定理を基本とする意味で、これらの方法は、非線型関数の正定値 2 次関数（放物線）近似を行なっているといえる。

§4.1. Powell の共役傾斜法 [10)]

関数の1階微分すらも計算しない方法で、直線上の直線探索をくり返すことによって共役方向を求めて、極小化を行なう。

それには次の性質を利用する。ある直線 d_1 の方向に沿っての極小点を X_2 とすれば、点 X_2 で求めた1階微分 g_2 は d_1 と直交している。すなわち、F が（2.1）の第3項までの2次関数なら、

$$g_2 = (\nabla F)_{X=X_2} = g_1 + G_1 \cdot (X_2 - X_1) \qquad (4.8)$$

であり、直交関係は次のように書ける。

$$d_1^T \cdot g_2 = 0 \qquad (4.9a)$$

d_1 と平行な他の直線上での極小点を X_4 とすると、同様にして、

$$d_1^T \cdot [g_1 + G_1 \cdot (X_4 - X_1)] = 0 \qquad (4.9b)$$

（4.9 a）と（4.9 b）の両式の差をとれば、

$$d_1^T \cdot G_1 \cdot (X_2 - X_4) = 0 \qquad (4.10)$$

となって、 $d_2=(X_2-X_4)$ は d_1 と共役である。例えば第3図で点2と4を結んだ線は x 軸と共役になっているというのである。もちろん、第3図は正確な2次関数ではないから、(4/0)を正確には満たしていない。何回も述べるように、極小点付近で $F(X)$ が正値2次関数で近似できるのであれば、この方法が有効なのは明らかである。

ただ、3次元であれば、第3軸 (d_3) が、他の二軸 (d_1 と d_2) の両方に共役であるためには、3回の直線探索が必要になり、N 変数の2次関数では全部で $N(N+1)/2$ 回の直線探索によって、N 個の共役方向が決定されることになる。この $N(N+1)/2$ とは G の独立な行列要素の数である。直線探索で極小点を求める数値的な誤差を考えれば、2次関数なら、G を直接求める方が見通しがよいのは明らかである。しかし、非線型関数の場合であれば、それぞれ独立した方向で極小化が進められており、極小点付近では2次関数で速い収束が保証されているという意味でNewton法より優れていると期待される。さらに、1階微分すら求める必要がないことも利点としてあげられよう。

ここで述べたやり方では、d_1 方向での直線探索は N 回、一方 d_N 方向は1回という不平等が発生するので、それも避けるように工夫されたのが Powell の共役傾斜法である。[10]

大阪大学の大型計算機センターのライブラリーには、この方法によるプログラムPOW(星野聰氏作成)が使用可能である。Rosenbrock(A) の例では189回 (0/66秒) であった。

§4.2. Davidon の計量変化法 [11)〜13), 5)]

変数 X の空間と関数 $F(X)$ の関係を微分幾何学との類推で考えると便利である。この節の始めに述べた、直交座標系から共役方向を軸とする系へ一般化したのもこの考え方であった。

2点 X と $X+dX$ の距離の2乗は、非ユークリッド幾何学での不変量として、

$$ds^2 = dX^T \cdot A \cdot dX \qquad (4.11)$$

で定義される。ここで A は考えている空間の性質を決定する共変計量テンソルである。もし A が単位行列であれば、(4.11)は N 次元Euclid空間の Pythagoras の定理を表す。ここでは、A の非対角要素が零でなく、X の関数であるヘシアン行列 G で与えられているという類推を用いようというのである。共分散行列 $V=G^{-1}$ は反変計量テンソルに対応する。(4.11) で行列 A を G と読みかえると、座標変換に関して不変なスカラー量 ds は、また変数 X_j を測る単位の選び方によらないスケール不変な量でもある。さらに有用なスケール不変なスカラー量は勾配ベクトル g_1 と共分散行列 V_1 による

$$\sigma = g_1^T \cdot V_1 \cdot g_1 \qquad (4.12)$$

である。極小値の近傍で、F が正値の放物線で近似できる場合には、点 X_1 と極小点 X_m との関数

値の差（$F(X_1)-F(X_m)$）は$\sigma/2$である。（これは、（2.4）を（2.1）に代入し、第3項で打ち切れば導かれる。Vを誤差行列と呼ぶ理由はここにある。）このσを極小値までの概算距離（Estimated Distance to Minimum）と呼ぶことにするが、スケールに無関係に収束条件を与えてくれるので大変有用である。

Fが2次関数であれば、GはX空間内で一定値をとり、一定の計量を持つ空間を対象にしていることになる。しかし実際の非線型関数では、（2.1）の第4項以下の高次の項が例え小さくても無視できないので、計量テンソルがXの関数であるX空間を問題にしていることになる。それ故、この方法を基礎にするものは計量変化法（Variable Metric Methods、可変計量法）と呼ばれている。Newton法では原理的にくり返し毎にVの値を必要とするが、前回のVを第一近似として、新しいくり返しに入る以前の情報を総合して、Vの補正項hを求めることができれば効率が良くなることが期待される。こうして、計量変化法は一般に次のような段階で進められる。

(1) 与えられた出発点X_1で、勾配g_1を計算し、V_1に対する近似式H_1を作る。原論文[12]ではH_1に単位行列にとっているが、筆者のプログラム（DAFLEP）では、（2.3）で求められるG_1の逆行列を用いた。

(2) 次の試行点は、ベクトル

$$d_j = -H_j \cdot g_j \qquad (j=1, 2, \cdots\cdots) \qquad (4.13)$$

を定義し、

$$X_{j+1} = X_j + \lambda_j d_j \qquad (4.14)$$

と選ぶ。λ_jとしては、もし$F(X)$が正定値2次関数で、$H_j=V_j$と正しく与えられていれば、（2.2）より、$\lambda_j=1$と置いて、X_{j+1}は極小点である。非線型関数では、これだけでは不十分なので、$F(X_{j+1})$が極小値をとるようなλ_jを探し、X_{j+1}が決められる。いずれにしても、$F(X_{j+1})$を極小にする点をX_{j+1}とし、そこでの勾配g_{j+1}を求める。

(3) 行列H_jを次式（Updating Formula）で補正する。

$$H_{j+1} = H_j + h_j(H_j, X_j, X_{j+1}, g_j, g_{j+1}) \qquad (4.15)$$

こうして求めたg_{j+1}とH_{j+1}を用いて段階(2)と(3)をくり返す。ただし、$F(X_{j+1})$等が、想定された収束条件を満たせば計算が停止するようにしておく。

行列H_jの補正項h_jを求める際に、$j=1, 2, \cdots$に応じて、すべてのd_jが共役方向になってくれれば好都合である。事実、こうすることが可能で、Gの値を知らなくても、前段階での1階微分

の知識があれば、新しい共役方向を求める近似法、共役勾配法（Conjugate Gradient）と呼ばれているものがある。

任意の二点 X_j と X_k での微分 g_j と g_k がわかっていて、$F(X)$ が2次関数で表されるなら、次の関係式が成立することは容易にわかる。

$$(g_j - g_k) = G \cdot (X_j - X_k) \tag{4.16}$$

$j=2$, $k=1$ の例が（4.8）にある。いいかえれば、

$$\Delta_k = X_j - X_k \tag{4.17}$$

$$y_k = g_j - g_k \tag{4.18}$$

とおいて、あるベクトル d が y と直交していると、Δ は d と共役であることを示している。

$$d^T \cdot y_k = d^T \cdot G \cdot \Delta_k = 0 \tag{4.19}$$

例えば（4.13）で $H_1=1$ とすると、X_2 は $d_1=-g_1$ 方向の極小点であるから、$g_2^T \cdot g_1 = 0$ の性質がある。このことを利用すると、すべてお互に共役なN個の共役ベクトルは次の簡単な式で与えられる。

$$d_{j+1} = -g_{j+1} + \frac{g_{j+1}^T \cdot g_{j+1}}{g_j^T \cdot g_j} d_j \tag{4.20}$$

ここで $j=1, \cdots\cdots, N-1$ で、$d_1 = -g_1$ とする。なお、g_{j+1} は d_j 方向での極小点 X_{j+1} で求めた勾配である。（証明は文献 9) か 5) をみられたい。）

この共役勾配法をそのまま用いたプログラムもあるが、[9] Davidon[11] と Fletcher-Powell[12] はより一般化して、上記段階(2)と(3)に用い、（4.13）の d_j がお互に共役であると要請して、

$$H_{j+1} = H_j + \frac{\Delta_j \Delta_j^T}{(\Delta_j^T \cdot y_j)} - \frac{(H_j \cdot y_j)(y_j \cdot H_j)}{(y_j^T \cdot H \cdot y_j)} \tag{4.21}$$

という補正項 h_j を見いだすのに成功した。この H_j を用いると、F が2次関数なら、N回勾配を計算し、d_j に沿ってN回直線探索を行なえば収束することや、H_j が正値行列であれば H_{j+1} も正値であることが証明されている。[12), 5)] また（4.16）に対応する

$$H_{j+1} \cdot y_j = \Delta_j \tag{4.22}$$

が満たされていることも明らかである。

1967年以降、この方法の良効率に刺戟されて、上記の H_j の補正項のより一般化が試みられ

た。特に注目されたのが、(4.22)のみでなく、

$$H_k^{(1)} \cdot y_j = \Delta_j, \qquad j=1, \cdots, k-1 \text{ に対し} \tag{4.23}$$

を満たすものとして、

$$H_{j+1}^{(1)} = H_j^{(1)} + \frac{(\Delta_j - H_j^{(1)} \cdot y_j)(\Delta_j - H_j^{(1)} \cdot y_j)^T}{(y_j^T \cdot \Delta_j) - (y_j^T \cdot H_j^{(1)} \cdot y_j)} \tag{4.24}$$

が発見されたことである。[14), 15)]

(4.23)が重要なのは、N回のくり返しをやれば、もしFが2次関数なら、(4.24)で導かれたH_Nは正しい共分散行列Vに一致すること($H_N=V=G^{-1}$)が示されたことである。このことから直線探索が原理的には不要となって、誤差の原因やXに制限条件を課すことが容易になるなど、大きな御利益があることがわかる。

さらに、この(4.24)において、Δとyを入れかえると、$[H_j^{(1)}]^{-1}$から$[H_{j+1}^{(1)}]^{-1}$を求めるのに全く同じ型の式を用いることができる。そして、

$$[H_{j+1}^{(1)}]^{-1} \cdot \Delta_j = y_j \tag{4.25}$$

が成立し、これは(4.16)そのものである。この対称性は

$$y \rightleftarrows \Delta \tag{4.26a}$$

$$H \cdot y = \Delta \rightleftarrows H^{-1} \cdot \Delta = y \tag{4.26b}$$

と記号的に示されよう。

不幸にして、この$H^{(1)}$は、補正項の分母にでてくる$y^T \cdot \Delta$も$y^T \cdot H \cdot y$も正の量で、その差が符号を変えると、数値計算が収束しなかったり、途中のH_jが大きく変動し、正値の保証がないなど、われわれの目的には不向である。

しかし、(4.26)に示された対称性はH_jを求める第三の方法の存在を暗示している。すなわち、H(4.21)が$H \cdot y = \Delta$(4.22)を満たすのであれば、$H^{-1} \cdot \Delta = y$を与えるH^{-1}の対応する表現はないであろうか。事実、次の式が発見された。[13)]

$$H_{j+1}^{-1} = \left[1 - \frac{y_j \Delta_j^T}{(\Delta_j^T \cdot y_j)}\right] \cdot H_j^{-1} \cdot \left[1 - \frac{\Delta_j y_j^T}{(\Delta_j^T \cdot y_j)}\right] + \frac{y_j y_j^T}{(\Delta_j^T \cdot y_j)} \tag{4.27}$$

対称性(4.26a)よりyとΔを入れかえて

$$H_{j+1}^{(2)} = \left[1 - \frac{\Delta_j \cdot y_j^T}{(\Delta_j^T \cdot y_j)}\right] \cdot H_j^{(2)} \cdot \left[1 - \frac{y_j \Delta_j^T}{(\Delta_j^T \cdot y_j)}\right] + \frac{\Delta_j \Delta_j^T}{(\Delta_j^T \cdot y_j)} \tag{4.28}$$

この$H^{(2)}$もHと同じ性質をすべて備えているのである。もちろんFが2次関数ならN回のくり返しで収束する性質を$H^{(2)}$も持っている。

こうして、われわれは共分散行列Vに対する近似式として、Hと$H^{(2)}$より作られた一般型

$$H_{\phi} = (1-\phi)H + \phi H^{(2)} \qquad (4.29)$$

を持つことになった。パラメーターϕを適当に選べば種々の可能な近似式Hが得られることになる。

例えば、

$$\phi = \frac{\Delta^T \cdot y}{\Delta^T \cdot y - y^T \cdot H \cdot y} \qquad (4.30)$$

と選べば、$H_{\phi}=H^{(1)}$で(4.24)に帰着する。

ところで、Hを選択する一つの基準として、Fが2次関数の場合に、Hが単調に正しいVに近づいて行くという条件を要請しよう。この条件は $1 \geq \phi \geq 0$ で与えられることがFletcherによって示された。[13] (4.30)から明らかなように、$H^{(1)}$は$\phi \geq 1$か$\phi \leq 0$かで、この条件を満たしていない。前に述べたように$H^{(1)}$(4.24)の補正項で分母が零になることがあるのは、$H^{(1)}$が正しいVになるにしても単調でないことを示していたのである。

こうして、計量変化法の有力な一つとして、(4.30)の分母に着目し、

$$\begin{aligned} &\Delta^T \cdot y \geq y^T \cdot H \cdot y \ (\phi \geq 1) \text{ なら} \qquad H_{\phi} = H^{(2)} \\ &\Delta^T \cdot y < y^T \cdot H \cdot y \ (\phi < 0) \text{ なら} \qquad H_{\phi} = H \end{aligned} \qquad (4.31)$$

と選んで、$H^{(1)}$に一番近い近似式Hを用いることにする。[13]

現在、大阪大学で使用可能なものとして、H(4.21)のみを用いたプログラムDAVID(星野聰氏作成)がある。Hと$H^{(2)}$を(4.31)に従って使い分けたプログラム(DAFLEP)を筆者が作成した。それらを用いたRosenbrockの関数での結果は次のようになった。勾配gを求めるのに、解析的に微分式を与えた場合と差分法で微分を計算した場合とを比較してみた。ただし差分は

$$\frac{\partial F}{\partial X_{\alpha}} = \frac{F(X+\Delta_{\alpha}) - F(X-\Delta_{\alpha})}{2\Delta_{\alpha}}$$

で求めたので、解析的に微分式を与えた場合も、微分式を用いれば、関数評価の回数を2回として数えてある。

プログラムの種類		Rosenbrock(A)	Rosenbrock(B)
微分式を与えた	DAFLEP	158回(0.101秒)	254回(0.123秒)
差分を用いた	DAFLEP	165〃(0.111〃)	247〃(0.134〃)
	DAVID	761〃(0.306〃)	

この結果でみると、この例に関しては、微分式で g の値を正確に求めても、F を2次式で近似する誤差のため、多少精度の落ちる差分で求めたものと大差がないことがわかる。なお、DAVIDの回数が多いのは、直線探索で微分値を用いてあるのが原因ではないかと思われる。

§5. 傾斜法（II）— 最小二乗法

関数 $F(X)$ が二乗和の形（1/1）の場合には、この性質を利用した特別な方法が考えられる。M 個の $f^{(l)}(X)$ を展開してその1次までとれば線型の最小二乗法になり、$F(X)$ は2次形式だが、そのヘシアン行列は正値行列であることが保証されている。すなわち、$f^{(l)}(X)$ の点 X_j での1階偏微分を

$$g_{j\alpha}^{(l)} \equiv g_{\alpha}^{(l)}(X_j) = \left(\frac{\partial f^{(l)}}{\partial X_\alpha}\right)_{X=X_j} \tag{5.1}$$

と定義すると、ヘシアン行列 G_j は

$$G_{j\alpha\beta} = 2\sum_{l=1}^{M} g_{j\alpha}^{(l)} g_{j\beta}^{(l)} + 2\sum_{l=1}^{M}\left[f^{(l)}(X)\frac{\partial^2 f^{(l)}(X)}{\partial X_\alpha \partial X_\beta}\right]_{X=X_j} \tag{5.2}$$

と書ける。この第2項を無視した近似を考えれば、この $G_{\alpha\beta}$ は正値行列である。

$F(X)$ の極小点が $X_m = X_1 + \Delta$ であるとして、$f^{(l)}(X_m)$ が零であれば、G_j で無視した第2項は Δ の程度であり、すべての $f^{(l)}$ が X の1次式なら正確な式である。極小点 X_m では $\nabla F = 0$ であるから、$\alpha = 1, \cdots, N$ に対して

$$\sum_{l=1}^{M} g_{\alpha}^{(l)}(X_1 + \Delta)\, f^{(l)}(X_1 + \Delta) = 0 \tag{5.3}$$

でなくてはならない。X_1 のまわりに展開して、上記近似を考慮すれば

$$\sum_{\beta=1}^{N}(1 + \mu\delta_{\alpha\beta})\left\{\sum_{l=1}^{M} g_{\alpha}^{(l)}(X_1)\, g_{\beta}^{(l)}(X_1)\right\}\Delta_\beta \approx -\sum_{l=1}^{M} g_{\alpha}^{(l)}(X_1)\, f^{(l)}(X_1) \tag{5.4}$$

となる。正確なTaylor展開では $\mu = 0$ であるが、§5.1で述べる理由で、おまじないとして、μ の項を入れておく。この N 個の連立方程式（$\alpha = 1, \cdots, N$）を解いて、Δ_α を求めれば、X_1 から極小値 X_m への方向とその際のステップの大きさ Δ_α が与えられることになる。

この方法は本質的に $f^{(l)}(X)$ の線型近似であるから、点 X_1 のまわりでのTaylor展開は

$$f^{(l)}(X) \approx f^{(l)}(X_1) + g_1^{(l)T}\cdot(X - X_1) \tag{5.5}$$

と表される。ここで、見方をかえて、$f^{(l)}(X_1)$ と $g_{1\alpha}^{(l)}$ を未知の係数と考えれば、原理的には $M(N+1)$ 点での F の値を知ればそれらの値が決定される。一方、二乗和であることを使わない

§ 4 の方法では、Fが 2 次関数で近似されるとして、対応する未知の係数は（ 21 ）式に出てくる $F(X_1)$、$g_{1\alpha}$ と $G_{1\alpha\beta}$ であるから、全部で $\frac{1}{2}(N+1)(N+2)$ 点でのFの知識が必要になる。例えば $N=10$、$M=2$ とすれば、前者では 22 点、後者では 66 点でのFの情報が必要になる。こうして、Mが小さくNが大きい場合に、この節で述べる方法は、§ 4 の方法より、効率のよいことが一般に期待できる。その代償として、最大の弱点は、極小点でも、正しいVの知識が得られないことである。

§ 5.1. Marguardt の方法 [16), 4), 5)]

上述の近似（ 55 ）は $f^{(l)}$ が X の 1 次式であることを意味するので、この方法は（ 54 ）を整理した Gauss の正規方程式を解いて、Δを求める問題に帰着される。極小点の近傍ではこの近似は有効であろうが、それから離れた領域では§ 4 より効率が悪いのも当然である。Marguardtはそれを防ぐために、極小点附近では $\mu \ll 1$ として 1 次近似をとり、離れた所ではμを大きくして最大傾斜法に近づけるような提案をした。[16)] 要するにμを一定の規則で増減させて、より広い領域で適用しようというのである。実用的なプログラムは文献 4) に CURFIT として与えられている。

§ 5.2. Powell の最小二乗法 [17), 5)]

極小点を指示するΔを求めるために（ 54 ）を解くに際し、Δを共役ベクトルと結びつけて解こうというのが Powell の提案である。[17)] X 空間での直交座標系から出発するが、Δが求まる毎に、座標軸のうち最大傾斜を与える軸をΔの方向で置き換えて、（ 54 ）の近似の範囲内でお互に共役な方向を決定して行く。Fが非線型関数であるため、出発点 X_1で Δ が定まると、$F(X_1+\lambda\Delta)$ が極小値をとるように直線探索を行なってλを決定する。この際に、次のくり返しに登場する新しい共役方向はこのΔであるから、その方向での偏微分係数 $g_\alpha^{(l)}$はλについての差分で求められる。こうして、始めにN直交方向について偏微分係数 $g_\alpha^{(l)}$を求める以外には、偏微分を陽に求めることなしにプログラムを作ることができる。

この方法は極小点付近では、各 $f^{(l)}(X)$の 2 階微分が小さいと予想されるので収束は極めて良い。しかし、極小点より離れた領域や、$f^{(l)}(X)$が急激に変化する場合には、うまく作動しないことが多い。もう一つの難点は、$f^{(l)}$ の偏微分係数を全部記憶させるため、実験の測定値の数Mが大きい場合、記憶容量の大きい（ $\geq M(N+3)+2N^2$ ）計算機が必要なことである。

筑波大学の小柳義夫氏が作成され、筆者が種々改良したプログラム（POWLS）では、Rosenbrock(A) の例で 76 回（ 0.087 秒）、(B) の例で 77 回（ 0.082 秒）であった。

§5.3. Peckham の方法[18), 注)]

Fが二乗和の型式で与えられた場合に、$f^{(l)}$を1次式で近似し、それにシンプレックス法の考えを利用する方法である。$(N+1)$以上の点、例えばp個の点X_kで、$f^{(l)}$を求め、それらを$f_k^{(l)} \equiv f^{(l)}(X_k)$とする。(5.5)での$f^{(l)}(X_1)$と$g_{1\alpha}^{(l)}$の$(N+1)$個を未知の係数とみなし、それらを$h^{(l)}$、$g_\alpha^{(l)}$と表して

$$\sum_{k=1}^{p} w_k^2 \left(h^{(l)} + \sum_{\alpha=1}^{N} g_\alpha^{(l)} (X_1 - X_k)_\alpha - f_k^{(l)} \right)^2$$

が極小になるように$h^{(l)}$と$g_\alpha^{(l)}$を決定する。ここで、重みw_kの選び方を工夫して、それらの係数を計算し易くしている。M個のlについて、これらの$h^{(l)}$と$g_\alpha^{(l)}$、すなわち$M(N+1)$個の係数を定め、(5.4)に対応するものに代入してΔを決定しようという方法である。

藪下氏が作成されたプログラムを Rosenbrock(A)の例に適用して、関数評価回数は48回であった。

Rosenbrock の関数の$f^{(2)}$、(2.7)、が1次式であることにもよるのであろうが、筆者が知る限りでは関数評価の回数は最小である。ただし、この方法は、極小点近くで、$f^{(l)}$の1次近似が有効であるような領域で使用するのが好ましいことは当然である。

§6.1. 実用上、どの方法を用いるか

これまで述べてきた種々の方法は、それぞれの特徴が判然とするように代表的なもののみを取り上げた。これら以外にも、いくつかの方法が提案されているが、紙数の関係もあってすべて割愛し、実用上の問題を考えてみよう。

§3と§4に述べた方法は、§4の方法が極小点附近で関数が二次関数で近似できると想定しているのが唯一の仮定で、きわめて一般的なものである。より効率のよい方法となると、現状では、一般性を多少あきらめて、§5の例のように関数Fの型の特徴を利用することである。そのような例は、例えば、最尤法で用いられる

$$F(X) = -\sum_{l=1}^{M} \ln f^{(l)}(X) \qquad (6.1)$$

である。これも本質的には最小二乗法の(5.2)でやったように$f^{(l)}$の2階偏微分項を落とす方法と同じ考えで処理できる。

また問題によっては、N個の変数間の相関が小さいような特徴があれば、§3.3の Hooke－Jeeves の方法が単純明快である。

注) この論文を指摘し、Rosenbrock(A) の例についての結果を教えて下さった基礎工学部大学院の藪下聰氏に感謝します。

筆者が専門とする素粒子物理学の実験データに関し、超越関数を複雑に含んだ$F(X)$で試みた経験によると、一番勝れている唯一の方法というものは存在しないようである。ある一つの関数Fに対してでさえ、極小点の近傍でも、遠く離れた領域でも、つまり、出発点X_1として選ぶ入力の値によらないで、効率よく働く方法はない。そこで、初期の段階では、直接探索法を用い、極小値に近い値になってから傾斜法を採用するのが、一番効果的であろう。前者としては、シンプレックス法が安定性もよく、意外と効率もよい。傾斜法でやると、狭い“くぼみ”に入り込んで出てこられないことがしばしば発生するが、この場合に、シンプレックス法でステップ巾Δを大小適当に変化させると、すらりと抜け出して、一気にFの値が小さくなることを経験した。傾斜法では、DAFLEP の方が POWLS よりも、このような“くぼみ”を抜け出すのがうまいようだが、これらは関数の性質によるので、どちらがいいとも断定できない。

二つの方法を直列に併用してどこで切り替えるか、また傾斜法のうちのどれを用いるかは、対象とする関数の性質や出発点X_1の選び方に依存している。現状では、使用可能なサブルーチンをいろいろ組みあわせて、試行錯誤で各自が問題とする関数に適したものを決めるという極めて常識的な答しか得られない。計算機に自動的に、これらの選択をさせる方法が開発されつつあるようだが、かなり大変な作業であろう。

実用上もう一つの難問は、§1の(4)でも述べた最小値を決定する問題である。これまでに述べた方法はすべて局所的な極小値を発見するように作られていて、それ以外に局所的極小値があるかとか、正しい最小値になっているかとかには無関係である。一般に、$F(X)$の最小値の位置が予想される場合には、そこから出発すればよいが、それでも注意すべきことがある。これまでに述べたどの方法でも、出発点から一番近い極小点が求まるとは限らないことである。それは次のような類推をとれば理解し易い。X空間を地表、$F(X)$を標高と考えると、山脈のどちら側から出発したか、また峡谷同志が接続しているかどうかもわからないからである。そういう意味で、例えば変数が正数に限られているなら、$X_\alpha > 0$ のようなXの変域に関する精細にわたらない制限を付しておけば、無駄な計算を省くことができる。筆者が作成したSIMPLEX、DAFLEP、POWLS の各プログラムにはそのような単純な制限条件が組み込んである。

非線型関数の場合、多数の極小点の存在が一般に予想される。それらを探し出す最も原始的な方法は、出発点X_1をいろいろと変えてみる方法である。X_1の選び方として、乱数表を用いるモンテ・カルロ法的なやり方、空間を適当に区切って順番に試みる方法などが考えられる。

ところで、局所的極小値が一つ発見された後、他の極小値を探し出す有力な考え方がGoldstein－Priceによって提案された。[19] 例えば、X_1が一つの極小値だったとする。(2/)でいえば、そこでは $g_1 = 0$ であるが、3階や4階微分を含む高次の項が次の極小値を決めるはずである。そこで高次の項のみを含むような次の関数F_1を定義する。

$$F_1(X\ ,X_1)=\frac{2[\ F(X)-F(X_1)\ ]}{[\ (X-X_1)^T\cdot G\cdot(X-X_1)]}=1+(\text{高次の項})\qquad(61)$$

ヘシアン行列Gの知識が必要だが、極小値であれば、この式の分母は正である。したがって、$F_1(X\ ,X_1)<0$となるXを発見できれば、その点を新しい出発点にして、改めてFについての極小化を再開すれば、最小値の発見に近づくことになる。$F_1>0$の場合でも、F_1を極小にするXがみつかれば、最小値ではないが、他の極小値X_2が発見されたことになる。このX_2に関して、(61)の変換を行なって、同様の手続きをくり返して行こうというのである。残念ながら、この方法をどこまでくり返せば、すべての極小値が発見されたといえるのかというプログラムを停止させる論拠はない。

極小点が二個以上発見されれば、それら二点を結ぶ直線上で、適当に新しい出発点を選んで、次の新しい極小点を探すという方法も考えられる。[20] 以上のように、最小点を決める問題や全部の極小点を探し出す問題は試行錯誤の域を出ず、新しい新鮮な発想が要求されている分野といえよう。

極小点を発見する方法そのものについても、将来さらに全く新しい発想がでてくるかもしれないし、それを期待したい。しかし、一方、ここで述べた方法にしても、計算機のプログラムの形になっていなければ、猫に小判である。このニュースの読者なら御存知のように、プログラムを書くことは案外難しく、注意力と根気が要求される重労働である。しかし、だれかが苦労して作ったものを、多くの人が使えるようにすれば、大変便利であろう。本文中に述べた、筆者が作成したプログラムSIMPLX、DAFLEP、POWLS の三つも、予想外に時間がかかって約二年を費した。なん人かの人に試用していただいて、種々の問題や異常事態の発生に対しても、うまく動作するようにしたつもりだが、それらを公開するので、多くの方々のお役に立てば幸だと思う。また、使用されて、改良についての御提案をお聞かせ願えれば幸である。多くの方々が独自に開発されたプログラムがあると思われるが、それらが公開されることを希望して筆を置くことにする。

最後にあたり、作成済のプログラムの使用を認めて下さった筑波大学の小柳義夫博士、ヨーロッパ連合原子核研究所のF．Schremp博士、 阪大の佐々木祥介博士、またプログラム作成のことでいろいろ御教示下さった阪大の渡部陽一博士・松尾武清博士、またカードのパンチや出入力でお世話になった田中瑞枝さんに心から御礼申しあげます。

（昭和５４年（1979年）１月２１日 記）

参 考 文 献

1) R. A. Fisher, *Statistical Methods for Research Workers* (Oliver and Boyd, Edimburgh and London, 1958)
(邦訳 : 遠藤健治, 鍋谷清治訳, 研究者のための統計的方法(森北出版, 1976))

2) M. Abramovity and I. Stegun, eds., *Handbook of Mathematical Functions* (National Bureau of Standards, Applied Mathematical Series, Vol. 55, Washington, 1964)

3) W. T. Eadie, D. Drijard, F. E. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics* (North-Holland, Amsterdam and London, 1971)

4) P. R. Bevinston, *Data Reduction and Analysis for the Physical Sciences* (McGraw-Hill, 1969). p. 215, 222.

5) J. Kowalik and M. R. Osborne, *Methods for Unconstrained Optimization Problems* (American Elsevier Publishing Co., New York 1968) (邦訳 : 山本善之, 小川建夫訳, 非線型最適化問題(培風館, 1971))

6) H. H. Rosenbrock, Computer Journal *3* (1960), 175.

7) R. Hooke and T. A. Jeeves, J. Assoc. Comput. Mach. **8** (1961), 212.

8) J. A. Nelder and R. Mead, Comput. J. **7** (1965), 308.

9) R. Fletcher and C. M. Reeves, Computer J. **7** (1964), 149.

10) M. J. D. Powell, Computer J. **7** (1964), 155.

11) W. C. Davidon, A. E. C. Research and Development Report, ANL-5990 (rev), 1959.

12) R. Fletcher and M. J. D. Powell, Computer J. **6** (1963), 163.

13) R. Fletcher, Computer J. **13** (1970), 317.

14) W. C. Davidon, Computer J. **10** (1968), 406.

15) M. J. D. Powell, A. E. R. E. Tech. Report, T. P. 383 (1969).

16) D. W. Marguardt, J. Soc. Indust. Appl. Math. **11** (1963), 431.

17) M. J. D. Powell, Computer J. **7** (1965), 303.

18) G. Peckham, Computer J. **13** (1970), 418.

19) A. A. Goldstein and J. F. Price, Math. Comput. **25** (1971), 569.

20) I. M. Gelfand and M. L. Tsetlin, Soviet Phys. Dokl. **6** (1961), 192.