数理解析研究所講究録
第 62 巻 1968 年 81-138

最適化手法のいろいろ
—総合報告—

慶応大・工・管理工学科
柳井浩

はじめに

‘函数方程式の近似解法’というシンポジウムで最適化手法—しかも，実際に計算機にかけて使う手法をテーマとして選んだのは，これらの手法を函数空間にやきなおして考えれば，たゞちに‘函数方程式’の解法にも役立てることができると考えるからである。しかし，こゝでは手法の‘やきなおし’を論ずることはせず，むしろ計算技術として使われているものを直接そのまゝの形で列挙して，その考え方，計算手法として具体化するやり方を示し，今後‘函数方程式’の近似解法のアルゴリズムを新しく作る上での参考としたい。

実際，‘方程式の解法’と‘最適化問題の解法’の間には密接な関係がある。すなわち，最小点や最大点をもとめるのに，目的函数 $f(x)$ を微分してゼロとおくという古典的操作を考えれば，たゞちに

$$\mathrm{grad}\ f(x) = 0$$

という連立方程式が得られる。従って，この方程式が解ければよい。しかし，このような連立方程式は，普通，そう簡単には解けないので，最適化問題をこのような操作によって解くにせよ，連立方程式を解く近似的な手法のどれかを動員しなければならないことになる。

また，逆に，

$$g_i(x) = 0 \qquad i = 1, \cdots, N$$

という連立方程式を解く問題も

$$\sum_{i=1}^{N} g_i^2(x)$$

という函数の最小点を求める問題と関係がつく。従って，最適化問題を直接解く良い方法があれば，これを方程式の解法に応用する可能性が生れてくる。

現状をみると，方程式の解法の方が何といっても伝統的な経験の蓄積があり，最適化手法の多くがこれに負っているといってよい。しかし，同時に，古典的な方程式の解法も，最適化問題という場所で考えなおしてみると，直観的なイメージが鮮明になり，有限個の変数に関する方程式の解法を函数方程式の解法に‘やきなおす’場合の手がかりが得られることが少くない。

また，一般に，最適化手法といえば，変数の領域を不等式で与え，その上で考えることの方が正統的と考えられている

うだが，本稿では，上に述べたような理由か [illegible] の制約件は，原則的に考えないことにする。

アルゴリズムの設計と評価

本稿では，いくつかのアルゴリズムを列挙するわけであるが，その各々に対する評価は避けたい。これは，第1に，本稿の興味の対象が主に手法の基礎をなす考え方にあること，第2に，アルゴリズムの評価そのものがきわめて困難だからである。そこで，こゝではたゞアルゴリズムの評価についてごく簡単な一般論だけを述べておくことにする。

ごく常識的にいって，最適化手法のアルゴリズムを評価する場合には次の4点が問題になる。

イ） 計算の簡単さ

ロ） 計算のはやさ

ハ） 計算の確実性

ニ） 計算の融通性

これらは，勿論，独立したものではない。2～3説明を加えると，ハ）では，そのアルゴリズムによって得られた点が本当にその函数の最小(大)点を与えているかいなかが問題である。ニ）で問題にされるのは，そのアルゴリズムがとりあつかえる函数の範囲である。いいかえれば，そのアルゴリズム

が特定のタイプの函数を頭において作られたものであっても，他のタイプの関数にも使えるかいなか？あるいは，多少の誤差を容認するとしたらどうかが問題にされる。

しかし，これら4項目も，アルゴリズムを設計する側と，使用する側とでは，考え方が微妙にくいちがってくる。すなわち，設計する側では，

(a-1) 式の簡単さ

(a-2) iteration の回数

(a-3) exactness (finite quadratic convergence)

(a-4) 凸函数に対する収束性

(a-5) 収束の 'order'

などを頭においているのに対し，使用する側では

(b-1) プログラムの簡単さ

(b-2) 計算時間

(b-3) 種々の 'test function' に対する成績

を頭においているといってよいだろう。なお，2〜3の点に注釈を加えておくことにすると，(a-3) で exact とあるが，目的函数の形が定符号の2次形式である場合に，計算誤差がない限り，有限回の iteration で最小(大)点に到達するようなアルゴリズムを exact あるいは finite quadratic convergent というのである。また，(a-4) で問題にされる点は

目的函数が最小点をもつ凸函数であるとき，そのアルゴリズムによって，(有限回あるいは無限回を問わず) iteration を行うことによって，最小点に収束するような点列が得られるか否かである。さらに，(a-5)は，iteration の回数と，それによって得られた最小(大)点との距離の間の函数関係の推定の一方法である。一方，(b-3)は，どちらかといえば性質のわるい関数から，いくつかの標準的な計算しやすいものをえらび，これらの標準問題を解くはやさ，信頼性等によってアルゴリズムを評価しようとするものである。今日，よくつかわれている test function 等を附録に示す。

問　題

本稿では，N 変数実関数

$$f(x) \qquad x=(x_1, x_2, \cdots, x_N) \in R^N$$

に最大または最小値を与える点 — 最大(小)点 $\hat{x}$ ($\check{x}$) を求めるアルゴリズムを考える。特別の場合を除き，変数には制約条件をつけず，また，函数は唯一つの極大(小)点をもつものとする。

函数が唯一つの極大(小)を持つことを仮定することには，多分，異論があろう。しかし，この条件をのぞいても，確実に最大(小)点に到達できるような手法は何一つないのである

。実用的にも，乱数を使って多くの点における函数の値を調べるとか，出発点をかえて，同じ手法を繰り返す以外に方法がないのである。

また，場合によっては，さらに，目的函数の微分可能性，2回連続微分可能性，あるいは目的函数が定符号の2次形式であることを仮定しなければならない。しかし，同時に，これらの前提条件を除くことが1つの努力目標ともなるわけだから，上の問題設定もいわば標準的な定式化として掲げておくわけである。

変数の数(N)についていえば，1変数の場合と多変数の場合では取扱いを異にするが，ここでは多変数の場合を中心に，1変数の場合は必要に応じて補足して行くことにする。また多変数の場合の説明の際には，主に2変数の場合の言葉——山，谷，尾根，峠，鞍部等を流用する。

以下，各節で順を追い，最大(小)点を求めるアルゴリズムを述べる。よく知られているように，最大点を求める問題と最小点を求める問題は，目的関数の符号を変えるだけで互に変換できるので，以下では，一々ことわらずに，いずれか一方をとりあげて説明する。

I 勾 配 法 (Gradient Metho

これは，最適化の手法として，最も基本的なもので，最も広く使われているものである。それだけに，いろいろな変形があり，いろいろな名称で呼ばれている。中でも，最も原始的な形のものが，次の Primitive Gradient Method である。

1 Primitive Gradient Method

この方法は，大ざっぱにいえば，ある点 $x^{(0)}$ のまわりで函数の値が大きくなる方向を求め，この方向に向っていくらか進み，その点を $x^{(1)}$ とし，又，その点を中心として函数の値が大きくなる方向を求めるという手順を繰返すものである。

目的函数がテーラー展開可能という作業仮説をたてゝ，第2項までをとって考えてみよう。

(1) $$f(x) \cong f(x^{(0)}) + \mathrm{grad}\, f(x^{(0)}) \cdot (x - x^{(0)})$$

すなわち，点 $x^{(0)}$ からあまりはなれていない点に於ける函数の値は，大体，右辺の値で近似されると考えるわけである。そこで，考えている点 $x^{(0)}$ を中心とする半径 $\delta\,(>0)$ の(超)球を考え，この球面上での函数の値が最も大きくなる点を，

上の近似によって求めれば，これが

(2) $$x = x^{(0)} + \delta\cdot(\mathrm{grad}\, f(x^{(0)}))^{T} / \|\mathrm{grad}\, f(x^{(0)})\|$$

であることがわかる。すなわち，この点 $x^{(0)}$ の近傍では，grad f の方向が，目的函数の値が最も大きくなる方向なのである。そこで，まずこの方向に向って進むことが考えられる。（これが gradient 法の名の由来である。）別のいゝ方をすれば，函数が（N＋1次元空間で）作る曲面に，接（超）平面をあてがい，この平面上での最大傾斜線に沿って登って行こうというものである。

この過程は，次の漸化式で示される。

(3) $$x^{(m+1)} = x^{(m)} + h_m (\mathrm{grad}\, f(x^{(m)}))^{T}$$

こゝで，問題となることがいくつかある。

オ1の問題は，初期点 $x^{(0)}$ のえらび方である。これは普通，経験的に最適な点があるだろうと思われる点がえらばれるが，山が2つ以上ある場合には，えらび方如何によって，高い山にのぼらずに，低い山にのぼってしまう可能性があるので，全過程を1回だけでなく，他の初期点についてもくりかえしてみる必要がある。（図 I-1）

オ2の問題は，各変数の単位のとり方である。一般に，各変数が共通の単位をもつ物理量であるわけではないし，また，仮にそうだとしても，同じ測定単位で測定されなければな

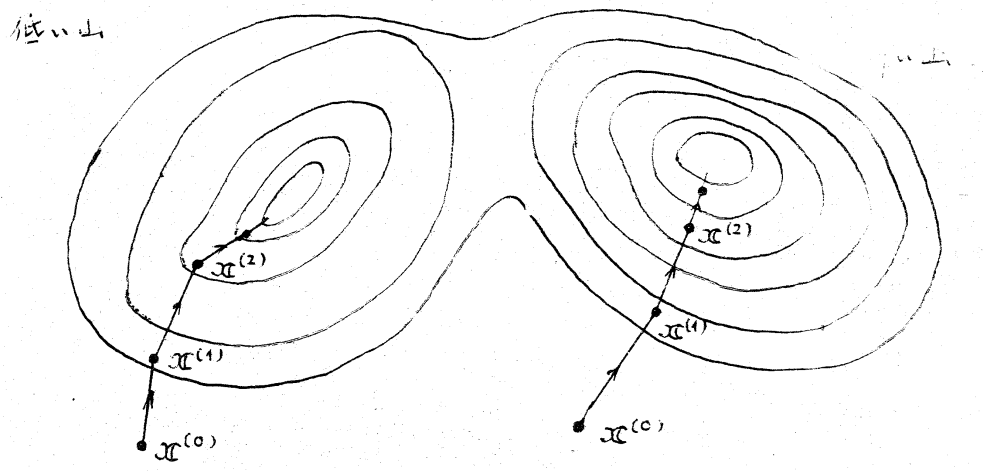


図 I − 1　　$x^{(0)}$のえらび方

ぬという論理的な必然性もない。そこで，変数ごとに，違う測定単位を用いたとしても，上と同様の議論が成立する筈である。しかし，これは，旧座標で，$x^{(0)}$を中心とする球をとることのかわりに，楕円体をとったにすぎない。元来，$x^{(0)}$を中心とする球をとること自体，特別に理由があるわけではなく，たゞ便宜上そうしたにすぎない。極端な話が，$x^{(0)}$を中心とする‘超’立方体をとって，その面上での最大値をとっても，一向に差支えないはずである。

‘立方体’をとるというのも，微分等をする都合上具合がわるいから，一応$x^{(0)}$を中心とする一般的な位置にある楕円体

(4)　$(x - x^{(0)})^T B (x - x^{(0)}) = \delta^2$

こゝに，

B：N×N正定符号行列

を考え，この面上で函数の値が最も大きくなる点を，やはり，テーラー展開の第2項までをとった近似によって求めてみれば，

(5) $$x = x^{(0)} + \frac{\delta B^{-1}(\mathrm{grad}\, f(x^{(0)}))^T}{(\mathrm{grad}\, f(x^{(0)}))B^{-1}(\mathrm{grad}\, f(x^{(0)}))^T}$$

が得られる。この式において，$B = I$（I：N×N単位行列）とおけば，(2)と同じものが得られるが，一般（$B \neq I$）には異る。従って，この方向をたどっても，接平面の最大傾斜線を登ることにはならない。$x^{(0)}$と(5)式をむすぶ方向によって，primitive gradient methodの一般的な漸化式を書けば次のようになる。

(6) $$x^{(m+1)} = x^{(m)} + h_m B^{-1}(\mathrm{grad}\, f(x^{(m)}))^T$$

ところで，こゝで，どのような楕円体（すなわちB）をとったらよいかという問題が生ずるが，これは後で論ずることにしよう。

次に，第3の問題は，'ステップ・サイズ'$\{h_m\}$の選定である。h_mが小さすぎれば，$\{x^{(m)}\}$の動きがにぶくなり，$\hat{x}$になかなか近づかない。大きすぎれば振動を起して，かえって具合が悪い（図I－2）。

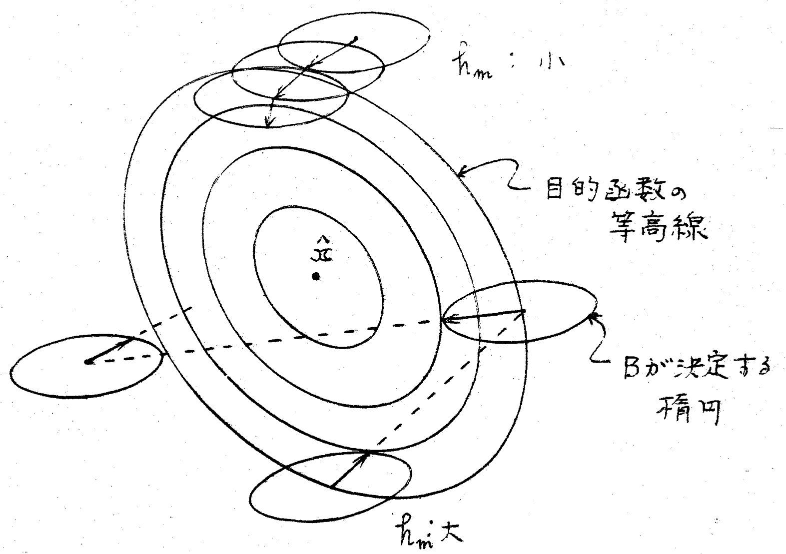


図 I − 2　　h_m のえらび方

それでは，h_m をどのような値に選んだらよいか？まず，どのような条件の下で $x^{(m)}$ が $\hat{x}$ に收束するのか？途中まで行った計算によって得られた，目的函数 $f(x)$ に関する情報を生かして，h_m が收束のための条件をみたすようにすることはできないか？さらに收束の速度をはやめることはできないか？という問題がでゝくる。

h_m のえらび方に関する理論的な研究は Crockett-Chernoff〈4〉 によって行われている。Crockett Chernoff の研究の主要な部分を以下にのべる。

Crockett-Chernoffの研究

目的函数$f(x)$はテーラー展開により，次のように近似できるものとする。

(7) $$f(x)=f(\check{x})+\frac{1}{2}(x-\check{x})^T H(x-\check{x})+o(\|x-\check{x}\|^2)$$

こゝに，

$\check{x}$ ： $f(x)$ の最小点，

$H=\left\{\frac{\partial^2 f}{\partial x_i \partial x_j}\right\}\Big|_{x=\check{x}}$ ：正定符号対称行列

$i, j = 1, 2, \cdots, N$

である。このとき，$\check{x}$を求めるのに，次の漸化式によって与えられる点列を用いる。

(8) $$x^{(m+1)}=x^{(m)}+h_m \delta(x^{(m)})$$

こゝに，

$\delta(x)=-B^{-1}g(x)$

$g(x)=(\mathrm{grad}\, f(x))^T$

B ： $N\times N$ 正定符号対称行列

である。

いま，

(9) $$\delta^{(m)}=\delta(x^{(m)})$$
$$\varepsilon^{(m)}=x^{(m)}-\check{x}$$

とおくと，次の関係式が成立する（証明略）。

$$
(10)\qquad
\begin{aligned}
\mathbf{e}^{(m+1)} &= (I - h_m B^{-1}H)\,\mathbf{e}^{(m)} + o(\mathbf{e}^{(m)}) \\
\boldsymbol{\delta}^{(m)} &= -B^{-1}H\,\mathbf{e}^{(m)} + o(\mathbf{e}^{(m)}) \\
\boldsymbol{\delta}^{(m+1)} &= (I - h_m B^{-1}H)\,\boldsymbol{\delta}^{(m)} + o(\mathbf{e}^{(m)})
\end{aligned}
$$

また，$B^{-1}H$は正定符号対称行列であるから，この行列の固有値はすべて正であり，これに属する固有ベクトルは線形独立である。これを

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N > 0$$

$$\boldsymbol{\mu}_1,\ \boldsymbol{\mu}_2,\ \cdots,\ \boldsymbol{\mu}_N$$

としよう。このとき，$\mathbf{e}^{(m)}$および$\boldsymbol{\delta}^{(m)}$をこの固有ベクトルの線形結合として次のように書こう。

$$
(11)\qquad
\begin{aligned}
\mathbf{e}^{(m)} &= k_1^{(m)}\boldsymbol{\mu}_1 + k_2^{(m)}\boldsymbol{\mu}_2 + \cdots + k_N^{(m)}\boldsymbol{\mu}_N \\
\boldsymbol{\delta}^{(m)} &= \bar{k}_1^{(m)}\boldsymbol{\mu}_1 + \bar{k}_2^{(m)}\boldsymbol{\mu}_2 + \cdots + \bar{k}_N^{(m)}\boldsymbol{\mu}_N
\end{aligned}
$$

$\mathbf{e}^{(m)}$および$\boldsymbol{\delta}^{(m)}$に関する上にあげた関係式(10)によって，これらの係数$k_i^{(m)}$および$\bar{k}_i^{(m)}$に関する次の関係式が導かれる。

$$
(12)\qquad
\begin{aligned}
k_i^{(m+1)} &= (1 - h_m\lambda_i)\,k_i^{(m)} + o(\mathbf{e}^{(m)}) \\
\bar{k}_i^{(m)} &= -\lambda_i k_i^{(m)} + o(\mathbf{e}^{(m)}) \\
\bar{k}_i^{(m+1)} &= (1 - h_m\lambda_i)\,\bar{k}_i^{(m)} + o(\mathbf{e}^{(m)})
\end{aligned}
$$

$\mathbf{e}^{(m)}$や$\boldsymbol{\delta}^{(m)}$がゼロに収束すれば，$\mathbf{x}^{(m)}$が解に収束する。そして，$\mathbf{e}^{(m)}$や$\boldsymbol{\delta}^{(m)}$がゼロに収束するためには，$k_i^{(m)}$や$\bar{k}_i^{(m)}$がゼロに収束すればよい。上の式からあきらかなように，

$k_i^{(m)}$や$\bar{k}_i^{(m)}$がゼロに収束するためには，

(13)　　$|1-h_m\lambda_i|<1 \qquad i=1,\cdots,N$

であればよい。さらに，(13)が成立するためには，

(14)　　$0<h_m<\dfrac{2}{\lambda_1+\lambda_N}$

であればよい。

また，このとき，mが十分大きくなると$|1-h_m\lambda_i|$の小さいものに関するk_iや$\bar{k}_i$は早くにゼロに近づいてしまうから，

(15)　　$\delta^{(m)}\fallingdotseq \bar{k}_N^{(m)}\mu_N$

(16)　　$\delta^{(m+1)}\fallingdotseq(1-h_m\lambda_N)\,\delta^{(m)} \qquad m:大$

が成立する。

理論的には，以上のようなことがわかるわけだが，実際の問題を解く場合には，最初からHやλ_iやμ_iがわかるわけでも，容易に推定できるわけでもない。したがって，上の条件をみたすようにh_mを定める方法が明確になったわけではない。そこで，Crockett-Chernoff は以上の理論をふまえてh_mのえらび方の指針を与えている。

(イ)　h_mはmの増加とともに次第に大きくとる。

(ロ)　$\delta^{(m)}$と$\delta^{(m+1)}$の方向が著しく異るときには，h_mの増加を手びかえる——あるいはh_mを小さく

する。

(ハ) $\delta^{(m+1)} \simeq \rho\, \delta^{(m)}$, ρ：正の定数で1にあまり近くない値

が成立する場合には

$$h_{m+1} = \frac{h_m}{1-\rho}$$

とする。

これらの指針の中で，(イ)は点列が山の長い尾根に沿って登る場合の加速的な効果をねらっており，(ロ)は $x^{(m)}$ が $\check{x}$ のまわりで振動に入るのを防いでいる。(ハ)では，m が大なるとき，(16)が成立するところから，

$$\rho \approx 1 - h_m \lambda_N$$
$$\Rightarrow\ h_{m+1} = \frac{h_m}{1-\rho} \approx \frac{1}{\lambda_N}$$

従って，$\delta^{(m+2)}$ が大巾に減少することが期待できる。

実際の問題を計算機にかけて解こうとする場合には，まだいろいろな問題点がある。例えば，漸化式

$$x^{(m+1)} = x^{(m)} + h_m B^{-1} (\mathrm{grad}\, f(x^{(m)}))^T$$

に於いて $\mathrm{grad}\, f(x)$ が解析的な形で得られることはまれである。従って，何等かの形で数値微分に類することを行わな

ければならない。しかし，各点で非常に精密な数値微分を行っても，全体としての効率があがるわけでないことはあきらかであろう。そのようなわけで，実際のアルゴリズムとしては，かなり大胆に簡略化を行った方がいろいろな点で具合がよい。

次に，実際の計算の場合によく使われているpattern searchとよばれるアルゴリズムを紹介しておく。これは，ある意味で，Crockett-Chernoffの指針を多くの点で具体化したものともいうことができる。

Pattern Search

この方法は大ざっぱにいうと，‘パターンの中心’と‘ベース・ポイント’とよばれる二つの点列を交互に定義して，山頂に向うのであるが，このとき，ベース・ポイントはパターンの中心の附近で函数の値が大なる点に，パターンの中心は二つのベース・ポイントを結ぶ直線上にえらぶのである。

簡単のため，2変数の場合について手順の詳細を説明しておく。3変数以上の場合も同様である。

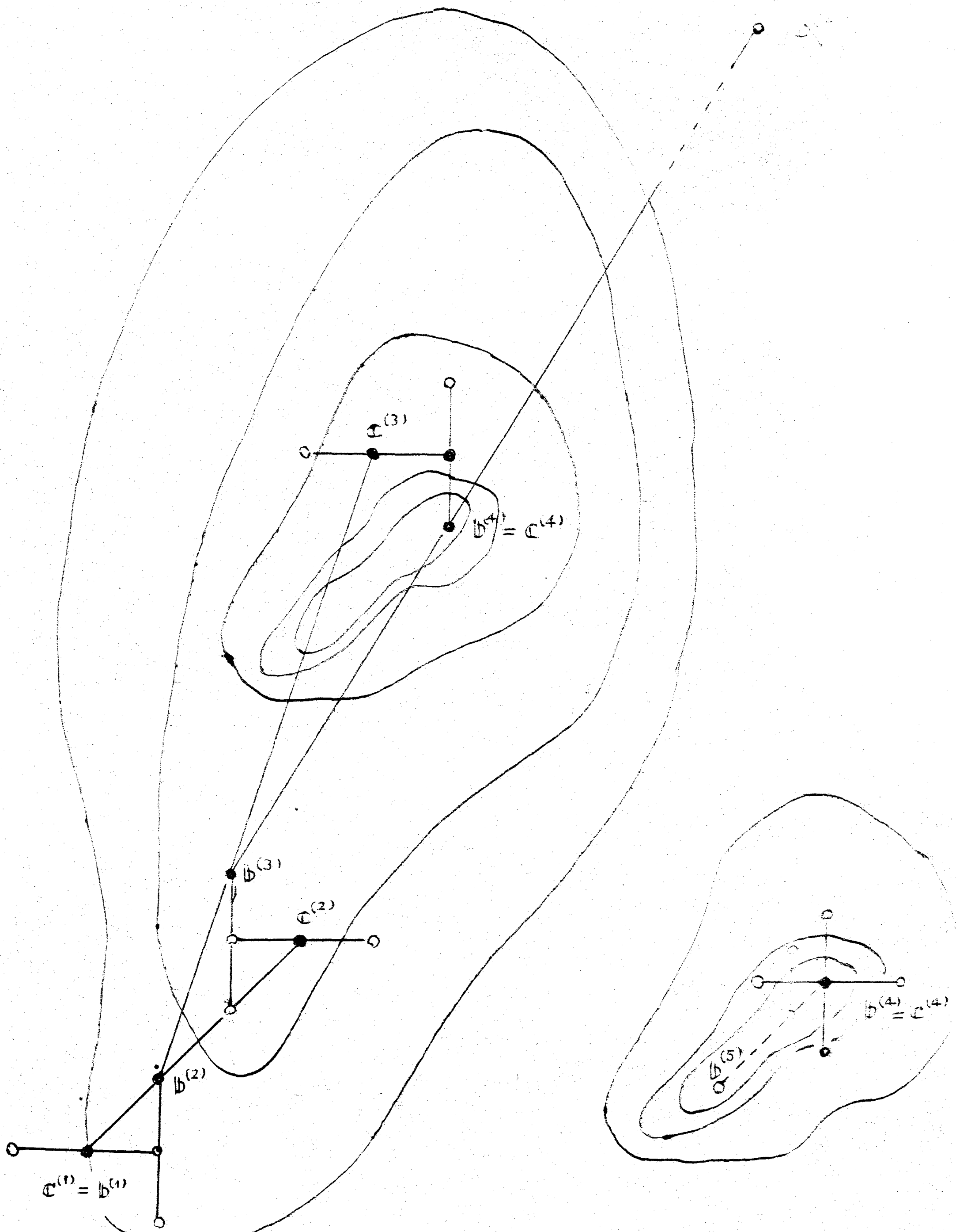


図 I-3 Pattern Search

(i)　二つの数 $\delta_1, \delta_2 > 0$ を適当に定め，これをステップ・サイズとよぶ。

(ii)　適当な1点をえらび，これで，最初のパターンの中心 $\mathbb{C}^{(1)}$ と最初のベース・ポイント $\mathbb{b}^{(1)}$ を兼ねるものとする。

(iii)　<u>パターン決定</u>　いま，m 番目のパターンの中心が $\mathbb{C}^{(m)} = (C_1^{(m)}, C_2^{(m)})$ にあるとき，3点，

$$\mathbb{C}^{(m)} = (C_1^{(m)}, C_2^{(m)}),\ (C_1^{(m)}+\delta_1, C_2^{(m)}),\ (C_1^{(m)}-\delta_1, C_2^{(m)})$$

において目的関数の値を求め，その値が最も大きい点を

$$\mathbb{C}^{(m1)} = (C_1^{(m1)}, C_2^{(m)})$$

と書こう。さらに，3点

$$\mathbb{C}^{(m1)} = (C_1^{(m1)}, C_2^{(m)}),\ (C_1^{(m1)}, C_2^{(m)}+\delta_2),\ (C_2^{(m1)}, C_2^{(m)}-\delta_2)$$

において目的函数の値を求め，その値が最も大きい点を新しいベース・ポイント $\mathbb{b}^{(m+1)}$ とする。パターンの中心から，新しいベース・ポイントを定めるこの手順を<u>パターン決定</u>とよぶ。

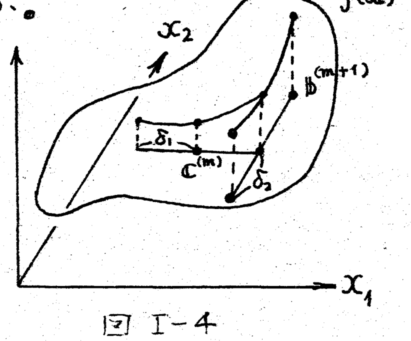


図 I-4

実際にパターン決定を行うのに，一度に3点を比較するのも面倒であるから，例えば，

$(c_1^{(m)}, c_2^{(m)})$ と $(c_1^{(m)}+\bar{\sigma}_1,\ c_2^{(m)})$ に

における函数の値を比較し，$(c_1^{(m)}, c_2^{(m)})$ における値が大きいときだけ $(c_1^{(m)}-\hat{\sigma}_1, c_2^{(m)})$ を調べることにしよう。$(c_1^{(m)}, c_2^{(m)})$ の附近で，函数が‘谷’になっている場合には，こうすると多少遠まわりになることもあるが，多くの場合に計算量をへらすことになる。

(iv) パターン移動　新しいパターンの中心は次のようにして定められる。(図 I-5)

(イ) $f(\mathbf{b}^{(m)}+\alpha(\mathbf{b}^{(m+1)}-\mathbf{b}^{(m)})) > f(\mathbf{b}^{(m+1)})$ のとき，

$$\mathbf{c}^{(m+1)} = \mathbf{b}^{(m)} + \alpha(\mathbf{b}^{(m+1)}-\mathbf{b}^{(m)})$$

(ロ) $f(\mathbf{b}^{(m)}+\alpha(\mathbf{b}^{(m+1)}-\mathbf{b}^{(m)})) \leq f(\mathbf{b}^{(m+1)})$ のとき，

$$\mathbf{c}^{(m+1)} = \mathbf{b}^{(m+1)}$$

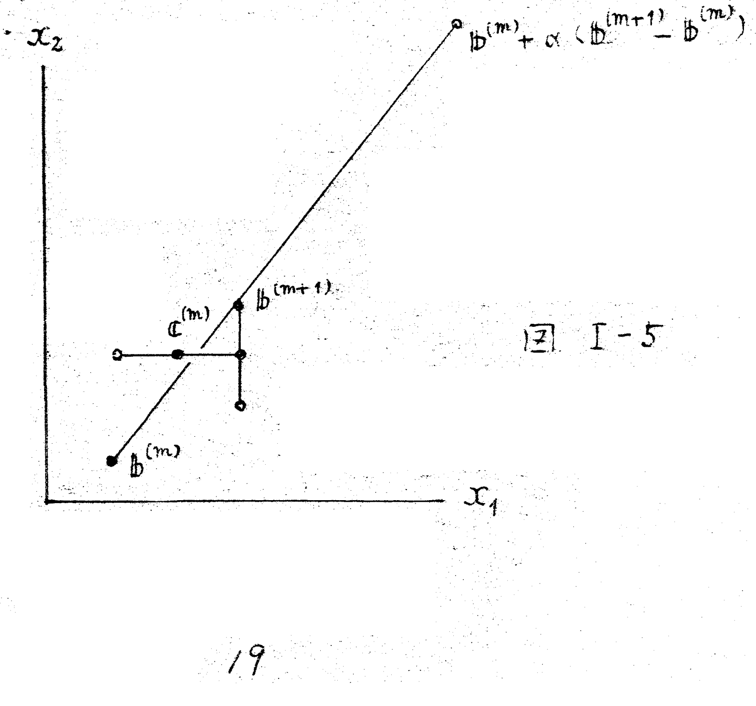


図 I-5

こゝに，αは1より大なる定数であり，通例，2.0にとられる。αが1.0より大きいことは，図Ⅰ-5からもわかるように，尾根にそって進むとき，次第に'歩み'が大きくなるという加速的な効果をもたらす。このようにして，新しいパターンの中心を求める手順をパターン移動という。

(v) <u>ステップ・サイズの縮小と探索の停止</u>　最初に与えられたパターンの中心とベース・ポイントから，パターン決定とパターン移動を交互にくりかえして，点列$\{\mathbb{c}^{(m)}\}$および$\{\mathbb{b}^{(m)}\}$を決定して行くのだが，パターン決定のとき，新しいベース・ポイントが

$$\mathbb{b}^{(m+1)} = \mathbb{c}^{(m)}$$

となったら，これをすてゝ，ステップ・サイズをたとえば$\sigma_1/2$，$\sigma_2/2$に縮小して探索を再開する。このようにして，探索をつゞけるうちにステップ・サイズがあらかじめ定めておいた値より小さくなったら探索をやめるのである。

2 Optimum Gradient Method

h_mをえらぶもう一つの方法は，上に述べた方法で方向がきまったら，その方向を向いた，$x^{(m)}$を始点とする半直線上で目的函数の値が最も大きくなる点をとり，次の近似点$x^{(m+1)}$とすることである。

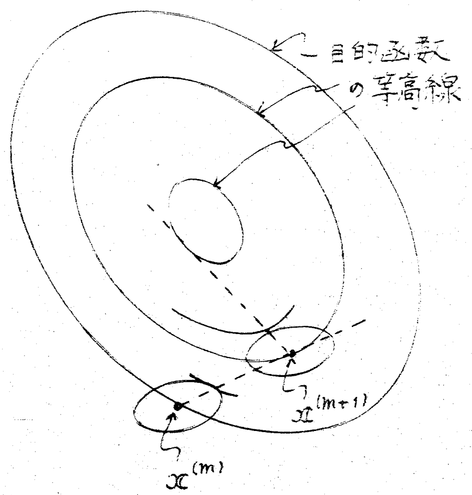


図 I－6

この方法は，Optimum Gradient Methodとよばれるもので，Chauchy＜1＞が1847年に，最初に勾配法を提唱したのもこの形である。Chauchyは，天体の軌道を計算するという応用上の目的から，方程式の一般的な解法としてこの方法を考えているが，文献＜1＞はごく短いもので，収束の証明等はされていない。（勿論，Chauchyが考えたのはB＝Iの場合である。）

その後も，一部の数学者達によって変分問題の解法等に，この方法が応用された。特に，Канторович＜12＞等はHilbert空間で定義された函数方程式の解法として，この方法に関する理論を発展させているが，工学上の応用としては，1944年にCurry＜2＞が応用上のデータはなく，むしろ，

収束の証明等に力が注がれているが，この方法の広い応用の可能性を示唆する一方，座標のとり方の問題等，それ以後他の人々によって研究されたいくつかの問題点が指摘されていることは注目に値しよう。

ここでもまた，目的函数 $f(x)$ を，$x^{(m)}$ を中心としてテーラー展開し，2階偏微係数を含む項までをとって，函数を近似してみよう。

$$f(x) \simeq f(x^{(m)}) + \mathrm{grad}\, f(x^{(m)})\cdot(x - x^{(m)}) + \frac{1}{2}(x - x^{(m)})^T H(x^{(m)})(x - x^{(m)})$$

こゝに，

$$H(x^{(m)}) = \left\{ \frac{\partial^2 f}{\partial x_i \partial x_j} \right\} \quad i, j = 1, \cdots, N$$

この近似にもとづいて，$x^{(m)}$ を通る直線

$$x = x^{(m)} + \xi p$$

上で，函数が最大値をとる点を求めれば，($p^T H(x^{(m)}) p < 0$ のとき

$$x = x^{(m)} - \frac{\mathrm{grad}\, f(x^{(m)})\cdot p}{p^T\cdot H(x^{(m)})\cdot p}\cdot p$$

となる。従って，この近似を用いることにすれば，optimum gradient method の手順は次の漸化式であらわされる。

$$x^{(m+1)} = x^{(m)} - \frac{\mathrm{grad}\, f(x^{(m)})\cdot p}{p^T H(x^{(m)})\cdot p}\cdot p$$

$$p = B^{-1}(\mathrm{grad}\, f(x^{(m)}))$$

なお，ここで直線 $x^{(m)}+\xi p$ が $x^{(m+1)}$ で $f(x)$ の等高線と接することにも注意しておく必要があろう。

ここでもまた，いくつかの問題があらわれる。

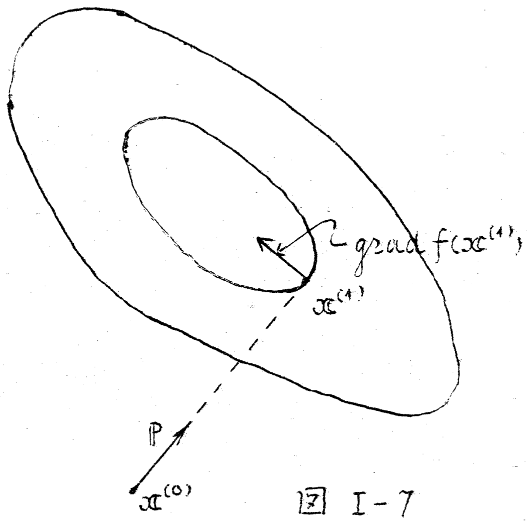


図 I-7

第1の問題は，$x^{(m)}$を1つの計算するたびに，$H(x^{(m)})$と$\mathrm{grad}\, f(x^{(m)})$ を計算しなければならないことである。もちろん，これが解析的な方法でできて，しかも，簡単な場合には問題はない。しかしこれは稀である。そこで，一々数値的に求めることになればこれは面倒である。

まず，

$$\mathrm{grad}\, f(x^{(m)}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_N}\right)\bigg|_{x=x^{(m)}}$$

であるが，この近似値を計算するには，たとえば，

$$\frac{\partial f}{\partial x_1}\bigg|_{x=x^{(m)}} \simeq \frac{1}{\Delta x_1}\Big[f(x_1^{(m)}+\Delta x_1, x_2^{(m)}, \cdots, x_N^{(m)}) - f(x_1^{(m)}, x_2^{(m)}, \cdots, x_N^{(m)})\Big]$$

というような近似を行なわなければならない。このような単純な方法でも，$\mathrm{grad}\,f$ を計算するたびに，目的函数 $f(x)$ の値を $N+1$ 回計算（又は測定）しなければならないので，N が大きいときには大変なことである。

primitive gradient method の場合の，この点に対する一つの解決法が pattern search であったわけだが，optimum gradient method を使おうとするなら，$\mathrm{grad}\,f(x^{(m)})$ の計算はいずれにせよまぬかれない。

次に，

$$H(x^{(m)}) = \left\{\frac{\partial^2 f}{\partial x_i \partial x_j}\right\} \quad i, j = 1, \cdots, N$$

であるが，これを，数値的に近似しようとすれば更に面倒なことになる。そこで，一つの逃げ道として考えられることは，直線 $x = x^{(m)} + \xi p$ 上に於ける函数 $f(x)$ の最大点を直接数値的な方法で求めることである。この方が，optimum gradient method 本来の形であるし，計算も簡単になることが少くないからである。

なかでも，この計算が解析的な方法で行われれば話は簡単であるが，そうでない場合には，一般的にいって，1変数函の最大点を数値的に求めることが問題になる。これにも，いろいろな方法が工夫されているが，ここでは，次のような方法があることにふれておくだけにとゞめよう。

(a) 差分をとり，符号の変化をしらべる。

(b) 2次関数をあてはめ（数値的に），その[illegible]大点を計算（解析的に）する

(c) 黄金分割法

実際，optimum gradient methodの場合に$H(x^{(m)})$が直接計算されることは非常に少ない。これは，せっかく，Hを行列の形で求めてみても$p^T H p$というスカラー量にして使うのでは計算に無駄が多い。行列の形でHを求めるのなら，むしろ，次節で述べるNewton法でも使った方がよいということになるからである。

第2の問題は，primitive gradient methodの場合と一部共通の問題であるが，最大方向を求める際に用いた楕円体

$$(x - x^{(m)})^T B (x - x^{(m)}) = \tilde{\sigma}^2$$

のえらび方が問題である。目的函数の形が，定符号行列によって与えられる2次形式である場合には，図 I - 8 からも明らかなように，

$$B = H$$

とすれば，一回で，最大点にまっすぐ到達することができる。この場合，Hを一度に求めてしまわなくても

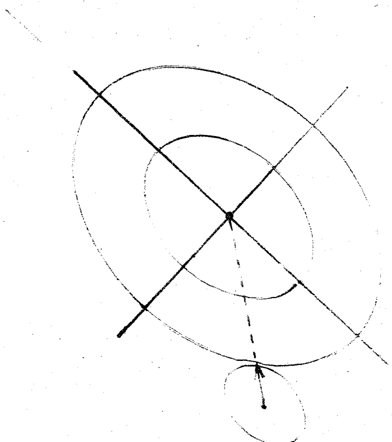


図 I - 8

近似の進行とともにBを変化させ，次第にHに近づけることも考えられる。

$$B^1, B^2, \cdots \longrightarrow H$$

また，目的函数が近似的に二次形式で表わされる場合でも，Bを変化させ，

$$B^{(m)} = H(x^{(m)})$$

とすれば，収束をはやめることができるであろう。しかし，これらの取扱いは，むしろ，次節で述べるNewton法の所の方が適当であろう。

Ⅱ　Newton 法

目的函数を点 $x^{(m)}$ を中心としてテーラー展開し，2階偏微係数を含む項までをとって近似しよう。

$$f(x) \approx f(x^{(m)}) + \mathrm{grad}\, f(x^{(m)}) \cdot (x - x^{(m)}) + \frac{1}{2}(x - x^{(m)})^T H(x^{(m)})(x - x^{(m)})$$

このとき，$H(x^{(m)})$ が正定符号であれば，右辺の函数に最小値を与える点は

$$x = x^{(m)} - H^{-1}(x^{(m)}) \cdot (\mathrm{grad}\, f(x^{(m)}))^T$$

で与えられる。従って，上の近似の精度が十分高く，しかも，H の正定符号性が確かめられゝば，目的函数の最小点もこの点の近くにあることが予想される。

そこで，これらの条件の下で

$$x^{(m+1)} = x^{(m)} - H^{-1}(x^{(m)})(\mathrm{grad}\, f(x^{(m)}))^T$$

という漸化式の作る点列によって，最小点への近似をすゝめて行くことができる。この方法を Newton 法という。(図Ⅱ-1)

この方法が，方程式の根を求める際に用いられる，いわゆる，Newton の近似法と同じものであることは明らかであろう。実際，連立方程式

$$(\mathrm{grad}\, f(x))^T = 0$$

10)

を解くのに，Newtonの近似法を適用してみれば，上と同じ漸化式が得られる。

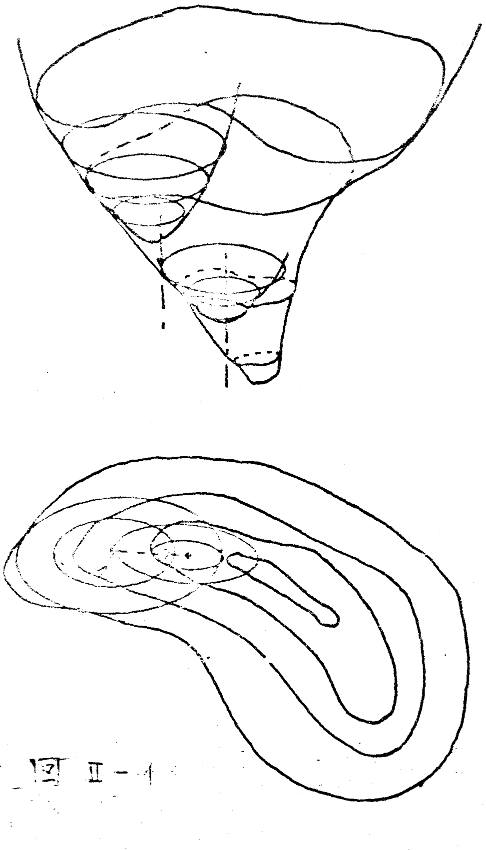
また，optimum gradient method に於いて，

$$B^{(m)} = H(x^{(m)})$$

とおいても，上と同じ漸化式が得られる。

このNewtonの方法には，勾配法の場合とは異る新しい問題がでてくる。

第1の問題は $H(x^{(m)})$ の正定符号性である。これが保証さ



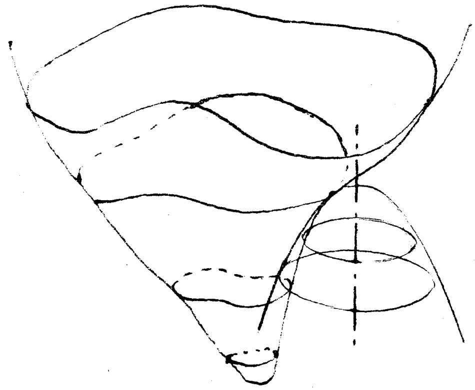


図 II-2

図 II-1

れていなければ，H^{-1}を計算できるかどうかも[illegible]ないし，仮に出来たとしても，$x^{(m)}$が鞍点上に停止してしまったり，振動を起してしまう可能性がでてくる。しかも，図Ⅱ-2からも明らかなように，1つの極小点をもつ函数であっても，Hが負定符号になったり，函数の作る曲面が鞍点をもち，Hについて，正負いずれの定符号性もいえなくなることが起りうるのである。

このような問題に取り組んでいるものとして，Goldfeld, Quand, Trotterの考え方がある〈24〉。彼等の提唱する方法は，目的函数が作る曲面の地形を調べて，適宜Newton法や，これを修正したものを使うものだが，以下にその理論の主要な点を述べる。

Goldfeld, Quandt, Trotterの研究

目的函数$f(x)$は2次函数とする。すなわち，点aを中心としてテーラー展開すれば

$$f(x)=f(a)+(\mathrm{grad}\,f(a))'(x-a)+\frac{1}{2}(x-a)^{T}H(x-a)$$

と書ける。こゝに，H は一定の(対称)行列であるが，Hの定符号性等は仮定しないでおく。

いま，

$$H-\alpha I \quad , \qquad I：単位行列$$

という行列を考えると，αを十分大きくとれば負定符号の行列になる。実際，Hの固有値（Hの対称性により実数）を

$$\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_N$$

とし，これに属する固有ベクトルを

$$u_1,\ u_2,\ \cdots,\ u_N \qquad \|u_i\| = 1 \quad i=1,\cdots,N$$

とすれば，明らかに，$H-\alpha I$の固有値，固有ベクトルは

$$\lambda_1-\alpha \geqq \lambda_2-\alpha \geqq \cdots \geqq \lambda_N-\alpha$$

$$u_1 \quad , \quad u_2 \quad , \quad \cdots \quad , \quad u_N$$

となるから，αを $\alpha > \lambda_1$ ととれば固有値はすべて負になる。以後，αはこのようにとる。

つぎに，点aを中心に次のような近傍を考える

$$B_\alpha = \{ x \mid \|x-a\| \leqq \|(H-\alpha I)^{-1}(\mathrm{grad}\, f(a))^T\| \}$$

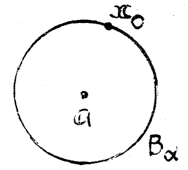
この球は $\alpha \in (\lambda_1, \infty)$ の範囲，すなわち，$H-\alpha I$ が負定符号になる範囲ではαの増加とともに小さくなる。（狭義の減少）

そこで，B_α内における函数$f(x)$の最大点をしらべてみると次のような結果が得られる。

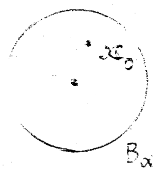
(1°) $\quad \mathrm{grad}\, f(a) \neq 0$ の場合

(イ) Hが負定符号でないとき（$\alpha > \lambda_1 \geqq 0$）

$$x_c = a - (H-\alpha I)^{-1}(\mathrm{grad}\, f(a))^T$$



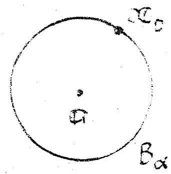
(ロ) Hが負定符号のとき,



(i) $\alpha < 0$ なら,

$$x_0 = a - H^{-1}(\mathrm{grad}\, f(a))^T$$

(ii) $\alpha \geq 0$ なら,

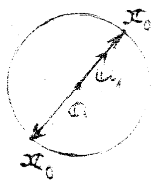


$$x_0 = a - (H - \alpha I)^{-1}(\mathrm{grad}\, f(a))^T$$

(2°) $\mathrm{grad}\, f(a) = 0$ の場合

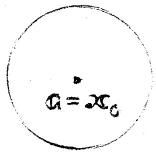
この場合には B_α は点 a だけになってしまうが, a を中心として任意の半径 $r > 0$ の球を考えれば, この球内での最大点は

(イ) $\lambda_1 > 0$ なら (aは鞍点または最小点,)



$$x_0 = a \pm r u_1$$

(ロ) $\lambda_1 \leq 0$ なら



$$x_0 = a$$

以上の結果をもとに, Goldfeld, Quandt, Trotterは実際に~~計算する~~最大点を求めるためのアルゴリズムとして次のような方法を提案している。(目的函数は2次函数とは限らない)

(a) $\|\mathrm{grad}\, f(x^{(m)})\| \gg 0$ の場合

$$\alpha = \lambda_1 + R\|\mathrm{grad}\, f(x^{(m)})\|$$

R : 正のパラメター

とおき, 点列 $\{x^{(m)}\}$ を次の漸化式で定義する。ここに, λ_1 は $H(x^{(m)}) = \left\{\frac{\partial^2 f}{\partial x_i \partial x_j}\right\}\Big|_{x = x^{(m)}}$ の最大固有値である。

$\alpha \geqq 0$ のとき，

$$x^{(m+1)} = x^{(m)} - (H(x^{(m)}) - \alpha I)^{-1} \cdot (\mathrm{grad}\, f(x^{(m)}))^T$$

$\alpha < 0$ のとき，

$$x^{(m+1)} = x^{(m)} - H^{-1}(x^{(m)}) \cdot (\mathrm{grad}\, f(x^{(m)}))^T$$

すなわち，目的函数の曲面をしらべて，その点の附近で函数が凹である場合（$\alpha < 0$）には，通常のNewton法を，これがいえない場合には，目的函数の$H(x^{(m)})$を$H(x^{(m)}) - \alpha I$という行列におきかえてNewton法を適用するわけである。

(b)　$\|\mathrm{grad}\, f(x^{(m)})\| \approx 0$　の場合

$H(x^{(m)})$の固有値

$$\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_N$$

を調べ

$\lambda_1 < 0$　のとき

$H(x^{(m)})$は負定符号であるから$x^{(m)}$はほゞ最大点に到達している。

$\lambda_1 \geqq 0$　のとき

$x^{(m)}$は鞍点である。そこで

$$x^{(m+1)} = x^{(m)} + \kappa u_1$$

とする。こゝにu_1はλ_1に属する固有ベクトル，κは正のパラメターである。

第2の問題は，$H(x^{(m)})$ とその逆行列 $H^{-1}(x^{(m)})$ の計算である。これらは普通，可成りの計算量になるので，できれば避けたい。

このためにも，いろいろなアイデアが出されている。たとえば，$H(x^{(m)})$ と $H(x^{(m)})^{-1}$ を一度計算したら，しばらくの間（$m+1, m+2, \cdots, m+\ell$）は，近似的に，これらをそのまゝ使おうとするのもある。また，次に述べる quasi-Newton 法の考え方もある。

<u>Quasi Newton 法</u>

これは，通常の Newton 法で，一挙に H や H^{-1} を計算しようとせず，適当な定符行列 $G^{(0)}$ から始めて，計算の途中で得た，目的函数に関する情報を使って，次第に H^{-1} に近づく行列の列

$$G^{(0)}, G^{(1)}, \cdots, G^{(m)}, \cdots \longrightarrow H^{-1}$$

を作ろうというのである。

特に，目的函数の形が本当に定符号行列によって与えられる2次曲面である場合には，H^{-1} が x に依存せず，一意的に定まっているから話ははっきりしている。また，見方をかえると，optimum gradient method の所で述べた，楕円体を与える行列 B を変化させて行く方法と，この方法は，実は，

同じものになるのである。

もちろん，こゝでも，点列 $x^{(m)}$ の収束性の問題がでてくる。しかし，この場合には，理論上，有限回で H^{-1} に合致するような行列の列 $\{G^{(m)}\}$ を作るアルゴリズムが知られている。だから，これに従えば，計算が誤差をともなわぬかぎり，$x^{(m)}$ も有限回で最大点や最小点に合致することになる。

前にもふれたが，目的函数が定符号行列によって与えられる2次曲面である場合に，$x^{(m)}$ が有限回で最大点や最小点に合致する性質があるとき，アルゴリズムは'exact'あるいは'finite quadratically convergent'といわれる。この性質は，実用上も大事なのでアルゴリズムを設計する上で重要視されている。

Quasi-Newton 法とよばれるものにも，いくつかの形があるようだがその詳細は参考文献<25>にゆずり，こゝでは，Davidon<6>によって提案され，Fletcher-Powell<46>によって整理されたものをあげておこう。

Fletcher Powell の方法

この方法は次のような手順によって点列 $x^{(m)}$ を与え，目的関数 $f(x)$ の最小点 $\hat{x}$ に近づこうとするものである。

1) $G^{(0)}=$ 任意の正定符号対称行列

2) $x^{(0)}=$ 初期点，任意

3) $s^{(m)}=-G^{(m)}g^{(m)}$ とする。こゝに

$$g^{(m)}=(\mathrm{grad}\,f(x^{(m)}))^T$$

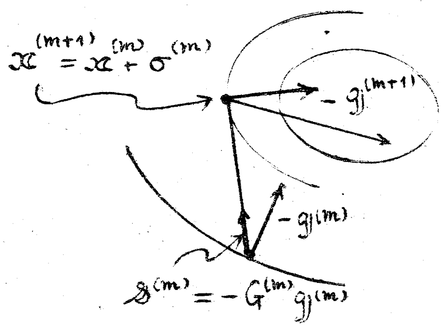
4) $x^{(m)}$ を始点とし，$s^{(m)}$ 方向を向いた半直線上で，目的函数の値が最も小さくなる点を求める。すなわち，

$$f(x^{(m)}+\alpha^{(m)}s^{(m)})=\min_{\lambda} f(x^{(m)}+\lambda s^{(m)})$$

を計算し，

$$\sigma^{(m)}=\alpha^{(m)}s^{(m)}$$

とおく。



5) $x^{(m+1)}=x^{(m)}+\sigma^{(m)}$

6) $G^{(m+1)}=G^{(m)}+A^{(m)}+B^{(m)}$

$$A^{(m)}=\sigma^{(m)}\cdot\sigma^{(m)T}/(\sigma^{(m)T}y^{(m)})$$

$$B^{(m)}=-(G^{(m)}\sigma^{(m)})(\sigma^{(m)T}G^{(m)})/(y^{(m)T}G^{(m)}y^{(m)})$$

$$y^{(m)}=g^{(m+1)}-g^{(m)}$$

116

このような手順で行列$G^{(m)}$を定めて行くと

$$G^N = H^{-1} \qquad (N：変数の数)$$

となることが証明されているが，これは文献<16>にゆずろう。また，この方法はHilbert空間でも全く同様に使えること，線形部分空間での最小点を求めるのにも，多少の修正をすれば使えることが最近Horwitz-Sarachikによって示されている<33>。

Ⅲ One at a time method

いさゝか妙な名前ではあるが，この方法の最も原始的な形は次のようなものである。

目的函数

$$f(x)=f(x_1,\cdots,x_N)$$

の最小点に近づく点列を作るのに，これまで述べてきた方法のように，grad f とか，$H=\left\{\dfrac{\partial^2 H}{\partial x_i \partial x_j}\right\}$ のような，局所的な情報は使わず，N個の単位ベクトル

$$\mathbb{P}_1,\cdots,\mathbb{P}_N$$

をあらかじめ定めておき，考えている点から，これらの方向に沿って，順ぐりに最小点を求めて行くのである。

すなわち

$$f(x^{(m,1)})=\min_{\xi} f(x^{(m-1,N)}+\xi\mathbb{P}_1)$$

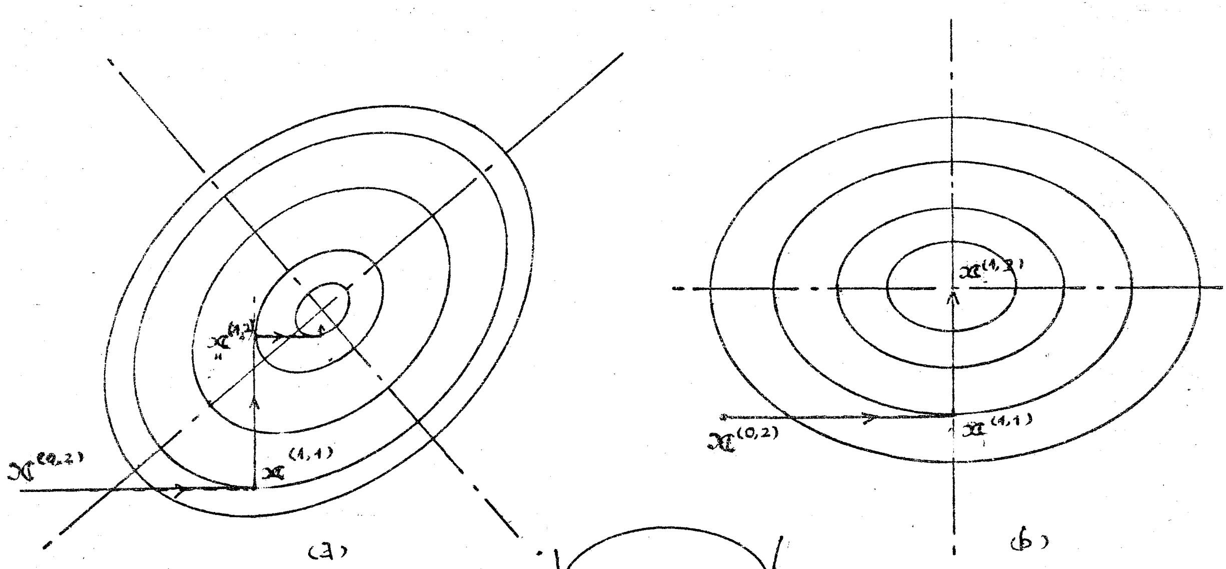
$$f(x^{(m,2)})=\min_{\xi} f(x^{(m,1)}+\xi\mathbb{P}_2)$$

$$\cdots\qquad\cdots$$

$$f(x^{(m,N)})=\min_{\xi} f(x^{(m,N-1)}+\xi\mathbb{P}_N)$$

という形で点列を定義して行くのである。

ベクトル $\mathbb{P}_1, \mathbb{P}_2, \cdots, \mathbb{P}_N$ のえらび方としては，座標軸方向にとるのが最も初等的なやり方である。実用上この方法で成功した例は少くないが，理論上は，収束性の点でかなり問題がある。このことは，図Ⅲ－1（うまく行く場合）と図Ⅲ



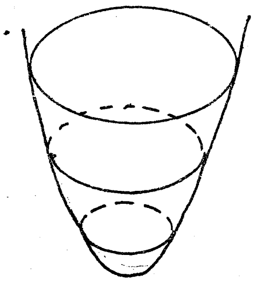


図 Ⅲ-1

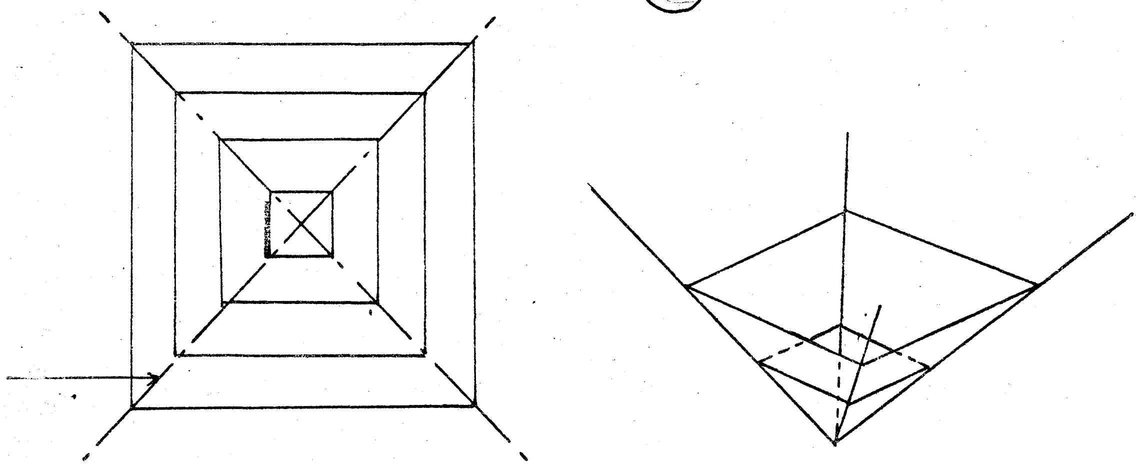
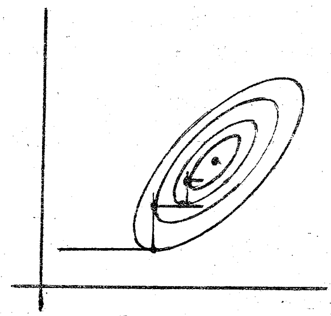


図 Ⅲ-2

－2（うまく行かない場合）を対比してみれば明らかであろう。一般的にいって，図Ⅲ－2にあるような，鋭い稜線があるときには，具合がわるい。目的函数が2次形式で書ける場合にも図Ⅲ－3のような細長い楕円体の等高線（面）があらわれる場合には，点列$x^{(m)}$はなかなか最小点に近づかない。ディジタル計算機で計算を行う場合には，量子化がなされるのでますます具合のわるいことになる。



図Ⅲ－3

また，この方法は連立方程式を解くGauß-Seidel法に対応しているので，点列を定義する漸化式が書ければ，Seidel過程の収束性吟味の方法を適用することができる。しかし，この考え方では，漸化式そのものの改良法はあまり得られない。

一方，図Ⅲ－1（b）のように，等高線の形によっては，有限回で最小点に到達してしまう。図Ⅲ－1（a）の楕円と図Ⅲ－2（b）の楕円は，実は，同じ形のものなので，このことは

$$\mathbb{P}_1, \cdots, \mathbb{P}_N$$

のえらび方が，収束性に大きく影響することを示している。

目的関数が，2次形式である場合には，$\mathbb{P}_1, \cdots, \mathbb{P}_N$をどのようにとればよいかという点については次のことが知ら

れている。

<u>共役ベクトル</u>

正定符号対称行列 $\underset{(N\times N)}{H}$ を考える。2つのベクトル $\mathbb{p}, \mathbb{q}$ のこの行列に関する双一次形式

$$\mathbb{p}^T H \mathbb{q}$$

がゼロになるとき，ベクトル $\mathbb{p}$ および $\mathbb{q}$ は互に共役であるといわれる。

いま，$n\ (\leqq N)$ 個の互に共役なベクトル

$$\mathbb{p}_1, \mathbb{p}_2, \cdots, \mathbb{p}_n$$

をとり，

$$f(x^{(1)}) = \min_{\xi} f(x^{(0)} + \xi \mathbb{p}_1)$$
$$f(x^{(2)}) = \min_{\xi} f(x^{(1)} + \xi \mathbb{p}_2)$$
$$\cdots$$
$$f(x^{(n)}) = \min_{\xi} f(x^{(n-1)} + \xi \mathbb{p}_n)$$

という手順によって $x^{(n)}$ を求めると，$x^{(n)}$ は，$\mathbb{p}_1, \cdots, \mathbb{p}_n$ の張る部分空間内での，2次形式

$$x^T H x$$

の最小点になる。

そこで，何等かの方法によって変数の数(=N)だけの線型独立で互に共役なベクトル

$$\mathbb{P}_1, \cdots, \mathbb{P}_N$$

が得られゝば，N回のiterationによって最小点に到達することができる。

しかし，共役であるかいなかを決定するのは，目的函数自身の性質であるから，結局のところ，計算の過程でこのようなベクトルをさがして行かなければならない。すなわち，最初，線形独立なベクトルの組

$$[\, \mathbb{P}_1 \,\vdots\, \mathbb{P}_2 \,\vdots\, \cdots \,\vdots\, \mathbb{P}_N \,]$$

から出発して，線形独立であると同時に共役なベクトルの組

$$[\, \mathbb{P}_1^{(M)} \,\vdots\, \mathbb{P}_2^{(M)} \,\vdots\, \cdots \,\vdots\, \mathbb{P}_N^{(M)} \,]$$

が得られるような過程を考えればよい。

このような過程として提案されているものがいくつかある。<3>,<14>,<25><34><30>また，上のベクトルの組をマトリクスとしてとらえればQuasi-Newton法とも関係がつく。

こゝでは，Zangwill <30> の手順をあげておく。

Zangwill の方法

この方法は，任意の相異るN個のベクトルから出発して，最小化の過程をくりかえしながら，互に共役なベクトルを作るもので，直接目的函数の偏微係数を求める手続きは含まれていない。目的函数が二次形式である場合を想定して作られた方法だが，その他の場合にも使えないことはない。(もちろんこの場合には最小点に有限回のステップで到達できるとはかぎらない。)手順は次の通りである。

(i) 座標軸方向の単位ベクトルを

$\mathbb{e}_i , \ i = 1, \cdots, N$

とする。

(ii) 第0ステップ

初期点 $x_N^{(0)}$

Nコの相異る単位ベクトル $\mathbb{P}_i^1 , \ i = 1, \cdots, N$

を与える。

$$f(x_N^{(0)} + \lambda_N^{(0)} \mathbb{P}_N^1) = \min_{\lambda} f(x_N^{(0)} + \lambda \mathbb{P}_N^1)$$

なる $\lambda_N^{(0)}$ を求める。

$$x_{N+1}^{(0)} = x_N^{(0)} + \lambda_N^{(0)} \mathbb{P}_N^{(1)}$$

とする。

$t = 1 , \ k = 1$ として第kステップへ

(iii) 第 k ステップ

第 k ステップへ入るときには

$$x_{N+1}^{(k-1)},\ \mathbb{P}_r^{(k)}\ (r=1,\cdots,n),\ t$$

が与えられている。

(イ) $$f(x_{N+1}^{k-1}+\alpha\,\mathbb{e}_t)=\min_{\mu} f(x_{N+1}^{k-1}+\mu\,\mathbb{e}_t)$$

なる α を求める。

$1\leqq t<n$ のとき $t\to t+1$

$t=n$ 〃 $t=n$

とする。

(ロ) (a) $\alpha\neq 0$ の場合

$$x_0^k=x_{N+1}^{k-1}+\alpha\,\mathbb{e}_t$$

とする。(ハ)へ

(b) $\alpha=0$ の場合

(イ)の最初へもどる。(イ)の手順がN回くりかえされゝば x_{N+1}^{k-1} は最小点。

(ハ) $r=1,\cdots,N$ について

$$f(x_{r-1}^k+\lambda_r^k\mathbb{P}_r^k)=\min_{\mu} f(x_{r-1}^k+\mu\mathbb{P}_r^k)$$

なる λ_r^k を求め，

$$x_r^k=x_{r-1}^k+\lambda_r^k\mathbb{P}_r^k$$

とする。また，

$$\mathbb{P}_{N+1}^k=(x_N^k-x_{N+1}^{k-1})/\|x_N^k-x_{N+1}^k\|$$

とする。

$$f(\mathbb{x}_N^k+\lambda_{N+1}^k \mathbb{P}_{N+1}^k)=\min_{\mu} f(\mathbb{x}_N^k+\mu \mathbb{P}_{N+1}^k)$$

なる λ_{N+1}^k を求め

$$\mathbb{x}_{N+1}^k = \mathbb{x}_N^k+\lambda_{N+1}^k \mathbb{P}_{N+1}^k$$

とし，

$$\mathbb{P}_r^{k+1}=\mathbb{P}_{r+1}^k \qquad r=1,\cdots,N$$

とおき，$k+1$ ステップへ

図Ⅲ-4にN=2の場合の手順を等高線図上に示す。

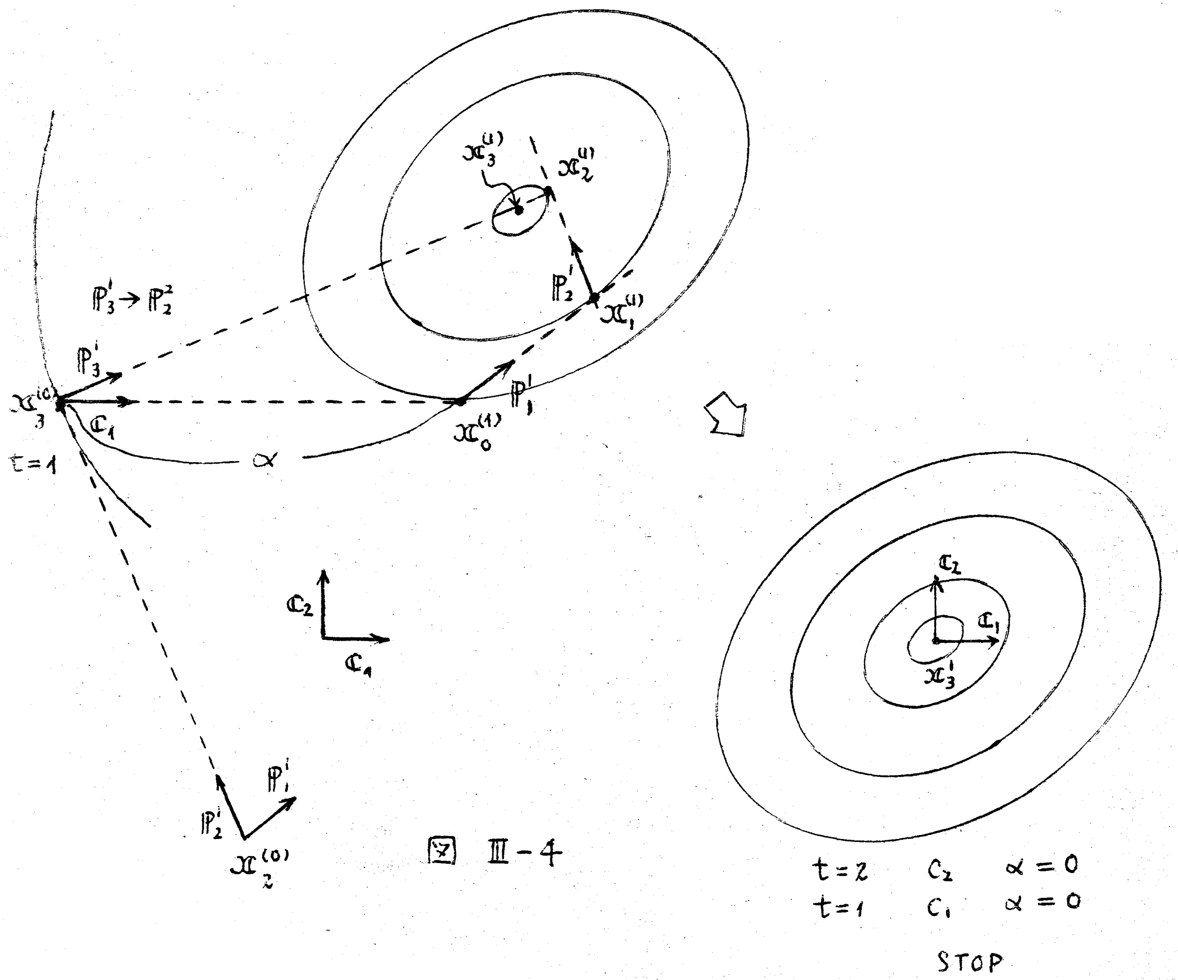


図 Ⅲ-4

Simplex 法

これは，線形計画の問題を解くのに用いられる有名なシンプレックス法とは別のもので，むしろ，非線形の連立方程式を解く際に用いられるsecant method〈5〉に対応するものである。

こゝでいうシンプレックス法は，N変数 $x_1, x_2, \cdots, x_N$ のつくるN次元空間の適当な位置に，適当な大きさのN次元シンプレックス（一次独立なN＋1個の点を頂点とする図形）をとり，その各頂点における目的函数の値を比較して，これが最も小さい値をとる頂点を，函数がもっと大きな値をとる新しい点と入れかえて，新しいシンプレックスを作る。この手順をくりかえして，シンプレックスを‘山の頂上’に近づけるのである。

シンプレックス法の手順は次の通りである。

(i) 変数 x のN次元空間の適当な位置に，適当な大きさのN次元のシンプレックス

$$S^{(0)} = \{ x_1^{(0)}, \cdots, x_{N+1}^{(0)} \}$$

をとる。

(ii) 第 m 番目のシンプレックス

$$S^{(m)} = \{ x_1^{(m)}, \cdots, x_{N+1}^{(m)} \}$$

の各頂点における函数の値 $f(x_1^{(m)}), f(x_2^{(m)}), \cdots, f(x_{N+1}^{(m)})$

を比較して次の3つのものを定める。

(イ) 函数の値が最も大きい頂点，$x_h^{(m)}$

$$y_h^{(m)} \equiv f(x_h^{(m)}) = \max_i \{f(x_i^{(m)})\}_{i=1}^{N+1}$$

(ロ) 函数の値が最も小さい頂点

$$y_l^{(m)} \equiv f(x_l^{(m)}) = \min_i \{f(x_i^{(m)})\}_{i=1}^{N+1}$$

(ハ) 函数の値が2番目に小さい頂点，

$$y_s^{(m)} \equiv f(x_s^{(m)}) = \min_{i \neq l} \{f(x_i^{(m)})\}_{i=1}^{N+1}$$

(iii) $x_l^{(m)}$を除くN個の頂点，の中心$x_m^{(m)}$を求める。

$$x_m^{(m)} = \frac{1}{N} \sum_{i \neq l} x_i^{(m)}$$

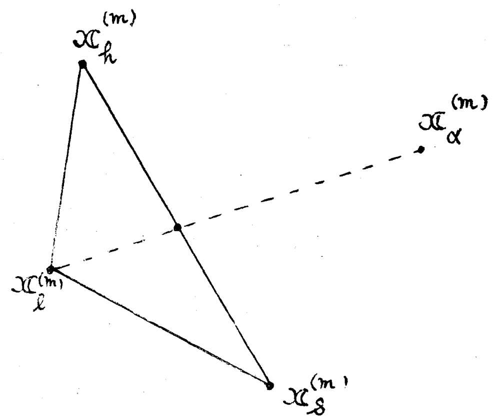
(iv) $x_l^{(m)}$と$x_m^{(m)}$をむすぶ直線上，$x_m^{(m)}$とは反対側に点，

$$x_\alpha^{(m)} = (1+\alpha)\, x_m^{(m)} - \alpha\, x_l^{(m)}\,, \qquad \alpha > 0$$

をとる。こゝに，αはパラメター（適当にえらぶ）であるが，

$$\alpha = \overline{x_m^{(m)} x_\alpha^{(m)}} \Big/ \overline{x_m\, x_l^{(m)}}$$

という関係をみたしている。



この点における函数の値を

$$y_\alpha^{(m)} = f(x_\alpha^{(m)})$$

とする。この $y_\alpha^{(m)}$ を(ii)で求めた $y_l^{(m)}, y_s^{(m)}, y_h^{(m)}$ と比較すると次の4つの場合がある。

$y_l^{(m)}$　　$y_s^{(m)}$　　$y_h^{(m)}$　　y

(ニ) $y_\alpha^{(m)} < y_l^{(m)}$ (ハ) $y_l^{(m)} \leqq y_\alpha^{(m)} < y_s^{(m)}$ (ロ) $y_s^{(m)} \leqq y_\alpha^{(m)} \leqq y_h^{(m)}$ (イ) $y_h^{(m)} < y_\alpha^{(m)}$

これらの場合に応じて，シンプレックス $S^{(m)}$ を次のように改める。

(イ) $y_h^{(m)} < y_\alpha^{(m)}$

この場合には，函数は $x_\alpha^{(m)}$ 方向に向って非常に急な登り坂になっているものと考えられる。そこで，この線をさらに延長して，点

$$x_\gamma^{(m)} = \gamma x_\alpha^{(m)} + (1-\gamma) x_m^{(m)}, \quad \gamma > 1$$

を定める。こゝに，γ はパラメター（適当にえらぶ）であるが，

$$\gamma = \overline{x_m^{(m)} x_\gamma^{(m)}} / \overline{x_m^{(m)} x_\alpha^{(m)}}$$

という関係をみたしている。

点 $x_\gamma^{(m)}$ における函数の値を計算し，

(a) $f(x_\gamma^{(m)}) > y_h^{(m)}$ の場合には $x_l^{(m)}$ をすてゝ $x_\gamma^{(m)}$ をとり新しいシンプレックス $S^{(m+1)}$ をつくる。

(b) $f(x_{\gamma}^{(m)}) \leqq y_{h}^{(m)}$ の場合には，$x_{l}^{(m)}$をすてゝ，$x_{\alpha}^{(m)}$をとり，新しいシンプレックス$S^{(m+1)}$をつくる。

(ロ) $y_{s}^{(m)} \leqq y_{\alpha}^{(m)} \leqq y_{h}^{(m)}$

この場合には，$x_{l}^{(m)}$をすてゝ$x_{\alpha}^{(m)}$をとり，新しいシンプレックス$S^{(m+1)}$を作る。

(ハ) $y_{l}^{(m)} \leqq y_{\alpha}^{(m)} < y_{s}^{(m)}$

この場合には2点，$x_{\alpha}^{(m)}$，$x_{m}^{(m)}$の間に次のような点$x_{\beta}^{(m)}$をとる。

$$x_{\beta}^{(m)} = \beta x_{\alpha}^{(m)} + (1-\beta) x_{m}^{(m)}$$

こゝに，βはパラメター（適当にえらぶ）であるが，

$$\beta = \overline{x_{m}^{(m)} x_{\beta}^{(m)}} / \overline{x_{(m)}^{(m)} x_{\alpha}^{(m)}}$$

という関係をみたしている。

点$x_{\beta}^{(m)}$における函数の値を計算し，

(a) $f(x_{\beta}^{(m)}) > y_{s}^{(m)}$ の場合には，$x_{l}^{(m)}$をすてゝ$x_{\beta}^{(m)}$をとり，新しいシンプレックス$S^{(m+1)}$をつくる。

(b) $f(x_{\beta}^{(m)}) \leqq y_{s}^{(m)}$ の場合には，シンプレックスが大きすぎると考えられるので，$x_{h}^{(m)}$を中心としてシンプレックスを次のように縮小す

る。すなわち

$$x_i^{(m+1)}=(x_h^{(m)}+x_i^{(m)})/2$$

$$x_h^{(m+1)}=x_h^{(m)}$$

を頂点とするシンプレックスを新しいシンプレックス $S^{(m+1)}$ とする。

(ニ)　$y_\alpha^{(m)}<y_l^{(m)}$

この場合には，2点 $x_l^{(m)}$, $x_m^{(m)}$ の間に次のような点 $x_{\beta'}^{(m)}$ をとる。

$$x_{\beta'}^{(m)}=\beta' x_l^{(m)}+(1-\beta')x_m^{(m)} \qquad 0<\beta'<1$$

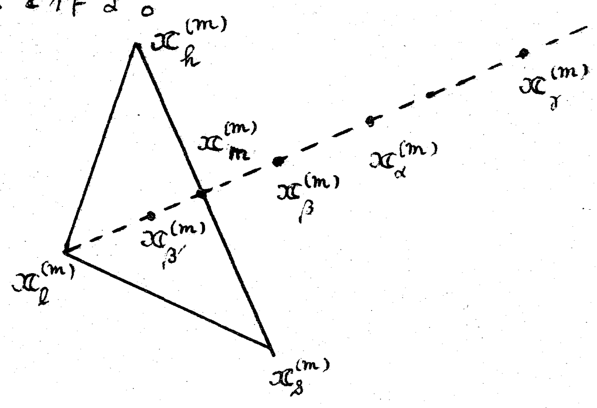
こゝに，β' はパラメター（適当にえらぶ）であるが

$$\beta'=\overline{x_m^{(m)}x_{\beta'}^{(m)}}/\overline{x_m^{(m)}x_l^{(m)}}$$

という関係をみたしている。

点 $x_{\beta'}^{(m)}$ における函数の値を計算し，

(a)　$f(x_{\beta'}^{(m)})>y_s$ の場合には，$x_l^{(m)}$ をすてて，$x_{\beta'}^{(m)}$ をとり，新しいシンプレックスを作る。



(ロ)　$f(x_j^{(m)}) \leq y_j^{(m)}$ の場合　には(い),(b)でのべた手続きによってシンプレックスを縮小する。

(v)　$S^{(m+1)}$ が得られたら，各頂点における函数の値を計算し，そのバラツキ

$$\frac{1}{N}\sum_{i=1}^{N+1}\left(f(x_i^{(m+1)})-\frac{1}{N}\sum_{j=1}^{N+1}f(x_j^{(m+1)})\right)^2$$

を調べ，これが，予め定められているある定数より小さくなっていたら計算を停止する。

以上の手順は，計算の都合上とりきめた部分が多いが，本質的な部分は，要するに，最初の所でもふれたように，シンプレックスを，函数の値が大きくなる方向に動かして行くことである。たゞ，注意すべきことは，シンプレックスが計算の途中で縮退しないようにすることである。(上の手順では，その心配がないようになっている。)

附録 I Test Functions

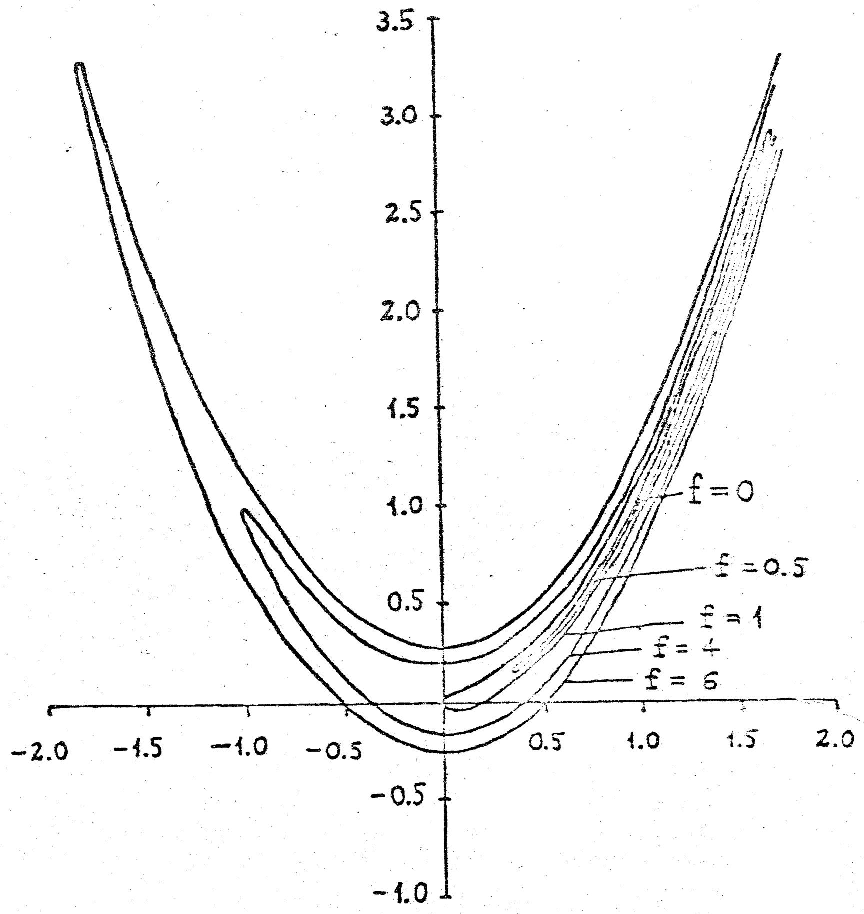
最適化手法の評価のため，標準問題として使われたTest functionのうち主なものを次に列挙する。

(1) Parabolic Valley, Rosenbrock's curved valley <7>

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

最小点および最小値 $f(1,1) = 0$

等高線図



(2) Helical valley <16>

$$f(x_1, x_2, x_3) = 100(x_3 - 10\theta)^2 + (r-1)^2 + x_3^2$$

$$x_1 = r\cos 2\pi\theta$$

$$x_2 = r\sin 2\pi\theta$$

(3) Function of 4 variables by Powell <11>

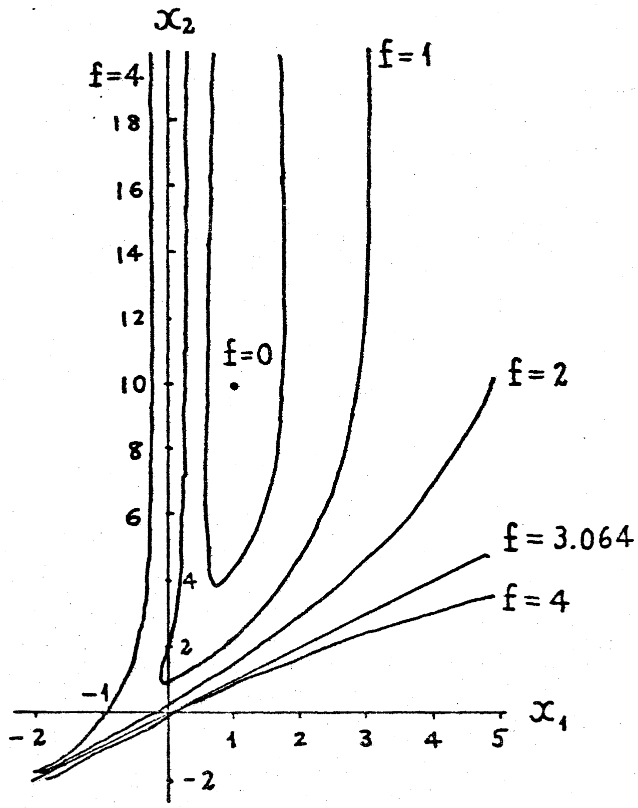
$$f(x_1, x_2, x_3, x_4) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$$

(4) 2 dimensional exponential function <21>

$$f(x_1, x_2) = \sum_a [(e^{-ax_1} - e^{-ax_2}) - (e^{-a} - e^{-10a})]^2$$

$$a = 0.1\,(0.1)\,1$$

最小点および最小値:　$f(1, 10) = 0$



(5) 3 - dimensional exponential function <21>

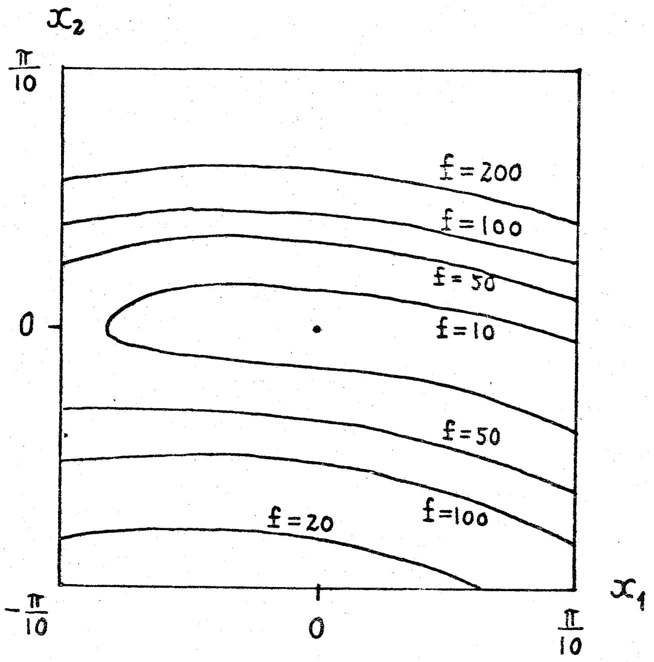
$$f(x_1, x_2, x_3) = \sum_a [(e^{-x_1 a} - e^{-x_2 a}) - x_3(e^{-a} - e^{-10a})]^2 \quad a = 0.1\,(0.1)\,1$$

最小点および最小値

$$f(1, 10, 1) = 0$$

(6) 2 - dimensional trigonometric function <21>

$$f(x_1, x_2) = \sum_{i=1}^{2} [E_i - \sum_{j=1}^{2} (A_{ij} \sin x_j + B_{ij} \cos x_j)]^2$$



(7) N - dimensional trigonometric functions <21>

$$f(x_1, \cdots, x_N) = \sum_{i=1}^{N} [E_i - \sum_{j=1}^{N} (A_{ij} \sin x_j + B_{ij} \cos x_j)]^2$$

(N = 5, 10, 20)

(8) Functions with saddle points (Crater function)<24>

(i) $f(x_1, x_2) = e^{-x_1 - x_2}(2x_1^2 + 3x_2^2)$

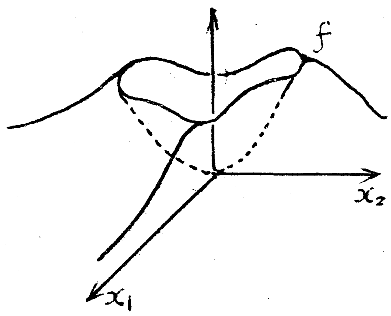
最大点および最大値　　$f(0,1) = 3e^{-1}$

　　　　　　　　　　　$f(0,-1) = 3e^{-1}$

鞍点およびその値　　　$f(1,0) = 2e^{-1}$

　　　　　　　　　　　$f(-1,0) = 2e^{-1}$

最小点および最小値　　$f(0,0) = 0$



(ii) $f(x_1, \cdots, x_5) = e^{-\sum_{i=1}^{5} x_i^2}(3.0x_1^2 + 2.0x_2^2 + 3.5x_3^2 + 4.0x_4^2 + 2.7x_5^2$

附録 II　参考文献

1> Cauchy,A.L. < Méthode Générale pour la Résolution des Systémes d'Équations Simultanée, C.R.Read Sci.,Paris,25,536-538,1847

2> Curry,H.B. < The Method of Steepest Descent for Non-Linear Minimization Problems > Quarterly of Applied Mathematics Vol.2,pp258-260 ,1944

3> Hestenes,M.R. and Stiefel,E. < Method of Conjugate Gradients for Solving Linear Systems > Journal of the Research of the National Bureau of Standards, Vol.49, No.6 pp409-436,1952

4> Crockett,J.B. and Chernoff, H. < Gradient Methods of Maximization > Pacific Journal of Mathematics,Vol.5,pp 33-50,1955

5> Wolfe,P. < The Secant Method for Simultaneous Nonlinear Equations > Communication of the ACM, Vol.2,pp 12-13,1959

6> Davidon,W.C. < Variable Metric Method for Minimization > Research and Development Rep.ANL-5990, Atomic Energy Commission 1959

7> Rosenbrock, H. H. < An Automatic Method for finding the Greatest or Least Value of a Function > The Computer Journal, Vol.3, pp 175-184, 1960-1

8> Martin,D.W. and Tee,G.J. < Iterative Methods for Linear Equations with Symmetric Positive Definite Matrix > The Computer Journal Vol.4,pp 242-254 ,1961

9> Spang,III,H.A. < A Review of Minimization Techniques for Nonlinear Functions > SIAM Review, Vol.4,pp343-365, 1962

<10> Bellman,R.E. and Dreyfus, S.E. < Applied Dynamic Programming >, Princeton Univ. Press,1962

<11> Powell,M.J.D.< An Iterative Method for finding Stationary Values of a Function of several Variables > The Computer Journal, Vol.5,pp 147-151,1962

<12> Kantrovich ,L.V. and Akilov,G.P. < Functional Analysis in Normed Spaces >, Pergamon Press, 1964

<13> Wilde,D.J. <Optimal Seeking Methods> Prentice Hall, 1964

<14> Fletcher ,R. and Reeves,C.M. < Function Minimization by Conjugate Gradients > The Computer Journal,Vol.7, pp 149-154,1964

<15> Powell,M.J.D. < An Efficient Method for finding the Minimum of a Function of Several Variables without calculating Derivatives.> The Computer Journal, Vol.7,pp 155-162 1964

<16> Fletcher ,R. and Powell, M.J.D. < A Rapidly Convergent Descent Method for Minimization > The Computer Journal, Vol.6 pp163-168, 1964

<17> Fletcher,R. < Function Minimization without evaluating Derivatives --- a Review > The Computer Journal, Vol.8, No.1 pp33-41,1965

<18> Nelder, J.A. and Mead,R. < A Simplex Method for Function Minimization > The Computer Journal,Vol. 7, No.4,pp 308-312 1965

<19> Faure,P. et Huard, P. <Resolution de Programmes Mathematiques a Fonction Non Lineare par la Method du Gradient Reduit > Revue Française de Recherche Operationelle,N°36,pp167-206, 1965

<20> Box,M.J. < A New Method of Constrained Optimization and a Comparison with other Methods > The Computer Journal, Vol. 8, No.1, pp 42-52,1965

<21> Box,M.J. < A Comparison of Several Current Optimization Methods, and the use of Transformations in Constrained problems >

<22> 柳井 浩 < 最大最小の探索の方法 >, 数学セミナー 1965,(4),(6), 1966 (1),(2)

<23> Greenstadt,J.L. < A Ricocheting Gradient Method for Nonlinear Optimization > ,SIAM, Appl.Math.,.Vol.14,No3,pp429-445, 1966

<24> Goldfeld,S.M. and Quandt,R.E. & Trotter,H.F. < Maximization by Quadratic Hill Climbing> Econometrica, Vol.34, No.3 pp 541-551,1966

<25> Broyden,C.G. < Quasi-Newton Method and their Application to Function Minimization > Mathematics of Computation,Vol. 21,pp 368-381 1966

<26> Greenstadt,J. < On the Relative Efficiencies of Gradient Methods> Mathematics of Computation,Vol.21, pp 360-367,1966

<27> Burrows,J.W. < Maximization of a Second-Degree Polynomial on the Unit Sphere> Mathematics of Computation,Vol.20, No.95,pp441-444,1966

<28> Wilde,D.J. and Avriel,M. < Optimal Search for a Maximum with Sequences of Simultaneous Function Evaluations > Management Science, Vol.12,No.9,pp722-731, 1966

<29> Lavi,A. and Vogl,T.P. (ed) < Recent Advances in Optimization Technique >, John Wiley & Sons, Inc., 1966

<30> Zangwill,W.I. < Minimizing a Function without Caluculating Derivatives > The Computer Journal,Vol.10,No.3,pp 293-296, 1967

<31> Stewart III,G.W. < A Modification of Davidon's Minimization Method to Accept Difference Approximation of Derivatives > Journal of the ACM,Vol.14,pp72-83 ,1967

<32> 柳井　浩 < 最適値問題とその定式化 > オペレーションズ・リサーチ, 1968, (1)

<33> Horwitz,L.B. and Sarachik,P.E. < Davidon's Method in Hilbert Space > SIAM J. Appl.Math. Vol.16,no.4, pp 676-695, 1968

<34> Zeleznik,F.J. < Quasi-Newton Method for Nonlinear Equations > Journal of the ACM, Vol.15,No.2, pp 265-271,1968

<35> Colville,A.R. < A Comparative Study on Nonlinear Programming Codes > IBM New York Scientific Center Report No.320-2949, 1968

<36> Wilde,D.J.and Avriel,M. < Golden Block Search for the Maximum of Unimodal Functions > Management Science,Vol.14,No.5, pp 307-319,1968

<37> Duffin,R.J. and Karlovitz,L.A. < Formulation of Linear Programs in Analysis I: Approximation Theory >, SIAM J.Appl.Math. Vol.16,No.4,July,1968

<38> Luenberger,D.G. < Convex Programming and Duality in Normed Space > IEEE Transaction on Systems Science and Cybernetics, Vol.SSC-4,No.2,1968

<39>