ニュートン法

はじめに

本稿では工学で用いられるツールの一つ，ニュートン法の使い方について述べる．ニュートン法の応用は幅広く，十分な解説のない専門書も多い．筆者が考えるところではモデル化には大きく2つの場面があり，モデルによりいずれかを選択することとなる．本稿では，それらを重点に解説する．

1変数のニュートン法

1変数のニュートン法とは(1)を満たすxを求める手法である．

$$f(x) = 0 \quad (1)$$

具体的には適当な初期値$x^{(0)}$から始めて(2)の漸化式で解を求める反復法である．

$$x^{(k+1)} = x^{(k)} - \frac{f\left(x^{(k)}\right)}{f'(x^{(k)})} \quad (2)$$

この様子を**図1**に示す．

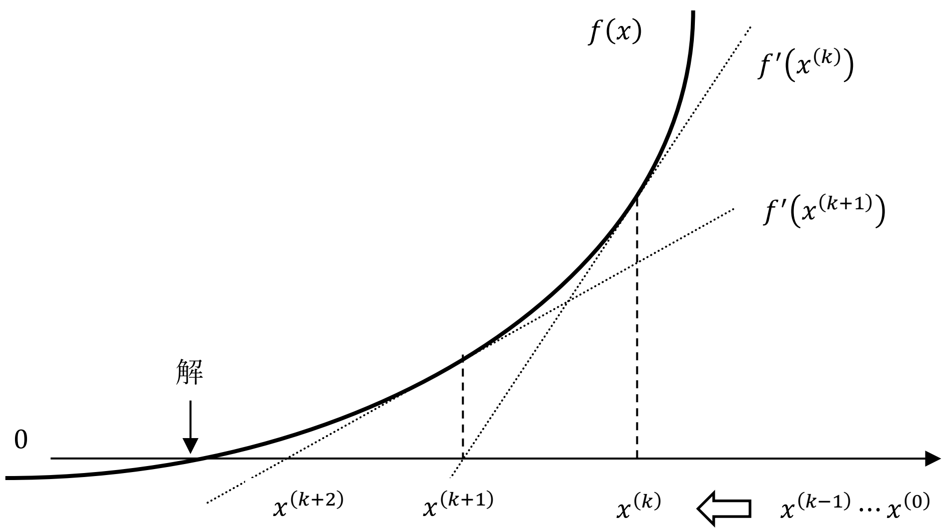

図1　1変数のニュートン法

漸化式(2)だが，これは$x = x^{(k)}$における接線の方程式(3)より導出できる．

$$y - f\left(x^{(k)}\right) = f'\left(x^{(k)}\right)\left(x - x^{(k)}\right) \quad (3)$$

(3)へ$y = 0, x = x^{(k+1)}$を代入することで(4)となる．

$$-f\left(x^{(k)}\right) = f'\left(x^{(k)}\right)\left(x^{(k+1)} - x^{(k)}\right) \quad (4)$$

変形すれば(4)は(2)で間違いない．専門書によっては以上の内容にとどまる．現実にこれだけでは，ほとんど利用価値のないものである．もちろん，第 1 の場面として(1)を直接扱える場面も存在する．しかし，それ以上に有用な場面があり，それこそが残差の 2 乗和の最小化(あるいは最大化)である．最小化と最大化については符号を反転すれば済むことなので，本稿では以降,「最小化」で統一して述べる．

さて，第 2 の場面である．まず関数fが何らかの関数の導関数であったとする(5)．

$$f(x) = r'(x) \quad (5)$$

すると, (1)は関数rの極値を求める問題とできる．さらに，rが 2 次の関数であれば極値の存在，さらには解の一意性まで保証される．例えば 2 次関数y(6)の係数を求める問題だが，通常なら近似曲線を求める問題は最小 2 乗法が用いられる．このとき，残差の 2 乗和は(7)で定義される．

$$y = ax^2 + bx + c \quad (6)$$

$$r = \frac{1}{2}\sum_{n=1}^{k}\left(y_n - y(x_n)\right)^2 \quad (7)$$

k: 計測データの個数

(7)の$\frac{1}{2}$は微分後の係数を消去するため，一般に付与する係数

係数aについて(7)を微分すれば(8)である．

$$\frac{d}{da}r = 2\sum_{n=1}^{k}\left(y_n - y(x_n)\right)\frac{d}{da}\left(y_n - y(x_n)\right) \quad (8)$$

つまり，ニュートン法としては(9)ということになる．

$$f(a) = \frac{d}{da}r = r'(a) = 0 \quad (9)$$

この例(6)では他にb, cがあり, 3 変数である．従って(8)と同様にして連立方程式となる．多変数のニュートン法については後述するとして, 1 変数の場合を述べる．仮に(6)を(10)とすれば, 1 変数のニュートン法(12)として記述できる．

$$y = c \quad (10)$$

$$\frac{d}{dc} r = 2\sum_{n=1}^{k}(y_n - c)\frac{d}{dc}(y_n - c) = 2\sum_{n=1}^{k}(c - y_n) \quad (11)$$

$$f(c) = \frac{d}{dc} r = 2\sum_{n=1}^{k}(c - y_n) = 0 \quad (12)$$

(12)の解は当然，平均値（13）となる．

$$c = \frac{1}{k}\sum_{n=1}^{k} y_n \quad (13)$$

以上の用例が 2 つ目の場面ということになる．

続けて，多変数のニュートン法を説明するため，補足しておく．まず，(14)とおけば，ニュートン法の漸化式(2)は(15)となる．

$$\Delta x = x^{(k+1)} - x^{(k)} \quad (14)$$

$$f'\left(x^{(k)}\right)\Delta x = -f\left(x^{(k)}\right) \quad (15)$$

よって残差を扱うモデルとするには, (5)を(15)へ代入して(16)となる．

$$r''\left(x^{(k)}\right)\Delta x = -r'\left(x^{(k)}\right) \quad (16)$$

多変数のニュートン法は(16)の形式である．この由来だが，数値計算が容易になるからである．第 2 の場面のイメージを**図 2** に示す．

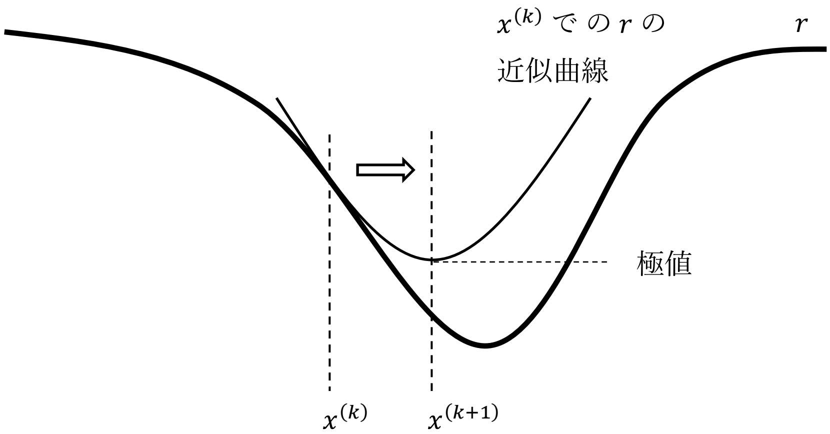

図 2　残差の最小化によるニュートン法のイメージ

テイラー展開から**図 2** を説明することもできる．まず(17)とおき，関数rを$x^{(k)}$でテイラー展開すると(18)である．これは**図 2** の近似曲線に相当する．

$$\Delta x = x^{(k+1)} - x^{(k)} \quad (17)$$

$$r\left(x^{(k)} + \Delta x\right) = r\left(x^{(k)}\right) + r'\left(x^{(k)}\right)\Delta x + \frac{1}{2}r''\left(x^{(k)}\right)(\Delta x)^2 + \frac{1}{3!}r'''\left(x^{(k)}\right)(\Delta x)^3 \cdots \quad (18)$$

$\Delta x \to 0$であれば 3 次以上の項は無視できる．またΔxで微分して，極値を求めるには(19)とする．

$$r'\left(x^{(k)}\right) + r''\left(x^{(k)}\right)\Delta x = 0 \quad (19)$$

変形すれば(16)と同一であると判る．

$$r''\left(x^{(k)}\right)\Delta x = -r'\left(x^{(k)}\right) \quad (16)$$

以上，工学で述べられる残差の最小化は頻繁に登場し，利用されるテクニックである．ニュートン法は，これと併せることで初めて威力を発揮する．既に述べたように 2 次形式は極値の存在が保証された上に 1 ケに限定される．これは極めて扱いやすい性質である．また微小な範囲(定義域)で俯瞰すれば，どのような曲線であろうと必ず 2 次関数で近似できるはずである．そして，工学においては，この感覚が大事である．結論だが，「ツールが使えるようなモデル化を考えよ」ということである．

　他に，ニュートン法では(5)の関数は線形である必要はない．非線形関数を扱うことも可能である．さらに，残差についてだが 2 次形式が絶対ということもない．もう一つ，微分(8)の段階で解析解を導出できない(微分可能な関数でない)場合がある，だが，これも数値微分

で対応できる．実際，非線形問題では，数値微分に頼らざるを得ないだろう．

多変数のニュートン法

ニュートン法も 1 変数しか扱えないのでは，利用価値に乏しい．ここでは 2 変数のニュートン法を説明する．2 変数のニュートン法を理解することで多変数のニュートン法の理解も，そのまま拡張できる．まず，それ以前に全微分(接平面)について解説する．

全微分（接平面）

全微分は(20)である．さらに一般化したn次元は(21)である．

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy \quad (20)$$

$$df = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 + \cdots + \frac{\partial f}{\partial x_n}dx_n \quad (21)$$

(20)を**図 3** で示す．

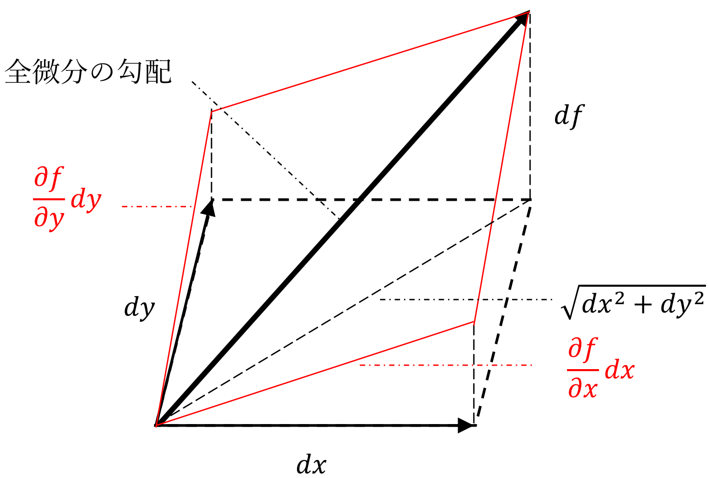

図 3　全微分のイメージ

図中，赤線の部分が接平面を表している．まず，全微分(20)の導出だが，本来，2 変数での微小差分を表すのは(22)である．そして，$\sqrt{dx^2 + dy^2} \to 0$の極限が全微分の定義となる．

$$\frac{\Delta f}{\sqrt{(\Delta x)^2+(\Delta y)^2}}=\frac{f(x+\Delta x,y+\Delta y)-f(x,y)}{\sqrt{(\Delta x)^2+(\Delta y)^2}} \quad (22)$$

1 変数のテイラー展開と同じくfを展開すると(23)である．

$$\begin{aligned}f(x+\Delta x,y+\Delta y)=f(x,y)+\frac{\partial f}{\partial x}\Delta x+\frac{\partial f}{\partial y}\Delta y+\\ \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(\Delta x)^2+\frac{1}{2}\frac{\partial^2 f}{\partial y^2}(\Delta y)^2+\frac{\partial^2 f}{\partial x\partial y}\Delta x\Delta y+\\ \frac{1}{3!}\frac{\partial^3 f}{\partial x^3}(\Delta x)^3+\frac{1}{3!}\frac{\partial^3 f}{\partial y^3}(\Delta y)^3+\frac{1}{2}\frac{\partial^3 f}{\partial x^2\partial y}(\Delta x)^2\Delta y+\frac{1}{2}\frac{\partial^3 f}{\partial x\partial y^2}\Delta x(\Delta y)^2+\cdots \quad (23)\end{aligned}$$

ここで，大雑把にΔxとΔyに比例する項目と$(\Delta x)^2$，$(\Delta y)^2$，$\Delta x\Delta y$，$(\Delta x)^3$，$(\Delta y)^3\cdots$の高次の項に分類される．$\Delta x\to 0,\Delta y\to 0$とすれば高次の項は 1 次の項より早く 0 に収束して無視できる(24)．

$$f(x+\Delta x,y+\Delta y)\fallingdotseq f(x,y)+\frac{\partial f}{\partial x}\Delta x+\frac{\partial f}{\partial y}\Delta y \quad (24)$$

これを(22)に代入すれば(25)である．

$$\frac{\Delta f}{\sqrt{(\Delta x)^2+(\Delta y)^2}}=\frac{\frac{\partial f}{\partial x}\Delta x+\frac{\partial f}{\partial y}\Delta y}{\sqrt{(\Delta x)^2+(\Delta y)^2}} \quad (25)$$

以上により(20)の導出となる．(25)は，ある点から$\sqrt{dx^2+dy^2}$だけ変位すれば，Δfだけ変化することを述べている．これが平面の表現そのものである．実際，点(p,q)における接平面の方程式は(26)であり，全微分の形式と一致する．

$$z-f(p,q)=\frac{\partial f(p,q)}{\partial x}(x-p)+\frac{\partial f(p,q)}{\partial y}(y-q) \quad (26)$$

2 変数のニュートン法

1 変数のニュートン法と同様に議論を進める．最初に，2 変数のニュートン法とは(27)と(28)を満たすx_1，x_2を求める手法である．

$$f_1(x_1,x_2)=0 \quad (27)$$

$$f_2(x_1,x_2)=0 \quad (28)$$

k回目の反復における接平面は(26)に当てはめて（29），（30）とできる．

$$z-f_1\left(x_1^{(k)},x_2^{(k)}\right)=\frac{\partial f_1\left(x_1^{(k)},x_2^{(k)}\right)}{\partial x_1}\left(x_1-x_1^{(k)}\right)+\frac{\partial f_1\left(x_1^{(k)},x_2^{(k)}\right)}{\partial x_2}\left(x_2-x_2^{(k)}\right) \quad (29)$$

$$z-f_2\left(x_1^{(k)},x_2^{(k)}\right)=\frac{\partial f_2\left(x_1^{(k)},x_2^{(k)}\right)}{\partial x_1}\left(x_1-x_1^{(k)}\right)+\frac{\partial f_2\left(x_1^{(k)},x_2^{(k)}\right)}{\partial x_2}\left(x_2-x_2^{(k)}\right) \quad (30)$$

1 変数のときと同様に$z=0$，$x_1=x_1^{(k+1)}$，$x_2=x_2^{(k+1)}$とすれば(31), (32)となる．

$$j_{11}\left(x_1^{(k+1)}-x_1^{(k)}\right)+j_{12}\left(x_2^{(k+1)}-x_2^{(k)}\right)=-f_1\left(x_1^{(k)},x_2^{(k)}\right) \quad (31)$$

$$j_{21}\left(x_1^{(k+1)}-x_1^{(k)}\right)+j_{22}\left(x_2^{(k+1)}-x_2^{(k)}\right)=-f_2\left(x_1^{(k)},x_2^{(k)}\right) \quad (32)$$

$$j_{11}=\frac{\partial f_1\left(x_1^{(k)},x_2^{(k)}\right)}{\partial x_1}, \quad j_{12}=\frac{\partial f_1\left(x_1^{(k)},x_2^{(k)}\right)}{\partial x_2}$$

$$j_{21}=\frac{\partial f_2\left(x_1^{(k)},x_2^{(k)}\right)}{\partial x_1}, \quad j_{22}=\frac{\partial f_2\left(x_1^{(k)},x_2^{(k)}\right)}{\partial x_2}$$

(33), (34), (35)とおいて，行列で表せば(36)となる．

$$\varDelta x_1=x_1^{(k+1)}-x_1^{(k)} \quad (33)$$

$$\varDelta x_2=x_2^{(k+1)}-x_2^{(k)} \quad (34)$$

$$X^{(k)}=\begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \end{bmatrix} \quad (35)$$

$$\nabla F^{(k)}\varDelta X=-F^{(k)} \quad (36)$$

$$\nabla F^{(k)}=\begin{bmatrix} j_{11} & j_{12} \\ j_{21} & j_{22} \end{bmatrix}, \quad \varDelta X=\begin{bmatrix} \varDelta x_1 \\ \varDelta x_2 \end{bmatrix}, \quad F^{(k)}=\begin{bmatrix} f_1\left({X^{(k)}}^{\top}\right) \\ f_2\left({X^{(k)}}^{\top}\right) \end{bmatrix}=\begin{bmatrix} f_1\left(x_1^{(k)},x_2^{(k)}\right) \\ f_2\left(x_1^{(k)},x_2^{(k)}\right) \end{bmatrix}$$

(36)は 1 変数の(4), (15)に相当する．$\nabla F^{(k)}$はヤコビ行列となり，多変数へそのまま拡張できる．

ここで，1 変数と同様にfがrの 1 階微分で，残差の 2 乗和を最小化する問題だったとする(37), (38).

$$\frac{\partial r(x_1, x_2)}{\partial x_1} = f_1(x_1, x_2) = 0 \quad (37)$$

$$\frac{\partial r(x_1, x_2)}{\partial x_2} = f_2(x_1, x_2) = 0 \quad (38)$$

これを(31), (32)に代入すれば(39), (40)となる.

$$h_{11}\left(x_1^{(k+1)} - x_1^{(k)}\right) + h_{12}\left(x_2^{(k+1)} - x_2^{(k)}\right) = -\frac{\partial}{\partial x_1} r\left(x_1^{(k)}, x_2^{(k)}\right) \quad (39)$$

$$h_{21}\left(x_1^{(k+1)} - x_1^{(k)}\right) + h_{22}\left(x_2^{(k+1)} - x_2^{(k)}\right) = -\frac{\partial}{\partial x_2} r\left(x_1^{(k)}, x_2^{(k)}\right) \quad (40)$$

$$h_{11} = \frac{\partial^2 r\left(x_1^{(k)}, x_2^{(k)}\right)}{\partial {x_1}^2}, \quad h_{12} = \frac{\partial^2 r\left(x_1^{(k)}, x_2^{(k)}\right)}{\partial x_1 \partial x_2}$$

$$h_{21} = \frac{\partial^2 r\left(x_1^{(k)}, x_2^{(k)}\right)}{\partial x_2 \partial x_1}, \quad h_{22} = \frac{\partial^2 r\left(x_1^{(k)}, x_2^{(k)}\right)}{\partial {x_2}^2}$$

これを行列で表せば(41)である.

$$H^{(k)} \Delta X = -\nabla r^{(k)} \quad (41)$$

$$H^{(k)} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}, \quad F^{(k)} = \begin{bmatrix} f_1\left(x_1^{(k)}, x_2^{(k)}\right) \\ f_2\left(x_1^{(k)}, x_2^{(k)}\right) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial r\left(x_1^{(k)}, x_2^{(k)}\right)}{\partial x_1} \\ \dfrac{\partial r\left(x_1^{(k)}, x_2^{(k)}\right)}{\partial x_2} \end{bmatrix} = \nabla r^{(k)}$$

(41)は 1 変数の(16)に相当する．$H^{(k)}$はヘッセ行列(42)となり，多変数へそのまま拡張できる.

$$H = \begin{bmatrix} \dfrac{\partial^2 r}{\partial {x_1}^2} & \dfrac{\partial^2 r}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 r}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 r}{\partial x_2 \partial x_1} & \dfrac{\partial^2 r}{\partial {x_2}^2} & \cdots & \dfrac{\partial^2 r}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 r}{\partial x_n \partial x_1} & \dfrac{\partial^2 r}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 r}{\partial {x_n}^2} \end{bmatrix} \quad (42)$$

数学としては関数が連続であれば微分の順序は入れ換えられ，ヘッセ行列は対称行列となる．数値計算でも，基本的には対称行列として非対角の一方のみを計算すればよい.

rから(41)の導出も可能である．rをn変数の関数としてテイラー展開すれば(43)である．

$$r(x_1+\Delta x_1, x_2+\Delta x_2, \cdots, x_n+\Delta x_n)$$
$$= r(x_1, x_2, \cdots, x_n) + \sum_{i=1}^{n} \frac{\partial r}{\partial x_i}\Delta x_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 r}{\partial x_i \partial x_j}\Delta x_i \Delta x_j + \cdots \quad (43)$$

これは2次の項までの表記としてある．1変数と同様に3次以上を切り捨て，Δx_iで偏微分すれば極値を与えることができる(44).

$$\frac{\partial r}{\partial x_i} + \sum_{j=1}^{n} \frac{\partial^2 r}{\partial x_i \partial x_j}\Delta x_j = 0$$
$$\sum_{j=1}^{n} \frac{\partial^2 r}{\partial x_i \partial x_j}\Delta x_j = -\frac{\partial r}{\partial x_i} \quad (44)$$
$$i = 1,2,\cdots,n：n\text{ケの式からなる連立方程式}$$

(44)は(41)の各行を表記するもので，同一と判る．以上，ニュートン法の現実的で最も利用される応用，一般形式になるのかと思う なお，(41)のΔXの求解についてはガウスの消去法が一般的である．巨大行列の求解に関しては，SDM(最急降下法)あるいは CGM(共役勾配法)のような反復法が適する．しかし，既述のような近似式の係数を求める問題では，変数は少数であり，数百，数千，あるいはそれ以上というような巨大行列にはならない．また，反復法の手順の中でさらに反復法を適用することは，(計算コストと精度に関して)現実的でない．このような場合, SDM あるいは CGM を直接使えるモデルを検討するのがよい.

多変数の勾配法

残差 2 乗和といった最小化の問題だが，これは**図 2** のように勾配に沿って解に至る手法である(**図 4** 参照).

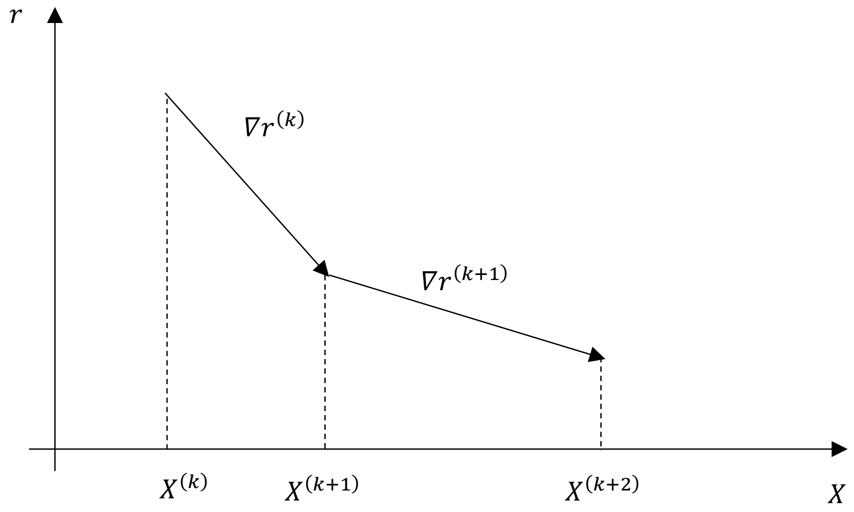

図 4　勾配法の探索

ここで，1 変数の勾配法を用いて，多変数の粗い探索を行う方法がある[1]．1 変数の勾配法にはニュートン法(2)を用いることとし，係数γを導入する(45)．

$$x^{(k+1)} = x^{(k)} - \frac{f\left(x^{(k)}\right)}{\gamma f'(x^{(k)})} \quad (45)$$

$$\gamma > 1$$

係数の目的は勾配を行き過ぎないようにすることにある．これをrで表記すれば(46)である．

$$x^{(k+1)} = x^{(k)} - \frac{r'\left(x^{(k)}\right)}{\gamma r''(x^{(k)})} \quad (46)$$

そのまま多変数へ拡張すれば(47)となる．

$$\begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ \vdots \\ x_n^{(k+1)} \end{bmatrix} = \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ \vdots \\ x_n^{(k)} \end{bmatrix} - \frac{1}{\gamma} \begin{bmatrix} \frac{1}{\frac{\partial^2 r^{(k)}}{\partial {x_1}^2}} & 0 & & \cdots & 0 \\ 0 & \frac{1}{\frac{\partial^2 r^{(k)}}{\partial {x_2}^2}} & 0 & \cdots & 0 \\ & 0 & & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{\frac{\partial^2 r^{(k)}}{\partial {x_n}^2}} \end{bmatrix} \begin{bmatrix} \frac{\partial r^{(k)}}{\partial x_1} \\ \frac{\partial r^{(k)}}{\partial x_2} \\ \vdots \\ \frac{\partial r^{(k)}}{\partial x_n} \end{bmatrix}$$

$$
=\begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ \vdots \\ x_n^{(k)} \end{bmatrix} - \frac{1}{\gamma}\begin{bmatrix} \frac{\partial^2 r^{(k)}}{\partial {x_1}^2} & 0 & & \cdots & 0 \\ 0 & \frac{\partial^2 r^{(k)}}{\partial {x_2}^2} & 0 & \cdots & 0 \\ & 0 & & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \frac{\partial^2 r^{(k)}}{\partial {x_n}^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial r^{(k)}}{\partial x_1} \\ \frac{\partial r^{(k)}}{\partial x_2} \\ \vdots \\ \frac{\partial r^{(k)}}{\partial x_n} \end{bmatrix} \quad (47)
$$

この手法で勾配を探索できることもある. (47)を行列で表現すると(48)である.

$$
X^{(k+1)} = X^{(k)} - \frac{1}{\gamma} D[H]^{-1} \nabla r \quad (48)
$$

$D[\cdot]$：対角成分を取り出すオペレータ

(48)を変形すると(49)となる.

$$
\gamma\left(X^{(k+1)} - X^{(k)}\right) = -D[H]^{-1}\nabla r
$$

$$
\gamma D[H] \Delta X = -\nabla r \quad (49)
$$

(49)は(41)の多変数のニュートン法と同形式である. 非対角成分を無視し, 対角成分だけで勾配を探索する手法と説明できる. 一般にこの手法を使うことはなく, この後のLM法を説明する準備である.

LM（Levenberg-Marquardt）法

逆行列を求めずして求解する手法では，対角成分にエネルギーが集中した段階，すなわち「非対角要素が全て０（ゼロ）となれば，解そのもの」である．これに準ずるのが上三角，下三角といった行列である．そして反復法では，対角成分の強度を調整して収束速度を調整する手法がある．また，非対角成分は変数間の相関を表す成分であるから，対角成分の強度を上げることは，その相関を排して勾配を探索しようということでもある．そして，その逆もまた然りである．

さて，LM 法[1]とは多変数のニュートン法(41)と(49)を組み合わせた手法(50)である．

$$(H+\gamma D[H])\Delta X=-\nabla r \quad (50)$$

一点，(50)では係数γの指定範囲に制限がなくなることに注意されたい．γについてだが，プログラミング上，対角成分への加算ではなく直接の乗数としてもよい．一般には対角成分の強度を調整することで収束性や収束速度を調整することができる．

初期値の決め方

初期値の決め方に定型の手法といったものは存在しない．基本的には，どれだけ解の近くに初期値を置けるかにかかっている．いかにニュートン法が 2 次収束するといえども，解から遠いところ(別の言い方をすれば，探索点の近傍で 2 次で近似できない曲面となっている)では，停滞も然りである．対象に合わせて工夫が必要になる．他に，変数がとりうる範囲(定義域)を前もって把握しておくことも重要である．定義域が判っていれば，目標とする解の精度に応じて分解能を決めて探索範囲に制限を課すといったことも考えられる．さらに，定義域を外れた場合，探索失敗とする(プログラムの停止)，あるいは異なる初期値の投入といった応用も可能である．

数値微分の刻み幅

ニュートン法ではその 1 階微分と 2 階微分が必要となる．モデル化の時点で微分可能な関数ならよいが，多くの応用では数値微分となる．1 変数の関数であれば**図 5** で示すように関数rをサンプリングすることとなる．**図 5** では 1 変数の微分に必要なサンプリング点を示している．

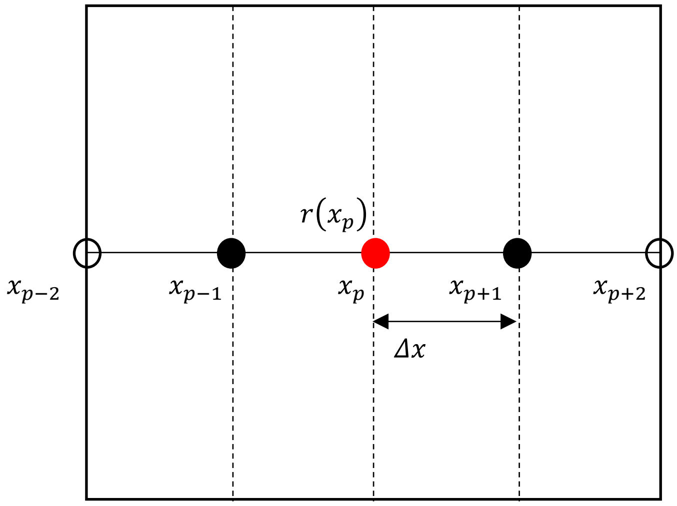

図 5　微分のサンプリング点(1 変数)

1 階微分, 2 階微分はそれぞれ(51), (52)となり, y方向も同様である.

$$\frac{\partial}{\partial x}r(x_p)=\frac{r(x_{p+1})-r(x_{p-1})}{2\Delta x} \qquad (51)$$

$$\frac{\partial^2}{\partial x^2}r(x_p)=\frac{r'(x_{p+1})-r'(x_{p-1})}{2\Delta x}=\frac{\frac{r(x_{p+2})-r(x_p)}{2\Delta x}-\frac{r(x_p)-r(x_{p-2})}{2\Delta x}}{2\Delta x} \qquad (52)$$

次に 2 変数の 2 階微分だが, x方向を先に微分する場合のサンプリング点は**図 6** である.

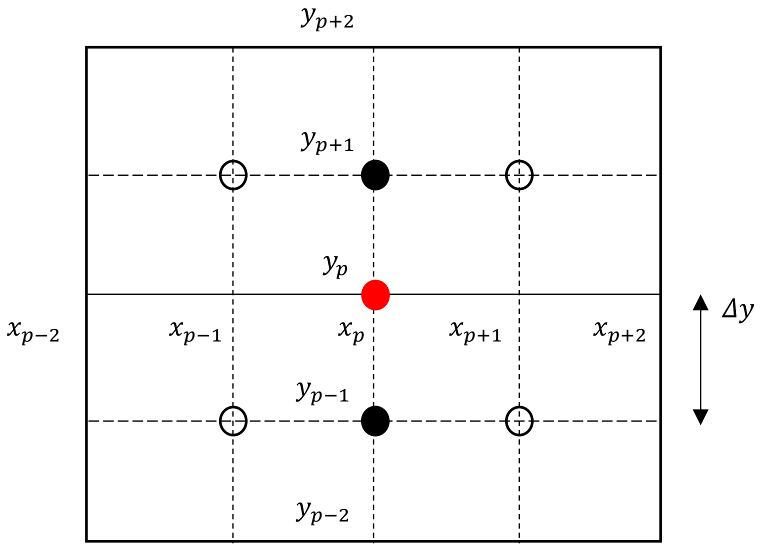

図 6　2 階微分(2 変数)

2 変数での数式は(53)となる.

$$\frac{\partial}{\partial y}\frac{\partial}{\partial x}r(x_{p,}y_p) = \frac{\frac{r(x_{p+1},y_{p+1}) - r(x_{p-1},y_{p+1})}{2\Delta x} - \frac{r(x_{p+1},y_{p-1}) - r(x_{p-1},y_{p-1})}{2\Delta x}}{2\Delta y} \tag{53}$$

以上，最も自然な数値微分の取り方である．他に CIP 法といったシミュレーションでみられる風上差分といったやり方もあるが本稿では割愛する．y方向を先に微分する場合，サンプリング点を変更することになる．ヘッセ行列を対称行列として扱う場合，非対角のいずれか一方で事足りるが，両方を計算することもできる．これは筆者の認識だが，それぞれの長所がある．

＊対称行列とした場合，反復法は収束しやすい．また，計算コストを抑えられる．

＊非対称とした場合，勾配を厳密に評価することができる．特に変数間の相関を厳密に反映できる．ただ，非対角成分に違いがありすぎる場合，もともと性質のよい問題ではなかった，とも言える．これは収束性にかかわることであり，手法が，そもそも機能しないということがある．

さて，問題は「数値微分の刻み幅をどの程度の大きさとするか？」である．本来，微分とは微小差分の極限，→0（ゼロ）での勾配である．とすれば，できるだけ小さい数値が望ましい．では「計算機で表現できる最小数，マシンイプシロンにすればよい,」ということにはならない．マシンイプシロンでは有意な変化量が生じない可能性がある．他に数値計算時の桁落ち誤差も問題となる．既述の数値微分のサンプリングに関して下記のガイドラインがある．

＊計算桁数の 1/3 くらいの桁のところに選ぶのがよい[2]

計算桁数とは計算における有効桁である．IEEE754 倍精度の仮数部は 52bit であるから 10 進での有効桁数はおよそ 15 桁である(54)．

$$52 \times \log 2 \fallingdotseq 15.65 \quad (54)$$

CPU に内蔵される FPU では当然，これ以上の桁数で計算が行われるため，計算誤差を基準とした評価[2]としては疑問もあろうが，ガイドラインどおり 1/3 とすれば数値微分の刻み幅は 5 桁(55)となる．

$$15.7 \times \frac{1}{3} \fallingdotseq 5 \quad (55)$$

例として，残差の 2 乗和 rが計測データを元に下記の大きさで算出されるものだったとする(56)．

$$r \approx 500 = 5 \times 10^2 \quad (56)$$

すると数値微分の刻み幅$\varDelta$としては(57)となる．

$$\varDelta = 5 \times 10^{2-5} = 5 \times 10^{-3} \quad (57)$$

筆者の経験上，これは目安にすぎない．現実には様々な挙動があり，下記のような問題に直面する．

① 1 階微分と 2 階微分で同じ刻み幅でよいのか？
② 反復を繰り返す過程でrの大きさが変化する場合，刻み幅は一定のままでよいのか？
③ 多変数の場合，変数の定義域における変数間の大きさの違いはどうすればよいのか？ 実際，変数間の大きさの違いで反復法が発散することもある．

④ 刻み幅が大きくても，問題なく収束することもある．さらに収束が早いこともある．

最初の①だが, 1 階微分と 2 階微分は同じ差分で評価すべきである．むしろ，そうしない理由があるのか？の逆説である．③では具体的な例(「ニュートン法の例」参照)を挙げると関数s(58)のカーブフィッティングである(**図 7** 参照).

$$s(x) = ax^2 e^{-bx} \quad (58)$$

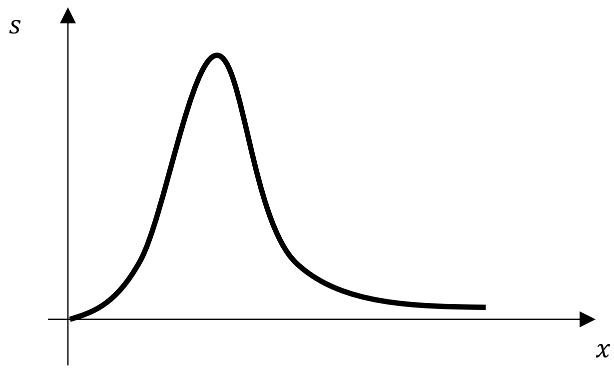

図 7　近似関数s

ここで，rが関数sによる残差の最小化問題とする(59).

$$r(a, b) = \frac{1}{2}\sum_{n=1}^{k}(s(x_n) - s_n)^2 \quad (59)$$

k: 計測データの個数

rは微分可能なため, 1 階微分, 2 階微分まで求めて算出式を得ることができる．ここでは説明のための例として了解いただきたい，とはいえ，理論式が微分可能であっても関数が複雑であれば，定式化された数値微分を利用した方が便利だし，プログラムのバグ回避ということにもなる．

(58)では，数式(57)から明らかなように指数bの方がaと比して，わずかな差分でrに与える変量が大きい．このような場合，変数間で解が振動して収束しない，ということもしばしばである．これは一方の変量で他方が大きく変動するからである．結局のところ，筆者は変数毎に刻み幅を調整することを常としている．他に，この例では対数をとった残差の評価も可能である．具体的にはsを(60)と書き換えて(61)とする．

$$s(x) = \log a + 2\log x - bx \quad (60)$$

$$r(c,b) = \frac{1}{2}\sum_{n=1}^{k}(s(x_n) - s_n)^2 \quad (61)$$
$$c = \log a$$

cの求解後，aを算出できる．これならば，変数間の刻み幅による変量を改善できる．なお，対数をとる手法にはデメリットもある．計測データが低 SN の場合，計測データの信号成分の重みが小さくなり，その一方でノイズの重みは大きくなる，結果としてフィッティングそのものの精度が甘くなってしまう．この対策がないこともないが，数値計算が簡単にはいかないところである．④については，対象の性質が良い場合(2 次の曲面の近似精度がよい→ **図 2** 参照)，刻み幅は大きくても問題ないことがある．とはいえ，解近傍での解そのものの精度(有効桁)の向上に難があるので注意されたい．以上，一例にすぎず，②を含めて筆者自身，ほとんどの場合で解を持ち合わせていない．結局のところ，問題毎に試行錯誤と工夫が必要になるのかと思う．

参考文献

[1] 金谷健一. これならわかる最適化数学. 共立出版. 2005.

[2] 伊理正夫, 藤野和建. 数値計算の常識. 共立出版. 1985.