UDC 65.012.1: 681.14: 519.281

電子計算機による実験データの整理

研究部電算研究室　森　安　正　司

Arrangement of Experimental Data with Digital Computer

Shōji Moriyasu
(Electrical Computer Sect., Research Dep't.)

Synopsis

When we perform scientific calculations with digital computer, the physical characteristics, e.g. B-H curve, i-s curves, Müschel curve, are often given with curve or surface. In such a case, interpolation method requires many memories, so that it is handy to express these characteristics by polynomial function.

In this paper, I explain curve and surface fitting of polynomial function.

The contents are as follows:

I Introduction
II Least square method
III Curve fitting
IV Surface fitting.

I. まえがき

電子計算機を用いて研究設計計算を行なうときに，計算に用いる物理特性（たとえば鉄心の B-H 特性，蒸気の i-s 線図，水車の効率曲線など）が曲線や曲面によって与えられることがよくある．このときに数表索引の方法もあるが非常に多くの記憶素子を占有するので，これを関数の形で与えておくと便利である。

最小二乗法は曲線を多項式で近似するときには有力な武器であり，さらに曲面に対する多項式近似にも拡張して考えることができるが，まだあまり例がないようである．しかし最小二乗法も各々の実験データの精度を考えて計算しないと結果が期待できない場合がある．

本稿ではまず最小二乗法，とくに各々の実験データの精度考慮した場合について説明し，それを曲線近似および曲面近似に適用した場合について述べる。

II. 最小二乗法

1. 最小二乗法の意味

ある実験を行なってN組の実験データ (x_1, y_1), (x_2, y_2),……(x_N, y_N) を得たとき，これらの間の関数関係として，次のような線形な方程式

$$y=c_1 g_1(x)+c_2 g_2(x)+\cdots\cdots+c_m g_m(x) \cdots(1)$$

を考える．ここで $g(x)$ は既知の関数であり，係数 c_1, c_2,……, c_m はこれから求めようとする係数（最確値）である。

式(1)を観測方程式というが，これは問題によって次の2種類がある．一つはたとえば落体の距離と時間のように変数間の数式が既知でただその中に現われる係数の最確値を求める場合（理論式）であり，他の一つは変数間に関数関係があることは確かであるが，その数式がわかっていない場合（実験式）である．後者の場合はまず式(1)の $g_1(x)$, $g_2(x)$,…, $g_m(x)$ の形をいかにすべきかが問題となるが，実際上はなるべく簡単で，しかも結果がよく実験に一致するような式を選ぶのであり，その実験に対する専門的知識による勘から主観的にきめることが多い。

さて式(1)は近似式にすぎないので，実験値 (x_k, y_k) に対して偏差（残差）

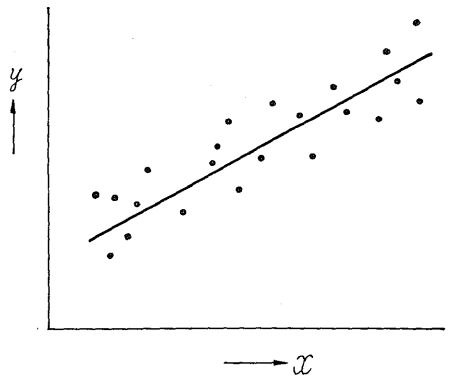
$$v_k=c_1 g_1(x_k)+c_2 g_2(x_k)+\cdots+c_m g_m(x_k)-y_k \cdots\cdots\cdots\cdots(2)$$

を生じる．この偏差の平方和

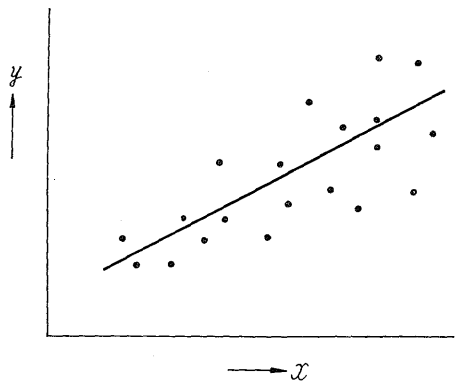
$$u=\sum_{k=1}^{N} v^2{}_k \cdots\cdots\cdots\cdots(3)$$

が最小になるように係数 c_1, c_2…,c_m をきめるのが最小二乗法である．これはおのおのの v_k の分布を正規分布とし，その分散がみな同一と考えられる場合に妥当な結果が得られるのである．すなわち第1図のように v_k が確率的に同等と考えられることができ，式(3)に対する寄与が同等に評価できる場合である．ところが第2図のようなばらつきのある実験結果に対して直接式(3)の最小を与える係数を求めるのは無理である．これは偏差 v_k が x

に比例して大きくなるため，式(3)に対して同等に評価できないからである．逆に式(3)を最小にするのでは，実験データを同一に評価せず，x の大きな値に対するデータほどそれだけウェイトを大きくかけていることになる．このような不都合をなくするのが重みをつけた最小二乗法である。



第1図 実 験 デ ー タ I
Fig. 1. Experimental data-I



第2図 実 験 デ ー タ II
Fig. 2. Experimental data-II

いま実験値 y の誤差の分散を $\alpha\sigma^2$（α は定数であるがどんな値でもよい）とすれば式(1)に対応する観測方程式

$$\frac{y}{\sigma}=\frac{c_1\,g_1(x)+c_2\,g_2(x)+\cdots+g_m(x)}{\sigma}\quad\cdots\cdots(4)$$

の誤差は各実験値に対して同一と考えることができるからこれに対して偏差の平方和

$$u=\sum_{k=1}^{N}\left(\frac{v_k}{\sigma_k}\right)^2=\sum_{k=1}^{N}p_k\,v^2{}_k\quad\cdots\cdots(5)$$

が最小になるように $c_1,\ c_2,\cdots,c_m$ を決めればよい。

$p_k=1/\sigma_k{}^2$ をウェイトといい，α の値によって異なる相対的な値であるが，これは式(4)すなわち $y/\sigma(x)$ のウェイトを1として基準に考えれば式(1)すなわち y のウェイトは p になることを意味する。

ウェイト p_k の値は物理的意味とか実験結果のばらつきなどからある程度直感的に決めるのが普通である．このようなウェイトを考慮しないで最小二乗法を適用している例をよくみるが，これでは意味がなく結果も期待できない。

式(5)の最小値を与える係数 $c_1,\ c_2,\cdots,c_m$ は，次の正規方程式の解である。

$$\begin{pmatrix} a_{11} & a_{12}\cdots\cdots & a_{1m} \\ a_{21} & a_{22}\cdots\cdots & a_{2m} \\ \vdots & \vdots & \vdots \\ a_{m1} & a_{m2}\cdots\cdots & a_{mm} \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}=\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}\quad\cdots\cdots(6)$$

ここに

$$\left.\begin{aligned} a_{ij}&=a_{ji}=\sum_{k=1}^{N}p_k\,g_i(x_k)\ g_j(x_k) \\ b_i&=\sum_{k=1}^{N}p_k\,y_k\,g_i(x_k) \end{aligned}\right\}\quad\cdots\cdots(7)$$

2. 誤 差 の 推 定

最小二乗法により求めた実験式に対して実験値は当然誤差が生じるが，基準のウェイト $p=1$ に対する実験値の誤差の分散は

$$V_y=\frac{1}{N-m}\sum_{k=1}^{N}p_k\,v_k{}^2=\frac{1}{N-m}\left(\sum_{k=1}^{N}p_k\,v_k{}^2-\sum_{i=1}^{m}c_i\,b_i\right)\quad\cdots\cdots(8)$$

として推定することができる．[1] これから一般のウェイト p_k に対する実験値 y_k に対する分散は V_y/p_k となる。

また式(8)で $N=m$ のときは実験データをすべてみたす実験式が決定でき，$v_k=0$ となるので，V_y は不定となる．つまり分散の推定が不可能であり，この場合最小二乗法は無意味である．データの数Nとして筆者は $3m$ 個以上を取っている．次に正規方程式の解として求まる係数 $c_1,\ c_2,\cdots,c_m$ も実験値によって変動する統計的な値であるからやはり誤差がある．c_i に対する誤差の分散の推定値 Vc_i は正規方程式の係数行列の逆行列の要素を d_{ij} とすれば

$$Vc_i=d_{ii}\,V_y\quad\cdots\cdots(9)$$

として求めることができる。

3. 正規方程式を解くときの問題点

式(6)の正規方程式を計算機で解くとき，ガウスの消去法が広く用いられている．これは l 回目の消去の段階での新しい係数の値を $a''_{ij}=a'_{ij}-a'_{lj}\,a'_{il}/a'_{ll}$（$a'_{ij}$, a'_{il}, a'_{lj}, a'_{ll} は $l-1$ 回目の消去によってできる係数）として計算を進めて行くが，分母の a'_{ll} を消去の軸という。

この軸が有効けたを失なって非常に小さな値になれば解の精度は期待できなくなる．このような悪い条件は特に多項式近似の場合によくおこるといわれている。

これをのがれるにはデータを1個取り除いたり，あるいは加えたり，さらにはわざと少し誤ったデータを与えるのも一つの方法であるが，データとして $(x_k,\ y_k)$ でなく

$$X_k=\frac{x_k-\bar{x}}{\sigma_x},\quad Y_k=\frac{y_k-\bar{y}}{\sigma_y} \quad\cdots\cdots\cdots\cdots(10)$$

を与えることによって等価的に有効けたを多くし，しかも正規方程式の係数を規格化するのが，もっとも有効な方法である．ここに $\bar{x}$, $\bar{y}$ σ_x, σ_y はデータ x, y の平均値および標準偏差である。

このときの実験式およびその誤差の推定値は

$$\left.\begin{aligned} y&=\bar{y}+\sigma_y\sum_{i=1}^{m}c_i\,g_i\left(\frac{x-\bar{x}}{\sigma_x}\right)\\ V_y&=\sigma_y{}^2\,V_Y\end{aligned}\right\}\quad\cdots\cdots\cdots\cdots(11)$$

となる．ここに V_Y はデータ $(X,\ Y)$ に対して式(8)から求まる値である．このようなデータの変換を行なっても a'_{ll} が小さくなれば，b_l を強制的に0において，つまり実験式のうち b_l の項がはじめからなかったものとして消去を進めて行けばよい．これは b_l に関係する係数 a_{li}, a_{lj} $(i,\ j=1,\ 2,\cdots\cdots m)$ は l 段階までの消去の過程で l 行，l 列以外には伝わらないから可能である。

III. 曲 線 近 似

1. 多 項 式 近 似

観測方程式が(1)に対応して

$$y=c_0+c_1\,x+c_2\,x^2+\cdots\cdots+c_m\,x^m \quad\cdots\cdots\cdots\cdots(12)$$

で表わせる場合である．このとき正規方程式の係数は

$$a_{ij}=\sum_{k=1}^{N}p_k\,x_k{}^{i+j},\quad b_i=\sum_{k=1}^{N}p_k\,y_k\,x^i{}_k\cdots\cdots\cdots\cdots\cdots(13)$$

となる．$p_k=1$ として計算するが普通であるが，しかし問題によってはウェイトを充分考慮しなければならない。

2. 直交多項式による近似[2]

$g(x)$ として直交多項式を用いる．すなわち

$$\sum_{k=1}^{N}p_k\,g_i(x_k)g_j(x_k)=\begin{cases}0:i\neq j\\ D_i\neq 0:i=j\end{cases}\quad\cdots\cdots\cdots(14)$$

をみたす多項式によって曲線近似を行なう。

$$y=c_0\,g_0(x)+c_1\,g_1(x)+\cdots\cdots+c_m\,g_m(x)\cdots\cdots(15)$$

この場合正規方程式の対角要素以外はすべて0になるので正規方程式を解く必要はなくしかも誤差の推定も簡単である．このような多項式の一つは

$$\left.\begin{aligned} &g_{i+1}(x)=(x-\alpha_i)\,g_i(x)-\beta_i\,g_{i-1}(x)\\ &\alpha_i=\sum_{k=1}^{N}p_k\,x_k\,g_i{}^2(x_k)/D_i\\ &\beta_i=D_i/D_{i-1}\\ &g_0(x)=1,\ g_1(x)=x-\bar{x}\\ &\bar{x}=\sum_{k=1}^{N}p_k\,x_k\Big/\sum_{k=1}^{N}p_k\end{aligned}\right\}\cdots\cdots(16)$$

なる繰返しの関係から求めることができる。

係数 c_i，ウェイト1の実験値 y の誤差の分散 V_y および c_i の誤差の分散 Vc_i は

$$\left.\begin{aligned} &c_i=\sum_{k=1}^{N}p_k\,y_k\,g_i(x_k)/D_i\\ &V_y=\frac{1}{N-(m+1)}\left(\sum_{k=1}^{N}p_k\,y_k{}^2-\sum_{i=0}^{m}c_i\,D_i\right)\\ &Vc_i=V_y/D_i\end{aligned}\right\}\cdots(17)$$

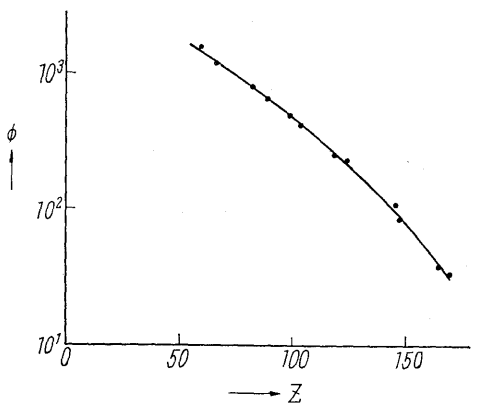
この方法のすぐれている点は近似多項式の次数 m が 1, 2, 3, ……について順次求まるので前項の多項式近似のように次数 m を最初に決めないで式(17)の分散 V_y との関係から最適な次数を求め得ることである。

3. 非線形模型に対する近似

曲線近似ではあるが未知の係数が式(1)のような線形関係になく一般に

$$y=f(x,\ c_1,\ c_2,\ c_3,\cdots\cdots)\cdots\cdots\cdots\cdots\cdots\cdots(18)$$

で表わせる場合である．この場合最確値 c_i を最確値に充分近い仮定値 $c^\circ{}_i$ と補正値 Δc_i の和と考え，テイラ展開によって，補正値 Δc_i についての線形な観測方程式を作って最小二乗法を適用する．第3図の未臨界実験について述べる[3]．軸方向 z，中性子束 ϕ に対する理論式は



第3図 未 臨 界 実 験

Fig. 3. Subcritical assembly experiments

$$\phi=c_1\ \sin\ h\ (c_2-c_3\,z)\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(19)$$

であるから，$c_1=c^\circ{}_1+\Delta c_1$, $c_2=c^\circ{}_2+\Delta c_2$, $c_3=c^\circ{}_3+\Delta c_3$ とおいて Δc_1, Δc_2, Δc_3 に対する観測方程式を作ると

$$\phi\doteqdot c^\circ{}_1\ \sinh\ (c^\circ{}_2-c^\circ{}_3\,z)+\sinh\ (c^\circ{}_2-c^\circ{}_3 z)\ \Delta c_1$$

$$+c^{\circ}{}_1 \cosh(c^{\circ}{}_2-c^{\circ}{}_3 z)\,\Delta c_2-c^{\circ}{}_1 z \cosh(c^{\circ}{}_2-c^{\circ}{}_3 z)\,\Delta c_3$$

より

$$m=\alpha\,\Delta c_1+\beta\,\Delta c_2+\gamma\,\Delta c_3 \cdots\cdots\cdots\cdots(20)$$

ここに $m=\phi-c^{\circ}{}_1 \sinh(c^{\circ}{}_2-c^{\circ}{}_3 z)$

$\alpha=\sinh(c^{\circ}{}_2-c^{\circ}{}_3 z)$

$\beta=c^{\circ}{}_1 \cosh(c^{\circ}{}_2-c^{\circ}{}_3 z)$

$\gamma=-c^{\circ}{}_1 z \cosh(c^{\circ}{}_2-c^{\circ}{}_3 z)$

となる．ところが第3図からわかるように，ϕ の値は $10\sim10^3$ まで大幅に変化するので，ϕ すなわち m の誤差の分散は一定とは考えられず，むしろ ϕ^2 に比例（相対誤差一定）すると考えるべきであるから，$p=1/\phi^2$ とすべきである．この場合正規方程式は

$$\left.\begin{aligned}
&\Delta c_1\sum_k\frac{\alpha_k{}^2}{\phi_k{}^2}+\Delta c_2\sum_k\frac{\alpha_k\beta_k}{\phi_k{}^2}+\Delta c_3\sum_k\frac{\alpha_k\gamma_k}{\phi_k{}^2}=\sum_k\frac{\alpha_k m_k}{\phi_k{}^2}\\
&\Delta c_1\sum_k\frac{\alpha_k\beta_k}{\phi_k{}^2}+\Delta c_2\sum_k\frac{\beta_k{}^2}{\phi_k{}^2}+\Delta c_3\sum_k\frac{\beta_k\gamma_k}{\phi_k{}^2}=\sum_k\frac{\beta_k m_k}{\phi_k{}^2}\\
&\Delta c_1\sum_k\frac{\alpha_k\gamma_k}{\phi_k{}^2}+\Delta c_2\sum_k\frac{\beta_k\gamma_k}{\phi_k{}^2}+\Delta c_3\sum_k\frac{\gamma_k{}^2}{\phi_k{}^2}=\sum_k\frac{\gamma_k m_k}{\phi_k{}^2}
\end{aligned}\right\}\cdots\cdots\cdots\cdots(21)$$

式 (8), (9) によって求まる誤差は線形化によって精度そのものは向上するわけではないから，$\{\phi-c_1 \sinh(c_2-c_3 z)\}/\phi=0$, c_1, c_2, c_3 に対する誤差と同じである。

IV. 曲 面 近 似

今近似する曲面が次の多項式

$$z=\sum_{i=0}^{m}\sum_{j=0}^{n} c_{ij}\,x^i\,y^j \cdots\cdots\cdots\cdots(22)$$

で表わせるものとする．これに対して偏差の平方和を最小にする係数 c_{ij} は式(6)に対応して次の正規方程式となる。

$$\begin{pmatrix}
a_{00}^{00} & a_{01}^{00} & \cdots\cdots & a_{0n}^{00} & a_{10}^{00} & \cdots\cdots & a_{mn}^{00}\\
a_{00}^{01} & a_{01}^{01} & \cdots\cdots & a_{0n}^{01} & a_{10}^{01} & \cdots\cdots & a_{mn}^{01}\\
\vdots & \vdots & & \vdots & \vdots & & \vdots\\
a_{00}^{mn} & a_{01}^{mn} & \cdots\cdots & a_{0n}^{mn} & a_{10}^{mn} & \cdots\cdots & a_{mn}^{mn}
\end{pmatrix}
\begin{pmatrix} c_{00}\\ c_{01}\\ \vdots\\ c_{mn}\end{pmatrix}
=\begin{pmatrix} b_{00}\\ b_{01}\\ \vdots\\ b_{mn}\end{pmatrix}\cdots(23)$$

ここに

$$\left.\begin{aligned}
&a_{ij}^{st}=a_{st}^{ij}=\sum_{k=1}^{N} x_k{}^{i+s}\,y_k{}^{i+t}\\
&b_{ij}=\sum_{k=1}^{N} z_k\,x_k{}^i\,y_k{}^j
\end{aligned}\right\}\cdots\cdots\cdots\cdots(24)$$

ウェイトは計算の複雑さを避けるため $p_k=1$ としたのでそれを補正するようにデータを与えなければならない．つまり p_k の大きい実験値の近辺はそれだけ多くデータを与えればよい．式(24)の係数 a_{ij}^{st} は $i+s$, $j+t$ の値が一定であれば同じ値であるから正規方程式の $\{(m+1)(n+1)\}^2$ 個の係数のうち独立なものは $(2m+1)(2n+1)$ 個にすぎない．だからこれらを配列するだけで正規方程式の係数はすべて求めることができる。

変数 z の誤差の分散の推定値 V_z は

$$V_z=\frac{1}{N-(m+1)(n+1)}\cdot\left(\sum_{k=1}^{N} z_k{}^2-\sum_{i=0}^{m}\sum_{j=0}^{n} c_{ij}\,b_{ij}\right)\cdots\cdots\cdots\cdots(25)$$

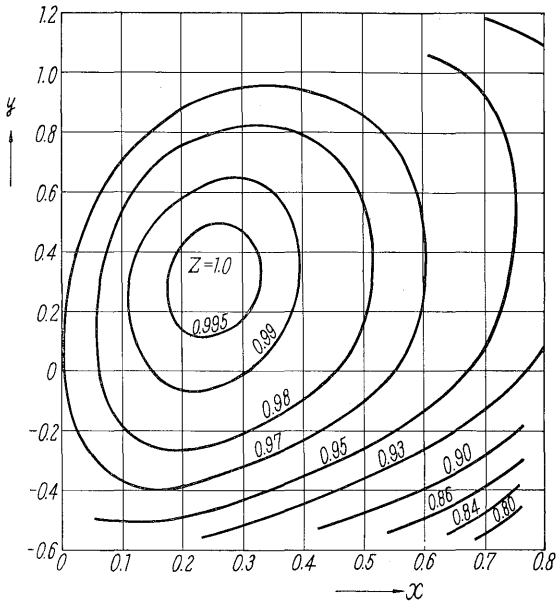
として与えることができる．曲面近似は未知係数の数が多くなるので，II．3．で述べた考慮が特に必要である。

第1表 第4図の計算結果

Table 1. Calculation results of Fig. 4

係 数	case 1	case 2	case 3
c_{00}	9.7886011−1	+9.6812189−1	7.8184389−1
c_{01}	5.9889258−2	−8.4130990−3	1.4972324−1
c_{02}	−7.7770803−2	−2.2230320−2	−3.8448447−1
c_{03}	−4.6973411−2	1.2276762−1	2.2974253−1
c_{04}	4.9360400−4	−2.2464544−1	−8.8386096−2
c_{10}	1.4056726−1	2.2717821−1	−3.9680574−1
c_{11}	−2.3107793−1	1.5112013−1	1.2411598−1
c_{12}	−2.1204863−1	−3.7933540−2	−1.9312630−1
c_{13}	4.5782987−1	−3.6430487−1	1.2205096−1
c_{14}	−5.5196320−2	9.4588682−1	−1.8895560−3
c_{20}	−3.3111822−1	−6.2430088−1	−2.3406284−1
c_{21}	1.4257700−3	−1.3552138−1	1.0511385−1
c_{22}	3.0658755−1	−1.1553688+0	−5.7699440−3
c_{23}	9.5600914−1	1.6256124+0	−5.3806046−2
c_{24}	−3.6346840−3	−1.9949172+0	−6.2126510−3
c_{30}	−7.3055300−3	5.2096555−1	6.0006606−2
c_{31}	9.3348703−1	1.5439433−1	3.2317758−2
c_{32}	1.0441359+0	2.0982042+0	−1.4425650−2
c_{33}	−9.2788260−1	−2.1076564+0	−3.1969745−3
c_{34}	−3.6419257+0	1.6896573+0	2.2634079−2
c_{40}	1.3385699−1	−2.2950134−1	−2.6784256−2
c_{41}	−4.8671238−1	1.2799400−3	−2.3269650−2
c_{42}	−1.9871858+0	−1.1874080+0	1.1958482−2
c_{43}	−1.2515326+0	9.5049160−1	1.2593075−2
c_{44}	5.2180631+0	−4.6495872−1	−1.1440900−2
$\bar{x}$			4.0172868−1
$\bar{y}$			1.4042636−1
$\bar{z}$			9.5662791−1
σ_x			2.3095141−1
σ_y			4.7812053−1
σ_z			4.2628205−2
V_z	6.0416673−5	3.7910329−6	2.0692665−6
M	7.7728163−3	1.9470575−3	1.4384945−3

（注） $-1.2515374-2=-1.2515374\times10^{-2}$



第4図　カプラン水車効率曲線

Fig. 4.　Efficiency curves of Kaplan type turbine

第4図はカプラン水車の特性で x, y, z 軸はそれぞれ回転数，流量および効率であるが，単位は使用上便利なように変えてある．この曲面について，$m=n=4$, $N=129$ で計算した結果を第1表に示す．M は $M=\sqrt{V_z}$ から求まる標準偏差で曲面の近似式の誤差を直接表わす値である．case 1 はデータを第 4 図から直接読み取って計算したものであり近似式は式 (22) で与えられる．case 2 は case 1 において，2組のデータをわざと誤って読み取って計算したものである．偶然ではあるが，V_z は case 1 にくらべて小さく結果はむしろよい。

case 3 は $(x-\bar{x})/\sigma_x$, $(y-\bar{y})/\sigma_y$, $(z-\bar{z})/\sigma_z$ をデータとして与えている．ここで

$$\bar{x}=\frac{1}{N}\sum_k x_k,\quad \bar{y}=\frac{1}{N}\sum_k y_k,\quad \bar{z}=\frac{1}{N}\sum_k z_k$$

$$\sigma_x=\sqrt{\frac{1}{N}\sum_k (x_k-\bar{x})^2},\quad \sigma_y=\sqrt{\frac{1}{N}\sum_k (y_k-\bar{y})^2},$$

$$\sigma_z=\sqrt{\frac{1}{N}\sum_k (z_k-\bar{z})^2}$$

である．近似式は

$$z=\bar{z}+\sigma_z\sum_{i=0}^{m}\sum_{j=0}^{n} c_{ij}\left(\frac{x-\bar{x}}{\sigma_x}\right)^i\left(\frac{y-\bar{y}}{\sigma_y}\right)^j \quad \cdots\cdots\cdots(26)$$

である．case 3 ではさらに誤差が小さくなっている。

なお曲面近似についても直交多項式を用いる方法もあるが，[2] 格子点でデータを与えなければならず実用上は不便な点があるのでここでは割愛した。

参考文献

(1) 森口：統計解析（現代応用数学講座）（昭 32）
(2) J.H. Cadwell: Computer Journal **3** 266 (1961)
(3) 菊池：富士時報 **35** 697（昭 37）

e-Front runners

＊本誌に記載されている会社名および製品名は，それぞれの会社が所有する
商標または登録商標である場合があります。