最小2乗法における不良データの影響*

石　川　甲 子 男**

The main purpose is the systematic study of the influences of inappropriate data in the method of least squares upon its results, where the inappropriate data mean not only the observation error but the unexpected error, for example, reading error at data producing, or punching error at input data of computer, and so on, which are not considered as the Ergodic time series.

The influences of inappropriate data are researched for each as follow;

1. a coefficent or observed value of observation equations at indirect observation,
2. a coefficent or observed value of condition equations at direct observation.

For this purpose, the analysis has been done through convenient formulas by matrices representation, and the efforts have become very economized for repeat computation.

Moreover, it is indicated that it is able to find out the positions of inappropriate data by calucuiation of the residuals in the method of the least squares.

Some examples are attached for application of this method.

1. まえがき

最小2乗法の手法は，物理実験における観測値の処理をはじめ，最近では需要予測など一般的な事象における観測値の処理に，しばしば用いられており，その適用範囲はかなり広範囲にわたっている．

最小2乗法は多数の観測データから比較的少数の事象の期待値を求めるという性質上，観測データに若干の誤差があっても結果には大きな誤差を生じなく，むしろ若干の誤差を観測誤差として予想しているのが本来の最小2乗法といえよう．

しかし，往々にして，種々の原因から観測データの一部に観測誤差以上の誤差がはいってくるおそれがあり，そのために結果が歪曲されることがある．

特に多量のデータを自動的に処理する場合には，その過程において特発的現象によって観測誤差以上の誤差が混入するおそれが多分にある．

たとえば，データ伝送過程の誤動作，計算機械にデータを送りこむまでに発生する過誤（いわゆるパンチ・ミスなど），あるいはもっと根本的には，データ採取時における誤り等々がある．

これらのことは最小2乗法のみのことではなく，一般的なデータ処理においても考えられ，このような不良データを自動的に除去する試みについては，ごく一般的なかつ概略的な事柄についてすでに一つの試みを行なった[1]．

しかし最小2乗法においては，不良データが結果におよぼす様子は比較的複雑なので，同様の方法で誤差を除去することはできない．したがって今回は，データを最小2乗法で処理する場合において不良データの結果におよぼす影響をまず調べ，次いでその除去の一方法を考えてみる．不良データが結果におよぼす影響を予め知っておくことは，結果の解釈の誤りを防ぐことにも役立てられ，またある実験において，観測データの精度を考えることにも適用される．

2. 最小2乗法の行列表示

最小2乗法では，観測値が独立でなく，いくつかの条件式に従う場合が最も一般的であるが，ここでは議論を簡単にするため，すべての観測値が独立な間接観測の場合と，観測値の間に従属関係のある直接観測の場合との二つに分けて別々に考察することにする．

一般的な場合については，これら二つが合成されたものと考えればよい．また関数関係はすべて線形であるとし，もし線形でない場合も，Taylor 展開をして線形になおしてから考えるとする．もちろん Taylor 展開できないような場合は最小2乗法の適用外であるから本論からも除外することにする．

以下の記述で一般に行列を次のように定義する．

いま r 行，t 列の行列 M を $\underset{rt}{M}$ と書く，すなわち

---

* The Influences of Inappropriate Data on The Method of Least Squares, by Kasio Isikawa (Geographical Survey Institute)

** 建設省国土地理院

$$\underset{rt}{M}=\begin{bmatrix} m_{11} & m_{12} & \cdots\cdots & m_{1t} \\ m_{21} & m_{22} & \cdots\cdots & m_{2t} \\ \cdots & \cdots & \cdots\cdots & \cdots \\ m_{r1} & m_{r2} & \cdots\cdots & m_{rt} \end{bmatrix}$$

特に

$$\underset{r1}{M}=\begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_r \end{bmatrix}$$

また $\underset{rt}{M}$ に対して $\underset{tr}{M^*}$ を $\underset{rt}{M}$ の転置行列，正方行列 $\underset{rr}{M}$ に対して $\underset{rr}{M^{-1}}$ をその逆行列とする．各行列とも特にことわらないかぎり，各要素は行列を表示する大文字に対する小文字に添字を付けたもので表わすこととする．

2.1 間接観測における最小2乗法

y_i $(i=1, 2, \cdots, m)$ なるm個の値に対して

$$\underset{mn}{A}\,\underset{n1}{T}=\underset{m1}{Y} \qquad (1)$$

なる式により t_i $(i=1, 2, \cdots, n)$ が間接に決定されるとする．ここに $\underset{mn}{A}$ は $\underset{n1}{T}$ を関係づける $m\times n$ 個の係数の行列とする．

いま y_i の代わりに観測値 l_j をいれて，それぞれの l_j が観測誤差 v_j をもっと考え，かつ t_i の期待値を x_i とすれば，

$$y_j=l_j+v_j$$

$$x_i=E(t_i)$$

これより

$$\underset{mn}{A}\,\underset{n1}{X}=\underset{m1}{L}+\underset{m1}{V}, \qquad (2)$$

ただし $m>n$，なる観測方程式ができる．

ここで l_j の各々の観測における重量（重み）を p_j とし最小2乗法により

$$\underset{1m}{V^*}\,\underset{mm}{P}\,\underset{m1}{V}=\min.$$

ただし

$$\underset{mm}{P}=\begin{bmatrix} p_1 & 0 & 0 & \cdots\cdots & 0 \\ 0 & p_2 & 0 & \cdots\cdots & 0 \\ \cdots & \cdots & \cdots & \cdots\cdots & \cdots \\ 0 & 0 & 0 & \cdots\cdots & p_m \end{bmatrix},$$

なる条件を入れて (2) 式を $\underset{n1}{X}$ について解くには，通常次の過程をへる．

すなわち

$$\underset{nn}{S}=\underset{nm}{A^*}\,\underset{mm}{P}\,\underset{mn}{A} \qquad (3)$$

および

$$\underset{n1}{K}=\underset{nm}{A^*}\,\underset{mm}{P}\,\underset{m1}{L} \qquad (4)$$

なる $\underset{nn}{S}$, $\underset{n1}{K}$ を求めて

$$\underset{nn}{S}\,\underset{n1}{X}=\underset{n1}{K} \qquad (5)$$

なる正規方程式を作り，これより $\underset{n1}{X}$ を求める．ここに $\underset{nn}{S}$ はよく知られているように対称の正方行列である．

2.2 直接観測における最小2乗法

t_j $(j=1, 2, \cdots, m)$ なる m 個の値に対して，

$$\underset{nm}{A}\,\underset{m1}{T}=\underset{n1}{L} \qquad (6)$$

ただし $m>n$ なるn個の関係があるとする．

ここに $\underset{nm}{A}$, $\underset{n1}{L}$ は $\underset{m1}{T}$ を関係づける係数の行列とする．

いまm個のt_jが直接観測されてm個のx_jを得，かつその誤差を v_j とし，t_j の期待値を改めて t_j とおくと，

$$t_j=x_j+v_j \qquad (7)$$

これより

$$\underset{nm}{A}\,(\underset{m1}{X}+\underset{m1}{V})=\underset{n1}{L} \qquad (8)$$

または

$$\underset{n1}{K}=\underset{n1}{L}-\underset{nm}{A}\,\underset{m1}{X} \qquad (9)^*$$

とおいて

$$\underset{nm}{A}\,\underset{m1}{V}=\underset{n1}{K} \qquad (10)$$

となる．

いま t_j の各々の観測における重量を p_j とし，最小2乗法により $\underset{m1}{V}$ に対して

$$\underset{1m}{V^*}\,\underset{mm}{P}\,\underset{m1}{V}=\min.$$

なるごとく解くためには

$$\underset{nn}{S}=\underset{nm}{A}\,\underset{mm}{P'}\,\underset{mn}{A^*} \qquad (11)$$

ただし $\underset{mm}{P'}$ は対角線の要素が $1/p_1, 1/p_2, \cdots, 1/p_m$ である対角行列，および (10) 式より

$$\underset{nn}{S}\,\underset{n1}{W}=\underset{n1}{K}$$

なる正規方程式を作成し，これを $\underset{n1}{W}$ について解いて，

$$\underset{mn}{A^*}\underset{n1}{W}=\underset{mm}{P}\,\underset{m1}{V} \qquad (12)$$

なる，いわゆる相関式に代入すれば，求める観測値の補正量としての $\underset{m1}{V}$ が求まる．この $\underset{m1}{V}$ を (7) 式に代入して所要の期待値のベクトル $\underset{m1}{T}$ が得られる．

以上を準備として次に本論に入る．

3. 最小2乗法における不良データ

念のために，ここでいう不良データとは，前に記したように，通常の観測誤差以上に大きな誤差をもつデ

* 議論を簡単にするため方程式は $\underset{m1}{X}+\underset{m1}{V}$ に関して線形としたが，実際には $\underset{m1}{V}$ に関してのみ線形であればこの方程式の最小2乗解は得られる．実際の計算にはしばしば $\underset{m1}{V}$ に関してのみ線形になるように工夫して複雑な最小2乗法の解を求めている．したがって，そのような場合のことを考えれば，(9) 式の $\underset{nm}{A}\,\underset{m1}{V}=\underset{n1}{K}$ の形から議論をはじめた方がよい場合もある．

ータをいう．これはたとえば，自動データ処理の過程における突発的現象によるものなどを考え，計算時に入力データとなるものは，すべてその対象となるので，必ずしも観測値のみではない．すなわち (2) 式における $\underset{mn}{A}, \underset{m1}{L}$ の各要素，(8) 式の $\underset{nm}{A}, \underset{n1}{L}$ の各要素のすべてに不良データが混入する可能性がある．

したがって，ここではその両方の場合について考えることとする．

3.1 間接観測の場合

3.1.1 係数に不良データがある場合

行列 $\underset{mn}{A}$ の一つの要素 a_{pq} が不良データであって，$a_{pq}+\mu$ となったと考える．このとき行列 $\underset{mn}{A}$ が

$$\underset{mn}{A}+\underset{mn}{\Delta A}$$

となったとする．ここに $\underset{mn}{\Delta A}$ は第 p 行第 q 列目に μ があって，他はすべて 0 を要素とする行列である．

これにより解である期待値のベクトル $\underset{n1}{X}$ が $\underset{n1}{X}+\underset{n1}{\Delta X}$ になり，同時に観測誤差 $\underset{m1}{V}$ が $\underset{m1}{V}+\underset{m1}{\Delta V}$ になったとすれば (2) 式は

$$(\underset{mn}{A}+\underset{mn}{\Delta A})(\underset{n1}{X}+\underset{n1}{\Delta X})=\underset{m1}{L}+\underset{m1}{V}+\underset{m1}{\Delta V} \tag{13}$$

となる．これより (3) (4) 式は

$$\underset{nn}{S}+\underset{nn}{\Delta S}=\underset{nm}{(A+\Delta A)^*}\underset{mm}{P}(\underset{mn}{A}+\underset{mn}{\Delta A}) \tag{14}$$

$$\underset{n1}{K}+\underset{n1}{\Delta K}=\underset{nm}{(A+\Delta A)^*}\underset{mm}{P}\,\underset{m1}{L} \tag{15}$$

となり，したがって (5) 式は

$$(\underset{nn}{S}+\underset{nn}{\Delta S})(\underset{n1}{X}+\underset{n1}{\Delta X})=\underset{n1}{K}+\underset{n1}{\Delta K} \tag{16}$$

となる．これらの式中 Δ のついている行列は誤差 μ によって各行列が変化した分である．

いま (16) 式より $\underset{n1}{X}$ の変化分 $\underset{n1}{\Delta X}$ の値を求めると

$$\underset{nn}{S}\,\underset{n1}{X}+\underset{nn}{S}\,\underset{n1}{\Delta X}+\underset{nn}{\Delta S}\cdot\underset{n1}{\Delta X}=\underset{n1}{K}+\underset{n1}{\Delta K}$$

(5) 式により

$$\underset{nn}{S}\underset{n1}{\Delta X}+\underset{nn}{\Delta S}\,\underset{n1}{X}+\underset{nn}{\Delta S}\cdot\underset{n1}{\Delta X}=\underset{n1}{\Delta K}$$

$$\underset{n1}{\Delta X}=(\underset{nn}{S}+\underset{nn}{\Delta S})^{-1}(\underset{n1}{\Delta K}-\underset{nn}{\Delta S}\cdot\underset{n1}{X}) \tag{17}$$

または

$$\underset{n1}{X}+\underset{n1}{\Delta X}=(\underset{nn}{S}+\underset{nn}{\Delta S})^{-1}(\underset{n1}{K}+\underset{n1}{\Delta K})$$

$$\underset{n1}{\Delta X}=(\underset{nn}{S}+\underset{nn}{\Delta S})^{-1}(\underset{n1}{K}+\underset{n1}{\Delta K})-\underset{n1}{K} \tag{18}$$

であるが，この (17) または (18) を計算するよりは普通の方法で (16) 式より $\underset{n1}{X}+\underset{n1}{\Delta X}$ を求めてから，$\underset{n1}{\Delta X}$ を求めた方がかえって簡単である．

ここで対称行列 $\underset{nn}{S}$ の変化分の行列 $\underset{nn}{\Delta S}$ およびベクトル $\underset{n1}{K}$ の変化分 $\underset{n1}{\Delta K}$ を求めてみる．まず (14) 式より

$$\begin{aligned}\underset{nn}{S}+\underset{nn}{\Delta S}&=\underset{nm}{(A+\Delta A)^*}\underset{mm}{P}(\underset{mn}{A}+\underset{mn}{\Delta A})\\&=(\underset{mm}{A^*}+\underset{nm}{\Delta A^*})\underset{mm}{P}(\underset{mn}{A}+\underset{mn}{\Delta A})\\&=\underset{nm}{A^*}\underset{mm}{P}\underset{mn}{A}+\underset{nm}{A^*}\underset{mm}{P}\underset{mn}{\Delta A}+\underset{nm}{\Delta A^*}\underset{mm}{P}\underset{mn}{A}\\&\quad+\underset{nm}{\Delta A^*}\underset{mm}{P}\underset{mn}{\Delta A}\end{aligned}$$

(3) 式を代入して

$$\underset{nn}{\Delta S}=\underset{nm}{A^*}\underset{mm}{P}\underset{mn}{\Delta A}+\underset{nm}{\Delta A^*}\underset{mm}{P}\underset{mn}{A}+\underset{nm}{\Delta A^*}\underset{mm}{P}\underset{mn}{\Delta A} \tag{19}$$

誤差 μ が行列 $\underset{mn}{A}$ の第 p 行第 q 列目に加わったと考えれば ΔA の要素は第 p 行第 q 列目のみで他はゼロであるから(19)式の第 1 項が影響を受けるのは q 列のみ，

第 2 項が影響を受けるのは q 行のみ，

第 3 項が影響を受けるのは第 q 行第 q 列目の要素一つのみであるから

$$\underset{nn}{S}=\begin{array}{c|c} & \begin{array}{ccc} & q & \end{array}\\ \hline \\ q & \left(\begin{array}{ccc} & \delta s_{1q} & \\ & \vdots & \\ 0 & \delta s_{2q} & 0\\ & \vdots & \\ \delta s_{q1}\cdots\cdots & \delta s_{qq} & \cdots\cdots\delta s_{qn}\\ 0 & \vdots & 0\\ & \delta s_{nq} & \end{array}\right)\end{array} \tag{20}$$

ただし

$\delta s_{kq}=\mu a_{pk},\quad k=1,2,\cdots q-1,\quad k\neq q$

$\delta s_{qq}=2\,\mu a_{pq}+\mu^2,$

$\delta s_{qk}=\mu a_{pk},\quad k=q+1,q+2,\cdots,n,\quad k\neq q$

となる．

同様に ΔK については (15) 式より

$$\begin{aligned}\underset{n1}{K}+\underset{n1}{\Delta K}&=\underset{nm}{(A+\Delta A)^*}\underset{mm}{P}\,\underset{m1}{L}\\&=\underset{nm}{A^*}\underset{mm}{P}\,\underset{m1}{L}+\underset{nm}{\Delta A^*}\underset{mm}{P}\,\underset{m1}{L}\end{aligned} \tag{21}$$

(4) 式により

$$\underset{n1}{\Delta K}=\underset{nm}{\Delta A^*}\underset{mm}{P}\,\underset{m1}{L}$$

で，

$$\underset{nm}{\Delta A^*}\underset{mm}{P}\,\underset{m1}{L}=\begin{array}{c}\\ \\ q\\ \\ \\ \\ \end{array}\left(\begin{array}{c}0\\ \vdots\\ 0\\ \delta k_q\\ 0\\ \vdots\\ 0\end{array}\right) \tag{22}$$

ただし $\delta k_q=\mu y_q$，となる．

(20)，(22) 式より誤差 μ のため $\underset{nn}{S}$ および $\underset{n1}{K}$ が影響を受けるのは，$\underset{nn}{S}$ では q 行 q 列，$\underset{n1}{K}$ では q 行目の要素のみで他は関係ない．この関係を利用すると，(16) 式で μ にいくつかの値を代入して，いくとおりかの $\underset{n1}{\Delta X}$ を求める場合に大変便利である．すなわち，

あらかじめμの入るべきp行q列の要素をすべて最右端，最下行のm行，n列に移行し，したがって行列 $\underset{n1}{\Delta S}$ の要素が０でない行と列をn行n列に移行し，かつ $\underset{n1}{\Delta K}$ のq番目の要素をn番目に移行しておくとよい．なぜなら実際に(16)式の消去を実行する場合は，左上端の消去からはじめて右方下段の方へ消去をしてゆくので，この場合は，$n-1$行，$n-1$列までの消去過程の情報は毎回そのままで，n行n列の消去に進められるからである．つまりμを変えた場合 n行n列の計算のみを入れ変えれば所要の結果が得られ，大幅に計算量の節約ができる．一般に(16)式を始めから解くのはnの約３乗に比例する労力を要しrとおりのμに対する解を求めるにはrn^3に比例する労力を必要とすることになる．しかし，いまのべた方法により，誤差 μ の増加に対する影響を計算するのであれば，$[n^3-(n-1)^3]r+(n-1)^3$ の労力ですみ，その差は

$$(r-1)(n-1)^3$$

となって比較的容易に計算できることになる．

特にこのような計算に便利な消去の方法として，Bjerhammer の計算表がある[3]．

3.1.2 観測値に不良データがある場合

ベクトル $\underset{m1}{Y}$ の一つの要素 y_q に不良データがあって $y_q+\mu$ となった場合を考える．すなわち $\underset{m1}{Y}$ が $\underset{m1}{Y}+\underset{m1}{\Delta Y}$ となったとする．ここに $\underset{m1}{\Delta Y}$ はq番目の要素がμ，他はすべて０である．したがって(4)式は

$$\begin{aligned}\underset{n1}{K}+\underset{n1}{\Delta K}&=\underset{nm}{A^*}\underset{mm}{P}(\underset{m1}{L}+\underset{m1}{\Delta L})\\&=\underset{nm}{A^*}\underset{mm}{P}\underset{m1}{L}+\underset{nm}{A^*}\underset{mm}{P}\underset{m1}{\Delta L}\\\underset{n1}{\Delta K}&=\underset{nm}{A^*}\underset{mm}{P}\underset{m1}{\Delta L}\end{aligned}\tag{23}$$

正規方程式は

$$\underset{nn}{S}(\underset{n1}{X}+\underset{n1}{\Delta X})=\underset{n1}{K}+\underset{n1}{\Delta K}$$

であるから，

$$\underset{nn}{S}\cdot\underset{n1}{\Delta X}=\underset{n1}{\Delta K}\tag{24}$$

$$\underset{n1}{\Delta X}=\underset{nn}{S^{-1}}\underset{n1}{\Delta K}\tag{25}$$

この場合，種々の μ に対して $\underset{n1}{\Delta K}$ が変化するのみで $\underset{nn}{S}$ は変わらない．したがってこの場合は，あらかじめ $\underset{nn}{S}$ の逆行列 $\underset{nn}{S^{-1}}$ を求めておけば，あとは行列の乗算のみで種々の $\underset{n1}{\Delta K}$ に対する種々の $\underset{n1}{\Delta X}$ を求めることができる．

はじめにのべたように，最小２乗法で不良データのはいる可能性が多く，かつ重要なのはこの場合であって，この場合における計算を行なう必要が最も多い．しかもこの場合の計算の反復は，前節にくらべてはるかに簡単である．ここに行列 $\underset{nn}{S^{-1}}$ は一つの観測系に対しては固定しているから，後述の直接観測の場合の行列 $\underset{mn}{H}$ とともに，ある観測系についての**特性行列**と名付けておくことにする．

3.2 直接観測の場合

3.2.1 係数に不良データがある場合

3.1.1 の場合と同様に，行列 $\underset{nm}{A}$ の一つの要素 a_{qp} が不良データであって $a_{qp}+\mu$ となったと考える．このとき行列 $\underset{nm}{A}$ が $\underset{nm}{A}+\underset{nm}{\Delta A}$ となったとする，ここに $\underset{nm}{\Delta A}$ は第q行第p列目のみμで，他はすべて０の要素をもつ行列である．これにともなって未定係数の行列 $\underset{n1}{W}$ が $\underset{n1}{W}+\underset{n1}{\Delta W}$ になったとする．これより(9)式は

$$(\underset{nm}{A}+\underset{nm}{\Delta A})^*(\underset{n1}{W}+\underset{n1}{\Delta W})=\underset{mm}{P}(\underset{m1}{V}+\underset{m1}{\Delta V})\tag{26}$$

となる．いま最小２乗法の原理にしたがって

$$(\underset{m1}{V}+\underset{m1}{\Delta V})^*(\underset{m1}{V}+\underset{m1}{\Delta V})=\min.$$

として解くと(10)式は

$$\underset{nn}{S}+\underset{nn}{\Delta S}=(\underset{nm}{A}+\underset{nm}{\Delta A})\underset{mm}{P'}(\underset{nm}{A}+\underset{nm}{\Delta A})^*$$

となり，また

$$\underset{n1}{K}+\underset{n1}{\Delta K}=\underset{n1}{L}-(\underset{nm}{A}+\underset{nm}{\Delta A})\underset{m1}{X}\tag{27}$$

となって，正規方程式(11)式は

$$(\underset{nn}{S}+\underset{nn}{\Delta S})(\underset{n1}{W}+\underset{n1}{\Delta W})=\underset{n1}{K}+\underset{n1}{\Delta K}\tag{28}$$

となる．これより形式的に

$$\underset{n1}{\Delta W}=\underset{n1}{\Delta K}-(\underset{nn}{S}+\underset{nn}{\Delta S})^{-1}\cdot\underset{nn}{\Delta S}\cdot\underset{n1}{W}$$

と書けるが，3.1.1 節と同様，$\underset{n1}{W}+\underset{n1}{\Delta W}$ を(28)式から求め，一方 $\underset{n1}{W}$ を独立に求めて差をとった方がかえって簡単であろう．しかしこの場合も 3.1.1 で行なった考察はそのままあてはまり，(16)式を消去法で解く場合と同様に，q行p列を入れかえて最右下端の行と列にして消去を行なえば，計算量を大幅に節減できる．

3.2.2 観測値に不良データがある場合

観測値に不良データがあった場合，すなわち第(8)式におけるベクトル $\underset{m1}{X}$ のq番目の要素 x_q が $x_q+\mu$ になって，

$$\underset{m1}{X}\to\underset{m1}{X}+\underset{m1}{\Delta X}$$

となった場合を考える．このためベクトル $\underset{m1}{V}$ が $\underset{m1}{V}+\underset{m1}{\Delta V}$ になったとすると(8)式より

$$\underset{nm}{A}(\underset{m1}{X}+\underset{m1}{\Delta X}+\underset{m1}{V}+\underset{m1}{\Delta V})=\underset{n1}{L}\tag{29}$$

これより

$$(\underset{m1}{V}+\underset{m1}{\Delta V})^*\underset{mm}{P'}(\underset{m1}{V}+\underset{m1}{\Delta V})=\min.$$

なるように解くと，結局(11)式の$\underset{nn}{S}$は不変，$\underset{n1}{K}$ は

$$\underset{n1}{K}+\underset{n1}{\Delta K}=\underset{n1}{L}-\underset{nm}{A}(\underset{m1}{X}+\underset{m1}{\Delta X})\tag{30}$$

となって第(12)式は

$$\begin{aligned}\underset{mm}{P}(\underset{m1}{V}+\underset{m1}{\Delta V})&=\underset{mn}{A^*}\underset{nn}{S^{-1}}(\underset{n1}{K}+\underset{n1}{\Delta K})\\&=\underset{mn}{A^*}\underset{nn}{S^{-1}}\{\underset{n1}{L}-\underset{nm}{A}(\underset{n1}{X}+\underset{n1}{\Delta X})\}\end{aligned}\tag{31}$$

となる．これより

$$\underset{mm}{P}\cdot\underset{m1}{\Delta V}=\underset{mn}{A^*}\cdot\underset{nn}{S^{-1}}\cdot\underset{n1}{\Delta K} \tag{32}$$

または
$$=-\underset{mn}{A^*}\cdot\underset{nn}{S^{-1}}\cdot\underset{nm}{A}\cdot\underset{m1}{\Delta X} \tag{33}$$

となる*.

いま (33) 式において $\underset{nm}{A}$ は $\underset{m1}{X}$ の変化にかかわらず不変，したがって $\underset{nn}{S}$ は不変，それから導びかれる $\underset{mn}{A^*}$, $\underset{nn}{S^{-1}}$ とも不変なので，結局 $\underset{m1}{X}$ の変化にかかわらず $-\underset{mn}{A^*}\underset{nn}{S^{-1}}\underset{nm}{A}$ は一定となる．いまこれを

$$\underset{mm}{H}=-\underset{mn}{A^*}\cdot\underset{nn}{S^{-1}}\cdot\underset{nm}{A} \tag{34}$$

とおいて 2.1.2 の $\underset{nn}{S^{-1}}$ と同様に元の方程式群の**特性行列**と呼ぶことにする．条件式における最小2乗法においても，この行列 $\underset{mm}{H}$ が分かれば，$\underset{m1}{X}$ の変化分 $\underset{m1}{\Delta X}$ に対してベクトル $\underset{m1}{\Delta V}$ を簡単に計算することができる．

4. いくつかの不良データがある場合

いままでは不良データとして特定の要素に μ なる値の誤差が加わったときの種々の場合について調べたものであった．一般に不良データの影響を調べるという目的には，特定のデータが不良になったと仮定して，その影響を調べるのが普通であるが，一方 Simulation の観点から一時に多数個の不良データが混入した場合の影響を調べておくこともしばしば必要となろう．

もしいくつかの要素に不良データが混入して，種々の μ_j $(j=1, 2, 3, \cdots, r)$ なる誤差が加わったときは，前節までの事柄が重ね合わされた形になると考えればよい．ただし，その場合は 3.1.1 または 3.2.1 におけるように消去の過程で計算量を節減することは望めない．しかし観測値そのものに不良データがあった場合なら 3.1.2, 3.2.2 のように，たとえ多数個の不良データがあったとしても，1個の場合とほとんど同様の計算量で処理できることは (25) 式または (33) 式により明らかである．Simulation が必要なのはむしろこの場合であって，多数個の誤差 μ_j を乱数表から与えてやることにより，観測値の乱れが結果に及ぼす影響を算出することができる．

* (33) 式の両辺に $\underset{nm}{A}$ をかけると

$$\underset{nm}{A}\cdot\underset{mm}{P}\cdot\underset{m1}{\Delta V}=-\underset{nm}{A}\cdot\underset{m1}{\Delta X}$$

となり，みかけ上 $\underset{mm}{P}\cdot\underset{m1}{\Delta V}=-\underset{m1}{\Delta X}$ となるが，矩形行列 $\underset{nm}{A}$ の逆行列を作るには特別の操作が必要[2]なので一般に $\underset{nm}{A}\,\underset{m1}{B}=\underset{nm}{A}\,\underset{m1}{C}$ のとき $\underset{m1}{B}=\underset{m1}{C}$ とはならないことが知られている．

5. 不良データによる標準偏差の増加

次に不良データによる標準偏差の変化について考える．このため第 (2) 式における $\underset{m1}{V}$, または第 (8) 式における $\underset{m1}{V}$ に対して，逆に最小2乗法を解いて得た結果を代入して求めたいわゆる残差または補正量なるものを改めて $\underset{m1}{V}$ とおき，その不良データによる増加分 $\underset{m1}{\Delta V}$ とおくことにする．

間接観測については，$\underset{mn}{A}$ が $\underset{mn}{A}+\underset{mn}{\Delta A}$ となって $\underset{n1}{X}$ が $\underset{n1}{X}+\underset{n1}{\Delta X}$ となったときは (2) 式より

$$\underset{m1}{V}+\underset{m1}{\Delta V}=(\underset{mn}{A}+\underset{mn}{\Delta A})(\underset{n1}{X}+\underset{n1}{\Delta X})-\underset{m1}{L}$$

これより

$$\underset{m1}{\Delta V}=\underset{mn}{\Delta A}\cdot(\underset{n1}{X}+\underset{n1}{\Delta X})+\underset{mn}{A}\cdot\underset{n1}{\Delta X} \tag{35}$$

$\underset{m1}{L}$ が $\underset{m1}{L}+\underset{m1}{\Delta L}$ となって $\underset{n1}{X}$ が $\underset{n1}{X}+\underset{n1}{\Delta X}$ となった場合は

$$\underset{m1}{V}+\underset{m1}{\Delta V}=\underset{mn}{A}(\underset{n1}{X}+\underset{n1}{\Delta X})-(\underset{m1}{L}-\underset{m1}{\Delta L})$$

これより

$$\underset{m1}{\Delta V}=\underset{mn}{A}\cdot\underset{n1}{\Delta X}-\underset{m1}{\Delta L} \tag{36}$$

次に直接観測については，$\underset{nm}{A}$ が $\underset{nm}{A}+\underset{nm}{\Delta A}$ となって $\underset{n1}{W}$ が $\underset{n1}{W}+\underset{n1}{\Delta W}$ となった場合は (12) 式より

$$\underbrace{(A+\Delta A)^*}_{mn}(\underset{n1}{W}+\underset{n1}{\Delta W})=\underset{mm}{P}(\underset{m1}{V}+\underset{m1}{\Delta V})$$

これより

$$\underset{mm}{P}\cdot\underset{m1}{\Delta V}=\underset{mn}{A^*}\underset{n1}{W}+\underset{mn}{\Delta A^*}(\underset{n1}{W}+\underset{n1}{\Delta W}) \tag{37}$$

また $\underset{m1}{X}$ が $\underset{m1}{X}+\underset{m1}{\Delta X}$ となった場合はすでに求めた (32) 式と同じになって

$$\underset{mm}{P}\cdot\underset{m1}{\Delta V}=-\underset{mn}{A^*}\underset{nn}{S^{-1}}\cdot\underset{nm}{A}\cdot\underset{m1}{\Delta X} \tag{38}$$

となる．これら (35) (36) (37) および (38) 式中の増加分 $\underset{m1}{\Delta X}$ または $\underset{n1}{\Delta W}$ は各要素が一様に変化する傾向にある中で，$\underset{mn}{\Delta A}$, $\underset{m1}{\Delta L}$ などは元になるデータでかつ不良データが混入する要素は特定の q 番目または pq 番目の要素のみであると仮定したので，これらの行列は特定の要素のみしかなく，したがって上の各式は何れも q 番目の要素のみが他より変化の度合が著しいと考えられる．したがって全体として $\underset{1m}{V^*}\underset{mm}{P}\underset{m1}{V}$ は増加し，その増加は不良データの変化分 μ に関係している．

いま標準偏差は間接観測に対しては

$$\sigma_I=\pm\sqrt{\nu/(m-n)}$$

直接観測に対しては

$$\sigma_D=\pm\sqrt{\nu/n}$$

ここに $\nu=\underset{1m}{V^*}\underset{m1}{P}\underset{m1}{V}$

で与えられ，μ の間接的な関数となる．

これらの事柄を逆に考えれば，いまある最小2乗解を求めたとき，σ_I または σ_D が予想よりはるかに大き

な値を示しているとすれば，元のデータに不良データが混入していたことが推定できる．かつ $\underset{m1}{V}$ の各要素を調べて q 番目の v_q が他の v_j と比べて著しく大きいときは，その行の要素を疑うことによって，不良データを速やかに発見することができよう．

一般に，ある計算から求めた解が不適当であって，もとのデータが疑われたとき，その不良データを発見するのは，なかなか容易ではないが，最小２乗解において以上の事柄を適用してよい効果を上げることができる．このことは従来からも経験的にある程度行なわれていたが，多くは人間の高度の判断を必要としていた．しかしすでに前節までに記した方法によって，予め不良データの影響を調べておくことにより，かなり正確に機械的判断の資料を提出できる．電子計算機などによって自動データ処理を行なう場合は，このような判定条件をあらかじめプログラムに組み込んでおくことによって不良データの発見または除去を自動的に行なうことができる．

6. 応　用　例

以上の事柄の応用例を 2，3 あげる．実際の問題は行列の次数 m, n とも相当大きなものが対象になるがここではいたずらに紙面を費やすので，各項目を代表する程度で，なるべく小さい次数の行列を例としてあげる．

6.1 間接観測の例

観測方程式が **Table 1.** ($m=24$, $n=4$) のようなも

Table 1. Example 1 for Observation Equations $m=24$, $n=4$

No.	p	a_1	a_2	a_3	a_4	l
1	1.0	0.5766	1.1780	0	0	0.42*
2	1.0	−0.6954	−0.5646	0	0	0.52
3	1.0	0.1188	−0.6133	0	0	1.74
4	1.0	1.1241	−1.2660	−1.0049	0.6533	−1.34
5	1.0	−0.1188	0.6133	0.2147	−0.3757	−0.53
6	1.0	−1.0053	0.6527	0.7902	−0.2776	−2.92
7	1.0	−1.7007	0.0880	1.0049	−0.6533	0.92
8	1.0	1.0053	−0.6527	−0.2500	0.7590	−0.25
9	1.0	0.6954	0.5646	−0.7549	−0.1057	0.05
10	1.0	0	0	0.5403	0.4814	−3.17
11	1.0	0	0	−0.7549	−0.1057	0.57
12	1.0	0	0	0.2147	−0.3757	1.21
13	1.0	0	0	1.1395	−0.4696	−5.33
14	1.0	0	0	−0.2147	0.3757	1.67
15	1˙0	0	0	−0.9249	0.0939	6.52
16	1.0	0	0	−1.0672	−0.2918	4.84
17	1.0	0	0	0.9249	−0.0939	−3.49
18	1.0	0	0	0.1424	0.3856	−2.66
19	1.0	0	0	−0.6126	0.2799	3.66
20	1.0	0	0	−0.1424	−0.3856	−0.70
21	1.0	0	0	0.7549	0.1057	1.45
22	1.0	0	0	−1.6798	−0.0118	8.50
23	1.0	0	0	0.9249	−0.0939	−3.75
24	1.0	0	0	0.7549	0.1057*	−2.06

のについて，不良データが表の＊印の所にあって誤差 μ が加わったための係数の変化に対する解の変化および観測値の変化に対する解の変化の状態を図示したのが **Fig. 1** および **Fig. 2** である．この場合の標準偏差 σ_{I} の変化の状態は **Fig. 3** に示す．

また同様の例であるが，典形的な曲線決定の問題

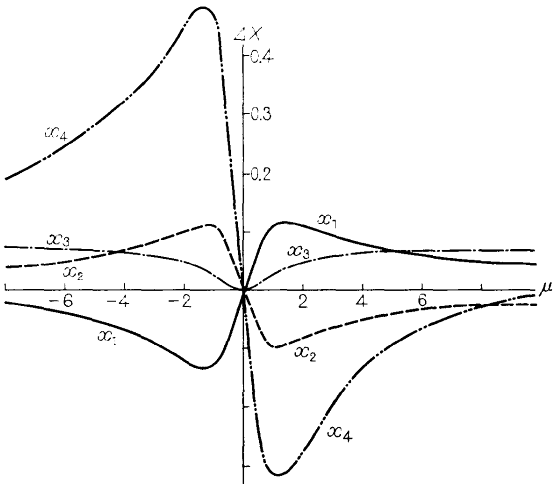

Fig. 1　Variation of Solutions due to Variation of a Coefficient at Example 1.

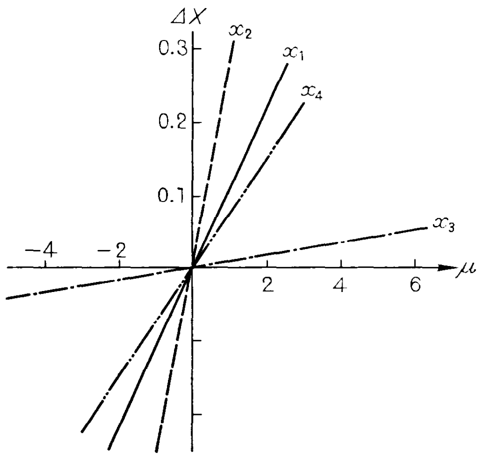

Fig. 2　Variation of the Solutions due to Variation of Observed Value at Example 1.

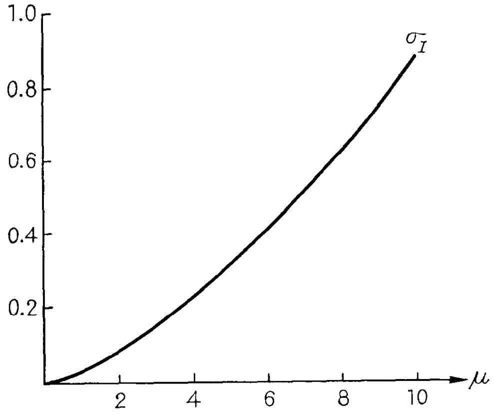

Fig. 3 Increasing of Standard Deviation due to Inappropriate Data at Example 1.

Table 2 Example 2 for Curve Fitting

No.	p	a	b	c	l
1	1	1	1	1	15.1
2	1	1	2	4	19.0
3	1	1	3	9	23.5
4	1	1	4	16	28.2
5	1	1	5	25	33.8
6	1	1	6	36	39.7
7	1	1	7	49	46.2
8	1	1	8	64	53.5
9	1	1	9	81	61.7
10	1	1	10	100	70.4

Table 2. について不良データの影響により曲線が如何に歪められるかを示したのが **Fig. 4** である．与えるデータの左端のデータを大きく変えることにより，決定される曲線が大きく歪曲されることが分かる．

6.2 直接観測の例

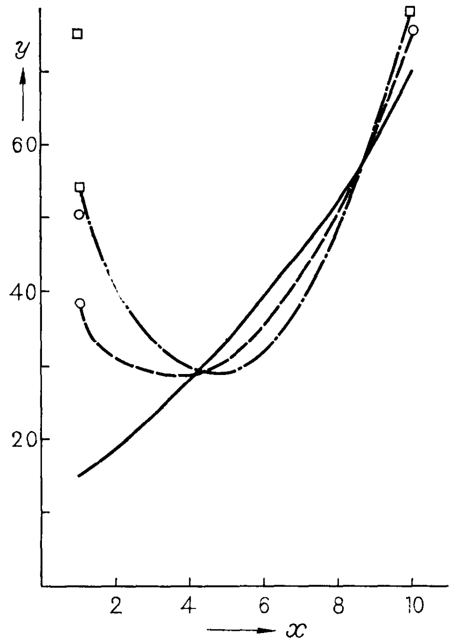

Fig. 4 Distortion of a Curve due to Inappropriate Data.

条件式が **Table 3.** ($m=14$, $n=3$) のようなものについて，観測値に不良データがあって誤差 μ が加わったときその μ の変化により解が如何に変化するかを図示したのが **Fig. 5** および **Fig. 6** である．またこの例での標準偏差 σ_D の変化の状態は **Fig. 7** に示してある．

6.3 やや大きな例

元数のやや大きなものの例として **Table 4.** にあげ

Table. 3. Example 3 for Condition Equations

No.	A														No.	X	L
	1	2	3	4	5	6	7	8	9	10	11	12	13	14			
1	−1	+1		−1	+1							−1	+1		1	0	3.141 619 425
2					−1	+1				−1	+1	−1		+1	2	1.084 016 123*	3.141 625 020
3			−1			+1	−1	+1	−1	+1					3	0	3.141 621 243
															4	3.661 643 868	
															5	4.662 245 603	
															6	0.592 912 941	
															7	0.643 068 163	
															8	1.684 652 924	
															9	0	
															10	1.507 120 069	
															11	2.425 320 122	
															12	0	
															13	1.056 996 925	
															14	1.195 394 011	

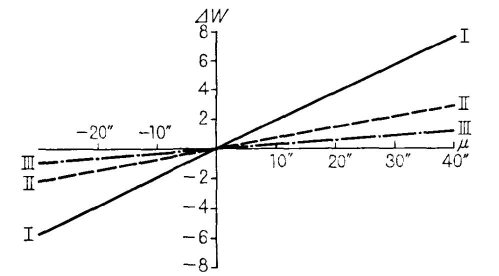

Fig. 5 Variation of Undetermined Coefficients due to Variation of Observed Value at Example 3.

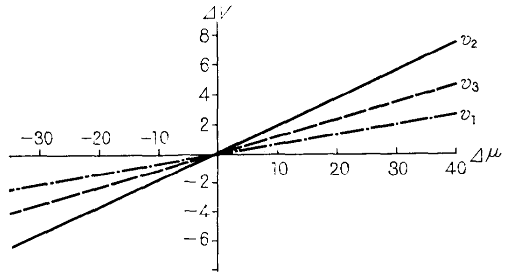

Fig. 6 Variations of Some Correction Terms due to Variation of Observed Value at Example 3.

たものは $m=16$, $n=10$ の観測方程式の例である．この程度の元数になると電子計算機を用いても多数回繰返すのは，それほど簡単ではないが，本論により消去過程を工夫することによって，合計計算時間を大幅に減らすことができる．この場合のいくつかの μ に対する解の値を **Table 5.** に示す．

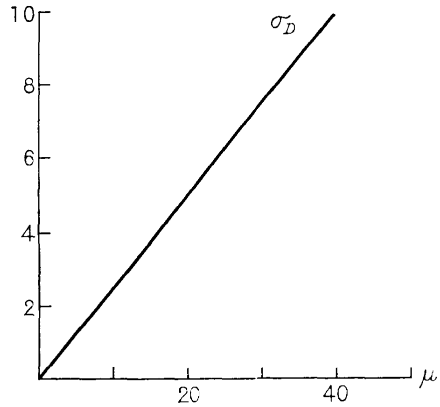

Fig. 7 Increase of Standard Deviation due to Inappropriate Data at Example 3.

Table 5. Increasing of Solution due to Inappropriate Data at Example 4

No.	X	ΔX		
		$\mu=+1.0$	$\mu=+2.0$	$\mu=+10.0$
1	0.1028	0.0249	0.0497	0.2486
2	−0.1719	−0.1203	−0.2405	−1.2025
3	−0.0054	0.0156	0.0493	0.1559
4	−0.2564	−0.0619	−0.1238	−0.6191
5	−0.0783	−0.0154	0.0308	0.1542
6	−0.4733	−0.0427	−0.0855	−0.4273
7	0.3137	0.0026	0.0052	0.0259
8	0.4326	−0.0120	−0.0240	−0.1198
9	−0.1677	0.0087	0.0174	0.0868
10	−0.5068	−0.0330	−0.0660	−0.3302

Table 4. Example 4 for Observation Equations. $m=16$, $n=10$

No.	P	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	L
1	0.2	1.62805	−6.21495	0	0	0	0	0	0	0	0	1.17109*
2	0.2	0.61496	4.58147	−0.61496	−4.58147	0	0	0	0	0	0	−0.00753
3	0.2	0	0	4.52758	2.94760	−4.52758	−2.94760	0	0	0	0	0.13739
4	0.2	0	0	0	0	0	0	0	0	1.61613	−2.28878	−0.28613
5	0.2	0	0	0	0	3.73985	4.20740	0	0	−3.73985	−4.20740	0.17053
6	0.2	0	0	0	0	0	0	1.35474	9.22402	0	0	4.11293
7	0.2	0	0	0	0	−0.19087	5.86162	0.19087	−5.86162	0	0	−5.70989
8	1.0	6.19672	1.62328	0	0	0	0	0	0	0	0	0.68506
9	1.0	−10.76461	−1.01014	4.56789	−0.61314	0	0	0	0	0	0	−0.61785
10	1.0	4.56789	−0.61314	−7.50667	5.12717	2.93878	−4.51404	0	0	0	0	1.21320
11	1.0	0	0	0	0	0	0	0	0	2.28192	1.61128	−1.23537
12	1.0	0	0	0	0	−4.19473	3.72859	0	0	1.91281	−5.33987	1.09434
13	1.0	0	0	−2.93878	4.51404	7.13351	−8.24262	0	0	−4.19473	3.72859	1.02499
14	1.0	0	0	0	0	0	0	−9.19598	1.35062	0	0	−2.26171
15	1.0	0	0	0	0	−5.84395	−0.19029	15.03992	−1.16033	0	0	4.80347
16	1.0	0	0	−2.93878	4.51404	8.78272	−4.32375	−5.84395	−0.19029	0	0	−1.67523

7. むすび

以上最小2乗法において不良データがどのように影響するか，またそれを計算するときの便法，さらに不良データを除去する方法などについて，最小2乗法の典形的ないくつかの形に分けて考えてみた．

一般に観測値を処理する場合，特に自動的に処理する場合，不良データの問題がしばしば隘路となる．著者自身も種々の場合についてその問題の解決に腐心しているが，今回は特に最小2乗法について焦点を当ててみた．

なお具体的な応用例として，測地計算における自動データ処理に適用した例[4,5]があることを付記する．

参考文献

1) 石川甲子男；自動データ処理における不良データの除去，経営科学，Vol. 4, No. 3 (1961), pp. 146.

2) A. Bjerhammer; Rectangular Reciprocal Matrices with Special Reference to Geodetic Calculatin, Bulletion Géodesique No. 19～22 (1951), pp. 188～220.

3) L.A. Gale; A Modified-equations method for the Least Squares Solution of Condition equations, Transaction Amer. Geophysical Union., Vol. 36, No. 36 (1955).

4) 石川甲子男・市原満；三角測量平均計算における不良データの影響，I. 座標平均計算，測地学会誌 Vol. 12, No. 3～4 (1967), pp. 127.

5) 石川甲子男；三角測量平均計算における不良データの影響 II. 網平均計算，測地学会誌 Vol. 13, No. 2 (1968).

(昭和43年2月22日受付，同3月30日再受付)