提供：
Optical Solutions
株式会社オプティカルソリューションズ
〒101-0032
東京都千代田区岩本町 2-15-8（MAS 三田ビル 3 階）
TEL：03-5833-1332　FAX：03-3865-3318
http://www.osc-japan.com/
e-mail:info@osc-japan.com

光学設計ノーツ　５３（ver.1.0）

最適化とは、最小二乗法について

株式会社タイコ
牛山善太

前回説明させていただいた様に、ニュートン-ラフソン法は収束も非常に早く、原理的にも理解しやすく有用な手法であるが、エンジニアリング的な分野では、繰り返し実行されなければならない、ヘッセ行列の計算、つまり２次微分の計算に困難が発生する。この困難を克服するために、ヘッセ行列の代わりの２次微分を必要としない適当な行列 B を用いる準ニュートンと呼ばれる手法も存在する。しかし、この代替えのヘシアンに相当する行列も、関数が複雑になれば、或いは関数自体を得ることが不可能な場合には、利用することが困難になる。今回はそうした場合に用いられる最小二乗法について触れさせて戴きたい。

１．　連立方程式の解

以下の様な n 元の連立方程式があるとき、

$$
\begin{gathered}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\
\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m
\end{gathered}
\tag{1}
$$

行列で表現すれば、

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \qquad (1B)$$

或いは、行列 A、ベクトル $\mathbf{x}$、B を用いて表現すれば、

$$\mathbf{Ax} = \mathbf{B} \qquad (1C)$$

であるが、n<m の場合にこの方程式を解くことを考えると、特別な場合を除き(1)式は解を持たない。特別な場合とは、線形独立な式がｎ個以下であって、残りの式は線形従属である場合である。線形独立とは互いに、線形結合で表現できないことである。線形結合とは、式を定数倍したり、或いは加え合わしたりすることを言う。

分かりやすい例では、3 次元空間を考えると、そこに独立の（線形独立した）２本のベクトル（位置ではなく、或いは直線ではなく、方向とその大きさだけを表すもの）があれば、ベクトルは自由に平行移動できるので、平面が表現できる。もし３本ベクトルがあったとしても、これら 3 本が同一平面上にあるとき、つまり平面が一つに決まるときは、他の２本の実数倍、或いはそれらの和で表現できるはずであるから、残り一本は線形従属していることになる。こうした場合には、n<m であっても方程式は解ける。

ところが、この範疇にない第三のベクトルが存在すれば、2 本のベクトルの示す平面内には残りの一本は存在しないことになるので、これらを全て含む平面は存在しない。平面と言う解は無い。上記の n<m の場合に(1)式は解を持たないというのはこうした意味である。

２． 最小二乗法の意味

n<m の時、具体的には測定誤差を含む測定データを扱う場合、離散化、或いは量子化誤差を含むデジタルデータを扱う場合、或いは計算誤差を含むデータを扱う場合にはこうした、解を持たない状態となる。こうした場合には、それぞれの方程式からの差があまり生じない、落としどころ（数学的でない表現で恐縮であるが）の解を探ることになる。これは、上記の 3 ベクトルの例でいえば、共通平面は存在しないとしても、３つのベクトルに、なるべく近いところにある平面を探ろうとすることである。(1)式で考えれば、下記の様な近似式を解くことに相当する。

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \approx b_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \approx b_2 \qquad (2)$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \approx b_m$$

この様な目的に最小２乗法という手段が頻繁に用いられる。

なるべく近傍の解を見つけるということは、

$$\phi = \sum_{i=1}^{m} \left(a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - b_i \right)^2 \qquad (3)$$

として、ある解ｘの組が決定した場合に（１）式左辺と右辺の差の２乗の和（誤差の２乗和）φを考えるとき、

$$\phi \rightarrow \text{最小} \qquad (3b)$$

となる、ベクトルｘを見つけるということである。

３． 最小二乗法の利用

さて、ここで、φを最小化することについて考えれば、(3)式を変数で微分して行って、

$$\frac{\partial \phi}{\partial x_1} = 0$$

$$\frac{\partial \phi}{\partial x_2} = 0 \qquad (4)$$

$$\vdots$$

$$\frac{\partial \phi}{\partial x_n} = 0$$

となる、連立方程式を解けば良いことになる。(3)式をｘj で偏微分すれば、

$$\frac{\partial \phi}{\partial x_j}\phi = 2\sum_{i=1}^{m}\left(a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - b_i\right)a_{ij}$$

$$= 2\left[x_1\sum_{i=1}^{m}a_{ij}a_{i1} + x_2\sum_{i=1}^{m}a_{ij}a_{i2} + \cdots + x_n\sum_{i=1}^{m}a_{ij}a_{in} - \sum_{i=1}^{m}a_{ij}b_i\right] \tag{5}$$

(5)式を０となる様に置いて、連立方程式が以下の様に得られる。

$$\begin{pmatrix} \sum_{i=1}^{m}a_{i1}a_{i1} & \sum_{i=1}^{m}a_{i1}a_{i2} & \cdots & \sum_{i=1}^{m}a_{i1}a_{in} \\ \sum_{i=1}^{m}a_{i2}a_{i1} & \sum_{i=1}^{m}a_{i2}a_{i2} & \cdots & \sum_{i=1}^{m}a_{i2}a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{m}a_{in}a_{i1} & \sum_{i=1}^{m}a_{in}a_{i2} & \cdots & \sum_{i=1}^{m}a_{in}a_{in} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m}a_{i1}b_i \\ \sum_{i=1}^{m}a_{i2}b_i \\ \vdots \\ \sum_{i=1}^{m}a_{in}b_i \end{pmatrix} \tag{6}$$

この(6)式を正規方程式と呼ぶ。この連立方程式は一般的に解ける。

ここで、（１C）式の表現を用いれば、（３ｂ）式は

$$\phi = \left\|\mathbf{Ax} - \mathbf{b}\right\|^2 \rightarrow \text{最小} \tag{7}$$

となり、ベクトルの内積の形で書けて

$$\phi = (\mathbf{Ax} - \mathbf{b}) \cdot (\mathbf{Ax} - \mathbf{b})$$

$$= (\mathbf{Ax} \cdot \mathbf{Ax}) - (\mathbf{Ax} \cdot \mathbf{b}) - (\mathbf{b} \cdot \mathbf{Ax}) + (\mathbf{b} \cdot \mathbf{b})$$

行列の関と異なり内積には交換則が成り立つので、

$$= (\mathbf{Ax} \cdot \mathbf{Ax}) - 2(\mathbf{Ax} \cdot \mathbf{b}) + (\mathbf{b} \cdot \mathbf{b}) \tag{8}$$

さらに、ここに、

$$\left(\mathbf{Ax}\cdot\mathbf{y}\right)=\left(\mathbf{x}\cdot\mathbf{A}^T\mathbf{y}\right) \tag{9}$$

なる性質があるので、

$$\begin{aligned}\phi &= \left(\mathbf{x}\cdot\mathbf{A}^T\mathbf{Ax}\right)-2\left(\mathbf{x}\cdot\mathbf{A}^T\mathbf{b}\right)+\left(\mathbf{b}\cdot\mathbf{b}\right)\\ &= \left(\mathbf{x}\cdot\mathbf{A}^T\mathbf{Ax}\right)-2\left(\mathbf{A}^T\mathbf{b}\cdot\mathbf{x}\right)+\left\|\mathbf{b}\right\|^2\end{aligned} \tag{10}$$

また、行列の偏微分の性質、

$$\nabla\left(\mathbf{x}\cdot\mathbf{Ax}\right)=2\mathbf{Ax} \tag{11}$$

$$\nabla\left(\mathbf{A}\cdot\mathbf{x}\right)=\mathbf{A} \tag{12}$$

より、(10)式を x で偏微分すると、

$$\nabla\phi=2\mathbf{A}^T\mathbf{Ax}-2\mathbf{A}^T\mathbf{b} \tag{13}$$

故に(6)式の正規方程式は

$$\mathbf{A}^T\mathbf{Ax}=\mathbf{A}^T\mathbf{b} \tag{14}$$

として表すことが出来る。

４．参考文献

1) 安達忠次：線形代数と解析幾何（森北出版、東京、1976）
2) 金谷健一：これなら分かる最適化数学（共立出版、東京、2008）
3) 今野浩、山下浩：非線形計画法（日科技連、東京、1978）
4) 松山実：基礎数値解析（昭晃堂、東京、1998）