DOI:10.13203/j.whugis1980.01.003

论最小二乘拟合

陶本藻

［提要］ 在大地测量、地震及地球动力学数据处理中，常常需用一个函数去拟合已被测定的一组数据，这就是函数的拟合。函数的拟合通常采用最小二乘准则，即最小二乘拟合。

在一般文献中，观测数据均假定具有独立性，但经常遇到的是处理相关数据。为了解决相关数据的拟合，本文按最小二乘相关估计的理论叙述了拟合的方法，并对这种方法的最优性作了全面的论证。考虑到处理数据时，经常要增添新的观测数据，为了解决数据的贮存便于电算和减少工作量，我们根据卡尔曼滤波的思想，推导了最小二乘拟合的递推公式。

（一）相关最小二乘拟合的理论

设在固定数组X情况下测得数据为Y，它们之间存在如下的线性模型：

$$Y=XB+\varepsilon \tag{1.1}$$

上式表明，数据Y划分为系统成分（信息）XB和随机成分（干扰）ε。B为参数，ε为误差（噪声）。

误差ε服从正态分布，其数学期望为零，即

$$E(\varepsilon)=0 \tag{1.2}$$

误差ε的协方差阵或观测值Y的协方差阵为

$$D(\varepsilon)=M(\varepsilon\varepsilon^T)=\begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots\cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots\cdots & \rho_{2n}\sigma_2\sigma_n \\ \cdots\cdots\cdots\cdots & & & \\ \rho_{1n}\sigma_1\sigma_n & \rho_{2n}\sigma_2\sigma_n & \cdots\cdots & \sigma_n^2 \end{pmatrix} \tag{1.3}$$

$$D(Y)=D(\varepsilon) \tag{1.4}$$

式中σ_i^2为y_i的方差，ρ_{ij}（$i\neq j$）为y_i与y_j间的相关系数。Y的协方差也可表成：

$$D(Y)=\sigma_0^2 Q_{YY} \tag{1.5}$$

式中σ_0^2为单位权方差（方差因子）。Q_{YY}为Y的协因数阵（权逆阵），其主对角线上的元素Q_{ii}是y_i的权倒数，非主对角线上的元素Q_{ij}是y_i与y_j间的协因数（相关权倒数）。

如果在数据X与Y间拟合一条直线，则（1.1）为

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}=\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}+\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix} \tag{1.6}$$

如果在数据X与Y间拟合超平面，则（1.1）为：

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots\cdots & x_{1t} \\ 1 & x_{21} & \cdots\cdots & x_{2t} \\ \cdots\cdots \\ 1 & x_{n1} & \cdots\cdots & x_{nt} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \cdots \\ b_t \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix} \tag{1.7}$$

如果在数据 X 与 Y 间拟合曲面（如二维曲面），则（1.1）为：

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & x_{11}x_{12} & x_{12}^2 \\ 1 & x_{21} & x_{22} & x_{21}^2 & x_{21}x_{22} & x_{22}^2 \\ \cdots\cdots \\ 1 & x_{n1} & x_{n2} & x_{n1}^2 & x_{n1}x_{n2} & x_{n2}^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \cdots \\ b_5 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix} \tag{1.8}$$

如果在数据 X 与 Y 间拟合周期函数，则（1.1）为：

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & \cos x_1 & \sin x_1 & \cos 2x_1 \cdots \sin kx_1 \\ 1 & \cos x_2 & \sin x_2 & \cos 2x_2 \cdots \sin kx_2 \\ \cdots\cdots \\ 1 & \cos x_n & \sin x_n & \cos 2x_n \cdots \sin kx_n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ b_1 \\ a_2 \\ \cdots \\ b_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix} \tag{1.9}$$

为了确定模型方程（1.1），需要估计参数 B，设 B 的估值为 b，代入（1.1）得

$$Y = Xb + V \tag{1.10}$$

V 为 ε 的估值，称为残差。

按最小二乘相关估计理论，要求在

$$V^T P_{YY} V = V^T Q_{YY}^{-1} V = \min$$

下求 B 的估值 b。为此，有

$$\frac{dV^T P_{YY} V}{db} = 2V^T P_{YY} \frac{dV}{db} = 2V^T P_{YY} X = 0$$

即

$$X^T P_{YY} V = 0 \tag{1.11}$$

将（1.10）代入得

$$X^T P_{YY} X b = X^T P_{YY} Y \tag{1.12}$$

解之

$$b = Q_{bb} X^T P_{YY} Y \tag{1.13}$$

式中

$$Q_{bb} = (X^T P_{YY} X)^{-1} \tag{1.14}$$

由（1.10）得残差计算式为：

$$V = Y - Xb = Y - XQ_{bb} X^T P_{YY} Y \tag{1.15}$$

残差平方和为：

$$V^T P_{YY} V = Y^T P_{YY} Y - b^T X^T P_{YY} Y \tag{1.16}$$

单位权方差 σ_0^2 的估值 S_0^2 由下式计算：

$$S_0^2=\frac{V^TP_{YY}V}{n-t} \tag{1.17}$$

由（1.13）可求得参数估值 b 的协方差阵为：

$$\begin{aligned}D(b)&=Q_{bb}X^TP_{YY}D(Y)P_{YY}XQ_{bb}\\&=\sigma_0^2Q_{bb}X^TP_{YY}XQ_{bb}\end{aligned}$$

即

$$D(b)=\sigma_0^2Q_{bb} \tag{1.18}$$

可见，由（1.14）确定的 Q_{bb} 就是 b 的协因数阵。则各参数 b_i 的方差估值和 b_i 与 b_j 间协方差估值为：

$$S_{b_i}^2=S_0^2Q_{b_ib_i},\quad S_{b_ib_j}=S_0^2\ Q_{b_ib_j} \tag{1.19}$$

因为 Y 是正态向量，b 是 Y 的线性函数，b 也是正态向量，故其元素 b_i 服从 N（Eb_i，$\sigma_0\sqrt{Q_{b_ib_i}}$）分布。$b_i$ 的标准化变量

$$\frac{b_i-Eb_i}{\sigma_0\sqrt{Q_{b_ib_i}}}\in N(0,1)$$

可以证明（[8]）统计量

$$\frac{V^TP_{YY}V}{\sigma_0^2}\in\chi^2_{(n-t)}$$

作统计量

$$t=\frac{\dfrac{b_i-Eb_i}{\sigma_0\sqrt{Q_{b_ib_i}}}}{\sqrt{\dfrac{V^TP_{YY}V}{\sigma_0^2(n-t)}}}=\frac{b_i-Eb_i}{S_0\sqrt{Q_{b_ib_i}}} \tag{1.20}$$

它是具有自由度 n－t 的 t 变量。

因此，具有下列概率表达式：

$$P(|t|<t_{\frac{\alpha}{2}})=1-\alpha \tag{1.21}$$

α为置信水平，$t_{\frac{\alpha}{2}}$ 由 t 分布表查出。将（1.20）代入即得 Eb_i 的区间估计公式：

$$P(b_i-t_{\frac{\alpha}{2}}S_0\sqrt{Q_{b_ib_i}}<Eb_i<b_i+t_{\frac{\alpha}{2}}\ S_0\sqrt{Q_{b_ib_i}})=1-\alpha \tag{1.22}$$

由（1.22）可见，当$S_0\sqrt{Q_{b_ib_i}}=S_{b_i}$越小时，所拟合的方程越稳定，故中误差 S_0 的大小是刻划拟合效果的一个重要指标。

（二）最小二乘拟合最优性的论证

最小二乘拟合具有最优性，是指拟合函数中参数估值 b 具有无偏和方差最小的统计性质。

参数估值 b 的统计性质

由（1.1）知，Y 的数学期望为

$$E(Y)=E(XB)+E(\varepsilon)=XB \tag{2.1}$$

Y 的协方差阵由（1.4）确定。

由（1.13）得

$$E(b)=Q_{bb}X^TP_{YY}E(Y)=Q_{bb}X^TP_{YY}XB$$

顾及（1.14）得

$$E(b)=B \tag{2.2}$$

即 b 为 B 的无偏估值。

设 b 是观测值 Y 的下列线性函数：

$$b=\alpha^TY \tag{2.3}$$

故有

$$E(b)=\alpha^TE(Y)=\alpha^TXB=B$$

因此，α^T 必须满足条件：

$$\alpha^TX=E \tag{2.4}$$

估值 b 具有最小方差，必须 b 的协方差阵 D(b)的迹或协因数阵 Q_{bb} 的迹为最小，即

$$tr(Q_{bb})=tr(\alpha^TQ_{YY}\alpha)=\min \tag{2.5}$$

对 Q_{YY} 进行分解，得

$$Q_{YY}=R^TR, \qquad P_{YY}=R^{-1}(R^{-1})^T \tag{2.6}$$

则

$$Q_{bb}=\alpha^TQ_{YY}\alpha=\alpha^TR^TR\alpha$$

令 $Q=(X^TP_{YY}X)^{-1}$，顾及（2.4）下对上式配方，可得

$$\begin{aligned}Q_{bb}=\alpha^TR^TR\alpha&=(QX^TR^{-1})(QX^TR^{-1})^T+\\&+(\alpha^TR^T-QX^TP_{YY}R^T)(\alpha^TR^T-QX^TP_{YY}R^T)^T\end{aligned} \tag{2.7}$$

按矩阵迹的运算规则（[6]）：

$$\begin{aligned}tr(Q_{bb})&=tr\{(QX^TR^{-1})(QX^TR^{-1})^T\}+\\&+tr\{(\alpha^TR^T-QX^TP_{YY}R^T)(\alpha^TR^T-QX^TP_{YY}R^T)^T\}\end{aligned}$$

上式右边第一项为已知常数，右边第二项的值随着 α 的不同而异。为使 $tr(Q_{bb})$ 为最小，必须

$$\alpha^TR^T-QX^TP_{YY}R^T=0 \tag{2.8}$$

由此求得 α^T 为

$$\alpha^T=QX^TP_{YY} \tag{2.9}$$

它满足（2.4），将上式代入（2.3）得

$$b=QX^TP_{YY}Y=(X^TP_{YY}X)^{-1}X^TP_{YY}Y \tag{2.10}$$

其结果与（1.13）同，故 D(b) 为最小。

残差 V 的统计性质

将（1.1）代入（1.15）得

$$V=(XB+\varepsilon)-XQ_{bb}X^TP_{YY}(XB+\varepsilon)=(E-XQ_{bb}X^TP_{YY})\varepsilon \tag{2.11}$$

因 $E(\varepsilon)=0$，故 V 的数学期望为零：

$$E(V)=0 \tag{2.12}$$

将 (1.1) 代入 (1.13) 得

$$b=Q_{bb}X^TP_{YY}(XB+\varepsilon)=B+Q_{bb}X^TP_{YY}\varepsilon \tag{2.13}$$

由 (2.11) 和 (2.13) 求得 V 与 b 的协方差阵为

$$\begin{aligned}D(b,\ V)&=Q_{bb}X^TP_{YY}D(\varepsilon)(E-XQ_{bb}X^TP_{YY})^T\\&=\sigma_0^2Q_{bb}X^TP_{YY}Q_{YY}(E-P_{YY}XQ_{bb}X^T)\\&=\sigma_0^2(Q_{bb}X^T-Q_{bb}X^T)=0\end{aligned} \tag{2.14}$$

即估值 b 与残差 V 互不相关。

因 b 与 V 互不相关，故可由 (1.10) 得

$$D(Y)=D(Xb)+D(V) \tag{2.15}$$

$$D(V)=\sigma_0^2(Q_{YY}-XQ_{bb}X^T) \tag{2.16}$$

即

$$Q_{vv}=Q_{YY}-XQ_{bb}X^T \tag{2.17}$$

(2.16) 和 (2.17) 为残差 V 的协方差阵和协因数阵的计算式。

残差平方和的统计性质

$$\begin{aligned}\varepsilon^TP_{YY}\varepsilon&=(Y-XB)^TP_{YY}(Y-XB)\\&=Y^TP_{YY}Y-B^TX^TP_{YY}Y-Y^TP_{YY}XB+B^TX^TP_{YY}XB\\&=Y^TP_{YY}Y-B^TX^TP_{YY}(Xb+V)-(Xb+V)^TP_{YY}XB+B^TX^TP_{YY}XB\end{aligned} \tag{2.18}$$

由 (1.16) 得

$$V^TP_{YY}V=Y^TP_{YY}Y-b^TX^TP_{YY}Xb-b^TX^TP_{YY}V \tag{2.19}$$

将 (2.19) 减去 (2.18)，顾及 $X^TP_{YY}V=0$，并取数学期望得

$$E(V^TP_{YY}V)=E(\varepsilon^TP_{YY}\varepsilon)-E\{(b-B)^TX^TP_{YY}X(b-B)\} \tag{2.20}$$

因为 ([6])：

$$\varepsilon^TP_{YY}\varepsilon=tr(\varepsilon\varepsilon^TP_{YY})$$

$$(b-B)^TX^TP_{YY}X(b-B)=tr\{(b-B)(b-B)^TX^TP_{YY}X\}$$

故

$$\begin{aligned}E(V^TP_{YY}V)&=E\{tr(\varepsilon\varepsilon^TP_{YY})\}-E\{tr((b-B)(b-B)^TX^TP_{YY}X)\}\\&=\sigma_0^2tr\{E(\varepsilon\varepsilon^TD^{-1}(Y)\}-\sigma_0^2tr\{E(b-B)(b-B)^TD^{-1}(b)\}\\&=\sigma_0^2tr\{D(Y)D^{-1}(Y)\}-\sigma_0^2tr\{D(b)D^{-1}(b)\}\\&=\sigma_0^2(n-t)\end{aligned} \tag{2.21}$$

即

$$E\left(\frac{V^TP_{YY}V}{n-t}\right)=E(S_0^2)=\sigma_0^2 \tag{2.22}$$

S_0^2 是 σ_0^2 的无偏估值。

(三) 最小二乘拟合的递推公式

在进行最小二乘拟合时，经常遇到要增加新数据，需要与原有数据一起重新拟合，为了减少工作量和便于电算，我们推导了最小二乘拟合的递推公式。

设原数据为 $\underset{n\times 1}{Y}$、$\underset{n\times t}{X}$ 新增加一次测量数据为 $(x_{n+1,1}\ x_{n+1,2}\cdots\cdots x_{n+1,t};\ y_{n+1})$，令

$G=(y_{n+1})$，$H=(x_{n+1,1}\ x_{n+1,2}\cdots\cdots x_{n+1,t})$

则观测方程为

$$\begin{pmatrix} Y \\ G \end{pmatrix}=\begin{pmatrix} X \\ H \end{pmatrix} b_{n+1}+\begin{pmatrix} V \\ v \end{pmatrix} \tag{3.1}$$

式中 $v=(v_{n+1})$，b_{n+1} 为由 $n+1$ 组测量数据求得的 B 的估值，相应地，由原数据求得的 B 的估值记为 b_n。

设新数据与原数据是相关的，它们的协因数阵为

$$Q_{n+1}=\begin{pmatrix} Q_{YY} & Q_{YG} \\ Q_{GY} & Q_{GG} \end{pmatrix} \tag{3.2}$$

相应的权阵为

$$P_{n+1}=\begin{pmatrix} P_{YY} & P_{YG} \\ P_{GY} & P_{GG} \end{pmatrix} \tag{3.3}$$

即

$$Q_{n+1}P_{n+1}=P_{n+1}Q_{n+1}=E \tag{3.4}$$

根据加边法求逆公式（[7]）知：

$$P_{YY}=Q_{YY}^{-1}+\frac{Q_{YY}^{-1}Q_{YG}Q_{GY}Q_{YY}^{-1}}{\alpha_n} \tag{3.5}$$

$$P_{YG}=-\frac{Q_{YY}^{-1}Q_{YG}}{\alpha_n} \tag{3.6}$$

$$P_{GY}=-\frac{Q_{GY}Q_{YY}^{-1}}{\alpha_n} \tag{3.7}$$

$$P_{GG}=\frac{1}{\alpha_n} \tag{3.8}$$

式中

$$\alpha_n=Q_{GG}-Q_{GY}Q_{YY}^{-1}Q_{YG} \tag{3.9}$$

Q_{YY} 为原数据 Y 的协因数阵，显然 $P_{YY}\neq Q_{YY}^{-1}$，令

$$Y_{n+1}=\begin{pmatrix} Y \\ G \end{pmatrix},\quad X_{n+1}=\begin{pmatrix} X \\ H \end{pmatrix},\quad V_{n+1}=\begin{pmatrix} V \\ v \end{pmatrix}$$

则 (3.1) 为

$$Y_{n+1}=X_{n+1}b_{n+1}+V_{n+1} \tag{3.10}$$

相应的法方程为

$$(X_{n+1}^{T}P_{n+1}X_{n+1})b_{n+1}=X_{n+1}^{T}P_{n+1}Y_{n+1} \tag{3.11}$$

将上式配成下列形式：

$$b_{n+1}=b_n+(X_{n+1}^TP_{n+1}X_{n+1})^{-1}X_{n+1}^TP_{n+1}(Y_{n+1}-X_{n+1}b_n) \tag{3.12}$$

式中 $(X_{n+1}^TP_{n+1}X_{n+1})^{-1}$ 为 t 阶逆阵，计算比较复杂。因为

$$(X_{n+1}^TP_{n+1}X_{n+1})^{-1}=(X^TP_{YY}X+X^TP_{YG}H+H^TP_{GY}X+H^TP_{GG}H)^{-1}$$
$$=(A+H^TP_{GG}H)^{-1} \tag{3.13}$$

式中 $A=X^TP_{YY}X+X^TP_{YG}H+H^TP_{GY}X$ 为可逆阵，按矩阵反演公式（见[4]之附录）得

$$(X_{n+1}^TP_{n+1}X_{n+1})^{-1}=A^{-1}-A^{-1}H^T(HA^{-1}H^T+P_{GG}^{-1})^{-1}HA^{-1} \tag{3.14}$$

再求 A^{-1}。令 $C=X^TP_{YY}X+X^TP_{YG}H$，它是可逆阵，并令 $W=P_{GY}X$，它是行向量，则

$$A^{-1}=(C+H^TW)^{-1}=C^{-1}-C^{-1}H^T(1+WB^{-1}H^T)^{-1}WC^{-1} \tag{3.15}$$

再求 C^{-1}。令 $Z^T=X^TP_{YG}$，它是列向量，则

$$C^{-1}=(X^TP_{YY}X+Z^TH)^{-1}=(X^TP_{YY}X)^{-1}-(X^TP_{YY}X)^{-1}Z^T(1+$$
$$+H(X^TP_{YY}X)^{-1}Z^T)^{-1}H(X^TP_{YY}X)^{-1} \tag{3.16}$$

但因 $P_{YY}\neq Q_{YY}^{-1}$，故 $X^TP_{YY}X$ 与原始数据中已求出的

$$Q_{bb}=(X^TQ_{YY}^{-1}X)^{-1} \tag{3.17}$$

不同。由（3.5）得

$$(X^TP_{YY}X)^{-1}=(X^TQ_{YY}^{-1}X+X^T\frac{Q_{YY}^{-1}Q_{YG}Q_{GY}Q_{YY}^{-1}}{\alpha_n}X)^{-1} \tag{3.18}$$

令

$$u^T=X^TQ_{YY}^{-1}Q_{YG}, \qquad V=\frac{1}{\alpha_n}Q_{GY}Q_{YY}^{-1}X \tag{3.19}$$

u^T 为列向量，V 为行向量，则（3.18）为

$$(X^TP_{YY}X)^{-1}=Q_{bb}-Q_{bb}u^T(1+VQ_{bb}u^T)VQ_{bb} \tag{3.20}$$

因 Q_{bb} 已求出，由（3.20）求 $(X^TP_{YY}X)^{-1}$，由（3.16）求 C^{-1}、由（3.15）求 A^{-1} 及由（3.14）求 $(X_{n+1}^TP_{n+1}X_{n+1})^{-1}$ 的计算过程中都只需求一个一阶逆阵（倒数），计算十分方便。

公式（3.20）、（3.16）、（3.15）、（3.14）、（3.12）就是测量数据相关时的递推计算公式，称为相关拟合的滤波公式。

特别地，当新增加数据与原数据不相关时，在（3.2）中的

$$Q_{YG}=0 \qquad Q_{GY}=0$$

则（3.5）、（3.6）、（3.7）、（3.8）为

$$P_{YY}=Q_{YY}^{-1}, \quad P_{YG}=P_{GY}^T=0, \quad P_{GG}=Q_{GG}^{-1} \tag{3.21}$$

此时，（3.12）可写成

$$b_{n+1}=b_n+(X^TP_{YY}X+H^TP_{GG}H)^{-1}H^TP_{GG}(G-Hb_n) \tag{3.22}$$

按矩阵反演公式得

$$(X^TP_{YY}X+H^TP_{GG}H)^{-1}=(Q_{bb}^{-1}+H^TP_{GG}H)^{-1}$$
$$=Q_{bb}-Q_{bb}H^T(Q_{GG}+HQ_{bb}H^T)^{-1}HQ_{bb} \tag{3.23}$$

此式也可直接由（3.14）得出。

公式（3.23）、（3.22）就是这种情况下的递推计算公式，即拟合的滤波公式。

运用上述滤波公式计算，不仅充分利用了原数据的作用，而且又不需要保存原数据，这样可使计算机的内存大为积省，从而可扩大计算范围，这在现代测量数据处理中是十分必须的。

参 考 文 献

［1］：森口繁一：统计分析 中译本 1961 年

［2］：H.WOLF：Das geodätische Gauβ-Helmert-Modell und seine Eigenschaften. Z.F.V. 1978 年 2 期

［3］：武汉大学数学系：轧钢自动化的数学方法讲义 1976 年

［4］：A.P.塞奇、J.L.梅尔萨：估计理论及其在通讯与控制中的应用 中译本1978年

［5］：H.PELZER：Zur Behandlung singularer Ausgleichungsaufgaben I Z.f.V 1974 年 5 期

［6］：D.E.Wells、E.J.Krakwiwsky：最小二乘法 中译本 1971 年

［7］：法捷也娃：线性代数计算法 中译本 1958 年

［8］：关学海：广义的等价观测 测绘学报 1963 年 2 期

On the Least Squares Simulation

Tao Benzao

Abstract

In geodetic, seismic and geodynamic data processing, a function is often required to simulate a group of measured data. That is the simulation of a function. The principle of the least squares being usually applied for this purpose, so it is called the least squares simulation.

In the current literatures, the measured data are assumed to be independent with one another. In pratice, however, we are frequently confronted with problems of handling correlated data. For solving such problems a method of simulation is suggested based upon the theory of least square correlation estimution. In addition, a comprehensive demonstration of the optimality of this method is made. In view of the fact that the new measured data are frequently to be supplemented to the data processing and also for economizing the data storage so as to reduce the workload in the computer, a recursion formula for the least squares simulation is derived according to the kalman filtering theory.