

广义岭估计的直接解法

游扬声¹ 王新洲² 刘 星³

(1 武汉大学测绘学院, 武汉市珞喻路 129 号, 430079)

(2 武汉大学科技部, 武汉市珞珈山, 430072)

(3 重庆大学土木工程学院, 重庆市沙坪坝区, 400045)

摘要: 提出了广义岭估计的直接解法。该解法不需计算岭参数, 可直接求得具有最小均方误差的解。并举例证实了该解法的优越性。

关键词: 病态方程; 广义岭估计; 岭参数

中图法分类号: P207.1; P207.2

1 广义岭估计的概念

设有平差线性模型:

$$L = BX + \Delta, E(\Delta) = 0, D_{\Delta} = \sigma_0^2 P^{-1} = \sigma_0^2 E \quad (1)$$

式中, 设 n 为观测值个数, $N = B^T B$ 满秩, 且设 $t = \text{rank}(N)$ 。则 X 的最小二乘解

$$X = N^{-1} B^T L \quad (2)$$

惟一, 而且是最优线性无偏估计。在有些实际平差问题中, 从模型(1)的物理意义而言, N 是满秩的, 但它的行列式的绝对值可能很小, 即 $\det(N) \approx 0$ 。数学中称这种矩阵为病态矩阵, 相应地称法方程为病态方程, 其解很不稳定。当系数阵 N 或常数项 L 有微小变化时, 可能导致 X 的最小二乘解 X 严重失真。1955 年史太恩(C. M. Stein)证明了当 $t > 2$ 时, 正态均值向量最小二乘估计 X 的不可允许性^[1], 即能找到一个估计在某种意义上一致优于最小二乘估计(以下简记为 LSE)。1970 年霍尔(A. E. Hoerl)和克纳德(R. W. Kennard)提出的岭估计就具有比 LSE 较小的均方误差。岭估计是广义岭估计的一个特例。广义岭估计的定义为:

$$X_k = (N + GK G^T)^{-1} B^T L \quad (3)$$

式中, G 为正交方阵, 使

$$G^T N G = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_t)$$

式中, $\lambda_i > 0 (i = 1, 2, \dots, t)$ 为 N 的 t 个特阵根。

而 $K = \text{diag}(k_1, k_2, \dots, k_t)$, $\text{diag}(\)$ 为对角阵。 $k_i (i = 1, 2, \dots, t)$ 称为广义岭估计的岭参数。若 $k_1 = k_2 = \dots = k_t = k$, 广义岭估计就退化为岭估计。若 $k_1 = k_2 = \dots = k_t = 0$, 则式(3)与式(2)相同。

2 确定岭参数的方法

尽管对于一切 $k_i > 0$, 均可以证明 X_k 是 X 的贝叶斯(Bayes)估计^[2], 但若选择不同的岭参数, 得到的估计结果可能大不相同, 易见岭参数的确定是广义岭估计的主要内容。虽然人们提出了确定岭参数的方法和原则, 但至今还没有一种公认的良好方法^[2]。例如, 文献[1]给出了 Hemmerle-Brantle 公式:

$$\hat{k}_i = \sigma_0^2 / (y_i^2 - \sigma_0^2 / \lambda_i) (i = 1, 2, \dots, t) \quad (4)$$

当 $y_i^2 - \sigma_0^2 / \lambda_i < 0$ 时, k_i 取无穷大; 文献[3]介绍了岭迹法, 也给出了一个迭代公式:

$$y_i(k_i) = y_i \lambda_i / (k_i + \lambda_i) \quad (5)$$

3 岭参数与均方误差的关系

由式(3)知, 广义岭估计是在 N 上加一个对角阵 K , 改善 N 的性态, 从而使 X_k 的均方误差小于最小二乘解 X 的均方误差, 即 $MSE(X_k) < MSE(X)$ 。于是, 确定岭参数的基本思路应是寻找使 $MSE(X_k)$ 达到最小的 K 。为此, 令

$$A = BG, Y = G^T X = (y_1, y_2, \dots, y_t)^T \quad (6)$$

则模型(1)的典则形式为:

$$L = AY + \Delta \quad (7)$$

再令 $C = A^T L = (c_1, c_2, \dots, c_n)^T$, 则典则参数 Y 的岭估计为:

$$Y_k = (A^T A + K)^{-1} A^T L = (\Lambda + K)^{-1} C \quad (8)$$

易证 $MSE(X_k) = MSE(Y_k)$, 则

$$MSE(X_k) = MSE(Y_k) =$$

$$\text{tr}(\text{cov}(Y_k)) + (E(Y_k) - Y)(E(Y_k) - Y)^T \quad (9)$$

顾及 $\text{cov}(Y_k) = \sigma_0^2 (\Lambda + K)^{-1} \Lambda (\Lambda + K)^{-1}$ (10)

$$\begin{aligned} E(Y_k) - Y &= (\Lambda + K)^{-1} \Delta Y - Y = \\ &= (\Lambda + K)^{-1} (\Lambda + K - K) Y - Y = \\ &= -K (\Lambda + K)^{-1} Y \end{aligned} \quad (11)$$

$$\begin{aligned} \text{有 } MSE(X_k) &= \sigma_0^2 (\Lambda + K)^{-1} \Lambda (\Lambda + K)^{-1} + \\ &+ Y^T (\Lambda + K)^{-1} K^T K (\Lambda + K)^{-1} Y = \\ &= \sigma_0^2 \sum_{i=1}^t \frac{\lambda_i}{(\lambda_i + k_i)^2} + \sum_{i=1}^t \frac{k_i^2 y_i^2}{(\lambda_i + k_i)^2} \end{aligned} \quad (12)$$

4 广义岭估计的直接解法

式(12)中, λ_i, σ_0^2 是与 k_i 无关的常量, y_i 是非随机参数。将 $MSE(X_k)$ 对 k_i 求导, 并令其为 0, 易于解得:

$$k_i = \sigma_0^2 / y_i^2 \quad (i = 1, 2, \dots, t) \quad (13)$$

此时的 $MSE(X_k)$ 达到最小。由于 y_i (及 σ_0^2) 未知, 须将式(13)写成迭代式:

$$k_i^{(j)} = \sigma_0^2 / y_i^{(j)} y_i^{(j)} \quad (14)$$

即用第 j 次的估计结果 $y_i^{(j)}$ 来估计同次的岭参数 k_i 。再根据式(8), 作第 $j+1$ 次估计:

$$y_i^{(j+1)} = \frac{c_i}{\lambda_i + k_i^{(j)}} \quad (15)$$

将式(14)代入式(15), 得:

$$\begin{aligned} y_i^{(j+1)} &= \frac{c_i}{\lambda_i + k_i^{(j)}} = \frac{c_i}{\lambda_i + \sigma_0^2 / y_i^{(j)} y_i^{(j)}} = \\ &= \frac{c_i y_i^{(j)} y_i^{(j)}}{\lambda_i y_i^{(j)} y_i^{(j)} + \sigma_0^2} \end{aligned} \quad (16)$$

命题 迭代式(16)收敛。

证明

对任意 $i = 1, 2, \dots, t$, 若 $c_i \geq 0$, 设 $j = 0, 1, 2,$

$\dots, \infty$, 有 $y_i^{(0)} = \frac{c_i}{\lambda_i} \geq 0$, 且易见 $y_i^{(0)} \geq y_i^{(1)} \geq \dots \geq$

$y_i^{(j)} \geq \dots \geq y_i^{(\infty)} \geq 0$; 反之, 若 $c_i < 0$, 则有 $y_i^{(0)} = \frac{c_i}{\lambda_i}$

< 0 , 且易见 $y_i^{(0)} \leq y_i^{(1)} \leq \dots \leq y_i^{(j)} \leq \dots \leq y_i^{(\infty)} \leq$

0, 即序列 $\{y_i^{(j)}\}$ 单调有界。则根据数列收敛准则知, 序列 $\{y_i^{(j)}\}$ 必然收敛。于是, 迭代式(16)收敛。证毕。

设 $y_i^{(\infty)} = y_i^*$, 则式(16)成为:

$$y_i^* = \frac{c_i y_i^{*2}}{\lambda_i y_i^{*2} + \sigma_0^2} \quad (17)$$

式(17)为一元三次方程。考虑到方程一定存在实数解, 以及 y_i^*, c_i 应同号, 可得此方程的解:

$y_i^* = 0$, 当 $c_i = 0$ 或 $d_i < 0$;

$y_i^* = (c_i + \sqrt{d_i}) / 2\lambda_i$, 当 $c_i > 0$ 且 $d_i \geq 0$; (18)

$y_i^* = (c_i - \sqrt{d_i}) / 2\lambda_i$, 当 $c_i < 0$ 且 $d_i \geq 0$ 。

式中, $d_i = c_i^2 - 4\lambda_i \sigma_0^2$ 。

若采用 σ_0^2 的先验值 σ_0^2 代入上式, 即可直接解得典则参数。再由 $X = GY$ 易于求得未知参数的广义岭估计值 X_k 。将该解法简记为 DSGRE (direct solution to generalized ridge estimate)。

5 广义岭估计的直接解法的性质

DSGRE 有较严格的理论根据, 不需先进行一次 LSE, 不需迭代, 不需计算岭参数。

将式(13)代入式(12), 得:

$$\begin{aligned} MSE(X_k) &= \sigma_0^2 \left(\sum_{i=1}^t \frac{\lambda_i}{(\lambda_i + k_i)^2} + \sum_{i=1}^t \frac{k_i^2 y_i^2}{(\lambda_i + k_i)^2} \right) = \\ &= \sigma_0^2 \left(\sum_{i=1}^t \frac{1}{\lambda_i + \sigma_0^2 / y_i^2} \right) < \sigma_0^2 \sum_{i=1}^t \frac{1}{\lambda_i} = MSE(X) \end{aligned} \quad (19)$$

这说明 DSGRE 的结果的均方误差比 LSE 的结果小, 即 DSGRE 改善了 LSE。改善的程度与 σ_0^2 、 λ_i, y_i (或 x_i) 有关。尽管 y_i^* 是 y_i 的有偏估计, 式(19)只能说是近似成立, 但总的说来, 当某个 (或多个) $\lambda \approx 0$ 或 $y \approx 0$ 时, DSGRE 将显著优于 LSE。

又若 $Y = 0$, 并代入式(19), 知 DSGRE 的均方误差为 0, 即不论 σ_0^2, λ_i 为多大, 不论观测值是多少, DSGRE 的结果 X_k 为真值。

6 算 例

以文献[3]的[例 6-1]为例来考察 DSGRE 的应用。

用 4 个历元、5 颗卫星的观测值来解算整周模糊度。误差方程系数阵 B 的资料见文献[3]。设参数的真值 $X^T = (\delta x, \delta y, \delta z, N_1, N_2, N_3, N_4) = (4.2, -2.1, 3.5, 48, 52, 31, 55)$, 观测值 $L^T = (-4.85, 9.25, -10.63, 12.20, -4.86,$

9.26, -10.64, 12.21, -4.84, 9.24, -10.62, 12.22, -4.84, 9.27, -10.63, 12.22)。观测误差 $\Delta = \mathbf{B}\mathbf{X} - \mathbf{L}$, $\sigma_0^2 = 1.46 \times 10^{-5}$, 法方程系数阵 $\mathbf{N} = \mathbf{B}^T \mathbf{B}$ 。

计算 $\mathbf{N}$ 的特征值分别为: $(4.8 \times 10^{-8}, 9.4 \times 10^{-10}, 1.8 \times 10^{-7}, 0.17, 0.52, 2.2, 30)$, 由于前 3 个特征值很小, 因而计算 $\mathbf{N}$ 的无穷大条件数^[4] $\text{condi}(\mathbf{N}) = 7.1 \times 10^{10}$, 可见 $\mathbf{N}$ 是病态矩阵。LSE 的解 $\mathbf{X}^T = (17.66, -28.37, 2.06, 78.21, -102.92, 133.88, -42.69)$, 与真值的差为: $\mathbf{X}^T - \mathbf{X}^T = (-13.46, 26.27, 1.44, -30.21, 154.92, -102.88, 97.69)$, 得 $\|\mathbf{X}^T - \mathbf{X}^T\| = 214.3$ 。可见, 无偏估计 LSE 的结果严重失真。

此时, 可采用有偏的广义岭估计。若采用岭

$$\mathbf{G} = \begin{bmatrix} 0.085 & -0.023 & 0.343 & -0.188 & 0.295 & 0.858 & -0.128 \\ 0.020 & 0.126 & 0.019 & -0.071 & 0.220 & 0.046 & 0.963 \\ 0.179 & -0.004 & -0.081 & -0.176 & 0.875 & -0.355 & -0.197 \\ 0.715 & -0.218 & -0.610 & -0.101 & -0.135 & 0.197 & 0.040 \\ 0.445 & 0.760 & 0.287 & -0.276 & -0.194 & -0.145 & -0.084 \\ 0.498 & -0.413 & 0.612 & 0.401 & -0.028 & -0.196 & 0.077 \\ 0.054 & 0.434 & -0.214 & 0.825 & 0.205 & 0.195 & -0.048 \end{bmatrix}$$

$$\mathbf{C} = \mathbf{B}^T \mathbf{G}^T \mathbf{L} = \begin{bmatrix} 3.532 \times 10^{-6} \\ -9.735 \times 10^{-8} \\ 3.436 \times 10^{-6} \\ 6.193 \\ -1.239 \\ 19.698 \\ -180.046 \end{bmatrix}$$

比较上述几种解法的结果, 可见 DSGRE 较优, 改善了最小二乘估计。

致谢: 衷心感谢陶本藻教授对本文的关心和指导!

参 考 文 献

1 黄维彬. 近代平差理论及其应用. 北京: 解放军出版社, 1992. 7

2 王松桂. 线性模型的理论及其应用. 合肥: 安徽教育出版社, 1987

3 崔希璋, 於宗伟, 陶本藻, 等. 广义测量平差(新版). 武汉: 武汉测绘科技大学出版社, 2001

4 易大义, 沈云宝, 李有法. 计算方法. 杭州: 浙江大学出版社, 1989

5 鲁国斌. 广义岭回归估计中关于 K 值选取的 $Q(c)$ 准则. 数理统计与应用概率, 1989. 64(2):

6 Cicci David A, Hall R L. Improved Estimation/ Prediction in Quick-look Orbit Determination. Journal of the Astro-

迹法, 文献[3] 取 $k = 0.7$, 相应的与真值的差为 $(-0.67, 3.96, 6.88, 47.49, 54.26, 29.87, 47.71)$, $\|\mathbf{X}^T - \mathbf{X}_k^T\| = 91.8$ 。又若采用 Hemmerle-Brantle 公式, 由于 $y_i^2 - \sigma_0^2/\lambda_i < 0$, 其解为 0 向量, 相应的与真值的差为 $(4.2, -2.1, 3.5, 48.52, 31.55)$, 得到 $\|\mathbf{X}^T - \mathbf{X}^T\| = 95.0$ 。

这里采用 DSGRE 解, 依式(18)解得: $\mathbf{Y}^T = (71.441, 0.000, 12.665, 37.365, -2.384, 8.830, -5.982)$, 再依 $\mathbf{X} = \mathbf{G}\mathbf{Y}$ 计算得 $\mathbf{X}_k^T = (11.05, -6.84, 1.14, 41.40, 24.80, 56.19, 33.51)$, 与真值的差为: $\mathbf{X}^T - \mathbf{X}_k^T = (-6.85, 4.74, 2.36, 6.60, 27.20, -25.19, 21.49)$, 得 $\|\mathbf{X}^T - \mathbf{X}_k^T\| = 44.2$ 。其中,

nautical Sciences, 1992. 91~106

7 Lu X S, Ou J K. A Method Using Norm Measure the Ill-conditioned Matrix. International Society for Mine Surveying 27th Presidium Meeting and Commissions Session. Tai' an, 1999

8 刘爱义, 王松桂. 一类相依回归系数的有偏估计. 应用概率统计, 1991. 8

9 Aldrin M. Length Modified Ridge Regression. Computational Statistics & Data Analysis. 1997. 377~398

10 Walker E, Birch J B. Influence Measures in Ridge Regression. Technometrics, 1988, 30(2):

11 Sjöberg L E. Unbiased vs. Biased Estimation of GPS Phase Ambiguities from Dual-frequency Code and Phase Observable. Journal of Geodesy, 1999

作者简介: 游扬声, 博士生。现从事现代测量数据处理理论及应用研究。代表成果: 信息扩散原理及其在测量数据处理中的应用。已经在本领域发表论文 3 篇。
E-mail: shengyy@263.net

Direct Solution to Generalized Ridge Estimate

YOU Yangsheng¹ WANG Xinzhou² LIU Xing³

(1 School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan, China, 430079)

(2 Projects and Consulting Office, Wuhan University, Luoja Hill, Wuhan, China, 430072)

(3 School of Civil Engineering, Chongqing University, Shapingba District, Chongqing, China, 400045)

Abstract: Generally, under the Gauss-Markov model $L = BX + \Delta$, the least square estimators of the parameters $X = N^{-1}B^TL$ possess some nice characters. In surveying, especially in dynamic GPS surveying, ill-conditioned problems may be encountered. When the surveying system is morbid and then the characters of the least square estimators become bad. Adding a diagonally matrix K to N will improve the state of N and then decrease the total variance of the estimates. Recent investigations of ill-conditioned problems have demonstrated that ridge-type estimation methods provide increased solution accuracy over conventional estimation techniques.

That the generalized ridge estimate has less mean square error than the least square estimation. Naturally, the gotten estimators is expected to reach the minimum of mean square error. Because the mean square error of estimators is the function of K , we should fix K depending on the minimum of the mean square error. On the basis of this idea, this paper brings about a method to solve the above problem, which is called direct solution to generalized ridge estimate (DSGRE). With DSGRE, we can obtain the optimal solution (that possesses the minimum of MSE) to the generalized ridge estimate directly and the ridge parameters K needn't to be calculated.

Key words: morbid equation; generalized ridge estimate; ridge parameter

About the author: YOU Yangsheng, Ph. D candidate. He is concentrated on the research on the theory and application of surveying data processing. His main achievement is the principium of information spreading and its application to surveying data processing. He has written 3 articles in this field.

E mail: sheng.yy@263.net