第 8 卷第 1 期 中国有色金属学报 1998 年 3 月
Vol. 8 No. 1 The Chinese Journal of Nonferrous Metals Mar. 1998

矩阵奇异值分解与广义岭估计及其在测量中的应用①

叶松林 朱建军
(中南工业大学资源环境与建筑工程学院, 长沙 410083)

摘 要 在测量等许多工程领域中, 存在因函数模型结构差即设计矩阵病态、致使未知数的最小二乘估计偏差太大且不稳定的问题, 因此, 研究了使用矩阵奇异值分解和广义岭估计进行数据处理的方法。首先, 简述了矩阵奇异值分解及广义岭估计的理论与性质; 然后, 重点比较研究了它们解算病态方程的思想、途径、对关键问题的处理、适应范围、工作量大小等; 最后, 通过摄影测量算例验证了所得结果。并且指出, 奇异值分解方法应用于病态方程的参数解算, 是一种易于操作、效果更好的方法, 有重要的应用价值。

关键词 奇异值分解 广义岭估计 病态方程

中图法分类号 P207.2

自著名数学家 Legendre 和 Gauss 先后于 1806 年和 1809 年独立地把最小二乘法应用于观测数据的误差分析, 以及 Markov 于 1900 年证明了著名的 Gauss- Markov 定理即最小二乘估计(简称 LS)在线性无偏估计类中具有最小方差后, 加上所刻划的 LS 估计在线性无偏估计类中的最优性, 使得人们长期以来把 LS 估计当做线性模型中的唯一最好的估计。测量平差作为测量数据处理的一门应用数学, 无疑也以最小二乘估计作为其研究及应用的基石。

随着现代计算技术的飞速发展, 使得人们考虑问题的全面性、复杂性的量化处理成为可能, 这时因为自变量很多, 不可避免地导致自变量之间存在近似的线性关系, 从而导致阵 $\boldsymbol{B}$ 的列向量的近似线性相关, 称这样的设计阵为病态。当 $\boldsymbol{B}$ 呈病态时, $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ 接近奇异, 这时虽然 LS 估计的方差在线性无偏估计类中最小, 但它的值却很大, 使得 LS 估计的精度很差。例如观测误差模型

$$\boldsymbol{L}=\boldsymbol{B}\boldsymbol{X}+\Delta \quad \Delta\approx N(0,\ \sigma^2\boldsymbol{I}) \tag{1}$$

因 $\boldsymbol{X}$ 的某种估计的均方误差为

$$\begin{aligned} MSE(\hat{\boldsymbol{X}}) &= E\,\|\hat{\boldsymbol{X}}-\boldsymbol{X}\|^2 \\ &= \mathrm{tr}(\mathrm{cov}(\hat{\boldsymbol{X}}))+\|\boldsymbol{E}(\hat{\boldsymbol{X}})-\boldsymbol{X}\|^2 \end{aligned} \tag{2}$$

则未知数 $\boldsymbol{X}$ 的 LS 估计 $\hat{\boldsymbol{X}}=(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{L}$ 的均方误差为[1]

$$MSE(\hat{\boldsymbol{X}})=\sigma^2\sum_{i=1}^{t}\frac{1}{\lambda_i} \tag{3}$$

式中 λ_i 为 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ 的特征值。当 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ 接近奇异时, 至少有一个特征根接近于零, 于是 $MSE(\hat{\boldsymbol{X}})$ 就很大。因此当设计矩阵病态时 LS 估计就不是一个好不好的问题, 而是一个能不能用的问题了。

于是问题的严重性及迫切性使得近 30 年来许多学者致力于改进 LS 估计, 提出了许多新的估计, 其中很重要的一类估计就是有偏估计, 即均值不等于参数 $\boldsymbol{X}$ 的估计, 在众多的有偏估计中影响最大的就是岭估计及广义岭估计

① 国家自然科学基金资助项目 49774209 收稿日期: 1996- 10- 18; 修回日期: 1997- 10- 20
叶松林, 男, 33 岁, 副教授, 硕士

(简称 RE)。

另一方面，借鉴最小二乘统一理论，如若求出广义 Gauss-Markov 模型设计矩阵的最小二乘最小范数逆，当然求解未知数也是显然的。矩阵奇异值分解(简称 SVD) 正为此开辟了崭新的途径。

1　RE 和 SVD 的定义及性质

1.1　RE 的定义和性质[2]

对于模型式(1)，参数的岭估计及广义岭估计定义为

$$\hat{\boldsymbol{X}}_k = (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + k\boldsymbol{I})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{L} \tag{4}$$

$$\hat{\boldsymbol{X}}_k = (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + \boldsymbol{U}\boldsymbol{K}\boldsymbol{U}^{\mathrm{T}})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{L} \tag{5}$$

式中　K 为岭参数 $K = \mathrm{diag}(k_1 \cdots k_t)$，$K \geqslant 0$，$\boldsymbol{U}$ 为正交阵，使

$$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{U} = \Lambda = \mathrm{diag}(\lambda_1 \cdots \lambda_t)$$

由式(4) 和(5) 可见岭估计是广义岭估计的特例。

广义岭估计具有下列性质：

① $\hat{\boldsymbol{X}}_k = \boldsymbol{B}_k\hat{\boldsymbol{X}}$

其中 $\boldsymbol{B}_k = (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + \boldsymbol{U}\boldsymbol{K}\boldsymbol{U}^{\mathrm{T}})^{-1}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})$ 即 RE 是 LS 的一个线性变换。

② $\boldsymbol{E}(\hat{\boldsymbol{X}}_k) = \boldsymbol{B}_k\boldsymbol{X}$

可见只要 $\boldsymbol{B}_k \neq \boldsymbol{I}$，等价地 $K \neq 0$，则广义岭估计就是有偏估计。

③对于任意

$$K = \mathrm{diag}(k_1 \cdots k_t),\quad k_i > 0,\quad \|\hat{\boldsymbol{X}}\| > 0$$

总有 $\|\hat{\boldsymbol{X}}_k\| < \|\hat{\boldsymbol{X}}\|$ 即 RE 是 LS 估计向原点的一种压缩。

④存在

$$K = \mathrm{diag}(k_1 \cdots k_t) > 0$$

使得

$$MSE(\hat{\boldsymbol{X}}_k) \leqslant MSE(\hat{\boldsymbol{X}})$$

它表明可在均方误差意义下，岭估计优于 LS 估计。

1.2　矩阵 SVD 及其广义逆

1.2.1　矩阵 SVD 的定义

定理 1[3]　若矩阵 $\boldsymbol{B} \in \boldsymbol{R}^{n \times t}$ 则有正交矩阵 $\boldsymbol{U} = (\boldsymbol{u}_1, \cdots \boldsymbol{u}_n) \in \boldsymbol{R}^{n \times n}$，$\boldsymbol{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_t) \in \boldsymbol{R}^{t \times t}$ 使

$$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{V} = \mathrm{diag}(\sigma_1 \cdots \sigma_p)\quad p = \min\{n, t\} \tag{6}$$

其中　$\sigma_1 \geqslant \cdots \geqslant \sigma_p \geqslant 0$，$\sigma_i$ 是 $\boldsymbol{B}$ 的奇异值。

推论：设 $\boldsymbol{B}$ 的 SVD 已由定理 1 给定，并且 $\sigma_1 \geqslant \cdots \geqslant \sigma_r \geqslant \sigma_{r+1} = \cdots = \sigma_p = 0$ 则

$$\mathrm{rank}(\boldsymbol{B}) = r \tag{7}$$

$$\boldsymbol{B} = \boldsymbol{U}_r\Sigma_r\boldsymbol{V}_r^{\mathrm{T}} \tag{8}$$

$$\|\boldsymbol{B}\|_2 = \sigma_1 \tag{9}$$

式中　$\boldsymbol{U}_r = (\boldsymbol{u}_1, \cdots \boldsymbol{u}_r)$，$\boldsymbol{V}_r = (\boldsymbol{v}_1, \cdots \boldsymbol{v}_r)$

$\Sigma_r = \mathrm{diag}(\sigma_1 \cdots \sigma_r)$

可以看出，SVD 最有价值的方面之一是能够在接近秩亏损的情况下，当舍入误差和模糊的数据使秩的确定成为一个困难问题时，允许我们定量地考虑近似秩亏损的概念。

定理 2　设矩阵 $\boldsymbol{B} \in \boldsymbol{R}^{n \times t}$ 的奇异值分解为 $\boldsymbol{B} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$，其中 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的意义同定理 1，$\Sigma = \mathrm{diag}(\sigma_1 \cdots \sigma_r\ 0 \cdots 0)$ 则 $\boldsymbol{B}$ 的广义逆表示为

$$\boldsymbol{B}^{+} = \boldsymbol{V}\Sigma^{-}\boldsymbol{U}^{\mathrm{T}} \tag{10}$$

其中

$$\Sigma^{-} = \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}_{t \times n}$$

且　$\Sigma_r^{-1} = \mathrm{diag}(\sigma_1^{-1} \cdots \sigma_r^{-1})$

定理 2 的证明按 $\boldsymbol{B}^{+}$ 的四个条件验证符合 Moore-Penrose 广义逆定义即可。

定理 2 把 $\boldsymbol{B}$ 的广义逆 $\boldsymbol{B}^{+}$ 与 $\boldsymbol{B}$ 的 SVD 直接联系起来了。这就是说可以用 SVD 解算线性方程组，对秩亏或近似秩亏的情况也适用。

1.2.2　矩阵 SVD 的一些特性：

(1) 奇异值分解基本唯一，即 Σ 唯一，而 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 不唯一。

(2) 奇异值与条件数的关系

矩阵谱条件数

$$\mathrm{cond}_2(\boldsymbol{B}) = \|\boldsymbol{B}\|_2 \cdot \|\boldsymbol{B}^{-1}\|_2 = \sqrt{\lambda_1(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}) / \lambda_t(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})}$$

已知 $\boldsymbol{B}$ 的奇异值及 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ 的特征值有如下关系[4]：

$$\sigma_i = \sqrt{\lambda_i}$$

代入上式并将其推广到长方阵有

$$\mathrm{cond}(\boldsymbol{B}) = \sigma_{\max} / \sigma_{\min} \tag{11}$$

可见，SVD 的又一有用之处是可顺便计算出矩阵的条件数，定量地刻划了 $\boldsymbol{B}$ 的病态程度。

(3) 奇异值分解的扰动性(稳定性)

若 $\boldsymbol{B}$ 与 $\boldsymbol{B}+\boldsymbol{E}$ 均属于 $\boldsymbol{R}^{n\times t}(n\geqslant t)$ 则对 $k=1, 2, \cdots, t$ 有

$$|\sigma_k(\boldsymbol{B}+\boldsymbol{E})-\sigma_k(\boldsymbol{B})|\leqslant\sigma_1(\boldsymbol{E})=\|\boldsymbol{E}\|_2 \tag{12}$$

这个性质说明了当矩阵有扰动 $\boldsymbol{E}$ 时，奇异值变化不会超过 $\|\boldsymbol{E}\|_2$，可见，奇异值分解法是求解线性方程组特别是病态线性方程组的数值稳定的方法。

2 SVD 解算误差方程及与 RE 法的比较研究

2.1 用 SVD 解算误差方程组

对于观测误差函数模型式(1)按广义 LS 原理有:

$$\boldsymbol{X}=\boldsymbol{B}^{+}\boldsymbol{L}$$

将式(10)代入有

$$\boldsymbol{X}=\boldsymbol{V}\Sigma^{-}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{L} \tag{13}$$

化简后纯量形式为

$$x_k=\sum_{j=1}^{r}\frac{v_{kj}}{\sigma_j}\sum_{i=1}^{n}u_{ij}l_i\quad(k=1, \cdots, t) \tag{14}$$

至于矩阵 SVD 的具体实现参见文献[3]。

2.2 二种方法的比较研究

2.2.1 改善病态方程解的思想及途径

(1) RE 方法

衡量某种估计的均方误差式(2)可知，改善方程解的途径之一是可以以偏差 $\|\boldsymbol{E}(\hat{\boldsymbol{X}})-\boldsymbol{X}\|^2$ 的适当增加，换取方差 $\mathrm{tr}(\mathrm{cov}(\hat{\boldsymbol{X}}))$ 的大量减少，以便使 $MSE(\hat{\boldsymbol{X}})$ 大幅度减小，即综合优度更好。

简单地，对于岭估计式(4)，容易求出[2]

$$\mathrm{MSE}(\hat{\boldsymbol{X}}_k)=\sigma^2\sum_{i=1}^{t}\lambda_i(\lambda_i+k)^{-2}+k^2\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}+k\boldsymbol{I})^{-2}\boldsymbol{X} \tag{15}$$

注意到式(15)右边第一项是 $\hat{\boldsymbol{X}}_k$ 的分量的方差的和(总方差)，而第二项是偏差的平方，对于 $k>0$，$\hat{\boldsymbol{X}}_k$ 是有偏的，且偏度随着 k 增加；另一方面，总方差是 k 的减函数。RE 的思想是挑选一个 k 值，使增加的偏差不超过减少的总方差。

(2) SVD 方法

改善病态方程解的途径之二是仍在 LS 估计下，舍弃小的甚至为零的法方程特征根，以使式(3)中的均方误差大为减少。

从设计矩阵的 SVD 解算误差方程的未知数结果式(14)来看，小的奇异值会使最小二乘问题失真。SVD 的优点在于舍弃较小的奇异值及其影响，就是说，决定 $r=\mathrm{rank}(\boldsymbol{B})$ 的值，从而把近似相关问题变为近似不相关。解决这个问题的一个办法是取参数 $\delta>0$，当 σ_i 使 $\sigma_1\geqslant\cdots\geqslant\sigma_r>\delta>\sigma_{r+1}\geqslant\cdots\geqslant\sigma_t$ 成立时，定 $\boldsymbol{B}$ 的“数值值”为 r，从而按式(14)计算时，消除了近似相关的影响，因而使估计有良好的准确度。

2.2.2 关键问题及其处理

(1) RE 方法的问题在于参数 k_i 的选择。在使 $MSE(\hat{\boldsymbol{X}})$ 达到最小的要求下，K 的最优值不但依赖于未知参数 σ_0^2，$\hat{\boldsymbol{X}}$，而且这种依赖没有显式表示。对于这个应用上的非常重要的问题，统计学家提出了十多种选 k_i 的方法，但还没有一种方法能够一致优于其它的方法。文献[1]介绍了几种方法，其中岭迹法和公式 Hoerl-Kennad 是大多数文献使用的方法。

(2) SVD 方法的关键在于式中的参数 δ 的选取。参数 δ 的选取应该是与机器精度 μ 相协调的常数，例如取 $\delta=\mu\|\boldsymbol{B}\|_\infty$。可是，如果在数据中一般的相对误差比 μ 大，那么 δ 取大一些，例如 $\delta=10^{-2}\|\boldsymbol{B}\|_\infty$。由于式中的“近似解”的 $\|\boldsymbol{X}\|_2\approx1/\sigma_r\leq1/\delta$，也可以按需要选取 δ，以便在范数适当小的意义下产生逼近于 $\boldsymbol{X}_{\mathrm{LS}}$ 的解。

2.2.3 二种方法适应问题的范围

从二种方法的解算思想上看，岭估计属于有偏估计类，一般讲使用与否需已知方程性态，更何况岭估计优于最小二乘估计的充分必要条件还属于理论上尚未解决的问题，岭估计

中的 k 的取值范围也有待理论上确定。而设计矩阵的 SVD 法解方程时，仍遵循广义最小二乘理论，无论方程良态或病态甚至秩亏都以方程秩的确定而使解的结果达到最优，适应性强。

2.2.4　解算工作的比较

显然 RE 方法首先要组成法方程式，对于已是病态的误差方程，构成法方程是很不利的。特别对于广义 RE，由于 K 的最优值依赖于未知参数且这种依赖没有显式，很多文献都是以最小二乘估值 σ_{LS}^2 的 $\boldsymbol{X}_{LS}$ 为 K 值函数的变量 σ_0^2 和 $\boldsymbol{X}$ 的初始值，进行广义 RE 的迭代计算。显然，这种初始值受方程的病态影响很大，得出最优的 K 值是极不容易的，计算工作量大且标准并不好把握。

SVD 方法从对设计矩阵 $\boldsymbol{B}$ 的正交化分解开始，通过 Housholder 双对角化和对双对角阵的带原点位移的隐式对称 QR 算法计算，使矩阵 B 的对角线元素迅速收敛于奇异值(一般立方收敛)，计算工作很快，对秩 $r = \mathrm{rank}(\boldsymbol{B})$ 的确定有一定标准，易于掌握，免除了解算过程受方程病态的影响，计算过程一般可"傻瓜化"。

3　算例

算例是参照文献[5]的办法设计一组摄影测量理想数据。在空间后方交会中，根据控制点布点情况分成两组：第一组中，控制点之间高程最大高差为摄影距离的 60%，在第二组中，控制点间最大高差为摄影距离的 0.1%。然后用严格的共线方程算出各控制点所对应的象点坐标(表 1)。假定象点坐标中误差，分别用计算机算出随机误差加入象点坐标中，就组成了两套不同的观测数据。在空间后方交会中，选择 7 个未知数即摄站坐标改正数 ΔX_s、ΔY_s、ΔZ_s，旋转角改正数 $\Delta\omega$、$\Delta\varphi$、$\Delta\kappa$，主距改正数 Δf。由于象点坐标误差较小且选定这 7 个待定值的原设计值为近似值，所以平差后这 7 个未知数的最小二乘估值应是接近于零的很小数值，这也是平差结果是否可靠的一个客观标准。用 LS 法，RE 法，SVD 法分别计算后的结果见表 2。

由表 2 结果可见，解算第一组方程时三种方法都可达到满意的结果；第二组方程因其病态程度较强(条件数 $\mathrm{cond}(\boldsymbol{B}) = 4.6 \times 10^{-5}$)，LS 法已克服不了病态的影响，特别是未知参数 ΔZ_s 和 Δf 的估值与理论值相差很大，说明两者之间有较强的相关。而用 RE 法，克服病态较好；使用 SVD 法，解算效果则最好。

4　结论

(1) 奇异值分解法与岭估计均能有效地改善病态方程的解，但奇异值分解法更易操作，克服病态的效果也较好。

表 1　控制点坐标

Table 1　Control-points coordinates

Group	Point name	Spatial coordinates/m			Photo coordinates/mm	
		X	Y	Z	x	y
I	1	−50	50	70	89.2857	−89.2857
	2	50	50	130	−48.0769	−48.0769
	3	−50	−50	100	62.5	62.5
	4	50	−50	90	−69.4444	69.4444
II	1	−50	50	99.9	62.5626	−62.5626
	2	50	50	100.1	−62.4376	−62.4376
	3	−50	−50	100	62.5	62.5
	4	50	−50	99.95	−62.5213	62.5313

* Suppose: $f = 125\text{mm}$, $X_s = Y_s = Z_s = 0$, $\omega = \varphi = \kappa = 0$　Mean square error of picture points $\sigma = 3.01 \times 10^{-4}$ mm

表 2　三种方法解算未知数估值

Table 2　Unknowns estimated by three methods

Group	Method	Unknowns						
		ΔX_s/m	ΔY_s/m	ΔZ_s/m	$\Delta\omega$/rad	$\Delta\varphi$/rad	$\Delta\kappa$/rad	Δf/rad
Ⅰ	LS	-6.63E-4	-2.41E-4	-9.84E-4	-7.16E-6	-1.16E-6	-9.22E-7	1.55E-3
	RE	-6.03E-4	-2.78E-4	-8.04E-4	-6.60E-6	-1.48E-6	-9.39E-7	1.28E-3
	SVD	9.65E-9	-5.05E-9	2.76E-9	-1.25E-6	7.41E-7	-9.76E-8	1.17E-9
Ⅱ	LS	-1.03E-3	-7.48E-4	-0.237	-9.37E-6	-4.98E-6	-4.53E-7	0.296
	RE	-8.26E-4	-6.67E-4	7.95E-5	-7.74E-6	-4.22E-6	-4.53E-7	9.18E-5
	SVD	1.07E-8	-7.02E-9	5.06E-12	-1.39E-6	9.12E-7	-4.51E-7	2.54E-12

* Suppose: Ridge parameter $k = 0.008$

(2) 奇异值分解法适用性强，无论方程有无病态，均能使用，数据处理程序可“傻瓜化”。它是最小二乘统一理论的一个重要体现。而岭估计一般要预先知道待解方程有无病态，这无疑增加了应用中的难度。

REFERENCES

1　Chen Xiru(陈希孺) and Wang Songgui(王松桂). Jindai Huigui Fenxi(近代回归分析). Hefei: Anhui Education Press, 1987: 15-148.

2　Wang Shiqing(王石青). Journal of North China Institute of Water Conservancy and Hydroelectric Power, 1995, 16(4): 76.

3　Golub G H and van Loan Charles F. Matrix Computations. Baltimore: The Johns Hopkins University Press, 1983: 15-400.

4　Li Qi(李　祺). Wutan Shuzhi Fangfa Daolun(物探数值方法导论). Beijing: Geological Publishing House, 1991: 87-119.

5　Huang Youcai(黄幼才). Journal of Wuhan Technical University of Surveying and Mapping (WTUSM), 1987, 12(4): 64.

APPLICATION OF SINGULAR VALUE DECOMPOSITION AND GENERALIZED RIDGE ESTIMATION IN SURVEYING

Ye Songlin and Zhu Jianjun

College of Resource, Environment and Civil Engineering,

Central South University of Technology, Changsha 410083, P. R. China

ABSTRACT　The methods of data processing in surveying were studied with Singular Value Decomposition(SVD) and Generalized Ridge Estimation(GRE) under the circumstances that the multicollinearity among the columns of the coefficient matrix makes deviation of estimator too great by least squares adjustment. The theory and their properties of SVD and GRE were narrated, then a comparison between these two methods in the thoughts, ways, key problems, amount of work, applicable limits of solving ill-conditioned equation series was made. At last a photogrammetrical example was used to give the verification for the conclusion reached, and the SVD method solving ill-conditioned equation series was pointed out to be easy to handle and effective, therefore SVD method will be of great value to surveying work.

Key words　singular value decomposition(SVD)　generalized ridge estimation(GRE)　ill-conditioned equation series

(编辑　何学锋)