

# 用 QR 分解拟合回归方程参数估计 和剩余的迭代加细<sup>\*</sup>

吕纯濂

(南京气象学院数学系, 南京, 210044)

## ITERATIVE REFINEMENT OF PARAMETER ESTIMATES AND RESIDUALS FROM A REGRESSION MODEL FITTED USING QR DECOMPOSITION METHODS

Lü Chun-lian

(Dept. of Mathematics, Nanjing Institute of Meteorology, Nanjing, 210044)

### Abstract

Algorithms for iteratively refining the parameter estimates and residuals from the fitting of a regression model using QR decomposition methods are described. It is shown that if square root free algorithms for performing the QR decomposition are used the related iterative refinement algorithms can also be square root free. Testing of the algorithms is carried out and comments made about accuracies of parameter estimates.

**Key words:** QR decomposition, iterative refinement, Householder transformations, Givens rotations, Gram-Schmidt methods

### § 1. 引 言

设回归方程为

$$Y = XB + E,$$

其中  $Y$  为依变量向量,  $B$  为回归系数向量,  $E$  为剩余误差向量. 在 QR(orthogonal-triangular) 分解中, 以上回归方程的  $(n \times p)$  设计矩阵  $X$  分解为

$$X = QR$$

---

\* 1999 年 1 月 11 日收到.

其中  $Q$  为  $(n \times p)$  的列正交矩阵,  $R$  为  $(p \times p)$  的上三角矩阵,  $n$  为观测数,  $p$  为独立变量数. 由于  $Q$  正交,  $Q^T Q = I$ , 正规方程  $(X^T X) \hat{B} = X^T Y$  变为

$$(R^T R) \hat{B} = R^T Q^T Y.$$

由于  $R^T$  为  $(p \times p)$  满秩矩阵, 故正规方程等价于

$$R \hat{B} = C,$$

其中  $C = Q^T Y$  为  $(p \times 1)$  向量, 这样就无需真正建立正规方程, 这正是 QR 分解比扫描法和 Cholesky 分解主要优越之处. 因而, 对共线性数据, 解的精度受  $X$  的统计量而不受  $X^T X$  的统计量的限制<sup>[1]</sup>. 若将矩阵  $(X|Y)$  分解, 含有上三角矩阵  $R$ , 则可定义一新矩阵  $R^*$ , 写为

$$R^* = \begin{bmatrix} R & C \\ 0 & s \end{bmatrix},$$

其中  $R$  和  $C$  如上所述, 标量  $s$  为拟合模型的剩余平方和, 该矩阵  $R^*$  包含拟合回归要获得的统计量的所有信息. 应注意,  $Q$  和  $R$  并不唯一, 改变  $R$  的第  $i$  行和  $Q$  的第  $i$  列所有元素的符号, 将不改变  $X$  的第  $i$  列的元素. 其实在实际中这并不重要, 因为元素的绝对值相同, 而在大部分计算时经常发生符号相消的现象.

所有生成  $R$  的算法都需要平方根, 由于计算时用这些算法较冗繁, 故若可能, 就不要用这些算法. 因而, 无需求平方根却能生成  $R$  变形的算法就应运而生了, 本文将考虑无需考虑计算平方根且紧密关联的  $R$  的两种变形: 其一用下标  $C$ <sup>[2]</sup>, 另一种用下标  $S$ <sup>[3]</sup>. 不失一般性, 设  $p = 3$ , 则有

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix} = K R_T,$$

其中

$$K = \begin{bmatrix} r_{11} & 0 & 0 \\ 0 & r_{22} & 0 \\ 0 & 0 & r_{33} \end{bmatrix}, \quad R_T = \begin{bmatrix} 1 & r_{12}/r_{11} & r_{13}/r_{11} \\ 0 & 1 & r_{23}/r_{22} \\ 0 & 0 & 1 \end{bmatrix}.$$

令  $D = K^T K$ , 则  $D$  为具有元素  $d_i = r_{ii}^2$  的对角矩阵.  $R$  的两种变形 (即  $R_C$  和  $R_S$ ) 皆来自  $R_T$ , 对  $R_C$  就用  $D$  代替  $R_T$  的对角元素, 对  $R_S$  用  $1/D$  代替  $R_T$  的对角元素, 故有

$$R_C = \begin{bmatrix} d_1 & r_{12}/r_{11} & r_{13}/r_{11} \\ 0 & d_2 & r_{23}/r_{22} \\ 0 & 0 & d_3 \end{bmatrix}, \quad R_S = \begin{bmatrix} 1/d_1 & r_{12}/r_{11} & r_{13}/r_{11} \\ 0 & 1/d_2 & r_{23}/r_{22} \\ 0 & 0 & 1/d_3 \end{bmatrix}$$

若

$$Q_T = Q K^{-1},$$

则

$$R_T = K^{-1} R, \quad X = Q R = Q_T D R_T, \quad Q_T = Q D^{-1/2}, \quad R_T = D^{-1/2} R. \quad (1)$$

$R$  的这两种变形是紧密相连的, 为什么要考虑这两种变形呢? 这与关于参数线性约束的最小二乘估计问题有关. 将这种约束处理为具有很大权重的观测值时, 就很容易将这种约束应用在构成  $R_C$  的方法中. 在不自觉地使用无限值为权重时, 又恰恰将这种约束应用在构成  $R_S$  的方法中, 在这种情况下用  $R_S$  比用  $R_C$  更好 [4].

## § 2. 迭代加细

迭代加细可以检查共线性问题 [5], 针对迭代加细收敛但对初始解却无改进或收敛到一个新解或不收敛等等不同的结论, 就可以提供设计矩阵共线性程度的一个指标 [6]. 设  $E$  为剩余误差向量,  $B$  为回归系数向量,  $u$  为一  $n$  维向量,  $z$  为一  $p$  维向量, 有关迭代加细的方程写为

$$\begin{bmatrix} u \\ z \end{bmatrix} = \begin{bmatrix} Y \\ 0 \end{bmatrix} - \begin{bmatrix} I & X \\ X^T & 0 \end{bmatrix} \begin{bmatrix} E \\ B \end{bmatrix}.$$

将该方程的两行分开, 则有

$$u = Y - E - XB, \quad z = -X^T E.$$

若  $E$  和  $B$  为回归方程精确解, 则  $u$  和  $z$  都为零, 而  $z$  等于零说明剩余向量  $E$  与观测矩阵  $X$  的各列正交, 这是模型所要求满足的条件. 故对用其它方法得到的近似的  $E$  和  $B$ , 向量  $u$  表示  $(Y - \hat{Y})$  与其它方法计算剩余之间的差异, 而  $z$  表示剩余与  $X$  的各列不正交的程度. 若  $E$  和  $B$  不是回归方程的精确解, 则  $u$  和  $z$  都不为零, 故向量  $\Delta E, \Delta B$  可分别加到  $E, B$  上, 因而有

$$\begin{bmatrix} u \\ z \end{bmatrix} = \begin{bmatrix} I & X \\ X^T & 0 \end{bmatrix} \begin{bmatrix} \Delta E \\ \Delta B \end{bmatrix} \quad (2)$$

或

$$u = \Delta E + X\Delta B, \quad z = X^T \Delta E. \quad (2')$$

由这两个联立方程可解出  $\Delta E$  和  $\Delta B$ , 求解的方法依赖于在 QR 分解中所用的方法.

Smith<sup>[7]</sup> 考虑并比较了应用在回归中构成 QR 分解的几种算法, 通常所用的三种方法中的两种方法是通过矩阵变换法, 将  $X$  变换为由变换矩阵相乘而得到  $R$  与  $Q$ . 另一种方法由  $X$  的正交化列产生  $Q$ , 而由  $Q$  的转置矩阵与  $X$  预先相乘产生  $R$ . 两种矩阵变换法即著名的 Householder 法和 Givens 旋转法, 而正交化法即著名的古典法和改进的 Gram-Schmidt 法. 若用其中某一种变换矩阵法, 用变换矩阵相乘构成一个等于  $[Q_1|Q_2]$  的  $(n \times n)$  维矩阵, 其中  $Q_1 = Q$ ,  $Q_2$  为一  $(n \times (n-p))$  矩阵, 矩阵  $Q_2$  在 Householder 法和 Givens 法中是不相同的.

迭代加细在 Householder 法和 Givens 法中是相同的. 对这两种方法, 估计剩余向量的方法皆用  $Q_2 Q_2^T Y$  代替  $(Y - \hat{Y})$ . 对于 Householder 法的迭代加细计算讨论如下: 在  $u$  和  $z$  的表达式 (2') 中, 用相应的矩阵替换  $X$ , 则有

$$u = \Delta E + [Q_1|Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} \Delta B, \quad z = [R^T|0] \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} \Delta E.$$

涉及  $z$  的方程可重写为

$$R^{-T}z = Q_1^T \Delta E =: h.$$

用  $Q^T$  左乘涉及  $u$  的方程, 有

$$\begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} u =: \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} \Delta E + \begin{bmatrix} R \\ 0 \end{bmatrix} \Delta B, \quad (3)$$

解  $\Delta B$  和  $\Delta E$ , 有

$$\Delta B = R^{-1}(a_1 - h), \quad \Delta E = Q \begin{bmatrix} h \\ a_2 \end{bmatrix}.$$

迭代过程为不断重新估计  $u$  和  $z$  的过程, 即重新估计  $h, a_1$  和  $a_2$  的过程, 若迭代过程从  $E = 0$  和  $B = 0$  开始 (对应于  $u = Y$  和  $z = 0$ ), 则第一次迭代将产生  $E$  和  $B$  正常的估计, 即

$$\Delta E = E = Q_2 Q_2^T Y, \quad \Delta B = B = R^{-1} Q_1^T Y.$$

迭代终止准则为 [8]

$$\begin{bmatrix} u \\ z \end{bmatrix} \leq p\varepsilon \|Y\|,$$

其中  $\varepsilon$  为在计算机上进行浮点 (加、减、乘除或平方根) 运算生成的数与精确数之间绝对差的上界. 而  $h$  可由方程  $R^T h = z$  向前替换得到,  $\Delta B$  由方程  $R \Delta B = (a_1 - h)$  向后替换得到 [17].

下面讨论改进的 Gram-Schmidt 方法的迭代加细, 相对于古典的 Gram-Schmidt 方法, 改进的 Gram-Schmidt 方法的迭代加细的重要改进在于  $\Delta B$  和  $\Delta E$  的计算上. 令  $Q = Q_1, a = a_1$ , 则由 (2') 式有

$$u = \Delta E + QR \Delta B, \quad z = R^T Q^T \Delta E,$$

故有

$$R^T h = R^T Q^T \Delta E = z, \quad a = Q^T u = Q^T \Delta E + R \Delta B. \quad (4)$$

这样就给出了

$$R \Delta B = a - h =: F, \quad \Delta E = u - QR \Delta B = u - QF.$$

令初始值  $u^0 = u$ , 用改进的 Gram-Schmidt 方法计算  $F$  和  $\Delta E = u - QF$ , 改进的 Gram-Schmidt 方法为: 对  $k = 1, 2, \dots, p$  有

$$f_k = q_k^T u^k - h_k, \quad u^{k+1} = u^k - f_k q_k, \quad \Delta E = u^{p+1},$$

向量  $\Delta B$  则在表达式  $R \Delta B = F$  中由向后替换得到 [8].

下面讨论用扫描法求解回归方程时, 用 QR 分解或 Cholesky 分解的迭代加细过程, 由迭代加细计算过程最初阶段的表达式 (2) 式, 有

$$\begin{bmatrix} \Delta E \\ \Delta B \end{bmatrix} = \begin{bmatrix} I & X \\ X^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} u \\ z \end{bmatrix} = \begin{bmatrix} I - X(X^T X)^{-1} X^T & X(X^T X)^{-1} \\ (X^T X)^{-1} X^T & -(X^T X)^{-1} \end{bmatrix} \begin{bmatrix} u \\ z \end{bmatrix},$$

该表达式的第二行为

$$\Delta B = (X^T X)^{-1} [X^T u - z].$$

利用上式, 则可将第一行表示为

$$\Delta E = u - X \Delta B.$$

这样, 用扫描法解回归方程, 就可得到迭代加细的回归系数和剩余. 在 QR 分解或 Cholesky 分解的情况下,  $\Delta B$  的表达式变为

$$\Delta B = (R^T R)^{-1} [X^T u - z] = R^{-1} R^{-T} [X^T u - z].$$

如在引言最后所述, 若  $R$  矩阵为无需求平方根却能生成  $R$  变形的两种算法中的一种, 矩阵  $X$  分解为 (1) 式, 若已构成  $R_C$  和  $R_S$ , 则  $R_T$  和  $D$  可直接由  $R_C$  和  $R_S$  得到, 并同时得到  $Q_T$ . 此时, 若用 Householder 变换或 Givens 旋转进行分解, 则有

$$QR = [Q_1 | Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} = [Q_{T1} D^{1/2} | Q_2] \begin{bmatrix} D^{1/2} R_T \\ 0 \end{bmatrix}.$$

由 (3) 式和 (1) 式, 有

$$a_1 = Q_1^T u = D^{1/2} Q_{T1}^T u.$$

令  $a_{T1} = Q_{T1}^T u, h_T = R_T^{-T} z$ , 则有

$$\Delta B = R_T^{-T} (a_{T1} - D^{-1} h_T), \quad \Delta E = [Q_{T1} | Q_2] \begin{bmatrix} h_T \\ a_2 \end{bmatrix}.$$

若用 Gram-Schmidt 过程进行分解, 由 (4) 式

$$a = Q^T u = D Q_T^T u, \quad h = R^{-T} z = D^{-1/2} R_T^{-T} z.$$

令  $a_T = Q_T^T u, h_T = R_T^{-T} z$ , 则结果为

$$\Delta B = R_T^{-T} (a_T - D^{-1} h_T) = F_T.$$

这样, 改进的 Gram-Schmidt 方法为: 对  $k = 1, 2, \dots, p$  有

$$f_{Tk} = q_{Tk}^T u_T^k - \frac{h_{Tk}}{d_k}, \quad u_T^{k+1} = u_T^k - d_k f_{Tk} q_{Tk}, \quad \Delta E = u_T^{p+1}.$$

向量  $\Delta B$  则在表达式  $R_T \Delta B = F_T$  中由向后替换得到, 且仍无需计算平方根.

### § 3. 讨论与比较

若所涉及到的矩阵共线性不是太强, 一般迭代加细过程是收敛的, 并且一个奇异矩阵将会被迭代过程的不收敛而显现出来<sup>[9]</sup>. 构成 QR 分解的各种方法间也存在着明显的差异, Givens 旋转是最差的, 若存在强共线性, 则用正交化的 Gram-Schmidt 方法是最好的, 而用  $R_C$  和  $R_S$  的形式比用对 Givens 和 Gram-Schmidt 方法  $R$  的形式给出更精确的估计. 总之, 若共线性不太强烈, 则迭代加细能够明显地改进估计的精确性, 而且用  $R_C$  和  $R_S$  的无需计算平方根的算法至少不会差于用  $R$  的算法.

SAS<sup>[10]</sup> 中的 ORTHOREG 过程和 REG 过程都是用来求解回归方程的, ORTHOREG 用的 QR 分解, REG 用 SWEEP 法<sup>[11]</sup>, GAUSS<sup>[12,13]</sup> 中的 OLS 过程用来计算最小二乘回归, OLSQR 过程用 QR 分解计算回归系数, SVD 过程计算矩阵的奇异值, SVDI 过程计算矩阵  $X$  的奇异值分解  $X = UDV^T$ .

奇异值分解估计和迭代加细估计都比由 SAS 中用 SWEEP 法的 REG 过程产生的有关估计更为精确. 在大多数情况下, SAS 中用 QR 分解的 ORTHOREG 过程产生又比用加细奇异值分解过程更为精确的估计, 但在严重共线性的系统中, SAS 中的 ORTHOREG 过程并不能得到所需要的正确结果<sup>[6]</sup>.

迭代加细可以改进由 QR 分解得到的回归估计的精度. 用 QR 分解可以获得有关共线性出现类型的更多的信息, 而且此时再用迭代加细就又可以得到共线性的诊断度量.

### 参 考 文 献

- [1] D. A. Belsley, E. Kuh, R. E. Welsch, Regression diagnostics. New York: Wiley, 1980.
- [2] M. R. B. Clarke, Algorithm AS163, A Givens Algorithm for moving from one linear model to another without going back to the data. Appl Statist, **30**(1981), 198–203.
- [3] W. D. Stirling, Algorithm AS164: Least square subject to linear constraints. Appl Statist, **30**(1981), 204–212 and 357.
- [4] R. W. Farebrother, Linear Least Squares Computations. New York: Marcel Dekker, 1988.
- [5] A. Bjoerck, Comment on the iterative refinement of least squares solutions. J Amer Statist Assoc, **73**(1978), 161–166.
- [6] 吕纯濂, 用奇异值分解求解回归方程的迭代加细. 数值计算与计算机应用, **19**:4(1998), 265–270.
- [7] D. M. Smith, A comparison of methods for forming the QR decomposition for use in regression. Comput Statist Data Anal, **14**(1992), 229–248.
- [8] W. B. Gragg, R. J. Leveque, J. A. Trangenstein, Numerically stable methods for updating regressions. J Amer Statist Assoc, **74**(1979), 161–168.
- [9] G. E. Forsythe, C. B. Moler, Computer Solution of Linear Algebraic Systems. Englewood Cliffs, N.J.:Prentice-Hall, 1967.
- [10] SAS/STAT. User's Guide, Version 6.4, Cary, NC, USA: SAS Institute Inc., 1990.
- [11] 吕纯濂, 线性约束和奇异 SS&CP 矩阵下 LS 回归的扫描算子, 南京气象学报, **21**:1(1998), 35–41.
- [12] GAUSS Software. GAUSS<sup>TM</sup>, Version 2.0 System and Graphics Manual, Washington, USA: Aptech Systems Inc., 1988.
- [13] 吕纯濂, 陈舜华, 矩阵语言与多元分析. 北京气象出版社, (1994), 14–177.