第9卷 增刊1 控制理论与应用 Vol.6 Suppl.I. 1
1989年1月 CONTROL THEORY AND APPLICATIONS Jan., 1989

参数辨识与建模中病态数据的处理方法及其应用*

毛剑琴 范耀祖 马超英

（北京航空学院自控系）

摘　要

本文分析了借助数字计算机进行辨识与建模时，舍入误差给数值计算带来的问题，并因此引入了问题的条件及算法的数值稳定性的概念。进而分析了几种不同的方法求最小二乘估计时问题的条件及其误差。

分析表明QR方法较解法方程方法精度高，但却难以处理病态数据形成的亏秩问题。只有通过奇异值分解才能求得亏秩问题的最小模的最小二乘解。

作为应用实例，文中对动力调谐陀螺在捷联状态下误差模型辨识中的病态数据进行了处理。从而说明了辨识与建模中研究病态数据处理的重要意义以及奇异值分解在解决这一问题中的重要作用。

一、问题的条件及算法的数值稳定性

数学模型的建立是对控制系统进行研究的前提。除了少数简单的系统外，一般的数学模型是根据系统的输入、输出实验数据，借助于系统的测试和辨识的方法，通过数值计算得到的。

由于计算机舍入误差的存在，对于同一个问题，所采用的算法不同，将产生不同的计算误差。其中，问题的条件及算法的数值稳定性是两个起着决定性作用的因素。因而下面首先介绍这两个重要的概念。

在数字计算机中，一般地数 a 以浮点数的形式表示为 $a^*=fl(a)$。由于机器的字长是有限的，于是有

$$|(a^*-a)/a|=\varepsilon.$$

这种误差，通常称为舍入误差。所有的算术运算都将引入舍入误差。

设有数学描述的问题 $f(\cdot)$，它作用于参数 $d\in D$ 上，得到解 $f(d)\in S$，其中 D 与 S 分别为某个参数与解的集合。由于舍入误差，在计算机中 d 通常只能近似地表示为 d^*。因而，通过计算机运算，得到的是 $f(d^*)$，而不是 $f(d)$。

---

*中国科学院科学基金资助的课题。

本文于1986年9月28收到。1987年12月14日收到修改稿。

定义 1[6] 对问题 $f(\cdot)$，若 $d^*\in D$ 并"接近于" d 时，其解 $f(d^*)\in S$ 也"接近于" $f(d)$，则问题 $f(\cdot)$ 称为"良态"问题，反之称为"病态"问题。

这里"接近"的概念，可定量地用相应的范数描述如下。

$$\operatorname{cond}(f)=\sup_{d,d^*\in D}\frac{\|f(d^*)-f(d)\|_s}{\|d^*-d\|_D},$$

其中，$\|\cdot\|_s$，$\|\cdot\|_D$ 表示分别在 S 与 D 空间取范数；$\operatorname{cond}(f)$ 称为问题 f 的条件。若 $\operatorname{cond}(f)$ 小，则问题为良态；否则问题为病态。

例 设有以下线性代数方程组

$$\begin{bmatrix}0.0161 & 0.1441\\ 1.2969 & 0.8648\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\begin{bmatrix}0.1440\\ 0.8642\end{bmatrix},$$

其解为，
$$x_1=2,\ x_2=-2.$$

若右端 b 有一个 10^{-8} 量级的小变化成为 $\tilde{b}$

$$\tilde{b}=\begin{bmatrix}0.1440000059\\ 0.86419993352\end{bmatrix},$$

其解为
$$\tilde{x}_1=0.991,\quad \tilde{x}_2=-0.4868501.$$

这种由系数矩阵或自由项的微小变化引起解的大变化的方程组称为病态方程组。

线性代数方程组的病态程度仅与其系数矩阵 A 的条件数 $P(A)$ 有关，$P(A)$ 有如下定义，

$$P(A)\triangleq\|A\|\cdot\|A^{-1}\|.$$

（当 A 为非方阵时，$P(A)$ 直接定义为 A 的最大奇异值与最小奇异值之比。）

不难得到 $P(A)$ 与解的相对误差的关系为

$$\frac{\|\Delta X\|}{\|X\|}\leqslant\frac{P(A)}{1-P(A)\dfrac{\|\Delta A\|}{\|A\|}}\left(\frac{\|\Delta A\|}{\|A\|}+\frac{\|\Delta b\|}{\|b\|}\right),\tag{1}$$

其中 $\|\cdot\|$ 表示矩阵或向量的范数。

（1）式具体地表示了：条件数 $P(A)$ 越大，则由系数矩阵或自由项的微小变化引起的相对误差也越大，即问题越病态。

可以算出上例中矩阵 A 的条件数 $P(A)>10^7$，由（1）式不难理解在 b 中仅有 10^{-8} 量级的变化，会引起解的如此大的变化。

在计算机进行数值计算时，必将对问题 $f(\cdot)$ 采用一定的算法，即将问题 $f(\cdot)$ 化作它的近似问题 $f^*(\cdot)$ 来解。

定义 2[6] 若算法 $f^*(\cdot)$ 没有引入比问题本身所固有的更多的对扰动的敏感性，则 $f^*(\cdot)$ 称为数值稳定的算法，否则为数值不稳定的。

由上得知，对一个具体的工程问题 $f(\cdot)$，$d\in D$；经过计算机的计算后，得到的解为 $f^*(d^*)$，而不是 $f(d)$。在分析解的计算误差时，必须分别考虑两个因素，首先，若

算法是数值稳定的，则$f^*(d^*)$“接近于”$f(d^*)$；其次，若问题是“良态”的，则$f(d^*)$“接近于”$f(d)$。因而计算得到的$f^*(d^*)$“接近于”所要求的解$f(d)$。

一、最小二乘估计中问题的条件及其误差分析

参数辨识问题往往归结为求解下列形式的矛盾方程组

$$AX=b, \tag{2}$$

其中，$A\in R^{m\times n}$，$m\geqslant n$，为输入构成的系数阵；$X\in R^n$，为要求辨识的模型参数向量；$b\in R^m$为输出构成的向量。

在通常使用的最小二乘估计中，一般采用解法方程的方法来求得（2）的解[3]。即

$$(A^TA)X=A^Tb, \tag{3}$$

则

$$X=(A^TA)^{-1}A^Tb.$$

于是问题化为求解一般线性代数方程组。

由（1）得出其误差估计为

$$\frac{\|\Delta X\|}{\|X\|}\leqslant\frac{P(A^TA)}{1-P(A^TA)\dfrac{\|\Delta(A^TA)\|}{\|A^TA\|}}\left(\frac{\|\Delta(A^TA)\|}{\|A^TA\|}+\frac{\|\Delta(A^Tb)\|}{\|A^Tb\|}\right), \tag{4}$$

可以证明当A满秩时，$P(A^TA)=K^2(A)$，其中$K(A)$为当$m\geqslant n$时，$K(A)=\|A\|\,\|A^+\|$，A^+为A的伪逆。所以，一般地本来不很病态的问题，用解法方程的方法求解时，往往使其化为病态问题。

特别是当A为亏秩时，$(A^TA)^{-1}$不存在，方程（3）无法解。因此有必要考虑数值稳定性更好的算法。

用QR分解求解（2）的方法就是一种较解法方程方法数值稳定性更好的算法。它基于矩阵的镜象映射变换或Householder变换，因此也称为镜象映射方法或Householder方法[4]。

用QR分解求解（2）的解的误差估计为

$$\frac{\|\Delta X\|}{\|X\|}\leqslant 1.6\left[K(A)+K^2(A)\frac{\|r\|}{\|A\|\|X\|}\right]\phi(m,n)2^{-t}, \tag{5}$$

其中，$\|r\|$为残差：

$$\|r\|=\|Y-AX\|;$$

$\phi(m,n)$为随着A之阶数m，n缓慢增长的函数，可视为与阶同量级的常数。

当（5）中的第二项比第一项小得可以忽略时，这时（5）成为：

$$\frac{\|\Delta X\|}{\|X\|}\leqslant 1.6K(A)\phi(m,n)2^{-t}. \tag{6}$$

显见，在这种情况下，QR分解方法较解法方程方法的精度高。当A为亏秩时，如所知，方程组（2）可以有无穷多组解。这时，工程中往往将所有解中范数最小者作为这一类问题的唯一解，即所谓的最小模的最小二乘解。在这种情况下，由于$(A^TA)^{-1}$不存在，

显然，用解法方程的方法无法求解。用QR分解方法虽然可以求出众多解中的一个解，却未必是最小模的最小二乘解。这时需要借助奇异分值解求（2）的解。

三、通过奇异值分解求病态数据的最小模的最小二乘解

定理 1[1]　设任意矩阵$A\in R^{m\times n}$（或$C^{m\times n}$），则存在正交矩阵（或酉阵）$U\in R^{m\times n}$（或$C^{m\times n}$）及正交矩阵$V\in R^{n\times n}$（或$C^{n\times n}$），使得

$$A=U\sum V^T\quad（或U\sum V^*），\tag{7}$$

其中，$\sum\in R^{m\times n}$，

$$\sum=\begin{pmatrix}S\\0\end{pmatrix},$$

其中，　$S=\mathrm{diag}(\sigma_1,\cdots,\sigma_n)$，

并且　$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n\geqslant 0$。

定义 3[1]　（7）称为对矩阵A的奇异值分解；$\{\sigma_i\}(i=1,\cdots,n)$称为$A$的奇异值；$U$，$V$分别为矩阵$A$的左、右奇异向量阵。

$\{\sigma_i\}$中不为零的奇异值的个数γ，即为矩阵A的秩。当$m\geqslant n$时最大奇异值σ_1与最小奇异值σ_n之比为A的条件数$P(A)=\sigma_1/\sigma_n$，而$K(A)=\sigma_1/\sigma_\gamma$。

将（7）代入（2），便有

$$U\sum V^TX=b。$$

令$z=V^TX$及$d=U^Tb$，则

$$\sum z=d。\tag{8}$$

方程组（8）是对角线方程组，很容易求得解[2]

$$\begin{cases}z_i=d_i/\sigma_i,i=1,\cdots,\gamma & (\sigma_i\neq 0时),\\ z_i=任意值,i=\gamma+1,\cdots,n & (\sigma_i=0时)。\end{cases}$$

显然，其最小模的最小二乘解所对应的是

$$\begin{cases}z_i=d_i/\sigma_i,i=1,\cdots,\gamma & (\sigma_i\neq 0时),\\ z_i=0,i=\gamma+1,\cdots,n & (\sigma_i=0时)。\end{cases}\tag{9}$$

令由（9）求出的方程组（2）的最小模的最小二乘解为X_0。

用奇异值分解求最小模的最小二乘解的误差估计公式为

$$\frac{\|\Delta X_0\|}{\|X_0\|}\leqslant 9\left[K(A)+K^2(A)\frac{\|\gamma\|}{\|A\|\|X_0\|}\right]\phi(m,n)2^{-t}\tag{10}$$

同样地，在（10）中当第二项与第一项比较小，小得可以忽略时，有

$$\frac{\|\Delta X_0\|}{\|X_0\|}\leqslant 9K(A)\phi(m,n)2^{-t}。\tag{11}$$

比较（5）、（7）、（11）可以看出，当A满秩时，QR分解方法的精度最高，奇异值分解方法的误差稍大，但二者相差不超过一个量级。而解法方程方法的误差最大，特别

当$P(A)$愈大时，其精度较前两个方法愈差。当A为亏秩时，奇异值分解方法的优越性则尤为突出。

四、动力调谐陀螺捷联状态下误差模型辨识中病态数据的处理

动力调谐陀螺捷联状态下误差模型辨识时[7]，测得一组数据：$A\in R^{13\times 4}$

$$A=\begin{pmatrix}
1.00000 & -156.91200 & 24621.37600 & -3863389.30000\\
1.00000 & -78.37220 & 6142.20170 & -481377.86000\\
1.00000 & -62.66430 & 3926.81450 & -246071.08000\\
1.00000 & -31.24830 & 976.45630 & -30512.59800\\
1.00000 & -7.68640 & 59.08070 & -454.11820\\
1.00000 & -1.14140 & 1.30280 & -1.48700\\
1.00000 & 0.16760 & 0.02810 & 0.00470\\
1.00000 & 1.47660 & 2.18030 & 3.21940\\
1.00000 & 8.02160 & 64.34610 & 516.15870\\
1.00000 & 31.58350 & 997.51750 & 31505.09400\\
1.00000 & 62.99950 & 3968.93700 & 250041.05000\\
1.00000 & 78.70740 & 6194.85480 & 487580.92000\\
1.00000 & 157.24720 & 24726.68200 & 3888201.50000
\end{pmatrix}$$

$$b^T=[8.481300,5.147850,4.139750,2.087500,0.527680,0.090820,0.002255,\\
-0.084395,-0.516200,-2.044250,-4.081450,-5.114100,-8.880950].$$

可算出A的条件数为$P(A)=1.83\times10^6$，说明这组数据相当病态。

由于M—160（Ⅱ）的单精度为$2^{-t}\doteq10^{-8}$，双精度为$2^{-t}\doteq10^{-15}$，因而对这样病态的问题采用单精度字长用解法方程的方法已不可能得出其正确解了。即使采用双精度字长，据（5）估其解的精度也将小于10^{-2}。

用双精度字长计算，采用QR分解方法，得出其解为

$$X^T=(0.0401280038,-0.0678309202,-0.387818896\times10^{-5},\\
0.508298172\times10^{-6}),$$

其残差为

$$r^T=(-0.0199991650,0.0908548236,0.00897225365;\ -0.0480594780,\\
-0.0330706276,-0.0267196596,-0.0265051983,-0.0243471563,\\
-0.0119054243,0.0508077033,0.0598894693,-0.00825484470,\\
-0.0116626769).$$

据（6）式估计，其解的精度约$\leqslant10^{-7}$。

进一步，若在A的右边加一列1.00000，即$A\in R^{13\times 5}$，并为亏秩阵。

可算得A的奇异值为

(5535175.00,36424.8203,120.323227,4.27260685,0.000256487867).

显然，相对最大的奇异值，其最小的奇异值可视为零。

若用QR分解方法，求得之解为

$$X^T=(-101518.625,-0.0673397183,0.463567631\times10^{-5},$$
$$0.404030459\times10^{-6},101518.937).$$

显然，这是一组（2）的解，但不是最小模的最小二乘解。对该工程问题是不适用的。

若用奇异值分解方法，则求得最小模的最小二乘解为

$$X_0^T=(0.0196958780,-0.0672304630,-0.876457489\times10^{-5}$$
$$0.442069904\times10^{-6},0.0196967721),$$

这一组解比前者对物理问题更为合理，是可以采用的。

以上全部数值计算是借助BLINAL[5]在M—160（Ⅱ）机上完成的。

五、结　　论

理论分析和数值计算表明，测试数据的病态程度直接影响着辨识与建模的精度。因而，对病态数据处理问题的研究在测试与辨识中具有重要的意义。

通过对三种方法的分析与比较，说明奇异值分解方法在保证辨识精度及处理病态数据方面有独到的优越性。因而进一步推广应用奇异值分解这一重要的数值工具是辨识和建模中值得注意和探讨的问题。

参 考 文 献

〔1〕 Stewart, G. W., Introduction to Matrix Computations. Academic Press, (1973).

〔2〕 Golub, G. H., C. Reinsch, Singular Value Decomposition and Least Squares Solutions, Humer, Math., Vol. 14, (1970).

〔3〕 夏天长著，熊光楞、李芳芸译，系统辨识，清华大学出版社，北京，(1983).

〔4〕 冯康等编著，数值计算方法，国防工业出版社，北京，(1978).

〔5〕 CCSCAD软件包联合设计组，CCSCAD（中国控制系统计算机辅助设计）软件包技术总结报告（Ⅰ）(1986).

〔6〕 Klema, V. C., A. J. Laub, The Singular Value Decomposition: Its Computation and Some Applications, IEEE Trans. On Auto Control, Vol. AC-25:2, (1980).

〔7〕 林士谔等编著，动力调谐陀螺仪，国防工业出版社，北京，(1983).

〔8〕 毛剑琴，奇异值分解在测试与辨识中的应用，惯性导航与仪表，2，(1984)，

On Approaches to Pathological Data for Parameter Identification and Modeling with Its Applications

Mao Jianqin, Fan Yaozu, Ma Chaoyin

(Department of Automatic Control, Beijing University of Aeronautics and Astronautics)

Abstract

It is analysed that the problem caused by rounding errors in parameter identification and modeling, which is solved by using a computer. The concepts of conditioning for a problem and the numerical stability for one algoriehm are introduced. Furthermore, the conditional numbers and numerical errors of several algorithms for solving linear least squares estimations are discussed.

By comparison, QR algorithm is more accuracy than solving the normal equation, thus generally for full rank problem it can handle the pathological data fairly well, but for rank-deficient problems it is not an ideal algorithm, In computational practice, for rank-deficient case by using the singular value decomposition to get the minimum norm solution is essentially the Only known method which is reliable.

As an applied example, the pathological data are computed for creat an error model of a dynamica-tuned gyroscope under the strapdown condition. Therefore, for identification and modeling the importance of approaches to pathological data and the important figure of singular value decomposition are obvious.