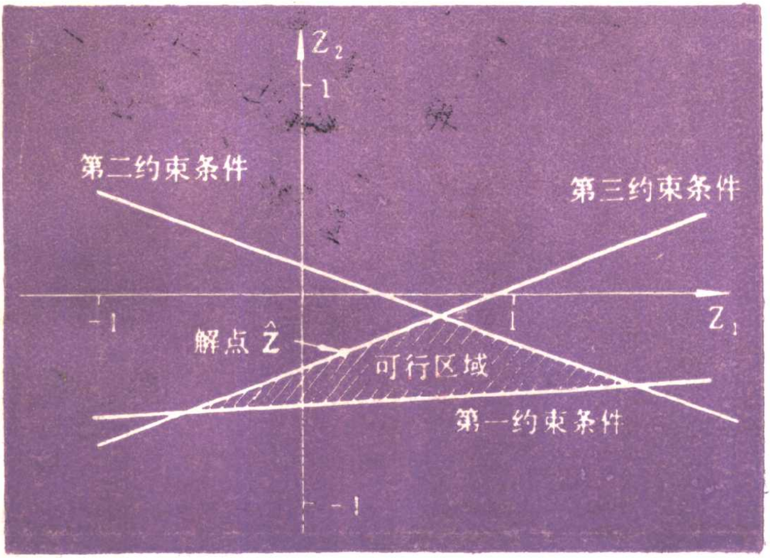
最小二乘问题计算方法

刘钦圣 编著



北京工业大学出版社

最小二乘问题计算方法

刘 钦 圣　　编著

北京工业大学出版社

内 容 简 介

本书系统地介绍了最小二乘问题的计算方法。全书共分五章，内容主要包括：线性最小二乘问题的古典和现代解法、实时最小二乘法、广义最小二乘法、广义逆矩阵、非线性最小二乘问题的 Gauss-Newton型方法、拟-Newton方法、线性-非线性可分离最小二乘法、带约束条件的最小二乘法等。

本书是国内介绍最小二乘法较为完整的一本新著，可供理工科大学生、研究生和广大科技人员学习和参考。

最小二乘问题计算方法

刘钦圣 编著

*

北京工业大学出版社出版

新华书店北京发行所发行

北京通县燕山印刷厂印刷

*

1989年10月第1版 第1次印刷

787×1092毫米 32开本 7.75印张 174千字

印数：1～3000册

ISBN7—5639—0038—1/O·5

定价：4.00元

前　言

1983年夏，在北京钢铁学院举办的全国穆斯堡尔谱计算机拟合讨论会上，编者做过几次最小二乘拟合计算方法的报告，会前编写了一本约10万字的讲义。本书就是在这本讲义的基础上，增补近年来的最新文献资料并参考读者的意见修改整理而成的。

全书分五章：一、引论；二、线性最小二乘问题与广义逆矩阵；三、非线性最小二乘问题；四、线性-非线性可分离最小二乘问题；五、带约束条件的最小二乘问题。编者力图在这本书中，把最小二乘问题的计算方法作一系统而简明的介绍。每种方法都讲述得相当完整，对其基本思想与收敛性等理论问题，给出必要的数学论证，其中许多方法还列出了详细的计算步骤或框图，并且尽可能使它们保持相对的独立性。这样可使不同水平和不同专业的读者都能接受并有所收益。搞实际应用的人，只需对感兴趣的章节，粗读有关的基本概念和公式，根据计算步骤或框图就可编制出程序；搞理论研究的人，可从本书系统地了解到这个专题的现代水平与发展方向。

最小二乘问题的计算方法在应用数学、物理、测绘、数理统计、数学规划、系统工程、控制论、经济与生物工程等领域中有着广泛的应用。而在一般计算方法与最优化方法的书籍中，对最小二乘法讲述的内容很简略，很多有效的新方法散

见于中、外文书刊上，查找非常不便，本书已尽可能地将它们收集整理和系统化，因此，它将是目前国内这一方面内容比较全面的一本书，可供理工科大学生、研究生和广大科技人员学习和参考。

本书从编写讲义到修改成书，一直得到我校金属物理学系马如璋教授的热心支持与鼓励；1983年李爱华同志（北京钢铁学院数力系当时的研究生）按照我指定的文献，曾整理了一份“带约束条件的最小二乘问题”的素材，对我编写第五章有一定参考价值；在此谨向他们表示衷心的感谢。

由于编者水平有限，实际计算经验不多，本书的缺点和错误在所难免，敬希读者批评指正。

编者

1988年6月于北京科技大学

本书使用的符号说明

R^n 实n维空间

$L(R^n, R^m)$ 从R^n到R^m的线性算子的线性空间

$R^n \times R^m$ 乘积空间

$f: D \subset R^n \to R^m$ 定义域D在R^n内值域在R^m的映射

$\boldsymbol{b}=(b_1, b_2, \cdots, b_n)^T$ 分量为b_i $(i=1,2,\cdots,n)$的列向量。所有黑体小写英文字母都表示向量，例如$\boldsymbol{b}, \boldsymbol{p}, \boldsymbol{x}$

$\boldsymbol{b}^T$ 向量$\boldsymbol{b}$的转置

$\{\boldsymbol{b}^{(k)}\}$ 向量的序列

$A=(a_{ij})$ 元素为a_{ij}的矩阵。大写英文字母一般表示矩阵，例如A, B, H

A^T 矩阵A的转置

A^{-1} 方阵A的逆矩阵

A^+ 矩阵A的广义逆矩阵

$R(A)$ 矩阵A的列空间

$N(A)$ 矩阵A的零空间

$\mathrm{rank}(A)$ 矩阵A的秩

$\rho(A)$ 矩阵A的谱半径

I 单位矩阵

$\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ 对角矩阵

$\|\cdot\|$ 向量或矩阵的某种范数

$R^{m\times n}$ 实$m\times n$阶矩阵的集合

$R_r^{m\times n}$　实$m\times n$阶且秩为r的矩阵的集合

$\in$　表示属于，例如$A\in R^{m\times n}$表示A是一个实$m\times n$阶矩阵

$\forall$　表示任意，例如$\forall K$表示任意一个K

$\exists$　表示存在，例如$\exists\delta$表示存在一个δ

#　表示定理证毕

∞　在本书中表示$+\infty$

目　　录

第一章　引论

1.1　曲线拟合

在科学技术的许多领域中，常会遇到下列类型的问题：设x和y都是被观测的量，且y是x的函数：

$$y=f(x;\ b_1,\ \cdots,\ b_n) \tag{1.1}$$

假定这个函数关系已由实际问题从理论上具体确定，因而(1.1)称为理论函数或理论曲线公式，但其中含有n个未知参数b_1，…，b_n。为了确定这n个参数，我们通过某种实验或观测得到m组数据：

$$(x_1,y_1),\ (x_2,y_2),\cdots,\ (x_m,y_m) \tag{1.2}$$

根据(1.2)来寻求参数的最佳估计值$\hat{b}_1,\cdots$，$\hat{b}_n$，即寻求最佳的理论曲线$y=f(x;\ \hat{b}_1,\ \cdots,\ \hat{b}_n)$，这就是通常所说观测数据的曲线拟合问题。也可称为观测数据的平滑问题。

实际问题中经常遇到的一种曲线拟合问题是需要从观测数据(1.2)求出y和x的一个经验公式。例如，用一个$n-1$次多项式去拟合观测数据$(x_i,\ y_i)$，$i=1,\cdots$，m，即假设理论函数为

$$y=f(x;\ b_1,\cdots,b_n)=b_1+b_2x+\cdots b_nx^{n-1} \tag{1.3}$$

用曲线拟合求出$\hat{b}_1,\cdots,\hat{b}_n$，就得到经验公式

$$y=\hat{b}_1+\hat{b}_2x+\cdots+\hat{b}_nx^{n-1}$$

必须指出，曲线拟合首先碰到的问题是函数关系(1.1)的具体确定，然后才能进行参数估值。对于变量x，y之间

已有较明确物理关系或一些简单的问题，要给出一个函数关系的具体表达式不太困难，但对某些复杂问题要建立一个有效的表达式就不容易了。本书的任务是讨论曲线拟合的最小二乘计算方法，因此假定函数关系的具体形式都是已知的，至于如何建立函数关系具体表达式（即构造数学模型）的问题将不涉及。

为弄清上述一些概念，试看一个最简单的例子。

设观测数据(x_k, y_k)，$k=1, 2, \cdots, 5$，如图1.1和表1.1所示。现在要求出x和y的关系式。由图可见，这些点大体分布在一条直线上，因此可用线性式

$$y=a+bx \tag{1.4}$$

来拟合。

表1.1

k	x_k	y_k
1	2	2.01
2	4	2.98
3	5	3.50
4	8	5.02
5	9	5.47

这时各观测值大体满足如下方程组

$$\begin{cases} a+2b=2.01 \\ a+4b=2.98 \\ a+5b=3.50 \\ a+8b=5.02 \\ a+9b=5.47 \end{cases} \tag{1.5}$$

式中a，b是待定参数，从上述5个方程求解2个未知数，不可能得到通常意义下的解，因为(1.5)中每二个方程都可求出一组解(a,b)，这些解一般是不相同的。

图1.1

一般，在实际问题中，只要观测点数m大于待定参数的个数n，则列出的方程组的个数m就大于未知数的个数n，即得出超定方程组。由于超定方程组得到的解会出现互相矛盾的现象，因此它又称为矛盾方程组。曲线拟合中参数的确定问题，实质上就是求解矛盾方程组的问题。

在上述例子中，如果用某种方法确定出参数a，b，则给出x后便可算出y，记作

$$y_k^* = a + bx_k, \quad k=1, 2, \cdots, 5 \tag{1.6}$$

称为y_k的估计值。由于数据的误差和表达式的不精确等原因，y_k^*与观测值y_k一般是不完全相同的。它们之间的差

$$r_k = y_k - y_k^* = y_k - (a + bx_k) \quad (k=1, 2, \cdots, 5) \tag{1.7}$$

称为残差。在原始数据给定的情况下，残差仅依赖于参数a，b的取值，因此残差的大小就是衡量被确定参数好坏的基本标志。

1.2 最小二乘准则

对于曲线拟合问题，我们可以采用各种原则来确定(1.1)

中的参数b_1，…，b_n，例如用残差作为拟合标准，此时

$$r_i=y_i-f(x_i;\ b_1,\ \cdots,\ b_n)$$

$$(i=1,\ 2,\ \cdots,\ m)$$

简记作

$$\boldsymbol{r}=\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{x};\ \boldsymbol{b}) \tag{1.8}$$

这里，$\boldsymbol{r}=(r_1, r_2,\ \cdots,\ r_m)^T$

$\boldsymbol{b}=(b_1,\ b_2,\ \cdots,\ b_n)^T$

$\boldsymbol{f}(\boldsymbol{x},\ \boldsymbol{b})=(f(x,\ \boldsymbol{b}),\ f(x_2, \boldsymbol{b}), \cdots,\ f(x_m, \boldsymbol{b}))^T$

残差向量r的三种范数记作

$$\|\boldsymbol{r}\|_\infty=\max_{1\leqslant i\leqslant m}|r_i|$$

$$\|\boldsymbol{r}\|_1=\sum_{i=1}^{m}|r_i|$$

$$\|\boldsymbol{r}\|_2=\left[\sum_{i=1}^{m}r_i^2\right]^{1/2}$$

很明显，残差可以表示拟合的误差，误差愈小拟合就愈好。虽然取前2种范数$\|\boldsymbol{r}\|_1$，或$\|\boldsymbol{r}\|_\infty$最小，比较直观和理想，但它们不便于计算，因此实用中是取欧氏范数$\|\boldsymbol{r}\|_2$最小，即求出参数$\boldsymbol{b}$，使

$$\sum_{i=1}^{m}[y_i-f(x_i;\ \boldsymbol{b})]^2=\min \tag{1.9}$$

这就是通常所谓的最小二乘法，几何语言也称最小二乘曲线拟合。

最小二乘法是一个比较古老的方法，早在十八世纪，Gauss就首先创立并成功地应用于天文观测和大地测量工作

中，此后的二百年来，它已广泛应用于科学实验与工程技术中。随着现代电子计算机的普及与发展，这个古老方法更加显示出其强大的生命力，它在曲线拟合、函数逼近、数据处理、方差分析与回归分析中都经常应用。特别是在当前非常活跃的一门新兴学科最优化技术中，最小二乘法大有用武之地，它是求一切平方和形式的目标函数的最优解的一个基本方法。

近二十年来，最小二乘法已有很大的改进与发展。例如解线性最小二乘问题，过去独一无二的经典方法是构造法方程组，但近年来提出的一些新算法，却主张直接从矛盾方程入手，以避免出现法方程组严重病态的困难；非线性最小二乘问题发展尤为迅速，新的算法不断产生。此外，在应用数学、物理、统计、数学规划、经济、控制论以及社会科学中提出的各种形式的带约束条件的最小二乘问题，其计算方法也在逐步建立和发展中。

从最优化方法的观点解释最小二乘法，就是要求理论函数$y=f(x;\ \boldsymbol{b})$中的参数$\boldsymbol{b}$在最小二乘意义下的最佳估计值$\boldsymbol{b}=(b_1,\cdots,\ b_n)^T$，也就是求使目标函数

$$Q=\sum_{i=1}^{m}[y_i-f(x_i;\ \boldsymbol{b})]^2 \tag{1.10}$$

取最小值时的参数$\boldsymbol{b}=(b_1,\ b_2,\cdots,b_n)^T$。有时，为了保证拟合的精度，常用各观测点残差的加权平方和作为目标函数，即求$\boldsymbol{b}$使

$$Q=\sum_{i=1}^{m}\omega_i r_i^2=\sum_{i=1}^{m}\omega_i[y_i-f(x_i;\ \boldsymbol{b})]^2 \tag{1.11}$$

为最小。其中$\omega_i>0$称为在观测点$(x_i;\ y_i)$处的权。关于权

我们可以粗略地理解为在进行实验观测时，有ω_i次重复得到这个观测点。

附带指出，在实验观测值的总体y服从正态分布的情况下，最小二乘法与数理统计中的最大拟然法是一致的。事实上，对应于每一个$x_i(i=1, 2, \cdots, m)$，y的观测值y_i是一随机变量，其理论值$y_i^*=f(x_i, \boldsymbol{b})$是$y_i$的数学期望值，理论方差为$\sigma_i^2$，观测值$(y_1, y_2, \cdots, y_m)$的似然函数为

$$L=\prod_{i=1}^{m}\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left\{-\frac{1}{2}\left[\frac{y_i-f(x_i,\boldsymbol{b})}{\sigma_i}\right]^2\right\}$$

$$=\frac{1}{(2\pi)^{m/2}\sigma_1\sigma_2\cdots\sigma_m}\exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\frac{[y_i-f(x_i,\boldsymbol{b})]^2}{\sigma_i^2}\right\} \tag{1.12}$$

最大似然法是求参数$\boldsymbol{b}$使似然函数L取最大值，即

$$L=\max$$

这等价于使

$$\sum_{i=1}^{m}\frac{1}{\sigma_i^2}[y_i-f(x_i, \boldsymbol{b})]^2=\min \tag{1.13}$$

由于观测值y_i的权重因子为$\omega_i=\frac{1}{\sigma_i^2}$，所以最大似然准则(1.12)与最小二乘准则(1.11)完全一致。

由(1.11)式定义的Q，若$y_i\sim N(y_i^*, \sigma_i)$，则$Q\sim x^2(m)$，由数学分析知，使

$$Q=\sum_{i=1}^{m}\frac{1}{\sigma_i^2}\left[y_i-f(x_i, \boldsymbol{b})\right]^2$$

为最小时的参数$\hat{\boldsymbol{b}}=(\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_n)^T$，必须满足

$$\left.\frac{\partial Q}{\partial b_j}\right|_{\boldsymbol{b}=\hat{\boldsymbol{b}}}=0 \quad (j=1, 2, \cdots, m) \tag{1.14}$$

因此$Q_{min}=\sum_{i=1}^{m}\frac{1}{\sigma_i^2}[y_i-f(x_i;\ \hat{b}_1,\hat{b}_2,\cdots,\hat{b}_n)]^2$

受(1.14)式m个方程的约束，所以

$$Q_{min}\sim\chi^2(m-n) \tag{1.15}$$

即最小的Q值服从自由度为$m-n$的χ^2分布。利用它可用来检验理论曲线$y=f(x;\ \boldsymbol{b})$的函数形式选取得是否合适。因为一个曲线拟合，应该要求Q_{min}的值在χ^2分布的1%点与99%点之间。由数理统计学知，当χ^2分布的自由度r很大时，$\sqrt{2\chi^2(r)}$近似服从正态分布N($\sqrt{2r-1}$,1)，亦即$\sqrt{2\chi^2}-\sqrt{2r-1}$近似地服从N(0, 1)，从而可得

$$\sqrt{2\chi_\alpha^2(r)}-\sqrt{2r-1}\approx u_\alpha$$

其中u_α是$N(0,1)$的上侧α分位点。由上式解出χ^2分布的上侧α分位点的近似值，即

$$\chi_\alpha^2(r)\approx\frac{1}{2}\ (u_\alpha+\sqrt{2r-1}\ )$$

实际应用中，当$r>45$时即可使用这个近似公式。经过查$N(0,1)$表和简单的计算，我们得到

$$\chi_{0.99}^2(r)\approx r+2.2-3.3\sqrt{r}$$

$$\chi_{0.01}^2(r)\approx r+2.2+3.3\sqrt{r}$$

由于曲线拟合中观测数据的个数m一般很大，且$m>>n$，因而$r=m-n$也很大，所以曲线拟合所得Q_{min}的值，应在区间

$$(r+2.2-3.3\sqrt{r},\ r+2.2+3.3\sqrt{r}\ )$$

之内。在实际应用中，常用Q_{min}的值是否靠近$r=m-n$作为

判别拟合好坏的标准（因$E(\chi^2(r))=r$），如果 Q_{min} 的值远远大于或远远小于r，即落在上述区间之外，则表明拟合结果不合理，应当检查出现这种情况的原因。

由于最小二乘法主要使用欧氏范数$\|\cdot\|_2$，为简便起见，本书省去它的下标而记作$\|\cdot\|$，如用其它范数时则写出下标，以防混淆。