

曲线拟合的最小二乘法

顾乐民

(同济大学 材料科学与工程学院, 上海 201804)

摘要: 最小一乘法的解, 由于存在着绝对值方程而不便于计算, 成为困扰数理界 200 多年悬而未决的难题. 基于对最小一乘准则下各种数学模型的大量计算和长期研究后发现, 若存在最小一乘最佳参数 $\mathbf{a} = \mathbf{a}^* \in R_n$ 使绝对偏差值和为极小的最小一乘准则 $\sum_{i=1}^m |y_i - f(x_i, \mathbf{a}^*)| = \min$ 成立, 则拟合函数 $f(x, \mathbf{a}^*)$ 的表征为: 至少存在 n 个点 x_1, x_2, \dots, x_n , 使 $y_i - f(x_i, \mathbf{a}^*) = 0, i = 1, 2, \dots, n (n \leq m)$ 成立, 从而最小一乘解可以实现.

关键词: 曲线拟合; 最小一乘; 逼近

中图分类号: O 241.5

文献标识码: A

Least Absolute Deviation Method of Curve Fitting

GU Lemin

(College of Material Science and Engineering, Tongji University, Shanghai 201804, China)

Abstract: The solution of least absolute deviation (LAD), a pending problem for more than 200 years in mathematics, is not easy to calculate because of the absolute value function. Based on a great deal of computing and long-term study of various mathematical models under LAD criteria, a conclusion is drawn that if there is a LAD parameter $\mathbf{a} = \mathbf{a}^* \in R_n$, and making the following LAD criterion tenable $\sum_{i=1}^m |y_i - f(x_i, \mathbf{a}^*)| = \min$, then the fitting function $f(x, \mathbf{a}^*)$ can be characterized that there are at least n points x_1, x_2, \dots, x_n , making $y_i - f(x_i, \mathbf{a}^*) = 0, i = 1, 2, \dots, n (n \leq m)$ valid, the problem of LAD solution can be achieved.

Key words: curve fitting; least absolute deviation; approximation

以得到更稳健的回归方程.

最小一乘准则是 1755—1757 年间数学家 Boscovitch 和 Laplace 在研究拟合直线时提出的, 比最小二乘法早 40 多年. 但由于计算上的困难, 最小一乘法的发展一直处于停滞状态. 1955 年, Charnes, Cooper 和 Ferguson 等人在研究一个特定的管理问题中使用了最小一乘法, 他们通过将偏差表示为 2 个非负变量之差的形式, 将其转化为线性规划问题来求解^[1], 局面才有所改变. 1987 年 9 月瑞士召开了有关这个准则研究的国际会议. 目前, 最小一乘法逐渐成为统计学研究领域的热点之一^[1-5].

设 Φ 表示定义在某闭区间上所有实值连续函数构成的集合, 或给定的某一函数类. 拟合函数 $f(x) \in \Phi$ 在所定义的区间内光滑连续且 1 阶可导. 对于给定 m 组离散数据 $(x_i, y_i), i = 1, 2, \dots, m$, 取曲线拟合方程 $y = f(x, \mathbf{a})$, 其中 $\mathbf{a} = (a_1, a_2, \dots, a_n)^T (n \leq m)$.

最小二乘法 原名 least squares method, 是依据使偏差平方和为极小的准则—— $\sum_{i=1}^m [y_i - f(x_i, \mathbf{a})]^2 = \min$ 来选择参数值 \mathbf{a} 而构成的一种曲线拟合法.

最小一乘法 原名 least absolute deviation method, 是依据使绝对偏差值和为极小的准则—— $\sum_{i=1}^m |y_i - f(x_i, \mathbf{a})| = \min$ 来选择参数值 \mathbf{a} 而构成的一种曲线拟合法.

最小一乘法具有直观和理想的特点. 但由于存在绝对值而不便于计算, 使相应方法的建立受到了限制. 为了去掉绝对值, 引入以下一个定理.

定理 若存在 $\mathbf{a} = \mathbf{a}^* \in R_n$, 使目标函数

$$Q = Q(\mathbf{a}^*) = \sum_{i=1}^m |y_i - f(x_i, \mathbf{a}^*)| = \min \quad (1)$$

成立. 则拟合函数 $f(x, \mathbf{a}^*)$ 的表征为: 至少存在 n

最小一乘法和最小二乘法是曲线拟合中常用的方法. 但是, 在近代关于数理统计中稳健性的研究发现, 用最小二乘估计有时不很理想. 例如某个异常数据会使回归方程有大的偏离^[1]. 于是, 人们提出其他方法来克服这个缺点, 比如使用最小一乘准则, 就可

个点的 x_1, x_2, \dots, x_n , 使

$$y_i - f(x_i, \mathbf{a}^*) = 0 \quad i = 1, 2, \dots, n \quad (2)$$

成立, 称这 n 个零偏差点为“0”点. 反之若式(2)构成的方程组不成立, 则式(1)目标函数 Q 的极小化也不成立.

对于该定理说明以下几点.

(1) 对于函数 $f(x) = f(x, \mathbf{a})$ 的类型不作明确的规定. 对函数 $f(x, \mathbf{a})$ 是 \mathbf{a} 的线性函数时的最小一乘问题, 称之为“线性函数类最小一乘问题”, 例如多项式函数等. 已有不少算法^[1-5]等都是十分有效的. 此时最佳逼近函数 $f(x)$ 并不一定唯一. 例如在直线型方程中, 有时可以找到多个 $f(x)$, 使目标函数 Q 的极小化得以成立, 但这不影响问题的讨论. 对函数 $f(x, \mathbf{a})$ 是 \mathbf{a} 的非线性函数时的最小一乘问题, 称之为“非线性函数类最小一乘问题”, 例如 Logistic 方程、Morgan-Mercer-Flodin 方程等.

(2) 定理中若令 $m = n$, 则只能存在一个由式(2)表示的方程组, 如果这个方程组无解, 则式(1)表示的目标函数极小化也不能存在. 如果有解, 则目标函数 Q 必为零, 最小一乘获得唯一最佳解; 反之, 其逆向也是存在的, 为使 Q 极小, 则方程组必须有解. 这与函数 $f(x, \mathbf{a})$ 的类型(线性或非线性的)无关. 当 $m > n$ 时可通过比较各基本解使目标函数 Q 极小化, 其过程在后文的“直接法”中介绍.

(3) 最小一乘解的具体实现和求解方式分为两个环节: ①第一环节用解析方式作为寻优过程. 通过建立一种近似关系, 将不可微问题转化为解析的方式进行, 解决了长期困扰数理界的因存在着绝对值方程而不便于计算的难点, 并得到最小一乘逼近的近似解. ②第二环节用解方程组方式获得最小一乘逼近的最佳解. 通过由定理构成的方法, 将近似解上升为最佳解. 这两个环节是通过优化判别法则来连接的.

这里的最佳解指的是一定精确度范围内的准确解.

1 最小一乘逼近问题的近似解

对于存在绝对值的式(1), 不能直接用解析的方法得到最佳参数 \mathbf{a}^* , 但给 \mathbf{a}^* 一个微偏离 $\mathbf{a}^* = \mathbf{a} + \Delta\mathbf{a}$, 当 $\Delta\mathbf{a}$ 充分小时, 可建立以下近似关系:

$$\|y - f(x, \mathbf{a}^*)\|_1 = \min_{\mathbf{a} + \Delta\mathbf{a} \in R_n} \|y - f(x, \mathbf{a} + \Delta\mathbf{a})\|_1 \approx$$

$$\min_{\mathbf{a} + \Delta\mathbf{a} \in R_n} \left\| y - f(x, \mathbf{a}) - \frac{\partial f(x, \mathbf{a})}{\partial \mathbf{a}} \Delta\mathbf{a} \right\|_1$$

如此可化为用解析方式求 $\mathbf{a} + \Delta\mathbf{a}$ 并得到最小一乘逼近的近似解, 这是为求最佳解 \mathbf{a}^* 过程中的第一环节, 称为“寻优”环节.

1.1 一般函数类

对于式(2), 由多元函数求极值的必要条件 $\frac{\partial Q}{\partial a_1}$

$= 0, \frac{\partial Q}{\partial a_2} = 0, \dots, \frac{\partial Q}{\partial a_n} = 0$, 可得方程组

$$\left. \begin{aligned} \frac{\partial Q}{\partial a_1} &= - \sum_{i=1}^m \frac{y_i - f(x_i)}{|y_i - f(x_i)|} \cdot \frac{\partial f}{\partial a_1} = 0 \\ \frac{\partial Q}{\partial a_2} &= - \sum_{i=1}^m \frac{y_i - f(x_i)}{|y_i - f(x_i)|} \cdot \frac{\partial f}{\partial a_2} = 0 \\ &\vdots \\ \frac{\partial Q}{\partial a_n} &= - \sum_{i=1}^m \frac{y_i - f(x_i)}{|y_i - f(x_i)|} \cdot \frac{\partial f}{\partial a_n} = 0 \end{aligned} \right\} \quad (3a)$$

或

$$\left. \begin{aligned} \sum_{i=1}^m \frac{r_i}{|r_i|} \cdot \frac{\partial f}{\partial a_1} &= 0 \\ \sum_{i=1}^m \frac{r_i}{|r_i|} \cdot \frac{\partial f}{\partial a_2} &= 0 \\ &\vdots \\ \sum_{i=1}^m \frac{r_i}{|r_i|} \cdot \frac{\partial f}{\partial a_n} &= 0 \end{aligned} \right\} \quad r_i \neq 0 \quad (3b)$$

式中, $r_i = y_i - f(x_i)$, $\frac{\partial f}{\partial a_j} = \frac{\partial f(x, \mathbf{a})}{\partial a_j}$, $j = 1, 2, \dots, n$. 函数 $f(x_i, \mathbf{a})$ 在点 $\mathbf{a}^{(0)} = (a_1^{(0)}, a_2^{(0)}, \dots, a_n^{(0)})^T$ 处展成 1 阶 Taylor 多元表式

$$f(x, \mathbf{a}) = f(x, \mathbf{a}^{(0)}) + \varphi_1 \Delta a_1 + \varphi_2 \Delta a_2 + \dots$$

$$+ \varphi_n \Delta a_n = f(x, \mathbf{a}^{(0)}) + \sum_{j=1}^n \varphi_j \Delta a_j$$

$$r_i = r_i(\mathbf{a}) = y_i - f(x_i, \mathbf{a}) = y_i - f(x_i, \mathbf{a}^{(0)}) -$$

$$\sum_{j=1}^n \varphi_{i,j} \Delta a_j, \quad 1 \leq i \leq m$$

其中

$$\varphi_j = \frac{\partial f(x, \mathbf{a})}{\partial a_j} \Big|_{\mathbf{a}=\mathbf{a}^{(0)}}$$

$$\varphi_{i,j} = \frac{\partial f}{\partial a_j} = \frac{\partial f(x_i, \mathbf{a})}{\partial a_j} \Big|_{\mathbf{a}=\mathbf{a}^{(0)}}$$

$$j = 1, 2, \dots, n$$

$$\Delta\mathbf{a} = \mathbf{a} - \mathbf{a}^{(0)} = (\Delta a_1, \Delta a_2, \dots, \Delta a_n)^T$$

$$\Delta a_1 = a_1 - a_1^{(0)}$$

$$\Delta a_2 = a_2 - a_2^{(0)}, \dots, \Delta a_n = a_n - a_n^{(0)}$$

$$r_i^{(0)} = y_i - f(x_i, a^{(0)}), \quad 1 \leq i \leq m$$

代入式(3),消去分子项 r_i ,有

$$\begin{cases} \sum_{i=1}^m \frac{1}{|r_i|} \varphi_{1,i} (\varphi_{1,i} \Delta a_1 + \varphi_{2,i} \Delta a_2 + \dots + \varphi_{n,i} \Delta a_n) \approx \sum_{i=1}^m \frac{r_i^{(0)}}{|r_i|} \varphi_{1,i} \\ \sum_{i=1}^m \frac{1}{|r_i|} \varphi_{2,i} (\varphi_{1,i} \Delta a_1 + \varphi_{2,i} \Delta a_2 + \dots + \varphi_{n,i} \Delta a_n) \approx \sum_{i=1}^m \frac{r_i^{(0)}}{|r_i|} \varphi_{2,i} \\ \vdots \\ \sum_{i=1}^m \frac{1}{|r_i|} \varphi_{n,i} (\varphi_{1,i} \Delta a_1 + \varphi_{2,i} \Delta a_2 + \dots + \varphi_{n,i} \Delta a_n) \approx \sum_{i=1}^m \frac{r_i^{(0)}}{|r_i|} \varphi_{n,i} \end{cases} \quad r_i \neq 0$$

为方便书写,记 $\Sigma = \sum_{i=1}^m$, $\varphi_j = \varphi_{i,j}$;将“ k ”代替上式中的“0”, k 表示迭代次数, $k = 0, 1, \dots$;用“ $=$ ”代替“ \approx ”,则这个线性方程组可用矩阵表示

$$\begin{pmatrix} \sum \frac{\varphi_1 \varphi_1}{|r_i|} & \sum \frac{\varphi_1 \varphi_2}{|r_i|} & \dots & \sum \frac{\varphi_1 \varphi_n}{|r_i|} \\ \sum \frac{\varphi_2 \varphi_1}{|r_i|} & \sum \frac{\varphi_2 \varphi_2}{|r_i|} & \dots & \sum \frac{\varphi_2 \varphi_n}{|r_i|} \\ \vdots & \vdots & & \vdots \\ \sum \frac{\varphi_n \varphi_1}{|r_i|} & \sum \frac{\varphi_n \varphi_2}{|r_i|} & \dots & \sum \frac{\varphi_n \varphi_n}{|r_i|} \end{pmatrix} \begin{pmatrix} \Delta a_1 \\ \Delta a_2 \\ \vdots \\ \Delta a_n \end{pmatrix} = \begin{pmatrix} \sum \frac{r_i^{(k)}}{|r_i|} \varphi_1 \\ \sum \frac{r_i^{(k)}}{|r_i|} \varphi_2 \\ \vdots \\ \sum \frac{r_i^{(k)}}{|r_i|} \varphi_n \end{pmatrix} \quad (4)$$

式中的 $\Delta a_1, \Delta a_2, \dots, \Delta a_n$ 作为迭代过程的微变量,服从

$$\begin{aligned} a_i^{(k+1)} &= a_i^{(k)} + \Delta a_i, \dots, a_n^{(k+1)} = a_n^{(k)} + \Delta a_n \\ k &= 0, 1, \dots \end{aligned} \quad (5)$$

作为 $k=0$ 时的迭代初始值 $a_1^{(0)}, a_2^{(0)}, \dots, a_n^{(0)}$,可自行设定,也可取最小二乘解的结果.这也就是说,可先进行最小二乘法计算,取得最小二乘估计值,再进行最小一乘法的计算.

记

$$D[x_1, x_2, \dots, x_n] :=$$

$$\begin{vmatrix} \sum \frac{\varphi_1 \varphi_1}{|r_i|} & \sum \frac{\varphi_1 \varphi_2}{|r_i|} & \dots & \sum \frac{\varphi_1 \varphi_n}{|r_i|} \\ \sum \frac{\varphi_2 \varphi_1}{|r_i|} & \sum \frac{\varphi_2 \varphi_2}{|r_i|} & \dots & \sum \frac{\varphi_2 \varphi_n}{|r_i|} \\ \vdots & \vdots & & \vdots \\ \sum \frac{\varphi_n \varphi_1}{|r_i|} & \sum \frac{\varphi_n \varphi_2}{|r_i|} & \dots & \sum \frac{\varphi_n \varphi_n}{|r_i|} \end{vmatrix}$$

当 $\varphi_j = \frac{\partial f(x, a)}{\partial a_j}$ ($j = 1, 2, \dots, n$) 线性无关, $r_i \neq 0$,

$D[x_1, x_2, \dots, x_n, x_{n+1}] \neq 0$, 式(4)有唯一解.从式(5)中解出 $a_j^{(k+1)}$ ($j = 1, 2, \dots, n$),直到精度符合要求便停止.取

$$a_j \approx a_j^{(k+1)}, \quad j = 1, 2, \dots, n \quad (6)$$

1.2 线性函数类

如果拟合的曲线方程为多项式 $f(x) = a_1 + a_2$

$x + a_3 x^2 + \dots + a_n x^{n-1} = \sum_{j=1}^n a_j x^{j-1}$, 则式(4)有

$$\begin{pmatrix} \sum \frac{1}{|r_i|} & \sum \frac{x_i}{|r_i|} & \dots & \sum \frac{x_i^{n-1}}{|r_i|} \\ \sum \frac{x_i}{|r_i|} & \sum \frac{x_i^2}{|r_i|} & \dots & \sum \frac{x_i^n}{|r_i|} \\ \vdots & \vdots & & \vdots \\ \sum \frac{x_i^{n-1}}{|r_i|} & \sum \frac{x_i^n}{|r_i|} & \dots & \sum \frac{x_i^{2n-2}}{|r_i|} \end{pmatrix} \begin{pmatrix} \Delta a_1 \\ \Delta a_2 \\ \vdots \\ \Delta a_n \end{pmatrix} = \begin{pmatrix} \sum \frac{r_i^{(k)}}{|r_i|} \\ \sum \frac{r_i^{(k)}}{|r_i|} x_i \\ \vdots \\ \sum \frac{r_i^{(k)}}{|r_i|} x_i^{n-1} \end{pmatrix}, \quad r_i \neq 0 \quad (7)$$

式(4)、式(7)即为最小一乘逼近解的行列式表示,由于分母不能为零,所以迭代只能有限 k 次进行.虽得到的是近似解,但为求得最佳解完成了必要的“寻优”环节.

2 优化集判别法则和最小一乘逼近的最佳解

2.1 直接法

由定理,通过解方程组式(2)获得最佳参数 a^* ,

从而使目标函数 Q 极小化得以实现的方法称为“最小一乘直接解法”,简称直接法.

直接法考虑由 m 个方程中任意不重复地取出 n 个方程构成一个满足式(2)的方程组,从而获得1个基本解.比较各基本解,其总数不超过 $m! / [(m-n)! n!]$ 个,必可找到一个最优解,该解即为最小一乘问题的最佳解.显然,当 $m = n$ 时,只存在1个方程组,如果这个方程组无解,则最小一乘逼近也不存在;如果有解,则目标函数 Q 必为零,最小一乘获得唯一最佳解.

当 $m > n$ 时,虽仅用了 n 组数据,剩下的 $m - n$ 组数据是否属于无用? 并非如此.因为究竟用哪套 n 组数据,需要由全部的 m 组数据来决定,是由 m 组数据中选出最好的 n 组构成的.

只需且至少存在1个有解的方程组,而剩下的 $m - n$ 组数据只要能够服从该方程,就必然可使目标函数 Q 的极小化得以实现.这与函数 $f(x, a)$ 的(线性或非线性的)类型无关,也为求解非线性函数类的最小一乘问题提供了解决的可行性.

这两种算法各有利弊.直接法在理论上可行,但常因计算量太大而不实用,且在 $m! / [(m-n)! n!]$ 个方程组中只要有1个无解,就可能造成整个计算的中断;近似解法采用了迭代方法使得连续计算成为可能,但由于分母不能为零的限制而无法获得最佳值,其解只是一种近似.

现考虑吸收这两种方法的各自长处:设法找到直接法中最佳解所对应的“0”点,就获得问题的解决,而寻优过程可用迭代方式进行.这样计算的总成本是有限次的迭代再加上解一个方程组.这是一个比较经济的算法.

由此引出了一个判别法则,用以判别“寻优”过程结束、找到了“0”点、转向求最佳解的节点.它由优化集、优化区、优化集判别法则等内容构成.

2.2 优化集与优化区

对于曲线拟合方程 $y = f(x, a)$, 设 a^* 是逼近的最佳参数,以 a^* 为核心, $a^* = a^* + \Delta a$ 为边界所构成的区域称为“优化区”,在优化区内,点 a 的全体构成的集合称为“优化集”,用符号 R_n^* 表示.

$$R_n^* = \{a^* \mid \max_{1 \leq j \leq n} |r_j(a^*)| \leq \min_{n+1 \leq i \leq m} |r_i(a^*)|\}$$

不讨论 R_n^* 的范围的大小等问题,而只需知道在优化集内的优化解 a^* 和最佳解 a^* 具有相同的特

征.在迭代 k 次后,当参数 a 一旦进入了优化集 R_n^* 后,无须再在计算上花“成本”,即没有必要继续迭代让 a^* 逐步趋于 a^* ,而是可以结束迭代,转向求解方程组得到 a^* ,并获得最小一乘法问题的最佳解.

2.3 优化集判别法则

在一定的 k 次迭代后,得到了 m 个绝对偏差值: $|r_1^{(k)}|, |r_2^{(k)}|, \dots, |r_n^{(k)}|, \dots, |r_m^{(k)}|$. 其中, $|r_1^{(k)}| \leq |r_2^{(k)}| \leq \dots \leq |r_n^{(k)}| \leq \dots \leq |r_m^{(k)}|$, 等号通常在 $n = m$ 时成立.分为两组:第一组由 $|r_j^{(k)}| (j = 1, 2, \dots, n)$ 组成,共有 n 个;第二组由余下的 $m - n$ 个 $|r_i^{(k)}| (i = n + 1, \dots, m)$ 组成.

继续迭代,当第一组中的最大绝对偏差 $\max |r_j| (j = 1, 2, \dots, n)$ 始终不再大于第二组中最小绝对偏差 $\min |r_i| (i = n + 1, \dots, m)$, 即 $\max_{1 \leq j \leq n} |r_j(a^{(k)})| \leq \min_{n+1 \leq i \leq m} |r_i(a^{(k)})|$, 同时第一组所有的 (x_j, y_j) 也不会再到第二组里去,则称第一组所对应的 n 个点为“0”点,称参数 a 为进入优化集的优化解 a^* .

2.4 具体算法一

以上假设了一个优化集,在实际应用中,并不一定要确定优化集的临界范围.这是因为迭代过程只是“寻优”并不能求得最佳解,用小的计算成本达到目的是很有必要的.

初始工作:设定初始值 $a^{(0)}$, 如可取最小二乘估计值.

(1) 迭代足够多 k 次,得第 k 次迭代结果 $a^{(k)}$ 和 $r_i^{(k)}$,

(2) 将总数为 m 个绝对偏差值作一由小至大的排列: $|r_1^{(k)}| \leq |r_2^{(k)}| \leq \dots \leq |r_n^{(k)}| \leq \dots \leq |r_m^{(k)}|$.

(3) 取前 $1, 2, \dots, n$ 共 n 个对应的 (x_j, y_j) 构成1个方程组,代入式(2)解出参数 a , 并记为 a^* .

(4) 分别计算 $\sum_{i=1}^m |r_i(a^*)|$ 和 $\sum_{i=1}^m |r_i(a^{(k)})|$;

(5) 比较大小.若 $\sum_{i=1}^m |r_i(a^*)| \leq \sum_{i=1}^m |r_i(a^{(k)})|$ 成立,可以转(6)再进行多次验证;若始终成立,则打印 $a^* = a^*$, 结束.

(6) 否则,置 $k: k + 1$, 转最小一乘继续迭代.在获得新 $a^{(k)}$ 后再转(1)判别.

2.5 具体算法二

对于“具体算法一”中的(3),多取1组数据,即取前

1,2,⋯,n+1 个对应的 (x_j, y_j) ,考虑由这 $n+1$ 个方程中任意不重复地取出 n 个方程构成 1 个方程组,从而获得 1 个基本解,基本解的总数不超过 $n+1$ 个. 比较各基本解,必可找到最小一乘问题的解.

例如,对于 3 参数方程,直接法确定的“0”点至少为 $n=3$ 个,不妨就设定为 3 个. 为防止假“0”点的出现(一种迭代次数不够多时有可能出现的情况),可以多取 1 组数据 $j=1,2,3,4$,于是构成了 $n+1=4$ 个基本解: $j=1,2,3, j=1,2,4, j=1,3,4$ 和 $j=2,3,4$. 比较各基本解,必可找到最小一乘问题的最佳解. 当然也可多取 2 组数据,而解必然在总数不超过 $(n+1)(n+2)/2$ 个基本解中找到. 具体算法二增大了计算工作量,但对于最佳解的获得起到一定保证作用.

3 应用实例

例 1 求解 MMF 模型: $f(x) = \frac{a_1 a_2 + a_3 x^{a_4}}{a_2 + x^{a_4}}$.

MMF 模型含有 4 个参数的 S 形生长模型,求解有一定的难度. 为简化表达,这里只用 a_1, a_2, a_3, a_4 , 4 个参数来表示,并略去对参数的具体介绍.

固结系数是反映土层固结特性的参数,在基础沉降计算中具有十分重要的意义. 表 1 第 2 列是文献[6]给出的某土体沉降 y 与时间 x 关系的 15 组原始数据,采用 MMF 方程对沉降变化进行描述,用最小一乘法对 MMF 方程加以拟合. 拟合值为 $f(x_i)$, 相对误差为 r_i/y_i .

表 1 最小一乘准则下的 MMF 曲线拟合
Tab.1 MMF curve fitting in the criterion of least absolute deviation

i	x_i/s	y_i/cm	$f(x_i)/cm$	$(r_i/y_i)/\%$
1	15	0.018	0.018	0
2	30	0.023	0.023 4	-1.57
3	60	0.036	0.032 3	10.19
4	120	0.049	0.046 7	4.75
5	240	0.068	0.068	0
6	540	0.100	0.102 0	-2.00
7	960	0.128	0.129 9	-1.45
8	1 500	0.151	0.151 5	-0.35
9	2 160	0.169	0.168 2	0.49
10	2 940	0.181	0.181	0
11	3 600	0.190	0.188 7	0.69
12	5 400	0.202	0.202 1	-0.07
13	7 200	0.211	0.210 1	0.41
14	12 600	0.222	0.222 3	-0.13
15	18 000	0.228	0.228	0

先作最小二乘法计算,得估计值并作为最小一乘迭代的初始值. 迭代 12 次得 $a^{(12)}$, 记 $a^* = a^{(12)}$. 继续迭代 5 次作为检验,仍有 $a^* = a^{(17)}$ 成立,即 a^* 与 $k \geq 12$ 后的迭代无关,可用 $a^* = a^*$ 表示,并得 $f(x) = (0.010\ 8 \times 340.331\ 1 + 0.246\ 3\ x^{0.835\ 9}) / (340.331\ 1 + x^{0.835\ 9})$. 数据处理见第 3 列. 它获得偏差绝对值和 $\sum |r_i| = 0.142$ 的极小化结果.

经计算知,曲线拟合误差 $E_{map} = (1/15) \times \sum |r_i| \times 100\% = 1.47\%$,也要小于最小二乘法中 $E_{map} = 2.06\%$ 的结果.

例 2 求解 Logistic 模型: $f(x) = \frac{a_1}{1 + a_2 e^{-a_3 x}}$.

Logistic 模型是含有 3 个参数的 S 型增长模型,为简化表达,这里只用 a_1, a_2, a_3 , 3 个参数来表示.

表 2 摘自《中国人口统计年鉴》中的 1990 年至 2008 年的 19 组数据,运用最小一乘法对 Logistic 模型加以拟合,并预测 2009 年以及 2010 年的人口数(已知 2009 年人口数为 133 474 万).

用最小二乘估计 $a^{(0)} = (141.7, 0.258, 0.07)$ 作为最小一乘迭代的初始值,由式(4)迭代 9 次得 $a^{(9)}$. 由判别法则知, $a^{(9)}$ 已进入了优化区,可用 $a^* = a^{(9)}$ 表示. 将位于 $i=2,7,18$ 的 3 组数据代入方程组,得最佳解 $a^* = a^*$ 及 Logistic 方程: $f(x) = 142.3 / (1 + 0.261\ 9\ e^{-0.068(x-1989)})$. 数据处理结果见表 2 第 3 列,它获得了绝对偏差值和为 1 245 万的最小结果.

表 2 1990 年至 2008 年我国人口数及其拟合结果
Tab.2 Population in China (1990—2008) and their fitting results

i	$x_i/\text{年}$	$y_i/\text{万}$	$f(x_i)/\text{万}$	$r_i/\text{万}$
1	1990	114 333	114 326	7
2	1991	115 823	115 823	0
3	1992	117 171	117 258	-87
4	1993	118 517	118 631	-114
5	1994	119 850	119 943	-93
6	1995	121 121	121 195	-74
7	1996	122 389	122 389	0
8	1997	123 626	123 526	100
9	1998	124 810	124 607	203
10	1999	125 909	125 634	275
11	2000	126 583	126 609	-26
12	2001	127 627	127 534	93
13	2002	128 453	128 411	42
14	2003	129 227	129 240	-13
15	2004	129 988	130 025	-37
16	2005	130 756	130 767	-11
17	2006	131 448	131 468	-20
18	2007	132 129	132 129	0
19	2008	132 802	132 753	49

预测 2009 年人数. 将 $x = 2009$ 代入, 得 $f(x) = 133\,341.3$ 万, 与已知数 133 474 万相差 “- 132.7 万”, 相对误差 0.099 5 %.

有两种途径可以预测 2010 年人数, 一是令 $x = 2010$ 代入式中, 得 $y = 133\,895.7$ 万.

另一种是重新列表, 计算 1990 年至 2009 年的 20 组数据, 并推算 2010 年的预测值. 经过计算得到的方程为: $f(x) = 142.47 / (1 + 0.263\,3e^{-0.067\,4(x-1989)})$. 将 $x = 2010$ 代入, 得预测值 $f(x) = 133\,908.1$ 万. 实际结果有待验证.

例 3 求解带有绝对值的 Richards 方程 $f(x) = \frac{a_1}{|1 + a_2e^{-a_3x}|^{1/a_4}}$.

Richards 模型是含有 4 个参数的 S 型增长模型, 为简化表达, 这里只用 a_1, a_2, a_3, a_4 , 4 个参数来表示.

这里给出 1 个将数值作为离散数据进行拟合的例子. 取 $x_i = 0, 1, \dots, 9$, 而 $y(x_i) = 10\,000 / |1 - 4e^{0.2x_i}|^{1/0.5}$ 并且四舍五入取整数部分作为离散数据 y_i , 将 $(x_i, y_i) i = 1, 2, \dots, 10$ 作为原始数据列于表 3 第 2 列.

表 3 最小一乘准则下的 Richards 曲线拟合				
Tab.3 Richards curve fitting in the criterion of least absolute deviation				
i	原始数据		拟合结果	
	x_i	y_i	$f(x_i)$	$r(x_i)$
1	0	1 111	1 111	0
2	1	662	662	0
3	2	405	405.06	- 0.06
4	3	253	252.72	0.28
5	4	160	160	0
6	5	103	102.43	0.57
7	6	66	66.15	- 0.15
8	7	43	43	0
9	8	28	28.10	- 0.10
10	9	19	18.44	0.56

用最小一乘法对 Richards 曲线方程拟合, 初始值取 $a = (10\,000, -4, -0.2, 0.5)^T$.

经式(4)迭代 20 次得 $a^{(20)}$. 由判别法则知, $a^{(20)}$ 已进

入了优化区, 因为通过继续迭代, 虽获得了 $a^{(30)}, a^{(40)}, \dots$, 但位于 $i = 1, 2, 5, 8$ 的 4 组数据为 “0” 点性质不改变, 可用 $a^* = a^{(20)}$ 表示. 将这些点组数据代入式(2)方程组, 得最佳解 $a^* = a^*$ 及 Richards 方程: $f(x_i) = \frac{9\,541.615\,84}{|1 - 4.127\,42e^{0.214\,6x_i}|^{1/0.530\,23}}$. 数据处理结果见表 3 第 3 列, 它获得了绝对偏差值和 $\sum |y_i - f(x_i)| = 1.71$ 的最小结果.

4 结语

一个 200 多年未能完全解决的问题, 不仅仅是因为绝对值的存在而不可微, 还涉及其他方面的一些理论和方法. 最小一乘法在过去实现不了, 但在科学高度发展的今日, 应该且能够回归到它应有的位置. 这个过程需要有传统理论来补充、客观事实来验证、创新理论来发展. 在不断的探讨中, 最小一乘法的基本理论和方法一定能更好地为社会服务.

参考文献:

[1] Charnes A, Cooper W G, Ferguson R O. Optimal estimation of executive compensation by linear programming [J]. Management Science, 1955(1): 138.

[2] 陈希孺. 最小一乘线性回归: 上[J]. 数理统计与管理, 1989. CHEN Xiru. Least absolute linear regression [J]. Application of Statistics and Management, 1989(5): 48.

[3] Fisher W D. A note on curve fitting with minimum deviations by linear programming [J]. Journal of American Statistical Association, 1961, 56: 359.

[4] Portnoy S, Koenker R. The Gaussian hare and the Laplacian tortoise, computability of squared-error versus absolute-error estimators [J]. Statistical Science, 1997, 12: 279.

[5] 李仲来. 最小一乘法介绍 [J]. 数学通报, 1992(2): 40. LI Zhonglai. Introduction to least absolute deviation method [J]. Bulletin of Maths, 1992(2): 40.

[6] 戴韬. MMF 曲线拟合模型在固结系数计算中的应用 [J]. 山西建筑, 2010, 24: 127. DAI Tao. The application of MMF curve fitting model for the consolidation coefficient evaluation [J]. Shanxi Architecture, 2010, 24: 127.