

第二章 城市供水量的预测模型

——插值与拟合算法

2.1 城市供水量的预测问题

2.1.1 实际问题与背景

为了节约能源和水源,某供水公司需要根据日供水量记录估计未来一时间段(未来一天或一周)的用水量,以便安排未来(该时间段)的生产调度计划。现有某城市7年用水量的历史记录,记录中给出了日期、每日用水量(吨/日)。如何充分地利用这些数据建立数学模型,预测2007年1月份城市的用水量,以制定相应的供水计划和生产调度计划。

表 2.1.1 某城市7年日常用水量历史记录(万吨/日)

日期	20000101	20000102	20061230	20061231
日用水量	122.1790	128.2410	150.40168	148.2064

表 2.1.2 2000-2006年1月城市的总用水量(万吨/日)

年份	2000	2001	2002	2003	2004	2005	2006
用水量	4032.41	4186.0254	4296.9866	4374.852	4435.2344	4505.4274	4517.6993

利用这些数据,可以采用时间序列、灰色预测等方法建立数学模型来预测2007年1月份该城市的用水量。如果能建立该城市的日用水量随时间变化的函数关系,则用该函数来进行预测非常方便。但是这一函数关系的解析表达式是没办法求出来的,那么能否根据历史数据求出该函数的近似函数呢?根据未知函数的已有数据信息求出其近似函数的常用方法有插值法和数据拟合。本章将介绍插值法和数据拟合,并用这两种方法对该城市的供水量进行预测。

2.2 求未知函数近似表达式的插值法

2.2.1 求函数近似表达式的必要性

一般地,在某个实际问题中,虽然可以断定所考虑的函数 $f(x)$ 在区间 $[a,b]$ 上存在且连续,但却难以找到它的解析表达式,只能通过实验和观测得到该函数在有限个点上的函数值(即一张函数表)。显然,要利用这张函数表来分析函数的性态,甚至直接求出其它一些点上的函数值是非常困难的。在有些情况下,虽然可以给出函数 $f(x)$ 的解析表达式,但由于结构相当复杂,使用起来很不方便。面对这些情况,希望根据所得函数表(或结构复杂的解析表达式),构造某个简单函数 $\phi(x)$ 作为未知函数 $f(x)$ 的近似。插值法是解决此类问题的一种常用的经典方法,它不仅直接广泛地应用于生产实际和科学研究中,而且也是进一步学习数值计算的基础。

定义 2.2.1 设函数 $y=f(x)$ 在区间 $[a,b]$ 上连续, 且在 $n+1$ 个不同的点 $a \leq x_0, x_1, \dots, x_n \leq b$ 上分别取值 y_0, y_1, \dots, y_n , 在一个性质优良、便于计算的函数类 Φ 中, 求一简单函数 $\phi(x)$, 使

$$\phi(x_i) = y_i \quad (i=0,1,\dots,n) \quad (2.2.1)$$

而在其它点 $x \neq x_i$ 上作为 $f(x)$ 的近似。称区间 $[a,b]$ 为插值区间, 点 x_0, x_1, \dots, x_n 为插值节点, 称 (2.2.1) 为 $f(x)$ 的插值条件, 称函数类 Φ 为插值函数类, 称 $\phi(x)$ 为函数在节点 x_0, x_1, \dots, x_n 处的插值函数。求插值函数 $\phi(x)$ 的方法称为插值法。

插值函数类 Φ 的取法不同, 所求得的插值函数 $\phi(x)$ 逼近 $f(x)$ 的效果就不同, 它的选择取决于使用上的需要。常用的有代数多项式、三角多项式和有理函数等。当选用代数多项式作为插值函数时, 相应的插值问题就称为多项式插值。这里主要介绍多项式插值。

在多项式插值中, 求一次数不超过 n 的代数多项式

$$P_n(x) = a_0 + a_1x + \dots + a_nx^n \quad (2.2.2)$$

使

$$P_n(x_i) = y_i \quad (i=0,1,\dots,n) \quad (2.2.3)$$

其中 a_0, a_1, \dots, a_n 为实数。满足插值条件 (2.2.3) 的多项式 (2.2.2), 称为函数 $f(x)$ 的 n 次插值多项式。

n 次插值多项式 $P_n(x)$ 的几何意义: 过曲线 $y=f(x)$ 上的 $n+1$ 个点 $(x_i, y_i) (i=0,1,\dots,n)$ 作一条 n 次代数曲线 $y=P_n(x)$ 作为曲线 $y=f(x)$ 的近似, 如图 2.2.1 所示。

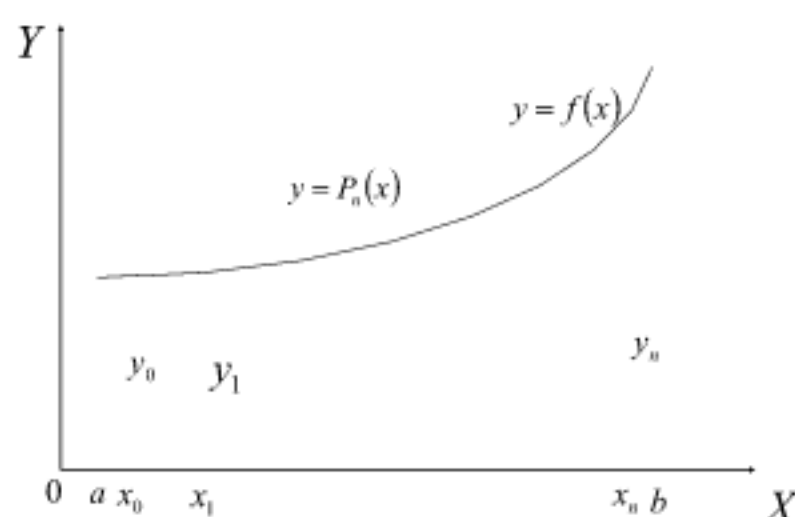


图 2.2.1 插值多项式的几何直观图

2.2.2 插值多项式的存在唯一性

由插值条件 (2.2.3) 知, $P_n(x)$ 的系数 $a_i (i=0,1,\cdots,n)$ 满足线性方程组:

[illegible]

由线性代数知, 线性方程组的系数行列式是 $n+1$ 阶范德蒙 (Vandermonde) 行列式, 且

$$V = \begin{vmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{vmatrix} = \prod_{i=1}^n \prod_{j=0}^{i-1} (x_i - x_j)$$

因 x_0, x_1, \dots, x_n 是区间 $[a, b]$ 上的不同点, 上式右端乘积中的每一个因子 $x_i - x_j \neq 0$, 于是系数行列式不等于 0, 即方程组 (2.2.4) 的解存在且唯一。从而得出插值多项式的存在唯一性定理。

定理 2.2.1 若插值节点 x_0, x_1, \cdots, x_n 互不相同, 则满足插值条件 (2.2.3) 的 n 次插值多项式 (2.2.2) 存在且唯一。

2.3 求插值多项式的 Lagrange (拉格朗日) 法

在上一节里,插值多项式存在唯一性的证明过程不仅指出了满足差值条件的 n 次插值多项式是存在唯一性的,而且也提供了插值多项式的一种求法,即通过解线性方程组 (2.2.4) 来确定其系数 a_i 。但是,当未知数个数多时,这种做法的计算工作量大,不便于实际应用。Lagrange 基于用简单插值问题的插值函数表示一般的插值函数思想,给出一种求插值函数的简便方法——Lagrange 插值法。

2.3.1 Lagrange 插值基函数

先考虑简单的插值问题：对节点 $x_i (i=0, 1, \cdots, n)$ 中任意一点 $x_k (0 \leq k \leq n)$ 做一 n 次多项式 $l_k(x)$ 使它在该点上取值为 1, 而在其余点 $x_i (i=0, 1, k-1, k+1, \cdots, n)$ 上取值为零, 即满足插值条件

$$l_k(x_i) = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases} \quad (2.3.1)$$

(2.3.1) 表明 n 个点 $x_i (i=0, 1, k-1, k+1, \cdots, n)$ 都是 n 次多项式 $l_k(x)$ 的零点, 故可

设 $l_k(x) = A_k(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)$ ，其中 A_k 为待定系数，由条件 $l_k(x_k)=1$ 可得

$$A_k = \frac{1}{(x_k - x_0)\cdots(x_k - x_{k-1})(x_k - x_{k+1})\cdots(x_k - x_n)}$$

故

$$l_k(x) = \frac{(x-x_0)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)} \quad (2.3.2)$$

对应于每一节点 x_k ($0 \leq k \leq n$) 都能求出一个满足插值条件 (2.3.1) 的 n 次插值多项式 (2.3.2)，这样，由 (2.3.2) 式可以求出 $n+1$ 个 n 次插值多项式 $l_0(x), l_1(x), \cdots, l_n(x)$ 。容易看出，这组多项式仅与节点的取法有关，称它们为在 $n+1$ 个节点上的 n 次基本插值多项式或 n 次插值基函数，即 Lagrange 插值基函数。

2.3.2 Lagrange (拉格朗日) 插值多项式

利用 Lagrange 插值基函数立即可以写出满足插值条件 (2.2.3) 的 n 次插值多项式

$$y_0 l_0(x) + y_1 l_1(x) + \cdots + y_n l_n(x) \quad (2.3.3)$$

事实上，由于每个插值基函数 $l_k(x)$ ($k=0,1,\cdots,n$) 都是 n 次多项式，故其线性组合 (2.3.3) 必是不高于 n 次的多项式，同时，根据条件 (2.2.1) 容易验证多项式 (2.3.3) 在节点 x_i 处的值 y_i ($i=0,1,\cdots,n$) 为 $P_n(x_i)$ ，因此，它就是待求的 n 次插值多项式 $P_n(x)$ 。形如 (2.3.3) 的插值多项式称为 Lagrange 插值多项式，记为

$$\begin{aligned} L_n(x) &= y_0 l_0(x) + y_1 l_1(x) + \cdots + y_n l_n(x) \\ &= \sum_{k=0}^n y_k \frac{(x-x_0)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)} \end{aligned} \quad (2.3.4)$$

令 $n=1$ ，由 (2.3.4) 即得两点插值公式

$$L_1(x) = y_0 \frac{x-x_1}{x_0-x_1} + y_1 \frac{x-x_0}{x_1-x_0} \quad (2.3.5)$$

即

$$L_1(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0} (x - x_0) \quad (2.3.6)$$

这是一个线性函数，用线性函数 $L_1(x)$ 近似代替函数 $f(x)$ ，在几何上就是通过曲

线 $y = f(x)$ 上两点 $(x_0, y_0), (x_1, y_1)$ 做一直线 $y = L_1(x)$ 近似代替曲线 $y = f(x)$ ，如见图 2.3.1 所示，故两点插值又称线性插值。

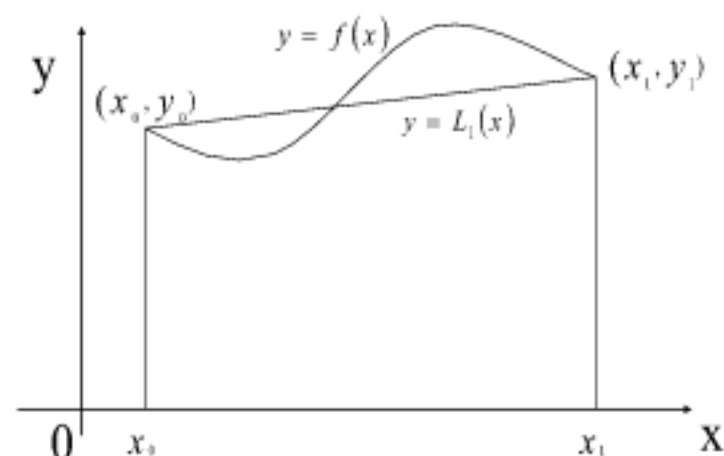


图 2.3.1 线性插值的几何直观图

令 $n = 2$ ，由 (2.3.4) 可得常用的三点插值公式：

$$L_2(x) = y_0 \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} + y_1 \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} + y_2 \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} \quad (2.3.7)$$

这是一个二次函数，用二次函数 $L_2(x)$ 近似代替函数 $f(x)$ ，在几何上就是通过曲线 $y = f(x)$ 上的三点 $(x_0, y_0), (x_1, y_1), (x_2, y_2)$ 作一抛物线 $y = L_2(x)$ ，近似地代替曲线 $y = f(x)$ ，如图 2.3.2 所示，故称为三点插值或二次插值，也称为抛物线插值。

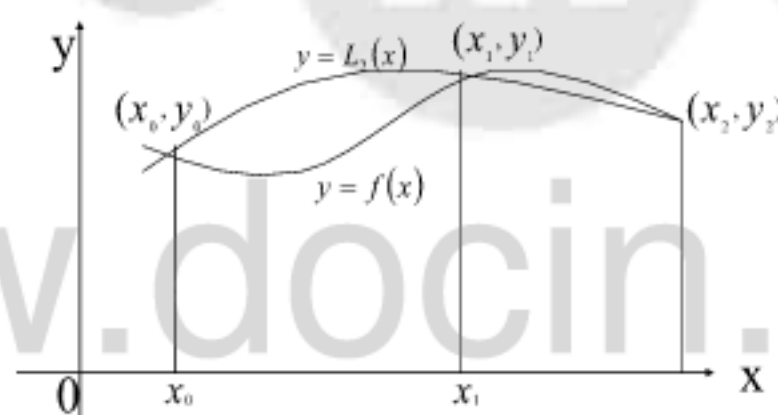


图 2.3.2 二次插值的几何直观图

例 2.3.1 已知 $\sqrt{100} = 10, \sqrt{121} = 11, \sqrt{144} = 12$ 分别用线性插值和抛物线插值求 $\sqrt{115}$ 的值。

解 因为 115 在 100 和 121 之间，故取节点 $x_0 = 100$ ， $x_1 = 121$ ，相应地有 $y_0 = 10$ ， $y_1 = 11$ ，于是，由线性插值公式 (2.3.5) 可得

$$L_1(x) = 10 \times \frac{x-121}{100-121} + 11 \times \frac{x-100}{121-100}$$

故用线性插值求得的近似值为：

$$\sqrt{115} \approx L_1(115) = 10 \times \frac{115-121}{100-121} + 11 \times \frac{115-100}{121-100} \approx 10.714$$

同理，用抛物线插值公式 (2.3.7) 所求得的近似值为：

$$\begin{aligned}\sqrt{115} \approx L_2(115) &= 10 \times \frac{(115-121)(115-144)}{(100-121)(144-121)} + 11 \times \frac{(115-100)(115-144)}{(121-100)(121-144)} \\ &\quad + 12 \times \frac{(115-100)(115-121)}{(144-100)(144-121)} \\ &\approx 10.732\end{aligned}$$

将所得结果与 $\sqrt{115}$ 的精确值10.7328...相比较,可以看出抛物线插值的精确度较好。为了便于上机计算,我们常将拉格朗日插值多项式(2.3.4)改写成公式(2.3.8)的对称形式

$$L_n(x) = \sum_{k=0}^n \left[y_k \prod_{\substack{j=0 \\ j \neq k}}^n \left(\frac{x-x_j}{x_k-x_j} \right) \right] \quad (2.3.8)$$

可用二重循环来完成 $L_n(x)$ 值的计算,先通过内循环,即先固定 k ,令 j 从0到

$n(j \neq k)$ 累乘求得 $l_k(x) = \prod_{\substack{j=0 \\ j \neq k}}^n \left(\frac{x-x_j}{x_k-x_j} \right)$,然后再通过外循环,即令 k 从0到 n ,累加得出插值结果 $L_n(x)$ 。

2.3.3 插值余项

在插值区间 $[a,b]$ 上用插值多项式 $P_n(x)$ 近似代替 $f(x)$,除了在插值节点 x_i 上没有误差以外,在其他点上一般有误差。若记

$$R_n(x) = f(x) - P_n(x)$$

则 $R_n(x)$ 就是用 $P_n(x)$ 近似代替 $f(x)$ 时所产生的截断误差,称 $R_n(x)$ 为插值多项式 $P_n(x)$ 的余项。关于插值多项式的误差有定理2.3.1的估计式。

定理 2.3.1 设 $f(x)$ 在区间 $[a,b]$ 上有直到 $n+1$ 阶导数, x_0, x_1, \dots, x_n 为区间 $[a,b]$ 上 $n+1$ 个互异的节点, $P_n(x)$ 为满足条件: $P_n(x_i) = f(x_i) (i=0,1,\dots,n)$ 的 n 次插值多项式,则对于任何 $x \in [a,b]$ 有

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega_{n+1}(x) \quad (2.3.9)$$

其中 $\omega_{n+1}(x) = \prod_{i=0}^n (x-x_i)$, $\xi \in (a,b)$ 且依赖于 x 。

证明 由插值条件 $P_n(x_i) = f(x_i)$ 知 $R_n(x_i) = 0 (i = 0, 1, \dots, n)$, 即插值节点都是 $R_n(x)$ 的零点, 故可设

$$R_n(x) = K(x)\omega_{n+1}(x) \quad (2.3.10)$$

其中 $K(x)$ 为待定函数。下面求 $K(x)$, 对区间 $[a, b]$ 上异于 x_i 的任意一点 $x \neq x_i$ 作辅助函数:

$$F(t) = f(t) - P_n(t) - K(x)\omega_{n+1}(t)$$

不难看出 $F(t)$ 具有如下特点

$$(1) F(x_i) = F(x_i) = 0 (i = 0, 1, \dots, n) \quad (2.3.11)$$

(2) 在 $[a, b]$ 上有直到 $n+1$ 阶导数, 且

$$F^{(n+1)}(t) = f^{(n+1)}(t) - K(x)(n+1)! \quad (2.3.12)$$

等式 (2.3.11) 表明 $F(t)$ 在 $[a, b]$ 上至少有 $n+2$ 个互异的零点, 根据罗尔 (Rolle) 定理, 在 $F(t)$ 的两个零点之间, $F'(t)$ 至少有一个零点, 因此, $F'(t)$ 在 (a, b) 内至少有 $n+1$ 个互异的零点, 对 $F'(t)$ 再应用罗尔定理, 推得 $F''(t)$ 在 (a, b) 内至少有 n 个互异的零点。继续上述讨论, 可推得 $F^{(n+1)}(t)$ 在 (a, b) 内至少有一个零点, 若记为 ξ , 则 $F^{(n+1)}(\xi) = 0$, 于是由 (2.3.12) 式得

$$f^{(n+1)}(\xi) - K(x)(n+1)! = 0$$

$$K(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

将它代入 (2.3.10) 即得 (2.3.9)。对于 $x = x_i$, (2.3.9) 显然成立。

例 2.3.2 在例 2.3.1 中分别用线性插值和抛物插值计算了 $\sqrt{115}$ 近似值, 试估计它们的截断误差。

解 用线性插值求 $f(x) = \sqrt{x}$ 的近似值, 其截断误差由插值余项公式 (2.3.9) 知

$$\begin{aligned} R_1(x) &= \frac{1}{2} f''(\xi) \omega_2(x) \\ &= -\frac{1}{8} \xi^{-3/2} (x - x_0)(x - x_1) \quad \xi \in [x_0, x_1] \end{aligned}$$

现在 $x_0=100$, $x_1=121$, $x=115$, 故

$$\begin{aligned} |R_1(115)| &\leq \frac{1}{8} \times |(115-100)(115-121)| \max_{\xi \in [100,121]} \xi^{-3/2} \\ &= \frac{1}{8} \times 15 \times 6 \times 10^{-3} = 0.01125 \end{aligned}$$

当用抛物插值求 $f(x)=\sqrt{x}$ 的近似值时, 其截断误差为

$$\begin{aligned} R_2(x) &= \frac{1}{3!} f'''(\xi) \omega_3(x) \\ &= \frac{1}{16} \xi^{-5/2} (x-x_0)(x-x_1)(x-x_2), \xi \in [x_0, x_2] \end{aligned}$$

将 $x_0=100, x_1=121, x_2=144, x=115$ 代入, 即得

$$|R_2(115)| \leq \frac{1}{16} |(115-100)(115-121)(115-144)| \times 10^{-5} < 0.0017$$

2.3.4 插值误差的事后估计法

在许多情况下, 要直接应用余项公式 (2.3.9) 来估计误差是很困难的, 下面将以线性插值为例, 介绍另一种估计误差的方法。

设 $x_0 < x < x_1 < x_2$ 且 $f(x_i)$ ($i=0,1,2$) 为已知, 若将用 x_0, x_1 两点做线性插值求得 $y=f(x)$ 的近似值为 y_1 , 用 x_0, x_2 两点作线性插值所求得 $y=f(x)$ 的近似值记为 y_2 , 则由余项公式 (2.3.9) 知:

$$\begin{aligned} y-y_1 &= \frac{1}{2} f''(\xi_1)(x-x_0)(x-x_1), \xi_1 \in [x_0, x_1] \\ y-y_2 &= \frac{1}{2} f''(\xi_2)(x-x_0)(x-x_2), \xi_2 \in [x_0, x_2] \end{aligned}$$

假设 $f''(x)$ 在区间 $[x_0, x_2]$ 中变化不大, 将上面两式相除, 即得近似式

$$\frac{y-y_1}{y-y_2} \approx \frac{x-x_1}{x-x_2}$$

即

$$y-y_1 \approx \frac{x-x_1}{x_2-x_1} (y_2-y_1) \quad (2.3.13)$$

近似式 (2.3.13) 表明, 可以通过两个结果的偏差 y_2-y_1 来估计插值误差, 这种

直接利用计算结果来估计误差的方法，称为事后估计法。

在例 2.3.1 中，用 $x_0=100, x_1=121$ 做节点，算得的 $\sqrt{115}$ 近似值为 $y_1=10.714$ ，

同样，用 $x_0=100, x_2=144$ 做节点，可算得 $\sqrt{115}$ 的另一近似值 $y_2=10.682$ 。

通过 (2.3.13) 可以估计出插值 y_1 的误差为：

$$\sqrt{115} - y_1 \approx \frac{115-121}{144-121}(10.682-10.714) = 0.00835$$

2.4 求插值多项式的 Newton 法

拉格朗日插值法通过求出拉格朗日基函数而方便地表示插值多项式，但是当增加节点时，所有的基函数都需重新计算，因此其计算过程不具继承性。由线性代数可知，任何一个不高于 n 次的多项式，都可表示成函数

$$1, x-x_0, (x-x_0)(x-x_1), \dots, (x-x_0)(x-x_1)\cdots(x-x_{n-1})$$

的线性组合，即可将满足插值条件 $P(x_i)=y_i (i=0,1,\dots,n)$ 的 n 次多项式写成形式

$$a_0 + a_1(x-x_0) + a_2(x-x_0)(x-x_1) + \cdots + a_n(x-x_0)(x-x_1)\cdots(x-x_{n-1})$$

其中 $a_k (k=0,1,\dots,n)$ 为待定系数。这种形式的插值多项式称为牛顿 (Newton) 插值多项式，把它记成 $N_n(x)$ ，即

$$N_n(x) = a_0 + a_1(x-x_0) + a_2(x-x_0)(x-x_1) + \cdots + a_n(x-x_0)(x-x_1)\cdots(x-x_{n-1}) \quad (2.4.1)$$

因此，牛顿插值多项式 $N_n(x)$ 是插值多项式 $P_n(x)$ 的另一种表示形式。在牛顿插值多项式中要用到的差分与差商等概念，与数值计算的其它方面有着密切的关系。

2.4.1 向前差分与 Newton 向前插值公式

设函数 $f(x)$ 在等距节点 $x_k = x_0 + kh (k=0,1,\dots,n)$ 处的函数值 $f(x_k) = y_k$ 为已知，其中 h 是正常数，称为步长，称两个相邻点 x_k 和 x_{k+1} 处函数值之差 $y_{k+1} - y_k$ 为函数 $f(x)$ 在点 x_k 处以 h 为步长的一阶向前差分（简称一阶差分），记 Δy_k ，即

$$\Delta y_k = y_{k+1} - y_k$$

于是，函数 $f(x)$ 在各节点处的一阶差分依次为

$$\Delta y_0 = y_1 - y_0, \Delta y_1 = y_2 - y_1, \dots, \Delta y_{n-1} = y_n - y_{n-1}$$

又称一阶差分的差分 $\Delta^2 y_k = \Delta(\Delta y_k) = \Delta y_{k+1} - \Delta y_k$ 为二阶差分。

一般地，定义函数 $f(x)$ 在点 x_k 处的 m 阶差分为：

$$\Delta^m y_k = \Delta^{m-1} y_{k+1} - \Delta^{m-1} y_k$$

为了便于计算与应用，通常采用表格形式计算差分，如表 2.4.1 所示。

表 2.4.1 差分计算表

x_k	y_k	Δy_k	$\Delta^2 y_k$	$\Delta^3 y_k$	$\Delta^4 y_k$
x_0	y_0	Δy_0	$\Delta^2 y_0$	$\Delta^3 y_0$	$\Delta^4 y_0$
x_1	y_1	Δy_1	$\Delta^2 y_1$	$\Delta^3 y_1$	
x_2	y_2	Δy_2	$\Delta^2 y_2$		
x_3	y_3	Δy_3			
x_4	y_4				

在等距节点 $x_k = x_0 + kh (k = 0, 1, \dots, n)$ 情况下，可以利用差分表示牛顿插值多项式 (2.4.1) 的系数，并将所得公式加以简化。

事实上，由插值条件 $N_n(x_0) = y_0$ 即可得 $a_0 = y_0$ 。

再由插值条件 $N_n(x_1) = y_1$ 可得：

$$a_1 = \frac{y_1 - y_0}{x_1 - x_0} = \frac{\Delta y_0}{h}$$

由插值条件 $N_n(x_2) = y_2$ 可得：

$$a_2 = \frac{y_2 - y_0 - \frac{\Delta y_0(x_2 - x_0)}{h}}{(x_2 - x_0)(x_2 - x_1)} = \frac{y_2 - 2y_1 + y_0}{2hh} = \frac{\Delta^2 y_0}{2!h^2}$$

一般地，由插值条件 $N_n(x_k) = y_k$ 可得

$$a_k = \frac{\Delta^k y_0}{k!h^k} (k = 1, 2, \dots, n)$$

于是，满足插值条件 $N_n(x_i) = y_i$ 的插值多项式为：

$$N_n(x) = y_0 + \frac{\Delta y_0}{h}(x-x_0) + \frac{\Delta^2 y_0}{2!h^2}(x-x_0)(x-x_1) + \cdots + \frac{\Delta^n y_0}{n!h^n}(x-x_0)(x-x_1)\cdots(x-x_{n-1})$$

令 $x = x_0 + th (t > 0)$ ，并注意到 $x_k = x_0 + kh$ 则牛顿插值多项式可简化为

$$N_n(x_0 + th) = y_0 + t\Delta y_0 + \frac{t(t-1)}{2!}\Delta^2 y_0 + \cdots + \frac{t(t-1)\cdots(t-n+1)}{n!}\Delta^n y_0 \quad (2.4.2)$$

这个用向前差分表示的插值多项式，称为牛顿向前插值公式，简称前插公式。它适用于计算 x_0 附近的函数值。

由插值余项公式 (2.3.9)，可得前插公式的余项为：

$$R_n(x_0 + th) = \frac{t(t-1)\cdots(t-n)}{(n+1)!}h^{n+1}f^{n+1}(\xi), \xi \in (x_0, x_n) \quad (2.4.3)$$

例 2.4.1 从给定的正弦函数表出发计算 $\sin(0.12)$ ，并估计截断误差。

表 2.4.2 $\sin x$ 的函数及差分表

x	$\sin x$	Δy	$\Delta^2 y$	$\Delta^3 y$
0.1	0.09983			
0.2	0.19867	0.09884	-0.00199	-0.00094
0.3	0.29552	0.09685	-0.00295	-0.00094
0.4	0.38942	0.09390	-0.00389	-0.00091
0.5	0.47943	0.09001	-0.00480	
0.6	0.56464	0.08521		

解 因为 0.12 介于 0.1 与 0.2 之间，故取 $x_0 = 0.1$ ，此时

$$t = \frac{x - x_0}{h} = \frac{0.12 - 0.1}{0.1} = 0.2,$$

为求 $\Delta y_0, \Delta^2 y_0, \Delta^3 y_0 \cdots$ ，构造差分表 2.4.2。

若用线性插值求 $\sin(0.12)$ 的近似值，则由前插公式 (2.4.2) 立即可得

$$\sin(0.12) \approx N_1(0.12) = 0.09983 + 0.2 \times 0.09884 = 0.11960$$

用二次插值得：

$$\begin{aligned} \sin(0.12) &\approx N_2(0.12) \\ &= 0.09983 + 0.2 \times 0.09884 + \frac{0.2 \times (0.2 - 1)}{2} \times (-0.00199) \\ &= N_1(0.12) + 0.00016 \\ &= 0.11976 \end{aligned}$$

用三次插值得：

$$\begin{aligned}
\sin(0.12) &\approx N_3(0.12) \\
&= N_2(0.12) + \frac{0.2 \times (0.2-1) \times (0.2-2)}{6} \times (-0.00096) \\
&= 0.11971
\end{aligned}$$

因 $N_3(0.12)$ 与 $N_2(0.12)$ 很接近, 且由差分表 2.4.2 可以看出, 三阶差分接近于常数 (即 $\Delta^4 y_0$ 接近于零), 故取 $N_3(0.12) = 0.11971$ 作为 $\sin(0.12)$ 的近似值, 此时由余项公式 (2.4.3) 可知其截断误差为:

$$R_3(0.12) \leq \left| \frac{0.2 \times (0.2-1) \times (0.2-2) \times (0.2-3)}{24} \right| \times (0.1)^4 \times \sin(0.4) < 0.000002$$

2.4.2 向后差分与 Newton (牛顿) 向后插值公式

在等距节点 $x_k = x_0 + kh (k=0, 1, \dots, n)$ 下, 除了向前差分外, 还可引入向后差分和中心差分, 其定义和记号分别如下:

$y = f(x)$ 在点 x_k 处以 h 为步长的一阶向后差分 and m 阶向后差分分别为:

$$\begin{aligned}
\nabla y_k &= y_k - y_{k-1} \\
\nabla^m y_k &= \nabla^{m-1} y_k - \nabla^{m-1} y_{k-1} \quad (m=2, 3, \dots)
\end{aligned}$$

$y = f(x)$ 在 x_k 点处以 h 为步长的一阶中心差分 and m 阶中心差分分别为:

$$\begin{aligned}
\delta y_k &= y_{k+\frac{1}{2}} - y_{k-\frac{1}{2}} \\
\delta^m y_k &= \delta^{m-1} y_{k+\frac{1}{2}} - \delta^{m-1} y_{k-\frac{1}{2}} \quad (m=2, 3, \dots)
\end{aligned}$$

其中 $y_{k+\frac{1}{2}} = f\left(x_k + \frac{h}{2}\right)$, $y_{k-\frac{1}{2}} = f\left(x_k - \frac{h}{2}\right)$, 各阶向后差分与中心差分的计算, 可通过构造向后差分表与中心差分表来完成。

利用向后差分, 可简化牛顿插值多项式 (2.4.1), 导出与牛顿前插公式 (2.4.2) 类似的公式。

事实上, 若将节点的排列次序看作 x_n, x_{n-1}, \dots, x_0 , 那么 (2.4.1) 可写成:

$$N_n(x) = b_0 + b_1(x - x_n) + b_2(x - x_n)(x - x_{n-1}) + \dots + b_n(x - x_n)(x - x_{n-1}) \cdots (x - x_1)$$

根据插值条件 $N_n(x_i) = y_i (i = n, n-1, \dots, 1, 0)$ 得到用向后差分表示的插值多项式:

$$N_n(x_n + th) = y_n + t \nabla y_n + \frac{t(t+1)}{2!} \nabla^2 y_n + \dots + \frac{t(t+1) \cdots (t+n-1)}{n!} \nabla^n y_n \quad (2.4.4)$$

其中 $t < 0$, 插值多项式 (2.4.4) 称为牛顿向后插值公式, 简称后插公式。它适

用于计算 x_n 附近的函数值。由插值余项公式 (2.3.9)，可写出后插公式的余项为：

$$R_n(x_n + th) = \frac{t(t+1)\cdots(t+n)}{(n+1)!} h^{n+1} f^{(n+1)}(\xi), (\xi \in (x_0, x_n)) \quad (2.4.5)$$

例 2.4.2 已知函数表如表 2.4.2，计算 $\sin(0.58)$ ，并估算其截断误差。

解 因为 0.58 位于表尾 $x_5 = 0.6$ 附近，故用后插公式 (2.4.4) 计算 $\sin(0.58)$ 的近似值。

一般的，为了计算函数在 x_5 处的各阶向后差分，应构造向后差分表。但由向前差分与向后差分的定义可以看出，对同一函数表来说，构造出来的向后差分表与向前差分表在数据上完全相同。因此，表 2.4.2 中用下划线标出的各数依次给出了 $\sin x$ 在 $x_5 = 0.6$ 处的函数值和向后差分值。因三阶向后差分接近于常数，故

用三次插值进行计算，且 $t = \frac{x - x_5}{h} = \frac{0.58 - 0.6}{0.1} = -0.2$ ，

于是由牛顿向后插值公式 (2.4.4) 得：

$$\begin{aligned} \sin(0.58) &\approx N_3(0.58) = 0.56464 + (-0.2) \times 0.08521 + \frac{(-0.2) \times (-0.2 + 1)}{2} \times (-0.00480) \\ &\quad + \frac{(-0.2) \times (-0.2 + 1) \times (-0.2 + 2)}{6} \times (-0.00091) \\ &= 0.54802 \end{aligned}$$

因为在整个计算中，只用到四个点 $x = 0.6, 0.5, 0.4, 0.3$ 上的函数值，固由余项公式 (2.4.5) 知其截断误差为：

$$|R_3(0.58)| \leq \left| \frac{-0.2 \times (-0.2 + 1) \times (-0.2 + 2) \times (-0.2 + 3)}{24} \right| \times (0.1)^4 \times \sin(0.6) < 0.000002$$

2.4.3 差商与牛顿基本插值多项式

当插值节点非等距分布时，就不能用差分来简化牛顿插值多项式，此时可用差商这个新概念来解决。

设函数 $f(x)$ 在一串互异的点 $x_{i_0}, x_{i_1}, x_{i_2}, \dots$ 上的值依次为

$$f(x_{i_0}), f(x_{i_1}), f(x_{i_2}), \dots$$

称函数值之差 $f(x_{i_1}) - f(x_{i_0})$ 与自变量之差 $x_{i_1} - x_{i_0}$ 的比值 $\frac{f(x_{i_1}) - f(x_{i_0})}{x_{i_1} - x_{i_0}}$ 为

函数 $f(x)$ 关于 x_{i_1}, x_{i_0} 点的一阶差商，记作 $f[x_{i_0}, x_{i_1}]$

例 2.4.3 $f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}, f[x_1, x_2] = \frac{f(x_2) - f(x_1)}{x_2 - x_1}, \dots$ 。

称一阶差商的差 $\frac{f[x_i, x_{i_2}] - f[x_{i_0}, x_{i_1}]}{x_{i_2} - x_{i_0}}$ 为函数 $f(x)$ 关于点 $x_{i_0}, x_{i_1}, x_{i_2}$ 的二阶差商，记作 $f[x_{i_0}, x_{i_1}, x_{i_2}]$ 。

例 2.4.4 $f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}$ 。

一般地，可通过函数 $f(x)$ 的 $m-1$ 阶差商定义的 m 阶差商如下：

$$f[x_{i_0}, x_{i_1}, \dots, x_{i_m}] = \frac{f[x_{i_1}, \dots, x_{i_m}] - f[x_{i_0}, \dots, x_{i_{m-1}}]}{x_{i_m} - x_{i_0}}$$

差商计算也可采用表格形式（称为差商表）进行，如表 2.4.3 所示。

表 2.4.3 差商计算表

x_k	$f(x_k)$	一阶差商	二阶差商	三阶差商
x_0	$f(x_0)$			
x_1	$f(x_1)$	$f[x_0, x_1]$		
x_2	$f(x_2)$	$f[x_1, x_2]$	$f[x_0, x_1, x_2]$	
x_3	$f(x_3)$	$f[x_2, x_3]$	$f[x_1, x_2, x_3]$	$f[x_0, x_1, x_2, x_3]$

差商具有下列重要性质（证明略）：

(1) 函数 $f(x)$ 的 m 阶差商 $f[x_0, x_1, \dots, x_m]$ 可由函数值 $f(x_0), f(x_1), \dots, f(x_m)$ 的线性组合表示，且

$$f[x_0, x_1, \dots, x_m] = \sum_{i=0}^m \frac{f(x_i)}{(x_i - x_0) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_m)}$$

(2) 差商具有对称性，即任意调换节点的次序，不影响差商的值。例如

$$f[x_0, x_1, x_2] = f[x_1, x_0, x_2] = f[x_1, x_2, x_0] = \dots$$

(3) 当 $f^{(m)}(x)$ 在包含节点 $x_j (j=0, 1, \dots, m)$ 的某个区间上存在时，在 $x_{i_0}, x_{i_1}, \dots, x_{i_m}$

之间必有一点 ξ ，使

$$f[x_0, x_1, \dots, x_m] = \frac{f^{(m)}(\xi)}{m!}$$

(4) 在等距节点 $x_k = x_0 + kh (k=0, 1, \dots, n)$ 情况下，可同时引入 $m (m \leq n)$ 阶差分与差商，且有下面关系：

$$f[x_0, x_1, \dots, x_m] = \frac{\Delta^m y_0}{m! \cdot h^m}$$

$$f[x_n, x_{n-1}, \dots, x_{n-m}] = \frac{\nabla^m y_n}{m! \cdot h^m}$$

引入差商的概念后，可利用差商表示牛顿插值多项式 (2.4.1) 的系数。事实上，从插值条件出发，可以像确定前插公式中的系数那样，逐步地确定 (2.4.1) 中的系数

$$\begin{cases} a_0 = f(x_0) \\ a_k = f[x_0, x_1, \dots, x_k] \quad (k=1, 2, \dots, n) \end{cases}$$

故满足插值条件 $N_n(x_i) = y_i (i=0, 1, \dots, n)$ 的 n 次插值多项式为：

$$N_n(x) = f(x_0) + f[x_0, x_1](x-x_0) + f[x_0, x_1, x_2](x-x_0)(x-x_1) + \dots + f[x_0, x_1, \dots, x_n](x-x_0)(x-x_1)\dots(x-x_{n-1}) \quad (2.4.6)$$

(2.4.6) 称为牛顿基本插值多项式，常用来计算非等距节点上的函数值。

例 2.4.5 试用牛顿基本差值多项式按例 1 要求重新计算 $\sqrt{115}$ 的近似值。

解 先构造差商表

x	\sqrt{x}	一阶差商	二阶差商
100	10	0.047619	
121	11		-0.000094
144	12	0.043478	

由上表可以看出牛顿基本插值多项式 (2.4.6) 中各系数依次为：

$$f(x_0) = 10$$

$$f[x_0, x_1] = 0.047619$$

$$f[x_0, x_1, x_2] = -0.000094$$

故用线性插值所得的近似值为：

$$\sqrt{115} \approx N_1(115) = 10 + 0.047619 \times (115 - 100) = 10.7143$$

用抛物插值所求得的近似值为：

$$\sqrt{115} \approx N_2(115) = N_1(115) + (-0.000094) \times (115 - 100) \times (115 - 121) = 10.7228$$

所得结果与例 2.4.1 一致。比较例 2.4.1 和例 2.4.5 的计算过程可以看出，与拉格朗日插值多项式相比较，牛顿插值多项式在增加节点时不需要重新计算前面的各项，只需在后面增加相应的项，即牛顿插值法具有很好的继承性。

由差值多项式的存在唯一性定理知，满足同一组插值条件的拉格朗日多项式 (2.3.4) 与牛顿基本插值多项式 (2.4.6) 是同一多项式。因此，余项公式 (2.3.9) 也适用于牛顿插值。但是在实际计算中，有时也用差商表示的余项公式：

$$R_n(x) = f[x_0, x_1, \dots, x_n, x] \omega_{n+1}(x) \quad (2.4.7)$$

来估计截断误差（证明略）

注意 上式中的 $n+1$ 阶差商 $f[x_0, x_1, \dots, x_n, x]$ 与 $f(x)$ 的值有关，故不能准确地计算出 $f[x_0, x_1, \dots, x_n, x]$ 的精确值，只能对它做估计。

例 2.4.6 当四阶差商变化不大时，可用 $f[x_0, x_1, x_2, x_3, x_4]$ 近似代替 $f[x, x, x, x, x]$ 。

2.5 求插值多项式的改进算法

2.5.1 分段低次插值

例 2.3.2、例 2.4.1 表明适当地提高插值多项式的次数，有可能提高插值的精度。但是决不可由此得出结论，认为插值多项式的次数越高越好。

例 2.5.1 对函数

$$f(x) = \frac{1}{1+25x^2} \quad (-1 \leq x \leq 1)$$

先以 $x_i = -1 + \frac{2}{5}i (i=0,1,\dots,5)$ 为节点作五次插值多项式 $P_5(x)$ ，再以 $x_i = -1 + \frac{1}{5}i (i=0,1,\dots,10)$ 为节点作十次插值多项式 $P_{10}(x)$ ，并将曲线 $f(x) = \frac{1}{1+25x^2}$ ， $y = P_5(x)$ ， $y = P_{10}(x)$ 描绘在同一坐标系中，如图 2.5.1

所示。可以看出，虽然在局部范围中，例如在区间 $[-0.2, 0.2]$ 中， $P_{10}(x)$ 比 $P_5(x)$ 较好地逼近 $f(x)$ ，但从整体上看， $P_{10}(x)$ 并非处处都比 $P_5(x)$ 较好地逼近，尤其是在区间 $[-1, 1]$ 的端点附近。进一步的分析表明，当 n 增大时，该函数在等距节点下的高次插值多项式，在 ± 1 两端会发生激烈的振荡。这种现象称为龙格(Runge)

现象。这表明，在大范围内使用高次插值，逼近的效果可能不理想。另一方面，插值误差除来自截断误差外，还来自初始数据的误差和计算过程中的舍入误差。插值次数越高，计算工作越大，积累误差也可能越大。

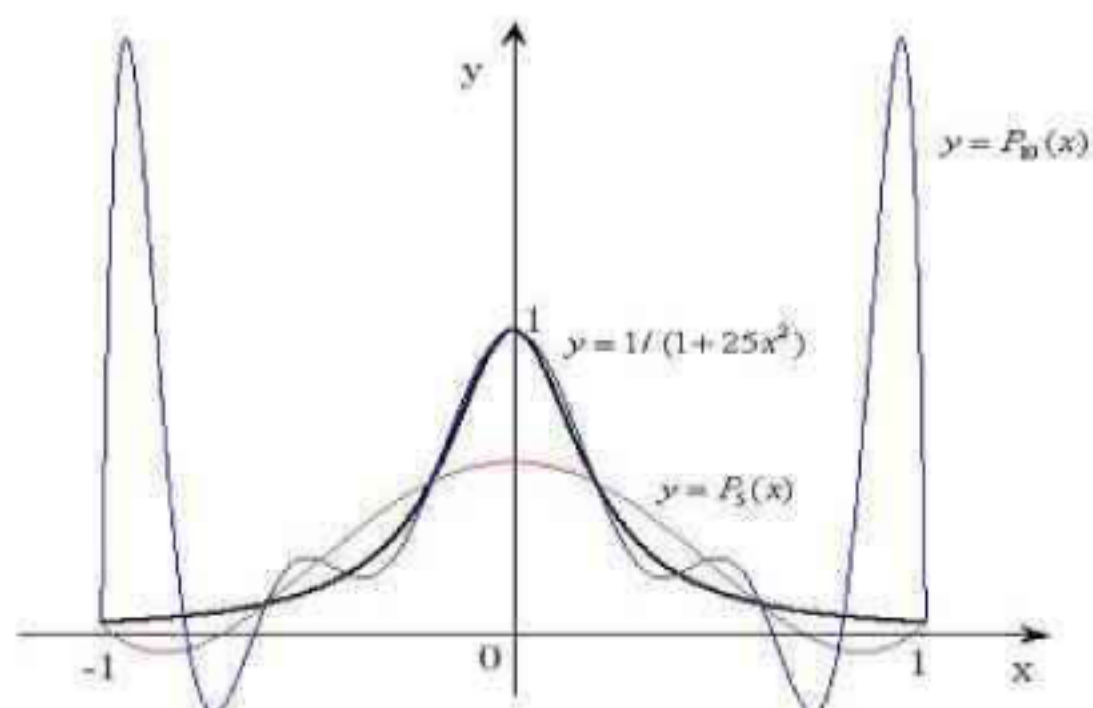


图 2.5.1 多次插值的比较图

因此，在实际计算中，常用分段低次插值进行计算，即把整个插值区间分成若干小区间，在每个小区间上进行低次插值。

例 2.5.2 当给定 $n+1$ 个点 $x_0 < x_1 < \dots < x_n$ 上的函数值 $y_0 < y_1 < \dots < y_n$ 后，若要计算点 $x \in [x_{i-1}, x_i]$ 处函数 $f(x)$ 的近似值，可先选取两个节点 x_{i-1}, x_i 使 $x \in [x_{i-1}, x_i]$ ，然后在小区间 $[x_{i-1}, x_i]$ 上作线性插值，即得

$$f(x) \approx P_1(x) = y_{i-1} \frac{x - x_i}{x_{i-1} - x_i} + y_i \frac{x - x_{i-1}}{x_i - x_{i-1}} \quad (2.5.1)$$

这种分段低次插值叫分段线性插值。在几何上就是用折线代替曲线，如图 2.5.2 所示。故分段线性插值又称折线插值。

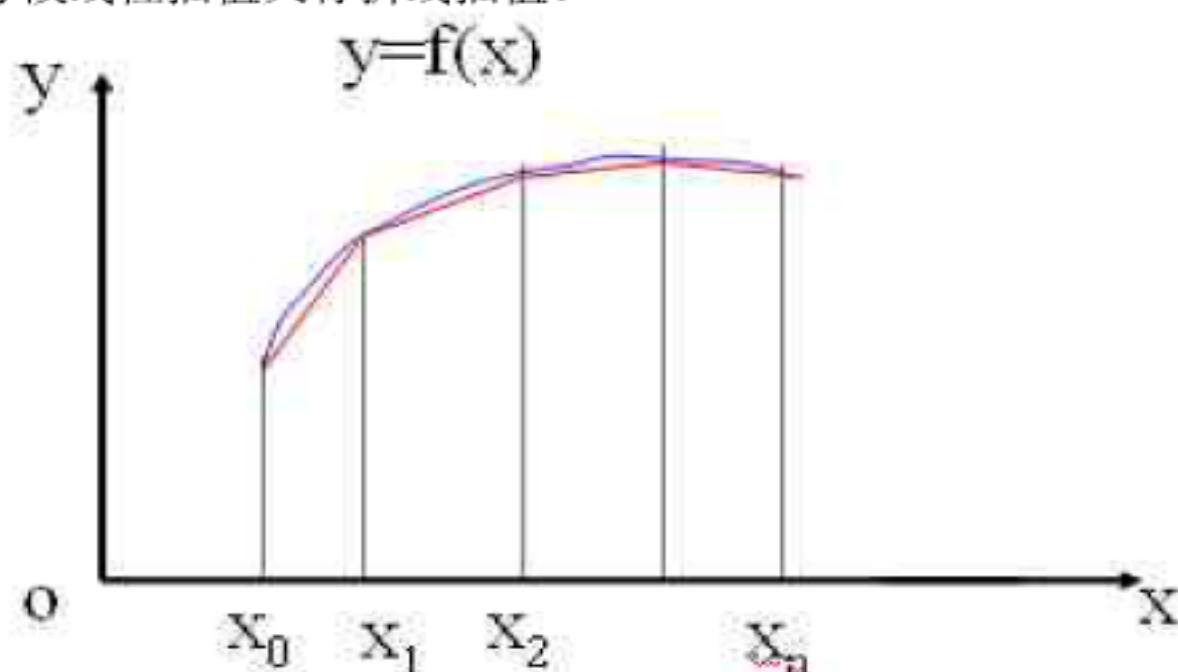


图 2.5.2 折线插值几何直观图

类似地，为求 $y = f(x)$ 的近似值，也可选取距点 x 最近的三个节点 x_{i-1}, x_i, x_{i+1} 进行二次插值，即取

$$f(x) \approx P_2(x) = \sum_{k=i-1}^{i+1} [y_k \prod_{\substack{j=i-1 \\ j \neq k}}^{i+1} (\frac{x-x_j}{x_k-x_j})] \quad (2.5.2)$$

这种分段低次插值叫分段二次插值。在几何上就是用分段抛物线代替曲线，故分段二次插值又称分段抛物插值。为了保证 x_{i-1}, x_i, x_{i+1} 是距点 x 最近的三个节点，

(2.5.2) 中的 i 可通过下面方法确定：

$$i = \begin{cases} 1 & x_0 \leq x \leq \frac{1}{2}(x_1 + x_2) \\ j & \frac{1}{2}(x_{j-1} + x_j) \leq x \leq \frac{1}{2}(x_j + x_{j+1}) \\ n-1 & \frac{1}{2}(x_{n-2} + x_{n-1}) \leq x \leq x_n \end{cases}$$

2.5.2 三次样条插值

分段低次插值虽然具有计算简单、稳定性好、收敛性有保证且易在计算机上实现等优点，但它只能保证各小段曲线在连接点上的连续性，却不能保证整条曲线的光滑性，这就不能满足某些工程技术上的要求。

自六十年代开始，由于航空、造船等工程设计的需要而发展起来的所谓样条 (Spline) 的插值方法，既保留了分段低次插值多项式的各种优点，又提高了插值函数的光滑性。今天，样条插值方法已成为数值逼近的一个极其重要的分支，在许多领域里得到越来越广泛的应用。本节介绍应用最广泛且具有二阶连续导数的三次样条插值函数。

三次样条插值函数的定义

对于给定的函数表

x	x_0	x_1	\cdots	x_n
$f(x)$	y_0	y_1	\cdots	y_n

其中 $a = x_0 < x_1 < \cdots < x_n = b$ ，若函数 $S(x)$ 满足：

- (1) 在每个子区间 $[x_{i-1}, x_i] (i=1, 2, \cdots, n)$ 上是不高于三次的多项式；
- (2) $S(x), S'(x), S''(x)$ 在 $[a, b]$ 上连续；
- (3) 满足插值条件 $S(x_i) = y_i (i=0, 1, \cdots, n)$ ；

则称 $S(x)$ 为函数 $f(x)$ 关于节点 x_0, x_1, \cdots, x_n 的三次样条插值函数。

边界条件问题的提出与类型

单靠一张函数表是不能完全确定一个三次样条插值函数的。事实上，由条件

(1) 知, 三次样条插值函数 $S(x)$ 是一个分段三次多项式, 若用 $S_i(x)$ 表示它在第 i 个子区间 $[x_{i-1}, x_i]$ 上的表达式, 则 $S_i(x)$ 形如:

$S_i(x) = a_{i0} + a_{i1}x + a_{i2}x^2 + a_{i3}x^3, x \in [x_{i-1}, x_i]$ 这里有四个待定系数 $a_{ij} (j=0,1,2,3)$ 。子区间共有 n 个, 确定 $S(x)$ 需要确定 $4n$ 个待定系数。

另一方面, 要求分段三次多项式 $S(x)$ 及其导数 $S'(x), S''(x)$ 在整个插值区间 $[a, b]$ 上连续, 只要在各子区间的端点 $x_i (i=1, 2, \dots, n-1)$ 连续即可。故由条件 (2),

(3) 可得待定系数应满足的 $4n-2$ 个方程为:

$$\begin{cases} S(x_i-0) = S(x_i+0) & (i=1, 2, \dots, n-1) \\ S'(x_i-0) = S'(x_i+0) & (i=1, 2, \dots, n-1) \\ S''(x_i-0) = S''(x_i+0) & (i=1, 2, \dots, n-1) \\ S(x_i) = y_i & (i=0, 1, \dots, n) \end{cases} \quad (2.5.3)$$

由此可以看出, 要确定个待定系数还缺少两个条件, 这两个条件通常在插值区间 $[a, b]$ 的边界点 a, b 处给出, 称为边界条件。边界条件的类型很多, 常见的有:

(1) 给定一阶导数值 $S'(x_0) = y'_0, S'(x_n) = y'_n$, 称为第一种边界条件;

(2) 给定二阶导数值 $S''(x_0) = y''_0, S''(x_n) = y''_n$, 称为第二种边界条件, 特别地, $S''(x_0) = S''(x_n) = 0$ 称为自然边界条件, 满足自然边界条件的三次样条插值函数称为自然样条插值函数。

(3) 当 $f(x)$ 是周期为 $b-a$ 的函数时, 则要求 $S(x)$ 及其导数都是以 $b-a$ 为周期的函数, 相应的边界条件称为第三种边界条件或周期边界条件:

$$\begin{aligned} S'(x_0+0) &= S'(x_n-0), \\ S''(x_0+0) &= S''(x_n-0). \end{aligned}$$

三次样条插值函数的求法

虽然可以利用方程组 (2.5.3) 和边界条件求出所有待定系数 a_{ij} , 从而得到三次样条插值函数 $S(x)$ 在各个子区间 $[x_{i-1}, x_i]$ 的表达式 $S_i(x)$ 。但是, 这种做法的计算工作量大, 不便于实际应用。下面介绍一种简便的方法。

设在节点 x_i 处 $S(x)$ 的二阶导数为

$$S''(x_i) = M_i (i=0, 1, \dots, n)$$

因为在子区间 $[x_{i-1}, x_i]$ 上 $S(x) = S_i(x)$ 是不高于三次的多项式，其二阶导数必是线性函数（或常数）。于是，有

$$S_i'(x) = M_{i-1} \frac{x - x_i}{x_{i-1} - x_i} + M_i \frac{x - x_{i-1}}{x_i - x_{i-1}}, x \in [x_{i-1}, x_i]$$

记 $h_i = x_i - x_{i-1}$ 则有

$$S_i'(x) = M_{i-1} \frac{x_i - x}{h_i} + M_i \frac{x - x_{i-1}}{h_i}$$

连续积分两次得：

$$S_i(x) = M_{i-1} \frac{(x_i - x)^3}{6h_i} + M_i \frac{(x - x_{i-1})^3}{6h_i} + A_i(x - x_{i-1}) + B_i \quad (2.5.3)$$

其中 A_i, B_i 为积分常数。利用插值条件

$$S_i(x_{i-1}) = y_{i-1}, S_i(x_i) = y_i$$

易得

$$A_i = \frac{y_i - y_{i-1}}{h_i} - \frac{1}{6}(M_i - M_{i-1}), \quad B_i = y_{i-1} - \frac{1}{6}M_{i-1}h_i^2$$

将它们代入 (2.5.3)，整理得

$$S_i(x) = M_{i-1} \frac{(x_i - x)^3}{6h_i} + M_i \frac{(x - x_{i-1})^3}{6h_i} + (y_{i-1} - \frac{M_{i-1}}{6}h_i^2) \frac{x_i - x}{h_i} + (y_i - \frac{M_i}{6}h_i^2) \frac{x - x_{i-1}}{h_i}$$

$$x \in [x_{i-1}, x_i], i = 1, 2, \dots, n \quad (2.5.4)$$

综合以上讨论可知，只要确定 $M_i (i = 0, 1, \dots, n)$ ，这 $n+1$ 个值，就可定出三次样条插值函数。

为了求出 $M_i (i = 0, 1, \dots, n)$ ，利用一阶导函数在子区间连接点上连续的条件即

$$\begin{aligned} S_i'(x_i - 0) &= S_{i+1}'(x_i + 0) \\ S_i'(x_i - 0) &= S_{i+1}'(x_i + 0) \end{aligned} \quad (2.5.5)$$

得

$$S_i'(x) = -M_{i-1} \frac{(x_i - x)^2}{2h_i} + M_i \frac{(x - x_{i-1})^2}{2h_i} \quad (2.5.6)$$

故

$$S_i'(x_i - 0) = \frac{y_i - y_{i-1}}{h_i} + \frac{h_i}{6}M_{i-1} + \frac{h_i}{3}M_i \quad (2.5.7)$$

将 (2.5.6) 中的 i 改为 $i+1$, 即得 $S'(x)$ 在子区间 $[x_i, x_{i+1}]$ 上的表达式 $S'_{i+1}(x)$, 并由此得

$$S'_{i+1}(x_i+0) = \frac{y_{i+1}-y_i}{h_{i+1}} - \frac{h_{i+1}}{3}M_i - \frac{h_{i+1}}{6}M_{i+1} \quad (2.5.8)$$

将 (2.5.7)、(2.5.8) 代入 (2.5.5) 整理后得

$$\frac{h_i}{6}M_{i-1} + \frac{h_i+h_{i+1}}{3}M_i + \frac{h_{i+1}}{6}M_{i+1} = \frac{y_{i+1}-y_i}{h_{i+1}} - \frac{y_i-y_{i-1}}{h_i}$$

两边同乘以 $\frac{6}{h_i+h_{i+1}}$, 即得方程组

$$\frac{h_i}{h_i+h_{i+1}}M_{i-1} + 2M_i + \frac{h_{i+1}}{h_i+h_{i+1}}M_{i+1} = \frac{6}{h_i+h_{i+1}}\left(\frac{y_{i+1}-y_i}{h_{i+1}} - \frac{y_i-y_{i-1}}{h_i}\right) \quad (i=1,2,\dots,n-1)$$

若记

$$\begin{cases} \mu_i = \frac{h_i}{h_i+h_{i+1}}, \lambda_i = \frac{h_{i+1}}{h_i+h_{i+1}} = 1 - \mu_i \\ g_i = \frac{6}{h_i+h_{i+1}}(f[x_i, x_{i+1}] - f[x_{i-1}, x_i]) \\ (i=1,2,\dots,n-1) \end{cases} \quad (2.5.9)$$

则所得方程组可简写成

$$\mu_i M_{i-1} + 2M_i + \lambda_i M_{i+1} = g_i \quad (i=1,2,\dots,n-1)$$

即

$$\begin{cases} \mu_1 M_0 + 2M_1 + \lambda_1 M_2 = g_1 \\ \mu_2 M_1 + 2M_2 + \lambda_2 M_3 = g_2 \\ \dots\dots\dots \\ \mu_{n-1} M_{n-2} + 2M_{n-1} + \lambda_{n-1} M_n = g_{n-1} \end{cases} \quad (2.5.10)$$

这是一个含有 $n+1$ 个未知数、 $n-1$ 个方程的线性方程组。要确定 M_i 的值, 还需用到边界条件。

在第(1)种边界条件下, 由于 $S'(x_0)=y'_0$ 和 $S'(x_n)=y'_n$ 已知, 可以得到包含 M_i 另外两个线性方程。由 (2.5.6) 知, $S(x)$ 在子区间 $[x_0, x_1]$ 上的导数为:

$$S'_1(x) = -M_0 \frac{(x_1-x)^2}{2h_1} + M_1 \frac{(x-x_0)^2}{2h_1} + \frac{y_1-y_0}{h_1} - \frac{h_1}{6}(M_1-M_0)$$

故由条件 $S'(x_0)=y'_0$ 立即可得

$$y'_0 = -M_0 \frac{h_1}{2} + \frac{y_1 - y_0}{h_1} - \frac{h_1}{6}(M_1 - M_0)$$

即

$$2M_0 + M_1 = \frac{6}{h_1} \left(\frac{y_1 - y_0}{h_1} - y'_0 \right) \quad (2.5.11)$$

同理，由条件 $S'(x_n) = y'_n$ ，可得

$$M_{n-1} + 2M_n = \frac{6}{h_n} \left(y'_n - \frac{y_n - y_{n-1}}{h_n} \right) \quad (2.5.12)$$

将(2.5.10)、(2.5.11)、(2.5.12)合在一起，即得确定 M_0, M_1, \dots, M_n 的线性方程组：

$$\begin{bmatrix} 2 & 1 & & & \\ \mu_1 & 2 & \lambda_1 & & \\ & \ddots & \ddots & \ddots & \\ & & \mu_{n-1} & 2 & \lambda_{n-1} \\ & & & 1 & 2 \end{bmatrix} \begin{bmatrix} M_0 \\ M_1 \\ \vdots \\ M_{n-1} \\ M_n \end{bmatrix} = \begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_{n-1} \\ g_n \end{bmatrix} \quad (2.5.13)$$

其中

$$\begin{cases} g_0 = \frac{6}{h_1} (f[x_0, x_1] - y'_0) \\ g_n = \frac{6}{h_n} (y'_n - f[x_{n-1}, x_n]) \end{cases} \quad (2.5.14)$$

在第(2)种边界条件下，由 $M_0 = S''(x_0) = y''_0$ ， $M_n = S''(x_n) = y''_n$ 已知，在方程组(2.5.14)中实际上只包含有 $n-1$ 个未知数 M_1, M_2, \dots, M_{n-1} ，并且可以改写成：

$$\begin{bmatrix} 2 & \lambda_1 & & & \\ \mu_2 & 2 & \lambda_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \mu_{n-2} & 2 & \lambda_{n-2} \\ & & & \mu_{n-1} & 2 \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \begin{bmatrix} g_1 - \mu_1 y''_0 \\ g_2 \\ \vdots \\ g_{n-2} \\ g_{n-1} - \lambda_{n-1} y''_n \end{bmatrix} \quad (2.5.15)$$

在第(3)种边界条件下，由 $S'(x_0+0) = S'(x_n-0)$ ，直接可得

$$M_0 = M_n \quad (2.5.16)$$

由条件 $S''(x_0+0) = S''(x_n-0)$ 可得

$$-M_0 \frac{h_1}{2} + \frac{y_1 - y_0}{h_1} - \frac{h_1}{6} (M_1 - M_0) = M_n \frac{h_n}{2} + \frac{y_n - y_{n-1}}{h_n} - \frac{h_n}{6} (M_n - M_{n-1})$$

注意到 $y_0 = y_n$ 和 $M_0 = M_n$ ，上式整理后得：

$$\frac{h_1}{h_1 + h_n} M_1 + \frac{h_n}{h_1 + h_n} M_{n-1} + 2M_n = \frac{6}{h_1 + h_n} \left(\frac{y_1 - y_0}{h_1} - \frac{y_n - y_{n-1}}{h_n} \right)$$

若记

$$\mu_n = \frac{h_n}{h_1 + h_n}, \lambda_n = \frac{h_1}{h_1 + h_n} = 1 - \mu_n, g_n = \frac{6}{h_1 + h_n} (f[x_1, x_n] - f[x_{n-1}, x_n])$$

则所得方程可简写成：

$$\lambda_n M_1 + \mu_n M_{n-1} + 2M_n = g_n \quad (2.5.17)$$

将 (2.5.10) (2.5.16) (2.5.17) 合在一起，即得确定 M_1, M_2, \dots, M_n 的线形方程组：

$$\begin{bmatrix} 2 & \lambda_1 & & & \mu_1 \\ \mu_2 & 2 & \lambda_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \mu_{n-1} & 2 & \lambda_{n-1} \\ \lambda_n & & & \mu_n & 2 \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_{n-1} \\ M_n \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{n-1} \\ g_n \end{bmatrix} \quad (2.5.18)$$

利用线性代数知识，可以证明方程组 (2.5.13)、(2.5.15) 及 (2.5.18) 的系数矩阵都是非奇异的，从而都有唯一确定的解。

针对不同的边界条件，解相应的方程组 (2.5.13)、(2.5.15) 或 (2.5.18)，求出 M_0, M_2, \dots, M_n 的值，将它们代入 (2.5.4)，就可以得到 $S(x)$ 在各子区间上的表达式。

综合分析，有如下定理

定理 2.5.1 对于给定的函数表

x	x_0	x_1	\cdots	x_n
$y = f(x)$	y_0	y_1	\cdots	y_n

$(a = x_0 < x_1 < \dots < x_n = b)$ ，满足第一、第二和第三种边界条件的三次样条插值函数是存在且唯一的。

三次样条插值函数 $S(x)$ 的求解过程在下面的例子中给出了详细的说明。

例 2.5.3 已知函数 $y = f(x)$ 的函数值如下

x	-1.5	0	1	2
y	0.125	-1	1	9

在区间 $[-1.5, 2]$ 上求三次样条插值函数 $S(x)$ ，使它满足边界条件：

$$S'(-1.5) = 0.75, S'(2) = 14$$

解 先根据给定数据和边界条件算出 μ_i, λ_i, g_i ，写出确定 M_i 的线性方程组。在本例中，给出的是第一种边界条件，确定 M_i ($i=0, 1, 2, 3$)的线性方程组，形如(2.5.13)。

由所给函数表知：

$$h_1 = 1.5 \quad h_2 = 1 \quad h_3 = 1 \quad f[x_0, x_1] = -0.75 \quad f[x_1, x_2] = 2 \quad f[x_2, x_3] = 8$$

于是由 μ_i, λ_i, g_i ($i=1, 2, \dots, n-1$)的算式(2.5.9)知：

$$\mu_1 = 0.6, \mu_2 = 0.5, \lambda_1 = 0.4, \lambda_2 = 0.5, g_1 = 6.6, g_2 = 18$$

由第一种边界条件下 g_0 与 g_n 的计算公式(2.5.14)知：

$$g_0 = \frac{6}{h_1}(f[x_0, x_1] - y'_0) = -6 \quad g_3 = \frac{6}{h_3}(y'_3 - f[x_2, x_3]) = 36$$

故确定 M_0, M_1, M_2 与 M_3 的方程组为：

$$\begin{bmatrix} 2 & 1 & 0 & 0 \\ 0.6 & 2 & 0.4 & 0 \\ 0 & 0.5 & 2 & 0.5 \\ 0 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} M_0 \\ M_1 \\ M_2 \\ M_3 \end{bmatrix} = \begin{bmatrix} -6 \\ 6.6 \\ 18 \\ 36 \end{bmatrix} \quad (2.5.19)$$

解此方程组得到 $S''(x)$ 在各节点 x_i 上的值 M_i 为：

$$M_0 = -5, M_1 = 4, M_2 = 4, M_3 = 16,$$

将 M_i 代入(2.5.4)，即得 $S(x)$ 在各子区间上的表达式 $S_i(x)$ ($i=1, 2, \dots, n$)。

由(2.5.4)知， $S(x)$ 在 $[x_0, x_1]$ 上的表达式为：

$$S_1(x) = M_0 \frac{(x_1 - x)^3}{6h_1} + M_1 \frac{(x - x_0)^3}{6h_1} + \left(y_0 - \frac{M_0}{6}h_1^2\right) \frac{x_1 - x}{h_1} + \left(y_1 - \frac{M_1}{6}h_1^2\right) \frac{x - x_0}{h_1}$$

在本例中，将

$$x_0 = -1.5, \quad x_1 = 0, \quad y_0 = 0.125, \quad y_1 = -1, \quad M_0 = -5, \quad M_1 = 4$$

代入，整理后得

$$S_1(x) = x^3 + 2x^2 - 1 \quad x \in [-1.5, 0],$$

同理可得：

$$S_2(x) = 2x^2 - 1 \quad x \in [0, 1]$$

$$S_3(x) = 2x^3 - 4x^2 + 6x - 3 \quad x \in [1, 2]$$

故所求三次样条插值函数为：

$$S(x) = \begin{cases} x^3 + 2x^2 - 1 & (-1.5 \leq x \leq 0) \\ 2x^2 - 1 & (0 \leq x \leq 1) \\ 2x^3 - 4x^2 + 6x - 3 & (1 \leq x \leq 2) \end{cases}$$

上述求三次样条插值函数的方法，其基本思路 and 特点是：

先利用一阶导数 $S'(x)$ 在内节点 $x_i (i=1, 2, \dots, n-1)$ 上的连续性以及边界条件，列出确定二阶导数 $M_i = S''(x_i) (i=0, 1, \dots, n)$ 的线性方程组（在力学上称为三弯矩方程组），并由此解出 M_i ，然后用 M_i 来表达 $S(x)$ 。

通过别的途径也可求三次样条插值函数。例如，可以先利用二阶导数在内节点上的连续性以及边界条件，列出确定一阶导数

$$m_i = S'(x_i) (i=0, 1, \dots, n)$$

的线性方程组（在力学上称为三转角方程组），并由此解出 m_i ，然后用 m_i 来表达 $S(x_i)$ 。在有些情况下，这种表达方法与前者相比较，使用起来更方便。读者可以参阅相关书籍，在此不赘述。

2.6 求函数近似表达式的拟合法

在科学实验和生产实践中，经常要从一组实验数据 $(x_i, y_i) (i=1, 2, \dots, m)$ 出发，寻求函数 $y = f(x)$ 的一个近似表达式 $y = \Phi(x)$ （称为经验公式）。从几何上，就是希望根据给出的 m 个点 (x_i, y_i) ，求曲线 $y = f(x)$ 的一条近似曲线 $y = \Phi(x)$ 。这就是曲线拟合的问题。

多项式插值虽然在一定程度上解决了由函数表求函数的近似表达式问题，但它来解决这里提出的问题，有明显缺陷。

首先，实验提供的数据通常带有测试误差。如要求近似曲线 $y = \Phi(x)$ 严格地通过所给的每个数据点 (x_i, y_i) ，就会使曲线保持原有的测试误差。当个别数据的误差较大时，插值效果显然是不理想的。

其次，由实验提供的数据往往较多（即 m 较大），用插值法得到的近似表达式，明显地缺乏实用价值。

因此，怎样从给定的一组数据出发，在某个函数类 Φ 中寻求一个“最好”的函数 $\Phi(x)$ 来拟合这组数据，是一个值得讨论的问题。

随着拟合效果“好”、“坏”标准的不同，解决此类问题的方法也不同。这里介绍一种最常用的曲线拟合方法，即最小二乘法。

2.6.1 曲线拟合的最小二乘法

如前所述，在一般情况下，我们不能要求近似曲线 $y = f(x)$ 严格地通过所有数据点 (x_i, y_i) ，亦不能要求所有拟合曲线函数在 x_i 处的偏差（亦称残差）

$\delta_i = \varphi(x_i) - y_i$ ($i = 1, 2, \dots, m$) 都严格地趋于零。但是，为了使近似曲线尽量反映所给数据点的变化趋势，要求 $|\delta_i|$ 都较小还是必要的。达到这一目标的途径很多，常见的有：

(1) 选取 $\Phi(x)$ ，使偏差绝对值之和最小，即

$$\sum_{i=1}^m |\delta_i| = \sum_{i=1}^m |\varphi(x_i) - y_i| = \min \quad (2.6.1)$$

(2) 选取 $\Phi(x)$ ，使偏差最大，绝对值最小，即

$$\max_{1 \leq i \leq m} |\delta_i| = \max_{1 \leq i \leq m} |\varphi(x_i) - y_i| = \min \quad (2.6.2)$$

(3) 选取 $\Phi(x)$ ，使偏差平方和最小，即

$$\sum_{i=1}^m \delta_i^2 = \sum_{i=1}^m [\varphi(x_i) - y_i]^2 = \min \quad (2.6.3)$$

为了方便计算、分析与应用，我们较多地根据“偏差平方和最小”的原则（称为最小二乘原则）来选取拟合曲线 $y = \Phi(x)$ 。按最小二乘原则选择拟合曲线的方法，称为最小二乘法。

本章要着重讨论的线性最小二乘问题，其基本提法是：对于给定数据表

x	x_0	x_1	\cdots	x_m
y	y_0	y_1	\cdots	y_m

要求在某个函数类 $\Phi = \{\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)\}$ （其中 $n < m$ ）中寻求一个函数

$$\varphi^*(x) = a_0^* \varphi_0(x) + a_1^* \varphi_1(x) + \cdots + a_n^* \varphi_n(x) \quad (2.6.4)$$

使 $\Phi^*(x)$ 满足条件

$$\sum_{i=1}^m [\varphi^*(x_i) - y_i]^2 = \min_{\varphi(x) \in \Phi} \sum_{i=1}^m [\varphi(x_i) - y_i]^2 \quad (2.6.5)$$

式中, $\varphi(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \dots + a_n\varphi_n(x)$ 是函数类 Φ 中任一函数。

满足关系式 (2.6.5) 的函数 $\varphi^*(x)$, 称为上述最小二乘问题的最小二乘解。

由此可知, 用最小二乘法解决实际问题包含两个基本环节: 先根据所给数据点的变化趋势与问题的实际背景确定函数类 Φ , 即确定 $\varphi(x)$ 所具有的形式; 然后按最小二乘法原则 (2.6.3) 求取最小二乘解 $\varphi^*(x)$, 即确定其系数 $a_k^* (k=0, 1, \dots, n)$ 。

由最小二乘解 (2.6.4) 应满足条件 (2.6.5) 知, 点 $a_0^*, a_1^*, \dots, a_n^*$ 是多元函数

$$S(a_0, a_1, \dots, a_n) = \sum_{i=1}^m \left[\sum_{k=0}^n a_k \varphi_k(x_i) - y_i \right]^2$$

的极小值点, 从而 $a_0^*, a_1^*, \dots, a_n^*$ 满足方程组:

$$\frac{\partial S}{\partial a_k} = 0, \quad (k=0, 1, 2, \dots, n)$$

即

$$\sum_{i=1}^m \varphi_k(x_i) [a_0\varphi_0(x_i) + a_1\varphi_1(x_i) + \dots + a_n\varphi_n(x_i) - y_i] = 0, \quad (k=0, 1, 2, \dots, n)$$

亦即

$$a_0 \sum_{i=1}^m \varphi_k(x_i) \varphi_0(x_i) + a_1 \sum_{i=1}^m \varphi_k(x_i) \varphi_1(x_i) + \dots + a_n \sum_{i=1}^m \varphi_k(x_i) \varphi_n(x_i) = \sum_{i=1}^m \varphi_k(x_i) y_i$$

若对任意的函数 $h(x)$ 和 $g(x)$, 引入记号

$$(h, g) = \sum_{i=1}^m h(x_i) g(x_i) \quad (2.6.6)$$

则上述方程组可以表示成

$$a_0(\varphi_k, \varphi_0) + a_1(\varphi_k, \varphi_1) + \dots + a_n(\varphi_k, \varphi_n) = (\varphi_k, f) \quad (k=0, 1, \dots, n)$$

写成矩阵形式即

$$\begin{bmatrix} (\varphi_0, \varphi_0) & (\varphi_0, \varphi_1) & \cdots & (\varphi_0, \varphi_n) \\ (\varphi_1, \varphi_0) & (\varphi_1, \varphi_1) & \cdots & (\varphi_1, \varphi_n) \\ \cdots & \cdots & \cdots & \cdots \\ (\varphi_n, \varphi_0) & (\varphi_n, \varphi_1) & \cdots & (\varphi_n, \varphi_n) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} (\varphi_0, f) \\ (\varphi_1, f) \\ \vdots \\ (\varphi_n, f) \end{bmatrix} \quad (2.6.7)$$

方程组(2.6.7)称为法方程组。

事实上,最小二乘法的方法方程可以用下面的方法形成。

在 $\varphi(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots + a_n\varphi_n(x)$ 中, 当 $x = x_i, i = 1, 2, \cdots, m$ 时, 令

$\varphi(x_i) = y_i$, 即得方程组

$$\begin{cases} a_0\varphi_0(x_1) + a_1\varphi_1(x_1) + \cdots + a_n\varphi_n(x_1) = y_1 \\ a_0\varphi_0(x_2) + a_1\varphi_1(x_2) + \cdots + a_n\varphi_n(x_2) = y_2 \\ \cdots \cdots \cdots \\ a_0\varphi_0(x_m) + a_1\varphi_1(x_m) + \cdots + a_n\varphi_n(x_m) = y_m \end{cases}$$

将其写成矩阵形式

$$\begin{pmatrix} \varphi_0(x_1) & \varphi_1(x_1) & \cdots & \varphi_n(x_1) \\ \varphi_0(x_2) & \varphi_1(x_2) & \cdots & \varphi_n(x_2) \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_0(x_m) & \varphi_1(x_m) & \cdots & \varphi_n(x_m) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{pmatrix}$$

令 $\begin{pmatrix} \varphi_0(x_1) & \varphi_1(x_1) & \cdots & \varphi_n(x_1) \\ \varphi_0(x_2) & \varphi_1(x_2) & \cdots & \varphi_n(x_2) \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_0(x_m) & \varphi_1(x_m) & \cdots & \varphi_n(x_m) \end{pmatrix} = A_{m \times (n+1)}, \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \vec{\alpha}, \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{pmatrix} = \vec{y}$, 则方程组可写为

$$A_{m \times (n+1)} \vec{\alpha} = \vec{y}$$

将方程两边同时乘以 A^T , 则可得到

$$A^T A_{m \times (n+1)} \vec{\alpha} = A^T \vec{y}$$

这就是最小二乘法的方法方程 (2.6.7)。

当 $\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)$ 线性无关时, 可以证明它有唯一解

$$a_0 = a_0^*, a_1 = a_1^*, \dots, a_n = a_n^*$$

并且相应的函数(2.6.4)就是满足条件(2.6.5)的最小二乘解。

定理 2.6.1 对任意给定的一组实验数据 $(x_i, y_i) (i = 1, 2, \cdots, m)$ (其中 x_i 互异), 在函数类

$$\Phi = \{\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)\} (n < m) \quad (\varphi_0, \varphi_1, \dots, \varphi_n \text{ 线性无关})$$

中, 存在唯一的函数

$$\varphi^* = a_0^* \varphi_0(x) + a_1^* \varphi_1(x) + \dots + a_n^* \varphi_n(x)$$

使得关系式(2.6.5)成立, 并且其系数 $a_i^* (i=0, 1, \dots, n)$ 可以通过解方程组(2.6.7)得到。

作为曲线拟合的一种常用的情况, 若讨论的是代数多项式拟合, 即取

$$\varphi_0(x) = 1, \varphi_1(x) = x, \dots, \varphi_n(x) = x^n,$$

则由(2.6.6)知:

$$(\varphi_j, \varphi_k) = \sum_{i=1}^m x_i^j x_i^k = \sum_{i=1}^m x_i^{j+k} \quad (j, k = 0, 1, \dots, n)$$

$$(\varphi_k, f) = \sum_{i=1}^m x_i^k y_i \quad (k = 0, 1, \dots, n)$$

故相应的法方程组为:

$$\begin{pmatrix} m & \sum_{i=1}^m x_i & \dots & \sum_{i=1}^m x_i^n \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 & \dots & \sum_{i=1}^m x_i^{n+1} \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^m x_i^n & \sum_{i=1}^m x_i^{n+1} & \dots & \sum_{i=1}^m x_i^{2n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m x_i y_i \\ \vdots \\ \sum_{i=1}^m x_i^n y_i \end{pmatrix} \quad (2.6.8)$$

下面通过两个具体的例子来说明用最小二乘法解决实际的问题的具体步骤与某些技巧。

例 2.6.1 某种铝合金的含铝量为 $x\%$, 其熔解温度为 $y^\circ\text{C}$, 由实验测得 x 与 y 的数据如表 2.6.1 左边三列。使用最小二乘法建立 x 与 y 之间的经验公式。

解 根据前面的讨论, 解决问题的过程如下:

(1) 将表中给出的数据点 $(x_i, y_i) (i=1, 2, \dots, 6)$ 描绘在坐标纸上, 如图 2.6.1 所示。

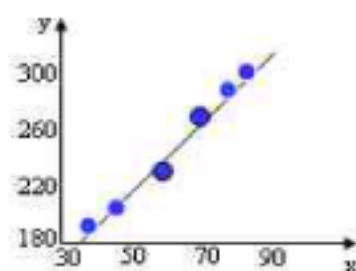


图 2.6.1 数据的散点图

(2) 确定拟合曲线的形式。由图 2.6.1 可以看出, 六个点位于一条直线的附近, 故可以选用线性函数 (直线) 来拟合这组实验数据, 即令

$$\varphi(x) = a + bx \quad (2.6.9)$$

其中 a, b 为待定常数。

(3) 建立法方程组。由于问题归结为一次多项式拟合问题, 故由 (2.6.8) 知, 相

应的法方程组形如:

$$\begin{bmatrix} 6 & \sum_{i=1}^6 x_i \\ \sum_{i=1}^6 x_i & \sum_{i=1}^6 x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^6 y_i \\ \sum_{i=1}^6 x_i y_i \end{bmatrix}$$

经过计算 (表 2.6.1) 即得确定待定系数 a, b 的法方程组:

$$\begin{cases} 6a + 396.9b = 1458 \\ 396.6a + 28365.28b = 101176.3 \end{cases} \quad (2.6.10)$$

(4) 解法方程组 (2.6.10) 得:

$$a = 95.3524, b = 2.2337$$

代入 (2.6.9) 即得经验公式:

$$y = 95.3524 + 2.2337x \quad (2.6.11)$$

表 2.6.1 多项式拟合法方程系数计算表

i	x_i	y_i	x_i^2	$x_i y_i$
1	36.9	181	1361.61	6678.9
2	46.7	197	2180.89	9199.9
3	63.7	235	4057.69	14969.5
4	77.8	270	6052.84	21006.0
5	84.0	283	7056.00	23772.0
6	87.5	292	7656.25	25550.0
Σ	396.6	1458	28365.28	101176.3

所得经验公式能否较好地反映客观规律, 还需通过实践来检验。由 (2.6.11) 式算出的函数值 (称为拟合值)

$$y_i = 95.3524 + 2.2337x_i (i = 1, 2, \dots, m)$$

与实际值有一定的偏差, 见表 2.6.2。

由表 2.6.2 可以看出, 偏差的平方和

$$\sum_{i=1}^6 \delta_i^2 = 26.6704$$

其平方根（称为均方误差）

$$\sqrt{\sum \delta_i^2} = 5.164$$

在一定程度上反映了所得经验公式的好坏。

同时，由表 2.6.2 还可以看出，最大偏差为 $\max_{1 \leq i \leq 6} |\delta_i| = 3.22$ 。

表 2.6.2 拟合偏差分析表

i	1	2	3	4	5	6
x_i	36.9	46.7	63.7	77.8	84.0	87.5
y_i	177.78	199.67	237.64	269.13	282.98	290.80
y_i	181	197	235	270	283	292
$\delta_i = y_i - y_i$	-3.22	2.67	2.64	-0.87	-0.02	-1.20
δ_i^2	10.37	7.13	6.97	0.76	0.0004	1.44
$\sum \delta_i^2$	26.6704					

如果认为这样的误差是允许的话，就可以用经验公式(2.6.11)来计算含铝量在36.9%~87.5%之间的溶解度。否则，就要用改变函数类型或者增加实验数据等方法来建立新的经验公式。

例 2.6.2 在某化学反应里，测得生成物的浓度 $y\%$ 与时间 t 的数据见表 2.6.3，

试用最小二乘法建立 t 与 y 的经验公式。

表 2.6.3 相应时间的浓度数据表

i	1	2	3	4	5	6	7	8
y	4.00	6.40	8.00	8.80	9.22	9.50	9.70	9.86
t	9	10	11	12	13	14	15	16
y	10.00	10.20	10.32	10.42	10.50	10.55	10.58	10.60

解 将已知数据点 $(t_i, y_i)(i=1,2,\cdots,16)$ 描述在坐标纸上，得散点图 2.6.2。

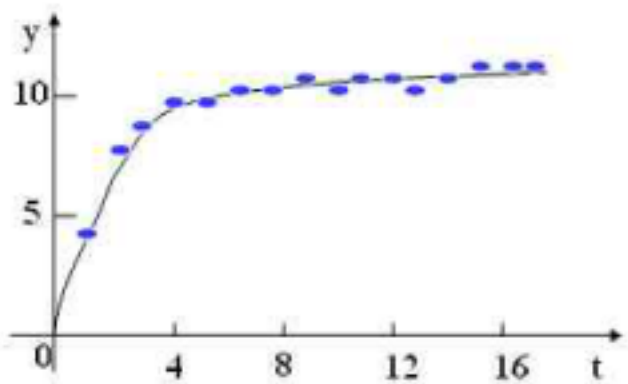


图 2.6.2 数据的散点图

由图 2.6.2 及问题的物理背景可以看出, 拟合曲线 $y = \varphi(x)$ 应具下列特点:

(1) 曲线随着 t 的增加而上升, 但上升速度由快到慢。

(2) 当 $t = 0$ 时, 反应还未开始, 即 $y = 0$; 当 $t = \infty$ 时, y 趋于某一常数。故曲线应通过原点(或者当 $t = 0$ 时以原点为极限点), 且有一条水平渐近线。

具有上述特点的曲线很多。选用不同的数学模型, 可以获得不同的拟合曲线与经验公式。

下面提供两种方案:

方案 1 设想 $y = \varphi(t)$ 是双曲线型的, 并且具有形式

$$y = \frac{t}{at+b} \quad (2.6.12)$$

此时, 若直接按最小二乘法原则去确定参数 a 和 b , 则问题归结为求二元函数

$$S(a, b) = \sum_{i=1}^{16} \left(\frac{t_i}{at_i + b} - y_i \right)^2 \quad (2.6.13)$$

的极小点, 这将导致求解非线性方程组:

$$\begin{cases} \frac{\partial S}{\partial a} = -2 \sum_{i=1}^{16} \frac{t_i^2}{(at_i + b)^2} \left(\frac{t_i}{at_i + b} - y_i \right) = 0 \\ \frac{\partial S}{\partial b} = -2 \sum_{i=1}^{16} \frac{t_i}{(at_i + b)^2} \left(\frac{t_i}{at_i + b} - y_i \right) = 0 \end{cases}$$

给计算带来了麻烦。可通过变量替换来将它转化为关于待定参数的线性形函数。为此, 将 (2.6.12) 改写成

$$\frac{1}{y} = a + \frac{b}{t}$$

于是, 若引入新变量

$$y^{(1)} = \frac{1}{y}, t^{(1)} = \frac{1}{t}$$

则 (2.6.12) 式就是

$$y^{(1)} = a + bt^{(1)}$$

同时, 由题中所给数据表 2.6.3 可以算出新的数据表 2.6.4。这样, 问题就归结为: 根据数据表 2.6.4, 求形如 $y^{(1)} = a + bt^{(1)}$ 的最小二乘解。参照例 2.6.1 的做法, 解方程组

$$\begin{bmatrix} 16 & \sum_{i=1}^{16} t_i^{(1)} \\ \sum_{i=1}^{16} t_i^{(1)} & \sum_{i=1}^{16} (t_i^{(1)})^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{16} y_i^{(1)} \\ \sum_{i=1}^{16} t_i^{(1)} y_i^{(1)} \end{bmatrix},$$

即得

$$a = 80.6621, b = 161.6822。$$

代入(2.6.12)，即得经验公式

$$y = \frac{t}{80.6621t + 161.6822} \quad (2.6.14)$$

表 2.6.4 算出数据表

i	1	2	3	...	16
$t_i^{(1)} = \frac{1}{t_i}$	1.00000	0.50000	0.33333	...	0.06250
$y_i^{(1)} = \frac{1}{y_i}$	0.25000	0.15625	0.12500	...	0.09434

方案 2 设想 $y = \varphi(t)$ 具有指数形式

$$y = ae^{bt} \quad (a > 0, b < 0) \quad (2.6.15)$$

为了求参数 a 和 b 时，避免求解一个非线性方程组，对上式两边取对数

$$\ln y = \ln a + \frac{b}{t},$$

此时，若引入新变量

$$y^{(2)} = \ln y, t^{(2)} = \frac{1}{t}$$

并记 $A = \ln a, B = b$ ，则上式就是

$$y^{(2)} = A + Bt^{(2)}$$

又由表 2.6.3 可算出新的数据表 2.6.5。

表 2.6.5 算出数据表

i	1	2	3	...	16
$t_i^{(2)} = \frac{1}{t_i}$	1.00000	0.50000	0.33333	...	0.06250
$y_i^{(2)} = \ln y_i$	1.38629	1.85630	2.07944	...	2.36085

于是将问题归为：根据数据表 2.6.5，求形如 $y^{(2)} = A + Bt^{(2)}$ 的最小二乘解。

参照方案 1，写出相应的法方程组并解之，即得

$$A = -4.4807, B = -1.0567$$

于是

$$a = e^{-4.4807} = 0.011325 \quad B = -$$

故得另一个经验公式

$$y = 0.011325e^{\frac{1.0567}{t}} \quad (2.6.16)$$

将两个不同的经验公式(2.6.14)和(2.6.16)的均方误差和最大偏差进行比较,见表2.6.6。从均方误差与最大偏差这两个不同角度看,后者均优于前者。因此,在解决实际问题时,常常要经过反复分析,多次选择,计算与比较,才能获得好的数学模型。

表 2.6.6 不同经验公式的误差比较表

经验公式	均方误差	最大偏差
(3.9) 式	1.19×10^{-3}	0.568×10^{-3}
(3.11) 式	0.34×10^{-3}	0.277×10^{-3}

下面以常用的多项式拟合为例,说明最小二乘法在计算机上实现的步骤。

设有一组实验数据 $(x_i, y_i) (i=1, 2, \dots, m)$, 今要用最小二乘法求一 $n (n < m)$ 次多项式曲线

$$\varphi_n(x) = a_0 + a_1x + \dots = a_nx^n$$

来拟合这组数据。显然,求 $\varphi_n(x)$ 的实质就是要确定其系数 $a_i (i=1, 2, \dots, n)$ 。

由前面讨论可知,问题可归结为建立和求解法方程组(2.6.8)。为了便于编程序并减少工作量,引入矩阵

$$C = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

则法方程组(2.6.8)的系数矩阵(用 A 表示)和右端向量(用 b 表示)分别为:

$$A = C^T C \quad b = C^T Y。$$

2.6.2 加权最小二乘法

在实际问题中测得的所有实验数据,并不是总是等精度、等地位的。显然,对于精度较高或地位较重要(这应根据具体情况来判定)的那些数据 (x_i, y_i) ,应当给予较大的权。在这种情况下,求给定数据的拟合曲线,就要采用加权最小二乘法。

用加权最小二乘法进行曲线拟合的要求与原则:对于给定的一组实验数据 $(x_i, y_i) (i=1, 2, \dots, m)$, 要求在某个函数类 $\Phi = \{\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)\} (n < m)$ 中, 寻求一个函数

$$\varphi^*(x) = a_0^* \varphi_0(x) + a_1^* \varphi_1(x) + \cdots + a_n^* \varphi_n(x)$$

使

$$\sum_{i=1}^m W_i [\varphi^*(x_i) - y_i]^2 = \min_{\varphi(x) \in \Phi} \sum_{i=1}^m W_i [\varphi(x_i) - y_i]^2$$

其中

$$\varphi(x) = a_0 \varphi_0(x) + a_1 \varphi_1(x) + \cdots + a_n \varphi_n(x)$$

为函数类 Φ 中任一函数； $W_i (i=1, 2, \cdots, m)$ 是一列正数，称为权，它的大小反映了

数据 (x_i, y_i) 地位的强弱。显然，求 $\varphi^*(x)$ 的问题可归结为求多元函数：

$$S(a_0, a_1, \cdots, a_n) = \sum_{i=1}^m W_i \left[\sum_{k=0}^n a_k \varphi_k(x_i) - y_i \right]^2$$

的极小点 $(a_0^*, a_1^*, \cdots, a_n^*)$ 。采用最小二乘解的求法，仍可得法方程组 (2.6.7)，但

其中

$$(\varphi_k, \varphi_j) = \sum_{i=1}^m W_i \varphi_k(x_i) \varphi_j(x_i) (k, j = 0, 1, \cdots, n),$$

$$(\varphi_k, f) = \sum_{i=1}^m W_i \varphi_k(x_i) y_i (k = 0, 1, \cdots, n)。$$

作为特例，如果选用的拟合曲线为 $\varphi_n(x) = a_0 + a_1 x + \cdots + a_n x^n$ ，那么相应的法方程组为

$$\begin{pmatrix} \sum_{i=1}^m W_i & \sum_{i=1}^m W_i x_i & \cdots & \sum_{i=1}^m W_i x_i^n \\ \sum_{i=1}^m W_i x_i & \sum_{i=1}^m W_i x_i^2 & \cdots & \sum_{i=1}^m W_i x_i^{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^m W_i x_i^n & \sum_{i=1}^m W_i x_i^{n+1} & \cdots & \sum_{i=1}^m W_i x_i^{2n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m W_i y_i \\ \sum_{i=1}^m W_i x_i y_i \\ \vdots \\ \sum_{i=1}^m W_i x_i^n y_i \end{pmatrix} \quad (2.6.17)$$

例 2.6.3 已知一组实验数据 (x_i, y_i) 及权 W_i 如表 2.6.7。若 x 与 y 之间有线性关系，

试用最小二乘法确定系数 a 和 b 。

表 2.6.7 实验数据表

i	1	2	3	4
W_i	14	27	12	1
X_i	2	4	6	8
Y_i	2	11	28	40

解 因为拟合曲线为一次多项式曲线（直线）

$$\varphi_1(x) = a + bx,$$

故相应的法方程组如 (2.6.17)。将表中各已知数据代入，即得方程组

$$\begin{cases} 54a + 216b = 701 \\ 216a + 984b = 3580 \end{cases}$$

解之得

$$a = -12.885, b = 6.467$$

2.6.3 利用正交函数作最小二乘法拟合

在前几节，虽然从原则上解决了最小二乘意义下的曲线拟合问题，但在实际计算中，由于当 n 较大，例如 $n \geq 7$ ，法方程组往往是病态的，因而给求解工作带来了一定困难。近年来，产生了许多解决这一困难的新方法。本节将简要介绍利用正交函数作最小二乘拟合的基本原理，以及利用正交多项式拟合的一种行之有效的方法。

对于 $\{x_i\}$ 和权 $\{W_i\}$ ($i = 1, 2, \dots, m$)，若一组函数

$$\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x) (n < m)$$

满足条件：

$$(\varphi_k, \varphi_j) = \sum_{i=1}^m W_i \varphi_k(x_i) \varphi_j(x_i) = \begin{cases} 0 & k \neq j \\ A_k > 0 & k = j \end{cases} \quad (k, j = 0, 1, \dots, n) \quad (2.6.18)$$

则称 $\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)$ 是关于点集 $\{x_i\}$ 带权 $\{W_i\}$ 的正交函数族。

特别，当 $\varphi_k(x) (k = 0, 1, \dots, n)$ 都是多项式时，就称 $\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)$ 是关于点集 $\{x_i\}$ 带权 $\{W_i\}$ 的一组正交多项式。

如果在提到正交函数或正交多项式时，没有提到权 W_i ，就意味着权都是 1。

若所考虑的函数类

$$\Phi = \{\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)\}$$

中的基函数是关于给定点集 $\{x_i\}$ 和权 $\{W_i\}$ ($i = 1, 2, \dots, m$) 的正交函数族，则由条件 (2.6.18) 知，在法方程组 (2.6.7) 的系数矩阵中，非对角线上元素

$$(\varphi_k, \varphi_j) = 0 (k \neq j),$$

此时法方程组简化为：

$$\begin{bmatrix} (\varphi_0, \varphi_0) & & & \\ & (\varphi_1, \varphi_1) & & \\ & & \ddots & \\ & & & (\varphi_n, \varphi_n) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} (\varphi_0, f) \\ (\varphi_1, f) \\ \vdots \\ (\varphi_n, f) \end{bmatrix} \quad (2.6.19)$$

只要由此解出

$$a_k = a_k^* (k=0,1,\cdots,n)$$

就可得到最小二乘法解:

$$\varphi^*(x) = a_0^* \varphi_0(x) + a_1^* \varphi_1(x) + \cdots + a_n^* \varphi_n(x) \quad (2.6.20)$$

由条件(2.6.18), 知

$$(\varphi_k, \varphi_k) \neq 0 (k=0,1,\cdots,n),$$

故易解方程组(2.6.19), 且有:

$$a_k^* = \frac{(\varphi_k, f)}{(\varphi_k, \varphi_k)} = \frac{\sum_{i=1}^m W_i \varphi_k(x_i) y_i}{\sum_{i=1}^m W_i [\varphi_k(x_i)]^2} \quad (k=0,1,\cdots,n) \quad (2.6.21)$$

这样就避免了求解病态方程组。

若函数类 Φ 的基函数 $\varphi_0(x), \varphi_1(x), \cdots, \varphi_n(x)$ 是关于给定点集 $\{x_i\}$ 和权 $\{W_i\}$ 的正交函数族, 则可以直接由(2.6.21)式算出待定参数 a_k^* , 进而写出最小二乘解(2.6.20)。因此, 问题归结为对给定的函数类 Φ , 寻求一组由正交函数族组成的基函数的问题。

构造正交函数组的方法很多。下面以多项式为例, 介绍一种具体的方法。这种做法是以下述定理为基础的。

定理 2.6.2 对于给定的点集 $\{x_i\}$ 和权 $\{W_i\}$ ($i=1,2,\cdots,m$) 利用递推公式:

$$\begin{cases} \varphi_0(x) = 1 \\ \varphi_1(x) = x - \alpha_1 \\ \varphi_{k+1}(x) = (x - \alpha_{k+1})\varphi_k(x) - \beta_k \varphi_{k-1}(x) \end{cases} \quad (2.6.22)$$

$$\alpha_{k+1} = \frac{\sum_{i=1}^m W_i x_i [\varphi_k(x_i)]^2}{\sum_{i=1}^m W_i [\varphi_k(x_i)]^2} \quad (2.6.23)$$

$$\beta_k = \frac{\sum_{i=1}^m W_i [\varphi_k(x_i)]^2}{\sum_{i=1}^m W_i [\varphi_{k-1}(x_i)]^2} \quad (k=1,2,\cdots,n-1; n < m) \quad (2.6.24)$$

构造的函数族 $\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)$ 是关于点集 $\{x_i\}$ 和权 $\{W_i\}$ 的一组正交多项式,

且 $\varphi_k(x) (k=1, 2, \dots, n)$ 是 k 次多项式, 其最高次项 x^k 的系数为 1。

例 2.6.4 已知一组实验数据如表 2.6.8, 试用最小二乘法求一条二次拟合曲线。

表 2.6.8 实验数据表

i	1	2	3	4	5	6
x_i	0.0	0.9	1.9	3.0	3.9	5.0
y_i	0.0	10.0	30.0	50.0	80.0	110.0

解 采用边构造正交多项式 $\varphi_k(x)$ 边求最小二乘解系数 a_k^* 的顺序来求拟合曲线。

由公式 (2.6.22) 及 (2.6.21) 立即可得:

$$\varphi_0(x) = 1, \quad a_0^* = \frac{\sum y_i}{6} = \frac{280}{6} = 46.667$$

由公式 (2.6.23)、(2.6.22) 及 (2.6.21) 依次可得:

$$a_1 = \frac{\sum x_i}{6} = \frac{14.7}{6} = 2.45;$$

$$\varphi_1(x) = x - 2.45;$$

$$a_1^* = \frac{\sum \varphi_1(x_i) y_i}{\sum [\varphi_1(x_i)]^2} = \frac{392}{17.615} = 22.254$$

由公式 (2.6.23)、(2.6.22)、(2.6.24) 及 (2.6.21) 依次可得:

$$a_2 = \frac{\sum x_i [\varphi_1(x_i)]^2}{\sum [\varphi_1(x_i)]^2} = \frac{44.359}{17.615} = 2.5183;$$

$$\beta_1 = \frac{\sum [\varphi_1(x_i)]^2}{\sum [\varphi_0(x_i)]^2} = \frac{17.615}{6} = 2.9358$$

$$\varphi_2(x) = (x - 2.5183)(x - 2.45) - 2.9385 = x^2 - 4.9683x + 3.2340$$

$$a_2^* = \frac{\sum \varphi_2(x_i) y_i}{\sum [\varphi_2(x_i)]^2} = \frac{82.8926}{36.8916} = 2.247$$

故所求拟合曲线为:

$$y = a_0^* \varphi_0(x) + a_1^* \varphi_1(x) + a_2^* \varphi_2(x) = 2.247x^2 + 110.9x - 0.5888$$

2.7 城市供水量预测的简单方法

2.7.1 供水量增长率估计与数值微分

供水量的增长率就是供水量函数的导数,因此,要估计供水量的增长率就需要求出供水量函数的导数。用供水量近似函数的导数作为其导数的近似估计是自然的想法。作为多项式插值的应用,本节介绍两种求函数导数近似值的方法。

2.7.2 利用插值多项式求导数

若函数 $f(x)$ 在节点 $x_i (i=0,1,\dots,n)$ 处的函数值已知,就可作 $f(x)$ 的 n 次插值多项式 $P_n(x)$, 并用 $P_n(x)$ 近似代替 $f(x)$, 即

$$f(x) \approx P_n(x) \quad (2.7.1)$$

由于 $P_n(x)$ 是多项式, 容易求导数, 故对应于 $f(x)$ 的每一个插值多项式 $P_n(x)$, 就易建立一个数值微分公式

$$f'(x) \approx P'_n(x),$$

这样建立起来的数值微分公式, 统称为插值型微分公式。

必须注意, 即使 $P_n(x)$ 与 $f(x)$ 的近似程度非常好, 导数 $f'(x)$ 与 $P'_n(x)$ 在某些点上的差别仍旧可能很大, 因而, 在应用数值微分公式时, 要重视对误差的分析。

由插值余项公式 (2.3.9) 知

$$f'(x) - p'_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega'_{n+1}(x) + \frac{\omega_{n+1}(x)}{(n+1)!} \frac{d}{dx} [f^{(n+1)}(\xi)] \quad (2.7.2)$$

由于式中 ξ 是 x 未知函数, 故 $x \neq x_i$ 时, 无法利用上式误差 $f'(x) - p'_n(x)$ 作出估计。

但是, 如果我们限定求某个节点 x_i 处的导数值, 那么 (2.7.2) 右端第二项之值应为零, 此时有

$$f'(x_i) - p'_n(x_i) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega'_{n+1}(x_i)$$

若将它写成带余项的数值微分公式, 即

$$f'(x_i) = p'_n(x_i) + \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega'_{n+1}(x_i) \quad (2.7.3)$$

其中 ξ 在 $x_0, x_1 \cdots x_n$ 之间。该式右端由两部分, 即导数的近似值和相应的截断误差组成。

由 (2.7.3), 当 $n=1$ 时, 插值节点为 x_0, x_1 , 记 $h = x_1 - x_0$, 得带余式的两点公式

$$\begin{cases} f'(x_0) = \frac{f(x_1) - f(x_0)}{h} - \frac{h}{2} f''(\xi) \\ f'(x_1) = \frac{f(x_1) - f(x_0)}{h} + \frac{h}{2} f''(\xi) \end{cases} \quad \xi \in [x_0, x_1] \quad (2.7.4)$$

前一公式的实质是用 $f(x)$ 在 x_0 处的向前差商作为 $f'(x_0)$ 的近似值, 后一公式则是用 $f(x)$ 在 x_1 处的向后差商作为 $f'(x_1)$ 的近似值。

当 $n=2$ 且节点为 $x_k = x_0 + kh$ ($k=0, 1, 2$) 时, 由 (2.7.3) 可得带余项的三点公式:

$$\begin{cases} f'(x_0) = \frac{1}{2h} [-3f(x_0) + 4f(x_1) - f(x_2)] + \frac{h^2}{3} f'''(\xi) \\ f'(x_1) = \frac{1}{2h} [-f(x_0) + f(x_2)] - \frac{h^2}{6} f'''(\xi) \\ f'(x_2) = \frac{1}{2h} [f(x_0) - 4f(x_1) + 3f(x_2)] + \frac{h^2}{3} f'''(\xi) \end{cases} \quad \xi \in [x_0, x_2] \quad (2.7.5)$$

中间一个公式的实质是用 $f(x)$ 在 x_1 处的中心差商作为 $f'(x_1)$ 的近似值, 它与前后两公式相比较, 其优越性是显然的。

用插值多项式 $P_n(x)$ 作为 $f(x)$ 的近似函数, 还可建立高阶的数值微分公式。

例 2.7.1 带余式的二阶三点公式

$$\begin{cases} f''(x_0) = \frac{1}{h^2} [f(x_0) - 2f(x_1) + f(x_2)] + \left[-hf'''(\xi_1) + \frac{h^2}{6} f^{(4)}(\xi_2) \right] \\ f''(x_1) = \frac{1}{h^2} [f(x_0) - 2f(x_1) + f(x_2)] - \frac{h^2}{12} f^{(4)}(\xi_2) \\ f''(x_2) = \frac{1}{h^2} [f(x_0) - 2f(x_1) + f(x_2)] + \left[hf'''(\xi_1) - \frac{h^2}{6} f^{(4)}(\xi_2) \right] \end{cases} \quad (2.7.6)$$

2.7.3 利用三次样条插值函数求导

由三次样条插值函数知, 对于给定函数表

x	x_0	x_1	\cdots	x_n
$y = f(x)$	y_0	y_1	\cdots	y_n

($a = x_0 < x_1 < \cdots < x_n = b$) 和适当的边界条件, 可以写出三次样条插值公式 $S(x)$,

并用 $S(x)$ 近似代替 $f(x)$, 即

$$f(x) \approx S(x), \quad x \in [a, b]$$

由于 $S(x)$ 是一个分段三次多项式, 在各子区间 $[x_{i-1}, x_i]$ ($i=1, 2, \dots, n$) 上容易求出导数, 故可建立数值微分公式

$$f'(x) \approx S'_i(x) = \frac{1}{2h_i} [M_i(x-x_{i-1})^2 - M_{i-1}(x_i-x)^2] + f[x_{i-1}, x_i] + \frac{h_i}{6}(M_{i-1} - M_i) \quad (2.7.7)$$

$$f''(x) \approx S''_i(x) = \frac{1}{h_i} [M_i(x-x_{i-1}) + M_{i-1}(x_i-x)] \quad x \in [x_{i-1}, x_i], i=1, 2, \dots, n \quad (2.7.8)$$

利用函数 $f(x) = \frac{1}{1+25x^2}$ 在节点 $x_i = -1 + 0.1i$ ($i=0, 1, \dots, 20$) 上的函数值和边界条件 $S'(-1) = 0.0740$, $S'(1) = -0.0740$, 构造三次样条插值公式 $S(x)$, 并用它来计算 $f(x)$ 和 $f'(x)$ 在下列点:

$$x_k = -1 + 0.02k \quad (k=0, 1, 2, \dots, 100)$$

处的近似值。计算结果见表 2.7.1。

表 2.7.1 近似值与准确值比较表

x	近似值		准确值	
	$S(x)$	$S'(x)$	$f(x)$	$f'(x)$
-1.00	0.03846	0.074	0.03846	0.07639
-0.92	0.04513	0.09369	0.04513	0.09367
-0.84	0.05365	0.1209	0.05365	0.1209
-0.76	0.06476	0.1594	0.06477	0.1594
-0.68	0.07961	0.2125	0.07962	0.2155
-0.60	0.1000	0.3000	0.1000	0.3000
-0.52	0.1289	0.4319	0.1289	0.4318
-0.44	0.1711	0.6457	0.1712	0.6451
-0.36	0.2359	1.003	0.2358	1.001
-0.28	0.3375	1.579	0.3378	1.598
-0.20	0.5000	2.563	0.5000	2.500
-0.12	0.7372	3.157	0.1353	3.244
-0.04	0.9594	1.885	0.9615	1.849

由表 2.7.1 可以看出, 利用三次样条插值函数 $S(x)$ 及其导数来逼近被插值函数 $f(x)$ 及其导数, 其效果很好。

2.7.4 城市供水量预测

现在来解决本章 2.1.1 节提出的城市供水量预测问题。

用插值方法预测 07 年 1 月份城市的用水量

预测 2007 年 1 月份城市的用水量有两种办法：一是先预测 1 月份每天的日用水量，求和后即得到 1 月用水总量；二是直接用 2000-2006 每年 1 月份的总用水量来预测。

(1) 用 06 年每天的日用水量预测

以预测 07 年 1 月 1 日为例，这里仅用 06 年全部 365 天的日用水量作为插值节点 $(x_i, y_i) (i=1, 2, \dots, 365)$ ，其中 $x_i = i$ ，表示第 i 天， y_i 即为第 i 天的日用水量（万吨/日）。其散点图如图 2.7.1 所示。设插值多项式为 $f(x)$ ，则 07 年 1 月 1 日的日用水量的预测值即为 $f(x)$ 在 $x=366$ 处的函数值。

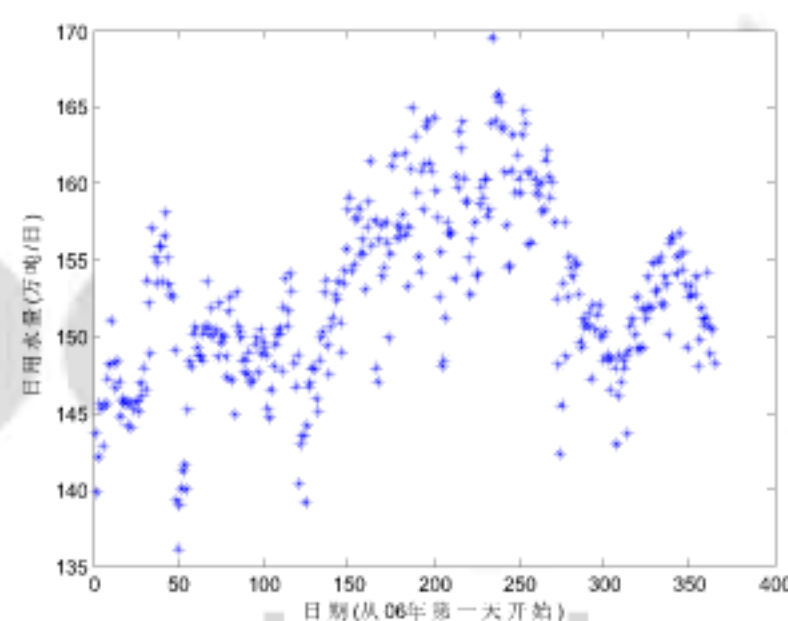


图 2.7.1 06 年城市日用水量散点图

分别用拉格朗日与牛顿插值方法得到 364 次插值多项式，并计算得 $f(366)$ 分别为 $1.8781e+109$ 、 $-9.3842e+167$ ，结果的误差显然太大。实际中，当节点数较多时，一般用分段低次插值，现在用三次样条插值法求解，其样条曲线如图 2.7.2 所示。对应的插值结果为 133.6641。

但若用这个三次样条函数来估计 $f(x), x=367, \dots, 396$ ，即该月后 30 天的日用水量，结果为 98.1477、33.0292、...、-4.8837，误差逐渐增大。由此可见，仅用 06 年一年数据预测 07 年 1 月份每天的日用水量时，即便使用低次插值，误差也会很大。

(2) 用 2000-2006 年 1 月份每天的日用水量预测

现用 2000-2006 年 1 月每天的日用水量预测 07 年 1 月对应天的日用水量。同样，以预测 07 年 1 月 1 日为例。以 2000-2006 每年 1 月 1 日的日用水量作为插值节点 $(x_i, y_i) (i=1, 2, \dots, 7)$ ，则 $f(8)$ 即为所要求的 07 年 1 月 1 日用水量的估计值。同理，其它 30 天的日用水量也对应地求出。

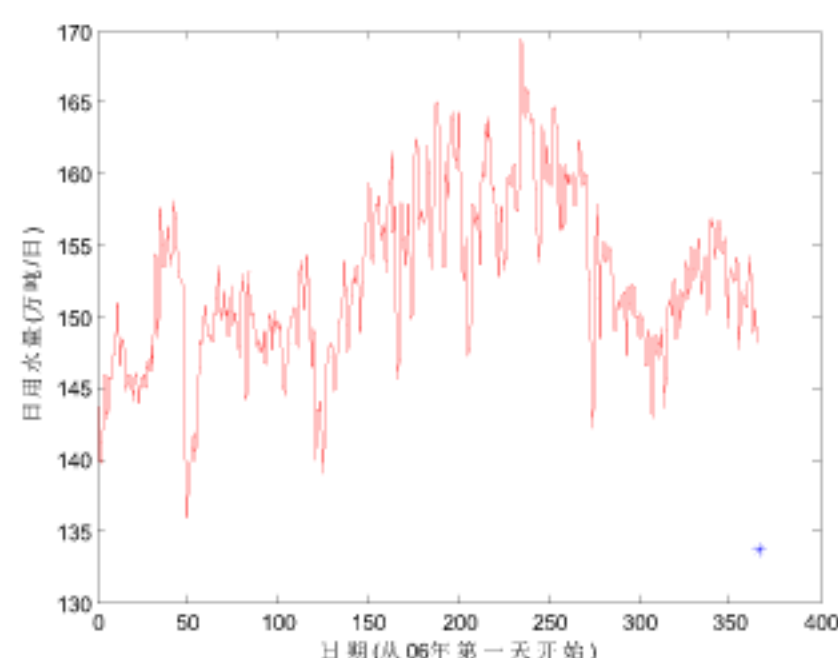


图 2.7.2 三次样条曲线

采用拉格朗日、牛顿和三次样条插值函数进行计算，三种方法计算的结果基本相同，求得07年1月每天日用水量分别为143.6962, 139.7505, ..., 146.5117。对它们求和得07年1月份总用水量的预测值为4517.6993（万吨/日）。其计算结果如图2.7.3所示（“+”号所示为217个已知的前七年日用水量节点，“*”号所示为预测的31天的用水量值）。

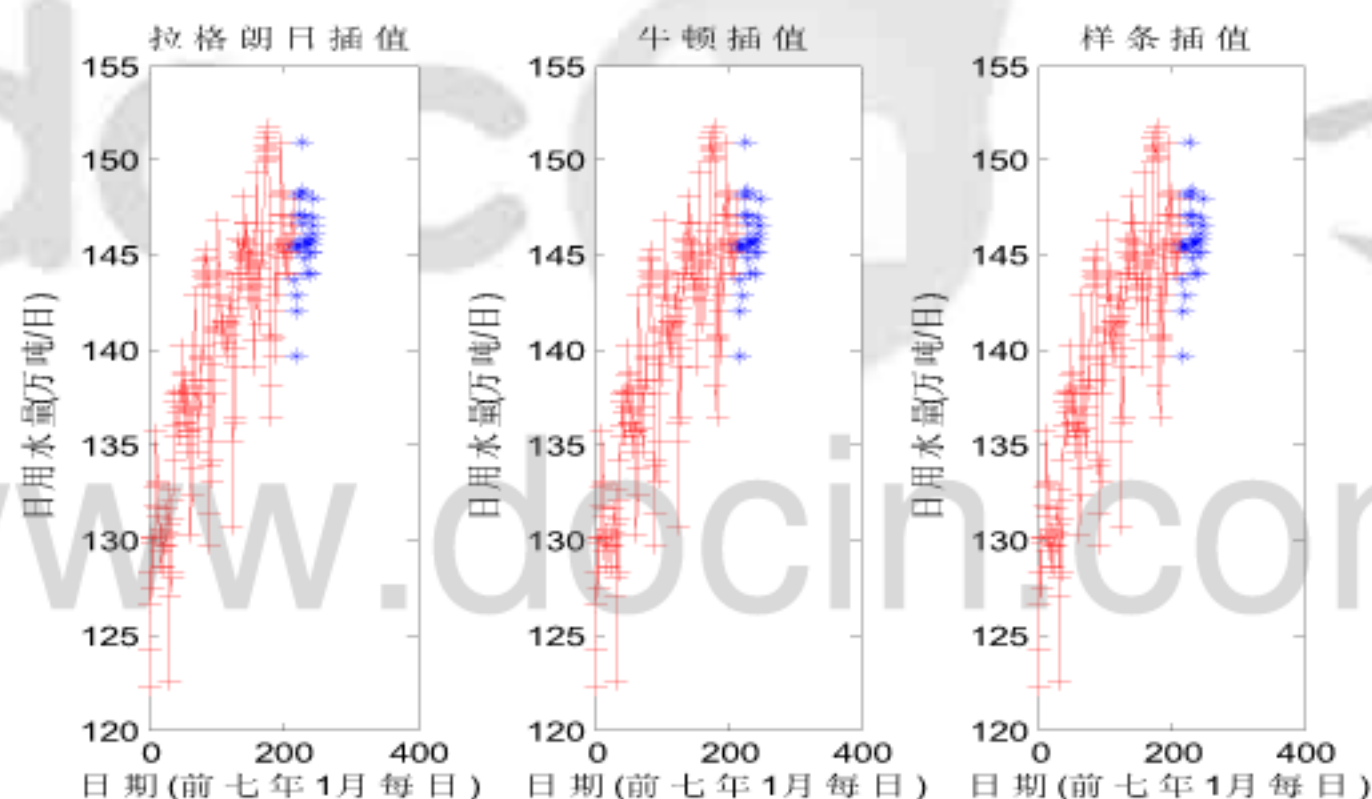


图 2.7.3 三种插值函数曲线

(4) 用 2000-2006 年 1 月城市的总用水量预测

由表 2.1.2 得到 7 个插值节点 (x_i, y_i) ，其中 $x_i = i, i = 1, 2, \dots, 7$ ，其散点图如 2.7.4 所示。用三次样条插值得 $f(8)$ 即所求的 07 年 1 月用水总量的估计值为 4378.1390（万吨/日）。同理得到的 07 年 1 月用水总量估计值如下表 2.7.2 所示。

表 2.7.2 插值法得到 07 年 1 月用水总量估计值（万吨/日）

插值节点	06 年日用水量	00-06 年 1 月每天日用水量	00-06 每年 1 月用水总量
预测结果	误差大	4517.6993（万吨/日）	4378.1390（万吨/日）

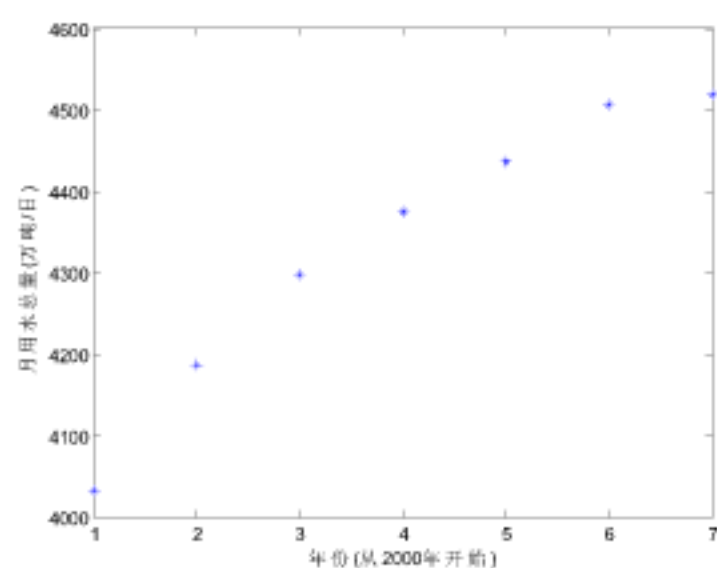


图 2.7.4 散点图

用数据拟合方法预测 2007 年 1 月份城市的用水量

(1) 用 06 年每天的日用水量预测

由 06 年全部 365 天的日用水量散点图 2.7.1 可知, 这些点并不简单地呈现线性或二次关系, 但有很强的聚集性。我们试图用几个多项式进行拟合。用 Matlab 工具箱得到如表 2.7.3 拟合结果。

表 2.7.3 拟合次数与残差的关系

	线性	二次	三次	四次	五次	六次	七次	八次	九次
均方根误差	5.366	4.689	4.55	4.4	3.993	3.897	3.781	3.731	3.701

拟合结果显示, 九次拟合的均方根误差是最小的, 但四次曲线拟合的效果已经与它很接近了, 拟合曲线如图 2.7.5 所示。

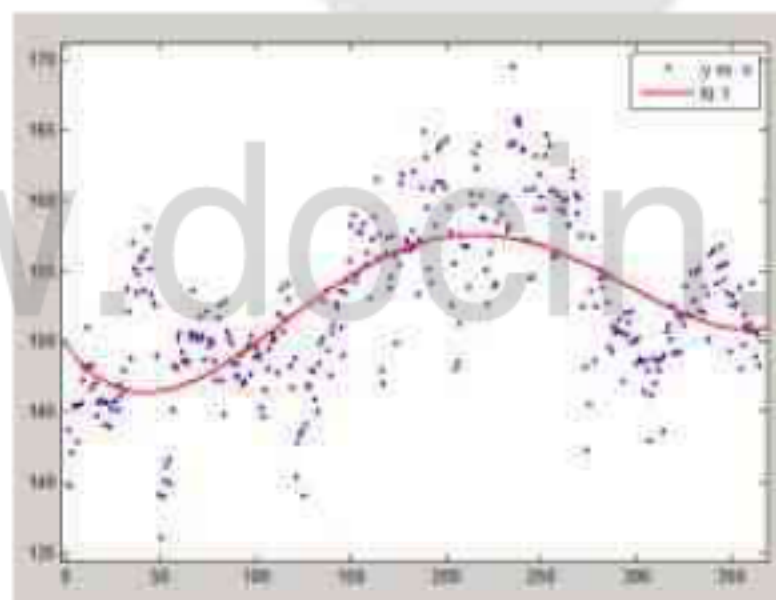


图 2.7.5 四次拟合曲线

拟合函数的表达式为:

$$f(x) = 1.389e-008x^4 - 1.142e-005x^3 + 0.00282x^2 - 0.18x + 149.7。$$

计算 $f(x)$ 在 $x = 366, \dots, 396$ 处的函数值分别得到 07 年 1 月 31 天的日用水量为 150.8524, 150.8718, \dots , 152.9422。对它们求和得到 1 月份总用水量的估计值为 4700.4297 (万吨/日)。

(2) 用 2000-2006 年 1 月份每天的日用水量预测

由 2000-2006 年 1 月 1 日的日用水量得到 7 个插值节点, 利用拟合函数计算 $f(8)$ 即 07 年 1 月 1 日日用水量估计值。

因插值节点只有7个，若用6次拟合曲线刚好可过所有插值节点。图2.7.6显示的是六次拟合结果。

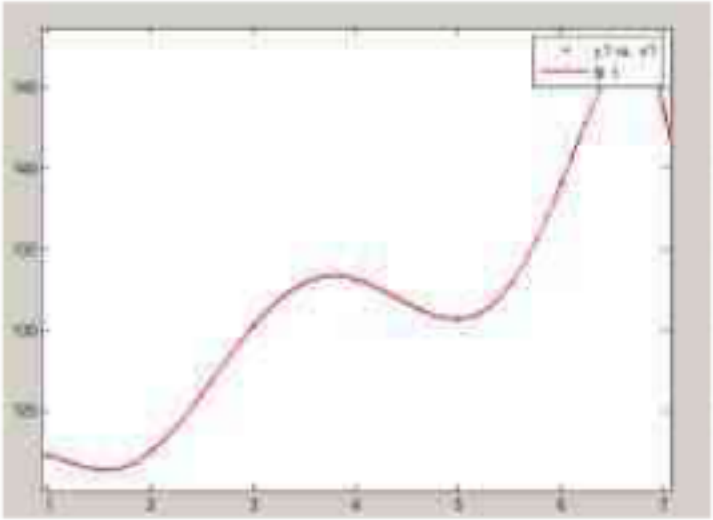


图 2.7.6 六次拟合曲线

但是计算 $f(8)$ 得07年1月1日日用水量为5.74456，产生了较大的误差。

通过1到5次拟合曲线的均方根误差指标值可知，用线性拟合结果最佳。拟合曲线表达式为 $f(x)=3.5088x +117.6164$ ，计算 $f(8)$ ，得07年1月1日日用水量估计值为145.6869。拟合曲线如图2.7.7所示

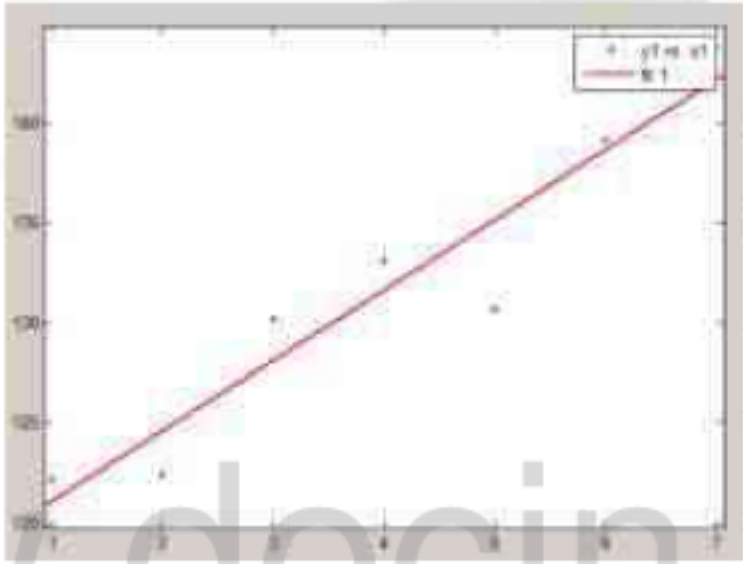


图 2.7.7 线性拟合曲线

同理用线性拟合得到第2到31天日用水量估计值为143.7937、…、148.3227。相加得到1月份用水总量的估计值为4654.8538（万吨/日）。

（3）用 2000-2006 年 1 月城市的总用水量预测

同理，尽管六次拟合曲线能完全拟合已知节点，但将对计算点产生较大误差，选用三次拟合，其拟合曲线如图 2.7.8 所示。

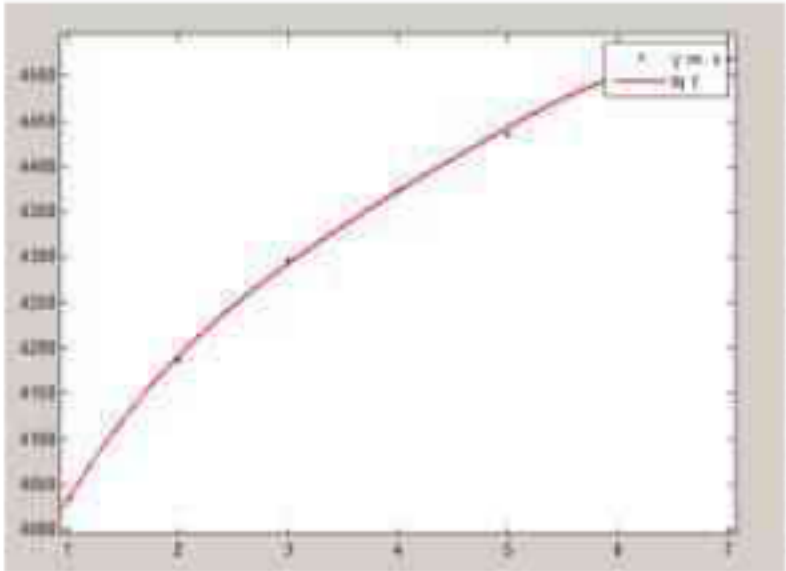


图 2.7.8 三次拟合曲线

拟合函数为: $f(x) = 0.7678x^3 - 20.47x^2 + 201.3x + 3854$, 计算得 $f(8)$ 的近似值为 4547.0724 (万吨/日)。

上述拟合方法得到的 07 年 1 月用水总量估计值如表 2.7.4 所示:

表 2.7.4 拟合法得到 07 年 1 月用水总量估计值 (万吨/日)

数据	06 年日用水量	00-06 年 1 月每天日用水量	00-06 每年 1 月用水总量
结果	4700.4297 (万吨/日)	4654.8538 (万吨/日)	4547.0724 (万吨/日)

习 题 2

1. 求经过下列已知点的最低次代数多项式

x	0	1.5	5.1
y	-1	4.25	35.21

2. 已知函数表如下

x	10	11	12	13
$\ln x$	2.3026	2.3979	2.4849	2.5649

试分别用线性插值与二次插值计算 $\ln(11.75)$ 的近似值, 并估计截断误差

3. 设 x_0, x_1, \dots, x_n 为任意给定的 $n+1$ 个互不相同的节点, 证明:

(1) 若 $f(x)$ 为不高于 n 次的多项式, 则 $f(x)$ 关于这组节点的 n 次插值多项式就是它自己;

(2) 若 $l_k(x)$ ($k=0, 1, \dots, n$) 是关于这组节点的 n 次基本插值多项式, 则有恒等式

$$\sum_{k=0}^n x_k^m l_k(x) \equiv x^m, (m=0, 1, \dots, n)$$

4. 已知函数表如下

x	0.0	0.2	0.4	0.6	0.8
e^x	1.0000	1.2214	1.4918	1.8221	2.2255

(1) 分别构造向前差分表与向后差分表;

(2) 分别用三点与四点前插公式计算 $e^{0.13}$ 的近似值, 并估计误差;

(3) 分别用三点与四点前插公式计算 $e^{0.72}$ 的近似值, 并估计误差;

(4) 构造差商表, 并分别用三点与四点牛顿基本插值公式计算 $e^{0.12}$ 的近似值.

5. 设 $f(x)$ 为 n 次多项式, 试证明当 $k \leq n$ 时差商 $f[x, x_0, x_1, \dots, x_k]$ (其中 x_0, x_1, \dots, x_k 互异) 为 $n-k$ 次多项式, 而当 $k > n$ 时其值恒为零.

6. 今要在区间 $[-4, 4]$ 上构造 $f(x) = e^x$ 在等距节点下的函数表. 问怎样选取函数表的步长, 才能保证用二次插值求 e^x 的近似值时, 其截断误差不超过 10^{-6}

7. 对于给定的插值条件

x	0	1	2	3
y	0	0	0	0

试分别求满足下列边界条件的三次样条插值函数:

(1) $S''(0) = 1, S''(3) = 0$

(2) $S'(0) = 1, S'(3) = 0$

8. 已知一组实验数据如下:

x	2	4	6	8
y	2	11	28	40

试用最小二乘法求一次和二次拟合多项式, 并分别算出均方误差与最大偏差。

9. 试根据下面五组测试数据, 用最小二乘法求出一个经验公式, 并计算均方误差。

x_i	293	313	343	363	383
y_i	28.98	30.2	32.9	35.9	38.8

10. 用最小二乘法求一形如 $W = Ct^\lambda$ 的经验公式 (其中 C 和 λ 为待定系数), 使之与下列数据相拟合。

t_i	1	2	4	8	16	32	64
W_i	4.22	4.02	3.85	3.59	3.44	3.02	2.59

11. 用最小二乘法求一形如 $y = a + bx^2$ 的多项式, 使之与下列数据相拟合。

X_i	19	25	31	38	44
Y_i	19.0	32.3	49.0	73.3	97.8

12. 利用正交多项式，对第 9 题给出的实验数据，拟合一条二次曲线。

利用 SPSS 软件对以上数据进行二次最优拟合，拟合优度分别达到 99.8%、99.1%，两个厂的拟合图形分别如图 2.1.1 和图 2.1.2. 根据以上处理后的数据，利用 MATLAB 软件产生了矩阵，然后进行二次拟合，得到了精度为 10 的两水厂一月份供水量 g_1, g_2 （吨）与时间 x （年）的变化函数表达式：

一水厂月供水量 g_1 函数关系式：

$$g_1 = -6.9173.41536x^2 + 1040417.918x + 23982938.23$$

二水厂月供水量 g_2 函数关系式：

$$g_2 = -44109.84679x^2 + 664616.9561x + 14816050.48$$

