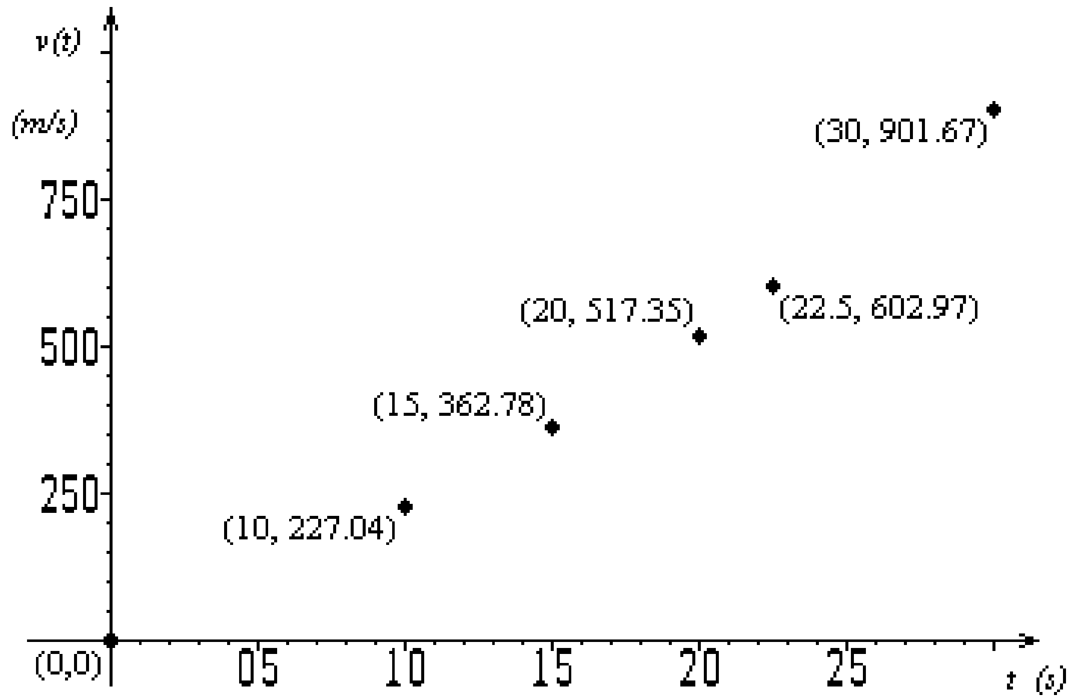
插值和拟合

浙江大学控制学院

举例

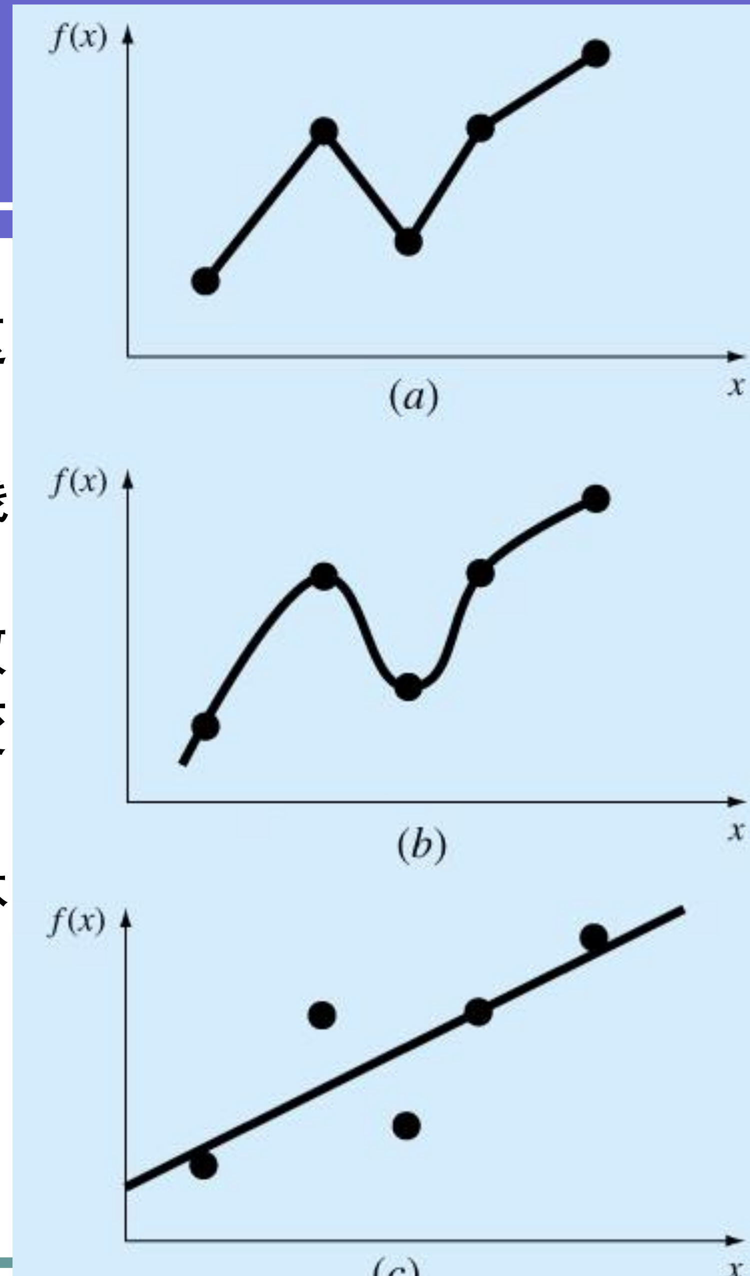
t(s)	$v(t)$(m/s)
0	0
10	227.04
15	362.78
20	517.35
22.5	602.97
30	901.67

- 已知火箭在几个不同时刻的速度如下表所示，求t=16时的速度？



插值和拟合

- (a) 用直线段连接这些点
  - 数据点分布接近线性或彼此距离很近
- (b) 用曲线描述了数据所具有的局部变化趋势
- (c) 体现了数据总体的趋势



线性插值

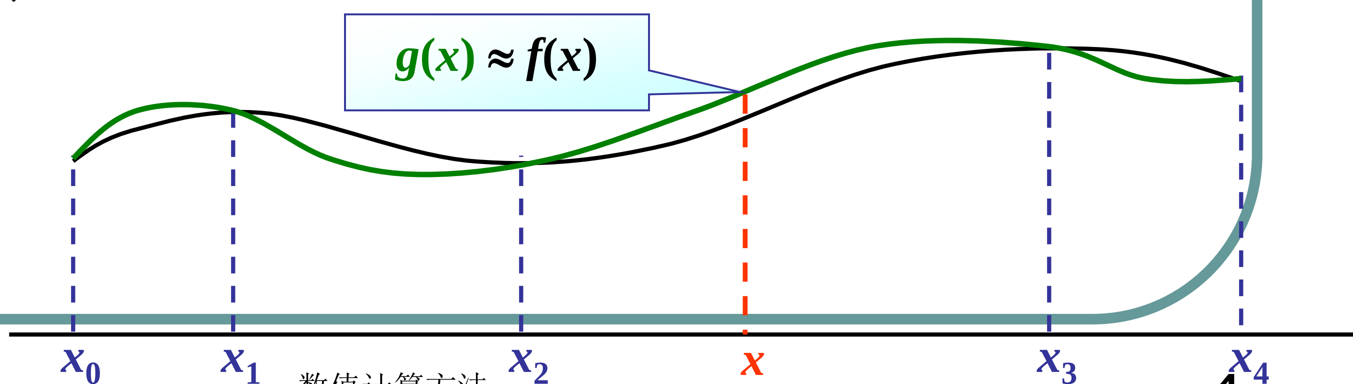
非线性插值

最小二乘回归

插值和拟合——概述

- 问题：
  - 有的函数虽有表达式，但**较复杂**，也可用简单的函数$g(x)$来逼近它
  - 某些变量之间的函数关系$y=f(x)$存在，但**没有$f(x)$的解析式**，$y=f(x)$以函数表格或曲线形式给出
- 要求：
  - 根据函数表推算该函数**在某些点上的函数值**。
  - **解决与该函数有关的一些问题**，如分析函数的性态，研究$y=f(x)$的变化规律，求导数、积分、零点与极值点等。

x	x_0	x_1	…	x_n
$f(x)$	y_0	y_1	…	y_n



插值和拟合——概述

- 用某个简单函数在满足一定条件下在某个范围内近似代替另一个较为复杂或者解析表达式未给出的函数，以便于简化对后者的各种计算或揭示后者的某些性质。

- 数学函数的逼近问题
  - 高精度
  - 快速计算
- 建立实验数据的数学模型
  - 只要求适度的精度

- 插值
- 拟合
  - 不要求过所有的点（可以消除误差影响）
  - 尽可能表现数据的趋势，靠近这些点

本章内容

- 插值
  - 插值问题
  - Lagrange插值多项式
  - Newton插值多项式
  - 分段低次插值
  - 样条插值
- 拟合
  - 最小二乘方法

插值问题

被插函数　插值区间　插值节点

- 插值
  - 函数$y=f(x)$ 在区间$[a,b]$上连续， $x_0,x_1,\ldots,x_n$是$[a,b]$上的$n+1$个互异的点，且在这些点处函数值$y_i=f(x_i)$已知
  - Φ为给定的某一函数类（性质优良、便于计算）
  - $g(x_i)=y_i \quad (i=0,1,\ldots,n)$

插值函数类　插值条件　插值函数　多项式插值

- 插值函数类
  - 代数多项式
  - 有理函数
  - 三角函数

$$g_n(x)=\sum_{i=0}^{n}a_ix^i$$

$$\Phi^{n,m}(x)=\frac{P_n(x)}{Q_m(x)}=\frac{p_0+p_1x+\cdots+p_nx^n}{q_0+q_1x+\cdots+q_mx^m}$$

- 基本问题
  - 是否存在唯一
  - 如何构造
  - 误差估计

插值问题

插值函数类

代数多项式：选取多项式$P_n(x)$作为插值函数，$P_n(x)$称为插值多项式。

$$P_n(x)=\sum_{i=0}^{n}a_ix^i$$

有理函数：选取有理函数（多项式的商）作为插值函数。

$$\Phi^{n,m}(x)=\frac{P_n(x)}{Q_m(x)}=\frac{p_0+p_1x+\cdots+p_nx^n}{q_0+q_1x+\cdots+q_mx^m}$$

有理插值可使区间内插值误差分布较为均匀，特别适用于某些被插函数具有无穷大间断点的附近，这种情况下若用多项式逼近效果很差。

三角函数：选取正弦和余弦等三角函数作为插值函数。

=> 选择不同的函数类作为插值函数逼近$f(x)$，其效果是不同的，因此需要根据实际问题选择合适的插值函数。

插值函数存在唯一定理

设 $g(x)=a_0\varphi_0(x)+\cdots+a_n\varphi_n(x)$

则 $g(x_i)=f(x_i)=a_0\varphi_0(x_i)+\cdots+a_n\varphi_n(x_i)$

$$\begin{bmatrix}\varphi_0(x_0) & \cdots & \varphi_n(x_0)\\ \vdots & \ddots & \vdots\\ \varphi_0(x_n) & \cdots & \varphi_n(x_n)\end{bmatrix}\begin{bmatrix}a_0\\ \vdots\\ a_n\end{bmatrix}=\begin{bmatrix}f(x_0)\\ \vdots\\ f(x_n)\end{bmatrix}$$

特点：

- 基函数个数与点个数相同
- 与插值节点的次序无关
- 与原函数$f(x)$无关

则 $\begin{bmatrix}a_0\\ \vdots\\ a_n\end{bmatrix}$ 有唯一解 $\Leftrightarrow$ $\begin{vmatrix}\varphi_0(x_0) & \cdots & \varphi_n(x_0)\\ \vdots & \ddots & \vdots\\ \varphi_0(x_n) & \cdots & \varphi_n(x_n)\end{vmatrix}\neq 0$

插值函数存在唯一定理

定理：$\{x_i\}_{i=0}^{n}$ 为$n+1$个节点，$\Phi = span\{\varphi_0, \varphi_1, \cdots \varphi_n\}$

$n+1$维空间，则插值函数存在唯一，当且仅当

$$\begin{vmatrix} \varphi_0(x_0) & \cdots & \varphi_n(x_0) \\ \vdots & \ddots & \vdots \\ \varphi_0(x_n) & \cdots & \varphi_n(x_n) \end{vmatrix} \neq 0$$

插值多项式存在唯一性

$$\Phi = P_n(x) = span\{1, x, x^2, \cdots x^n\}$$

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

n+1阶范德蒙(Vandermonde)行列式

- 若插值节点互异，则存在唯一的多项式$P_n(x)$使

$f(x_i)=P_n(x_i)$　　　　$i=0,1,\ldots,n$

通过求解线性方程组的方式确定插值多项式的系数，工作量大，不实用。

本章内容

- 插值
  - 插值问题
  - **Lagrange插值多项式**
  - Newton插值多项式
  - 分段低次插值
  - 样条插值
- 拟合
  - 最小二乘方法

插值多项式

$$P_n(x)=\sum_{i=0}^{n}a_i l_i(x)$$

不大于n次的多项式，插值基函数

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

x	x_0	x_1	…	x_k	…	x_n
$l_k(x)$	0	0	…	1	…	0

- 可设 $$l_k(x)=\alpha_k(x-x_0)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)$$
- 则 $$\alpha_k=\frac{1}{(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}$$

$$l_k(x)=\frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}=\prod_{\substack{i=0\\ i\neq k}}^{n}\frac{(x-x_i)}{(x_k-x_i)}$$

n次基本插值多项式或n次插值基函数

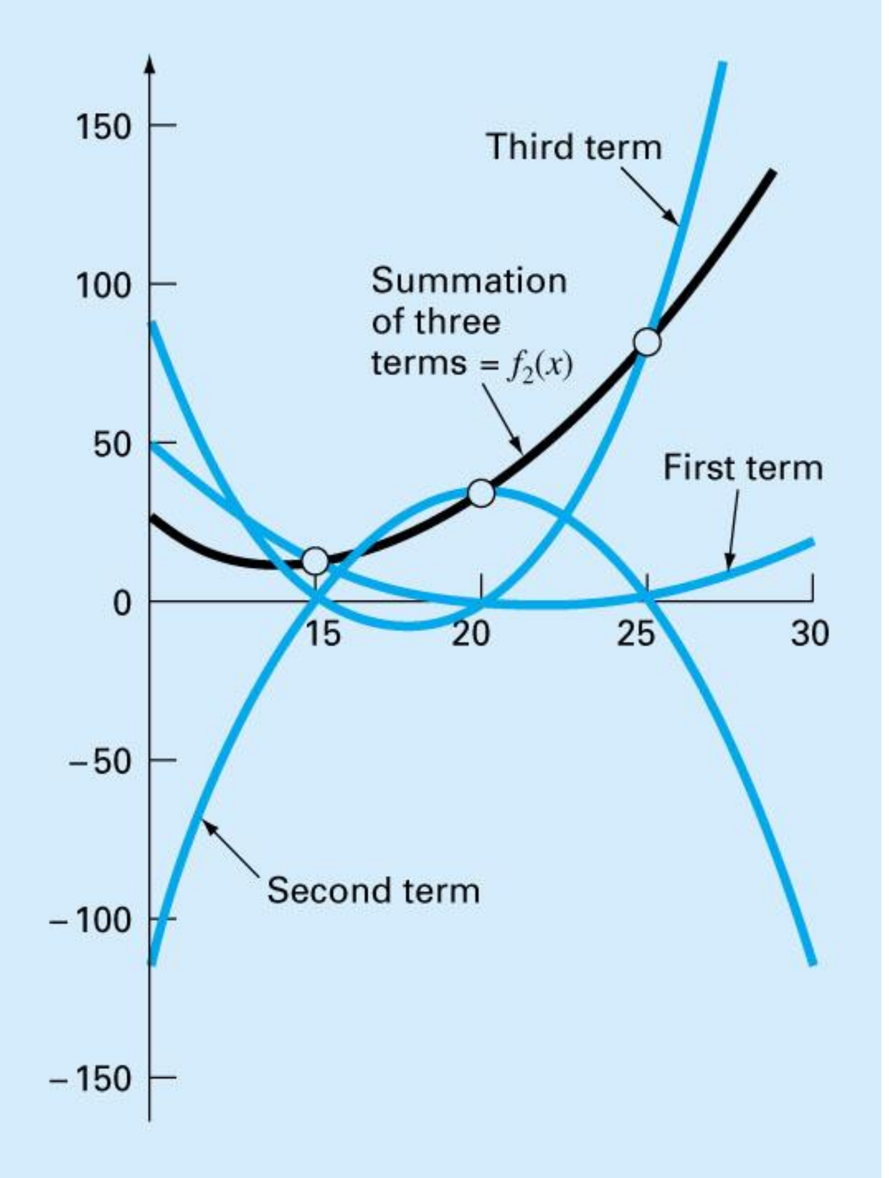
Lagrange插值多项式：取$l_i(x)$为基函数，令$a_i=y_i$

$$L_n(x)=y_0l_0(x)+y_1l_1(x)+\cdots+y_nl_n(x)=\sum_{i=0}^{n}l_i(x)y_i$$

Lagrange插值多项式原理的直观描述

$$f_2(x) = y_0 l_0(x) + y_1 l_1(x) + y_2 l_2(x)$$

- 以三点为例
  - 每一项都经过其中一个数据点，而在另外两个数据点处为0
  - 三项之和必定是唯一正好经过三点的二阶插值多项式

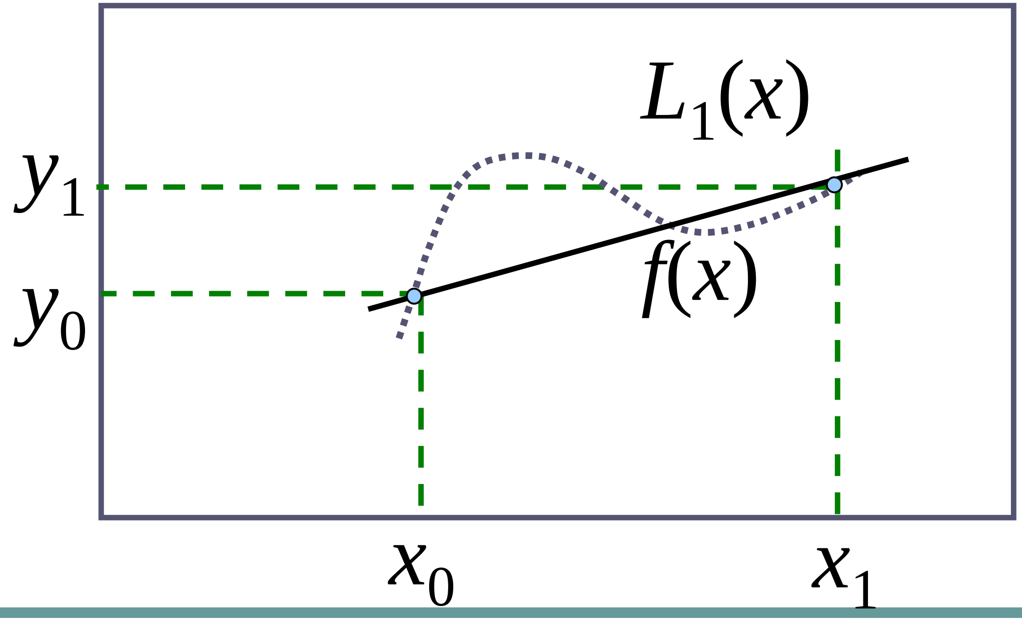


Lagrange插值多项式

- 线性插值（n=1）

$$l_0(x)=\frac{x-x_1}{x_0-x_1} \qquad l_1(x)=\frac{x-x_0}{x_1-x_0}$$

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



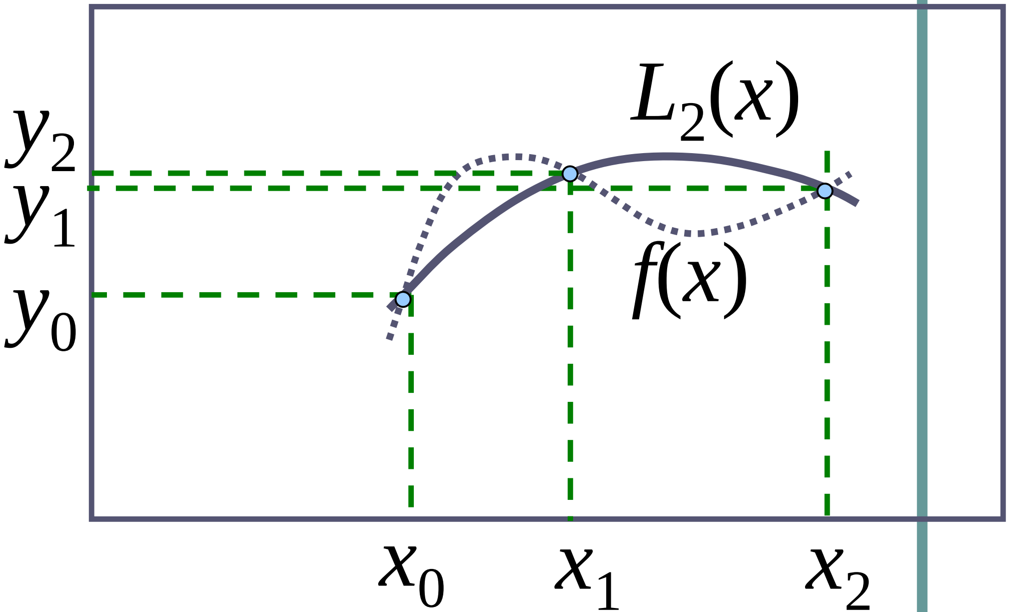
Lagrange插值多项式

- 二次（抛物线）插值（n=2）

$$l_0(x)=\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}$$

$$l_1(x)=\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}$$

$$l_2(x)=\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}$$



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

插值余项（截断误差）

- $R_n(x) = f(x) - P_n(x)$ 表示用$P_n(x)$表示$f(x)$时，在点x处产生的误差。
- 定理：
  - 若$f(x)$在区间$[a,b]$上有直到$n+1$阶导数$f^{(n+1)}(x)$存在，$P_n(x)$为$f(x)$在$n+1$个节点x_i上的n次插值多项式，则$\forall x \in [a,b]$有

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{i=0}^{n}(x - x_i) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\omega_{n+1}(x)$$

$\xi \in (a,b)$, 且依赖于x

➢$f^{(n+1)}(\xi)$对余项的影响

➢在实际问题中，往往不知道$f(x)$的具体解析表达式，因此要估计$f^{(n+1)}(x)$的上界是不现实的。在这种情形下，可以采用**误差的事后估计法**。

$$\omega_{n+1}(x) = \prod_{i=0}^{n}(x - x_i)$$

➢当插值点x位于插值区间的节点附近时，插值误差较小，而距离节点较远处插值误差较大。

插值误差的事后估计

- $\{x_i\}_{i=0}^{n+1}$ 任取$n+1$个构造$P_n(x)$

$$i=0,\cdots,n \quad \Rightarrow \quad P_n(x)$$

$$i=1,\cdots,n+1 \quad \Rightarrow \quad \tilde{P}_n(x)$$

$$f(x)-P_n(x)=\frac{f^{(n+1)}(\xi_1)}{(n+1)!}(x-x_0)\cdots(x-x_n)$$

$$f(x)-\tilde{P}_n(x)=\frac{f^{(n+1)}(\xi_2)}{(n+1)!}(x-x_1)\cdots(x-x_{n+1})$$

- 假设

$$f^{(n+1)}(\xi_1)\approx f^{(n+1)}(\xi_2) \quad \Rightarrow$$

$$\frac{f(x)-P_n(x)}{f(x)-\tilde{P}_n(x)}\approx\frac{x-x_0}{x-x_{n+1}}$$

$$\Rightarrow f(x)\approx\frac{x-x_{n+1}}{x_0-x_{n+1}}P_n(x)+\frac{x-x_0}{x_{n+1}-x_0}\tilde{P}_n(x)$$

$$\Rightarrow f(x)-P_n(x)\approx\frac{x-x_0}{x_0-x_{n+1}}\left(P_n(x)-\tilde{P}_n(x)\right)$$

通过两个结果的偏差来估计插值误差

Lagrange插值多项式

x	$\ln x$
1	0
4	1.386294
6	1.791759

- 例：用线性插值估计ln2（实际值为0.6931472）
- 解：采用x_0=1到x_1=6进行线性插值

$$\ln 2 = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) = 0 + \frac{1.791759 - 0}{6 - 1}(2 - 1) = 0.3583519$$

$$\varepsilon_t = 48.3\%$$

误差估计：$R_1(x) = \frac{f''(\xi)}{2!}\prod_{i=0}^{n}(x - x_i) = \frac{2}{\xi^2}$ $1 < \xi < 6$ ⇨ $0.05555 < R_1(x) < 2$

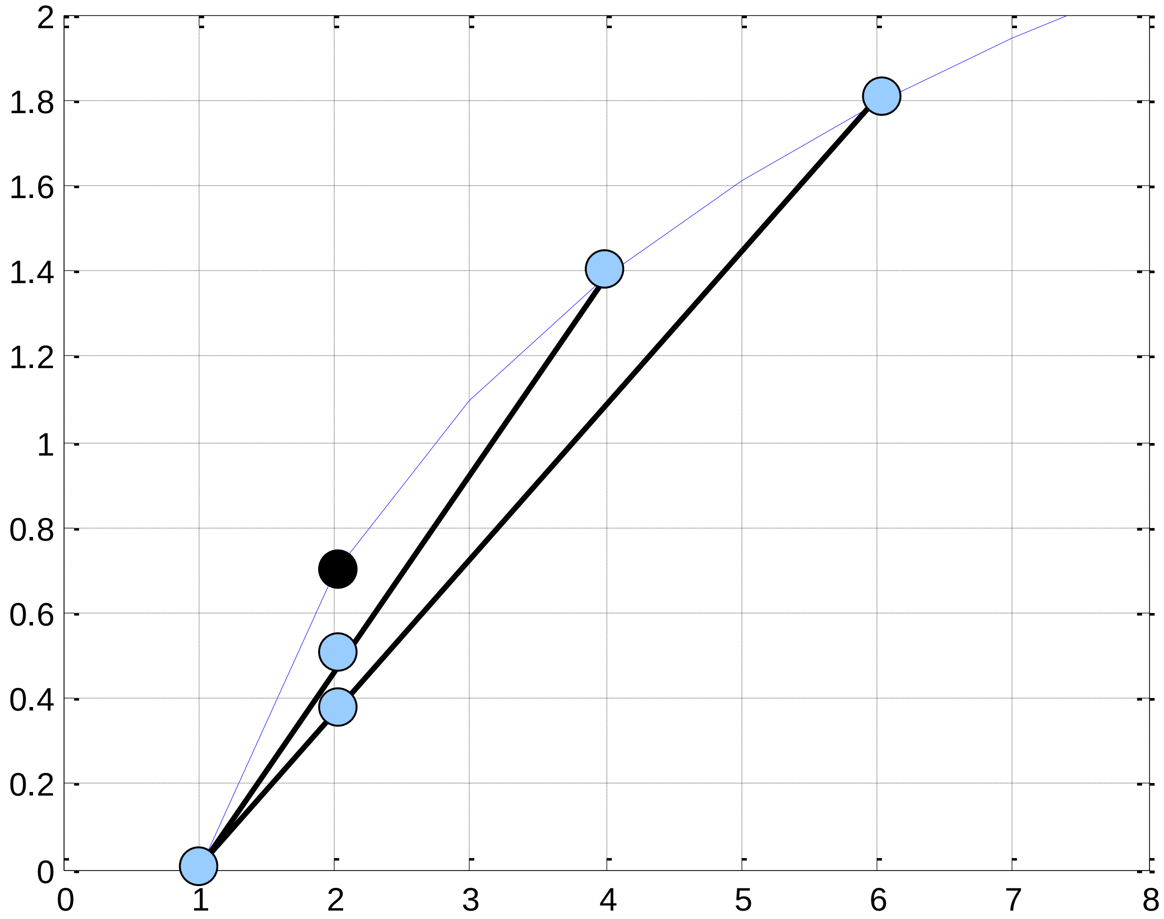
采用x_0=1到x_1=4进行线性插值

$$\ln 2 = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) = 0 + \frac{1.386294 - 0}{4 - 1}(2 - 1) = 0.4620981$$

$$\varepsilon_t = 33.3\%$$

- 采用较小的区间可以得到更好的估计值

Lagrange插值多项式——线性插值



Lagrange插值多项式

x	$\ln x$
1	0
4	1.386294
6	1.791759

- 例：用二次插值估计ln2（实际值为0.6931472）
- 解：

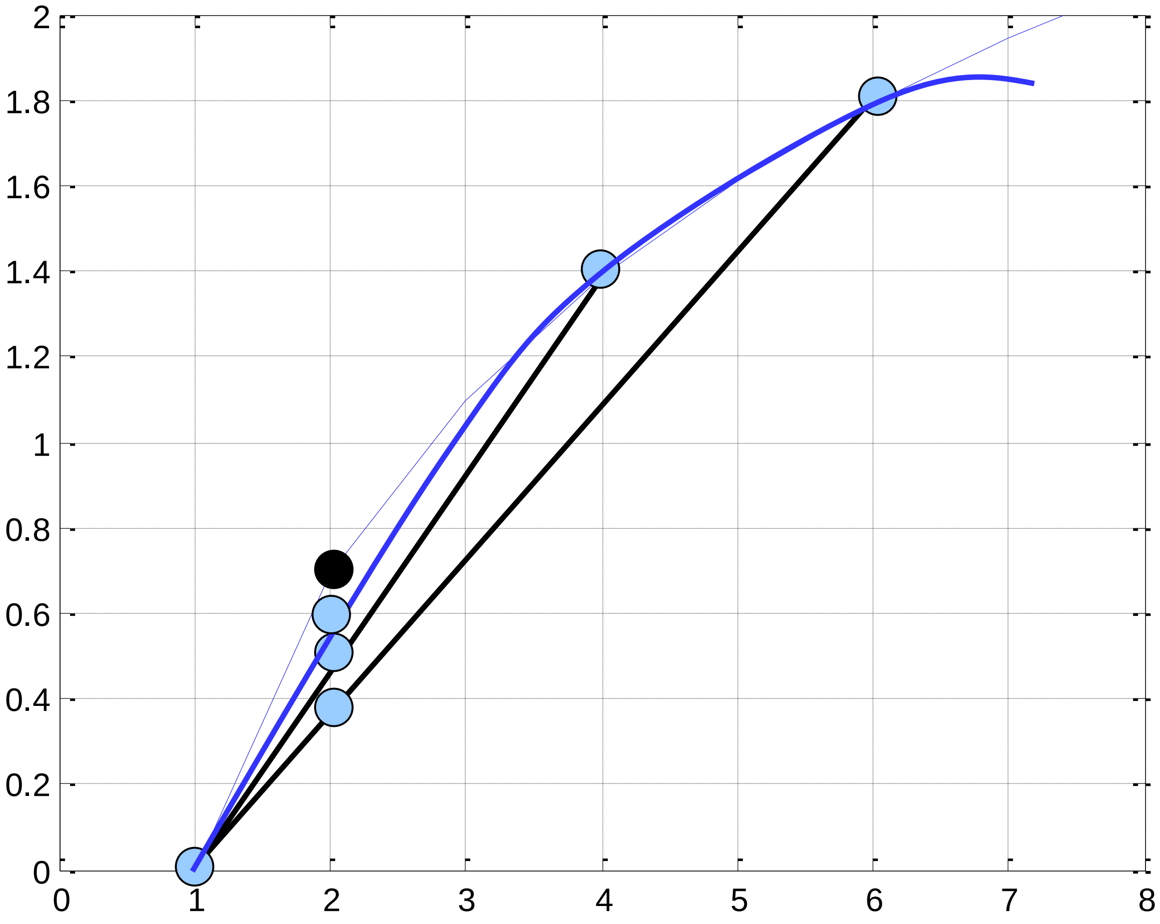
$$\ln(2) = y_0 \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} + y_1 \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} + y_2 \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}$$

$$= 0 + 1.386294 \frac{(2-1)(2-6)}{(4-1)(4-6)} + 1.791759 \frac{(2-1)(2-4)}{(6-1)(6-4)} = 0.5658442$$

$$\varepsilon_t = 18.4\%$$

- 高次插值通常优于低次插值

Lagrange插值多项式



内插和外推

x	$\ln x$
1	0
2	0.6931472
4	1.386294
6	1.791759

- 内插（ interpolation ）
  - 插值点x位于插值区间内，即$a<x<b$
- 外推（ extrapolation ）
  - 插值点x位于插值区间外，即$x<a$或$x>b$ ，但又较接近于插值区间端点时
- 例：分别采用x_0=1到x_1=6和x_0=1到x_1=2进行线性插值估计ln4

$$\ln 4 = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) = 0 + \frac{1.791759 - 0}{6 - 1}(4 - 1) = 1.075055$$

$$\varepsilon_t = 22.4\%$$

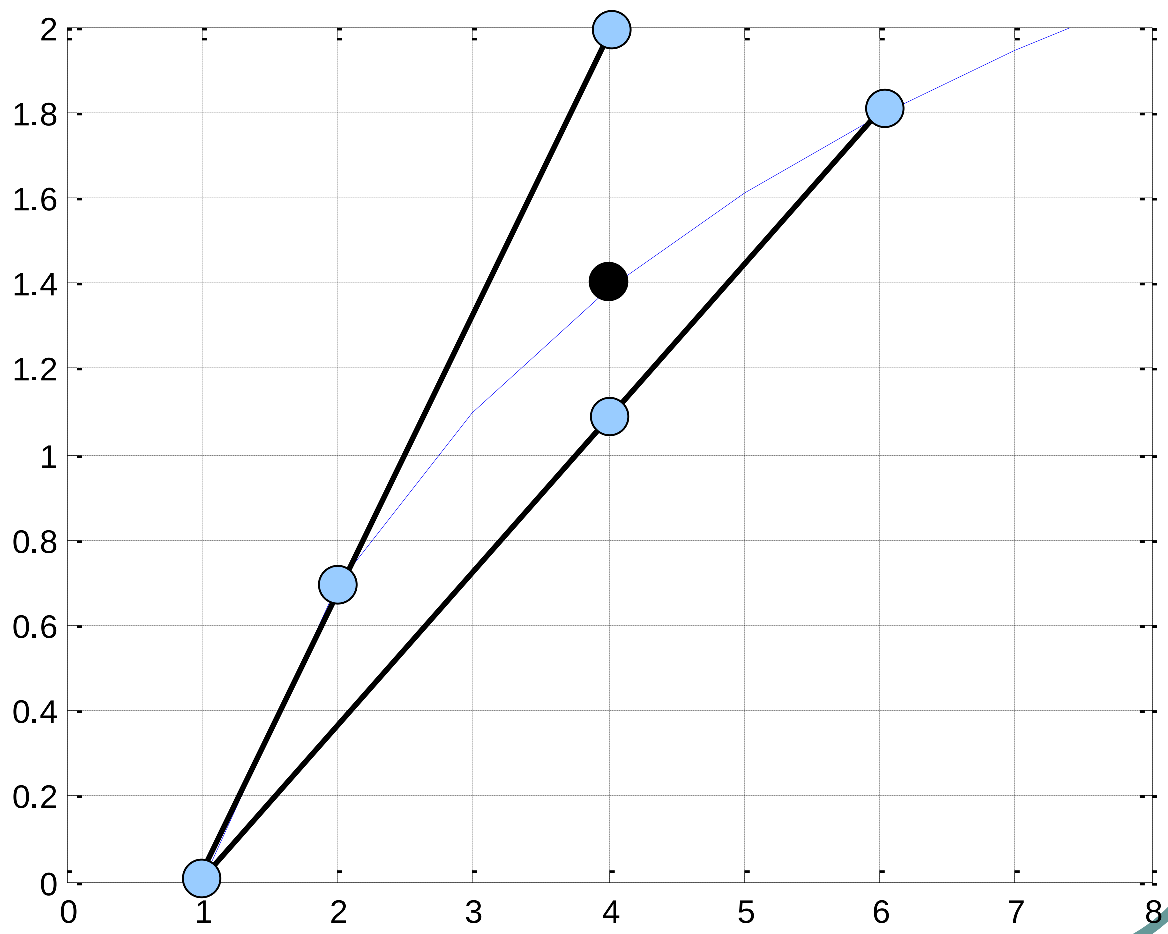
$$\ln 4 = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) = 0 + \frac{0.6931472 - 0}{2 - 1}(4 - 1) = 2.079442$$

$$\varepsilon_t = 50\%$$

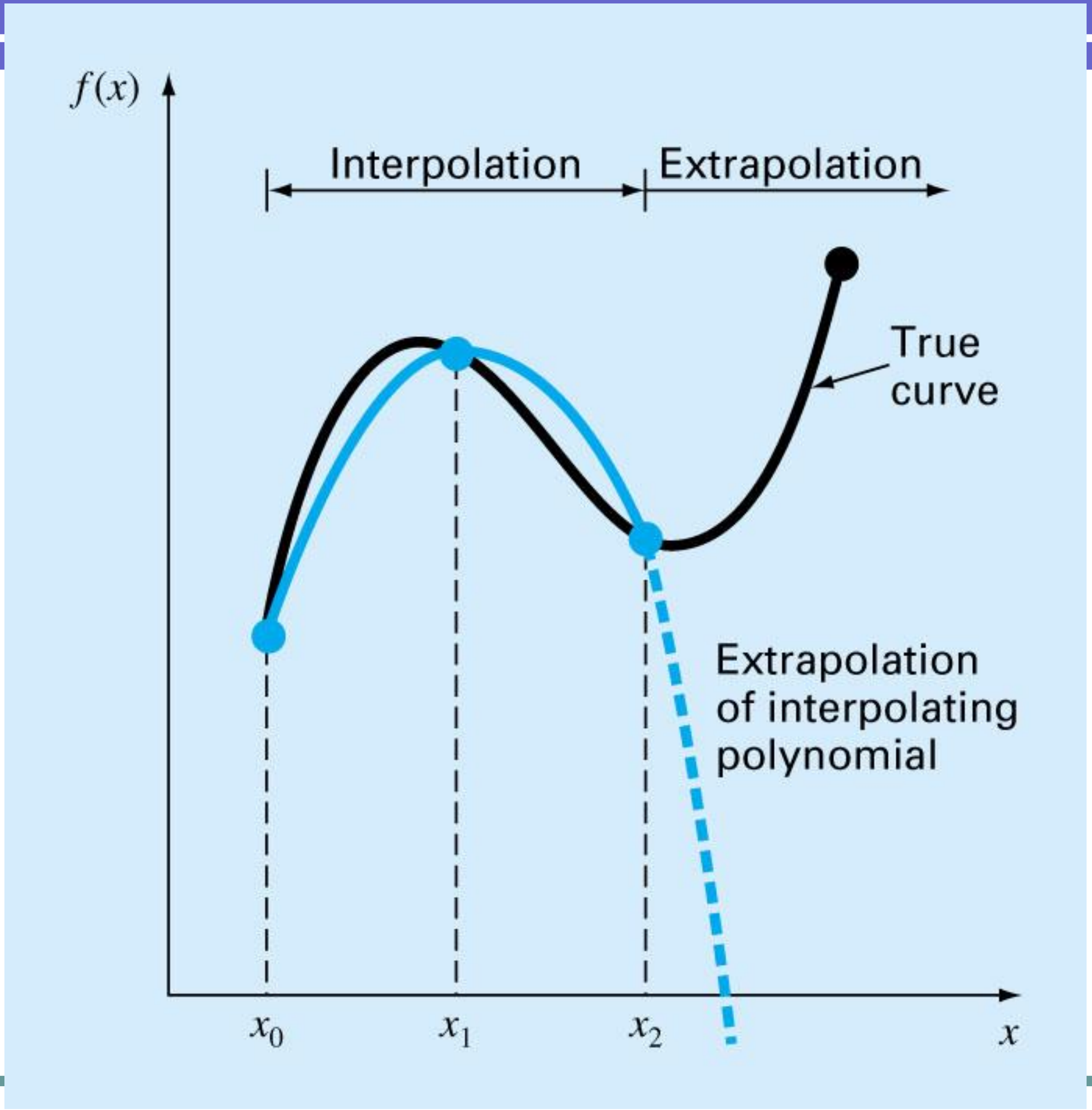
内插和外推

$$R_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{i=0}^{n}(x-x_i)$$

➢采用外推法时插值点x不宜位于插值区间之外的远处。



内插和外推



$f(x)$
Interpolation
Extrapolation
True
curve
Extrapolation
of interpolating
polynomial
x_0
x_1
x_2
x

插值误差

- **插值多项式仅与已知数据有关，与$f(x)$原来形式无关，但余项与 $f(x)$密切相关**
- **若$f(x)$本身是一个不超过n次多项式，则$P_n(x)=f(x)$**
- **对多项式插值，增加阶数不一定能提高精度**
  - 一般推荐3~4个节点的插值，5~6个节点也可行。除非对估计的误差控制得相当严格，一般极少使用更多的节点。
- **插值点x不能位于插值区间之外的远处**
  - 如果数据表中包含的点多于插值所需的数目，则在每次插值时，就必须选取表中合适位置的点。

Lagrange 插值的缺点

- 无承袭性
  - 增加一个节点，所有的基函数都要重新计算

$$N_{n+1}(x) = N_n(x) + q_{n+1}(x)$$

当插值节点在区间$[a,b]$上，适当加密，其误差在多数情况下会小些，但节点过多，不仅高次$L_n(x)$计算量大，精度也不一定好。$\lim_{n\to\infty} L_n(x) = f(x)$ 一般不成立。

本章内容

- 插值
  - 插值问题
  - Lagrange插值多项式
  - **Newton插值多项式**
  - 分段低次插值
  - 样条插值
- 拟合
  - 最小二乘方法

Newton插值多项式

- 承袭性 $N_{n+1}(x)=N_n(x)+q_{n+1}(x) \quad \in P^{n+1}$

$\{x_0,x_1\cdots,x_{n+1}\}$ $\{x_0,x_1\cdots,x_n\}$

$$N_n(x_i)=N_{n+1}(x_i)=f(x_i) \quad i=0,\cdots n \Rightarrow q_{n+1}(x)=a_{n+1}(x-x_0)\cdots(x-x_n)$$

$$N_n(x)=N_{n-1}(x)+q_n(x) \Rightarrow q_n(x)=a_n(x-x_0)\cdots(x-x_{n-1})$$

$$\Rightarrow N_n(x)=a_0+a_1(x-x_0)+\cdots+a_n(x-x_0)\cdots(x-x_{n-1})$$

$$\begin{cases} N_n(x_0)=a_0=f(x_0) \\ N_n(x_1)=a_0+a_1(x_1-x_0)=f(x_1) \\ N_n(x_2)=a_0+a_1(x_2-x_0)+a_2(x_2-x_0)(x_2-x_1)=f(x_2) \\ \quad\vdots \\ N_n(x_n)=a_0+a_1(x_n-x_0)+\cdots+a_n(x_n-x_0)\cdots(x_n-x_{n-1})=f(x_n) \end{cases}$$

Newton插值多项式

连接x_0,x_1两点直线的斜率

$$a_0 = f(x_0)$$

$$a_1 = \frac{f(x_1)-f(x_0)}{x_1-x_0}$$

$$a_2 = \frac{1}{x_2-x_1}\left(\frac{f(x_2)-f(x_0)}{x_2-x_0}-a_1\right) \Rightarrow a_2 = \frac{\frac{f(x_2)-f(x_1)}{x_2-x_1}-\frac{f(x_1)-f(x_0)}{x_1-x_0}}{x_2-x_0}$$

$$a_3 = \frac{1}{x_3-x_2}\left(\left(\frac{f(x_3)-f(x_0)}{x_3-x_0}-a_1\right)\frac{1}{x_1-x_0}-a_2\right)$$

$$\vdots$$

有限差商（finite divided difference）

n阶差商可表示为函数值的线性组合

$$f[x_0,\cdots,x_n]=\sum_{i=0}^{n}\frac{f(x_i)}{\prod_{\substack{j=0\\ j\neq i}}^{n}(x_i-x_j)}$$

- 设实数$x_0,x_1,\ldots,x_n$互异，$f(x)$关于点的
  - 零阶差商 $f[x_i]=f(x_i)$
  - 一阶差商 $f[x_i,x_j]=\dfrac{f(x_i)-f(x_j)}{x_i-x_j}$
  - 二阶差商 $f[x_i,x_j,x_k]=\dfrac{f[x_i,x_j]-f[x_j,x_k]}{x_i-x_k}$
  - k阶差商 $f[x_i,x_{i+1},\cdots,x_{i+k}]=\dfrac{f[x_{i+1},\cdots,x_{i+k}]-f[x_i,\cdots,x_{i+k-1}]}{x_{i+k}-x_i}$

对称性 $f[x_0,\cdots,x_k]=f[x_{i_0},\cdots,x_{i_k}]$ $i_0,\cdots,i_k$是$0,\cdots,k$的任意排列

Newton插值多项式

$$a_0 = f(x_0)$$
$$a_1 = f[x_0, x_1]$$
$$a_2 = f[x_0, x_1, x_2]$$
$$\vdots$$
$$a_n = f[x_0, x_1, \cdots, x_n]$$

牛顿基本插值多项式

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

差商表

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	$f(x_3)$	$f[x_2,x_3]$	$f[x_1,x_2,x_3]$	$f[x_0,x_1,x_2,x_3]$

利用差商表的最外一行，构造插值多项式

Newton插值多项式——另一种理解

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\ldots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+R_n$$

$$f[x,x_0]=\frac{f(x)-f(x_0)}{x-x_0} \Rightarrow f(x)=f(x_0)+\left(x-x_0\right)f[x,x_0]$$

代入

$$f[x,x_0,x_1]=\frac{f[x,x_0]-f[x_0,x_1]}{x-x_1} \Rightarrow f[x,x_0]=f[x_0,x_1]+\left(x-x_1\right)f[x,x_0,x_1]$$

$$\begin{aligned} f(x)&=f(x_0)+\left(x-x_0\right)\left(f[x_0,x_1]+\left(x-x_1\right)f[x,x_0,x_1]\right)\\ &=f(x_0)+\left(x-x_0\right)f[x_0,x_1]+\left(x-x_0\right)\left(x-x_1\right)f[x,x_0,x_1]\end{aligned}$$

为了提高精度，可以继续增加节点

Newton插值多项式——例

- 例：估计ln2 （实际值为0.6931472）

k	x_k	$f(x_k)$	一阶差商	二阶差商	三阶差商
0	1	0			
1	4	1.386294	0.4620981		
2	6	1.791759	0.2027326	-0.05187311	
3	5	1.609438	0.1823216	-0.02041100	0.007865529

$$N_2(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1)$$

$$\ln(2) = 0.5658442$$

$$N_3(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + a_3(x - x_0)(x - x_1)(x - x_2)$$

$$\ln(2) = 0.6287686$$

$$\varepsilon_t = 9.3\%$$

拉格朗日插值多项式与牛顿插值多项式是同一个n次多项式，因此，两者的误差分析也是一致的。

Newton插值多项式的误差估计

$$R_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{i=0}^{n}(x-x_i)=\frac{f^{(n+1)}(\xi)}{(n+1)!}\omega_{n+1}(x)$$

- 增加一个节点a

$$N_{n+1}(x)=N_n(x)+f[x_0,\cdots,x_n,a](x-x_0)\cdots(x-x_n)$$

$$N_{n+1}(a)=f(a)$$

$$f(a)-N_n(a)=f[x_0,\cdots,x_n,a](a-x_0)\cdots(a-x_n)$$

$$f[x_0,\cdots,x_n,a]=\frac{f^{n+1}(\xi)}{(n+1)!}$$

➢如果有额外的数据点$f(x_{n+1})$，近似计算

➢n阶多项式的误差估计等于$n+1$阶和n阶估计值之差

$$R_n(x)=N_{n+1}(x)-N_n(x) \Rightarrow N_{n+1}(x)=N_n(x)+R_n(x)$$

选择节点的重要性：应尽可能集中或靠近未知点

Newton插值多项式的误差估计

- 例：估计二阶插值多项式$f_2(2)$的误差
- 可以利用额外的数据点$f(x_3)=f(5)=1.609438$

$$R_2(x)=f[x_0,x_1,x_2,x_3](x-x_0)(x-x_1)(x-x_2)$$

$$R_2(x)=0.0629242$$

- 真实误差为

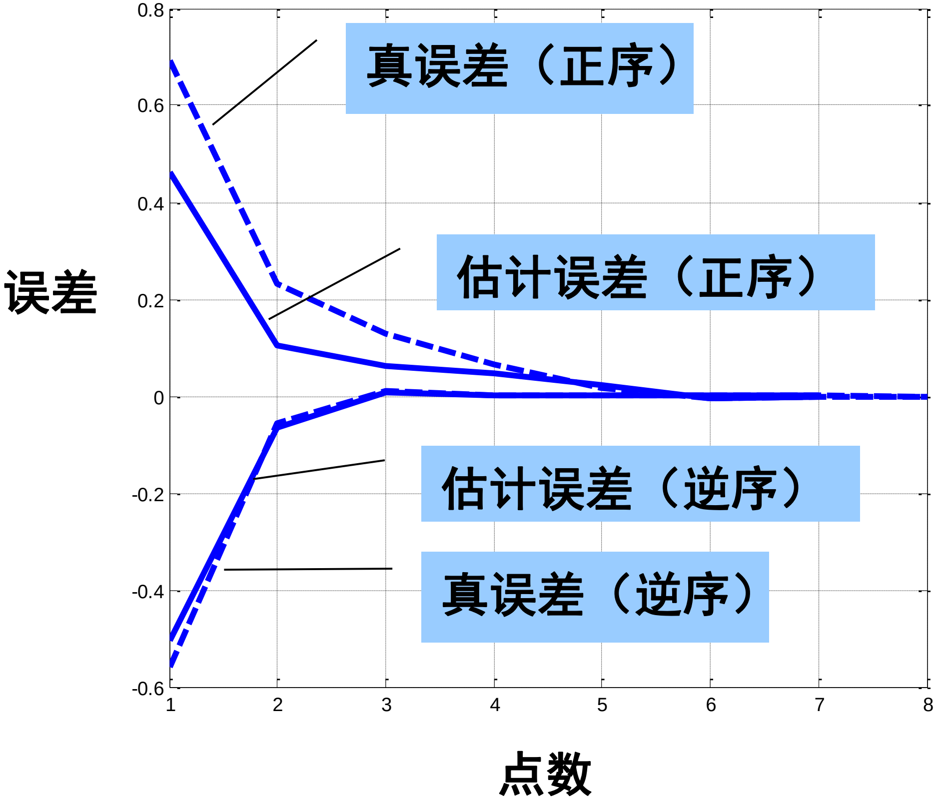
$$R_t(x)=0.6931472-0.5658442=0.1283030$$

$n+1$阶和n阶估计值之差小于真实误差，因此采用两次迭代结果之差无法作为迭代的停止准则。实际上，高次多项式插值往往会发散。

Newton插值多项式的误差估计

- 例：利用下表中的数据估计ln(2)的值。

x	lnx
1	0
4	1.3862944
6	1.7917595
5	1.6094379
3	1.0986123
1.5	0.4054651
2.5	0.9162907
3.5	1.2527630



等距节点的插值

- 数据在空间上间距相等
  - $x_0<\ldots<x_n$，其中$x_i-x_{i-1}=h$， $x_i=x_0+ih$，h称作**步长**
  - **有限差分**

$\Delta f(x)=f(x+h)-f(x)$ **一阶向前差分**

$\nabla f(x)=f(x)-f(x-h)$ **一阶向后差分**

$\delta f(x)=f(x+h/2)-f(x-h/2)$ **一阶中心差分**

Δ称为前差算子，∇称为后差算子，δ称为中差算子

n阶前差 $\Delta^n f(x)=\Delta[\Delta^{n-1}f(x)]=\Delta^{n-1}f(x+h)-\Delta^{n-1}f(x)$

n阶后差 $\nabla f(x)=\nabla[\nabla^{n-1}f(x)]=\nabla^{n-1}f(x)-\nabla^{n-1}f(x-h)$

n阶中差 $\delta f(x)=\delta[\delta^{n-1}f(x)]=\delta^{n-1}f(x+h/2)-\delta^{n-1}f(x-h/2)$

规定 $\Delta^0 f(x)=f(x)$，$\nabla^0 f(x)=f(x)$，$\delta^0 f(x)=f(x)$

等距节点的插值

- 差商与有限差分的关系
  - 前差

$$f[x_0,x_1,\cdots,x_n]=\frac{\Delta^n f(x_0)}{n!h^n}$$

  - 后差

$$f[x_0,x_1,\cdots,x_n]=\frac{\nabla^n f(x_n)}{n!h^n}$$

差分表

x_k	$f(x_k)$	一阶差分	二阶差分	三阶差分
x_0	$f(x_0)$			
x_1	$f(x_1)$	$\Delta f(x_0)$		
x_2	$f(x_2)$	$\Delta f(x_1)$	$\Delta^2 f(x_0)$	
x_3	$f(x_3)$	$\Delta f(x_2)$	$\Delta^2 f(x_1)$	$\Delta^3 f(x_0)$

等距节点的插值——牛顿向前插值公式

x在节点x_0的附近($x>x_0$)，将节点按$x_0,\ldots,x_n$的顺序排列

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

$$f[x_0,x_1,\cdots,x_n]=\frac{\Delta^n f(x_0)}{n!h^n}$$

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

令$x=x_0+th$

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

- 常用来计算表头x_0附近的函数值，$x\in[x_0,x_1]$
- 余项

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

$$\xi\in(x_0,x_n)$$

等距节点的插值——牛顿向后插值公式

x在节点x_n的附近($x<x_n$)，将节点按$x_n,\ldots,x_0$的顺序排列

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

$$f[x_n,x_{n-1},\cdots,x_0]=f[x_0,x_1,\cdots,x_n]=\frac{\nabla^n f(x_n)}{n!h^n}$$

$$N_n(x)=f(x_n)+\frac{x-x_n}{h}\nabla f(x_n)+\frac{(x-x_n)(x-x_{n-1})}{2!h^2}\nabla^2 f(x_n)+\cdots+\frac{(x-x_n)\cdots(x-x_1)}{n!h^n}\nabla^n f(x_n)$$

令$x=x_n+th$（一般$t<0$）

$$N_n(x)=f(x_n+th)=f(x_n)+t\nabla f(x_n)+\frac{t(t+1)}{2}\nabla^2 f(x_n)+\cdots+\frac{t(t+1)\cdots(t+n-1)}{n!}\nabla^n f(x_n)$$

常用来计算表尾x_n附近的函数值，即$x\in[x_n,x_{n-1}]$

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

Newton前插(后插)公式——例

- 例：用牛顿插值公式计算sin$x$在 x=0.12 及 x=0.58 处的函数近似值

x_k	$f(x_k)$	Δf	$\Delta^2 f$	$\Delta^3 f$	$\Delta^4 f$
0.1	0.09983				
0.2	0.19867	0.09884			
0.3	0.29552	0.09685	-0.00199		
0.4	0.38942	0.09390	-0.00295	-0.00096	
0.5	0.47943	0.09001	-0.00389	-0.00094	0.00002
0.6	0.56464	0.08521	-0.00480	-0.00091	0.00003

三阶差分接近于常数，四阶差分接近于零

- x=0.12时，采用前插公式，$t=(x-x_0)/h=0.2$
  - 线性插值：

$$\sin(0.12)\approx N_1(0.12)=0.09983+0.2\times0.09884=0.11960$$

Newton前插(后插)公式——例

- 二次插值：

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

- 三次插值：

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

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

区间内单调，取上界

Newton前插(后插)公式——例

- x=0.58时，采用后插公式，$t=(x-x_5)/h$=-0.2
  - 三次插值：

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

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

Hermite插值多项式

- 构造插值函数除了函数值的条件以外，还需要一定的连续性条件，如一阶导数值等

定义：满足$\{x_i, f(x_i)\}_{i=0}^n$，和$\{(x_i, f^k(x_i), k=1,\cdots,k_i),\}_{i=0}^n$条件的插值称为Hermite插值

- 以所有k_i=1为例，即每个点上还要满足一阶导数条件

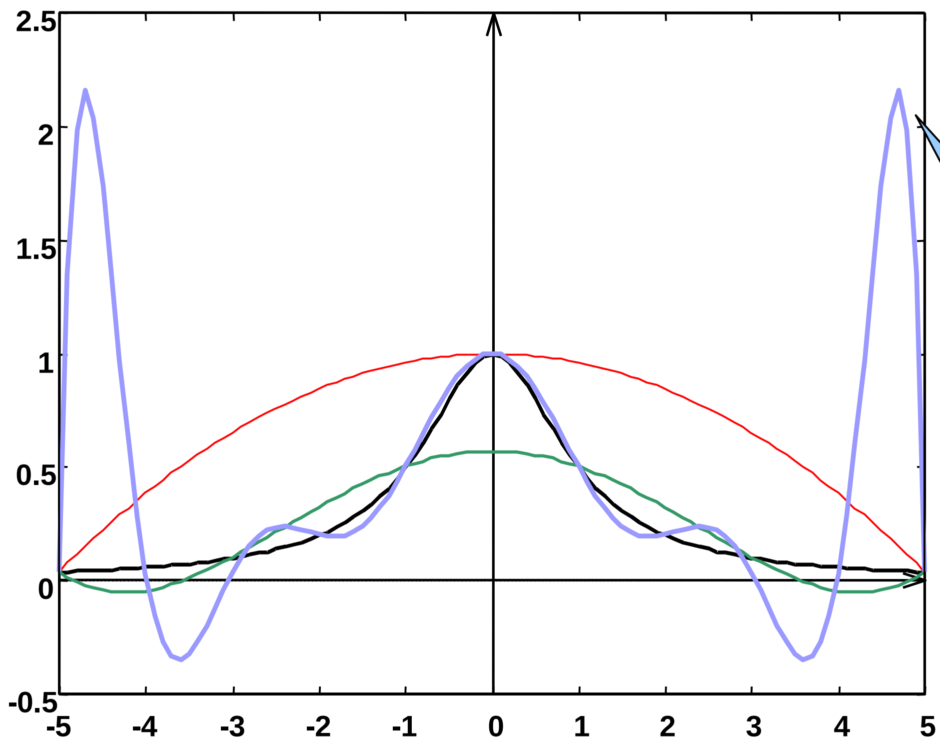
$$\{x_i, f(x_i)\}_{i=0}^n, \text{和}\{x_i, f'(x_i)\}_{i=0}^n$$

- 在设计交通工具的外形时，参照海豚的标本上已知点及已知点的导数，做插值在计算机上模拟海豚的外形制成飞机、汽车等外形。

Runge 现象(是否次数越高越好?)

- 例：在[–5, 5]上考察 $f(x)=\dfrac{1}{1+x^2}$ 的$L_n(x)$。取

$$x_i=-5+\frac{10}{n}i \quad (i=0,\ldots,n)$$



n 越大，端点附近抖动越大，称为Runge 现象

Runge 现象(是否次数越高越好?)

适当提高插值多项式的次数，可能提高计算结果的准确程度。然而次数越高意味着参加插值节点数越多，

(1) 计算量大；

(2) 插值函数曲线在部分区间上（两端）发生激烈振荡，插值多项式**截断误差/计算余项**偏大。

龙格/Runge 现象说明加密节点并不能保证所得到的插值多项式能更好地逼近$f(x)$，**高次插值效果不一定比低次插值好**。

插值的误差

- 截断误差是插值的收敛性问题，Runge现象
- 舍入误差（在插值计算过程中可能被扩散或放大）是插值的稳定性问题。

➢ 随插值多项式次数的提高，计算误差的影响也会增大。假设给出$f(x)$的一组等距节点上x_i的函数值y_i 有初始误差变为$\tilde{y}_i$，误差限为ε，则拉格朗日插值多项式$L_n(x)$的计算值 $\tilde{L}_n(x)$ 也会有误差：

$$\left|L_n(x)-\tilde{L}_n(x)\right|=\left|\sum_{i=0}^{n} l_i(x)\,y_i-\sum_{i=0}^{n} l_i(x)\,\tilde{y}_i\right|$$

$$\le\sum_{i=0}^{n}\left|l_i(x)\right|\left|y_i-\tilde{y}_i\right|\le\sum_{i=0}^{n}\left|l_i(x)\right|\varepsilon=\varepsilon\cdot\sum_{i=0}^{n}\left|l_i(x)\right|$$

n很大时，此项>>1，插值多项式的计算误差大，即当n增大时，拉格朗日插值法不稳定。

- 为避免龙格现象和不稳定，常限定$n<7$。当n较大时不采用高次多项式插值，而采用分段低次插值、样条插值或用次数较低的最小二乘逼近。

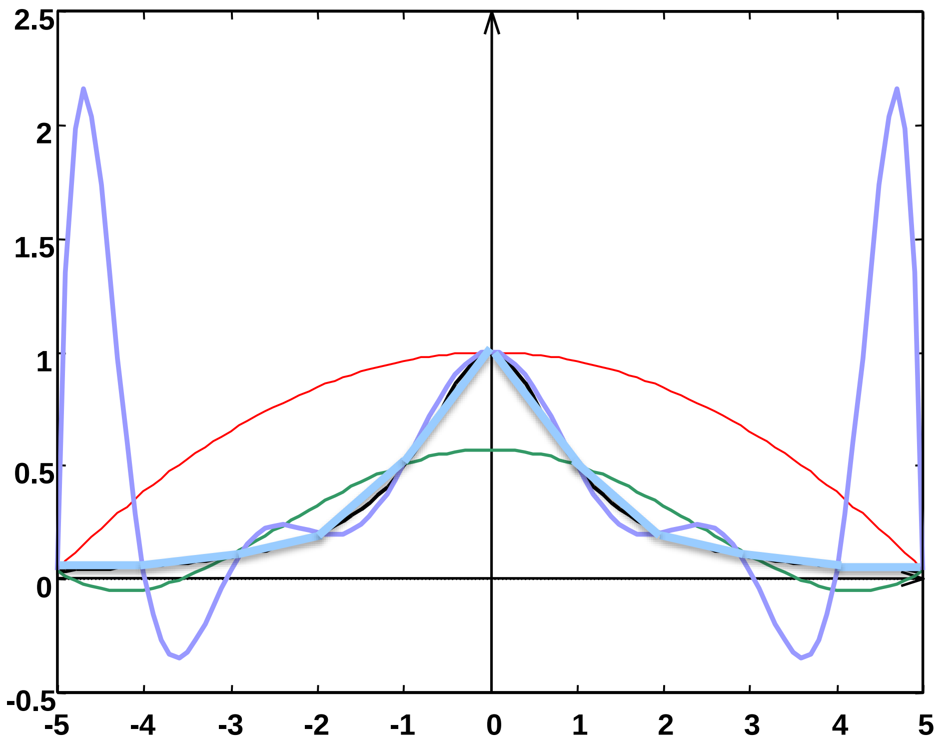
本章内容

- 插值
  - 插值问题
  - Lagrange插值多项式
  - Newton插值多项式
  - 分段低次插值
  - 样条插值
- 拟合
  - 最小二乘方法

分段低次插值

- 把整个插值区间分成若干个小区间，在每个小区间上，进行低次插值
- **设$a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b$，则节点把$[a,b]$分成$n$个小区间**
- 当插值点x在第i个小区间$[x_{i-1},x_i]$上时，采用一次线性插值公式，称为分段线性插值。也称为折线插值，即通过曲线$n+1$个点(x_i,y_i)的折线去近似替代曲线
- 当求插值点x的插值（即求$f(x)$的近似值）时，可选取距点x最近的三个插值节点x_{i-1}，x_i，x_{i+1}，在小区间$[x_{i-1},x_i,x_{i+1}]$上使用二次插值公式，称为分段抛物插值或分段二次插值

分段低次插值



分段低次插值

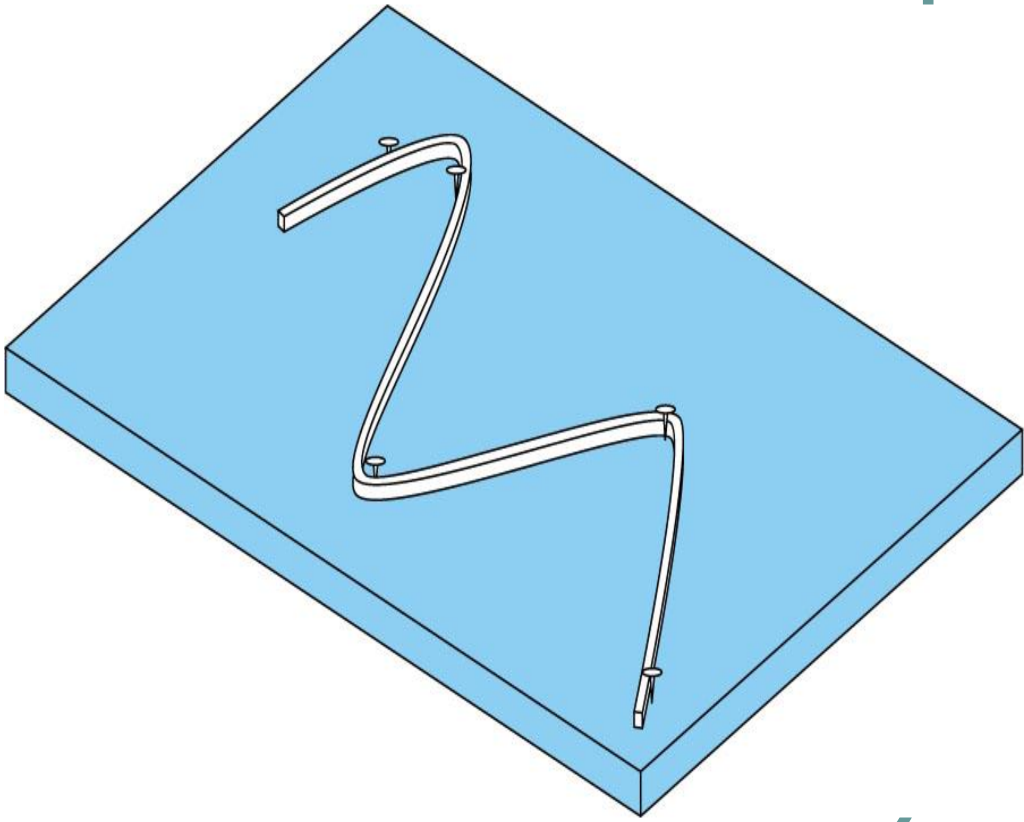
- 优点
  - 分段低次插值函数公式简单，只要区间充分小，就能保证误差要求
  - 它的一个显著优点是它的局部性质，如果修改了某个节点x_i的值，仅在相邻的两个区间$[x_{i-1}, x_i]$，$[x_i, x_{i+1}]$受到影响
- 缺点
  - 不能保证节点处插值函数的导数连续，因而不能满足某些工程上技术上曲线光滑性的要求

本章内容

- 插值
  - 插值问题
  - Lagrange插值多项式
  - Newton插值多项式
  - 分段低次插值
  - **样条插值**
- 拟合
  - 最小二乘方法

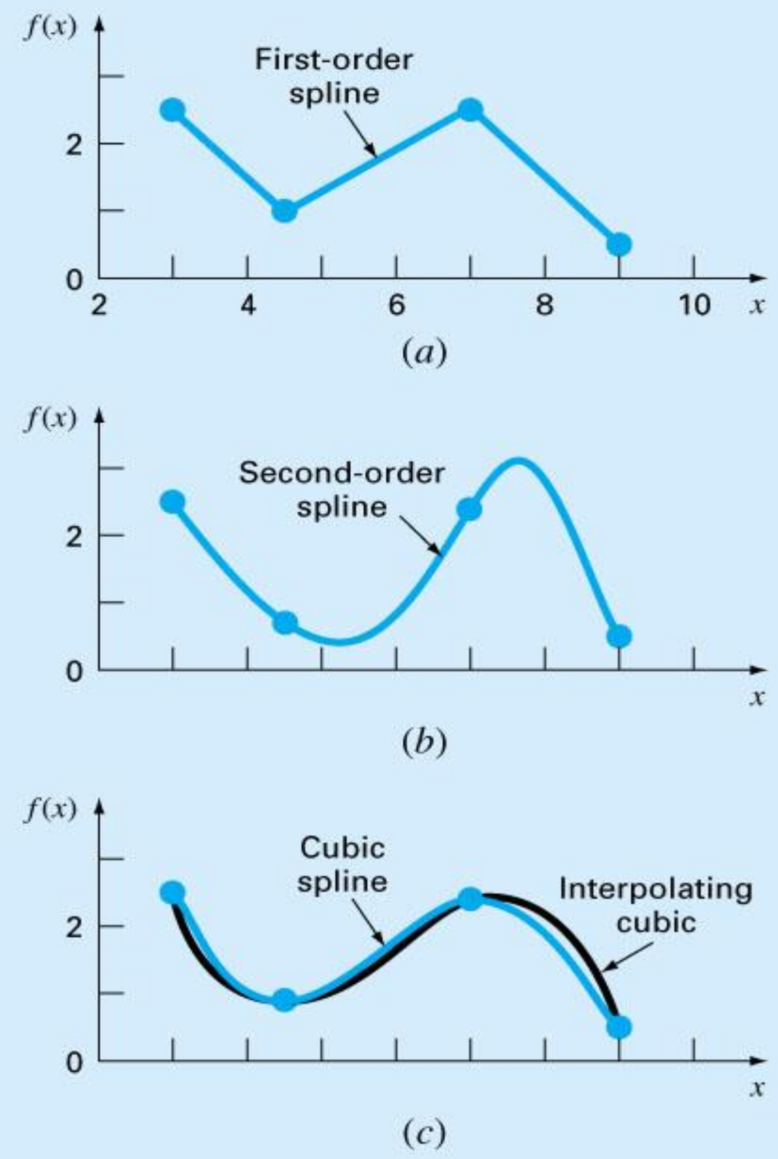
样条曲线

- 样条本来是指在飞机和船舶制造过程中，为了描绘光滑的外形所用的一种工具。
- 它是一种富有弹性的细长条，使用时用压铁固定在一些给定的节点上，其他地方任由它自由变化，然后依样画下的光滑曲线成为样条曲线。



样条插值

- 线性（一次）样条
  - 分段线性插值
  - 节点处一阶导数不连续（不光滑）
- 二次样条
  - 在节点处一阶导数连续
  - 每两个相邻节点组成的区间中推导一个二阶多项式
- 三次样条
  - 节点处具有连续的一阶、二阶导数
  - 每两个相邻节点组成的区间中推导一个三阶多项式
  - 三阶导数和高阶导数可能不连续



样条插值

- 定义
  - 设$f(x)$ 是区间$[a,b]$上的一个二次(m-1次)连续可微函数。在区间上给定一组基点：$a=x_0<\ldots<x_n=b$，设函数：

$$S(x)=\begin{cases} S_1(x), & x\in[x_0,x_1] \\ \ldots & \\ S_i(x), & x\in[x_{i-1},x_i] \\ \ldots & \\ S_n(x), & x\in[x_{n-1},x_n] \end{cases}$$

满足条件：

(1) $S(x)$在区间$[a,b]$上存在二阶(m-1阶)的连续导数；

(2) 每个子区间$[x_{i-1},x_i]$上$S_i(x)$都是一个不高于3次(m次)的多项式；

(3) 满足插值条件$S(x_i)=f(x_i)$，$i=0,\ldots,n$。

则称$S(x)$为函数$f(x)$关于基点x_0、x_1、…、x_n的三次(m次)样条插值函数，简称三次(m次)样条。

二次样条

$$f_i(x)=a_ix^2+b_ix+c_i \quad x\in[x_{i-1},x_i] \quad i=1,2,\cdots n$$

- 相邻多项式在内部节点处函数值相等

$$a_{i-1}x_{i-1}{}^2+b_{i-1}x_{i-1}+c_{i-1}=f(x_{i-1}) \quad a_ix_{i-1}{}^2+b_ix_{i-1}+c_i=f(x_{i-1})$$

$$i=2,3,\cdots n$$

- 第一和最后一个函数通过端点

$$a_1x_0{}^2+b_1x_0+c_1=f(x_0)$$

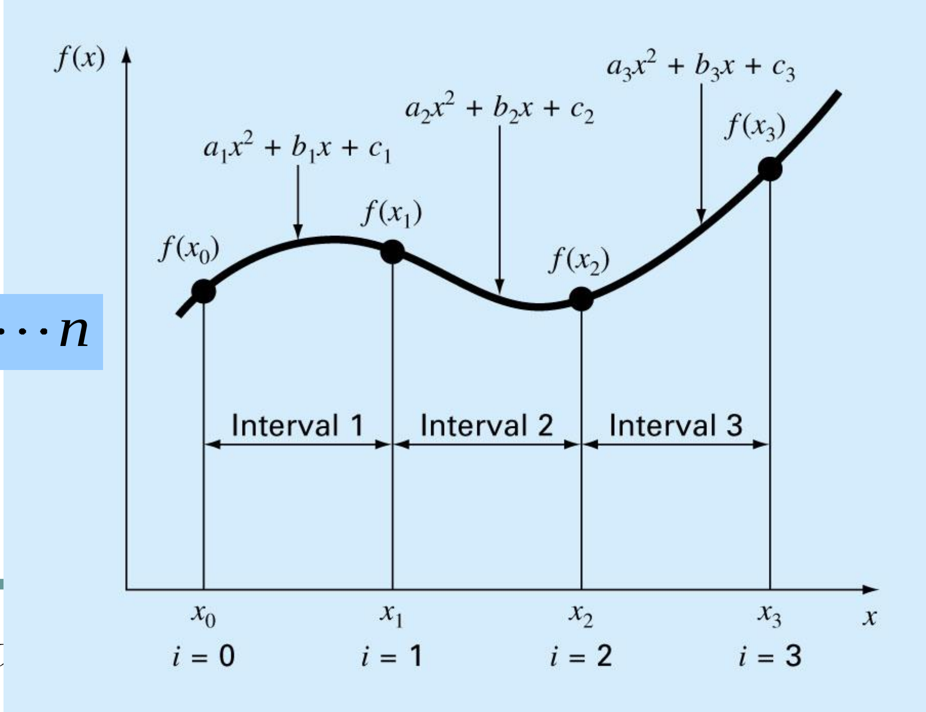
$$a_nx_n{}^2+b_nx_n+c_n=f(x_n)$$

- 内部节点的一阶导数相等

$$2a_{i-1}x_{i-1}+b_{i-1}=2a_ix_{i-1}+b_i \quad i=2,3,\cdots n$$

- 在第一个节点处二阶导数为0

$a_1=0$ ——连接前两点的为一条直线



二次样条插值——例

- 例：4个数据点拟合一个二次样条函数
- 解：$a_1 = 0$

x	$f(x)$
3.0	2.5
4.5	1.0
7.0	2.5
9.0	0.5

$$
\begin{bmatrix}
4.5 & 1 & & & & & & \\
 & & 20.25 & 4.5 & 1 & & & \\
 & & 49 & 7 & 1 & & & \\
 & & & & & 49 & 7 & 1 \\
3 & 1 & & & & & & \\
 & & & & & 81 & 9 & 1 \\
1 & & -9 & -1 & & & & \\
 & & 14 & 1 & & -14 & -1 &
\end{bmatrix}
\begin{bmatrix} b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \\ a_3 \\ b_3 \\ c_3 \end{bmatrix}
=
\begin{bmatrix} 1 \\ 1 \\ 2.5 \\ 2.5 \\ 2.5 \\ 0.5 \\ 0 \\ 0 \end{bmatrix}
$$

$$
\begin{aligned}
f_1(x) &= -x + 5.5 & 3.0 \le x \le 4.5 \\
f_2(x) &= 0.64x^2 - 6.76x + 18.46 & 4.5 \le x \le 7.0 \\
f_3(x) &= -1.6x^2 + 24.6x - 91.3 & 7.0 \le x \le 9.0
\end{aligned}
$$

- 连接前两点的为直线
- 最后一段看上去凸得有点高

三次样条插值函数

$S_i(x)=a_ix^3+b_ix^2+c_ix+d_i, x\in[x_{i-1},x_i]$ $i=1,\cdots,n$

- 有4n个未知数
- 满足条件

$$\left.\begin{array}{ll} S(x_i-0)=S(x_i+0) & i=1,\cdots,n-1 \\ S'(x_i-0)=S'(x_i+0) & i=1,\cdots,n-1 \\ S''(x_i-0)=S''(x_i+0) & i=1,\cdots,n-1 \\ S(x_i)=f(x_i) & i=0,\cdots,n \end{array}\right\} \text{4}n\text{-2个}$$

- 边界条件

1. 假设在两端点的导数$f'(a)=f_0'$和$f'(b)=f_n'$为已知， $S'(x_0)=f_0'$, $S'(x_n)=f_n'$
2. 假设在两端点的二阶导数$f''(a)=f_0''$和$f''(b)=f_n''$为已知， $S''(x_0)=f_0''$, $S''(x_n)=f_n''$。若$S''(x_0)=S''(x_n)=0$，称为自然边界条件
3. 给定函数具有周期特性，$S^{(t)}(x_0+0)=S^{(t)}(x_n-0)$, $t=0,1,2$

三弯矩法

- 利用$S(x)$在节点x_i处的二阶导数值$M_i=S''(x_i)$（M_i称为$S(x)$的矩）表示$S_i(x)$。用$S'(x)$在内节点x_i上的连续性和边界条件来确定M_i。

$S(x)$是3次多项式

- $S''(x_i)$在每一个子区间上是线性函数

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

- 积分两次，并利用插值条件$S_i(x_{i-1})=f(x_{i-1}),S_i(x_i)=f(x_i)$可得

$$S_i(x)=\frac{1}{6h_i}[M_{i-1}(x_i-x)^3+M_i(x-x_{i-1})^3]+(f_{i-1}-\frac{M_{i-1}h_i^2}{6})\frac{x_i-x}{h_i}+(f_i-\frac{M_ih_i^2}{6})\frac{x-x_{i-1}}{h_i}$$

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

$n+1$个未知数，$n-1$个方程

- 由$S'(x_i+0)=S'(x_i-0)$

$$\mu_iM_{i-1}+2M_i+\lambda_iM_{i+1}=g_i\qquad i=1,2,\cdots,n-1$$

$$\mu_i=\frac{h_i}{h_i+h_{i+1}}\qquad \lambda_i=\frac{h_{i+1}}{h_i+h_{i+1}}=1-\mu_i\qquad g_i=\frac{6}{h_i+h_{i+1}}\left(f[x_i,x_{i+1}]-f[x_{i-1},x_i]\right)=6f[x_{i-1},x_i,x_{i+1}]$$

三弯矩法

边界条件1：

$S'(x_0)=f_0' \Rightarrow 2M_0+M_1=6([f[x_0,x_1]-f_0'])/h_1$

$S'(x_n)=f_n' \Rightarrow M_{n-1}+2M_n=6(f_n'-f[x_{n-1},x_n])/h_n$

边界条件

边界条件

$$\begin{bmatrix} 2 & \lambda_0 & & & & \\ \mu_1 & 2 & \lambda_1 & & & \\ & \mu_2 & 2 & \lambda_2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \mu_{n-1} & 2 & \lambda_{n-1} \\ & & & & \mu_n & 2 \end{bmatrix} \begin{bmatrix} M_0 \\ M_1 \\ \vdots \\ M_{n-1} \\ M_n \end{bmatrix} = \begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_{n-1} \\ g_n \end{bmatrix}$$

$\lambda_0=\mu_n=1$

$$g_0=6\frac{(f[x_0,x_1]-f_0')}{h_1}$$

$$g_n=6\frac{(f_n'-f[x_{n-1},x_n])}{h_n}$$

三弯矩法

- 边界条件2： $S''(x_0)=f_0''$, $S''(x_n)=f_n'' \Rightarrow M_0=f_0''\quad M_n=f_n''$

$$\begin{bmatrix} 2 & \lambda_0 & & & & \\ \mu_1 & 2 & \lambda_1 & & & \\ & \mu_2 & 2 & \lambda_2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \mu_{n-1} & 2 & \lambda_{n-1} \\ & & & & \mu_n & 2 \end{bmatrix} \begin{bmatrix} M_0 \\ M_1 \\ \vdots \\ M_{n-1} \\ M_n \end{bmatrix} = \begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_{n-1} \\ g_n \end{bmatrix}$$

$$\lambda_0=\mu_n=0$$

$$g_0=2f_0''$$

$$g_n=2f_n''$$

$$\Downarrow$$

$$\begin{bmatrix} 2 & \lambda_1 & & & \\ \mu_2 & 2 & \lambda_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \mu_{n-2} & 2 & \lambda_{n-2} \\ & & & \mu_{n-1} & 2 \end{bmatrix} \begin{bmatrix} M_1 \\ \vdots \\ M_{n-1} \end{bmatrix} = \begin{bmatrix} g_1-\mu_1 f_0'' \\ g_2 \\ \vdots \\ g_{n-2} \\ g_{n-1}-\lambda_{n-1} f_n'' \end{bmatrix}$$

三弯矩法

- 边界条件3：　$S^{(t)}(x_0+0) = S^{(t)}(x_n-0)$, $t=1,2$
  - $M_0 = M_n$　　　　$\lambda_n M_1 + \mu_n M_{n-1} + 2M_n = g_n$

$$\lambda_n = \frac{h_1}{h_n + h_1}$$

$$\lambda_n = 1 - \mu_n$$

$$g_n = 6\frac{\left(f[x_0, x_1] - f[x_{n-1}, x_n]\right)}{h_n + h_1}$$

$$\begin{bmatrix} 2 & \lambda_1 & & & & \mu_1 \\ \mu_2 & 2 & \lambda_2 & & & \\ & \mu_3 & 2 & \lambda_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \mu_{n-1} & 2 & \lambda_{n-1} \\ \lambda_n & & & & \mu_n & 2 \end{bmatrix} \begin{bmatrix} M_1 \\ \vdots \\ M_{n-1} \\ M_n \end{bmatrix} = \begin{bmatrix} g_1 \\ \vdots \\ g_{n-1} \\ g_n \end{bmatrix}$$

三次样条插值——例

x	$f(x)$	$f[x_{i-1},x_i]$	$f[x_{i-2},x_{i-1},x_i]$
3.0	2.5		
4.5	1.0	-1	
7.0	2.5	1.5/2.5	1/2.5
9.0	0.5	-1	-4/11.25

- 例：采用自然边界条件，

利用4个数据点拟合一个三次样条函数

- 解：

$$\begin{bmatrix} 2 & \lambda_1 \\ \mu_2 & 2 \end{bmatrix}\begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \qquad \mu_i = \frac{h_i}{h_i + h_{i+1}} \qquad \lambda_i = \frac{h_{i+1}}{h_i + h_{i+1}} = 1 - \mu_i$$

$$g_i = \frac{6}{h_i + h_{i+1}}\left(f[x_i, x_{i+1}] - f[x_{i-1}, x_i] \right) = 6 f[x_{i-1}, x_i, x_{i+1}]$$

$$\begin{bmatrix} 8 & 2.5 \\ 2.5 & 9 \end{bmatrix}\begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \begin{bmatrix} 9.6 \\ -9.6 \end{bmatrix} \Longrightarrow \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \begin{bmatrix} 1.67909 \\ -1.53308 \end{bmatrix}$$

$$S_i(x) = \frac{1}{6h_i}[M_{i-1}(x_i - x)^3 + M_i(x - x_{i-1})^3] + (f_{i-1} - \frac{M_{i-1}h_i^2}{6})\frac{x_i - x}{h_i} + (f_i - \frac{M_i h_i^2}{6})\frac{x - x_{i-1}}{h_i}$$

$$S_1(x) = 0.186566(x-3)^3 + 1.666667(4.5 - x) + 0.246894(x-3)$$

$$S_2(x) = 0.111939(7-x)^3 - 0.102205(x-4.5)^3 - 0.299621(7-x) + 1.638783(x-4.5)$$

$$S_3(x) = -0.127757(9-x)^3 + 1.761027(9-x) + 0.25(x-7)$$

样条插值与其它插值

- 简单插值的次数与节点个数有关，n+1个节点上要用n次多项式来插值，而样条插值多项式的次数与节点个数无关，便于在多个节点上用低次插值。
- 一般的分段低次插值，每段上插值多项式不同，各段表达式之间没有内在联系，而样条插值多项式各段之间有联系。
- 样条多项式往往只需已知m-1个边界节点上的导数值就够了，其余节点上的导数值是自然形成的。

逆插值（反插值）（inverse interpolation）

- 根据$(x_i, f(x_i))$

x	1	2	3	4	5	6	7
$f(x)$	1	0.5	0.333	0.25	0.2	0.167	0.143

- 确定对应某个$f(x)$的x值——**逆插值**
- **交换变量**：

$f(x)$	0.143	0.167	0.333	0.2	0.25	0.5	1
x	7	6	3	5	4	2	1

  - 横坐标在空间分布的不均匀性会导致插值多项式产生振荡（即使是低阶多项式）
- **利用初始数据进行插值，得到插值多项式后，寻找对应的x值，相当于方程求根**

插值法的缺陷

- 由实验提供的数据带有测量误差，插值函数会保留数据的全部测量误差
- 当插值函数的阶数较高时，曲线摆动很大，而求得的插值函数与实验规律可能偏离甚远
- 实验数据往往很多，用插值法得到的近似表达式缺乏实用价值

本章内容

- 插值
  - 插值问题
  - Lagrange插值多项式
  - Newton插值多项式
  - 分段低次插值
  - 样条插值
- 拟合
  - 最小二乘方法

曲线拟合

- 实际问题
  - 从一组实验测量数据中(x_i,y_i) $(i=1,\ldots,m)$找出实验规律的数学表达式，即求函数$y=f(x)$的一个近似表达式$y=\varphi(x)$（经验公式）
- 要求
  - 构造一个能逼近列表数据的近似的数学表达式，使各数据点从总体上最贴近，而不一定要求构造的函数曲线通过所给数据点

从几何上来看，就是通过给定的m个数据点(x_i,y_i)，求曲线$y=f(x)$的一条近似曲线$y=\varphi(x)$。

最小二乘法

- 一般情况下，不要求近似曲线$y=\varphi(x)$严格地经过所有数据点$(x_i,y_i)(i=1,2\ldots,m)$，即不要求拟合函数$\varphi(x)$在x_i处的偏差（残差）$\boldsymbol{\delta_i=\varphi(x_i)-y_i}$严格地等于0 。
- 为使近似曲线能尽量反映所给数据点的变化趋势，要求偏差δ_i适当地小
  - 偏差绝对值之和最小 $\sum|\delta_i|$
  - 最大偏差绝对值最小 $\max|\delta_i|$
  - 偏差平方和最小 $\sum|\delta_i^2|$——**最小二乘原则**
- 对于给定数据(x_i,y_i)，要求在某个函数类Φ中寻求一个函数$\varphi^*(x)$，使

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

最小二乘解

最小二乘法的基本环节

- 确定 $\varphi(x)$ 的形式
  - 通常的作法是描绘出数据点 (x_i, y_i)，然后根据这些点的分布情况来选择 $\varphi(x)$ 的形式。
- 求最小二乘解
  - 求满足条件的近似函数 $\varphi^*(x)$。

最小二乘问题的解法

- 设 $\varphi(x)=F(a_0,\ldots,a_n, x)$ $(n<m)$
- 选择适当参数$a_k=a_k^*$，使相应的函数$\varphi^*(x)=F(a_0^*,\ldots,a_n^*, x)$满足最小二乘条件，即点$(a_0^*,\ldots,a_n^*)$是多元函数

$$S(a_0,\cdots,a_n)=\sum_{i=1}^{m}[F(a_0,\cdots,a_n,x_i)-y_i]^2$$

的极小点，从而$a_0^*,\ldots,a_n^*$满足方程组

$$\frac{\partial S}{\partial a_k}=0$$

法方程组

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

相应的法方程组必是线性方程组

最小二乘问题的解法

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

$$S(a_0,a_1,\ldots\ldots,a_n)=\sum_{i=1}^{m}[a_0\varphi_0(x_i)+a_1\varphi_1(x_i)+\ldots\ldots+a_n\varphi_n(x_i)-y_i]^2$$

$$\frac{\partial S}{\partial a_k}=0 \Longrightarrow \sum_{i=1}^{m}\varphi_k(x_i)[a_0\varphi_0(x_i)+a_1\varphi_1(x_i)+\ldots\ldots+a_n\varphi_n(x_i)-y_i]=0$$

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

$$a_0(\varphi_k,\varphi_0)+a_1(\varphi_k,\varphi_1)+\ldots\ldots+a_n(\varphi_k,\varphi_n)=(\varphi_k,f)$$

$$k=0,1,\cdots n$$

函数**h,g**的关于离散点列$\{x_i\}_{i=1}^{m}$的**离散内积**

最小二乘问题的解法

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

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

当$\varphi_0(x),\varphi_1(x),\ldots\varphi_n(x)$线性无关时，存在唯一解

最小二乘解存在的唯一性

- 对于给定的一组实验数据(x_i,y_i)(x_i互异;$i=1,2\ldots,m$),在函数类 $\Phi=\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)\}$ ($n<m$且$\varphi_0(x),\varphi_1(x)$, …$\varphi_n(x)$线性无关)中,存在唯一的函数 $\varphi^*(x)=a_0^*\varphi_0(x)+a_1^*\varphi_1(x)+\cdots+a_n^*\varphi_n(x)$

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

系数 a_k^* 可以通过解法方程组得到。

代数多项式拟合

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

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

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

Aa=b

$$\begin{bmatrix} m & \sum_{i=1}^{m} x_i & \cdots & \sum_{i=1}^{m} x_i^{\,n} \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^{\,2} & \cdots & \sum_{i=1}^{m} x_i^{\,n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{m} x_i^{\,n} & \sum_{i=1}^{m} x_i^{\,n+1} & \cdots & \sum_{i=1}^{m} x_i^{\,2n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \\ \vdots \\ \sum_{i=1}^{m} x_i^{\,n} y_i \end{bmatrix}$$

当 $m=n+1$ 时，所得的拟合多项式就是Newton或Lagrange插值多项式，对数据拟合效果并不好，对次数较高的多项式进行拟合，方程组的系数矩阵是病态的。

代数多项式拟合

记

$$C=\begin{bmatrix}1 & x_1 & \cdots & x_1{}^n\\ 1 & x_2 & \cdots & x_2{}^n\\ \vdots & \vdots & \ddots & \vdots\\ 1 & x_m & \cdots & x_m{}^n\end{bmatrix}\qquad Y=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_m\end{bmatrix}\qquad a=\begin{bmatrix}a_0\\ a_1\\ \vdots\\ a_n\end{bmatrix}$$

代数多项式拟合条件为 $Ca=y$

故法方程组**Aa=b**中 $A=C^TC$ $b=C^TY$

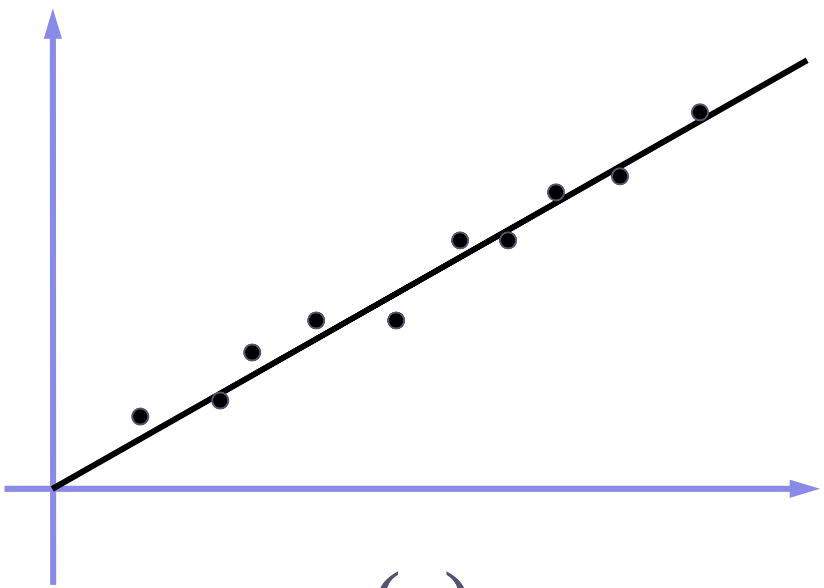
即法方程组为 $C^TCa=C^TY$

则 $a=\left(C^TC\right)^{-1}C^Ty$

采用广义逆表示的代数多项式拟合解。
实际计算时，通常不进行求逆计算，而是通过解法方程组得到。

选择合适的拟合数学公式

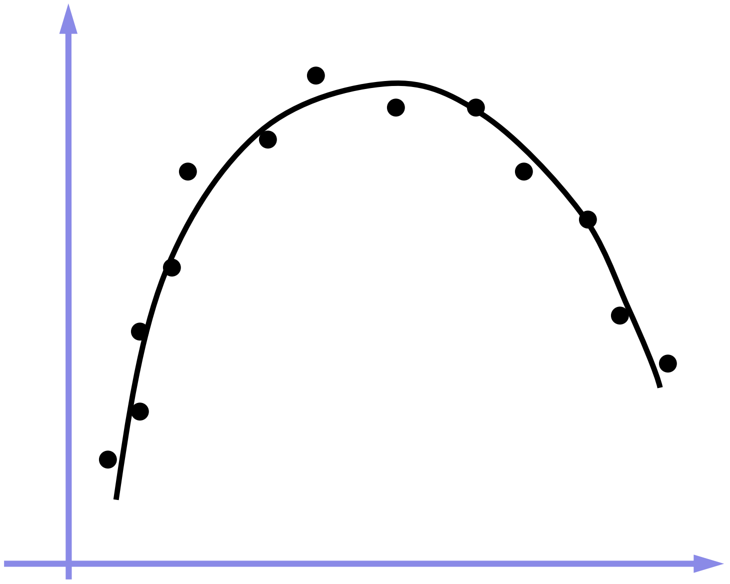
在作数据拟合时，选择合适的拟合数学公式是很重要的。通常最好是选用所给数据预先作出列表函数的曲线图，再根据曲线的大致形状，选择和确定合适的拟合数学公式形状，一味采用多项式拟合并不一定可取。



(a)

- 采用线性拟合

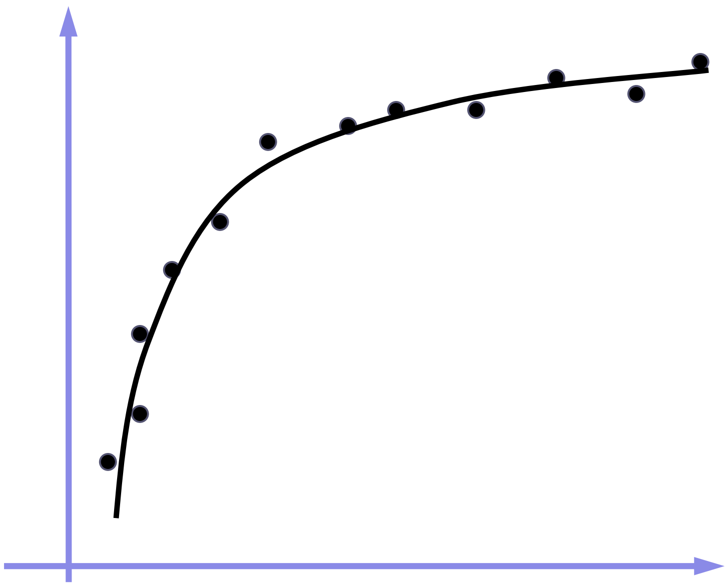
选择合适的拟合数学公式



(b)

- 采用多项式拟合

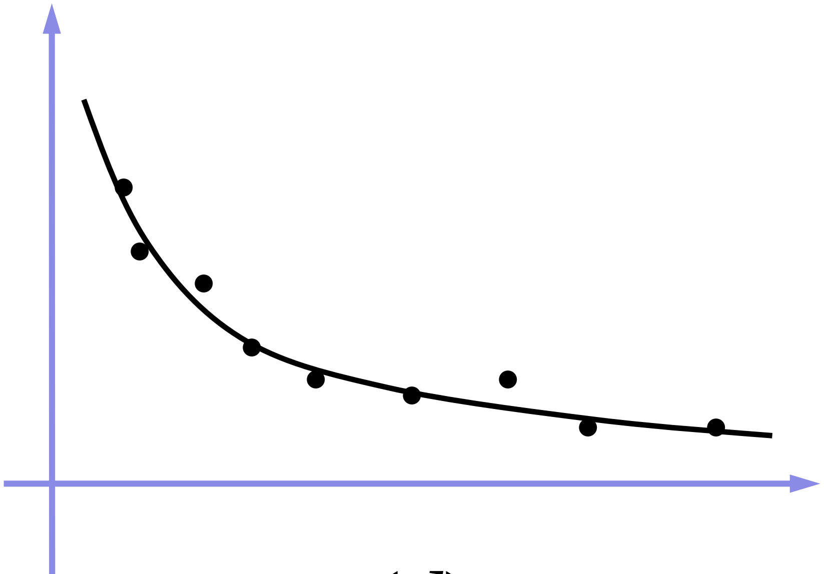
选择合适的拟合数学公式



(c)

- 采用双曲线型函数 $\varphi(x)=x/(ax+b)$
- 或指数型函数 $\varphi(x)=ae^{-b/x}$

选择合适的拟合数学公式



(d)

- 采用
  - $\varphi(x)=1/(ax+b)$
  - $\varphi(x)=1/(a+bx^2)$
  - $\varphi(x)=ae^{-bx}$

等函数拟合

最小二乘拟合——例1

- 已知数据，试用最小二乘法拟合x和y之间的经验公式

i	x_i	y_i	x_i^2	x_iy_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

- 绘出散点图
- 确定曲线形式——线性$\varphi(x)=a+bx$
- 建立法方程组

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

$$a = 95.3524, b = 2.2337$$

最小二乘拟合——例1

- 结果检验
- 均方误差

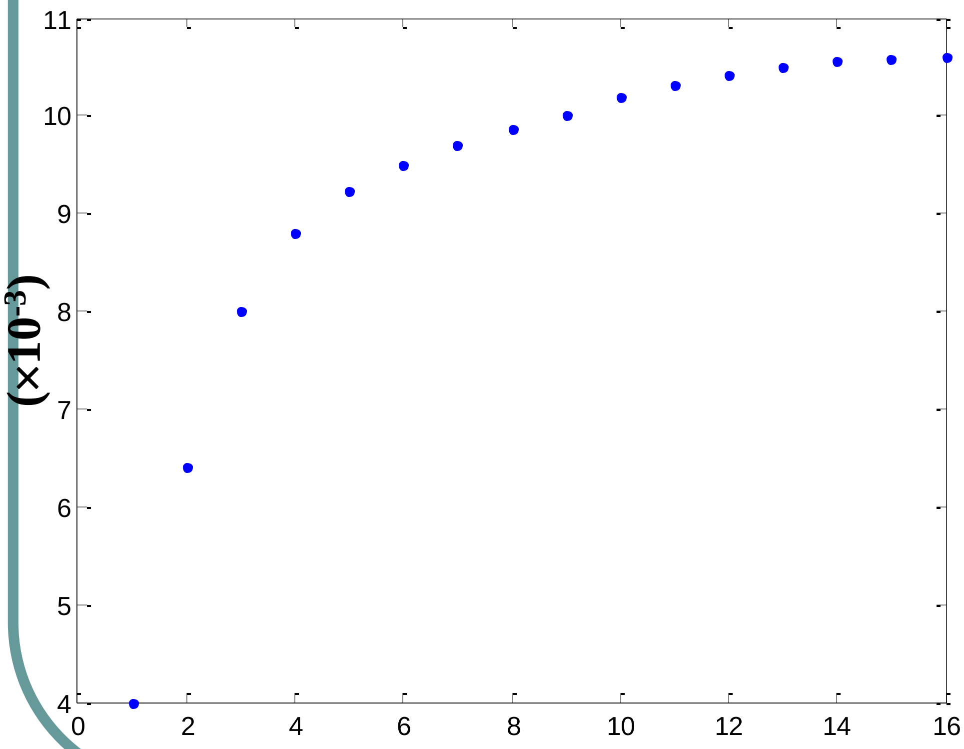
$$\frac{1}{n}\Sigma\delta_i^{\,2} = 4.445$$

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

如果认为误差不能满足要求，需要改变函数类型或增加实验数据等方法建立新的拟合曲线

最小二乘拟合——例2

- 例：已知数据，试用最小二乘建立y和t之间的关系



t	y $(\times10^{-3})$	t	y $(\times10^{-3})$
1	4.00	9	10.00
2	6.40	10	10.20
3	8.00	11	10.32
4	8.80	12	10.42
5	9.22	13	10.50
6	9.50	14	10.55
7	9.70	15	10.58
8	9.86	16	10.60

最小二乘拟合——例2

- (1) $y=\dfrac{t}{at+b}$ ⇒ $\dfrac{1}{y}=a+\dfrac{b}{t}$

  $y^{(1)}=\dfrac{1}{y}, t^{(1)}=\dfrac{1}{t}$ ⇒ $y^{(1)}=a+bt^{(1)}$

  $y=\dfrac{t}{80.6621t+161.6822}$

- (2) $y=ae^{b/t}$ ⇒ $\ln y=\ln a+\dfrac{b}{t}$

  $y^{(2)}=\ln y, t^{(2)}=\dfrac{1}{t}$ ⇒ $y^{(2)}=A+Bt^{(2)}$

  $A=\ln a, B=b$

  $y=0.011325e^{-\frac{1.0567}{t}}$

拟合方案	均方误差 ($\times 10^{-8}$)	最大偏差 ($\times 10^{-3}$)
(1)	8.83	0.568
(2)	0.73	0.277

扩展内容

- 加权最小二乘

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

- 利用正交函数作最小二乘拟合

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

- 拟合结果的置信区间

MATLAB中的函数

函数	描述
polyfit	根据数据用多项式进行最小二乘拟合
interp1	一维插值（查表）
interp1q	快速一维线性插值
interp2	二维插值
interpn	*n*维插值
spline	三次样条插值
ppval	分段多项式估计函数
cftool	Curve fitting Tool

第四章 总结——重要内容

- 相同点
  - 从函数角度看，插值法与最小二乘法都是一种根据函数表求函数的近似表达式的问题，属于函数逼近问题。
  - 从几何上看，二者都是根据一列数据点求曲线的近似曲线问题，是曲线拟合问题。
- 不同点
  - 插值法根据插值条件来选择近似函数；最小二乘法根据“偏差平方和最小”原则选择近似函数。

第四章 总结——重要内容

- 插值
  - 多项式插值
    - Lagrange插值
    - Newton插值
  - 分段低次插值
  - 样条插值
- 拟合
  - 最小二乘拟合
    - 代数多项式拟合