目录

正交多项式

- **定义3.9** 设$g_n(x)$是首项系数$a_n \neq 0$的n次多项式，如果多项式序列$g_0(x), g_1(x), \cdots$满足

$$(g_j, g_k) = \int_a^b \rho(x) g_j(x) g_k(x) \mathrm{d}x = \begin{cases} 0, & j \neq k \\ A_k > 0, & j = k \end{cases} \quad (j, k = 0,1, \dots)$$

  则称多项式序列$g_0(x), g_1(x), \cdots$在$[a, b]$上带权$\rho(x)$正交，并称$g_n(x)$是$[a, b]$上带权$\rho(x)$的n次正交多项式

- 一般来说，当权函数$\rho(x)$及区间$[a, b]$给定以后，可以由线性无关的一组基$\{1, x, x^2, \dots, x^n\}$并利用正交化方法构造出正交多项式

正交化

- 正交化方法

$$g_0(x) = 1,$$

$$g_n(x) = x^n - \sum_{k=0}^{n-1} \frac{(x^n, g_k(x))}{(g_k(x), g_k(x))} g_k(x) \qquad (n = 1,2,\ldots)$$

- 这样构造的正交多项式有以下性质

1. $g_n(x)$是最高项系数为1的n次多项式
2. 任一n次多项式$P_n(x) \in H_n$均可表示为$g_0(x), g_1(x), \cdots, g_n(x)$的线性组合
3. 当$n \neq m$时，$(g_n, g_m) = 0$且$g_n(x)$与任一次数小于n的多项式正交

正交化（续）

4. 有递推关系

$$g_{n+1}(x) = (x-\alpha_n)g_n(x) - \beta_n g_{n-1}(x), \quad (n = 0,1,\ldots)$$

其中 $g_0(x) = 1, \quad g_{-1}(x) = 0$

$$\alpha_n = \frac{(xg_n, g_n)}{(g_n, g_n)}, \quad (n = 0,1,\ldots)$$

$$\beta_n = \frac{(g_n, g_n)}{(g_{n-1}, g_{n-1})}, \quad (n = 1,2, \ldots)$$

5. 设$g_0(x), g_1(x), \cdots$是在$[a,b]$上带权$\rho(x)$的正交多项式序列，则$g_n(x)(n \geq 1)$的n个根都是单重实根，且都在区间(a,b)内

Legendre（勒让德）多项式

Legendre多项式

- 当区间为$[-1,1]$、权函数$\rho(x)\equiv 1$时，由$\{1,x,x^2,\ldots,x^n\}$正交化得到的多项式

$$P_0(x)=1 \qquad P_n(x)=\frac{1}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}x^n}[(x^2-1)^n] \quad (n=1,2,\cdots)$$

- 由于$(x^2-1)^n$是$2n$次多项式，求n阶导数后得

$$P_n(x)=\frac{1}{2^2 n!}(2n)(2n-1)\cdots(n+1)x^n+a_{n-1}x^{n-1}+\cdots+a_0$$

最高项系数为1的Legendre多项式

$$\tilde{P}_n(x)=\frac{n!}{(2n)!}\frac{\mathrm{d}^n}{\mathrm{d}x^n}[(x^2-1)^n]$$

Legendre多项式的性质

- 性质1 正交性

$$\int_{-1}^{1} P_n(x)P_m(x)\mathrm{d}x = \begin{cases} 0, & m \neq n \\ \dfrac{2}{2n+1}, & m = n \end{cases}$$

- 性质2 奇偶性

$$P_n(-x) = (-1)^n P_n(x)$$

  - n为偶数时$P_n(x)$为偶函数，n为奇数时为奇函数

- 性质3 递推关系

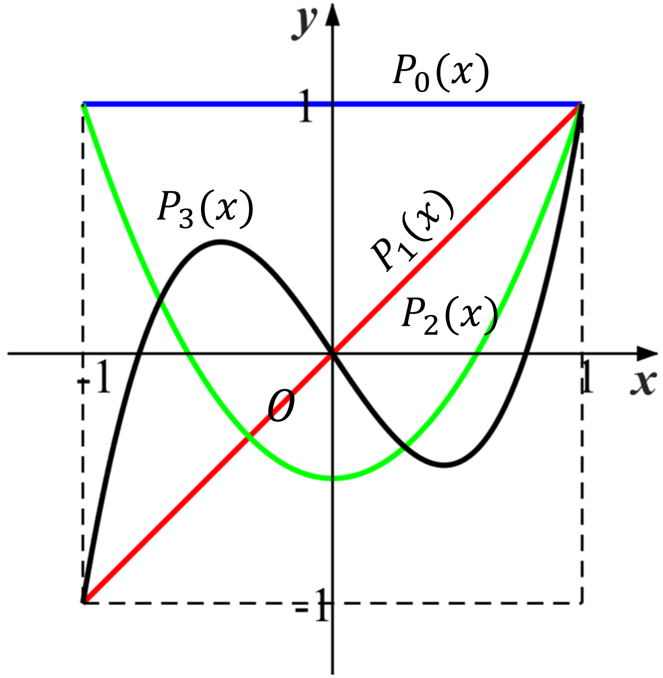
$$P_0(x) = 1 \qquad P_1(x) = x$$

$$(n+1)p_{n+1}(x) = (2n+1)xp_n(x) - np_{n-1}(x) \ (n = 1,2\cdots)$$

Legendre多项式的性质（续）

- 根据递推公式

$$
\begin{aligned}
&P_0(x) = 1\\
&P_1(x) = x\\
&P_2(x) = (3x^2 - 1)/2\\
&P_3(x) = (5x^3 - 3x)/2\\
&P_4(x) = (35x^4 - 30x^2 + x)/8\\
&P_5(x) = (63x^5 - 70x^3 + 15x)/8\\
&P_6(x) = (231x^6 - 315x^4 + 105x^2 - 5)/16\\
&\cdots\cdots
\end{aligned}
$$



- **性质4** 在所有最高项系数为1的n次多项式中，$\tilde{P}_n(x)$在$[-1,1]$上与零的平方误差最小
  - 由于权重函数为1
- **性质5** $P_n(x)$在区间$(-1,1)$内有n个不同实零点

Chebyshev多项式

- Chebyshev多项式
  - 当权函数$\rho(x)=\frac{1}{\sqrt{1-x^2}}$、区间为$[-1,1]$时，由序列$\{1,x,x^2,\dots,x^n\}$正交化得到的正交多项式

$$T_n(x)=\cos(n\arccos x),|x|\le 1$$

  - 若令$x=\cos\theta$，则

$$T_n(x)=\cos(n\theta),0\le\theta\le\pi$$

- Chebyshev多项式并不是三角函数
- 权函数不同、正交多项式也不同

Chebyshev多项式的性质

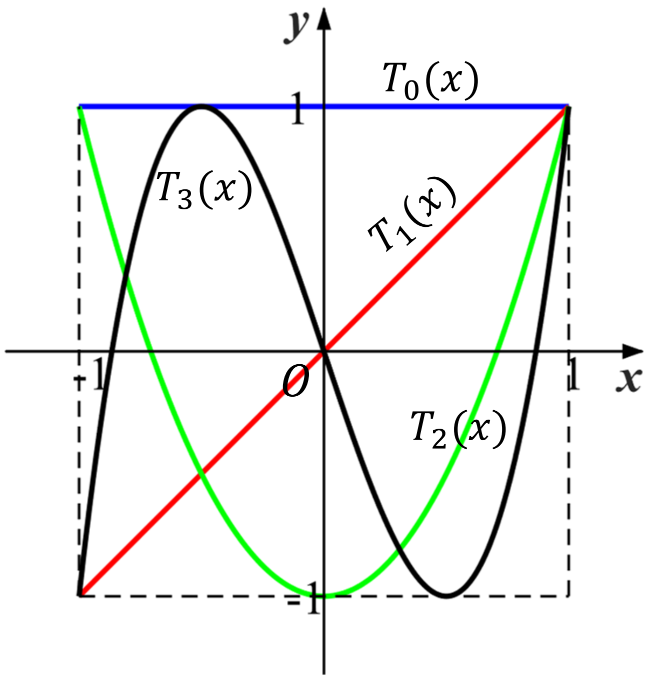
□ 性质1 递推关系

$$T_0(x) = 1 \qquad T_1(x) = x$$

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \ \ (n = 1,2, \dots)$$

- 由递推关系式可得

$$\begin{aligned}
&T_0(x) = 1\\
&T_1(x) = x\\
&T_2(x) = 2x^2 - 1\\
&T_3(x) = 4x^3 - 3x\\
&T_4(x) = 8x^4 - 8x^2 + 1\\
&T_5(x) = 16x^5 - 20x^3 + 5x\\
&T_6(x) = 32x^6 - 48x^4 + 18x^2 - 1\\
&\dots \dots
\end{aligned}$$



Chebyshev多项式的性质（续）

- **性质2** $T_n(x)$对零的偏差最小
  - **定理3.7** 在区间$[-1,1]$上所有最高项系数为1的n次多项式中，$\omega_n(x)=\frac{1}{2^{n-1}}T_n(x)$与零的偏差最小，其偏差为$\frac{1}{2^{n-1}}$
  - $T_n(x)$的最高项系数是$2^{n-1}(n\geq 1)$
  - 在区间$[-1,1]$上，将x^n在H_{n-1}中的“最佳一致逼近多项式”记为$P_{n-1}^*(x)$，则误差

$$x^n-P_{n-1}^*(x)=\omega_n(x)=\frac{1}{2^{n-1}}T_n(x)$$

例3.3

- 求$f(x) = 2x^3 + x^2 + 2x - 1$ 在$[-1,1]$上的最佳二次逼近多项式
  - 考虑“最佳一致逼近多项式”， $P_2^*(x)$应满足

$$\max_{-1\le x\le 1} |f(x) - P_2^*(x)| = \min$$

  - 注意到，$x^2 + 2x - 1$能够被二次函数完美表达，因此只有$2x^3$会导致逼近误差
  - 由定理3.7可知，

$$\frac{1}{2}[f(x) - P_2^*(x)] = x^3 - \hat{P}_2^*(x) = \frac{1}{2^{3-1}} T_3(x)$$

例3.3（续）

- 代入 $$T_3(x) = 4x^3 - 3x$$

- 得到 $$P_2^*(x) = f(x) - \frac{1}{2}T_3(x)$$
$$= x^2 + \frac{7}{2}x - 1$$

就是$f(x)$在$[-1, 1]$上的最佳二次逼近多项式

Chebyshev多项式的性质（续）

- 性质**3** Chebyshev多项式$\{T_n(x)\}$在区间$[-1,1]$上带权$\rho(x)=\frac{1}{\sqrt{1-x^2}}$正交，且

$$\int_{-1}^{1}\frac{T_n(x)T_m(x)\mathrm{d}x}{\sqrt{1-x^2}}=\begin{cases}0, & n\neq m\\ \frac{\pi}{2}, & n=m\\ \pi, & n=m=0\end{cases}$$

- 性质**4** $T_{2k}(x)$只含x的偶次幕，$T_{2k+1}(x)$只含x的奇次幕

Chebyshev多项式的性质（续）

- **性质5** $T_n(x)$在区间$[-1,1]$上有n个零点

$$x_k = \cos\frac{2k-1}{2n}\pi, k = 1,2,\cdots$$

- 此外，x^n可以用$T_0, T_1, \dots, T_n$的线性组合表示

$$x^n = 2^{1-n}\sum_{k=0}^{\left\lfloor\frac{n}{2}\right\rfloor}\binom{n}{k}T_{n-2k}(x)$$

其中规定$T_0 = 1/2$

第二类Chebyshev多项式

- 第二类Chebyshev多项式
  - 在区间[−1,1]上带权$\rho(x)=\sqrt{1-x^2}$的正交多项式

$$U_n(x)=\frac{\sin[(n+1)\arccos x]}{\sqrt{1-x^2}}$$

- 正交关系

$$\int_{-1}^{1}U_n(x)U_m(x)\sqrt{1-x^2}\mathrm{d}x=\begin{cases}0, & m\neq n\\ \frac{\pi}{2}, & m=n\end{cases}$$

- 递推关系式

$$U_0(x)=1 \qquad U_1(x)=2x$$

$$U_{n+1}(x)=2xU_n(x)-U_{n-1}(x)\ \ (n=1,2,\ldots)$$

Laguerre（拉盖尔）多项式

Laguerre多项式

- 在区间$[0,\infty)$上带权$\rho(x) = e^{-x}$的正交多项式

$$L_n(x) = e^n \frac{\mathrm{d}^n}{\mathrm{d}x^n}(x^n e^{-x})$$

正交关系

$$\int_0^\infty L_n(x) L_m(x) e^{-x} \mathrm{d}x = \begin{cases} 0, & m \neq n \\ (n!)^2, & m = n \end{cases}$$

递推关系式

$$L_0(x) = 1 \qquad L_1(x) = 1 - x$$

$$L_{n+1}(x) = (1 + 2n - x)L_n(x) - n^2 U_{n-1}(x) \ \ (n = 1,2, \dots)$$

Hermite（埃尔米特）多项式

Hermite多项式

- 在区间$(-\infty,\infty)$上带权$\rho(x)=e^{-x^2}$的正交多项式

$$H_n(x)=(-1)^n e^{x^2}\frac{\mathrm{d}^n}{\mathrm{d}x^n}\left(e^{-x^2}\right)$$

正交关系

$$\int_{-\infty}^{+\infty}H_n(x)H_m(x)e^{-x^2}\mathrm{d}x=\begin{cases}0, & m\neq n\\ 2^n n!\sqrt{\pi}, & m=n\end{cases}$$

递推关系式

$$H_0(x)=1\qquad H_1(x)=2x$$

$$H_{n+1}(x)=2xH_n(x)-2nH_{n-1}(x)\ \ (n=1,2,\dots)$$

目录

- 引言与预备知识
- 最佳一致逼近多项式
- 最佳平方逼近
- 正交多项式
- 函数按正交多项式展开
- 曲线拟合的最小二乘法

函数按正交多项式展开

- 设$f(x) \in C[a,b]$用正交多项式$\{g_0(x), g_1(x), \cdots, g_n(x)\}$作基，求最佳平方逼近多项式

$$S_n(x) = a_0 g_0(x) + a_1 g_1(x) + \cdots + a_n g_n(x) \qquad (3.5.1)$$

- 由$\{g_k(x)\}$的正交性及式(3.3.13),

$$\sum_{j=0}^{n} (\varphi_k, \varphi_j) a_j = (f, \varphi_k) (k = 0,1, \dots, n) \qquad (3.3.13)$$

可求得系数

$$a_k = \frac{(f, g_k)}{(g_k, g_k)} (k = 0,1, \dots, n) \qquad (3.5.2)$$

函数按正交多项式展开（续）

- $f(x)$的最佳平方逼近多项式为

$$S_n(x)=\sum_{k=0}^{n}\frac{(f,g_k)}{(g_k,g_k)}g_k(x) \qquad (3.5.3)$$

- 由式(3.3.15)

$$\|\delta\|_2^2=(f,f)-(f,S^*)=\|f\|_2^2-\sum_{k=0}^{n}a_k^*(\varphi_k,f) \qquad (3.3.15)$$

可得均方误差为

$$\|\delta_n\|_2=\|f-S_n\|_2=\sqrt{\|f\|_2^2-\sum_{k=0}^{n}\frac{(f,g_k)}{(g_k,g_k)}(f,g_k)} \qquad (3.5.4)$$

函数按正交多项式展开（续）

- 若$f(x)$在$[a,b]$上按正交多项式$\{g_k(x)\}$展开，系数$a_k(k=0,1,\dots)$按式(3.5.2)计算，这样可得到$f(x)$的展开式

$$f(x)\sim\sum_{k=0}^{\infty}a_k g_k(x) \qquad (3.5.5)$$

  - 式(3.5.5)右端级数称为广义Fourier级数
  - 系数a_k称为广义Fourier系数
- 任何$f(x)\in C[a,b]$均可展开成广义Fourier级数，其部分和$S_n(x)$是$f(x)$的最佳平方逼近
  - 系数a_k与n无关：n增加时，只要计算增加的系数
  - 在$f(x)$满足一定条件下也可一致收敛到$f(x)$

举例

$$\|\delta_n\|_2 = \sqrt{\|f\|_2^2 - \sum_{k=0}^{n} \frac{(f,g_k)}{(g_k,g_k)}(f,g_k)} \qquad (3.5.4)$$

- 考虑函数$f(x) \in C[-1,1]$按Legendre多项式$\{P_0(x), P_1(x), \cdots\}$展开的最佳平方逼近多项式$S_n^*$。 由式(3.5.1)和式(3.5.2)可得

$$S^*(x) = a_0^* P_0(x) + a_1^* P_1(x) + \cdots + a_n^* P_n(x) \qquad (3.5.6)$$

  - 其中

$$a_k^* = \frac{(f,P_k)}{(P_k,P_k)} = \frac{2k+1}{2}\int_{-1}^{1} f(x)P_k(x)\mathrm{d}x \ \ (k = 0,1,\ldots,n) \qquad (3.5.7)$$

- 根据式(3.5.4)平方误差为

$$\|\delta_n\|_2^2 = \int_{-1}^{1} f^2(x)x - \sum_{k=0}^{n} \frac{2}{2k+1} {a_k^*}^2 \qquad (3.5.8)$$

例3.4

- 求$f(x)=e^x$在$[-1,1]$上的三次最佳平方逼近多项式
  - 先计算$(f,P_k)(k=\ 0,1,2,3)$，即

$$(f,P_0)=\int_{-1}^{1}e^x\mathrm{d}x\approx 2.3504\quad (f,P_1)=\int_{-1}^{1}xe^x\mathrm{d}x\approx 0.7358$$

$$(f,P_2)=\int_{-1}^{1}\left(\frac{3}{2}x^2-\frac{1}{2}\right)e^x\mathrm{d}x\approx 0.1431$$

$$(f,P_3)=\int_{-1}^{1}\left(\frac{5}{2}x^2-\frac{3}{2}\right)e^x\mathrm{d}x\approx 0.02013$$

例3.4（续）

$$a_k^* = \frac{(f, P_k)}{(P_k, P_k)} \qquad (3.5.7)$$

- 由式(3.5.7)得

$$a_0^* = 1.1752, a_1^* = 1.1036, a_2^* = 0.3578, a_3^* = 0.07046$$

- 代入式(3.5.6)，得

$$S_3^*(x) = 0.9963 + 0.9979x + 0.5367x^2 + 0.1761x^3$$

- 均方误差为

$$\|\delta_n\|_2 = \|e^x - S_3^*(x)\|_2 = \sqrt{\int_{-1}^{1} e^{2x}\mathrm{d}x - \sum_{k=0}^{3} \frac{2}{2k+1} {a_k^*}^2} \le 0.0084$$

- 最大误差为

$$\|\delta_n\|_\infty = \|e^x - S_3^*(x)\|_\infty \le 0.0112$$

任意区间上的函数逼近

- 如果$f(x) \in C[a,b]$，求$f(x)$在区间$[a,b]$上的最佳平方逼近多项式

  1. 作变换
  $$x = \frac{b-a}{2}t + \frac{b+a}{2} (-1 \le t \le 1)$$
  得到$F(t) = f(x) = f\left(\frac{b-a}{2}t + \frac{b+a}{2}\right)$为定义在$[-1,1]$上的函数

  2. 用Legendre多项式，求$F(t)$的最佳平方逼近多项式$S_n^*(t)$

  3. 回代，得到结果$S(x) = S_n^*\left(\frac{1}{b-a}(2x-a-b)\right)$

讨论

- Legendre多项式$\{P_k(x)\}$是在区间$[-1,1]$上由$\{1, x, x^2, \dots, x^k, \dots\}$正交化得到的
  - 利用函数的Legendre展开部分和得到的最佳平方逼近多项式与解法方程得到的H_n中的最佳平方逼近多项式是一致的
- 当n较大时求法方程会出现病态方程，计算误差较大，不能使用
- 而用Legendre展开不用解线性方程组，不存在病态问题，计算公式使用起来也较方便，因此通常都用此法求最佳平方逼近多项式

目录

- 引言与预备知识
- 最佳一致逼近多项式
- 最佳平方逼近
- 正交多项式
- 函数按正交多项式展开
- 曲线拟合的最小二乘法

数据拟合

- 在科学实验的统计方法研究中，往往要从一组实验数据$(x_i, y_i)(i = 0,1,\dots,m)$中寻找自变量$x$与因变量$y$之间的函数关系$y = F(x)$
  - 由于观测数据往往不准确，因此不要求$y = F(x)$经过所有点(x_i, y_i)而只要求在给定点x_i上误差$\delta_i = F(x_i) - y_i\ (i = 0,1,\dots,m)$按某种标准最小
- 若记$\boldsymbol{\delta} = (\delta_0, \delta_1, \dots, \delta_m)^{\mathrm{T}}$，就是要求向量$\boldsymbol{\delta}$的范数$\|\boldsymbol{\delta}\|$最小
  - 如果用最大范数，计算上困难较大
  - 通常就采用Euclid范数$\|\boldsymbol{\delta}\|_2$作为误差度量的标准

一般的最小二乘逼近

- 对于给定的一组数据$(x_i, y_i)(i = 0,1,\dots,m)$，要求在函数空间$\Phi = \mathrm{span}\{\varphi_0, \varphi_1, \dots, \varphi_n\}$中找一个函数$y = S^*(x)$，使误差平方和满足

$$\|\boldsymbol{\delta}\|_2^2 = \sum_{i=0}^{m} \delta_i^2 = \sum_{i=0}^{m} [S^*(x_i) - y_i]^2 = \min_{S(x)\in\Phi} \sum_{i=0}^{m} [S(x_i) - y_i]^2 \quad (3.6.1)$$

  - 这里

$$S(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots + a_n\varphi_n(x) \quad (3.6.2)$$

  - 若$\varphi_k(x)$是k次多项式，$S(x)$就是n次多项式

- 几何语言说，就称为曲线拟合的最小二乘

一般的最小二乘逼近（续）

- 确定$S(x)$的具体形式
  - 不是单纯数学问题，还与所研究问题的运动规律及所得观测数据(x_i, y_i)有关
  - 通常要从问题的运动规律及给定数据描图来确定$S(x)$的形式，并通过实际计算选出较好的结果
- 把最小二乘法中$\|\boldsymbol{\delta}\|_2^2$考虑为加权平方和

$$\|\boldsymbol{\delta}\|_2^2 = \sum_{i=0}^{m} \omega(x_i)[S^*(x_i) - f(x_i)]^2 \tag{3.6.3}$$

  - $\omega(x)$是$[a,b]$上的权函数，它表示不同点$(x_i, f(x_i))$处的数据比重
  - $\omega(x_i)$可表示在点$(x_i, f(x_i))$处重复观测的次数

$$S(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots + a_n\varphi_n(x) \tag{3.6.2}$$

求解过程

用最小二乘法求拟合曲线的问题

- 在形如式(3.6.2)的$S(x)$中求一函数$y = S^*(x)$，使式(3.6.3)取得最小

$$\|\boldsymbol{\delta}\|_2^2 = \sum_{i=0}^{m} \omega(x_i)[S^*(x_i) - f(x_i)]^2 \tag{3.6.3}$$

- 等价于求多元函数极小点$(a_0^*, a_1^*, \cdots, a_n^*)$

$$I(a_0, a_1, \cdots, a_n) = \sum_{i=0}^{m} \omega(x_i)\left[\sum_{i=0}^{n} a_i\varphi_i(x_i) - f(x_i)\right]^2 \tag{3.6.4}$$

与3.3节讨论的“函数的最佳平方逼近”类似

求解过程（续）

- 由求多元函数极值的必要条件，有

$$\frac{\partial I}{\partial a_k} = 2\sum_{i=0}^{m}\omega(x_i)\left[\sum_{j=0}^{n}a_j\varphi_j(x_i) - f(x_i)\right]\varphi_k(x_i) = 0\ (k = 0,1,\dots,n)$$

- 若记

$$(\varphi_j, \varphi_k) = \sum_{i=0}^{m}\omega(x_i)\,\varphi_j(x_i)\varphi_k(x_i) \qquad (3.6.5)$$

$$(f, \varphi_k) = \sum_{i=0}^{m}\omega(x_i)\,f(x_i)\varphi_k(x_i) \equiv d_k \quad (k = 0,1,\cdots,n)$$

- 则必要条件可改写为

$$\sum_{j=0}^{n}(\varphi_k, \varphi_j)a_j = d_k \quad (k = 0,1,\cdots,n) \qquad (3.6.6)$$

求解过程（续）

- 此方程称为法方程它也可写成矩阵形式

$$\boldsymbol{G}\boldsymbol{a} = \boldsymbol{d}$$

其中

$$\boldsymbol{a} = (a_0, a_1, \cdots, a_n)^{\mathrm{T}} \qquad \boldsymbol{d} = (d_0, d_1, \cdots, d_n)^{\mathrm{T}}$$

$$\boldsymbol{G} = \begin{bmatrix} (\varphi_0, \varphi_0) & (\varphi_0, \varphi_1) & \cdots & (\varphi_0, \varphi_n) \\ (\varphi_1, \varphi_0) & (\varphi_1, \varphi_1) & \cdots & (\varphi_1, \varphi_n) \\ \vdots & \vdots & & \vdots \\ (\varphi_n, \varphi_0) & (\varphi_n, \varphi_1) & \cdots & (\varphi_n, \varphi_n) \end{bmatrix} \tag{3.6.7}$$

- 该方程是否有解？
- 该方程是否有唯一解？

求解过程（续）

要求$m > n$

- 由于$\varphi_0, \varphi_1, \ldots, \varphi_n$线性无关，故系数行列式$|\boldsymbol{G}| \neq 0$，于是方程组(3.6.6)有唯一解

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

- 从而得到函数$f(x)$的最小二乘解为

$$S^*(x) = a_0^*\varphi_0(x) + a_1^*\varphi_1(x) + \cdots + a_n^*\varphi_n(x) \qquad (3.3.14)$$

- 同样可以证明充分性：对于任何$S(x) \in \Phi$，都有

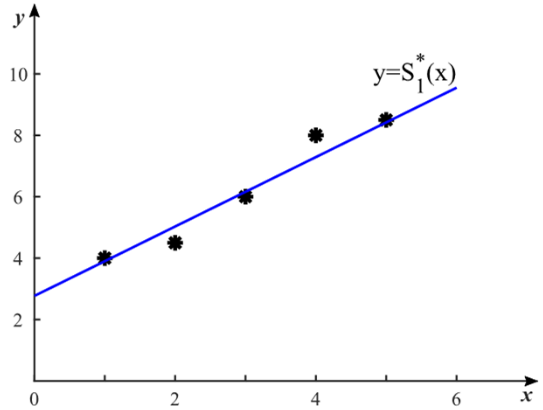
$$\sum_{i=0}^{m} \omega(x_i)[S^*(x_i) - f(x_i)]^2 \leq \sum_{i=0}^{m} \omega(x_i)[S(x_i) - f(x_i)]^2$$

故$S^*(x)$确是所求最小二乘解

例3.5

□ 已知一组实验数据如下表，求它的拟合曲线

x_i	1	2	3	4	5
f_i	4	4.5	6	8	8.5
ω_i	2	1	3	1	1



- 在坐标纸上标出所给数据，观测到各点分布在一条直线附近，故可选择线性函数作拟合曲线
- 令$S_1(x) = a_0 + a_1 x$，这里$m = 4, n = 1, \varphi_0(x) = 1, \varphi_1(x) = x$故

$$(\varphi_0, \varphi_0) = \sum_{i=0}^{4} \omega_i = 8 \qquad (\varphi_0, \varphi_1) = (\varphi_1, \varphi_0) = \sum_{i=0}^{4} \omega_i x_i = 22$$

例3.5（续）

$$\sum_{j=0}^{n}(\varphi_k,\varphi_j)a_j = d_k \qquad (k=0,1,\cdots,n) \tag{3.6.6}$$

$$(\varphi_1,\varphi_1)=\sum_{i=0}^{4}\omega_i x_i^2 = 74$$

$$(\varphi_0,f)=\sum_{i=0}^{4}\omega_i f_i = 47 \qquad (\varphi_1,f)=\sum_{i=0}^{4}\omega_i x_i f_i = 145.5$$

- 由式(3.6.6)得方程组

$$\begin{cases} 8a_0 + 22a_1 = 47 \\ 22a_0 + 74a_1 = 145.5 \end{cases}$$

- 解得

$$a_0 = 2.77, a_1 = 1.13$$

- 于是所求拟合曲线为

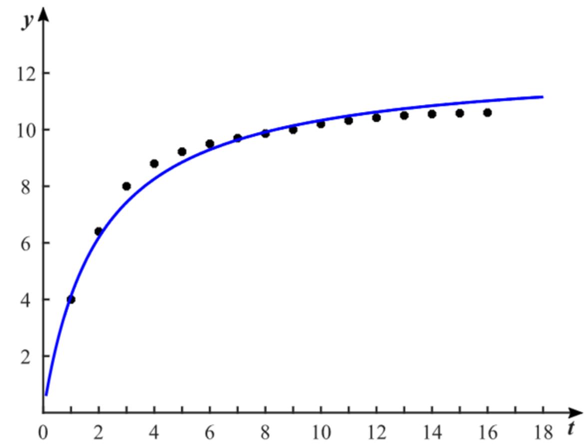
$$S_1^*(x) = 2.77 + 1.13x$$

例3.6

□ 在某化学反应过程中，根据实验所得生成物的质量分数与时间的关系如下表所示，求质量分数y与时间t的拟合曲线$y = F(t)$

t/min	1	2	3	……	14	15	16
$y/\times 10^{-3}$	4.00	6.40	8.00	……	10.55	10.58	10.60

- 质量分数开始时增加较快，后来逐渐减慢，到一定时间就基本稳定在一个水平上，即当$t \to \infty$时，y趋于某个数，故$y = F(t)$有一水平渐近线
- $t = 0$时，反应未开始，质量分数为零



例3.6（续）

1. 设想$y = F(t)$是双曲线型$1/y = a + b/t$，即$y = t/(at + b)$，与给定数据的规律大致符合

- 为了确定a, b的值，先将上述关系改写为

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

- 令$\bar{y} = 1/y, x = 1/t$，因此可以用x的线性函数$S_1(x) = a + bx$拟合数据$(x_i, \bar{y}_i)$ $(i = 1,2, \dots, 16)$
- $(x_i, \bar{y}_i)$由原始数据(t_i, y_i)变换而来
- 与例3.5类似，得到下面的方程组

$$\begin{cases} 16a + 3.38073b = 1.8372 \times 10^3 \\ 3.38073a + 1.58435b = 0.52886 \times 10^3 \end{cases}$$

例3.6（续）

- 得 $a = 80.6621, b = 161.6822$
- 从而得到
$$y = t/(80.6621t + 161.6822) = F^{(1)}(t)$$
- 其误差为
$$\delta_i^{(1)} = y_i - F^{(1)}(t_i)(i = 1,2,\ldots,16)$$

2. 符合给定数据的函数还可选为指数形式，此时令曲线为
$$y = ae^{b/t}$$

- 当$t \to \infty$时，$y \to a$；当$t \to 0$时，若$b < 0$，则$y \to 0$且t增加时y增加，与给出数据规律相同

例3.6（续）

- 为了确定a与b，对上式两端取对数

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

- 令$\hat{y} = \ln y, A = \ln a, x = 1/t$
- 由(t_i, y_i)计算出$(x_i, \hat{y}_i)$，拟合数据$(x_i, \hat{y}_i)$的曲线仍为$S_1(x) = A + bx$
- 用例3.5的方法计算出$A = -4.48072, b = -1.0567$，从而$a = e^A = 11.3253 \times 10^{-3}$
- 最后求得

$$y = 11.3253 \times 10^{-3} e^{-1.0567/t} = F^{(2)}(t)$$

- 误差为

$$\delta_i^{(2)} = y_i - F^{(2)}(t_i)(i = 1,2, \dots, 16)$$

例3.6（续）

- 怎样比较这两个数学模型的好坏呢？
  - 分别计算各点误差，从中挑选误差较小的模型
  - 经过计算可得

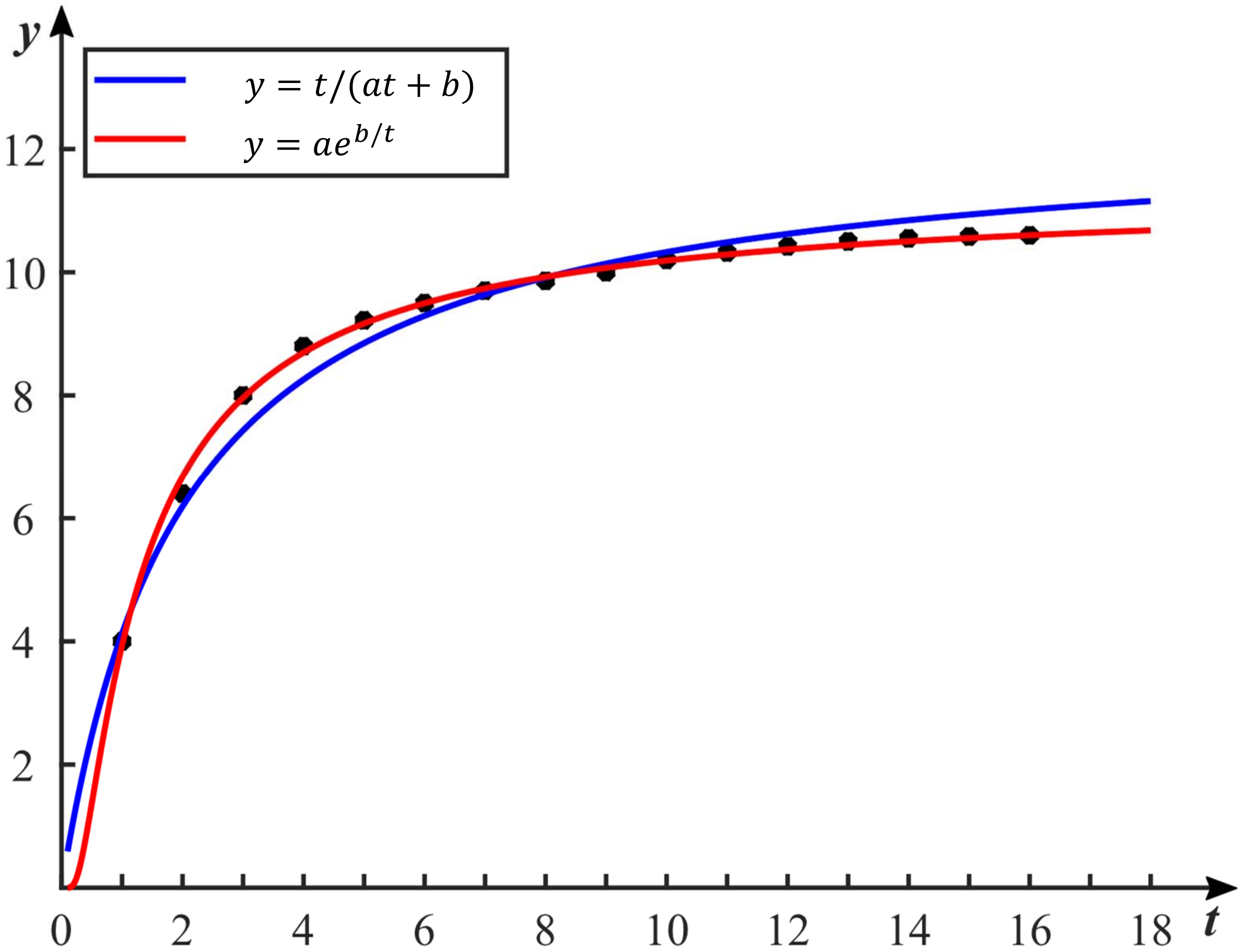
$$\left\|\boldsymbol{\delta}^{(1)}\right\|_\infty = 0.568\times 10^{-3} \qquad \left\|\boldsymbol{\delta}^{(2)}\right\|_\infty = 0.277\times 10^{-3}$$

$$\left\|\boldsymbol{\delta}^{(1)}\right\|_2 = \sqrt{\sum_{i=1}^{m}\left(\delta_i^{(1)}\right)^2} = 1.19\times 10^{-3}$$

$$\left\|\boldsymbol{\delta}^{(2)}\right\|_2 = \sqrt{\sum_{i=1}^{m}\left(\delta_i^{(2)}\right)^2} = 0.34\times 10^{-3}$$

$\left\|\boldsymbol{\delta}^{(2)}\right\|_2$及$\left\|\boldsymbol{\delta}^{(2)}\right\|_\infty$都比较小，所以用$y=F^{(2)}(t)$作拟合曲线比较好

例3.6（续）



数据拟合流程

1. 在坐标系中画出数据点，观察分布特点
2. 由数据点的分布，取基函数系 $\{\varphi_0, \varphi_1, \dots, \varphi_n\}$ $\begin{cases} \text{多项式} \\ \text{指数函数} \\ \text{三角函数} \end{cases}$
3. 构成法方程并求解，得到拟合曲线，分析误差

- 拟合曲线的数学模型并不是一开始就能选得好的，往往要通过分析确定若干模型后，再经过实际计算，才能选到较好的模型
  - 存在过拟合的风险！

用正交函数作最小二乘拟合

- 最小二乘法得到的法方程组(3.6.6)，其系数矩阵$\boldsymbol{G}$可能是病态的

$$\boldsymbol{G}\boldsymbol{a} = \boldsymbol{d}$$

$$\boldsymbol{G} = \begin{bmatrix} (\varphi_0, \varphi_0) & (\varphi_0, \varphi_1) & \cdots & (\varphi_0, \varphi_n) \\ (\varphi_1, \varphi_0) & (\varphi_1, \varphi_1) & \cdots & (\varphi_1, \varphi_n) \\ \vdots & \vdots & & \vdots \\ (\varphi_n, \varphi_0) & (\varphi_n, \varphi_1) & \cdots & (\varphi_n, \varphi_n) \end{bmatrix} \qquad (3.6.7)$$

- 考虑$\varphi_0, \varphi_1, \dots, \varphi_n$是关于点集$\{x_i\}$ $(i = 0,1, \dots, m)$带权$\omega(x)$ $(i = 0, 1, \dots, m)$正交的函数族

$$(\varphi_j, \varphi_k) = \sum_{i=0}^{m} \omega(x_i)\varphi_j(x_i)\varphi_k(x_i) = \begin{cases} 0, & j \neq k \\ A_k > 0, & j = k \end{cases} \qquad (3.6.8)$$

$$a_k = \frac{(f, g_k)}{(g_k, g_k)} (k = 0,1, \ldots, n) \quad (3.5.2)$$

用正交函数作最小二乘拟合（续）

- 根据(3.6.6)

$$\sum_{j=0}^{n} (\varphi_k, \varphi_j) a_j = d_k \qquad (k = 0,1, \cdots, n) \qquad (3.6.6)$$

可得

$$a_k^* = \frac{(f, \varphi_k)}{(\varphi_k, \varphi_k)} = \frac{\sum_{i=0}^{m} \omega(x_i) f(x_i) \varphi_k(x_i)}{\sum_{i=0}^{m} \omega(x_i) \varphi_k^2(x_i)} \ (k = 0,1, \cdots, n) \quad (3.6.9)$$

✓ 与(3.5.2)一致

- 经过化简，易得平方误差为

$$\|\boldsymbol{\delta}\|_2^2 = \|f\|_2^2 - \sum_{i=0}^{m} A_k (a_k^*)^2$$

✓ 与(3.5.4)一致

正交化

与3.4.1节函数正交化手续一致

- 根据给定节点$x_0, x_1, \dots, x_m$及权函数$\omega(x) > 0$，构造带权$\omega(x)$正交的多项式$\{P_n(x)\}$
  - 注意$n \le m$，递推公式

$$\begin{cases} P_0(x) = 1 \\ P_1(x) = (x - \alpha_1)P_0(x) \\ P_{k+1}(x) = (x - \alpha_k)P_k(x) - \beta_k P_{k-1}(x)(k = 1,2 \cdots, n - 1) \end{cases} \tag{3.6.10}$$

  - $P_k(x)$是首项系数为1的k次多项式，根据正交性得

$$\begin{cases} \alpha_{k+1} = \dfrac{\sum_{i=0}^{m} \omega(x_i) x_i P_k^2(x_i)}{\sum_{i=0}^{m} \omega(x_i) P_k^2(x_i)} = \dfrac{(xP_k, P_k)}{(P_k, P_k)} \\ \beta_k = \dfrac{\sum_{i=0}^{m} \omega(x_i) P_k^2(x_i)}{\sum_{i=0}^{m} \omega(x_i) P_{k-1}^2(x_i)} = \dfrac{(P_k, P_k)}{(P_{k-1}, P_{k-1})} \end{cases} (k = 1,2 \cdots, n - 1) \tag{3.6.11}$$

计算过程

与3.5节函数按正交多项式展开一致

- 用正交多项式$\{P_k(x)\}$的线性组合作最小二乘
  - 只要在根据式(3.6.10)及式(3.6.11)逐步求$P_k(x)$的同时，相应计算出系数

$$a_k=\frac{(f,P_k)}{(P_k,P_k)}=\frac{\sum_{i=0}^{m}\omega(x_i)f(x_i)P_k(x_i)}{\sum_{i=0}^{m}\omega(x_i)P_k^2(x_i)}\quad(k=0,1,\cdots,n)$$

  - 逐步把$a_k^*P_k(x)$累加到$F(x)$中去，最后就可得到所求的拟合曲线

$$y=F(x)=a_0^*P_0(x)+a_1^*P_1(x)+\cdots+a_n^*P_n(x)$$

- n可事先给定或在计算过程中根据误差确定
  - 当逼近次数增加一次，只要把程序中循环数加1

多元最小二乘拟合

- 已知多元函数$y=f(x_1,x_2,\dots,x_l)$的一组测量数据$(x_{1i},x_{2i},\dots,x_{li},y_i)$， 以及一组权系数$\omega_i>0(i=0,1,\dots,m)$，要求函数

$$S_n(x_1,x_2,\dots,x_l)=\sum_{k=0}^{n}a_k\varphi_k(x_1,x_2,\dots,x_l)\,,n\le m$$

使得

$$F(a_0,a_1,\dots,a_n)=\sum_{i=1}^{m}\omega_i[y_i-S_n(x_{1i},x_{2i},\dots,x_{li})]^2$$

最小

多元最小二乘拟合（续）

- 这与式(3.6.4)的极值问题完全一样，系数$a_0, a_1, \dots, a_n$同样满足法方程组(3.6.6)，

$$\sum_{j=0}^{n}(\varphi_k, \varphi_j)a_j = d_k \quad (k = 0,1,\cdots,n) \tag{3.6.6}$$

只是这里

$$(\varphi_k, \varphi_j) = \sum_{i=1}^{m} \omega(x_i)\, \varphi_k(x_{1i}, x_{2i}, \dots, x_{li})\varphi_j(x_{1i}, x_{2i}, \dots, x_{li})$$

- 求解法方程组(3.6.6)就可得到$a_k(k = 0,1,\cdots,n)$，从而得到$S_n(x_1, x_2, \dots, x_l)$

数据拟合 vs 数据插值

- 数据拟合
  - 对拟合函数不要求它通过所给的数据点
  - 所处理的数据量大且不能保证每一个数据没有误差
  - 要求一个函数严格通过每一个数据点是不合理的
  - 数据拟合方法求拟合函数

- 数据插值
  - 插值函数则必须通过每一个数据点
  - 插值方法求插值函数

总结

- 引言与预备知识
  - 函数逼近与计算、一致逼近、均方逼近
  - 一致逼近的存在性、连续函数空间
- 最佳一致逼近多项式
  - 定义、偏差、最佳逼近多项式的偏差点性质
  - Chebyshev定理，推论、最佳一次逼近多项式
- 最佳平方逼近
  - 定义、函数内积、函数范数、正交函数
  - 线性无关函数、最佳平方逼近函数、法方程

总结（续）

- 正交多项式
  - 带权$\rho(x)$正交、正交化
  - Legendre多项式、Chebyshev多项式
- 函数按正交多项式展开
  - 简化的计算过程、任意区间上的函数逼近
- 曲线拟合的最小二乘法
  - 数据拟合、最小二乘逼近、法方程
  - 用正交函数作最小二乘拟合、正交化