第七章　数据拟合与函数逼近

7.1　拟合与逼近

本章继续讨论用简单函数近似代替较复杂函数的问题. 上章提到的插值就是近似代替的方法之一, 插值的近似标准是在插值点处误差为零. 但在实际应用中, 有时不要求具体某些点误差为零, 而要求考虑整体的误差限制, 这就引出了拟合和逼近的概念.

7.1.1　数据拟合

对离散型函数(即数表形式的函数)考虑数据较多的情况. 若将每个点都当作插值节点, 则插值函数是一个次数很高的多项式, 比较复杂. 而且由于龙格振荡现象, 这个高次的插值多项式可能并不接近原函数. 同时由于数表中的点一般是由观察测量所得, 往往带有随机误差, 要求近似函数过所有的点既不现实也不必要.

如果不是要求近似函数过所有的数据点,而是要求它反映原函数整体的变化趋势,可得到更简单更适用的近似函数,这样的方法称为数据拟合.数据拟合最常用的近似标准是最小二乘法则：设$f(x)$为原函数,$\varphi(x)$为近似函数,$(x_i, f(x_i))(i=0,1,\ldots,n)$为数据点,要求选择$\varphi(x)$使

$$\sum_{i=0}^{n}\left[f(x_i)-\varphi(x_i)\right]^2$$

为最小.

当$\varphi(x)$选择为多项式时,称为多项式拟合.最小二乘拟合,特别是多项式拟合,是最流行的数据处理方法之一.它常用于把实验数据(离散的数据)归纳总结为经验公式(连续的函数),以利于进一步的推演分析或应用.

7.2 超定方程组的最小二乘解

设$Ax=b$中$A=(a_{ij})_{m\times n}$,b是m维已知向量,x是n维解向量.当$m>n$时,即方程组中方程的个数多于未知量个数时,称此方程组为超定方程组或矛盾方程组.一般说,超定方程组无解.但有时需要寻找一个“最近似”的解.记$r=b-Ax$,定义使$\|r\|_2$为最小的解x^*为$Ax=b$的最小二乘解.关于超定方程组的最小二乘解有如下定理:

定理7.1 x^*为$Ax=b$的最小二乘解的充要条件为 $A^TAx^*=A^Tb$.

证明(略)

以上定理说明求解超定方程组$Ax=b$的最小二乘解可转化为求解它对应的正规方程组$A^TAx^*=A^Tb$.A^TA是对称正定的系数阵,此方程组可用平方根法或SOR方法求解.

7.3 多项式拟合

仍假设有已知数据组$(x_i, y_i)(i=0,1,2,\ldots,m)$.现求作一个不超过$n(n<m)$次多项式

$$P_n(x)=\sum_{k=0}^{n} a_k x^k$$

使得 $\sum_{i=0}^{m}\left[y_i-\sum_{k=0}^{n} a_k x_i^k\right]^2$ 取最小 .

记 $r_i=y_i-P_n(x_i)$ $(i=0,1,2,\ldots,m)$, $r=(r_0,r_1,\ldots,r_m)^T$,不难看出以上多项式最小二乘拟合问题就是求解关于$a_k(k=0,1,\ldots,n)$的超定方程组.把a_k当作变量,上述方程组的矩阵记法为

$$\begin{pmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{pmatrix}\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}=\begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{pmatrix}$$

$$
\begin{pmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{pmatrix}
$$

是一个超定方程组.由定理7.1可得对应的正规方程组

$$
\begin{pmatrix} m+1 & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^n \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{n+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum x_i^n & \sum x_i^{n+1} & \sum x_i^{n+2} & \cdots & \sum x_i^{2n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \cdots \\ a_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \cdots \\ \sum x_i^n y_i \end{pmatrix}
$$

以上的$\sum$记号均为从**0**到m求和,记

$$
S_k = \sum_{i=0}^{m} x_i^k, \qquad t_k = \sum_{i=0}^{m} x_i^k\, y_i. \qquad (i = 0,1,2,\cdots,m, \qquad m > n)
$$

则上式可改写为

$$\begin{pmatrix} S_0 & S_1 & S_2 & \cdots & S_n \\ S_1 & S_2 & S_3 & \cdots & S_{n+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ S_n & S_{n+1} & S_{n+2} & \cdots & S_{2n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \cdots \\ a_n \end{pmatrix} = \begin{pmatrix} t_0 \\ t_1 \\ \cdots \\ t_n \end{pmatrix}$$

通过解该正规方程组便可解出a_k,从而确定出拟合多项式$P_n(x)$.
多项式拟合的一般方法可归纳为:

(1)根据具体问题,确定拟合多项式的次数n;

(2)计算 $$S_k = \sum_{i=0}^{m} x_i^k, \qquad t_k = \sum_{i=0}^{m} x_i^k y_i.$$

(3)写出正规方程组
(4)解正规方程组,求出$a_0,a_1,\ldots,a_n$;
(5)写出拟合多项式$P_n(x)$

设正规方程组的解为:

$$a_0 = a_0^*, a_1 = a_1^*, a_2 = a_2^*, \cdots, a_n = a_n^*$$

则以此解为系数的多项式

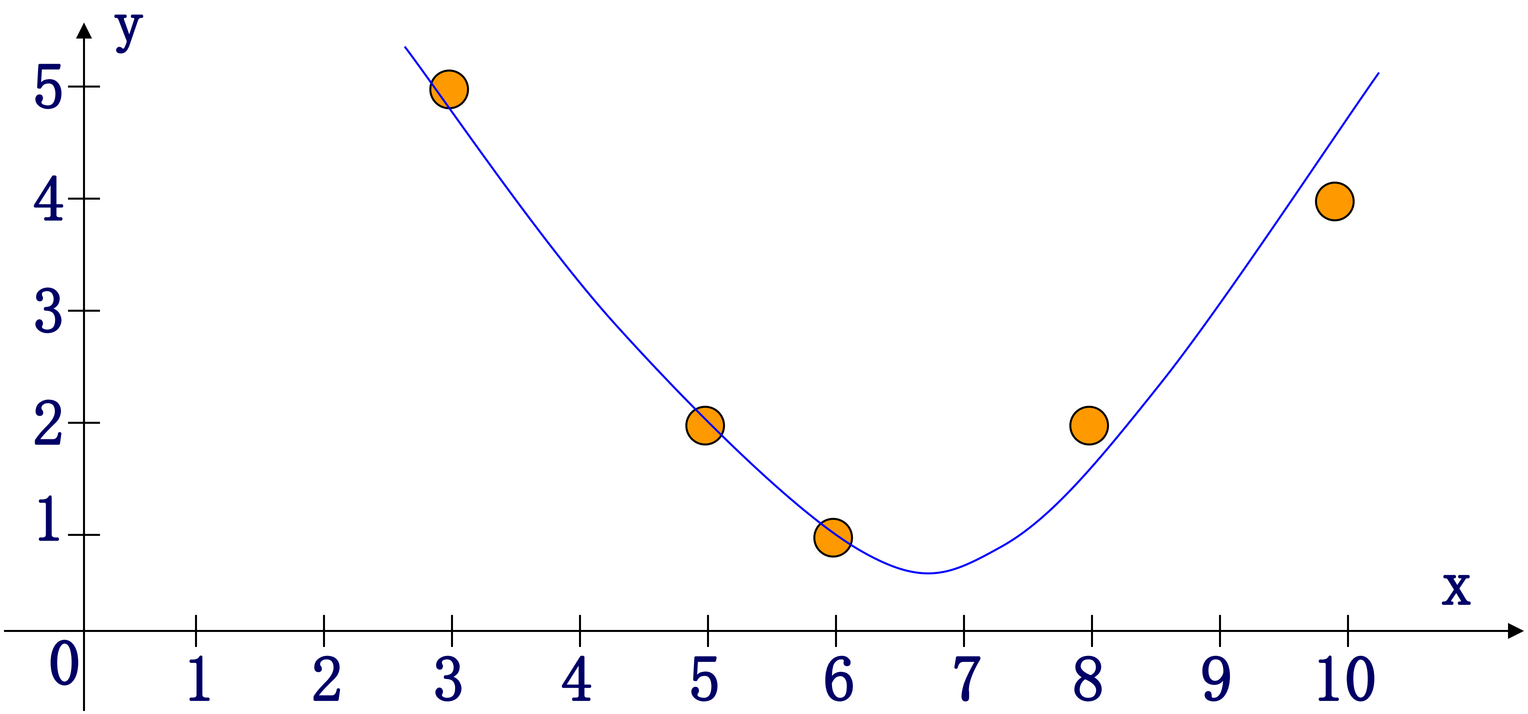
$$p_n^*(x) = a_0^* + a_1^* x + a_2^* x^2 + \cdots + a_n^* x^n$$

就是最小二乘拟合多项式。

例1　设5组数据如下表,用一多项式对其进行拟合。

x	3	5	6	8	10
y	5	2	1	2	4

解　首先作平面散点图如下：



从图中观察，这5个点大致在一条抛物线的附近，可考虑用二次多项式 $p_2(x)=a_0+a_1x+a_2x^2$ 进行拟合。

然后计算正则方程组(m=4)

$$
\begin{cases}
5a_0 + \left(\sum x_i\right)a_1 + \left(\sum x_i^2\right)a_2 = \sum y_i \\
\left(\sum x_i\right)a_0 + \left(\sum x_i^2\right)a_1 + \left(\sum x_i^3\right)a_2 = \sum x_i y_i \\
\left(\sum x_i^2\right)a_0 + \left(\sum x_i^3\right)a_1 + \left(\sum x_i^4\right)a_2 = \sum x_i^2 y_i
\end{cases}
$$

的系数如下表：

i	x_i	y_i	$x_i y_i$	x_i^2	$x_i^2 y_i$	x_i^3	x_i^4
0	3	5	15	9	45	27	81
1	5	2	10	25	50	125	625
2	6	1	6	36	36	216	1296
3	8	2	16	64	128	512	4096
4	10	4	40	100	400	1000	10000
Σ	32	14	87	234	659	1880	16098

正则方程组为
$$\begin{cases} 5a_0 + 32a_1 + 234a_2 = 14 \\ 32a_0 + 234a_1 + 1880a_2 = 97 \\ 234a_0 + 1880a_1 + 16098a_2 = 659 \end{cases}$$

用高斯-若当无回代消去法解此方程组, 得a_0=13. 454, a_1=−3. 657, a_2=0. 272。

最小二乘拟合多项式为:

$$y = p_2(x) = 13.454 - 3.657x + 0.272x^2$$

3. 非线性曲线转化为线性:

有些非线性曲线可以转化为线性, 从而用线性拟合进行处理, 比如:

$$y = \alpha \mathrm{e}^{\beta x} \to \ln y = \ln \alpha + \beta x$$

$$令 Y = \ln y, A = \ln \alpha \to Y = A + \beta x$$

又如：$y=\dfrac{x}{ax+b} \to \dfrac{1}{y}=a+\dfrac{b}{x}$

令 $Y=\dfrac{1}{y}, X=\dfrac{1}{x} \to Y=a+bX$

又如：$y=a+bx^2$

令 $X=x^2 \to y=a+bX$

例3：已知数据为

x	1	2	3	4	5	6	7	8
y	15.3	20.5	27.4	36.6	49.1	65.6	87.8	117.6

求一个形如 $y=ae^{bx}$ 的经验公式（a,b 为常数）.

解:两边取对数得: $\ln y = \ln a + bx$

$Y = \ln y, A = \ln a \Rightarrow Y = A + bx$

i	x_i	y_i	Y_i	x_i^2	x_iY_i
0	1	15.3	2.7279	1	2.7279
1	2	20.5	3.0204	4	6.0408
2	3	27.4	3.3105	9	9.9315
3	4	36.6	3.6000	16	14.4000
4	5	49.1	3.8939	25	19.4695
5	6	65.6	4.1836	36	25.1016
6	7	87.8	4.4751	49	31.3257
7	8	117.6	4.7673	64	38.1384
$\sum$	36		29.9787	204	147.1354

法方程组:

$$\begin{pmatrix} 8 & 36 \\ 36 & 204 \end{pmatrix}\begin{pmatrix} A \\ b \end{pmatrix} = \begin{pmatrix} 29.9787 \\ 147.1354 \end{pmatrix}$$

解该方程组的 A=2. 4368, b=0. 2912

由A=ln$a$, 即得a=e^A=11. 436 9
所以, 经验公式为：y=11. 4369e$^{0.2912x}$

§ 7.7 正交多项式系

不超过n次的多项式可以看成是幂函数基底｛$1,x,x^2,\ldots,x^n$｝的线性组合.如果选择基函数为正交多项式系.可以有更好的性质.
正交多项式系数值积分中也有重要用途.下面介绍正交多项式系.

7.7.1 正交函数系

定义7.3 某函数系｛$\varphi_k(x)$｝$k=0,\ldots,m$.中每一个函数$\varphi_k(x)$都在［a,b］上连续且不恒为零，如果满足

$$\left(\varphi_i(x),\varphi_j(x)\right)=\int_a^b \varphi_i(x)\varphi_j(x)dx\begin{cases}=0 & (i\neq j)\\ >0 & (i=j)\end{cases},\qquad(7.13)$$

则称此函数系｛$\varphi_k(x)$｝为区间［a,b］上的正交函数系.

更一般地，若有权函数$\rho(x)\geq 0$，即

$$\left(\varphi_i(x),\varphi_j(x)\right)=\int_a^b \rho(x)\varphi_i(x)\varphi_j(x)dx\begin{cases}=0 & (i\neq j)\\ >0 & (i=j)\end{cases},\qquad(7.14)$$

则称此函数系｛$\varphi_k(x)$｝,k=0,…,m,为区间［a,b］上关于权函数$\rho(x)$的正交函数系.

例11 三角函数系：1，cos$x$，sin$x$，cos2x，sin2x，…是区间［-π，π］上的正交函数系，因为

$$\int_{-\pi}^{\pi} \sin kx \sin jx dx = 0, \qquad (j \neq k)$$

$$\int_{-\pi}^{\pi} \cos kx \cos jx dx = 0, \qquad (j \neq k)$$

$$\int_{-\pi}^{\pi} \sin kx \cos jx dx = 0,$$

$$\int_{-\pi}^{\pi} \sin kx = \int_{-\pi}^{\pi} \cos kx = 0,$$

$$\int_{-\pi}^{\pi} \sin^2 kx = \int_{-\pi}^{\pi} \cos^2 kx = \pi > 0,$$

实际上，这就是付里叶(Fourier)逼近的基函数.

7.7.2 正交多项式系

如果正交函数系为多项式系｛$P_i(x)$｝i=0,…,m，则称之为正交多项式系，即

$$\left(P_i(x), P_j(x)\right) = \int_a^b \rho(x) P_i(x) P_j(x) dx \begin{cases} = 0 & (i \neq j) \\ > 0 & (i = j) \end{cases}, \qquad (7.15)$$

则称此多项式系为区间[a,b]上关于权函数$\rho(x)$的正交多项式系.

常见的正交多项式系有：

1.切比雪夫(Чебыщев)多项式系｛$T_k(x)$｝令$x=\cos\theta$,则
$T_n(x)=\cos n\theta$，它也可以用以下递推公式写出：

$T_0(x)=1$，$T_1(x)=x$

$T_{k+1}(x)=2xT_k(x)-T_{k-1}(x)$，$k\geq 1$　　(7.16)

它在[-1,1]上关于权函数$\rho(x)=1/(1-x)^2$正交，即

$$\left(T_i(x), T_j(x)\right) = \int_{-1}^{1} \frac{1}{1-x^2} T_i(x) T_j(x) dx \begin{cases} 0 & i \neq j \\ \pi/2 & i = j > 0 \\ \pi & i = j = 0 \end{cases} \qquad (7.16)$$

它的前7个多项式如下表：

$T_0(x)=1$

$T_1(x)=x$

$T_2(x)=2x^2-1$

$T_3(x)=4x^3-3x$

$T_4(x)=8x^4-8x^2+1$

$T_5(x)=16x^5-20x^3+5x$

$T_6(x)=32x^6-48x^4+18x^2-1$

2.勒让德(Legendre)多项式系 $\{P_k(x)\}$

$$P_n(x)=\frac{1}{2^n\times n!}\frac{d^n}{dx^n}\left(x^2-1\right)^n,$$

它也可由如下递推公式写出：

$P_0(x)=1$，$P_1(x)=x$

$P_{k+1}(x)=[(2k+1)/(k+1)]\,x\,P_k(x)-[k/(k+1)]\,P_{k-1}(x)$，$k\geq 1$ (7.17)

它在［-1,1］上关于权函数$\rho(x)\equiv 1$正交，即

$$\left(P_i(x), P_j(x)\right) = \int_{-1}^{1} P_i(x)P_j(x)dx \begin{cases} = 0 & (i \neq j) \\ = \dfrac{1}{2n+1} & (i = j)' \end{cases} \qquad (7.17)$$

它的前7式如下表：

$P_0(x)=1$

$P_1(x)=x$

$P_2(x)=(3x^2-1)/2$

$P_3(x)=(5x^3-3x)/2$

$P_4(x)=(35x^4-30x^2+3)/8$

$P_5(x)=(63x^5-70x^3+15x)/8$

$P_6(x)=(231x^6-315x^4+105x^2-5)/16$

3.拉盖尔(Laguerre)多项式系｛$L_k(x)$｝

$$L_k(x)=\mathrm{e}^x\frac{d^k}{dx^k}\left(x^n\mathrm{e}^{-x}\right),$$

它也可用如下递推公式写出：$L_0(x)=1$，$L_1(x)=1-x$

$L_{k+1}(x)=(2n+1-x)L_k(x)-n^2L_{k-1}(x)$，$k\geq 1$ (7.18)

它在[0，+∞]上关于权函数$\rho(x)=\mathrm{e}^{-x}$正交，即

$$\left(L_i(x),L_j(x)\right)=\int_0^{+\infty}\mathrm{e}^{-x}L_i(x)L_j(x)dx\begin{cases}=0 & (i\neq j)\\ =(n!)^2 & (i=j)\end{cases},$$

它的前7式如下：

$L_0(x)=1$

$L_1(x)=1-x$

$L_2(x)=2-4x+x^2$

$L_3(x)=6-18x+9x^2-x^3$

$L_4(x)=24-96x+72x^2-16x^3+x^4$

$L_5(x)=120-600x+600x^2-200x^3+25x^4-x^5$

$L_6(x)=720-4320x+5400x^2-2400x^3+450x^4-36x^5+x^6$

4. 埃尔米特(Hermite)多项式系 $\{H_k(x)\}$

$$H_k(x) = (-1)^k \mathrm{e}^{x^2} \frac{d^k}{dx^k} \mathrm{e}^{-x^2},$$

它可由如下**递推公式**写出

$H_0(x)=1$， $H_1(x)=2x$

$H_{k+1}(x)=2xH_k(x)-2kH_{k-1}(x)$， $k\geqslant 1$ (7.19)

它在$(-\infty, +\infty)$上关于权函数$\rho(x)=\mathrm{e}^{-x^2}$正交，即

$$\big(H_i(x), H_j(x)\big) = \int_{-\infty}^{+\infty} \mathrm{e}^{-x^2} H_i(x) H_j(x) dx \begin{cases} = 0 & (i \neq j) \\ = 2n \times n! \times \sqrt{\pi} & (i = j) \end{cases},$$

它的前7式如下表：

$H_0(x)=1$，　$H_1(x)=2x$，$H_2(x)=4x^2-2$

$H_3(x)=8x^3-12x$

$H_4(x)=16x^4-48x^2+12$

$H_5(x)=32x^5-160x^3+120x$

$H_6(x)=64x^6-480x^4+720x^2-120$

正交多项式还有一个重要性质：以上各种正交多项式的零点全部是单实根，且都分布在它的正交定义区间内.这个性质在数值积分中有应用.

7.5 最佳一致逼近多项式

7.5.1 线性赋范空间

定义7.1 在线性空间X中,对每一个元x引进一个度量$\|x\|$, 称为x的范数,这时线性空间X就称为线性赋范空间,若 $y\in X$,则称$\|x-y\|$为元素x与y的距离.

例7 在R^3中,任一$x\in R^3$定义范数

$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2}$$

$$\|x - y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$$

常称之为三维欧几里德空间,就是一个线性赋范空间.

例8 在连续函数空间C[*a*,*b*]中.定义范数,设$f\in C[a,b]$,

$$\|f\|_\infty = \max_{x\in[a,b]} |f(x)|$$

则C[*a*,*b*]成为线性赋范空间.

设$M\subset C[a,b]$,M是C[*a*,*b*]的一个子空间,

寻找$\varphi(x)\in M$,使得

$$\|f-\varphi\|_\infty = \max_{x\in[a,b]} |f(x)-\varphi(x)|$$

为最小, 则称$\varphi(x)$为M中对$f(x)$的最佳一致逼近函数.

7.5.2 最佳一致逼近多项式

定义7.2 设$f(x)\in C[a,b]$, $\|f\|_\infty$已定义, M_n为不超过n次的多项式的集合,M_n显然是$C[a,b]$的子空间,寻找$P_n(x)\in M_n$,使

$$\|f(x)-P_n(x)\|_\infty=\max_{x\in[a,b]}|f(x)-P_n(x)|$$

为最小， 则称$P_n(x)$为M_n中对$f(x)$的n次最佳一致逼近多项式.

以下定理保证了最佳一致逼近多项式的存在唯一性.

定理7.1 设$f(x)\in C[a,b]$,M_n为不超过n次的多项式的集合,则存在唯一的

$$P_n^*(x)\in M_n,\text{使得}\|f-P_n^*\|_\infty=\inf_{P_n\in M_n}\|f-P_n\|_\infty=\min_{P_n\in M_n}\|f-P_n\|_\infty$$

$$=\min_{P_n\in M_n}\max_{x\in[a,b]}|f(x)-P_n(x)|$$

这个定理的证明比较复杂,本书从略.

7.5.3 最佳一致逼近多项式的特征

为讨论最佳一致逼近多项式的构造方法,先讨论它的特征.

先给出偏差点的概念:

对连续函数$f(x)$和$p(x)$,其差的绝对值也必连续,故在$[a,b]$上存在x_0,使

$$|f(x_0)-p(x_0)|=\max_{x\in[a,b]}|f(x)-p(x)|=\|f-p\|_\infty$$

x_0称为$p(x)$对$f(x)$的偏差点.

若$p(x_0)-f(x_0)>0$,称x_0为正偏差点

若$p(x_0)-f(x_0)<0$,称x_0为负偏差点

切比晓夫给出了最佳一致逼近多项式的特征定理.

定理7.2 $P_n(x) \in \mathrm{M_n}$, $P_n(x)$为最佳一致逼近多项式的充要条件是

$P_n(x)$在$[a,b]$上至少有$n+2$个交错偏差点,即

$a \le x_0 < x_1 < \ldots < x_n < x_{n+1} \le b$ 使得

$$P_n(x_i) - f(x_i) = (-1)^i \sigma \|f - P_n\|_\infty$$

σ为1或−1.

这个定理本书不证明.

定理7.2说明$P_n(x_i) - f(x_i)$ 至少在$n+2$个点上交错变号,都达到最大偏差幅度$\|f - P_n\|_\infty$, 因此在整个$[a,b]$上误差分布比较均匀.

求最佳一致逼近多项式是十分困难的,以下定理可解决在某些简单情况下的求解.