《数值分析》之

函数逼近

徐岩、夏银华

中国科学技术大学数学系

yxu@ustc.edu.cn, yhxia@ustc.edu.cn

https://bb.ustc.edu.cn

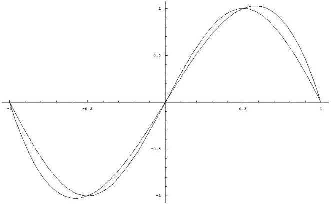
$$f(x) = \frac{1}{1+x^2}, x \in [-5, 5]$$

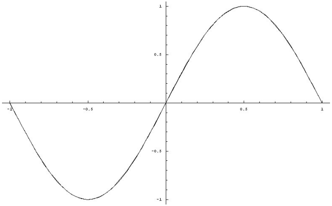
等距插值函数$p_n(x)$.

Degree n	Max error
2	0.65
4	0.44
6	0.61
8	1.04
10	1.92
12	3.66
14	7.15
16	14.25
18	28.74
20	58.59
22	121.02
24	252.78

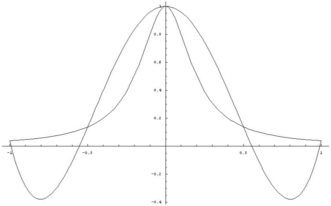
Runge现象

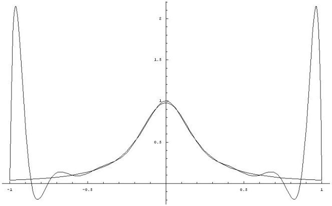
- $f(x) = \sin \pi x$, $[-1, 1]$上的5个和16个等距结点





- $f(x) = \frac{1}{(25x^2+1)}$, $[-1, 1]$上的5个和16个等距结点





此例由Runge在1901年给出

Runge现象（续）

- n 越大，端点附近抖动越大
- 等距高次插值，数值稳定性差，本身是病态的。
- 考虑使用分段插值

对给定区间$[a,b]$作分割

$$a = x_0 < x_1 < \cdots < x_n = b$$

在每个小区间$[x_i, x_{i+1}]$上，作线性插值函数$p(x) = p_i(x)$

$$p_i(x) = \frac{x - x_{i+1}}{x_i - x_{i+1}} f(x_i) + \frac{x - x_i}{x_{i+1} - x_i} f(x_{i+1}), \; x \in [x_i, x_{i+1}]$$

满足

- $p(x)$连续，即保证函数在$x = x_0, x_1, \cdots, x_n$点连续
- $p(x)$在每个小区间上为一个不高于1次的多项式

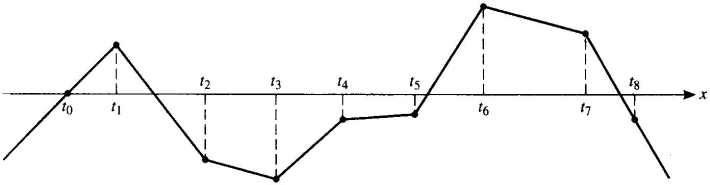
当$x \in [x_i, x_{i+1}]$时，

$$f(x) - p(x) = f(x) - p_i(x) = \frac{f^{(2)}}{2}(x - x_i)(x - x_{i+1})$$
$$\leq \frac{M_2}{8}(x_{i+1} - x_i)^2, \quad M_2 = \max_{a \leq x \leq b} |f''(x)|$$

由此得

$$|f(x) - p(x)| = \max_n \{\frac{M_2}{8}(x_{i+1} - x_i)^2\} = \frac{M_2}{8} h^2$$

当$h \to 0$时，$p(x)$收敛于$f(x)$.



- 局部性质，如果修改了某节点$(x_i, f(x_i))$的值，仅在相邻的两个区间$[x_{i-1}, x_i]$, $[x_i, x_{i+1}]$需要改动
- 插值节点处仅连续，不光滑
- 类似，可以作二重Hermite插值

样条函数

- 样条函数是一类分段（片）光滑、并且在各段交接处也有一定光滑性的函数。简称样条(spline)。
- 样条一词来源于工程绘图人员为了将一些指定点连接成一条光顺曲线所使用的工具，即富有弹性的细木条或薄钢条。由这样的样条形成的曲线在连接点处具有连续的坡度与曲率。分段低次多项式、在分段处具有一定光滑性的函数插值就是模拟以上原理发展起来的，它克服了高次多项式插值可能出现的振荡现象，具有较好的数值稳定性和收敛性，由这种插值过程产生的函数就是多项式样条函数。

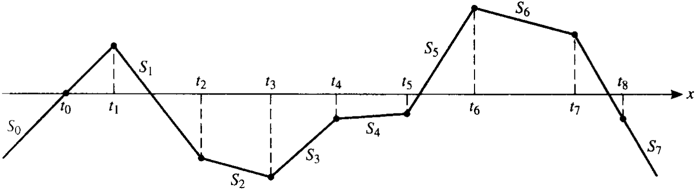
- 样条函数的研究始于20世纪中叶的概率论和数理统计中的数据拟合，到了60年代它与计算机辅助设计相结合，在外形设计方面得到成功的应用。
- 样条理论已成为函数逼近的有力工具。它的应用范围也在不断扩大，不仅在数据处理、数值微分、数值积分、微分方程和积分方程数值解等数学领域有广泛的应用，而且与最优控制、变分问题、统计学、计算几何与泛函分析等学科均有密切的联系

- 给定$n+1$个点$t_0 < t_1 < \cdots < t_n$ (称为结点, knots), 并指定一个非负整数$k \geqslant 0$. 在这些结点上定义的一个k次样条函数(spline function)是指满足下列条件的函数S:
  1. 在每个区间$[t_{i-1}, t_i)$上, S是一个次数不超过k的多项式。
  2. 在$[t_0, t_n]$上S有$(k-1)$阶连续导数。
- 当固定结点以及次数k, 那么所有的S在通常函数运算的意义下构成一个线性空间。自然的问题是这个样条空间的维数是多少? 基函数是什么? 这就是样条理论的核心。其中对基函数的一般要求是:
  1. 非负性
  2. 单位剖分
  3. 局部支集

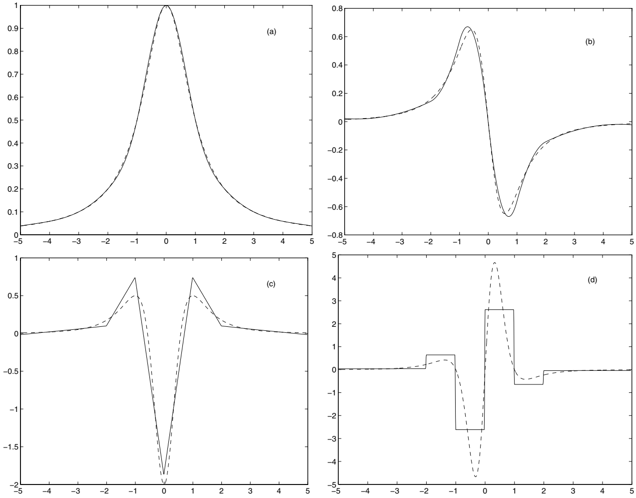
- $k=0$: 分段常数

$$
S(x)=\begin{cases}
S_0(x)=c_0 & x\in[t_0,t_1)\\
S_1(x)=c_1 & x\in[t_1,t_2)\\
\vdots & \vdots\\
S_{n-1}(x)=c_{n-1} & x\in[t_{n-1},t_n]
\end{cases}
$$

- $k=1$: 分段折线函数，零阶连续



$\frac{1}{1+x^2}$：三次样条插值



函数值，一阶导数，二阶导数，三阶导数。

- 给定数据点$(x_0, y_0), \ldots, (x_n, y_n)$，其中$x_0 < \cdots < x_n$. 计算一个以$x_0, \ldots, x_n$为结点的三次样条函数$S(x)$使得，
  - $S(x_i) = y_i, \quad i = 0, 1, \cdots, n$
  - 在每个小区间$[x_i, x_{i+1}]$上至多为三次多项式
  - $S(x)$在$[a, b]$上有连续的二阶导数
- 条件分析：
  1. 待定每个区间上的三次多项式形式，共有$4n$个自由系数
  2. 每个子区间上的三次多项式在两个端点有两个函数值约束，即$S(x_i) = y_i$, $S(x_{i+1}) = y_{i+1}$, 因此共有$2n$个约束
  3. 在每一个内结点上，S的一阶和二阶导数连续，分别给出$n-1$个约束

$2n + n - 1 + n - 1 = 4n - 2$

因此还有两个(??，有待确认)自由度，通常在边界处给出

- 记待求样条在结点的二阶导数值为$M_i = S''(x_i)$
- $S_i''(x)$是区间$[x_i, x_{i+1}]$上的线性函数，因此

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

- 积分两次，再结合$S_i(x_i) = y_i$和$S_i(x_{i+1}) = y_{i+1}$:

$$\begin{aligned} S_i(x) =& \frac{M_i}{6h_i}(x_{i+1} - x)^3 + \frac{M_{i+1}}{6h_i}(x - x_i)^3 \\ &+ \left(\frac{y_{i+1}}{h_i} - \frac{M_{i+1}h_i}{6}\right)(x - x_i) + \left(\frac{y_i}{h_i} - \frac{M_i h_i}{6}\right)(x_{i+1} - x) \end{aligned}$$

- 可以用S'的连续性确定M_i, $i=1,\ldots,n-1$. 由前页结果,

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

根据$S_{i-1}'(x_i) = S_i'(x_i)$可得

$$h_{i-1}M_{i-1}+2(h_i+h_{i-1})M_i+h_iM_{i+1} = \frac{6}{h_i}(y_{i+1}-y_i)-\frac{6}{h_{i-1}}(y_i-y_{i-1})$$

对$i=1,\ldots,n-1$成立

- 从而对$M_0,\ldots,M_n$给出了$n-1$个线性条件。可以任意指定M_0和M_n以求得$M_1,\ldots,M_{n-1}$, 从而确定相应的样条

- 当取$M_0 = M_n = 0$时，得到的样条称为三次自然样条(natural cubic spline)
- 此时求解$M_1, \ldots, M_{n-1}$对应的线性方程组为

$$\begin{pmatrix} u_1 & h_1 & & & & \\ h_1 & u_2 & h_2 & & & \\ & h_2 & u_3 & h_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & h_{n-3} & u_{n-2} & h_{n-2} \\ & & & & h_{n-2} & u_{n-1} \end{pmatrix} \begin{pmatrix} M_1 \\ M_2 \\ M_3 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_{n-2} \\ v_{n-1} \end{pmatrix}$$

其中$h_i = t_{i+1} - t_i$, $u_i = 2(h_i + h_{i+1})$, $b_i = 6(y_{i+1} - y_i)/h_i$, $v_i = b_i - b_{i-1}$. 系数阵是对称的、三对角的、对角占优的

- 固定边界: 当取$S'(x_0) = m_0$, $S'(x_n) = m_n$时，此时边界处满足

$$2M_0 + M_1 = \frac{6}{h_0}\left(\frac{y_1 - y_0}{h_0} - m_0\right) = v_0$$
$$M_{n-1} + 2M_n = \frac{6}{h_{n-1}}\left(m_n - \frac{y_n - y_{n-1}}{h_{n-1}}\right) = v_n$$

此时求解$M_0, \ldots, M_n$对应的$n+1$阶线性方程组为

- 周期边界： $m_0 = m_n$. $M_0 = M_n$.

当确定系数$M_0, \ldots, M_n$后，对应的三次样条函数在任意点x处的值可如下计算：

1. 确定x是在下述哪个区间中(最有效方法是什么？)：

$$(-\infty, t_1), [t_1, t_2), \ldots, [t_{n-2}, t_{n-1}), [t_{n-1}, +\infty)$$

   假设区间指标为i

2. 重写$S_i(x)$的表达式为

$$S_i(x) = y_i + (x - t_i)[C_i + (x - t_i)[B_i + (x - t_i)A_i]]$$

   其中

$$A_i = \frac{1}{6h_i}(M_{i+1} - M_i)$$

$$B_i = \frac{M_i}{2}$$

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

Theorem (三次自然样条最优性定理)

设$f \in C^2[a,b]$, $a = t_0 < t_1 < \cdots < t_n = b$。若$S$是$f$在结点$t_0, \ldots, t_n$上的三次自然插值样条，则

$$\int_a^b \left(S''(x)\right)^2 dx \leqslant \int_a^b \left(f''(x)\right)^2 dx$$

证明：令$g = f - S$，则$g(t_i) = 0$, $i = 0, \ldots, n$, 并且

$$\int_a^b (f'')^2 dx = \int_a^b (S'')^2 dx + \int_a^b (g'')^2 dx + 2\int_a^b S''g''dx$$

根据分部积分：

$$\begin{aligned}\int_a^b S''g''dx &= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} S''g''dx \\ &= \sum_{i=1}^n \left\{(S''g')(t_i) - (S''g')(t_{i-1}) - \int_{t_{i-1}}^{t_i} S'''g'dx\right\}\end{aligned}$$

$$
\begin{aligned}
&= -\sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} S''' g' dx \\
&= -\sum_{i=1}^{n} c_i \int_{t_{i-1}}^{t_i} g' dx \\
&= -\sum_{i=1}^{n} c_i [g(t_i) - g(t_{i-1})] \\
&= 0
\end{aligned}
$$

$\square$

自然样条最优性分析

- 由于$y=f(x)$定义曲线的曲率为$|f''(x)|\left[1+\left(f'(x)\right)^2\right]^{-3/2}$，因此可以认为三次自然样条是一条具有最小近似曲率的曲线
- 定理证明中应用到了下述和式：

$$\sum_{i=1}^{n}\left\{(S''g')(t_i)-(S''g')(t_{i-1})\right\}=(S''g')(b)-(S''g')(a)$$

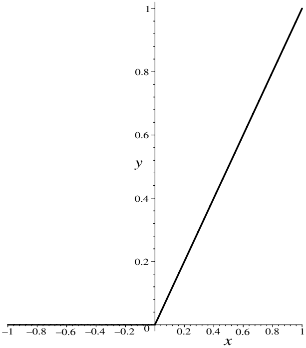
证明中这个和式为零。实际上只要它非负，定理仍然成立。令$S'(a)=f'(a)$, $S'(b)=f'(b)$也是一种选择的方法。

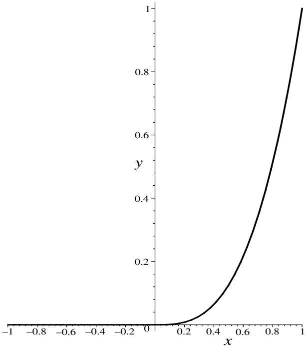
- 三次自然样条是完全来自于工程实际，我们可以对此进行推广，定义奇数次自然样条
- 给定结点集$t_0 < t_1 < \cdots < t_n$，一个$2m+1$次自然样条是一个函数$S \in C^{2m}(\mathbb{R})$，在每一个区间$[t_i, t_{i+1}]$内它是一个次数不超过$2m+1$的多项式，而在区间$(-\infty, t_0)$和$(t_n, +\infty)$内为一个次数至多为m的多项式。在给定结点上的全体$2m+1$次自然样条构成的线性空间记为$\mathcal{N}^{2m+1}(t_0, t_1, \ldots, t_n)$或$\mathcal{N}_n^{2m+1}$
- 三次自然样条符合上述定义，此时在区间$(-\infty, t_0)$和$(t_n, +\infty)$内延拓定义为线性多项式

- 截断幂函数：

$$x_+^n = \begin{cases} x^n & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

它是属于C^{n-1}函数类的





Theorem

$\mathcal{N}_n^{2m+1}$中的每个元素S可以表示为

$$S(x) = \sum_{i=0}^{m} a_i x^i + \sum_{i=0}^{n} b_i (x - t_i)_+^{2m+1}$$

其中对于$0 \leqslant j \leqslant m$, $\sum_{i=0}^{n} b_i t_i^j = 0$.

证明

- 在$(-\infty, t_0)$内，$a_0, \ldots, a_m$被唯一确定。
- 在$[t_i, t_{i+1})$内S为一个$2m+1$次多项式，其中t_i点的直至$2m$阶导数已确定，因此存在b_i使得

$$S(x) = \sum_{i=0}^{m} a_i x^i + \sum_{i=0}^{n} b_i (x - t_i)_+^{2m+1}$$

- 在区间$(t_n, +\infty)$上，S为次数$\leqslant m$的多项式，因此

$$0 = S^{(m+1)}(x) = \sum_{i=0}^{n} b_i (2m+1)(2m)\cdots(m+1)(x-t_i)^m, \quad x > t_n$$

$$0 = \sum_{i=0}^{n} b_i (x - t_i)^m = \sum_{i=0}^{n} b_i \sum_{j=0}^{m} \binom{m}{j} x^{m-j} (-t_i)^j$$

$$= \sum_{j=0}^{m} \left(\sum_{i=0}^{n} b_i t_i^j \right) (-1)^j \binom{m}{j} x^{m-j}$$

Theorem

给定结点$t_0 < t_1 < \cdots < t_n$, $0 \leqslant m \leqslant n$, 则存在唯一的$2m+1$次自然样条在这些结点上取给定值。

证明：根据自然样条表示的定理得到如下方程组

$$\begin{cases} S(t_i) = \sum_{j=0}^{m} a_j t_i^j + \sum_{j=0}^{n} b_j (t_i - t_j)_+^{2m+1} = \lambda_i, \quad 0 \leqslant i \leqslant n \\ \sum_{j=0}^{n} b_j t_j^i = 0, \quad 0 \leqslant i \leqslant m \end{cases}$$

有$m+n+2$个方程，$m+n+2$个未知数。为证方程组有唯一解，只需要证对应的齐次方程组仅有零解。

假设对$i=0,1,\ldots,n$, $S(t_i)=0$，下证

$$I=\int_{t_0}^{t_n}\left(S^{(m+1)}(x)\right)^2 dx=0 \tag{1}$$

实际上，

$$I=S^{(m+1)}(x)S^{(m)}(x)\Big|_{t_0}^{t_n}-\int_{t_0}^{t_n}S^{(m)}(x)S^{(m+2)}(x)dx$$

$$=-\int_{t_0}^{t_n}s^{(m)}(x)S^{(m+2)}(x)dx=\cdots=(-1)^m\int_{t_0}^{t_n}S'(x)S^{(2m+1)}(x)dx$$

$$=(-1)^m\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}c_iS'(x)dx=(-1)^m\sum_{i=1}^{n}c_i(S(t_i)-S(t_{i-1}))=0$$

因(1)式可知$S^{m+1}(x)\equiv 0$, 即S为一个次数至多m的多项式，其有零点$t_0,\ldots,t_n$，$n+1>m$，因此$S(x)$为零函数 □

Theorem

设$m \leqslant n$, $f \in C^{m+1}[a, b]$, S为在结点$a = t_0 < t_1 < \cdots < t_n = b$上插值$f$的$2m+1$次自然样条，则

$$\int_a^b \left(S^{(m+1)}(x)\right)^2 dx \leqslant \int_a^b \left(f^{(m+1)}(x)\right)^2 dx$$

证明：令$g = f - S$，则可以证明

$$\int_a^b g^{(m+1)}(x) S^{(m+1)}(x) dx = 0$$

根据这一正交性可以证明所需结论。 $\square$

在结点$a = t_0 < t_1 < \cdots < t_n = b$ 上定义函数$f \in C^1[a, b]$的Hermite三次样条$S_H(x) \in C^1[a, b]$，使得

1. $S_H(t_i) = f(t_i),\ S_H'(t_i) = f'(t_i),\ i = 0, \cdots n,$
2. S_H 在区间$[t_{i-1}, t_i],\ i = 1, \cdots n$是三次多项式。

Theorem

设$f \in C^4[a, b]$, S_H为在结点$a = t_0 < t_1 < \cdots < t_n = b$上关于$f$的Hermite三次样条，则

$$||f - S_H||_\infty \le \frac{1}{384} h^4 ||f^{(4)}||_\infty,$$

其中$h = \max_i (x_i - x_{i-1})$.

- B样条(B-splines)是给定样条空间的一组特殊基，从而其它样条函数都可以写成它的线性组合
- B样条具有许多很好的性质，如：
  - 局部支集
  - 非负性
  - 单位剖分，即所有的基函数的和恒等于1
  - 定义和计算简单
- 在B样条理论中，有许多方法可以作为定义B样条，如借助于截断幂函数的均差方法；借助于多重线性组合的递归方法和Blossoming方法[1]，等等

[1]在R. Goldman所著《金字塔算法》一书的前言中提到，“太早介绍blossoming，会出现一些问题。一是blossoming太有用了。晚接触它，可以多学一些其它的方法；二是只有亲眼看到blossoming可以取代那些毫无关联的方法技巧后，才会真正体会到blossoming的用处。”

- 0次B样条：

$$B_i^0(x) = \begin{cases} 1 & t_i \leqslant x < t_{i+1} \\ 0 & \text{其它} \end{cases}$$

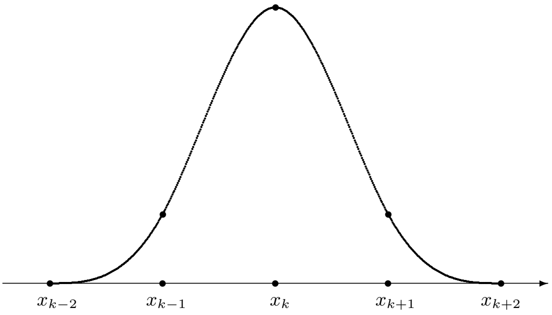
- $k \geqslant 1$:

$$B_i^k(x) = \frac{x - t_i}{t_{i+k} - t_i} B_i^{k-1}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} B_{i+1}^{k-1}(x)$$

例：一次B样条

$$B_i^1(x)=\begin{cases}0 & x<t_i\text{或}x\geqslant t_{i+2}\\ \frac{x-t_i}{t_{i+1}-t_i} & t_i\leqslant x<t_{i+1}\\ \frac{t_{i+2}-x}{t_{i+2}-t_{i+1}} & t_{i+1}\leqslant x<t_{i+2}\end{cases}$$

例：三次B样条



B样条性质

- $B_i^k(x)$的支集为$[t_i, t_{i+k+1}]$
- $B_i^k(x) \geqslant 0.$
- $\sum\limits_{i=-\infty}^{\infty} B_i^k(x) = 1$
- $$\frac{d}{dx}B_i^k(x) = \frac{k}{t_{i+k}-t_i}B_i^{k-1}(x) - \frac{k}{t_{i+k+1}-t_{i+1}}B_{i+1}^{k-1}(x)$$
- $$\int_{-\infty}^{x} B_i^k(s)ds = \frac{t_{i+k+1}-t_i}{k+1}\sum_{j=i}^{\infty} B_j^{k+1}(x)$$
- $\{B_j^k, \ldots, B_{j+k}^k\}$在$(t_{k+j}, t_{k+j+1})$上线性无关，
  $\{B_{-k}^k, \ldots, B_{n-1}^k\}$在$(t_0, t_n)$上线性无关

样条函数求值

- B样条函数构成相应样条空间的一组基，若

$$S(x) = \sum_i c_i^k B_i^k(x)$$

  则称c_i^k为相应于$B_i^k(x)$的控制系数
- 为了计算函数$S(x)$在一点的值，先计算出每个基函数在点x的值，再进行线性组合。这是一种效率很低的算法。实际上B样条函数标准的求值算法，是de Boor算法

de Boor算法

确定指标m使得$t_m \leqslant x < t_{m+1}$

$$\begin{array}{lllll}
c_m^k & c_m^{k-1} & \cdots & c_m^1 & c_m^0 = S(x) \\
c_{m-1}^k & c_{m-1}^{k-1} & \cdots & c_{m-1}^1 & \\
\vdots & \vdots & \iddots & & \\
c_{m-k+1}^k & c_{m-k+1}^{k-1} & & & \\
c_{m-k}^k & & & &
\end{array}$$

其中

$$c_i^{j-1} = \frac{1}{t_{i+j} - t_i}\Big((x - t_i)c_i^j + (t_{i+j} - x)c_{i-1}^j\Big)$$

上机作业1

对函数

$$f(x) = e^x, x \in [0,1]$$

构造等距节点的样条插值函数，对以下两种类型的样条函数

1. 一次分片线性样条
2. 满足$S'(0) = 1$, $S'(1) = e$的三次样条

并计算如下误差

$$\max_i \{|f(x_{i-\frac{1}{2}}) - S(x_{i-\frac{1}{2}})|, \quad i = 1, \cdots, N\}$$

这里$x_{i-\frac{1}{2}}$为每个小区间的中点。对$N = 5, 10, 20, 40$比较以上两组节点的结果。讨论你的结果。利用公式计算算法的收敛阶。

$$Ord = \frac{\ln(Error_{old}/Error_{now})}{\ln(N_{now}/N_{old})}$$

输出形式如下：

n	Method (1) error	order	Method (2) error	order
5		–		–
10				
20				
40				

上机作业2

对函数

$$f(x)=1/(1+x^2), x\in[0,5]$$

- 在节点$x=0,5/3,10/3,5$ 构造线性样条、自然三次样条和三次Hermite样条插值函数，画出误差对比图并解释你所看到的现象。
- 类似上机作业1，加密网格给出三种样条插值函数的误差表。