

第四章作业

1 问题叙述

由于钢水对材料的侵蚀作用，随着使用次数的增加，炼钢厂出钢时所用盛钢水的钢包的容积不断增大。现希望获得增大容积 y 与使用次数 x 之间的函数关系，实测得到如下数据：

使用次数 x	2	3	4	5	6	7	8	9
增大容积 y	6.7	8.2	9.58	9.5	9.7	10	9.96	9.99
使用次数 x	10	11	12	13	14	15	16	
增大容积 y	10.49	10.59	10.6	10.8	10.6	10.9	10.76	

图 1: 问题叙述

- 请用插值方法建立 y 与 x 之间的函数关系，画出散点图和插值函数曲线，从结果说明插值方法是否适合该问题。
- 请分别按从 ① $\frac{1}{y} = a + \frac{b}{x}$ ② $y = ae^{\frac{b}{x}}$ 两种形式拟合建立 y 与 x 之间的函数关系，画出散点图和拟合函数曲线，并根据你选定的适当指标判断哪一种形式更好。

2 问题分析

首先绘制出本问题的散点图：

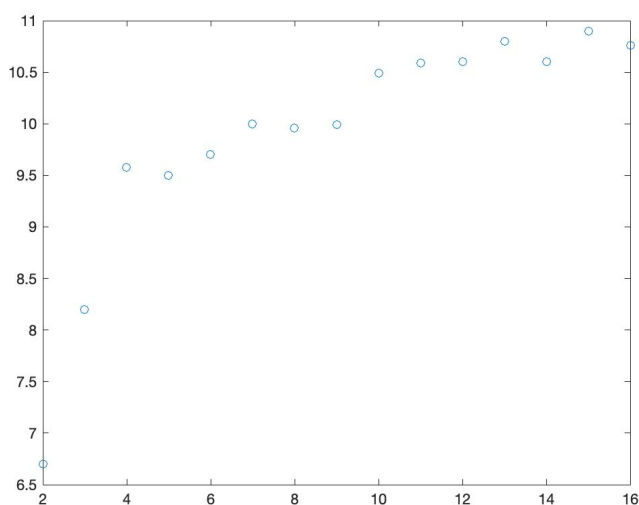


图 2: 散点图

可以看出, y 大致随 x 单调递增, 同时 x 越大, 增速越慢。

对于第一问, 使用多项式插值(拉格朗日插值、牛顿插值)、分段低次插值、样条插值三种方法对原函数进行估计, 其中分段低次插值尝试使用分段一次插值和分段二次插值, 样条插值使用二次样条插值和三次样条插值。通过绘制出对应函数图像来判断插值方法的优劣程度, 同时判断差之方法是否适合该问题。

对于第二问, 可对两个形式的函数进行适当变形转换为一次函数拟合问题。通过计算 MSE (均方误差):

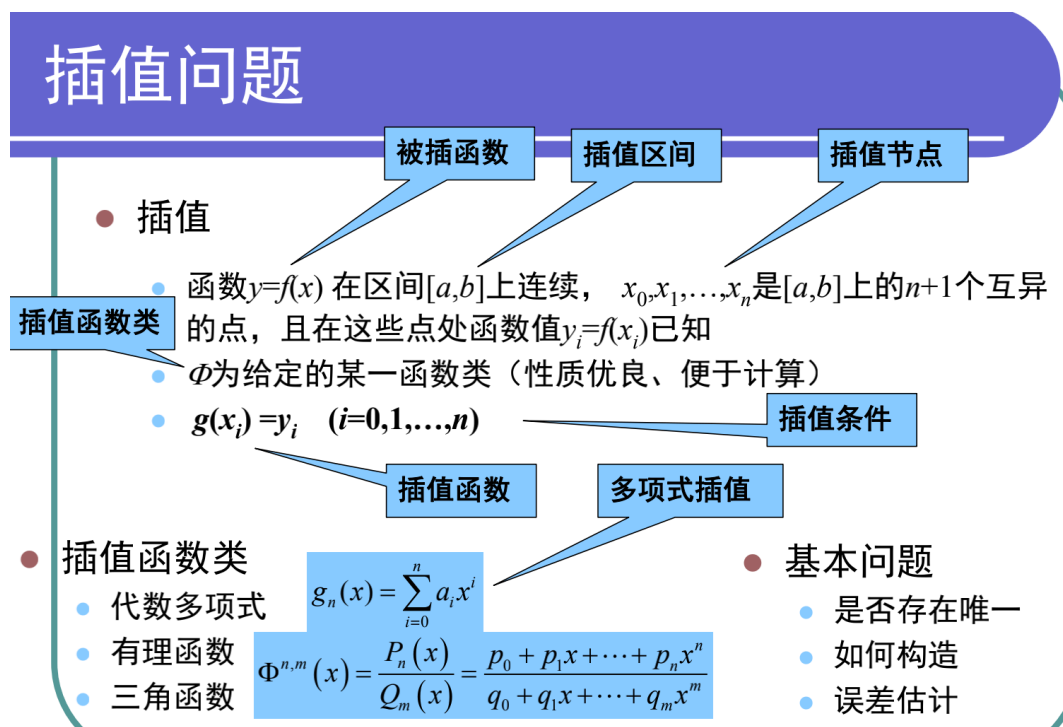
$$MSE = \frac{1}{n} \sum (\hat{y}_i - y_i)^2 \quad (1)$$

来估计拟合的优劣程度, 均方误差越小的形式拟合效果越好。

目录

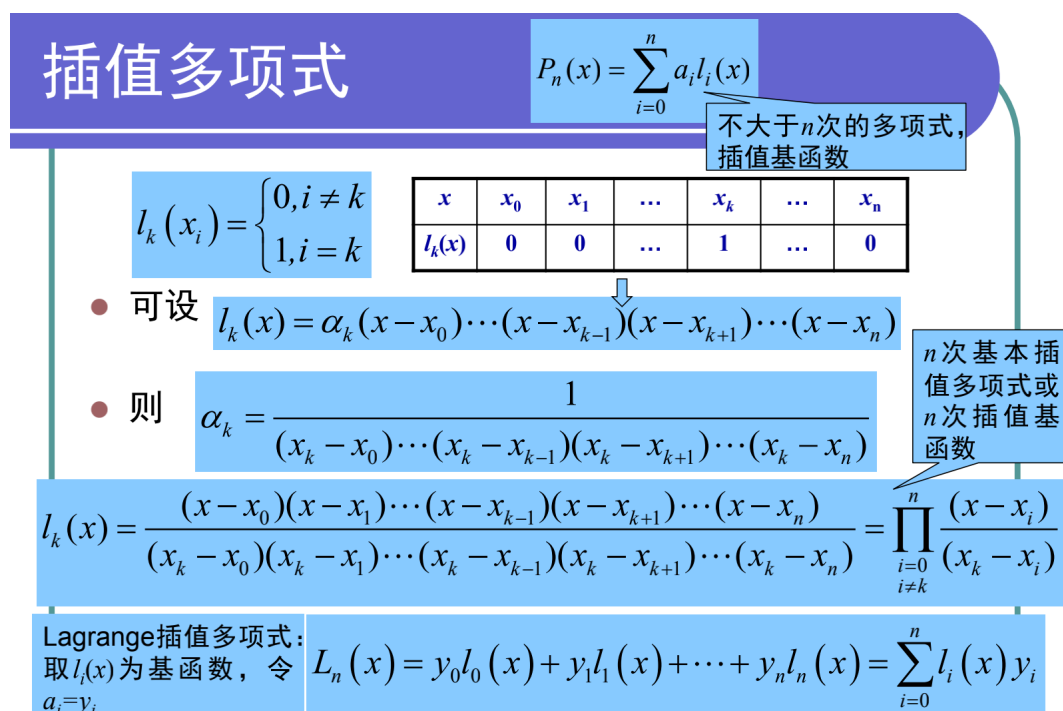
1 问题叙述	1
2 问题分析	1
3 插值法	3
3.1 多项式插值(拉格朗日插值法)	3
3.2 分段低次插值	5
3.2.1 分段一次插值	6
3.2.2 分段二次插值	7
3.3 样条插值	9
3.3.1 二次样条插值	10
3.3.2 三次样条插值	12
3.4 插值法总结	15
4 拟合法	15
4.1 形式 ①	15
4.2 形式 ②	17
4.3 拟合法总结	20
5 小结	20

3 插值法



3.1 多项式插值 (拉格朗日插值法)

由于拉格朗日插值和牛顿插值仅仅是在数学表达式的形式上不同, 其插值后的函数本质上是完全一样的, 因此仅以拉格朗日插值为例来解答本题。



```
1  clc , clear
2  %% 多项式插值
3  syms x;
4  % 数据
5  x0=2:16;
6  y0=[6.7  8.2  9.58  9.5  9.7  10  9.96  9.99  10.49  10.59  10.6
      10.8  10.6  10.9  10.76];
7  y=0;
8  % 拉格朗日插值
9  for i=1:15
10     l=1;
11     for j=1:15
12         if j~=i
13             l=l*(x-x0(j));
14             l=l/(x0(i)-x0(j));
15         end
16     end
17     l=simplify(y0(i)*l);
18     y=y+l;
19 end
20 % 绘图
21 plot(x0,y0,'o');
22 hold on;
23 fplot(y);
24 xlim([1,17]);
25 ylim([0,15]);
```

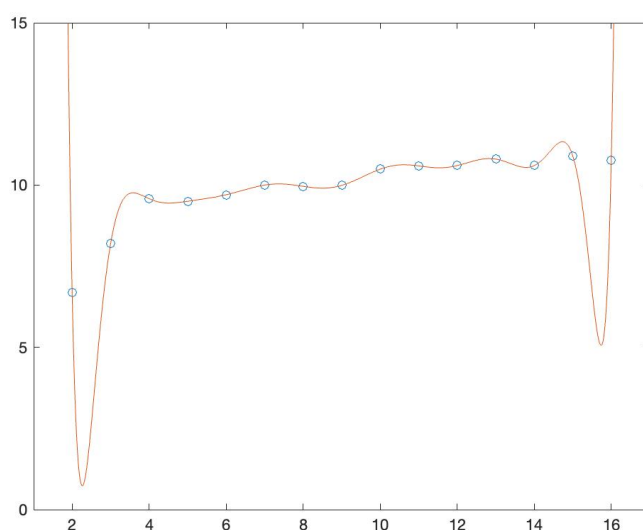


图 3: 多项式插值 (拉格朗日插值法)

可以看出, 拉格朗日插值后出现了明显的 Runge 现象, 这是因为本题的多项式次数很高, 此时端点附近抖动很大, 因此拉格朗日插值不适合用于估计本题。

3.2 分段低次插值

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 $a = x_0 < x_1 < x_2 < \dots < x_{n-1} < x_n = b$, 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 $[x_{i-1}, x_{i+1}]$ 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

● 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

图 4: 分段低次插值优缺点

3.2.1 分段一次插值

```
1 %% 分段一次插值
2 x=2:step:16-step;
3 count=1;
4 for t=2:step:16-step
5     y(count)=y_value(floor(t)-1)+(t-x_value(floor(t)-1)
6         *(y_value(floor(t))-y_value(floor(t)-1))/(
7         x_value(floor(t))-x_value(floor(t)-1));
8     count=count+1;
9 end
10 %绘图
11 plot(x,y,'r');
12 hold on
13 plot(x_value,y_value,'o');
14 title('一次分段插值函数')
```

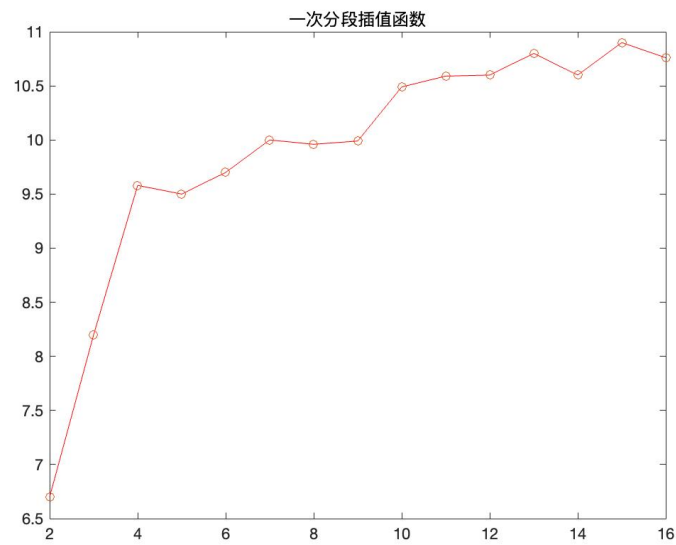


图 5: 分段一次插值

3.2.2 分段二次插值

```
1 %% 分段二次插值
2 x_bond=zeros(1,length(x_value)-1); % Calculate boundary
   value
3 x_bond(1)=2;
4 for i=2:length(x_value)-2
5     x_bond(i)=0.5*(x_value(i)+x_value(i+1));
6 end
7 x_bond(end)=16;
8 count=1;
9 x=2:step:16;
10 for t=2:step:16
11     for j=1:length(x_bond)
12         if t<x_bond(j)
13             i=j-1;break;
14         end
15     end
16     y0=y_value(i);y1=y_value(i+1);y2=y_value(i+2);
17     x0=x_value(i);x1=x_value(i+1);x2=x_value(i+2);
18     % Quadratic Lagrange interpolation
```

```
19     y(count)=y0*(t-x1)*(t-x2)/(x0-x1)/(x0-x2)+y1*(t-x0)  
        *(t-x2)/(x1-x0)/(x1-x2)+y2*(t-x0)*(t-x1)/(x2-x0)  
        /(x2-x1);  
20     count=count+1;  
21 end  
22 %绘图  
23 plot(x,y,'r');  
24 hold on  
25 plot(x_value,y_value,'o');  
26 title('二次分段插值函数')
```

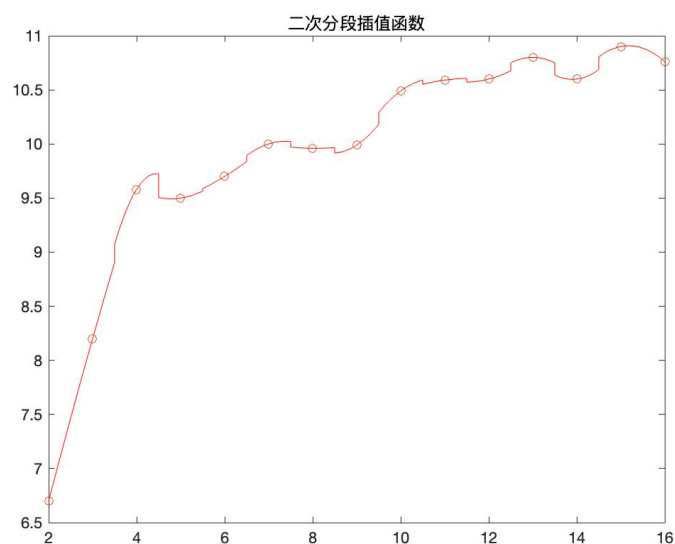


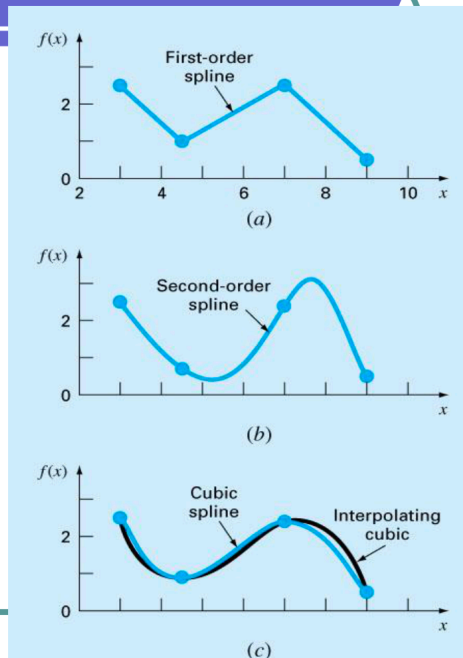
图 6: 分段二次插值

从结果可以看出，分段低次插值在端点之间过度过于僵硬，分段一次插值在端点处导数不连续，二次分段插值甚至函数不连续，无法作为估计本题的合适方法。

3.3 样条插值

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



2022/3/20

数值计算方法

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： $a=x_0 < \dots < x_n=b$ ，设函数：

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

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 $S(x_i) = f(x_i)$ ， $i=0, \dots, n$ 。

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

3.3.1 二次样条插值

二次样条

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

- 相邻多项式在内部节点处函数值相等

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

- 第一和最后一个函数通过端点

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

$$a_1 x_0^2 + b_1 x_0 + c_1 = f(x_0)$$

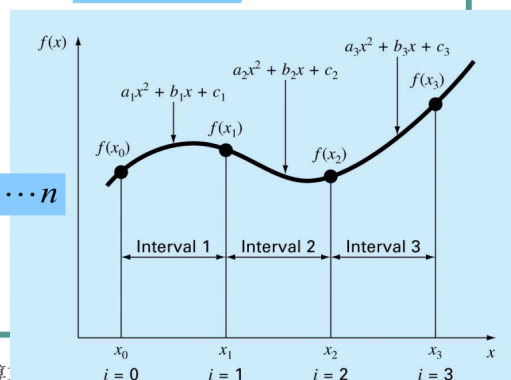
$$a_n x_n^2 + b_n x_n + c_n = f(x_n)$$

- 内部节点的一阶导数相等

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

- 在第一个节点处二阶导数为0

$$a_1 = 0 \quad \text{——连接前两点的为一条直线}$$



2022/3/20

数值计算

```

1  clc , clear
2  %% 二次样条插值
3  % 数据
4  x0=2:16;
5  y0=[6.7 8.2 9.58 9.5 9.7 10 9.96 9.99 10.49 10.59 10.6
      10.8 10.6 10.9 10.76];
6  % 构造方程组, X为(a1,b1,c1,a2,b2,c2...a14,b14,c14)
7  A=zeros(42,42);
8  b=zeros(42,1);
9  A(1,1)=x0(1)^2;A(1,2)=x0(1);A(1,3)=1;
10 b(1)=y0(1);
11 for i=1:13
12     A(2*i,3*i-2)=x0(i+1)^2;A(2*i,3*i-1)=x0(i+1);A(2*i,
        3*i)=1;
13     A(2*i+1,3*i+1)=x0(i+1)^2;A(2*i+1,3*i+2)=x0(i+1);A
        (2*i+1,3*i+3)=1;
14     b(2*i)=y0(i+1);
15     b(2*i+1)=y0(i+1);

```

```

16 end
17 b(28)=y0(15);
18 A(28,40)=x0(15)^2;A(28,41)=x0(15);A(28,42)=1;
19 A(29,1)=1;
20 for i=2:14
21     A(28+i,3*i-5)=2*x0(i);A(28+i,3*i-2)=-2*x0(i);A(28+i
        ,3*i-4)=1;A(28+i,3*i-1)=-1;
22 end
23 X=A\b;
24 % 绘图
25 step=1e-5;
26 x=2:step:16-step;
27 count=1;
28 for t=2:step:16-step
29     i=floor(t)-1;
30     y(count)=X(3*i-2)*t^2+X(3*i-1)*t+X(3*i);
31     count=count+1;
32 end
33 plot(x0,y0,'o');
34 hold on;
35 plot(x,y);

```

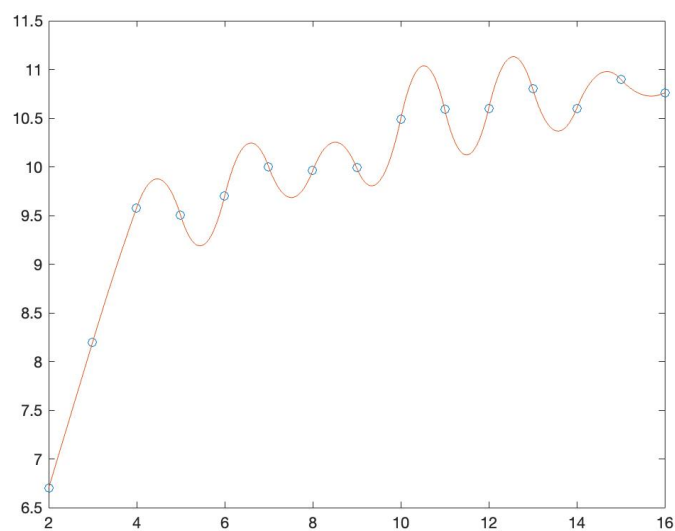


图 7: 二次样条插值

3.3.2 三次样条插值

三次样条插值函数

$$S_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i, x \in [x_{i-1}, x_i] \quad i = 1, \dots, n$$

- 有 $4n$ 个未知数

- 满足条件

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

- 边界条件

- 假设在两端点的导数 $f'(a)=f'_0$ 和 $f'(b)=f'_n$ 为已知, $S'(x_0)=f'_0$, $S'(x_n)=f'_n$
- 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
- 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 。

- $S''(x_i)$ 在每一个子区间上是线性函数

$S(x)$ 是3次
多项式

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

- 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

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

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

$n+1$ 个未知
数, $n-1$ 个
方程

- 由 $S'(x_i+0)=S'(x_i-0)$ $\mu_i M_{i-1} + 2M_i + \lambda_i M_{i+1} = g_i \quad i = 1, 2, \dots, n-1$

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

三弯矩法

边界条件1:

$$S'(x_0)=f'_0 \Rightarrow 2M_0+M_1=6([f[x_0,x_1]-f'_0])/h_1$$

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

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

边界条件

边界条件

$$\lambda_0 = \mu_n = 1$$

$$g_0 = 6 \frac{(f[x_0, x_1] - f'_0)}{h_1}$$

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

```

1  clc , clear
2  %% 三次样条插值
3  x0=2:16;
4  y0=[6.7  8.2  9.58  9.5  9.7  10  9.96  9.99  10.49  10.59  10.6
      10.8  10.6  10.9  10.76];
5  % 三弯矩法方程组构造
6  for i=1:14
7      h(i)=x0(i+1)-x0(i);
8  end
9  for i=1:13
10     u(i)=h(i)/(h(i)+h(i+1));
11     ld(i)=1-u(i);
12     g(i)=6*((y0(i+2)-y0(i+1))/(x0(i+2)-x0(i+1))-(y0(i+1)-y0(i))/(x0(i+1)-x0(i)))/(h(i)+h(i+1));
13 end
14 A=zeros(15,15);
15 b=zeros(15,1);
16 A(1,1)=2;A(1,2)=1;
17 for i=2:14
18     A(i,i-1)=u(i-1);A(i,i)=2;A(i,i+1)=ld(i-1);
19 end

```

```

20 A(15,14)=1;
21 A(15,15)=2;
22 b(1)=6*((y0(2)-y0(1))/(x0(2)-x0(1))-0)/h(1);
23 for i=2:14
24     b(i)=g(i-1);
25 end
26 b(15)=6*(0-(y0(15)-y0(14))/(x0(15)-x0(14)))/h(14);
27 X=A\b;
28 % 绘图
29 step=1e-5;
30 x=2:step:16-step;
31 count=1;
32 for t=2:step:16-step
33     i=floor(t)-1;
34     y(count)=1/(6*h(i))*(X(i)*(x0(i+1)-t)^3+X(i+1)*(t-
        x0(i))^3)+(y0(i)-X(i)*h(i)^2/6)*(x0(i+1)-t)/h(i)
        +(y0(i+1)-X(i+1)*h(i)^2/6)*(t-x0(i))/h(i);
35     count=count+1;
36 end
37 plot(x0,y0,'o');
38 hold on;
39 plot(x,y);

```

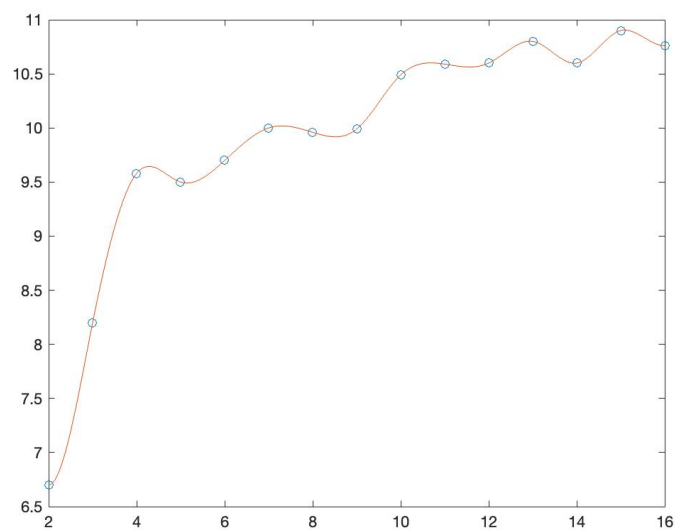


图 8: 三次样条插值

从结果图像可以看出，二次样条插值在每两个点之间有明显的凸起，很显然这种估计效果不好。而三次样条插值在头尾之间的估计效果看起来比较好，然而有个致命的缺点，那就是无法估计头尾数据点以外的数据，因此均不能作为估计本题的方法。

3.4 插值法总结

插值法中多项式插值可以达到外推的目的，然而在本题中出现了严重的 Runge 现象，在头尾端点外的范围明显外推值误差巨大；而分段插值、样条插值无法外推，同时在取样点内部分段差值有不连续的问题，二次插值在相邻端点中波动幅度较大。

因此插值法不适合该问题。

4 拟合法

4.1 形式 ①

```
1  clc , clear
2  %% 1/y=a+b/x
3  % 数据
4  x0=2:16;
5  y0=[6.7 8.2 9.58 9.5 9.7 10 9.96 9.99 10.49 10.59 10.6
      10.8 10.6 10.9 10.76];
6  % 数据变换
7  x1=1./x0;
8  y1=1./y0;
9  % 方程组构造
10 A=zeros(2,2);
11 b=zeros(2,1);
12 A(1,1)=15;
13 A(1,2)=sum(x1);
14 A(2,1)=sum(x1);
15 A(2,2)=norm(x1)^2;
16 b(1)=sum(y1);
17 b(2)=dot(x1,y1);
18 X=A\b;
19 % 绘图
20 step=1e-5;
```

```
21 x=0.05:step:0.6;
22 count=1;
23 for t=0.05:step:0.6
24     y(count)=X(1)+t*X(2);
25     count=count+1;
26 end
27 plot(x1,y1,'o');
28 hold on;
29 plot(x,y);
30
31 x=1:step:17;
32 count=1;
33 for t=1:step:17
34     y(count)=t/(X(1)*t+X(2));
35     count=count+1;
36 end
37 count=1;
38 for t=2:16
39     y_value(count)=t/(X(1)*t+X(2));
40     delta(count)=y_value(count)-y0(count);
41     count=count+1;
42 end
43 hold off;
44 plot(x0,y0,'o');
45 hold on;
46 plot(x,y);
47 MSE=1/15*norm(delta)^2;
```

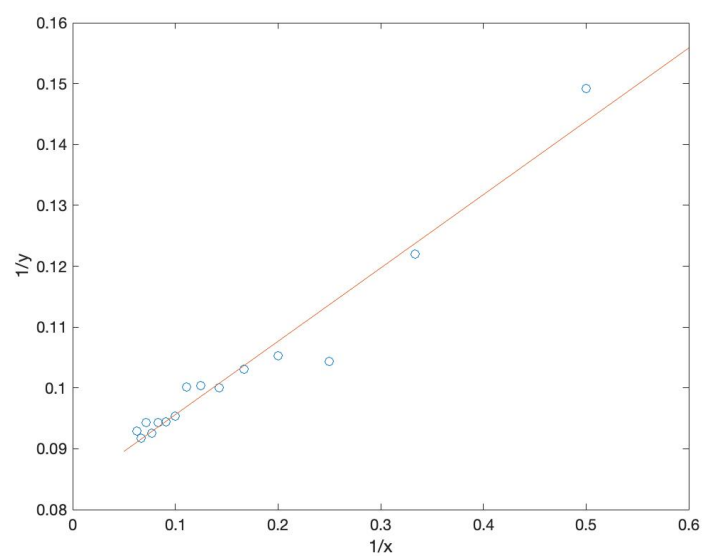



图 9: 线性形式

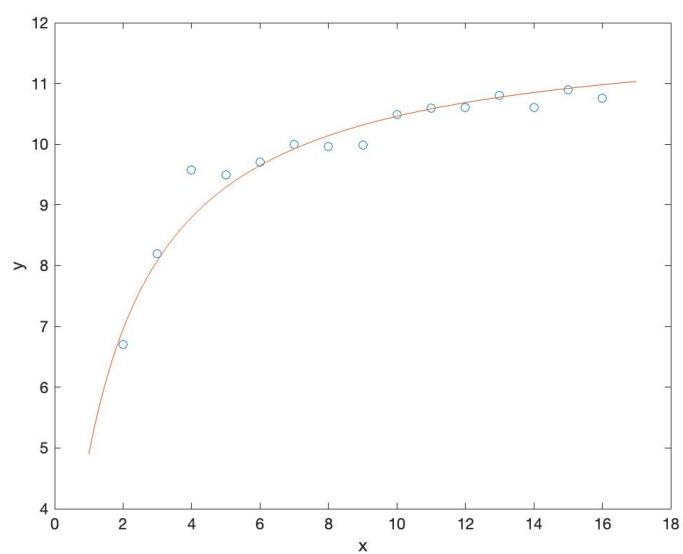


图 10: 原函数

计算得 $MSE = 0.066884156776709$

4.2 形式 ②

```
1  clc , clear
2  %% 1/y=a+b/x
3  % 数据
4  x0=2:16;
```

```
5 y0=[6.7 8.2 9.58 9.5 9.7 10 9.96 9.99 10.49 10.59 10.6
    10.8 10.6 10.9 10.76];
6 % 数据变换
7 x1=1./x0;
8 y1=log(y0);
9 A=zeros(2,2);
10 b=zeros(2,1);
11 A(1,1)=15;
12 A(1,2)=sum(x1);
13 A(2,1)=sum(x1);
14 A(2,2)=norm(x1)^2;
15 b(1)=sum(y1);
16 b(2)=dot(x1,y1);
17 X=A\b;
18 % 绘图
19 step=1e-5;
20 x=0.05:step:0.6;
21 count=1;
22 for t=0.05:step:0.6
23     y(count)=X(1)+t*X(2);
24     count=count+1;
25 end
26 plot(x1,y1,'o');
27 hold on;
28 plot(x,y);
29 x=1:step:17;
30 count=1;
31 for t=1:step:17
32     y(count)=exp(X(1))*exp(X(2)/t);
33     count=count+1;
34 end
35 count=1;
36 for t=2:16
37     y_value(count)=exp(X(1))*exp(X(2)/t);
38     delta(count)=y_value(count)-y0(count);
39     count=count+1;
40 end
```

```
41 hold off;  
42 plot(x0,y0,'o');  
43 hold on;  
44 plot(x,y);  
45 % 计算均方误差  
46 MSE=1/15*norm(delta)^2;
```

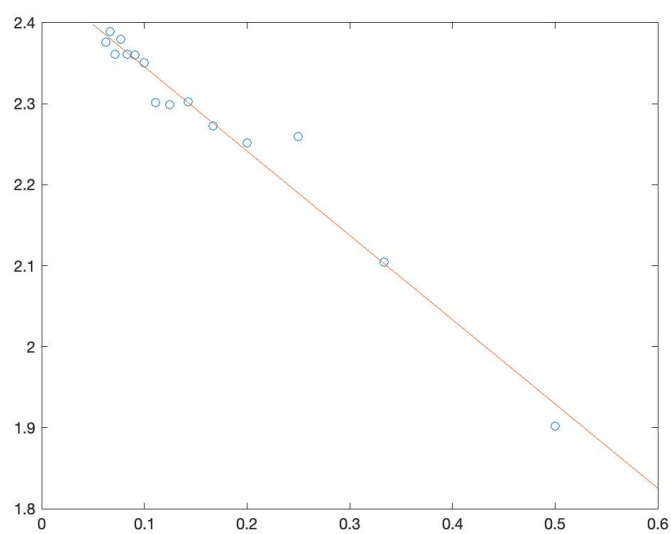


图 11: 线性形式

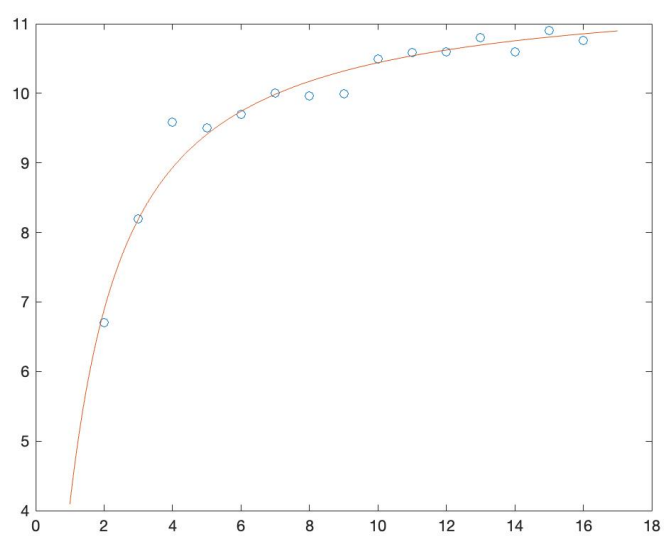


图 12: 原函数

计算得 $MSE = 0.045191947620288$

4.3 拟合法总结

综合对比两种拟合形式，指数形式的 MSE 明显更小，因此形式二拟合效果更好。同时，拟合法解决了插值法无法外推的问题，是适合该问题的估计方法。

5 小结

本文对钢水对材料侵蚀实际问题中的增大容积 y 和使用次数 x 之间的函数关系进行了分析，从插值法和拟合法两大方向进行了讨论，最后得出结论：插值方法不适合该问题，拟合法较为适合该问题，且拟合曲线形式 ②（即指数关系）的 MSE 更小，更加适合该问题。