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：

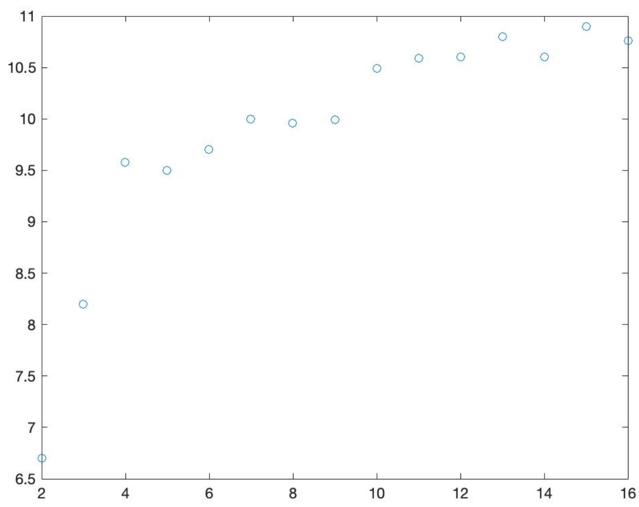


图 2: 散点图

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。

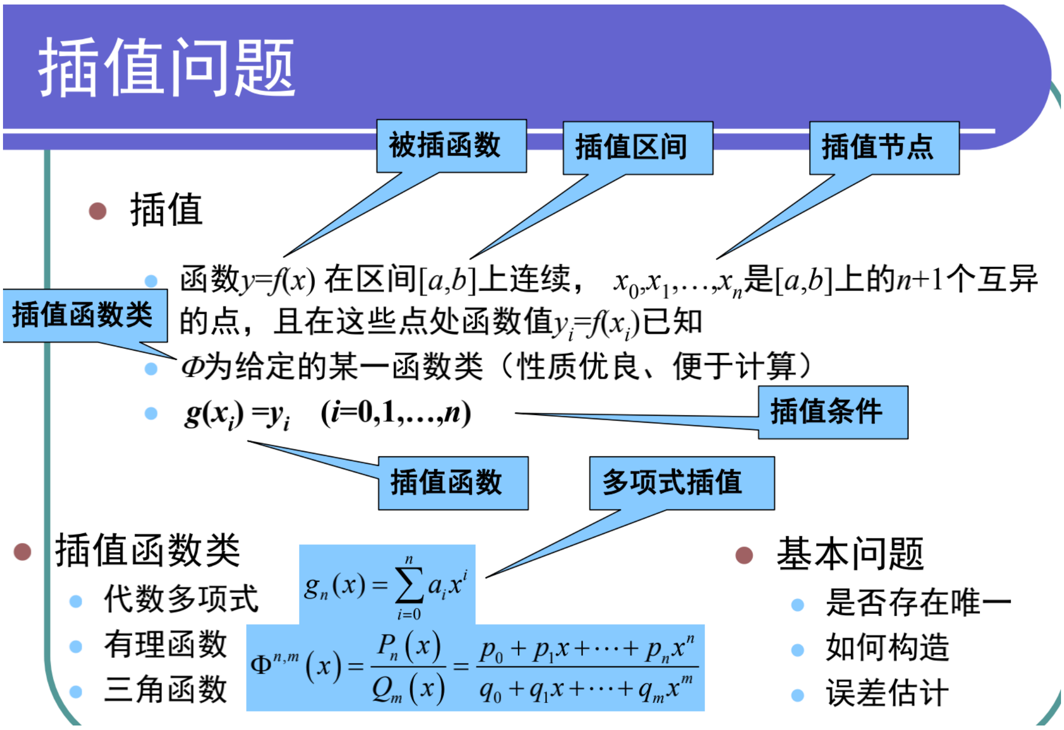
对于第二问，可对两个形式的函数进行适当变形转换为一次函数拟合问题。通过计算 MSE（均方误差）:

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

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。

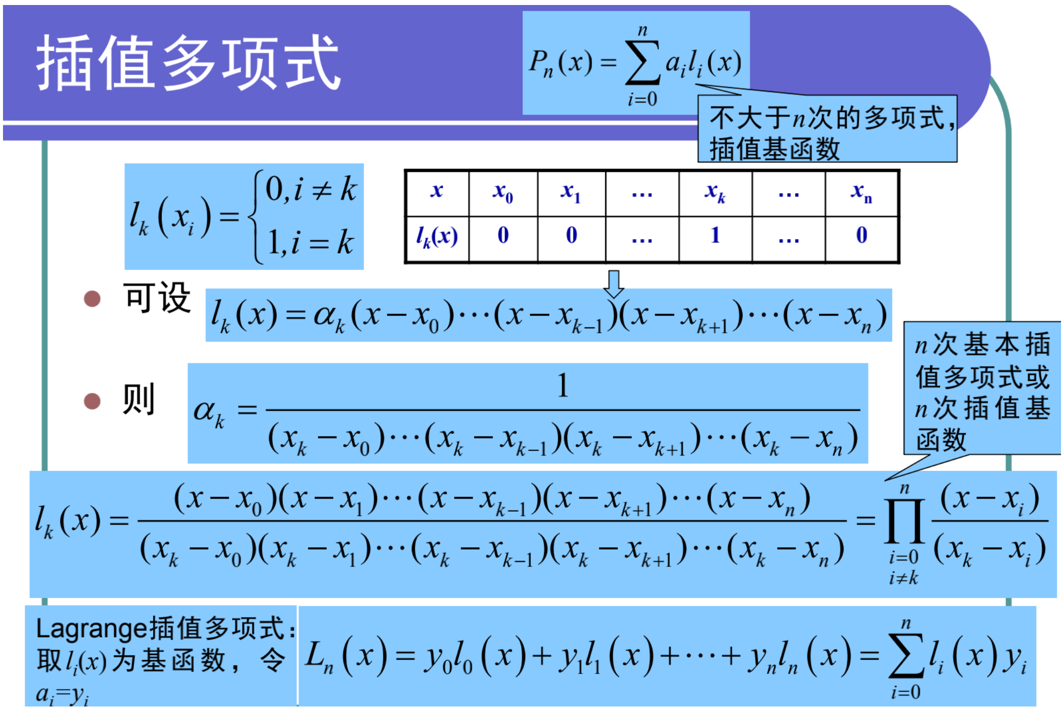
目录

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。



```
clc,clear
%% 多项式插值
syms x;
% 数据
x0=2:16;
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];
y=0;
% 拉格朗日插值
for i=1:15
    l=1;
    for j=1:15
        if j~=i
            l=l*(x-x0(j));
            l=l/(x0(i)-x0(j));
        end
    end
    l=simplify(y0(i)*l);
    y=y+l;
end
% 绘图
plot(x0,y0,'o');
hold on;
fplot(y);
xlim([1,17]);
ylim([0,15]);
```

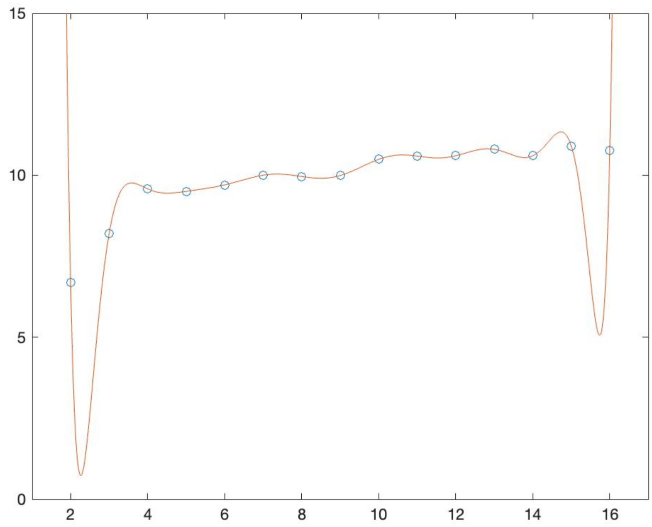
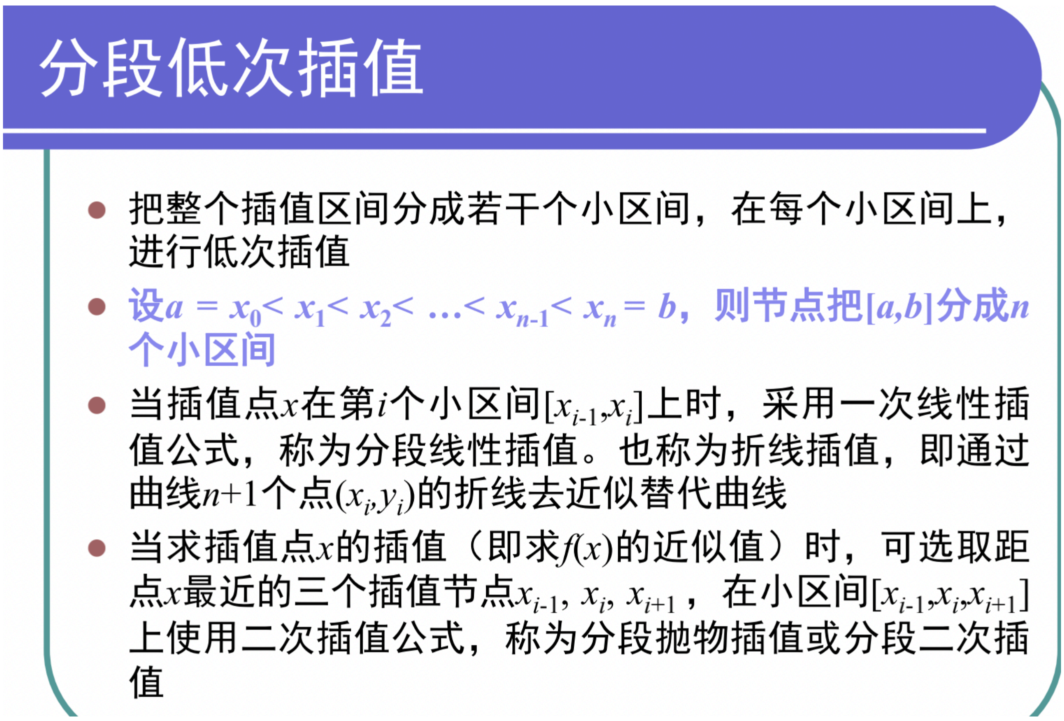


图 3: 多项式插值（拉格朗日插值法）

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



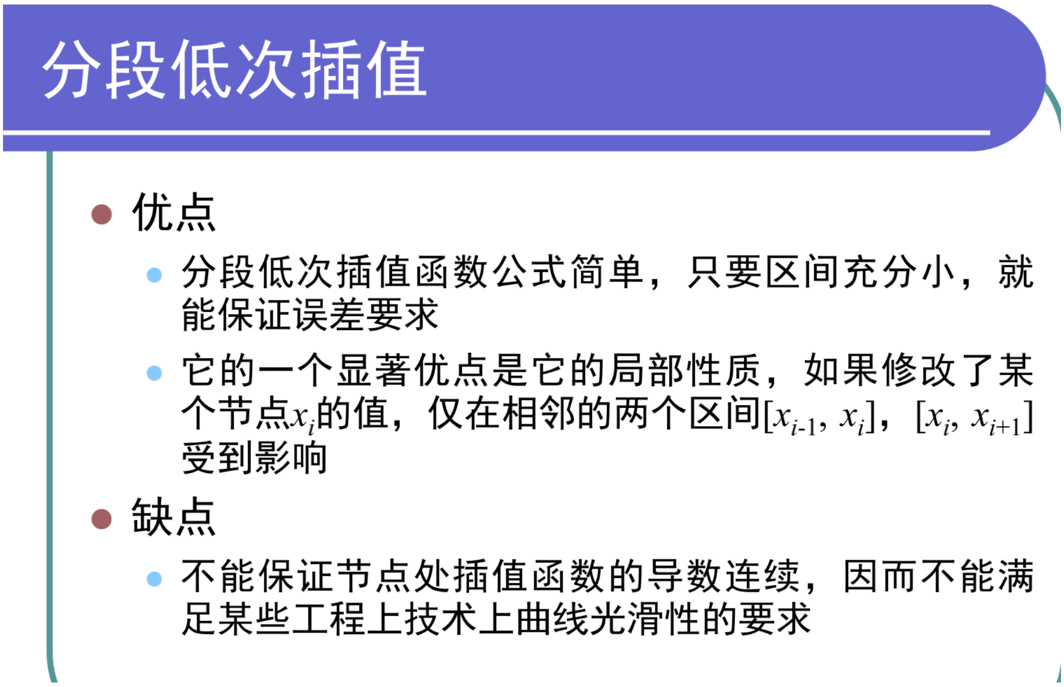


图 4: 分段低次插值优缺点

3.2.1 分段一次插值

```
%% 分段一次插值
x=2:step:16-step;
count=1;
for  t=2:step:16-step
     y(count)=y_value(floor(t)-1)+(t-x_value(floor(t)-1))*(y_value(floor(t))-y_value(floor(t)-1))/(x_value(floor(t))-x_value(floor(t)-1));
     count=count+1;
end
%绘图
plot(x,y,'r');
hold on
plot(x_value,y_value,'o');
title('一次分段插值函数')
```

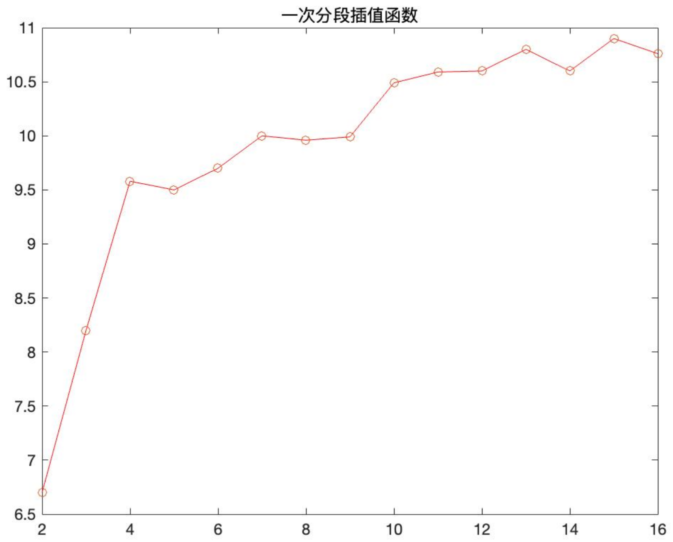


图 5: 分段一次插值

3.2.2 分段二次插值

```
%% 分段二次插值
x_bond=zeros(1,length(x_value)-1); % Calculate boundary value
x_bond(1)=2;
for i=2:length(x_value)-2
    x_bond(i)=0.5*(x_value(i)+x_value(i+1));
end
x_bond(end)=16;
count=1;
x=2:step:16;
for t=2:step:16
    for j=1:length(x_bond)
        if t<x_bond(j)
            i=j-1;break;
        end
    end
    y0=y_value(i);y1=y_value(i+1);y2=y_value(i+2);
    x0=x_value(i);x1=x_value(i+1);x2=x_value(i+2);
    % Quadratic Lagrange interpolation
```

```
    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);
    count=count+1;
end
%绘图
plot(x,y,'r');
hold on
plot(x_value,y_value,'o');
title('二次分段插值函数')
```

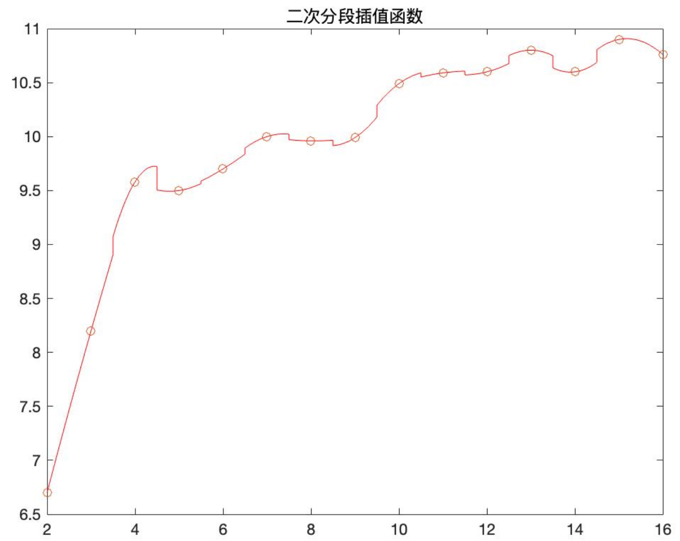


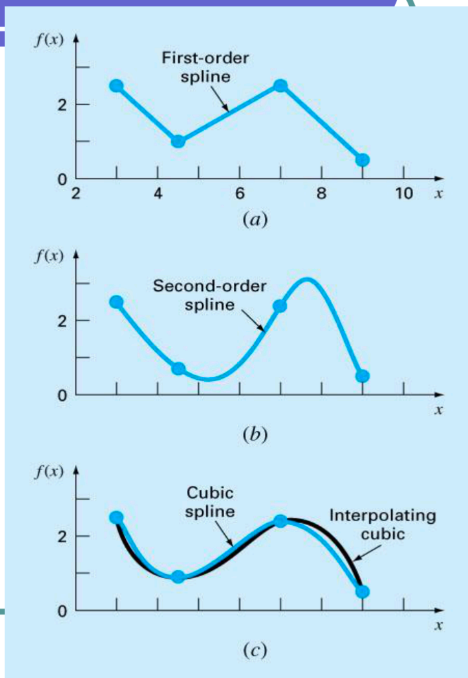
图 6: 分段二次插值

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：$a=x_0<\ldots<x_n=b$，设函数：

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

满足条件：

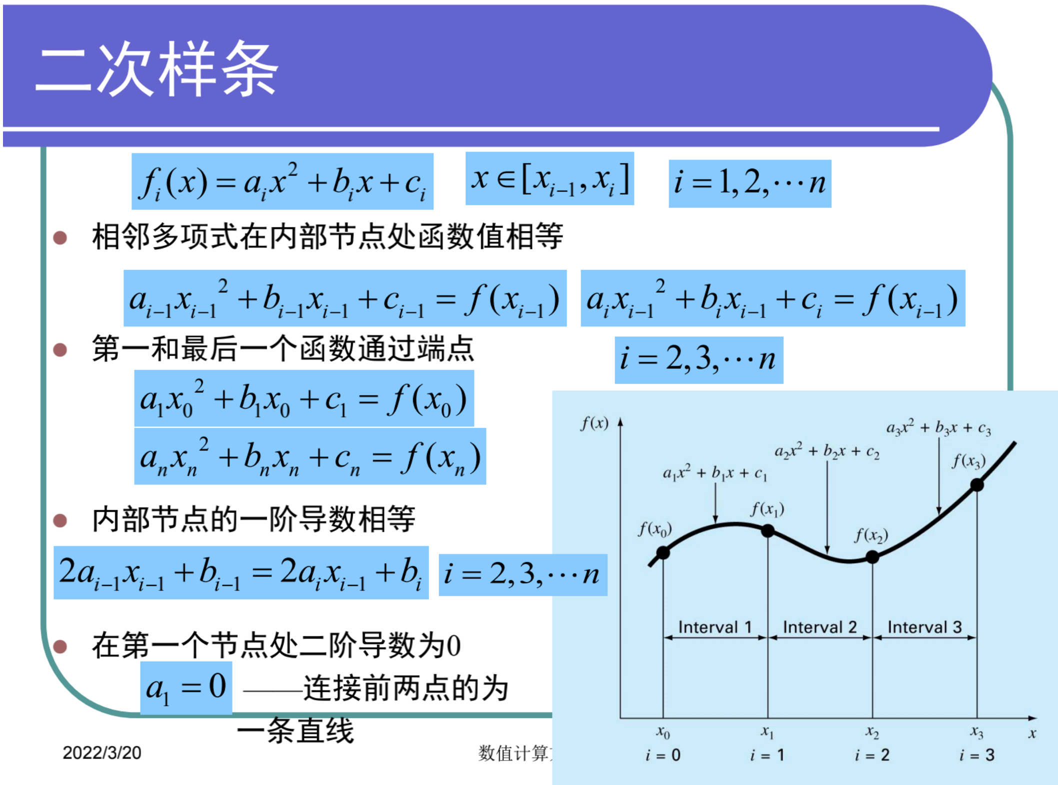
(1)　$S(x)$在区间$[a,b]$上存在二阶(m-1阶)的连续导数；

(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。

3.3.1　二次样条插值




```
clc,clear
%% 二次样条插值
% 数据
x0=2:16;
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];
% 构造方程组，X为(a1,b1,c1,a2,b2,c2...a14,b14,c14)
A=zeros(42,42);
b=zeros(42,1);
A(1,1)=x0(1)^2;A(1,2)=x0(1);A(1,3)=1;
b(1)=y0(1);
for i=1:13
    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;
    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;
    b(2*i)=y0(i+1);
    b(2*i+1)=y0(i+1);
```

```
end
b(28)=y0(15);
A(28,40)=x0(15)^2;A(28,41)=x0(15);A(28,42)=1;
A(29,1)=1;
for i=2:14
    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;
end
X=A\b;
% 绘图
step=1e-5;
x=2:step:16-step;
count=1;
for t=2:step:16-step
    i=floor(t)-1;
    y(count)=X(3*i-2)*t^2+X(3*i-1)*t+X(3*i);
    count=count+1;
end
plot(x0,y0,'o');
hold on;
plot(x,y);
```

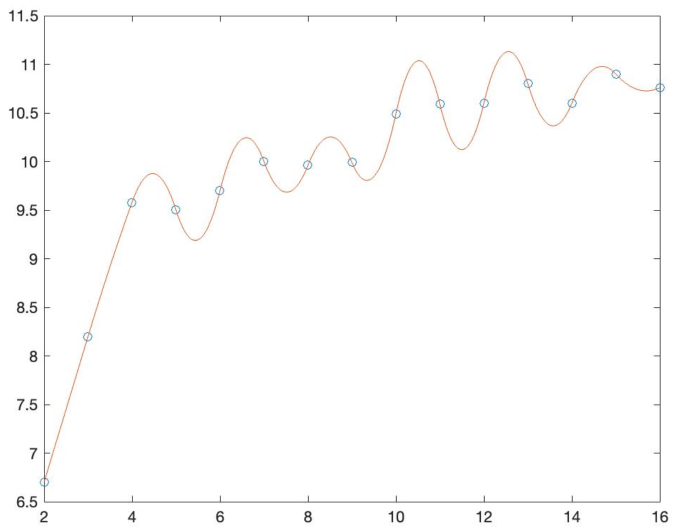


图 7: 二次样条插值

3.3.2 三次样条插值

三次样条插值函数

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

- 有$4n$个未知数
- 满足条件

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

- 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。
- $S''(x_i)$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

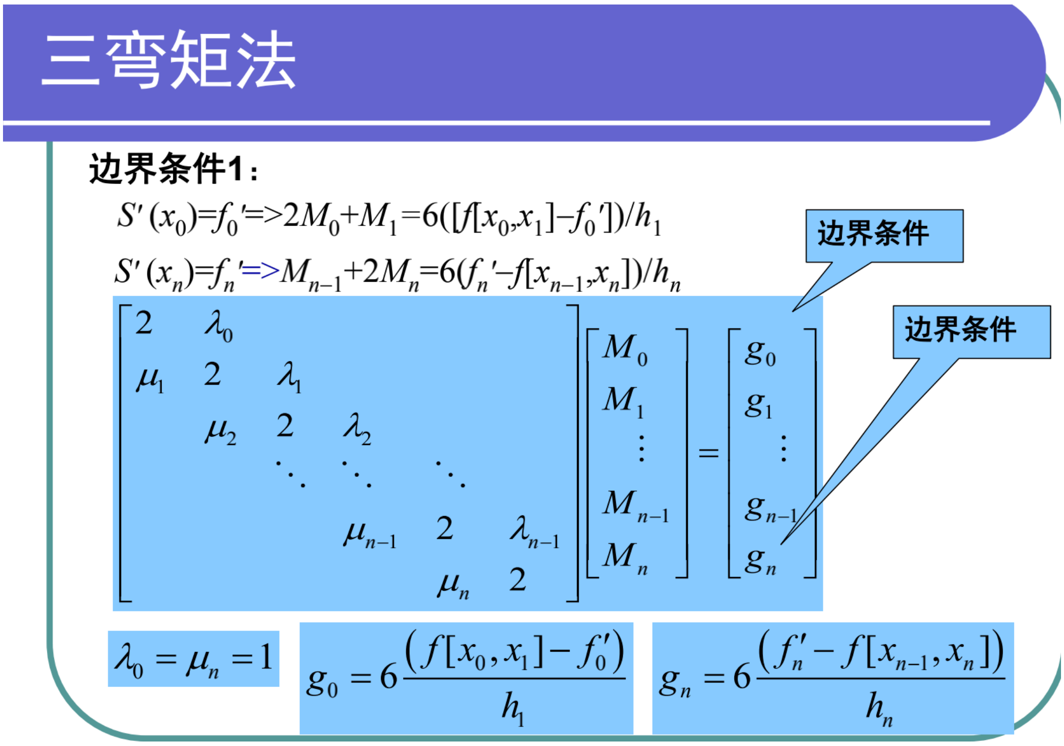
$$S_i(x)=\frac{1}{6h_i}[M_{i-1}(x_i-x)^3+M_i(x-x_{i-1})^3]+(f_{i-1}-\frac{M_{i-1}h_i^2}{6})\frac{x_i-x}{h_i}+(f_i-\frac{M_ih_i^2}{6})\frac{x-x_{i-1}}{h_i}$$

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

$n+1$个未知数，$n-1$个方程

- 由$S'(x_i+0)=S'(x_i-0)$ $\mu_iM_{i-1}+2M_i+\lambda_iM_{i+1}=g_i \quad i=1,2,\cdots,n-1$

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




```
clc,clear
%% 三次样条插值
x0=2:16;
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];
% 三弯矩法方程组构造
for i=1:14
    h(i)=x0(i+1)-x0(i);
end
for i=1:13
    u(i)=h(i)/(h(i)+h(i+1));
    ld(i)=1-u(i);
    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));
end
A=zeros(15,15);
b=zeros(15,1);
A(1,1)=2;A(1,2)=1;
for i=2:14
    A(i,i-1)=u(i-1);A(i,i)=2;A(i,i+1)=ld(i-1);
end
```

```
A(15,14)=1;
A(15,15)=2;
b(1)=6*((y0(2)-y0(1))/(x0(2)-x0(1))-0)/h(1);
for i=2:14
    b(i)=g(i-1);
end
b(15)=6*(0-(y0(15)-y0(14))/(x0(15)-x0(14)))/h(14);
X=A\b;
% 绘图
step=1e-5;
x=2:step:16-step;
count=1;
for t=2:step:16-step
    i=floor(t)-1;
    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);
    count=count+1;
end
plot(x0,y0,'o');
hold on;
plot(x,y);
```

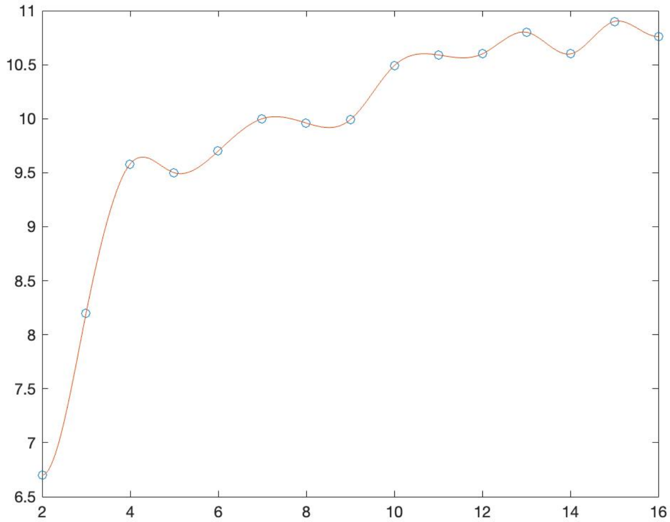


图 8: 三次样条插值

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

```
clc,clear
%% 1/y=a+b/x
% 数据
x0=2:16;
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];
% 数据变换
x1=1./x0;
y1=1./y0;
% 方程组构造
A=zeros(2,2);
b=zeros(2,1);
A(1,1)=15;
A(1,2)=sum(x1);
A(2,1)=sum(x1);
A(2,2)=norm(x1)^2;
b(1)=sum(y1);
b(2)=dot(x1,y1);
X=A\b;
% 绘图
step=1e-5;
```

```
x=0.05:step:0.6;
count=1;
for t=0.05:step:0.6
    y(count)=X(1)+t*X(2);
    count=count+1;
end
plot(x1,y1,'o');
hold on;
plot(x,y);

x=1:step:17;
count=1;
for t=1:step:17
    y(count)=t/(X(1)*t+X(2));
    count=count+1;
end
count=1;
for t=2:16
    y_value(count)=t/(X(1)*t+X(2));
    delta(count)=y_value(count)-y0(count);
    count=count+1;
end
hold off;
plot(x0,y0,'o');
hold on;
plot(x,y);
MSE=1/15*norm(delta)^2;
```

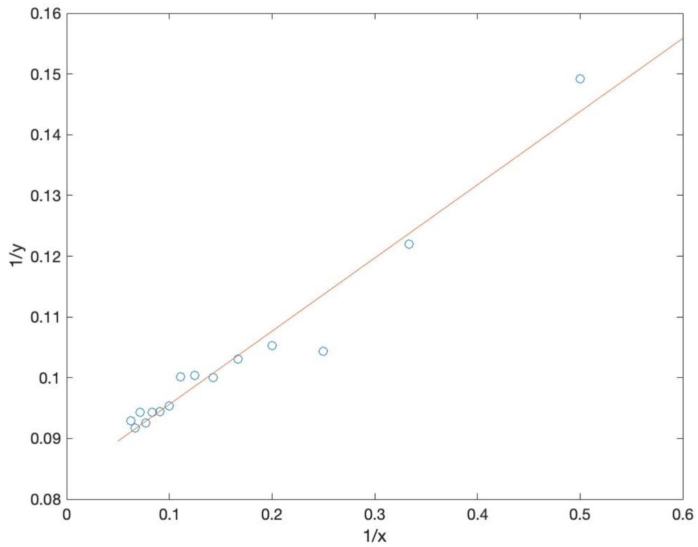


图 9: 线性形式

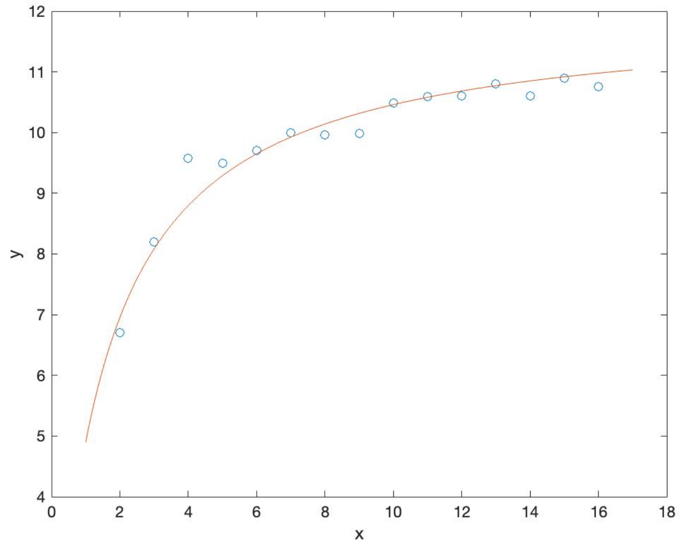


图 10: 原函数

计算得 $MSE = 0.066884156776709$

4.2 形式 ②

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

```
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];
% 数据变换
x1=1./x0;
y1=log(y0);
A=zeros(2,2);
b=zeros(2,1);
A(1,1)=15;
A(1,2)=sum(x1);
A(2,1)=sum(x1);
A(2,2)=norm(x1)^2;
b(1)=sum(y1);
b(2)=dot(x1,y1);
X=A\b;
% 绘图
step=1e-5;
x=0.05:step:0.6;
count=1;
for t=0.05:step:0.6
    y(count)=X(1)+t*X(2);
    count=count+1;
end
plot(x1,y1,'o');
hold on;
plot(x,y);
x=1:step:17;
count=1;
for t=1:step:17
    y(count)=exp(X(1))*exp(X(2)/t);
    count=count+1;
end
count=1;
for t=2:16
    y_value(count)=exp(X(1))*exp(X(2)/t);
    delta(count)=y_value(count)-y0(count);
    count=count+1;
end
```

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

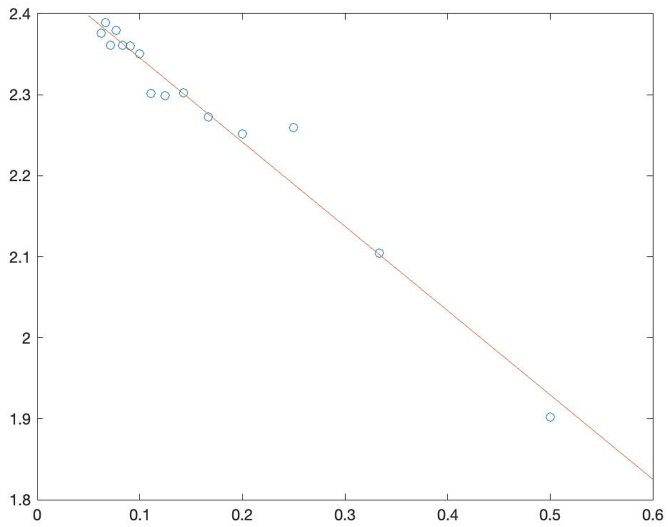


图 11: 线性形式

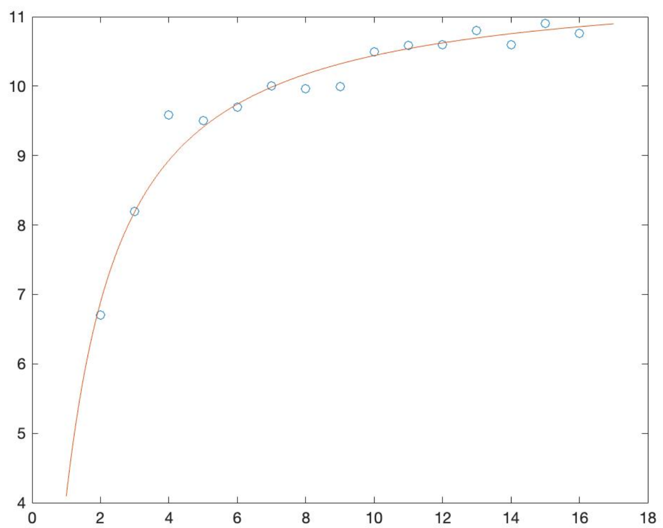


图 12: 原函数

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。