二次样条插值的探究

——wcy1122

1 介绍

插值问题是数值计算的一类经典问题，常用的方法有拉格朗日插值，牛顿插值等。不过以上的多项式插值方法存在一个弊端，如果点数较多，很容易因为此处过高而出现多项式震荡的现象，即机器学习中的过拟合现象。由此看来，当数据点较多时，次数过高的多项式插值并不适用。样条插值因此产生。

样条插值是一种很优秀的解决方法，通过控制表达式在两个相邻数据点间的多项式次数，使得产生的曲线既能通过所有数据点，又不会出现多项式震荡的现象。一次样条和三次样条插值是两种比较常见的插值，而本文将对二次样条插值进行一些探究。

我们将首先对二次样条插值进行公式推导，并随后使用一些函数观察二次样条插值的拟合误差。

2 二次样条的推导

我们重述一下二次样条插值问题。给定 k 个数据点(x_1,y_1)，(x_2,y_2)，……，(x_k,y_k)，要求出一个满足如下形式的二次函数，使其能够通过全部 k 个数据点。

$$S_1(x) = a_1 * (x - x_1)^2 + b_1 * (x - x_1) + c_1 \qquad x \in [x_1, x_2]$$
$$S_2(x) = a_2 * (x - x_2)^2 + b_2 * (x - x_2) + c_2 \qquad x \in [x_2, x_3]$$
$$\cdots$$
$$S_{k-1}(x) = a_{k-1} * (x - x_{k-1})^2 + b_{k-1} * (x - x_{k-1}) + c_{k-1} \qquad x \in [x_{k-1}, x_k]$$

二次样条插值可以使用的几个性质

$$\mathrm{S}(x_k) = y_k$$
$$S_k(x_{k+1}) = S_{k+1}(x_{k+1})$$
$$S_k'(x_{k+1}) = S_{k+1}'(x_{k+1})$$

方便推导，定义以下几个参数，

$$h_i = x_{i+1} - x_i$$
$$M_i = S'(x_i)$$

对$S_i(x)$求导，可以得到，

$$S_i'(x) = 2 * a_i * (x - x_i) + b_i \qquad x \in [x_i, x_{i+1}]$$

根据一阶导数连续性和插值性，将$\mathrm{x} = x_i$和$\mathrm{x} = x_{i+1}$依次代入$s_i'(x)$，将$\mathrm{x} = x_i$代入$S(x)$，可以解出a_i，b_i，c_i，

$$a_i = \frac{M_{i+1} - M_i}{2 * h_i}$$
$$b_i = M_i$$
$$c_i = y_i$$

根据连续性，将$\mathrm{x} = x_{i+1}$代入 $S_i(x)$，可以得到，

$$\mathrm{y}_{i+1} = \frac{M_{i+1} - M_i}{2 * h_i} * h_i^2 + M_i * h_i + y_i$$

化简可得，

$$M_i + M_{i+1} = 2 * \frac{\mathrm{y}_{i+1} - y_i}{h_i}$$

因此，我们可以根据以上形式构造方程组，

$$M_1 + M_2 = 2 * \frac{y_2 - y_1}{h_1}$$

$$M_2 + M_3 = 2 * \frac{y_3 - y_2}{h_2}$$

$$\cdots$$

$$M_{k-1} + M_k = 2 * \frac{y_k - y_{k-1}}{h_{k-1}}$$

写成矩阵乘法的形式，

$$\begin{bmatrix} 1 & 1 & 0 & & 0 & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & & 0 & 0 & 0 \\ & \vdots & & \ddots & & \vdots & \\ 0 & 0 & 0 & & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 1 & 0 \\ 0 & 0 & 0 & & 0 & 1 & 1 \end{bmatrix} * \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ \vdots \\ M_{k-2} \\ M_{k-1} \\ M_k \end{bmatrix} = 2 * \begin{bmatrix} \frac{y_2 - y_1}{h_1} \\ \frac{y_3 - y_2}{h_2} \\ \frac{y_4 - y_3}{h_3} \\ \vdots \\ \frac{y_{k-2} - y_{k-3}}{h_{k-3}} \\ \frac{y_{k-1} - y_{k-2}}{h_{k-2}} \\ \frac{y_k - y_{k-1}}{h_{k-1}} \end{bmatrix}$$

这里有 k 个变量，k-1 个方程，是欠定的，有无穷多解。可以增加条件，使得这个方程组变成恰定的方程组。

求解出M_i，再代回a_i, b_i, c_i，即可得到二次样条插值函数。

3 实验结果

根据以上推导结果，我们得到了二次样条插值的计算方法。先使用矩阵乘法就出 M 向量的值，再带回求出 a，b，c 向量的值，即得到了二次样条插值多项式。根据不同的边界条件，我们可以得到不同的插值结果。我们将根据不同类型的函数，使用均方误差评估二次样条插值的准确性，并绘出图像。

为了统一标准，以下为测试函数，

$$f_1(\mathrm{x}) = x^2 - 2 * x + 3 \quad x \in [-3:1:3]$$

$$f_2(\mathrm{x}) = x^3 - 3 * x^2 + 2 * x + 4 \quad x \in [-3:1:3]$$

$$f_3(\mathrm{x}) = \sin x \quad x \in [-3:1:3]$$

$$f_4(\mathrm{x}) = \mathrm{e}^x \quad x \in [-3:1:3]$$

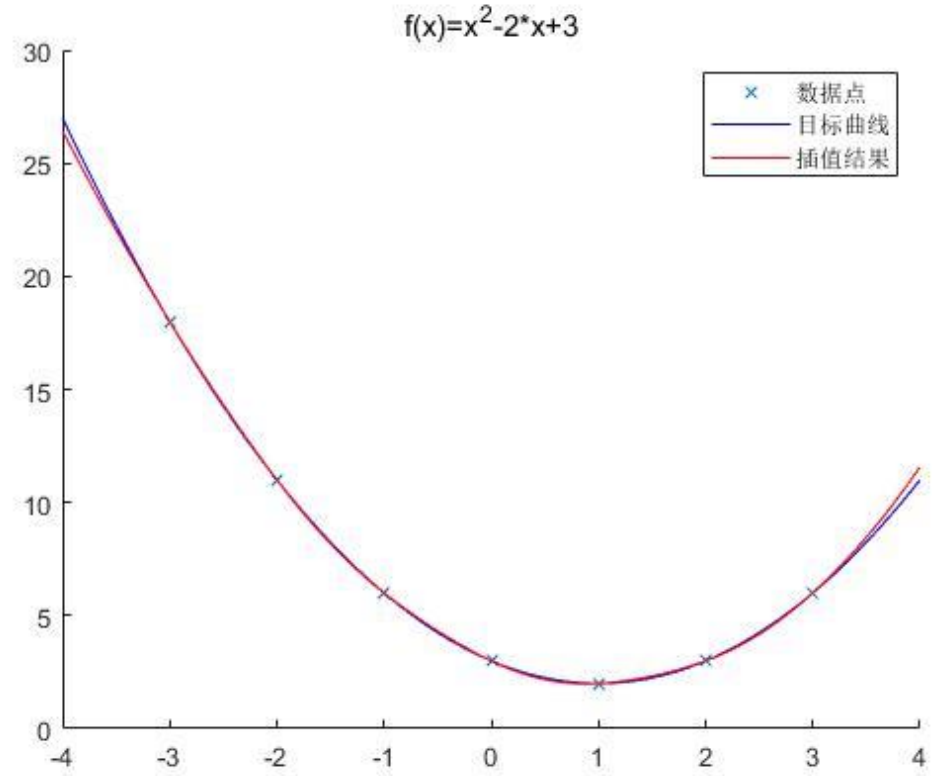
$$f_5(\mathrm{x}) = x^3 - 3 * x^2 + 2 * x + 4 \quad x \in [-3, -1, 0, 0.5, 2, 2.1, 3]$$

以下为分析误差，

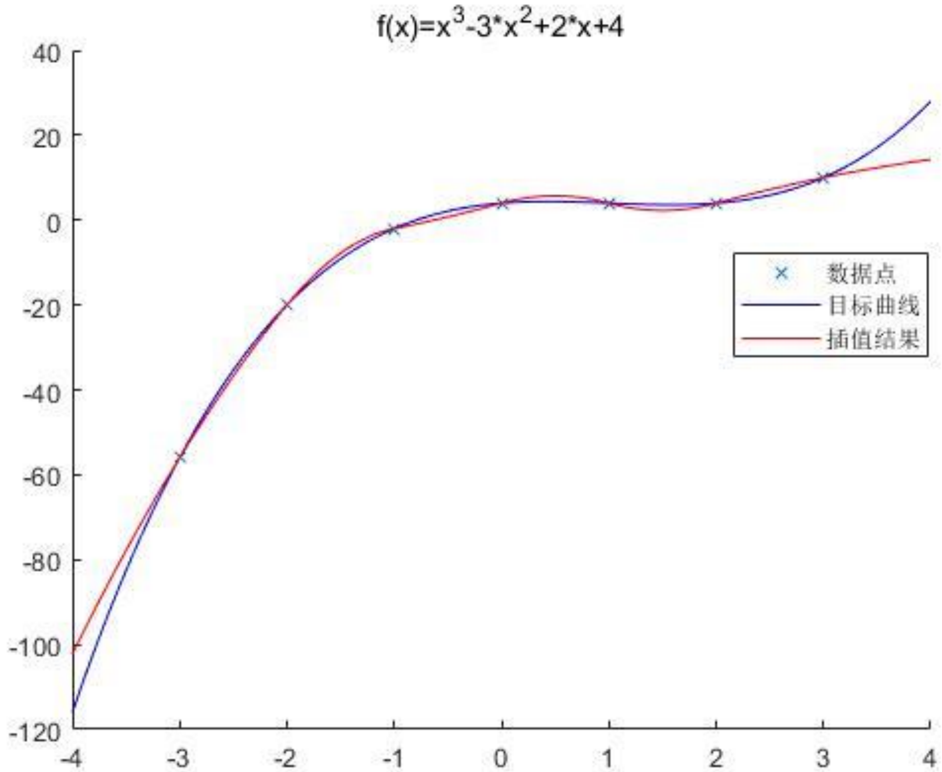
$$\mathrm{Loss} = \frac{\sum_{x=L-1}^{R+1} (f(x) - y_x)^2}{len} \quad x \in [L-1: 0.01: R+1], \mathrm{len} = \mathrm{sizeof(x)}$$

3.1 伪逆

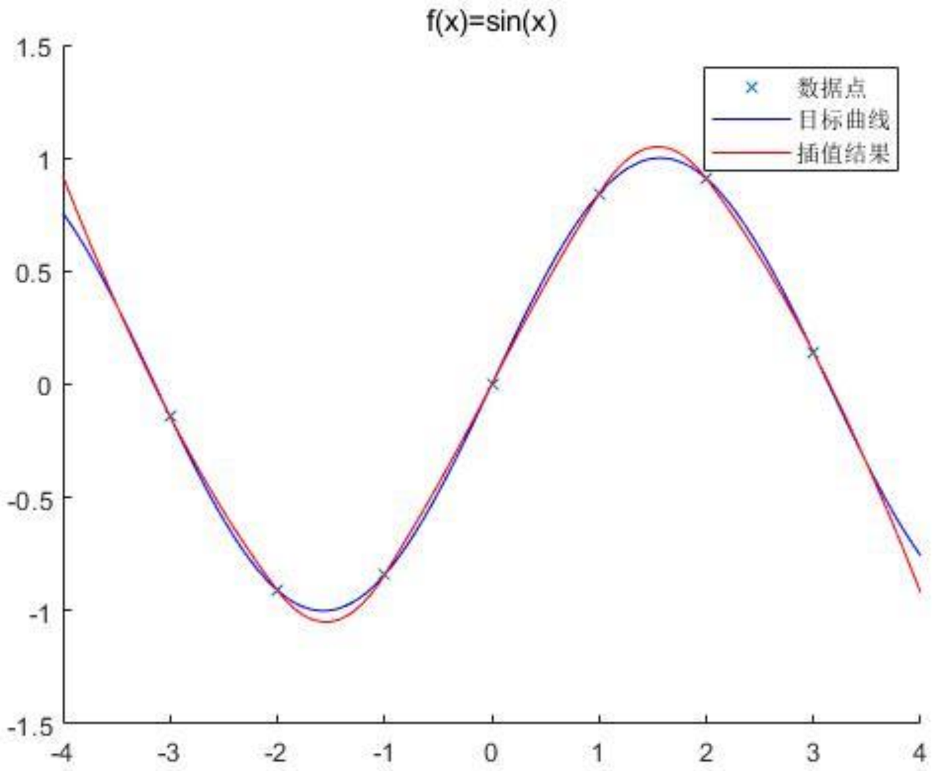
我尝试不引入其他约束条件，直接采用矩阵求伪逆的方法求出 M 向量。



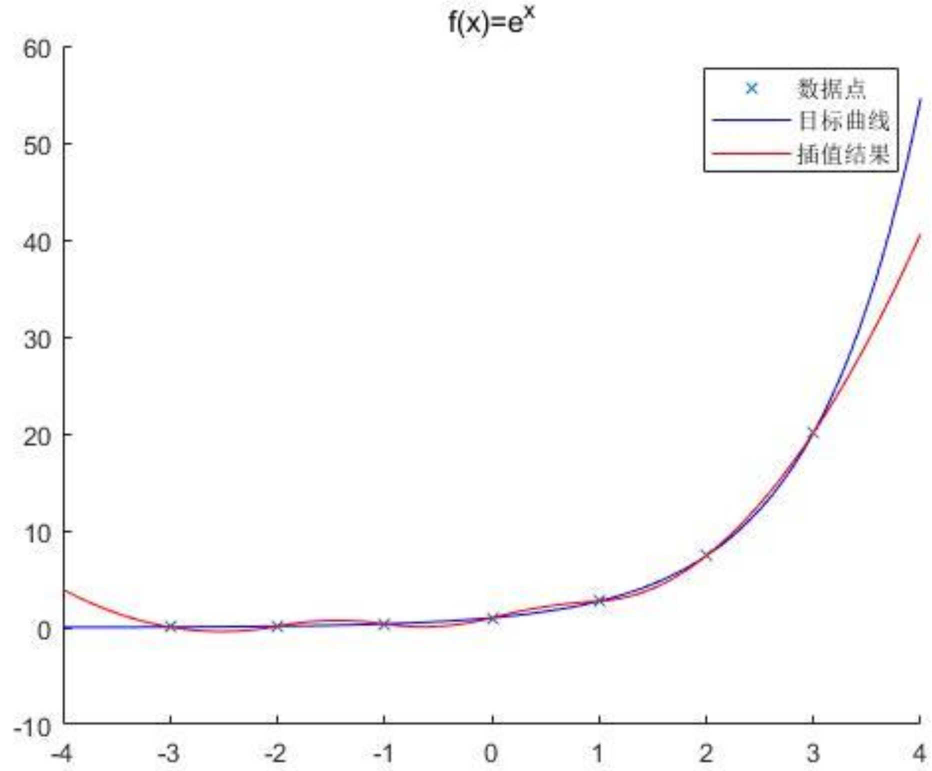
f(x)=x^2-2*x+3
数据点
目标曲线
插值结果
30
25
20
15
10
5
0
-4
-3
-2
-1
0
1
2
3
4



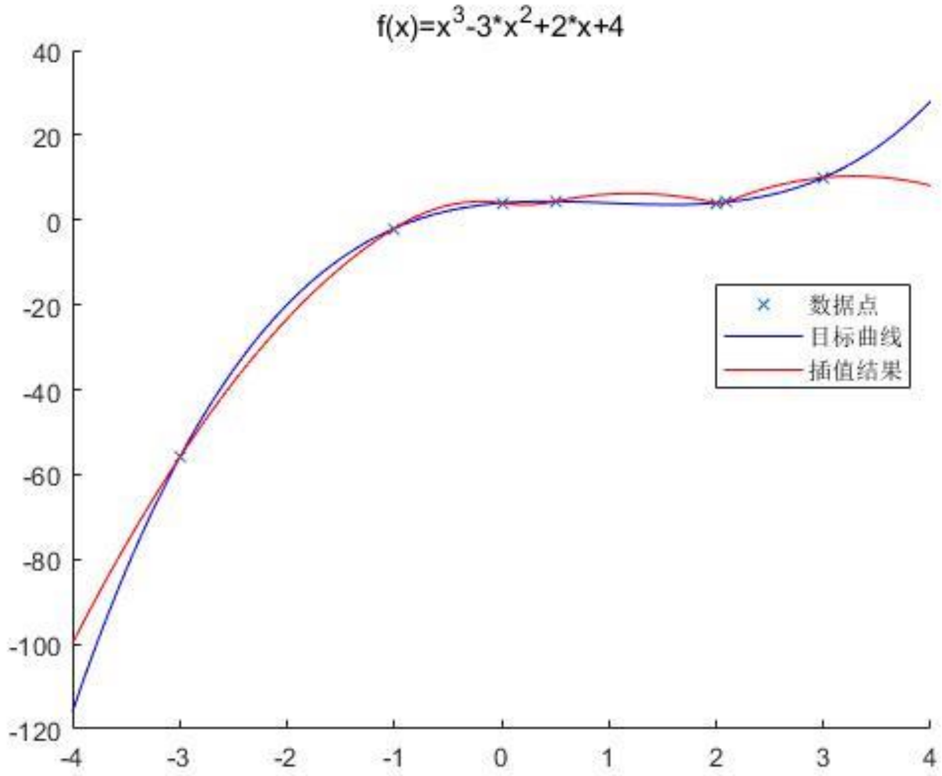
f(x)=x^3-3*x^2+2*x+4
数据点
目标曲线
插值结果
40
20
0
-20
-40
-60
-80
-100
-120
-4
-3
-2
-1
0
1
2
3
4



f(x)=sin(x)
1.5
1
0.5
0
-0.5
-1
-1.5
-4
-3
-2
-1
0
1
2
3
4
数据点
目标曲线
插值结果



f(x)=e^x
60
50
40
30
20
10
0
-10
-4
-3
-2
-1
0
1
2
3
4
数据点
目标曲线
插值结果

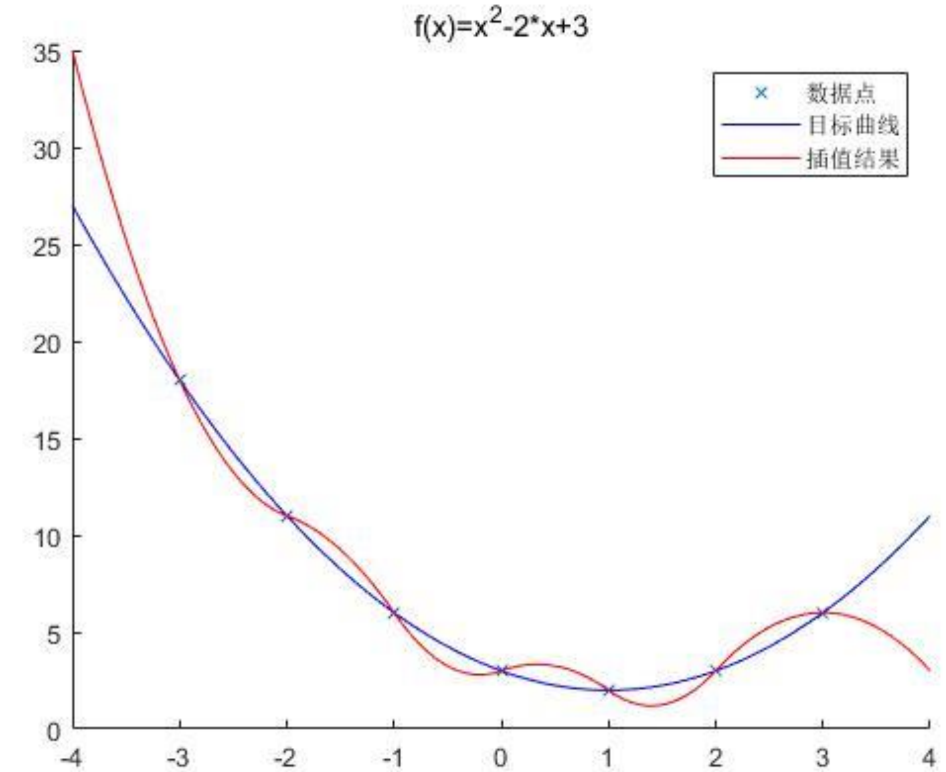


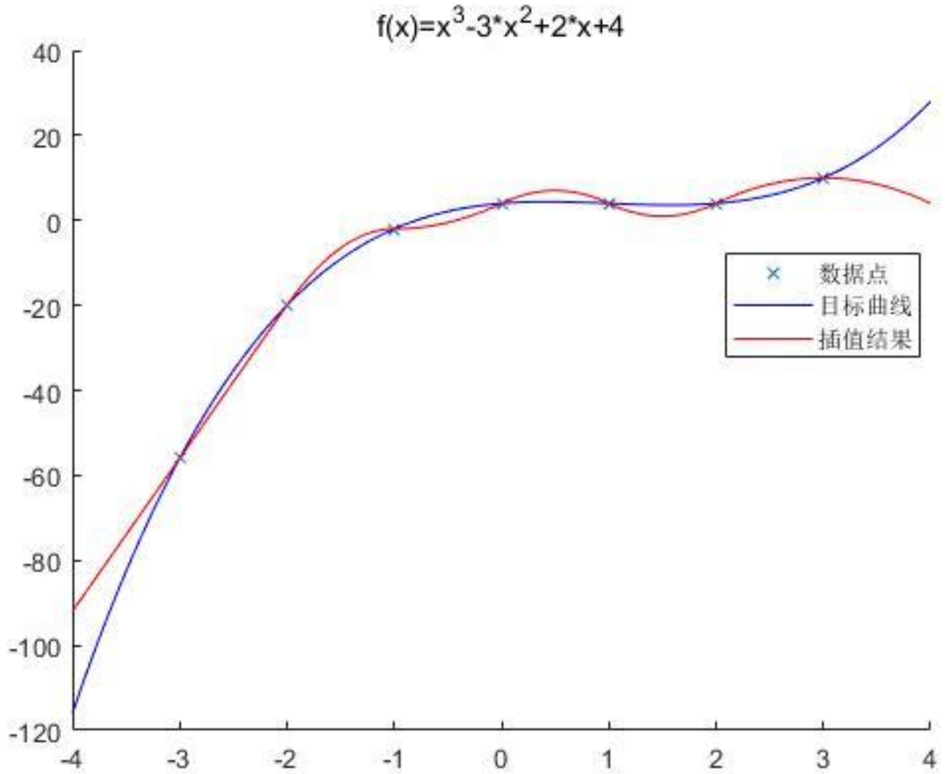
函数	内插值误差	外插值误差
$f_1(\mathrm{x})$	0.0026941	0.085143
$f_2(\mathrm{x})$	0.94834	46.1763
$f_3(\mathrm{x})$	0.00094217	0.003367
$f_4(\mathrm{x})$	0.13188	21.7828
$f_5(\mathrm{x})$	3.2647	82.8024

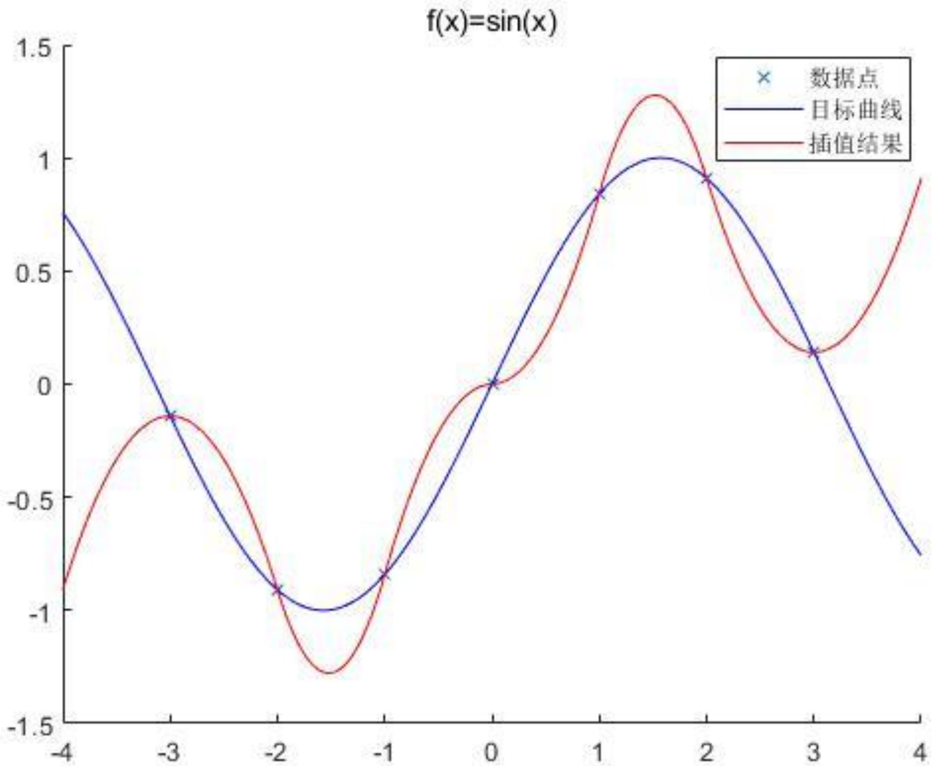
3.2 限制条件 1

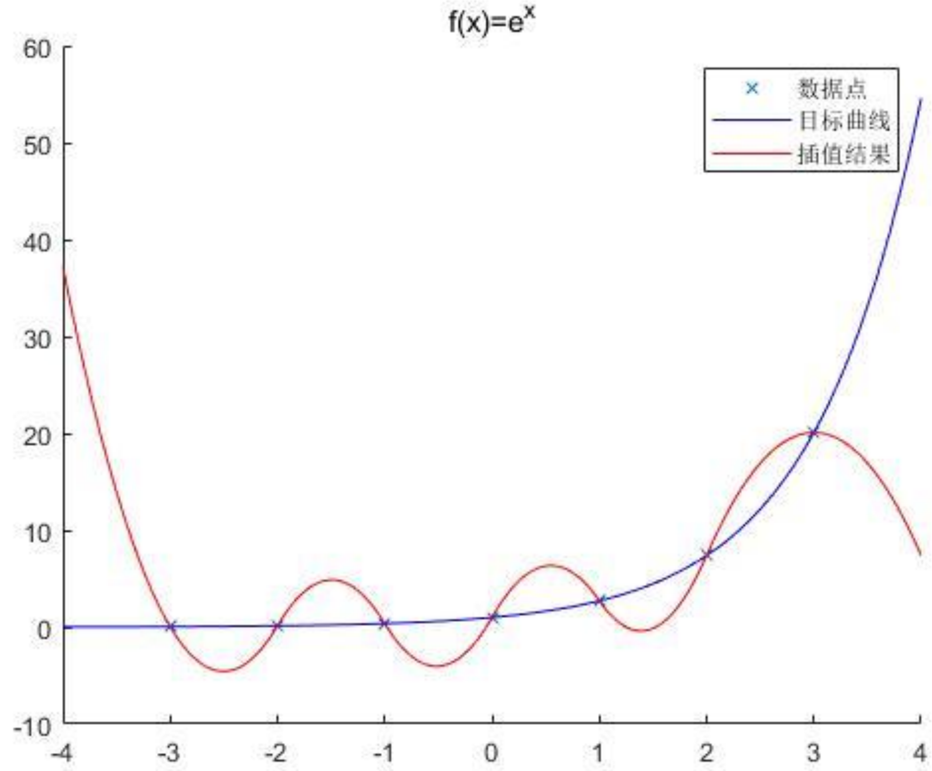
类比三次样条，在二次样条中，加入附加条件$m_k = 0$，得到如下$R^{(k-1)*(k-1)}$的方程组并进行测试。加入$m_1 = 0$的结果与之类似，不再进行实验。

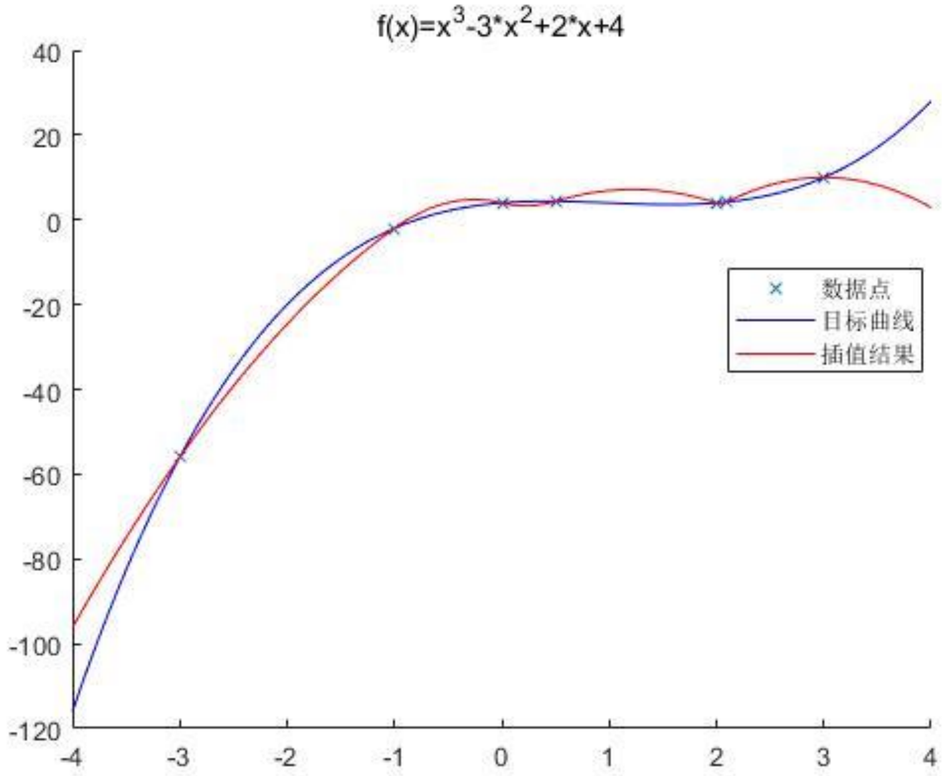
$$\begin{bmatrix} 1 & 1 & 0 & & 0 & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & & 0 & 0 & 0 \\ & \vdots & & \ddots & & \vdots & \\ 0 & 0 & 0 & & 1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 1 \\ 0 & 0 & 0 & & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ \vdots \\ M_{k-3} \\ M_{k-2} \\ M_{k-1} \end{bmatrix} = 2 * \begin{bmatrix} \frac{y_2 - y_1}{h_1} \\ \frac{y_3 - y_2}{h_2} \\ \frac{y_4 - y_3}{h_3} \\ \vdots \\ \frac{y_{k-2} - y_{k-3}}{h_{k-3}} \\ \frac{y_{k-1} - y_{k-2}}{h_{k-2}} \\ \frac{y_k - y_{k-1}}{h_{k-1}} \end{bmatrix}$$


f(x)=x^2-2*x+3
数据点
目标曲线
插值结果
35
30
25
20
15
10
5
0
-4
-3
-2
-1
0
1
2
3
4


f(x)=x^3-3*x^2+2*x+4
数据点
目标曲线
插值结果
40
20
0
-20
-40
-60
-80
-100
-120
-4
-3
-2
-1
0
1
2
3
4


f(x)=sin(x)
数据点
目标曲线
插值结果
1.5
1
0.5
0
-0.5
-1
-1.5
-4
-3
-2
-1
0
1
2
3
4


f(x)=e^x
数据点
目标曲线
插值结果
60
50
40
30
20
10
0
-10
-4
-3
-2
-1
0
1
2
3
4

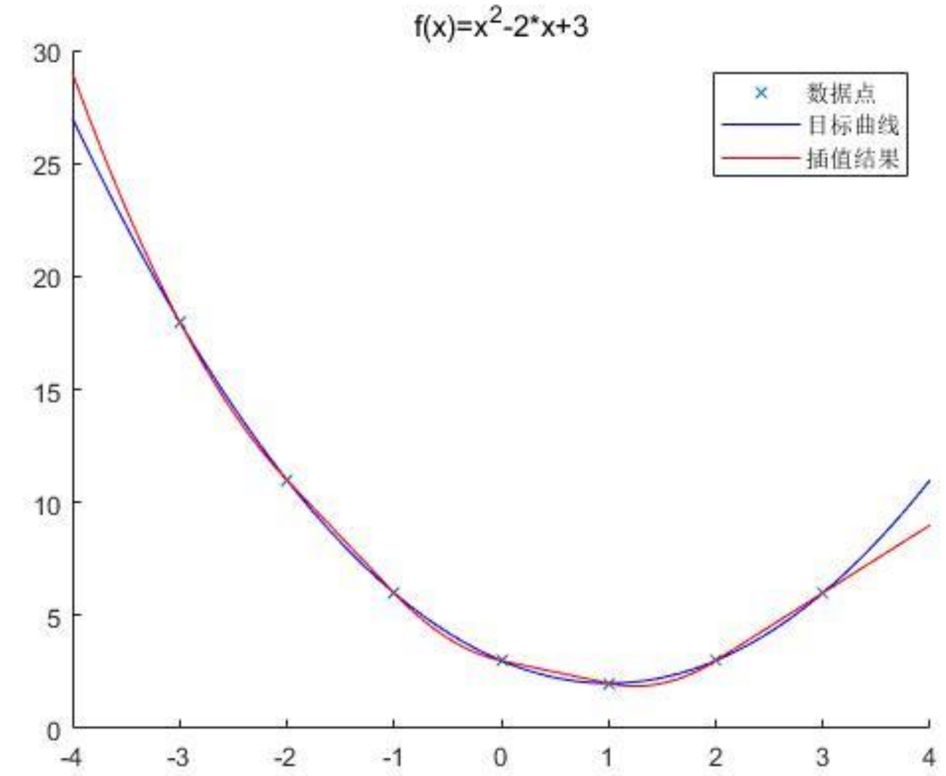


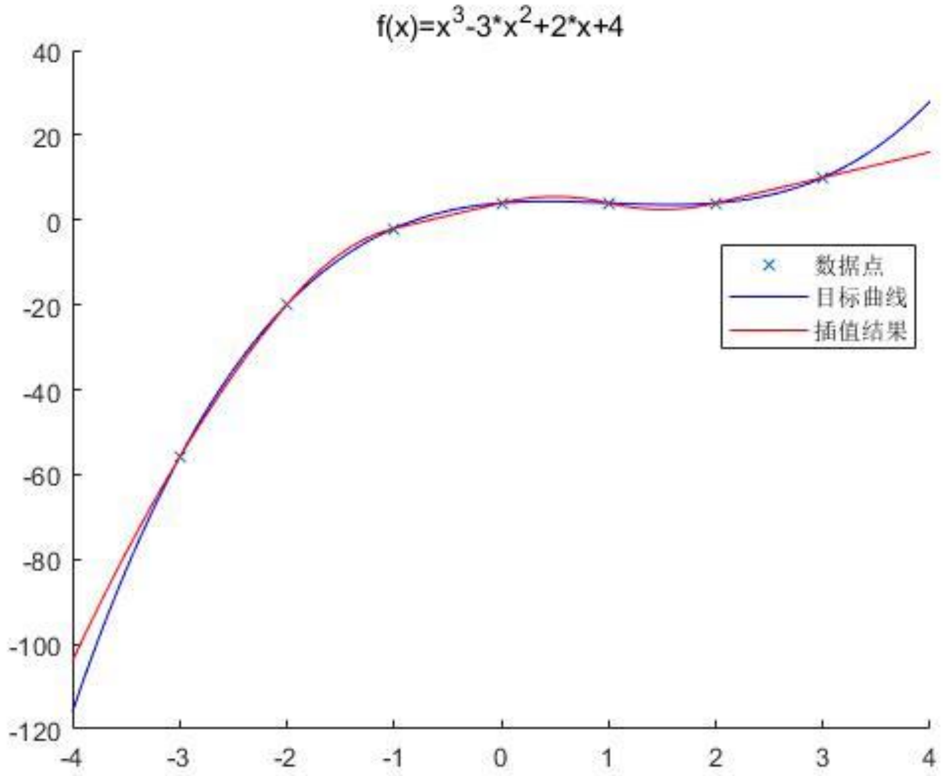
函数	内插值误差	外插值误差
$f_1(x)$	0.52805	16.6881
$f_2(x)$	3.6398	145.1202
$f_3(x)$	0.038583	0.78177
$f_4(x)$	11.4303	450.0108
$f_5(x)$	6.1039	130.7351

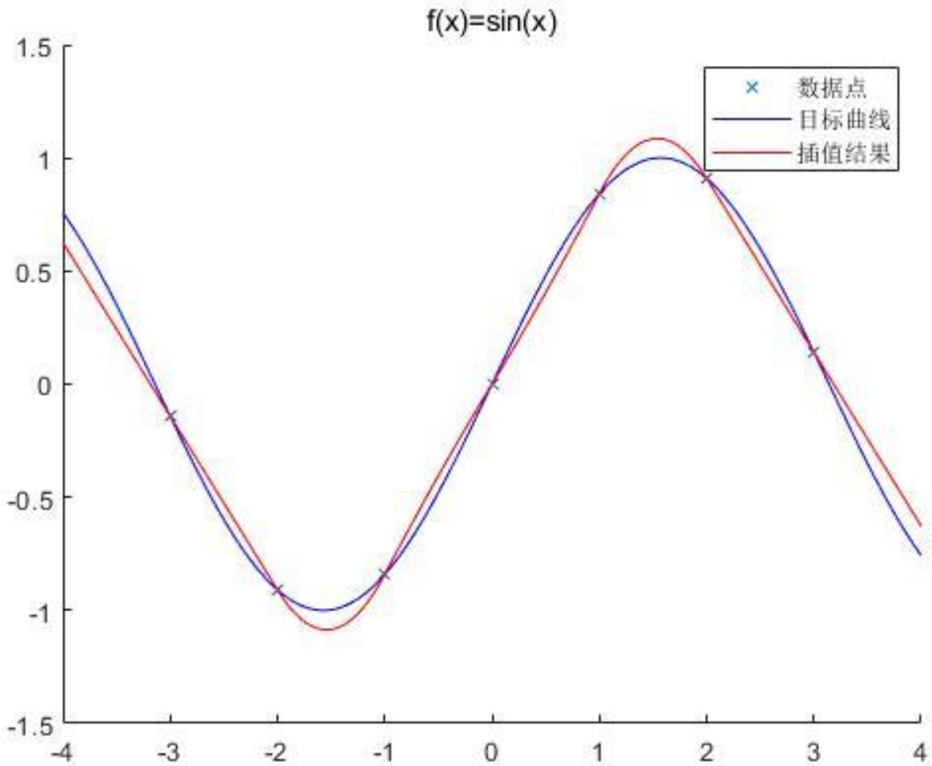
3.3 限制条件 2

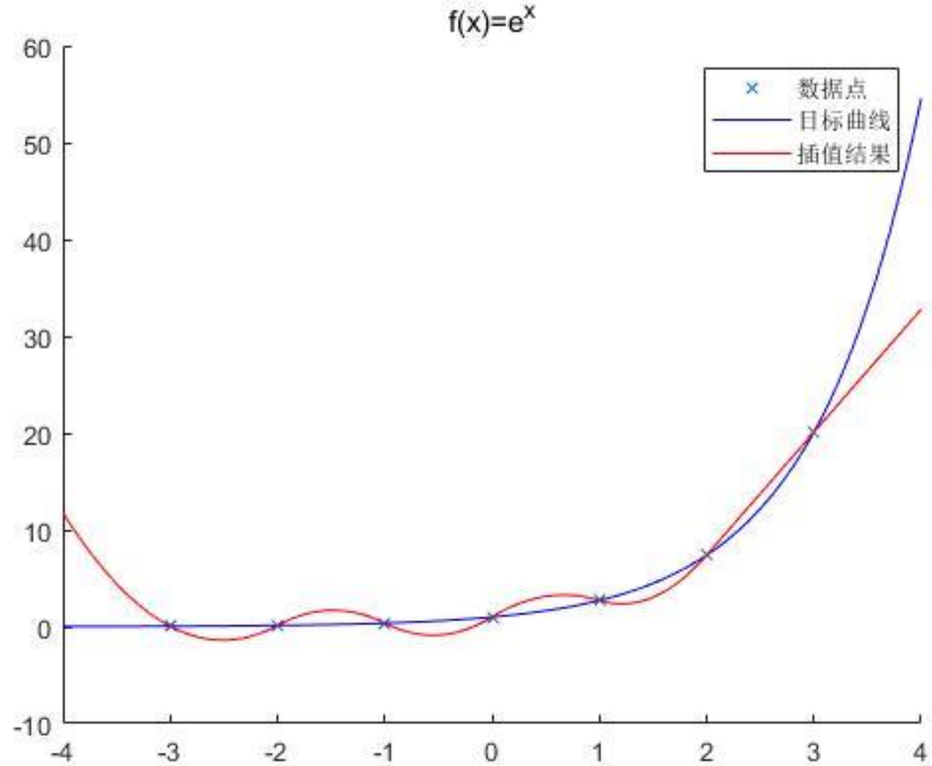
类比三次样条，在二次样条中，加入附加条件$m_k = m_{k-1}$，得到如下$R^{(k-1)*(k-1)}$的方程组并进行测试。加入$m_1 = m_2$的结果与之类似，不再进行实验。

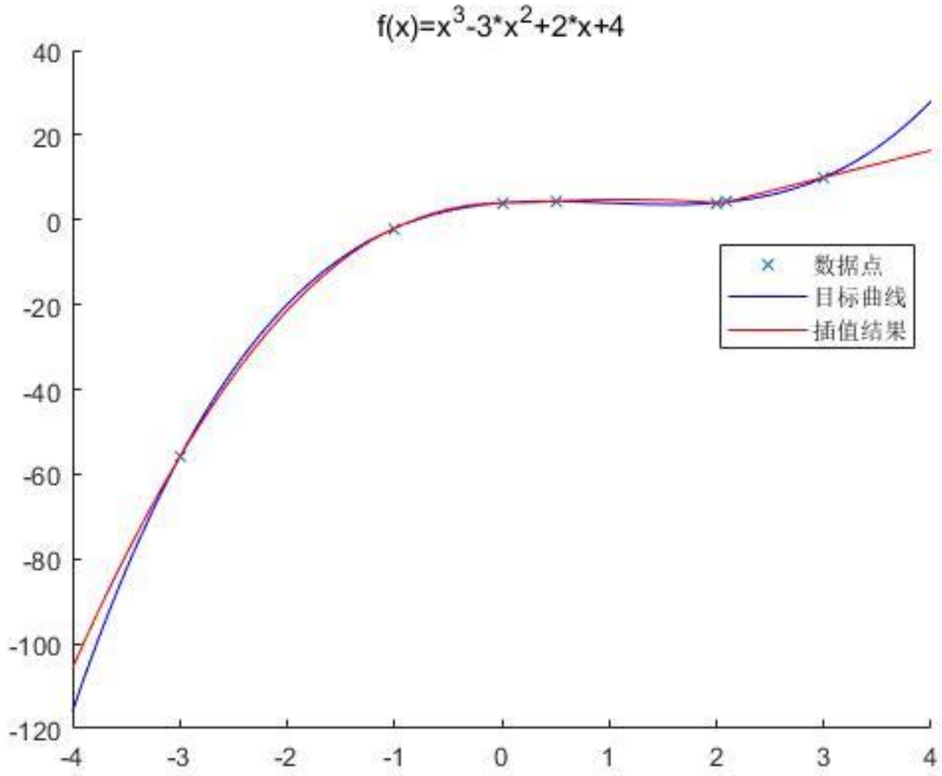
$$\begin{bmatrix} 1 & 1 & 0 & & 0 & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & & 0 & 0 & 0 \\ & \vdots & & \ddots & & \vdots & \\ 0 & 0 & 0 & & 1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 1 \\ 0 & 0 & 0 & & 0 & 0 & 2 \end{bmatrix} * \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ \vdots \\ M_{k-3} \\ M_{k-2} \\ M_{k-1} \end{bmatrix} = 2 * \begin{bmatrix} \frac{y_2 - y_1}{h_1} \\ \frac{y_3 - y_2}{h_2} \\ \frac{y_4 - y_3}{h_3} \\ \vdots \\ \frac{y_{k-2} - y_{k-3}}{h_{k-3}} \\ \frac{y_{k-1} - y_{k-2}}{h_{k-2}} \\ \frac{y_k - y_{k-1}}{h_{k-1}} \end{bmatrix}$$


f(x)=x2-2*x+3
数据点
目标曲线
插值结果


f(x)=x3-3*x2+2*x+4
数据点
目标曲线
插值结果


f(x)=sin(x)
数据点
目标曲线
插值结果
1.5
1
0.5
0
-0.5
-1
-1.5
-4
-3
-2
-1
0
1
2
3
4


$f(x)=e^x$
数据点
目标曲线
插值结果
60
50
40
30
20
10
0
-10
-4
-3
-2
-1
0
1
2
3
4

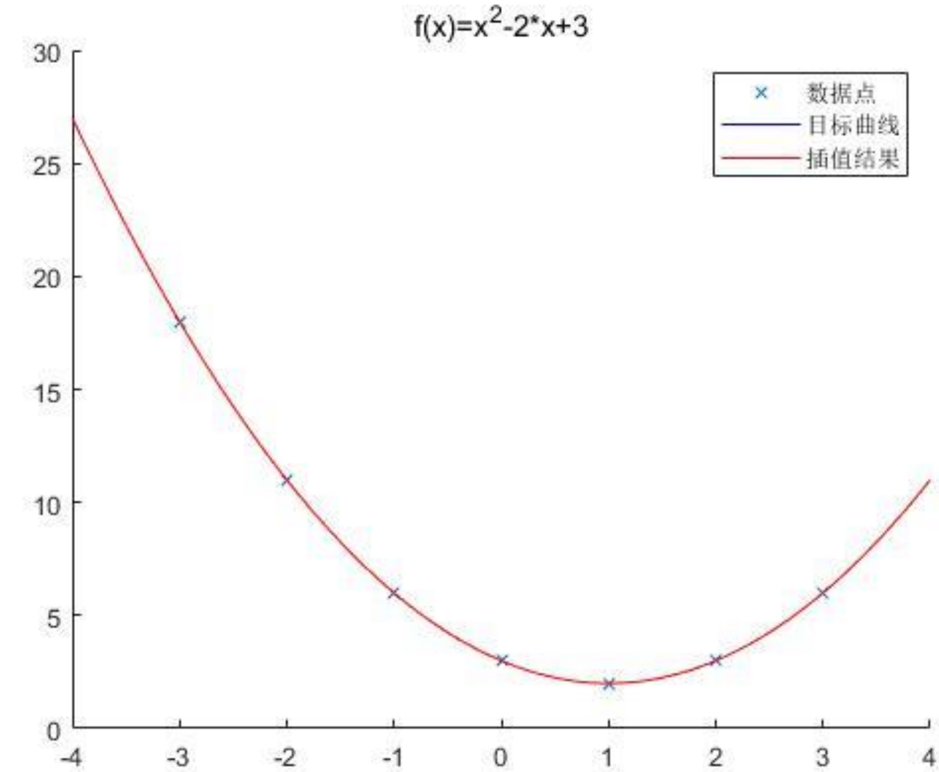


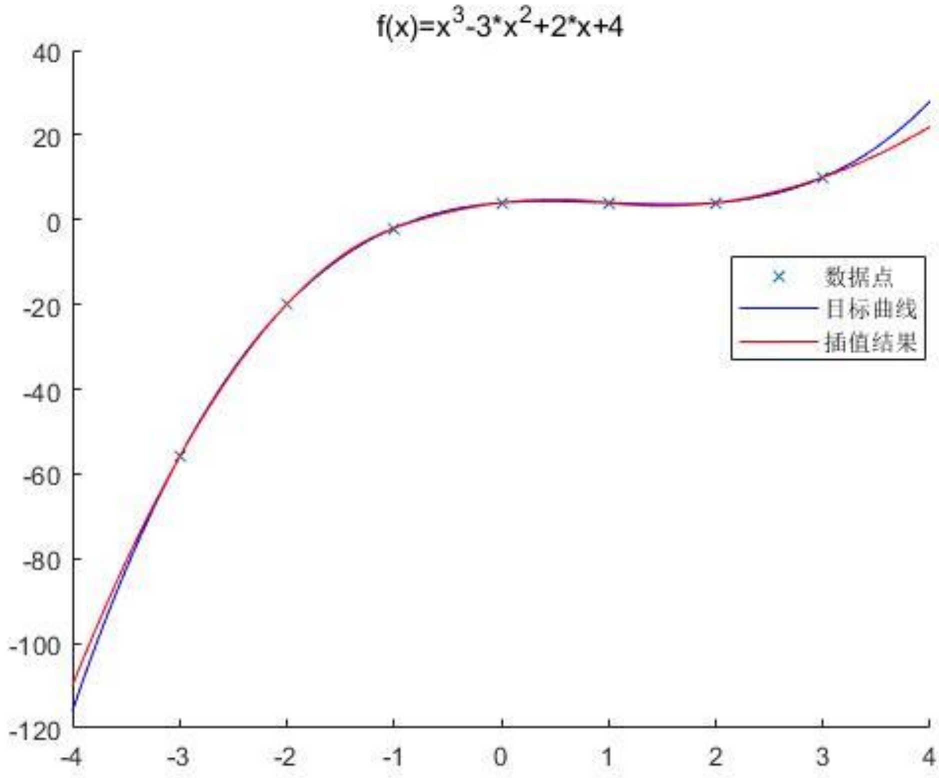
函数	内插值误差	外插值误差
$f_1(x)$	0.033003	1.043
$f_2(x)$	0.6695	35.0497
$f_3(x)$	0.0032609	0.010625
$f_4(x)$	1.1554	70.9392
$f_5(x)$	0.58455	29.0741

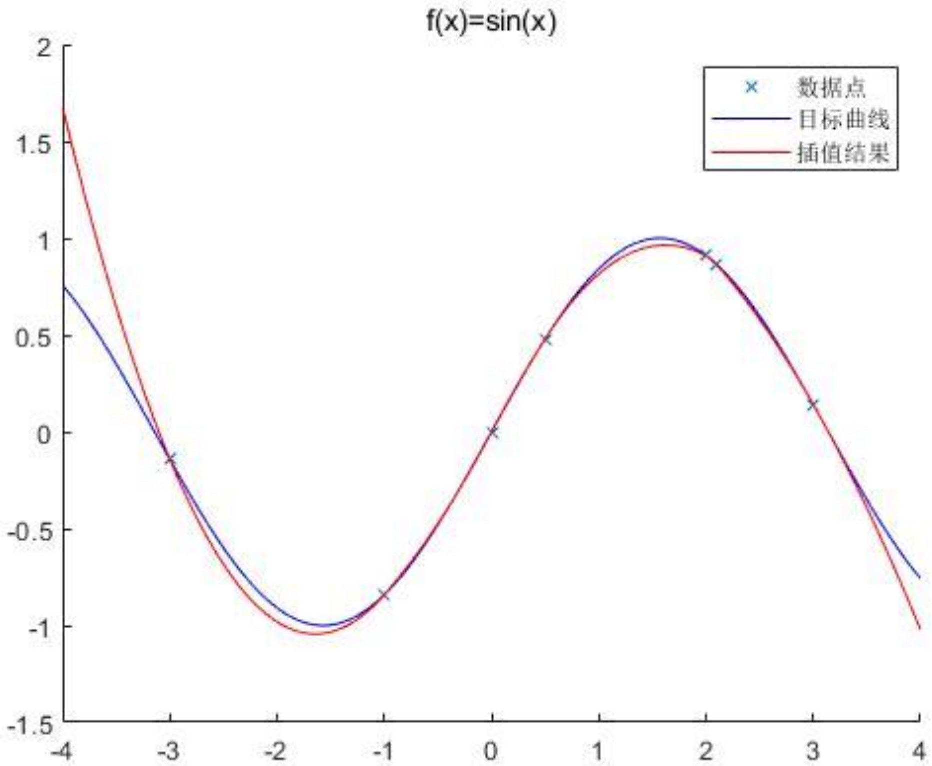
3.4 限制条件 3

类比三次样条，在二次样条中，加入附加条件$m_k = 2 * m_{k-1} - m_{k-2}$，得到如下$R^{(k-1)*(k-1)}$的方程组并进行测试。

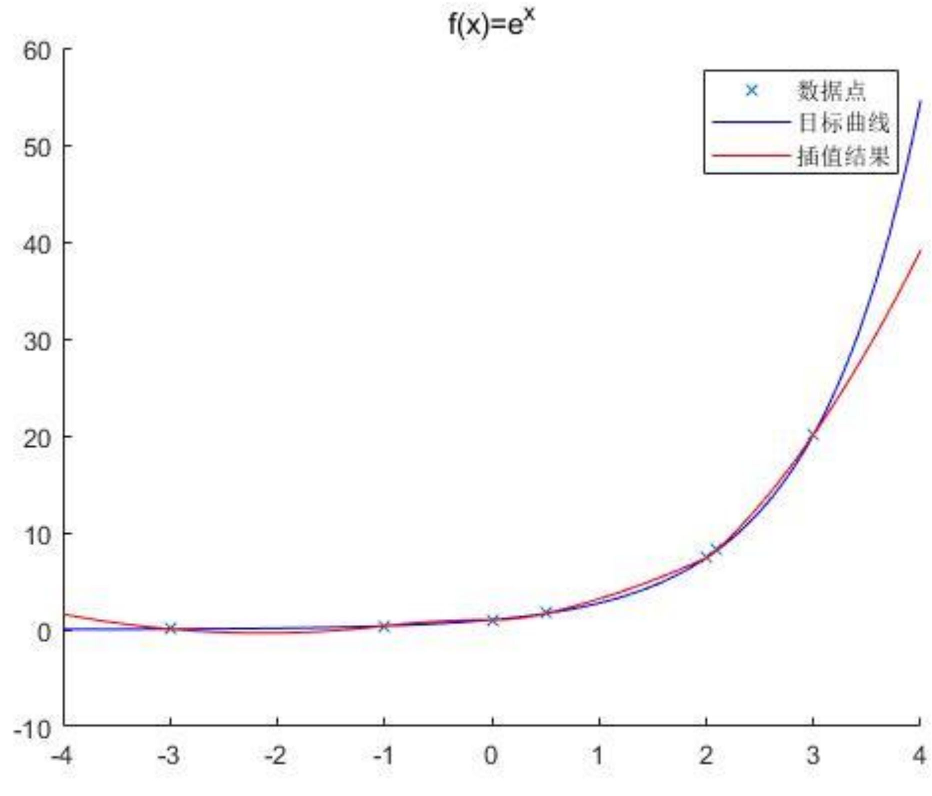
$$\begin{bmatrix} 1 & 1 & 0 & & 0 & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & & 0 & 0 & 0 \\ & \vdots & & \ddots & & \vdots & \\ 0 & 0 & 0 & & 1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 1 \\ 0 & 0 & 0 & & 0 & -1 & 3 \end{bmatrix} * \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ \vdots \\ M_{k-3} \\ M_{k-2} \\ M_{k-1} \end{bmatrix} = 2 * \begin{bmatrix} \frac{y_2 - y_1}{h_1} \\ \frac{y_3 - y_2}{h_2} \\ \frac{y_4 - y_3}{h_3} \\ \vdots \\ \frac{y_{k-2} - y_{k-3}}{h_{k-3}} \\ \frac{y_{k-1} - y_{k-2}}{h_{k-2}} \\ \frac{y_k - y_{k-1}}{h_{k-1}} \end{bmatrix}$$


f(x)=x2-2*x+3
数据点
目标曲线
插值结果

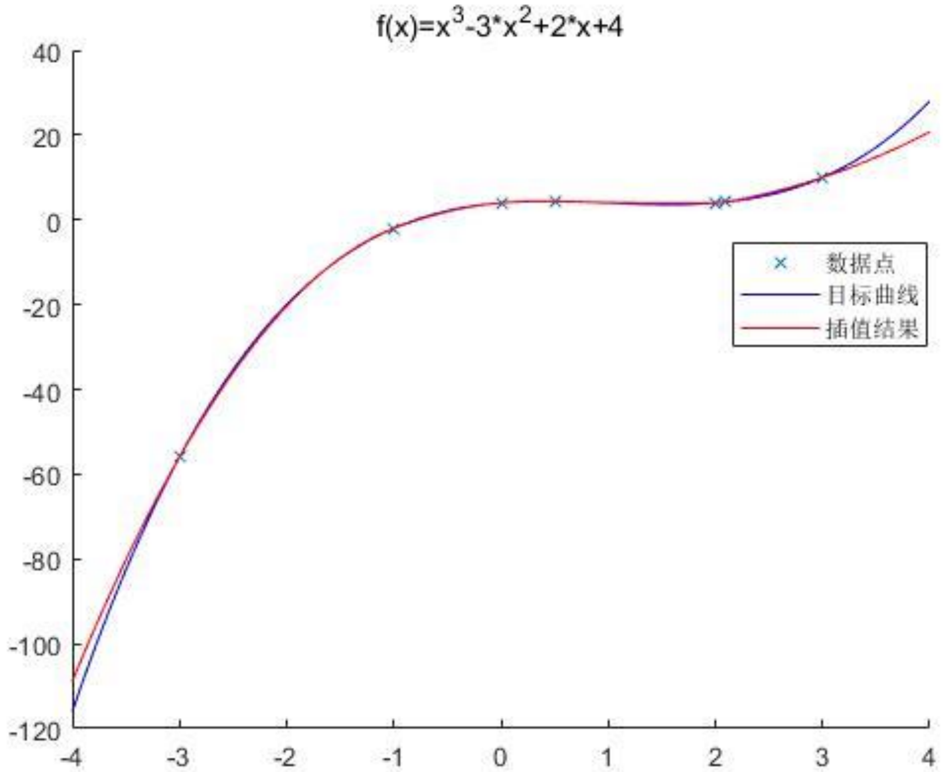

f(x)=x3-3*x2+2*x+4
数据点
目标曲线
插值结果



f(x)=sin(x)
数据点
目标曲线
插值结果
2
1.5
1
0.5
0
-0.5
-1
-1.5
-4
-3
-2
-1
0
1
2
3
4



f(x)=e^x
数据点
目标曲线
插值结果
60
50
40
30
20
10
0
-10
-4
-3
-2
-1
0
1
2
3
4



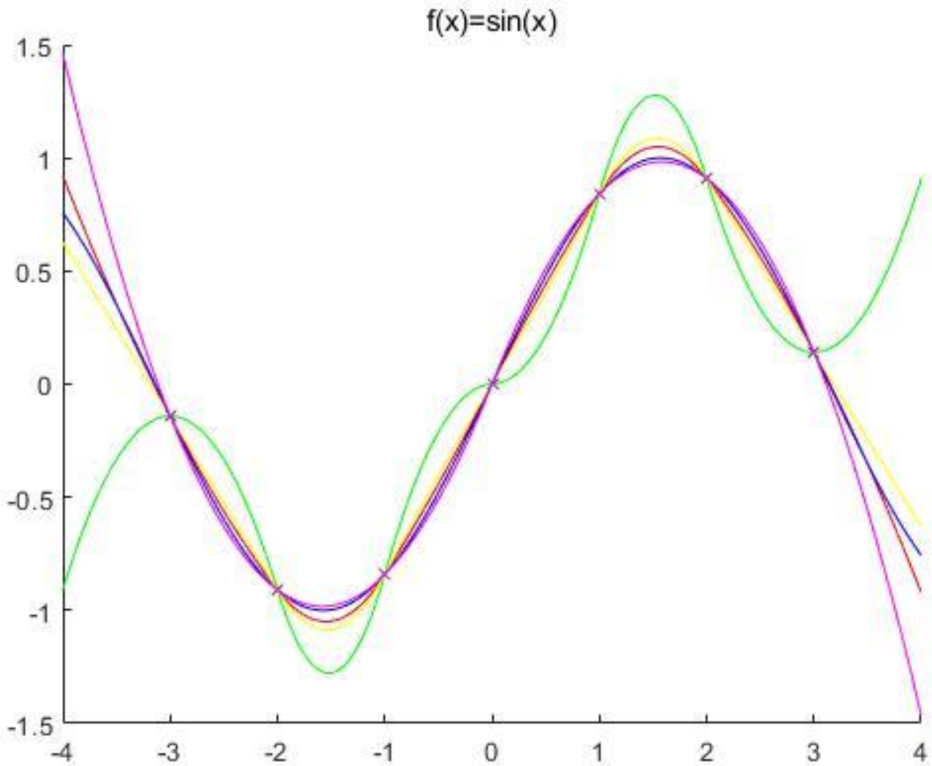
函数	内插值误差	外插值误差
$f_1(x)$	9.7957e-32	1.0778e-30
$f_2(x)$	0.075436	8.1755
$f_3(x)$	0.0013613	0.10348
$f_4(x)$	1.1554	70.9392
$f_5(x)$	0.13413	25.232

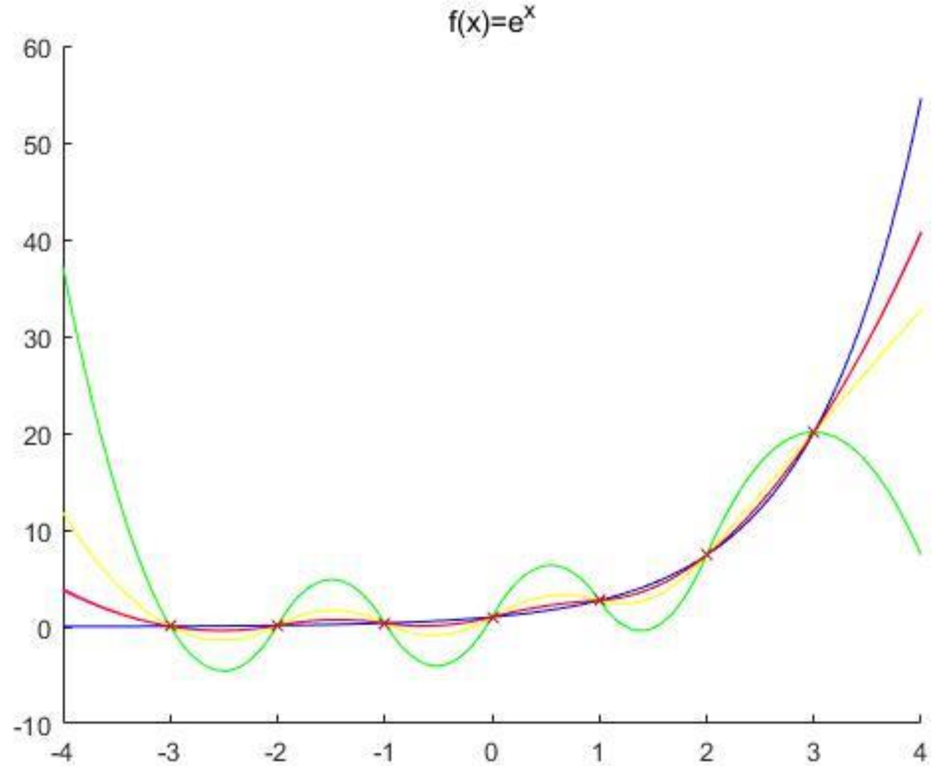
4 结论

实验的总体表现见下表。我们观察一下总体的表现，无论是内插值还是外插值，限制条件 4 的表现都是最好的，其次是限制条件 1，即不加限制直接求伪逆。如果目标函数已知，那么可以根据目标函数构造二次样条的附加条件，比如限制条件 2 就是根据二次函数求导直接构造的，因此它在二次函数的插值中表现优异，几乎完全收敛。构造的限制条件一般比伪逆的效果更好。

不过在现实问题中，一般很难能够提前知道目标函数的类型，甚至很难分析数据。在目标函数未知的情况下，其实直接求伪逆是一种很棒的选择，由上面的实验结果我们可以发现求伪逆的效果是很棒的。比如指数函数和三角函数，我们没有根据这两个函数构造限制条件，使用其他限制条件的运行结果都不如伪逆。

函数	内插值 best	外插值 best
$f_1(x)$	Condition 4	Condition 4
$f_2(x)$	Condition 4	Condition 4
$f_3(x)$	Condition 1	Condition 1
$f_4(x)$	Condition 1	Condition 1
$f_5(x)$	Condition 4	Condition 4


f(x)=sin(x)
1.5
1
0.5
0
-0.5
-1
-1.5
-4
-3
-2
-1
0
1
2
3
4


f(x)=e^x
60
50
40
30
20
10
0
-10
-4
-3
-2
-1
0
1
2
3
4