**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。

1. **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，，……，，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。

二次样条插值可以使用的几个性质

方便推导，定义以下几个参数，

对求导，可以得到，

根据一阶导数连续性和插值性，将和依次代入，将代入，可以解出，，，

根据连续性，将代入，可以得到，

化简可得，

因此，我们可以根据以上形式构造方程组，

写成矩阵乘法的形式，

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，再代回，即可得到二次样条插值函数。

1. **实验结果**

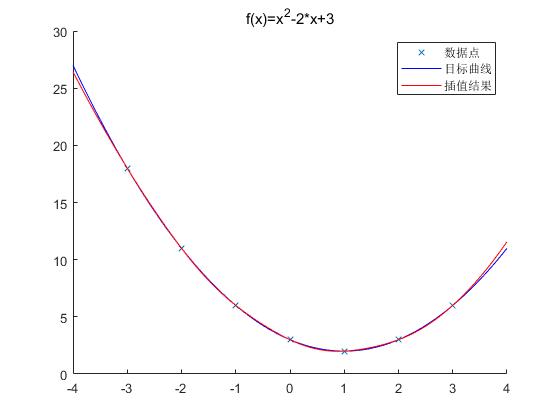
根据以上推导结果，我们得到了二次样条插值的计算方法。先使用矩阵乘法就出M向量的值，再带回求出a，b，c向量的值，即得到了二次样条插值多项式。根据不同的边界条件，我们可以得到不同的插值结果。我们将根据不同类型的函数，使用均方误差评估二次样条插值的准确性，并绘出图像。

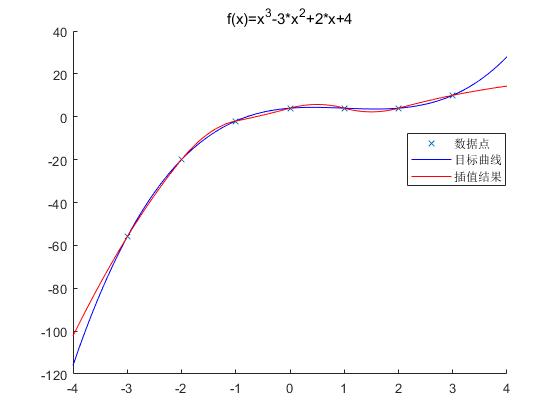
为了统一标准，以下为测试函数，

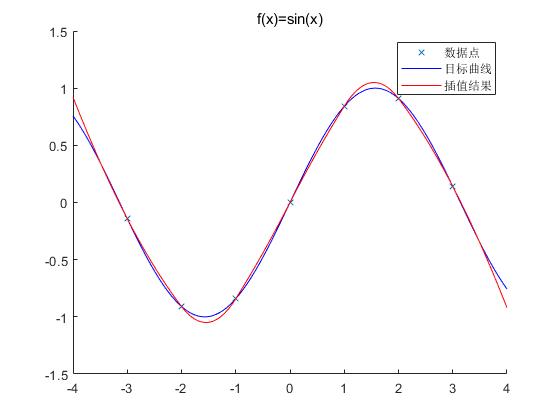
以下为分析误差，

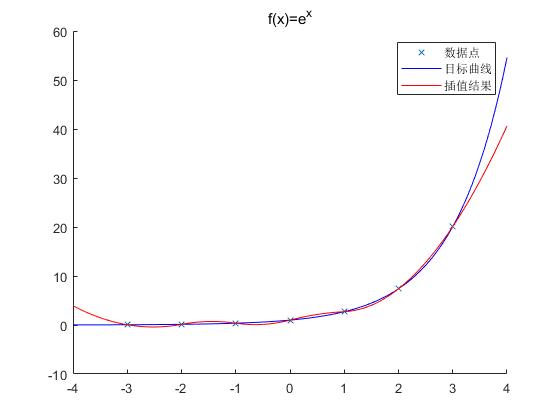
* 1. **伪逆**

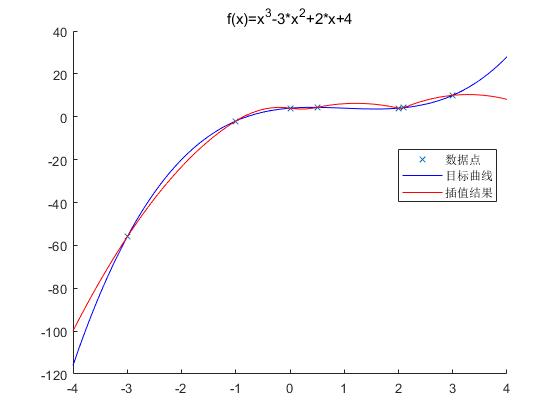
我尝试不引入其他约束条件，直接采用矩阵求伪逆的方法求出M向量。







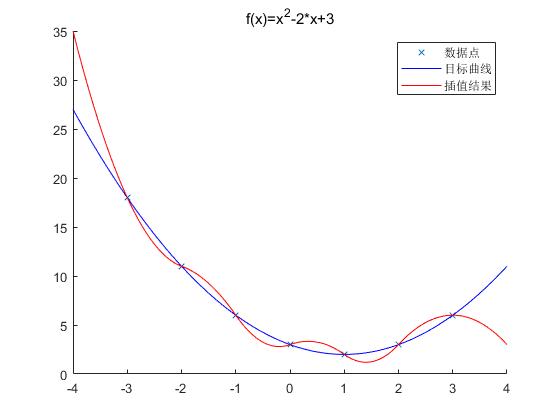


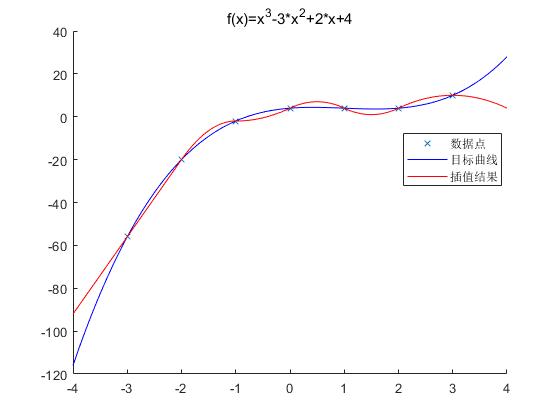


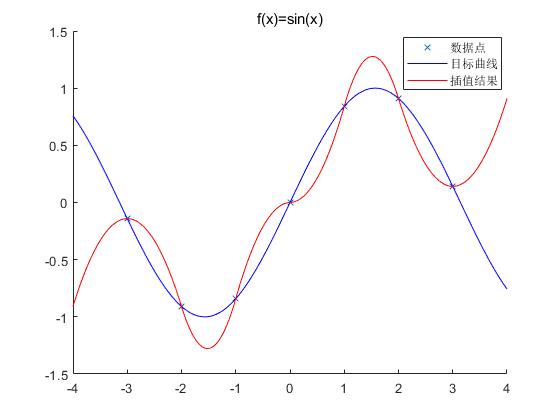
|  |  |  |
| --- | --- | --- |
| 函数 | 内插值误差 | 外插值误差 |
|  | 0.0026941 | 0.085143 |
|  | 0.94834 | 46.1763 |
|  | 0.00094217 | 0.003367 |
|  | 0.13188 | 21.7828 |
|  | 3.2647 | 82.8024 |

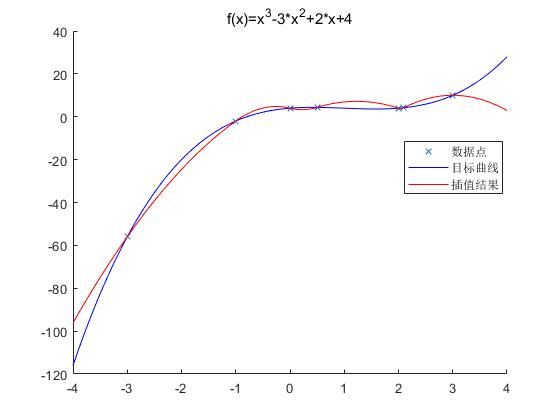
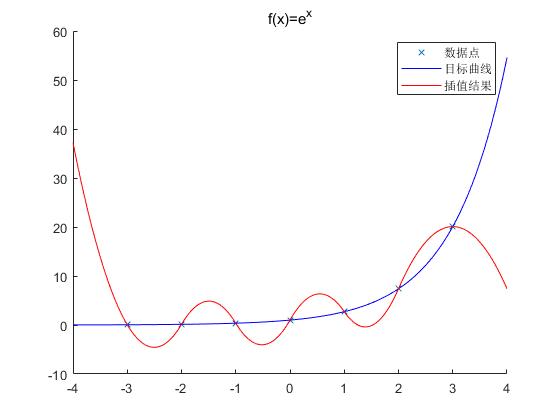
* 1. **限制条件1**

类比三次样条，在二次样条中，加入附加条件，得到如下的方程组并进行测试。加入的结果与之类似，不再进行实验。





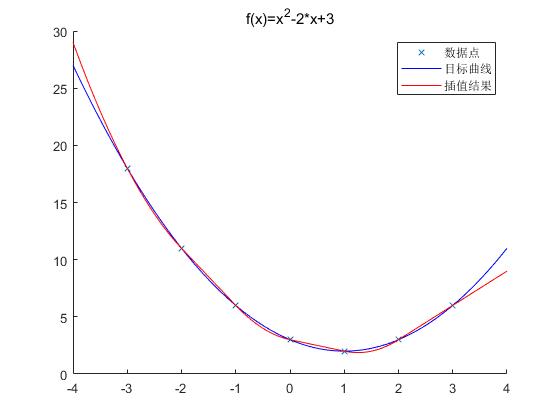


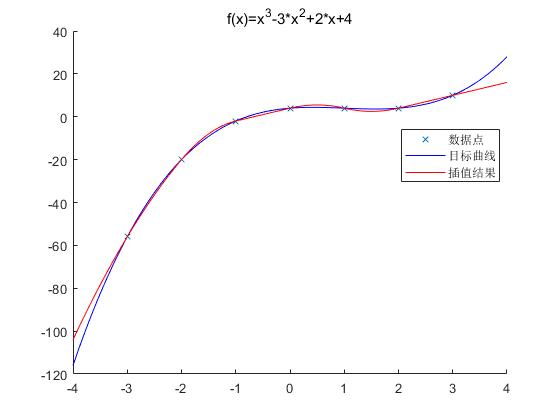


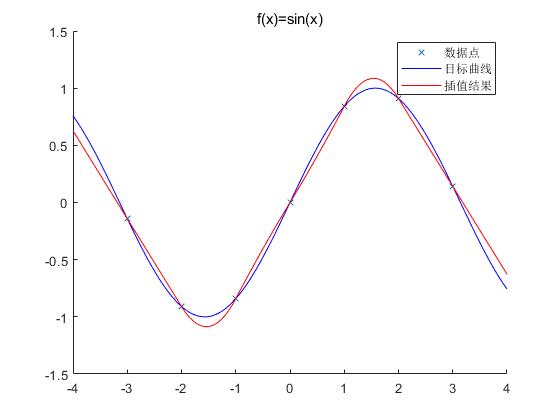
|  |  |  |
| --- | --- | --- |
| 函数 | 内插值误差 | 外插值误差 |
|  | 0.52805 | 16.6881 |
|  | 3.6398 | 145.1202 |
|  | 0.038583 | 0.78177 |
|  | 11.4303 | 450.0108 |
|  | 6.1039 | 130.7351 |

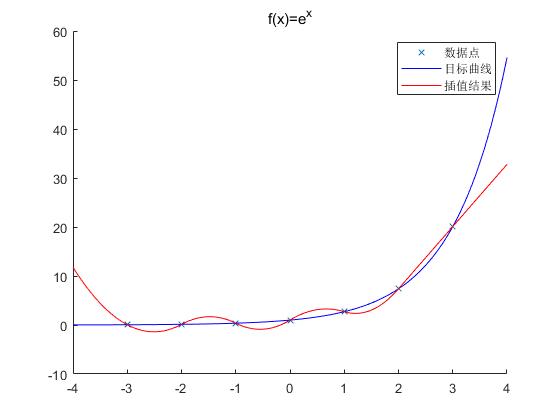
* 1. **限制条件2**

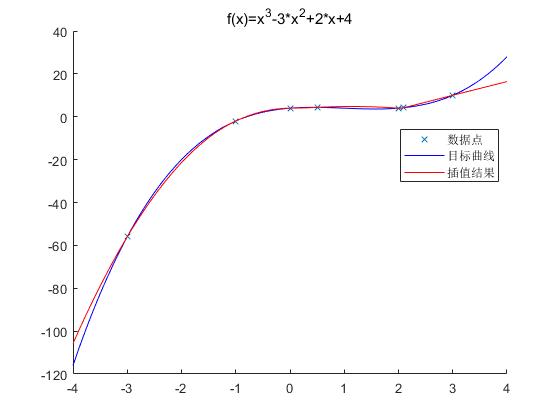
类比三次样条，在二次样条中，加入附加条件，得到如下的方程组并进行测试。加入的结果与之类似，不再进行实验。







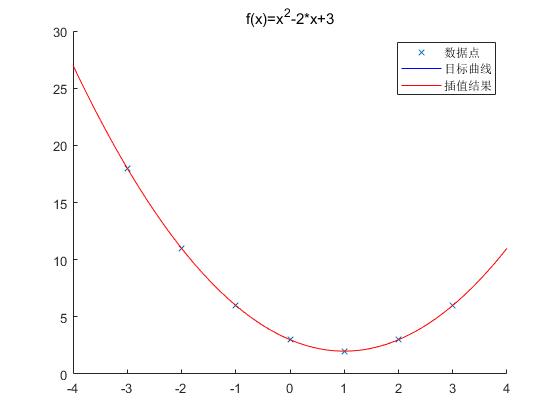


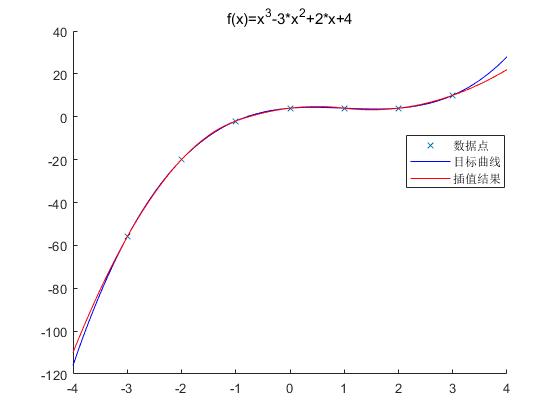


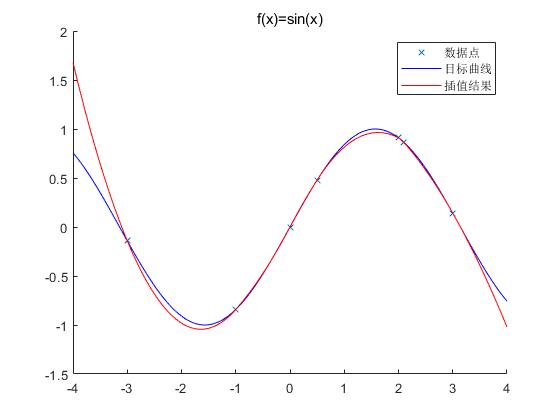
|  |  |  |
| --- | --- | --- |
| 函数 | 内插值误差 | 外插值误差 |
|  | 0.033003 | 1.043 |
|  | 0.6695 | 35.0497 |
|  | 0.0032609 | 0.010625 |
|  | 1.1554 | 70.9392 |
|  | 0.58455 | 29.0741 |

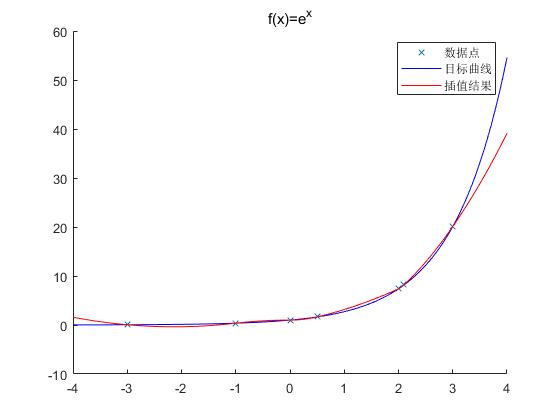
* 1. **限制条件3**

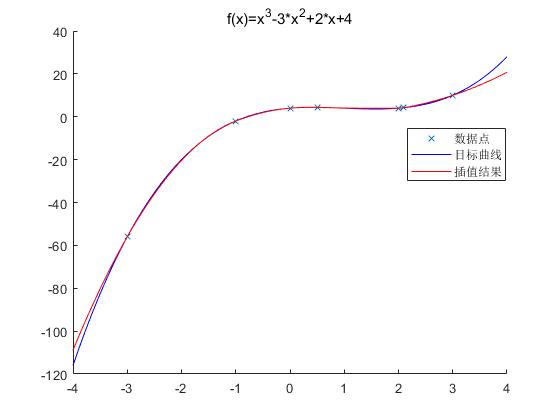
类比三次样条，在二次样条中，加入附加条件，得到如下的方程组并进行测试。











|  |  |  |
| --- | --- | --- |
| 函数 | 内插值误差 | 外插值误差 |
|  | 9.7957e-32 | 1.0778e-30 |
|  | 0.075436 | 8.1755 |
|  | 0.0013613 | 0.10348 |
|  | 1.1554 | 70.9392 |
|  | 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 |
|  | Condition 4 | Condition 4 |
|  | Condition 4 | Condition 4 |
|  | Condition 1 | Condition 1 |
|  | Condition 1 | Condition 1 |
|  | Condition 4 | Condition 4 |

