Lagrange 插值多项式外推时的误差公式与数值实验

谷正阳 [1]

（1.学号 18308045）

摘　要：拉格朗日插值法是一种常用于外推的方法，但是教材中并未给出其误差项的证明，所以给出*证明*。并直接验证，和通过*大 O 表示*，*含测量误差的误差项*推论间接验证。

关键词：拉格朗日外推估计；误差项；大 O 表示；含测量误差的误差项

Error Formula and Numerical Experiment of Lagrange Interpolation Polynomial Extrapolation

Zhengyang Gu[1]

(1.Student ID 18308045)

Abstract：Lagrange interpolation is a method commonly used for extrapolation, but the textbook does not give a proof of its error term, so the proof is given here. And directly verify it and indirectly verify it using corollaries related to big O notation and error term involving measure error.

Key words：Lagrange extrapolation; error term; big O notation; error term involving measure error

引　言

在数值计算领域中，外推法是一类用于根据变量之间的关系，估计超出原始观察范围值的估计。其中拉格朗日插值法是一种常用于外推的方法。

1　拉格朗日插值法

在数值分析领域中，拉格朗日插值法是以法国十八世纪数学家约瑟夫・拉格朗日命名的一种多项式插值方法。

1.1 拉格朗日系数多项式

假设 $(x_i, y_i), 0 \le i \le N$ 为已知的 $N+1$ 个点，并且假设 $(x, P_N(x))$ 为待估计的点，则拉格朗日系数多项式[2]有如下形式：

$$P_N(x) = \sum_{k=0}^{N} y_k L_{N,k}(x) \tag{1.1.1}$$

其中：

$$L_{N,k}(x) = \frac{\prod_{i=0,i\ne k}^{N}(x - x_i)}{\prod_{i=0,i\ne k}^{N}(x_k - x_i)} \tag{1.1.2}$$

可以利用该多项式，进行插值估计或外推估计[2]。

1.2 拉格朗日多项式估计

应用(1.1.1)和(1.1.2)进行插值估计或外推估计，有如下定理[2]。

定理1.　假设 $f \in C^{N+1}[a,b]$，且有：

$x_0, x_1, \cdots, x_N \in [a,b]$。

如果 $x \in [a,b]$，那么有：

$$f(x) = P_N(x) + E_N(x) \tag{1.2.1}$$

这里 $P_N(x)$ 是一个可以被用来估计 $f(x)$ 的多项式：

$$f(x) \approx P_N(x) = \sum_{k=0}^{N} f(x_k) L_{N,k}(x) \tag{1.2.2}$$

误差项 $E_N(x)$ 有如下形式：

$$E_N(x) = \frac{\prod_{i=0}^{N}(x - x_i) f^{(N+1)}(c)}{(N+1)!} \tag{1.2.3}$$

下面着重对于误差项 $E_N(x)$ 进行证明。

1.2.1 拉格朗日估计误差项在 $N=1$ 且插值情况下的证明

由于拉格朗日插值误差项 $E_N(x)$ 在 $N=1$ 的情况[2]比较简单，可能可以对一般的情况做出启发，所以先考虑这种特殊的情况。

证明1.　考虑 $N=1$ 的情况，定义如下特殊的函数 $g(t)$：

$$g(t)=f(t)-P_1(t)-E_1(x)\frac{(t-x_0)(t-x_1)}{(x-x_0)(x-x_1)} \tag{1.2.4}$$

注意，其中 x,x_0,x_1 对于 t 是常数，且这 x,x_0,x_1 分别是 $g(t)$ 的零点：

$$\begin{aligned} g(x) &= f(x)-P_1(x)-E_1(x)\frac{(x-x_0)(x-x_1)}{(x-x_0)(x-x_1)} \\ &= f(x)-P_1(x)-E_1(x) \\ &= 0 \end{aligned} \tag{1.2.5}$$

$$\begin{aligned} g(x_0) &= f(x_0)-P_1(x_0)-E_1(x)\frac{(x_0-x_0)(x_0-x_1)}{(x-x_0)(x-x_1)} \\ &= f(x_0)-P_1(x_0) \\ &= 0 \end{aligned} \tag{1.2.6}$$

$$\begin{aligned} g(x_1) &= f(x_1)-P_1(x_1)-E_1(x)\frac{(x_1-x_0)(x_1-x_1)}{(x-x_0)(x-x_1)} \\ &= f(x_1)-P_1(x_1) \\ &= 0 \end{aligned} \tag{1.2.7}$$

假设 x 在开区间 (x_0,x_1) 中。对 $g(t)$ 在区间 $[x_0,x]$ 应用罗尔定理，产生一个值 d_0，满足 $x_0<d_0<x$ 且有：

$$g'(d_0)=0 \tag{1.2.8}$$

对 $g(t)$ 在区间 $[x,x_1]$ 应用罗尔定理，则可以得到存在一个 d_1，满足 $x<d_1<x_1$ 且有：

$$g'(d_1)=0 \tag{1.2.9}$$

(1.2.8)(1.2.9)表示 $g'(t)$ 有零点 d_0 和 d_1。对 $g'(t)$ 在区间 $[d_0,d_1]$ 应用罗尔定理，得到存在一个值 c 满足 $d_0<c<d_1$ 且有：

$$g''(c)=0 \tag{1.2.10}$$

利用(1.2.4)计算 $g'(t)$ 和 $g''(t)$：

$$g'(t)=f'(t)-P_1'(t)-E_1(x)\frac{(t-x_0)+(t-x_1)}{(x-x_0)(x-x_1)} \tag{1.2.11}$$

求解 $g''(t)$ 时，考虑一个隐含条件 $P_1(t)$ 是次数为 $N=1$ 的多项式，因此其二阶导数 $P_1''(t)=0$。

$$g''(t)=f''(t)-0-E_1(x)\frac{2}{(x-x_0)(x-x_1)} \tag{1.2.12}$$

(1.2.12)带入 $t=c$ 并联立(1.2.10)得：

$$0=f''(c)-E_1(x)\frac{2}{(x-x_0)(x-x_1)} \tag{1.2.13}$$

(1.2.13)解出 $E_1(x)$ 得到与(1.2.3)对应的误差项形式：

$$E_1(x)=\frac{\prod_{i=0}^{1}(x-x_i)f^{(2)}(c)}{2!} \tag{1.2.14}$$

1.2.2 拉格朗日估计误差项在一般情况下的证明

根据上面特殊情况的启发，发现每次对一个函数的若干区间使用罗尔定理，得到该函数导数的零点数是原函数零点数减1，考虑也可以多次使用罗尔定理，最终剩下一个零点，来求得 $E_N(x)$。

定义如下特殊的函数 $g(t)$：

$$g(t)=f(t)-P_N(t)-E_N(x)\frac{\prod_{i=0}^{N}(t-x_i)}{\prod_{i=0}^{N}(x-x_i)} \tag{1.2.15}$$

设 m 是 $x,x_0,x_1,\cdots,x_N$ 中的最小值，M 是最大值。

引理1. 对自然数 n 且 $n\le N+1$ 存在 $N+2-n$ 个数 $\xi_0,\xi_1,\cdots,\xi_{N+1-n}\in[m,M]$，且是 $g^{(n)}(t)$ 的零点。

证明2. 对 n 作归纳证明。

归纳基 当 $n=0$ 时，有：

$$
\begin{aligned}
g(x) &= f(x) - P_N(x) - E_N(x)\frac{\prod_{i=0}^{N}(x-x_i)}{\prod_{i=0}^{N}(x-x_i)} \\
&= f(x) - P_N(x) - E_N(x) = 0
\end{aligned}
\tag{1.2.16}
$$

对任意整数 j，满足 $0 \le j \le N$，有：

$$
\begin{aligned}
g(x_j) &= f(x_j) - P_N(x_j) - E_N(x)\frac{\prod_{i=0}^{N}(x_j-x_i)}{\prod_{i=0}^{N}(x-x_i)} \\
&= f(x_j) - P_N(x_j) \\
&= 0
\end{aligned}
\tag{1.2.17}
$$

即存在 $N+2$ 个数 $x, x_0, x_1, \cdots, x_N \in [m, M]$ 是 $g(t)$ 的零点。

归纳步 对自然数 k 且 $k \le N$，假设 $n = k$ 时存在 $N+2-k$ 个数 $a_0, a_1, \cdots, a_{N+1-k} \in [m, M]$，是 $g^{(k)}(t)$ 的零点

对于 $n = k+1$，对 $a_0, a_1, \cdots, a_{N+1-k}$ 从小到大排序产生一个新的序列 $c_0, c_1, \cdots, c_{N+1-k}$。则这 $N+2-k$ 个零点可以把区间 $[c_0, c_{N+1-k}]$ 分成 $N+1-k$ 个区间 $[c_i, c_{i+1}]$，其中自然数 $i \le N-k$。对 $g^{(k)}(t)$ 每个区间 $[c_i, c_{i+1}]$ 应用罗尔定理，产生 $N+1-k$ 个数 d_i，满足 $c_i < d_i < c_{i+1}$ 且有：

$$
g^{(k+1)}(d_i) = 0 \tag{1.2.18}
$$

存在 $N+1-k$ 个数 $d_0, d_1, \cdots, d_{N-k} \in [m, M]$ 是 $g^{(k+1)}(t)$ 的零点。

引理2. 对自然数 n，有：

$$
\frac{\mathrm{d}^{n+1}(\prod_{i=0}^{n}(t-x_i))}{\mathrm{d}t^{n+1}} = (n+1)! \tag{1.2.19}
$$

证明3. 设：

$$
\prod_{i=0}^{n}(t-x_i) = a_0 t^{n+1} + a_1 t^n + \cdots + a_n \tag{1.2.20}
$$

当且仅当连乘中每一项贡献 t，才能凑出 t^{n+1}，因而易得 $a_0 = 1$，而其他含 t 项 t 的次数均为正整数因而有：

$$
\frac{\mathrm{d}^{n+1}(\prod_{i=0}^{n}(t-x_i))}{\mathrm{d}t^{n+1}} = 1 \times (n+1) \times n \times \cdots \times 1 = (n+1)! \tag{1.2.21}
$$

下面通过引理 1 引理 2 证明 $E_N(x)$ 满足(1.2.3)形式。

证明4. 由引理 1 当 $n = N+1$ 时，存在 1 个数：$\xi_0 \in [m, M]$，是 $g^{(N+1)}(t)$ 的零点。

现在回到(1.2.15)计算 $g^{(N+1)}(t)$，由引理 2：

$$
g^{(N+1)}(t) = f^{(N+1)}(t) - 0 - E_N(x)\frac{(N+1)!}{\prod_{i=0}^{N}(x-x_i)} \tag{1.2.22}
$$

在(1.2.22)中，我们使用了隐含条件 $P_N(t)$ 是次数为 N 的多项式，因此其 $N+1$ 阶导数 $P^{(N+1)}(t) = 0$。在(1.2.22)中代入 $t = \xi_0$：

$$
0 = f^{(N+1)}(\xi_0) - E_N(x)\frac{(N+1)!}{\prod_{i=0}^{N}(x-x_i)} \tag{1.2.23}
$$

(1.2.23)解出 $E_N(x)$ 得到(1.2.3)对应的误差项形式：

$$
E_N(x) = \frac{\prod_{i=0}^{N}(x-x_i) f^{(N+1)}(\xi_0)}{(N+1)!} \tag{1.2.24}
$$

1.2.3 拉格朗日估计误差项的验证

考虑(1.2.3) $E_N(x)$ 的形式中，存在不确定的 $f^{(N+1)}(\xi_0)$。所以为了便于验证，将实验中被估计的函数设为 $N+1$ 次多项式，给 $N+1$ 个已知点，这样 $f^{(N+1)}(\xi_0)$ 是一个可知的常数。以下结果均保留四位小数。

表1. 拉格朗日估计误差项的验证

	$f(x)$	$P_N(x)$	$E_N(x)$
Test1	-0.0492	-0.3352	0.2860
Test2	-0.8809	-0.8817	8.6758e-04
Test3	0.5866	-13.2286	13.8152
Test4	-0.3605	-0.0663	-0.2941
Test5	-1.2095	-1.2088	-6.1920e-04

可以看出，满足 $f(x) \approx P_N(x) + E_N(x)$。

2 拉格朗日估计误差项的性质及推论

2.1 拉格朗日多项式误差的大 O 表示

观察(1.2.3)的形式，可以发现用拉格朗日估计在一定条件下，$E_N(x) = O(x^{N+1})$。

性质1. 若 $x \in [a,b]$，且对任意整数 i，满足 $0 \le i \le N$，有 $x, x_i \in [a,b]$，且在 $\left[a,b\right]$ 内 $f^{(N+1)}(\xi_0)$ 有界，则有：

$$E_N(x) = O(x^{N+1}) \tag{2.1.1}$$

2.1.1 拉格朗日多项式误差的大 O 表示证明

证明5. 设 x_i 中最小值为 m，最大值为 M，所以有：

$$\begin{aligned} |x - x_i| &\le \max\{|x-m|,|x-M|\} \\ &\le |x| + \max\{|m|,|M|\} \end{aligned}$$

所以有当 $|x| \ge 1$ 时：

$$\begin{aligned} \left|\prod_{i=0}^{N}(x-x_i)\right| &\le (|x| + \max\{|m|,|M|\})^{N+1} \\ &\le \max\{|m|,|M|,1\}^{N+1}(|x|+1)^{N+1} \\ &\le (2\max\{|m|,|M|,1\})^{N+1}|x|^{N+1} \end{aligned}$$

又因为 $f^{(N+1)}(\xi_0)$ 有界，设其中一个界 B。所以有：

$$|E_N(x)| \le \frac{(2\max\{|m|,|M|,1\})^{N+1}B}{(N+1)!}|x|^{N+1}$$

即 $E_N(x) = O(x^{N+1})$。

2.1.2 拉格朗日多项式误差的大 O 表示验证

因为余弦函数 $\cos x$ 的 n 阶导数为 $\cos(x + n\pi/2)$，其有界1，方便运算，所以选用 $\cos x$ 进行验证。

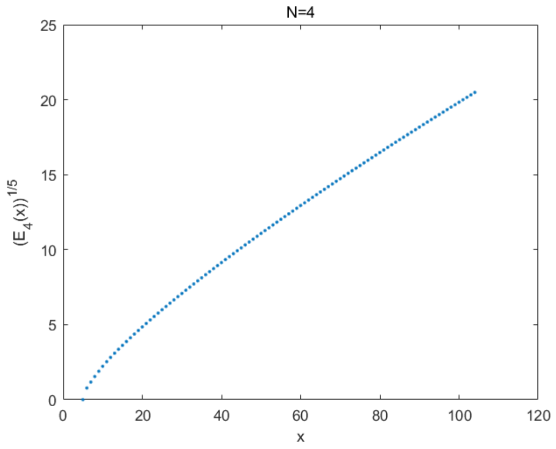

图1. $N=4$ 时的 x - $(E_4(x))^{(1/5)}$ 图像

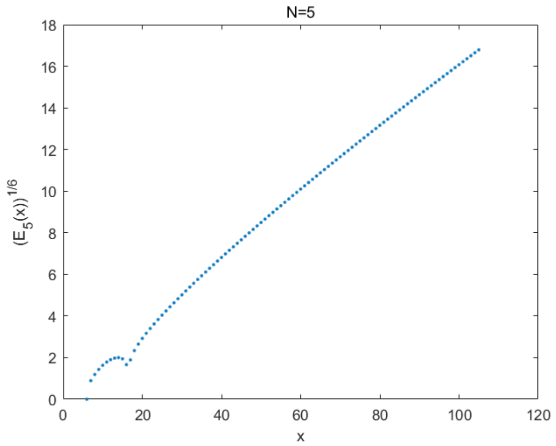

图2. $N=5$ 时的 x - $(E_5(x))^{(1/6)}$ 图像

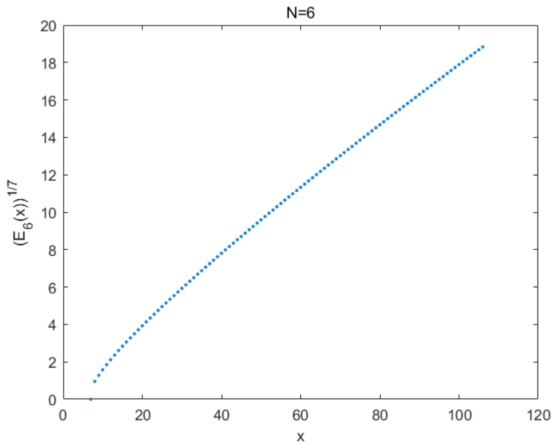

图3. $N=6$ 时的 x - $(E_r(x))^{(1/7)}$ 图像

可以看出，都是很接近线性的次线性，验证了大 O 表示的正确性。

2.2 含测量误差的点进行拉格朗日估计的误差项

考虑到拉格朗日估计的使用场景，有的时候是通过测量值获得测量得到，可能会有测量误差。所以在此引入误差。

推论1. 若 $N+1$ 个点有 $M+1$ 个点函数值精确知道，剩下点不精确知道。即设 $M \le N$，且设 $N+1$ 个点 $(x_0, y_0), \cdots, (x_M, y_M), (x_{M+1}, y_{M+1}), \cdots, (x_N, y_N)$，且有：

$$f(x_k) = \begin{cases} y_k, & 0 \le k \le M \\ y_k + {}_i, & M+1 \le k \le N \end{cases} \qquad (2.2.1)$$

则有：

$$f(x) = \sum_{k=0}^{N} y_k L_{N,k}(x) + \sum_{k=M+1}^{N} {}_k L_{N,k}(x) + E_N(x) \qquad (2.2.2)$$

设:

$$P_N'(x) = \sum_{k=0}^{N} y_k L_{N,k}(x) \qquad (2.2.3)$$

$$E_N'(x) = \sum_{k=M+1}^{N} {}_k L_{N,k}(x) + E_N(x) \qquad (2.2.4)$$

则有：

$$f(x) = P_N'(x) + E_N'(x) \qquad (2.2.5)$$

证明6. 结合(1.2.1)(1.2.2)(2.2.1)，易证。

2.2.1 含测量误差进行拉格朗日估计的误差项验证

为验证该推论，可以找一个导数易算且易算界的函数。在此选用 $\ln x$，其 n 阶导数是 $(-1)^{(n-1)}(n-1)!/x^n$，其为单调函数，存在界在 $\min\{\min_{0\le i\le N} x_i, x\}$ 和 $\max\{\max_{0\le i\le N} x_i, x\}$ 之中取。假设测量相对误差不超过 0.001。假设 E_{11} 和 E_{22} 为考虑了测量误差的边界，E_1 和 E_2 是未考虑测量误差的边界。

考虑 $-0.001 \le e = (y - f(x))/f(x) \le 0.001$，根据(2.2.1)就有：

$${}_i = \frac{y_i}{e+1} - y_i$$

所以 ${}_i / y_i \in [1/1.001-1, 1/0.999-1]$。根据(2.2.4)，可知考虑了误差的边界应为：

$$E_{11} = \sum_{k=M+1}^{N} e_{11}^k + E_1$$

$$E_{22} = \sum_{k=M+1}^{N} e_{22}^k + E_2$$

其中：

$$e_{11}^k = \begin{cases} (1/1.001-1) y_k L_{N,k}, & y_k L_{N,k} > 0 \\ (1/0.999-1) y_k L_{N,k}, & y_k L_{N,k} \le 0 \end{cases}$$

$$e_{22}^k = \begin{cases} (1/0.999-1) y_k L_{N,k}, & y_k L_{N,k} > 0 \\ (1/1.001-1) y_k L_{N,k}, & y_k L_{N,k} \le 0 \end{cases}$$

表2. 不考虑测量误差

	$f(x)$	$P_N'(x)+E_1$	$P_N'(x)+E_2$
Test1	3.1189	1.1958e+03	1.2444e+03
Test2	3.0603	3.7799	5.5529
Test3	2.7875	2.7704	2.8213

表3. 考虑测量误差

	$f(x)$	$P_N'(x)+E_{11}$	$P_N'(x)+E_{22}$
Test1	3.1189	-9.9691e+03	1.2409e+04
Test2	3.0603	0.2483	9.0845
Test3	3.2015	2.5235	3.0682

结果发现，Test2，Test3 表明不考虑测量误差给出范围可能是错的也可能是对的，而考虑测量误差则真实值在其范围内。Test1 表明细微的测量误差，可能造成极大的影响。

3 结论

通过直接的验证，和用推论间接验证，均符合推导出的误差公式(1.2.3)。所以验证误差公式结果是没有错误的。

参考文献:

[1] Varga R S . Extrapolation methods: theory and practice[J]. Numerical Algorithms, 1993, 4(2):305-305.

[2] Mathews J H, Fink K D. Numerical methods using MATLAB[M]. Upper Saddle River, NJ: Pearson prentice hall, 2004.