数值分析中期报告

陈高明　梁浩贤　汪泓宇

2022 年 11 月 13 日

摘要

最佳平方逼近是函数逼近论中的重要课题，本质上它研究函数在 L^2 范数意义下的收敛序列。在多项式最佳平方逼近中，以正交多项式作为逼近函数空间的基有许多好处，如在增加多项式的次数后，最佳逼近多项式只需增加一项新的正交多项式的常数倍，而对于之前的正交多项式的组合系数并不改变。最大误差研究函数在 L^∞ 范数意义下的收敛性。

然而不幸的是，函数在 L^2 意义下的收敛性无法推出 L^∞ 收敛，这往往需要单独的证明。我们研究在区间 $[-1,1]$ 上，用勒让德多项式函逼近数 $f(x)=e^x$。在编程中使用 python 的 sympy 包，计算过程都为解析的。本文利用递推关系计算了 38 次勒让德多项式的解析解，并求解 f 的各次最佳平方逼近函数的解析解。随后计算了这些最佳平方逼近多项式的最大误差，并利用数值试验猜测他们和 n 的函数关系，并进行理论证明。最终，我们的数值试验和理论证明表明，$f(x)=e^x$ 在勒让德多项式的平方逼近下，最大误差收敛速度为 $O((\frac{e}{2n})^n)$.

关键词:

1　计算最佳平方逼近的 m 次多项式

本次实验，我们需要利用勒让德正交多项式 $\varphi_n(x)=\frac{1}{n!2^n}\frac{d^n}{dx^n}(x^2-1)^n$，构造最佳逼近函数。

1.1　构造勒让德多项式

我们首先需要利用所学知识，根据勒让德多项式的递推关系 $(k+1)\varphi_{k+1}(x)=(2k+1)x\varphi_k(x)-k\varphi_{k-1}(x)$，构造出 $k+1$ 阶勒让德多项式的表达式。这样做的目的是为了避免求微分的运算过于繁琐，增加运算量，若使用递推公式，则只涉及到 $O(n)$ 次的乘法运算。

十次及以下勒让德多项式如下:

$$
\begin{aligned}
\varphi_0(x) &= 1 \\
\varphi_1(x) &= x \\
\varphi_2(x) &= \frac{3x^2}{2} - \frac{1}{2} \\
\varphi_3(x) &= \frac{x\left(5x^2-3\right)}{2} \\
\varphi_4(x) &= \frac{35x^4}{8} - \frac{15x^2}{4} + \frac{3}{8} \\
\varphi_5(x) &= \frac{x\left(63x^4-70x^2+15\right)}{8} \\
\varphi_6(x) &= \frac{231x^6}{16} - \frac{315x^4}{16} + \frac{105x^2}{16} - \frac{5}{16} \\
\varphi_7(x) &= \frac{x\left(429x^6-693x^4+315x^2-35\right)}{16} \\
\varphi_8(x) &= \frac{6435x^8}{128} - \frac{3003x^6}{32} + \frac{3465x^4}{64} - \frac{315x^2}{32} + \frac{35}{128} \\
\varphi_9(x) &= \frac{x\left(12155x^8-25740x^6+18018x^4-4620x^2+315\right)}{128} \\
\varphi_{10}(x) &= \frac{46189x^{10}}{256} - \frac{109395x^8}{256} + \frac{45045x^6}{128} - \frac{15015x^4}{128} + \frac{3465x^2}{256} - \frac{63}{256}
\end{aligned}
$$

为了避免分数转化为小数的误差，在求勒让德多项式的过程中，我们使用了准确的分数表示勒让德多项式的各项系数（以下称其为解析的）。

1.2 指数函数在勒让德多项式下的组合系数

首先计算 k 阶勒让德多项式的模长 $(\varphi_k(x), \varphi_k(x))$，$k$ 阶勒让德多项式与目标函数的内积 $(\varphi_k(x), f(x))$。两者的计算均使用软件上相应的程序包，该程序包可以对初等函数简单组合的积分求解析的原函数。

得到以上结果后，利用课件上的公式 $P_n(x) = \sum_{k=0}^{n} \frac{(f,\varphi_k)}{(\varphi_k,\varphi_k)}\varphi_k(x)$，得到 $P_n(x)$ 的解析表达式（此时未将目标函数与多项式内积中的 e 的次方项化为小数形式）。随后可计算误差如下

$$
\begin{aligned}
P_0(x) &= \sinh(1) \\
P_1(x) &= \frac{1}{2e}(6x - 1 + e^2) \\
P_2(x) &= \frac{1}{4e}(12x + 5\left(-7 + e^2\right)\left(3x^2 - 1\right) - 2 + 2e^2) \\
P_3(x) &= \frac{1}{4e}(-7x\left(-37 + 5e^2\right)\left(5x^2 - 3\right) + 12x + 5\left(-7 + e^2\right)\left(3x^2 - 1\right) - 2 + 2e^2) \\
P_4(x) &= \frac{5}{8e}(-8379x^4 + 1134x^4e^2 - 70x^3e^2 + 518x^3 - 966x^2e^2 + 7140x^2 - 306x + 42xe^2 - 705 + 96e^2) \\
P_5(x) &= \frac{3}{16e}(-75999x^5e^2 + 561561x^5 - 27930x^4 + 3780x^4e^2 - 622230x^3 + 84210x^3e^2 - 3220x^2e^2 \\
&\quad + 23800x^2 - 17955xe^2 + 132685x - 2350 + 320e^2) \\
P_6(x) &= \frac{7}{32e}(-11586003x^6 + 1567995x^6e^2 - 65142x^5e^2 + 481338x^5 - 2134935x^4e^2 + 15775155x^4 \\
&\quad - 533340x^3 + 72180x^3e^2 - 5245965x^2 + 709965x^2e^2 - 15390xe^2 + 113730x - 33665e^2 + 248765) \\
P_7(x) &= \frac{1}{32e}(-307876140x^7e^2 + 2274914070x^7 - 81102021x^6 + 10975965x^6e^2 - 3671491824x^5 \\
&\quad + 496882386x^5e^2 - 14944545x^4e^2 + 110426085x^4 - 225557640x^3e^2 + 1666658070x^3 \\
&\quad - 36721755x^2 + 4969755x^2e^2 - 184802940x + 25010370xe^2 - 235655e^2 + 1741355) \\
P_8(x) &= \frac{9}{256e}(-64784168735x^8 + 8767583825x^8e^2 - 273667680x^7e^2 + 2022145840x^7 - 16356400060x^6e^2 \\
&\quad + 120858357620x^6 - 3263548288x^5 + 441673232x^5e^2 - 69669409810x^4 + 9428729310x^4e^2 \\
&\quad - 200495680x^3e^2 + 1481473840x^3 - 1712312140x^2e^2 + 12652370500x^2 - 164269280x \\
&\quad + 22231440xe^2 - 350813575 + 47477465e^2) \\
P_9(x) &= \frac{5}{256e}(-568595781611x^9e^2 + 4201386127939x^9 - 116611503723x^8 + 15781650885x^8e^2 \\
&\quad - 8893413114300x^7 + 1203592582764x^7e^2 - 29441520108x^6e^2 + 217545043716x^6 - 842064617394x^5e^2 \\
&\quad + 6222062696850x^5 - 125404937658x^4 + 16971712758x^4e^2 - 1594240291644x^3 + 215756961420x^3e^2 \\
&\quad - 3082161852x^2e^2 + 22774266900x^2 - 14695291611xe^2 + 108584334243x - 631464435 + 85459437e^2) \\
P_{10}(x) &= \frac{11}{256e}(-76432856441487x^{10} + 10344062275092x^{10}e^2 - 258452628005x^9e^2 + 1909720967245x^9 \\
&\quad - 24491921384385x^8e^2 + 180972181079820x^8 - 4042460506500x^7 + 547087537620x^7e^2 \\
&\quad - 148980681084690x^6 + 20162342671380x^6e^2 - 382756644270x^5e^2 + 2828210316750x^5 \\
&\quad - 6717527311950x^4e^2 + 49636186154100x^4 - 724654678020x^3 + 98071346100x^3e^2 - 5723477491095x^2 \\
&\quad + 774588447360x^2e^2 - 6679678005xe^2 + 49356515565x - 14070053529e^2 + 103964414904)
\end{aligned}
$$

最后得到误差表达式 $f(x) - P_n(x)$.

2 误差收敛性估计

在 $[-1,1]$ 中均匀的取 201 个点代入上述误差求得误差 ϵ_n，并画出误差曲线。

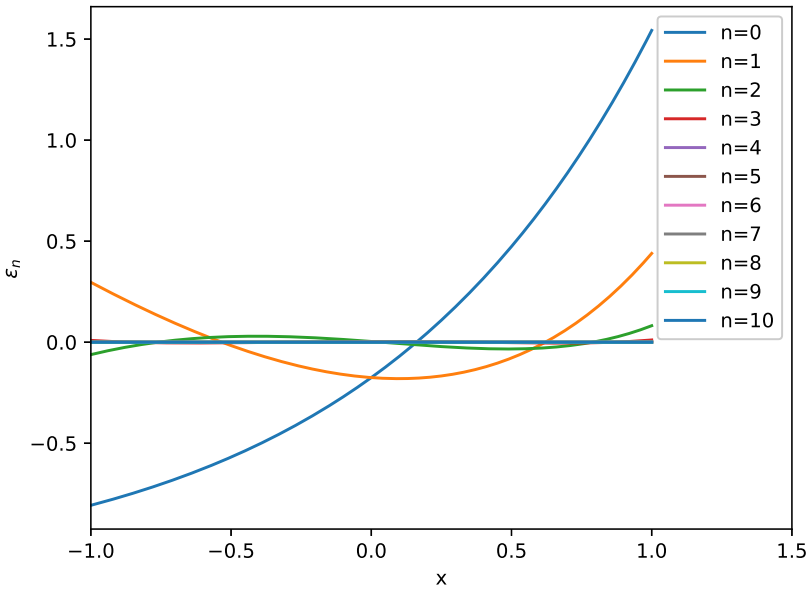

图 1: 最佳逼近多项式误差

计算每个逼近函数误差的最大值得到 $\epsilon_n = \max_{x\in[-1,1]} |\epsilon_n(x)|$，并画出图像

n=0	n=1	n=2	n=3
$\epsilon_0 = 1.54308$	$\epsilon_1 = 0.439442$	$\epsilon_2 = 0.0816280$	$\epsilon_3 = 0.0111723$
n=4	n=5	n=6	n=7
$\epsilon_4 = 0.00120720$	$\epsilon_5 = 0.000107613$	$\epsilon_6 = 8.15837 \cdot 10^{-6}$	$\epsilon_7 = 5.37829 \cdot 10^{-7}$
n=8	n=9	n=10	
$\epsilon_8 = 3.13566 \cdot 10^{-8}$	$\epsilon_9 = 1.63844 \cdot 10^{-9}$	$\epsilon_{10} = 7.75540 \cdot 10^{-11}$	

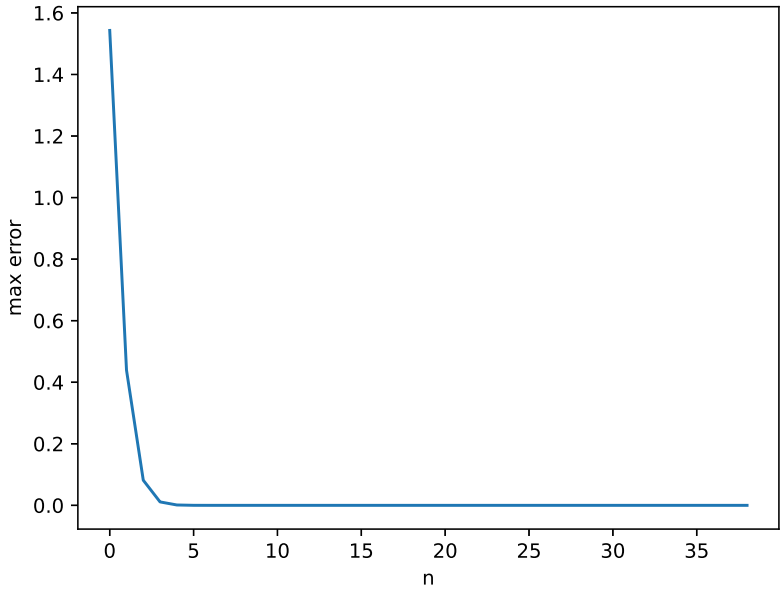

图 2: 最佳逼近多项式最大误差

通过观察勒让德多项式逼近结果的误差与多项式空间的维数 n 的关系，我们对收敛速度进行猜测。从图中可看出，误差以较快的速度衰减到 0。

这里我们采用最小二乘法 (线性回归) 来拟合误差:

一般模型: $\epsilon \sim \beta_0 + \beta_1 f_1(n) + \beta_2 f_2(n) + \cdots + \beta_k f_k(n)$

设 $f_i(k), (k = 1, 2, \cdots, n)$ 为向量 $x_i = (\xi_i^1, \xi_i^2, \cdots, \xi_i^n)^T$，$\epsilon$ 的样本点为 $Y = (y_1, y_2, \cdots, y_n)^T$，

$X = (1_n, x_1, x_2, \cdots, x_k)$，$\beta = (\beta_0, \beta_1, \cdots, \beta_k)^T$，

$\therefore \quad Y = X\beta + u$，$\beta = (X^T X)^{-1} X^T Y$. 其中

$$X = \begin{bmatrix} 1 & f_1(0) & f_2(0) & \cdots & f_k(0) \\ 1 & f_1(1) & f_2(1) & \cdots & f_k(1) \\ 1 & f_1(2) & f_2(2) & \cdots & f_k(2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & f_1(n) & f_2(n) & \cdots & f_k(n) \end{bmatrix}$$

一般情况下 $k < n$，当 $\{f_1, f_2, \cdots, f_n\}$ 不相关时，X 为列满秩，故 $X^T X$ 为 $k+1$ 阶可逆矩阵. 因此系数估计 β 存在.

2.1 以指数速度收敛

即 $error \sim e^{-n}$，仅对 error 求对数得 $ln(error) \sim n$，

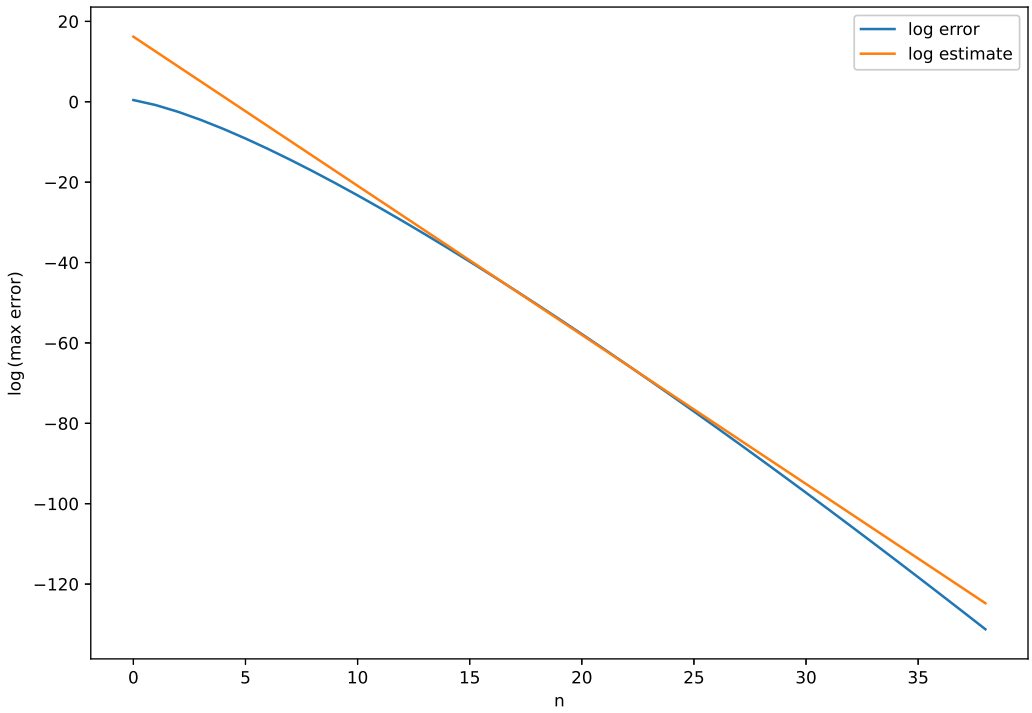

图 3: 对数误差线性回归

回归结果为 $\ln(\epsilon_n) = -3.70952n + 16.1966$。

2.2 以指数的多项式速度衰减

即 $error \sim e^{n^k}$ 对 error 求两次对数得 $\ln(\ln(error)) \sim \ln(n)$

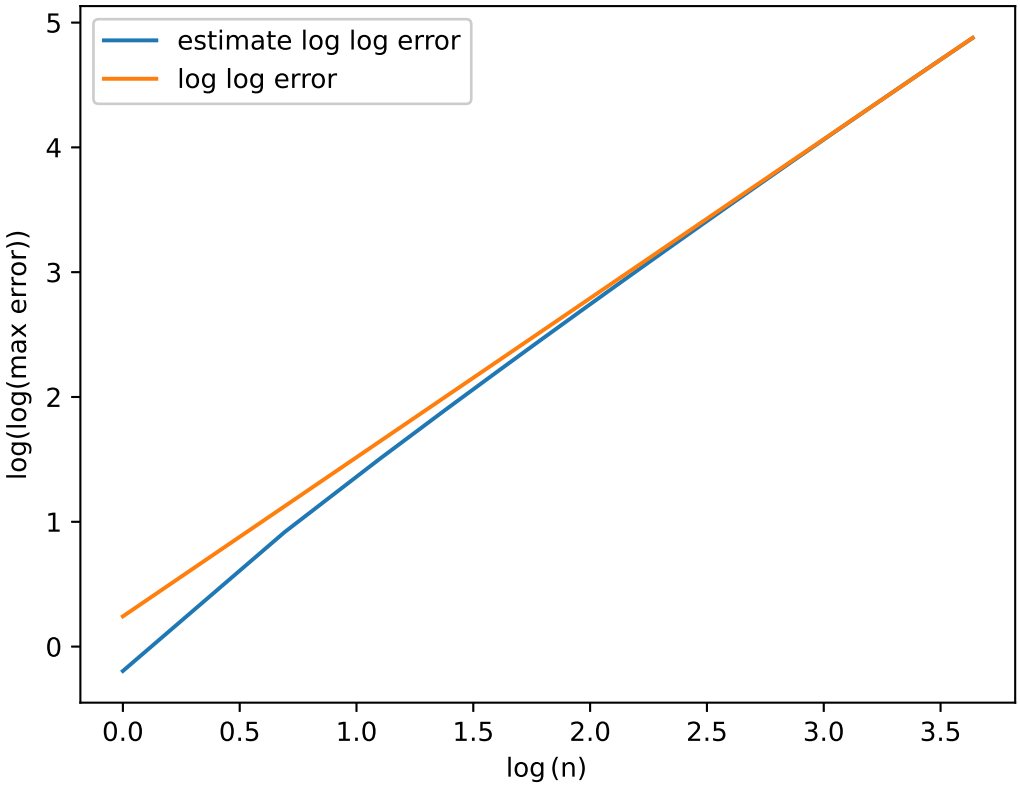

图 4: 对数误差关于 n 的次数

知 $\ln(\ln(\epsilon_n)) = 0.242014 + 1.27460\ln(n)$，故推测 $\ln(\epsilon_n)$ 和 $n^1.26$ 同阶，将 $\ln(\epsilon_n)$ 关于 $\beta_1 n^{1.27} + \beta_2 n + \beta_3$ 做最小二乘，得 $\ln(\epsilon_n) = -1.24295n^{1.27} - 0.193834n + 1.89277$，并作出图像如下

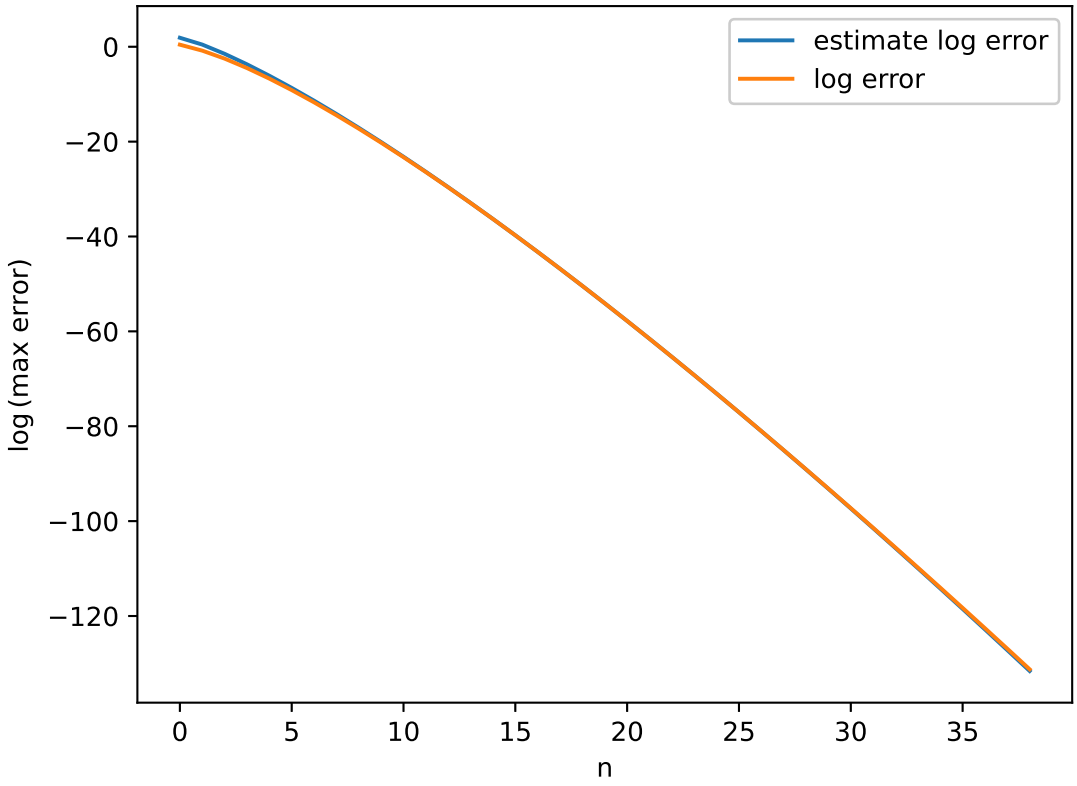

图 5: 对数误差关于 n 的小数次多项式回归

2.3 最终模型

易见，上述的几种模型都不能很好地拟合数据，当多项式空间维数足够高的时候会出现较大偏差，并且从图像的趋势可以看出实际收敛速度更快，因此，我们小组尝试采用先理论推导后数值验证的方法进行探索：

首先设函数 $f(x)$ 在 $[-1,1]$ 上连续，因此 $\exists M > 0, s.t. |f(x)| \leqslant M$. 记勒让德多项式空间为 $span\{\varphi_1(x), \varphi_2(x), \cdots, \varphi_n(x)\}$, $f(x)$ 的 n 阶勒让德多项式逼近函数为 $P_n(x) = \sum_{k=1}^{n} a_k \varphi_k(x)$, 其中

$$a_k = \frac{\int_{-1}^{1} f(x)\varphi_k(x)\,dx}{\int_{-1}^{1} \varphi_k^2(x)\,dx} = \frac{2k+1}{2} \int_{-1}^{1} f(x)\varphi_k(x)\,dx$$

记 $\epsilon_n = \max\limits_{-1 \leqslant x \leqslant 1} |f(x) - P_n(x)|$.

我们小组通过数值实验得出，在 $f(x) = e^x$ 时，ϵ_n 在边界上取得，即 $x = \pm 1$ 时误差取得最大值，而 $P_n(1) = \sum_{k=0}^{n} a_k$, $P_n(-1) = \sum_{k=0}^{n} (-1)^n a_k$ 因此 $\epsilon_n = |f(1) - P_n(1)| = |e - \sum_{k=0}^{n} a_k|$ 或

$\epsilon_n = |f(-1) - P_n(-1)| = |e^{-1} - \sum_{k=0}^{n}(-1)^n a_k|$, 故 $|\epsilon_n| \leqslant \sum_{k=0}^{n}|a_k|$, 因此 ϵ_n 的收敛性可以被 $\sum_{k=0}^{n}|a_k|$ 所控制.
下面我们给出 a_n 的估计:

引理 2.1 记 $g_n = (x^2-1)^n$, 则对 $\forall 1 \leqslant k \leqslant n-1$, $g_n^{(k)}(1) = g_n^{(k)}(-1) = 0$.

Proof:
$\because g_n(x) = (x-1)^n(x+1)^n$, -1 和 1 是 n 重根, 故在求 k 阶导数后为 $n-k$ 重根 ($1 \leqslant k \leqslant n-1$), 因此 $g_n^{(k)}(1) = g_n^{(k)}(-1) = 0$.

$$\because a_n = \frac{2n+1}{2}\int_{-1}^{1} f(x)\varphi_n(x)\,dx = \frac{2n+1}{n!2^{n+1}}\int_{-1}^{1} e^x \frac{d^n}{dx^n}[(x^2-1)^n]\,dx$$

利用分部积分:

$$\int_{-1}^{1} e^x \frac{d^n}{dx^n}[(x^2-1)^n]dx = e^x \frac{d^{n-1}}{dx^{n-1}}[(x^2-1)^n]|_{-1}^{1} - \int_{-1}^{1} e^x \frac{d^{n-1}}{dx^{n-1}}[(x^2-1)^n]dx$$

由引理 2.1 知, 右端第一项为零. 反复使用分部积分公式可得到:

$$a_k = (-1)^n \frac{2n+1}{n!2^{n+1}}\int_{-1}^{1} e^x (x^2-1)^n\,dx,$$

$$\because \quad \forall -1 \leqslant x \leqslant 1, \quad |(x^2-1)^n| \leqslant 1,$$

$$\therefore |a_k| = \frac{2n+1}{n!2^{n+1}}\int_{-1}^{1} e^x (x^2-1)^n\,dx| \leqslant \frac{2n+1}{n!2^{n+1}}\int_{-1}^{1} e\,dx \leqslant \frac{e(2n+1)}{n!2^n}.$$

由 stirling 公式, $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$ 带入上式可得:

$$|a_n| \sim e\sqrt{\frac{2n+1}{2\pi n}}(\frac{e}{2n})^n \quad (n \to \infty)$$

$\therefore \sum_{k=0}^{\infty} a_k$ 绝对收敛, 对于足够大的 n, 其余项

$$R_n = \sum_{k=n+1}^{\infty}|a_k| < \sum_{k=n+1}^{\infty} C_1(\frac{e}{2n})^k = C\frac{(\frac{e}{2n})^{n+1}}{1-\frac{e}{2n}} \sim C(\frac{e}{2n})^n. \quad (n \to \infty)$$

其中 C_1, C 为常数.
由上述论证知, ϵ_n 可由 $\sum_{k=0}^{n} a_k$ 的收敛性控制, 后者的收敛性由余项 R_n 表征, 而 R_n 以接近 $(\frac{e}{2n})^n$ 的速度趋于零, 因此我们小组采用如下形式的回归:

$$\ln(\epsilon_n) \sim \beta_0 + \beta_1 n + \beta_2 n\ln(n)$$

得到结果: $\ln(\epsilon_n) = -0.946895n\ln(n) + 0.0438809n - 1.94602$

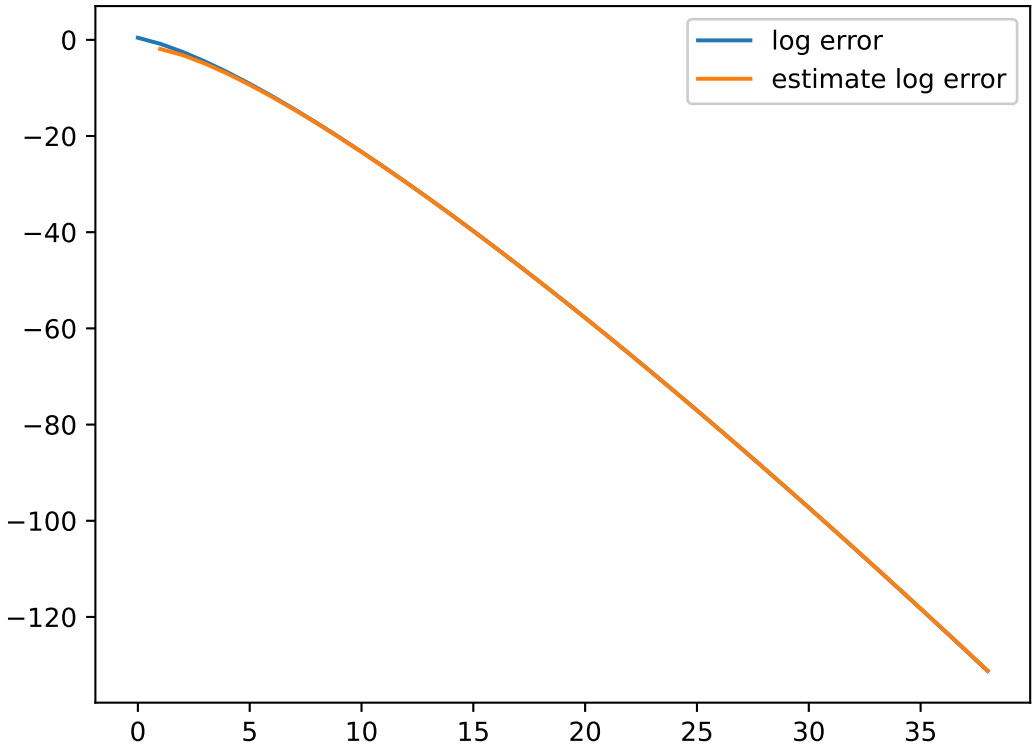

图 6: $\ln(\epsilon_n)$ 对 $n\ln(n)$ 和 n 回归

该结果 t 检验和 F 检验高度显著。

由上述的误差近似形式可知，其收敛速度主要受到 n^{-n} 的影响，而对于级数的一些放缩步骤也可能导致收敛速度的估计出现一些偏差，因此我们主要关注 β_2 的回归系数的显著性，经检验知，在 95% 置信度下不能拒绝 $\beta_2 = -1$ 的假设. 然而 β_1 和 $\frac{e}{2}$ 有稍微的偏差，说明在放缩过程中的确将收敛速度放慢了，然而该项并不是影响收敛速度的主要因素，因此 β_1 和预期值相差一个小常数属于可接受范围。综上所述，该回归结果与我们小组的理论推导基本吻合.

3 结论

1. L^∞ 范数意义下的收敛性：对于 $f(x) = e^x$，其 n 次勒让德逼近多项式 $P_n(x) = \sum_{k=0}^{n} a_k\varphi_k(x)$ 在 L^∞ 范数意义下以 $O((\frac{e}{n})^n)$ 的速度收敛到 $f(x)$.

2. L^2 范数意义下的收敛性：由于各阶勒让德正交多项式在-1 或 1 处取得最大最小值 (1 或-1)，故对 $\forall x \in [-1,1],\quad |P_n(x)| \leqslant \sum_{k=0}^{n} |a_k||\varphi_k(x)| \leqslant \sum_{k=0}^{n} |a_k|$，而上面证明了级数 $\sum_{k=0}^{\infty} |a_k|$ 收敛，故函数列 $P_n(x)$ 收敛. 在 $L^2(\mathbb{R})$ 空间中，采用 $\langle f(x), g(x)\rangle = \int_{-1}^{1} f(x)g(x)dx$ 作为内积，则 $P_n(x)$ 是 $f(x)$ 在有限维函数空间 $H_n = span\{\varphi_1(x), \varphi_2(x), \cdots, \varphi_n(x)\}$ 中的投影，且 $H_n \subseteq H_{n+1}, \forall n \in \mathbb{N}^*$. 故随着 n 的增大，误差项 $Rn(x) = f(x) - P_n(x)$ 的模长单调递降。又由于 $[-1,1]$ 是紧集，且多项式函数空间 $P[-1,1]$ 在连续函数空间 $C[-1,1]$ 中按无穷范数诱导的度量稠密，故也按 L^2 度量稠密. 注意到 $P[-1,1] = \bigcup_{n=1}^{\infty} H_n = \lim_{n\to\infty} H_n$，因此 $f(x)$ 在 H_n 中的投影按 L^2 度量随 $n \to \infty$ 收敛到 $f(x)$.