-1- (Taylor Series.) Let

$$f(x) = e^x \quad \text{and} \quad g(x) = \ln(x+1), \tag{1}$$

and let p_n and q_n be the Taylor polynomials of degree n for f and g, respectively, about

$$x_0 = 0. \tag{2}$$

Plot the graphs of f, g, p_n and q_n, for some small values of n, and comment on your results. Discuss in particular how well f and g are approximated by their Taylor polynomials. Explain your observations in terms of a suitable expression for the error in the approximation.

$$f(x) = \sum_{i=0}^{\infty} f^{(i)}(a) \cdot \frac{(x-a)^i}{i!}$$

$$f(x) = e^x \Longrightarrow f^{(i)}(0) = 1 \quad \forall i \in \mathbb{N}$$

$$\underline{f(x) = e^x} = \sum_{i=0}^{\infty} \frac{x^i}{\underline{i!}}$$

$$g(x) = \ln(x+1) = 0 \quad \text{for } x=0$$

$$g'(x) = \frac{1}{x+1} = 1 = 1!$$

$$g''(x) = \frac{-1}{(x+1)^2} = -1 = -1!$$

$$g'''(x) = \frac{2}{(x+1)^3} = 2 = 2!$$

$$g^{(4)}(x) = \frac{-6}{(x+1)^4} = -6 = -3!$$

$$\vdots$$

$$g^{(n)}(x) = \frac{(-1)^{n-1}(n-1)!}{(x+1)^n} = (-1)^{n-1}(n-1)! \quad \text{for } x=0 \quad n>0$$

$$\Longrightarrow \underline{\sum_{n=1}^{\infty} (-1)^{n-1}(n-1)! \cdot \frac{x^n}{n!}}$$

Insert problem1graphs.py (g, f)?

Qualitively we can see the taylor polynomial, P_n, begins to better approximate e^x as n gets larger. However, the T.P. q_n, tends to be problematic for larger n. It seems that T.P. q_n doesn't approximate $g(x)$ very well for $x>1$, despite increasing values of n.

We have the error functions to be,

$$E_n^f(x) = \left| e^x - \sum_{k=0}^{n} \frac{x^k}{k!} \right| \quad \longleftarrow \text{Error b/w } f \text{ \& } P_n$$

$$E_n^g(x) = \left| \ln(x+1) - \sum_{k=1}^{n} (-1)^{k-1}(k-1)! \cdot \frac{x^k}{k!} \right| \quad \longleftarrow \text{Error b/w } g \text{ \& } q_n$$

Look at graph for error at x values less than 1 & larger than 1.

Insert Problem1errors.py

-2- (A "simple" program.) Write a program that reads n and the entries x_1, x_2, $\cdots$, x_n of a vector $x \in \mathbb{R}^n$ from standard input and prints

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

to standard output. Note: Later in the semester it will be instructive to look at production codes e.g., in `lapack`, and compare them with the (presumably) simple minded code generated for this assignment. **Mail me your code before the lecture on March 3.**

Done in Python

-3- (Some Iteration.) Consider the iteration

$$x_{n+1} = F(x_n) = \sin x_n, \quad x_0 = 1 \tag{3}$$

(where of course the angle is measured in radians). What does our theory tell us about convergence? Show that the iteration does converge! What is the limit? How fast does the iteration converge? **Carefully explain the effects of rounding errors.**

We have that the fixed point iteration $x_{n+1} = g(x_n)$ converges if we have, that

$$|g'(\alpha)| < 1$$

Where $g(\alpha) = \alpha$, and x_0 is suff. close to α.

In our case we have $g'(\alpha) = \cos(0) = 1$. Which on its own tells us nothing. So our convergence theory is not particularly helpful.

However, we can say that it indeed converges. $\sin(x)$ is monotonically decreasing and bounded below by 0. So it must converge. In fact since $\sin(x)$ is cont. then as $x_n \to 0$ as $n \to \infty$ we can simply say $\lim_{x \to 0^+} \sin(x) = \sin(0) = 0$. The limit is zero.

This converges even slower than linearly, since $\underbrace{g'(\alpha) \neq 0}_{=1}$. In fact this takes a very long time.

: Insert some code example (Cool little animation??)

The issues of rounding errors can have catastrophic effects on the ability for our sequence to converge. If we have machine that only computes to say the first 3 decimal points (i.e it rounds), then are sequence will never converge. The iteration is so slow to converge that 3 decimal points of information is much to small to maintain accuracy. In fact it becomes constant when running it on python:

Insert graphic running for 3 decimal

The iteration settles on ____ after ____ steps. More examples? Now of course for this simple example we know that $\sin(x_n) = x_{n+1}$ for $x_0 = 1$ converges to zero from our deeper analysis. If we simply went to iteration via the computer, then we may get misleading results.

-4- (Newton's Method.) Suppose f has a root of multiplicity $p > 1$ at $x = \alpha$, i.e.,

$$f(\alpha) = f'(\alpha) = \ldots = f^{(p-1)}(\alpha) = 0, \quad f^{(p)}(\alpha) \neq 0. \qquad (4)$$

a. Show that Newton's method applied to $f(x) = 0$ converges linearly to α.

b. Show that this modification of Newton's Method:

$$x_{k+1} = x_k - p\frac{f(x_k)}{f'(x_k)} \qquad (5)$$

converges quadratically to α. **Hint:** You probably are thinking of using the Rule of L'Hôpital, but the problem is much easier if you think of f as being defined by $f(x) = (x-\alpha)^p F(x)$ where $F(\alpha) \neq 0$.

a) $$g = x - \frac{f}{f'} = x - \frac{(x-\alpha)^p F(x)}{p(x-\alpha)^{p-1}F(x) + (x-\alpha)^p F'(x)} \iff g(\alpha) = \alpha$$

$$= x - \frac{(x-\alpha)F(x)}{pF(x) + (x-\alpha)F'(x)}$$

$$g' = 1 - \frac{(pF(x) + (x-\alpha)F'(x))(F(x) + (x-\alpha)F'(x)) - (x-\alpha)F(x) \cdot \text{Denom}'}{(pF(x) + (x-\alpha)F'(x))^2}$$

$$g'(\alpha) = 1 - \frac{p \cdot F(\alpha) \cdot F(\alpha)}{p^2 \cdot F(\alpha)^2} = 1 - \frac{1}{p} = \left|\frac{p-1}{p}\right| < 1 \quad \text{for all } p > 1$$

Absolute Value of derivative at α is less than 1 $\forall p > 1$. Converges linearly.

b). $$g(x) = x - p\frac{f}{f'}$$

$$g'(x) = 1 - p\left(\frac{f}{f'}\right)' \implies g'(\alpha) = 1 - p \cdot \frac{1}{p} = 0 \quad \text{from above}$$

$\longrightarrow$ Converges at least Quadratically...

... ??

-5- (Division without division.) Suppose you have a computer or calculator that has no built-in division. Come up with a fixed point iteration that converges to $1/r$ for any given non-zero number r, and that only uses addition, subtraction, and multiplication. **Hint:** Write down an equation satisfied by $1/r$, apply Newton's method to that equation, and then modify Newton's method so that it doesn't use division. Your resulting method should converge of order 2.

Let $f(x) = \frac{1}{x} - r$. Now f is satisfied by $x = \frac{1}{r}$ for $r \neq 0$.

We have,

$$f(x) = \frac{1}{x} - r$$

$$f'(x) = \frac{-1}{x^2}$$

$$f''(x) = \frac{2}{x^3}$$

Both f' & f'' are non-zero at $x = \frac{1}{r}$ meaning the Newton's method converges quadratically.

Perhaps

We now look at Newton's method:

$$g(x) = x - \frac{f}{f'} = x - \frac{\frac{1}{x} - r}{\frac{-1}{x^2}}$$

$$= x - (-x + rx^2)$$

$$g(x) = 2x - rx^2$$

Desmos

-6- (A cubically convergent method.) Consider the iteration

$$x_{k+1} = g(x_k) \quad \text{where} \quad g(x) = x - \frac{f(x)}{f'(x)} - \frac{1}{2}\frac{f^2(x)f''(x)}{\left(f'(x)\right)^3}.$$

(We assume f is sufficently differentiable, and $f'(\alpha) \neq 0$.) Suppose that $g(\alpha) = \alpha$. Show that

$$g'(\alpha) = g''(\alpha) = 0.$$

(Thus the fixed point method will converge of order at least 3 if we start sufficiently close to α.)

$$g = x - \frac{f}{f'} - \frac{1}{2}\frac{f^2 \cdot f''}{f'^3}$$

$$g' = 1 - \frac{f'^2 - ff''}{f'^2} - \frac{1}{2}\cdot\frac{f'^3(2\cdot f\cdot f'\cdot f'' + f^2\cdot f''') - 3\cdot f'^2\cdot f'' f^2 \cdot f''}{f'^6}$$

$$= \frac{ff''}{f'^2} - \frac{1}{2}\frac{2ff'^4f''}{f'^6} - \frac{1}{2}\frac{f^2f'^3f'''}{f'^6} + \frac{1}{2}\frac{3f^2f'^2f''^2}{f'^6}$$

$$= \frac{ff''}{f'^2} - \frac{ff''}{f'^2} - \frac{1}{2}\frac{f^2f'''}{f'^3} + \frac{1}{2}\frac{3f^2f''^2}{f'^4}$$

$$= \frac{1}{2}f^2\left(\frac{f''^2}{f'^4} - \frac{f'''}{f'^3}\right) = 0 \quad \text{at} \quad x = \alpha \quad \text{since } f(\alpha) = 0 \text{ and } f'(\alpha) \neq 0$$

$$g'' = f\cdot f'\left(\frac{f''^2}{f'^4} - \frac{f'''}{f'^3}\right) + \frac{1}{2}f^2\left(\frac{f''^2}{f'^4} - \frac{f'''}{f'^3}\right)'$$

$= 0$ since each term has a multiple of f which is 0 at α.

$\Longrightarrow$ Thus our fixed point iteration converges of order at least 3

-7- (Polynomial Interpolation.) Suppose you want to interpolate to the data (x_i, y_i), $i = 0, \ldots, n$ by a polynomial of degree n. Recall that the interpolating polynomial p can be written in its Lagrange form as

$$p(x) = \sum_{i=0}^{n} y_i L_i(x) \quad \text{where} \quad L_i(x) = \frac{\Pi_{i \neq j}(x - x_j)}{\Pi_{i \neq j}(x_i - x_j)}. \tag{6}$$

Show that

$$\sum_{i=0}^{n} x_i^j L_i(x) = x^j \quad \text{for} \quad j = 0, \ldots, n. \tag{7}$$

When initially looking at this problem I was stuck in the head space that I needed to write a formal problem, like some form of induction.

However, this can statement can be shown by the uniqueness of interpolation.

We first note that we have $n+1$ data points, so any polynomial interpolation up to degree n will be unique.

We have the points (x_i, x_i^d) for $i=0,1,2,\ldots,n$ and where d is fixed power that can take on the values $d=0,1,\ldots,n$. Also assuming all points are distinct

↳ This is crucial as it lets us interpolate uniquely

But by the constructing of these points, the polynomial x^d interpolates our data points. Since poly. Inter. is unique then it must be the case.

$$P(x) = \sum_{i=0}^{n} x_i^d L_i(x) = x^d \quad \text{for } d=0,1,\ldots,n$$

-8- (Uniqueness of the interpolating polynomial.) Assume you are given the data

$$\begin{array}{lcccc} x_i: & 1 & 2 & 4 & 8 \\ y_i: & 1 & 2 & 3 & 4 \end{array} \tag{8}$$

Construct the interpolating polynomial using

a. the power form obtained by solving the Vandermonde system,

b. the Lagrange form,

c. the Newton form,

and show that they all yield the same polynomial.

a) Vandermond

We have 4 nodes so we inter with poly. of deg 3

Correct word

have:

$$a + bx + cx^2 + dx^3 = y$$

$$\Longrightarrow \begin{aligned} a+b+c+d &= 1 \\ a+2b+4c+8d &= 2 \\ a+4b+16c+64d &= 3 \\ a+8b+64c+512d &= 4 \end{aligned} \iff \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 4 & 16 & 64 \\ 1 & 8 & 64 & 512 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

Solving the coresponding Augmented matrix reveals.

$$[a \quad b \quad c \quad d]^T = [-\tfrac{10}{21} \quad \tfrac{7}{4} \quad \tfrac{-7}{24} \quad \tfrac{1}{56}]^T$$

$$\Longrightarrow \underline{P(x) = -\tfrac{10}{21} + \tfrac{7}{4}x - \tfrac{7}{24}x^2 + \tfrac{1}{56}x^3}$$

b). Lagrange Form

$$L_0(x) = \frac{(x-2)(x-4)(x-8)}{(-1)(-3)(-7)} = \frac{x^3 - 14x^2 + 56x - 64}{-21}$$

$$L_1(x) = \frac{(x-1)(x-4)(x-8)}{(1)(-2)(-6)} = \frac{x^3 - 13x^2 + 44x - 32}{12}$$

$$L_2(x) = \frac{(x-1)(x-2)(x-8)}{(3)(2)(-4)} = \frac{x^3 - 11x^2 + 26x - 16}{-24}$$

$$L_3(x) = \frac{(x-1)(x-2)(x-4)}{(7)(6)(4)} = \frac{x^3 - 7x^2 + 14x - 8}{168}$$

$$\Longrightarrow P(x) = \sum_{i=0}^{3} y_i L_i(x) = \tfrac{1}{-21}(x^3 - 14x^2 + 56x - 64) + \tfrac{2}{12}(x^3 - 13x^2 + 44x - 32)$$

$$- \tfrac{3}{24}(x^3 - 11x^2 + 26x - 16) + \tfrac{4}{168}(x^3 - 7x^2 + 14x - 8)$$

$$= \tfrac{1}{168}\big(-8x^3 + 112x^2 - 448x + 512 + 28x^3 - 364x^2 + 1232x - 896$$

$$-21x^3 + 231x^2 - 546x + 336 + 4x^3 - 28x^2 + 56x - 32\big)$$

$$= \tfrac{1}{168}(3x^3 - 49x^2 + 294x - 80)$$

$$= \frac{1}{56}x^3 - \frac{7}{24}x^2 + \frac{7}{4}x - \frac{10}{21}$$

c). Newton Form

x	f	$F[x_i, x_j]$	$F[x_i, x_j, x_k]$	$F[x_i, x_j, x_k, x_l]$
1	1	$F[x_0,x_1] = \frac{f_1 - f_0}{x_1 - x_0} = 1$	$F[x_0,x_1,x_2] = \frac{F[x_1,x_2] - F[x_0,x_1]}{x_2 - x_0} = -\frac{1}{6}$	$F[x_0,x_1,x_2,x_3] = \frac{F[x_1,x_2,x_3] - F[x_0,x_1,x_2]}{x_3 - x_0} = \frac{1}{56}$
2	2	$F[x_1,x_2] = \frac{f_2 - f_1}{x_2 - x_1} = \frac{1}{2}$	$F[x_1,x_2,x_3] = \frac{F[x_2,x_3] - F[x_1,x_2]}{x_3 - x_1} = -\frac{1}{24}$	
4	3	$F[x_2,x_3] = \frac{f_3 - f_2}{x_3 - x_2} = \frac{1}{4}$		
8	4			

$$P(x) = f_0 + (x-x_0)F[x_0,x_1] + (x-x_0)(x-x_1)F[x_0,x_1,x_2] + (x-x_0)(x-x_1)(x-x_2)F[x_0,x_1,x_2,x_3]$$

$$= 1 + (x-1)\cdot 1 + (x-1)(x-2)\left(-\frac{1}{6}\right) + (x-1)(x-2)(x-4)\frac{1}{56}$$

$$= 1 + x - 1 - \frac{1}{6}(x^2 - 3x + 2) + \frac{1}{56}(x^3 - 7x^2 + 14x - 8)$$

$$= x - \frac{1}{6}x^2 + \frac{1}{2}x - \frac{1}{3} + \frac{1}{56}x^3 - \frac{1}{8}x^2 + \frac{1}{4}x - \frac{1}{7}$$

$$= \frac{1}{56}x^3 - \frac{7}{24}x^2 + \frac{7}{4}x - \frac{10}{21}$$

All of these interpolating methods yield the same polynomial.

-9- (The infamous Runge-Phenomenon.) It is not generally true that higher degree interpolation polynomials yield more accurate approximations. This is illustrated in this problem. Let

$$f(x) = \frac{1}{1+x^2} \quad \text{and} \quad x_j = -5 + jh, \quad j = 0, 1, \cdots, n, \quad h = \frac{10}{n}.$$

For

$$n = 1, 2, 3, \cdots, 20$$

plot the graph (in the interval [-5,5]) of the interpolant

$$p(x) = \sum_{i=0}^{n} \alpha_i x^i$$

defined by

$$p(x_i) = f(x_i), \quad i = 0, 1, \cdots, n.$$

Also list the approximate maximum error in the interval [-5,5] for each polynomial degree. To approximate the maximum error sample the error at 200 evenly spaced points (at least!) in the interval.

Program is done. Error increases in size as degree of inter. poly. increases.

-10- (Judicious interpolation.) Repeat the above except that you interpolate at the roots of the Chebycheff polynomials, i.e.,

$$x_i = 5\cos\frac{i\pi}{n}, \quad i = 0, 1, \ldots, n. \tag{9}$$

Note: The last three problems illustrate three approaches to the approximation of functions by simpler functions, specifically by polynomials. The first leads to the concept of Least Squares approximation and together the three problems suggest one reason why Least Squares Approximation is so popular: because it's so simple. They also illustrate the major point that a slight change in the mathematical problem can make a big difference in the difficulty of the problem. The emphasis is on what can do you by yourself, not what might be able to find in the literature. All three problems can be solved from first (Calculus) principles.

Also done! But it looks like it actually converges...? Very crazy.

-11- (Least Squares approximation of functions.) Find a linear function $l(x)$ such that

$$\int_0^1 (e^x - l(x))^2 \, dx = \min . \qquad (10)$$

If we let $l(x) = mx + b$ for $m, b \in \mathbb{R}$, then we can think of eqn (10) as a function of m & b.

$$F(m,b) = \int_0^1 (e^x - mx - b)^2 \, dx$$

We want to minimize $F(m,b)$, so we take partial derivatives.

$$\frac{\partial F}{\partial m} = \frac{\partial}{\partial m} \int_0^1 (e^x - mx - b)^2 \, dx$$

partial

$$= \int_0^1 \frac{\partial}{\partial m} (e^x - mx - b)^2 \, dx$$

two

$$= \int_0^1 2(e^x - mx - b)(-x) \, dx$$

$$= 2 \int_1^0 x(e^x - mx - b)$$

IBP-Tabular

	D	I
+	x	$e^x - mx - b$
−	1	$e^x - \frac{m}{2}x^2 - bx$
+	0	$e^x - \frac{m}{6}x^3 - \frac{b}{2}x$

$$= 2 \left[x e^x - \frac{m}{2}x^3 - bx^2 - e^x + \frac{m}{6}x^3 + \frac{b}{2}x \right]_1^0$$

$$= 2 \left[-1 - \left(e - \frac{m}{2} - b - e + \frac{m}{6} + \frac{b}{2} \right) \right]$$

$$= 2 \left[-1 - \left(-\frac{b}{2} - \frac{m}{3} \right) \right]$$

$$= 2 \left(\frac{b}{2} + \frac{m}{3} - 1 \right)$$

$$= b + \frac{2m}{3} - 2$$

$$\frac{\partial F}{\partial b} = \frac{\partial}{\partial b}\int_0^1 (e^x - mx - b)^2\,dx$$

$$= \int_0^1 \frac{\partial}{\partial b}(e^x - mx - b)^2\,dx$$

$$= 2\int_1^0 (e^x - mx - b)\,dx$$

$$= 2\left[e^x - \frac{m}{2}x^2 - bx\right]_1^0$$

$$= 2\left[1 - \left(e - \frac{m}{2} - b\right)\right]$$

$$= 2\left(\frac{m}{2} + b + 1 - e\right)$$

$$= m + 2b + 2(1-e)$$

To minimize we set partial derivatives equal to zero & solve for m & b.

$$m + 2b + 2(1-e) = 0$$
$$b + \frac{2m}{3} - 2 = 0$$

$$\Longrightarrow \begin{array}{c} \\ \times 3 \end{array}\left[\begin{array}{cc|c} 1 & 2 & 2(e-1) \\ \frac{2}{3} & 1 & 2 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 2 & 2(e-1) \\ 2 & 3 & 6 \end{array}\right] \times -2$$

$$\times 2 \left[\begin{array}{cc|c} 1 & 2 & 2(e-1) \\ 0 & -1 & 10-4e \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0 & 18-6e \\ 0 & 1 & 4e-10 \end{array}\right]$$

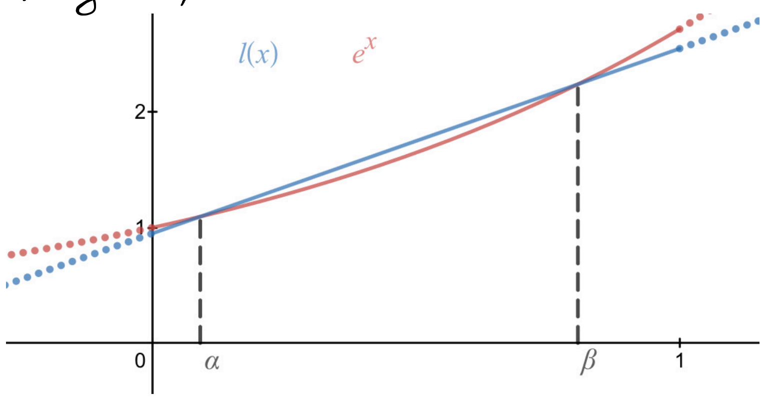
$$\Longrightarrow b = 4e - 10$$
$$m = 18 - 6e$$

Thus $l(x) = \underline{(18-6e)x + (4e-10)}$ minimizes (10)

-12- (An alternative approximation problem.) Find a linear function $l(x)$ such that

$$\int_0^1 |e^x - l(x)| \mathrm{d}x = \min . \tag{11}$$

We know that our linear function $l(x)$ must actually intersect e^x twice on the interval $[0, 1]$. This is just from a qualitative inspection of the graph. $l(x)$ must intersect e^x at some arbitrary values, say α & β.



So what we want to minimize the following

$$I = \int_0^1 |e^x - l(x)|\,dx = \int_0^\alpha (e^x - l(x))\,dx + \int_\alpha^\beta (l(x) - e^x)\,dx + \int_\beta^1 (e^x - l(x))\,dx$$

If we let $l(x) = mx + b$, the folling evaluates to,

$$I = \left[e^x - \frac{m}{2}x^2 - bx\right]_0^\alpha + \left[\frac{m}{2}x^2 + bx - e^x\right]_\alpha^\beta + \left[e^x - \frac{m}{2}x^2 - bx\right]_\beta^1$$

$$= \left(e^\alpha - \frac{\alpha^2}{2}m - \alpha b - 1\right) + \left(\frac{\beta^2}{2}m + \beta b - e^\beta\right) - \left(\frac{\alpha^2}{2}m + \alpha b - e^\alpha\right)$$

$$+ \left(e - \frac{1}{2}m - b\right) - \left(e^\beta - \frac{\beta^2}{2}m - \beta b\right)$$

$$= e^\alpha - \frac{\alpha^2}{2}m - \alpha b - 1 + \frac{\beta^2}{2}m + \beta b - e^\beta - \frac{\alpha^2}{2}m - \alpha b + e^\alpha$$
$$+ e - \frac{1}{2}m - b - e^\beta + \frac{\beta^2}{2}m + \beta b$$

$$F(m, b) = \left(\beta^2 - \alpha^2 - \frac{1}{2}\right)m + (2\beta - 2\alpha - 1)b + (2e^\alpha - 2e^\beta + e - 1)$$

$$\frac{\partial F}{\partial m} = \beta^2 - \alpha^2 - \frac{1}{2} = 0 \implies (\beta - \alpha)(\beta + \alpha) = \frac{1}{2}$$

$$\frac{\partial F}{\partial b} = 2\beta - 2\alpha - 1 = 0 \implies \beta - \alpha = \frac{1}{2}$$

$$\implies \underline{\alpha = \frac{1}{4}, \quad \beta = \frac{3}{4}}$$

$$\Longrightarrow \quad m = \frac{e^{\beta} - e^{\alpha}}{\beta - \alpha} = 2\left(e^{\frac{3}{4}} - e^{\frac{1}{4}}\right)$$

$$y = 2\left(e^{\frac{3}{4}} - e^{\frac{1}{4}}\right)x + b \quad \Longrightarrow \quad e^{\frac{1}{4}} = \tfrac{1}{2}e^{\frac{3}{4}} - \tfrac{1}{2}e^{\frac{1}{4}} + b$$

$$b = \tfrac{1}{2}\left(3e^{\frac{1}{4}} - e^{\frac{3}{4}}\right)$$

$\ell(x) = 2\left(e^{\frac{3}{4}} - e^{\frac{1}{4}}\right)x + \frac{1}{2}\left(3e^{\frac{1}{4}} - e^{\frac{3}{4}}\right)$ minimizes (11)

-13- (Another alternative approximation problem.) Find a linear function $l(x)$ such that

$$\max_{0\le x\le 1} |e^x - l(x)| = \min\,. \tag{12}$$

We know that our function $l(x)$ will intersect e^x twice, say α & β, however we wont focus on these values all that much. The idea that the max error will occur one of these intervals. Now since we wish to minimize the max error, the max error on each interval $[0,\alpha]$, $[\alpha,\beta]$, and $[\beta,1]$ must be identical. otherwise, we can always adjust the line to reduce the error. We can see what this looks like, qualitatively below where E denotes our minimized max error.

Desmos Image

We know that based the shape of e^x on the interval, the minimized max error will occur at $x=0$, $x=1$ & some $\tilde{x}$ on the interval $[\alpha,\beta]$. If we let $l(x)=mx+b$ then we have the following system.

$$e^0 - (m(0)+b) = E \qquad\qquad e^{\tilde{x}} = m \iff \ln(m)$$

$$e^{\tilde{x}} - (m\tilde{x}+b) = -E$$

$$e^1 - (m(1)+b) = E$$

We also note that $E'(\tilde{x}) = e^{\tilde{x}} - m = 0 \iff e^{\tilde{x}} = m \iff \tilde{x} = \ln(m)$. Solving the system we have,

$$\Rightarrow \begin{array}{l} 1-b=E \\ e-m-b=E \\ e^{\tilde{x}}-m\tilde{x}-b=-E \end{array} \qquad \begin{array}{l} \Rightarrow e-1-m=0 \iff \underline{m=e-1} \\ \\ \Rightarrow e+e^{\tilde{x}}-m-m\tilde{x}=2b \end{array}$$

$$\begin{aligned} b &= \tfrac{1}{2}(e-m\tilde{x}) \\ &= \tfrac{1}{2}(e-(e-1)\ln(e-1)) \\ &= \underline{\tfrac{1}{2}(e-e\ln(e-1)+\ln(e-1))} \end{aligned}$$

Thus $l(x) = (e-1)x + \frac{1}{2}(e-e\ln(e-1)+\ln(e-1))$ minimizes our function.