James Amidei
September 8, 2023

1 Problem 1

(Excercise 1.1-1) Find the second order Taylor polynomial for $f(x) = e^x \sin x$ about $x_0 = 0$.

a) Compute $P_2(0.4)$ to approximate $f(0.4)$. Use the remainder term $R_2(0.4)$ to find an upper bound for the error $|P_2(0.4) - f(0.4)|$. Compare the upper bound with the actual error.

b) **(MATH 5660 ONLY)** Compute $\int_0^1 P_2(x)dx$ to approximate $\int_0^1 f(x)dx$. Find an upper bound for the error using $\int_0^1 R_2(x)dx$, and compare it to the actual error.

a.) $f(x) = e^x \sin(x)$; $f(x_0) = e^0 \sin(0) = 0$

$f'(x) = e^x(\sin(x) + \cos(x))$; $f'(x_0) = e^0(\sin(0) + \cos(0))$

$f'(x_0) = (1)(0 + 1) = 1$

$f''(x) = 2e^x \cos(x)$; $f''(x_0) = 2e^0 \cos(0) = 2$

$P_2(x) = x + x^2$

$P_2(0.4) = (0.4) + (0.4)^2 = 0.4 + 0.16 = 0.56$

$f(0.4) = e^{(0.4)} \sin(0.4) \approx 0.58$

```
>>> np.e**(0.4)*np.sin(0.4)
0.5809439007705672
>>>
```

$$|P_2(0.4) - f(0.4)| = \left| f^{(3)}(\xi) \frac{(0.4 - 0)^3}{3!} \right|$$

$$= \left| 2e^{\xi}(\cos(\xi) - \sin(\xi)) \left(\frac{(0.4)^3}{3!} \right) \right|$$

$$= \left| \frac{8}{375} e^{\xi}(\cos(\xi) - \sin(\xi)) \right| \leq$$

$0 < \xi < 0.4$

Over the interval $(0, 0.4)$

$-\sin(\xi)$ starts at 0 and increases

$-\cos(\xi)$ starts at its maximum of 1 and decreases

and decreases

- e^{ξ} starts at 1 and increases

So, the maximum point of $f^{(3)}(\xi)$ over the interval $(0, 0.4)$ is at 0.

$$|P_2(0.4) - f(0.4)| \leq \left|\frac{8}{375} e^0 (\cos(0) - \sin(0))\right|$$

$$\leq \left|\frac{8}{375}(1)((1) - 0)\right|$$

$$|P_2(0.4) - f(0.4)| \leq \left|\frac{8}{375}\right| \approx 0.0213$$

$$|0.56 - 0.58| = 0.02 < 0.0213$$

2 Problem 2

(Exercise 1.3-1) Consider the following toy model for a normalized floating-point representation in base 2: $x = (-1)^s(1.a_2a_3)_2 \times 2^e$ where $-1 \leq e \leq 1$. Find all positive machine numbers (there are 12 of them) that can be represented in this model. Convert the numbers to base 10, and then carefully plot them on the number line, by hand, and comment on how the numbers are spaced.

$s = 0$, $0 \leq a_2 \leq 1$, $0 \leq a_3 \leq 1$, $-1 \leq e \leq 1$

8	4	2	1	1/2	1/4	1/8

$X_1 = (-1)^0(1.00)_2 \times 2^{-1} = 0.100 \Rightarrow 0.5$

$X_2 = (-1)^0(1.01)_2 \times 2^{-1} = 0.101 \Rightarrow 0.625$

$X_3 = (-1)^0(1.10)_2 \times 2^{-1} = 0.110 \Rightarrow 0.750$

$X_4 = (-1)^0(1.11)_2 \times 2^{-1} = 0.111 \Rightarrow 0.875$

$X_5 = (-1)^0(1.00)_2 \times 2^0 = 1.00 \Rightarrow 1.00$

$X_6 = (-1)^0(1.01)_2 \times 2^0 = 1.01 \Rightarrow 1.25$

$X_7 = (-1)^0(1.10)_2 \times 2^0 = 1.10 \Rightarrow 1.50$

$X_8 = (-1)^0(1.11)_2 \times 2^0 = 1.11 \Rightarrow 1.75$

$X_7 = (-1)^0 (1.10)_2 \times 2^0 = 1.10 \Rightarrow 1.50$

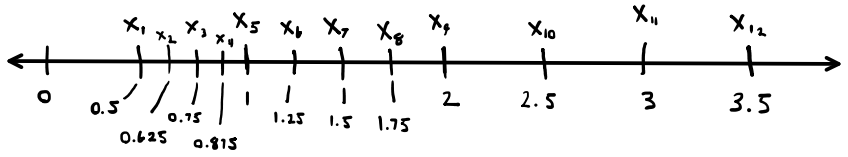
$X_8 = (-1)^0 (1.11)_2 \times 2^0 = 1.11 \Rightarrow 1.75$

$X_9 = (-1)^0 (1.00)_2 \times 2^1 = 10.0 \Rightarrow 2.00$

$X_{10} = (-1)^0 (1.01)_2 \times 2^1 = 10.1 \Rightarrow 2.50$

$X_{11} = (-1)^0 (1.10)_2 \times 2^1 = 11.0 \Rightarrow 3.00$

$X_{12} = (-1)^0 (1.11)_2 \times 2^1 = 11.1 \Rightarrow 3.50$



The spacing in larger between larger machine numbers and smaller between smaller machine numbers.

3 Problem 3

(Exercise 1.3-2) The x-intercept of the line passing through the points (x_1, y_1) and (x_2, y_2) can be computed using either one of the following formulas:

$$x = \frac{x_1 y_2 - x_2 y_1}{y_2 - y_1}$$

or,

$$x = x_1 - \frac{(x_2 - x_1) y_1}{y_2 - y_1}$$

with the assumption $y_1 \neq y_2$.

a) Show that the formulas are equivalent to each other.

b) Compute the x-intercept using each formula when $(x_1, y_1) = (1.02, 3.32)$ and $(x_2, y_2) = (1.31, 4.31)$. Use three-digit rounding arithmetic.

c) Use Python (or a calculator) to compute the x-intercept using the full-precision of the device (you can use either one of the formulas). Using this result, compute the relative and absolute errors of the answers you gave in part (b). Discuss which formula is better and why.

a.)

(1) $X = \frac{x_1 y_2 - x_2 y_1}{y_2 - y_1}$

(2) $X = x_1 - \frac{(x_2 - x_1) y_1}{y_2 - y_1}$

$$= \frac{x_1 (y_2 - y_1) - y_1 (x_2 - x_1)}{y_2 - y_1}$$

$$= \frac{x_1 y_2 - x_1 y_1 - x_2 y_1 + x_1 y_1}{y_2 - y_1}$$

$$= \frac{x_1y_2 - x_1y_1 - x_2y_1 + x_1y_1}{y_2 - y_1}$$

$$x = \frac{x_1y_2 - x_2y_1}{y_2 - y_1} = \frac{x_1y_2 - x_2y_1}{y_2 - y_1}$$

b.) $(1.02, 3.32)$ $(1.31, 4.31)$
(x_1, y_1) (x_2, y_2)

(1)

$$x = \frac{(1.02)(4.31) - (1.31)(3.32)}{4.31 - 3.32} = \frac{4.40 - 4.35}{0.99}$$

$$x = \frac{0.05}{0.99} = 0.0556$$

(2)

$$x = (1.02) - \frac{(1.31 - 1.02)(3.32)}{4.31 - 3.32} = 1.02 - \frac{0.963}{0.99}$$

$$x = 1.02 - 0.973 = 0.047$$

c.)

```
>>> ((1.02)*(4.31)-(1.31)*(3.32))/(4.31-3.32)
0.047474747474474719
```

(1) $x = \frac{x_1y_1 - x_2y_1}{y_2 - y_1}$

Abs

```
>>> 0.0556-0.047474747474474719
0.008125252525252806
```

Rel

```
>>> 0.008125252525252806/0.047474747474474719
0.1711489361702197
```

(2) $x = x_1 - \frac{(x_2 - x_1)y_1}{y_2 - y_1}$

Abs

```
>>> 0.047474747474474719-0.047
0.00047474747474719053
```

Rel

```
>>> 0.00047474747474719053/0.047474747474474719
0.009999999999994073
```

Looking at the absolute and relative errors for the rounded answers from part b),

for the rounded answers from part b), the second equation (2) seems to be much more accurate.

If I had to make some assumption as to why this is the case, I would say that the two multiplications in the numerator in equation (1) cause more to be lost in the rounding. By reducing the number of multiplications, we are able to reduce the amount of information lost on this front.

4 Problem 4

(Exercise 1.3-4) Polynomials can be evaluated in a nested form (also called Horner's method) that has two advantages: the nested form has significantly less computation, and it can reduce roundoff error. For

$$p(x) = a_0 + a_1x + a_2x^2 + ... + a_{n-1}x^{n-1} + a_nx^n$$

its nested form is

$$p(x) = a_0 + x(a_1 + x(a_2 + ... + x(a_{n-1} + x(a_n))...)).$$

Consider the polynomial $p(x) = x^2 + 1.1x - 2.8$.

a) Compute $p(3.5)$ using three-digit rounding, and three-digit chopping arithmetic. What are the absolute errors? (Note that the exact value of $p(3.5)$ is 13.3.)

b) Write $x^2 + 1.1x - 2.8$ in nested form by these simple steps:

$$x^2 + 1.1x - 2.8 = (x^2 + 1.1x) - 2.8 = (x + 1.1)x - 2.8.$$

Then compute $p(3.5)$ using three-digit rounding and chopping using the nested form. What are the absolute errors? Compare the errors with the ones you found in (a).

a.) $p(3.5) \approx (3.5)^2 + (1.1)(3.5) - 2.8$

$= 12.25 + 3.85 - 2.8$

Rounding $\begin{cases} = 12.3 + 3.85 - 2.8 = 13.35 \\ = 13.4 \end{cases}$

Abs

$|13.4 - 13.3| = 0.1$

Cutoff $= 12.2 + 3.85 - 2.8$

$= 13.25 = 13.2$

Abs

$|13.2 - 13.31| = 0.1$

b.) $p(x) = x^2 + 1.1x - 2.8$

$= x(x + 1.1) - 2.8$

$p(3.5) = (3.5)(3.5 + 1.1) - 2.8$

$= (3.5)(4.6) - 2.8$

$= 16.1 - 2.8 = 13.3$

Note: The result of chopping & rounding is the same in this case.

Abs

$|13.3 - 13.31| = 0$

The absolute error we get as a result of using the nested form is 0, which is obviously smaller than the absolute error(s) we get from using both chopping & rounding in the un-nested form.

5 Problem 5

(Exercise 1.3-5) Consider the polynomial written in standard form: $5x^4 + 3x^3 + 4x^2 + 7x - 5$.

a) Write the polynomial in its nested form.

b) (**MATH 5660 ONLY**) How many multiplications does the nested form require when we evaluate the polynomial at a real number? How many multiplications does the standard form require? Can you generalize your answer to any nth degree polynomial?

a.) $p(x) = 5x^4 + 3x^3 + 4x^2 + 7x - 5$

a.) $p(x) = 5x^4 + 3x^3 + 4x^2 + 7x - 5$

$= x(5x^3 + 3x^2 + 4x + 7) - 5$

$= x(x(5x^2 + 3x + 4) + 7) - 5$

$= x(x(x(5x + 3) + 4) + 7) - 5$

b.) The nested form requires 4 multiplications as opposed to the 10 in the standard form

m_n - # of multiplications for an n-degree polynomial

L - lowest number of multiplications

U - highest number

$n = 1$

$a_1 x + a_0$

$L = 1$

$U = 1$

$n = 2$

$a_2 x^2 + a_1 x + a_0$

$a_2 x x + a_1 x + a_0$

$L = 2$

$U = 3$

$n = 3$

$a_3 x^3 + a_2 x^2 + a_1 x + a_0$

$a_3 x x x + a_2 x x + a_1 x + a_0$

$L = 3$

$U = 6$

$n = 4$

$a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0$

$a_4 x x x x + a_3 x x x + a_2 x x + a_1 x + a_0$

$L = 4$

$U = 10$

$n = 5$

$a_5 x^5 + a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0$

$a_5 x x x x x + a_4 x x x x + a_3 x x x + a_2 x x + a_1 x + a_0$

$L = 5$

$U = 15$

$n = 6$

$a_6 x^6 + a_5 x^5 + a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0$

$a_6 x x x x x x + a_5 x x x x x + a_4 x x x x + a_3 x x x + a_2 x x + a_1 x + a_0$

$L = 6$

$L = 6$

$U = 21$

$$n \leq m_n \leq n + (n-1) + (n-2) + (n-3) + \ldots + 2 + 1$$

Lowest number of possible multiplications in equal to the degree of the polynominal.

$$m_n \leq n + (n-1) + (n-2) + \ldots + (n-n+2) + (n-n+1)$$

$$\leq n + n - 1 + n - 2 + \ldots + n - n + 1 + n - n + 1$$

$$\leq kn$$

n	kn	k
1	1	1
2	3	1.5
3	6	2
4	10	2.5
5	15	3
6	21	3.5

$$k = 0.5 + 0.5n$$

$$kn = (0.5 + 0.5n)n$$

$$kn = \frac{1}{2}(1+n)n$$

The number of multiplications required to solve an nth degree polynomial is

$$n \leq m_n \leq \frac{1}{2}n(n+1).$$

In other words, the number of

In other words, the number of multiplications required to solve an nth degree polynomial in its nested form is equal to the degree n of the polynomial and the number of multiplications required to solve an nth degree polynomial in its standard form is $\frac{1}{2}n(n+1)$.