APPM 4600 — HOMEWORK # 1 Solutions

General Information, applicable to all homework assignments

Homework will be assigned on a regular basis. Generally the problems will be based on the text, but occasionally they will require you to fill in details from class discussions, or further explore a topic outside of class. Problems assigned during any given week will be due the following Friday in class. The solutions should include the following:

- *Clear, brief restatement of the problem or question.*
- *Statement of important assumptions.*
- *Neat, detailed, step–by–step solution including sufficient comments to make the solution "read" well.*
- *When appropriate, a discussion of the accuracy (or lack thereof) of final and partial results. Is the answer consistent with the "physics of the problem"? What does the answer mean? Is it reasonable? How many digits can you trust (are significant)?*

You are encouraged to work together on the assignments, however, the major portion should be done on your own. In all cases your submission should demonstrate that you understand the problem and its solution.

1. Consider the polynomial
$p(x) = (x-2)^9 = x^9 - 18x^8 + 144x^7 - 672x^6 + 2016x^5 - 4032x^4 + 5376x^3 - 4608x^2 + 2304x - 512.$

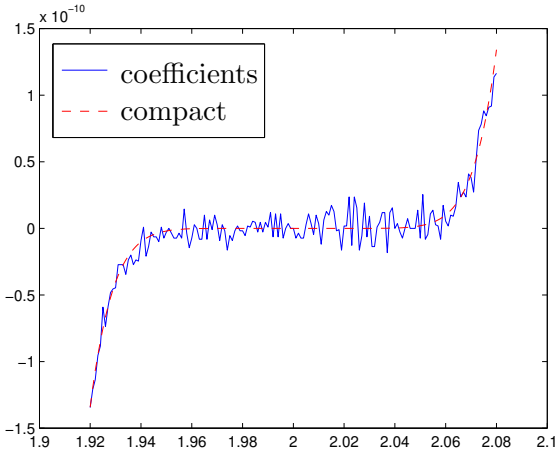
   i. Plot $p(x)$ for $x = 1.920, 1.921, 1.922, \ldots, 2.080$ (i.e. $x = [1.920 : 0.001 : 2.080]$;) evaluating p via its coefficients.

   ii. Produce the same plot again, now evaluating p via the expression $(x-2)^9$.

   iii. What is the difference? What is causing the discrepancy? Which plot is correct?

Soln:

i. & ii



iii. The plot using the compact form of the function is correct. This is clear because we know the function only has one root. The function plotted using the coefficients has many more than one root. The reason for lack of accuracy in the expanded form is that the evaluation of the many of the monomials is of the order 10^4. Thus important significant digits are loss in the resulting addition and subtraction.

2. How would you perform the following calculations to avoid cancellation? Justify your answers.

   i. Evaluate $\sqrt{x+1}-1$ for $x \simeq 0$.

   ii. Evaluate $\sin(2(x+a)) - \sin(2a)$ for $x \approx 0$.

   iii. Evaluate $\frac{1-\cos(x)}{\sin(x)}$ for $x \simeq 0$.

Soln:

i. **Soln:** For $x \simeq 0$, $\sqrt{x+1} \sim 1$. Thus

$$\text{fl}\left(\sqrt{\text{x}+1}-1\right) = 0.$$

However if we multiply by $1 = \dfrac{\sqrt{x+1}+1}{\sqrt{x+1}+1}$, we get

$$\sqrt{x+1}-1 = \frac{x}{\sqrt{x+1}+1}.$$

The numerator is accurate to machine precision and the denominator is accurate to machine precision. Thus the division can be executed to machine accuracy.

ii. The subtraction here leads to instability. We can avoid it by using a trig identity

$$\begin{aligned}\sin(2(x+a)) - \sin(2a) &= 2\cos\left(\frac{2x+4a}{2}\right)\sin\left(\frac{2x}{2}\right)\\ &= 2\cos(x+2a)\sin(x).\end{aligned}$$

This avoids the subtraction of nearly equal numbers unless $x \approx -2a$. In this case, $x \approx 0$ and $\cos(x+2a) \approx 1 + \mathcal{O}(x^2)$ (the error in cosine is much smaller than x) and we maintain precision.

iii. For this problem, we will multiply by 1.

$$\begin{aligned}\frac{1-\cos(x)}{\sin(x)} &= \frac{1-\cos(x)}{\sin(x)}\left(\frac{1+\cos(x)}{1+\cos(x)}\right)\\ &= \frac{1-\cos^2(x)}{\sin(x)(1+\cos(x))}\\ &= \frac{\sin^2(x)}{\sin(x)(1+\cos(x))}\\ &= \frac{\sin(x)}{1+\cos(x)}\end{aligned}$$

Now there is no subtraction of items that are the same size.

3. Find the second degree Taylor polynomial $P_2(x)$ for $f(x) = (1 + x + x^3)\cos(x)$ about $x_0 = 0$.

   (a) Use $P_2(0.5)$ to approximate $f(0.5)$. Find an upper bound for the error $|f(0.5) - P_2(0.5)|$ using the error formula and compare it to the actual error.

   (b) Find a bound for the error $|f(x) - P_2(x)|$ when $P_2(x)$ is used to approximate $f(x)$. This will be a function of x.

   (c) Approximate $\int_0^1 f(x)dx$ using $\int_0^1 P_2(x)dx$.

   (d) Estimate the error in the integral.

Soln:

(a) $P_2(0.5) = 1.375$. $f(0.5) \sim 1.4261$. So the error is $5.11x10^{-2}$.
From Calc II, we know

$$|f(x) - P_2(x)| \leq \frac{\max_{\eta\in(0,0.5)} |f^{(3)}(\eta)|}{3!}x^3.$$

The third derivative of f is $f^{(3)}(x) = (3 - 9x^2)\cos(x) + (1 - 17x + x^3)\sin(x)$. Now let's maximize it over all $x \in (0, 0.5)$

$$\begin{aligned}|(3 - 9x^2)\cos(x) + (1 - 17x + x^3)\sin(x)| &\leq |3 - 9x^2| + |1 - 17x + x^3| \\ &\leq 3 + |1 - 17(0.5) + (0.5)^3| = 10.375\end{aligned}$$

since $|3 - 9x^2|$ is largest at $x = 0$ and $|1 - 17x + x^3|$ is largest at $x = 0.5$. This means

$$|f(0.5) - P_2(0.5)| \leq \frac{10.375}{3!}(0.5)^3 = 0.216145833.$$

Notice that the error bound is larger than the actual error. This is to be expected since it is an upper bound.

(b)

$$|f(x) - P_2(x)| \leq \frac{\max_{\eta\in(0,\alpha)} |f^{(3)}(\eta)|}{3!}x^3$$

where $\alpha > 0$.

(c)

$$\int_0^1 P_2(x)dx = \int_0^1 (1 + x - 0.5x^2)dx = (x + \frac{x^2}{2} - \frac{x^3}{6})\Big|_0^1 = \frac{4}{3}$$

(d) The error can be approximated by integrating the error bound in the approximation. In other words, we need to evaluate

$$\frac{\max_{\eta\in(0,1)} |f^{(3)}(\eta)|}{3!}|\int_0^1 x^3dx \leq \frac{\max_{\eta\in(0,\alpha)} |f^{(3)}(\eta)|}{3!}\left(\frac{x^4}{4}\right)\Big|_0^1 = \frac{\max_{\eta\in(0,1)} |f^{(3)}(\eta)|}{3!}\frac{1}{4}$$

Now we need to create an upper bound on the $|f^{(3)}(x)|$ on the interval $(0, 1)$.

$$\max_{\eta\in(0,1)} |f^{(3)}(\eta)| \leq |3 - 9x^2| + |1 - 17x + x^3| \leq 6 + 15 = 21$$

. Thus

$$\frac{\max_{\eta\in(0,1)}|f^{(3)}(\eta)|}{3!}\int_0^1 x^3dx \leq \frac{21}{24} \sim 0.875.$$

4. Consider the quadratic equation $ax^2 + bx + c = 0$ with $a = 1, b = -56, c = 1$.

   (a) Assume you can calculate the square root with 3 correct decimals (e.g. $\sqrt{2} \approx 1.414 \pm \frac{1}{2}10^{-3}$) and compute the relative errors for the two roots to the quadratic when computed using the standard formula

$$r_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

   (b) A better approximation for the "bad" root can be found by manipulating $(x - r_1)(x - r_2) = 0$ so that r_1 and r_2 can be related to a, b, c. Find such relations (there are two) and see if either can be used to compute the "bad" root more accurately.

Soln:

- The following code rounds the square root to the third decimal place before computing each root using the quadratic formula. The result is that the root close to 0 only has 3 digits of accuracy.

  We find that

$$\sqrt{56^2 - 4 \cdot 1 \cdot 1} = 55.96427431853289,$$

  however, we can only keep the first three digits after the decimal, so we have

$$\sqrt{56^2 - 4 \cdot 1 \cdot 1} \approx 55.964.$$

  Continuing we find that the roots are given by

$$r_1 \approx 55.9820, \qquad r_2 \approx 0.0180$$

  Using e_1 as the exact value for the first root (computed using full precision at all steps) we find

$$\text{Relative Error} = \left| \frac{r_1 - e_1}{e_1} \right| \approx \frac{1.3716e - 04}{55.982} \approx 2.450E - 06.$$

  Using e_2 as the exact value for the second root (computed using full precision at all steps) we find

$$\text{Relative Error} = \left| \frac{r_2 - e_2}{e_2} \right| \approx \frac{1.3716e - 04}{0.01786} \approx 7.678E - 03$$

  and we can see the relative error is much larger for the second root.

- The two formulae are $r_2 = r_1 - \frac{b}{a}$ and $r_2 = \frac{c}{ar_1}$. The first formula has subtraction of nearly equal numbers. The second formula is stable. Using it we find the correct root to six digits of accuracy.

  This results in

$$r_2 = \frac{1}{r_1} = \frac{1}{55.982} \approx 0.01786288.$$

5. **Cancellation of terms**. Consider computing $y = x_1 - x_2$ with $\tilde{x}_1 = x_1 + \Delta x_1$ and $\tilde{x}_2 = x_2 + \Delta x_2$ being approximations to the exact values. If the operation $x_1 - x_2$ is carried out exactly we have $\tilde{y} = y + \underbrace{(\Delta x_1 - \Delta x_2)}_{\Delta y}$.

   Play with different values of x. One really small value (< 1) and one large value $> 10^5$.

   (a) Find upper bounds on the absolute error $|\Delta y|$ and the relative error $|\Delta y|/|y|$, when is the relative error large?

   (b) First manipulate $\cos(x + \delta) - \cos(x)$ into an expression without subtraction. Pick two values of x; say $x = \pi$ and $x = 10^6$. Then for each x, tabulate or plot the difference between your expression and $\cos(x + \delta) - \cos(x)$ for $\delta = 10^{-16}, 10^{-15}, \cdots, 10^{-2}, 10^{-1}, 10^{0}$ (note that you can use your logx command to uniformly distribute δ on the x-axis).

   (c) Taylor expansion yields $f(x + \delta) - f(x) = \delta f'(x) + \frac{\delta^2}{2!} f''(\xi), \quad \xi \in [x, x + \delta]$. Use this expression to create your own algorithm for approximating $\cos(x + \delta) - \cos(x)$. Explain why you chose the algorithm. Then compare the approximation from your algorithm with the techniques in part (b). Use the same values for x and δ.

Soln:

(a)

$$|\Delta y| = |\Delta x_1 - \Delta x_2| \leq |\Delta x_1| + |\Delta x_2|$$
$$\frac{|\Delta y|}{|y|} \leq \frac{|\Delta x_1| + |\Delta x_2|}{|x_1 - x_2|}$$

This means that the closer x_1 and x_2 are the more you see the impact of the approximations.

(b) Using trig identities, we can rewrite the expression as follows:

$$\cos(x + \delta) - \cos(x) = -2 \sin\left(\frac{2x + \delta}{2}\right) \sin\left(\frac{\delta}{2}\right).$$

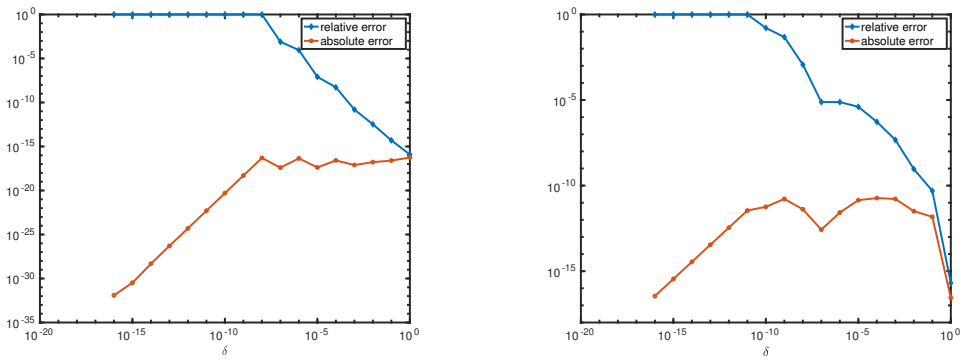
We know from part (a) that trouble will arise when $\cos(x + \delta)$ is close to $\cos(\delta)$. Also, we will need x to be large relative to δ because evaluating $\cos(x + \delta)$ will result in a loss of accuracy. The code for testing the two techniques is below.

Figure 1 illustrates the absolute and relative error for two different values of x; $x = \pi$ and $x = 10^6$. Note that when $x = \pi$, the absolute error is not bad for any values of δ but the relative error is only accurate when $\delta = 1$. When $x = 10^6$, the absolute error is not very good until δ is small. The relative error is still poor for almost all choices of δ.

(c) There are many choices of algorithms that you could make with the Taylor expansion.

option 1: If δ is really small, the second term does not contribute much so we can toss it. Then our algorithm is to approximate $f(x + \delta) - f(x)$ with $\delta f'(x)$.

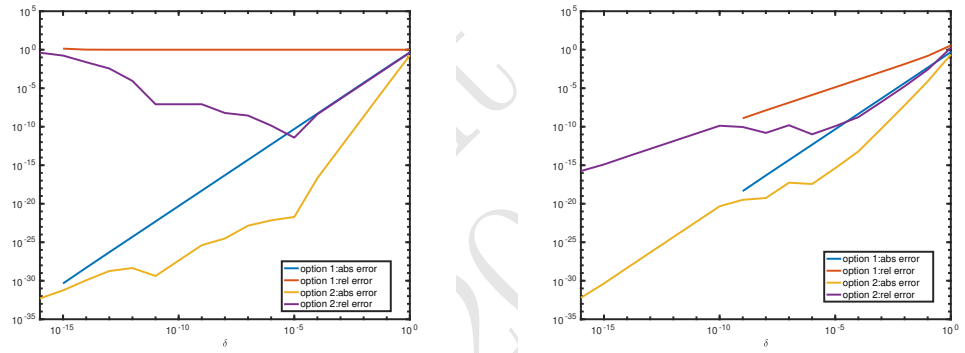
option 2: Since we know our function, $f''(x) = -\cos(x)$. On the interval $(x - \delta, x + \delta)$, $\cos(x)$ does not change much for small δ, thus we can chose to take an average value of $\cos(x)$ on the interval. i.e.



(a) (b)

Figure 1: Illustration of the relative and absolute error when (a) $x = \pi$ and (b)$x = 10^6$.

$$f(x+\delta) - f(x) \sim \delta f'(x) - \frac{\delta^2}{4}\left(\cos(x+\delta) + \cos(x-\delta)\right)$$



(a) (b)

Figure 2: Illustration of the relative and absolute error for the two different algorithms when (a) $x = \pi$ and (b)$x = 10^6$.

Figure 2 reports on the performance of the two algorithms. For large x, both algorithms are not accurate in terms of relative error. Option 1 is not very accurate but option 2 does a good job as $\delta \to 0$.