Numerical Analysis (MATH-411-01)
Homework #1: Due Thursday, 9/15

Please show all your work clearly and list the names of other students with whom you collaborated. If the assignment involves computer code, please turn in your code and figures as well.

1. **Break math**:

   Break math using a computer. To be a bit more specific, demonstrate a numerical calculation using the computer language of your choice where the answer is demonstrably wrong. I'll want to see the code you used, preferably something brief and punchy, and then the result. For full credit, fix math again by demonstrating an alternate method of calculating the result that previously broke math but yields the correct answer.

2. **Hello World!**

   Write a program that greets the world. Have it read in a number and then write it back out. Then, have it write out the value of π, *using only inverse trig functions and multiplication/division.*

3. **Matrix inversion via libraries and research:**

   Write a code that can invert a given matrix, *using as much pre-existing code as possible.* Through legal use of numerical libraries/packages, see how little code you have to write yourself to do this task (we'll need it later in the course anyway). Make sure to find a good code to do this, with pivoting, not just a straight Gauss-Jordan elimination code, or, heaven forfend, Cramer's rule. Please submit your code, noting where you got it from.

4. **Binary and Floating point notation**

   Convert the following numbers to binary, using overbar notation for repeated decimals, and then give the 64-bit floating point representation of the number:

   $$a.)10.5; \quad b.)\frac{1}{3}; \quad c.)\frac{22}{7}$$

5. **Head-to-head**

   Consider the following methods to calculate $\sqrt[3]{2}$:

   - Bisection for $f(x) = x^3 - 2$ on $[0, 2]$
   - Secant method for $f(x) = x^3 - 2$ on $[0, 2]$
   - False position method for $f(x) = x^3 - 2$ on $[0, 2]$

- FPI for $g(x) = \frac{x}{2} + \frac{1}{x^2}$, Initial guess 1
- FPI for $g(x) = \frac{2x}{3} + \frac{2}{3x^2}$, Initial guess 1
- FPI for $g(x) = x - \alpha(x^3 - 2)$, Initial guess 1; try a couple different values of α to determine a useful one!

How many iterations of each do we need to reach 8 digits of accuracy? Show the output from each method. Please discuss the overall rates of convergence.

6. **More FPI**

   Derive three different forms for $g(x)$ for finding roots of the cubic $f(x) = 2x^3 - 8x - 1$, such that at least one form can stably find each of the three different real roots. Show the performance of each equation near all three roots, and comment on stability.

7. **Secants behaving badly**

   This is derived from an actual example I encountered once. Define the following two functions

$$\Omega_1(x) = \frac{k_1}{A_1x^2 + B_1x + C_1}$$
$$\Omega_2(x) = \frac{k_2}{A_2x^2 + B_2x + C_2}$$

   Consider the following values: $k_1 = 1,\ A_1 = 1.0,\ B_1 = 1.5,\ C_1 = -0.25,\ k_2 = 1.0,\ A_2 = 4.0,\ B_2 = -6.4,\ C_2 = 1.5$. Explain why finding the solution to $\Omega_1(x) = \Omega_2(x)$ on the domain $-1 \leq x \leq 1$ via the relation $\Omega_1 - \Omega_2 = 0$ using the secant method with initial points $x = \pm 1$ can fail rather miserably. Is there a safer way to solve this particular equation?

8. **Convergence, or the lack thereof**

   If we try to evaluate exponentials via the Taylor series, e.g.,

$$e^{3x} \equiv \sum_{n=0}^{\infty} \frac{(3x)^n}{n!}$$

   we expect convergence in the continuous regime. Discrete floating-point math works a bit differently.

   - Beginning from $x_0 = 2, -2$, and -12, calculate how many terms you need before the relative error is less than 10^{-9}. Note that to avoid overflows, the $n+1$st term in the series is $\frac{3x}{n+1}$ times the nth term. Stop at 100 terms, to avoid working forever.
   - If you let the program run until the $n+1$st partial sum is no more than 10^{-15} different than the previous one, with no other stopping criteria, how accurate are the answers using the same starting points as the previous part?

- Repeat the calculation for $x_0 = 20, -20$ as in the previous part. Comment on how accurate the summation is, and why this is so.

9. **Newton's method beyond the first dimension**

   Consider the following system of equations in the two unknowns V and W, and the known quantities $B,\ S,\ Q,\ D,\ N$:

   $$\begin{aligned} V(B+W)^2 - S\frac{B+2W}{W^2} &= Q \\ -\frac{B}{2}(1+V) + \frac{S}{2}W^{-2} - W + \frac{1}{2}\left[(1-V)W - D\sqrt{1-V}\right] &= N \end{aligned}$$

   Given values $B = 0.04,\ S = 0.0011111,\ Q = 0.083333,\ D = 1.29074,\ N = -1.774598$, please solve for V and W using:

   - A 2-D Newton-Raphson solver
   - A 1-D Newton-Raphson solver where you solve for $V(W)$ using the first of the two equations, and plug it into the second.

   Show the sequence of steps from an initial guess to the solution. Does it differ between the two methods, for the same initial guess?

   This is the actual method used to solve the conservative to primitive variable conversion for general relativistic magnetohydrodynamics for a polytropic fluid, as described by former RIT postdoc Scott Noble in a paper from several years ago.

10. **Physics example – Blackbody radiation**

    Planck's law states that the spectral energy density of a blackbody is given by

    $$u(\nu, T) = \frac{dU}{d\nu} = \frac{8\pi h\nu^3}{c^3}\frac{1}{e^{h\nu/kT} - 1}$$

    where h and k are Planck's and Boltzmann's constants, respectively, ν is the frequency of the radiation, T is the temperature, and u is the energy density per unit frequency.

    i.) For small values of frequency, specifically $h\nu \ll kT$, we can Taylor expand the exponential term in powers of the frequency ν. The first order expansion yields the Rayleigh-Jeans law, which you can look up if you'd like. Derive this famous expression.

    What is the next, higher-order polynomial correction term, i.e., the next one not included in the Rayleigh-Jeans law? You'll want to compute the second order Taylor polynomial for the exponential, and then make an approximation to move the correction factor to the numerator. Remember that for $x \ll 1$, $1/(1-x) \approx 1 + x$. Your final answer should be a quadratic in ν.

ii.) Since astronomy used to be performed in terms of wavelength $\lambda \equiv \frac{c}{\nu}$, where c is the (constant) speed of light, what is $u(\lambda, T) = \frac{dU}{d\lambda}$? Remember the chain rule!

iii.) Graph both $u(\nu, T)$ and $u(\lambda, T)$ [separate plots] with appropriately labeled axes for $T = 500, 1000, 5000K$. Try linear, semilog (log u vs. ν) and log-log (log u vs. log ν) to see which is the most informative!

iv.) If we define $T_0 = 1000K$ as a characteristic temperature, and

$$\begin{aligned}\nu_0 &\equiv \frac{kT_0}{h} = 2.083 \times 10^{13}\ \text{Hz} \\ \lambda_0 &\equiv \frac{c}{\nu_0} = 1.438 \times 10^{-5} m\end{aligned}$$

we can rewrite the relationships above in dimensionless form. Define $\hat{T} \equiv \frac{T}{T_0}$, $\hat{\nu} \equiv \frac{\nu}{\nu_0}$, and $\hat{\lambda} \equiv \frac{\lambda}{\lambda_0}$, and we find

$$u(\hat{\nu}, \hat{T}) = \frac{8\pi h \nu_0^3}{c^3} \frac{\hat{\nu}^3}{e^{\hat{\nu}/\hat{T}} - 1}$$

Finally, defining a variable $x \equiv \frac{\hat{\nu}}{\hat{T}} = \frac{h\nu}{kT}$, and we find

$$u(x; T) = \frac{8\pi h \nu_0^3}{c^3} \hat{T}^3 \frac{x^3}{e^x - 1}$$

where the parameter x contains all of the frequency dependence in the blackbody equation. Written this way, it is clear that for a given temperature T, or equivalently $\hat{T}$, there is a value of x that yields a maximum value for $u(x;T)$, which you can determine by setting $\frac{\partial u}{\partial x} = 0$, and then using a numerical root-finder. What is the proper value of x? Note which method you use, and show the steps.

v.) Astronomers typically need to classify the widths of frequency distributions, and one simple form is the "Full-width, half-maximum", of FWHM. If we define x_{max} to be the value of x that maximizes u, that you determined above, determine the values [plural] of x_h for which $u(x_h;T)/u(x_{max};T) = 0.5$.