Math 4610 – HW 6

Jake Daniels A02307117

November 7, 2022

1) Define relative and absolute errors and give examples where relative error is a better measure and examples where absolute error may be a better measure of differences.
Response: The following equations are the equations for absolute and relative error:

$$e_{abs} = |exact - approx|,\ e_{rel} = \frac{|exact - approx|}{|exact|}$$

Let $u = exact$ and $v = approximation$:

a) If $u, v = \mathcal{O}(1)$ then both e_{abs} and e_{rel} are valid

b) If $u, v << 1$ then use e_{abs}

c) If $u, v >> 1$ then use e_{rel}

2) Describe the difference between the concepts of accuracy, efficiency, and robustness in the development of algorithms for the approximation of solutions of mathematical problems.
Response: The definitions for accuracy, efficiency and robustness are as follows:

(a) *accuracy:* Decreasing the size of your error, the smaller the error the more accurate your approximation is.

(b) *efficiency:* This describes how our code should run as fast as possible by decreasing the amount of computations or iterations must take place such that we still successfully accomplish the task.

(c) *robustness:* The ability of code to apply to a wide variety of different problems and still accomplish its task.

3) Define the rounding unit (or the machine precision) and explain the importance of the rounding unit for computation.
Response: Machine precision is defined to be the upper bound of the relative error due to rounding in floating point arithmetic. This is important as this is what causes round off error. This may not be a big issue for a single calculation, but will begin to add up as the number of calculations increases.

4) What is a nonlinear equation? Compare this to linear equations.
Response: A nonlinear equations is one where:

a) The independent variable has an exponent not equal to 1

b) The independent variable has a non-constant coefficient

c) The independent variable is involved in a function such as sin, ln, etc.

Similarly a linear equation is one such that the opposite of the above are all true. Another difference of the two is that a linear combination of linear equations is always linear, but a linear combination of nonlinear equations is not necessarily linear.

5) Is the bisection (i) efficient, (ii) accurate, (iii) robust? What smoothness conditions on the function are needed for Bisection to work?
Response: The bisect method is one where we find a root by taking nested intervals that converge to the root. It is...

i. The conditions for this problem are simply that your function, f, must be continuous as well for your starting interval [a,b], $f(a)f(b) < 0$.

i. It is pretty efficient, since we are halving the intervals we are looking at, this method will converge linearly at a rate of 2^n.

ii. It is very accurate and the error is guaranteed to be bounded.

iii. It is also robust because as long as the above conditions are met it will always converge to a root.

6) Does the bisection method provide a robust platform for the development of algorithms for the solution of systems of nonlinear equations?
Response: No it does not. Let's consider a function, $f : \mathcal{R}^2 \to \mathcal{R}$. If we have two points (a, b) and (c, d) such that $f(a, b)f(c, d) < 0$, then we can't really do anything with this knowledge. Even if we were able to find a root, that

wouldn't tell us anything about the behavior of the function around that root, and we would only be finding a single root out of a possible infinite amount. As far as developing a robust method that will apply to lots of different problems, we will want to look towards Secant and Newton's Method rather than the Bisection Method.

7) What are basic conditions for fixed point iteration to converge when searching for the root of a nonlinear function of a single variable. How are these conditions related to the iteration function, $g(x)$, defined in terms of the original function, f, defined as the input of a root finding problem?
Response: For functional iteration we define our function g(x) as follows:

$$g(x) = x - f(x)$$

In some cases it is beneficial to multiply the f in the above function such that it doesn't change where the roots are but it allow it to converge where it wouldn't before. The main two basic conditions are as follows:

(a) $|g(x^*)| < 1$ where $f(x^*) = 0$, or functional iteration will not converge.

(b) We don't require our function f to be continuous, but in order for this method to work $f(x)$ must exist for all x within the interval we are looking at.

8) State two advantages and two disadvantages of Newton's method for finding roots of nonlinear functions.
Response: Two advantages are...

1. This method converges fastest of all the methods we have learned and converges quadratically.
2. This method only needs an initial guess and is quite easy to code

Two disadvantages are...

1. This method requires us to calculate a derivative, so if the user does not want to do that then this is not a great method.
2. In order for this method to actually converge to a root we need an initial guess that is sufficiently close to the root, as well as at the root, x^*, $f'(x^*) \neq 0$, and lastly $|f''(x)|$ (second derivative is bounded).

9) Why would a person use the Secant method in place of Newton's method?
Response: The main reason we would use the Secant method in place of Newton's method is when the user does not want to calculate the derivative. In the secant method we replace the derivative with its finite difference approximation. When we switch to the Secant method we lose the quickness of convergence as the secant method only converges super-linearly.

10) Distinguish between the terms data fitting, interpolation, and polynomial iteration.
Response: Here are the following definitions...

(a) *data fitting:* process of fitting models to data

(b) *interpolation:* type of data fitting, it is the process of approximating the value of a given function at a given set of discrete points

(c) *polynomial interpolation:* the interpolation of a given data set by the polynomial of the lowest possible degree that passes through the points of the data set

11) State one advantage and two disadvantages of using the monomial basis for polynomial interpolation. **Response:** The main advantage is...

1. It is easy to work with since there are not as many calculations required when comparing to something like Lagrange polynomials, these are also easy to take integrals and derivatives of.

However there are the following disadvantages...

1. When calculating this, we will have to work with a Vandermonde matrix which will become worse to work with as the size of the matrix increases.

2. Prone to error due to oscillations for higher degree polynomials. Consider if we had a tenth degree polynomial where all ten roots were within $[0, 1]$ then our polynomial would be oscillating very quickly over that integral as it has to cross the x-axis 10 times in that small interval.

12) Define Lagrange polynomials (the cardinal functions) and how are they used in the development of algorithms for numerical integration.
Response: Lagrange polynomials are defined as follows:

$$L_i = \begin{array}{c} L_i(x_i) = 1 \\ L_i(x_j) = 0, \text{ if } j \neq i \end{array}$$

This is equivalent to saying:

$$L_i = \prod_{k=0, k\neq i}^{n} \frac{x - x_k}{x_i - x_k}$$

This is used for numerical integration by rewriting the function we are taking the integral of as a polynomial which we can again rewrite as the sum of coefficients multiplied by Lagrange polynomials. This allows us to approximate an integral that we wouldn't have been able to calculate before hand.

13) We have bumped into errors in the approximating roots of functions, approximating derivatives using difference quotients, approximations of solutions of differential equations and approximations of definite integrals.

$$|error| \leq C\, h^p$$

Write a brief explanation of the formula in terms the increment, h, $the constant$,C, and how to compute these parameters. Use an example like Newton's method for finding roots of functions.
Response: We can define C and h as follows:

C: This is our variable that is dependant on the smoothness of the function. It is found by looking at the remainder of our Taylor series expansion for our function. Which based on how we define it is usually the $(n+1)^{st}$ derivative evaluated at some point in our interval over $(n+1)!$.

h: As $h \longrightarrow 0$ our error will be order p and will behave like $C\, h^p$.

Looking at Newton's Method we calculate these values by taking the Taylor series expansion of our root. This allows us to show that for Netwon's Method the error is going to be bounded by a constant C times $h = e_k$ where $p = 2$. Meaning its error is bounded by the error at the iteration before it.

14) Discuss the pros and cons of using the Trapezoid rule for approximating definite integrals.
Response: The pros are the following...

1. More accurate then the left and right Riemann Sums
2. It has a composite rule that allows us to do the calculation in only $n + 1$ calculations

The cons are as follows...

1. It is less accurate then Simpson's Rule
2. If $f'' > 0$ then the integral will be overestimated and it will be underestimated if $f'' < 0$

15) 15. Compare the explicit and implicit Euler methods for approximate solution of initial value problems. You can use the logistic equation to illustrate your explanations.
Response: The explicit Euler method uses tangent lines to approximate an ODE which leads to it over estimating convex functions and under estimating concave functions. Whereas, the implicit Euler method uses secant lines to approximate an ODE which leads to an over approximation if the function is concave and an under approximation if the function is convex. The main difference between these methods is where we evaluate the ODE. For example looking at the logistic equation $P' = \alpha P - \beta P^2$, for explicit Euler method we evaluate the right hand side at the t-value given in our initial condition, t_0 which allows us to solve for $P(t_0 + h)$ on the right hand side and gives us a recursion relation that we can iterate over to get an approximation. However, for the implicit Euler method we evaluate the right hand side at $t = t_0 + h$ which means we have to use our root finding methods, specifically Euler's method, to find $P(t_0 + h)$. The reason we use Implicit Euler method over Explicit occasionally is because it approximates ODE's whose solutions change rapidly in a defined interval.