UGA CSCI 4150/6150
Caleb Ashmore Adams

1 Generally Useful Maths

Trig Properties

$\sin^2 x + \cos^2 x = 1 \quad \sec x = \frac{1}{\cos x}$
$2 \sin x = \sin x \cos x$
$\tan x = \frac{\sin x}{\cos x} = \frac{1}{\cot x} \quad \csc x = \frac{1}{\sin x}$
$\frac{d}{dx} \sin x = \cos x \quad \frac{d}{dx} \cos x = -\sin x$
$\frac{d}{dx} \tan x = \sec^2 x \quad \frac{d}{dx} \cot x = -\csc^2 x$
$\frac{d}{dx} \arcsin x = \frac{1}{\sqrt{1-x^2}} \quad \frac{d}{dx} \arccos x = \frac{-1}{\sqrt{1-x^2}}$
$\frac{d}{dx} \arctan x = \frac{1}{1+x^2} \quad \frac{d}{dx} \sec x = \sec x \tan x$

Log & Exp Properties

$\log x^n = n \log x \quad \log(\frac{1}{x}) = -\log x$
$\log_a x = \frac{\log_b x}{\log_a x} \quad \frac{d}{dx} e^{ax} = ae^{ax}$
$x^0 = 1 \quad x^n \cdot x^m = x^{n+m} \quad x^{-n} = \frac{1}{x^n}$
$\log_a x^n = n \log_a x \quad \frac{e^{-nx}}{e^x} = e^{-(n+1)x}$
$\log_a(\frac{x}{y}) = \log_a x - \log_a y$
$\log_a(xy) = \log_a x + \log_a y$
$\frac{d}{dx} \ln x = \frac{1}{x} \quad \frac{d}{dx} a^{g(x)} = \ln(a) a^{g(x)} g'(x)$
$\frac{d}{dx} a^{g(x)} = \ln(a) a^{g(x)} g'(x) \quad \frac{d}{dx} b^x = b^x \ln x$
$\frac{d}{dx} e^{g(x)} = g'(x) e^{g(x)} \quad \frac{d}{dx} a^x = a^x \ln a$
$\frac{d}{dx} \log_a(g(x)) = \frac{g'(x)}{ln(a)g(x)}$

Other Derivative Rules

$\frac{d}{dx} f(g(x)) = f'(g(x))g(x)$
$\frac{d}{dx} f(x)/g(x) = \frac{(f'(x)g(x) - g'(x)f(x))}{g(x)^2}$

Useful Series

$r^0 + r^1 + r^2 + r^3 = \frac{r^n - 1}{r-1}$
for an alternating series the following will work to start:
$\sum_{n=0}^{\infty}(-1)^n$ or $\sum_{n=0}^{\infty}(-1)^{n+1}$

In Class Terminology

the relative error formula: $\frac{|x-\hat{x}|}{x}$
more generally, with $\hat{x}, \hat{y}$ being rounded terms we get relative error as:
$\frac{(x-y)-(\hat{x}-\hat{y})}{(x-y)} = relative \quad error$
these were represed strangely in class:
$x' = f(t,x) \quad x(2) = 1 \to t = 2, x = 1$
If $x'' = xx'$ then $x''' = xx'' + x'x'$
When adding small number, it was mentioned in class that a >= or <= is preferable to a == when checking for values in a loop.

2 Base Conversion

Decimal to Binary

For this simply find the place of the largest binary number that (of the form 2^n) that is within the number. Successivley subtract these numbers while keeping track of their place to generate the binary number.

Binary to Decimal

For this notice that each place in the decimal number has a corresponding power of 2. If the decimal number has a floating point then the power is negative counting from zero. This generates a sum of the form:
$2^n + \ \dots \ + 2^2 + 2^1 + 2^{-1} + 2^{-2} + \ \dots \ + 2^{-m}$
Where n is the most significant digit and m is the least. The 2^{-1} term is the beginning of the floating point numbers.

Binary to Octal

Simply follow the table: $000 \to 0$ $001 \to 1$ $002 \to 2$ $003 \to 3$ $004 \to 4$ $005 \to 5$ $006 \to 6$ $007 \to 7$

Binary to Hex

This identical to the Octal method, the Hex symbols range from 0 to F and binary from 0000 to 1111. Simply count up un binary and there is a simple conversion.

One & Two's Complement

The one's complement of a bitstring is, simply, the inverse of that bitstring. i.e. all 1ś become 0ś and vice versa. The two's complement of a bitstring is the one's complement $+1$ at the end, so that (sometimes) there is a cascade of digit flips that occur.

3 IEEE Floating Points

Definitions

s = signed bit, c = based exponent, F = fraction. The general form for this is $(-1)^s \cdot 2^{c-127} \cdot 1.F$, for both $|s| = 1$
For single precision: $|c| = 8$, $|F| = 23$
For double precision: $|c| = 11$, $|F| = 52$
Machine Numbers are numbers which can be represted perfectly (no error) in an IEEE floating point format.
$\epsilon_{single} = 2^{-23}$ and $\epsilon_{double} = 2^{-52}$, floating points have about 6 digits of accuracy because $2^{-23} \approx 1.19 \cdot 10^{-7}$ and double has about 15 digits of accuracy becase $2^{-52} \approx 2.22 \cdot 10^{-16}$

IEEE Format

recall the above formula:
$(-1)^s \cdot 2^{c-127} \cdot 1.F$
A number will have the form $D_n \dots \ D_1 D_0 . F_0 F_1 \ \dots F_m$, to start we need to shift the values left (normalize) so that the number is now of the form: $D_n . F_0 F_1 \ \dots F_{m+(n-1)} \cdot 10^{n-1}$. note that the c term is, by definition, solved from $2^{c-127} = 2^{n-1}$.

4 Loss of Significance

Loss of Precision Theorem

The general form of the theorem is as follows:
x and y are floating point numbers such that $x > y > 0$, the theorem states that given: $2^{-p} \le 1 - \frac{y}{x} \le 2^{-q}$
there are at most p and at least q digits lost in the subtraction $x - y$.

practically speaking, we view the equation as: $E(x) = f(x) - g(x)$. If we notice this approaches 0 we have a concern of loss of precision at that point. to find that point we typically view the max loss acceptable as 1, so we set the euqation to $\frac{g(x)}{f(x)} = \frac{1}{2}$. We find the $x = z$ values that cause the $\frac{1}{2}$ flip and use a Taylor method there and use the normal formula elsewhere. We're just avoiding the loss of precision as $x \to z$.

Rationalizing Numerators

In some cases we want to rationalize a numerator to avoid a loss of significance. The general form form for radicals in a demonitor is:
$\sqrt[k]{x^n + r} + c \cdot \frac{\sqrt[k]{x^n + r - c}}{\sqrt[k]{x^n + r - c}} = \frac{x^n + r - 2c}{\sqrt[k]{x^n + r - c}}$

Small Numbers

If a set of small numbers $\{s_0, s_1, \dots, s_i\}$ is each on the order of 10^{n+1} decimal places but a large large number l is on the order of 10^n decimal places, it is better to add $\sum_{k=0}^{i} s_n$ small numbers there are before adding an l large number.

5 Taylor, Maclaurin, & Euler

Taylor Series

The Taylor series is a sum of derivatives of increasing order that equate to a function. The formula for the Taylor series of $f(x)$ evaluated at a is:
$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3$

Alternating Series Theorem

If $a_x \ge a_2 \ge \dots \ge a_n \ge 0$ for all n and $lim_{n \to \infty} a_n = 0$ then the alternating series $a_1 - a_2 + a_3 - a_4 + \dots$ so,
$S = lim_{n \to \infty} S_n = \sum_{k=1}^{\infty} (-1)^{k-1} a_k = lim_{n \to \infty} \sum_{k=1}^{\infty} (-1)^{k-1} a_k$
Here S is the sum and S_n is a partial sum. we note that, for all n, $|S - S_n| \le a_{n+1}$

Taylor's Method for ODEs

This method takes advantage of the previously mentioned series. here this is some step size h that we take from some $f(x)$ value. This is the Initial Value Problem (IVP).
$f(x+h) = f(x) + f'(x)h + \frac{1}{2!} f''(x)h^2 + \frac{1}{3!} f'''(x)h^3 + \quad \dots$

Maclaurin Series

The Maclaurin series is just the Taylor series at the special case where $x = 0$. This gives the following:
$f(x) = f(0) + f'(0) + \frac{x^2}{2!} f'''(0) + \frac{x^3}{3!} f'''(0) + \frac{x^4}{4!} f''''(0) + \quad \dots = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$

Euler's Method for ODEs

This method is just a Taylor series of order 1 with the same step term h, though many steps can be taken:
$f(x+h) = f(x) + f'(x)h$

Modified Euler's Method

This is the equation:
$x_{n+1} = x_n + h \frac{f(t_n, x_n) + f(t_{n+1}, x_{n+1})}{2}$
$= x_n = \frac{h}{2}[x'_n + x'_{n+1}]$

Error Terms

We note that Taylor's theorem in terms of $x + h$ is:
$f(x+h) = \sum_{k=0}^{n} \frac{f^k(x)}{k!} h^k + E_{n+1}$
Thus, error terms are of the form:
$E_{n+1} = \frac{f^{n+1}(\xi)}{(n+1)!} h^{n+1}$

It pays off to look at the term more specifically for the problem. A lot of times the error term takes the form $\frac{(n+1)^2}{(n+1)!}$ or $\frac{(n+1)^2}{(n+1)}$.
It is important to note that we only care about the $0.5 \cdot 10^n$ if our desired accuracy is to the nth decimal. Thus we set $E_{n+1} < 0.5 \cdot 10^n$ the $n+1$ portion of this is **very important!** To reiterate:
$E_{n+1} = \frac{f^{n+1}(\xi)}{(n+1)!} h^{n+1}$ or $E_{n+1} = \frac{x^{n+1}}{(n+1)!}$

First Derivative Formulas

For Taylor's Theorem, the forward difference formula is:
$f'(x) = \frac{f(x+h) - f(x)}{h} +$ error of $O(h)$
For Taylor's Theorem, the backwards difference formula is:
$f'(x) = \frac{f(x) - f(x-h)}{h} +$ error of $O(h)$
For Taylor's Theorem, the central difference formula is:
$f'(x) = \frac{f(x+h) - f(x-h)}{2h} +$ error of $O(h^2)$

6 Runge-Kutta Methods

RK4

This is the 4th order (RK4) Runge-Kutta method for the Initial Value Problem (IVP):
$x(t+h) = x(t) + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4)$
where the following are values of K_n:
$K_1 = hf(t, x)$
$K_2 = hf(t + \frac{1}{2}h, x + \frac{1}{2}K_1)$
$K_3 = hf(t + \frac{1}{2}h, x + \frac{1}{2}K_2)$
$K_4 = hf(t + h, x + K_3)$
first, the K_n values are calculated in succession. They the K_n values are filled into the first formula above.