Keith On . . .
Essential Mathematics

K.E. Schubert

Founder
Renaissance Research Labs

Professor
Department of Computer Science and Engineering
California State University, San Bernardino

Contents

List of Figures

Part I

Algebra

Chapter 1

Algebra

1.1 Functions

Functions map values of the domain (independent variable) to values of the range (dependent variable).

There are three ways to define functions

Symbolic The usual algebraic formulas we all think of. Typically these are good for "first principle" scientific equations $E = mc^2$, $F = ma$, $V = Ir$.

Visual Another one sometimes thought of is a graph. Typically we get this from lab experiments or other measurement processes, such as: EKG, or seismograph.

Numeric Table The final way is a table. This is the common way to express statistics (say temperature data) or discrete processes (say computer output).

Note that not every thing is a function. Essentially each independent variable must go to only one value of the dependent variable. Graphically this means you must be able to draw a vertical line anywhere and not hit more than one point on the graph. See below.

The domain and range are important characteristics of our functions. The domain is the values of the independent variable, for which the function is defined. The range is the values of the dependent variable for which the function is defined.

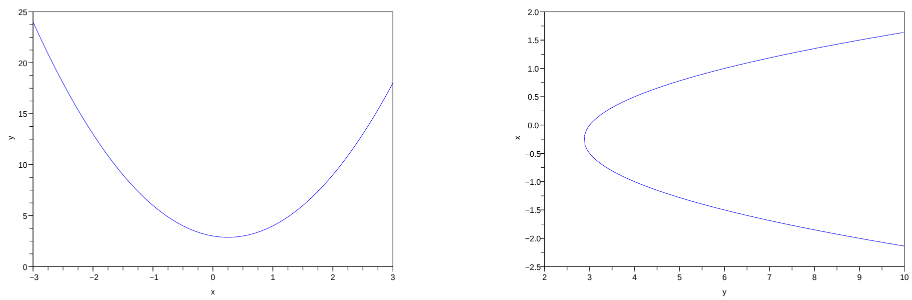
For example, consider the following and reference Figure 1.1.

(a) The domain of the function $y = x^2$ is all x and the range is $y \geqslant 0$.

(b) In contrast $y = \sqrt{x}$ is defined for the domain of $x \geqslant 0$ and has the range of $y \geqslant 0$. Notice square root is only defined from positive numbers to positive numbers[1].

(c) $y = -|x|$ is defined for the domain of all x, and the range of $y \leqslant 0$.

[1] We are right now assuming we do not have imaginary numbers. When we add them later the domain and range will be the complex numbers, which is one great reason to love complex numbers.



(a) $y = ax^2 + bx + c$ (b) $x = \frac{-b \pm \sqrt{b^2 + 4a(y-c)}}{2a}$

Figure 1.1: Vertical line test examples, (a) passes, (b) doesn't.

(d) $y = \frac{|x|+20}{x-10}$ is defined for the domain of all x except $x = 10$ (can't divide by zero), and the range of $|y| > 1$.

1.2 Linear

The slope, m, is the rate of change of the linear function. When you become familiar with calculus, you will know it as the derivative of the linear function.

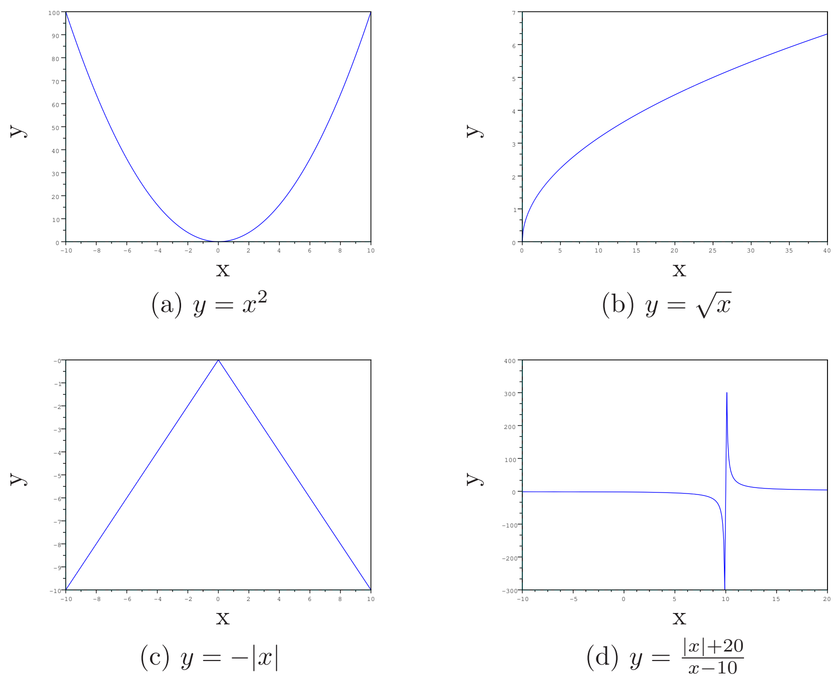
$$m = \frac{\Delta y}{\Delta x} \tag{1.1}$$

$$= \frac{y_1 - y_0}{x_1 - x_0} \tag{1.2}$$

Where (x_0, y_0) and (x_1, y_1) are two points on the line, recall that two points determine a line.

The intercept (also called the y-intercept), is the value of the dependent variable (height or y value if you like) when the independent variable is 0. In basic graphing you can think of it as the value of y when the graph crosses the y-axis. There is also an x-intercept, which can be similarly defined, but we usually call the x-intercept to be the "root" of the equation. It will turn out that finding the roots of an equation is essential to solving almost every interesting problem out there, and in fact we can reformulating any problem into a root finding one[2]. I will leave this off till the part of my notes on scientific computing[3], which is also called numerical analysis, as root finding is an essential technique in solving problems on computers. Both the x and y intercepts are just points on our linear function, albeit particularly useful ones.

[2]This is not always the best thing to do, but it is an option, and the more options the better.
[3]KONA



(a) $y = x^2$

(b) $y = \sqrt{x}$

(c) $y = -|x|$

(d) $y = \frac{|x|+20}{x-10}$

Figure 1.2: Four examples of algebraic functions to illustrate the domain and range.

The two basic forms we use are the slope-intercept form and the point slope form. In reality they are the same formula, the slope-intercept just using the point $(0, b)$, i.e. the intercept. Note that the root is $x = -\frac{b}{m}$.

Slope-Intercept Form

$$y = mx + b \tag{1.3}$$

Point-Slope Form

$$y - y_0 = m(x - x_0) \tag{1.4}$$

I have already shown how to calculate the slope given two points, you can then pick either point to solve for the rest of the equation. In practice, pick the one that makes your calculations easy. The most general form is the point slope, so you have either

$$\begin{aligned}
y - y_0 &= m(x - x_0) && (1.5)\\
y - y_0 &= \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) && (1.6)\\
y &= \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) + y_0 && (1.7)\\
y &= \frac{y_1 - y_0}{x_1 - x_0}x - \frac{y_1 - y_0}{x_1 - x_0}x_0 + y_0 && (1.8)\\
y &= \frac{y_1 - y_0}{x_1 - x_0}x + \left(y_0 - \frac{y_1 - y_0}{x_1 - x_0}x_0\right) && (1.9)\\
&\Rightarrow b = y_0 - \frac{y_1 - y_0}{x_1 - x_0}x_0 && (1.10)
\end{aligned}$$

or

$$\begin{aligned}
y - y_1 &= m(x - x_1) && (1.11)\\
y - y_1 &= \frac{y_1 - y_1}{x_1 - x_1}(x - x_1) && (1.12)\\
y &= \frac{y_1 - y_1}{x_1 - x_1}(x - x_1) + y_1 && (1.13)\\
y &= \frac{y_1 - y_0}{x_1 - x_0}x - \frac{y_1 - y_0}{x_1 - x_0}x_1 + y_1 && (1.14)\\
y &= \frac{y_1 - y_0}{x_1 - x_0}x + \left(y_1 - \frac{y_1 - y_0}{x_1 - x_0}x_1\right) && (1.15)\\
&\Rightarrow b = y_1 - \frac{y_1 - y_0}{x_1 - x_0}x_1 && (1.16)
\end{aligned}$$

Note that the two definitions of b are equivalent

$$\begin{aligned}
b &= y_0 - \frac{y_1 - y_0}{x_1 - x_0} x_0 && (1.17)\\
&= y_0 - \frac{y_1 - y_0}{x_1 - x_0}(x_0 - x_1 + x_1) && (1.18)\\
&= y_0 - \frac{y_1 - y_0}{x_1 - x_0}(x_0 - x_1) - \frac{y_1 - y_0}{x_1 - x_0} x_1 && (1.19)\\
&= y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x_1 - x_0) - \frac{y_1 - y_0}{x_1 - x_0} x_1 && (1.20)\\
&= y_0 + (y_1 - y_0) - \frac{y_1 - y_0}{x_1 - x_0} x_1 && (1.21)\\
&= y_1 - \frac{y_1 - y_0}{x_1 - x_0} x_1 && (1.22)
\end{aligned}$$

Example 1 *Consider an electric thermometer. It works by using a material, whose electrical properties change with temperature. In particular the resistance to electrical current is proportional to the temperature.*

$$T = mR + b \qquad (1.23)$$

Usually this is calibrated at the factory. Let's say for our thermometer, we were told that the following points were measured

Temperature	*Resistance*
$10°F$	50Ω
$120°F$	270Ω

Using these two points in our formula for the slope, we find

$$\begin{aligned}
m &= \frac{T_1 - T_0}{R_1 - R_0} && (1.24)\\
&= \frac{120 - 10}{270 - 50} && (1.25)\\
&= \frac{110}{220} && (1.26)\\
&= \frac{1}{2} && (1.27)
\end{aligned}$$

Then we put this in our formula, and pick a point to solve for b

$$\begin{aligned}
T &= .5R + b && (1.28)\\
10 &= .5 \cdot 50 + b && (1.29)\\
10 &= 25 + b && (1.30)\\
-15 &= b && (1.31)
\end{aligned}$$

So our formula is

$$T = R/2 - 15 \qquad (1.32)$$

We are not quite done yet. Now we want to use the resistance of the material to find the temperature. To measure the resistance we apply a voltage to the material, say 5 volts, and measure the current that flows. Then by the famous Ohm's law

$$V = iR \tag{1.33}$$
$$R = \frac{V}{i} \tag{1.34}$$
$$R = \frac{5}{i} \tag{1.35}$$

We can substitute this back in the original formula to find

$$T = R/2 - 15 \tag{1.36}$$
$$= \frac{5}{2i} - 15 \tag{1.37}$$
$$= \frac{2.5}{i} - 15 \tag{1.38}$$

While this is not linear anymore, both equations it came from where, which goes to show that things don't always stay simple. Let's say we measure 0.025 *amperes of current. What temperature is it?*

$$T = \frac{2.5}{i} - 15 \tag{1.39}$$
$$= \frac{2.5}{0.025} - 15 \tag{1.40}$$
$$= 100 - 15 \tag{1.41}$$
$$= 85^\circ F \tag{1.42}$$

1.3 Exponentials

Consider for a moment the decay of radioactive isotopes. At any given period of time a fraction of the radioisotopes decays. A common time period we use is to pick the amount of time needed for half of the radioactive material to decay, this is called the half-life of the material, and it is different for different materials. Let's denote the half-life by τ. Let's say we start with a quantity p_0, then one half life later we have

$$p_1 = \frac{1}{2}p_0. \tag{1.43}$$

Another half-life later we have

$$p_2 = \frac{1}{2}p_1 \tag{1.44}$$
$$= \frac{1}{2}\frac{1}{2}p_0 \tag{1.45}$$
$$= \left(\frac{1}{2}\right)^2 p_0. \tag{1.46}$$

We can extend this to n half-lives later.

$$p_n = \left(\frac{1}{2}\right)^n p_0 \tag{1.47}$$

We now have a simple expression of an exponential equation. To make our formula generic, we will use a instead of $\frac{1}{2}$. We call a our base. We now have $p_n = a^n p_0$. Let's see what we can do with this.

Let's imagine we go to time period $m+n$. We will also create a shorthand for multiplying a series of numbers in a pattern, $\prod_{i=1}^{4} a = a \cdot a \cdot a \cdot a$. We will consider $\frac{p_{m+n}}{p_0}$.

$$\frac{p_{n+m}}{p_0} = \frac{a^{n+m} p_0}{p_0} \tag{1.48}$$

$$= a^{n+m} \tag{1.49}$$

$$= \left(\prod_{i=1}^{n+m} a\right) \tag{1.50}$$

$$= \left(\prod_{i=1}^{n} a\right)\left(\prod_{i=1}^{m} a\right) \tag{1.51}$$

$$= (a^n)(a^m) \tag{1.52}$$

Thus

$$a^{n+m} = a^n a^m. \tag{1.53}$$

This is one of our basic properties of exponents, multiplication in the base means addition of the exponent.

Let's continue with our half-life example. Since the time elapsed, t is just the number of half-lives times the time of one half life, we have that $t = n\tau$. Solving for n, we find $n = \frac{t}{\tau}$, and

$$p(t) = \left(\frac{1}{2}\right)^{\frac{t}{\tau}} p(0). \tag{1.54}$$

I have switched from subscript notation to parenthesis notation because the relation is different, and the notational difference is used to try and underscore this. Our choice of $\frac{1}{2}$ was just convenient for the example used, but if we had picked a different problem, or defined our time period differently we would have a different number.

Sometimes we want to express our base in terms of the amount changed from the original, which we will call the rate, r. Since no change would happen if the base was 1 (1 multiplying itself any finite number of times is just 1 and multiplying 1 by any number is just the number you multiplied it by). The base is the ratio of the new amount to the old amount, but that is just 1 plus the rate, $a = \frac{p_{new}}{p_{old}} = 1 + r$. We can thus replace a by $1 + r$, noting

rate	meaning
$r < -2$	the magnitude increases but the sign changes each time period
$r = -2$	the sign of the amount changes each time period
$-2 < r < -1$	amount is decreasing in magnitude by the sign changes each time period
$r = -1$	amount goes to zero in one time period
$-1 < r < 0$	amount is decreasing
$r = 0$	amount is staying the same
$r > 0$	amount is increasing

Given the discussions above, we have two general forms of exponential equations

1. $P = P_0 a^n$,

2. $P = P_0(1 + r)^n$.

On another note, we could have a value of t which is not evenly divided by τ, and thus we could have a fractional part. When the exponent is a whole number, the computation is obvious, but what do we mean when the number has a fractional part? Or how about a negative one? Later we can even ask, what we mean when we have imaginary or complex exponents. Definitions of values:

n	$a^n = \prod_{i=1}^n a$	Multiply a by itself n times.
0	$a^0 = 1$	Not multiplying by a.
-1	$a^{-1} = \frac{1}{a}$	
$\frac{1}{n}$	$a^{(1/n)} =$ n^{th} root of a	

Properties

$$a^x a^t = a^{x+t} \tag{1.55}$$

$$\frac{a^x}{a^t} = a^{x-t} \tag{1.56}$$

$$(a^x)^t = a^{xt} \tag{1.57}$$

Can always switch between bases

$$P_0 e^{kt} = P_0(e^k)^t \tag{1.58}$$

$$= P_0 a^n \tag{1.59}$$

With $a = e^k$ thus $k = \ln(a)$

1.3.1 Exponential Properties

To keep our discussion as general as possible I will use a, b, and c to refer to any number or expression.

1. $a^b a^c = a^{b+c}$

2. $\frac{a^b}{a^c} = a^{b-c}$

3. $(a^b)^c = a^{bc}$

1.4 Logs

We want to look at logarithmic functions, which the book covers in sections 1.6 and 1.7. Rather than present these as two different subjects I will discuss them as one idea. We would like to develop the concepts and properties of logs from the properties of exponential functions. Lets recall three main properties of exponential functions.

1.4.1 Motivational Problem

Say we wanted to find the half-life of carbon-14. We measure a sample of c-14 and find it has 10000mg of c-14. We wait a year and measure again and find we have 9998.79mg. We recall that this is an exponential decay so we calculate that our decay factor is 0.999879. We thus have the equation:

$$c = 10000(0.999879)^t$$

Now we want to find when the amount of c-14 has dropped to 5000mg. The equation is thus

$$\begin{aligned} 5000 &= 10000(0.999879)^t \\ 0.5 &= (0.999879)^t \end{aligned}$$

How do I find the half-life? I could plot the right side and see when it was 1/2. I could make a table of values and find out when it is 1/2. Both of these would only approximate the solution. I want better.

1.4.2 Defining the Log

We will define the log to answer this problem.

$$\log_a(c) = t \Leftrightarrow a^t = c$$

To use our definition of the log we need to develop some properties. First we want to look at the graph of a^t and note from the definition of the log that I am just flipping the dependent and independent variables.

Now I look at where I can calculate log. From the definition of the log I can notice that if $c = a^x$ then

$$\log_a(a^x) = t \Leftrightarrow a^t = a^x$$

I notice that this is only true if $t = x$. I also notice this applies to all x since that is the domain of the exponential, and the range of the exponential is the domain of the log. I also notice that if I look at $t = \log_a(x)$ then

$$a^{\log_a(x)} = c \Leftrightarrow \log_a(c) = \log_a(x)$$

thus c must be x. Note that since the logarithm is only defined for positive x, this only works for $x > 0$. We have thus shown:

1. Definition: $\log_a(c) = t \Leftrightarrow a^t = c$
2. $\log_a(a^x) = x$
3. $a^{\log_a(x)} = x$

1.4.3 Multiplication Becomes Addition

We want to use the definition and the basic properties to find some more interesting results. First lets see what it means to take the log of a product. We note that for the product to be positive we need both terms to have the same sign. We want to right the product in terms of the logs of each term so we will need to further restrict the terms to be positive. The problem is thus

$$\log_a(bc) = ?$$

for $b > 0$ and $c > 0$. We will take advantage of one of our elementary properties.

$$\begin{aligned} b &= a^{\log_a b} \\ c &= a^{\log_a c} \end{aligned}$$

thus we have

$$\begin{aligned} \log_a(bc) &= \log_a\left(a^{\log_a b} a^{\log_a c}\right) \\ &= \log_a\left(a^{\log_a b + \log_a c}\right) \end{aligned}$$

But now we can use the other property of logs we have developed to show that

$$\begin{aligned} \log_a(bc) &= \log_a\left(a^{\log_a b + \log_a c}\right) \\ &= \log_a b + \log_a c \end{aligned}$$

And we are done.

1.4.4 Division Becomes Subtraction

This motivates us to look at subtraction with the same requirements

$$\log_a(\frac{b}{c}) = ?$$

for $b > 0$ and $c > 0$. We will again take advantage of one of our elementary properties.

$$\begin{aligned} b &= a^{\log_a b} \\ c &= a^{\log_a c} \end{aligned}$$

thus we have

$$\begin{aligned}\log_a(\frac{b}{c}) &= \log_a\left(\frac{a^{\log_a b}}{a^{\log_a c}}\right)\\ &= \log_a\left(a^{\log_a b - \log_a c}\right)\\ &= \log_a b - \log_a c\end{aligned}$$

And we are again done.

1.4.5 Exponentiation Becomes Multiplication

Now consider the following

$$\log_a(b^c) = ?$$

for $b > 0$. We will take advantage of one of our elementary properties.

$$b = a^{\log_a b}$$

thus we have

$$\begin{aligned}\log_a(b^c) &= \log_a\left(\left(a^{\log_a b}\right)^c\right)\\ &= \log_a\left(a^{c\log_a b}\right)\\ &= c\log_a b\end{aligned}$$

And we are done

1.4.6 Solving our problem

We can now use our log function to solve for the half-life of c-14

$$\begin{aligned}0.5 &= (0.999879)^t\\ \log(0.5) &= \log\left((0.999879)^t\right)\\ \log(0.5) &= t\log(0.999879)\\ t &= \frac{\log(0.5)}{\log(0.999879)}\\ &\approx 5728.1425\end{aligned}$$

1.4.7 Time constant

One final area of interest is the time constant of a system. We have often heard of half-lives and doubling times. One ratio used in physics and engineering is the time constant. For our purposes the time constant is the time for a quantity to decay to 1/e of its initial amount.

This turns out to have more physical significance in most systems. Consider for instance a resistor and capacitor system. The charge on the capacitor at any time is given by

$$q = q_0 e^{-\frac{t}{rc}}$$

thus the time constant is

$$\begin{aligned} q_0 e^{-1} &= q_0 e^{-\frac{t}{rc}} \\ -1 &= -\frac{t}{rc} \\ t &= rc \end{aligned}$$

1.4.8 Changing Bases

Theorem 1

$$\log_a c = \log_a b \log_b c$$

Proof:

By definition $\log_b c = x$ is the same as $c = b^x$. Take the $\log_a$ of both sides of $c = b^x$

$$\begin{aligned} \log_a c &= \log_a b^x \\ &= x \log_a b \end{aligned}$$

thus

$$x = \frac{\log_a c}{\log_a b}.$$

We already know that $\log_b c = x$ so

$$\begin{aligned} \log_b c &= \frac{\log_a c}{\log_a b} \\ \log_a c &= \log_a b \log_b c. \end{aligned}$$

♢ SDG ♢

Chapter 2

Trigonometry

Radians: 1 rad is the angle at the center of a unit circle that cuts off an arc length of 1 in the counter-clockwise direction.

Why use radians? There are many reasons which stem from the basic fact that radians are the natural measurement system. Here is one:

From the definition we know that for a unit circle 1 rad corresponds to an arc length of 1. We can see by bisecting the angle that we bisect the arc and 1/2 rad corresponds to an arc length of 1/2. This can be extended to see that for arc length (s), radius (r), and angle (θ) we have $s = r\theta$.

What is the circumference of a circle? Note that $2\pi(1)$ used in the definition tells us the internal angle of circle is 2π. Using the unit circle we can see the meanings of the two most important trig functions:

$$\sin(\theta) = y \qquad \& \qquad \cos(\theta) = x$$

Thus for any point on the circle we can express its location as $(\cos(\theta), \sin(\theta))$. Also we know the equation of a unit circle is $x^2 + y^2 = 1$ or $(\cos(\theta))^2 + (\sin(\theta))^2 = 1$. This gives us one of the most famous trig relations, and is related to one of the most famous mathematical relations of all time: Pythagorean Theorem.

Amplitude (A) half the height difference from the peak to the trough of a wave.

Period (p) The time to taken to complete one full wave cycle.

Phase () The amount the wave is shifted on the independent axis (angle).

This is written:

$$A\sin\left(\left(\frac{2\pi}{p}\right)(\theta - \phi)\right) = A\sin\left(\left(\frac{2\pi}{p}\right)\theta - \left(\frac{2\pi}{p}\right)\phi\right) \tag{2.1}$$

$$= A\sin(B\theta - C) \tag{2.2}$$

The third trig function we will look at is tan. Tangent is defined as the ratio of $sin(\theta)$ over $cos(\theta)$. Graphically mention inverse functions, to show why you can't find any angle. Give domain and range of inverse functions.

Chapter 3

Function Transformations

3.1 Old from New

The first thing we would like to consider is how we can shift, stretch, reflect an existing function. Since this should be somewhat familiar we will do is recall in table form all the transforms.

	Horizontal			Vertical		
Shift	$f(x+n)$	$n>0$	Left	$f(x)+n$	$n>0$	Up
		$n<0$	Right		$n<0$	Down
Scale	$f(kx)$	$k>1$	Compress	$kf(x)$	$1>k>0$	Compress
	$f(kx)$	$1>k>0$	Stretch	$kf(x)$	$k>1$	Stretch
Reflect	$f(-x)$			$-f(x)$		

Let's consider a couple examples.

Example 2 *Consider the function $3sin(2x+\pi/4)$. We have three transforms. In the horizontal direction we have a shift of $\pi/4$ which will shift the function to the right and a scaling of 2, which will compress the graph so twice as many peaks will occur in the same area. In the vertical direction we have a scaling of 3, which will stretch the graph so it is $3x$ the height (and depth). Look at figure 3.1 and you can see each of these effects.*

Example 3 *Now consider $-sin(2x)+2$. It also has three transforms. In the horizontal direction it is only compressed by the factor of 2. In the vertical direction it is reflected about the x-axis and then shifted up by 2. If you look at figure 3.2, all of these transforms can be readily seen.*

The key to multiple transforms is to do them in sequence, starting from x and working your way out. Keep the order of operations. Don't confuse yourself, once you have done a transform don't do it again or make yourself wonder. If it is done, it is already handled.

Figure 3.1: Plot of $3sin(2x + \pi/4)$ (pink) and $sin(x)$ (green).

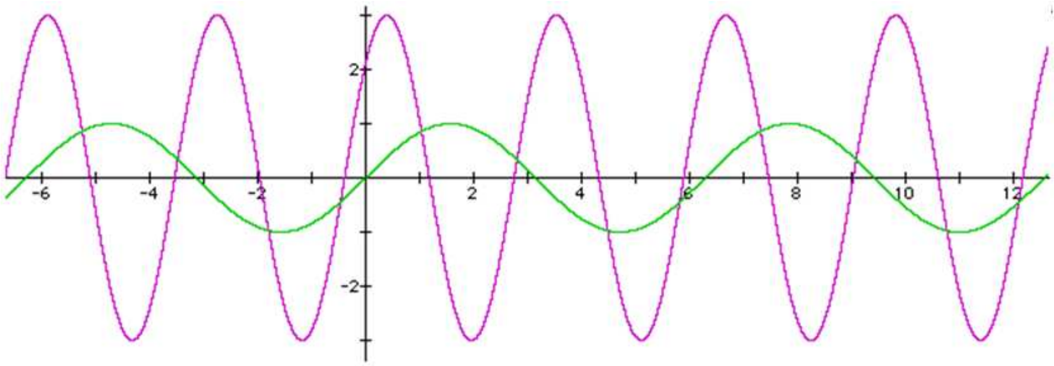
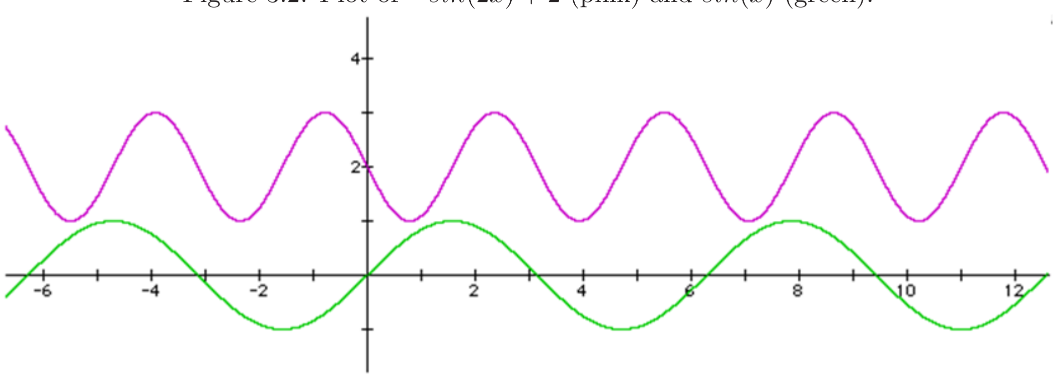


Figure 3.2: Plot of $-sin(2x) + 2$ (pink) and $sin(x)$ (green).



3.2 Even & Odd Functions

From the reflection transforms we can notice that some graphs stay the same when reflected. For instance consider $f(x) = x^2$, with a horizontal reflection. You notice the graph is the same after the reflection! Since this is an interesting and important result we give it a name, even symmetry and even functions. A function is even if $f(-x) = f(x)$. This is also referred to as symmetric about the y-axis. Now consider what happens to $f(x) = x^3$, and perform both horizontal and vertical reflections. You notice the graph again remains the same! This is a nice result also, so we call it odd symmetry or symmetry through the origin. A function is odd if $f(-x) = -f(x)$. Notice that there is no symmetry about the x-axis, as this would result in multiple values if $y \neq 0$. We thus limit ourselves to the use of even and odd symmetry.

3.3 Sums & Compositions

How do we calculate the sum of two functions? For instance, consider $e^x + e^{-x}$. We just stack them. Note that as x approaches infinity e^x is large while e^{-x} is small so the sum approaches e^x from above asymptotically. The reverse is true as x approaches minus infinity.

What is a composition? Simply put a composition is a function within a function. We have seen compositions in the context of the chain rule and calculating inverses.

3.4 Polynomials

Many things could be said about polynomials, but we will limit ourselves to a few key ideas. The first is that of the degree of a polynomial. The degree is the power of the largest monomial. A nice relation between the degree of a polynomial and the roots of the polynomial exists. Note that a root of a function is where the function is zero. A polynomial has as many roots as its degree, if you use complex numbers. Polynomials are usually written in one of two forms: standard or factored.

$$a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$$

or

$$a_n (x - r_1)(x - r_2) \ldots (x - r_n)$$

The second form is the easiest to use to get the formula from a graph. You can read the roots from graph, and then solve for k by using another point, such as the y-intercept. Note that a double root does not cross the axis. One final comment, the sign of the leading coefficient can be recognized by which direction the graph goes as x goes to infinity. If the leading coefficient is positive the graph will go to infinity, if it is negative the graph will go to negative infinity.

3.5 Rational Functions

Rational functions are those that can be expressed as a ratio of two polynomials, Provided there is no cancelation of roots (look at the polynomials in factored form) then the following rules can be used to sketch a graph of the function.

1. The roots of the numerator are the roots of the rational function.
2. The roots of the denominator are the vertical asymptotes of the rational function.
3. If the polynomials have the same degree then the horizontal asymptote is given by the ratio of the leading coefficients.
4. If the denominator has a higher degree then the horizontal asymptote is given by 0.
5. If the numerator has a higher degree then the function does not have a horizontal asymptote.
6. To see if the function approaches the horizontal asymptote from above or below, look at which is larger the numerator or denominator. This can be done by examining the next highest terms and seeing how much they add or subtract.

Chapter 4

Complex Numbers

This might seem an odd place to start for some people, after all they are called "complex numbers" for a reason, right? Of course they are! They turn complex problems into easy ones. They are the numbers to turn to when the problem is hard and you want something to help you out of the tough spots. I will not go into complex analysis, though if you get the chance to you should, it is not only useful it is beautiful...

4.1 Complex Simplifications

One way to start off with complex numbers is to define

$$j = \sqrt{-1}$$

thus we have

$$j^2 = -1$$

Note that mathematicians use i instead of j, but engineers use j because i is current. Anyway, it is then reasonable to ask questions like what is the square root of -4?

$$\begin{aligned} \sqrt{-4} &= \sqrt{(-1)(4)} \\ &= \sqrt{-1}\sqrt{4} \\ &= 2j \end{aligned}$$

We can quickly see that we have more than just one oddball answer. We call the number line of a real number times j the imaginary number line. We often have to deal with a real number added to an imaginary number, so we more generally deal with the complex numbers

$$c = a + jb$$

The real numbers and the imaginary numbers can be handled as a line, but the complex numbers result in a plane. A natural question arises in how do I add and multiply complex numbers. For addition, we add the real parts and the imaginary parts separately

$$(a + bj) + (c + dj) = (a + c) + (b + d)j$$

Multiplication is slightly more complex, but still not bad. We multiply like we would polynomials.

$$\begin{aligned} (a + bj)(c + dj) &= ac + adj + bcj + bdj^2 \\ &= ac + adj + bcj - bd \\ &= (ac - bd) + (ad + bc)j \end{aligned}$$

This is all lovely, but how does this help me with trig? I notice first that I have a plane and I can reference the angle and distance to a point. Thus I can look at the unit circle in the complex plane. I note that this is given by (from the coordinates of a circle)

$$\cos(\theta) + j\sin(\theta)$$

The nice result that helps us is from a Swiss mathematician named Leonhard Euler. Euler proved that

$$e^{j\theta} = \cos(\theta) + j\sin(\theta)$$

The easiest way to prove this involves knowing the derivative of the trig functions and so I will not cover it (at the moment). It is not hard and builds on the intuition we have developed with derivatives of exponentials. It would be nice to take complex numbers and then express them in Euler's form. In general we are not on a unit circle so we need to account for the radius. To find the radius remember that for a point, (a, b), on the circle we can use the Pythagorean Theorem $r^2 = a^2 + b^2$ to find the radius. Divide both terms by this and use the inverse trig functions on your calculator. Note that you must know what quadrant you are in to make sure you specify the correct angle.

4.2 Euler's Formula

Probably one of the most useful features of complex numbers is that they have the ability to turn trigonometry into an application of exponential functions. This will seem an odd claim to most people, particularly as functions like sine and cosine do not look like an exponential. Well at least not exponentials with real exponents. When we define the complex numbers we have to explain what we mean by the various functions. I am going to develop Euler's equation from a heuristic standpoint to try to encourage some intuition for complex numbers. I hope as intuition increases that fear will decrease. If you want a rigorous treatment, there are many fine texts on the subject of complex analysis.

Consider a complex number $a = \alpha_r + j\alpha_c$. In any case, this describes a point in cartesian space (α_r, α_c). The number can then be thought of as a point, and the distance of this point from the origin, say r is given by the pythagorean identity,

$$r = \sqrt{\alpha_r^2 + \alpha_c^2}.$$

We could also find the angle between the horizontal and the line from the origin to the point to be

$$\theta = \arctan \frac{\alpha_c}{\alpha_r}.$$

Make sure to use the four quadrant version of arctangent, that requires the specification of the numerator and denominator (not just the ratio). Consider a particular number in the complex plane, $(1,0)$. This has radius, $r=1$, and angle, $\theta=0$. What happens when, we square this? We know the answer must be one, but there are many ways to calculate this, which one makes sense? Lets consider a couple more points. In defining the complex plane we have used the point $j=\sqrt{-1}=(0,1)$. This point has radius, $r=1$ and an angle[1] of $\theta=\frac{\pi}{2}$. Now the square of j must be $-1=(-1,0)$, which has radius, $r=1$ and angle, $\theta=\pi$. This last relation suggest that we want to multiply the radiuses and add the angles. That would mean

$$\begin{aligned} j^2 &= (0,1)\times(0,1) \\ &= 1\angle\frac{\pi}{2}\times 1\angle\frac{\pi}{2} \\ &= 1\times 1\angle(\frac{\pi}{2}+\frac{\pi}{2}) \\ &= 1\angle\pi \\ &= -1 \end{aligned}$$

Following this definition we have

$$\begin{aligned} (-1)^2 &= (-1,0)\times(-1,0) \\ &= 1\angle\pi\times 1\angle\pi \\ &= 1\times 1\angle(\pi+\pi) \\ &= 1\angle 2\pi \\ &= 1 \end{aligned}$$

This works quite nicely, but now begs the question of why we add to do multiplication. Does anything else have this property? Exponentials! Recall that

$$e^a \times e^b = e^{a+b}$$

[1] Note that we will be using radians to measure all angles. This may bug some people who are used to degrees. I am sorry you were taught wrong originally. Degrees were introduced by the Babylonians who used base 60 numbers to simplify calculations (particularly division of angles). It is fine to use this when all you are concerned with is division. With the advent of computers and calculators this is not an issue so a better basis can be used to pick your measurement system. Radians were designed to make calculation easy. This is kind of a mantra with me. Make life and calculation easy. What calculation you may ask. Besides a bunch of calculus and transform problems I will deal with later, it simplifies problems like the finding the length of an arc. In radian angle measure, the length of an arc is just the radius of the arc times the internal angle, $l=r\theta$. This will not work if you use degrees. You could include an ugly scale factor. Each time you had a new equation that was easy in radian measure you could put in a factor to convert it to degrees, but this can involve a bunch of different constants (all related to the basic conversion of $\frac{\pi}{180}$). Soon you would realize that you don't have to use all those ugly factors if you just used radians. I first leaned using degrees. Radians are better.

Note that the addition happens in the exponent, so the angle must be in the exponent, but the exponent will change the magnitude, right? For real exponents that is true, but what about on the imaginary numbers? We can define the function any consistent way we want, so we define $|e^{j\theta}| = 1$. This implies the "phasor" notation of engineering

$$\begin{aligned}(a, b) &= \sqrt{a^2 + b^2} \angle \arctan \frac{b}{a} \\ &= \sqrt{a^2 + b^2} e^{j \arctan \frac{b}{a}} \\ &= r e^{j\theta}\end{aligned}$$

Now note this polar coordinate version must be equal to the cartesian version

$$\begin{aligned}(a, b) &= a + jb \\ &= \sqrt{a^2 + b^2} \cos \arctan \frac{b}{a} + j \sqrt{a^2 + b^2} \sin \arctan \frac{b}{a} \\ &= r \cos \theta + jr \sin \theta\end{aligned}$$

By equating Eq 4.1 and Eq 4.1 we find

$$\begin{aligned}r \cos \theta + jr \sin \theta &= r e^{j\theta} \\ \cos \theta + j \sin \theta &= e^{j\theta}\end{aligned}$$

This is the famous Euler's (pronounced oiler's) equation. Now what can we do with it. First we want to solve for cosine and sine.

$$\begin{aligned}e^{j\theta} &= \cos \theta + j \sin \theta \\ e^{-j\theta} &= \cos -\theta + j \sin -\theta \\ &= \cos \theta - j \sin \theta\end{aligned}$$

Adding Eq 4.1 and Eq 4.1 we find

$$\begin{aligned}e^{j\theta} + e^{-j\theta} &= 2 \cos \theta \\ \cos \theta &= \frac{e^{j\theta} + e^{-j\theta}}{2}\end{aligned}$$

Similarly, subtracting yields

$$\begin{aligned}e^{j\theta} - e^{-j\theta} &= 2j \sin \theta \\ \sin \theta &= \frac{e^{j\theta} - e^{-j\theta}}{2j}\end{aligned}$$

Example 4 *Let's use our new equations to calculate something. So you can verify it, I will* *prove the famous equation* $\cos^2\theta + \sin^2\theta = 1$.

$$
\begin{aligned}
f(\alpha,\beta) &= \cos^2\theta + \sin^2\theta \\
&= \left(\frac{e^{j\theta}+e^{-j\theta}}{2}\right)^2 + \left(\frac{e^{j\theta}-e^{-j\theta}}{2j}\right)^2 \\
&= \left(\frac{e^{j\theta}e^{j\theta}+e^{-j\theta}e^{j\theta}+e^{j\theta}e^{-j\theta}+e^{-j\theta}e^{-j\theta}}{4}\right) + \left(\frac{e^{j\theta}e^{j\theta}-e^{-j\theta}e^{j\theta}-e^{j\theta}e^{-j\theta}+e^{-j\theta}e^{-j\theta}}{-4}\right) \\
&= \left(\frac{e^{2j\theta}+e^{0}+e^{0}+e^{-2j\theta}}{4}\right) - \left(\frac{e^{2j\theta}-e^{0}-e^{0}+e^{-2j\theta}}{4}\right) \\
&= \frac{e^{2j\theta}+e^{0}+e^{0}+e^{-2j\theta}-e^{2j\theta}+e^{0}+e^{0}-e^{-2j\theta}}{4} \\
&= \frac{4e^0}{4} \\
&= e^0 \\
&= 1
\end{aligned}
$$

It is just algebra. No more special trig games. That is cool.

Example 5 *Now let us try to use this for simplification.*

$$
\begin{aligned}
g(\alpha,\beta) &= \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) \\
&= \left(\frac{e^{j\alpha}+e^{-j\alpha}}{2}\right)\left(\frac{e^{j\beta}+e^{-j\beta}}{2}\right) - \left(\frac{e^{j\alpha}-e^{-j\alpha}}{2j}\right)\left(\frac{e^{j\beta}-e^{-j\beta}}{2j}\right) \\
&= \frac{e^{j\alpha}e^{j\beta}+e^{-j\alpha}e^{j\beta}+e^{j\alpha}e^{-j\beta}+e^{-j\alpha}e^{-j\beta}}{4} - \frac{e^{j\alpha}e^{j\beta}-e^{-j\alpha}e^{j\beta}-e^{j\alpha}e^{-j\beta}+e^{-j\alpha}e^{-j\beta}}{-4} \\
&= \frac{e^{j\alpha}e^{j\beta}+e^{-j\alpha}e^{j\beta}+e^{j\alpha}e^{-j\beta}+e^{-j\alpha}e^{-j\beta}+e^{j\alpha}e^{j\beta}-e^{-j\alpha}e^{j\beta}-e^{j\alpha}e^{-j\beta}+e^{-j\alpha}e^{-j\beta}}{4} \\
&= \frac{2e^{j\alpha}e^{j\beta}+2e^{-j\alpha}e^{-j\beta}}{4} \\
&= \frac{e^{j(\alpha+\beta)}+e^{-j(\alpha+\beta)}}{2} \\
&= \cos(\alpha+\beta)
\end{aligned}
$$

A nice result and it gives us something we will use when we are working on the derivative of cosine.

Example 6 *Try the following*

$$
\begin{aligned}
h(\alpha,\beta) &= \cos(\alpha)\sin(\beta)+\sin(\alpha)\cos(\beta) \\
&= \left(\frac{e^{j\alpha}+e^{-j\alpha}}{2}\right)\left(\frac{e^{j\beta}-e^{-j\beta}}{2j}\right)+\left(\frac{e^{j\alpha}-e^{-j\alpha}}{2j}\right)\left(\frac{e^{j\beta}+e^{-j\beta}}{2}\right) \\
&= \frac{e^{j\alpha}e^{j\beta}+e^{-j\alpha}e^{j\beta}-e^{j\alpha}e^{-j\beta}-e^{-j\alpha}e^{-j\beta}}{4j}+\frac{e^{j\alpha}e^{j\beta}-e^{-j\alpha}e^{j\beta}+e^{j\alpha}e^{-j\beta}-e^{-j\alpha}e^{-j\beta}}{4j} \\
&= \frac{e^{j\alpha}e^{j\beta}+e^{-j\alpha}e^{j\beta}-e^{j\alpha}e^{-j\beta}-e^{-j\alpha}e^{-j\beta}+e^{j\alpha}e^{j\beta}-e^{-j\alpha}e^{j\beta}+e^{j\alpha}e^{-j\beta}-e^{-j\alpha}e^{-j\beta}}{4j} \\
&= \frac{2e^{j\alpha}e^{j\beta}-2e^{-j\alpha}e^{-j\beta}}{4j} \\
&= \frac{e^{j(\alpha+\beta)}-e^{-j(\alpha+\beta)}}{2j} \\
&= \sin(\alpha+\beta)
\end{aligned}
$$

A nice result and it gives us something we will use when we are working on the derivative of sine.

Chapter 5

Averages

5.1 Average Velocity

We want to consider the ratio of how the output changes when the input is changed. We will do this by approximating with secant lines. For the function, $Q = f(t)$, this is written mathematically as:

$$\frac{\Delta Q}{\Delta t} = \frac{Q_2 - Q_1}{t_2 - t_1} = \frac{f(t_2) - f(t_1)}{t_2 - t_1} \tag{5.1}$$

often referred to as rise over run. The "run" is the step size of the secant line. Let $t_2 - t_1 = h$ then

$$\frac{\Delta Q}{\Delta t} = \frac{f(t_2) - f(t_1)}{t_2 - t_1} \tag{5.2}$$

$$= \frac{f(t_1 + h) - f(t_1)}{t_1 + h - t_1} \tag{5.3}$$

$$= \frac{f(t_1 + h) - f(t_1)}{h} \tag{5.4}$$

5.2 Outline

Here is what I plan on adding once I have time.

1. Sketch linear function and label the points. Note what this means.
   (a) Add one more point (t3,Q3) and discuss the (average) rate of change on the new interval
2. Now sketch $y = x2 - 2x + 1$
   (a) Calculate over the intervals:

0,3
1,3
0,2
0,1 , [1,2], [2,3]

(b) Note which intervals where increasing. What is going on with (i) and (iii)?

(c) Trying to approximate slope on small intervals.

3. What if we took limit? - Tangent line!

Graph tangent lines
Show how to approximate a function with them
Given info on slope at various points show how to approximate graph
Discuss Supply and demand again

Part II

Calculus

Chapter 6

Limits

The fundamental concept in calculus is that of the limit. The limit is a mathematical operation, which returns the value that a sequence is tending to. Note that the sequence does not have to ever reach the point. The sequence can be defined by an equation, in which case the the limit must be expressed in terms of what the independent variable is approaching and sometimes from what direction it approaches.

Consider the following

- $\lim_{x\to\infty}\frac{1}{x}=0$ This is true even though the sequence never is equal to zero. Zero is the value the sequence is approaching.

- $\lim_{x\to0^+}\frac{1}{x}=\infty$ As x approaches zero from the positive side, the function gets larger and larger. Again the function is going to infinity by infinity is not part of the reals, so it does not reach it. The point we all think of as "reaching it" is $x=0$, but the function is undefined for that value, as division by zero is undefined (because infinity is not part of the reals). You could add infinity, in which case you get the extended reals, but that is not the point here.

- $\lim_{x\to0^-}\frac{1}{x}=-\infty$ As x approaches zero from the negative side, the function gets more and more negative. As above it approaches negative infinity, but never reaches it. This is one of those weird things about infinity. If you think of the function as a continuous string on the extended reals, the function gets more and more negative as $x\to0^-$ till it reaches negative infinity. It then flips sign to positive infinity and comes back down. To have a continuous string, both positive and negative infinity must be next to each other in some sense. I always liked that. The extremes bend around to each other.

Note we can't speak of the limit as $x\to0$ as the limit from the left and right are different. This becomes a major idea in calculus.

6.1 Formalizing the limit

The function, $f(x)$, has limit, L, at $x = a$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $|L - f(x)| < \epsilon$ if $0 < |a - x| < \delta$. This is often referred to as $\epsilon - \delta$ (epsilon-delta) notation. How can we use it[1]? We must find delta as a function of epsilon ($\delta = g(\epsilon)$) so that the definition is satisfied and then you have proven that the function converges to the stated limit at that point. Note that if some δ_1 works then any smaller (but not zero or negative) will also work. What would such a proof look like?

Example 7 *Consider $y = x^2$ as $x \to 3$. This is trivial to calculate, but this should make it easier to see what is happening. We believe that $x^2 \to 9$ as $x \to 3$, but we want to prove it. In this case, $L = 9$ and $a = 3$, thus given $\epsilon > |9 - x^2|$ we need to find a δ such that $0 < |3 - x| < \delta$*

1. *Factor our ϵ expression: $\epsilon > |9 - x^2| = |(3 + x)(3 - x)|$*

2. *Note that $|(3 + x)(3 - x)| \leqslant |3 + x||3 - x| = |6 + x - 3||3 - x| \leqslant (6 + |x - 3|)|3 - x| < (6 + \delta)\delta < \epsilon$*

3. *Consider the equation $\delta^2 + 6\delta - \epsilon = 0$*

4. *Roots are: $\delta = \frac{-6\pm\sqrt{36+4\epsilon}}{2} = -3 \pm \sqrt{9 + \epsilon}$*

5. *We can thus pick $\delta = \sqrt{9 + \epsilon} - 3 > 0$ and we have δ as a function of ϵ, and the limit is proven.*

Seems like a lot of work, but it is rigorous (i.e. there are no holes with our proof).

Example 8 *Consider $y = \mathrm{sinc}x = \frac{sin(x)}{x}$ as $x \to 0$. We evaluate the function a few points approaching zero, and it looks like the answer is 1. How do we know? We must show that given $\epsilon > |1 - \mathrm{sinc}x|$ that there exists δ such that $0 < |x| < \delta$*

1. *$|1 - \mathrm{sinc}x| = \frac{|x-sin(x)|}{|x|} = \frac{|x-(x-x^3/6+\ldots)|}{|x|} < \frac{|x^3/6|}{|x|} = \frac{|x|^2}{6} = \frac{\delta^2}{6} < \delta^2 = \epsilon$. Note the final equality is my choice. All intermediate steps were to get a nice upper bound in terms of δ. Note also that I used the Taylor series expansion of $\sin(x)$. I will explain Taylor series later, but it vastly simplifies the proof to borrow it now.*

2. *$\delta = \sqrt{\epsilon} > 0$ thus satisfies our requirements and the limit of 1 is proven.*

What else can we do with it? Prove some properties of limits!

[1] Besides making Yasha happy that we are being rigorous.

6.2 Properties of Limits

Lemma 2 *The limit of a constant is the constant*

Proof:

This may seem trivial, only because it is. But even simple things must be proven. If it is easy, then no use leaving it undone. Let $\lim_{x\to a} c$ be the limit we wish to prove goes to c. Then given $\epsilon > |c - c|$, we must show $\exists\delta$ such that $0 < |a - x| < \delta$. Note $\exists$ is mathematical short-hand for "there exists".

$|c - c| = 0 < |a - x| < \delta = \epsilon$, and we have our function, and thus our proof.

◇ SDG ◇

Theorem 3 *The limit of a sum is the sum of the limits*

Proof:

I want to show that

$$\lim_{x\to a}(f(x) + g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$$

Let

$$F = \lim_{x\to a} f(x)$$
$$G = \lim_{x\to a} g(x)$$

Then given $\epsilon > |F + G - (f(x) + g(x))|$, we must show $\exists\delta$ such that $0 < |a - x| < \delta$.

$|F + G - (f(x) + g(x))| = |(F - f(x)) + (G - g(x))| = |(\epsilon_f) + (\epsilon_g)| = \epsilon_f + \epsilon_g \leqslant 2\max(\epsilon_f, \epsilon_g) < \epsilon$, since the indicated limits exist. Now since the limits exist there must be a function $\delta = h_f(\epsilon_f)$ and a function $\delta = h_g(\epsilon_g)$, for $f(x)$ and $g(x)$ respectively. Further, $\delta = \min(h_f(\epsilon), h_g(\epsilon)) \geqslant |x - a| \geqslant 0$ is trivially a function, which satisfies our requirements. Thus again we have our function and proof.

◇ SDG ◇

Theorem 4 *The limit of a product is the product of the limits*

Proof:

I want to show that

$$\lim_{x\to a}(f(x)g(x)) = \left(\lim_{x\to a} f(x)\right)\left(\lim_{x\to a} g(x)\right)$$

Let

$$F = \lim_{x\to a} f(x)$$
$$G = \lim_{x\to a} g(x)$$

Then given $\epsilon > |FG - (f(x)g(x))|$, we must show $\exists\delta$ such that $0 < |a - x| < \delta$.

$|FG - f(x)g(x)| = |FG - Fg(x) + Fg(x) - f(x)g(x)| = |F(G - g(x)) + (F - f(x))g(x)| = |F(\epsilon_g) + (\epsilon_f)g(x)| \leqslant |F(\epsilon_g) + (\epsilon_f)(|G| + \epsilon_g)| \leqslant |F|\epsilon_g + \epsilon_f|G| + \epsilon_f\epsilon_g$. Now let $\epsilon_m = \max(\epsilon_f, \epsilon_g)$ then $|FG - f(x)g(x)| \leqslant (F + G)\epsilon_m + \epsilon_m^2 = \epsilon$. Thus $\delta = \sqrt{\min(h_f(\epsilon), h_g(\epsilon))} \geqslant |x - a| \geqslant 0$ and again we have our function and proof.

$\diamondsuit$ SDG $\diamondsuit$

Example 9

$$\lim_{x\to 2} \frac{5x^3 + 4}{x - 3} = -44 \tag{6.1}$$

6.3 Limits at Infinity

Note properties still hold for limits at infinity.

Theorem 5

$$\lim_{x\to\pm\infty} \frac{1}{x} = 0$$

Simple technique: divide numerator and denominator by highest power

Example 10

$$\begin{aligned}
\lim_{x\to\infty} \frac{3x+5}{6x-8} &= \lim_{x\to\infty} \frac{3+5/x}{6-8/x} && (6.2)\\
&= \frac{3 + 5\lim_{x\to\infty}(1/x)}{6 - 8\lim_{x\to\infty}(1/x)} && (6.3)\\
&= \frac{3 + 5\cdot 0}{6 - 8\cdot 0} && (6.4)\\
&= \frac{3}{6} && (6.5)\\
&= \frac{1}{2} && (6.6)
\end{aligned}$$

Example 11

$$\begin{aligned}
\lim_{x\to\infty} \frac{3x+5}{6x^2-8} &= \lim_{x\to\infty} \frac{3/x+5/x^2}{6-8/x^2} && (6.7)\\
&= \frac{3\lim_{x\to\infty}(1/x) + 5\lim_{x\to\infty}(1/x^2)}{6 - 8\lim_{x\to\infty}(1/x^2)} && (6.8)\\
&= \frac{3\cdot 0 + 5\cdot 0}{6 - 8\cdot 0} && (6.9)\\
&= \frac{0}{6} && (6.10)\\
&= 0 && (6.11)
\end{aligned}$$

Example 12

$$\lim_{x\to\infty} \frac{3x^2+5}{6x-8} = \lim_{x\to\infty} \frac{3+5/x^2}{6/x-8/x^2} \tag{6.12}$$

$$= \frac{3+5\lim_{x\to\infty}(1/x^2)}{6\lim_{x\to\infty}(1/x)-8\lim_{x\to\infty}(1/x^2)} \tag{6.13}$$

$$= \lim_{d\to 0} \frac{3+5\cdot 0}{6\cdot d-8\cdot d} \tag{6.14}$$

$$= \lim_{d\to 0} \frac{3}{-2d} \tag{6.15}$$

$$= -1.5 \lim_{d\to 0} \frac{1}{d} \tag{6.16}$$

$$= \infty \tag{6.17}$$

Behaves like ratio of highest powers as $x \to \infty$

Chapter 7

Calculus

7.1 Derivative

Let's say we wanted to know what the slope was at some point on a curve. Slopes are important in a lot of things we do, and we are often working with curves rather than straight lines so the case is fairly common.

For instance we could be trying to find the velocity from a table of time vs. position data. The first idea that comes to mind is to take two points from the table and calculate the change in position divided by the change in time. This is the "average velocity", and it implicitly assumes zero acceleration[1]. We are have from the table[2] that $p = f(t)$ and thus

$$\bar{v} = \frac{p_1 - p_0}{t_1 - t_0} \tag{7.1}$$

$$= \frac{f(t_1) - f(t_0)}{t_1 - t_0} \tag{7.2}$$

Were you to draw this, you would have some curve, $p = f(t)$ and a line of slope, $\bar{v}$, which crosses the curve at two points. Such a line is called a secant line in geometry. If the spacing between the time intervals is constant, say Δt then

$$\bar{v} = \frac{f(t_1) - f(t_0)}{t_1 - t_0} \tag{7.3}$$

$$= \frac{f(t_0 + \Delta t) - f(t_0)}{t_0 + \Delta t - t_0} \tag{7.4}$$

$$= \frac{f(t_0 + \Delta t) - f(t_0)}{\Delta t} \tag{7.5}$$

Our assumption of no acceleration has occurred becomes closer to true as the time interval becomes smaller and smaller. Let's consider what would happen if the spacing between the

[1] This is different that the acceleration being zero. Zero acceleration means the velocity has not changed (for instance no time might have elapsed), while the acceleration being zero, means the velocity will not change. The difference is subtle but important.

[2] Recall a table defines a function. This is all we are using, as the function is unspecified at this point.

points approached zero[3]. Right, we have to take a limit.

$$v = \lim_{\Delta t \to 0} \frac{f(t_0 + \Delta t) - f(t_0)}{\Delta t} \tag{7.6}$$

Notice that I have written v and not $\bar{v}$. This is not a typo, it is deliberate. When we calculate the derivative we are calculating the velocity at a point, also called "instantaneous velocity". In one sense "instantaneous velocity" makes no physical sense. How can you have velocity with a change in position? This concept has puzzeled man since at least Zeno's dialog with Socrates[4]. We say this limit exists when three conditions are met

1. $-\infty < \lim_{\Delta t \to 0^-} \frac{f(t_0+\Delta t)-f(t_0)}{\Delta t} < \infty$
2. $-\infty < \lim_{\Delta t \to 0^+} \frac{f(t_0+\Delta t)-f(t_0)}{\Delta t} < \infty$
3. $\lim_{\Delta t \to 0^-} \frac{f(t_0+\Delta t)-f(t_0)}{\Delta t} = \lim_{\Delta t \to 0^+} \frac{f(t_0+\Delta t)-f(t_0)}{\Delta t}$

In words, the left and right limit must exist (be finite) and be equal to each other. When this happens we say the derivative exists and we define the derivative to be the value of this limit. I told you the limit was essential to us. We write the derivative as

$$f'(t) = \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t} \tag{7.7}$$

We also can see that the velocity is the derivative of position. This is one of many relations used in physics and engineering that is a direct application of calculus.

As an important side note, we will often deal with piecewise functions (defined on intervals, but discontinuous between them). We must have a continuous interval in which to define the derivative. This is true for discontinuities in x and y.

7.1.1 Derivative at a Point

We will deal with an important sub-topic to derivatives: derivatives at a point. Sometimes we don't care about the derivative at every point, rather we care about some point of interest. It often is easier to handle just this case.

Assume we want to examine the derivative of a function $f(x)$ at $x = x_0$. As we saw in the case of the instantaneous velocity, we will consider the limit of secant line defined by the two points $(x_0, f(x_0))$ and $(x_0 + h, f(x_0 + h))$, that yields the tangent line. Since the derivative is the slope of this line, we have:

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{(x_0 + h) - (x_0)} \tag{7.8}$$

$$= \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} \tag{7.9}$$

[3]When the points merge, our secant line is now only touching the curve at 1 point, so it is now a tangent line. This observation is also important to calculus, as we speak of the derivative as the tangent of the curve.

[4]Zeno is one of the rare few, who got the better of Socrates, and like Meno, who also did, the point was lost on the otherwise brilliant Socrates, because he had no answer. Let this be a warning to us, as denying the problem set math back 2500 years in this case.

Example 13 *Given* $f(x) = x^3 + 5x + 1$ *find* $f'(2)$

$$f'(2) = \lim_{h \leftarrow 0} \frac{f(2+h) - f(2)}{h} \quad (7.10)$$

$$= \lim_{h \leftarrow 0} \frac{(2+h)^3 + 5(2+h) + 1 - 19}{h} \quad (7.11)$$

$$= \lim_{h \leftarrow 0} \frac{(8 + 12h + 6h^2 + h^3) + (10 + 5h) + 1 - 19}{h} \quad (7.12)$$

$$= \lim_{h \leftarrow 0} \frac{17h + 6h^2 + h^3}{h} \quad (7.13)$$

$$= \lim_{h \leftarrow 0} (17 + 6h + h^2) \quad (7.14)$$

$$= 17 \quad (7.15)$$

7.1.2 Derivative Graphically

Example 14 *Consider a chart of healing rate for a cut. The size of the scab is measured by taking a digital photo at a set distance and the pixels added up and scaled to give the area. The data is listed below.*

Area [mm^2]	*400*	*360*	*180*	*120*	*90*	*72*	*60*
Days	*0*	*1*	*2*	*3*	*4*	*5*	*6*

Approximating the derivative numerically using secant formula on successive points. Note this will give us one less slope point as we are finding the slope between two points.

Secant	*40*	*180*	*60*	*30*	*18*	*8*

The secant calculations can be used to give us the slope approximations for each point on the left (-) and right (+).

Area [mm^2]	*400*	*360*	*180*	*120*	*90*	*72*	*60*
Days	*0*	*1*	*2*	*3*	*4*	*5*	*6*
m^+	*40*	*180*	*60*	*30*	*18*	*8*	
m^-		*40*	*180*	*60*	*30*	*18*	*8*

Note that one value for both the left and right derivatives is missing. We can approximate the missing value by assuming it is the same as the slope next to it.

Area [mm^2]	*400*	*360*	*180*	*120*	*90*	*72*	*60*
Days	*0*	*1*	*2*	*3*	*4*	*5*	*6*
m^+	*40*	*180*	*60*	*30*	*18*	*8*	*8*
m^-	*40*	*40*	*180*	*60*	*30*	*18*	*8*

We can improve this by averaging the left and right secant line slopes

Area [mm^2]	*400*	*360*	*180*	*120*	*90*	*72*	*60*
Days	*0*	*1*	*2*	*3*	*4*	*5*	*6*
m^+	*40*	*180*	*60*	*30*	*18*	*8*	*8*
m^-	*40*	*40*	*180*	*60*	*30*	*18*	*8*
mavg	*40*	*110*	*120*	*45*	*24*	*13*	*8*

7.1.3 Understanding the Derivative

Significance of the derivative

1. The sign of the derivative tells us if the function is increasing or decreasing at the point
2. The magnitude of the derivative tells us how fast the function is changing (rate)
3. When the derivative is zero then we are at a critical point, which gives us the potential locations for minima or maxima (a key observation for optimization)

Four cases when derivative won't exist:

1. undefined point/function (can't even write expression)
2. not continuous (goes to infinity, thus no limit)
3. vertical tangent (no limit)
4. not smooth (left limit $\neq$ right limit, ambiguous)

7.1.4 Derivative Review

Let's Review:

- We have talked about limits
- We have discussed the need to find an interval in x so that $f(x)$ lies within an interval of the limit on y

Where are we going and how will we use it? We want to take a derivative, for example, to find the instantaneous velocity from distance measures.

- Instantaneous velocity is not possible to measure. What does it mean? How does it work? Basic problem since Zeno postulated his paradoxes 2500 years ago
- We can talk meaningfully about average speed, and measure it.
- Distance traveled over time - secant lines
- Take smaller intervals - limits
- $\lim_{\Delta t \to 0} \frac{f(t+\Delta t)-f(t)}{\Delta t}$
- How do we calculate this? Must know limit
- We will develop short-cuts and rules, but they are based on what we do know.

7.2 Calculus in Business

Marginal cost Rate of change in cost per unit change in production In economics and business, marginal refers to a derivative.

Example 15 *Cost to produce (in hundreds of dollars) x windsurfers per day is* $3+10x-x^2$ *for* $0 \leqslant x \leqslant 4$. *Find the marginal cost at x= 1 and x=3.*

7.3 Properties of the Derivative

7.3.1 constants

Theorem 6 *The derivative of a constant is zero.*

Proof:

$$
\begin{aligned}
f(x) &= c && (7.16)\\
f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} && (7.17)\\
&= \lim_{h\to 0} \frac{c-c}{h} && (7.18)\\
&= \lim_{h\to 0} \frac{0}{h} && (7.19)\\
&= \lim_{h\to 0} 0 && (7.20)\\
&= 0 && (7.21)
\end{aligned}
$$

♢ SDG ♢

7.3.2 Monomials

Theorem 7 *The derivative of a line is the slope.*

Proof:

$$
\begin{aligned}
f(x) &= mx+b && (7.22)\\
f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} && (7.23)\\
&= \lim_{h\to 0} \frac{(m(x+h)+b)-(mx+b)}{h} && (7.24)\\
&= \lim_{h\to 0} \frac{mh}{h} && (7.25)\\
&= \lim_{h\to 0} m && (7.26)\\
&= m && (7.27)
\end{aligned}
$$

♢ SDG ♢

Theorem 8 (Power Rule) *The derivative of a monomial, ax^n, is anx^{n-1}.*

Proof:

$$
\begin{aligned}
f(x) &= ax^n && (7.28)\\
f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} && (7.29)\\
&= \lim_{h\to 0}\frac{a(x+h)^n-ax^n}{h} && (7.30)\\
&= \lim_{h\to 0}\frac{a\left(x^n+\binom{n}{1}hx^{n-1}+\ldots+h^n\right)-ax^n}{h} && (7.31)\\
&= \lim_{h\to 0}\frac{a\left(\binom{n}{1}hx^{n-1}+\ldots+h^n\right)}{h} && (7.32)\\
&= \lim_{h\to 0} a\left(\binom{n}{1}x^{n-1}+\ldots+h^{n-1f}\right) && (7.33)\\
&= a\binom{n}{1}x^{n-1} && (7.34)\\
&= anx^{n-1} && (7.35)
\end{aligned}
$$

♢ SDG ♢

7.3.3 Exponentials

Theorem 9 *The derivative of the exponential, e^x, is e^x.*

Proof:

$$
\begin{aligned}
f(x) &= e^x && (7.36)\\
f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} && (7.37)\\
&= \lim_{h\to 0}\frac{e^{(x+h)}-e^x}{h} && (7.38)\\
&= \lim_{h\to 0}\frac{e^xe^h-e^x}{h} && (7.39)\\
&= \lim_{h\to 0} e^x\frac{e^h-1}{h} && (7.40)\\
&= \lim_{h\to 0} e^x\frac{(1+h+h^2/2\ldots)-1}{h} && (7.41)\\
&= \lim_{h\to 0} e^x\frac{h+h^2/2\ldots}{h} && (7.42)\\
&= \lim_{h\to 0} e^x(1+h/2\ldots) && (7.43)\\
&= e^x && (7.44)
\end{aligned}
$$

♢ SDG ♢

Note that I have again taken advantage of Taylor's expansion. On another note, this theorem shows what is so natural about e.

Corollary 10

$$\frac{d}{dx}\left[e^{kx}\right] = ke^{kx}$$

Corollary 11

$$\frac{d}{dx}\left[a^{kx}\right] = \ln(a)ka^{kx}$$

Proof:

$$\frac{d}{dx}\left[a^{kx}\right] = \frac{d}{dx}\left[\left(e^{\ln(a)}\right)^{kx}\right] \quad (7.45)$$

$$= \frac{d}{dx}\left[e^{\ln(a)kx}\right] \quad (7.46)$$

$$= \ln(a)ke^{\ln(a)kx} \quad (7.47)$$

$$= \ln(a)k\left(e^{\ln(a)}\right)^{kx} \quad (7.48)$$

$$= \ln(a)ka^{kx} \quad (7.49)$$

♢ SDG ♢

7.3.4 Product and Quotient Rules

Theorem 12 (Product Rule)

$$\frac{d}{dx}[f(x)g(x)] = f(x)g'(x) + f'(x)g(x)$$

Proof:

$$\frac{d}{dx}[f(x)g(x)] = \lim_{h\to 0}\frac{f(x+h)g(x+h) - f(x)g(x)}{h} \quad (7.50)$$

$$= \lim_{h\to 0}\frac{f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x)}{h} \quad (7.51)$$

$$= \lim_{h\to 0}\frac{f(x+h)(g(x+h) - g(x)) + (f(x+h) - f(x))g(x)}{h} \quad (7.52)$$

$$= \lim_{h\to 0} f(x+h)\lim_{h\to 0}\frac{g(x+h) - g(x)}{h} + g(x)\lim_{h\to 0}\frac{f(x+h) - f(x)}{h} \quad (7.53)$$

$$= f(x)g'(x) + g(x)f'(x) \quad (7.54)$$

♢ SDG ♢

Theorem 13 (Quotient rule)

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$$

Proof:

Let $Q(x) = \frac{f(x)}{g(x)}$ thus $f(x) = Q(x)g(x)$. We will take the derivative of this using the product rule.

$$f'(x) = Q(x)g'(x) + Q'(x)g(x) \quad (7.55)$$
$$= \frac{f(x)}{g(x)}g'(x) + Q'(x)g(x) \quad (7.56)$$

Now solve for $Q'(x)$.

$$f'(x) = \frac{f(x)}{g(x)}g'(x) + Q'(x)g(x) \quad (7.57)$$
$$Q'(x)g(x) = f'(x) - \frac{f(x)}{g(x)}g'(x) \quad (7.58)$$
$$Q'(x) = \frac{f'(x)}{g(x)} - \frac{f(x)g'(x)}{g^2(x)} \quad (7.59)$$
$$Q'(x) = \frac{f'(x)g(x)}{g^2(x)} - \frac{f(x)g'(x)}{g^2(x)} \quad (7.60)$$
$$Q'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)} \quad (7.61)$$

♢ SDG ♢

7.3.5 Chain Rule

We want to consider composition of functions, $f(x) = g(s(x))$. This result is called the chain rule, because it is used to "chain" functions together.

Theorem 14 (Chain Rule)

$$\frac{d}{dx}[g(s(x))] = g'(s(x))s'(x)$$

Proof:

I start with my standard first step of writing the derivative algebraically

$$\frac{d}{dx}[g(s(x))] = \lim_{h\to 0}\frac{g(s(x+h)) - g(s(x))}{h} \quad (7.62)$$

I need to write this into combinations of:

$$\frac{d}{dx}[g(x)] = \lim_{\Delta u \to 0} \frac{g(u+\Delta u) - g(u)}{\Delta u} \tag{7.63}$$

and

$$\frac{d}{dx}[s(x)] = \lim_{h \to 0} \frac{s(x+h) - s(x)}{h} \tag{7.64}$$

This means I need to rewrite $s(x+h)$, so I recall that the secant approximation of the derivative can be expressed as

$$s'(x) = \frac{s(x+h) - s(x)}{h} - E_s(h) \tag{7.65}$$

to get

$$s(x+h) = s(x) + (s'(x) + E_s(h))h \tag{7.66}$$

Substitute this into my limit definition

$$\frac{d}{dx}[g(s(x))] = \lim_{h \to 0} \frac{g(s(x) + (s'(x) + E_s(h))h) - g(s(x))}{h} \tag{7.67}$$

I note that if I call $u = s(x)$ and $\Delta u = (s'(x) + Es(h))h$, that I almost have what I wanted. The only remaining challenge is to get the denominator to be Δu. To get this we use the trick[5] of multiplying by 1.

$$\begin{aligned}
\frac{d}{dx}[g(s(x))] &= \lim_{h \to 0} \frac{g(s(x) + (s'(x) + E_s(h))h) - g(s(x))}{h} \frac{(s'(x) + E_s(h))}{(s'(x) + E_s(h))} && (7.68)\\
&= \lim_{h \to 0} \frac{g(s(x) + (s'(x) + E_s(h))h) - g(s(x))}{(s'(x) + E_s(h))h} (s'(x) + E_s(h)) && (7.69)\\
&= \lim_{\Delta u \to 0} \frac{g(u+\Delta u) - g(u)}{\Delta u} \lim_{h \to 0} (s'(x) + E_s(h)) && (7.70)\\
&= g'(u)s'(x) && (7.71)\\
&= g'(s(x))s'(x) && (7.72)
\end{aligned}$$

♢ SDG ♢

7.4 Using the Chain Rule

7.4.1 Derivative of $x^{(1/n)}$

Up till now we have been using the fact that the square root of x falls under the power rule, but we haven't proven it. We know have the tools to be able to show this. We will not prove this directly, but rather we will use the chain rule to find it. Start by defining $f(x)$.

$$f(x) = x^{\left(\frac{1}{n}\right)}$$

[5] Most algebraic tricks in proof can be reduced to multiplying by 1 or adding 0. Of course we choose a cleaver way to express 1 or 0, such as $f(x) - f(x) = 0$ or $f(x)/f(x) = 1$.

Raise both sides to the nth power.

$$f(x)^n = x$$

Now take the derivative of both sides and use the Chain Rule on the left side.

$$\frac{d}{dx}f(x)^n = 1$$

We define the intermediate terms

$$\begin{aligned} u &= f(x) \\ g(u) &= u^n \\ g'(u) &= nu^{n-1} \end{aligned}$$

Thus we have

$$\begin{aligned} \frac{d}{dx}f(x)^n &= 1 \\ n(f(x))^{n-1}f'(x) &= \\ n(x^{\left(\frac{1}{n}\right)})^{n-1}f'(x) &= \\ nx^{\left(\frac{n-1}{n}\right)}f'(x) &= \end{aligned}$$

Now solve for $f'(x)$, which is what we wanted.

$$f'(x) = \frac{1}{n}x^{\frac{1-n}{n}}$$

We note that indeed fractional powers do obey the power rule.

7.5 Derivative of $log_a(x)$

Since we have been looking at logs it would be nice to find their derivatives.

Theorem 15 (Derivative of Logarithms)

$$\frac{d}{dx}\log_a(x) = \frac{1}{x\ln(a)}$$

Proof:

Note that there is no nice way to rewrite $\log_a(x+h)$ as a function of x and a function of h. We think of the trick we just used and maybe we can use it here. Recall:

$$a^{\log_a x} = x$$

Thus

$$\frac{d}{dx}a^{\log_a x} = 1$$

Now we can use the chain rule on this. The inner function is $\log_a(x)$ and the outer function is a^x. For ease of notation we will define $f(x) = \log_a(x)$ and $g(u) = a^u$, thus $g'(u) = \ln(a)a^u$. Then we find that

$$\begin{aligned} 1 &= \frac{d}{dx} a^{\log_a x} \\ &= g'(f(x))f'(x) \\ &= \ln(a)a^{\log_a(x)} f'(x) \\ &= \ln(a)xf'(x) \end{aligned}$$

We can then find the derivative of $\log_a(x)$ by solving for $f'(x)$.

$$\frac{d}{dx}\log_a(x) = f'(x) = \frac{1}{\ln(a)x}$$

♢ SDG ♢

Note that we usually only deal with the derivative of ln(x) because in that case we have:

Corollary 16 (Derivative of the Natural Logarithms)

$$\frac{d}{dx}\log_a(x) = \frac{1}{x}$$

Note also that we can rewrite any base of log into another, so we could just use the derivative for $\ln(x)$ to find them all.

7.6 Trigonometric Derivatives

7.6.1 Derivative at Zero

Example 16

$$\begin{aligned} \frac{d}{dx}(\sin(x))_{x=0} &= \lim_{h\to 0} \frac{\sin(0+h)-\sin(0)}{h} && (7.73) \\ &= \lim_{h\to 0} \frac{\sin(h)}{h} && (7.74) \\ &= 1 && (7.75) \end{aligned}$$

Example 17

$$\begin{aligned} \frac{d}{dx}(\cos(x))_{x=0} &= \lim_{h\to 0} \frac{\cos(0+h)-\cos(0)}{h} && (7.76) \\ &= \lim_{h\to 0} \frac{\cos(h)-1}{h} && (7.77) \\ &= 0 && (7.78) \end{aligned}$$

7.6.2 General Derivative

Theorem 17

$$\frac{d}{dx}\sin(x) = \cos(x) \tag{7.79}$$

Proof:

$$\begin{aligned}
\frac{d}{dx}(\sin(x)) &= \lim_{h\to 0}\frac{\sin(x+h)-\sin(x)}{h} &(7.80)\\
&= \lim_{h\to 0}\frac{\cos(h)\sin(x)+\sin(h)\cos(x)-\sin(x)}{h} &(7.81)\\
&= \lim_{h\to 0}\frac{\sin(h)\cos(x)-\sin(x)(1-\cos(h))}{h} &(7.82)\\
&= \lim_{h\to 0}\frac{\sin(h)}{h}\cos(x)-\lim_{h\to 0}\frac{(1-\cos(h))}{h}\sin(x) &(7.83)\\
&= 1\cos(x)+0\sin(x) &(7.84)\\
&= \cos(x) &(7.85)
\end{aligned}$$

♢ SDG ♢

Theorem 18

$$\frac{d}{dx}\cos(x) = -\sin(x) \tag{7.86}$$

Proof:

$$\begin{aligned}
\frac{d}{dx}(\cos(x)) &= \lim_{h\to 0}\frac{\cos(x+h)-\cos(x)}{h} &(7.87)\\
&= \lim_{h\to 0}\frac{\cos(h)\cos(x)-\sin(h)\sin(x)-\cos(x)}{h} &(7.88)\\
&= \lim_{h\to 0}\frac{-\sin(h)\sin(x)-\cos(x)(1-\cos(h))}{h} &(7.89)\\
&= -\lim_{h\to 0}\frac{\sin(h)}{h}\sin(x)-\lim_{h\to 0}\frac{(1-\cos(h))}{h}\cos(x) &(7.90)\\
&= -1\sin(x)+0\cos(x) &(7.91)\\
&= -\sin(x) &(7.92)
\end{aligned}$$

♢ SDG ♢

Theorem 19

$$\frac{d}{dx}\tan(x) = \frac{1}{\cos(x)^2} \tag{7.93}$$

Proof:

$$\frac{d}{dx}(\tan(x)) = \frac{d}{dx}\left(\frac{\sin(x)}{\cos(x)}\right) \tag{7.94}$$
$$= \frac{\frac{d}{dx}(\sin(x))\cos(x) - \sin(x)\frac{d}{dx}(\cos(x))}{\cos(x)^2} \tag{7.95}$$
$$= \frac{(\cos(x))\cos(x) - \sin(x)(-\sin(x))}{\cos(x)^2} \tag{7.96}$$
$$= \frac{1}{\cos(x)^2} \tag{7.97}$$

$\diamondsuit$ SDG $\diamondsuit$

7.7 Derivatives of the inverse trigs

One final thing we would like to find the derivative of is the inverse trig functions. Akin to what we have been doing we will use the chain rule to solve for them.

Theorem 20

$$\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2} \tag{7.98}$$

Proof:

Let $f(x) = \arctan(x)$ and $g(u) = \tan(u)$, so that $\tan'(x) = \frac{1}{\cos(x)^2}$. We will use the identity that $\tan(\arctan(x)) = x$ when x lies in $-\pi/2$ to $\pi/2$.

$$\begin{aligned} 1 &= \frac{d}{dx}\tan(\arctan(x)) \\ &= \frac{1}{\cos^2(\arctan(x))} f'(x) \\ f'(x) &= \cos^2(\arctan(x)) \end{aligned}$$

While this is an answer it is not a pretty one. We then recall the identity $1 + \tan^2(x) = 1/\cos^2(x)$, and use it to rewrite

$$\begin{aligned} \frac{d}{dx}\arctan(x) &= \cos^2(\arctan(x)) \\ &= \frac{1}{1+\tan^2(\arctan(x))} \\ &= \frac{1}{1+x^2} \end{aligned}$$

♢ SDG ♢

Theorem 21

$$\frac{d}{dx}\arccos(x) = \frac{-1}{\sqrt{1-x^2}} \quad (7.99)$$

Proof:

Let $f(x) = \arccos(x)$ and $g(u) = \cos(u)$, so $g'(u) = -\sin(u)$. We will use the identity that $\cos(\arccos(x)) = x$ when x lies in 0 to π.

$$\begin{aligned} 1 &= \frac{d}{dx}\cos(\arccos(x)) \\ &= -\sin(\arccos(x))f'(x) \\ f'(x) &= \frac{-1}{\sin(\arccos(x))} \end{aligned}$$

This is another not-so-pretty. We then recall the identity $\sin^2(x) + \cos^2(x) = 1$, thus $\sin(x) = \sqrt{1-\cos^2(x)}$. We can rewrite the expression as

$$\begin{aligned} \frac{d}{dx}\arccos(x) &= \frac{-1}{\sin(\arccos(x))} \\ &= \frac{-1}{\sqrt{1-\cos^2(\arccos(x))}} \\ &= \frac{-1}{\sqrt{1-x^2}} \end{aligned}$$

♢ SDG ♢

Chapter 8

Implicit Functions

Mathematicians have many ways of breaking functions into general categories, but one of them is particularly important to us at the moment. Namely we are interested in what it means for a function to be explicit or implicit. If I said, "Tell me explicitly what you did." What would I mean? Explicit means direct and exact with nothing hidden or left out. The mathematical meaning is similar. An explicit function is one which is in the form

$$y = f(x).$$

The relation is thus explicitly stated. This is the predominate method used thus far. This is not the most general way of expressing a function. We can also express a function implicitly. An implicit function is one for which the relation is not directly stated it is thus in the form

$$f(x, y) = g(x, y).$$

Implicitly defined functions cannot always be expressed implicitly, so it would be useful to know how to take the derivative of an implicit function.

8.1 Implicit Differentiation

The main idea of implicit differentiation is to treat the dependent variable, say y, as a function of x. In explicit differentiation we find dy/dx by evaluating $df(x)/dx$. For implicit functions we will not have y so easy to access, but we can still take the derivative of both sides of the expression. In implicit differentiation, we then need to use algebra to find dy/dx, which will now be a function of both x and y. Lets consider an example. Consider the following

$$\ln(x) + \ln(y^3) = 5xy - 5.$$

We want to find dy/dx. To do this we take the derivative with respect to x of both sides and solve for dy/dx.

$$\frac{d}{dx}\left(\ln(x)+\ln(y^3)\right) = \frac{d}{dx}(5xy-5) \tag{8.1}$$

$$\frac{1}{x}+\frac{1}{y^3}(3y^2)\frac{dy}{dx} = 5y+5x\frac{dy}{dx} \tag{8.2}$$

$$\frac{1}{y^3}(3y^2)\frac{dy}{dx}-5x\frac{dy}{dx} = 5y-\frac{1}{x} \tag{8.3}$$

$$\left(\frac{1}{y^3}(3y^2)-5x\right)\frac{dy}{dx} = 5y-\frac{1}{x} \tag{8.4}$$

$$\frac{dy}{dx} = \frac{5y-\frac{1}{x}}{\frac{1}{y^3}(3y^2)-5x} \tag{8.5}$$

$$\frac{dy}{dx} = \frac{5xy^2-y}{3x-5x^2y} \tag{8.6}$$

We can thus find the rate y is changing with respect to x without even finding an explicit function which defines y in terms of x. It is thus reasonable to ask what the tangent line to an implicit function is at a particular point.

To find a tangent line to a graph we need to know two things, the point where we want the tangent line, and the slope (derivative) of the line at the point. Say that we want to find the tangent line at some point $(x0, y0)$. Then say we found the slope, $dy/dx = h(x, y)$ (how y changes with respect to x). We could then express the tangent line in point-slope form as

$$k(x) = y_0 + h(x_0, y_0)(x - x_0)$$

Lets find the tangent line to $\ln(x) + \ln(y^3) = 5xy - 5$ at the point (1,1).

$$k(x) = y_0 + h(x_0, y_0)(x - x_0) \tag{8.7}$$

$$= y_0 + \frac{5x_0y_0^2-y_0}{3x_0-5x_0^2y_0}(x-x_0) \tag{8.8}$$

$$= 1 + \frac{5-1}{3-5}(x-1) \tag{8.9}$$

$$= 1 - 2(x-1) \tag{8.10}$$

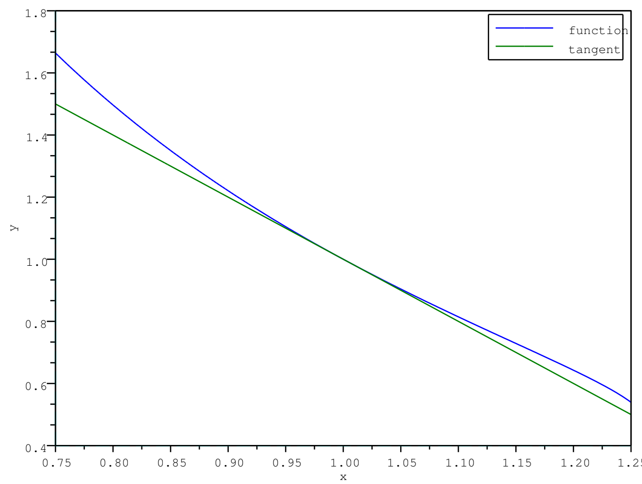
$$= 3 - 2x \tag{8.11}$$

Note that since the tangent line is close to the actual function for x close to 1 we can use this nice linear equation to approximate the more difficult non-linear one. See figure 8.1.

8.2 Inverse Functions

"There And Back Again, A Hobbit's Holiday", that is the name Bilbo gave to his memoirs in the book "The Hobbit". That is also a good way to think of inverse functions. A normal

Figure 8.1: Function $\ln(x) + \ln(y^3) = 5xy - 5$ and the tangent line $3 - 2x$.



function takes its domain and sends it there, to the range. An inverse function takes the range and maps it back again to the domain. In particular an inverse function is defined by

$$f^{-1}(d) = t \Leftrightarrow f(t) = d.$$

Note in particular, we defined the log base a in terms of our inverse notation.

$$\log_a(d) = t \Leftrightarrow a^t = d$$

We can thus note that since the inverse maps a function back, it must be true that

1. $f^{-1}(f(t)) = t$
2. $f(f^{-1}(d)) = d$

We showed this in particular for log base a, and used these forms to establish the other properties.

8.2.1 Derivative of a logarithm

We can use this to calculate the derivative of a logarithm. Note that

$$a^{\log_a(x)} = x. \tag{8.12}$$

Now take the derivative of this with respect to x.

$$\frac{d}{dx}a^{\log_a(x)} = \frac{d}{dx}x \tag{8.13}$$

$$\ln(a)a^{\log_a(x)}\frac{d}{dx}\log_a(x) = 1 \tag{8.14}$$

$$\ln(a)x\frac{d}{dx}\log_a(x) = 1 \tag{8.15}$$

$$\frac{d}{dx}\log_a(x) = \frac{1}{x\ln(a)} \tag{8.16}$$

The above proves what we showed earlier, but using a different method. There are often many ways to prove something, and each way provides insight into the problem.

8.3 One Function's Domain

Since the inverse is defined as the function which takes an element in the range and maps it back to the element in the domain of the original function, which is associated with the element in the range. In other words the range of the original function is the domain of the inverse function, and the domain of the original function is the range of the inverse.

8.4 An Inverse Or Not An Inverse

That is the question. Does an inverse always exist? Sadly, not every function has an inverse. Recall that to be a function every element in the domain had to have a unique element in the range which was associated with it. Well since the inverse is the function which reverses this we will need each value in the range of the original function to have one and only one element in the domain associated with it. The graphical test for a function was the vertical line test. In similar style the graphical test for an inverse is the horizontal line test. Many important functions do not have true inverses because of this, but the original function is so important that we make a sliced inverse. Consider the inverse of x^2. Technically, it does not exist, but x^2 is so important we need to define the square root, so we pick only the positive branch. By picking only part we can now define an "inverse", which is functional though not entirely true. A similar thing must be done for the trig functions, whose "inverses" are defined on a small region.

I would be remiss if I did not mention that if a function is strictly increasing or decreasing then trivially an inverse exists. Note that a function can be mixed if it is piecewise defined.

8.5 Here Inverse, Inverse

So how do we find inverses? Well we saw one way when we defined the logarithm. Another method that often comes up is strait algebraic solution. Thus if $y = f(x)$ we try to solve for x in terms of y using the rules of algebra. This is not always an easy thing (or even doable). One case exists where it is easy to find the inverse though. Graphically we can find the inverse by flipping the original function around the line $y = x$. We could also find the inverse numerically.

Chapter 9

L'Hopital's Rule

9.1 A Problem With Limits

While we all love taking limits, sometimes they can be difficult. Consider again one of our favorites

$$\lim_{x \to 0} \frac{\sin(x)}{x} \tag{9.1}$$

We can numerically approximate it and find it is tending towards 1. This is still not pleasing though.

We can graphically find the derivative of $sin(x)$ at 0 to be 1. We then note by the definition of the derivative at a point that:

$$\begin{aligned} f(x) &= \sin(x) \\ f'(0) &= \lim_{h \to 0} \frac{\sin(0+h) - \sin(0)}{h} \\ &= \lim_{h \to 0} \frac{\sin(h) - 0}{h} \\ &= \lim_{h \to 0} \frac{\sin(h)}{h} \end{aligned}$$

Since we know $f'(0) = 1$, this proves the result we were seeking. It gets us thinking though, could there be a way to do this more generally

9.2 Local Linearization

To start off we want to recall what the local linearization (tangent line approximation) of a function is.

$$f(x) \approx f(a) + f'(a)(x-a) \tag{9.2}$$

At the point, a, the slopes of the two are exact, thus we are able to use the two forms interchangeably when we are at the point. This was implicitly what we were doing when we found the slope of $\sin(x)$ at zero.

9.3 A Clever Idea

This sparks an idea, since the local linearization behaves like the real function at the point, I could use it in place of the original function in the evaluation of the limit. Thus for the function we have been considering we would have

$$\begin{aligned}\lim_{x\to 0}\frac{\sin(x)}{x} &= \lim_{x\to 0}\frac{\sin(0)+cos(0)(x-0)}{x}\\ &= \lim_{x\to 0}\frac{0+1(x)}{x}\\ &= \lim_{x\to 0}\frac{x}{x}n = 1\end{aligned}$$

It works here but does it work in general?

9.4 Generalizing Our Idea

Note that the denominator was already in a "local linear form", namely because it is linear all the time. Additionally the value of the functions at the point we were interested in was zero ($f(a) = 0 = g(a)$). It ended up that the limit of the ratio was the ratio of the derivatives. Let's see if that works.

$$\begin{aligned}\frac{f'(a)}{g'(a)} &= \frac{\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}}{\lim_{h\to 0}\frac{g(a+h)-g(a)}{h}}\\ &= \frac{\lim_{h\to 0}\frac{f(a+h)}{h}}{\lim_{h\to 0}\frac{g(a+h)}{h}}\\ &= \lim_{h\to 0}\frac{f(a+h)}{g(a+h)}\\ &= \frac{f(a)}{g(a)}\end{aligned}$$

It works nicely! But what if both derivatives were zero, we would be back to square one. Or would we?

9.5 L'Hopital's Rule

The general form of L'Hopital's rule can be stated as follows. If $f(a) = g(a) = 0$ then

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \lim_{x\to a}\frac{f'(x)}{g'(x)} \tag{9.3}$$

It could easily be shown by taking limit as x approaches a of the previous argument instead of taking it at the point a. What is the benefit? Well we can take L'Hopital's rule multiple times to evaluate a difficult expression. Consider the following

That works great! Can it get any nicer than this?

9.6 Extending L'Hopital's Rule

We have considered L'Hopital's rule from a perspective of problems involving 0/0. But we also could have considered the case of ∞/∞ . We would have found that L'Hopital's rule works for

$$\lim_{x \to a} f(x) = \pm\infty \tag{9.4}$$

$$\lim_{x \to a} g(x) = \pm\infty \tag{9.5}$$

And even when $a = \pm\infty$. This idea can help us when we are considering what functions dominate other functions as x tends to infinity. Consider the following.

Example 18

$$\lim_{x \to \infty} \frac{x^2 + 2}{x^3 + 2} \tag{9.6}$$

$$= \lim_{x \to \infty} \frac{2x}{3x^2} \tag{9.7}$$

$$= \lim_{x \to \infty} \frac{2}{6x} \tag{9.8}$$

$$= 0 \tag{9.9}$$

Example 19

$$\lim_{x \to \infty} \frac{x^2 + 2}{2x^2 + x} \tag{9.10}$$

$$= \lim_{x \to \infty} \frac{2x}{4x + 1} \tag{9.11}$$

$$= \lim_{x \to \infty} \frac{2}{4} \tag{9.12}$$

$$= \frac{1}{2} \tag{9.13}$$

Note the last example demonstrates the idea behind our Big-O notation.

Part III

Matrix Algebra

Chapter 10

Introduction to Matrices

10.1 Why We Need Them

Consider that we have to find the intersection of two lines. We need to find both the x and y coordinate of the intersection. The equations of the lines are

$$y = m_0 x + b_0 \tag{10.1}$$
$$y = m_1 x + b_1 \tag{10.2}$$

Since both x and y need to be solved for we should rearrange so the constants are on one side and the variables on the other.

$$y - m_0 x = b_0 \tag{10.3}$$
$$y - m_1 x = b_1 \tag{10.4}$$

We can put this into matrix notation as

$$\begin{bmatrix} -m_0 & 1 \\ -m_1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \tag{10.5}$$

As an interesting[1] side note the rows of the constant matrix on the right are the vectors that are perpendicular to the lines we are considering. This can be easily seen by considering[2] the case where the intercepts are zero, $b_i = 0$ for $i \in \{0, 1\}$, then you have $c^T v = 0$, where c^T is a row of the constant matrix, and v is the variable vector. Trivially, we can see that this is the inner product and thus shows they are orthogonal. Because of this we know the

[1]What is interesting will change in your life. The more you know a topic, usually the more you like it. The more you like a topic the more interesting you find it. If it is not interesting now, keep working, you will like it more eventually.

[2]without loss of generality (wlog) - which basically means this does not remove the generality of the statement, though it usually reduces the work significantly

slope of the line perpendicular to another line has a slope of

$$slope = \frac{rise}{run} \tag{10.6}$$

$$= \frac{1}{-m} \tag{10.7}$$

Note how easy it was to show this in matrix notation. Sometime go back and look at what you must do in your early geometry classes.

Anyway back to our original problem.To find the point of intersection, we just need the value of x and y that works for the two equations. There exists a large number of solutions because the problem of finding a vector $\vec{x}$ that satisfies the matrix equation $A\vec{x} = \vec{b}$ is very important, and our problem is in this form with

$$A = \begin{bmatrix} -m_0 & 1 \\ -m_1 & 1 \end{bmatrix} \tag{10.8}$$

$$\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{10.9}$$

$$\vec{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \tag{10.10}$$

One well know solution technique is Gaussian elimination[3]. The idea is to cause the lower left entry of the A matrix to go to zero (this is done by adding a scalar multiple of the top row). The bottom row gives an expression for y and this can then be substituted back into the top row to find x. Implementing these steps, we first zero out the lower left element of A by adding $-\frac{-m_1}{-m_0}$ of the first row of A and$\vec{b}$ to the second rows respectively.

$$\begin{bmatrix} -m_0 & 1 \\ -m_1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \tag{10.11}$$

$$\begin{bmatrix} -m_0 & 1 \\ -m_1 - \frac{-m_1}{-m_0}(-m_0) & 1 - \frac{-m_1}{-m_0}(1) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 - \frac{-m_1}{-m_0}(b_0) \end{bmatrix} \tag{10.12}$$

$$\begin{bmatrix} -m_0 & 1 \\ 0 & 1 - \frac{-m_1}{-m_0} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 - \frac{-m_1}{-m_0}(b_0) \end{bmatrix} \tag{10.13}$$

Now the second row tells us that

$$\left(1 - \frac{-m_1}{-m_0}\right) y = b_1 - \frac{-m_1}{-m_0}(b_0) \tag{10.14}$$

$$y = \frac{b_1 - \frac{-m_1}{-m_0}(b_0)}{1 - \frac{-m_1}{-m_0}} \tag{10.15}$$

$$= \frac{-m_0 b_1 + m_1 b_0}{-m_0 + m_1} \tag{10.16}$$

[3]See Keith On Numerical Analysis for more techniques.

This is then substituted into the first row

$$
\begin{aligned}
y - m_0 x &= b_0 && (10.17)\\
\frac{-m_0 b_1 + m_1 b_0}{-m_0 + m_1} - m_0 x &= b_0 && (10.18)\\
-m_0 x &= b_0 - \frac{-m_0 b_1 + m_1 b_0}{-m_0 + m_1} && (10.19)\\
x &= \frac{b_0 - \frac{-m_0 b_1 + m_1 b_0}{-m_0 + m_1}}{-m_0} && (10.20)\\
&= \frac{b_0 \frac{-m_0 + m_1}{-m_0 + m_1} - \frac{-m_0 b_1 + m_1 b_0}{-m_0 + m_1}}{-m_0} && (10.21)\\
&= \frac{\frac{-m_0 b_0 + m_1 b_0 + m_0 b_1 - m_1 b_0}{-m_0 + m_1}}{-m_0} && (10.22)\\
&= \frac{\frac{-m_0 b_0 + m_0 b_1}{-m_0 + m_1}}{-m_0} && (10.23)\\
&= \frac{-m_0 \frac{b_0 - b_1}{-m_0 + m_1}}{-m_0} && (10.24)\\
&= \frac{b_0 - b_1}{-m_0 + m_1} && (10.25)
\end{aligned}
$$

Here I have done quite a bit of simplifying, mainly because I like to on such simple problems. In a real problem I would just do it numerically with a matrix algorithm, and I would get a number as an answer, rather than an expression.

Consider the following example.

$$
\begin{aligned}
y &= 3x - 1 && (10.26)\\
y &= -x + 3 && (10.27)
\end{aligned}
$$

We can stick these values into our equations

$$
\begin{aligned}
x &= \frac{b_0 - b_1}{-m_0 + m_1} && (10.28)\\
y &= \frac{-m_0 b_1 + m_1 b_0}{-m_0 + m_1} && (10.29)\\
x &= \frac{-1 - 3}{-3 - 1} && (10.30)\\
y &= \frac{-3 \cdot 3 + 1}{-3 - 1} && (10.31)\\
x &= \frac{-4}{-4} && (10.32)\\
y &= \frac{-8}{-4} && (10.33)\\
x &= 1 && (10.34)\\
y &= 2 && (10.35)
\end{aligned}
$$

Which is indeed the answer (check it yourself by substituting it back into the two original problems.

10.2 Vector Spaces

10.2.1 Dependence and Independence

A group of vectors v_i for $i \in \{0, 1, \cdots, n\}$ is independent if

$$\begin{aligned} \sum_{i=0}^{n} \alpha_i v_i &= 0 &(10.36)\\ \iff \alpha_i &= 0, \quad \forall i \in \{0, 1, \cdots, n\} &(10.37) \end{aligned}$$

otherwise they are dependent. Consider an example

Example 20 *Are the following vectors independent or dependent?*

$$a = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \qquad b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \qquad c = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} \tag{10.38}$$

Answer
Consider

$$\begin{aligned} 0 &= \alpha_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \alpha_2 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + \alpha_3 \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} &(10.39)\\ &= \begin{bmatrix} 1\alpha_1 + 4\alpha_2 + 7\alpha_3 \\ 2\alpha_1 + 5\alpha_2 + 8\alpha_3 \\ 3\alpha_1 + 6\alpha_2 + 9\alpha_3 \end{bmatrix}. &(10.40) \end{aligned}$$

Thus from the first row $\alpha_1 = -4\alpha_2 - 7\alpha_3$ and so we have (ignoring the first row)

$$\begin{aligned} 0 &= \begin{bmatrix} 2(-4\alpha_2 - 7\alpha_3) + 5\alpha_2 + 8\alpha_3 \\ 3(-4\alpha_2 - 7\alpha_3) + 6\alpha_2 + 9\alpha_3 \end{bmatrix} &(10.41)\\ &= \begin{bmatrix} -3\alpha_2 - 6\alpha_3 \\ -6\alpha_2 - 12\alpha_3 \end{bmatrix}. &(10.42) \end{aligned}$$

From the first row of this matrix (old second row), $\alpha_2 = -2\alpha_3$. Putting this in the final row we find that $-6(-2\alpha_3) - 12\alpha_3 = 0\alpha_3 = 0$. This is true for all finite values of α_3, and thus there are non-zero values of α_i for which the equality holds and they are dependent. For instance let $\alpha_3 = 1$, thus $\alpha_2 = -2(1) = -2$ and $\alpha_1 = -4(-2) - 7(1) = 1$. Note that

$$\begin{aligned} 0 &= 1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - 2 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + 1 \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} &(10.43)\\ &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -8 \\ -10 \\ -12 \end{bmatrix} + \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}. &(10.44) \end{aligned}$$

Example 21 *What is required for two, 2-dimensional vectors to be independent?*

Answer

Let our two vectors be $[a \quad b]^T$ *and* $[c \quad d]^T$. *Now consider the equation for independence.*

$$0 = \alpha_1 \begin{bmatrix} a \\ b \end{bmatrix} + \alpha_2 \begin{bmatrix} c \\ d \end{bmatrix} \tag{10.45}$$

$$= \begin{bmatrix} a\alpha_1 + c\alpha_2 \\ b\alpha_1 + d\alpha_2 \end{bmatrix} \tag{10.46}$$

From the first row either a *or* b *is non-zero, because if both were the vectors would be trivially dependent*[4]. *Assume without loss of generality that* $a \neq 0$ *and thus,* $\alpha_1 = -\alpha_2 \frac{c}{a}$. *Putting this in the second equation we obtain,*

$$0 = -\alpha_2 \frac{bc}{a} + \alpha_2 d \tag{10.47}$$

$$= \alpha_2 \frac{ad - bc}{a}. \tag{10.48}$$

Independence requires $\alpha_2 = 0$*, and thus* $\frac{ad-bc}{a} \neq 0$. *Since the denominator is not zero (shown above), we only need the numerator to be non-zero. Thus if* $ad - bc \neq 0$ *then the vectors are independent. This should look familiar. Recall that the determinant of* $\begin{bmatrix} a & c \\ b & d \end{bmatrix}$ *is* $ad - bc$*, thus our requirement says the vectors are independent if the matrix composed of the vectors is non-singular. Indeed this can be extended to larger dimensions and more vectors.*

10.2.2 Sets and Spaces

10.2.3 Inner Product

A vector space with an inner product is called a Hilbert space. Inner products have a variety of notations, here are a few of the more well known varieties of the inner product of a and b

1. The old math notation $\langle a, b \rangle$
2. For real vectors $a^T b$
3. For complex vectors $a^H b$
4. Componentwise version for real vectors $\sum_{i=1}^{n} a_i b_i$
5. For complex vectors $a^H b$ or $a^* b$
6. Componentwise version for complex vectors $\sum_{i=1}^{n} a_i^* b_i$, where x^* is the complex conjugate (note some people use $\bar{x}$ for this)
7. For geometric views $\|a\| \, \|b\| \cos(\theta)$, where θ is the angle between a and b

[4] They would both be of the form $[0 \quad 1]^T$.

8. For continuous spaces $\int_{-\infty}^{\infty} a(t)b(t)d(t)$
9. For quantum physics (Dirac Notation- bra-ket) $< a|b >$ or the non-condensed form $< a||b >$

Note that this induces a norm (see the next section) by $\|a\|^2 =< a, a >$. Note also that the pythagorean identity is a trivial implication of this.

Example 22 *Let's calculate the inner product of*

$$\begin{bmatrix}1\\2\end{bmatrix} \textit{ and } \begin{bmatrix}-3\\4\end{bmatrix}. \tag{10.49}$$

For real vectors we know the inner product is

$$\begin{bmatrix}1\\2\end{bmatrix}^T \begin{bmatrix}-3\\4\end{bmatrix} = (1)(-3) + (2)(4) \tag{10.50}$$

$$= 5. \tag{10.51}$$

10.2.4 Norms

A vector space with a norm is called a Banach space. A norm is a measure, and there are many different ones that share some common properties. A *semi-norm* has

1. Positive scalability (positive homogeneity) $\|\alpha a\| = |\alpha|\|a\|$
2. Triangle inequality (subadditivity) $\|a + b\| \leqslant \|a\| + \|b\|$

As a result of these two properties, we have that the norm has *positivity*, $\|a\| \geqslant 0$[5].A norm is a semi-norm with

1. $\|a\| = 0 \iff a = 0$

Often people just stick all the properties together as a list, though this is just a matter of presentation, not content.

10.2.5 Orthogonality

10.3 Four Fundamental Spaces

10.4 Planes

A plane is defined by its normal. Literally we are defining it by the null space (or kernel if you prefer). For a plane that passes through the origin with normal vector, $N \in \Re^n$, which has been normalized, $\|N\| = 1$

$$N^T a = 0, \tag{10.52}$$

[5]First we can show that $\|\vec{0}\| = 0$ by noting $0 \cdot a = \vec{0}$ then by positive scalability, $\|\vec{0}\| = \|0 \cdot a\| = |0|\|a\| = 0\|a\| = 0$. Then we can show positivity by $\|\vec{0}\| = \|a - a\| \leqslant \|a\| + \|-a\| = \|a\| + |-1|\|a\| = \|a\| + \|a\| = 2\|a\|$, but that means $0 \leqslant 2\|a\|$ so $\|a\| \geqslant 0$.

where $a \in \Re^n$ is any vector in the plane. What this equation is telling us is that every vector that is orthogonal to the normal vector is in the plane. Now if we want our plane to not go through the origin we can easily handle his by changing the 0 to c where c is the shortest distance from the plane to the origin. This can easily be found since it is in the direction of the normal, and any point in the plane can be substituted into the equation to find c.

Say we wanted to find the angle between two planes. This can easily be found, as it is just the angle between the normals of the planes. We can find this by using the inner product, and noting that since the normals are normalized we have $N_1^T N_2 = \cos(\theta)$. It is trivial to solve this.

Part IV

Number Theory

Chapter 11

Introduction

11.1 Divisions of Number Theory

1. Elementary number theory (only arithmetic)
2. Analytic number theory (calculus and complex analysis)
3. Algebraic number theory (roots of integer polynomials)
4. Geometric number theory (triangular, square, ...)
5. Combinatorial number theory (uses combinatorial ideas)
6. Computational number theory (algorithms)

Chapter 12

Elementary Number Theory

12.1 Modular Arithmetic

12.1.1 Congruence

We say a is congruent to b modulus n when $a - b$ is divisible by n. In mathematical notation, we write $a \equiv b \pmod{n} \quad \Leftrightarrow \quad a - b = kn$ for some integer k. Several important properties of congruence are

1. $a \equiv a \pmod{n}$
2. $a \equiv b \pmod{n} \quad \Rightarrow \quad b \equiv a \pmod{n}$
3. $\{\{a \equiv b \pmod{n}\} \cdot \{b \equiv c \pmod{n}\}\} \quad \Rightarrow \quad a \equiv c \pmod{n}$

Example 23

$$\begin{aligned} 8 &\equiv 29 \pmod{7} \\ 8 - 29 &= -21 \\ &= (-3)7 \end{aligned}$$

$$\begin{aligned} 9 &\equiv -15 \pmod{6} \\ 9 - (-15) &= 24 \\ &= (4)6 \end{aligned}$$

12.1.2 Modulus

Invariably confusion happens with integer division, modulus, and remainder involving negative numbers. The problem arises in the basic definition. For a dividend, $a \in \mathbb{Z}$ and a divisor, $b \in \mathbb{Z}$, the quotient, q and remainder r must satisfy

1. $\{r, q\} \in \mathbb{Z}$,
2. $a = b * q + r$,
3. $|r| < |d|$.

The problem comes with the last requirement, because many choices can be made. The three most justifiable definitions are below[1]

1. Truncate division preserves the magnitudes of the quotient and remainder, independent of the signs of the dividend and divisor. This forces the remainder to have the same sign as the dividend.
2. Floor division forces the remainder to have the same sign as the divisor.
3. Euclidean division defines $r \geqslant 0$ and thus ensures $b \times q \leqslant a$.

Each is defensible.

Truncate

Remainder's definition is based off the definition of integer division. Integer division, a/b, is defined for positive a and b to be the number q such that

1. $b \times q \leqslant a$,
2. $b \times (q + 1) \geqslant a$.

When negative numbers are allowed the following requirement is added

3. $(-a)/b = a/(-b) = -(a/b)$,

still for a and b positive. One could summarize this as:

$$c/d \quad = \quad \mathrm{sgn}(c)\mathrm{sgn}(d)(|c|/|d|)$$

Given we now have quotient or integer division defined we can then define remainder such that

$$\begin{aligned} a &= (a/b)b + (a \text{ rem } b) \\ a \text{ rem } b &= a - (a/b)b. \end{aligned}$$

Note that the sign of the remainder is the same as the dividend because

[1] other definitions exist such as ceiling division and rounding division, but they do not correspond to the what most people think of division for positive numbers. Note, from the requirements nothing says $5/2 = 3r{-}1$ but this is hardly what most people would think of, and thus would probably not be programmed very well.

$$\begin{aligned} a \text{ rem } b &= a - (a/b)b \\ &= \text{sgn}(a)|a| - (\text{sgn}(a)|a|/(\text{sgn}(b)|b|))\text{sgn}(b)|b| \\ &= \text{sgn}(a)|a| - \text{sgn}(a)\text{sgn}(b)\text{sgn}(b)(|a|/|b|)|b| \\ &= \text{sgn}(a)|a| - \text{sgn}(a)(|a|/|b|)|b| \\ &= \text{sgn}(a)(|a| - (|a|/|b|)|b|). \end{aligned}$$

Note that $|a| - (|a|/|b|)|b| \geqslant 0$ from item 1 of the definition above, with equality holding only when b divides a.

Example 24 *Consider the following:*

$$\begin{aligned} 5/2 &= 2 & 5 \text{ rem } 2 &= 1 \\ (-5)/2 &= -2 & (-5) \text{ rem } 2 &= -1 \\ 5/(-2) &= -2 & 5 \text{ rem } (-2) &= 1 \\ (-5)/(-2) &= 2 & (-5) \text{ rem } (-2) &= -1 \end{aligned}$$

12.1.3 Addition

$$\{\{a \equiv b \pmod{n}\} \cdot \{c \equiv d \pmod{n}\}\} \quad \Rightarrow \quad a + c \equiv b + d \pmod{n}$$

12.1.4 Additive Inverse

$$\begin{aligned} a + \bar{a} &\equiv 0 \pmod{n} \\ a + \bar{a} &= kn, \qquad k \in \mathbb{Z} \\ \bar{a} &= kn - a, \qquad k \in \mathbb{Z} \end{aligned}$$

Example 25 *Find the additive inverse(s) of 3 mod 7.*

$$\begin{aligned} \bar{a} &= kn - a, \qquad k \in \mathbb{Z} \\ &= 7k - 3, \qquad k \in \mathbb{Z} \end{aligned}$$

k	$\bar{a}$	$(3 + \bar{a}) \mod 7$
1	*4*	$(3 + 4) \mod 7 = 0$
2	*11*	$(3 + 11) \mod 7 = 0$
3	*18*	$(3 + 18) \mod 7 = 0$
4	*25*	$(3 + 25) \mod 7 = 0$
⋮	⋮	

12.1.5 Multiplication

$\{\{a \equiv b \pmod{n}\} \cdot \{c \equiv d \pmod{n}\}\} \quad \Rightarrow \quad ac \equiv bd \pmod{n}$

12.1.6 Multiplicative Inverse

$$\begin{aligned} a\bar{a} &\equiv 1 \pmod{n} \\ a\bar{a} &= 1 + kn, \qquad k \in \mathbb{Z} \\ \bar{a} &= \frac{1+kn}{a}, \qquad k \in \mathbb{Z} \end{aligned}$$

Let $k_1 + ak_2 = k$ for k_1 and k_2 positive integers.

$$\begin{aligned} \bar{a} &= \frac{1+kn}{a}, \qquad k \in \mathbb{Z} \\ &= \frac{1+k_1n+ak_2n}{a}, \qquad k_1, k_2 \in \mathbb{Z}^+ \\ &= \frac{1+k_1n}{a} + k_2n, \qquad k_1, k_2 \in \mathbb{Z}^+ \end{aligned}$$

We need a to divide $1+k_1n$, which means it divides with no remainder (aka divides evenly). Consider what would happen if $gcd(a,n) = a_1 > 1$, thus $a = a_1a_2$ and $n = a_1n_2$ for a_1, a_2, and n_2 positive integers. If a_1 is a factor of n then it is also a factor of k_1n If a_1 is a factor of k_1n then it cannot be a factor of k_1n+1 (it evenly divides k_1n and k_1n+k_1 but nothing in between).

Now assume $gcd(a,n) = 1$. For a to divide $1+k_1n$ implies $ak_3 = 1+k_1n$ for some positive integer k_3.

Example 26 *Find the multiplicative inverse(s) of 3 mod 7.*

$$\begin{aligned} \bar{a} &= \frac{1+kn}{a}, \qquad k \in \mathbb{Z} \\ &= \frac{1+7k}{3}, \qquad k \in \mathbb{Z} \end{aligned}$$

k	$\bar{a}$	$(3+\bar{a}) \mod 7$
1	$\frac{8}{3}$ *no*	
2	$\frac{15}{3} = 5$	$(3 \times 5) \mod 7 = 1$
3	$\frac{22}{3}$ *no*	
4	$\frac{29}{3}$ *no*	
5	$\frac{36}{3} = 12$	$(3 \times 12) \mod 7 = 1$
$\vdots$	$\vdots$	

Chapter 13

Geometric Number Theory

13.1 Triangular Numbers

13.2 Square Numbers

Lemma 22 *Any odd number, q, when squared, q^2, has the following properties*

1. $q^2 \mod 8 = 1$

2. $q^2 - 1 = 8 \times \left(\sum_{i=1}^{\frac{q-1}{2}} i\right)$

Proof:

Let n be any whole number, then $q = 2n + 1$ is an odd number, so we are interested in

$$\begin{aligned} (2*n+1)^2 - 1 &= (4*n^2 + 4*n + 1) - 1 && (13.1)\\ &= 4(n^2+n) && (13.2)\\ &= 4(n(n+1)) && (13.3)\\ &= 8\frac{n(n+1)}{2} && (13.4) \end{aligned}$$

Since either n or $n+1$ is even the quantity $\frac{n(n+1)}{2}$ is a whole number. Thus we have that an odd number squared minus one can be factored into 8 and $\frac{n(n+1)}{2}$, which means that it is divisible by 8 and $\frac{n(n+1)}{2}$, which is a famous sequence, namely the sum of the first n whole numbers. Since the square of an odd number minus one is divisible by eight, we must have the square of an odd number is congruent to 1 mod eight (by definition).

$\diamondsuit$ SDG $\diamondsuit$

So who cares? I stumbled on this when I was working on a cellular automata problem. I was considering square regions around a point of some radius, say n. Thus the area I had to sum up was a square minus the center, and the square had side length of twice the radius plus one (radius distance in each direction plus the center square). I wanted an easy formula for finding out what this could be if all surrounding squares were full. I started by writing a little piece of code to look at the numbers, and I noticed they were always divisible by eight. I then modified the code to find what the quotient was when I divided the eight out. I noticed the sequence was the famous sum of the first n integers, where n was my radius. This intrigued me even more, so I wrote the algebraic proof above. I then discovered that it was already known, oh well, it was fun discovering again.

Chapter 14

Encryption

14.1 Affine Encryption Program

Affine encryption is one of the simplest methods for doing encryption. Let P_i be the i^{th} character in the plain text message, and let C_i be the corresponding encoded character. Let there be n possible characters to encode, then the basic idea is to pick two numbers (a, b) to encode a message such that $\gcd(a, n) = 1$ (so a has an inverse). No requirement on b is needed if your modulus function has been encoded correctly. The encoded character can then be found by

$$a \times P_i + b \quad = \quad C_i \mod n.$$

Note that the " $\mod n$" at the end says the equation holds in $\mathbb{Z}_n$, the set of integers mod n with appropriately defined arithmetic.

To decrypt the message, the equation

$$\bar{a} \times (C_i + d) \quad = \quad P_i \mod n$$

is used. The term $\bar{a}$ is the inverse of a in $\mathbb{Z}_n$, which is found by solving

$$a \times \bar{a} \quad = \quad 1 \mod n$$

or

$$a \times \bar{a} \quad = \quad m \times n + 1.$$

Note that m is any whole number. The term d is the additive inverse of b in $\mathbb{Z}_n$, which is found by solving

$$d = n - (b \mod n).$$

We can summarize this by saying an affine cipher is an encryption technique that encodes using three integers: a, b, and n. If *plain* is the character to be encoded (with 'A'=0 and

‘Z’=25) then $code = (a * plain + b) \mod n$. Decoding is also done using three integers: c, d, and n. If $code$ is the character to be encoded (with ‘A’=0 and ‘Z’=25) then $plain = (c * (code + d)) \mod n$. The requirements on (a, b, c, d, n) are:

- $\gcd(a, n) = 1$
- $(ac) \mod n = 1$
- $(b + d) \mod n = 0$

Below is C code to implement a particular case of affine cyphers.

```
char affine_encode(char plain){
    // affine codes capital letter in plain using a=5, b=12 thus this is modulo 26
    int iCode, iPlain, a=3,b=0;

    // convert char to integer and shift so A=0
    iPlain=int(plain)-65;

    // do the encoding
    iCode = (a*iPlain+b)%26;

    // return the result as a char
    return char(iCode+65);
}

char affine_decode(char code){
    // affine decodes capital letter in plain using c=21, d=8 thus this is modulo 26
    int iCode, iPlain, c=9, d=0;

    // convert char to integer and shift so A=0
    iCode=int(code)-65;

    // do the decoding
    iPlain = (c*(iCode+d))%26;

    // return the result as a char
    return char(iPlain+65);
}
```

Chapter 15

Joy of Number Theory

15.1 Patterns

I get lots of emails from friends showing me something funny or cool. Recently a friend sent me a series of beauty of math problems. While the problems were neat I had to know why this worked or it was not fun for me. Below are my explanations.

15.1.1 Multiplication By 8

$$\begin{aligned}
1 \cdot 8 + 1 &= 9 \\
12 \cdot 8 + 2 &= 98 \\
123 \cdot 8 + 3 &= 987 \\
1234 \cdot 8 + 4 &= 9876 \\
12345 \cdot 8 + 5 &= 98765 \\
123456 \cdot 8 + 6 &= 987654 \\
1234567 \cdot 8 + 7 &= 9876543 \\
12345678 \cdot 8 + 8 &= 98765432 \\
123456789 \cdot 8 + 9 &= 987654321
\end{aligned}$$

This one can be understood by realizing it is multiplication by 10. Consider the following

$$a \cdot 8 + y = c \tag{15.1}$$

and then

$$a \cdot 8 \cdot 10 + y \cdot 10 = c \cdot 10. \tag{15.2}$$

It is probably easiest to see this by looking at the original pattern multiplied by 10:

$$
\begin{array}{rclcrcl}
1 \cdot 8+1 & = & 9 & & 10 \cdot 8+10 & = & 90 \\
12 \cdot 8+2 & = & 98 & & 120 \cdot 8+20 & = & 980 \\
123 \cdot 8+3 & = & 987 & & 1230 \cdot 8+30 & = & 9870 \\
1234 \cdot 8+4 & = & 9876 & \rightarrow & 12340 \cdot 8+40 & = & 98760 \\
12345 \cdot 8+5 & = & 98765 & & 123450 \cdot 8+50 & = & 987650 \\
123456 \cdot 8+6 & = & 987654 & & 1234560 \cdot 8+60 & = & 9876540 \\
1234567 \cdot 8+7 & = & 9876543 & & 12345670 \cdot 8+70 & = & 98765430 \\
12345678 \cdot 8+8 & = & 98765432 & & 123456780 \cdot 8+80 & = & 987654320
\end{array}
$$

Now note that there are many ways to multiply by 10, one of which is to shift the numbers to the left one place in base 10 representation (i.e. units become 10's, 10's become hundreds, etc.). This can be easily seen in the sequence, $1 \rightarrow 12 \rightarrow 123$, etc. The term $y \cdot 10$ is handled by the units term which is added, for example from the first to second line, the $+1$ in row 1 needs to become a $+10$. The only way to achieve this is by the units digit times 8. Let's formalize this. I will use concatenation to show the digits in a number. For instance, $x0$ means the most significant digits of the number are the same as x, and the ones digit is zero, so if $x = 123$ then $x0 = 1230$. Let $z = y + 1$, then

$$
\begin{aligned}
a \cdot 8 \cdot 10 + y \cdot 10 &= c \cdot 10 && (15.3) \\
a0 \cdot 8 + y \cdot 8 + y \cdot 2 &= c0 && (15.4) \\
a0 \cdot 8 + z \cdot 8 - 8 + y \cdot 2 &= && (15.5) \\
az \cdot 8 - 8 + y \cdot 2 &= c0. && (15.6)
\end{aligned}
$$

Now we let $x = 9 - y$ and add x to both sides giving us

$$
\begin{aligned}
az \cdot 8 - 8 + y \cdot 2 + x &= c0 + x && (15.7) \\
az \cdot 8 - 8 + y \cdot 2 + 9 - y &= cx && (15.8) \\
az \cdot 8 + y \cdot 2 - y + 9 - 8 &= && (15.9) \\
az \cdot 8 + y + 1 &= cx. && (15.10)
\end{aligned}
$$

15.1.2 Multiplication by 9 Part I

$$
\begin{aligned}
1 \cdot 9 + 2 &= 11 \\
12 \cdot 9 + 3 &= 111 \\
123 \cdot 9 + 4 &= 1111 \\
1234 \cdot 9 + 5 &= 11111 \\
12345 \cdot 9 + 6 &= 111111 \\
123456 \cdot 9 + 7 &= 1111111 \\
1234567 \cdot 9 + 8 &= 11111111 \\
12345678 \cdot 9 + 9 &= 111111111 \\
123456789 \cdot 9 + 10 &= 1111111111
\end{aligned}
$$

15.1.3 Multiplication by 9 Part II

$$
\begin{aligned}
9 \cdot 9 + 7 &= 88 \\
98 \cdot 9 + 6 &= 888 \\
987 \cdot 9 + 5 &= 8888 \\
9876 \cdot 9 + 4 &= 88888 \\
98765 \cdot 9 + 3 &= 888888 \\
987654 \cdot 9 + 2 &= 8888888 \\
9876543 \cdot 9 + 1 &= 88888888 \\
98765432 \cdot 9 + 0 &= 888888888
\end{aligned}
$$

15.1.4 Overlapping

$$
\begin{aligned}
1 \cdot 1 &= 1 \\
11 \cdot 11 &= 121 \\
111 \cdot 111 &= 12321 \\
1111 \cdot 1111 &= 1234321 \\
11111 \cdot 11111 &= 123454321 \\
111111 \cdot 111111 &= 12345654321 \\
1111111 \cdot 1111111 &= 1234567654321 \\
11111111 \cdot 11111111 &= 123456787654321 \\
111111111 \cdot 111111111 &= 12345678987654321
\end{aligned}
$$

Part V

Probability & Statistics

Chapter 16

RMSE and Standard Deviation

Let x be an n-vector of "true" measurements and $\hat{x}$ be a vector of estimates of x, then $\tilde{x}$ is the error in the estimates. The mean of the errors is given by μ.

$$\mu = \frac{\sum_{i=1}^{n} \tilde{x}}{n} \tag{16.1}$$

The standard deviation of a sample is given by σ.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (\tilde{x} - \mu)^2}{n}} \tag{16.2}$$

Alternately, the standard deviation of a distribution is also denoted σ.

$$\sigma = \sqrt{E[\tilde{x}^2] - E[\tilde{x}]^2} \tag{16.3}$$

$$\approx \sqrt{MSE - \mu^2} \tag{16.4}$$

The mean squared error (MSE), is given by

$$MSE = \frac{\|\tilde{x}\|_2^2}{n}, \tag{16.5}$$

thus

$$\sigma = \sqrt{MSE - \mu^2} \tag{16.6}$$

$$= \sqrt{\frac{\|\tilde{x}\|_2^2}{n} - \left(\frac{\sum_{i=1}^{n} \tilde{x}}{n}\right)^2} \tag{16.7}$$

To add to the confusion, the square root of the MSE (called the root MSE or RMSE or just RMS) is also called the standard deviation.

So what should be done? In reality all will be fairly close. I would calculate Eq 16.2 and Eq 16.7, note the first should be slightly larger. The following is Scilab code, which you can easily convert for use in Matlab.

```
n=20;

x=rand(n,1,'normal');
mu=0;
for i=1:n
  mu=mu+x(i);
end
mu=mu/n

mse=(norm(x)^2)/n

rms=sqrt(mse)

stddev=sqrt(mse-(mu^2)/n)

sample_stddev=0;
for i=1:n
  sample_stddev=sample_stddev+ (x(i)-mu)^2;
end
sample_stddev=sqrt(sample_stddev/n)
```

16.1 Variance Estimation

For a sample drawn from a distribution, the estimator S^2 is an unbiased estimate of the sample variance σ^2

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \tag{16.8}$$

16.2 Chi-Squared

Consider the following statistic, which has a chi-squared distribution with $n-1$ degrees of freedom:

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \tag{16.9}$$

Chapter 17

Bayes

17.1 Terminology

Upper case will be used to represent events but not necessarily fixed (i.e. it is a variable). Lower case represents fixed events (i.e. a constant).

$P(A)$ Prior/marginal probability of A. We call it prior because it is before you know anything about B.

$P(A|B)$

17.2 Theory

$$\begin{aligned} P(A \cap B) &= P(A|B)P(B) && (17.1)\\ &= P(B|A)P(A) && (17.2) \end{aligned}$$

Relation between probability and likelihood

$$\begin{aligned} P(A|B) &= \alpha L(A|B=b)P(A) && (17.3)\\ &\propto L(A|b) && (17.4) \end{aligned}$$

where α is some constant[1].

17.3 Multiple

$$\begin{aligned} P(A \cap B \cap C) &= P(A|B \cap C)P(B \cap C) && (17.5)\\ &= P(A|B \cap C)P(B|C)P(C) && (17.6) \end{aligned}$$

[1]The value rarely matters as we usually deal with likelihood ratios, so the constant cancels out.

Bibliography