Principles of Mathematics 12

Your Guide To Everything

by Nicholas Kim

Table of Contents

Copyright Notice

Introduction

Morpheus: You take the blue pill, the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill, you stay in Wonderland, and I show you how deep the rabbit hole goes.

Neo: [begins to reach for the red pill]

Morpheus: Remember, all I'm offering is the truth.

– The Matrix

Behold the successor to the Math 12 help articles on my website (http://technetia.ca/): the Math 12 Guide to Everything. (Math 12 is shorthand for "Principles of Mathematics 12", an advanced course in elementary algebra designed for senior students in British Columbia, Canada.)

Despite the title's claims, this "textbook" (if indeed such a document deserves such a title) is *very* minimalist, with few examples and no practice problems offered. It is quite theory-oriented, meaning that it focuses mainly on explaining concepts (for example, why use radians at all?) rather than how to put them into practice. Therefore, I highly recommend that you continue to make use of your own textbook and your teacher's notes.

This textbook is organized almost exactly as my own Math 12 classes were.

- I initially begin with function transformations, since the concept turns out to be quite fundamental in many aspects of other chapters.
- I then begin trigonometry, first starting by reviewing the basics and introducing radian measure, then moving on to graphing trigonometric functions and then solving trigonometric equations.)
- Exponential functions and logarithms are covered next. Geometric sequences and series, while they were also part of this chapter in the Math 12 textbook I used, are included in their own separate chapter here, since [a] the chapter would otherwise grow too long, [b] my Math 12 teacher treated it as such.
- Finally, I look at permutations and combinations, as well as probability.
- Appendix A is a formula sheet, similar to the ones Math 12 students are given on exams.
- Appendix B is a table of special trigonometric angles, for which exact values can be found.
- Appendix C is a review of solving polynomial equations, with a little bit of additional material (the existence of cubic and quartic formulae, the bisection method for *approximating* solutions). It is *not* a full tutorial and should not be treated as such.
- Appendix D is a review of solving systems of linear equations. As with Appendix C, it is not intended to be a full tutorial – the emphasis is on the word *review*.

Let me end by saying that, no matter what you want to use this textbook for, I hope you find it to be interesting (and useful) reading. I would appreciate feedback very much, so please feel free to leave comments on my website. (For the record, I have plans to write a *real* textbook in the future, but that will have to wait.)

– Nicholas Kim

Acknowledgements

People

Surprise! There are no acknowledgements of anyone in this section. While I would love to credit those who had a positive impact on my studies of mathematics, doing so would elevate the status of this document above what it really is – just a simple document that's supposed to be better than the old math help articles on my website. So, nothing will be written here. (Names gladly given on personal request.)

Tools

This document was written on OpenOffice.org (version 3.2) Writer, run on Mac OS X 10.6 ("Snow Leopard").

All graphs were produced with Mac OS X's Grapher tool. Any edits, as well as original images, were made (by me) using The GIMP and Inkscape.

Function Transformations

Eragon: I don't understand.

Brom: Of course you don't. That's why I'm teaching you and not the other way around. Now stop talking or we'll never get anywhere.

– The Inheritance Cycle

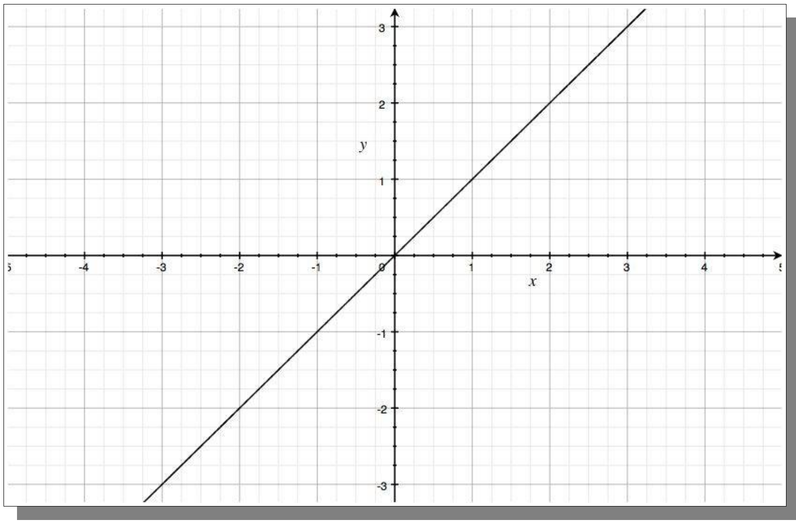
Overview

In previous years, you will undoubtedly have looked at linear and quadratic functions, and studied what tweaking the various constants do. For example, given the linear function $y=mx+b$, m controls the slope ("steepness") of the line, and b controls where the function crosses the y-axis. By tweaking these two modifiers correctly, we can achieve any (non-vertical) line we want.

In Math 12, however, we look at a much larger variety of functions, and memorizing specific details for tweaking every type of function would be tiresome, to put it mildly. A more promising approach is to study some *general* techniques for transforming (the formal word for "tweaking") functions, which is what this beginning chapter of Math 12 looks at. The techniques taught here will be applied throughout the rest of the Math 12 curriculum, so it is very important to master them well.

Some Basic Functions

Before we begin studying transformations, it would be a good idea to build a small library of some "basic" functions and analyze them a bit. This way, we have a foundation to build on when we begin learning about transformations.

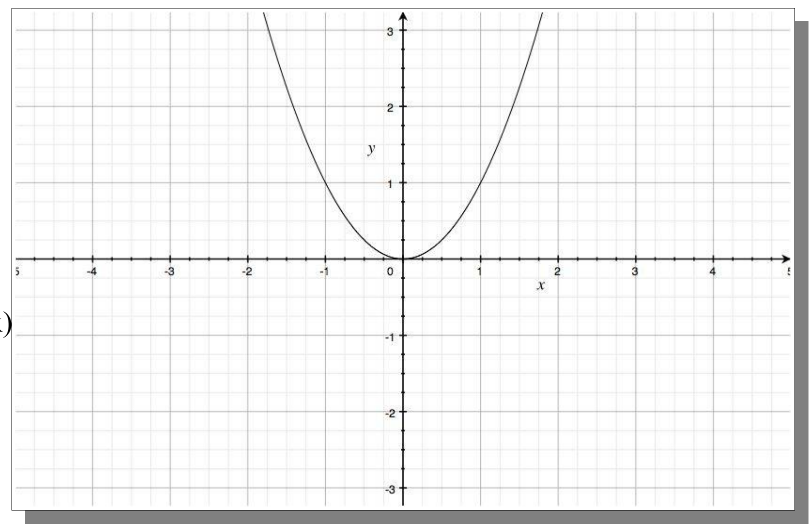


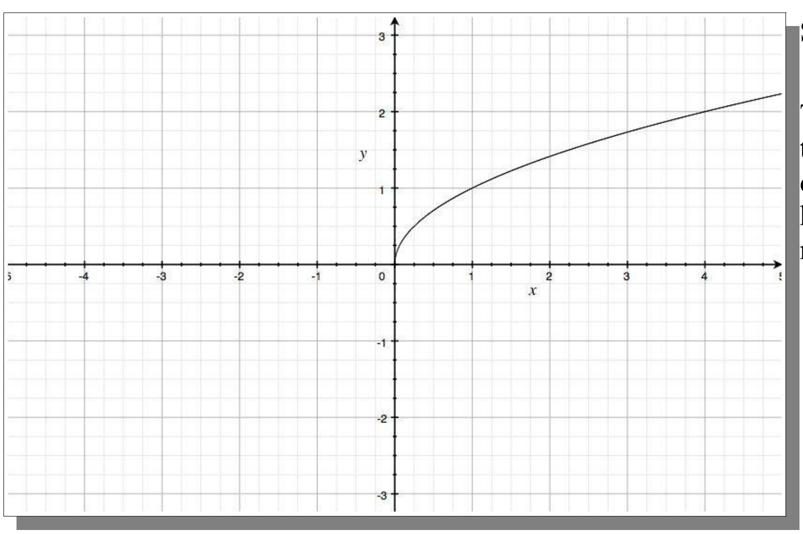
Linear Functions

Arguably the simplest type of function, linear functions are obviously named for the straight line they trace when graphed. The linear function pictured here is y=x, which is also known as the *identity function* (so called because the function will output exactly the same as what you give it). Also, linear functions that can be graphed as a straight horizontal line are called *constant functions*, due to their output being the same regardless of input.

Quadratic Functions

Named after the Latin word for square, quadratic functions trace a parabolic curve when graphed. Unlike linear functions, quadratics have an absolute maximum/minimum point (called the vertex) and need not touch the x-axis at all.



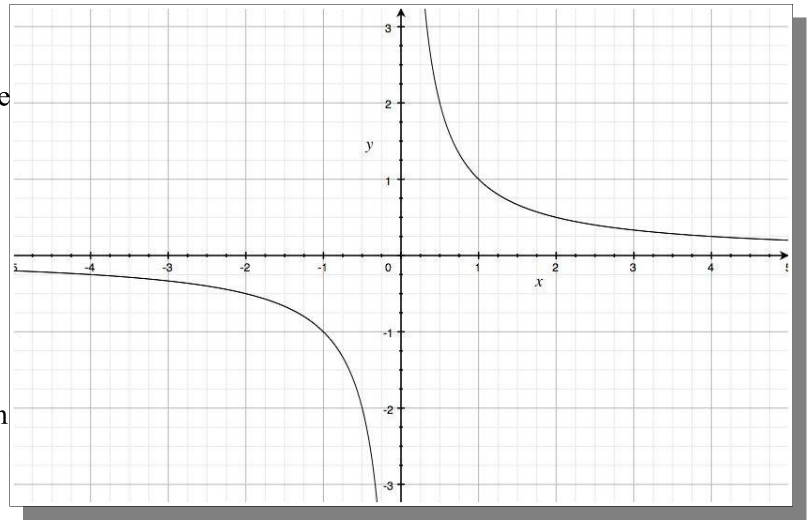


Square Root Functions

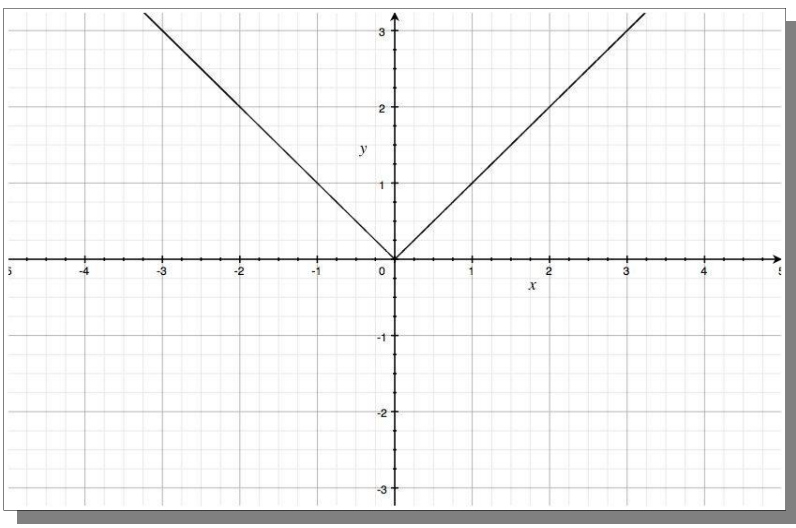
Taking form $y=\sqrt{x}$ (in its most basic form), these can be considered the inverses of quadratic functions. An important point to keep in mind is that the expression under the radical symbol *must* be ≥ 0.

Reciprocal Functions

Not to be confused with inverse functions, reciprocal functions are formed by taking the reciprocal of every y-value in the original function. The example here is $y=1/x$.



If you compare this graph to the graph of $y=x$, notice the vertical asymptote here formed as x approaches 0 (reciprocal functions are not defined where the original has a y-value of 0), and the horizontal asymptote formed as x approaches infinity in either direction.



Absolute Value Functions

In the set of real numbers, the absolute value function takes its argument and makes it positive if it isn't already – geometrically, this means that all parts of the graph below the x-axis are mirrored vertically. The graph shown here is $y=|x|$. Compare to the graph of $y=x$ above.

Translations

Perhaps the simplest type of transformation to perform is a *translation* ("shift"), which basically means to take the graph of a function and move it around – or, more formally, to move every point on the graph x units left/right and y units up/down.

Horizontal translations may be expressed mathematically as

$$y=f(x-k)$$

where k is the number of units the graph has been shifted to the *right*. (If k is negative, it will represent a shift to the left.)

Example: Graph $y=(x+3)^2$.

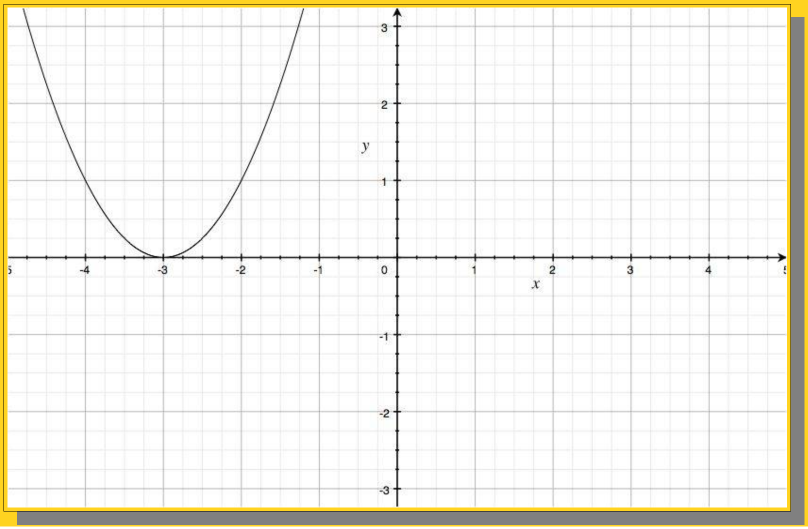
Solution: Since this is equivalent to

$$y=f(x-(-3))$$

where

$f(x)=x^2$

the graph is simply the parabola shifted -3 units right (or +3 units left):



Note carefully that the shift is to the *left*, even though intuitively, we think of a "plus" as moving to the right. **Remember, horizontal translations *k* units right are described by *x-k*, not *x+k*.**

Similarly, vertical translations may be expressed mathematically as

$$y-k=f(x)$$

where k is the number of units the graph has been shifted *up*. However, because it is customary to write the equations of functions with y completely on its own, a vertical shift k units up will usually be written as

$$y=f(x)+k$$

Notice that writing the vertical shift in this way makes it more intuitive than horizontal shifts, because "adding" some value to every y-coordinate logically should cause a shift "up", as expected. **In general, vertical transformations can be interpreted intuitively, whereas horizontal transformations must be examined carefully. This is due to the fact that when we modify *y*, we usually move the modification over to the other side of the equation.**

Performing translations is a simple matter – take every point on the graph and move it a certain number of units vertically and horizontally, depending on what your equation tells you.

Example: Describe the translations performed on *f(x)* given

$$y+3=f(x-1)+2$$

Solution: Combine the separate vertical transformations:

$$y=f(x-1)-1$$

From this we see that we have shifts of 1 right and 1 down.

Reflections

A *reflection* is essentially the mirroring of a graph across a certain line. For the purposes of Math 12, we only consider reflection along three lines: the x-axis, the y-axis, and the line described by the equation *y=x*.

Reflection across the y-axis (horizontal mirroring) is expressed mathematically as

$$y = f(-x)$$

Note that it is the x variable that is modified, not y, even though we are mirroring across the y-axis.

For certain graphs, horizontal mirroring makes no difference in their graphs (such as $y=x^2$). If this is the case, the function describing the graph is said to be an *even* function. Formally, an even function is one where

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

for all x.

(For comparison, an *odd* function is one with point symmetry about the origin – that is, rotating it 180 degrees about the origin will make no difference. Mathematically,

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

for all x.)

Similarly, reflection across the x-axis (vertical mirroring) is expressed mathematically as

$$-y = f(x)$$

or, as we usually write it,

$$y = -f(x)$$

Reflections across the axes are performed by taking every x- (or y-, or perhaps both) coordinate, and taking the negative of it.

Reflection across the line $y=x$ is quite different from reflection across an axis, and is expressed mathematically as taking the inverse of the function. In symbols,

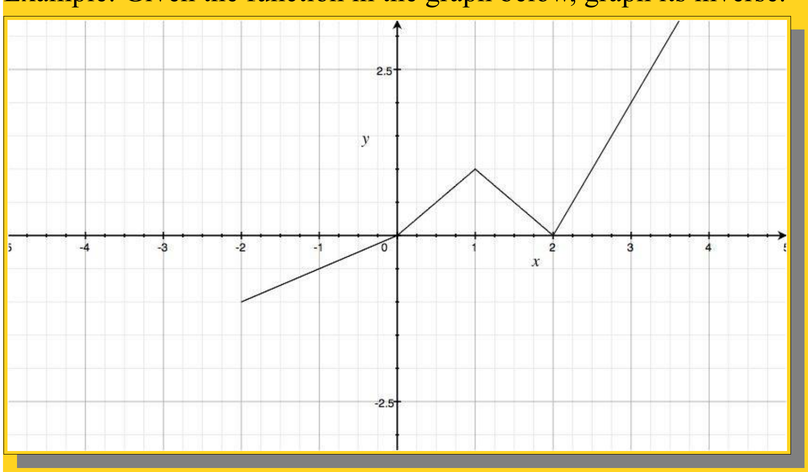
$$x = f(y)$$

or writing y on its own,

$$y = f^{-1}(x)$$

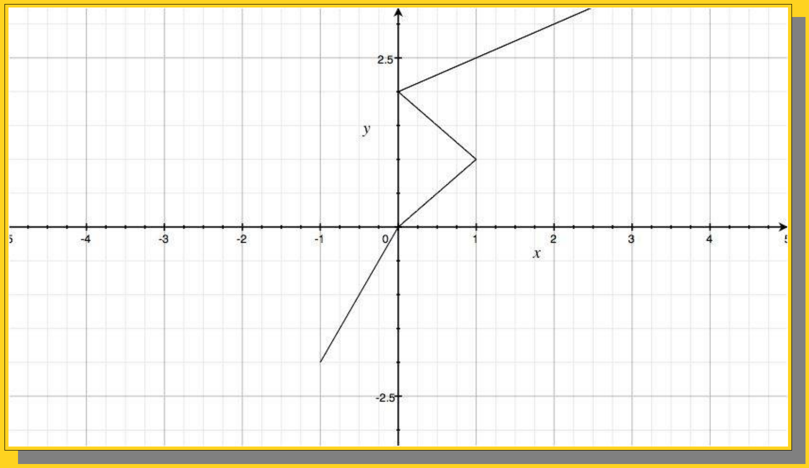
Geometrically, to graph the inverse of a function, take every coordinate pair *(x, y)*, and swap the x- and y-coordinates.

Example: Given the function in the graph below, graph its inverse.



Solution: Locate the “key points” on the graph – usually the points when the function changes slope, but basically whatever points you need to construct the inverse by drawing lines between them. In this case, the points chosen include (-2, -1), (0, 0), (1, 1), (2, 0), and (3, 2).

Now, the inverse will have all (x, y) coordinate pairs swapped, so the points on our new graph include (-1, -2), (0, 0), (1, 1), (0, 2), (2, 3).



Keep in mind that the inverse presented here is not a function, due to it failing the vertical line test. (There exists a test known as the *horizontal line test*, which is a test to see if the inverse of a function is also a function. Clearly, the original fails this test, so this inverse cannot be a function.)

Expansions and Compressions

More generally called scaling, *expansions* and *compressions* are the act of multiplying the x- and/or y-coordinates of every point on a graph by a certain constant factor, effectively “stretching” or “squeezing” the graph.

Horizontal scaling is expressed as

$$y=f(kx)$$

where *1/k* (not *k*, but *1/k*) is the stretching factor.

For example, if we have *y=f(2x)*, the graph is horizontally *compressed* by a factor of 1/2. But if we have *y=f((1/2)x)*, the graph is horizontally *expanded* by a factor of 2.

In general, when *k* is greater than 1, we have a compression, and when *k* is between 0 and 1, we have an expansion. (When *k=1*, nothing changes; when *k=0*, the function reduces to a straight line *y=f(0)*.)

Similarly, vertical scaling is expressed as

$$ky=f(x)$$

or usually as

$$y=\frac{1}{k}f(x)$$

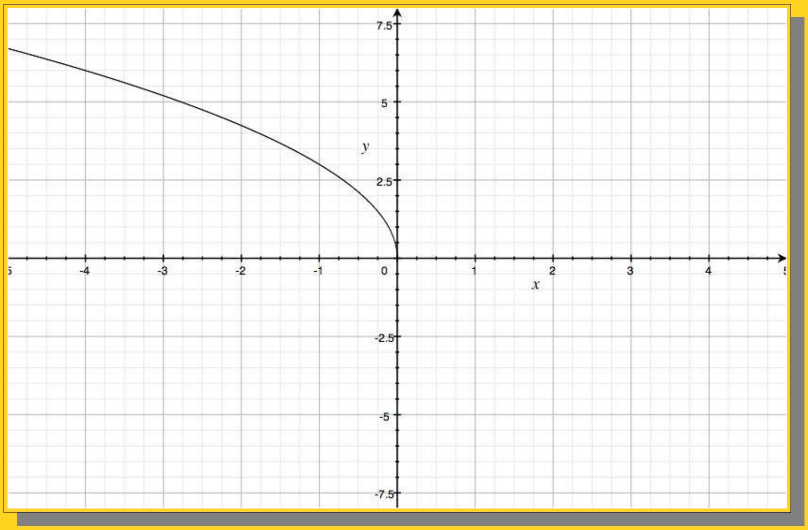
Note that if *k* is negative, it merely ends up being a combination of a reflection and a scale.

Example: Graph the function

$$y=3\sqrt{-x}$$

Solution: Since this is a vertical transformation, we may interpret it literally as a vertical expansion by a factor of

3. Also note the reflection across the y-axis.



Special Transformations

Two last transformations worthy of note are the *absolute value* and the *reciprocal* transformations.

An absolute value transformation is expressed as

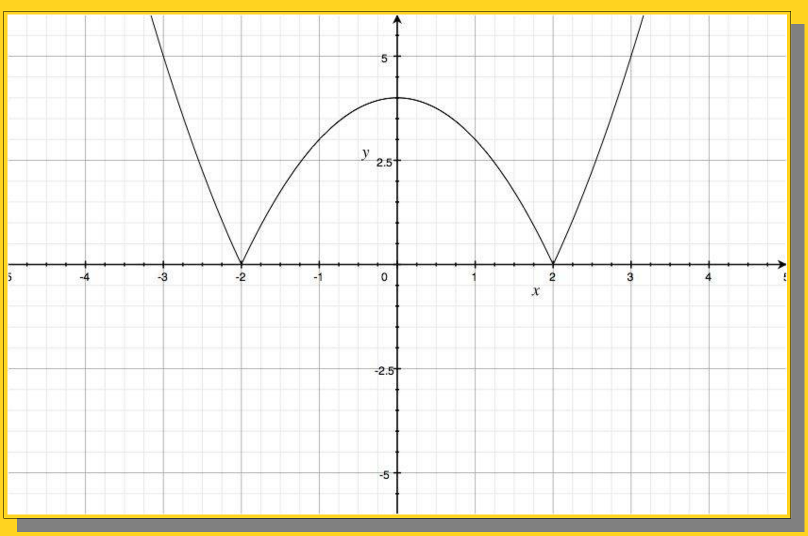
$$y=|f(x)|$$

and is performed by taking all parts of the graph below the x-axis and mirroring them above it.

Example: Graph the function

$$y=|x^2-4|$$

Solution: $x^2 - 4 < 0$ when x is between -2 and 2 (can you see why?), so we must take that portion of the parabola and reflect it vertically.



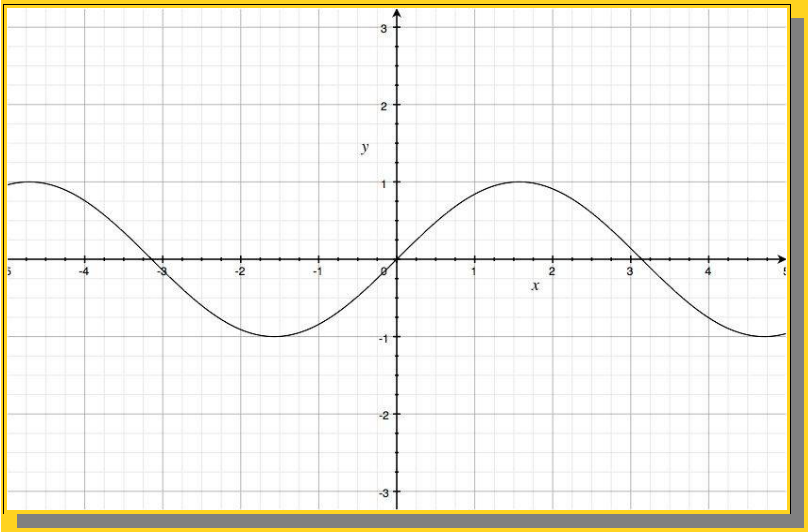
A reciprocal transformation is expressed as

$$y = \frac{1}{f(x)}$$

and is performed by taking each point on the graph, calculating the reciprocal of its y-coordinate, and plotting the new point.

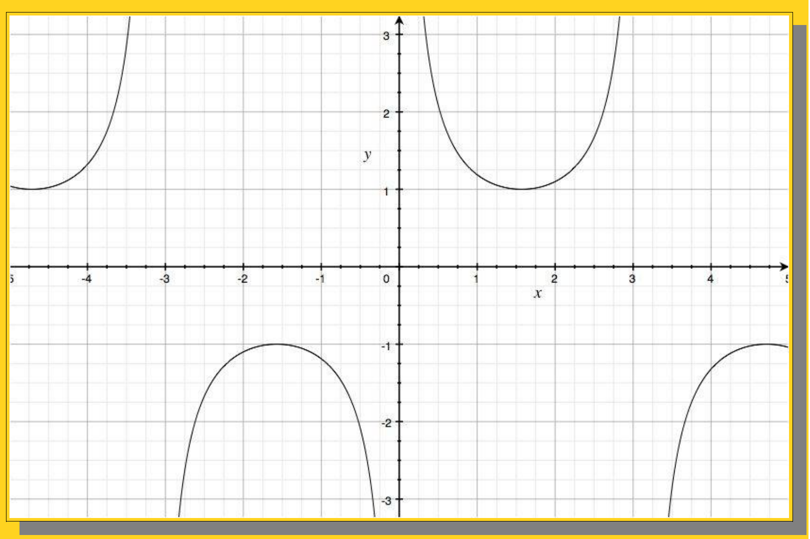
Note that when *f(x)=0*, a vertical asymptote will usually form, and when *f(x)=1* or *f(x)=-1*, the reciprocal transformation leaves that point unchanged.

Example: Graph the reciprocal of the function in the graph shown.



Solution: Mark where *f(x)=1* or *f(x)=-1*, and leave those points as they are. Also, whenever *x=0*, mark vertical asymptotes at those points. Elsewhere, simply approximate the reciprocal values, remembering that *larger* y-values yield *smaller* ones when taking the reciprocal.

Incidentally, if you weren't already aware, the graph of the original function was the graph of *y = sin x*. While we will discuss transforming graphs of trigonometric functions later, the point is to be able to perform the transformation regardless of what you know (or don't know) about the function at hand.



Other Notes on Transformations

When multiple transformations are involved, the general order to follow is:

1. Special transformations
2. Reflections, expansions, compressions
3. Translations

(Basically, follow the order of operations.)

Also, when interpreting translations involving x, make sure it stands on its own. Given

$$y=f(3x+4)$$

you might think the translation is 4 left, but actually, when rewritten so that x stands on its own as such:

$$y=f\left(3\left(x+\frac{4}{3}\right)\right)$$

then we can see that the translation is actually 4/3 left.

Trigonometry: Fundamentals

Xander: Giles lived for school. He's actually still bitter that there are only twelve grades.
Buffy: He probably sat in math class thinking, "There should be more math. This could be mathier."
– Buffy the Vampire Slayer

Overview

This chapter is primarily a review of some of the fundamental concepts of trigonometry learned in previous years. The major difference this time around is the introduction of the *radian*, a new measure of angle.

Periodic Functions

A *periodic function* is a function whose graph oscillates with a fixed pattern – that is, the graph will repeat a certain part of itself over and over.

Formally, a function is said to be *periodic* with period p if

$$f(x+p)=f(x)$$

for all x.

As an example, think about the sine of 90°, which is 1. If you add/subtract any multiple of 360°, you will still end up with 1 when you take the sine of it, since adding 360° returns you to the exact same angle. Therefore, the sine function is periodic with period 360°, because $\sin(x+360^{\circ})=\sin(x)$ for all x.

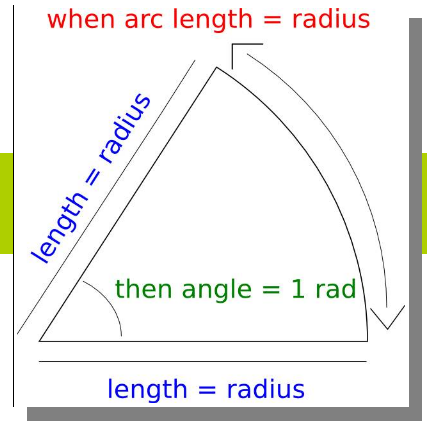
Naturally, all the trigonometric functions are periodic, and in the next chapter, we will see this in their graphs, as well as learn how to take advantage of their periodic nature in order to model some situations.

Radian Measure

A *radian* is a measure of angle, just like the degree. Some of you may be surprised to learn that there are other ways of measuring angles, and may even wonder why another measure of angle is even necessary, especially when degrees are so widespread and commonly used.

It's true that in practical applications, degrees are almost always preferred, simply because we're so familiar with them. But we are now a little past the level of practical geometry (this course is practically a direct predecessor to calculus), and it is time we asked: why divide the circle into 360 parts? Because, from a mathematical perspective, the number 360 is *random* – a number somebody just pulled out of thin air.

Keeping this in mind, a more promising measure of angle would be one that is somehow based upon the properties of the circle itself. And this is, in fact, how the radian is defined:



Given a circle and two radii (plural of radius), the angle between them is exactly one radian if the arc subtending the angle is exactly the length of the radius. (See adjacent picture.)

Notice how the radian is defined without any "magic numbers" dropped anywhere. All we are doing is noting how big of an angle we need to make the arc length exactly equal to the radius, and defining that to be a single radian.

This definition also gives us a useful formula immediately: given the arc length s and the radius r, if the angle θ is in radians, then

$$\theta = \frac{s}{r}$$

because when θ is one radian, s and r are equal. Rearranging this gives us a formula for the length of an arc on a circle:

$$s = r\theta$$

Obviously, **this formula only works when θ is in radians** – so already, we have a compelling reason for choosing radians over degrees: it gives us a simple formula for arc length on a circle.

Now, since the circumference of a circle is defined as

$$C = 2\pi r$$

it follows from the arc length formula (where $s=C$) that a complete revolution (360 degrees) is defined as 2π radians. Or, dividing by two:

π radians = 180 degrees

This gives us a conversion factor from degrees to radians.

Example: Convert the following angles from degrees to radians:

[a] 0° [b] 30° [c] 90° [d] 180° [e] 225° [f] 57.296°

Solution:

[a] Trivial. 0 deg = 0 rad, obviously.

[b] Following the physics/chemistry method of unit conversion,

$$30^{\circ} \times \frac{\pi \, rad}{180^{\circ}} = \frac{\pi}{6} rad$$

Since π radians is exactly 180°, we are essentially multiplying by 1, which does not change anything (but we do get new units).

[c] By a similar method, 90° = π/4 rad.

[d] By a similar method, 180° = π rad. (Obviously.)

[e] By a similar method, 225° = 5π/4 rad. (It is customary to express radian angle measurements as improper fractions, as opposed to mixed numbers.)

[f] By a similar method, 57.296° ≈ 1 rad. (One radian is a little under 60°.)

Note that the arc length formula (mentioned previously) has an equivalent for degrees:

$$s = \frac{\pi}{180} r \theta$$

But just by looking at the addition of the ugly "conversion factor", you can see why radians might be preferable in this case.

As one final example of the usefulness of radians, this is the formula for the area of a sector (the area of a "slice of pie" if you imagine the circle as a pie).

$$A = \frac{1}{2} r^2 \theta$$

Notice that when $\theta = 2\pi$ (a full circle), $A = \pi r^2$, as expected. The equivalent when θ is in degrees:

$$A = \frac{\pi}{360} r^2 \theta$$

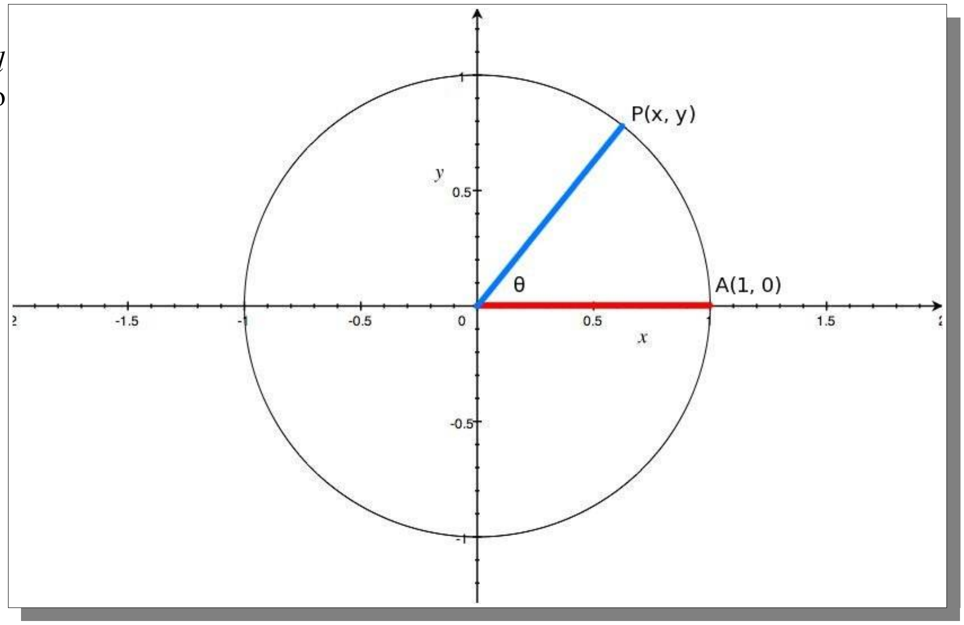
More advanced branches of mathematics (notably calculus) use radians almost exclusively for reasons such as these – there is a mathematical "naturalness" in their usage that is not seen in any other form of angular measurement.

Trigonometric Functions: Sine and Cosine

When the trigonometric functions are introduced, they are typically defined as ratios in a right triangle: given a reference angle A, sin A = opposite/hypotenuse, cos A = adjacent/hypotenuse. (The tangent function will be covered in the next section.)

Unfortunately, these definitions are hopelessly inadequate for any non-acute angle, so what we need is a way of defining the trigonometric functions for *all* angles. To do this, consider the *unit circle*, a circle with radius 1 and center at the origin.

The red line is called the *initial arm*, and extends from (0, 0) to (1, 0). The blue line is called the *terminal arm*, and unlike the initial arm, it is free to be moved around the circle. The angle between these two arms is denoted θ.



If you draw a right triangle, using the terminal arm as the hypotenuse, you will find, using the old trigonometric ratios, that

$$\sin\theta = \frac{x}{r}$$

$$\cos\theta = \frac{y}{r}$$

where r is the radius of the circle.

But since the radius is defined to be 1:

$$\cos\theta = x$$

$$\sin\theta = y$$

In other words, the cosine and sine of an angle are defined to be the x- and y-coordinates of the point P on the terminal arm, respectively.

Because we derived this result using the old definitions of ratios in right triangles, we can be certain that it matches with the old definition for acute angles – but, as you can see in the diagram above, we can swing the terminal arm past 90°, and we will still have definitions for the sine and cosine of θ. Note, however, that the answer may turn out to be *negative*, depending on where exactly the terminal arm is located.

(The angle between the terminal arm and the x-axis is called the *reference angle*, and is always acute. The sine/cosine of any angle is always equal to the sine/cosine of the reference angle, give or take a minus sign.)

Notice that swinging the terminal arm around gives us the same results eventually, repeating every 360°. We can say that angles are *coterminal* if they result in the exact same positioning of the terminal arm – for example, 0° and 360°. In general, we can add (or subtract, since angles can be negative as well) any integer multiple of 360° to form yet another coterminal angle.

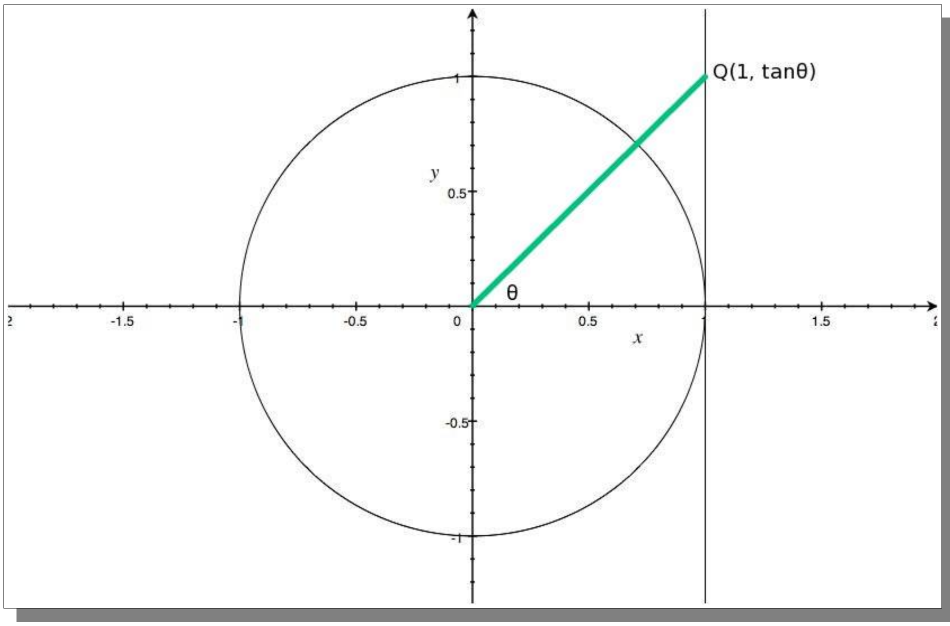
Using radians instead of degrees gives us 2π instead of 360, but the concepts are identical. (However, we will be using radians almost exclusively henceforth, so you will have to get used to seeing multiples of π thrown around.)

Two more important notes:

- The sine function yields positive numbers in quadrants I and II, where the y-coordinate of the terminal arm is positive. Similarly, the cosine function yields positive numbers in quadrants I and IV, where the x-coordinate of the terminal arm is positive.
- For a few key angles, it is possible to find the *exact* value of the sine/cosine of that angle. For a full list, see Appendix B.

Trigonometric Functions: Tangent

The tangent function is defined as the ratio opposite/adjacent, but as with the sine and cosine functions, we will need a different way of defining the tangent of any non-acute angle. Have a look at the following diagram:



If we draw a *tangent* line to the unit circle at its right edge, described by equation $x=1$, we can extend the terminal arm until it touches the terminal line. The point of intersection (Q) has an x-coordinate of 1 and, by using the tangent ratio on this right triangle, we have $\tan\theta=y/1=y$. Hence the y-coordinate is $\tan\theta$.

(Based on this definition, it should be obvious where the tangent function got its name from.)

Notice that, unlike the sine and cosine functions, the tangent function's output is not restricted in between -1 and 1 – in fact, as θ approaches 90°, tanθ approaches infinity. (The tangent of 90° is undefined, because the terminal arm will never intersect the tangent line.) Once the tangent function passes 90°, however, its value then becomes *negative*, as the terminal arm must be extended the other way in order to intersect the tangent line. (Past 180°, the value becomes positive again, and so forth.)

One last interesting item of note is the *quotient identity*:

$$\tan\theta=\frac{\sin\theta}{\cos\theta}$$

This suggests that the tangent function yields positive numbers where the sine and cosine functions have the same sign – that is, in quadrants I (both positive) and III (both negative).

Trigonometric Functions: Reciprocals

If you think about the sine ratio, opposite/hypotenuse, you might wonder why we couldn't have a ratio for hypotenuse/opposite. And, for that matter, why don't we have a hypotenuse/adjacent ratio, or an adjacent/opposite ratio?

Well, as it turns out, there *do* exist functions with the reciprocal ratios.

cosecant: $csc\,x=\frac{1}{\sin x}$

secant: $sec\,x=\frac{1}{\cos x}$

cotangent: $\cot x=\frac{1}{\tan x}$

Be careful not to get cosecant paired with cosine, even though they both have the prefix “co” – cosecant is the reciprocal of *sine*.

As for making use of these functions, your calculator is unlikely to have the keys for these reciprocal functions, but there is no need for them – simply make use of the identities above.

Example: Calculate csc(4.3).

Solution:

$$csc(4.3)=\frac{1}{\sin(4.3)}\approx-1.092$$

(Notice that, in the absence of explicit units, the assumption is that arguments are in radians.)

Trigonometry: Graphing

Kaiba: I'm here for your Blue-Eyes, old man, and I won't take no for an answer. Now give it to me.

Trusdale: No.

Kaiba: Curses. Foiled again. I'm going to go hire some thugs to kidnap you now. I'm a billionaire, so...nobody will even think about pressing charges. [leaves]

Trusdale: That Kaiba kid needs to get laid.

– Yu-Gi-Oh: The Abridged Series

Overview

With the fundamentals of trigonometry under our belt, it is now time to study the graphs of the trigonometric functions and perform transformations on them. All the transformation work is an application of the transformations chapter, but there is a little terminology exclusive to trig functions, and a few techniques specifically designed for quickly sketching trig curves.

It is important to note that when graphing trig functions, the arguments are assumed to be in radians.

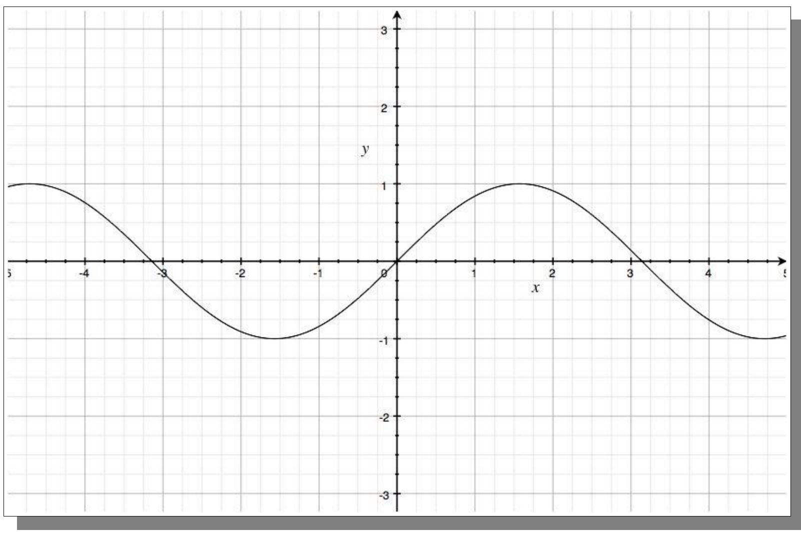
Trigonometric Graphs: Sine and Cosine

Graphing the sine function may seem like a difficult task if you have no idea what it looks like, but we can always fall back on the old method of making a table of values.

x	$y=\sin(x)$
0	0

π/6	1/2
π/2	1
5π/6	1/2
π	0
7π/6	-1/2
3π/2	-1
11π/6	-1/2
2π	0

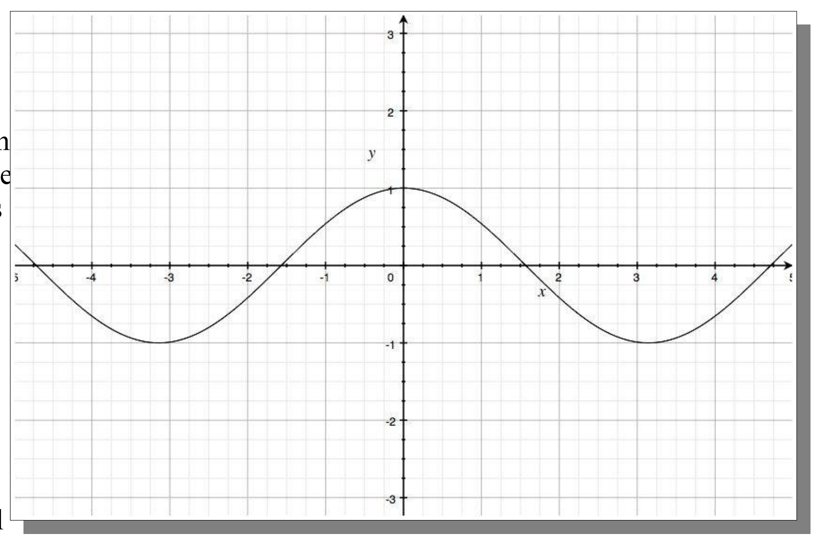
(Obviously, the list is composed of angles that lead to convenient numbers for graphing.) Past 2π, we know that the function repeats itself (since the function is *periodic*), so we can duplicate the results of a single cycle across the board. Eventually, we will get something like:



This type of wave is called a *sinusoid*, and the sine function is known as a *sinusoidal function*. Because of the regularly repeating pattern, it becomes useful for modelling a wide variety of situations.

Also notice that the sine function is odd (the technical term, not "weird"). (Refer to the transformations chapter for details.)

The cosine function is similar, and is pictured in the adjacent graph.



The cosine function is also sinusoidal, and in fact is almost completely identical to the sine function in many regards. In fact, the graphs themselves suggest the identity

$$\cos(x)=\sin(x+\frac{\pi}{2})$$

In other words, the cosine function is the sine function shifted π/2 units to the left. Notice how it ends up making the sinusoidal function *even*, instead of odd.

Trigonometric Graphs: Sinusoidal Transformations

In principle, transformations on graphs of sinusoidal functions are no different from the transformations of any general function. However, transformations on sinusoidal functions are so common that specific terminology has been developed for it, as well as various methods to quickly graph the needed transformations.

The general sinusoidal equation has form

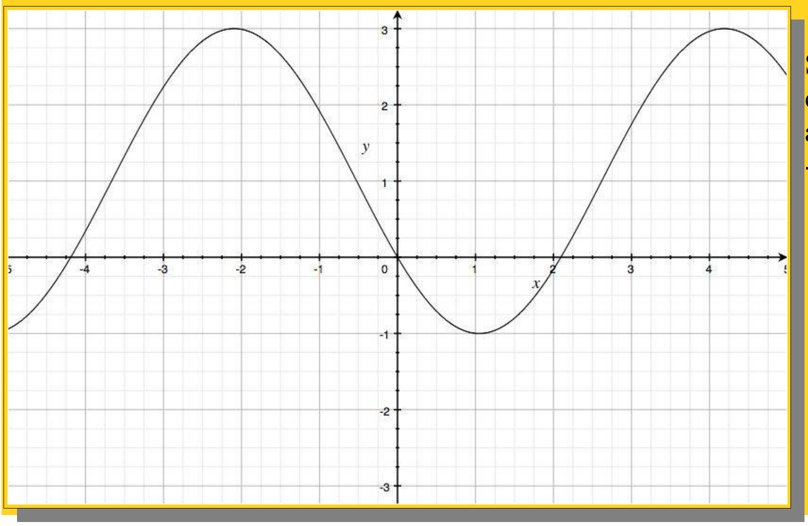
$$y=a\sin(b(x-c))+d$$

(Cosines may be expressed as sines, and vice versa, so whatever is mentioned here about sines applies to cosines as well.)

The quantity $|a|$ (notice the absolute value signs) is called the *amplitude* of the sinusoid, and is basically a measure of how "extreme" the peaks and valleys of the graphs are. It is defined as the vertical distance from the midpoint of the wave to either a peak or a valley. (The equation expresses it in an alternate form, using half the distance from a peak to a valley.)

$$|a|=\frac{y_{max}-y_{min}}{2}$$

Example: Identify the amplitude of the sinusoid in the following graph.



Solution: Although we don't have the equation on hand, we can identify the peak and valley y-coordinates, which are 3 and -1, respectively. Thus, the amplitude is:

$$\frac{3+(-1)}{2}=2$$

The quantity *2π/b* is called the *period* of the sinusoid, and is essentially a fancy term for the horizontal stretch factor – modifying *b* has the same effects as any normal horizontal stretching. It is also a description of how long a single full cycle of the function is (see the chapter on trigonometric basics).

Usually, *b* will itself be a multiple of π, so that the period works out to be a convenient, rational number. This is especially useful in applications, such as in the following example:

Example: The velocity of a particle as a function of time (in meters per second), *v(t)*, can be modelled with a sinusoid. Initially, *v(t)* is at its minimum value of -2, and it reaches its maximum value of 2 in 3 seconds. Find an equation for *v(t)*.

Solution: We know the function's minimum and maximum y-values, so using those, we can find the amplitude:

$$|a|=\frac{2-(-2)}{2}=2$$

To determine whether *a* is positive or negative, we have to first decide on whether to use the sine or cosine function to model the velocity. Choosing sine gives us an initial value of 0, whereas choosing cosine gives us an initial value of 1.

Since the initial value is -2, choosing the cosine function and taking *a* to be -2 (since the cosine function's initial value is 1 by default, we need to multiply by -2) works out to be the simplest option. (It is entirely possible to use the sine function, but that would require some shifting, which complicates matters.) Therefore, we have determined that *a=-2*.

Now, shifting our attention to the period, we note that it reaches its maximum value in 3 seconds. This means that it will take another 3 seconds to reach its minimum value again, resulting in a full cycle taking 6 seconds. Thus the period is 6, and to find *b*, we use our formula

$$\frac{2\pi}{b}=6$$

and solve for *b*, which works out to be π/3 (when simplified). Thus, the final equation is

$$v(t)=-2\cos\left(\frac{\pi}{3}t\right)$$

The variables *c* and *d* are referred to as the *phase shift* and *vertical displacement* of the sinusoid, respectively, and are normal translation variables, behaving like any other. Note, however, that there is a formula for the vertical displacement, expressing it as the average of the peak and the valley:

$$d=\frac{y_{max}+y_{min}}{2}$$

d itself also represents the y-coordinate of the midpoint of the sinusoidal wave.

We can also derive formulae for the peak and valley y-coordinates using the equations for the amplitude and vertical displacement:

$$y_{max}=|a|+d$$
$$y_{min}=-|a|+d$$

By this point, all this terminology may seem overwhelming, especially given that most of it is composed of special cases of general function transformations. After all, wasn't the whole point of the transformations chapter to free us from the work of having to memorize a new method of transformations for each function?

Well, don't worry – the work from that chapter hasn't been wasted. In fact, we applied it here, in order to quickly figure out the effects of each variable in the general sinusoidal equation. It just so happens that transforming trig graphs is common enough that developing specific terminology is useful. As an example, let us now use this new terminology to derive a quick way to graph transformed sinusoidal equations.

Since the sinusoids are periodic, if we can figure out how to graph one full cycle, we can duplicate the result to the left and right of it, so we only need to focus on a single cycle. The idea is to create a box marking the boundaries of the cycle, so that when we draw a full cycle within it, it is automatically the right size and in the right place.

The top and bottom of the box are obviously the peak and valley y-coordinates, respectively, and can be calculated using the formulae above for y_{max} and y_{min}. The left of the box is precisely equal to the phase shift value, because when *x=c* (the phase shift), the function will look like

$$y=a\sin(b(c-c))+d=a\sin(0)+d=d$$

or

$$y = a\cos(0) + d = a + d$$

meaning that the “starting point” of the wave begins. As for the right of the box, we take the phase shift and we add the value of the period to it, which makes sense – the period is the length of a full cycle.

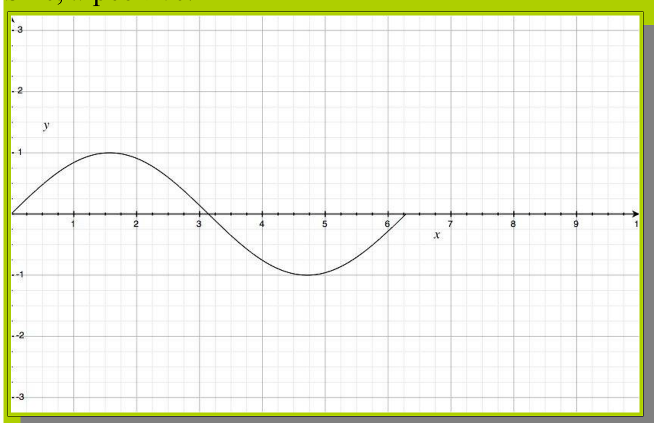
So, to recap:
Left bound: c (phase shift)
Right bound: $c + 2\pi/b$ (phase shift + period)
Upper bound: $|a| + d$ (y_{max})
Lower bound: $-|a| + d$ (y_{min})

Now you simply draw the box with these boundaries, fit a single cycle of the wave inside, and then duplicate the results.

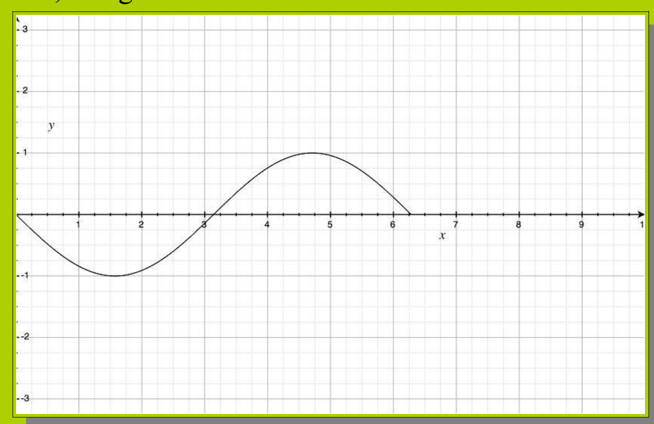
As for what the cycle should actually look like, that depends on two things:

1. whether we're using the sine or cosine function
2. whether a is positive or negative

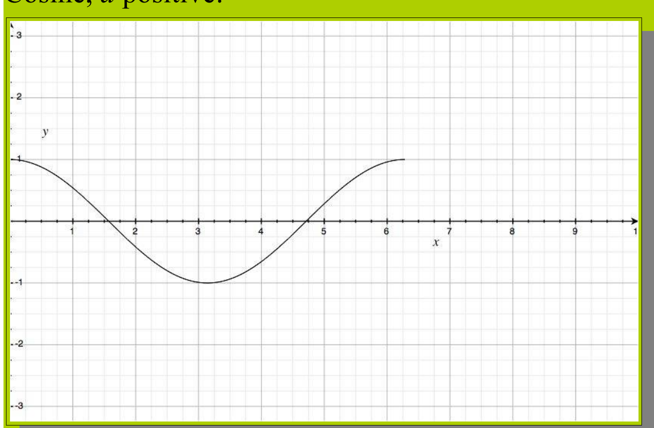
Sine, a positive:



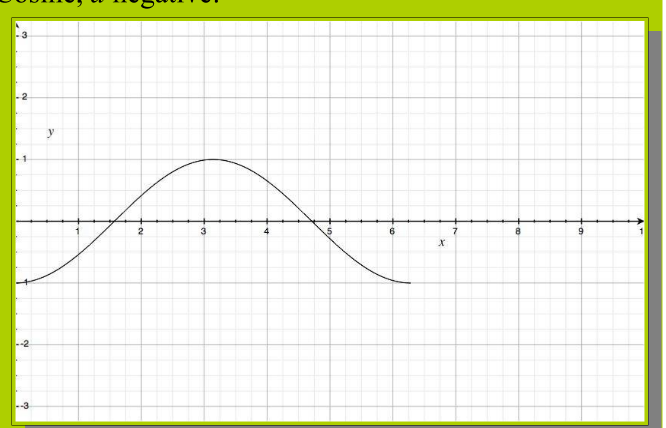
Sine, a negative:



Cosine, a positive:



Cosine, a negative:



An example will probably clarify things.

Example: Graph the function

$$y=-\cos(2(x-\frac{\pi}{3}))+1$$

for two full cycles.

Solution: We first draw a box with the boundary conditions listed.

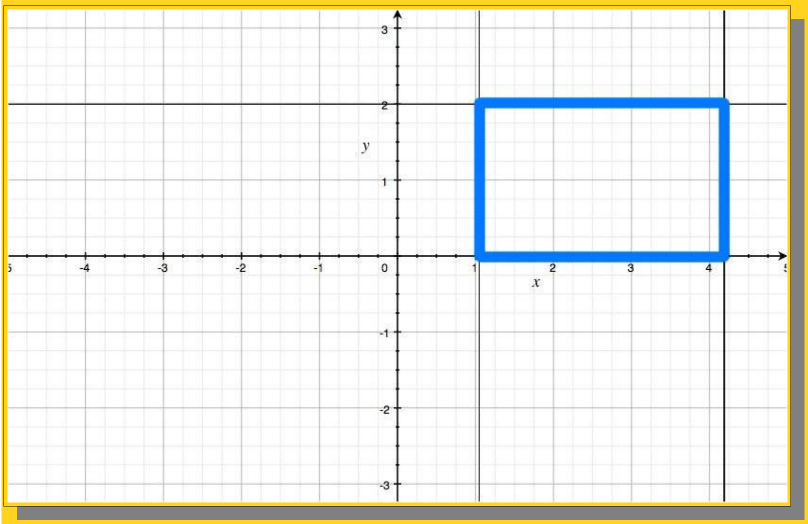
Left bound: $c = \pi/3$

Right bound: $c + 2\pi/b = \pi/3 + 2\pi/2 = 4\pi/3$

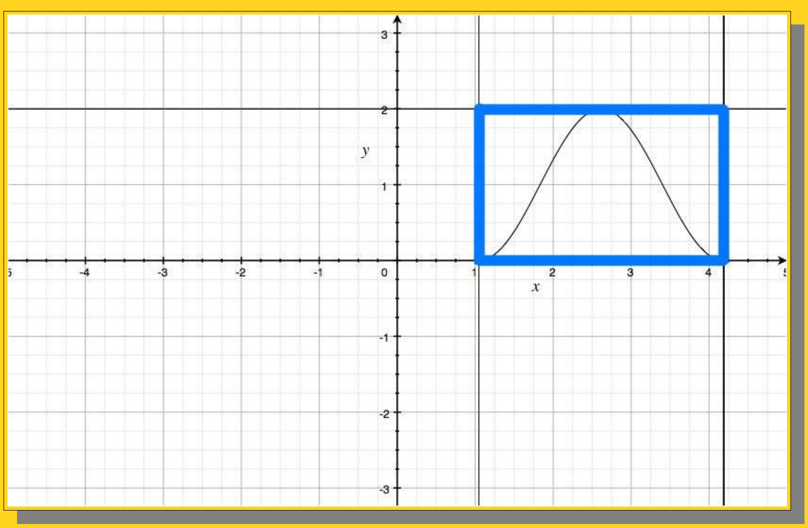
Upper bound: $|a| + d = |-1| + 1 = 2$

Lower bound: $-|a| + d = -|-1| + 1 = 0$

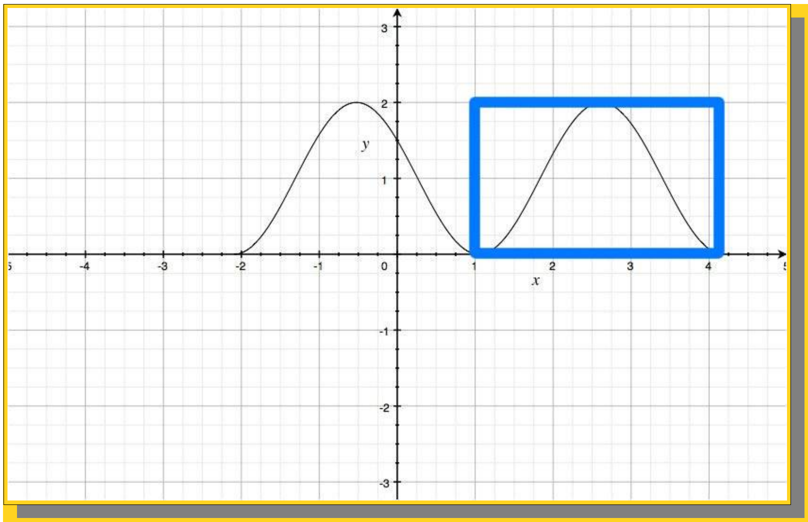
On the graph paper, we mark the box like so:



Then we graph a single cycle. Since a is negative, and we are using the cosine function, we pick the appropriate cycle from the list above:



Finally, we duplicate the cycle once, since we wanted two full cycles:



As you can see, this method is considerably faster than taking the graph of the basic cosine function and performing all the transformations one at a time. **It is by no means a *requirement* for you to use this faster method, but it will have considerable payoff.**

Note that, unlike in the graphing examples given, you will usually scale the grid to some multiple of π when graphing by hand, in order to enable easier calculation of boundaries. A common scaling used is having each grid space = π/6.

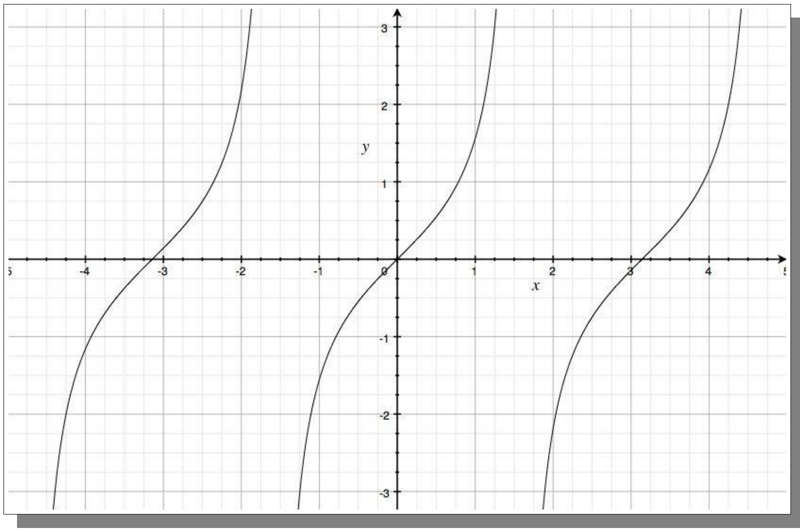
Trigonometric Graphs: Tangent

Unlike the sine and cosine functions, the tangent function is not sinusoidal, but it does share their periodic property, meaning that, as before, we only need to figure out how to graph a single "cycle" of the tangent function, then duplicate that.

Let us begin by first figuring out what the basic tangent graph looks like. Using a table of values:

x	$y=\tan(x)$
0	0
π/6	1/√3
π/4	1
π/3	√3
π/2	undefined
2π/3	-√3
3π/4	-1
5π/6	-1/√3
π	0

Past π radians, we find that the tangent function repeats itself, suggesting a period of π. The graph of it confirms this:



Notice the vertical asymptotes. We know one occurs at π/2, since the table of values tells us that *tan(x)* is undefined at that point, but what about the others?

To answer the question, we make use of the quotient identity:

$$\tan x = \frac{\sin x}{\cos x}$$

By looking at this equation, we can see that wherever *cos(x)=0*, the tangent function has vertical asymptotes (since a fraction is not defined when its denominator is 0).

We will examine how to solve trigonometric equations in the next chapter. For now, suffice to say that the tangent function has asymptotes at *x=nπ/2*, where *n* is an odd integer.

As for transforming graphs of the tangent function, there is no equivalent of the “box” method for the sinusoidal functions. Instead, we use our normal transformation techniques directly on this graph.

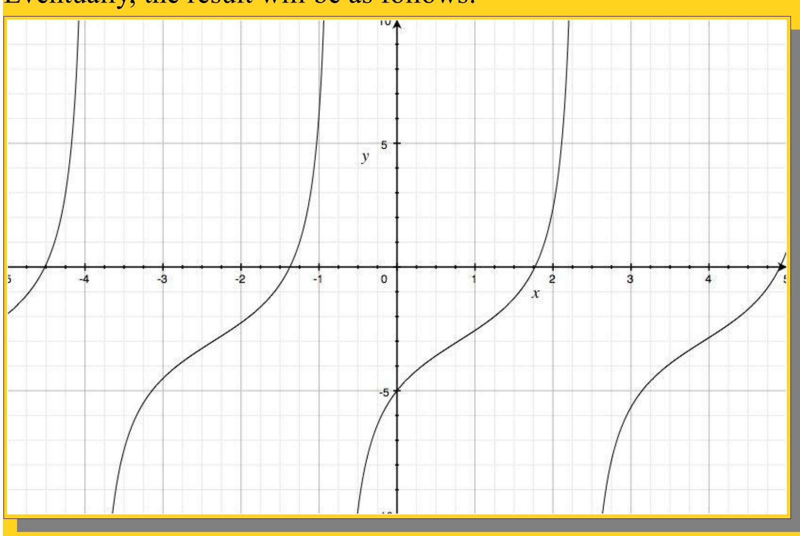
Example: Graph the function

$$y = 2\tan(x - \frac{\pi}{4}) - 3$$

and give a general formula for its asymptotes.

Solution: We note down the various transformations: a vertical expansion by a factor of 2, a horizontal shift π/4 right, and a vertical shift 3 down.

We then take the tangent graph and perform the transformations (remember, expansions before shifts). Eventually, the result will be as follows:



As for the vertical asymptotes of this function, we should first note that *vertical* transformations have no effect on their position – only horizontal ones do. (If you find this difficult to accept, think about the fact that asymptotes are essentially vertical lines, meaning that no amount of vertical stretching or shifting will change them.)

Therefore, we have to note the horizontal transformations in this graph. In this case, we have one, a shift right by π/4. Therefore, we take the default equation for asymptotes, which is *x=nπ/2* (where *n* is an odd integer), and perform our horizontal transformations on it. In this case, that means moving it π/4 to the right, so:

The equation for the asymptotes is

$$x = \frac{n\pi}{2} + \frac{\pi}{4}$$

where n is an odd integer.

Trigonometric Graphs: Reciprocals

Graphing the reciprocal trigonometric functions follows the same process as graphing the reciprocal of any general function.

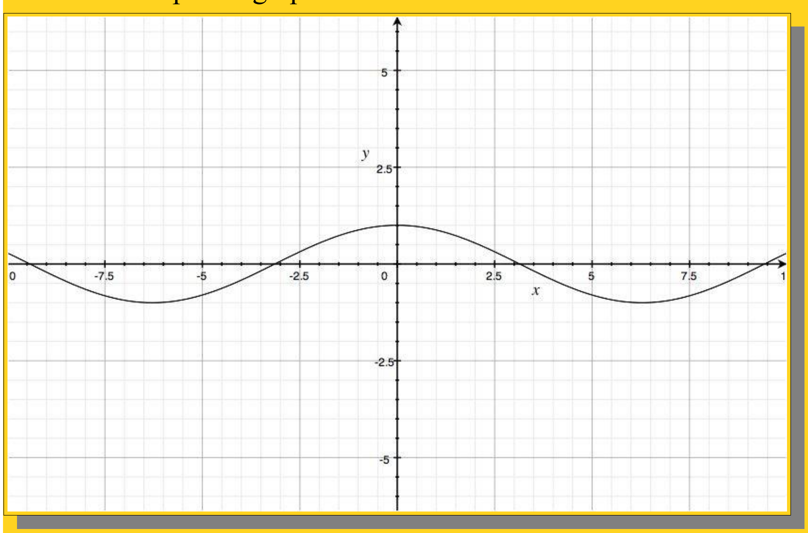
Example: Graph the function

$$y = 4\sec\left(\frac{1}{2}x\right) + 1$$

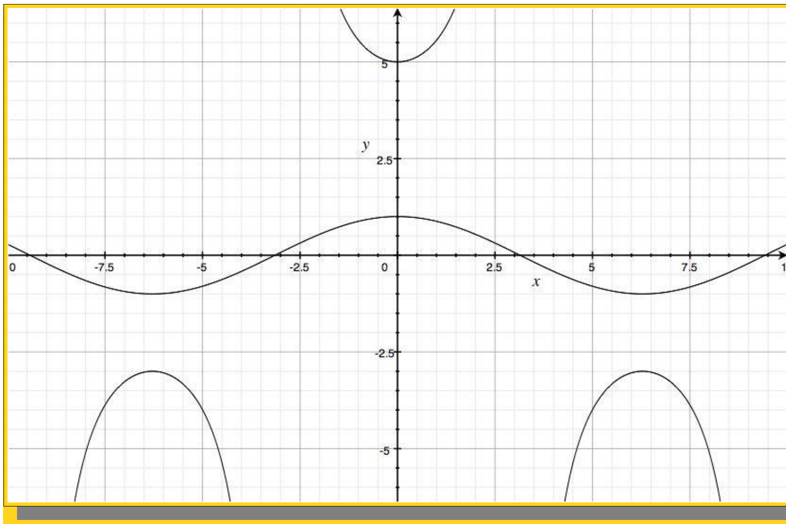
Solution: This can be written as

$$y = 4\left(\frac{1}{\cos\left(\frac{1}{2}x\right)}\right) + 1$$

So the first step is to graph the cosine function:



Then perform the reciprocal transformation and finish with the vertical expansion and shift:



(The cosine function is left as a reference.)

Notice the vertical asymptotes that form. Again, these are when the denominator – in this case, *cos((1/2)x)* – equals 0. (Solving trigonometric equations will be covered throughly in the next chapter.)

Note that when graphing the cotangent function, the asymptotes of the tangent function become x-intercepts of the cotangent function. This is because we basically have 1/∞ (the asymptotes can be considered to have a y-value of ∞), which is 0. (This is not mathematically precise, but it should make sense intuitively.)

Trigonometry: Equations and Identities

Contractor: We have a flat tire. Can you help us?

James: Sure, yeah. Do you have spares?

Contractor: We have spares, but we've used up our wrench.

James: How do you use up a wrench?

Contractor: Well, uh, the guy over there, Feisal, he threw it at someone.

– The Hurt Locker

Overview

This final chapter of trigonometry will now reconcile trigonometry with algebra, establishing some fundamental identities and employing them to solve equations with trigonometric functions.

As always, all angles are measured in radians.

Proving Trigonometric Identities

As the graphs of the trigonometric functions suggest, there are many relationships between all six trigonometric functions. By use of the quotient identity, we can express all six trig functions in terms of the sine and cosine functions.

It is now time to look at the *Pythagorean identity*, one of the most important identities in trigonometry. The identity states

$$\sin^2 x + \cos^2 x = 1$$

Two notes about the Pythagorean identity:

- $\sin^2 x$ is shorthand for $(\sin x)^2$. The former notation, while possibly confusing, is so widespread that there is little use in attempting to fight it. (Similar comments apply for any function “raised to a power”.)

- The Pythagorean identity establishes a relationship between the sine and cosine functions, enabling us to express all six trig functions in terms of only *one* of the trig functions.

A full list of identities (trigonometric and otherwise) can be found in Appendix A.

Now, by making use of these basic trig identities, we may prove many other ones. It may surprise you to find so many identities involving the trigonometric functions, but it illustrates how closely linked they all are. However, note that only a few, such as the Pythagorean identity, are worthy of special attention. The identities we prove in this chapter are run-of-the-mill.

Example: Prove the identity

$$\frac{1-\cos x}{\sin x}=\frac{\sin x}{1+\cos x}$$

Solution: On the left-hand side, we have 1 – cos x, which looks very similar to the 1 + cos x seen in the denominator on the right-hand side. So, making an effort to set the denominators equal to each other, if we multiply the left fraction by 1 + cos x on the bottom and top (so as to not alter anything), we obtain:

$$\frac{(1-\cos x)(1+\cos x)}{(\sin x)(1+\cos x)}=\frac{\sin x}{1+\cos x}$$

The numerator of the left fraction is a difference of squares factored:

$$\frac{1-\cos^2 x}{(\sin x)(1+\cos x)}=\frac{\sin x}{1+\cos x}$$

We're almost there, but not quite; we still have a sin x in the denominator of the left fraction. To get rid of it, we need to call upon the Pythagorean identity to simplify the numerator of the left fraction:

$$\frac{\sin^2 x}{(\sin x)(1+\cos x)}=\frac{\sin x}{1+\cos x}$$

(We simply rearrange the Pythagorean identity.) Now finish simplifying:

$$\frac{\sin x}{1+\cos x}=\frac{\sin x}{1+\cos x}$$

And since the left side now equals the right side, we are done.

Two notes:

- The more mathematically observant among you will notice that non-permissible values are altered throughout the proof, as cancellations in fractions happen. This is fine; when we say that the two expressions are equal, we mean that they are equal *only for the values for which both are well-defined.*
- You may ask “why can't we just cross-multiply and show it that way?” The answer (perhaps surprisingly, perhaps not) is that the Math 12 curriculum says you can't – that is, it is an arbitrary restriction created to make your life more difficult. (Did you expect to see *that* inside a math textbook?)

There is no fixed algorithm for proving trig identities; it is a technique that is mastered mostly through practice and requires one to develop some intuition. However, here are general guidelines:

- Try working with the "harder" or "more complicated" side first. It might seem strange, but it seems to be easier to start with a complicated expression and simplify it, rather than start with a simple expression and complicate it (correctly).
- Write everything in terms of sine and cosine. Tangent should be converted to sin/cos, secant to 1/cos, and so forth.
- When dealing with fractions, try and combine individual fractions into one single fraction. For example, given sin/cos + 1, change that to sin/cos + cos/cos = (sin + cos)/cos.
- **Memorize the Pythagorean identity (see Appendix A) and make use of it.** Its importance in proving trigonometric identities is difficult to exaggerate.
- If you see a fraction with a sine/cosine on top and a cosine/sine on bottom +/- 1 (or vice versa), you can manipulate it, using differences of squares and the Pythagorean identity. (The example given above illustrates this.)

Solving Trigonometric Equations

The other part of this chapter, and the more difficult one, is solving equations involving trigonometric functions (called *trigonometric equations*). An example would be

$$2\cos x - \sin^2 x + 2 = 0$$

It is important to note that only the simplest trigonometric equations admit exact solutions. An equation as simple as

$$\sin x = 0.8$$

has *no* exact solutions because we do not know the exact angle that will yield 0.8 when the sine function is applied to it. (We can, of course, approximate it – your calculator performs such an approximation.)

For the purposes of Math 12, all trigonometric equations will be constructed in such a way that exact solutions are possible to find. You should keep in mind, however, that this construction is artificial and hardly representative of more general equations.

So, let us begin by attempting to solve the example problem above, which should highlight some of the important details in solving trigonometric equations.

Example: Solve

$$2\cos x - \sin^2 x + 2 = 0$$

Solution: In general, working with a *single* type of trigonometric function is desirable. Here, for example, we have both a sine and a cosine function, but it would be nice to have only sines, or only cosines.

In this case, we can use the Pythagorean identity to convert the sine into a cosine:

$$2\cos x - (1 - \cos^2 x) + 2 = 0$$

Now that we only have cosines, we can continue. If we introduce a temporary variable

$$u = \cos x$$

then the equation becomes

$$2u - (1 - u^2) + 1 = u^2 + 2u + 1 = 0$$

which is an ordinary quadratic equation. Solving this gives us

$$u=-1$$
(the only distinct solution since the polynomial factors into a perfect square). Therefore, by reverse substitution:
$$\cos x=-1$$
which means
$$x=\cos^{-1}(-1)=\pi$$
which is the only solution between 0 and 2π. Of course, this is not the only solution; indeed, since *cos(x)* is periodic with period 2π, we can say that the complete set of solutions is described by
$$x=\pi+2\pi n$$
where *n* is any integer.

Although the solving of trigonometric equations appears complicated, it is actually a very systematic procedure, and resembles the solving of a normal algebraic equation in many ways. The key difference here is that, because we have periodic functions, there are an *infinite* number of solutions to the problem. (Usually, however, we are only interested in solutions between 0 and 2π.)

Example: Solve the trigonometric equation
$$3\sin x-4\sin^3 x=1$$
Find only the answers in the range $0 \le x < 2\pi$.

Solution: Since we only have sines present in this equation, we have no need to make conversions. Now, let us introduce a temporary variable
$$u=\sin x$$
which gives us
$$3u-4u^3=1$$
which resolves into a cubic equation:
$$4u^3-3u+1=0$$
By use of your method of choice (the rational root theorem is a good candidate), you can solve this cubic equation to find
$$u=-1,\frac{1}{2}$$
Therefore
$$\sin x=-1,\frac{1}{2}$$
and we have
$$x=\sin^{-1}(-1)=\frac{3\pi}{2}$$
or
$$x=\sin^{-1}\left(\frac{1}{2}\right)=\frac{\pi}{6,}\frac{5\pi}{6}$$

Since the question asks for all solutions between 0 and 2π, we can be certain we have found all the solutions, because we have explicitly searched for all solutions in that range manually.

Incidentally, there exists an identity that simplifies this question considerably:

$$\sin 3x = 3\sin x - 4\sin^3 x$$

(This identity is not required for the purposes of Math 12, but it is easy to derive by finding *sin(2x + x)* and using trigonometric identities.)

Using this identity gives us

$$\sin 3x = 1$$

which means

$$3x = \sin^{-1}(1) = \frac{\pi}{2}$$

and therefore

$$x = \frac{\pi}{6}$$

We know, however, that the sine of 5π/2 radians is also 1 (since adding 2π gives us the same angle), so *3x* is also equal to 5π/2. Normally, this additional solution would be of no concern to us, since it is a duplicate of π/2 (with 2π added). However, in this case, because we are dividing our solutions by 3 at the very end, dividing 5π/2 by 3 gives us 5π/6, which is *inside* our desired range. Therefore, we must add another solution:

$$x = \frac{5\pi}{6}$$

By similar reasoning, the sine of 9π/2 radians is also 1, so the final solution to be added is

$$x = \frac{9\pi}{6} = \frac{3\pi}{2}$$

and we note that these are the exact same answers as those which we found earlier.

Now that we've looked at a couple of examples of trigonometric equations, what can we take from it?

- Convert all trigonometric functions into a single trigonometric function. Various identities exist to assist in this goal, ensuring that it is always possible to express all trigonometric functions using only one.
- If the equation involves more than a single instance of a trigonometric function, it may help to introduce a temporary variable to represent that function, then solve the equation as though it were a normal algebraic equation.
- **It is important not to cancel any variables.** Cancellation of variables leads to a loss of solutions.
- At the final step, when you are attempting to find the angle itself, it is important to look for *two* answers between 0 and 2π, because there are almost always two angles within that range that satisfy the answer. (There will only be one answer if you are taking the inverse sine/cosine of 1/-1, because in a single revolution, the x- and y-coordinates only reach those values once.) (Using the inverse trigonometric function buttons on your calculator only gives one answer, so it is a good idea to search for the answers manually.)
- If you are asked to find *general* solutions, make sure to note that for each of the solutions, adding any integer multiple of the period is also a solution.

As a final note, questions become more interesting when non-trigonometric functions are thrown into the mix. For example:

$$x - \cos x = 0$$

However, this equation has *no* exact solutions. Therefore, equations of this nature are studied in calculus, when methods of *approximating* solutions are discussed. (If you're curious, look to Appendix C, where a technique for approximation that does not require calculus is discussed.)

Exponentials and Logarithms

Sarge: Get over here, give me a boost.
Caboose: Ok. [walks over] You are a good person, and people say nice things about you.
Sarge: Not a morale boost, moron, a physical one. I need to see what's in that window.
Caboose: That window is very high. I don't think you are tall enough.
Sarge: I know. I need you to help me look through it.
Caboose: I don't think I am tall enough either.
– Red vs Blue

Overview

Having just finished three chapters on trigonometry, a return to algebra is long overdue, and in this chapter, we return to a very old topic: exponents and powers. The difference here is that we pay special attention to the *exponent*, as opposed to the *base*; whereas you previously looked at functions with a *variable* base and a *constant* exponent (i.e. polynomials), you will now see functions with a *constant* base and a *variable* exponent.

Exponential Functions

An *exponential function* is a function that takes the form

$$y=ab^x$$

where a and b are constants.

(Remember that you *cannot* combine a and b, because the order of operations dictates that exponentiation happens before multiplication. Although this may seem obvious, consider

$$y=2(3)^x$$

which is most certainly *not* equivalent to

$$y=6^x$$

despite initial appearances.)

Although this doesn't look like anything new, let us examine it more closely. (For now, let us assume $a=1$, so as to simplify our analysis.)

Unlike with the similar-looking polynomial functions, the *base* is now *constant*. Our variable has now become the *exponent* instead of the *base*, and make no mistake: this is a considerable difference. The functions

$$y=2^x$$

and

$$y=x^2$$

are *very* different, even if the only change was having the positions of the x and the 2 swapped.

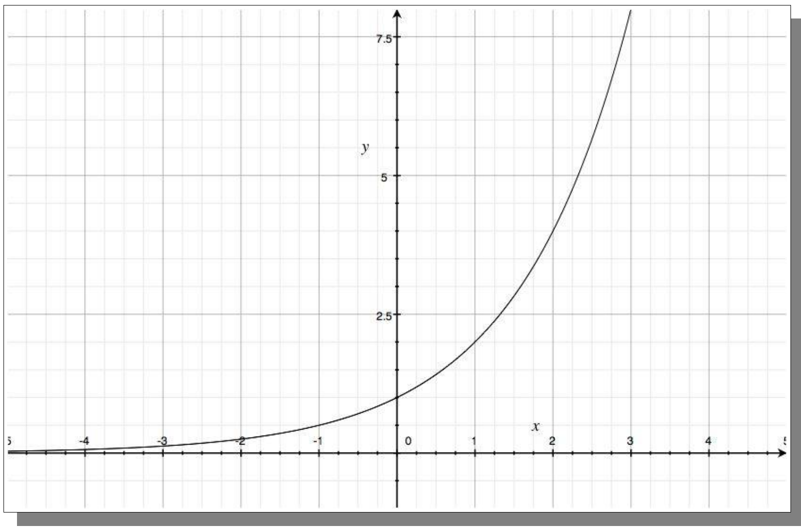
But how are they different? To answer that, let us first make a graph and then make some notes on it.

A table of values for the function $y=2^x$ is easy enough to generate:

x	$y=2^x$

-2	1/4
-1	1/2
0	1
1	2
2	4
3	8
4	16

And we can plot these points and draw ourselves a curve passing through these points:



Now, what can we notice about the graph itself?

- It passes through the point (0, 1). In fact, *every* exponential function $y=b^x$ does, since b to the power of 0 equals to 1 for any b.
- As x keeps going farther left along the x-axis, the curve appears to get closer and closer to 0, but never quite reaches it. Thus, the x-axis is a horizontal asymptote of the exponential function.
- As x keeps going farther right along the x-axis, the curve appears to grow faster and faster. (In fact, exponential functions grow faster than *any* polynomial function given enough time, but that's a topic for calculus.)
- The domain of the function includes all real numbers, as previously mentioned.
- The range of the function includes all real numbers greater than 0.

Now, suppose we want every exponential function to look something like this. In that case, what restrictions must we place on the choice of base?

- The base cannot be 1, because $1^x = 1$ for any x. Thus we will only have ourselves a constant function.
- The base cannot be negative because the definition of raising negative numbers to fractional powers is extremely tricky. For example, if we had the function $y=(-1)^x$, the function is well defined when $x=1/3$, but not defined at all when $x=1/2$. Issues like these lead us to remove negative numbers as possible choices for a base.
- The base cannot be 0, for similar reasons (constant for positive powers, undefined otherwise).

If the base is greater than 1, it will essentially look like the graph shown above. On the other hand, if the base is between 0 and 1, the graph will look like a mirror image across the y-axis. As a demonstration of why this follows:

$$y=\left(\frac{1}{2}\right)^x$$

is equivalent to

$$y=(2^{-1})^x$$

which, by the laws of exponents, gives us

$$y=2^{-x}$$

which, by the transformation techniques we found in the transformations chapter, is simply the graph of

$$y=2^{x}$$

mirrored across the y-axis.

Example: Graph the function

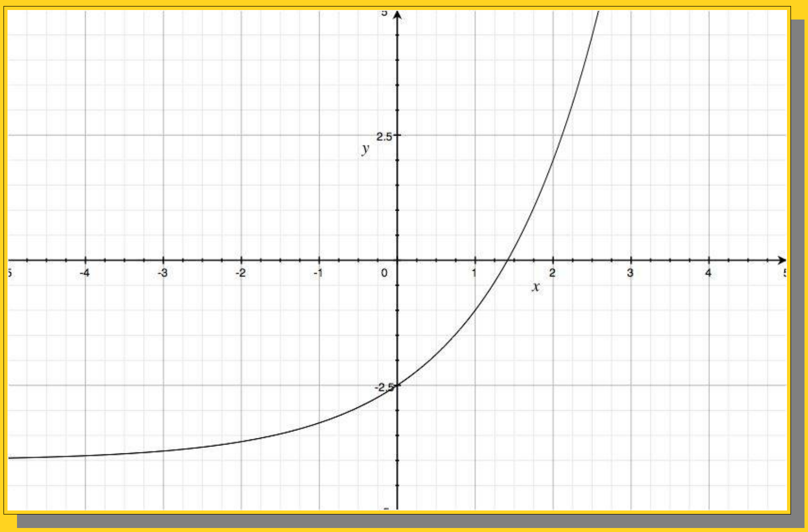
$$y=3(2^{x-1})-4$$

and find the equation of the horizontal asymptote.

Solution: We know how to graph the function

$$y=2^{x}$$

so we only have to perform some transformations on it: a vertical expansion by a factor of 3, a horizontal shift of 1 right, and a vertical shift of 4 down. Applying these results in



As for the horizontal asymptote, our graph clearly shows the curve approaching -4 as x decreases, so we can say that the equation is

$$y=-4$$

(Notice that it is simply the result of the original asymptote, $y=0$, with the vertical shift transformation applied to it. This works in general, which saves the trouble of graphing.)

Logarithmic Functions

Given the equation

$$a=b^{x}$$

we obviously know how to calculate a, through the use of exponentiation. We can also calculate b, through the use of radicals:

$$b=\sqrt[x]{a}$$

But how to calculate x? Unlike addition and multiplication, exponentiation is *not* commutative, meaning that we *cannot* rewrite the equation as

$$a=x^{b}$$

and use radicals again. Therefore, we need *another* inverse operation for exponentiation, one that solves for the *exponent* (radicals solve for the *base*). This inverse operation is a function known as a *logarithm*, and will prove to be extremely useful.

Logarithmic functions are defined as the inverse of exponential functions. That is, given the exponential function

$$y = a^x$$

the corresponding logarithmic function is

$$x = a^y$$

which, when solving for *y*, becomes

$$y = \log_a x$$

which reads "*y* is equal to the base-*a* logarithm of *x*".

As with exponential functions, the choice of base is important, and cannot be negative, 0, or 1. (We will later see that certain bases are considered very special and are used very frequently.)

Also, since exponential and logarithmic functions are inverses, they cancel each other out. That is:

$$\log_a a^x = x$$

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

Of course, this is still just a bunch of theory, and is likely to be confusing. Let's consider a few examples:

- 10 to the power of 1 is 10, so the base-10 logarithm of 10 is 1.
  - $10^1 = 10 \;\; \Rightarrow \;\; \log_{10} 10 = 1$
- 10 to the power of 2 is 100, so the base-10 logarithm of 100 is 2.
  - $10^2 = 100 \;\; \Rightarrow \;\; \log_{10} 100 = 2$
- 10 to the power of 3 is 1000, so the base-10 logarithm of 1000 is 3.
  - $10^3 = 1000 \;\; \Rightarrow \;\; \log_{10} 1000 = 3$
- 10 to the power of 4 is 10000, so the base-10 logarithm of 10000 is 4.
  - $10^4 = 10000 \;\; \Rightarrow \;\; \log_{10} 10000 = 4$
- 10 to the power of 0 is 1, so the base-10 logarithm of 1 is 0.
  - $10^0 = 1 \;\; \Rightarrow \;\; \log_{10} 1 = 0$
- 10 to the power of -1 is 0.1, so the base-10 logarithm of 0.1 is -1.
  - $10^{-1} = 0.1 \;\; \Rightarrow \;\; \log_{10} 0.1 = -1$

In essence, logarithms are the exponent a certain number (the base) needs to be raised to, in order to obtain another number.

Example: Find

$$\log_3 81$$

Solution: We need to find the power that 3 needs to be raised to in order to obtain 81. In other words, we are solving this equation for *x*:

$81 = 3^x$

By trial and error:

$3^1 = 3$

$3^2 = 9$

$3^3 = 27$

$3^4 = 81$

so the base-3 logarithm of 81 is 4.

Of course, the numbers thus far have been deliberately chosen for simplicity. For example, the base-2 logarithm of 10 is the solution to the equation

$$2^x = 10$$

However, a little trial and error will convince you that this equation has no easy solution. It would be nice to be able to obtain a decimal approximation, at least, but we have no means of doing even that. Therefore, we first need to study some properties of logarithms, which we will do in the next section.

More on Logarithmic Functions

Thus far, we have only introduced the logarithmic functions formally. We used little more than guessing to solve the equation

$$3^x = 81$$

which we could have done without talking about logarithms beforehand.

It is now time to demonstrate why these functions are actually *useful*, and how to make use of them. We begin with the fundamental property of a logarithm (which holds for any choice of base):

$$\log(ab) = \log(a) + \log(b)$$

In other words, the logarithmic function reduces multiplication to addition.

This can be proven in many ways, but intuitively, it follows because the exponential function, its inverse, does the exact opposite:

$$x^{a+b} = x^a x^b$$

Similarly, logarithms reduce division to subtraction:

$$\log\left(\frac{a}{b}\right) = \log(a) - \log(b)$$

Logarithms also reduce exponentiation to multiplication:

$$\log(a^b) = b \log a$$

These properties are often used to solve equations like the following.

Example: Solve the equation

$$3(2^{x-1}) = 48$$

Solution: First we divide both sides by 3, like usual:

$$2^{x-1} = 16$$

As the variable is present in an exponent, we have no methods of dealing with it directly. However, suppose we take the base-2 logarithm of both sides, like so:

$$\log_2(2^{x-1}) = \log_2 16$$

By the laws of logarithms, the exponentiation on the left side can be reduced to multiplication, so that the variable becomes easy to solve for:

$$(x-1)\log_2 2 = \log_2 16$$

Now by inspection, we know that

$$\log_2 2 = 1$$

because

$$2^1 = 2$$

and we also know that

$$\log_2 16 = 4$$

by similar reasoning. Therefore, we have

$$(x-1)(1) = 4$$

which, when solving for x, gives us

$$x = 5$$

which is the solution we were looking for.

Now, at this point, you may still not be impressed by the properties of logarithms. "So what? I could have guessed that solution too!" Well, yes, you could have, but that was only because the numbers were carefully chosen to give nice, integer solutions. As an illustration of how questions can suddenly get much harder, consider the equation

$$2^x = 10$$

which, of course, looks simple at first glance. But, after some thought, you should notice that there is *no* nice solution. We know that it must be equal to

$$\log_2 10$$

but we have no way of figuring out what that number is.

So, the obvious solution is to get a decimal approximation for it. And indeed, if you pull out your scientific calculator and take a look at it, you should notice a button labelled "log". (If you do not, you need a better scientific calculator.) Obviously, this button is for calculating logarithms, but with one catch: it only calculates logarithms to the base 10.

Enter another rule of logarithms to the rescue: the *change-of-base rule*. This rule is extremely crucial for calculating logarithms of any base using a calculator. It states that

$$\log_b x = \frac{\log_a x}{\log_a b}$$

In other words, if we want to calculate a logarithm using an inconvenient base b, we instead use a more convenient base a, then divide by the base-a logarithm of b.

Example: Solve the equation

$$2^x = 10$$

and give a decimal approximation to six decimal places.

Solution: We know from before that

$$x = \log_2 10$$

However, we know from the change-of base rule that

$$x = \frac{\log_{10} 10}{\log_{10} 2}$$

which, when we use our calculator, gives us

$$x = 3.321928$$

using six decimal places.

The base-10 logarithm is so convenient that it is customary to omit the subscript of 10. In other words, we define

$$\log x = \log_{10} x$$

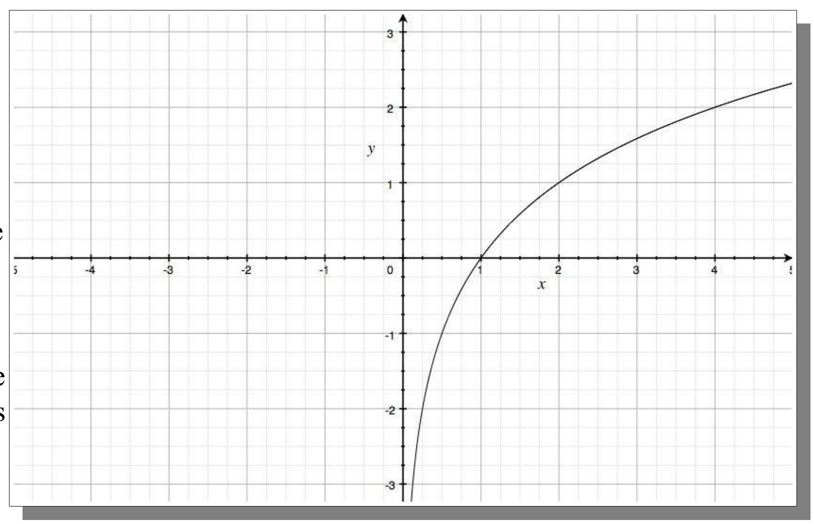
to save us some writing. (The base-10 logarithm is also called the *common logarithm* for this reason.)

(Note that in some disciplines, omitting the base does not mean a base-10 logarithm. For example, advanced mathematics usually refers to a base known as *e* (which we will discuss later), and computer science refers to the base 2.)

Graphs of Logarithmic Functions

As the inverses of exponential functions, the graph of a logarithmic function is simply the graph of an exponential function mirrored across the line *y=x*.

An example would be:



We can draw many parallels between this graph and the graph of an exponential function:

- The x-intercept is always 1, because *log(1)=0* for any base. (Compare with exponential curves always having a y-intercept of 1.)
- There is no y-intercept, although the curve appears to be heading towards “negative infinity” as it nears the y-axis. (Compare with exponential curves having no x-intercept.)
- As *x* moves to the right, the curve is growing, but it appears to be continuously *slowing down* in its growth (compare with exponential curves, which appear to be continuously speeding up). In fact, it can be shown (in calculus) that logarithmic curves grow more slowly than *any* polynomial curve. (Nevertheless, the curve does not have any horizontal asymptotes.)
- The domain of the function includes all real numbers greater than 0, and the range of the function includes all real numbers. (Compare with exponential functions having the exact opposite properties.)

Although not shown here, when the base of the logarithm is between 0 and 1, the curve is simply vertically reflected across the x-axis.

Some Applications of Exponential Functions

There are two main uses for exponential functions that are covered in Math 12:

- Exponential functions are used to model exponential growth or decay (that is, growth/decay that is directly proportional to the amount of "stuff" in the system).
- Exponential functions are used to model compounding interest.

Both will be discussed in this section.

As previously mentioned, exponential functions are often used to model problems involving rates of growth or decay *that are directly proportional to the amount of "stuff" in the system*.

For example, suppose we have some bacteria, and we wish to model its growth. Bacteria reproduce by splitting themselves into two, and a little thought should convince you that the more bacteria are present, the more reproduction happens – and thus *the faster the population grows*. In other words, the growth rate of the bacteria population is directly related to how many bacteria there are.

Continuing with this example, suppose we start with a population of 100 bacteria, and the population doubles every two hours. How long would it take before we reach a population of 5000 bacteria?

Well, first, we must model the situation. Since the population starts at 100 bacteria, we begin with:

$$A=100$$

where A is the amount of bacteria present in the system at any given time.

Now, the population *doubles*, so we add in an exponential function of base 2:

$$A=100(2^t)$$

(t denotes the time that has elapsed, in hours. When $t=0$, $A=100$, as expected – we initially begin with 100 bacteria.) The choice of base 2 was made because the bacteria population *doubles*.

Since the population doubles *every two hours*:

$$A=100(2^{t/2})$$

The reason for the division by two is because we only want a doubling to occur every *two* hours. So, when we use division by two, then when *two* hours have elapsed, $t=2$, the exponent becomes 1, and *one* doubling occurs, as expected.

And now we have the function that models the growth rate of the bacteria. Now, to find out how long it takes to get the population to 5000 bacteria, we set A equal to 5000 and solve for t, which can easily be accomplished using logarithms:

$$5000=100(2^{t/2})$$

$$5=2^{t/2}$$

$$\log_2 5=\frac{t}{2}$$

$$t=2\log_2 5$$

We can use the change-of-base rule to manipulate this expression into a more convenient form:

$$t = 2\frac{\log 5}{\log 2}$$

(Remember that we assume base 10 if no base is specified.) Punching these numbers into our calculator gives us

$$t = 4.644$$

to three decimal places. Therefore, we can say that it takes roughly 4.644 hours for the population to reach 5000 bacteria.

In general, exponential growth/decay rate problems can be modelled with the equation

$$A = A_0 r^{t/d}$$

where

- A is the amount of "stuff" in the system at any given time
- A_0 is the amount of "stuff" in the system initially (when $t=0$)
- r is the growth rate (> 1 for growth, < 1 for decay)
- t is the amount of time elapsed
- d is the length of one growth/decay period, and **it must have the same units of time as *t*** (for example, if a population doubles every 4 days, d would be 4, assuming t is also in days)

Example: Radon-216 has a half-life of about 455 microseconds. How long will it take for a sample of Radon-216 to decay to 10% of its original amount?

Solution: By definition, the *half-life* of a substance is the amount of time required for the substance to decay to one-half of its original amount. Therefore, our choice of exponent will be 0.5.

Since the half-life of Radon-216 is about 455 microseconds, we will adopt microseconds as our unit of time and use $d=455$. We do not know the initial amount of Radon-216 that we have, but we don't need to know – we only need to know that after a certain amount of time, 10% is left. Therefore,

$$A = 0.1 \times A_0$$

and our equation becomes

$$0.1 = 0.5^{t/455}$$

Taking the base-10 logarithm of both sides:

$$\log 0.1 = \log(0.5^{t/455})$$

Reduce the exponentiation to multiplication by the laws of logarithms:

$$\log 0.1 = \frac{t}{455}\log 0.5$$

Finishing solving for t:

$$t = 455\frac{\log 0.1}{\log 0.5}$$

As a numerical value, this is about 1511 microseconds.

(If you were wondering why we chose the base-10 logarithm instead of, say, the base 0.5 logarithm, it just has to do with the fact that we can work with base-10 logarithms directly on our calculators.)

Another use of exponential functions is to model compound interest problems. *Compound interest* refers to adding interest on previous interest.

As an example, let us take \$1000, with an interest rate of 1%. With *simple interest*, we would continuously add 1% of the initial \$1000 (which is \$10) – so, we would first have \$1010, then \$1020, then \$1030, and so on – **the interest is only calculated on the initial amount**.

Contrast this to compound interest, where we add 1% of *whatever we currently have* – so, given the same initial \$1000, we first have \$1010, **but then we add 1% of this new \$1010**, not \$1000, so we then have \$1020.10, then \$1030.30, and so on.

Compound interest can be modelled with the general equation

$$A = A_0\left(1+\frac{r}{n}\right)^{nt}$$

where

- A is the amount at any given time
- A_0 is the initial amount (when $t=0$)
- r is the interest rate (as a decimal, so 1% would be 0.01, for example)
- n is the number of times the interest is compounded per unit time (using the same unit of time as t, so if t were in years, for example, then n would be the number of times the interest is compounded per year)
- t is the amount of time that has passed (using any unit of time)

Example: Given an initial investment of \$5000 compounded twice a year at a rate of 4%, how many years will it take for the investment to double?

Solution: We first use the equation:

$$10000 = 5000\left(1+\frac{0.04}{2}\right)^{2t}$$

(The value for n comes from the fact that the investment compounds twice a year.)

We then use logarithms to solve for t:

$$2=(1+0.02)^{2t}$$
$$\log 2=\log(1.02^{2t})$$
$$\log 2 = 2t \log 1.02$$
$$t=\frac{\log 2}{2\log 1.02}$$

As a decimal approximation, it takes roughly 17.5 years.

Natural Logarithms

The most-often used base for logarithms is 10, simply as a matter of convenience: our numbering system is base 10, and our calculators have buttons for computing base-10 logarithms directly.

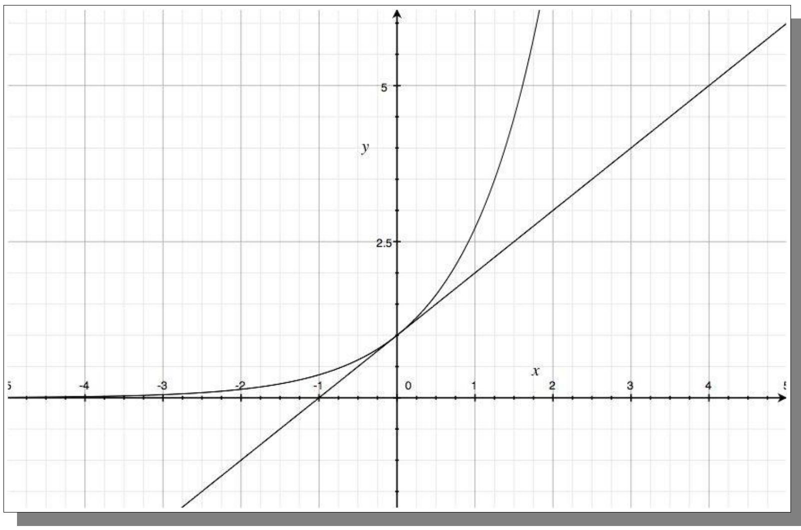
From a mathematical perspective, though, the number 10 is nothing special. Therefore, as with the case of angle measurement, we should ask ourselves if there exists a more *natural* (mathematically speaking) choice of base, instead of our arbitrary choice of 10.

Naturally (no pun intended), such a base exists – the base e. e is a special mathematical constant, like π, and its numerical value is approximately 2.71828. (Your calculator should have a button somewhere with this value.)

As with radians, the inevitable question is: why is base e so special? There are many reasons, most of which require knowledge of calculus to fully appreciate, but here is one reason: consider the exponential curve

$$y = e^x$$

which is a normal exponential curve, except with e as our choice of base. Graphing this gives us the graph shown below:



Notice the additional tangent line drawn in. All exponential curves (with base > 1) can have a similar tangent line drawn at the y-intercept, but when the base is e, the tangent line has a very special property:

Its slope is exactly 1.

(If we graph the base-e logarithm, we will find that it also has a tangent line with a slope of exactly 1 at its x-intercept.)

For the purposes of Math 12, however, the concept of slope is irrelevant. Instead, the choice of base e is introduced because of its use in applications – as an example, consider the formula for compounding interest:

$$A = A_0\left(1 + \frac{r}{n}\right)^{nt}$$

What happens when n gets larger? Or, in other words, what happens when the compounding period gets smaller?

It turns out that when n is extremely large, approaching infinity (and thus the compounding period extremely short, approaching zero), the function becomes

$$A = A_0 e^{rt}$$

which is simpler to work with than the previous formula. Of course, this assumes that the compounding period really is extremely short, but this is often a reasonable approximation.

Example: Given an initial investment of \$5000 compounded continuously at a rate of 4%, how many years will it take for the investment to double?

Solution: The problem is exactly the same as the one given in an earlier section ("Some Applications of Exponential Functions"), but now we assume *continuous* compounding instead of only twice a year. Setting up the equation:

$$10000 = 5000 e^{0.04t}$$

Solving for t eventually gives us

$$t = \frac{\log 2}{0.04 \log e}$$

which, as a decimal approximation, is about 17.3 years – fairly close to the value of 17.5 years we found before.

The base-e logarithm is also called the *natural logarithm*, and is denoted "ln" (pronounced "lawn"). It is also often found as a separate button on most scientific calculators, eliminating the need for the change-of-base rule when using base e.

Geometric Sequences and Series

O'Brien: It's an old naval tradition. Whoever's in command of a ship, regardless of rank, is referred to as "captain".

Nog: You mean if I had to take command, I would be called "captain", too?

O'Brien: Cadet, by the time you *took command, there'd be nobody left to call you anything.*

– Star Trek Deep Space Nine

Overview

A sequence of numbers that follows a certain pattern is likely to be nothing new to you – you've probably looked at many over the course of your mathematics education, in a question such as:

Fill in the missing numbers:
1, 3, 5, __, __, __, 13

And you would probably look at that and instantly say "7, 9, 11".

Now, Math 10 was your first *formal* introduction to number sequences, or, more specifically, one very common type of sequence: *arithmetic sequences*, or in other words, sequences whose terms were separated by a common *difference*. (For example, the example sequence above is an arithmetic sequence with a common difference of 2, since each number is 2 more than the one before it.)

Fast forward to Math 12, where we now look at another common type of sequence: *geometric sequences*. These, like arithmetic sequences, follow specific patterns, but in the case of a geometric sequence, the terms are separated by a common *ratio*. An example would be 1, 2, 4, 8, 16... where the numbers are separated by a common ratio of 2, since each number is twice as large as the one before it.

Naturally, you can also sum the elements of a geometric sequence, turning it into a *geometric series*. We will, of course, find a formula for doing so, and also find that in certain cases, you can add an *infinite* geometric sequence – and still get a finite answer! This finding enables us to answer *Zeno's paradoxes*, which are a set of ancient philosophical questions about the possibility of motion.

Geometric Sequences

As previously mentioned, a *geometric sequence* is a sequence where the terms are separated by a common ratio. The most basic example is the sequence 1, 2, 4, 8, 16... which has a common ratio of 2 (since each term is twice as large as the one before it). (Of course, the common ratio need not be 2, and for that matter, it need not even be a number.)

Geometric sequences have two fundamental properties: their *common ratio*, which was already explained, and their *starting term* (or *initial term*), which is the first term in the sequence. Both properties are needed to completely describe a geometric sequence.

Now, how do we describe the terms of a geometric sequence? The first term is obviously the starting term, which we will denote with a. But what about the second term? Well, since we know that there is a common ratio (which we will denote r) between each term, the second term must be the first term multiplied by the common ratio, which works out to be ar – that is, a multiplied by r, as previously stated.

And what of the third term? We know that it is simply the second term multiplied by the common ratio, so we have $(ar)r$, which simplifies to ar^2. By similar reasoning, we can find that the fourth term is ar^3, and so on.

In general, the nth term in a sequence takes the form

$$t_n = ar^{n-1}$$

where a is the starting term and r is the common ratio. (The reason for the -1 in the exponent is because the first term is simply a, without any multiplications by r; therefore, the exponent will always end up being one less than the term number.)

Note that because of the presence of exponents, logarithms are required whenever we wish to solve for n.

Example: If the second term in a geometric sequence is 12 and the sixth term is 15552, what is the starting term and the common ratio of the sequence?

Solution: A few guesses will likely fail miserably, so we seem to have no choice but to search for an algebraic solution. We write down the information given:

$$t_2 = ar = 12$$

$$t_6 = ar^5 = 15552$$

We have a system of two (nonlinear) equations and two unknowns, which we can solve by substitution. Alternatively, we can divide t_6 by t_2:

$$\frac{t_6}{t_2} = \frac{ar^5}{ar} = r^4 = \frac{15552}{12} = 1296$$

Taking the fourth root of r gives us

$$r = \pm 6$$

Notice that there are *two* possible answers! This often occurs when solving problems such as these, and you must take care to consider both solutions.

Now, if we take $r=6$, then by substitution, we can easily find that $a=2$. On the other hand, if we take $r=-6$, then $a=-2$, so the solutions are either

$$a = 2, r = 6$$

or

$$a = -2, r = -6$$

Of course, you may still be wondering how it is possible to have two possible answers. This is best answered with a demonstration, by showing both possible sequences:

2, 12, 72, 432, 2592, 15552, ...

-2, 12, -72, 432, -2592, 15552, ...

Geometric Series

We can, of course, always turn a geometric sequence into a series by summing the elements. However, manual adding is obviously too time-consuming, so a general formula is desired.

So, consider the sum

$$S_n = a + ar + ar^2 + \ldots + ar^{n-1}$$

Now consider what happens when everything is multiplied through by *r*:

$$rS_n = ar + ar^2 + ar^3 + \ldots + ar^n$$

Subtract the second equation from the first:

$$rS_n - S_n = (ar^n + ar^{n-1} + \ldots + ar^2 + ar) - (ar^{n-1} + ar^{n-2} + \ldots + ar + a) = ar^n - a$$

(Notice how most of the terms simply cancel out.) Now, if we factor both sides:

$$S_n(r-1) = a(r^n - 1)$$

This means that the formula for the sum of a geometric series is

$$S_n = a\frac{r^n - 1}{r-1}$$

which can also be expressed as

$$S_n = a\frac{1-r^n}{1-r}$$

by dividing the fraction by -1.

(We should observe that the formula fails when $r=1$, due to division by zero. But, if *r* were actually equal to 1, then all the terms would be equal, making addition of them straightforward.)

Example: Suppose Alice invests \$500 every year for 10 years. If the bank pays 3% interest per year, how much money will she have at the end of 10 years?

Solution: Rather than trying to plug numbers into a formula immediately, let us consider a few initial cases.

- In the first year, she invests \$500, and at the end of it, she will have 3% interest on top of that, or 500(1.03).
- In the second year, she invests another \$500, and at the end of it, she will have 3% interest on top of that, or $(500(1.03) + 500)(1.03) = 500(1.03) + 500(1.03)^2$.
- In the third year, she invests another \$500, and the ultimate result is $(500(1.03) + 500(1.03)^2 + 500)(1.03) = 500(1.03) + 500(1.03)^2 + 500(1.03)^3$.

We can see that this works out to be a geometric series, but the initial term is 500(1.03) and not just 500, because we're counting using the *ends* of the years, when the interest has been applied.

In any case, plugging these numbers in gives us

$$S_{10} = 515\frac{1.03^{10} - 1}{1.03 - 1}$$

which works out to equal 5903.90 when rounded to two decimal places.

Infinite Geometric Series

The concept of infinity is something of an enigma; every student *thinks* he or she understands what it means, because it is such an intuitively understood concept.

But consider the addition of an *infinite* series of numbers, such as

$$1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+...$$

which is, of course, a geometric series with initial term 1 and common ratio 0.5.

At first glance, you may be tempted to say that the sum of the series is infinity, since we are adding an infinite number of numbers. However, the problem is that **we don't really know what it means to add an infinite number of numbers**. It's easy to say "just add numbers an infinite amount of times", but if that were really what we had to do, we would be sitting around here forever.

So instead, let us consider what we know about *finite* geometric series. Taking the above geometric series, we know that the sum for *n* terms must be

$$S_n=\frac{0.5^n-1}{-0.5}=-2(0.5^n-1)$$

Now let us consider what happens as we sum more and more terms in the series. When we take the first eight sums:

$$S_1=-2(0.5^1-1)=1$$
$$S_2=-2(0.5^2-1)=1.5$$
$$S_3=-2(0.5^3-1)=1.75$$
$$S_4=-2(0.5^4-1)=1.875$$
$$S_5=-2(0.5^5-1)=1.9375$$
$$S_6=-2(0.5^6-1)=1.96875$$
$$S_7=-2(0.5^7-1)=1.984375$$
$$S_8=-2(0.5^8-1)=1.9921875$$

It shouldn't take a genius to realize that as we add more and more terms in the series, the sum gets closer and closer to 2. **Therefore, we can reasonably claim that the *infinite* series sums to 2** – a result that is perhaps surprising: *an infinite series (in this case) sums to a finite number.*

We can now make a reasonable definition for what it means to sum an infinite geometric series: by defining it as the sum that partial, *finite* series appear to approach as we consider more and more terms. (A mathematical definition requires calculus, but our definition says almost exactly what the math would say.)

Of course, not all infinite geometric series have a finite sum; the series

$$1+2+4+8+16+...$$

obviously sums to infinity (which can be shown by considering partial sums, as before, and showing that the partial sums do not approach any one number, but rather just grow indefinitely).

So, at this point, we have two major questions to answer:

- When does an infinite geometric series have a finite sum?
- If an infinite geometric series has a finite sum, what *is* that sum?

Looking at earlier examples, it seems that the primary criteria for determining whether an infinite geometric series has a finite sum or not is the common ratio. When it was 0.5, the series had a finite sum, but when it was 2, the series had no finite sum. So what differentiates these numbers? Consider the expression

$$0.5^n$$

What happens as *n* gets larger? It seems (as can be shown if you punch numbers into your calculator) that increasing *n* has the effect of *decreasing* the value of the entire expression. We can even make the reasonable statement that **when *n* is infinity, 0.5^n is actually equal to 0**.

Compare this with the expression

$$2^n$$

where as *n* gets larger, 2^n grows larger, never reaching any particular finite value.

Recall from the previous chapter that when the base of an exponential function is between 0 and 1, the graph is reflected across the y-axis, meaning that as *x* grows larger, the y-coordinate approaches 0. Therefore, we can conclude that the common ratio must be between 0 and 1. (Unlike exponential functions, however, the common ratio can be negative or 0, so instead we will say that **the common ratio must be between -1 and 1**.)

Now, what *is* the actual sum, then? Well, consider the formula for the sum of a finite series:

$$S_n = a\frac{1-r^n}{1-r}$$

If *r* is between -1 and 1, then r^n goes to 0 as *n* increases, so we can say that

$$S = a\frac{1-0}{1-r} = \frac{a}{1-r}$$

for an infinite geometric series.

Remember that the above formula only works when *r* is between -1 and 1 – otherwise, the formula is not valid, and the infinite series has no finite sum.

Example: When a particular ball bounces on the ground, it bounces back up to 80% of its previous height. If we drop this ball from a height of 1 meter, how much total distance does the ball cover?

Solution: While it would be tempting to immediately apply the formula for summing an infinite geometric series with *a=1* and *r=0.8*, this is not quite accurate. The ball travels both up *and down* a certain distance, so ignoring the initial drop, every drop is accompanied by a bounce of the same height. Thus, we must set up an infinite series with the *second drop* as the initial term and multiply the result by 2 (because every drop *except the initial drop* has a bounce associated with it), then add the initial drop of 1 at the end.

$$d = 2\left(\frac{0.8}{1-0.8}\right) + 1 = 2(4) + 1 = 9$$

Therefore, the ball covers a total distance of 9 meters.

Zeno's Paradoxes

Note: this material is completely optional.

Zeno was an ancient Greek philosopher, famous for a set of logical paradoxes he is credited for. Here, we will examine one of his paradoxes of motion, and show how geometric series play a role in the resolution of it.

The story involves a race between Achilles and the tortoise. Suppose that the tortoise is given a head start. Achilles, of course, runs faster, so we would expect him to catch up to the tortoise at some point. But when he catches up to where the tortoise was at the start of the race, the tortoise has moved forward a bit. When Achilles reaches *that new point*, the tortoise has moved forward yet again. And so on, and so on, **meaning that Achilles will never catch up to the tortoise**!

But how is this possible? We know from intuition and experience that Achilles *must* catch up to the tortoise and overtake him eventually.

So to make things easier on ourselves, let's try a concrete example with numbers. Suppose Achilles runs at 10 m/s and the tortoise at 1 m/s. Further suppose that the tortoise is given a head start of 10 m.

Achilles then has to run 10 m to catch up to the tortoise, so he runs 10 m in 1 s. The tortoise moves 1 m in that time, so the tortoise is at the 11 m mark. So Achilles has moved 10/11 of the distance to the tortoise, with a total time of 1 s.

Achilles then has to run 1 m to catch up to the tortoise, so he runs 1 m in 0.1 s. The tortoise moves 0.1 m in that time, so the tortoise is at the 11.1 m mark. So Achilles has moved 11/11.1 of the distance to the tortoise, with a total time of 1.1 s.

Achilles then has to run 0.1 m to catch up to the tortoise, so he runs 0.1 m in 0.01 s. The tortoise moves 0.01 m in that time, so the tortoise is at the 11.11 m mark. So Achilles has moved 11.1/11.11 of the distance to the tortoise, with a total time of 1.11 s.

By this point, we can see that the total distance Achilles travels to catch the tortoise is

$$10+1+0.1+0.01+...$$

and the total time he takes to do so is

$$1+0.1+0.01+0.001+...$$

where both series are infinite.

But these series are *geometric*, meaning that we can use our knowledge of summing infinite geometric series to look at these! The series representing the distance he covers has an initial term of 10 and a common ratio of 0.1, while the series representing the time needed has an initial term of 1 and a common ratio of 0.1.

Since the common ratio is between -1 and 1, these infinite series sum to a *finite number*, as our formula shows:

$$distance=\frac{10}{1-0.1}=11.111\,...\,meters$$

$$time=\frac{1}{1-0.1}=1.111\,...\,seconds$$

Therefore, Achilles travels a *finite* distance in a *finite* amount of time to catch up to the tortoise. Thus the paradox is resolved, with the mathematics of infinite geometric series. (The fallacy in Zeno's paradox is in assuming that an infinite series must sum to infinity.)

Sigma Notation

Mathematicians are as lazy as anyone else, and so they came up with notation to compactly express a series, called *sigma notation*. The name refers to the Greek capital letter sigma, Σ.

Sigma notation is quite difficult to describe simply by using words, so let us consider an example: a geometric series with initial term *a* and common ratio *r*, and *n* terms:

$$\sum_{k=0}^{n-1} ar^k$$

Now to examine this expression:

- We begin by looking at the sigma symbol itself, and what is below it: $k=0$.
- 0 is called the *lower bound of summation*, and it marks where the series begins.
- Above the sigma symbol, $n-1$ is called the *upper bound of summation*, which marks where the series ends. (If the series is infinite, the upper bound is marked with an infinity symbol.)
- *k* is called the *index of summation*, **which is a variable that starts at the lower bound of summation and increases by 1 each time until it reaches the upper bound of summation**. (This is the most crucial aspect of sigma notation.)
- The expression to the right side of the sigma symbol (in this case ar^k) represents how each term looks, which of course changes as *k* takes on new values.

So if we wanted to expand this expression to verify that it indeed represents a geometric series, let us start with the lower bound of summation, $k=0$. Then the first term is

$$ar^0=a$$

which is exactly what we wanted. The second term, $k=1$, is

$$ar^1=ar$$

and we continue doing this until we reach the upper bound of summation, $k=n-1$, which gives us

$$ar^{n-1}$$

The sigma symbol says to *add* the terms together (remember that the sigma symbol represents a series with its terms added), so we have

$$a+ar+\ldots+ar^{n-1}$$

which is exactly what a geometric series should look like.

Note that there is also a formula for the *number of* terms in a series expressed by sigma notation:

$$n=(\mathit{upper\ bound})-(\mathit{lower\ bound})+1$$

Because of its compact nature, sigma notation is often used with geometric series, so you should get used to seeing it being used.

Example: Express the infinite geometric series

$$0.9+0.09+0.009+\ldots$$

in sigma notation and find the sum.

Solution: The initial term is clearly 0.9. As for the common ratio, we know that the second term is of form *ar*, and the first term of form *a*, so dividing the second term by the first should give us *r*. Doing this gives us $r=0.1$.

Therefore, by following previous examples, we can express this as

$$\sum_{k=0}^{\infty}(0.9)(0.1^k)$$

and the sum, by the formula for infinite geometric series, is

$$S=\frac{0.9}{1-0.1}=1$$

Notice something very interesting, however. We know that 0.9 + 0.09 = 0.99, and we also know that 0.9 + 0.09 + 0.009 = 0.999, and so on. Therefore, the final number should be 0.999... with an infinitely long string of 9s – but according to our formula, we found that the sum should equal 1! (What do you think is the answer to this paradox?)

One last interesting thing to note is that sigma notation obeys the rules of linear operators – in other words,

$$\sum_{k=0}^{n-1} ar^k = a\sum_{k=0}^{n-1} r^k$$

(you can move a **constant** factor – that is, one not dependent on the index of summation – outside of the sigma symbol if you feel like it) and

$$\sum_{k=0}^{n-1}(r^k+s^k)=\sum_{k=0}^{n-1} r^k+\sum_{k=0}^{n-1} s^k$$

(you can split a sum into individual sums). These properties are not necessary to know for the purposes of Math 12, but they are presented here for your interest.

Permutations and Combinations

Zuko: How can you forgive me so easily?! I thought you'd be furious with me!

Iroh: I was never angry with you. I was sad, because I was afraid you'd lost your way.

Zuko: I did lose my way.

Iroh: But you found it again! And you did it by yourself! And I'm so happy *you found your way here.*

Zuko: It wasn't that hard, Uncle. You have a pretty strong scent.

– Avatar: The Last Airbender

Overview

The final area of study in Math 12 is the study of probability, a new topic in the Principles of Mathematics curriculum. It is not really algebra in the usual sense of the word, and it is certainly not trigonometry, but it is an extremely useful branch of mathematics, one whose applications are immediately obvious when we play poker or the like.

A prerequisite for the study of probability, however, is the study of permutations and combinations – in other words, the study of counting. And this, of course, is what this chapter is all about – counting the number of ways something can be done.

The Fundamental Counting Principle

Suppose we place an order for a pepperoni pizza combo. Further suppose that there are three sizes, two dipping sauces, and five drinks to choose from. Then the question is, how many different possible combos are there?

To answer this question, let us first examine only one of the choices: size. There are three sizes, so clearly we must have at least three possibilities – one for each size.

If we then look at the dipping sauces, there are two of them, so we must have two possibilities, one for each dipping sauce. However, we know that each dipping sauce can be paired up with a different size, so we must have two dipping sauces for each size – or a total of

(3 sizes)(2 dipping sauces) = 6 possibilities

Similarly, we know that each of the five drinks has the three sizes and two dipping sauces attached to it, so the grand total becomes

(3 sizes)(2 dipping sauces)(5 drinks) = 30 possibilities

Notice how exactly the total number of possibilities were counted – by multiplying the number of choices for every item together. This principle holds as a general result known as *the Fundamental Counting Principle*, which states:

If you have *a* ways of choosing one item and *b* ways of choosing another item, you have *ab* ways of choosing both.

The rest of the work we do in this chapter is based on this simple principle.

Factorials

Suppose we want to know how many ways there are to arrange the 26 letters of the alphabet.

Well, according to the Fundamental Counting Principle, we have 26 ways of choosing the first letter, 25 ways of choosing the second, 24 ways of choosing the third, and so on. Thus the principle says that the number of ways is equal to

$$26\times25\times24\times...\times3\times2\times1$$

Here, the important thing isn't the actual answer, but rather the multiplication itself – starting at the number 26 and multiplying by every number below it until we reach 1. This sort of multiplication is so common that mathematicians call it a *factorial*, and denote it with an exclamation point. For example, the above multiplication would be expressed as 26!.

We can therefore define the factorial of a number n as

$$n!=(n)(n-1)(n-2)...(3)(2)(1)$$

It is important to note that the concept of a factorial is only defined for whole numbers. (0! is defined as 1, for reasons shown later.)

An alternate way to express the factorial of any number is

$$n!=(n)(n-1)!$$

which is called a *recursive definition* because it uses itself as part of its definition. Of course, this must stop somewhere, so we add in the condition $n > 1$ to the above definition, with 1! = 1.

The recursive definition has one advantage: it gives us a definition for 0!. If we choose to express 1! recursively, we get:

$$1!=(1)(0!)$$

Since 1! = 1, solving for 0! gives us 0! = 1, which suggests that **0! should be defined as 1**. Although this definition may seem counterintuitive, there are many strong reasons for this definition, so for now we will just accept it.

Note that we already have a practical use for factorials – by the fundamental counting principle, if we want to know the number of ways to arrange *n* objects, we know that there are *n* ways to select the first object, then *n-1* ways to select the second, and so on, meaning that there are *n!* ways to arrange *n* objects.

Permutations

It is now time for us to tackle bigger problems – the sort of problems that form the core of this chapter. To solve these problems, we will start **by figuring out how to count the number of ways there are to arrange a group of objects** – a rather useful technique.

Let us take the example of the alphabet again. Suppose we would now like to how many possible ways there are to select twenty letters from the group of twenty-six. Of course, the Fundamental Counting Principle tells us that the answer is

$$26\times 25\times 24\times \ldots\times 7$$

(count the numbers if you want to verify that there are twenty numbers in the above multiplication). The problem is, this is rather tiresome to write out, and we can't use factorials directly – but luckily, we can get around this, simply by using factorials in a rather clever manner:

$$\frac{26!}{6!}=\frac{26\times 25\times 24\times \ldots\times 2\times 1}{6\times 5\times \ldots\times 2\times 1}=26\times 25\times 24\times \ldots\times 7$$

Notice that 6 = 26-20, where the former number is the number of letters to choose from and the latter number is the number of letters we want to select.

Now let's generalize this trick. Suppose we want to know how many possible ways there are select *r* objects from a group of *n* objects (where $n \geq r$). Following the example with the alphabet, the Fundamental Counting Principle tells us the answer must be

$$(n)(n-1)(n-2)\ldots(n-(r-1))$$

(the reason why we end with *r-1* is because we begin with *n-0*, so we only need to get to *r-1* to reach *r* objects). However, as before, we can express this in a much more concise manner:

$$\frac{n!}{(n-r)!}$$

This expression is so common that there is shorthand for this shorthand; therefore, we can sum up this entire section as follows:

If we wish to find out how many ways we can select *r* objects from a group of *n* of them, the formula is

$$P(n,r)={}_nP_r=\frac{n!}{(n-r)!}$$

where any one particular selection of objects is called a *permutation*. **It is important to note that this formula implies that the *order of selection of objects* is important – that is, *selecting objects in a different order is a different permutation*.**

Example: We have a race with 10 runners.

[a] How many ways are there of awarding the gold, silver, and bronze medals?

[b] Suppose half of the runners are boys and half are girls. How many ways are there of awarding the medals so that a girl gets the silver medal?

Solution: For part [a], obviously the order matters (it makes a big difference as to *which* medal you get), so this question is one involving permutations. Since there are ten runners with three medals to award, the answer is

$$_{10}P_3 = \frac{10!}{(10-3)!} = 720\,ways$$

For part [b], the order still matters, but now we have a new restriction – a girl must obtain the silver medal. Since half the runners are girls, we have five choices for the silver medal, with nine choices left over for the other two medals. According to the Fundamental Counting Principle, the total is then the *product* of these two:

$$({_5P_1})({_9P_2}) = \left(\frac{5!}{(5-4)!}\right)\left(\frac{9!}{(9-2)!}\right) = (5)(72) = 360\,ways$$

More on Permutations

There are a couple of twists often thrown into permutation problems that are worth mentioning:

- The presence of duplicates
- The grouping of objects

Let's first examine what effect duplicates have. Clearly, they must *reduce the total number of permutations*, since reordering duplicates makes absolutely no difference. But since we want to figure out exactly what effect they have, it would be best to start with a simple example: the number of permutations of the letters AABCD.

Now, if the letters were all different, the problem would be simple – there would simply be 5! permutations. However, since we have 2 As, our answer is 2! times too big, because there are 2! ways of arranging two letters. Thus, the answer is

$$\frac{5!}{2!} = 60$$

It can be shown that in general, if we have a group of *n* objects, where there are *a* duplicates of one object, *b* duplicates of another object, *c* duplicates of another object, and so forth, the number of permutations is

$$\frac{n!}{a!b!c!\ldots}$$

Now, as for the grouping of objects, what exactly do we mean by that? Well, we mean that we always want certain objects next to each other – for example, if we were sorting a class, what if we wanted all the boys together and all the girls together?

The trick here is to treat the objects you want grouped together as one super-object, and work with this super-object as you would any normal object. However, because the objects inside the super-object can be ordered among themselves, you must multiply the result by the number of ways to arrange the objects *inside* the super-object. An example will probably clarify things.

Example: If five adults and three children are in a line, how many ways can they be arranged if the children must all stay together?

Solution: Since the children *must* remain all together, it will be convenient to group the three children into one super-object. We then have five adults and one super-object, making a total of six objects. So our total so far is 6!.

However, because the super-object consists of three children, we know that there are 3! ways to arrange the children among themselves. Thus, we must multiply the result by 3!, and the final answer is

$$6!3!=4320$$

Combinations

Suppose that we want to select *r* objects from a group of *n*, as before. But this time, suppose that *we don't care about the order of selection.* In that case, we can use our permutations formula from before – but we must divide out the number of ways there are to order the things we selected.

To make the point clear, let's try a concrete example: suppose that we want to know how many ways there are to draw five cards from a standard deck (with 52 of them). Using our permutations formula gives us

$$\frac{52!}{47!}=311875200$$

However, the permutations formula assumes that order matters to us. If it doesn't – and in the case of a deck of cards, it almost always doesn't – this number is too big, because it counts different orderings separately. In fact, we know that there are 5! ways to order five cards, so this number is 5! too big. Therefore, the real result is

$$\frac{\frac{52!}{47!}}{5!}=2598960$$

Note that this can be rewritten (by some algebraic manipulation) as

$$\frac{52!}{5!47!}=2598960$$

These results suggest that we make the following definition:

If we wish to find out how many ways we can select *r* objects from a group of *n* of them, the formula is

$$\binom{n}{r}=C(n,r)={}_nC_r=\frac{n!}{r!(n-r)!}$$

where any one selection of objects is called a *combination*. **It is important to note that this formula implies that the *order of selection of objects* is irrelevant – that is, *selecting objects in a different order is the same combination*.**

Example: How many different combinations exist when playing Lotto 6/49?

Solution: The Lotto 6/49 lottery involves choosing 6 numbers in the range of 1-49, and since order is unimportant, the answer is

$${}_{49}C_6=\frac{49!}{6!43!}=13983816$$

which is the magic number often quoted about the odds of this lottery.

Pascal's Triangle

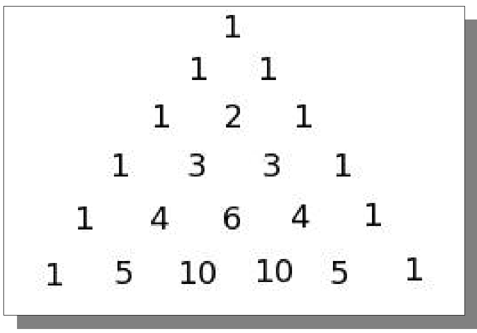
Although this chapter is not really about algebra, combinations play a useful role in what is known as the *binomial theorem*, which is a formula for expanding

$$(a+b)^n$$

where n is any whole number.

Because the binomial theorem is a little complicated, we will look at it slowly, over two sections. This first section will be about *Pascal's Triangle*, a fascinating mathematical triangle, and how it relates to combinations.

So what is Pascal's Triangle? Have a look at the following image (which shows only the first few rows):



If you stare at the triangle for long enough, you should notice the key properties of this triangle:

- The outer edges are all 1s.
- Every other number is created by taking the two numbers above it and adding them together.

But while this triangle is indeed fascinating, we have to ask how this relates to combinations in any way.

Well, as it turns out, *the numbers in Pascal's Triangle can be generated by combinations* – in fact, the $(r+1)$th number in the $(n+1)$th row is the exact same number as ${}_nC_r$. (The reason why the +1's exist in the expressions is that we are calling the top row the first row, and the left column the first column. If we agree to call those the zeroth row and column instead, the expressions become the much cleaner "rth number in the nth row". But for the purposes of Math 12, we will stick with the uglier expressions.)

Example: What is the 4th number in the 20th row of Pascal's Triangle?

Solution: Because we are adopting the convention of referring to the $(r+1)$th number in the $(n+1)$th row as ${}_nC_r$, we note that r must equal 3 and n must equal 19 – this way, $r+1=4$, and $n+1=20$, and we then refer to the 4th number in the 20th row, as desired. So then we have ${}_{19}C_3$, which is 969.

As a bonus, the relationship between Pascal's Triangle and combinations enables us to formulate some laws about combinations using the triangle (these laws can also be derived algebraically, but using the triangle is much easier):

- The triangle is symmetrical, so ${}_nC_r = {}_nC_{n-r}$.
- The triangle is generated recursively (each number is generated by adding the numbers above it), so ${}_nC_r = {}_{n-1}C_{r-1} + {}_{n-1}C_r$.

The Binomial Theorem

With the knowledge obtained about Pascal's Triangle, we are now ready to examine the *binomial theorem*, which combinations play a prominent role in.

To refresh your memory, the binomial theorem is a formula for expanding an expression of the form

$$(a+b)^n$$

where n is any whole number.

Let us first start by trying a few concrete examples. We know that

$$(a+b)^0=1$$

and we also know that

$$(a+b)^1=a+b$$

A quick application of FOIL gives us

$$(a+b)^2=a^2+2ab+b^2$$

and another expansion gives us

$$(a+b)^3=a^3+3a^2b+3ab^2+b^3$$

More tedious algebra will give us

$$(a+b)^4=a^4+4a^3b+6a^2b^2+4ab^3+b^4$$

and

$$(a+b)^5=a^5+5a^4b+10a^3b^2+10a^2b^3+5ab^4+b^5$$

By this point, multiplying everything out by hand will be extremely tiresome. Let us instead try and figure out how we can derive a general formula using these examples.

To begin with, the most obvious patterns are that a^n and b^n are the first and last terms, respectively. In between them, it seems that the powers of a and b are random – but why don't we have a closer look at, for example, the powers of the final expansion? We have

$$a^5$$
$$a^4b$$
$$a^3b^2$$
$$a^2b^3$$
$$ab^4$$
$$b^5$$

which we can rewrite as

$$a^5b^0$$
$$a^4b^1$$
$$a^3b^2$$
$$a^2b^3$$
$$a^1b^4$$
$$a^0b^5$$

because any number to the power of 0 is simply 1.

Now let us examine this column of powers closely. After a while, it should be apparent that *the powers of a keep decreasing by one each term, and the powers of b keep increasing by one each term.* Furthermore, *the powers of a and b always add up to 5, which is the exponent of the whole expansion.*

So now we know how to generate the powers in every expansion – by starting a with the value of the exponent and starting b at zero, and increasing/decreasing them until b and a swap places from where they started.

But what about the coefficients of the expansion? Well, if we simply write out the coefficients for the expansions, we will find that *the coefficients generate Pascal's Triangle* – in other words, *the coefficients can be created using combinations*.

We can now state the *binomial theorem*:

$$(a+b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \ldots + \binom{n}{n-1} a b^{n-1} + \binom{n}{n} b^n$$

where n is any whole number. Although this expression looks complicated, we are only applying the patterns we found in earlier examples. Notice that the powers of a and b still obey the rules we found earlier, and the coefficients are simply combination expressions with r beginning at 0 and increasing all the way to n – which matches Pascal's Triangle.

Sigma notation can make this quite compact:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

Notice that the sigma notation version also makes it clear exactly how each individual term looks in the expansion.

Example: What is the 8th term in the expansion of $(a-2)^{10}$?

Solution: We note that $a-2$ is equivalent to $a+(-2)$, so we simply have b replaced by -2.

As for the 8th term, we previously agreed (in the section on Pascal's Triangle) that ${}_nC_r$ would represent the $(r+1)$th number in the $(n+1)$th row. Therefore, since we want the 8th number, we need r to equal 7. n equals 10 since that is the power we are raising our binomial to (if we said the 10th *row*, then n would equal 9, but here we are not referencing rows).

If we look at the binomial theorem as written in sigma notation, it is clear that the kth term takes the form

$$\binom{n}{k} a^{n-k} b^k$$

With $n=10$, $a=a$, $b=-2$, $k=r=7$, we can now say that the 8th term is

$$\binom{10}{7} a^{10-7} (-2)^7 = (120) a^3 (-128) = -15360 a^3$$

Probability

Sang: I want twenty million in fifties.
Carter: Ok, twenty million in fifties.
Sang: Twenty million in twenties.
Carter: Ok, twenty million in twenties.
Sang: And ten million in tens.
Carter: Ten million in tens. Ok. Want any fries with that?

– *Rush Hour*

Overview

As previously mentioned, the study of permutations and combinations is a prerequisite for the study of probability, which we can now do here, in this final chapter of Math 12.

We obviously cannot look at the entire subject here in a single chapter; probability, as a branch of mathematics, could fill up many books on its own, to say nothing of a few pages in a document. Nevertheless, we can spend enough time to get a feel for the subject and understand some basic, fundamental ideas.

Experimental and Theoretical Probability

We can divide the concept of probability into two: experimental and theoretical.

Experimental probability is the probability found by experiment. So, as an example, if you flip a coin ten times and eight of the flips result in heads, the experimental probability for heads is 8/10.

Contrast this to *theoretical probability*, which is the "mathematically calculated" probability. Continuing with the coin example, the theoretical probability of heads is 5/10, because we *expect* half of the flips to be heads. (We are, of course, assuming we have an ideal coin, with no thickness and even weight on both sides.)

Obviously, experimental probability does not have to agree with theoretical probability, and in fact, it often doesn't – at least, not exactly. However, **the larger the number of tests, the closer the experimental probability gets to the theoretical probability** – a fact known as the *law of large numbers*.

(Note that for the purposes of Math 12, we deal in theoretical probability, so experimental probability and the law of large numbers are not really relevant. Therefore, all references to "probability" mean theoretical probability.)

We now begin with the most basic of probability statements:

If there are n different events that are all equally likely to occur, and r of them are favorable outcomes (meaning they are what you want), the probability of a favorable outcome is r/n. Mathematically, if the event(s) we want is/are called A, we say that

$$P(A)=\frac{r}{n}$$

For example, if we want to know the chance of an even number on a dice roll, the probability is 3/6, since there are six possibilities and three favorable ones. (Of course, this simplifies to 1/2.)

Example: What is the probability of drawing an ace from a deck of cards?

Solution: There are 52 cards in a deck, with 4 aces. Each card is equally likely to be drawn, so the probability is 4/52, which simplifies to 1/13.

Notice that a probability value is always between 0 and 1 (because you can't have more favorable outcomes than possible outcomes), where 0 indicates "it will never happen" and 1 means "it is guaranteed".

Sample Space

The *sample space* of an experiment is simply a list of all the possible outcomes. For example, if we roll two dice, the sample space would simply be a chart of 36 entries, with 6 possibilities for each individual die.

As it is rather tedious to list out the sample space of many experiments, we oftentimes simply use a *Venn diagram* to illustrate the sample space. However, we will not discuss such diagrams here.

Complementary Events

Every event has a *complement*, which is essentially the “opposite” of the event. For example, if “rain tomorrow” is an event, its complement is “no rain tomorrow”.

It should be obvious that if an event doesn't happen, then its complement happens. In the above example, it will either rain tomorrow or it will not, and there is no other possible outcome. Therefore, we can say that

$$P(A)+P(\bar{A})=1$$

where $\overline{A}$ denotes the complement of A.

Rearranging the above expression gives us

$$P(\bar{A})=1-P(A)$$

which gives us a formula for calculating the probability of any complement of an event.

Example: What is the chance of rolling at least one 6 in 10 rolls of a six-sided die?

Solution: Although we could calculate the probability directly, it is easier to calculate the probability of the complement – in other words, the chances of *not* getting *any* 6s.

The chance of *not* getting a 6 in a single roll is 5/6, since only one of the possible outcomes is a 6. The chances of not getting a 6 in *two* rolls is 5/6 * 5/6 (see the section “A and B”), and so on, so we have

$$P(\text{no 6 in 10 rolls})=\frac{5}{6}\times\frac{5}{6}\times\ldots\times\frac{5}{6}=\left(\frac{5}{6}\right)^{10}$$

and thus

$$P(\text{at least one 6 in 10 rolls})=1-\left(\frac{5}{6}\right)^{10}$$

which is roughly equal to 0.838.

Notice that as we roll the dice more and more often, the chances of getting at least one 6 constantly gets closer to 100%, which supports the law of large numbers (see the section “Experimental and Theoretical Probability”).

Independent Events

Two events are said to be *independent* if neither affects the outcome of the other. For example, if you roll a die multiple times, the chance of getting a 6 (for example) on any given roll is completely independent of previous results. Thus, the rolls are said to be *independent events*.

On the other hand, if you draw cards from a deck (without replacing them afterwards), the chance of getting an ace (for example) is *not* independent of previous results – with each new result, the probability of drawing an ace changes. Thus, the draws are said to be *dependent events*.

The probability of an event A, *given that another event B has occurred*, is denoted by

$$P(A|B)$$

which is usually read "the probability of A given B".

If A is independent of B, then

$$P(A|B)=P(A)$$

which states that whether B happened or not, the probability of A is always the same.

A and B

Given an event A and an event B, the chance that *both* will occur is given by

$$P(A \text{ and } B)=P(A)\times P(B|A)$$

Of course, if A and B are independent, this becomes

$$P(A \text{ and } B)=P(A)\times P(B)$$

If both *cannot* occur simultaneously (for example, it is impossible to roll a single die and get two numbers at the same time), then

$$P(B|A)=0$$

meaning that

$$P(A \text{ and } B)=0$$

and the events are said to be *mutually exclusive*.

A or B

Defining "A or B" is slightly trickier than "A and B", so let us begin with an example.

Suppose we have the statement "if I have cash or a credit card, I can pay for lunch". Here, having only cash, or only a credit card, is sufficient to pay for lunch, but having both still works. Thus, "or" here is used *inclusively*, meaning that as long as *at least one* condition is satisfied, the whole statement is true.

On the other hand, if we have the statement "I can walk to school or get a ride", you can do either, but doing both is not an option. Thus, "or" here is used *exclusively*, meaning that *only one* condition can be satisfied for the whole statement to be true – if *both* are satisfied, the statement is *false*.

In logic, "or" is always defined in the *inclusive* sense of the term. Therefore, when we say "A or B", we really mean "A or B or both".

We can now make a definition: given two events A and B, the chance that *at least one* will occur is given by

$$P(A \text{ or } B)=P(A)+P(B)-P(A \text{ and } B)$$

If the events are mutually exclusive, then

$$P(A \text{ and } B)=0$$

and the expression simplifies to

$$P(A \text{ or } B)=P(A)+P(B)$$

Example: What is the chance of drawing a heart or an ace from a standard deck?

Solution: The probability of a heart is 13/52 and the probability of an ace is 4/52. The probability of *both* (that is, an ace of hearts) is 1/52, since only one card is both an ace and a heart. Therefore

$$P(\text{heart or ace}) = P(heart) + P(ace) - P(\text{heart and ace}) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

Bayes' Law

It should be intuitively obvious that

$$P(B \text{ and } A) = P(A \text{ and } B)$$

which means that

$$P(B) \times P(A|B) = P(A) \times P(B|A)$$

and upon rearranging, we get

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$$

a result known as *Bayes' Law*. This law is quite useful in many probability questions.

Example: Suppose there exists a camera that detects whether any given person is a criminal or not, with 99% accuracy – that is, 1% of the time it triggers a false alarm for a non-criminal, and 1% of the time it triggers no alarm for a criminal.

Now, in a city of one million people, suppose 100 are criminals. If a random person is detected by the camera and the alarm is raised, what is the chance that they are really a criminal?

Solution: Naively, it's easy to think that the chance is 99%, because the camera has 99% accuracy. In other words, the easy assumption is that

$$P(criminal|alarm) = P(alarm|criminal)$$

the latter of which is 99% – if someone is a criminal, there is a 99% chance the alarm will be raised.

However, this is false, because the correct expression uses Bayes' Law:

$$P(criminal|alarm) = \frac{P(criminal) \times P(alarm|criminal)}{P(alarm)}$$

The chance of the person being a criminal is 100/1000000 (there are 100 in a city of 1000000 people), which is 1/10000, or 0.01%. The chance of the alarm being raised for a criminal is 99%, of course.

As for the chances of the alarm being raised in general, we can rewrite it as

$$P(alarm) = P(criminal\ alarm \text{ or } non-criminal\ alarm)$$

which is equivalent to

$$P(alarm) = P(criminal\ alarm) + P(non-criminal\ alarm)$$

where we can omit the "and" part of the expression because the two events are mutually exclusive (there is no way a person is a criminal and non-criminal simultaneously). Now, the chances of the "criminal alarm" are 99% for 0.01% of the population – that is, 0.99 * 0.0001 – and the chances of the "non-criminal alarm" are 1% for 99.99% of the population – that is, 0.01 * 0.0009.

So we have

$$P(criminal|alarm) = \frac{0.0001 \times 0.99}{0.99 \times 0.0001 + 0.01 \times 0.9999} = \frac{1}{102}$$

which is about 0.98%.

Clearly, the “99% accuracy” camera doesn't even come close to its advertised accuracy, due to the fact that there are far more non-criminals than criminals. (This is known as the *base rate fallacy*.)

Using Permutations and Combinations

Recall the basic method for calculating probabilities: if we have r favorable outcomes out of a total of n, then

$$P(A) = \frac{r}{n}$$

However, counting outcomes is not always easy to do manually, so we often use permutation and combination expressions to help us.

Example: If you are dealt five cards from a standard deck, what is the probability of being dealt a full house?

Solution: The total number of ways we can be dealt five cards is simply

$${}_{52}C_5$$

because ordering is irrelevant.

A full house is composed one three-of-a-kind and one pair, so we can look at them separately. The number of ways we can receive one three-of-a-kind is

$${}_{13}C_1 \times {}_4C_3$$

because we choose one of the thirteen possible ranks (A, 2, ...Q, K) and three of the four cards that are in each rank.

Then the number of ways we can receive one pair is

$${}_{12}C_1 \times {}_4C_2$$

and the reason we have 12 instead of 13 is because one rank was already used up for the three-of-a-kind (we cannot repeat ranks).

The total is then

$${}_{13}C_1 \times {}_4C_3 \times {}_{12}C_1 \times {}_4C_2$$

and the probability becomes

$$\frac{{}_{13}C_1 \times {}_4C_3 \times {}_{12}C_1 \times {}_4C_2}{{}_{52}C_5}$$

which is approximately 0.00144, or 0.144% – not exactly great odds.

Using the Binomial Theorem

Although perhaps surprising, even the binomial theorem can assist in the calculation of certain probabilities.

Suppose the following conditions are satisfied for some event:

- There are only two possible outcomes (success or failure).
- The probability of success/failure is independent of previous trials.

Then if the experiment is performed n times, the chances of having x successes is

$${}_nC_x(p)^x(1-p)^{n-x}$$

where p is the chance of success for any given trial.

Example: Suppose a golfer has a 80% chance of sinking a 10-foot putt. What is the chance that he will sink at most 2 putts if he tries 10 times?

Solution: If the golfer can sink at most 2 putts, then he can sink either 0, 1, or 2. Since these are all mutually exclusive, we can simply add the probabilities of each.

The probability of sinking no putts at all is

$${}_{10}C_0(0.8)^0(0.2)^{10}$$

Similarly, the probability of sinking only a single putt is

$${}_{10}C_1(0.8)^1(0.2)^9$$

and the probability of sinking exactly two putts is

$${}_{10}C_2(0.8)^2(0.2)^8$$

and so the total probability is

$${}_{10}C_0(0.8)^0(0.2)^{10}+{}_{10}C_1(0.8)^1(0.2)^9+{}_{10}C_2(0.8)^2(0.2)^8$$

which is approximately 0.0000779, or 0.00779%.

Notice that the adding gets tiresome as we increase the number of trials. Therefore, we usually leave the mundane work of adding to graphing calculators, which have functions for these sorts of calculations built into them.

Appendix A: Formula Sheet

This is a list of many of the identities and formulae used in Math 12.

Trigonometry

Quotient identity:

$$\tan x=\frac{\sin x}{\cos x}$$

Reciprocal trigonometric functions:

$$csc\ x=\frac{1}{\sin x}$$

$$sec\ x=\frac{1}{\cos x}$$

$$\cot x = \frac{1}{\tan x} = \frac{\cos x}{\sin x}$$

Pythagorean identity:

$$\sin^2 x + \cos^2 x = 1$$

Corollaries to the Pythagorean identity:

$$\tan^2 x + 1 = sec^2 x$$
$$\cot^2 x + 1 = csc^2 x$$

Odd-even identities:

$$\sin(-x) = -\sin(x)$$
$$\cos(-x) = \cos(x)$$

Sum-and-difference identities:

$$\sin(x+y) = \sin x \cos y + \cos x \sin y$$
$$\sin(x-y) = \sin x \cos y - \cos x \sin y$$
$$\cos(x+y) = \cos x \cos y - \sin x \sin y$$
$$\cos(x-y) = \cos x \cos y + \sin x \sin y$$

Double-angle identities:

$$\sin(2\mathrm{x}) = 2\sin x \cos x$$
$$\cos(2\mathrm{x}) = \cos^2 x - \sin^2 x = 2\cos^2 x - 1 = 1 - 2\sin^2 x$$

Exponential Functions and Logarithms

Strength-reduction laws:

$$\log(xy) = \log x + \log y \quad \text{for} \quad x > 0, y > 0$$
$$\log\left(\frac{x}{y}\right) = \log x - \log y \quad \text{for} \quad x > 0, y > 0$$
$$\log(x^y) = y \log x \quad \text{for} \quad x > 0$$

Change-of-base rule:

$$\log_b x = \frac{\log_a x}{\log_a b}$$

Geometric Sequences and Series

nth term in a geometric sequence:

$$t_n = ar^{n-1}$$

Sum of a finite geometric series:

$$S_n = a\frac{r^n - 1}{r - 1} = a\frac{1 - r^n}{1 - r} \quad \text{for} \quad r \neq 1$$

Sum of an infinite geometric series:

$$S = \frac{a}{1 - r} \quad \text{for} \quad |r| < 1$$

Permutations and Combinations

Definition of a permutation:

$${}_nP_r = \frac{n!}{(n - r)!}$$

Definition of a combination:

$${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n - r)!}$$

Pascal's triangle – law of symmetry:

$${}_nC_r = {}_nC_{n-r}$$

Pascal's triangle – law of recursion:

$${}_nC_r = {}_{n-1}C_{r-1} + {}_{n-1}C_r$$

Binomial theorem:

$$(a + b)^n = \sum_{k=0}^{n} {}_nC_k\, a^{n-k} b^k$$

Probability

Logical operators:

$$P(A \text{ and } B) = P(A) \times P(B|A)$$
$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$
$$P(\bar{A}) = 1 - P(A)$$

Bayes' Law:

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$$

Binomial probability:

$P(A) = {}_nC_x (p)^x (1 - p)^{n-x}$ for an event A with probability p, with x successes in n trials

Appendix B: Trigonometric Functions of Special Angles

This is a table of the sine, cosine, and tangent of every special angle from 0 to 2π radians. Angles are also listed in degrees for convenience.

Angle (rad)	Angle (deg)	sin(angle)	cos(angle)	tan(angle)
0	0	0	1	0
$\pi/6$	30	1/2	$\sqrt{3}/2$	$1/\sqrt{3}$
$\pi/4$	45	$1/\sqrt{2}$	$1/\sqrt{2}$	1
$\pi/3$	60	$\sqrt{3}/2$	1/2	$\sqrt{3}$
$\pi/2$	90	1	0	undefined
$2\pi/3$	120	$\sqrt{3}/2$	-1/2	$-\sqrt{3}$
$3\pi/4$	135	$1/\sqrt{2}$	$-1/\sqrt{2}$	-1
$5\pi/6$	150	1/2	$-\sqrt{3}/2$	$-1/\sqrt{3}$
π	180	0	-1	0
$7\pi/6$	210	-1/2	$-\sqrt{3}/2$	$1/\sqrt{3}$
$5\pi/4$	225	$-1/\sqrt{2}$	$-1/\sqrt{2}$	1
$4\pi/3$	240	$-\sqrt{3}/2$	-1/2	$\sqrt{3}$
$3\pi/2$	270	-1	0	undefined
$5\pi/3$	300	$-\sqrt{3}/2$	1/2	$-\sqrt{3}$
$7\pi/4$	315	$-1/\sqrt{2}$	$1/\sqrt{2}$	-1
$11\pi/6$	330	-1/2	$\sqrt{3}/2$	$-1/\sqrt{3}$
2π	360	0	1	0

Appendix C: Solving Polynomial Equations

Overview

The solution of polynomial equations in a single variable – that is, solving equations of the form

$$f(x)=0$$

where *f(x)* is a polynomial – has been a prominent mathematical problem for thousands of years.

There are two issues in solving polynomial equations. One is the assurance that solutions exist in the first place, and the other is the actual finding of the solutions.

At first glance, it seems that the first should follow from the second. After all, if you can find the solutions, does that not automatically prove they exist? Well, not really. As an example, let us look at the ancient Greeks. Although their mathematical achievements were nothing short of incredible (considering the time period), they refused to consider negative numbers as solutions to a mathematical problem. So, for example, the equation

$$x + 1 = 0$$

would have no solution in their eyes.

Now of course, you would look at the above equation and immediately say that the solution is -1. The problem is, you have implicitly assumed that negative numbers were permissible. To illustrate the issue, consider the equation

$$x^2 + 1 = 0$$

The quadratic formula (which can be found in the next section) assures us that the solution must be

$$x = \pm\sqrt{-1}$$

Now if you look at that and say, "That's ridiculous! You can't take the square root of a negative number!", then the point has been illustrated nicely: *you have restricted yourself to the set of real numbers*. But, supposing we were to permit ourselves to use a *larger* set of numbers – one that contains the solutions to this equation – then the solution would, in fact, exist. And this is why the issue of the *existence* of solutions is distinct from the issue of *finding* the solutions.

So, what mathematicians eventually did was define a constant i:

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

(Although no *real* number is equal to i, it doesn't matter, because we are not restricting ourselves to the set of real numbers any longer.) They then formulated the general *complex number*:

$$a + bi$$

where a and b are real numbers. (Notice that if $b=0$, the complex number reduces to a real number.) If $a=0$, the number is also called an *imaginary number* (or a *pure imaginary number*, since there is no real part to it).

Clearly, the complex numbers contain the solutions to the equation

$$x^2 + 1 = 0$$

And you begin to wonder: do the complex numbers contain the solutions to *every* polynomial equation? The answer, of course, is yes, and it was proved in 1797 by Gauss, in a proof known as the *fundamental theorem of algebra*.

The fundamental theorem of algebra states that every polynomial equation of degree n has, at most, n solutions, and in fact is guaranteed n solutions if we count multiplicities and complex solutions.

For the purposes of Math 12, we will only concern ourselves with real solutions. It is important to remember, however, that complex solutions will always exist, even when real ones do not.

Linear and Quadratic Equations

The general *linear equation*

$$ax + b = 0$$

has the solution

$$x = -\frac{b}{a}$$

a formula so trivial that it is hardly ever used (instead, linear equations are usually solved directly).

More interesting to note is the general *quadratic equation*

$$ax^2+bx+c=0$$

which can be solved using a wide variety of methods. Some were known as far back as the time of the ancient Babylonians, although they were usually geometric in nature and did not admit negative or irrational solutions.

Perhaps the simplest *algebraic* technique for solving a quadratic equation is to consider the general *monic quadratic equation*:

$$x^2+mx+n=0$$

(In fact, *every* quadratic equation can be reduced to the monic form simply by dividing all terms by the leading coefficient.) We then attempt to *factor* the equation by changing it into the form

$$(x+r)(x+s)=0$$

where

$$rs=n$$
$$r+s=m$$

(You must guess to find r and s. Attempting to find them algebraically only gives you another quadratic equation.) The solutions are then

$$x=-r$$
$$x=-s$$

because if two numbers multiply to give 0, at least one of them must be 0.

Of course, since factoring the quadratic relies on guessing, it is quite unreliable and usually only done with the simplest quadratic equations. More general ones are usually solved with the *quadratic formula*, which gives the exact solutions to any quadratic equation. The derivation is omitted, but it can be done fairly easily by completing the square.

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

The expression under the radical

$$\Delta=b^2-4ac$$

is of particular interest, and is called the *discriminant*. It tells us about the solutions to the quadratic equation without needing to resort to fully solving the equation.

- If it is greater than zero, we have two distinct real solutions.
- If it is equal to zero, we have two identical real solutions.
- If it is less than zero, we have two distinct complex (but non-real) solutions (which we will usually consider as having *no solution in the real numbers*).

Beyond the Quadratic

Although linear and quadratic equations can be solved very neatly, the situation for higher-degree equations changes drastically. Math 11 does not present any formulae for higher-degree equations, and instead teaches students to *guess* at the solutions by means of the following theorem:

The *rational root theorem* says that if the coefficients of a polynomial equation are all integers, then all rational solutions, assuming there are any, must be of form

$$x=\frac{p}{q}$$

where p is any factor of the constant term and q is any factor of the leading coefficient. (Either can be negative, of course.)

It should go without saying that such a technique is extremely unreliable. For starters, the coefficients *must* be integers, and there must exist rational solutions in the first place – otherwise, the rational root theorem is of no help whatsoever, and we are once again back at square one, with absolutely no methods of solving polynomial equations of degree three or greater.

Luckily for us, the *general* cubic equation

$$ax^3+bx^2+cx+d=0$$

is indeed solvable, and the ultimate result, known as the *cubic formula*, is

$$x=-\frac{b}{3a}+\sqrt[3]{\left(\frac{-b^3}{27a^3}+\frac{bc}{6a^2}-\frac{d}{2a}\right)+\sqrt{\left(\frac{-b^3}{27a^3}+\frac{bc}{6a^2}-\frac{d}{2a}\right)^2+\left(\frac{c}{3a}-\frac{b^2}{9a^2}\right)^3}}$$
$$+\sqrt[3]{\left(\frac{-b^3}{27a^3}+\frac{bc}{6a^2}-\frac{d}{2a}\right)-\sqrt{\left(\frac{-b^3}{27a^3}+\frac{bc}{6a^2}-\frac{d}{2a}\right)^2+\left(\frac{c}{3a}-\frac{b^2}{9a^2}\right)^3}}$$

(The derivation is omitted. Note that this only gives one solution; the other two are given with very similar formulae.)

Similarly, the general *quartic equation* can also be solved, but the result, the *quartic formula*, is horrendously complicated and will not be shown here for a lack of space. (If you still wish to see it, a Google search should satisfy your curiosity.)

It is important to note that the analytic solutions to the cubic and quartic equations are only important *theoretically*. In practice, the formulae are far too complicated to be of any practical use (there is a *reason* that these formulae are never presented to students in almost all mathematics courses), so instead, we favor the *approximation* of the solutions (see the section “Methods of Approximation”).

(There is *no* similar formula for polynomial equations of degree five or higher – as in, it is literally impossible to find them, a result known as the *Abel-Ruffini theorem*.)

Methods of Approximation

Analytic solutions to polynomial equations are fascinating (at least, from a theoretical perspective), but beyond the quadratic, there is little practical value in them, as the complexity of the required formula increases exponentially. Instead, it is in our interest to *approximate* the solutions to an arbitrary degree of accuracy, which, for practical purposes, is just as useful as an analytic formula, and considerably cheaper computationally.

Various approximation methods exist, but most of them rely on tools from calculus, so only one approximation method will be presented here – one that does not require any calculus. It is called the *bisection method*.

It relies on a theorem known as the *intermediate value theorem*, or more precisely its corollary, which says that, given two x-coordinates a and b, if

$$f(a)>0$$

and

$$f(b)<0$$

(or vice versa), there is a solution between *a* and *b*. So, once we find ourselves a value for *a* and a value for *b*, we guess that the midpoint between them is the solution. Depending on the result we get, we adjust *a* or *b* to shrink the boundaries of the solution, and repeat as many times as we want.

Example: Find the solutions to the polynomial equation

$$x^3+x-1=0$$

Solution: As a graphing calculator can show you, there is only one real solution, and an irrational one at that. Short of applying the cubic formula (which is probably something you'd rather avoid), there is no easy way to solve this equation, so this is a good candidate for the bisection method.

We first have to discover a range the solution is in. For example, we find that

$$f(0)=-1$$

and

$$f(1)=1$$

so the intermediate value theorem says that there must be a solution between 0 and 1, since $f(0) < 0$ and $f(1) > 0$. (A suitable range may take a little while to find. However, as long as one point is less than 0, and one point greater than 0, that is all that is needed.)

Let our first guess be the midpoint, 0.5. We find

$$f(0.5)=-0.375$$

Since this is less than zero, we readjust our lower boundary to 0.5. Because $f(0.5) < 0$ and $f(1) > 0$, we know the solution must be between 0.5 and 1.

Let our second guess be the midpoint, 0.75. We find

$$f(0.75)=0.171875$$

Since this is greater than zero, we readjust our upper boundary to 0.75. Because $f(0.5) < 0$ and $f(0.75) > 0$, we know the solution must be between 0.5 and 0.75.

Let our third guess be the midpoint, 0.625. We find

$$f(0.625)\approx-0.130859$$

Since this is less than zero, we readjust our lower boundary to 0.625. Because $f(0.625) < 0$ and $f(0.75) > 0$, we know the solution must be between 0.625 and 0.75.

By this point, you should be seeing the pattern. Unless we are lucky, we will never hit on the exact solution, but we can get as close as we like by applying this method as many times as we like. The solution to six decimal places is $x=0.682328$.

As an aside, if you're curious, the exact answer (using the cubic formula) is

$$x=\sqrt[3]{\frac{9+\sqrt{93}}{18}}-\sqrt[3]{\frac{2}{3(9+\sqrt{93})}}$$

which indeed evaluates to roughly 0.682328.

It is important to note that this approximation method works for *any* equation, not just polynomial equations – yet another advantage: they generalize easily to any equation we care for.

Note that the bisection method suffers from two major drawbacks:

- It cannot detect double roots, because the curve of the polynomial will be either greater than zero or less than zero on both sides of the root.
- It's slow. (The convergence is linear, meaning that the accuracy of our solution increases at a constant rate.)

Both of these drawbacks can be addressed by another method known as *Newton's method*, but that requires techniques from calculus, which are currently unavailable to us.

Appendix D: Solving Systems of Linear Equations

Overview

The solution of *multiple* linear equations is a topic of *linear algebra*, and is one of the most important topics covered in the study of it. While the full theory is beyond the scope of a textbook on elementary algebra, the basics of solving a two- or three-equation system is important, as it has wide applications in elementary algebra.

One basic theory of importance is that for a unique solution to exist, there *must* be at least as many equations as unknown variables. (It is possible to have more equations than variables, but it will always be the case that either the extra equations are redundant, or they cause the system to have no solutions at all.)

Keep in mind that this section is only a brief summary of the techniques, as it is assumed that you are at least somewhat familiar with the theory of solving systems of linear equations.

Systems of Two Equations

A system of two equations takes the general form

$$a_1 x + b_1 y = c_1$$
$$a_2 x + b_2 y = c_2$$

Two basic methods exist to solve such a system: the *method of substitution*, and the *method of elimination*.

Method of substitution. The idea behind the method of substitution is to take one of the two equations, solve it for a single variable, then substitute the resulting expression into the other equation. That other equation should now be a single-variable equation, so the solution can be found for that variable. The other value can then be calculated by reverse substitution.

Example: Solve the system

$$4x + 2y = 14$$
$$2x - y = 1$$

using the method of substitution.

Solution: Arbitrarily, let's take the second equation and solve it for y:

$y = 2x - 1$

We then substitute this into the first equation:

$4x + 2(2x - 1) = 14$

Solve this equation for x:

$x = 2$

Then since we know that

$y = 2x - 1$

We know that

$y = 2(2) - 1 = 3$

Naturally, we could have solved the first system for y instead, or even chosen x as the variable of substitution. The important thing is the method; the choice of variable is arbitrary (usually based only on convenience).

Method of elimination. The idea behind the method of elimination is to add the two equations together directly, in such a way that one of the variables will cancel out. The other variable can then be solved for easily.

Example: Solve the system

$$4x + 3y = 38$$
$$6x - 3y = 12$$

using the method of elimination.

Solution: Add the two equations together.

$4x + 3y + 6x - 3y = 38 + 12$

Despite initial appearances, this does *not* violate the golden rule of equations, because the second equation says that $6x - 3y = 12$, so we really are adding the same thing to both sides.

Now, if we simplify this a bit:

$10x = 50$

because the $3y$ and the $-3y$ cancel. Therefore, the solution ends up being

$x = 5, y = 6$

If coefficients do *not* work out so nicely, then simply multiply either of the equations by whatever number is required to get the coefficients to work out.

Example: Solve the previous system

$$4x + 2y = 14$$
$$2x - y = 1$$

by the method of elimination.

Solution: If we multiply the bottom equation by 2, we get

$$4x + 2y = 14$$
$$4x - 2y = 2$$

Now adding the two equations gives us

$8x = 16$

and we eventually end up with

$$x=2, y=3$$
as before.

A general formula can be found for the solution of these systems:

$$x=\frac{c_1 b_2 - b_1 c_2}{a_1 b_2 - b_1 a_2}$$

$$y=\frac{a_1 c_2 - c_1 a_2}{a_1 b_2 - b_1 a_2}$$

Systems of Three Equations

A system of three equations takes the general form

$$a_1 x + b_1 y + c_1 z = d_1$$
$$a_2 x + b_2 y + c_2 z = d_2$$
$$a_3 x + b_3 y + c_3 z = d_3$$

The methods that worked for systems of two equations work here as well, but the method of elimination is almost always preferred, due to it being less algebraically intensive. However, unlike with before, the elimination cannot happen all at once; instead, it must be carried out in small steps.

Example: Solve the system

$$3x+2y-z=1$$
$$2x-2y+4z=-2$$
$$-x+\frac{1}{2}y-z=0$$

Solution: The idea here is to try and eliminate some of the variables present. While the steps may seem like "magic", they are clear if you have a goal in mind. In this case, the goal is to take two of the equations, and eliminate their x variables. We can then use those two equations and solve for y and z as with any normal two-variable system, which can then be used to find x.

So we begin by multiplying the bottom equation by 2:

$$3x+2y-z=1$$
$$2x-2y+4z=-2$$
$$-2x+y-2z=0$$

Now if we add the third equation to the second one, it looks like:

$$2x-2y+4z-2x+y-2z=-2+0$$

which simplifies to

$$-y+2z=-2$$

Therefore, our new system of equations is

$$3x+2y-z=1$$
$$-y+2z=-2$$
$$-2x+y-2z=0$$

which is a considerable improvement; the second equation has no x present.

Now if we multiply the bottom equation by 3/2, we get

$$-3x+\frac{3}{2}y-3z=0$$

Adding the first one gives us a new first equation:

$$3x+2y-z-3x+\frac{3}{2}y-3z=1+0$$

which simplifies to

$$\frac{7}{2}y-4z=1$$

Our system is now

$$\frac{7}{2}y-4z=1$$
$$-y+2z=-2$$
$$-2x+y-2z=0$$

and now we can solve the top two equations like any normal two-variable system. We find

$$y=-2, z=-2$$

and plugging these values into the third equation gives us

$$x=1$$

and we are done.

A general solution for a three-variable system does exist, but the result is horribly complicated:

$$x=\frac{d_1 b_2 c_3+b_1 c_2 d_3+c_1 d_2 b_3-c_1 b_2 d_3-d_1 c_2 b_3-b_1 d_2 c_3}{a_1 b_2 c_3+b_1 c_2 a_3+c_1 a_2 b_3-c_1 b_2 a_3-a_1 c_2 b_3-b_1 a_2 c_3}$$

$$y=\frac{a_1 d_2 c_3+d_1 c_2 a_3+c_1 a_2 d_3-c_1 d_2 a_3-a_1 c_2 d_3-d_1 a_2 c_3}{a_1 b_2 c_3+b_1 c_2 a_3+c_1 a_2 b_3-c_1 b_2 a_3-a_1 c_2 b_3-b_1 a_2 c_3}$$

$$z=\frac{a_1 b_2 d_3+b_1 d_2 a_3+d_1 a_2 b_3-d_1 b_2 a_3-a_1 d_2 b_3-b_1 a_2 d_3}{a_1 b_2 c_3+b_1 c_2 a_3+c_1 a_2 b_3-c_1 b_2 a_3-a_1 c_2 b_3-b_1 a_2 c_3}$$

Notice that the denominators are all the same (the same holds with the formulae for a two-equation system). This particular quantity is very special, and while the study of it is outside of the scope of elementary algebra, here it has one use: it is equal to zero if and only if the system has no *unique* solution. (That is, if this number is zero, the system will either have no solutions, or an infinite number of them.)

End of Document